

ANTONIO DI CRESCENZO
VIRGINIA GIORNO
AMELIA GIUSEPPINA NOBILE
LUIGI MARIA RICCIARDI

Un primo corso in probabilità

per scienze pure e applicate

Antonio Di Crescenzo
Virginia Giorno
Amelia Giuseppina Nobile
Luigi Maria Ricciardi

Un primo corso in probabilità

per scienze pure e applicate

Liguori Editore

Questa opera è protetta dalla Legge sul diritto d'autore

(Legge n. 633/1941: http://www.giustizia.it/cassazione/leggi/1633_41.html).

Tutti i diritti, in particolare quelli relativi alla traduzione, alla citazione, alla riproduzione in qualsiasi forma, all'uso delle illustrazioni, delle tabelle e del materiale software a corredo, alla trasmissione radiofonica o televisiva, alla registrazione analogica o digitale, alla pubblicazione e diffusione attraverso la rete Internet sono riservati, anche nel caso di utilizzo parziale.

La riproduzione di questa opera, anche se parziale o in copia digitale, è ammessa solo ed esclusivamente nei limiti stabiliti dalla Legge ed è soggetta all'autorizzazione scritta dell'Editore.

La violazione delle norme comporta le sanzioni previste dalla legge.

Il regolamento per l'uso dei contenuti e dei servizi presenti sul sito della

Casa Editrice Liguori è disponibile al seguente indirizzo:

http://www.liguori.it/politiche_contatti/default.asp?c=legal

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

Liguori Editore - I 80123 Napoli

<http://www.liguori.it/>

© 2009 by Liguori Editore, S.r.l.

Tutti i diritti sono riservati

Prima edizione italiana Luglio 2009

Di Crescenzo, Antonio :

Un primo corso in probabilità per scienze pure e applicate/Antonio Di Crescenzo, Virginia

Giorno, Amelia Giuseppina Nobile, Luigi Maria Ricciardi

Napoli : Liguori, 2009

ISBN-13 978 - 88 - 207 - 4845 - 6

1. Matematica 2. Casualità I. Titolo.

Aggiornamenti:

17 16 15 14 13 12 11 10 09

10 9 8 7 6 5 4 3 2 1 0

Indice

Premessa	ix
1 Introduzione alla probabilità	1
1.1 Introduzione	1
1.2 Nota storica	3
1.3 Spazio campione ed eventi	5
1.4 Prime definizioni di probabilità	10
1.4.1 Definizione classica	10
1.4.2 Definizione frequentista	12
1.4.3 Definizione soggettiva	14
1.5 Probabilità geometriche	15
1.6 Problemi di calcolo combinatorio	18
2 La teoria assiomatica	25
2.1 Caratterizzazione degli eventi	25
2.2 Definizione assiomatica di probabilità	30
2.3 Disuguaglianza di Boole e formula di inclusione-esclusione	34
2.4 Indipendenza di eventi	40
2.5 Probabilità condizionata	47
2.6 Legge delle alternative	53
2.7 Teorema di Bayes	56
3 Le variabili aleatorie	59
3.1 Variabili aleatorie unidimensionali	59
3.2 La funzione di distribuzione	62
3.3 Classificazione delle variabili aleatorie unidimensionali	65
3.3.1 Variabili aleatorie discrete	65
3.3.2 Variabili aleatorie assolutamente continue	68
3.3.3 Variabili aleatorie miste	70
3.4 Trasformazioni di variabili aleatorie	72
3.5 Vettori aleatori	78
3.6 Funzione di distribuzione congiunta	80
3.7 Classificazione dei vettori aleatori	82
3.7.1 Vettori aleatori bidimensionali discreti	82

3.7.2	Vettori aleatori bidimensionali assolutamente continui	84
3.7.3	Vettori aleatori multidimensionali	86
3.8	Indipendenza di variabili aleatorie	89
3.9	Trasformazioni di vettori aleatori	94
4	Distribuzioni notevoli	105
4.1	Famiglie parametriche di distribuzioni	105
4.2	Variabili aleatorie discrete	105
4.2.1	Distribuzione uniforme	106
4.2.2	Distribuzione di Bernoulli	107
4.2.3	Distribuzione binomiale	108
4.2.4	Distribuzione ipergeometrica	112
4.2.5	Distribuzione geometrica	116
4.2.6	Distribuzione binomiale negativa	118
4.2.7	Distribuzione di Poisson	122
4.3	Vettori aleatori discreti	125
4.3.1	Distribuzione multinomiale	126
4.3.2	Distribuzione ipergeometrica multivariata	130
4.4	Variabili aleatorie assolutamente continue	132
4.4.1	Distribuzione uniforme	132
4.4.2	Distribuzione esponenziale	135
4.4.3	Distribuzioni di Erlang e gamma	140
4.4.4	Distribuzione iperesponenziale	144
4.4.5	Distribuzione di Weibull	146
4.4.6	Distribuzione normale	148
5	Valore medio, momenti e funzioni generatrici	153
5.1	Introduzione	153
5.2	Valore medio	153
5.3	Valore medio di una funzione di variabile aleatoria	166
5.4	Momenti di una variabile aleatoria	173
5.5	Valore medio e momenti di una funzione di vettore aleatorio	186
5.6	Covarianza e coefficiente di correlazione	189
5.7	Momenti centrali di somme di variabili aleatorie	196
5.8	Funzioni generatrici	202
5.8.1	Funzione generatrice dei momenti	203
5.8.2	Funzioni generatrici di probabilità	205
5.8.3	Esempi di funzioni generatrici	208
6	Distribuzioni e momenti condizionati	219
6.1	Introduzione	219
6.2	Distribuzioni condizionate per variabili discrete	219
6.3	Distribuzioni condizionate per variabili continue	224
6.4	Distribuzioni condizionate per vettori aleatori misti	228
6.4.1	X discreta e Y assolutamente continua	228

6.4.2	X assolutamente continua e Y discreta	231
6.5	Probabilità condizionate	233
6.5.1	Variabile condizionante discreta	233
6.5.2	Variabile condizionante assolutamente continua	234
6.6	Legge delle alternative e teorema di Bayes per variabili aleatorie	234
6.7	Medie e momenti condizionati	240
6.8	Valori medi delle medie condizionate	245
6.9	La densità normale bivariata	249
7	Densità di probabilità speciali e loro proprietà	255
7.1	Introduzione	255
7.2	Statistiche ordinate e distribuzione beta	255
7.3	Distribuzione chi-quadrato	260
7.4	Distribuzione di Fisher	263
7.5	Distribuzione di Student	267
8	Disuguaglianze notevoli	271
8.1	Introduzione	271
8.2	Alcune semplici disuguaglianze	271
8.3	Disuguaglianze coinvolgenti i soli momenti	284
8.4	Limitazioni per somme di variabili aleatorie indipendenti	288
8.5	Altre disuguaglianze rilevanti	291
8.6	Momenti di variabili aleatorie stocasticamente ordinate	296
9	Teoremi asintotici	299
9.1	Successioni di variabili aleatorie	299
9.2	Convergenza di variabili aleatorie	299
9.3	Teorema centrale di convergenza	307
9.4	Cenni alle leggi dei grandi numeri	312
9.5	Convergenze di distribuzioni binomiali	317
Appendici		
A	Variabili aleatorie discrete	323
B	Variabili aleatorie continue	327
C	Distribuzione normale standard	333
D	Fattoriali e coefficienti binomiali	335
E	Formule notevoli	337
F	Disuguaglianze	341
	Alcuni personaggi	347

Indice analitico**349****Indice delle abbreviazioni e dei simboli****353**

Premessa

Ben diversamente dagli scorsi decenni, nel panorama delle discipline oggetto d'insegnamento nei corsi di laurea triennali e specialistici sempre maggiore spazio trova il Calcolo delle Probabilità; ciò non soltanto per la rilevanza che esso riveste per gli studenti delle Facoltà di Scienze, d'Ingegneria e di Economia, ma anche per il riconosciuto ruolo formativo che svolge, dischiudendo allo studente, desiderabilmente fin dagli inizi del percorso universitario, l'esistenza di strumenti e modi non convenzionali di guardare ai fenomeni naturali e, in breve, al "mondo dell'incerto" che, dopotutto, è proprio quello nel quale siamo tutti in larga parte immersi e che molte delle nostre azioni condiziona. Non trascurabile è poi la circostanza che una delle principali applicazioni del Calcolo delle Probabilità è la Statistica Inferenziale, ormai pane quotidiano per studenti e ricercatori anche in altre aree disciplinari quali Sociologia e Medicina. Va peraltro sottolineato che una corretta e non deviante utilizzazione delle tecniche e, in generale, della metodologia statistica, non può avvenire in assenza di una sufficientemente approfondita conoscenza del Calcolo delle Probabilità, unitamente alla consapevolezza dei limiti di applicabilità di questa nel trarre affidabili conclusioni sui fenomeni in considerazione.

A differenza ad esempio della Geometria, il cui sistema assiomatico-deduttivo affonda le proprie radici nell'opera di Euclide del III secolo a.C., o del Calcolo Differenziale il cui sviluppo, legato ai nomi di Newton e Leibniz, risale al XVII secolo, il Calcolo delle Probabilità è una conquista del XX secolo. Invero, a superamento dei precedenti risultati concernenti fondamentalmente giochi d'azzardo ed altre situazioni a questi in ultima analisi riconducibili, il Calcolo delle Probabilità, o, più propriamente, la Teoria della Probabilità, doveva attendere Andrey Nikolaevich Kolmogorov per trovare negli anni trenta una sistemazione assiomatica rigorosa. Questa, tuttavia, non è oggetto del presente volume, il cui scopo è di fornire un "primo corso" sulla materia avente la funzione di informare, e soprattutto di formare, lo studente interessato sui vari modi con i quali avvicinarsi alla risoluzione di problemi concreti; ciò anche attraverso ampia esemplificazione e svolgimento dettagliato di numerosi esercizi, tutti da riguardarsi come parte integrante degli argomenti trattati. Gli strumenti dei quali si richiede lo studente sia in possesso consistono in una qualche dimestichezza con gli elementi base dell'analisi matematica, tipicamente oggetto d'insegnamento in un paio di moduli di un corso di studi universitari di durata triennale. Va tuttavia osservato che la comprensione di una significativa parte del testo prescinde anche da siffatti prerequisiti, e che le formule particolarmente necessarie, quando non ricavate esplicitamente, sono state inserite in un'apposita Appendice.

Pur non trattandosi di un volume a carattere avanzato, non si rifugge da precisione nell'esposizione dei concetti: i risultati sono infatti tutti ricavati in modo rigoroso e, quando non

dimostrati a causa di indesiderabili appesantimenti o della necessità di far ricorso a strumenti troppo avanzati, essi sono corredati da specifici commenti e considerazioni esplicative.

I primi sette capitoli sono dedicati agli elementi di base ed alle proprietà salienti della probabilità, ivi inclusa una sistematica, ampia trattazione delle distribuzioni di probabilità con enfasi sulla loro utilizzazione per la risoluzione di problemi suggeriti da svariati contesti disciplinari. L'ottavo capitolo è interamente dedicato alla presentazione di talune disuguaglianze probabilistiche il cui ruolo è ampio e riconosciuto anche nei contesti applicativi. Il nono capitolo è infine incentrato sui cosiddetti teoremi asintotici della probabilità, la cui trattazione avviene, nel presente volume, attraverso l'utilizzazione della funzione generatrice dei momenti invece che della più potente funzione caratteristica, peraltro meno accessibile allo studente di destinazione. Sono altresì presenti alcune Appendici sinotticamente elencanti le distribuzioni trattate nel volume unitamente alle loro principali caratteristiche, nonché varie formule e disuguaglianze notevoli delle quali è stato fatto sistematico od occasionale uso.

Napoli, 25 marzo 2009

Gli Autori

Capitolo 1

Introduzione alla probabilità

1.1 Introduzione

Il Calcolo delle Probabilità, del quale la Statistica costituisce la più popolare applicazione, oltre a rivestire elevato interesse sotto il profilo matematico fornisce utili strumenti di indagine in numerose e svariate aree tra le quali vanno annoverate le scienze fisiche e naturali nonché discipline quali tecnologia, psicologia, sociologia, economia, medicina. Grazie al Calcolo delle Probabilità ed alla Statistica è risultata possibile l'estensione di rigorosi metodi quantitativi all'analisi ed all'interpretazione di esperimenti i cui risultati non appaiono univocamente determinati nel senso che ripetizioni di tali esperimenti in condizioni per quanto possibile identiche non conducono a risultati identici. Gli esempi più familiari sono forniti dai giochi d'azzardo, riconducibili spesso a lanci di dadi o monete, scelta di carte da mazzi ben mescolati, lotterie; altrettanto significativi sono peraltro i dati tratti da sondaggi d'opinione, indagini epidemiologiche, sperimentazioni di nuovi farmaci, emissione di particelle da sorgenti radioattive, numero di utenti in file di attesa alle casse di un supermercato o ai caselli autostradali.

A differenza di altre branche della matematica, quali ad esempio la geometria il cui sistema assiomatico-deduttivo affonda le proprie radici nell'opera di Euclide, la teoria della probabilità ha origini piuttosto recenti dal momento che i contributi più antichi a questa disciplina risalgono ai secoli XV e XVI. La corrispondenza tra Pierre de Fermat e Blaise Pascal iniziata nel 1654, relativa a problemi specifici di giochi d'azzardo, che non venivano comunque annoverati tra le finalità della matematica dell'epoca, condusse all'introduzione di concetti di base quali "probabilità" e "aspettazione". Un primo studio matematico sistematico, sempre con riferimento ai giochi d'azzardo, risale allo stesso periodo (1657) ed è dovuto a Christian Huygens (*De Ratiociniis in Ludo Aleae*).

Di questi iniziali sviluppi vanno sottolineate due caratteristiche fondamentali: anzitutto l'analisi dei giochi d'azzardo è pressoché l'unico punto di riferimento concreto nella costruzione della teoria della probabilità, presumibilmente a causa dello scarso sviluppo delle scienze naturali in quell'epoca; in secondo luogo gli strumenti usati si riducono essenzialmente all'aritmetica ed al calcolo combinatorio.

Un vero e proprio trattato a carattere analitico, che costituì una fondamentale svolta rispetto a tale tendenza, fu pubblicato da Pierre-Simon Laplace nel 1812 con il titolo di *Théorie Analytique des Probabilités*. Contributi allo sviluppo di un apparato analitico sempre più complesso, dovuti anche a Abraham De Moivre, a Karl Friedrich Gauss e a Simeon Denis Poisson, furono sviluppati a seguito di pressanti, specifiche richieste da parte della comunità scientifica e politica: necessità di formulare una teoria degli errori, risolvere problemi di balistica e ideare metodi di statistica, specialmente di statistica delle popolazioni. La formulazione moderna del calcolo delle probabilità deve comunque riguardarsi come una grande conquista del XX secolo. Grazie ai progressi registrati nei settori tecnologici e delle scienze fisiche e naturali, lo sviluppo della teoria della probabilità si è poi ulteriormente accentuato allargando enormemente il campo delle sue applicazioni prima confinate esclusivamente ai giochi d'azzardo o a situazioni a questi riconducibili.

Il calcolo delle probabilità nasce dall'osservazione di fenomeni casuali sistematicamente riscontrabili in natura e si prefigge di studiarne le leggi che li governano. Nel suo sviluppo possono individuarsi alcuni filoni principali che riguardano rispettivamente

- (i) la natura della probabilità,
- (ii) la teoria matematica del calcolo delle probabilità,
- (iii) le applicazioni.

Come suggerito da Rudolf Carnap (*Logical Foundations of Probability*, University of Chicago Press, 1950), nell'ambito della problematica concernente la natura della probabilità (punto (i)) possono individuarsi due aree, l'una a carattere filosofico, l'altra privilegiante l'aspetto sperimentale. La prima è incentrata sulla cosiddetta "inferenza induttiva"; questa si prefigge di ricavare delle proprietà o delle indicazioni, da tradurre in assiomi di una teoria matematica, a partire da considerazioni ed osservazioni imprescindibili da un contesto induttivo. La seconda area, che privilegia l'aspetto sperimentale, è incentrata sul problema dell'individuazione di relazioni che permettano di collegare i possibili risultati di esperimenti casuali con grandezze intrinseche dei fenomeni coinvolti. In particolare, essa è rivolta allo studio di eventi ripetitivi caratterizzati dalla validità della cosiddetta *legge empirica del caso*, ossia dalla proprietà che le loro frequenze empiriche di occorrenza in un gran numero di prove ripetute in condizioni sempre macroscopicamente identiche; all'aumentare del numero delle prove appaiono stabilizzarsi intorno a dei valori limite. La teoria matematica del calcolo delle probabilità (punto (ii)) è invece fondata sulla formulazione di un insieme di assiomi dai quali sia possibile pervenire ad una teoria matematica intrinsecamente coerente. La validità di questa va poi collaudata mettendo a confronto risultati teorici con dati sperimentali. È bene peraltro menzionare che gli assiomi stessi sono formulati in guisa da tradurre in forma quantitativa talune proprietà suggerite da osservazioni di natura empirica. La teoria generale costruita in base a tali assiomi nel suo insieme prescinde, invece, dai fatti sperimentali, e quindi proprio attraverso questi essa va valutata a posteriori; la necessità di sostituzioni o di ritocchi del sistema assiomatico diventerebbe invero ineludibile qualora si riscontrassero contraddizioni o incongruenze con fatti sperimentali.

L'esigenza di una costruzione assiomatica della teoria della probabilità nasce dalla circostanza che in molti casi sono da prendersi in esame situazioni ben più complesse di quelle che, ad esempio, caratterizzano esperimenti collegati ai classici giochi d'azzardo, per i quali è spesso sufficiente l'uso della nozione di frequenza empirica. Un fondamentale contributo in tal senso è costituito dalla celeberrima monografia del 1933 del matematico russo An-

drey Nikolaevich Kolmogorov, disponibile in traduzione inglese col titolo *Foundations of the Theory of Probability*, Chelsea, N.Y., 1950. La struttura di base della teoria assiomatica della probabilità da questi elaborata è da allora rimasta pressoché invariata.

Per quanto attiene, infine, al settore concernente le applicazioni del calcolo delle probabilità (punto (iii)), va sottolineato che tra queste emerge per notorietà ed utilità la Statistica. Essa comprende un insieme di metodi di natura logica e matematica atti a raccogliere, elaborare, analizzare ed interpretare dati allo scopo di descrivere fenomeni collettivi e di estendere la descrizione di certi fenomeni osservati ad altri fenomeni dello stesso tipo non ancora osservati. La Statistica ha quindi per oggetto l'analisi di dati tratti da esperimenti effettivi, costituendo così un naturale collegamento tra la teoria matematica della probabilità e la realtà.

Concludiamo questo paragrafo con una considerazione di un Premio Nobel per la Fisica, il danese Max Born, che puntualizza¹ un aspetto della rilevanza della teoria della probabilità nella Scienza: “...Il concetto di casualità interviene già nei primi passi dell'attività scientifica in virtù del fatto che nessuna misura è corretta in maniera assoluta. Io ritengo che la casualità sia un concetto più fondamentale della causalità; invero, se in un caso concreto la relazione di causa-effetto sussista o meno può solo giudicarsi applicando alle misure le leggi della casualità ...”.

1.2 Nota storica

L'assenza di una nozione quantitativa di probabilità nel mondo pregalileiano è presumibilmente da attribuirsi all'inesistenza del metodo sperimentale. Fu infatti con la nascita di questo metodo, gloria tutta italiana, che riguardando all'antico gioco dei dadi con la nuova mentalità scientifica si scoprì l'esistenza di talune regolarità nei risultati di esperimenti i cui esiti non sono prevedibili con certezza in quanto legati in qualche misura al caso. Di ciò viene fatta menzione nell'opera *De Ludo Aleae* di Gerolamo Cardano (1501–1576), medico, matematico e filosofo italiano, completata forse nel 1526 ma pubblicata postuma nel 1663. Precedentemente il concetto di probabilità era stato oggetto d'interesse del matematico Luca Pacioli (1445–1517) e, successivamente, dello stesso Galileo. Ma fu soltanto con Blaise Pascal (1623–1662) che il calcolo delle probabilità venne alla luce soprattutto per merito del Cavaliere Antoine Gombaud de Mére (1607–1684), personaggio molto in vista nell'ambiente bene parigino. Costui, accanito giocatore, accusava la matematica, della quale possedeva qualche conoscenza, di essere responsabile delle sue perdite ai dadi dal momento che le frequenze dei risultati del gioco non corrispondevano alle sue valutazioni teoriche sulle quali egli basava la determinazione della posta. Proprio il suo porre a Pascal quesiti del tipo “quanti lanci occorrono per poter sperare di ottenere un 6 doppio lanciando una coppia di dadi” fu all'origine della prima richiamata corrispondenza tra questi e Pierre de Fermat (1601–1665) che si estese successivamente a numerosi problemi di probabilità combinatoria. Seguirono l'opera dell'olandese Christian Huygens (1629–1695), della quale si è già fatta menzione, ed i lavori dei Bernoulli, membri di una famiglia di matematici svizzeri. A Jacques Bernoulli (1654–1705), autore della *Ars Coniectandi*, che apparve postuma nel 1713, sono dovuti un celebre teorema ed un famoso modello matematico che portano il suo nome. Daniele Bernoulli (1700–1782) si preoccupò invece di applicare il calcolo delle probabilità a svariati

¹M. Born. *Natural Philosophy of Cause and Chance*, Oxford University Press, 1949.

problemi concreti, introducendo anche la nozione di “utilità” nella teoria della probabilità. A lui è dovuta la formulazione del cosiddetto paradosso di San Pietroburgo, che val qui la pena riassumere. Si considera il gioco d’azzardo consistente in una successione di lanci indipendenti di una moneta non truccata. Un giocatore viene ammesso al gioco previo pagamento di una certa somma, diciamo di s Euro. Si suppone che il giocatore riceve 2 Euro se si verifica testa al primo lancio, 4 Euro se testa si verifica per la prima volta al secondo lancio, 8 Euro se testa si verifica per la prima volta al terzo lancio e, in generale, 2^n Euro se testa si verifica per la prima volta all’ n -esimo lancio. Ci si chiede quale sia un valore “equo” di s , ossia quale sia un’equa somma da richiedersi al giocatore per consentirgli di partecipare al gioco. Intuitivamente si sarebbe portati ad identificare s con la somma che in media il giocatore vince, ma questa si dimostra valere $+\infty$. Tale risultato è appunto paradossale in quanto si esigerebbe una somma infinitamente grande per consentire la partecipazione ad un gioco dal quale non può che ricavarsi una vincita limitata.

Ritorniamo al nostro breve excursus di carattere storico.² Ruolo importante per lo sviluppo della teoria della probabilità fu svolto dal matematico inglese Abraham de Moivre (1667–1754) che con la sua *Doctrine of Chance* del 1718 gettò le basi della convergenza della distribuzione binomiale alla legge normale, dimostrando così, sia pure in un caso particolare, una legge riscontrata molto spesso nello studio sperimentale dei cosiddetti fenomeni casuali o aleatori, ossia dei fenomeni retti dalle leggi del caso. Sempre in Inghilterra, soprattutto a seguito del crescente sviluppo delle società di assicurazione, specialmente delle assicurazioni marittime a copertura dei rischi legati ai commerci d’oltremare, Thomas Bayes (1702–1761) formulò il cosiddetto “problema della probabilità delle cause”, fornendone una prima soluzione attraverso una nota formula che porta il suo nome. Al matematico francese Antoine Deparcieux (1703–1768) va fatta poi risalire la nascita della statistica moderna grazie all’utilizzazione che egli sistematicamente fece di concetti e metodi della probabilità in indagini di natura statistica. Seguirono Pierre-Simon Laplace (1749–1827), anch’egli francese, il cui trattato sulla teoria analitica della probabilità è già stato richiamato, ed il tedesco Karl Friedrich Gauss (1777–1855), grazie ai quali trovarono ulteriore sviluppo e sistemazione importanti questioni quali il teorema centrale di convergenza, l’analisi degli errori nelle misure, il trattamento dei dati sperimentali. È significativo che il grafico della famosa curva di Gauss, che estrema rilevanza riveste nel contesto della probabilità, abbia fatto bella mostra di sé sulle banconote tedesche dei vecchi 10 marchi!

Notevoli ulteriori contributi alla teoria della probabilità furono apportati nel Secolo XIX soprattutto da matematici francesi e inglesi. Basti ricordare le opere di Simeon Denis Poisson (1781–1840) al cui nome è legata una fondamentale legge della teoria, e di George Boole (1815–1864) la cui intuizione doveva poi, tra l’altro, contribuire ad individuare la struttura algebrica della teoria della probabilità.

Dalla metà del XIX secolo sino agli anni venti di quello appena trascorso lo sviluppo della teoria della probabilità è saldamente legato ai nomi dello studioso sovietico Pafnuty Lvovich Chebyshev (1821–1894) e dei suoi allievi Andrei Andreyevich Markov (1856–1922) e Aleksandr Mikhailovich Lyapunov (1857–1918) ai quali va anche attribuito il merito di aver introdotto e sistematicamente utilizzato il concetto di variabile aleatoria. Essi molto contribuirono a dar lustro alla celebre scuola matematica di San Pietroburgo alla quale sono

²Una trattazione approfondita della storia della probabilità fino alla nascita della teoria assiomatica è, ad esempio, presente nel classico trattato L.E. Maistrov, *Probability Theory – A Historical Sketch*, Academic Press, 1974.

ascrivibili fondamento logico e impostazione astratta della teoria della probabilità, nonché la nascita stessa di una nuova branca di questa teoria, oggi nota quale teoria dei processi stocastici.

Lo sviluppo della moderna teoria della probabilità ed il suo successivo allargamento a molteplici, diversificati settori di grande interesse applicativo è avvenuto attraverso l'opera di numerosissimi studiosi di svariati Paesi. Va comunque detto che ruolo centrale ha svolto la cosiddetta Scuola Russa, particolarmente con Sergei Natanovich Bernstein (1880–1968), Andrey Nikolaevich Kolmogorov (1903–1987) e Aleksandr Yakovlevich Khinchin (1894–1959). Fu però durante la prima decade del '900 che Émile Borel (1871–1956) formulò il primo collegamento tra teoria della probabilità e aspetti della teoria astratta della misura delle funzioni di variabile reale. Successivamente, negli anni venti, queste idee furono ulteriormente sviluppate ad opera dei sopra menzionati Khinchin e Kolmogorov oltre che da Evgeny Evgenievich Slutsky (1880–1948), Paul Lévy (1886–1971), Richard von Mises (1883–1953) e molti altri, con risultati davvero fecondi per lo sviluppo della teoria. Ulteriori importanti sviluppi si ebbero anche con Jarl Waldemar Lindeberg (1876–1932), Bernstein, William Feller (1906–1970) ed innumerevoli altri studiosi, tra cui l'eclettico Norbert Wiener (1894–1964). Gli strumenti della teoria della misura e dell'analisi funzionale hanno poi condotto ad estensioni formidabili della teoria della probabilità, soprattutto per quanto attiene alla sopra ricordata teoria dei processi stocastici che, già presente in embrione in lavori di Louis Bachelier (1870–1946), Adriaan Daniel Fokker (1887–1972) e Max Planck (1858–1947), ha trovato fondamento matematico rigoroso agli inizi degli anni trenta ad opera di Khinchin e di Kolmogorov. Proprio a quest'ultimo, come si è già detto, è dovuta la formulazione assiomatica della teoria della probabilità.

1.3 Spazio campione ed eventi

Convenzionalmente con la locuzione “esperimento casuale” si indica ogni atto o processo, spontaneamente verificantesi o artificialmente realizzato, di cui non sia prevedibile con certezza il risultato o lo sviluppo, ma che sia ripetibile o, quantomeno, concepibile come tale. Ogni singola esecuzione dell'esperimento casuale viene detta *prova*. I possibili esiti, o “risultati”, di un esperimento casuale devono intendersi invece sempre ben definiti o precisabili. L'impossibilità di previsione può essere di duplice natura: si può trattare di fenomeni deterministici di cui, però, sono mal note o non sono affatto note leggi atte a descriverli, oppure di fenomeni di natura aleatoria, ossia intrinsecamente retti da leggi probabilistiche. Per quanto riguarda fenomeni del primo tipo si fissi per un attimo l'attenzione sul problema della descrizione dell'evoluzione dinamica delle molecole di un gas. In linea di principio sarebbe possibile formulare le equazioni del moto delle singole molecole, ma tali equazioni sarebbero in realtà improduttive; ciò non solo per il loro numero astronomico (si ricordi che ad esempio un grammo di idrogeno contiene circa $6 \cdot 10^{23}$ molecole), ma anche perché in realtà il moto delle molecole individuali non offre interesse essendo molto più importante ottenere informazioni sul comportamento *macroscopico* del gas. In questo caso tecniche di tipo statistico vengono in aiuto per risolvere problemi che pure sono di natura deterministica. Ricordiamo, poi, che vi sono fenomeni a carattere intrinsecamente aleatorio, quali quelli connessi con il comportamento della materia a livello subatomico. A carattere intrinsecamente aleatorio sono

altresì annoverati tutti gli esperimenti concernenti estrazioni di carte da mazzi ben mescolati, di biglie da urne, o in generale riguardanti giochi d'azzardo.

Si consideri un esperimento casuale con caratteristiche e condizioni ambientali ben definite; si definisce *spazio campione*, e lo si indica tradizionalmente con Ω , l'insieme dei possibili risultati dell'esperimento casuale.³ Ad esempio, nell'esperimento consistente nell'estrazione di una biglia da un'urna contenente n biglie in parte bianche e in parte rosse, lo spazio campione potrebbe essere costituito da n elementi, ciascuno identificante una ed una sola biglia. Se invece dovesse essere d'interesse solo il colore della biglia estratta, lo spazio campione consisterebbe dei due soli elementi "bianco", "rosso". A seconda delle situazioni, lo spazio campione può essere discreto (finito o numerabile) o continuo. Se, ad esempio, l'esperimento casuale consiste nel lanciare una moneta per un prefissato numero di volte e nel registrare la successione di teste e di croci ottenuta, Ω è finito; se, invece, l'esperimento consiste nel lanciare ripetutamente un dado fino a quando il risultato è un numero pari, Ω è numerabile; infine, Ω è continuo se ad esempio si assume che l'esperimento fornisca come esito un numero reale appartenente ad uno specificato intervallo. Si noti che esperimenti casuali distinti possono dar luogo allo stesso spazio campione.

I possibili risultati dell'esperimento casuale vengono detti *eventi elementari*. Nel seguito il generico evento elementare sarà denotato con ω . Lo spazio campione Ω è costituito dunque dalla totalità degli eventi elementari.

L'associare ad ogni esperimento casuale uno spazio campione permette di introdurre il concetto di *evento*. Un *evento* è un sottoinsieme dell'insieme Ω , ivi compreso l'insieme vuoto e l'intero Ω . Si noti che con tale definizione un evento elementare non è un evento, mentre evento può essere un *singleton*, ossia un sottoinsieme di Ω costituito da un solo evento elementare. Nel seguito un generico evento sarà usualmente denotato con lettere romane maiuscole A, B, \dots .

È importante sottolineare che non sempre tutti i sottoinsiemi dello spazio campione Ω possono essere considerati eventi. Il concetto di evento è, invero, inscindibilmente legato alle nozioni di osservabilità e misurabilità. Perché un sottoinsieme E di Ω possa riguardarsi come evento, deve anzitutto potersi stabilire se il risultato dell'esperimento casuale, quando effettuato, è, oppure no, un elemento di E ; inoltre, deve anche essere possibile "misurare" l'evento, ossia associare ad E una misura $P(E)$ che viene interpretata come probabilità di E . Si pensi, ad esempio, all'esperimento casuale consistente nel lancio di una moneta ripetuto quattro volte. Lo spazio campione Ω consta di 16 eventi elementari ciascuno dei quali è formato da una sequenza di 4 elementi, ognuno di questi essendo T (Testa) o C (Croce). Sia E il sottoinsieme di Ω costituito dagli eventi elementari contenenti almeno tre teste:

$$E = \{(TTTT), (CTTT), (TCTT), (TTCT), (TTTC)\}.$$

Si supponga ora che i risultati di questo esperimento casuale siano registrati da un dispositivo fisico in grado di trascrivere soltanto 2 dei 4 risultati dei lanci. L'osservatore è dunque nell'impossibilità di stabilire, sulla base dei dati registrati dal dispositivo, se il risultato dell'esperimento casuale appartiene, oppure no, ad E . Quindi E non è osservabile, così che non può riguardarsi come evento. D'altra parte, si comprende come questa non osservabilità

³Talvolta per spazio campione si intende un qualsiasi insieme i cui elementi siano in corrispondenza biunivoca con i possibili risultati dell'esperimento casuale considerato.

comporti l'impossibilità di associare ad E una misura (ad esempio attraverso la registrazione delle frequenze empiriche di occorrenza di E) che sia interpretabile come la sua probabilità. In realtà, anche prescindendo da considerazioni di "osservabilità", è ben noto che vi sono situazioni in cui non è possibile associare una misura (ossia una probabilità in questo contesto) a tutti i sottoinsiemi di Ω . Ad esempio se Ω coincide con \mathbb{R}^n si è quasi sempre costretti da ragioni di consistenza matematica a considerare come famiglia degli eventi la sola classe dei cosiddetti insiemi di Borel (dei quali si dirà nel Capitolo 2) in luogo della collezione di tutti i sottoinsiemi di Ω .

Si dice che un evento E si verifica, o "occorre", quando il risultato ω dell'esperimento casuale effettuato appartiene ad E . Ad esempio, nell'esperimento consistente nel lanciare una sola volta un dado, può assumersi $\Omega = \{1, 2, \dots, 6\}$, così che $E = \{2, 4, 6\}$ è l'evento "numero pari". Se il lancio dà come esito il numero 2, oppure il numero 4 oppure il numero 6, diciamo che l'evento E si è verificato. Si noti che il verificarsi di un evento non esclude il verificarsi anche di altri eventi. Ad esempio, con riferimento al lancio del dado, l'uscita del numero 6 indica l'occorrenza sia dell'evento "uscita di un numero pari" sia dell'evento "uscita di un numero maggiore di 4".

Lo spazio Ω , costituito dalla totalità degli eventi elementari, è detto *evento certo* ("certo" perché qualunque sia l'evento elementare ω che si verifica, risulta $\omega \in \Omega$). Quindi, Ω è un insieme di eventi elementari *necessari* (nel senso che uno di essi si deve verificare necessariamente) ed *incompatibili* (ossia più eventi elementari non possono verificarsi simultaneamente). L'evento che non contiene nessun evento elementare viene detto *evento impossibile* e denotato con \emptyset .

Si noti che esiste un'analogia tra il linguaggio del calcolo delle probabilità e quello della teoria degli insiemi. Infatti, agli elementi di un insieme corrispondono gli eventi elementari; al termine "sottoinsieme" corrisponde il termine "evento"; all'insieme Ω corrisponde l'evento certo; all'insieme vuoto \emptyset corrisponde l'evento impossibile.

Se A è un evento, \bar{A} (*complemento* o *complementare* di A), denota l'evento costituito dall'insieme degli eventi elementari di Ω che non appartengono ad A . Ovviamente il complemento di \bar{A} è l'evento A ; quindi $\overline{\bar{A}} = A$ e $\overline{\emptyset} = \Omega$.

Se A e B sono eventi, $A \cup B$ denota l'evento che consiste di tutti gli eventi elementari che appartengono ad almeno uno degli eventi A , B . Quindi $A \cup B$ si verifica se A si verifica e B non si verifica, oppure se B si verifica e A non si verifica, oppure se entrambi A e B si verificano.

Se A e B sono eventi, $A \cap B$ denota l'evento che consiste di tutti gli eventi elementari che appartengono sia ad A che a B . Inoltre, se $A \cap B = \emptyset$ gli eventi A e B non si possono verificare contemporaneamente; infatti, se $\omega \in A$ allora $\omega \notin B$, e se $\omega \in B$ allora $\omega \notin A$. In questo caso gli eventi A e B sono detti *incompatibili* oppure *mutuamente esclusivi*.

Con la scrittura $A \subset B$, che si legge *A implica B*, si indica che il verificarsi dell'evento A implica il verificarsi dell'evento B . Quindi, se $A \subset B$ e se $\omega \in A$, allora $\omega \in B$. Ovviamente $A \subset B$ equivale a $\bar{B} \subset \bar{A}$. Inoltre, la scrittura $A = B$ indica che ogni evento elementare in A è un evento elementare in B ed ogni evento elementare in B è un evento elementare in A . Si noti che $A = B$ equivale a richiedere che $A \subset B$ e $B \subset A$.

Più in generale, l'evento $A_1 \cup A_2 \cup \dots \cup A_n$ si verifica quando almeno uno degli eventi A_k ($k = 1, 2, \dots, n$) si verifica, mentre l'occorrenza dell'evento $A_1 \cap A_2 \cap \dots \cap A_n$ consiste nel verificarsi di tutti gli eventi A_k ($k = 1, 2, \dots, n$).

Gli eventi A_1, A_2, \dots, A_n sono detti incompatibili se e solo se essi sono incompatibili a due a due, cioè se e solo se $A_i \cap A_j = \emptyset$ ($i, j = 1, 2, \dots, n; i \neq j$).

Le relazioni tra eventi sono spesso rappresentate mediante *diagrammi di Venn*, in cui lo spazio campione Ω è solitamente indicato con un rettangolo, mentre gli eventi sono rappresentati da regioni del rettangolo, ad esempio cerchi o parti di cerchi (v. Figura 1.1).

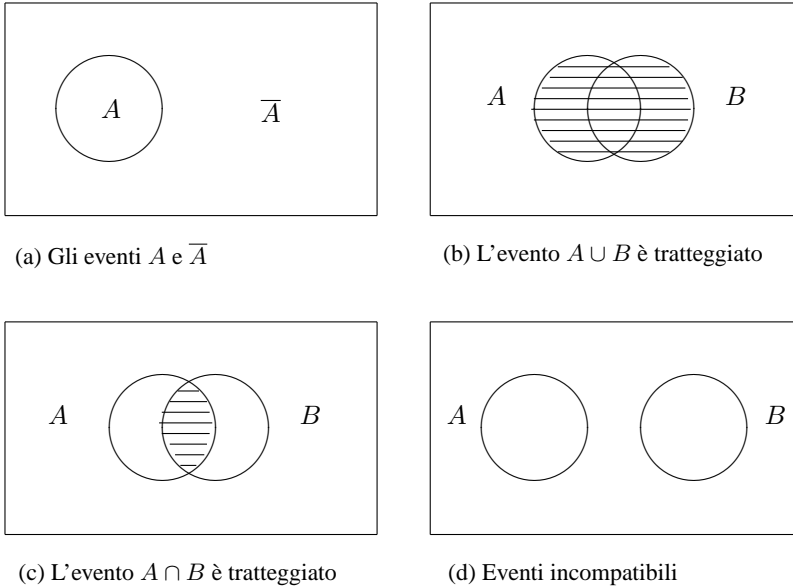


Figura 1.1 – Diagrammi di Venn indicanti alcuni eventi.

Esempio 1.1 Si supponga di lanciare per due volte un dado e si assuma come spazio campione l'insieme delle 36 coppie di possibili risultati: $\Omega = \{(i, j): i, j = 1, 2, \dots, 6\}$. Si denoti con $A = \{(i, j): i = 1, 2; j = 1, 2, \dots, 6\}$ l'evento che si verifica quando il primo dado fornisce un numero minore di 3, con $B = \{(i, j): i = 1, 2, \dots, 6; j = 4, 5, 6\}$ l'evento che si verifica quando il secondo dado dà un numero maggiore di 3 e con $C = \{(i, j): i = 4, 5, 6; j = 1, 2\}$ l'evento che si verifica quando il primo dado fornisce un numero maggiore di 3 e il secondo un numero minore di 3. In Figura 1.2 sono indicati gli eventi A , B e C mediante un diagramma di Venn. Si noti che gli eventi A e B non sono incompatibili, poiché $A \cap B = \{(i, j): i = 1, 2; j = 4, 5, 6\}$, mentre A e C sono incompatibili, così come incompatibili sono B e C . \diamond

Alcune semplici relazioni tra eventi sono le seguenti:

$$(I) \quad \emptyset \cap A = \emptyset \quad \text{e} \quad \emptyset \cup A = A$$

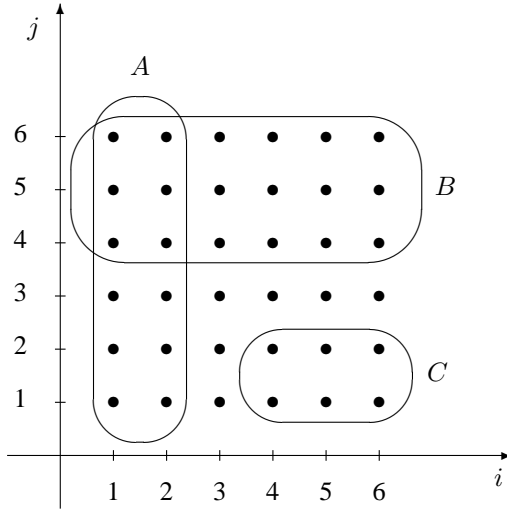


Figura 1.2 – Diagramma di Venn che indica i tre eventi A , B e C dell'Esempio 1.1.

$$(2) \quad \Omega \cap A = A \quad \text{e} \quad \Omega \cup A = \Omega$$

$$(3) \quad A \cap \bar{A} = \emptyset \quad \text{e} \quad A \cup \bar{A} = \Omega$$

$$(4) \quad (A \cup B) \cup C = A \cup (B \cup C) \quad (\text{proprietà associativa dell'unione di eventi})$$

$$(5) \quad (A \cap B) \cap C = A \cap (B \cap C) \quad (\text{proprietà associativa dell'intersezione di eventi})$$

$$(6) \quad A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \quad (\text{distributività dell'intersezione rispetto all'unione})$$

$$(7) \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \quad (\text{distributività dell'unione rispetto all'intersezione}).$$

Altre relazioni, di leggermente più riposta dimostrazione, sono le seguenti:

$$(8) \quad A \cap (\bar{A} \cap B) = \emptyset.$$

Infatti, per la proprietà associativa (5) si ha $A \cap (\bar{A} \cap B) = (A \cap \bar{A}) \cap B = \emptyset \cap B = \emptyset$.

$$(9) \quad A = (A \cap B) \cup (A \cap \bar{B}).$$

Invero, dire che l'evento A si verifica equivale a dire che si verificano entrambi gli eventi A e B , oppure che si verifica l'evento A senza che si verifichi B ; infatti, per le proprietà (2) e (3) e per la proprietà distributiva (6) risulta $A = A \cap \Omega = A \cap (B \cup \bar{B}) = (A \cap B) \cup (A \cap \bar{B})$.

$$(10) \quad A = A \cup (A \cap B).$$

Infatti, dire che l'evento A si verifica equivale a dire che si verifica l'evento A oppure che si verificano entrambi gli eventi A e B . Ciò è immediata conseguenza dell'osservazione che l'unione di un evento con un suo sottoinsieme riproduce l'evento stesso.

$$(11) \quad A = A \cap (A \cup B).$$

Dire che l'evento A si verifica equivale a dire che si verifica l'evento A e che si verifica almeno uno degli eventi A, B ; infatti, per la proprietà distributiva (6) si ha che $A \cap (A \cup B) = (A \cap A) \cup (A \cap B) = A \cup (A \cap B) = A$.

$$(12) \quad A \cup B = A \cup (\overline{A} \cap B).$$

Infatti, dire che almeno uno degli eventi A, B si verifica equivale a dire che o si verifica l'evento A oppure che si verifica l'evento B senza che si verifichi A ; invero, usando la proprietà distributiva (7) si ha $A \cup (\overline{A} \cap B) = (A \cup \overline{A}) \cap (A \cup B) = \Omega \cap (A \cup B) = A \cup B$.

$$(13) \quad A \cup B = \overline{\overline{A} \cap \overline{B}} \quad \text{e} \quad A \cap B = \overline{\overline{A} \cup \overline{B}} \quad (\text{formule di De Morgan}).$$

Concludendo, la famiglia degli eventi deve possedere una struttura adeguatamente ricca. Infatti, è ragionevole richiedere che se un sottoinsieme E di Ω è un evento, deve risultare possibile stabilire inequivocabilmente se il generico risultato ω dell'esperimento casuale è, oppure no, un elemento di E . Se ciò è possibile, è certamente anche possibile stabilire se ω appartiene, oppure no, al complemento \overline{E} di E ; se, inoltre, si può stabilire se ω appartiene, oppure no, a ciascun sottoinsieme E_i ($i = 1, 2, \dots, n$) di Ω , si può anche stabilire se ω appartiene, oppure no, all'unione $\bigcup_{i=1}^n E_i$ (e, quindi, anche all'intersezione $\bigcap_{i=1}^n E_i$). È dunque naturale richiedere che la classe degli eventi sia chiusa rispetto alle operazioni di complementazione, unione e intersezione finita. Infine, poiché è sempre affermativa la risposta al quesito se ω appartenga o meno ad Ω , va richiesto che anche l'intero spazio campione Ω sia un evento. In realtà, come vedremo in seguito, si suppone che la classe degli eventi sia chiusa anche rispetto alle operazioni di complementazione, unione ed intersezione numerabili; ciò non risulta giustificabile su base intuitiva e va, per il momento, interpretato come un requisito atto a pervenire ad una maggiormente ricca teoria matematica della probabilità.

1.4 Prime definizioni di probabilità

Per costruire una teoria utilizzabile in problemi concreti è opportuno partire da una definizione di probabilità che ne rispecchi il contenuto intuitivo e che allo stesso tempo sia operativa nel senso di contenere in sé le regole di calcolo che sono alla base dei necessari sviluppi matematici. Poiché l'interpretazione degli aspetti intuitivi non è unica, si sono sviluppate nel corso dei tempi definizioni diverse di probabilità: *classica*, *frequentista* e *soggettiva*.

1.4.1 Definizione classica

La prima definizione di probabilità, che chiameremo "classica", si ritrova già in Pascal e viene utilizzata anche da Laplace.

Definizione 1.1 (Probabilità classica) Dato uno spazio campione Ω finito, si definisce probabilità $P(A)$ di un evento $A \subset \Omega$ il rapporto tra il numero $N(A)$ di casi favorevoli al verificarsi dell'evento A ed il numero $N(\Omega)$ dei casi possibili, purché questi ultimi siano “ugualmente possibili”:

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

Tale definizione, sebbene operativa, contiene in realtà una tautologia perché affermare che gli eventi sono “ugualmente possibili” equivale ad affermare che sono equiprobabili, ossia ugualmente probabili. Quindi, tale definizione presuppone che si sappia stabilire a priori quando eventi hanno uguale probabilità di occorrenza. Ciò è peraltro intuitivamente fattibile quando ci si riferisce, ad esempio, a risultati di esperimenti quali giochi nei quali l'equiprobabilità di taluni risultati (uscita di rosso o nero in roulette non truccate, estrazioni del lotto, ecc.) è realisticamente conseguenza delle condizioni in cui si opera.

Dalla Definizione 1.1 segue immediatamente che la probabilità è un numero compreso tra 0 (quando nessun caso è favorevole) e 1 (quando tutti i casi sono favorevoli); in particolare si ha il valore 1 quando l'evento si verifica certamente. Un'altra conseguenza della definizione è la *legge di additività finita della probabilità* che può così formularsi: se A_1 e A_2 sono eventi incompatibili dello stesso spazio campione Ω , la probabilità della loro unione è uguale alla somma delle rispettive probabilità:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2). \quad (1.1)$$

Infatti, poiché per ipotesi i due eventi sono incompatibili, il numero $N(A_1 \cup A_2)$ dei casi favorevoli all'evento $A_1 \cup A_2$ è uguale a $N(A_1) + N(A_2)$, ossia alla somma del numero dei casi favorevoli ad A_1 e del numero di quelli favorevoli ad A_2 . Usando la Definizione 1.1, si ha:

$$P(A_1 \cup A_2) = \frac{N(A_1 \cup A_2)}{N(\Omega)} = \frac{N(A_1) + N(A_2)}{N(\Omega)} = \frac{N(A_1)}{N(\Omega)} + \frac{N(A_2)}{N(\Omega)} = P(A_1) + P(A_2).$$

Abbiamo così ricavato, come diretta conseguenza della definizione data, alcune proprietà della probabilità.

Da quanto detto discende che l'area nella quale è utilizzabile la definizione classica di probabilità è quella dei giochi d'azzardo e simili. Invero in tal caso le regole stesse individuano con esattezza le diverse possibili alternative che, come si è già sottolineato, possono spesso assumersi ugualmente probabili. Si tratta allora di determinare il numero dei casi possibili ed il numero dei casi favorevoli, nel che gioca ruolo importante il calcolo combinatorio.

Va qui menzionato che nel caso in cui Ω non contiene un numero finito di elementi, il concetto di equiprobabilità deve essere espresso in altro modo, come si vedrà nel seguito.

Esempio 1.2 Si supponga di lanciare un dado⁴ e di voler calcolare la probabilità che il risultato sia un numero pari. Per motivi di simmetria, le uscite di ciascuna delle sei facce del dado sono da considerarsi ugualmente probabili. Essendo 3 il numero di casi favorevoli e 6 il numero di casi possibili, per la definizione classica la probabilità richiesta è $3/6 = 1/2$. Si supponga ora di lanciare il dado per due volte e di essere interessati alla probabilità che

Tabella 1.1 – Somma dei risultati in due lanci di un dado.

Primo lancio	Secondo lancio					
	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

la somma dei risultati sia 10 oppure 11. Si considerino i seguenti eventi: $B = \{\text{la somma dei risultati è 10 oppure 11}\}$, $A_1 = \{\text{la somma dei risultati è 10}\}$, $A_2 = \{\text{la somma dei risultati è 11}\}$. Si noti che $B = A_1 \cup A_2$ e, poiché gli eventi A_1 e A_2 sono incompatibili, dalla (1.1) risulta $P(B) = P(A_1) + P(A_2)$. Il numero dei casi possibili è $N = 36$, ossia uguale al numero delle coppie (i, j) , $i, j = 1, \dots, 6$. Come si nota dalla Tabella 1.1, le coppie che forniscono come somma 10 sono $(4, 6)$, $(5, 5)$ e $(6, 4)$, così che il numero dei casi favorevoli all'evento A_1 è 3. Ne segue $P(A_1) = 3/36 = 1/12$. Invece, le coppie che forniscono come somma 11 sono $(5, 6)$ e $(6, 5)$, di modo che il numero dei casi favorevoli all'evento A_2 è 2. Ne segue $P(A_2) = 2/36 = 1/18$. La probabilità richiesta è in conclusione $P(B) = 1/12 + 1/18 = 5/36$. \diamond

1.4.2 Definizione frequentista

È noto fin dall'antichità che per un gran numero di fenomeni traducibili in esperimenti casuali consistenti in prove ripetute (quali ad esempio ripetizioni del lancio di una moneta) il verificarsi o meno di un prefissato evento esibisce talune regolarità. Invero, indicato con $\nu_n(A)$ il numero di volte in cui l'evento A si verifica in n prove ripetute nelle stesse condizioni (*frequenza assoluta* di occorrenza dell'evento A), la *frequenza relativa* $f_n(A)$ di occorrenza di A , ossia il rapporto $\nu_n(A)/n$, al crescere di n appare stabilizzarsi intorno ad un qualche "valore limite". Questa osservazione, di natura esclusivamente empirica, ha condotto alla formulazione del seguente postulato, noto come Legge Empirica del Caso: *In una successione di prove effettuate nelle stesse condizioni, al crescere del numero delle prove la frequenza relativa di ogni prefissato evento si avvicina alla probabilità dell'evento stesso.*

La legge empirica del caso mette in relazione la frequenza relativa, determinata sperimentalmente, con la nozione teorica di probabilità, che risulta pertanto indirettamente definita. Si giunge così alla definizione frequentista di probabilità.

Definizione 1.2 (Probabilità frequentista) *In una successione di prove effettuate nelle stesse condizioni la probabilità di un evento è misurata dalle frequenze relative di occorrenza dell'evento quando il numero delle prove cresce indefinitamente.*

⁴A meno di esplicito avviso contrario, assumeremo sempre tacitamente che dadi, monete e mazzi di carte siano "equi", ossia non truccati, e che gli esperimenti coinvolti non siano effettuati ad opera di bari.

Va subito detto che perché questa definizione sia utilizzabile è necessario disporre di successioni di prove ripetute tutte effettuate nelle medesime condizioni. Ciò ne restringe l'applicabilità a situazioni ben definite, quali lanci successivi di una moneta.

Nemmeno la definizione frequentista è esente da critiche. Infatti, le frequenze relative di cui si dice costituiscono successioni costruite sperimentalmente così che i concetti classici, quali quelli di limite, non sono appropriati. Come la definizione classica, la definizione frequentista è comunque operativa.

La frequenza relativa delle prove in cui l'evento considerato si verifica possiede le stesse caratteristiche esibite dal rapporto tra il numero di casi favorevoli e il numero dei casi possibili. Infatti si ha $0 \leq f_n(A) \leq 1$, con l'uguaglianza a 0 se l'evento non si è verificato nelle n prove e l'uguaglianza ad 1 se l'evento si è verificato in ognuna delle n prove. È ragionevole assumere che tali proprietà sussistano anche nell'ideale passaggio al limite, ossia quando il numero delle prove viene assunto infinitamente grande. Quindi anche per la definizione frequentista la probabilità è un numero compreso tra 0 e 1. Inoltre, se in un esperimento consistente in n prove ripetute si considerano due eventi incompatibili A_1 e A_2 , allora per la definizione di frequenza relativa si ha:

$$f_n(A_1 \cup A_2) = \frac{\nu_n(A_1 \cup A_2)}{n} = \frac{\nu_n(A_1)}{n} + \frac{\nu_n(A_2)}{n} = f_n(A_1) + f_n(A_2).$$

È ragionevole assumere che tale proprietà sussista al crescere indefinito del numero delle prove. Usando quindi la definizione frequentista si giunge di nuovo alla legge (1.1) di additività finita della probabilità.

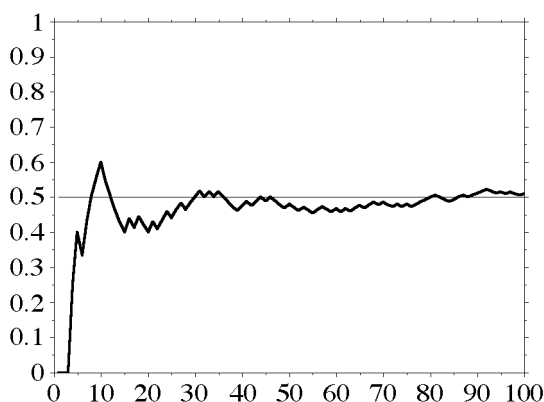


Figura 1.3 – Frequenze relative in una particolare successione di lanci di una moneta.

La Figura 1.3 mostra l'andamento della frequenza relativa di occorrenza dell'evento “Testa” in un particolare esperimento consistente in 100 lanci di una moneta equa. Si noti come al crescere del numero dei lanci effettuati tale frequenza si avvicini al valore $1/2$ coincidente con la probabilità a priori dell'evento considerato.

Esempio 1.3 Si consideri l'esperimento consistente nel lanciare una moneta che si sa essere non equa, ossia truccata. La definizione classica di probabilità non può in questo caso essere

utilizzata poiché le uscite ora non sono equiprobabili. Se in 1000 lanci il numero di volte in cui appare Testa è 750, le frequenze relative all'uscita di Testa e Croce sono rispettivamente 0.75 e 0.25; esse possono ragionevolmente essere riguardate come valutazioni delle corrispondenti probabilità. \diamond

1.4.3 Definizione soggettiva

Esistono delle situazioni in cui non è possibile ricorrere alle due precedenti definizioni di probabilità: da un lato poiché non sembra avere alcuna giustificazione l'ipotesi di equiprobabilità degli eventi elementari, dall'altro perché non è possibile effettuare ripetizioni dell'esperimento nelle medesime condizioni. Un esempio tipico è costituito dalle scommesse sul risultato di un incontro di calcio in cui non si può ritenere che i tre possibili risultati (vincita della squadra di casa, pareggio, vincita della squadra in trasferta) siano equiprobabili. Inoltre i precedenti incontri sostenuti dalle due squadre non possono essere riguardati come prove ripetute nelle medesime condizioni (perché non è detto che i giocatori siano sempre nelle stesse condizioni atletiche, perché i campi di gioco non sono caratterizzati da condizioni immutabili meteorologiche o di manto erboso, perché l'effetto della tifoseria può essere variamente condizionante, e così via).

In situazioni analoghe a quella descritta si può ricorrere all'impostazione "soggettiva" (o "personale") della probabilità. Questa, già accennata in Pascal ed attribuibile a D. Bernoulli, è stata ripresa e sviluppata in epoca recente soprattutto da Bruno de Finetti (1906–1985) e da Leonard Jimmie Savage (1917–1971). Nell'approccio soggettivo la probabilità di un evento viene identificata con il "grado di fiducia" che una persona ripone nel verificarsi dell'evento. Più precisamente, si dà la seguente definizione:

Definizione 1.3 (Probabilità soggettiva) *La probabilità di un evento A è il prezzo $P(A)$ che un individuo ritiene equo pagare per ricevere 1 se l'evento si verifica e ricevere 0 se l'evento non si verifica.*

È appena il caso di menzionare che tale prezzo deve intendersi minore dell'unità.

La Definizione 1.3 trova un'immediata interpretazione quando ci si riferisca al contesto delle scommesse, acquistando la seguente formulazione: *La probabilità di un evento A è l'importo $P(A)$ che uno scommettitore è disposto a puntare per ricevere 1 in caso di vincita e 0 in caso di perdita.*

La definizione data poggia su principi di *equità* e di *coerenza*. L'individuo, infatti, deve essere in grado di valutare "in modo equo", nel senso che deve essere disposto ad accettare la scommessa senza mutare la somma puntata quando da scommettitore diventa banco: scommettitore e banco devono quindi potersi scambiare i rispettivi ruoli senza alterare le probabilità di vincita e di perdita associate al gioco. L'individuo deve inoltre caratterizzarsi come "coerente" nel senso che l'insieme delle probabilità da lui assegnate non deve consentire di realizzare vincita certa o perdita certa attraverso un insieme di scommesse simultanee.

La probabilità di un evento si interpreta in sostanza come l'importo che un individuo, in base alle proprie personali valutazioni, giudica equo pagare (farsi pagare) per riscuotere (pagare) l'importo unitario se l'evento si verifica e l'importo nullo se l'evento non si verifica. In tali condizioni l'individuo è dunque disposto a pagare (ricevere) $s P(A)$ per ricevere (pagare) s se l'evento si verifica.

Dalla definizione soggettiva è possibile dedurre alcune regole per la probabilità. Infatti, in base alla definizione, in una scommessa su A si paga $P(A)$ per ricevere 1 se A si verifica e ricevere 0 se A non si verifica. Poiché non sarebbe coerente pagare più di 1 per ricevere 1 se l'evento A si verifica, ne segue che è $P(A) \leq 1$. D'altra parte non sarebbe coerente pagare meno di 1 se l'evento si verifica con certezza. Quindi, deve essere $P(A) = 1$ se l'evento è certo. Inoltre, se l'evento A non si verifica mai non sarebbe coerente pagare più di 0 per ricevere 0 e quindi $P(A) = 0$.

Dalla condizione di coerenza è possibile inoltre far discendere la legge di additività finita (1.1) della probabilità. Siano A_1, A_2, \dots, A_n degli eventi necessari ed incompatibili; si considerino n scommesse, una su ciascuno di tali eventi, in cui si paga $P(A_i)$ per ricevere 1 se A_i si verifica e ricevere 0 se A_i non si verifica ($i = 1, 2, \dots, n$). Quindi, in totale si paga $\sum_{i=1}^n P(A_i)$. Poiché certamente uno ed uno solo degli eventi si verifica, si riceve 1 nella scommessa relativa a quell'evento e si riceve 0 nelle altre scommesse. Pertanto, dall'insieme delle scommesse si ottiene certamente 1. Per la condizione di coerenza deve risultare $\sum_{i=1}^n P(A_i) = 1$. Infatti, se tale somma fosse diversa da 1, si realizzerebbe in ogni caso possibile un guadagno pari a $1 - \sum_{i=1}^n P(A_i)$, che risulterebbe positivo se $\sum_{i=1}^n P(A_i) < 1$ e negativo se $\sum_{i=1}^n P(A_i) > 1$, contraddicendo così la condizione di coerenza. Si considerino ora due eventi incompatibili A_1 e A_2 . I tre eventi A_1, A_2 e $\overline{A_1 \cup A_2}$ sono necessari e incompatibili. Quindi, $P(A_1) + P(A_2) + P(\overline{A_1 \cup A_2}) = 1$. D'altra parte anche gli eventi $A_1 \cup A_2$ e $\overline{A_1 \cup A_2}$ sono necessari e incompatibili e pertanto $P(A_1 \cup A_2) + P(\overline{A_1 \cup A_2}) = 1$. Queste due ultime uguaglianze implicano che $P(A_1 \cup A_2) = P(A_1) + P(A_2)$.

Nemmeno la definizione soggettiva è esente da critiche. In molti casi l'individuo non è indifferente di fronte al rischio; in alcuni casi lo cerca (in lotterie, giochi d'azzardo) pagando un prezzo che certamente non è equo ed in altri casi (nelle assicurazioni) paga proprio per evitare il rischio. La critica maggiore è che la probabilità soggettiva è legata alla persona coinvolta: la probabilità di un dato evento può invero essere notevolmente diversa per persone diverse, riflettendo le differenti informazioni, esperienze e atteggiamenti su cui esse basano le proprie convinzioni.

1.5 Probabilità geometriche

La definizione classica di probabilità, che richiede che tutti gli eventi costituiti da singoli punti dello spazio campione siano equiprobabili, trova un'analogia di carattere geometrico nel caso in cui lo spazio campione Ω consiste ad esempio di figure geometriche (intervalli, figure piane, figure solide). In tal caso i punti di Ω non sono più in numero finito, così che il concetto di "ugualmente probabile" sta ora semplicemente ad indicare che la probabilità che un punto appartenga ad un sottoinsieme dello spazio campione Ω è proporzionale alla misura di questo sottoinsieme. In generale, quindi, se M è la misura di Ω (lunghezza, area, volume) e m è l'analogia misura di un evento E di pari dimensionalità (riguardato come sottoinsieme di Ω), allora la probabilità $P(E)$ di tale evento viene posta uguale a m/M . Ad esempio, con riferimento alla Figura 1.4, sia AB il segmento di lunghezza L e sia CD un segmento di AB di lunghezza ℓ . Ci si può chiedere quale sia la probabilità che un punto scelto a caso su AB appartenga a CD . In questo caso Ω è costituito dall'insieme dei punti di AB . All'evento $E = \{\text{un punto scelto a caso su } AB \text{ appartiene a } CD\}$ va pertanto associata la probabilità $P(E) = \ell/L$.

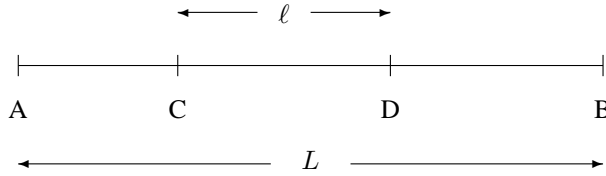


Figura 1.4 – Esempio di probabilità geometriche in una dimensione.

Come secondo esempio si consideri un quadrato Q di lato d e sia C il cerchio inscritto in Q (v. Figura 1.5). Si intende determinare la probabilità che un punto scelto a caso in Q

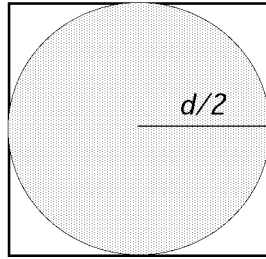


Figura 1.5 – Esempio di probabilità geometriche nel piano.

appartenga a C . In questo caso Ω è l'insieme dei punti di Q così che all'evento $E = \{\text{un punto scelto a caso in } Q \text{ cade in } C\}$ va associata la probabilità

$$P(E) = \frac{\text{area di } C}{\text{area di } Q} = \frac{\pi d^2/4}{d^2} = \frac{\pi}{4} \approx 0.7853. \quad (1.2)$$

Giova osservare che in taluni casi la definizione geometrica di probabilità può dar luogo ad ambiguità. È ad esempio celebre il cosiddetto paradosso di Bertrand che nasce dalla considerazione del seguente problema: tracciata a caso una corda di una circonferenza, calcolare la probabilità che la sua lunghezza sia maggiore di quella del lato del triangolo equilatero inscritto nella circonferenza. Questo problema apparentemente ammette più di una soluzione in conseguenza delle diverse concrete traduzioni in termini operativi della procedura con la quale è possibile tracciare una corda “a caso”. Tra i vari possibili criteri, qui considereremo i tre sotto elencati (v. Figura 1.6).

1. Per ragioni di simmetria si può assegnare a priori la direzione della corda da tracciare. Con riferimento alla figura (a), sceglieremo come direzione quella orizzontale. Consideriamo poi il diametro perpendicolare a tale direzione e su questo fissiamo a caso il punto attraverso il quale far passare la corda. Nel caso della figura (a) la corda tracciata con tale scelta casuale sia AB .

2. Sempre per ragioni di simmetria si può fissare uno degli estremi della corda sulla circonferenza. Nella figura (b) tale estremo è stato denotato con A. La corda verrà considerata come tracciata a caso se l'altro suo estremo (P nel caso della figura) è un punto scelto a caso sulla circonferenza.
3. Si può scegliere a caso un punto interno al cerchio e considerarlo come punto medio della corda da tracciare, come indicato in figura (c).

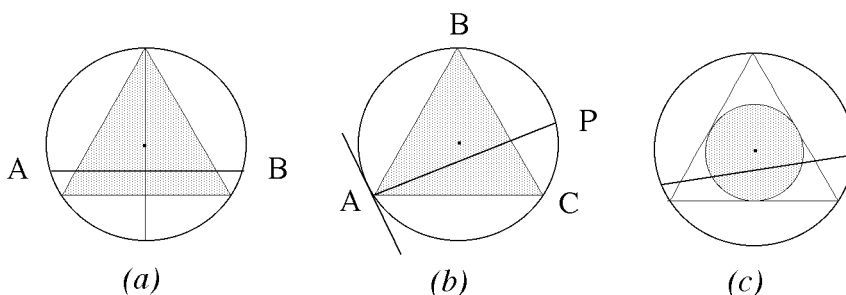


Figura 1.6 – Illustrazioni del paradosso di Bertrand.

Anzitutto ricordiamo che se r denota il raggio della circonferenza, il lato del triangolo equilatero in essa inscritto ha lunghezza $r\sqrt{3}$ e che l'altezza di tale triangolo è $3r/2$. Per ognuno dei tre criteri precedentemente descritti calcoliamo la probabilità dell'evento $E = \{\text{la lunghezza della corda scelta a caso è maggiore della lunghezza del lato del triangolo equilatero inscritto nella circonferenza}\}$.

La sopra specificata diversità delle procedure di costruzione della corda, conduce a tre differenti risultati. Infatti, nel caso (a) lo spazio campione Ω è costituito dall'insieme dei punti del diametro individuato e l'evento E si verifica per quei punti del diametro la cui distanza dal centro della circonferenza è minore di $r/2$. Quindi la probabilità richiesta è $P(E) = (r/2)/r = 1/2$. Nel caso (b) la tangente alla circonferenza nel punto A ed i due lati del triangolo equilatero con vertice in questo punto individuano tre angoli di $\pi/3$ ciascuno che insistono sugli archi \widehat{AB} , \widehat{BC} , \widehat{CA} , ognuno di lunghezza $2\pi r/3$. In tal caso lo spazio campione Ω è costituito dall'insieme dei punti della circonferenza e l'evento E si verifica se e solo se l'estremo P cade sull'arco \widehat{BC} . La probabilità di E risulta pertanto data dal rapporto $(2\pi r/3)/(2\pi r) = 1/3$. Infine, nel caso (c) lo spazio Ω è costituito dai punti interni alla circonferenza ed E si verifica se e solo se il punto medio della corda cade nel cerchio di raggio $r/2$ concentrico a quello di partenza. La probabilità di E può essere pertanto calcolata come rapporto tra le aree dei due cerchi: $P(E) = [\pi (r/2)^2]/(\pi r^2) = 1/4$.

Questo risultato, soltanto in apparenza paradossale, ammette una spiegazione semplice: le diverse soluzioni ottenute sono in realtà soluzioni di problemi diversi caratterizzati da spazi campione differenti. Il paradosso è in ultima analisi dovuto alla circostanza che l'enunciato del problema non definisce in modo univoco cosa debba intendersi per "tracciare a caso una corda".

Il paradosso di Bertrand costituisce un esempio paradigmatico di come generiche affermazioni quali “scegliere a caso”, “scegliere con equiprobabilità” o simili sono talora inadeguate a specificare in modo non ambiguo condizioni e criteri alla base dell’esperimento casuale considerato.

Come la formula (1.2) suggerisce, la teoria della probabilità può essere utilizzata anche per effettuare calcoli relativi a problemi non necessariamente connessi con fenomeni aleatori. Nel caso dell’esempio precedentemente considerato del cerchio inscritto nel quadrato, se si denota con n il numero di punti scelti a caso nel quadrato e con $\nu_n(E)$ il numero di questi punti che capitano nel cerchio, dalla definizione frequentista segue che per n sufficientemente grande la frequenza relativa $f_n(E) = \nu_n(E)/n$ tende a stabilizzarsi intorno al valore limite costante $P(E)$. Dalla (1.2) si ricava così una stima del numero π avendosi $\pi \approx 4 f_n(E)$. Questo costituisce un semplice esempio di applicazione di alcune tecniche, oggi largamente utilizzate, note come *metodi di Monte Carlo*. Tali tecniche si rivelano spesso utili in svariati problemi di natura non esclusivamente probabilistica, quali la valutazione di integrali definiti per funzioni di più variabili di forma anche complicata, la risoluzione di equazioni differenziali o di sistemi di equazioni algebriche.

1.6 Problemi di calcolo combinatorio

Se lo spazio campione Ω è finito e se è inoltre ragionevole assumere che gli eventi costituiti dai singleton dei suoi elementi sono equiprobabili, è possibile considerare una classe di problemi in cui, facendo uso della definizione classica, in modo naturale si assegnano probabilità ad eventi più complessi. In problemi di questo tipo ruolo fondamentale riveste il calcolo combinatorio.

Si assuma che l’insieme Ω consiste di n oggetti. In problemi coinvolgenti la scelta di oggetti da questo insieme occorre distinguere il caso in cui questa è effettuata *con rimpiazzamento* dal caso in cui essa è effettuata *senza rimpiazzamento*. Si può inoltre porre o meno l’attenzione sull’ordine con cui gli oggetti si presentano nella selezione.

Definizione 1.4 *Dicesi disposizione senza ripetizione (o, semplicemente, disposizione) di n oggetti distinguibili su k posti ogni selezione ordinata di k oggetti di Ω senza rimpiazzamento. Dicesi permutazione degli n oggetti ogni disposizione senza ripetizione degli n oggetti su n posti. Dicesi disposizione con ripetizione degli n oggetti su k posti ogni selezione ordinata con rimpiazzamento di k elementi di Ω .*

Due disposizioni si considerano distinte quando differiscono o per la scelta degli n oggetti o per l’ordine con cui essi sono distribuiti sui k posti.

Proposizione 1.1 *Il numero $D_{n,k}$ di disposizioni senza ripetizione di n oggetti su k posti è*

$$D_{n,k} = n(n-1)(n-2) \cdots (n-k+1) = \frac{n!}{(n-k)!} \quad (1 \leq k \leq n); \quad (1.3)$$

il numero P_n di permutazioni di n oggetti è

$$P_n = n! \quad (n \geq 1); \quad (1.4)$$

il numero $\hat{D}_{n,k}$ di disposizione con ripetizione di n oggetti su k posti è

$$\hat{D}_{n,k} = n^k \quad (k \geq 1). \quad (1.5)$$

Dimostrazione La (1.3) segue osservando che la scelta del primo elemento può effettuarsi in n modi diversi, quella del secondo elemento in $n - 1$ modi diversi, e così via, fino a che si perviene all'elemento k -esimo che può scegliersi in $n - k + 1$ modi diversi. La (1.4) segue immediatamente dalla (1.3) ponendo $k = n$. Infine, la (1.5) si ricava notando che la scelta di ognuno dei k oggetti può essere effettuata in n modi diversi. \square

Definizione 1.5 *Dicesi combinazione senza ripetizione (o, semplicemente, combinazione) di n oggetti a gruppi di k ogni selezione non ordinata di k oggetti tratti da Ω senza rimpiazzamento. Dicesi combinazione con ripetizione di n oggetti a gruppi di k ogni selezione non ordinata di k elementi tratti da Ω con rimpiazzamento.*

Proposizione 1.2 *Il numero $C_{n,k}$ di combinazioni di n oggetti a gruppi di k è*

$$C_{n,k} = \binom{n}{k} \quad (1 \leq k \leq n); \quad (1.6)$$

il numero $\hat{C}_{n,k}$ di combinazioni con ripetizione di n oggetti a gruppi di k è

$$\hat{C}_{n,k} = \binom{n+k-1}{k} \quad (k \geq 1). \quad (1.7)$$

Dimostrazione Calcoliamo in primo luogo $C_{n,k}$. Se si considerano tutte le possibili permutazioni degli oggetti di ciascuna combinazione senza ripetizione si ottengono tutte le disposizioni senza ripetizione di n elementi su k posti; in altri termini deve risultare $P_k C_{n,k} = D_{n,k}$. La (1.6) segue di qui facendo uso delle (1.3) e (1.4). Dimostriamo ora la (1.7). Fissati n e k , sia $\hat{C}_{n,k}$ il numero di combinazioni con ripetizione di n oggetti a gruppi di k . Il numero complessivo di oggetti presenti nelle $\hat{C}_{n,k}$ combinazioni con ripetizione è pertanto $k \hat{C}_{n,k}$. Per motivi di simmetria (nessuno degli n oggetti è privilegiato rispetto agli altri) ognuno degli n oggetti compare $k \hat{C}_{n,k}/n$ volte nell'insieme delle $\hat{C}_{n,k}$ combinazioni con ripetizione. La dimostrazione della (1.7) consiste ora nel pervenire ad un'equazione alle differenze la cui soluzione fornirà proprio $\hat{C}_{n,k}$. A tal fine, esprimiamo in una diversa forma il rapporto $k \hat{C}_{n,k}/n$. Considerato nuovamente l'insieme di tutte le combinazioni con ripetizione degli n elementi a gruppi di k , fissiamo l'attenzione su quelle che contengono almeno una volta un prefissato elemento. Indipendentemente dal numero di volte che tale elemento vi compare, lo si elimini una sola volta da ciascuna combinazione. Ciò che rimane dopo tale operazione è un insieme costituito dalle $\hat{C}_{n,k-1}$ combinazioni con ripetizione di n elementi a gruppi di $k - 1$, contenenti in totale $(k - 1) \hat{C}_{n,k-1}$ oggetti. Ognuno degli n oggetti comparirà in tale insieme $(k - 1) \hat{C}_{n,k-1}/n$ volte. Quanti elementi abbiamo eliminato in totale? Evidentemente, $\hat{C}_{n,k-1}$. Pertanto sussiste l'uguaglianza

$$\frac{k \hat{C}_{n,k}}{n} = \frac{(k - 1) \hat{C}_{n,k-1}}{n} + \hat{C}_{n,k-1},$$

ossia

$$\widehat{C}_{n,k} = \frac{n+k-1}{k} \widehat{C}_{n,k-1}.$$

Di qui, per iterazione, si ottiene:

$$\widehat{C}_{n,k} = \frac{(n+k-1)(n+k-2)\cdots(n+1)}{k(k-1)\cdots 1} \widehat{C}_{n,1}.$$

Poiché si ha evidentemente $\widehat{C}_{n,1} = n$, otteniamo in definitiva

$$\widehat{C}_{n,k} = \frac{(n+k-1)!}{k! (n-1)!} = \binom{n+k-1}{k}.$$

□

I numeri di disposizioni e di combinazioni con e senza ripetizione sono riassunti in Tabella 1.2.

Tabella 1.2 – Numero di disposizioni e di combinazioni con e senza ripetizione.

Disposizioni (l'ordine è importante nella selezione)		Combinazioni (l'ordine non è importante nella selezione)	
Senza ripetizione $1 \leq k \leq n$	Con ripetizione $k \geq 1$	Senza ripetizione $1 \leq k \leq n$	Con ripetizione $k \geq 1$
$D_{n,k} = \frac{n!}{(n-k)!}$	$\widehat{D}_{n,k} = n^k$	$C_{n,k} = \binom{n}{k}$	$\widehat{C}_{n,k} = \binom{n+k-1}{k}$

Esempio 1.4 Si scelgano due lettere dall'insieme $\Omega = \{x, y, z\}$. Le disposizioni senza ripetizione delle 3 lettere su 2 posti sono le seguenti: (x, y) , (x, z) , (y, x) , (y, z) , (z, x) , (z, y) . Il loro numero è ottenibile dalla (1.3) ponendovi $n = 3$, $k = 2$. Invece le disposizioni con ripetizione delle 3 lettere su 2 posti sono (x, x) , (x, y) , (x, z) , (y, x) , (y, y) , (y, z) , (z, x) , (z, y) , (z, z) , quindi in numero di $\widehat{D}_{3,2} = 9$, come segue dalla (1.5). Inoltre, come si ricava dalla (1.6), esistono $C_{3,2} = 3$ combinazioni di lettere dell'insieme Ω : (x, y) , (x, z) , (y, z) . Esistono poi $\widehat{C}_{3,2} = 6$ combinazioni con ripetizione di due lettere di Ω , come si deduce dalla (1.7); queste sono (x, x) , (x, y) , (x, z) , (y, y) , (y, z) , (z, z) . Infine, le permutazioni delle 3 lettere di Ω sono (x, y, z) , (x, z, y) , (y, x, z) , (y, z, x) , (z, x, y) , (z, y, x) , il cui numero è $P_3 = 6$, come si ricava dalla (1.4) per $n = 3$. ◇

Esempio 1.5 Si calcoli la probabilità che in un'estrazione del lotto esca l'ambo (20, 50) su una fissata ruota.

Lo spazio campione Ω è costituito dalle cinque che si possono formare con i numeri interi da 1 a 90. Poiché non interessa l'ordine in cui appaiono gli elementi della cinquina, gli elementi di Ω sono le combinazioni di 90 numeri a gruppi di 5. Il numero di tali cinque, che è ragionevole ritenere equiprobabili, è $N(\Omega) = \binom{90}{5}$. Si consideri ora l'evento $A = \{\text{uscita dell'ambo (20, 50) sulla fissata ruota}\}$. I casi ad esso favorevoli sono le cinque che contengono, in posizioni qualsiasi, i numeri 20 e 50. Il numero di tali cinque si ottiene pertanto calcolando il numero $N(A)$ di combinazioni dei rimanenti 88 numeri a gruppi di 3, che è $\binom{88}{3}$. Quindi, in definitiva, si ottiene:

$$P(A) = \frac{N(A)}{N(\Omega)} = \frac{\binom{88}{3}}{\binom{90}{5}} = \frac{88!}{3!85!} \frac{85!5!}{90!} = \frac{2}{89 \cdot 9} = 0.0025.$$

Evidentemente tale probabilità non cambia se al posto dei numeri 20 e 50 si pone una qualsiasi altra coppia di numeri possibili, in particolare la coppia (1, 2). \diamond

Esempio 1.6 Ci proponiamo di calcolare la probabilità che in quattro successivi lanci di un dado i risultati si presentano in ordine strettamente crescente.

Lo spazio Ω dei casi possibili è quello delle disposizioni con ripetizione di 6 elementi su 4 posti poiché in ogni lancio i sei possibili risultati sono 1, 2, 3, 4, 5, 6. Quindi il numero di casi possibili è $N(\Omega) = 6^4$ che, per ragioni di simmetria, possiamo giudicare equiprobabili. Si consideri l'evento $A = \{i \text{ risultati dei quattro lanci si presentano in ordine strettamente crescente}\}$. Il numero di casi favorevoli all'occorrenza di tale evento è $N(A) = 1 + \binom{4}{3} + \binom{5}{3} = 15$. Infatti, affinché i risultati siano in ordine strettamente crescente, l'ultimo lancio del dado deve fornire come risultato 4, 5 oppure 6; se esso fornisce come risultato 4, esiste un'unica sequenza possibile, cioè (1, 2, 3, 4); se fornisce come risultato 5 occorre considerare tutte le possibili combinazioni senza ripetizione dei quattro numeri $\{1, 2, 3, 4\}$ su tre posti, cioè le $\binom{4}{3}$ sequenze (1, 2, 3, 5), (1, 2, 4, 5), (1, 3, 4, 5), (2, 3, 4, 5); infine, se fornisce come risultato 6 occorre considerare tutte le possibili combinazioni senza ripetizione dei cinque numeri $\{1, 2, 3, 4, 5\}$ su tre posti, cioè le $\binom{5}{3}$ sequenze (1, 2, 3, 6), (1, 2, 4, 6), (1, 2, 5, 6), (1, 3, 4, 6), (1, 3, 5, 6), (1, 4, 5, 6), (2, 3, 4, 6), (2, 3, 5, 6), (2, 4, 5, 6), (3, 4, 5, 6). Pertanto risulta

$$P(A) = \frac{N(A)}{N(\Omega)} = \frac{15}{6^4} = 0.0115.$$

\diamond

Esempio 1.7 Si consideri un'urna contenente n biglie numerate da 1 a n e si supponga di estrarre k biglie dall'urna effettuando estrazioni con rimpiazzamento. Si è interessati a calcolare la probabilità che il campione estratto di k biglie, detto anche "campione di taglia k ", non contenga ripetizioni, cioè che ogni numero appaia esattamente una volta.

La cardinalità di Ω è $N(\Omega) = n^k$, pari al numero delle disposizioni con ripetizione delle n biglie su k posti. Si consideri l'evento $A = \{\text{nel campione di } k \text{ biglie non ci sono ripetizioni}\}$. Il numero $N(A)$ dei casi favorevoli ad A è uguale al numero delle disposizioni

di n biglie su k posti, ossia $N(A) = n(n-1) \cdots (n-k+1) = n!/(n-k)!$. Pertanto si ha:

$$P(A) = \frac{N(A)}{N(\Omega)} = \frac{n(n-1) \cdots (n-k+1)}{n^k} = \prod_{i=1}^{k-1} \left(1 - \frac{i}{n}\right).$$

◇

Esempio 1.8 Si consideri un ordinamento casuale dei numeri $1, 2, \dots, n$. Si calcoli la probabilità che i numeri 1 e 2 risultino consecutivi e la probabilità che i numeri 1, 2 e 3 siano anch'essi consecutivi.

Lo spazio Ω consiste di tutte le possibili permutazioni di n elementi, il cui numero è $N(\Omega) = n!$. Consideriamo i seguenti eventi: $A = \{i \text{ numeri } 1, 2 \text{ sono consecutivi nella sequenza}\}$ e $B = \{i \text{ numeri } 1, 2 \text{ e } 3 \text{ sono consecutivi nella sequenza}\}$. Il numero dei casi favorevoli ad A è $N(A) = (n-1)(n-2)!$ poiché i numeri 1, 2 possono apparire consecutivamente nella sequenza di n numeri in $n-1$ posizioni e, fissate le posizioni dei numeri 1 e 2, occorre considerare tutte le possibili permutazioni dei rimanenti $n-2$ numeri. Analogamente, il numero dei casi favorevoli a B è $N(B) = (n-2)(n-3)!$ poiché i numeri 1, 2, 3 possono essere sistemati consecutivamente nella sequenza di n numeri in $n-2$ modi e, fissate le tre posizioni dei numeri 1, 2 e 3, occorre considerare tutte le possibili permutazioni degli altri $n-3$ numeri. In conclusione, si ha:

$$P(A) = \frac{N(A)}{N(\Omega)} = \frac{(n-1)(n-2)!}{n!} = \frac{(n-1)!}{n!} = \frac{1}{n},$$

$$P(B) = \frac{N(B)}{N(\Omega)} = \frac{(n-2)(n-3)!}{n!} = \frac{1}{n(n-1)}.$$

◇

Esempio 1.9 In fila in un negozio vi sono m uomini e n donne. Supponendo che tutti siano giunti in ordine casuale, si determini la probabilità che gli uomini nella fila occupino tutti posizioni consecutive.

La cardinalità dell'insieme Ω è $N(\Omega) = (m+n)!$ essendo uguale al numero delle permutazioni delle $m+n$ persone nella fila. Consideriamo l'evento $A = \{gli \text{ uomini presenti nella fila sono tutti in posizioni consecutive}\}$. Il numero di casi favorevoli a tale evento è $N(A) = (n+1)n!m! = (n+1)!m!$. Infatti gli m uomini possono essere considerati come un unico blocco che può essere sistemato nella fila lunga $m+n$ in $n+1$ modi distinti. Fissate le m posizioni del blocco occorre considerare tutte le possibili permutazioni degli m uomini e delle n donne. In definitiva si ha:

$$P(A) = \frac{N(A)}{N(\Omega)} = \frac{(n+1)!m!}{(m+n)!}.$$

◇

Esempio 1.10 Un mazzo di n chiavi contiene la chiave che apre una determinata serratura. Ci proponiamo di calcolare la probabilità che, scegliendo ogni volta a caso una chiave diversa, sia la k -esima quella giusta.

Lo spazio Ω consiste di tutte le disposizioni senza ripetizione di n chiavi su k posti. La sua cardinalità è quindi $N(\Omega) = n(n-1) \cdots (n-k+1)$. Si consideri l'evento $A = \{\text{la } k\text{-esima chiave scelta apre la serratura}\}$. Il numero di casi ad esso favorevoli è $N(A) = (n-1)(n-2) \cdots (n-k+1)$. Infatti occorre considerare tutte le disposizioni senza ripetizione che hanno all'ultimo posto la chiave giusta mentre le rimanenti $k-1$ chiavi vanno scelte tra le $n-1$ chiavi rimanenti. Pertanto risulta:

$$P(A) = \frac{N(A)}{N(\Omega)} = \frac{(n-1)(n-2) \cdots (n-k+1)}{n(n-1) \cdots (n-k+1)} = \frac{1}{n}.$$

Si noti che il risultato non dipende da k , come è intuitivo: con uguali probabilità, ciascuna pari a $1/n$, la chiave giusta è la prima, la seconda, \dots , l' n -esima. \diamond

Esempio 1.11 Si scelgano casualmente k numeri dall'insieme $\{0, 1, \dots, 9\}$ con rimpiazzamento. Si è interessati a calcolare la probabilità che nella sequenza così costituita non siano presenti i numeri 0 e 1, nonché la probabilità che nella sequenza, supposta di lunghezza non inferiore a 3, il numero 0 appaia esattamente 3 volte.

La cardinalità dell'insieme Ω è $N(\Omega) = 10^k$ in quanto l'insieme Ω consiste di tutte le disposizioni con ripetizione di 10 elementi su k posti. Sia $A = \{\text{nella sequenza di lunghezza } k \text{ non sono presenti il numero 0 ed il numero 1}\}$. Il numero di casi favorevoli a questo evento è $N(A) = 8^k$, pari al numero di tutte le disposizioni con ripetizione degli otto elementi rimanenti su k posti. Pertanto si ha:

$$P(A) = \frac{N(A)}{N(\Omega)} = \frac{8^k}{10^k} = \left(\frac{8}{10}\right)^k \quad (k = 2, 3, \dots).$$

Si consideri ora l'evento $B = \{\text{nella sequenza di lunghezza } k \text{ il numero 0 appare esattamente 3 volte}\}$. Il numero di casi favorevoli a B è $N(B) = \binom{k}{3} 9^{k-3}$. Infatti esistono $\binom{k}{3}$ modi distinti in cui si può collocare tre volte il numero 0 in una sequenza di lunghezza k , rimanendo così disponibile una sottosequenza di lunghezza $k-3$ nella quale i rimanenti numeri $1, 2, \dots, 9$ possono collocarsi, con ripetizione, in 9^{k-3} modi distinti. Risulta quindi:

$$P(B) = \frac{N(B)}{N(\Omega)} = \frac{\binom{k}{3} 9^{k-3}}{10^k} = \binom{k}{3} \left(\frac{1}{9}\right)^3 \left(\frac{9}{10}\right)^k \quad (k \geq 3).$$

\diamond

Esempio 1.12 Due persone, denotate con U e V, sono in una fila costituita in totale da n persone. Si calcoli la probabilità che tra U e V vi siano k persone.

La cardinalità dell'insieme Ω è $N(\Omega) = n!$, pari al numero di permutazioni delle n persone nella fila. Sia $A = \{k \text{ persone tra U e V sono nella fila}\}$. Il numero di casi favorevoli ad A è $N(A) = 2(n-k-1)(n-2)!$. Il coefficiente 2 deriva dal fatto che le posizioni di U e V possono essere scambiate (U può precedere V oppure U può seguire V). Se U precede V, la scelta della posizione di U nella sequenza (in maniera tale da avere k persone che separano U da V) può essere effettuata in $n-k-1$ modi; infine, fissate le posizioni di U e V, esistono $(n-2)!$ modi di sistemare le rimanenti $n-2$ persone in fila. Si ha quindi:

$$P(A) = \frac{N(A)}{N(\Omega)} = \frac{2(n-k-1)(n-2)!}{n!}.$$

\diamond

Esempio 1.13 In un'aula vi sono k ($k \leq 365$) studenti convocati indipendentemente dalle loro date di nascita, che sono supposte equidistribuite nei 365 giorni dell'anno (ipotesi certo semplificatrice). Ci proponiamo di calcolare la probabilità che tutti gli studenti presenti abbiano distinti compleanni, nonché la probabilità che almeno 2 studenti abbiano lo stesso compleanno.

Lo spazio campione Ω è costituito da tutte le disposizioni con ripetizione di 365 elementi su k posti, in numero di $N(\Omega) = 365^k$. Sia $A = \{\text{tutti i } k \text{ studenti compiono gli anni in giorni differenti}\}$. Il numero di casi favorevoli a tale evento è $N(A) = 365 \cdot 364 \cdots (365 - k + 1)$, pari al numero delle disposizioni di 365 elementi su k posti. Pertanto si ha:

$$P(A) = \frac{N(A)}{N(\Omega)} = \frac{365 \cdot 364 \cdots (365 - k + 1)}{365^k} = \frac{364 \cdot 363 \cdots (365 - k + 1)}{365^{k-1}}.$$

Si consideri ora l'evento $B = \{\text{almeno 2 studenti compiono gli anni nello stesso giorno}\}$. Evidentemente, $B = \overline{A}$, così che il numero di casi favorevoli a B è $N(B) = N(\overline{A}) = N(\Omega) - N(A) = 365^k - 365 \cdot 364 \cdots (365 - k + 1)$. Quindi:

$$P(B) = 1 - P(A) = 1 - \frac{364 \cdot 363 \cdots (365 - k + 1)}{365^{k-1}}.$$

Tabella 1.3 – Probabilità che almeno 2 studenti tra k compiano gli anni nello stesso giorno.

k	$P(B)$	k	$P(B)$
5	0.0271	24	0.5383
10	0.1169	25	0.5687
15	0.2529	30	0.7063
20	0.4114	40	0.8912
21	0.4437	50	0.9704
22	0.4757	60	0.9941
23	0.5073	70	0.9992

La Tabella 1.3 mostra che per $k = 23$ la probabilità di trovare in aula almeno due studenti che hanno lo stesso compleanno è 0.5073 (massima incertezza!), mentre per $k = 70$ tale probabilità è molto prossima all'unità (quasi certezza!). \diamond

Capitolo 2

La teoria assiomatica

2.1 Caratterizzazione degli eventi

Se all'esperimento casuale è associato uno spazio campione Ω finito, è ragionevole richiedere che la famiglia degli eventi contenga l'evento certo e che sia chiusa sotto le operazioni di unione finita (e, quindi, anche di intersezione finita) e di complementazione. Spesso, tuttavia, vanno considerati esperimenti casuali caratterizzati da uno spazio campione infinito per i quali gli eventi di interesse sono espressi mediante operazioni di unione, intersezione e complementazione ripetute anche infinite volte, come è indicato nel seguente esempio.

Esempio 2.1 Si consideri l'esperimento consistente in una successione di lanci di una moneta, ognuno dei quali fornisce come risultato *testa* oppure *croce*. Lo spazio campione Ω è costituito da tutte le sequenze infinite di T (Testa) e C (Croce). Esso prende il nome di *spazio di Bernoulli*. Per ogni fissato intero positivo n , si considerino gli eventi $T_n = \{\text{al lancio } n\text{-esimo esce testa}\}$ ed il suo complementare $C_n = \{\text{al lancio } n\text{-esimo esce croce}\}$. T_n è dunque l'insieme di tutte le sequenze infinite di T e C aventi T nella posizione n -esima, mentre C_n è costituito da tutte le sequenze infinite di T e C aventi C nella posizione n -esima. Si definiscono *eventi dello spazio di Bernoulli* tutti i sottoinsiemi di Ω che possono ottenersi a partire dagli eventi T_n e C_n ($n = 1, 2, \dots$) mediante le operazioni di unione, intersezione e complementazione, applicate anche un numero infinito di volte. Si supponga ora di essere interessati ai seguenti eventi: $A = \{\text{testa si presenta per la prima volta in corrispondenza di un numero pari di lanci}\}$, $B = \{\text{testa si presenta per la prima volta in corrispondenza di un numero dispari di lanci}\}$, $E = \{\text{testa non appare mai}\}$.

Notiamo che l'evento $E_k = \{\text{testa si presenta per la prima volta al lancio } k\text{-esimo}\}$ ($k = 1, 2, \dots$) può esprimersi in termini degli eventi T_n e C_n nel seguente modo:

$$E_1 = T_1, \quad E_k = C_1 \cap C_2 \cap \dots \cap C_{k-1} \cap T_k \quad (k = 2, 3, \dots).$$

Gli eventi E_1, E_2, \dots così costruiti sono incompatibili essendo $E_i \cap E_j = \emptyset \quad \forall i, j: i \neq j$. È

poi immediato convincersi che si ha:

$$A = \bigcup_{k=1}^{+\infty} E_{2k} \quad B = \bigcup_{k=0}^{+\infty} E_{2k+1} \quad E = \bigcap_{k=1}^{+\infty} \overline{E_k}.$$

Gli eventi A e B sono pertanto rappresentati mediante unioni numerabili di eventi incompatibili ciascuno dei quali è a sua volta espresso mediante intersezioni degli eventi T_n e C_n . Infine, l'evento E è rappresentato mediante intersezione numerabile di eventi non incompatibili.

◇

L'Esempio 2.1 suggerisce che la famiglia degli eventi deve possedere una struttura sufficientemente ricca da consentire che componendo eventi mediante operazioni di unione, intersezione e complementazione, ripetute anche un numero infinito di volte, si ottengano ancora eventi. Occorre a tal proposito osservare che l'iterazione infinita dell'operazione di unione o intersezione di elementi di una famiglia di insiemi non sempre fornisce come risultato un elemento della famiglia stessa, come è indicato nell'esempio seguente.

Esempio 2.2 Si consideri la famiglia $\mathcal{I} = \{I_n; n = 1, 2, \dots\}$ di intervalli chiusi di \mathbb{R} così definiti:

$$I_n = \left[0, 1 - \frac{1}{n}\right].$$

Evidentemente risulta:

$$\bigcup_{n=1}^k I_n = \left[0, 1 - \frac{1}{k}\right], \quad \bigcup_{n=1}^{+\infty} I_n = [0, 1).$$

Quindi, mentre una qualsiasi unione finita di intervalli I_n appartiene alla famiglia \mathcal{I} , l'unione infinita degli I_n non è un elemento di \mathcal{I} .

◇

La circostanza che operazioni quali l'unione numerabile di elementi di una preassegnata famiglia possano generare elementi non appartenenti alla medesima famiglia suggerisce l'introduzione nella teoria della probabilità di famiglie di insiemi che siano chiuse rispetto alle operazioni di unione e complementazione. In tal modo, operando sugli insiemi di una siffatta famiglia mediante operazioni di unione e complementazione ripetute anche infinite volte, si generano insiemi anch'essi appartenenti alla famiglia stessa.

Diamo ora la seguente

Definizione 2.1 Una famiglia \mathcal{F} di sottoinsiemi di Ω costituisce una σ -algebra se sussistono le seguenti proprietà:

- (i) $\Omega \in \mathcal{F}$;
- (ii) se $A_n \in \mathcal{F}$ per $n = 1, 2, \dots$, allora $\bigcup_{n=1}^{+\infty} A_n \in \mathcal{F}$;
- (iii) se $A \in \mathcal{F}$, allora $\overline{A} \in \mathcal{F}$.

Proposizione 2.1 Se \mathcal{F} è una σ -algebra di sottoinsiemi di Ω , si ha:

(a) $\emptyset \in \mathcal{F}$;

(b) se A_1, A_2, \dots è un insieme numerabile di elementi di \mathcal{F} , allora $\bigcap_{n=1}^{+\infty} A_n \in \mathcal{F}$;

(c) se A_1, A_2, \dots, A_n è un insieme finito di elementi di \mathcal{F} , allora $\bigcup_{k=1}^n A_k \in \mathcal{F}$ e $\bigcap_{k=1}^n A_k \in \mathcal{F}$.

Dimostrazione La (a) discende immediatamente dalle (i) e (iii) della Definizione 2.1 avendosi $\emptyset = \overline{\Omega}$. La (b) segue osservando che $\overline{A_n} \in \mathcal{F}$ per la (iii), che $\bigcup_{n=1}^{+\infty} \overline{A_n} \in \mathcal{F}$ per la (ii) e che, infine, $\overline{\bigcup_{n=1}^{+\infty} \overline{A_n}} \in \mathcal{F}$ di nuovo per la (iii). D'altro canto da una delle leggi di De Morgan segue $\bigcup_{n=1}^{+\infty} \overline{A_n} = \bigcap_{n=1}^{+\infty} A_n$, il che completa la dimostrazione della (b). Notiamo, infine, che la (c) riguarda unioni ed intersezioni finite. Per dimostrarla si consideri la successione (infinita) di insiemi $\{B_n\}$, dove $B_i = A_i$ ($i = 1, 2, \dots, n$) mentre $B_{n+1} = B_{n+2} = \dots = \emptyset$. Si osservi che $\{B_n\}$ è una successione di elementi di \mathcal{F} . Infatti, per ipotesi $A_i \in \mathcal{F}$ ($i = 1, 2, \dots, n$) ed inoltre $\emptyset \in \mathcal{F}$ in virtù della (a). Per la (ii) della Definizione 2.1 si ha allora $\bigcup_{k=1}^n A_k = \bigcup_{k=1}^{+\infty} B_k \in \mathcal{F}$. Analogamente, si consideri la successione (infinita) di insiemi $\{C_n\}$, dove $C_i = A_i$ ($i = 1, 2, \dots, n$) mentre $C_{n+1} = C_{n+2} = \dots = \Omega$. Nuovamente risulta che $\{C_n\}$ è una successione di elementi di \mathcal{F} . Infatti, per ipotesi $A_i \in \mathcal{F}$ ($i = 1, 2, \dots, n$) ed inoltre $\Omega \in \mathcal{F}$ in virtù della (i) della Definizione 2.1. Dalla (b) si ha pertanto $\bigcap_{k=1}^n A_k = \bigcap_{k=1}^{+\infty} C_k \in \mathcal{F}$. \square

Dalla Definizione 2.1 e dalla Proposizione 2.1 discende che una σ -algebra \mathcal{F} , costruita a partire da uno spazio campione Ω , è una famiglia di sottoinsiemi di Ω chiusa rispetto alle operazioni di unione numerabile, intersezione numerabile e complementazione. Ciò garantisce che mediante operazioni di unione, intersezione e complementazione di elementi di \mathcal{F} , ripetute anche un numero infinito di volte, si generano elementi ancora appartenenti a \mathcal{F} .

Osservazione 2.1 Si supponga che lo spazio campione Ω sia costituito da n elementi. Poiché i sottoinsiemi di un insieme di n elementi (compreso l'insieme vuoto e l'insieme totale) sono $\sum_{k=0}^n \binom{n}{k} = 2^n$, il numero di sottoinsiemi di Ω che è possibile considerare è 2^n . La famiglia di tali sottoinsiemi, nota quale *insieme potenza*, è una σ -algebra poiché sono certamente soddisfatte le proprietà (i) ÷ (iii) della Definizione 2.1.

A partire da una famiglia \mathcal{G} di sottoinsiemi di Ω si può costruire una classe di σ -algebre ognuna contenente \mathcal{G} . Si può dimostrare che in questa classe esiste una σ -algebra minima \mathcal{F}_0 coincidente con l'intersezione di tutte le σ -algebre contenenti \mathcal{G} ; essa, che è quindi la minima σ -algebra contenente tutti gli elementi della famiglia \mathcal{G} , è denominata “ σ -algebra generata da \mathcal{G} ”.

Definizione 2.2 Dato un esperimento casuale ed individuata una famiglia \mathcal{G} di sottoinsiemi di Ω , chiameremo “famiglia degli eventi” la σ -algebra generata da \mathcal{G} e diremo “eventi” gli elementi di questa.

Osservazione 2.2 Indicheremo qui un metodo per costruire la minima σ -algebra contenente n preassegnati eventi A_1, A_2, \dots, A_n . A tal fine, consideriamo gli eventi della forma $A_1^* \cap A_2^* \cap \dots \cap A_n^*$, dove A_k^* ($k = 1, 2, \dots, n$) è o A_k oppure $\overline{A_k}$. Due eventi di questo tipo sono distinti se differiscono per almeno un A_k^* , ossia se nell'uno compare A_k e nell'altro $\overline{A_k}$. Quindi, la loro intersezione è l'evento impossibile. Ne segue che gli eventi considerati sono tra loro incompatibili ed inoltre che uno ed uno solo di essi si deve necessariamente verificare. Indichiamo con E_1, E_2, \dots, E_m i nuovi eventi non impossibili così costruiti. (Si noti che risulta $m \leq 2^n$). L'evento impossibile, gli eventi E_1, E_2, \dots, E_m e le unioni di tali eventi costituiscono tutti insieme la minima σ -algebra contenente gli eventi A_1, A_2, \dots, A_n assegnati; essa consta di 2^m eventi.

Esempio 2.3 Si consideri l'esperimento del lancio di un dado e sia $\Omega = \{1, 2, 3, 4, 5, 6\}$. Dall'Osservazione 2.1 segue che la minima σ -algebra contenente gli eventi elementari è costituita da $2^6 = 64$ eventi. La minima σ -algebra di sottoinsiemi di Ω generata dall'evento $A = \{\text{uscita di un numero pari}\}$ consta invece di soli $2^2 = 4$ eventi avendosi $\mathcal{F}_0 = \{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \Omega\}$. Se, invece, si desidera la minima σ -algebra di sottoinsiemi di Ω generata dagli eventi $A_1 = \{\text{uscita di un numero pari}\}$ e $A_2 = \{\text{uscita di un numero divisibile per } 3\}$, in analogia con il procedimento descritto nell'Osservazione 2.2 occorre considerare gli eventi

$$\begin{aligned} E_1 &= A_1 \cap A_2 = \{2, 4, 6\} \cap \{3, 6\} = \{6\}, \\ E_2 &= A_1 \cap \overline{A_2} = \{2, 4, 6\} \cap \{1, 2, 4, 5\} = \{2, 4\}, \\ E_3 &= \overline{A_1} \cap A_2 = \{1, 3, 5\} \cap \{3, 6\} = \{3\}, \\ E_4 &= \overline{A_1} \cap \overline{A_2} = \{1, 3, 5\} \cap \{1, 2, 4, 5\} = \{1, 5\}. \end{aligned}$$

La minima σ -algebra \mathcal{F}_0 contenente A_1 e A_2 è quindi costituita dai $2^4 = 16$ seguenti eventi:

$$\mathcal{F}_0 = \left\{ \emptyset, \{3\}, \{6\}, \{1, 5\}, \{2, 4\}, \{3, 6\}, \{1, 3, 5\}, \{1, 5, 6\}, \{2, 3, 4\}, \{2, 4, 6\}, \{1, 2, 4, 5\}, \right. \\ \left. \{1, 3, 5, 6\}, \{2, 3, 4, 6\}, \{1, 2, 3, 4, 5\}, \{1, 2, 4, 5, 6\}, \{1, 2, 3, 4, 5, 6\} \right\}.$$

◇

Esempio 2.4 Si supponga che lo spazio campione sia l'insieme dei numeri reali, cioè $\Omega = \mathbb{R}$, e sia $\mathcal{G} = \{(-\infty, x], x \in \mathbb{R}\}$ la famiglia degli intervalli infiniti a sinistra e chiusi a destra. La σ -algebra generata da \mathcal{G} , che in questo caso solitamente si indica con \mathcal{B} , prende il nome di *classe di Borel*. La σ -algebra generata da \mathcal{G} si ottiene componendo gli elementi di \mathcal{G} mediante operazioni di complementazione, unione (finita e numerabile) e intersezione (finita e numerabile). La classe di Borel \mathcal{B} è molto ampia; in particolare essa contiene gli insiemi costituiti da singoli punti, gli intervalli aperti, chiusi, semiaperti a destra o a sinistra, finiti o infiniti.

◇

Sia $\{A_1, A_2, \dots\}$ un insieme numerabile di eventi di \mathcal{F} . La successione A_1, A_2, \dots è non crescente se e solo se per ogni n risulta $A_{n+1} \subset A_n$, mentre è non decrescente se e solo se per ogni n si ha $A_n \subset A_{n+1}$. La successione è monotona se è non crescente oppure non decrescente. Per successioni monotone si possono definire le consuete operazioni di limite.

Definizione 2.3 Il limite di una successione monotona A_1, A_2, \dots di eventi di \mathcal{F} si definisce al seguente modo:

$$\lim_{n \rightarrow +\infty} A_n = \bigcap_{n=1}^{+\infty} A_n \quad (\text{se la successione è non crescente}),$$

$$\lim_{n \rightarrow +\infty} A_n = \bigcup_{n=1}^{+\infty} A_n \quad (\text{se la successione è non decrescente}).$$

Dalla Definizione 2.1 e dalla Proposizione 2.1 si nota che il limite di una successione monotona di eventi di \mathcal{F} è un evento poiché appartiene a \mathcal{F} . Inoltre, per successioni generali di eventi di \mathcal{F} si possono definire le consuete operazioni di limite inferiore e di limite superiore.

Definizione 2.4 Si definiscono rispettivamente limite inferiore e limite superiore di una successione A_1, A_2, \dots di eventi di \mathcal{F} i seguenti limiti:

$$\liminf_{n \rightarrow +\infty} A_n = \lim_{n \rightarrow +\infty} \bigcap_{k=n}^{+\infty} A_k, \quad \limsup_{n \rightarrow +\infty} A_n = \lim_{n \rightarrow +\infty} \bigcup_{k=n}^{+\infty} A_k.$$

Anche il limite inferiore ed il limite superiore di una successione di eventi di \mathcal{F} sono eventi in quanto appartengono a \mathcal{F} . Infatti, poiché $\{\bigcap_{k=n}^{+\infty} A_k; n = 1, 2, \dots\}$ è una successione non decrescente, utilizzando la Definizione 2.3 risulta

$$\liminf_{n \rightarrow +\infty} A_n = \lim_{n \rightarrow +\infty} \bigcap_{k=n}^{+\infty} A_k = \bigcup_{n=1}^{+\infty} \bigcap_{k=n}^{+\infty} A_k.$$

Quindi dalla Definizione 2.1 e dalla Proposizione 2.1 segue che $\liminf_{n \rightarrow +\infty} A_n \in \mathcal{F}$. Analogamente, poiché $\{\bigcup_{k=n}^{+\infty} A_k; n = 1, 2, \dots\}$ è una successione non crescente, utilizzando la Definizione 2.3 segue:

$$\limsup_{n \rightarrow +\infty} A_n = \lim_{n \rightarrow +\infty} \bigcup_{k=n}^{+\infty} A_k = \bigcap_{n=1}^{+\infty} \bigcup_{k=n}^{+\infty} A_k.$$

Utilizzando ancora la Definizione 2.1 e la Proposizione 2.1, si ricava che $\limsup_{n \rightarrow +\infty} A_n \in \mathcal{F}$.

Occorre osservare che $\liminf_{n \rightarrow +\infty} A_n$ è l'evento che si verifica se e solo se esiste almeno un indice n tale che tutti gli eventi A_n, A_{n+1}, \dots si verificano, ossia se e solo se si verificano tutti gli eventi della successione tranne al più un numero finito. Inoltre $\limsup_{n \rightarrow +\infty} A_n$ è l'evento che si verifica se e solo per ogni n almeno uno degli eventi A_n, A_{n+1}, \dots si verifica, ossia se e solo se si verificano infiniti eventi della successione. In generale si ha quindi $\liminf_{n \rightarrow +\infty} A_n \subset \limsup_{n \rightarrow +\infty} A_n$. Se risulta $\liminf_{n \rightarrow +\infty} A_n = \limsup_{n \rightarrow +\infty} A_n = A$, si dice che la successione $\{A_n; n = 1, 2, \dots\}$ ammette limite e si scrive $\lim_{n \rightarrow +\infty} A_n = A$.

Nell'esempio che segue esamineremo alcune proprietà di successioni di eventi delle quali si farà uso nel seguito per la dimostrazione di un'importante disuguaglianza (v. Paragrafo 2.3).

Esempio 2.5 Data una successione $\{A_n; n = 1, 2, \dots\}$ di eventi di \mathcal{F} , si definisca una nuova successione $\{B_n; n = 1, 2, \dots\}$, con $B_1 = A_1$, $B_n = \overline{A_1} \cap \overline{A_2} \cap \dots \cap \overline{A_{n-1}} \cap A_n$ ($n = 2, 3, \dots$). Si vuole dimostrare che risulta

$$\bigcup_{n=1}^{+\infty} A_n = \bigcup_{n=1}^{+\infty} B_n \quad (2.1)$$

ed inoltre che gli eventi della successione $\{B_n; n = 1, 2, \dots\}$ sono incompatibili.

In primo luogo dimostriamo per induzione che sussiste la relazione $\bigcup_{n=1}^k B_n = \bigcup_{n=1}^k A_n$ ($k = 1, 2, \dots$). Tale uguaglianza è vera per $k = 1$; supponendo che sia valida per k , utilizzando le leggi di De Morgan dimostriamo che essa sussiste per $k + 1$. A tal fine, osserviamo che

$$\begin{aligned} \bigcup_{n=1}^{k+1} B_n &= \bigcup_{n=1}^k B_n \cup B_{k+1} = \bigcup_{n=1}^k A_n \cup \left[A_{k+1} \cap \left(\bigcup_{n=1}^k \overline{A_n} \right) \right] \\ &= \bigcup_{n=1}^{k+1} A_n \cap \left[\bigcup_{n=1}^k A_n \cup \left(\bigcup_{n=1}^k \overline{A_n} \right) \right] = \bigcup_{n=1}^{k+1} A_n \cap \Omega = \bigcup_{n=1}^{k+1} A_n. \end{aligned}$$

La (2.1) segue immediatamente osservando che le successioni $\{\bigcup_{n=1}^k A_n; k = 1, 2, \dots\}$ e $\{\bigcup_{n=1}^k B_n; k = 1, 2, \dots\}$ sono non decrescenti e che quindi ammettono limite. Notiamo, infine, che gli eventi della successione $\{B_n; n = 1, 2, \dots\}$ sono incompatibili risultando $B_1 \cap B_i = A_1 \cap (\overline{A_1} \cap \overline{A_2} \cap \dots \cap \overline{A_{i-1}} \cap A_i) = \emptyset$ ($i = 2, 3, \dots$) ed inoltre $B_i \cap B_j = (\overline{A_1} \cap \overline{A_2} \cap \dots \cap \overline{A_{i-1}} \cap A_i) \cap (\overline{A_1} \cap \overline{A_2} \cap \dots \cap \overline{A_{j-1}} \cap A_j) = \emptyset$ ($i, j = 2, 3, \dots; i < j$). \diamond

2.2 Definizione assiomatica di probabilità

Come mostrato nel Paragrafo 1.4, le diverse definizioni di probabilità date condividono le seguenti tre proprietà: (i) $P(A) \geq 0$ per ogni evento A ; (ii) $P(\Omega) = 1$; (iii) se A e B sono eventi incompatibili, $P(A \cup B) = P(A) + P(B)$. Passeremo ora brevemente in rassegna gli elementi essenziali dell'impostazione assiomatica della teoria della probabilità, nell'ambito della quale le suddette proprietà costituiranno la base intuitiva di partenza. Occorrerà, peraltro, riferirsi a famiglie di eventi caratterizzate da una struttura che consenta di assegnare ad ogni evento una misura, che qui verrà identificata con la probabilità dell'evento. Come si è già detto, l'impostazione assiomatica della teoria della probabilità è dovuta principalmente ad A.N. Kolmogorov.

Definizione 2.5 Si dice spazio probabilizzabile ogni coppia (Ω, \mathcal{F}) , dove Ω è uno spazio campione e \mathcal{F} è una σ -algebra generata da sottoinsiemi di Ω .

Definizione 2.6 Dato uno spazio probabilizzabile (Ω, \mathcal{F}) , si dice misura di probabilità (o probabilità) una funzione di insieme $P : \mathcal{F} \rightarrow \mathbb{R}$ che gode delle seguenti proprietà:

- (i) $P(A) \geq 0 \quad \forall A \in \mathcal{F}$;

(ii) $P(\Omega) = 1$;

(iii) se $\{A_n; n = 1, 2, \dots\}$ è una successione di eventi incompatibili di \mathcal{F} , ossia una famiglia di elementi di \mathcal{F} tali che $A_i \cap A_j = \emptyset \forall i, j = 1, 2, \dots, i \neq j$, allora

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

La proprietà (iii) esprime l'*additività completa* o *additività numerabile* della probabilità. D'ora innanzi ci riferiremo alle proprietà (i), (ii) e (iii) come agli *assiomi della probabilità*, che, per brevità, chiameremo semplicemente *assiomi* e che denoteremo rispettivamente come “primo assioma”, “secondo assioma” e “terzo assioma”.

Definizione 2.7 Dicesi spazio di probabilità ogni tripla (Ω, \mathcal{F}, P) , dove Ω è uno spazio campione, \mathcal{F} è una σ -algebra generata da sottoinsiemi di Ω e P è una misura di probabilità su (Ω, \mathcal{F}) .

Dalla Definizione 2.6 seguono alcune proposizioni di semplice dimostrazione, particolarmente significative dal punto di vista probabilistico.

Proposizione 2.2

$$P(\emptyset) = 0. \quad (2.2)$$

Dimostrazione Si consideri la successione di eventi $\{A_n; n = 1, 2, \dots\}$ tale che $A_n = \emptyset$ per ogni n . Poiché $\bigcup_{n=1}^{+\infty} A_n = \emptyset \in \mathcal{F}$, per il terzo assioma si ha

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

che, per il primo assioma, è verificata se e solo se $P(\emptyset) = 0$. □

La (2.2) esprime la circostanza che l'evento impossibile ha probabilità nulla di verificarsi. È bene però precisare che non è detto che un evento a probabilità zero coincida con l'evento impossibile. Ad esempio, nell'esperimento consistente nella scelta casuale di un numero reale, con $\Omega = \mathbb{R}$, l'evento $A = \{\text{uscita di un prefissato numero reale } \alpha\}$ deve avere probabilità zero, pur non identificandosi con l'evento impossibile. Analogamente, un evento avente probabilità 1 non coincide necessariamente con l'evento certo Ω . Ad esempio, l'evento $\overline{A} = \{\text{non uscita di un prefissato numero reale } \alpha\}$ ha probabilità 1, pur non coincidendo con Ω .

Definizione 2.8 Un evento A si dice quasi certo se è $P(A) = 1$; se invece risulta $P(A) = 0$, l'evento A viene detto quasi impossibile.

Dimostriamo ora che l'additività numerabile della misura P , di cui alla (iii) della Definizione 2.1, implica la sua additività finita.

Proposizione 2.3 Siano A_1, A_2, \dots, A_k eventi incompatibili di \mathcal{F} . Si ha:

$$P\left(\bigcup_{n=1}^k A_n\right) = \sum_{n=1}^k P(A_n). \quad (2.3)$$

Dimostrazione Si consideri una successione di eventi $\{B_n; n = 1, 2, \dots\}$ tali che

$$B_n = A_n \quad (n = 1, 2, \dots, k), \quad B_{k+1} = B_{k+2} = \dots = \emptyset.$$

Si noti che risulta $B_i \cap B_j = \emptyset$ per $i \neq j$, così che la successione $\{B_n; n = 1, 2, \dots\}$ è costituita da eventi incompatibili. Utilizzando il terzo assioma della probabilità e la (2.2) si ha:

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

che esprime proprio l'additività finita della funzione P . \square

Proposizione 2.4 Per ogni $A \in \mathcal{F}$ si ha:

$$P(\bar{A}) = 1 - P(A). \quad (2.4)$$

Dimostrazione Dalla Definizione 2.1 segue che se $A \in \mathcal{F}$ allora anche $\bar{A} \in \mathcal{F}$; inoltre risulta $\Omega = A \cup \bar{A}$. Poiché A e \bar{A} sono eventi incompatibili, dalla Proposizione 2.3 e ricordando il secondo assioma si ha:

$$1 = P(\Omega) = P(A \cup \bar{A}) = P(A) + P(\bar{A}),$$

da cui segue la (2.4). \square

Dalla Proposizione 2.4 discende che il complementare di un evento quasi certo è un evento quasi impossibile, e viceversa.

Proposizione 2.5 Per ogni $A \in \mathcal{F}$ si ha $0 \leq P(A) \leq 1$.

Dimostrazione Se $A \in \mathcal{F}$, dal primo assioma si ha $P(A) \geq 0$; inoltre, poiché $\bar{A} \in \mathcal{F}$, dalla Proposizione 2.4 e dal primo assioma segue $P(\bar{A}) = 1 - P(A) \geq 0$, e quindi $P(A) \leq 1$. \square

La probabilità P è dunque una funzione $P : \mathcal{F} \rightarrow [0, 1]$ che ad ogni evento $A \in \mathcal{F}$ assegna un reale dell'intervallo $[0, 1]$ nel rispetto della Definizione 2.6.

Proposizione 2.6 Siano A e B eventi appartenenti ad \mathcal{F} . Se $A \subset B$ allora $P(A) \leq P(B)$.

Dimostrazione Anzitutto osserviamo che, essendo $A \subset B$, è possibile esprimere B come unione degli eventi incompatibili A e $\bar{A} \cap B$ rappresentanti rispettivamente la parte di B comune ad A e la parte di B che non è comune ad A :

$$B = A \cup (\bar{A} \cap B), \quad \text{con} \quad A \cap (\bar{A} \cap B) = \emptyset.$$

Dal primo assioma e dalla Proposizione 2.3 segue allora:

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

\square

Una conseguenza di quanto testé dimostrato è che se A e B sono eventi di \mathcal{F} , risulta:

$$P(A \cap B) \leq \min\{P(A), P(B)\} \leq \max\{P(A), P(B)\} \leq P(A \cup B).$$

Osservazione 2.3 Se $\{A_n; n = 1, 2, \dots\}$ è una famiglia finita o numerabile di eventi di \mathcal{F} e se $\bigcup_{n=1}^{+\infty} A_n = \Omega$, allora $P\left(\bigcup_{n=1}^{+\infty} A_n\right) = 1$.

Esempio 2.6 In un esperimento casuale con spazio campione Ω si consideri la σ -algebra generata dagli eventi A_1 e A_2 , e si ponga

$$P(A_1) = a, \quad P(A_2) = b, \quad P(A_1 \cap A_2) = c.$$

Si desidera determinare delle condizioni su a, b, c in modo tale che siano soddisfatti gli assiomi della probabilità.

Osserviamo in primo luogo che dalla Proposizione 2.5 segue $0 \leq a \leq 1$, $0 \leq b \leq 1$ e $0 \leq c \leq 1$. Tali condizioni però non sono sufficienti a garantire che siano soddisfatti gli assiomi della probabilità. In base all'Osservazione 2.2 consideriamo gli eventi

$$E_1 = A_1 \cap A_2, \quad E_2 = A_1 \cap \overline{A_2}, \quad E_3 = \overline{A_1} \cap A_2, \quad E_4 = \overline{A_1} \cap \overline{A_2}.$$

Poiché essi sono incompatibili, dalle ipotesi fatte e dalla Proposizione 2.3 segue:

$$\begin{aligned} P(E_1) &= P(A_1 \cap A_2) = c, \\ P(E_2) &= P(A_1 \cap \overline{A_2}) = P(A_1) - P(A_1 \cap A_2) = a - c, \\ P(E_3) &= P(\overline{A_1} \cap A_2) = P(A_2) - P(A_1 \cap A_2) = b - c, \\ P(E_4) &= 1 - P(E_1) - P(E_2) - P(E_3) = 1 - a - b + c. \end{aligned}$$

Quindi, affinché siano soddisfatti gli assiomi della probabilità occorre imporre che risulti $P(E_i) \geq 0$ ($i = 1, 2, 3, 4$), ossia $c \geq 0$, $a - c \geq 0$, $b - c \geq 0$ e $1 - a - b + c \geq 0$. Quindi,

$$0 \leq a \leq 1, \quad 0 \leq b \leq 1, \quad \max(0, a + b - 1) \leq c \leq \min(a, b).$$

Così, se $P(A_1) = 1/2$ e $P(A_2) = 3/4$ allora risulta $\max(0, a + b - 1) = 1/4$ e $\min(a, b) = 1/2$ e pertanto deve essere $1/4 \leq c \leq 1/2$. \diamond

Esempio 2.7 Con riferimento all'esperimento dell'Esempio 2.1 si calcolino le probabilità dei seguenti eventi: $E_k = \{\text{testa si presenta per la prima volta al lancio } k\text{-esimo}\}$ ($k = 1, 2, \dots$), $A = \{\text{testa si presenta per la prima volta in corrispondenza di un numero pari di lanci}\}$, $B = \{\text{testa si presenta per la prima volta in corrispondenza di un numero dispari di lanci}\}$, $E = \{\text{testa non si presenta mai}\}$.

Se i lanci sono k , i 2^k possibili risultati per ragioni di simmetria possono considerarsi ugualmente probabili. Uno solo di essi realizza la sequenza di $k - 1$ croci seguite da una testa, verificando con ciò E_k . Quindi $P(E_k) = 1/2^k$. Essendo gli eventi E_1, E_2, \dots incompatibili, dal terzo assioma segue:

$$P(A) = P\left(\bigcup_{k=1}^{+\infty} E_{2k}\right) = \sum_{k=1}^{+\infty} P(E_{2k}) = \sum_{k=1}^{+\infty} \frac{1}{2^{2k}} = \sum_{k=1}^{+\infty} \left(\frac{1}{4}\right)^k = \frac{1}{3}.$$

Ragionando in maniera analoga, si ottiene poi:

$$P(B) = P\left(\bigcup_{k=0}^{+\infty} E_{2k+1}\right) = \sum_{k=0}^{+\infty} P(E_{2k+1}) = \sum_{k=0}^{+\infty} \frac{1}{2^{2k+1}} = \frac{2}{3}.$$

Infine, essendo $A \cup B \cup E = \Omega$, con A , B ed E incompatibili, segue:

$$P(E) = 1 - \frac{1}{3} - \frac{2}{3} = 0.$$

L'evento E è pertanto quasi impossibile. ◇

La funzione probabilità gode dell'importante proprietà di essere una funzione di insieme continua come emerge dal seguente teorema che ci limitiamo ad enunciare.

Teorema 2.1 *Se $\{A_n; n = 1, 2, \dots\}$ è una successione di eventi di \mathcal{F} dotata di limite, posto*

$$\lim_{n \rightarrow +\infty} A_n = A,$$

risulta

$$\lim_{n \rightarrow +\infty} P(A_n) = P(A).$$

2.3 Disuguaglianza di Boole e formula di inclusione-esclusione

Il terzo assioma della probabilità esprime la proprietà di additività completa per eventi incompatibili da cui, come si è visto, segue anche la proprietà di additività finita. Esaminiamo ora il caso in cui si abbandona l'ipotesi di incompatibilità degli eventi.

Teorema 2.2 (*Disuguaglianza di Boole*) *Se $\{A_n; n = 1, 2, \dots\}$ è una successione di eventi di \mathcal{F} , si ha:*

$$P\left(\bigcup_{n=1}^{+\infty} A_n\right) \leq \sum_{n=1}^{+\infty} P(A_n) \quad (2.5)$$

Dimostrazione Consideriamo la successione ausiliaria di eventi $\{B_n; n = 1, 2, \dots\}$ con $B_1 = A_1, B_n = \overline{A_1} \cap \overline{A_2} \cap \dots \cap \overline{A_{n-1}} \cap A_n$ ($n = 2, 3, \dots$). Come si è dimostrato nell'Esempio 2.5 gli eventi di tale successione sono incompatibili ed inoltre risulta $\bigcup_{n=1}^{+\infty} A_n = \bigcup_{n=1}^{+\infty} B_n$. Per il terzo assioma si ha quindi:

$$P\left(\bigcup_{n=1}^{+\infty} A_n\right) = P\left(\bigcup_{n=1}^{+\infty} B_n\right) = \sum_{n=1}^{+\infty} P(B_n) \leq \sum_{n=1}^{+\infty} P(A_n), \quad (2.6)$$

dove l'ultima disuguaglianza segue dall'essere $A_1 = B_1, B_n \subset A_n$ per $n = 2, 3, \dots$ e dalla Proposizione 2.6. □

La disuguaglianza di Boole fornisce una maggiorazione della probabilità dell'unione di un numero finito o numerabile di eventi che, ovviamente, è significativa solo se risulta $\sum_{n=1}^{+\infty} P(A_n) \leq 1$. L'esempio seguente indica un'applicazione della disuguaglianza di Boole.

Esempio 2.8 (*Problema degli insiemi monocromatici*) Sia $\{S_1, S_2, \dots, S_n\}$ una famiglia di sottoinsiemi di un insieme \mathcal{U} , ciascuno dei quali contiene k elementi. Diremo colorazione di \mathcal{U} ogni funzione che a ciascun elemento di \mathcal{U} associa uno di due colori, che qui supporremo essere rosso e blu. Diremo che S_i è monocromatico se tutti i suoi elementi hanno lo stesso colore. Nel 1963 Paul Erdős dimostrò il seguente risultato: “Se è $n < 2^{k-1}$, allora esiste una colorazione tale che nessuno degli insiemi S_1, S_2, \dots, S_n è monocromatico”. Per dimostrarlo, si consideri una colorazione casuale degli elementi di \mathcal{U} , ossia si supponga che ogni elemento di \mathcal{U} abbia colore rosso o blu con probabilità $1/2$. Il numero di possibili colorazioni distinte è $N(\mathcal{U}) = 2^{|\mathcal{U}|}$, dove $|\mathcal{U}|$ denota la cardinalità dell'insieme \mathcal{U} . Si consideri ora l'evento $A_i = \{\text{l'insieme } S_i \text{ è monocromatico}\}$; il numero di casi favorevoli a tale evento è $N(A_i) = 2 \cdot 2^{|\mathcal{U}|-k}$. Infatti i k elementi di S_i devono avere tutti colore rosso oppure tutti colore blu ed, una volta fissato tale colore, ciascuno dei rimanenti $|\mathcal{U}| - k$ elementi di \mathcal{U} può essere colorato in due differenti modi. Pertanto, per la definizione classica di probabilità, si ha $P(A_i) = 2 \cdot 2^{|\mathcal{U}|-k} / 2^{|\mathcal{U}|} = 2^{-(k-1)}$ ($i = 1, 2, \dots, n$). Se si considera l'evento $A = \{\text{almeno uno degli insiemi } S_1, S_2, \dots, S_n \text{ è monocromatico}\}$, risulta $A = \cup_{i=1}^n A_i$. Quindi, per la disuguaglianza di Boole si ha:

$$P(A) = P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i) = \sum_{i=1}^n \left(\frac{1}{2}\right)^{k-1} = \frac{n}{2^{k-1}} < 1,$$

dove l'ultima disuguaglianza segue dall'aver ipotizzato che risulta $n < 2^{k-1}$. Dall'essere $P(A) < 1$, segue che l'evento complementare $\bar{A} = \{\text{nessuno degli insiemi } S_1, S_2, \dots, S_n \text{ è monocromatico}\}$ ha probabilità positiva. Dunque esiste almeno una colorazione tale che nessuno degli insiemi S_1, S_2, \dots, S_n è monocromatico. \diamond

Il precedente esempio è particolarmente istruttivo anche perché indica l'utilizzazione di considerazioni probabilistiche per la risoluzione di un problema di natura non probabilistica.

Teorema 2.3 Se A_1 e A_2 sono eventi di \mathcal{F} , si ha:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2). \quad (2.7)$$

Dimostrazione Dalle proprietà (9) e (12) del Paragrafo 1.3 risulta:

$$A_1 \cup A_2 = A_1 \cup (\bar{A}_1 \cap A_2), \quad A_2 = (A_1 \cap A_2) \cup (\bar{A}_1 \cap A_2). \quad (2.8)$$

Si noti che gli eventi $A_1 \cup A_2$ e A_2 sono così stati entrambi espressi come unioni di eventi incompatibili poiché risulta $A_1 \cap (\bar{A}_1 \cap A_2) = \emptyset$ e $(A_1 \cap A_2) \cap (\bar{A}_1 \cap A_2) = \emptyset$. Facendo uso della Proposizione 2.3 in (2.8) si ha:

$$P(A_1 \cup A_2) = P(A_1) + P(\bar{A}_1 \cap A_2), \quad P(A_2) = P(A_1 \cap A_2) + P(\bar{A}_1 \cap A_2) \quad (2.9)$$

da cui, per eliminazione di $P(\bar{A}_1 \cap A_2)$, segue la tesi. \square

Si noti che dalla (2.7) discende $P(A_1 \cup A_2) \leq P(A_1) + P(A_2)$, in accordo con quanto previsto dalla disuguaglianza di Boole.

Esempio 2.9 Un imprenditore si avvale dell'opera di due collaboratori C_1 e C_2 . Con riferimento ad un prefissato giorno lavorativo, si considerino gli eventi $A_i = \{\text{il collaboratore } C_i \text{ è assente nel giorno fissato}\}$ ($i = 1, 2$) e si supponga che $P(A_1) = 3/100$, $P(A_2) = 4/100$ e $P(A_1 \cap A_2) = 1/100$. Si calcolino le probabilità dei seguenti eventi: $E_1 = \{\text{l'imprenditore si avvale al più di un collaboratore}\}$, $E_2 = \{\text{l'imprenditore si avvale almeno di un collaboratore}\}$, $E_3 = \{\text{l'imprenditore si avvale di un solo collaboratore}\}$.

Poiché $E_1 = A_1 \cup A_2$, dal Teorema 2.3 segue $P(E_1) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = 6/100$. Inoltre, essendo $E_2 = \overline{A_1 \cap A_2}$, si ha $P(E_2) = 1 - P(A_1 \cap A_2) = 99/100$. Infine, osservando che $E_3 = (A_1 \cap \overline{A_2}) \cup (\overline{A_1} \cap A_2)$ (ossia che è esprimibile come unione di eventi incompatibili), risulta $P(E_3) = P(A_1 \cap \overline{A_2}) + P(\overline{A_1} \cap A_2) = 5/100$, dove l'ultima uguaglianza segue in quanto $P(A_1 \cap \overline{A_2}) = P(A_1) - P(A_1 \cap A_2)$ e $P(\overline{A_1} \cap A_2) = P(A_2) - P(A_1 \cap A_2)$. \diamond

La seguente proposizione ha delle interessanti implicazioni.

Proposizione 2.7 Se A_1 e A_2 sono eventi di \mathcal{F} , si ha:

- (a) $P(A_1) = 0 \Rightarrow P(A_1 \cap A_2) = 0$ e $P(A_1 \cup A_2) = P(A_2)$;
- (b) $P(A_1) = 1 \Rightarrow P(A_1 \cup A_2) = 1$ e $P(A_1 \cap A_2) = P(A_2)$.

Dimostrazione Dimostriamo in primo luogo l'implicazione (a). Poiché $A_1 \cap A_2 \subset A_1$, applicando la Proposizione 2.6 si ha $P(A_1 \cap A_2) \leq P(A_1)$. Poiché $P(A_1) = 0$ per ipotesi, ricordando il primo assioma si ha $0 \leq P(A_1 \cap A_2) \leq P(A_1) = 0$, da cui segue $P(A_1 \cap A_2) = 0$. Inoltre, dalla (2.7) risulta $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = P(A_2)$. Dimostriamo ora l'implicazione (b). Poiché $A_1 \subset A_1 \cup A_2$, in virtù della Proposizione 2.6 si ha $P(A_1) \leq P(A_1 \cup A_2)$. Poiché $P(A_1) = 1$ per ipotesi, ricordando la Proposizione 2.5 si ottiene $1 = P(A_1) \leq P(A_1 \cup A_2) \leq 1$, da cui segue $P(A_1 \cup A_2) = 1$; dalla (2.7) segue infine $P(A_1 \cap A_2) = P(A_1) + P(A_2) - P(A_1 \cup A_2) = P(A_2)$. \square

Notiamo che un evento A rimane invariato se unito all'evento impossibile avendosi $A \cup \emptyset = A$. Se si sostituisce a \emptyset un evento quasi impossibile l'uguaglianza non sussiste più, pur continuando a sussistere l'uguaglianza tra le rispettive probabilità, come mostrato nella Proposizione 2.7. Pertanto si può affermare che un evento quasi impossibile unito ad un evento non modifica la probabilità dell'evento unione. Analogamente, un evento quasi certo intersecato con un evento non modifica la probabilità dell'evento intersezione.

La (2.7) è suscettibile di generalizzazione al caso di un numero finito di eventi. L'estensione del risultato (2.7) al caso di n eventi arbitrari è detta *formula di Poincaré* o, più significativamente, *formula di inclusione-esclusione*. Infatti la probabilità dell'unione di n eventi è data dalla somma delle probabilità di ciascuno dei singoli eventi diminuita della somma delle probabilità delle intersezioni a due a due aumentata della somma delle probabilità delle intersezioni a tre a tre e così via, fino ad arrivare alla probabilità dell'intersezione di tutti gli n eventi da aggiungersi con il segno $(-1)^{n+1}$. Sussiste infatti il seguente teorema:

Teorema 2.4 Se $A_1, A_2, \dots, A_n \in \mathcal{F}$, allora

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{i=1}^n P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1} P(A_1 \cap A_2 \cap \dots \cap A_n). \quad (2.10)$$

Dimostrazione Si può procedere per induzione. Osserviamo anzitutto che, in virtù del Teorema 2.3, la (2.10) sussiste per $n = 2$. Supponendo che essa sia valida per $n = r$, dimostriamone la validità per $n = r + 1$. A tal fine, posto

$$B_r = \bigcup_{i=1}^r A_i,$$

in virtù del Teorema 2.3 segue:

$$P\left(\bigcup_{i=1}^{r+1} A_i\right) = P(B_r \cup A_{r+1}) = P(B_r) + P(A_{r+1}) - P(B_r \cap A_{r+1}).$$

Poiché la (2.10) è stata assunta valida per $n = r$, si ha anche:

$$\begin{aligned} P\left(\bigcup_{i=1}^{r+1} A_i\right) &= P\left(\bigcup_{i=1}^r A_i\right) + P(A_{r+1}) - P\left[\left(\bigcup_{i=1}^r A_i\right) \cap A_{r+1}\right] \\ &= \left[\sum_{i=1}^r P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \dots \right. \\ &\quad \left. + (-1)^{r+1} P(A_1 \cap A_2 \cap \dots \cap A_r) \right] + P(A_{r+1}) - P\left[\bigcup_{i=1}^r (A_i \cap A_{r+1})\right]. \end{aligned} \quad (2.11)$$

Osserviamo ora che, applicando di nuovo la formula (2.10), risulta:

$$\begin{aligned} P\left[\bigcup_{i=1}^r (A_i \cap A_{r+1})\right] &= \sum_{i=1}^r P(A_i \cap A_{r+1}) - \sum_{i<j} P[(A_i \cap A_{r+1}) \cap (A_j \cap A_{r+1})] \\ &\quad + \sum_{i<j<k} P[(A_i \cap A_{r+1}) \cap (A_j \cap A_{r+1}) \cap (A_k \cap A_{r+1})] - \dots \\ &\quad + (-1)^{r+1} P[(A_1 \cap A_{r+1}) \cap (A_2 \cap A_{r+1}) \cap \dots \cap (A_r \cap A_{r+1})]. \end{aligned}$$

Facendo uso della proprietà associativa dell'intersezione, si ha poi:

$$\begin{aligned} P\left[\bigcup_{i=1}^r (A_i \cap A_{r+1})\right] &= \sum_{i=1}^r P(A_i \cap A_{r+1}) - \sum_{i<j} P(A_i \cap A_j \cap A_{r+1}) \\ &\quad + \sum_{i<j<k} P(A_i \cap A_j \cap A_k \cap A_{r+1}) - \dots + (-1)^{r+1} P(A_1 \cap A_2 \cap \dots \cap A_{r+1}). \end{aligned} \quad (2.12)$$

Sostituendo la (2.12) nell'ultima delle (2.11), si ricava infine la (2.10) scritta per $n = r + 1$. \square

Si osservi, in particolare, che se $A_1, A_2, A_3 \in \mathcal{F}$, la (2.10) dà:

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3).$$

Vale poi la pena di menzionare che se A_1, A_2, \dots, A_n sono eventi incompatibili, dalla (2.10) segue la legge (2.3) di additività finita della probabilità.

Esempio 2.10 (*Il problema delle concordanze*) In un'urna sono contenute n biglie numerate da 1 a n . Si consideri l'esperimento consistente nell'estrarre le biglie dall'urna l'una dopo l'altra fino ad esaurimento. Si dice che esiste una *concordanza* all'estrazione i -esima se la biglia contrassegnata con il numero i si presenta all' i -esima estrazione. Si calcolino le probabilità dei seguenti eventi: $B_n = \{\text{si ha almeno una concordanza nelle } n \text{ estrazioni}\}$, $C_n = \{\text{non si ha nessuna concordanza nelle } n \text{ estrazioni}\}$, $F_{n,r} = \{\text{si hanno esattamente } r \text{ concordanze nelle } n \text{ estrazioni}\}$.

Osserviamo in primo luogo che l'evento $A_i = \{\text{si ha una concordanza alla } i\text{-esima estrazione}\}$ ha probabilità

$$P(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n} \quad (i = 1, 2, \dots, n).$$

Infatti, l'esperimento considerato ha come risultato una qualsiasi delle $n!$ possibili permutazioni degli interi $1, 2, \dots, n$; di queste ve ne sono $(n-1)!$ nelle quali si verifica la concordanza esattamente all' i -esima estrazione. (Più semplicemente, le uscite della biglia i -esima alla prima, seconda, \dots , n -esima estrazione sono equiprobabili e quindi ciascuna, ivi compresa l'uscita della biglia i -esima proprio all' i -esima estrazione, ha probabilità $1/n$). Analogamente, la probabilità che si abbiano r concordanze nelle estrazioni i_1, i_2, \dots, i_r è data da

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_r}) = \frac{(n-r)!}{n!},$$

poiché tra le $n!$ possibili permutazioni degli interi $1, 2, \dots, n$ ve ne sono $(n-r)!$ nelle quali si verificano le concordanze esattamente nelle estrazioni i_1, i_2, \dots, i_r . Evidentemente risulta $B_n = \bigcup_{i=1}^n A_i$ e quindi, facendo uso della formula di inclusione-esclusione, si ottiene:

$$\begin{aligned} P(B_n) &= P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cap A_j) \\ &\quad + \sum_{i < j < k} P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1} P(A_1 \cap A_2 \cap \dots \cap A_n) \\ &= \sum_{i=1}^n \frac{1}{n} - \sum_{i < j} \frac{(n-2)!}{n!} + \sum_{i < j < k} \frac{(n-3)!}{n!} - \dots + (-1)^{n+1} \frac{1}{n!} \\ &= 1 - \binom{n}{2} \frac{(n-2)!}{n!} + \binom{n}{3} \frac{(n-3)!}{n!} - \dots + (-1)^{n+1} \frac{1}{n!} \end{aligned}$$

$$= 1 - \frac{1}{2!} + \frac{1}{3!} - \dots + (-1)^{n+1} \frac{1}{n!} = \sum_{i=1}^n (-1)^{i+1} \frac{1}{i!}, \quad (2.13)$$

poiché il numero di intersezioni distinte di k eventi è $\binom{n}{k}$. Si noti ora che $C_n = \bigcap_{i=1}^n \overline{A_i}$ così che, facendo uso delle formule di De Morgan, risulta:

$$C_n = \overline{\left(\bigcup_{i=1}^n A_i \right)} = \overline{B_n}.$$

Pertanto, ricordando (2.13), si ha:

$$P(C_n) = 1 - P(B_n) = \sum_{i=0}^n (-1)^i \frac{1}{i!}.$$

È interessante notare che quando $n \rightarrow +\infty$ la probabilità di non avere nessuna concordanza è

$$\lim_{n \rightarrow +\infty} P(C_n) = \frac{1}{e} = 0.3678795.$$

Il reale e^{-1} fornisce dunque un'approssimazione per la probabilità di non ottenere nessuna concordanza nelle estrazioni al crescere del numero n delle biglie.

La Tabella 2.1 mostra che le probabilità di non avere nessuna concordanza nelle n estrazioni variano di poco con n e raggiungono molto rapidamente il valore limite e^{-1} .

Tabella 2.1— Probabilità di non avere nessuna concordanza nelle n estrazioni. (Le approssimazioni sono alla settima cifra decimale).

n	$P(C_n)$	n	$P(C_n)$
2	0.5	7	0.3678572
3	0.3333333	8	0.3678820
4	0.3750000	9	0.3678792
5	0.3666667	10	0.3678795
6	0.3680556	$+\infty$	0.3678795

Per calcolare la probabilità di avere esattamente r concordanze nelle n estrazioni, consideriamo l'evento $D_{n,r} = \{\text{si hanno esattamente } r \text{ fissate concordanze nelle } n \text{ estrazioni delle } n \text{ biglie}\}$. In questo caso vi sono $n!$ permutazioni possibili delle n biglie e tra queste sono favorevoli quelle che, tenendo fisse le r biglie nei posti dove esiste concordanza, permutano le altre $n - r$ biglie senza fornire alcuna concordanza. Il numero di tali permutazioni, corrispondente al numero di casi favorevoli all'evento $D_{n,r}$, è pertanto $(n - r)! P(C_{n-r})$. Utilizzando la definizione classica di probabilità, si ha quindi:

$$P(D_{n,r}) = \frac{(n - r)! P(C_{n-r})}{n!} = \frac{(n - r)!}{n!} \sum_{i=0}^{n-r} (-1)^i \frac{1}{i!}.$$

Poiché esistono $\binom{n}{r}$ modi distinti ed equiprobabili di scegliere tra le n biglie r di esse che forniscono concordanze, risulta:

$$P(F_{n,r}) = \binom{n}{r} P(D_{n,r}) = \frac{1}{r!} \sum_{i=0}^{n-r} (-1)^i \frac{1}{i!}. \quad (2.14)$$

Verifichiamo, infine, che sussiste l'uguaglianza $\sum_{r=0}^n P(F_{n,r}) = 1$. Dalla (2.14) si ha

$$\sum_{r=0}^n P(F_{n,r}) = \sum_{r=0}^n \frac{1}{r!} \sum_{i=0}^{n-r} \frac{(-1)^i}{i!} = \sum_{r=0}^n \frac{1}{r!} \sum_{j=r}^n \frac{(-1)^{j-r}}{(j-r)!}, \quad (2.15)$$

dove l'ultima uguaglianza è stata ottenuta ponendo $i = j - r$. Invertendo l'ordine delle somme nell'ultimo termine della (2.15) e facendo uso della formula del binomio di Newton si ha

$$\sum_{r=0}^n P(F_{n,r}) = \sum_{j=0}^n \frac{1}{j!} \sum_{r=0}^j \binom{j}{r} (-1)^{j-r} = \sum_{j=0}^n \frac{1}{j!} (1-1)^j = 1,$$

essendo $(1-1)^j = 1$ per $j = 0$ e $(1-1)^j = 0$ per $j \neq 0$.

Una situazione tipica in cui è possibile applicare tali risultati è la seguente. Un computer prepara addebiti mensili per n utenti e stampa per ciascuno l'indirizzo su di una busta. A causa di un errore, gli addebiti vengono inseriti ciascuno in una busta a caso. L'avere almeno un addebito all'indirizzo giusto corrisponde proprio al problema di avere almeno una concordanza. \diamond

2.4 Indipendenza di eventi

Fondamentale nella teoria della probabilità è il concetto di *indipendenza* di eventi. Prima di darne la definizione formale ed esplicitarne alcune proprietà, è opportuno illustrarne il significato intuitivo.

Si considerino due esperimenti casuali, il primo caratterizzato da spazio campione Ω_1 con $N(\Omega_1)$ eventi elementari equiprobabili ed il secondo caratterizzato da spazio campione Ω_2 con $N(\Omega_2)$ eventi elementari anch'essi equiprobabili. Se A è un evento che può o meno verificarsi nel primo esperimento, allora dalla definizione classica segue che $P(A) = N(A)/N(\Omega_1)$, avendo denotato con $N(A)$ il numero di casi favorevoli ad A . Analogamente, se B denota un evento che può o meno verificarsi nel secondo esperimento, nuovamente dalla definizione classica si trae $P(B) = N(B)/N(\Omega_2)$, dove $N(B)$ è il numero di casi favorevoli a B . Si supponga ora che ognuno dei due esperimenti possa essere effettuato in maniera tale che i suoi risultati non dipendano dai risultati dell'altro. In questa situazione l'occorrenza di A nel primo esperimento non dipende e non influenza l'occorrenza di B nel secondo esperimento. In altre parole A e B sono eventi "indipendenti". Si consideri l'esperimento globale costituito dall'insieme dei due precedenti esperimenti e sia Ω lo spazio campione ad esso associato. Si faccia ora riferimento alle probabilità di A , di B e di $A \cap B$. Il numero di casi possibili è ora $N(\Omega) = N(\Omega_1) N(\Omega_2)$, il numero di casi favorevoli ad A è $N(A) N(\Omega_2)$,

il numero di casi favorevoli a B è $N(\Omega_1) N(B)$ ed il numero di casi favorevoli ad $A \cap B$ è $N(A) N(B)$. Pertanto

$$P(A) = \frac{N(A)}{N(\Omega_1)}, \quad P(B) = \frac{N(B)}{N(\Omega_2)}, \quad P(A \cap B) = \frac{N(A) N(B)}{N(\Omega)} = \frac{N(A)}{N(\Omega_1)} \frac{N(B)}{N(\Omega_2)}.$$

Quindi sussiste la relazione $P(A \cap B) = P(A) P(B)$. In altre parole, la circostanza che i due eventi sono indipendenti comporta che la probabilità della loro intersezione è uguale al prodotto delle probabilità dei singoli eventi.

Anche se l'assenza di legame tra i due esperimenti costituisce il punto di partenza per definire l'indipendenza tra i due eventi, è bene osservare che possono aversi eventi A e B riferiti ad uno stesso esperimento per i quali sussiste la relazione $P(A \cap B) = P(A) P(B)$. Ad esempio, se $A = \{\text{uscita di un numero pari}\}$ e $B = \{\text{uscita di un numero divisibile per 3}\}$ nel lancio di un dado, dalla definizione classica si ottiene $P(A) = 3/6$, $P(B) = 2/6$ e $P(A \cap B) = 1/6$, così che risulta $P(A \cap B) = P(A) P(B)$, anche se l'esperimento considerato non può essere scisso in due distinti esperimenti.

Si giunge così alla seguente definizione di indipendenza.

Definizione 2.9 *Siano A e B eventi di \mathcal{F} . Essi si dicono indipendenti se risulta*

$$P(A \cap B) = P(A) P(B). \quad (2.16)$$

A differenza dell'incompatibilità, che è una proprietà intrinseca degli eventi, l'indipendenza è una proprietà che dipende esclusivamente dalle probabilità degli eventi e non, in generale, dalla struttura di questi. Come emerge dalla Definizione 2.9, l'indipendenza è una relazione simmetrica: se A è indipendente da B , allora B è indipendente da A . Affermare che gli eventi A e B sono indipendenti significa dunque che il verificarsi di A non influenza la probabilità che B si verifichi, e viceversa.

Si osservi che la locuzione "indipendenza" nell'ambito della teoria della probabilità denota una proprietà ben distinta da quella di indipendenza logica o indipendenza causale.

Esempio 2.11 Si consideri l'esperimento consistente nell'estrarre una carta da un mazzo ben mescolato (ossia si assume valida l'equiprobabilità delle 52 carte). Gli eventi $A = \{\text{la carta estratta è una figura}\}$ e $B = \{\text{la carta estratta è di cuori}\}$ sono indipendenti. Infatti, $N(\Omega) = 52$ è il numero dei casi possibili, $N(A) = 12$ il numero dei casi favorevoli ad A , $N(B) = 13$ il numero di casi favorevoli a B e $N(A \cap B) = 3$ il numero di casi favorevoli ad $A \cap B$, così che

$$P(A) = \frac{N(A)}{N(\Omega)} = \frac{3}{13}, \quad P(B) = \frac{N(B)}{N(\Omega)} = \frac{1}{4}, \quad P(A \cap B) = \frac{N(A \cap B)}{N(\Omega)} = \frac{3}{52} = P(A) P(B).$$

Quindi A e B sono eventi indipendenti. \diamond

Proposizione 2.8 *Se A è un evento di \mathcal{F} tale che $P(A) = 0$ oppure $P(A) = 1$ e se B è un qualsiasi altro evento di \mathcal{F} , si ha che A e B sono indipendenti.*

Dimostrazione Se $P(A) = 0$, dalla Proposizione 2.7 segue che $P(A \cap B) = 0$ così che la relazione (2.16) è banalmente verificata. Analogamente, dalla Proposizione 2.7 e dall'essere $P(A) = 1$ segue $P(A \cap B) = P(A)$, restando così nuovamente verificata la (2.16). \square

La Proposizione 2.8 mostra che gli eventi quasi impossibili e gli eventi quasi certi sono indipendenti da qualsiasi altro evento.

Proposizione 2.9 *Se A e B sono eventi indipendenti di \mathcal{F} , allora*

(i) *A e \overline{B} sono indipendenti;*

(ii) *\overline{A} e B sono indipendenti;*

(iii) *\overline{A} e \overline{B} sono indipendenti.*

Dimostrazione *Caso (i)* Poiché $A = (A \cap B) \cup (A \cap \overline{B})$, con $A \cap B$ e $A \cap \overline{B}$ eventi incompatibili, facendo uso della Proposizione 2.3 e dell'ipotesi di indipendenza di A e B si ha:

$$P(A) = P(A \cap B) + P(A \cap \overline{B}) = P(A)P(B) + P(A \cap \overline{B})$$

e quindi:

$$P(A \cap \overline{B}) = P(A) [1 - P(B)] = P(A)P(\overline{B}),$$

ossia la tesi.

Caso (ii) Per dimostrare che \overline{A} e B sono indipendenti, basta semplicemente scambiare i ruoli di A e B nel *Caso (i)*.

Caso (iii) Per dimostrare, infine, che \overline{A} e \overline{B} sono indipendenti è sufficiente applicare il procedimento descritto nel *Caso (i)* agli eventi indipendenti \overline{B} ed A . \square

La Proposizione 2.9 permette di affermare che la relazione di indipendenza non è transitiva: se A è indipendente da B e B lo è da C , questo non comporta che A e C siano indipendenti. Per convincersene basta considerare un evento A non impossibile e diverso da Ω , e porre $C = \overline{A}$. Se si considera un evento B indipendente da A , in virtù della Proposizione 2.9 si nota che B è anche indipendente da C . Nonostante ciò, A e C non sono indipendenti in quanto $P(A \cap C) = 0$ e $P(A)P(C) > 0$.

Proposizione 2.10 *Se A e B sono eventi indipendenti di \mathcal{F} , allora $P(A \cup B) = 1 - P(\overline{A})P(\overline{B})$.*

Dimostrazione Dall'identità $A \cup B = \overline{\overline{A} \cap \overline{B}}$ (legge di De Morgan) e dalla Proposizione 2.4 si ricava $P(A \cup B) = 1 - P(\overline{A} \cap \overline{B})$. La tesi segue immediatamente dalla Proposizione 2.9 osservando che, essendo gli eventi A e B indipendenti, tali sono anche \overline{A} e \overline{B} . \square

Esempio 2.12 Da un'urna contenente 10 biglie rosse e 5 biglie nere si estraggono due biglie con rimpiazzamento. Per calcolare la probabilità dell'evento $C = \{\text{almeno una delle biglie estratte è nera}\}$, è conveniente passare all'evento complementare $\overline{C} = \{\text{nessuna delle due biglie estratte è nera}\}$. L'evento \overline{C} è l'intersezione di due eventi indipendenti $A = \{\text{la prima biglia estratta è rossa}\}$ e $B = \{\text{la seconda biglia estratta è rossa}\}$. Dalla definizione classica di probabilità si ricava $P(A) = P(B) = 2/3$, e quindi dalla Proposizione 2.10 segue $P(C) = 1 - P(\overline{C}) = 1 - (2/3)^2 = 5/9$. \diamond

Il concetto intuitivo di indipendenza precedentemente descritto nel caso di un esperimento che può essere scisso in due distinti esperimenti si può estendere al caso di un esperimento che può essere decomposto in un arbitrario numero finito di esperimenti effettuati in maniera tale che i risultati di ognuno non dipendono dai risultati di ogni altro. Infatti, si considerino n esperimenti casuali E_1, E_2, \dots, E_n rispettivamente caratterizzati da spazi campione $\Omega_1, \Omega_2, \dots, \Omega_n$ e si supponga che il generico Ω_j consista di $N(\Omega_j)$ risultati semplici equiprobabili. Si consideri poi un esperimento $E = \{E_{i_1}, E_{i_2}, \dots, E_{i_k}\}$ ($k = 2, 3, \dots, n$) consistente in k dei differenti esperimenti precedenti e sia Ω il suo spazio campione, costituito quindi da $N(\Omega) = N(\Omega_{i_1}) N(\Omega_{i_2}) \cdots N(\Omega_{i_k})$ risultati semplici. Rivolgeremo qui la nostra attenzione alla probabilità degli eventi $A_{i_1}, A_{i_2}, \dots, A_{i_k}$, dove A_{i_r} è un evento che può o meno verificarsi nell'esperimento i_r -esimo, ed inoltre alla probabilità dell'evento $A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}$. Utilizzando la definizione classica di probabilità si ricava:

$$P(A_{i_1}) = \frac{N(A_{i_1})}{N(\Omega_{i_1})}, \quad P(A_{i_2}) = \frac{N(A_{i_2})}{N(\Omega_{i_2})}, \quad \dots, \quad P(A_{i_k}) = \frac{N(A_{i_k})}{N(\Omega_{i_k})},$$

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \frac{N(A_{i_1}) N(A_{i_2}) \cdots N(A_{i_k})}{N(\Omega_{i_1}) N(\Omega_{i_2}) \cdots N(\Omega_{i_k})}.$$

Si noti che sussiste la relazione $P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \cdots P(A_{i_k})$, così che la probabilità dell'intersezione di eventi relativi a distinti esperimenti, effettuati in maniera tale che i risultati di ognuno non dipendano dai risultati di ogni altro, è uguale al prodotto delle probabilità dei singoli eventi.

In accordo con l'interpretazione intuitiva di indipendenza, si giunge così alla seguente definizione.

Definizione 2.10 *Gli eventi $A_n \in \mathcal{F}$, in numero finito o numerabile, si dicono indipendenti se, comunque scelti k di essi $A_{i_1}, A_{i_2}, \dots, A_{i_k}$ ($k = 2, 3, \dots$), la probabilità dell'intersezione di questi si fattorizza nel prodotto delle rispettive probabilità:*

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \cdots P(A_{i_k}). \quad (2.17)$$

Nel caso di tre eventi A, B, C la condizione di indipendenza è espressa complessivamente da quattro relazioni, ossia dalle seguenti tre relazioni tra coppie di eventi

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

ed inoltre dalla relazione seguente:

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

Esempio 2.13 Si consideri l'esperimento consistente nel lanciare 2 volte una moneta. Lo spazio campione è $\Omega = \{TT, TC, CT, CC\}$ e ciascuno dei suoi eventi elementari ha probabilità $1/4$. Consideriamo i seguenti eventi: $A = \{\text{al primo lancio esce testa}\}$, $B = \{\text{al secondo lancio esce testa}\}$ e $C = \{\text{nei due lanci si ha lo stesso risultato}\}$. Poiché $A = \{TT, TC\}$, $B = \{TT, CT\}$ e $C = \{TT, CC\}$, risulta $P(A) = P(B) = P(C) = 1/2$. Inoltre, poiché $A \cap B = A \cap C = B \cap C = \{TT\}$, si ha $P(A \cap B) = P(A \cap C) = P(B \cap C) = 1/4$. Da ciò si trae che $P(A \cap B) = P(A)P(B)$, $P(A \cap C) = P(A)P(C)$ e $P(B \cap C) =$

$P(B)P(C)$, così che gli eventi A , B e C sono indipendenti a due a due. Tuttavia, poiché $A \cap B \cap C = \{TT\}$, risulta $P(A \cap B \cap C) = 1/4$, mentre $P(A)P(B)P(C) = 1/8$. Essendo $P(A \cap B \cap C) \neq P(A)P(B)P(C)$, i tre eventi A , B e C non sono indipendenti. \diamond

Esempio 2.14 Si prenda in esame l'esperimento che consiste nella scelta a caso di una sequenza costituita da terne di cifre binarie. Lo spazio campione $\Omega = \{000, 001, 010, 011, 100, 101, 110, 111\}$ consiste di 8 eventi elementari che assumeremo equiprobabili. Si considerino i seguenti eventi: $A = \{\text{la prima cifra è 1}\}$, $B = \{\text{al più una cifra è 1}\}$, $C = \{\text{la seconda ed la terza cifra sono uguali}\}$. Risulta quindi $A = \{100, 101, 110, 111\}$, $B = \{000, 001, 010, 100\}$ e $C = \{000, 011, 100, 111\}$, così che $P(A) = P(B) = P(C) = 1/2$. Inoltre, poiché $A \cap B \cap C = \{100\}$, si ha che $P(A \cap B \cap C) = 1/8 = P(A)P(B)P(C)$. Si noti che la sussistenza di tale relazione non garantisce che gli eventi A , B , C sono indipendenti. Infatti, essendo $A \cap B = \{100\}$, si ha $P(A \cap B) = 1/8 \neq P(A)P(B) = 1/4$. Si osservi infine che gli eventi A e C , così come gli eventi B e C , sono indipendenti, risultando $P(A \cap C) = 1/4 = P(A)P(C)$ e $P(B \cap C) = 1/4 = P(B)P(C)$. \diamond

La Proposizione 2.9 può essere generalizzata al caso di più di due eventi come emerge dalla seguente proposizione che ci limitiamo ad enunciare.

Proposizione 2.11 *Se si considera una famiglia di eventi indipendenti di \mathcal{F} e se per ogni evento di qualche sua sottofamiglia si opera la sostituzione con il suo complemento, allora la nuova classe è ancora formata da eventi indipendenti.*

La Proposizione 2.10 può essere estesa al caso di più di due eventi indipendenti.

Proposizione 2.12 *Se A_1, A_2, \dots, A_n sono eventi indipendenti di \mathcal{F} , allora*

$$P\left(\bigcup_{i=1}^n A_i\right) = 1 - \prod_{i=1}^n P(\overline{A_i}).$$

Dimostrazione Dalla legge di De Morgan $\bigcup_{i=1}^n A_i = \overline{\bigcap_{i=1}^n \overline{A_i}}$ e dalla Proposizione 2.4 si trae $P\left(\bigcup_{i=1}^n A_i\right) = 1 - P\left(\bigcap_{i=1}^n \overline{A_i}\right)$. La tesi segue immediatamente dalla Proposizione 2.11 osservando che, essendo gli eventi A_1, A_2, \dots, A_n indipendenti, tali sono anche $\overline{A_1}, \overline{A_2}, \dots, \overline{A_n}$. \square

Esempio 2.15 Il Cavaliere de Méré, noto appassionato di giochi d'azzardo, era erroneamente convinto che fossero uguali le probabilità dei seguenti eventi: $A = \{\text{lanciando quattro dadi si ottiene 6 almeno una volta}\}$ e $B = \{\text{lanciando ventiquattro coppie di dadi si ottiene un 6 doppio almeno una volta}\}$. Determiniamo le probabilità di questi eventi.

Per calcolare la probabilità di A osserviamo che l'evento complementare \overline{A} si può esprimere come intersezione dei quattro eventi indipendenti $C_i = \{\text{nel lancio } i\text{-esimo non si ottiene 6}\}$ ($i = 1, 2, 3, 4$). Pertanto, per la Proposizione 2.4 e per l'indipendenza degli eventi C_i si ha

$$P(A) = 1 - P(\overline{A}) = 1 - P(C_1)P(C_2)P(C_3)P(C_4) = 1 - \left(\frac{5}{6}\right)^4 = 0.5177.$$

Analogamente risulta

$$P(B) = 1 - P(\overline{B}) = 1 - \left(\frac{35}{36}\right)^{24} = 0.4914,$$

che, sebbene dimostri l'incorrettezza della congettura del de Méré, pur tuttavia ne manifesta la notevole intuizione in vista della differenza di solo circa il 2.5% dei valori esatti di dette probabilità. \diamond

Proposizione 2.13 Se A_1, A_2, \dots, A_n sono eventi indipendenti di \mathcal{F} , risulta:

$$1 - \exp\left\{-\sum_{i=1}^n P(A_i)\right\} \leq P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i). \quad (2.18)$$

Dimostrazione L'ultima disuguaglianza segue dalla disuguaglianza di Boole. Dimostriamo ora l'altra disuguaglianza. Facendo uso della Proposizione 2.12 e della nota disuguaglianza¹ $1 - x \leq e^{-x}$ ($x \geq 0$) si ha:

$$P\left(\bigcup_{i=1}^n A_i\right) = 1 - \prod_{i=1}^n [1 - P(A_i)] \geq 1 - \prod_{i=1}^n e^{-P(A_i)} = 1 - \exp\left\{-\sum_{i=1}^n P(A_i)\right\}.$$

□

A partire dalla (2.18), si può dimostrare che la probabilità che non si verifichi nessuno di n eventi indipendenti A_1, A_2, \dots, A_n di \mathcal{F} soddisfa la disuguaglianza

$$1 - \sum_{i=1}^n P(A_i) \leq P(\overline{A_1} \cap \overline{A_2} \cap \dots \cap \overline{A_n}) \leq \exp\left\{-\sum_{i=1}^n P(A_i)\right\}.$$

Va menzionato che molti esperimenti consistono in prove indipendenti, tali cioè che il risultato di ciascuna prova non influenza e non è influenzato dal risultato di ogni altra prova. Tipicamente, prove consistenti nel lanciare una moneta un certo numero di volte sono tra loro indipendenti; estrarre biglie da un'urna significa eseguire prove indipendenti a condizione che la biglia estratta venga reinserita ogni volta nell'urna; estrarre carte da un mazzo significa anche eseguire prove indipendenti a condizione che la carta estratta venga rimessa ogni volta nel mazzo e che il mazzo sia sempre accuratamente mescolato.

Esempio 2.16 Si consideri un'apparecchiatura costituita da tre dispositivi che lavorano indipendentemente l'uno dall'altro. Si supponga che durante un fissato intervallo di tempo l'affidabilità (probabilità di corretto funzionamento) del primo dispositivo sia p_1 , del secondo sia p_2 e del terzo sia p_3 . Si assuma che il corretto funzionamento del primo dispositivo sia indispensabile per il corretto funzionamento dell'apparecchiatura e che il guasto simultaneo del secondo e del terzo dispositivo metta fuori uso l'apparecchiatura. Per determinare l'affidabilità dell'apparecchiatura durante l'intervallo di tempo considerato, si considerino gli eventi

¹Da $e^{-y} \leq 1$ per $y \geq 0$, segue, per integrazione, $0 \leq \int_0^x (1 - e^{-y}) dy = x + e^{-x} - 1$.

$A = \{\text{perfetto funzionamento dell'apparecchiatura}\}$ e $A_i = \{\text{perfetto funzionamento del dispositivo } i\text{-esimo}\}$ ($i = 1, 2, 3$). Si ha $A = (A_1 \cap A_2 \cap A_3) \cup (A_1 \cap \bar{A}_2 \cap A_3) \cup (A_1 \cap A_2 \cap \bar{A}_3)$. Poiché gli eventi $A_1 \cap A_2 \cap A_3$, $A_1 \cap \bar{A}_2 \cap A_3$ e $A_1 \cap A_2 \cap \bar{A}_3$ sono incompatibili, si ha:

$$P(A) = P(A_1 \cap A_2 \cap A_3) + P(A_1 \cap \bar{A}_2 \cap A_3) + P(A_1 \cap A_2 \cap \bar{A}_3).$$

Poiché per ipotesi i tre dispositivi lavorano tra loro indipendentemente, risulta:

$$P(A) = p_1 p_2 p_3 + p_1 (1 - p_2) p_3 + p_1 p_2 (1 - p_3) = p_1 (p_2 + p_3 - p_2 p_3).$$

◇

Questo esempio dà un'idea di alcuni problemi che si affrontano nella teoria dell'affidabilità allorché si desidera calcolare le probabilità di funzionamento di apparecchiature complesse.

Esempio 2.17 Due giocatori G_1 e G_2 competono in un torneo. Si supponga che la probabilità che G_1 vinca una partita sia $1/2$, la probabilità che G_2 vinca una partita sia $1/3$ e la probabilità che una partita termini in pareggio sia $1/6$. Se G_1 e G_2 partecipano ad un torneo di tre partite riguardate come indipendenti, si intende calcolare la probabilità dei seguenti eventi: $C = \{G_1 \text{ vince tutte e tre le partite}\}$, $D = \{\text{due delle tre partite terminano in pareggio}\}$, $F = \{G_1 \text{ e } G_2 \text{ vincono alternativamente le tre partite}\}$ e $H = \{G_2 \text{ vince almeno una delle tre partite}\}$.

Si considerino i seguenti eventi: $A_i = \{G_1 \text{ vince la partita } i\text{-esima}\}$ ($i = 1, 2, 3$), $B_i = \{G_2 \text{ vince la partita } i\text{-esima}\}$ ($i = 1, 2, 3$), $E_i = \{\text{la partita } i\text{-esima termina in pareggio}\}$ ($i = 1, 2, 3$). Si noti che è possibile esprimere gli eventi di interesse C , D , F e H in termini di A_i , B_i e E_i ($i = 1, 2, 3$). Infatti, essendo $C = A_1 \cap A_2 \cap A_3$, per l'ipotesi di indipendenza tra le partite si ha $P(C) = P(A_1) P(A_2) P(A_3) = 1/8$. Inoltre, poiché $D = (E_1 \cap E_2 \cap \bar{E}_3) \cup (E_1 \cap \bar{E}_2 \cap E_3) \cup (\bar{E}_1 \cap E_2 \cap E_3)$ e $F = (A_1 \cap B_2 \cap A_3) \cup (B_1 \cap A_2 \cap B_3)$, per la proprietà di additività finita della probabilità e per l'ipotesi di indipendenza tra le partite risulta

$$P(D) = P(E_1) P(E_2) P(\bar{E}_3) + P(E_1) P(\bar{E}_2) P(E_3) + P(\bar{E}_1) P(E_2) P(E_3) = \frac{5}{72},$$

$$P(F) = P(A_1) P(B_2) P(A_3) + P(B_1) P(A_2) P(B_3) = \frac{1}{4}.$$

Infine, essendo $\bar{H} = \bar{B}_1 \cap \bar{B}_2 \cap \bar{B}_3$, si ha:

$$P(H) = 1 - P(\bar{H}) = 1 - P(\bar{B}_1) P(\bar{B}_2) P(\bar{B}_3) = 1 - \left(\frac{2}{3}\right)^3 = \frac{19}{27}.$$

◇

Esempio 2.18 In un tiro al bersaglio due concorrenti sparano a turno un colpo finché uno dei due non fa centro. Quando ciò accade la competizione termina ed il concorrente che ha fatto centro vince. Si supponga che i risultati dei tiri siano indipendenti e che la probabilità di fare centro sia costante in ogni tiro per ciascun concorrente; questa sia p per colui che

inizia a sparare e q per l'altro. Per calcolare la probabilità di vittoria di ciascun concorrente, si considerino i seguenti eventi: $A = \{\text{il concorrente che inizia a sparare vince}\}$, $B = \{\text{il concorrente che spara per secondo vince}\}$, $A_{2i+1} = \{\text{il concorrente che inizia a sparare vince al tiro } 2i+1\}$ ($i = 0, 1, \dots$), $B_{2i} = \{\text{il concorrente che spara per secondo vince al tiro } 2i\}$ ($i = 1, 2, \dots$). Per l'ipotesi di indipendenza dei tiri si ha:

$$P(A_{2i+1}) = [(1-p)(1-q)]^i p \quad (i = 0, 1, \dots)$$

$$P(B_{2i}) = (1-p) [(1-p)(1-q)]^{i-1} q \quad (i = 1, 2, \dots).$$

Gli eventi A_1, A_3, \dots e gli eventi B_2, B_4, \dots sono tutti tra loro incompatibili. Pertanto, poiché $A = \bigcup_{i=0}^{+\infty} A_{2i+1}$ e $B = \bigcup_{i=1}^{+\infty} B_{2i}$, applicando il terzo assioma della probabilità si ottiene:

$$P(A) = P\left(\bigcup_{i=0}^{+\infty} A_{2i+1}\right) = \sum_{i=0}^{+\infty} P(A_{2i+1}) = \frac{p}{p+q-pq},$$

$$P(B) = P\left(\bigcup_{i=1}^{+\infty} B_{2i}\right) = \sum_{i=1}^{+\infty} P(B_{2i}) = \frac{(1-p)q}{p+q-pq}.$$

Si osservi che (supponendo che la competizione possa protrarsi indefinitamente) la probabilità che la gara termini è unitaria essendo $P(A) + P(B) = 1$. Quindi, quasi certamente vi sarà un vincitore. \diamond

2.5 Probabilità condizionata

Può talora accadere che si acquisisca una conoscenza parziale sullo svolgimento dell'esperimento casuale prima che il risultato finale diventi noto. Le nuove informazioni permettono di escludere alcuni risultati dell'esperimento casuale che erano inizialmente possibili. In tale circostanza si pone il problema di definire una nuova misura di probabilità che chiameremo *probabilità condizionata*. Con la notazione $P(A|B)$ si indica la probabilità dell'evento A condizionata dal verificarsi di B , ossia la probabilità che ha l'evento A di verificarsi quando si sappia che l'evento B si è verificato.

Il concetto di probabilità condizionata riveste un interessante contenuto intuitivo che emerge nelle tre definizioni classica, frequentista e soggettiva di probabilità.

Nella definizione classica si definisce probabilità condizionata $P(A|B)$ il rapporto tra il numero di casi favorevoli all'evento $A \cap B$ ed il numero dei casi favorevoli all'evento B . Quindi, se si denota con $N(\Omega)$ il numero dei casi possibili (supposti equiprobabili), con $N(A \cap B)$ il numero di casi favorevoli all'evento $A \cap B$ e con $N(B)$ il numero di casi favorevoli all'evento B , se è $N(B) > 0$ si ha:

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

Nella definizione frequentista la probabilità condizionata $P(A|B)$ è calcolata restringendo l'attenzione solo alle prove in cui l'evento B si verifica ed è definita come limite della

frequenza relativa (calcolata solo sulle prove in cui B si verifica) delle prove in cui l'evento $A \cap B$ si verifica quando il numero delle prove tende all'infinito. Quindi, secondo la definizione frequentista dalla successione di prove nelle quali si osserva il verificarsi dell'evento A occorre eliminare le prove in cui non si verifica B . Nel caso in cui si effettuano n prove, denotando con $\nu_n(A \cap B)$ il numero di prove in cui si verifica $A \cap B$ e con $\nu_n(B)$ il numero delle prove in cui si verifica B , se risulta $\nu_n(B) > 0$ si ha:

$$f_n(A|B) = \frac{\nu_n(A \cap B)}{\nu_n(B)} = \frac{\nu_n(A \cap B)}{n} \frac{n}{\nu_n(B)} = \frac{f_n(A \cap B)}{f_n(B)}.$$

Per la definizione frequentista di probabilità, al crescere del numero delle prove $f_n(A \cap B)$ e $f_n(B)$ costituiscono rispettivamente misure di $P(A \cap B)$ e $P(B)$; ne segue che $f_n(A|B)$ misura il rapporto $P(A \cap B)/P(B)$, che denotiamo con $P(A|B)$.

Nella definizione soggettiva la probabilità $P(A|B)$ è una misura del grado di fiducia che un individuo coerente ripone nel verificarsi di A , calcolata nell'ipotesi che B si verifichi. Quindi, la probabilità $P(A|B)$ è il prezzo che un individuo ritiene equo pagare per ricevere 1 se l'evento A si verifica e 0 se l'evento A non si verifica, prezzo valutato nell'ipotesi che B si verifichi.

Esempio 2.19 Calcoliamo la probabilità che il lancio di un dado dia come risultato 2 sapendo che il risultato del lancio è pari.

Consideriamo gli eventi $A = \{\text{il risultato del lancio è } 2\}$ e $B = \{\text{il risultato del lancio è pari}\}$. In questo caso $N(A \cap B) = 1$ e $N(B) = 3$ di modo che, usando la definizione classica, si ha $P(A|B) = 1/3$. Invece, secondo la definizione frequentista, nella successione di lanci del dado consideriamo soltanto quelli che danno un risultato pari; la frequenza relativa del risultato 2, condizionatamente all'ipotesi che il risultato sia pari, ci si attende sia "asintoticamente" prossima a $1/3$ al crescere del numero di lanci. \diamond

In generale, si dà la seguente definizione di probabilità condizionata.

Definizione 2.11 Siano A e B eventi di \mathcal{F} con $P(B) > 0$. La probabilità di A condizionata dal verificarsi di B , denotata con $P(A|B)$ (si legge "probabilità di A dato B "), è così definita:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (2.19)$$

Esempio 2.20 Nell'esperimento consistente nel lancio di un dado ripetuto due volte, si calcoli la probabilità che la somma dei risultati sia minore di 7 sapendo che il primo dado ha fornito come risultato j ($j = 1, 2, \dots, 6$).

Si considerino i seguenti eventi: $A = \{\text{la somma dei risultati è minore di } 7\}$ e $B_j = \{\text{il risultato del primo lancio è } j\}$ ($j = 1, 2, \dots, 6$). Dalla definizione classica di probabilità si ha (v. Tabella 1.1) $P(A \cap B_j) = (6 - j)/36$ ($j = 1, 2, \dots, 6$). Essendo $P(B_j) = 1/6$, dalla definizione di probabilità condizionata segue

$$P(A|B_j) = \frac{P(A \cap B_j)}{P(B_j)} = 1 - \frac{j}{6} \quad (j = 1, 2, \dots, 6).$$

Si noti che, sebbene gli eventi B_1, B_2, \dots, B_6 siano equiprobabili, la probabilità richiesta $P(A|B_j)$ varia con j . \diamond

Dimostriamo ora che la (2.19) è effettivamente una probabilità, ossia che essa soddisfa i tre assiomi caratterizzanti la probabilità.

Teorema 2.5 *Se $B \in \mathcal{F}$ e $P(B) > 0$, allora $P(\cdot | B)$ è una probabilità su \mathcal{F} nel senso che*

(i) $P(A|B) \geq 0$ per ogni $A \in \mathcal{F}$,

(ii) $P(\Omega|B) = 1$,

(iii) *Se $\{A_n; n = 1, 2, \dots\}$ è una successione di eventi incompatibili di \mathcal{F} allora*

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

Dimostrazione La Definizione 2.11 implica evidentemente che la probabilità condizionata soddisfa la (i). Si ha inoltre $P(\Omega|B) = P(\Omega \cap B)/P(B) = 1$, ossia la (ii). Infine, essendo per ipotesi $\{A_n; n = 1, 2, \dots\}$ una successione di eventi incompatibili di \mathcal{F} , usando nuovamente la Definizione 2.11 si ha:

$$P\left(\bigcup_{n=1}^{+\infty} A_n \middle| B\right) = \frac{P\left[\left(\bigcup_{n=1}^{+\infty} A_n\right) \cap B\right]}{P(B)} = \frac{P\left[\bigcup_{n=1}^{+\infty} (A_n \cap B)\right]}{P(B)}.$$

Essendo $\{A_n \cap B; n = 1, 2, \dots\}$ una successione di eventi incompatibili, segue infine:

$$P\left(\bigcup_{n=1}^{+\infty} A_n \middle| B\right) = \sum_{n=1}^{+\infty} \frac{P(A_n \cap B)}{P(B)} = \sum_{n=1}^{+\infty} P(A_n|B),$$

ossia la (iii). □

Esempio 2.21 Si consideri l'esperimento consistente in 4 prove indipendenti, in ognuna delle quali si inserisce una biglia in un'urna scelta a caso tra quattro urne. Si vuole calcolare la probabilità di trovare tre biglie nella medesima urna sapendo che le prime due biglie sono state inserite in urne diverse.

Si noti che la cardinalità dello spazio campione è $N(\Omega) = 4^4$, poiché ogni biglia può essere inserita in una qualsiasi delle 4 urne. Si considerino i seguenti eventi: $A = \{\text{tre biglie finiscono nella stessa urna}\}$, $B = \{\text{le prime due biglie sono inserite in urne diverse}\}$. Il numero dei casi favorevoli all'evento B è $N(B) = 4 \cdot 3 \cdot 4^2 = 192$; infatti la prima biglia può essere inserita in una qualsiasi delle quattro urne, la seconda in una qualsiasi delle tre rimanenti ed ognuna delle ultime due biglie può essere inserita in una qualsiasi delle quattro urne. Pertanto, per la definizione classica di probabilità, si ha $P(B) = N(B)/N(\Omega) = 3/4$. Il numero dei casi favorevoli all'evento $A \cap B$ è invece $N(A \cap B) = (4 \cdot 3) \cdot 2 = 24$ poiché la prima biglia può essere inserita in una qualsiasi delle quattro urne, la seconda in una qualsiasi delle rimanenti tre urne mentre le ultime due devono essere inserite nell'urna dove era stata inserita la prima biglia oppure nell'urna dove era stata inserita la seconda biglia. Utilizzando nuovamente la definizione classica di probabilità, si ha $P(A \cap B) = N(A \cap B)/N(\Omega) = 3/32$. Infine, dalla (2.19) si ricava $P(A|B) = N(A \cap B)/N(B) = 1/8$. ◇

Esempio 2.22 Potendo effettuare due sole telefonate ad un abbonato una cifra del cui numero è ignota, si calcoli la probabilità di comporre il numero giusto scegliendo a caso l'incognita cifra.

Si considerino all'uopo i seguenti eventi: $A_i = \{\text{si compone il numero giusto alla } i\text{-esima telefonata}\}$ ($i = 1, 2$), $A = \{\text{si compone il numero giusto in una delle due telefonate}\}$. Poiché risulta $A = A_1 \cup (\overline{A_1} \cap A_2)$, si ha:

$$P(A) = P(A_1) + P(\overline{A_1} \cap A_2) = P(A_1) + P(\overline{A_1}) P(A_2|\overline{A_1}) = \frac{1}{10} + \frac{9}{10} \cdot \frac{1}{9} = \frac{1}{5}.$$

Vale la pena osservare che il quesito proposto in questo esempio può, più semplicemente, ma equivalentemente, formularsi al seguente modo: calcolare la probabilità che scegliendo a caso senza rimpiazzamento due tra dieci oggetti, tra i due oggetti scelti se ne ritrovi uno che è stato inizialmente fissato. Tale probabilità p è evidentemente la somma delle probabilità che il primo oggetto scelto sia quello giusto ($1/10$) e della probabilità congiunta che non essendo giusto il primo ($1 - 1/10$) risulti giusto il secondo tra i nove oggetti rimanenti ($1/9$): $p = (1/10) + (1 - 1/10) \cdot (1/9) = 1/5$. \diamond

Va menzionato che da parte di vari studiosi è stata proposta un'assiomatizzazione della teoria della probabilità che parte dal concetto di probabilità condizionata. Una delle ragioni è di superare l'impossibilità di definire la probabilità condizionata se l'evento condizionante ha probabilità nulla, venendo così incontro a talune esigenze intuitive. Ad esempio, si consideri l'esperimento consistente nella scelta a caso di un numero reale nell'intervallo $(0, 1)$ e siano $A_1 = \{\text{il numero scelto è } x_0\}$ e $A_2 = \{\text{il numero scelto è } y_0\}$, con $x_0, y_0 \in (0, 1)$, $x_0 \neq y_0$. In tali condizioni è abbastanza intuitivo richiedere che risulti $P(A_1|A_1 \cup A_2) = 1/2$. Si nota, invece, che la Definizione 2.11 non permette di calcolare tale probabilità condizionata dal momento che $P(A_1 \cup A_2) = 0$.

Osserviamo ora che se è $P(A) > 0$, scambiando i ruoli di A e B nella (2.19) si definisce la probabilità condizionata di B dato A tramite la posizione:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}. \quad (2.20)$$

Dalle (2.19) e (2.20) segue immediatamente che se A e B sono eventi di \mathcal{F} con $P(A) > 0$ e $P(B) > 0$, allora

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

Si può dare una caratterizzazione dell'indipendenza di due eventi tramite tre proprietà equivalenti, come qui di seguito indicato.

Proposizione 2.14 Se A e B sono eventi di \mathcal{F} tali che $P(A) > 0$ e $P(B) > 0$, allora le seguenti proprietà sono equivalenti:

- (i) $P(A \cap B) = P(A) P(B)$;
- (ii) $P(A|B) = P(A)$;
- (iii) $P(B|A) = P(B)$.

Dimostrazione Per dimostrare questa proposizione basta mostrare che sussistono le seguenti implicazioni: $(i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (i)$.

Dimostriamo che $(i) \Rightarrow (ii)$. Si ha infatti:

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

Dimostriamo che $(ii) \Rightarrow (iii)$. Sempre tenendo conto della Definizione 2.11 si ha $P(A \cap B) = P(A|B)P(B)$, e quindi per la (2.20) e per la (ii) segue:

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

Dimostriamo, infine, che $(iii) \Rightarrow (i)$. A tale scopo notiamo che ancora dalla (2.20) segue:

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

Ciò completa la dimostrazione. \square

La definizione di probabilità condizionata è suscettibile di immediata estensione. Infatti, se A_1, A_2, \dots, A_n sono eventi di \mathcal{F} tali che $P(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$, si pone:

$$P(A_n | A_1 \cap A_2 \cap \dots \cap A_{n-1}) = \frac{P(A_1 \cap A_2 \cap \dots \cap A_n)}{P(A_1 \cap A_2 \cap \dots \cap A_{n-1})}. \quad (2.21)$$

Il seguente teorema, detto *regola moltiplicativa* o *legge delle probabilità composte*, permette di calcolare la probabilità dell'intersezione di n assegnati eventi facendo uso delle probabilità condizionate.

Teorema 2.6 Siano A_1, A_2, \dots, A_n una collezione di eventi di \mathcal{F} tali che $P(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$. Si ha:

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) P(A_2 | A_1) \dots P(A_n | A_1 \cap A_2 \cap \dots \cap A_{n-1}). \quad (2.22)$$

Dimostrazione Poiché risulta

$$A_1 \cap A_2 \cap \dots \cap A_{n-1} \subset A_1 \cap A_2 \cap \dots \cap A_{n-2} \subset \dots \subset A_1 \cap A_2 \subset A_1,$$

segue:

$$0 < P(A_1 \cap A_2 \cap \dots \cap A_{n-1}) \leq P(A_1 \cap A_2 \cap \dots \cap A_{n-2}) \leq \dots \leq P(A_1 \cap A_2) \leq P(A_1), \quad (2.23)$$

avendo fatto uso della positività ipotizzata per la probabilità dell'intersezione degli $n - 1$ eventi. Essendo tali probabilità non nulle, dalla (2.21) e dalla (2.23) si ricava poi:

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1 \cap A_2 \cap \dots \cap A_{n-1}) P(A_n | A_1 \cap A_2 \cap \dots \cap A_{n-1}). \quad (2.24)$$

Continuando iterativamente ad applicare la (2.21) al secondo membro della (2.24), si giunge finalmente alla (2.22). \square

Esempio 2.23 Tre carte sono estratte senza rimpiazzamento da un mazzo ben mescolato di 52 carte. Si calcoli la probabilità di non ottenere nessuna carta di cuori.

Si considerino i seguenti eventi: $A = \{\text{nessuna delle tre carte estratte è di cuori}\}$, $A_i = \{\text{la carta } i\text{-esima estratta non è di cuori}\}$ ($i = 1, 2, 3$). Poiché $A = A_1 \cap A_2 \cap A_3$, usando il Teorema 2.6 si ricava $P(A) = P(A_1) P(A_2|A_1) P(A_3|A_1 \cap A_2)$. Essendo $P(A_1) = 39/52$, $P(A_2|A_1) = 38/51$ e $P(A_3|A_1 \cap A_2) = 37/50$, segue immediatamente che $P(A) = (39 \cdot 38 \cdot 37)/(52 \cdot 51 \cdot 50) = 0.4135$. \diamond

Negli esempi che seguono è mostrato il calcolo di probabilità condizionate mediante l'uso del calcolo combinatorio.

Esempio 2.24 Un'urna contiene quattro biglie contrassegnate con le lettere a, b, c, d . Si estraggono due biglie, una dopo l'altra, e con reinserimento. Si è interessati a calcolare la probabilità che nessuna delle due biglie sia contrassegnata con b sapendo che la stessa biglia non risulta selezionata in entrambe le estrazioni.

Lo spazio campione Ω è costituito da $4^2 = 16$ eventi elementari, cioè da tutte le disposizioni con ripetizione di 4 oggetti su due posti. Si considerino i seguenti eventi: $A = \{\text{nessuna delle due biglie è contrassegnata con } b\}$, $B = \{\text{la stessa biglia non è selezionata in entrambe le estrazioni}\}$. Il numero dei casi possibili è $N(\Omega) = 16$, il numero dei casi favorevoli all'evento A è $N(A) = 9$, il numero di casi favorevoli all'evento B è $N(B) = 12$ ed il numero di casi favorevoli all'evento $A \cap B$ è $N(A \cap B) = 6$. Pertanto risulta $P(A) = 9/16$, $P(B) = 3/4$ e $P(A \cap B) = 3/8$. Dalla (2.19) si ottiene quindi $P(A|B) = P(A \cap B)/P(B) = 1/2$. \diamond

Esempio 2.25 Ci proponiamo di calcolare la probabilità che una mano di poker sia costituita solo da carte di cuori sapendo che essa consiste unicamente di carte rosse (cuori e quadri).

La mano di poker consiste di cinque carte scelte a caso da un mazzo di 52 carte. Si considerino gli eventi: $A = \{\text{una mano di poker è formata solo da carte di cuori}\}$, $B = \{\text{una mano di poker consiste solo di carte rosse}\}$. Il numero dei casi possibili è $N(\Omega) = \binom{52}{5}$, il numero dei casi favorevoli all'evento A è $N(A) = \binom{13}{5}$ ed il numero di casi favorevoli all'evento B è $N(B) = \binom{26}{5}$. Dalla definizione classica si ha dunque $P(A) = \binom{13}{5} / \binom{52}{5}$, $P(B) = \binom{26}{5} / \binom{52}{5}$. Poiché $A \subset B$ segue $A = A \cap B$ e quindi $P(A \cap B) = P(A) = \binom{13}{5} / \binom{52}{5}$. Pertanto, dalla (2.19), si ottiene $P(A|B) = P(A \cap B)/P(B) = \binom{13}{5} / \binom{26}{5}$. Si nota che $P(A|B) = 0.0196$ è di circa dieci volte maggiore di $P(A) = 0.0021$, cioè della probabilità che una mano di poker consista solo di carte di cuori. \diamond

Esempio 2.26 Un'urna contiene 10 biglie contrassegnate con le lettere b_1, b_2, \dots, b_{10} . Cinque biglie sono estratte senza reinserimento. Si è interessati a calcolare la probabilità che tra le biglie estratte siano incluse quelle contrassegnate con b_i e b_j ($i, j = 1, 2, \dots, 10, i \neq j$) ed inoltre la probabilità che tra le biglie estratte sia inclusa quella contrassegnata con b_i dato che è stata già inclusa la biglia contrassegnata con b_j ($i, j = 1, 2, \dots, 10, i \neq j$).

Lo spazio campione Ω consiste di tutte le possibili combinazioni delle 10 biglie a gruppi di 5; quindi $N(\Omega) = \binom{10}{5}$. Consideriamo i seguenti eventi: $A_i = \{\text{tra le biglie estratte compare quella contrassegnata con } b_i\}$ ($i = 1, 2, \dots, 10$). Il numero di casi favorevoli ad A_i è $N(A_i) = \binom{9}{4}$ poiché dopo che la biglia contrassegnata con b_i è stata scelta debbono essere

scelte altre 4 biglie tra le rimanenti 9. Ne segue che $P(A_i) = \binom{9}{4} / \binom{10}{5}$ ($i = 1, 2, \dots, 10$). Inoltre, il numero di casi favorevoli all'evento $A_i \cap A_j$ è $N(A_i \cap A_j) = \binom{8}{3}$ poiché dopo aver scelto le biglie contrassegnate con b_i e b_j debbono essere scelte altre 3 biglie dalle rimanenti 8. Quindi, si ha $P(A_i \cap A_j) = \binom{8}{3} / \binom{10}{5}$ ($i, j = 1, 2, \dots, 10, i \neq j$). Essendo $P(A_i \cap A_j) \neq P(A_i) P(A_j)$ ($i \neq j$), gli eventi A_i e A_j non sono indipendenti. Infine, dalla Definizione 2.11, si ottiene $P(A_i|A_j) = P(A_i \cap A_j)/P(A_j) = \binom{8}{3} / \binom{9}{4} = 4/9$. \diamond

Esempio 2.27 Si consideri l'esperimento consistente nel lanciare tre volte un dado. Ci si propone di calcolare la probabilità che uno ed uno solo dei lanci dia come risultato 1 sapendo che nessuno dei lanci fornisce lo stesso risultato.

La cardinalità dell'insieme Ω è $N(\Omega) = 6^3$, pari al numero delle possibili disposizioni con ripetizione di 6 oggetti su tre posti. Consideriamo i seguenti eventi: $A = \{\text{uno ed uno solo dei tre dadi fornisce come risultato 1}\}$, $B = \{\text{nessuno dei tre dadi fornisce lo stesso risultato}\}$. Il numero di casi favorevoli all'evento B è $N(B) = 6 \cdot 5 \cdot 4 = 120$ e il numero di casi favorevoli all'evento $A \cap B$ è $N(A \cap B) = 3(5 \cdot 4) = 60$. Pertanto si ha $P(A \cap B) = 5/18$ e $P(A|B) = N(A \cap B)/N(B) = 1/2$. \diamond

Esempio 2.28 Due carte sono estratte a caso da un mazzo ben mescolato di 52 carte. Si intende calcolare la probabilità che vengano scelti il Re di Picche ed il Re di Cuori distinguendo il caso in cui l'estrazione è effettuata con rimpiazzamento dal caso in cui l'estrazione è effettuata senza rimpiazzamento.

Consideriamo i seguenti eventi: $A_i = \{\text{la carta } i\text{-esima estratta è il Re di Picche}\}$ ($i = 1, 2$), $B_i = \{\text{la carta } i\text{-esima estratta è il Re di Cuori}\}$ ($i = 1, 2$), $C = \{\text{le carte scelte nelle due estrazioni sono il Re di Picche ed il Re di Cuori}\}$. L'evento C si può rappresentare come unione di due eventi incompatibili: $C = (A_1 \cap B_2) \cup (A_2 \cap B_1)$; segue quindi $P(C) = P(A_1 \cap B_2) + P(A_2 \cap B_1)$. Se l'estrazione è effettuata con rimpiazzamento, per l'indipendenza si ha $P(C) = P(A_1)P(B_2) + P(B_1)P(A_2) = 2 \cdot (1/52)^2$. Invece, se l'estrazione è effettuata senza rimpiazzamento, risulta $P(C) = P(A_1)P(B_2|A_1) + P(B_1)P(A_2|B_1) = 2 \cdot (1/52) \cdot (1/51)$. \diamond

2.6 Legge delle alternative

Oggetto di questo paragrafo sono alcuni importanti risultati coinvolgenti le probabilità condizionate, molto utili dal punto di vista applicativo. In primo luogo osserviamo che se A e B sono elementi di \mathcal{F} , la relazione $A = (A \cap B) \cup (A \cap \overline{B})$ conduce ad affermare che $P(A) = P(A \cap B) + P(A \cap \overline{B})$ e, se si suppone $0 < P(B) < 1$, la regola moltiplicativa comporta $P(A) = P(B)P(A|B) + P(\overline{B})P(A|\overline{B})$. Gli eventi B e \overline{B} , che costituiscono una partizione di Ω , possono essere riguardati come due alternative poiché, per ogni fissato $\omega \in \Omega$ si ha che $\omega \in B$ oppure che $\omega \in \overline{B}$. Quanto detto può essere generalizzato introducendo la nozione di insieme di alternative.

Definizione 2.12 Sia (Ω, \mathcal{F}, P) uno spazio di probabilità e sia B un evento di \mathcal{F} . Una famiglia $\{B_n; n = 1, 2, \dots, K\}$ di eventi di \mathcal{F} costituisce un insieme di alternative per l'evento B se sono soddisfatte le seguenti proprietà:

(i) $B_i \cap B_j = \emptyset$ ($i, j = 1, 2, \dots, K$; $i \neq j$);

(ii) $\bigcup_{n=1}^K B_n = B$;

(iii) $P(B_n) > 0$ per $n = 1, 2, \dots, K$.

Inoltre, se $B = \Omega$, le condizioni (i), (ii) e (iii) si dicono individuare un “insieme completo” di alternative.

In particolare la (i) afferma che gli eventi B_1, B_2, \dots, B_K sono incompatibili, e quindi alternativi, mentre la (ii) esprime la circostanza che tali eventi sono anche necessari per B . Sussiste il seguente

Teorema 2.7 (Legge delle alternative) Sia $\{B_n; n = 1, 2, \dots, K\}$ un insieme completo di alternative e sia A un evento di \mathcal{F} . Risulta:

$$P(A) = \sum_{n=1}^K P(A|B_n) P(B_n). \quad (2.25)$$

Dimostrazione Poiché per ipotesi gli eventi dell'insieme $\{B_n; n = 1, 2, \dots, K\}$ sono a due a due incompatibili, tali sono anche gli eventi dell'insieme $\{A \cap B_n; n = 1, 2, \dots, K\}$. Inoltre, poiché $\bigcup_{n=1}^K B_n = \Omega$, si ha:

$$P(A) = P(A \cap \Omega) = P\left(A \cap \bigcup_{n=1}^K B_n\right) = P\left(\bigcup_{n=1}^K (A \cap B_n)\right) = \sum_{n=1}^K P(A \cap B_n),$$

da cui, facendo uso della regola moltiplicativa di cui al Teorema 2.6, segue la (2.25). \square

La legge delle alternative permette di calcolare la probabilità a priori di A conoscendo le probabilità $P(B_n)$ delle alternative e le probabilità condizionate $P(A|B_n)$.

Esempio 2.29 Data un'urna contenente biglie numerate con gli interi da 1 a 9 si estragga con equiprobabilità una biglia e successivamente, senza reinserire la prima biglia nell'urna, se ne estragga una seconda. Si calcoli la probabilità che la seconda biglia estratta abbia numero pari.

Si considerino gli eventi $A_i = \{\text{la } i\text{-esima biglia estratta ha numero pari}\}$ ($i = 1, 2$). Dalla legge delle alternative segue:

$$P(A_2) = P(A_1) P(A_2|A_1) + P(\overline{A_1}) P(A_2|\overline{A_1}) = \frac{4}{9} \cdot \frac{3}{8} + \frac{5}{9} \cdot \frac{4}{8} = \frac{4}{9}.$$

\diamond

Esempio 2.30 Supponiamo che quotidianamente vengano esaminate le condizioni meteorologiche in una prefissata regione. Si consideri l'evento $A_j = \{\text{il } j\text{-esimo giorno in esame non è piovoso}\}$ ($j = 1, 2, \dots$) e si supponga che $P(A_{j+1}|A_j) = \beta$ e $P(A_{j+1}|\overline{A_j}) = \alpha$ con α e β numeri arbitrari interni all'intervallo $(0, 1)$. Si intende calcolare $P(A_j)$ ($j = 1, 2, \dots$).

Dalla legge delle alternative segue:

$$P(A_{j+1}) = P(A_j) P(A_{j+1}|A_j) + P(\overline{A_j}) P(A_{j+1}|\overline{A_j}) = P(A_j) \beta + [1 - P(A_j)] \alpha,$$

da cui si trae:

$$P(A_{j+1}) = (\beta - \alpha) P(A_j) + \alpha \quad (j = 1, 2, \dots).$$

Abbiamo ottenuto un'equazione ricorsiva, la cui soluzione, ricavata in modo iterativo, è la seguente:

$$\begin{aligned} P(A_j) &= (\beta - \alpha)^{j-1} P(A_1) + \alpha \sum_{i=0}^{j-2} (\beta - \alpha)^i \\ &= (\beta - \alpha)^{j-1} P(A_1) + \alpha \frac{1 - (\beta - \alpha)^{j-1}}{1 - (\beta - \alpha)} \quad (j = 1, 2, \dots). \end{aligned}$$

Si noti che il limite

$$\lim_{j \rightarrow +\infty} P(A_j) = \frac{\alpha}{1 - \beta + \alpha}$$

non dipende da $P(A_1)$; ciò indica che con il trascorrere dei giorni la probabilità di non avere pioggia tende a non dipendere dalla probabilità con cui ha piovuto il primo giorno. \diamond

Scelto un insieme di alternative per un sottoinsieme proprio B di Ω , sussiste la cosiddetta *legge delle alternative condizionate*, espressa dal seguente teorema:

Teorema 2.8 *Sia $\{B_n; n = 1, 2, \dots, K\}$ un insieme di alternative per l'evento $B \in \mathcal{F}$ e sia A un evento di \mathcal{F} . La probabilità di A condizionata da B è esprimibile al seguente modo:*

$$P(A|B) = \sum_{n=1}^K P(B_n|B) P(A|B_n). \quad (2.26)$$

Dimostrazione Poiché per ipotesi gli eventi dell'insieme $\{B_n; n = 1, 2, \dots, K\}$ sono a due a due incompatibili e tali che $P(B_n) > 0$ per $n = 1, 2, \dots, K$, usando la proprietà di additività delle probabilità si ha $P(B) = P\left(\bigcup_{n=1}^K B_n\right) = \sum_{n=1}^K P(B_n) > 0$. Si noti che gli eventi dell'insieme $\{A \cap B_n; n = 1, 2, \dots, K\}$ sono incompatibili; poiché per ipotesi $\bigcup_{n=1}^K B_n = B$, si ricava:

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} = \frac{P\left[A \cap \left(\bigcup_{n=1}^K B_n\right)\right]}{P(B)} = \frac{P\left[\bigcup_{n=1}^K (A \cap B_n)\right]}{P(B)} \\ &= \sum_{n=1}^K \frac{P(A \cap B_n)}{P(B)} = \sum_{n=1}^K \frac{P(B_n)}{P(B)} P(A|B_n). \end{aligned}$$

Osservando che per ogni n si ha $B_n = B_n \cap B$, risulta infine:

$$P(A|B) = \sum_{n=1}^K \frac{P(B_n \cap B)}{P(B)} P(A|B_n) = \sum_{n=1}^K P(B_n|B) P(A|B_n).$$

□

Si noti che se nel Teorema 2.8 si pone $B = \Omega$, la (2.26) si identifica con la (2.25).

2.7 Teorema di Bayes

Introdurremo ora un teorema di particolare rilievo concernente le probabilità condizionate; esso riveste grande importanza sia concettuale che applicativa, ma la sua utilizzazione, se non correttamente effettuata, può dar luogo — ed ha invero spesso generato — equivoci e risultati paradossali. Si tratta del celebre *teorema o legge di Bayes*, che trovò la sua prima chiara formulazione ad opera appunto di Thomas Bayes nel 1763.

Teorema 2.9 (*Legge di Bayes*) Sia $\{B_n; n = 1, 2, \dots, K\}$ un insieme di eventi incompatibili di \mathcal{F} tali che $P(B_n) > 0$ per $n = 1, 2, \dots, K$ e sia $A \in \mathcal{F}$ un evento con $P(A) > 0$. Se $A \subset \bigcup_{n=1}^K B_n$, per $n = 1, 2, \dots, K$ si ha:

$$P(B_n|A) = \frac{P(B_n) P(A|B_n)}{\sum_{i=1}^K P(B_i) P(A|B_i)}. \quad (2.27)$$

Dimostrazione Essendo per ipotesi $P(A) > 0$ e $P(B_n) > 0$ per $n = 1, 2, \dots, K$, dalla Definizione 2.11 segue:

$$P(B_n|A) = \frac{P(A \cap B_n)}{P(A)} = \frac{P(B_n) P(A|B_n)}{P(A)}. \quad (2.28)$$

Inoltre, poiché per ipotesi $A \subset \bigcup_{i=1}^K B_i$, si ha $A = A \cap \left(\bigcup_{i=1}^K B_i\right) = \bigcup_{i=1}^K (A \cap B_i)$. Per ipotesi gli eventi B_i ($i = 1, 2, \dots, K$) sono a due a due incompatibili e tali che $P(B_i) > 0$ per $i = 1, 2, \dots, K$; pertanto risulta:

$$P(A) = P\left(\bigcup_{i=1}^K (A \cap B_i)\right) = \sum_{i=1}^K P(A \cap B_i) = \sum_{i=1}^K P(B_i) P(A|B_i). \quad (2.29)$$

Sostituendo la (2.29) nella (2.28) segue la (2.27). \square

Si noti che la (2.27) è valida in particolare se gli eventi B_n ($n = 1, 2, \dots, K$) sono a due a due incompatibili e tali che $\bigcup_{n=1}^K B_n = \Omega$. Infatti, in tal caso la condizione $A \subset \bigcup_{n=1}^K B_n$ è soddisfatta per ogni evento $A \in \mathcal{F}$.

La legge di Bayes assume un significato particolarmente importante se agli eventi coinvolti si assegna il ruolo di *cause* ed *effetto*. Precisamente, gli eventi B_n possono essere considerati come possibili cause del verificarsi dell'evento A (effetto delle cause considerate) o essere riguardati come ipotesi che ne rendono conto. La condizione $A \subset \bigcup_{n=1}^K B_n$ comporta che se si verifica A , si verifica necessariamente uno ed uno solo degli eventi B_n , essendo questi ultimi incompatibili. Quindi, se l'evento A si verifica, una ed una sola delle cause, o delle ipotesi, B_n deve essere intervenuta. Una volta osservato l'evento A , ci si può dunque chiedere quale sia stata la causa o l'ipotesi responsabile della sua occorrenza. Poiché la legge di Bayes fornisce in termini probabilistici la risposta a tale domanda, ad essa viene spesso attribuita la denominazione di legge delle “probabilità delle cause” o delle

“probabilità delle ipotesi”. Si noti che nella (2.27) $P(B_n)$ è la probabilità che si verifichi B_n indipendentemente dal presentarsi o meno dell’evento A ; essa è pertanto detta *probabilità a priori*. La probabilità condizionata $P(B_n|A)$ è la probabilità di B_n valutata sapendo che si è verificato l’evento A , ed è denominata *probabilità a posteriori*. La legge di Bayes permette quindi di calcolare la probabilità a posteriori degli eventi B_n supponendo note le probabilità a priori delle “cause” o “ipotesi” $P(B_n)$ e le probabilità condizionate $P(A|B_n)$ dette anche “probabilità probative” o “verosimiglianze”. Naturalmente tale interpretazione può generare errori ed equivoci se utilizzata ad esempio nel contesto di fenomeni deterministici e non con specifico riferimento ad uno spazio di probabilità. È a tal proposito opportuno sottolineare che l’utilizzazione di formule probabilistiche va effettuata esclusivamente con riferimento ad uno spazio di probabilità specificato in relazione ad un esperimento casuale; mai per rendere conto di singoli fatti compiuti specifici che, proprio in quanto tali, non hanno nulla di probabilistico.

Esempio 2.31 Si considerino due urne indistinguibili, l’una contenente una biglia bianca e una biglia rossa e l’altra tre biglie rosse e una biglia verde. L’esperimento consiste nello scegliere a caso un’urna e da essa estrarre una biglia. Si vuole calcolare la probabilità che, scegliendo a caso un’urna ed estraendo da essa una biglia che risulta essere rossa, sia stata scelta l’urna i -esima ($i = 1, 2$).

Si considerino i seguenti eventi: $A = \{\text{è stata estratta una biglia rossa}\}$, $B_i = \{\text{è stata scelta l’urna } i\text{-esima}\}$ ($i = 1, 2$). Poiché non vi è alcuna preferenza nella scelta dell’una o dell’altra urna, $P(B_i) = 1/2$ ($i = 1, 2$); inoltre, dai dati forniti, risulta $P(A|B_1) = 1/2$ e $P(A|B_2) = 3/4$. Dalla (2.27) segue allora:

$$P(B_1|A) = \frac{P(A|B_1)}{P(A|B_1) + P(A|B_2)} = \frac{2}{5},$$

e quindi $P(B_2|A) = 3/5$. ◇

Esempio 2.32 Due tiratori sparano indipendentemente un colpo ciascuno su di un bersaglio. Siano rispettivamente p e q le probabilità che essi colpiscano il centro del bersaglio. Si è interessati a calcolare la probabilità che l’autore del centro sia stato il primo (secondo) tiratore sapendo che il centro del bersaglio è stato colpito una sola volta.

Si definiscano i seguenti eventi: $A = \{\text{il centro del bersaglio è stato colpito una sola volta}\}$, $B_1 = \{\text{i due tiratori sparano un colpo a vuoto}\}$, $B_2 = \{\text{i due tiratori fanno entrambi centro}\}$, $B_3 = \{\text{solo il primo tiratore colpisce il bersaglio}\}$, $B_4 = \{\text{solo il secondo tiratore colpisce il bersaglio}\}$. Si nota che gli eventi B_1, B_2, B_3, B_4 sono necessari ed incompatibili e tali che $P(B_1) = (1 - p)(1 - q)$, $P(B_2) = pq$, $P(B_3) = p(1 - q)$, $P(B_4) = (1 - p)q$. Dal Teorema 2.9 si ricava:

$$P(B_n|A) = \frac{P(B_n)P(A|B_n)}{\sum_{i=1}^4 P(B_i)P(A|B_i)} \quad (n = 1, 2, 3, 4). \quad (2.30)$$

Dai dati forniti risulta evidentemente $P(A|B_1) = P(A|B_2) = 0$, $P(A|B_3) = P(A|B_4) = 1$, così che dalla (2.30) segue $P(B_1|A) = P(B_2|A) = 0$, mentre per le probabilità richieste si

ottiene:

$$P(B_3|A) = \frac{p(1-q)}{p+q-2pq}, \quad P(B_4|A) = \frac{(1-p)q}{p+q-2pq}.$$

◇

È infine qui opportuno menzionare che la legge delle alternative e il teorema di Bayes si estendono facilmente al caso di insiemi numerabili di alternative.

Capitolo 3

Le variabili aleatorie

3.1 Variabili aleatorie unidimensionali

Lo spazio di probabilità associato ad un esperimento casuale ne descrive i risultati in termini probabilistici. Questi sono spesso espressi da numeri, incerti prima dell'effettuazione dell'esperimento e successivamente determinati. Talora si è interessati proprio nei risultati dell'esperimento, altre volte in quantità ottenute mediante talune elaborazioni dei risultati. Naturalmente, qualora i risultati non siano direttamente espressi da numeri, potendo ad esempio essere colori, forme, specie zoologiche, ecc., mediante opportune corrispondenze biunivoche ci si può sempre ricondurre al caso numerico. In generale l'interesse è rivolto ad una funzione dei risultati dell'esperimento casuale. Tale funzione è detta *variabile aleatoria* oppure *variabile casuale*. Esempi di variabili aleatorie sono il numero di prove necessarie affinché si verifichi per la prima volta un determinato evento in un esperimento di prove ripetute, il numero di biglie di un fissato colore in un campione estratto a caso da un'urna, il numero di chiamate che pervengono ad un centralino telefonico in un generico intervallo di tempo di durata fissata, il numero di clienti in fila in un supermercato in un fissato istante oppure il tempo di attesa di un generico cliente, il profitto dell'attività commerciale di un'azienda in un generico esercizio finanziario, il numero di componenti difettosi in un dispositivo elettronico scelto a caso in un ampio stock, la durata di corretto funzionamento di una complessa apparecchiatura, il numero di punti distinti attraverso cui transita una particella che si muove in modo erratico su di un asse coordinato, e così via. Questi esempi sono tutti caratterizzati da un elemento comune: la variabile che esprime la grandezza in esame non assume un unico valore, ma più valori in corrispondenza dei risultati degli esperimenti casuali esplicitamente invocati o implicitamente considerati.

Se Ω è lo spazio campione e $\omega \in \Omega$ è il generico risultato dell'esperimento casuale, una variabile aleatoria X associa ad ω un numero reale $X(\omega)$. Quindi una variabile aleatoria è semplicemente una funzione a valori reali dei risultati dell'esperimento casuale, ossia una funzione $X: \Omega \rightarrow \mathbb{R}$. Affinché possa essere di interesse nel calcolo delle probabilità, tale funzione deve soddisfare taluni requisiti tali da consentire l'individuazione di appropriate relazioni tra certi sottoinsiemi di numeri reali e gli elementi di \mathcal{F} , ossia gli eventi, ai quali sono

già assegnate le probabilità di occorrenza tramite la misura di probabilità P . I sottoinsiemi numerici che rivestono importanza, ed ai quali risulta utile estendere la nozione di probabilità, sono quelli che si ottengono mediante operazioni di unione, intersezione e complementazione di intervalli dell'asse reale, ossia gli insiemi della cosiddetta classe di Borel \mathcal{B} . Di qui la necessità di introdurre un nuovo concetto di variabile che coinvolga sia la struttura degli eventi che la misura di probabilità definita su di essa.

Dire che una variabile aleatoria X assume valori nell'insieme $B \in \mathcal{B}$ equivale ad individuare l'insieme $\{\omega \in \Omega: X(\omega) \in B\}$. Perché sia definita la seguente probabilità

$$P(X \in B) = P(\{\omega \in \Omega: X(\omega) \in B\}),$$

è necessario che l'insieme $\{\omega \in \Omega: X(\omega) \in B\}$ sia un evento, ossia un elemento della σ -algebra \mathcal{F} . Quanto finora esposto conduce alla seguente

Definizione 3.1 *Sia assegnato uno spazio di probabilità (Ω, \mathcal{F}, P) . Una variabile aleatoria, o variabile casuale, è una funzione $X: \Omega \rightarrow \mathbb{R}$ tale che per ogni sottoinsieme di Borel $B \in \mathcal{B}$ risulti $\{\omega \in \Omega: X(\omega) \in B\} \in \mathcal{F}$.*

Poiché le semirette sinistre $(-\infty, x]$, con $x \in \mathbb{R}$, sono elementi generatori della classe di Borel \mathcal{B} su \mathbb{R} , una funzione $X: \Omega \rightarrow \mathbb{R}$ è una variabile aleatoria se $\{\omega \in \Omega: X(\omega) \leq x\} \in \mathcal{F}$ per ogni reale x . D'altra parte, un noto risultato della teoria della misura asserisce che X è una funzione misurabile su un insieme Ω se e solo se per ogni $x \in \mathbb{R}$ risulta $\{\omega \in \Omega: X(\omega) \leq x\} \in \mathcal{F}$. La definizione di funzione misurabile su Ω equivale pertanto a dire che per ogni $x \in \mathbb{R}$ è possibile assegnare agli insiemi $\{\omega \in \Omega: X(\omega) \leq x\}$ una misura di probabilità, ossia riguardare tali insiemi come eventi. Quindi, in alternativa, è possibile dare la seguente

Definizione 3.2 *Una variabile aleatoria è una funzione $X: \Omega \rightarrow \mathbb{R}$ a valori reali definita sullo spazio campione Ω ed ivi misurabile, ossia tale che $\{\omega \in \Omega: X(\omega) \leq x\} \in \mathcal{F}$ per ogni $x \in \mathbb{R}$.*

Nel seguito indicheremo una variabile aleatoria con $X(\omega)$, $Y(\omega)$,... o, più semplicemente, con le lettere X, Y, \dots ; potrà trattarsi di una variabile aleatoria adimensionale, ossia descrivente quantità scalari, oppure di una variabile aleatoria che descrive una grandezza dimensionale (ad esempio un peso, una lunghezza, un tempo, un importo,...), e che quindi deve essere espressa mediante opportune unità di misura.

Esempi di variabili aleatorie sono riportati qui di seguito.

Esempio 3.1 Si consideri l'esperimento consistente nel lancio di una moneta, con $\Omega = \{T, C\}$ e $\mathcal{F} = \{\emptyset, \{T\}, \{C\}, \Omega\}$. Introduciamo una funzione $X: \Omega \rightarrow \mathbb{R}$ tramite le posizioni $X(C) = 0$ e $X(T) = 1$. Pertanto $\{\omega \in \Omega: X(\omega) = 0\} = \{C\}$ e $\{\omega \in \Omega: X(\omega) = 1\} = \{T\}$. Poiché

$$\{\omega \in \Omega: X(\omega) \leq x\} = \begin{cases} \emptyset, & x < 0 \\ \{C\}, & 0 \leq x < 1 \\ \{T, C\} = \Omega, & x \geq 1, \end{cases}$$

$\{\omega \in \Omega: X(\omega) \leq x\} \in \mathcal{F}$ per ogni $x \in \mathbb{R}$ e quindi dalla Definizione 3.2 segue che X è una variabile aleatoria. Si prenda ora in esame l'esperimento consistente nel lancio ripetuto

due volte di una moneta con spazio campione $\Omega = \{CC, CT, TC, TT\}$ e sia \mathcal{F} la σ -algebra costituita dall'insieme potenza di Ω . Si fissi poi l'attenzione sul numero totale di T ottenuto nei due lanci; in tal caso si può introdurre una funzione $Y : \Omega \rightarrow \mathbb{R}$ tramite le posizioni $Y(CC) = 0, Y(CT) = 1, Y(TC) = 1, Y(TT) = 2$. Essendo

$$\{\omega \in \Omega : Y(\omega) \leq y\} = \begin{cases} \emptyset, & y < 0 \\ \{CC\}, & 0 \leq y < 1 \\ \{CC, CT, TC\}, & 1 \leq y < 2 \\ \{CC, CT, TC, TT\} = \Omega, & y \geq 2, \end{cases}$$

si ha che $\{\omega \in \Omega : Y(\omega) \leq y\} \in \mathcal{F}$ per ogni $y \in \mathbb{R}$, così che Y è una variabile aleatoria in virtù della Definizione 3.2. \diamond

Esempio 3.2 Si consideri l'esperimento consistente nel lancio di un dado con spazio campione $\Omega = \{1, 2, 3, 4, 5, 6\}$ e sia \mathcal{F} la σ -algebra costituita dall'insieme potenza di Ω . Introduciamo una funzione $X : \Omega \rightarrow \mathbb{R}$ tramite la posizione $X(\omega) = \omega$ ($\omega = 1, 2, 3, 4, 5, 6$) e sia $Y = |X - 3|$. Dimostriamo che Y è una variabile aleatoria.

Si noti che la funzione $Y : \Omega \rightarrow \mathbb{R}$ è esplicitamente definita al seguente modo:

$$\begin{aligned} \{\omega \in \Omega : Y(\omega) = 0\} &= \{3\}, & \{\omega \in \Omega : Y(\omega) = 1\} &= \{2, 4\}, \\ \{\omega \in \Omega : Y(\omega) = 2\} &= \{1, 5\}, & \{\omega \in \Omega : Y(\omega) = 3\} &= \{6\}. \end{aligned}$$

Essendo

$$\{\omega \in \Omega : Y(\omega) \leq y\} = \begin{cases} \emptyset, & y < 0 \\ \{3\}, & 0 \leq y < 1 \\ \{2, 3, 4\}, & 1 \leq y < 2 \\ \{1, 2, 3, 4, 5\}, & 2 \leq y < 3 \\ \{1, 2, 3, 4, 5, 6\} = \Omega, & y \geq 3, \end{cases}$$

dalla Definizione 3.2 segue che Y è una variabile aleatoria. \diamond

Esempio 3.3 Si supponga che pervenga una chiamata ad un centralino telefonico in un istante a caso dell'intervallo $(0, \tau)$. Lo spazio campione è costituito da tutti gli istanti t di questo intervallo, mentre la famiglia degli eventi \mathcal{F} consiste negli insiemi della classe di Borel appartenenti all'intervallo $(0, \tau)$. Sia

$$X(t) = \begin{cases} 1, & 0 < t \leq t_1 \\ 0, & t_1 < t < \tau. \end{cases}$$

Poiché risulta

$$\{t \in (0, \tau) : X(t) \leq x\} = \begin{cases} \emptyset, & x < 0 \\ (t_1, \tau), & 0 \leq x < 1 \\ (0, \tau), & x \geq 1, \end{cases}$$

si ha $\{t \in (0, \tau) : X(t) \leq x\} \in \mathcal{F}$ per ogni $x \in \mathbb{R}$, così che X è una variabile aleatoria. \diamond

3.2 La funzione di distribuzione

Uno strumento essenziale per lo studio delle variabili aleatorie è la *funzione di distribuzione*, detta anche *funzione di ripartizione*.

Definizione 3.3 Sia $X : \Omega \rightarrow \mathbb{R}$ una variabile aleatoria. La funzione

$$F_X(x) = P(X \leq x) = P(\{\omega \in \Omega: X(\omega) \leq x\}), \quad x \in \mathbb{R} \quad (3.1)$$

prende il nome di *funzione di distribuzione*, o *funzione di ripartizione*, della variabile aleatoria X .

Dalla Definizione 3.3 segue che $F_X : \mathbb{R} \rightarrow [0, 1]$ e che ogni variabile aleatoria possiede un'unica funzione di distribuzione. Si noti anche che variabili aleatorie distinte possono avere la medesima funzione di distribuzione. In tal caso si dice anche che esse sono identicamente distribuite. Ad esempio, sia $\Omega = \{\omega_1, \omega_2\}$ con $P(\{\omega_1\}) = P(\{\omega_2\}) = 1/2$. La variabile aleatoria X definita dalle relazioni $X(\omega_1) = -1$ e $X(\omega_2) = 1$ è caratterizzata da funzione di distribuzione

$$F_X(x) = P(\{\omega \in \Omega: X(\omega) \leq x\}) = \begin{cases} 0, & x < -1, \\ 1/2, & -1 \leq x < 1, \\ 1, & x \geq 1. \end{cases}$$

Se si considera la variabile aleatoria $Y = -X$, si verifica agevolmente che risulta $F_Y(x) = F_X(x)$ per ogni $x \in \mathbb{R}$, il che mostra che le variabili aleatorie distinte X e Y possiedono la stessa funzione di distribuzione, ovvero X e Y sono identicamente distribuite.

Osserviamo esplicitamente che la funzione di distribuzione $F_X(x)$ di una variabile aleatoria X non fornisce le probabilità di *tutti* gli eventi, ossia di tutti gli elementi di \mathcal{F} , ma solo di quelli le cui immagini tramite X sono intervalli del tipo $(-\infty, x]$ con $x \in \mathbb{R}$. Vedremo tuttavia che mediante la funzione di distribuzione risulta possibile esprimere la probabilità di qualsiasi evento $E \in \mathcal{F}$ la cui immagine tramite X è un elemento $B \in \mathcal{B}$; ciò in virtù della circostanza che gli intervalli del tipo $(-\infty, x]$ con $x \in \mathbb{R}$ generano la classe di Borel \mathcal{B} su \mathbb{R} .

Teorema 3.1 La funzione di distribuzione $F_X(x)$ di una variabile aleatoria $X : \Omega \rightarrow \mathbb{R}$ gode delle seguenti proprietà:

- (i) $F_X(x)$ è non decrescente in \mathbb{R} , ossia per ogni $x_1, x_2 \in \mathbb{R}$ con $x_1 < x_2$ risulta $F_X(x_1) \leq F_X(x_2)$;
- (ii) $F_X(x)$ è continua a destra in ogni $x \in \mathbb{R}$, ossia per ogni $\varepsilon > 0$ si ha $\lim_{\varepsilon \downarrow 0} F_X(x + \varepsilon) = F_X(x)$;
- (iii) $\lim_{x \rightarrow -\infty} F_X(x) = 0$, $\lim_{x \rightarrow +\infty} F_X(x) = 1$.

Dimostrazione (i) Siano x_1 e x_2 reali arbitrari con $x_1 < x_2$. Poiché

$$\{\omega \in \Omega: X(\omega) \leq x_1\} \subset \{\omega \in \Omega: X(\omega) \leq x_2\},$$

applicando il Teorema 2.6, segue:

$$P(\{\omega \in \Omega: X(\omega) \leq x_1\}) \leq P(\{\omega \in \Omega: X(\omega) < x_2\}),$$

cosicché, per la Definizione 3.3, si ha $F_X(x_1) \leq F_X(x_2)$.

(ii) Essendo

$$\lim_{\varepsilon \downarrow 0} F_X(x + \varepsilon) = \lim_{n \rightarrow +\infty} F_X\left(x + \frac{1}{n}\right),$$

è sufficiente mostrare che in ogni $x \in \mathbb{R}$ si ha:

$$\lim_{n \rightarrow +\infty} F_X\left(x + \frac{1}{n}\right) = F_X(x).$$

A tal fine, osserviamo che poiché

$$\left\{\omega \in \Omega: X(\omega) \leq x + \frac{1}{n}\right\} = \{\omega \in \Omega: X(\omega) \leq x\} \cup \left\{\omega \in \Omega: x < X(\omega) \leq x + \frac{1}{n}\right\},$$

con i due eventi al secondo membro incompatibili, risulta:

$$F_X\left(x + \frac{1}{n}\right) = F_X(x) + P\left(\left\{\omega \in \Omega: x < X(\omega) \leq x + \frac{1}{n}\right\}\right). \quad (3.2)$$

Posto

$$A_n = \left\{\omega \in \Omega: x < X(\omega) \leq x + \frac{1}{n}\right\} \quad (n = 1, 2, \dots),$$

si nota che $\{A_n; n = 1, 2, \dots\}$ è una successione non crescente di eventi il cui limite è l'evento impossibile. Dal Teorema 2.1 segue pertanto

$$\lim_{n \rightarrow +\infty} P(A_n) = P(\emptyset) = 0,$$

così che passando al limite per $n \rightarrow +\infty$ nella (3.2) si ottiene:

$$\lim_{n \rightarrow +\infty} F_X\left(x + \frac{1}{n}\right) = F_X(x) + \lim_{n \rightarrow +\infty} P(A_n) = F_X(x).$$

La funzione $F_X(x)$ è dunque continua a destra in ogni $x \in \mathbb{R}$.

(iii) Essendo

$$\lim_{x \rightarrow -\infty} F_X(x) = \lim_{n \rightarrow +\infty} F_X(-n), \quad \lim_{x \rightarrow +\infty} F_X(x) = \lim_{n \rightarrow +\infty} F_X(n),$$

è sufficiente mostrare che risulta:

$$\lim_{n \rightarrow +\infty} F_X(-n) = 0, \quad \lim_{n \rightarrow +\infty} F_X(n) = 1.$$

A tal fine poniamo:

$$B_n^- = \{\omega \in \Omega: X(\omega) \leq -n\}, \quad B_n^+ = \{\omega \in \Omega: X(\omega) \leq n\} \quad (n = 0, 1, 2, \dots).$$

Poiché $\{B_n^-; n = 1, 2, \dots\}$ è una successione non crescente di eventi il cui limite è l'evento impossibile, dal Teorema 2.1 segue:

$$\lim_{n \rightarrow +\infty} F_X(-n) = \lim_{n \rightarrow +\infty} P(B_n^-) = P(\emptyset) = 0.$$

Analogamente, essendo $\{B_n^+; n = 1, 2, \dots\}$ una successione non decrescente di eventi il cui limite è l'evento certo, dal Teorema 2.1 si ha:

$$\lim_{n \rightarrow +\infty} F_X(n) = \lim_{n \rightarrow +\infty} P(B_n^+) = P(\Omega) = 1.$$

□

È interessante osservare che le proprietà (i), (ii) e (iii) caratterizzano completamente le funzioni di distribuzione. Sussiste infatti la seguente proposizione che ci limitiamo ad enunciare:

Proposizione 3.1 *Sia $F: \mathbb{R} \rightarrow \mathbb{R}$ una funzione che soddisfa le proprietà (i), (ii) e (iii) del Teorema 3.1. Esistono allora uno spazio di probabilità (Ω, \mathcal{F}, P) ed una variabile aleatoria $X: \Omega \rightarrow \mathbb{R}$ avente funzione di distribuzione $F_X(x) = F(x)$ per ogni $x \in \mathbb{R}$.*

In virtù di questo risultato, per verificare se una funzione è interpretabile come funzione di distribuzione è sufficiente verificare che essa soddisfi le suddette proprietà (i), (ii), (iii).

La conoscenza della funzione di distribuzione di una variabile aleatoria permette di calcolare la probabilità che essa assuma valori in un qualsiasi insieme di Borel B . La proposizione che segue mostra come sia possibile esprimere in termini di funzione di distribuzione le probabilità che una variabile aleatoria assuma valori in intervalli qualsiasi dell'asse reale o in singoli punti.

Proposizione 3.2 *Sia X una variabile aleatoria di funzione di distribuzione $F_X(x)$ e siano x_1, x_2 e x numeri reali con $x_1 < x_2$. Posto $F_X(x^-) = \lim_{\varepsilon \downarrow 0} F_X(x - \varepsilon)$, si ha:*

- (i) $P(X < x) = F_X(x^-)$;
- (ii) $P(X = x) = F_X(x) - F_X(x^-)$;
- (iii) $P(x_1 < X < x_2) = F_X(x_2^-) - F_X(x_1)$;
- (iv) $P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)$;
- (v) $P(x_1 \leq X < x_2) = F_X(x_2^-) - F_X(x_1^-)$;
- (vi) $P(x_1 \leq X \leq x_2) = F_X(x_2) - F_X(x_1^-)$;
- (vii) $P(X > x) = 1 - F_X(x)$;
- (viii) $P(X \geq x) = 1 - F_X(x^-)$.

Se $F_X(x)$ non è continua in un punto $x = x_0 \in R$, la differenza

$$F_X(x_0) - F_X(x_0^-) = P(X = x_0) > 0,$$

rappresenta l'ampiezza del salto di $F_X(x)$ in x_0 . Tale ampiezza fornisce la probabilità che X assuma il valore x_0 .

È possibile mostrare che ogni funzione di distribuzione possiede al più un'infinità numerabile di punti di discontinuità. Tali discontinuità hanno tutte la forma di salti, cioè sono discontinuità di prima specie. Infatti, essendo $0 \leq F_X(x) \leq 1$, una funzione di distribuzione non può esibire più di un salto ciascuno di ampiezza superiore ad $1/2$, non più di tre salti ciascuno di ampiezza compresa tra $1/4$ a $1/2$ e, in generale, non può possedere più di $2^n - 1$ salti di ampiezza ciascuno compresa tra $1/2^n$ e $1/2^{n-1}$.

Occorre menzionare, infine, che spesso la funzione di distribuzione non viene definita tramite la (3.1) ma attraverso la posizione $F_X^*(x) = P(X < x)$, così che essa risulta continua a sinistra invece che a destra. Non è difficile vedere come vadano in tal caso modificati gli enunciati del Teorema 3.1 e della Proposizione 3.2.

Esempio 3.4 Si riconsideri la variabile aleatoria X introdotta nell'Esempio 3.1 e si supponga che Testa si presenti con probabilità p e Croce con probabilità $q = 1 - p$. In tal caso si ha

$$F_X(x) = P(X \leq x) = \begin{cases} 0, & x < 0 \\ q, & 0 \leq x < 1 \\ 1, & x \geq 1, \end{cases} \quad F_X^*(x) = P(X < x) = \begin{cases} 0, & x \leq 0 \\ q, & 0 < x \leq 1 \\ 1, & x > 1. \end{cases}$$

Le due funzioni di distribuzione differiscono soltanto nei punti $x = 0$ e $x = 1$, dove risulta $P(X = 0) = F_X(0) - F_X(0^-) = F_X^*(0^+) - F_X^*(0)$ e $P(X = 1) = F_X(1) - F_X(1^-) = F_X^*(1^+) - F_X^*(1)$. \diamond

3.3 Classificazione delle variabili aleatorie unidimensionali

Se X è una variabile aleatoria definita su Ω ed a valori in \mathbb{R} , si è generalmente interessati a calcolare probabilità di eventi che coinvolgono X , ossia del tipo $P(\{\omega \in \Omega: X(\omega) \in B\})$ per specificati insiemi B di Borel. A tal fine è conveniente considerare separatamente variabili aleatorie di tre diversi tipi: discrete, assolutamente continue e miste.

3.3.1 Variabili aleatorie discrete

Definizione 3.4 Una variabile aleatoria X si dice discreta se esiste un insieme $S = \{x_1, x_2, \dots\}$ al più numerabile di reali distinti tale che $P(X \in S) = 1$.

Poiché

$$\{\omega \in \Omega: X(\omega) \in S\} = \bigcup_{n: x_n \in S} \{\omega \in \Omega: X(\omega) = x_n\},$$

per l'incompatibilità degli eventi di cui si è effettuata l'unione, segue:

$$\sum_{n: x_n \in S} P(X = x_n) = P\left(\bigcup_{n: x_n \in S} \{X = x_n\}\right) = P(X \in S) = 1. \quad (3.3)$$

Una variabile aleatoria discreta è quindi individuata dall'insieme di coppie $\{x_n, p_n; n = 1, 2, \dots\}$, dove $x_n \in S$ sono i valori assunti da X e dove si è posto $p_n = P(X = x_n)$.

Alla variabile aleatoria discreta X è possibile associare la funzione $p_X : \mathbb{R} \rightarrow [0, 1]$ così definita:

$$p_X(x) = P(X = x) = \begin{cases} p_n, & x = x_n \quad (n = 1, 2, \dots) \\ 0, & \text{altrimenti.} \end{cases}$$

Ad essa si dà il nome di *funzione di probabilità* o *distribuzione di probabilità*. La probabilità che X assuma valori in un insieme B di Borel può allora essere calcolata nel seguente modo:

$$P(X \in B) = \sum_{n: x_n \in B} p_n. \quad (3.4)$$

Dalla (3.4) segue che la funzione di distribuzione di una variabile aleatoria discreta X può essere così espressa:

$$F_X(x) = P(X \leq x) = \sum_{n: x_n \leq x} p_n \quad (x \in \mathbb{R}). \quad (3.5)$$

Dalla (3.5) si ricava poi:

$$F_X(x) - F_X(x^-) = \sum_{n: x_n \leq x} p_n - \sum_{n: x_n < x} p_n = \begin{cases} p_k, & x = x_k \quad (k = 1, 2, \dots) \\ 0, & \text{altrimenti,} \end{cases}$$

che mostra come la funzione di distribuzione di una variabile aleatoria discreta sia discontinua e crescente a salti nei punti $x_n \in S$ nei quali risulta $p_X(x_n) > 0$. L'ampiezza del salto di $F_X(x)$ in corrispondenza del punto di ascissa x_n è dunque uguale a p_n . Se due valori assunti dalla variabile aleatoria discreta X sono separati da un intervallo in cui X non assume altri valori con probabilità non nulle, allora la funzione di distribuzione $F_X(x)$ è costante in tale intervallo. Se la variabile aleatoria discreta X assume un numero finito r di valori x_k con probabilità p_k positiva, allora il grafico di $F_X(x)$ è a gradini e costante in $r + 1$ intervalli; se, invece, X assume valori in un insieme numerabile, questo potrebbe anche essere ovunque denso, nel qual caso non esiste nessun intervallo in cui la funzione di distribuzione è costante. Ad esempio, si consideri una variabile aleatoria X che assume valori nell'insieme numerabile x_1, x_2, \dots costituito dai numeri razionali e sia $P(X = x_k) = 1/2^k$ ($k = 1, 2, \dots$). In tal caso ognuno di questi è un punto di discontinuità della funzione di distribuzione; poiché ogni intervallo comunque piccolo contiene infiniti punti di ascissa razionale, ne segue che la funzione di distribuzione presenta infiniti salti.

Esempio 3.5 Si consideri una variabile aleatoria X che assume il valore a con probabilità unitaria. La sua funzione di distribuzione è pertanto:

$$F_X(x) = P(X \leq x) = \begin{cases} 0, & x < a \\ 1, & x \geq a. \end{cases}$$

In tal caso X è detta *degenere* o *concentrata in un unico valore*; essa perde il carattere aleatorio per assumere un unico valore con probabilità unitaria, ossia quasi certamente. Si tratta evidentemente di un caso limite, che peraltro riveste importanza nel calcolo delle probabilità.

◇

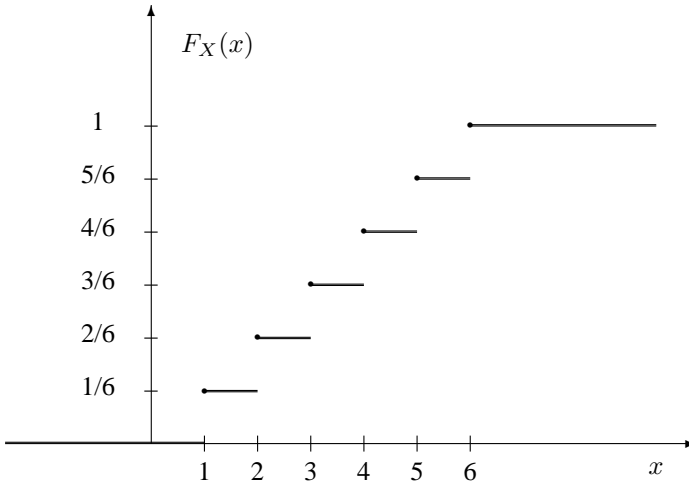


Figura 3.1 – La funzione di distribuzione di cui all'Esempio 3.6.

Esempio 3.6 Nell'esperimento consistente nel lancio di un dado si consideri la variabile aleatoria X che assume i valori $1, 2, \dots, 6$ con probabilità $P(X = n) = 1/6$ ($n = 1, 2, \dots, 6$). La sua funzione di distribuzione è pertanto la seguente:

$$F_X(x) = P(X \leq x) = \begin{cases} 0, & x < 1 \\ \frac{k}{6}, & k \leq x < k+1 \quad (k = 1, 2, \dots, 5) \\ 1, & x \geq 6, \end{cases}$$

il cui grafico, a gradini di lunghezza unitaria e di altezza $1/6$, è mostrato in Figura 3.1. \diamond

Si può dare un'interessante ed utile rappresentazione delle variabili aleatorie discrete mediante la cosiddetta *funzione indicatrice*. La funzione indicatrice I_A di un evento A è definita al seguente modo:

$$I_A(\omega) = \begin{cases} 0, & \omega \in \bar{A} \\ 1, & \omega \in A. \end{cases} \quad (3.6)$$

Essa è pertanto una variabile aleatoria discreta che assume i valori 0 e 1, di funzione di distribuzione

$$F(x) = \begin{cases} 0, & x < 0 \\ P(I_A = 0) = P(\bar{A}), & 0 \leq x < 1 \\ 1, & x \geq 1. \end{cases}$$

Si consideri ora una variabile discreta X che assume valori in $S = \{x_1, x_2, \dots\}$. Poiché gli insiemi $A_n = \{\omega \in \Omega: X(\omega) = x_n\}$ ($n = 1, 2, \dots$) possono essere riguardati come eventi

necessari ed incompatibili, si ha

$$X(\omega) = \sum_n x_n I_{A_n}(\omega). \quad (3.7)$$

Infatti, $I_{A_k}(\omega) = 1$ se $\omega \in A_k$, mentre le altre funzioni indicatrici nella somma (3.7) si annullano, così che $X(\omega) = x_k$. La (3.7) mostra che ogni variabile aleatoria discreta può esprimersi mediante una combinazione lineare di variabili aleatorie che assumono i soli valori 0 e 1.

3.3.2 Variabili aleatorie assolutamente continue

Definizione 3.5 Una variabile aleatoria X di funzione di distribuzione $F_X(x)$ si dice assolutamente continua se esiste una funzione non negativa $f_X(x)$ tale che $F_X(x)$ è esprimibile al seguente modo:

$$F_X(x) = \int_{-\infty}^x f_X(z) dz \quad \forall x \in \mathbb{R}. \quad (3.8)$$

La funzione $f_X(x)$ prende il nome di *densità di probabilità* della variabile aleatoria X . In ogni punto di continuità di $f_X(x)$, dal teorema fondamentale del calcolo integrale segue:

$$f_X(x) = \frac{d}{dx} F_X(x) = \frac{d}{dx} \int_{-\infty}^x f_X(z) dz. \quad (3.9)$$

In ogni punto di continuità la densità di probabilità coincide, dunque, con la derivata della funzione di distribuzione.

Per la proprietà di non decrescenza della funzione di distribuzione (v. la (i) del Teorema 3.1) deve risultare $f_X(x) \geq 0$ almeno nei punti $x \in \mathbb{R}$ in cui tale funzione è continua. D'altro canto, poiché sarebbe non significativo definire una densità di probabilità negativa in qualche punto isolato, nella Definizione 3.5 va esplicitamente richiesto che risulti:

$$f_X(x) \geq 0, \quad \forall x \in \mathbb{R}. \quad (3.10)$$

Inoltre, dalla proprietà (iii) del Teorema 3.1 si ricava:

$$\int_{-\infty}^{+\infty} f_X(z) dz = 1. \quad (3.11)$$

La densità di probabilità di una variabile aleatoria X assolutamente continua deve quindi soddisfare le proprietà (3.10) e (3.11).

La probabilità che una variabile aleatoria X assuma valori in un assegnato insieme B di Borel può essere espressa nel seguente modo:

$$P(X \in B) = \int_B f_X(z) dz. \quad (3.12)$$

A differenza del caso discreto, la funzione di distribuzione di una variabile aleatoria assolutamente continua X è una funzione continua; infatti, per ogni $x \in \mathbb{R}$ risulta:

$$\begin{aligned} F_X(x^-) &= \lim_{\varepsilon \downarrow 0} F_X(x - \varepsilon) = \lim_{\varepsilon \downarrow 0} \int_{-\infty}^{x-\varepsilon} f_X(z) dz = F_X(x), \\ F_X(x^+) &= \lim_{\varepsilon \downarrow 0} F_X(x + \varepsilon) = \lim_{\varepsilon \downarrow 0} \int_{-\infty}^{x+\varepsilon} f_X(z) dz = F_X(x). \end{aligned}$$

Pertanto, in virtù della (ii) della Proposizione 3.2, se X è assolutamente continua risulta:

$$P(X = x) = F_X(x) - F_X(x^-) = 0 \quad \forall x \in \mathbb{R}. \quad (3.13)$$

L'evento $\{X = x\} = \{\omega \in \Omega: X(\omega) = x\}$ è quindi quasi impossibile, ossia ha probabilità nulla di verificarsi. Come conseguenza della (3.13), nel caso assolutamente continuo si trae:

$$\begin{aligned} P(a \leq X \leq b) &= P(a < X \leq b) = P(a \leq X < b) = P(a < X < b) \\ &= \int_a^b f_X(z) dz = F_X(b) - F_X(a). \end{aligned} \quad (3.14)$$

L'integrale di $f_X(x)$ esteso ad un intervallo esprime dunque la probabilità che X assuma valori in quell'intervallo.

Si noti che a differenza del caso discreto, nel caso assolutamente continuo la funzione di probabilità perde di interesse, risultando $P(X = x) = 0$ per ogni $x \in \mathbb{R}$; occorre quindi riferirsi ad intervalli e non a singoli punti.

È opportuno sottolineare che, a differenza della funzione di probabilità del caso discreto, la densità di probabilità non è una probabilità, come del resto la sua denominazione indica. Infatti, se x è un punto di continuità per $f_X(x)$, dalle (3.9) e (3.14) segue:

$$f_X(x) = \lim_{\Delta x \downarrow 0} \frac{F_X(x + \Delta x) - F_X(x)}{\Delta x} = \lim_{\Delta x \downarrow 0} \frac{P(x < X \leq x + \Delta x)}{\Delta x}.$$

Osserviamo che il rapporto $P(x < X \leq x + \Delta x)/\Delta x$ ha le dimensioni di $(\Delta x)^{-1}$ in quanto $P(x < X \leq x + \Delta x)$ è uno scalare; per ottenere una probabilità, che è una quantità scalare, occorre quindi moltiplicare $f_X(x)$ per Δx .

Talora nelle applicazioni si fa uso della scrittura:

$$f_X(x) dx \simeq P(x < X \leq x + dx);$$

a meno di infinitesimi di ordine superiore, $f_X(x) dx$ rappresenta quindi la probabilità che X assuma valori nell'intervallo infinitesimo $(x, x + dx]$.

Giova infine osservare che la densità di probabilità di una variabile aleatoria assolutamente continua non è necessariamente continua o limitata, come evidenziato nel seguente esempio.

Esempio 3.7 Si consideri la variabile aleatoria X di funzione di distribuzione

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

Poiché $F_X(x)$ è una funzione continua che soddisfa le proprietà (i), (ii) e (iii) del Teorema 3.1, X è assolutamente continua ed ha densità di probabilità:

$$f_X(x) = \begin{cases} \frac{1}{2\sqrt{x}}, & 0 < x < 1 \\ 0, & \text{altrimenti.} \end{cases}$$

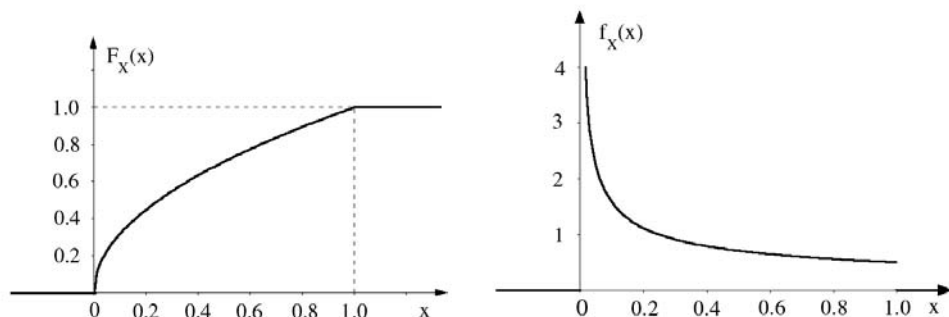


Figura 3.2 – La funzione di distribuzione e la densità di probabilità di cui all'Esempio 3.7.

In Figura 3.2 sono rappresentate le funzioni $F_X(x)$ e $f_X(x)$. Si noti che $f_X(x)$ non è né continua né limitata. \diamond

3.3.3 Variabili aleatorie miste

Nelle applicazioni si incontrano talora variabili aleatorie che non sono né discrete né assolutamente continue. Si può dimostrare che la funzione di distribuzione $F_X(x)$ di una variabile aleatoria X può infatti sempre rappresentarsi come somma di tre componenti:

$$F_X(x) = a_1 F_1(x) + a_2 F_2(x) + a_3 F_3(x), \quad x \in \mathbb{R}, \quad (3.15)$$

con $a_i \geq 0$ ($i = 1, 2, 3$) costanti vincolate dalla relazione $a_1 + a_2 + a_3 = 1$, mentre F_1, F_2, F_3 sono funzioni di distribuzione tali che

- (i) $F_1(x)$ è la funzione di distribuzione di una variabile aleatoria assolutamente continua;
- (ii) $F_2(x)$ è la funzione di distribuzione di una variabile aleatoria discreta;
- (iii) $F_3(x)$ è la *componente singolare* di $F_X(x)$; essa è una funzione ovunque continua a derivata quasi ovunque nulla che caratterizza variabili aleatorie continue ma non assolutamente continue.

Inoltre, le tre componenti $a_1 F_1(x)$, $a_2 F_2(x)$, $a_3 F_3(x)$ sono univocamente determinate da $F_X(x)$.

Le variabili aleatorie continue ma non assolutamente continue, ossia quelle per le quali la (3.15) sussiste con $a_1 = a_2 = 0$ e quindi $a_3 = 1$, non rivestono particolare rilevanza

pratica. Pertanto d'ora innanzi ci riferiremo sempre a variabili aleatorie le cui funzioni di distribuzione sono esprimibili nella forma

$$F_X(x) = a_1 F_1(x) + a_2 F_2(x), \quad (3.16)$$

con a_1, a_2 reali non negativi tali da aversi $a_1 + a_2 = 1$. In particolare, $a_1 = 1$ e $a_2 = 0$ se la variabile aleatoria è assolutamente continua mentre $a_1 = 0$ e $a_2 = 1$ se essa è discreta.

Definizione 3.6 Una variabile aleatoria X si dice *mista* se la sua funzione di distribuzione si può esprimere come in (3.16) con le due componenti $a_1 F_1(x)$ e $a_2 F_2(x)$ entrambe non identicamente nulle.

Esempio 3.8 Si supponga che i costi di riparazione di una determinata apparecchiatura dipendano sia dal tipo di guasto, sia dall'entità dello stesso, sia dal tempo di riparazione. Il costo di riparazione in presenza di un guasto può ragionevolmente ritenersi descritto da una variabile aleatoria assolutamente continua che assume valori non negativi con probabilità unitaria, caratterizzata da una qualche funzione di distribuzione $F_1(x)$. Ovviamente, in assenza di guasti non si incorre in costi di riparazione, ed inoltre il costo, per sua stessa natura, non può risultare negativo. Denotando con p la probabilità che l'apparecchiatura risulti guasta, la funzione di distribuzione della variabile aleatoria X che descrive il costo dell'eventuale riparazione è così esprimibile:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - p + p F_1(x), & x \geq 0. \end{cases} \quad (3.17)$$

La (3.17) esprime quantitativamente le seguenti circostanze:

- (a) la probabilità di incorrere in un costo negativo è nulla;
- (b) fissato un x positivo, la probabilità che il costo sia non superiore ad x quando x tende a zero è $1 - p$, ossia si identifica con la probabilità che l'apparecchiatura non risulti guasta;
- (c) per ogni $x > 0$ la probabilità di incorrere in un costo non superiore ad x cresce monotonamente verso l'unità al tendere di x all'infinito (in quanto $F_1(x)$ tende ad 1 al divergere di x), con una legge di crescita che evidentemente dipende dalla funzione $F_1(x)$ ipotizzata.

Si noti che la variabile aleatoria X non è assolutamente continua poiché la sua funzione di distribuzione non è continua in $x = 0$, presentando in tal punto un salto di ampiezza $1 - p$. Non è poi difficile rendersi conto che X è mista; infatti $F_X(x)$ è suscettibile della seguente rappresentazione:

$$F_X(x) = p F_1(x) + (1 - p) F_2(x), \quad \forall x \in \mathbb{R}, \quad (3.18)$$

dove $F_2(x)$ è la funzione di distribuzione di una variabile aleatoria discreta di tipo degenere che assume il valore 0 quasi certamente (v. Esempio 3.5). \diamond

3.4 Trasformazioni di variabili aleatorie

In problemi sia teorici che applicativi occorre spesso considerare variabili che sono funzioni di una variabile aleatoria assegnata. In tal caso nascono spontaneamente due domande:

- (a) data una variabile aleatoria X ed una funzione $g: \mathbb{R} \rightarrow \mathbb{R}$, la posizione $Y = g(X)$ definisce una variabile aleatoria Y ?
- (b) se $Y = g(X)$ è una variabile aleatoria, quale legame sussiste tra le funzioni di distribuzione di X e di Y ?

La proposizione che segue fornisce una risposta affermativa alla prima domanda quando la funzione g è Borel-misurabile¹.

Proposizione 3.3 *Sia $X: \Omega \rightarrow \mathbb{R}$ una variabile aleatoria e sia $g(x)$ con $g: \mathbb{R} \rightarrow \mathbb{R}$ una funzione Borel-misurabile. La funzione $Y = g(X)$ è allora essa stessa una variabile aleatoria.*

Dimostrazione Essendo $Y(\omega) = g[X(\omega)]$ per ogni $\omega \in \Omega$, la variabile Y è evidentemente definita su Ω ed assume valori in \mathbb{R} , ossia $Y: \Omega \rightarrow \mathbb{R}$. In accordo con la Definizione 3.2 occorre ora dimostrare che per ogni sottoinsieme di Borel $B \in \mathcal{B}$ si ha $\{\omega \in \Omega : Y(\omega) \in B\} \in \mathcal{F}$, dove \mathcal{F} è la sigma algebra costruita sullo spazio campione Ω . Ricordando che $g^{-1}(B) \in \mathcal{B}$ poiché g è Borel-misurabile e che X è una variabile aleatoria, per ogni $B \in \mathcal{B}$ si ha:

$$\{\omega \in \Omega : Y(\omega) \in B\} = \{\omega \in \Omega : g[X(\omega)] \in B\} = \{\omega \in \Omega : X(\omega) \in g^{-1}(B)\} \in \mathcal{F},$$

ossia Y è misurabile su Ω . □

Affrontiamo ora la questione di cui alla precedente seconda domanda. Dal momento che Y è funzione di X , per ogni sottoinsieme di Borel $B \in \mathcal{B}$ risulta:

$$P(Y \in B) = P(g(X) \in B) = P(X \in g^{-1}(B)). \quad (3.19)$$

In particolare, per la funzione di distribuzione di Y si ha:

$$F_Y(y) = P(Y \leq y) = P[g(X) \leq y] = P\{X \in g^{-1}((-\infty, y])\}. \quad (3.20)$$

La formula (3.20) permette di ottenere la funzione di distribuzione di Y operando sulla funzione di distribuzione di X . Il punto centrale di questo passaggio consiste nella seconda uguaglianza che sposta l'attenzione dalla variabile Y alla variabile X , risolvendo così il problema di cui alla domanda. Le difficoltà nell'ottenere effettivamente la funzione di distribuzione (3.20) di Y nei casi specifici, ossia le difficoltà che si incontrano nel calcolo di $P\{X \in g^{-1}((-\infty, y])\}$, vanno affrontate di volta in volta con procedimenti ed accorgimenti particolari dipendenti dalla natura della variabile aleatoria X e dalle caratteristiche della funzione g .

Nel seguito sono analizzate alcune delle situazioni che ricorrono sovente nelle applicazioni.

¹Una funzione $g: \mathbb{R} \rightarrow \mathbb{R}$ è Borel-misurabile se per ogni sottoinsieme di Borel $B \in \mathcal{B}$ si ha $g^{-1}(B) = \{x \in \mathbb{R} : g(x) \in B\} \in \mathcal{B}$.

Teorema 3.2 Sia X una variabile aleatoria discreta $\{x_r, p_X(x_r); r = 1, 2, \dots\}$ e sia $g : \mathbb{R} \rightarrow \mathbb{R}$ una funzione Borel-misurabile. Se g è strettamente monotona, allora $Y = g(X)$ è una variabile aleatoria discreta $\{y_k, p_Y(y_k); k = 1, 2, \dots\}$ tale che

$$p_Y(y_k) = P(Y = y_k) = \begin{cases} P[X = h(y_k)], & y_k = g(x_k) \quad (k = 1, 2, \dots) \\ 0, & \text{altrimenti,} \end{cases}$$

dove h denota la funzione inversa di g .

Dimostrazione L'ipotesi fatta su g consente di affermare che tale funzione è invertibile, così che esiste una corrispondenza biunivoca tra i valori assunti da Y ed i valori assunti da X ; quindi per ogni $y_k = g(x_k)$ si ha $p_Y(y_k) = p_X(x_k)$. \square

Esempio 3.9 Sia X una variabile aleatoria discreta $\{k, p_k; k = 0, 1, \dots, n\}$ con

$$p_k = P(X = k) = \frac{2^k}{2^{n+1} - 1} \quad (k = 0, 1, \dots, n).$$

Determiniamo la funzione di probabilità della variabile aleatoria $Y = n - X$.

In tal caso $g(x) = n - x$ è una funzione strettamente monotona e la variabile aleatoria Y assume i valori $0, 1, \dots, n$. Pertanto dal Teorema 3.2 si ricava:

$$p_Y(y) = P(Y = y) = P(X = n - y) = \begin{cases} \frac{2^{n-y}}{2^{n+1} - 1}, & y = 0, 1, \dots, n \\ 0, & \text{altrimenti.} \end{cases}$$

◇

Teorema 3.3 Sia X una variabile aleatoria discreta $\{x_r, p_X(x_r); r = 1, 2, \dots\}$ e sia $g : \mathbb{R} \rightarrow \mathbb{R}$ una funzione Borel-misurabile. Si assuma inoltre che il dominio di g sia esprimibile come l'unione di intervalli disgiunti I_1, I_2, \dots, I_n e che g sia strettamente monotona nell'intervallo I_j ($j = 1, 2, \dots, n$). Se si denota con h_j la funzione inversa di g nell'intervallo I_j , allora $Y = g(X)$ è una variabile aleatoria discreta $\{y_k, p_Y(y_k); k = 1, 2, \dots\}$ tale che

$$p_Y(y_k) = P(Y = y_k) = \sum_{j=1}^n P[X = h_j(y_k)], \quad (3.21)$$

dove il termine j -esimo della somma è nullo se y_k non appartiene al dominio di h_j .

Dimostrazione L'ipotesi fatta sulla funzione g permette di affermare che in ogni intervallo I_j ($j = 1, 2, \dots, n$) tale funzione è invertibile così che esiste una corrispondenza biunivoca tra i valori assunti da Y ed i valori assunti da X . Quindi si ha:

$$p_Y(y_k) = P[g(X) = y_k] = \sum_{j=1}^n P[X = h_j(y_k)].$$

□

Esempio 3.10 Sia X una variabile aleatoria discreta che assume i valori $-1, 0, 1$ con rispettive probabilità $q, 1 - p - q, p$. Determiniamo la funzione di probabilità di $Y = X^2$, i cui possibili valori sono 0 e 1 .

Essendo $g(x) = x^2$, il dominio di g può essere rappresentato come l'unione dei due intervalli disgiunti $I_1 = (-\infty, 0]$ e $I_2 = (0, +\infty)$ in ognuno dei quali g è strettamente monotona. Dal Teorema 3.3 si ricava:

$$\begin{aligned} P(Y = 0) &= P(X^2 = 0) = P(X = 0) = 1 - p - q, \\ P(Y = 1) &= P(X^2 = 1) = P(X = -1) + P(X = 1) = p + q. \end{aligned}$$

Pertanto la desiderata funzione di probabilità è la seguente:

$$p_Y(y) = \begin{cases} 1 - p - q, & y = 0 \\ p + q, & y = 1 \\ 0, & \text{altrimenti.} \end{cases}$$

◇

Osserviamo che se X è una generica variabile aleatoria e se $Y = g(X)$, allora ad ogni valore assunto da X corrisponde un unico valore assunto da Y , mentre un valore assunto da Y può provenire da più valori assunti da X . Da ciò segue immediatamente che se X è discreta tale è anche Y .

Teorema 3.4 Sia X una variabile aleatoria assolutamente continua con densità di probabilità $f_X(x)$ e sia $g : \mathbb{R} \rightarrow \mathbb{R}$ una funzione strettamente monotona dotata di derivata prima continua. Denotata con h la funzione inversa di g e con D il suo dominio, $Y = g(X)$ è una variabile aleatoria assolutamente continua di densità di probabilità

$$f_Y(y) = \begin{cases} f_X[h(y)] \left| \frac{dh(y)}{dy} \right|, & y \in D \\ 0, & \text{altrimenti.} \end{cases} \quad (3.22)$$

Dimostrazione Poiché per ipotesi g è strettamente monotona, esiste una corrispondenza biunivoca tra i valori di X e di Y . Inoltre, poiché la derivata prima di g è continua, la sua inversa h è derivabile. Occorre distinguere due casi: (a) g è strettamente crescente, (b) g è strettamente decrescente.

(a) Se g è strettamente crescente si ha:

$$F_Y(y) = P(Y \leq y) = P[g(X) \leq y] = P[X \leq h(y)] = \int_{-\infty}^{h(y)} f_X(z) dz. \quad (3.23)$$

Effettuando il cambiamento di variabile $z = h(u)$, dalla (3.23) si ricava:

$$F_Y(y) = \int_{-\infty}^y f_X[h(u)] \frac{dh(u)}{du} du.$$

Si vede che $F_Y(y)$ è della forma (3.8), così che Y è una variabile aleatoria assolutamente continua con densità di probabilità:

$$f_Y(y) = f_X[h(y)] \frac{dh(y)}{dy}. \quad (3.24)$$

(b) Se g è strettamente decrescente si ha

$$F_Y(y) = P(Y \leq y) = P[g(X) \leq y] = P[X \geq h(y)] = \int_{h(y)}^{+\infty} f_X(z) dz$$

che, mediante il cambiamento di variabile $z = h(u)$, diviene:

$$F_Y(y) = \int_{-\infty}^y \left\{ -f_X[h(u)] \frac{dh(u)}{du} \right\} du.$$

Anche in questo caso $F_Y(y)$ è della forma (3.8), di modo che Y è una variabile aleatoria assolutamente continua con densità di probabilità:

$$f_Y(y) = -f_X[h(y)] \frac{dh(y)}{dy}. \quad (3.25)$$

Notiamo infine che è $dh(y)/dy > 0$ se g è strettamente crescente ed è $dh(y)/dy < 0$ se g è strettamente decrescente. Quindi dalle (3.24) e (3.25) segue la (3.22). \square

Esempio 3.11 Sia X un'arbitraria variabile aleatoria. Per determinare la funzione di distribuzione di $Y = aX + b$, con a e b reali, occorre distinguere i seguenti casi: (i) $a > 0$, (ii) $a = 0$ e (iii) $a < 0$.

(i) Per $a > 0$, si ha:

$$F_Y(y) = P(aX + b \leq y) = P\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right). \quad (3.26)$$

(ii) Se è $a = 0$, allora Y è degenere ed assume il valore b quasi certamente.

(iii) Per $a < 0$, infine, si ottiene:

$$F_Y(y) = P(aX + b \leq y) = P\left(X \geq \frac{y-b}{a}\right) = 1 - P\left(X < \frac{y-b}{a}\right) = 1 - F_X\left[\left(\frac{y-b}{a}\right)^-\right].$$

Esaminiamo ora il caso in cui X è assolutamente continua con $a \neq 0$. Per $a > 0$, operando il cambiamento di variabile $z = (t-b)/a$, dalla (3.26) si ottiene:

$$F_Y(y) = \int_{-\infty}^{(y-b)/a} f_X(z) dz = \int_{-\infty}^y \frac{1}{a} f_X\left(\frac{t-b}{a}\right) dt,$$

da cui segue:

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right).$$

Se, invece, è $a < 0$, attraverso il cambiamento di variabile $z = (t-b)/a$, si ricava:

$$F_Y(y) = 1 - \int_{-\infty}^{(y-b)/a} f_X(z) dz = 1 - \int_y^{+\infty} \frac{1}{|a|} f_X\left(\frac{t-b}{a}\right) dt = \int_{-\infty}^y \frac{1}{|a|} f_X\left(\frac{t-b}{a}\right) dt.$$

Pertanto se la variabile aleatoria X è assolutamente continua risulta:

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right) \quad (a \neq 0). \quad (3.27)$$

La (3.27) può essere anche ottenuta direttamente facendo uso del Teorema 3.4 dal momento che se è $a \neq 0$ la funzione $g(x) = ax + b$ è strettamente monotona e dotata di inversa $h(y) = (y - b)/a$. \diamond

Teorema 3.5 *Detta X una variabile aleatoria assolutamente continua di densità di probabilità $f_X(x)$, si consideri una funzione $g : \mathbb{R} \rightarrow \mathbb{R}$. Si supponga che il dominio di g consiste nell'unione di intervalli disgiunti I_1, I_2, \dots, I_n e che nell'intervallo I_j ($j = 1, 2, \dots, n$) la funzione g sia strettamente monotona e dotata di derivata prima continua tranne, al più, che negli estremi degli intervalli. Se si denota con $h_j(y)$ la funzione inversa di g nell'intervallo I_j ($j = 1, 2, \dots, n$) e con D_j il suo dominio, allora $Y = g(X)$ è una variabile aleatoria assolutamente continua di densità di probabilità data da*

$$f_Y(y) = \sum_{j=1}^n f_X[h_j(y)] \left| \frac{dh_j(y)}{dy} \right| \quad (3.28)$$

se y appartiene ad almeno uno dei domini D_j . Se invece y è esterno all'unione dei D_j , risulta $f_Y(y) = 0$.

Esempio 3.12 Consideriamo una variabile aleatoria X caratterizzata da funzione di distribuzione

$$F_X(x) = \begin{cases} 0, & x < -1 \\ \frac{(1+x)^2}{2}, & -1 \leq x < 0 \\ 1 - \frac{(1-x)^2}{2}, & 0 \leq x < 1 \\ 1, & x \geq 1. \end{cases} \quad (3.29)$$

Determiniamo la funzione di distribuzione e la densità di probabilità di $Y = X^2$.

Se è $y < 0$, allora $F_Y(y) = 0$, mentre per $y \geq 1$ si ha $F_Y(y) = 1$. Analizziamo ora il caso in cui risulta $0 \leq y < 1$. Evidentemente si ha:

$$F_Y(y) = P(X^2 \leq y) = P(-\sqrt{y} < X < \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}).$$

Essendo $0 \leq y < 1$, risulta $0 \leq \sqrt{y} < 1$ e $-1 < -\sqrt{y} \leq 0$; quindi dalla (3.29) si ricava:

$$F_Y(y) = 1 - \frac{(1 - \sqrt{y})^2}{2} - \frac{(1 + \sqrt{y})^2}{2} = 1 - (1 - \sqrt{y})^2, \quad 0 \leq y < 1.$$

Pertanto, in conclusione si ha:

$$F_Y(y) = \begin{cases} 0, & y < 0 \\ 1 - (1 - \sqrt{y})^2, & 0 \leq y < 1 \\ 1, & y \geq 1, \end{cases}$$

da cui segue la densità di probabilità:

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{y}} - 1, & 0 < y < 1 \\ 0, & \text{altrimenti.} \end{cases} \quad (3.30)$$

Tale risultato si può far discendere direttamente dal Teorema 3.5. Infatti, il dominio di $g(x) = x^2$ può essere rappresentato come l'unione dei due intervalli disgiunti $I_1 = (-\infty, 0)$ e $I_2 = [0, +\infty)$. La funzione inversa di g in I_1 è $h_1(y) = -\sqrt{y}$, mentre in I_2 essa risulta essere $h_2(y) = \sqrt{y}$. Pertanto, facendo uso della (3.28), se è $y \leq 0$ oppure $y \geq 1$ risulta $f_Y(y) = 0$, mentre per $0 < y < 1$ si ha:

$$f_Y(y) = \sum_{j=1}^2 f_X[h_j(y)] \left| \frac{dh_j(y)}{dy} \right| = \frac{1}{2\sqrt{y}} [f_X(-\sqrt{y}) + f_X(\sqrt{y})].$$

Poiché dalla (3.29) si trae

$$f_X(x) = \begin{cases} 1 - |x|, & -1 < x < 1 \\ 0, & \text{altrimenti,} \end{cases}$$

segue

$$f_Y(y) = \frac{1}{2\sqrt{y}} [f_X(-\sqrt{y}) + f_X(\sqrt{y})] = \frac{1}{2\sqrt{y}} (1 - |\sqrt{y}| + 1 - |\sqrt{y}|), \quad 0 < y < 1,$$

che conduce alla (3.30). \diamond

Concludiamo questo paragrafo osservando esplicitamente che mentre la stretta monotonia della funzione g non altera la natura della variabile aleatoria $Y = g(X)$ (nel senso che se X è discreta tale è anche Y , e se X è assolutamente continua tale è anche Y) la stessa cosa non accade se si rimuove l'ipotesi di stretta monotonia di g . Invero, se g non è strettamente monotona e se X è una variabile aleatoria assolutamente continua, $Y = g(X)$ non è necessariamente tale, come evidenziato nell'esempio seguente.

Esempio 3.13 Sia X una variabile aleatoria assolutamente continua di funzione di distribuzione $F_X(x)$ e sia $\{x_1, x_2, \dots\}$ una successione strettamente crescente di numeri reali tale che risulti $\lim_{n \rightarrow +\infty} F_X(x_n) = 1$. Si consideri poi la seguente trasformazione $g: \mathbb{R} \rightarrow \mathbb{R}$:

$$Y = g(X) = \begin{cases} y_1, & X \leq x_1 \\ y_2, & x_1 < X \leq x_2 \\ \dots & \dots\dots\dots \\ y_k, & x_{k-1} < X \leq x_k \\ \dots & \dots\dots\dots \end{cases} \quad (3.31)$$

con $y_1 < y_2 < \dots$. Va esplicitamente notato che g è una funzione definita in \mathbb{R} ed assumente i soli valori y_1, y_2, \dots , così che la variabile aleatoria Y assume valori soltanto nell'insieme

$\{y_1, y_2, \dots\}$. La trasformazione (3.31) non soddisfa le ipotesi del Teorema 3.5. Infatti, poiché g è costante a tratti, il suo dominio non è rappresentabile mediante unioni di intervalli disgiunti in ciascuno dei quali g risulti strettamente monotona. Essendo

$$\begin{aligned} P(Y = y_1) &= P(X \leq x_1) = F_X(x_1), \\ P(Y = y_k) &= P(x_{k-1} < X \leq x_k) = F_X(x_k) - F_X(x_{k-1}) \quad (k = 2, 3, \dots), \end{aligned}$$

si nota immediatamente che risulta:

$$\begin{aligned} P(Y = y_k) &\geq 0 \quad (k = 1, 2, \dots), \\ \sum_{k=1}^{+\infty} P(Y = y_k) &= \lim_{n \rightarrow +\infty} \sum_{k=1}^n P(Y = y_k) = \lim_{n \rightarrow +\infty} F_X(x_n) = 1. \end{aligned}$$

Ciò indica che Y è una variabile aleatoria discreta di funzione di distribuzione:

$$F_Y(y) = P(Y \leq y) = \begin{cases} 0, & y < y_1 \\ F_X(x_1), & y_1 \leq y < y_2 \\ \dots\dots\dots & \dots\dots\dots \\ F_X(x_k), & y_k \leq y < y_{k+1} \\ \dots\dots\dots & \dots\dots\dots \end{cases}$$

La trasformazione (3.31) si rivela particolarmente efficace per la simulazione di variabili aleatorie discrete a partire da variabili aleatorie assolutamente continue. \diamond

3.5 Vettori aleatori

Talora ad ogni risultato di un esperimento casuale è necessario associare più numeri reali onde descriverne molteplici caratteristiche. Ciò si effettua ricorrendo a più variabili aleatorie tutte definite nello stesso spazio di probabilità, collettivamente denotate quale *variabile aleatoria multidimensionale* o *vettore aleatorio*.

Ad esempio, se lo spazio campione Ω è costituito dagli individui di un'assegnata popolazione, l'altezza ed il peso di un individuo ivi scelto a caso possono essere descritti tramite una variabile aleatoria bidimensionale $\mathbf{X} = (X_1, X_2)$, con X_1 e X_2 variabili aleatorie unidimensionali denotanti, rispettivamente, altezza e peso. Analogamente, nel caso di un tiro al bersaglio il risultato potrebbe essere rappresentato mediante una coppia (X_1, X_2) di coordinate cartesiane ortogonali individuanti il punto colpito piuttosto che mediante la sua distanza dal centro del bersaglio, ovvero ricorrendo ad una coppia (R, Φ) individuante detta posizione in un sistema di coordinate polari con origine nel centro del bersaglio.

Sia (Ω, \mathcal{F}, P) lo spazio di probabilità associato ad un esperimento casuale e siano X_1, X_2, \dots, X_n variabili aleatorie unidimensionali ivi definite. Il vettore aleatorio n -dimensionale $\mathbf{X} = (X_1, X_2, \dots, X_n)$ associa allora ad ogni $\omega \in \Omega$ la n -upla di reali $(X_1(\omega), X_2(\omega), \dots, X_n(\omega))$. Un vettore aleatorio n -dimensionale è dunque semplicemente una funzione $\mathbf{X}: \Omega \rightarrow \mathbb{R}^n$. Come nel caso unidimensionale, questa deve soddisfare opportuni requisiti atti a consentire l'individuazione di appropriate relazioni tra certi sottoinsiemi di \mathbb{R}^n e gli elementi di \mathcal{F} , ossia gli eventi, le cui probabilità sono assunte specificate tramite la misura di probabilità P .

I sottoinsiemi numerici che rivestono importanza, ai quali risulta utile estendere la nozione di probabilità, sono gli insiemi della classe di Borel \mathcal{B} su \mathbb{R}^n , ossia gli elementi della minima σ -algebra contenente gli intervalli n -dimensionali².

Quanto finora esposto conduce alla seguente

Definizione 3.7 Sia assegnato uno spazio di probabilità (Ω, \mathcal{F}, P) . Un vettore aleatorio n -dimensionale, o variabile aleatoria n -dimensionale, $\mathbf{X} = (X_1, X_2, \dots, X_n)$ è una funzione $\mathbf{X}: \Omega \rightarrow \mathbb{R}^n$ tale che per ogni sottoinsieme di Borel $B \in \mathcal{B}$ su \mathbb{R}^n risulti $\{\omega \in \Omega: (X_1(\omega), X_2(\omega), \dots, X_n(\omega)) \in B\} \in \mathcal{F}$.

Poiché gli intervalli n -dimensionali $\{\mathbf{z} = (z_1, z_2, \dots, z_n) \in \mathbb{R}^n : -\infty < z_i \leq x_i, i = 1, 2, \dots, n\}$ sono elementi generatori della classe di Borel \mathcal{B} su \mathbb{R}^n , in alternativa è possibile dare la seguente

Definizione 3.8 Una variabile aleatoria n -dimensionale $\mathbf{X} = (X_1, X_2, \dots, X_n)$ è una funzione $\mathbf{X}: \Omega \rightarrow \mathbb{R}^n$ a valori in \mathbb{R}^n definita sullo spazio campione Ω ed ivi misurabile, ossia tale che

$$\{\omega \in \Omega: X_1(\omega) \leq x_1, X_2(\omega) \leq x_2, \dots, X_n(\omega) \leq x_n\} = \bigcap_{i=1}^n \{\omega \in \Omega: X_i(\omega) \leq x_i\} \in \mathcal{F} \quad (3.32)$$

per ogni n -upla $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$.

Esempio 3.14 Si consideri l'esperimento consistente nel lancio ripetuto due volte di una moneta con spazio campione $\Omega = \{CC, CT, TC, TT\}$ e sia \mathcal{F} la σ -algebra costituita dall'insieme potenza di Ω . Introduciamo una funzione $\mathbf{X} = (X_1, X_2): \Omega \rightarrow \mathbb{R}^2$, dove X_1 descrive il numero di T nei due lanci e X_2 descrive il numero di variazioni del risultato nei due lanci. Pertanto \mathbf{X} è definita tramite le posizioni

$$\mathbf{X}(CC) = (0, 0), \quad \mathbf{X}(CT) = (1, 1), \quad \mathbf{X}(TC) = (1, 1), \quad \mathbf{X}(TT) = (2, 0)$$

e quindi $\{\omega \in \Omega: X_1(\omega) = 0, X_2(\omega) = 0\} = \{CC\}$, $\{\omega \in \Omega: X_1(\omega) = 1, X_2(\omega) = 1\} = \{CT, TC\}$ e $\{\omega \in \Omega: X_1(\omega) = 2, X_2(\omega) = 0\} = \{TT\}$. Poiché risulta

$$\{\omega \in \Omega: X_1(\omega) \leq x_1, X_2(\omega) \leq x_2\} = \begin{cases} \emptyset, & (x_1 < 0, x_2 \in \mathbb{R}) \text{ oppure } (x_1 \in \mathbb{R}, x_2 < 0) \\ \{CC\}, & (0 \leq x_1 < 1, x_2 \geq 0) \text{ oppure } (1 \leq x_1 < 2, 0 \leq x_2 < 1) \\ \{CC, CT, TC\}, & 1 \leq x_1 < 2, x_2 \geq 1 \\ \{CC, TT\}, & x_1 \geq 2, 0 \leq x_2 < 1 \\ \Omega, & x_1 \geq 2, x_2 \geq 1, \end{cases}$$

si ha che $\{\omega \in \Omega: X_1(\omega) \leq x_1, X_2(\omega) \leq x_2\} \in \mathcal{F}$ per ogni $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$, così che, in virtù della Definizione 3.8, \mathbf{X} è un vettore aleatorio bidimensionale. \diamond

²Siano $\mathbf{a} = (a_1, a_2, \dots, a_n)$ e $\mathbf{b} = (b_1, b_2, \dots, b_n)$ n -uple di numeri reali tali che $a_i < b_i$, $i = 1, 2, \dots, n$ (si può anche eventualmente avere $a_i = -\infty$ oppure $b_i = +\infty$). L'intervallo n -dimensionale semichiuso a destra $(\mathbf{a}, \mathbf{b}]$ è definito al seguente modo: $(\mathbf{a}, \mathbf{b}] = \{\mathbf{z} = (z_1, z_2, \dots, z_n) \in \mathbb{R}^n : a_i < z_i \leq b_i, i = 1, 2, \dots, n\}$. Gli intervalli aperti, chiusi e semichiusi a sinistra si definiscono in modo analogo.

3.6 Funzione di distribuzione congiunta

Uno strumento essenziale per lo studio dei vettori aleatori è la *funzione di distribuzione congiunta*, detta anche *funzione di ripartizione congiunta*.

Definizione 3.9 Sia $\mathbf{X} = (X_1, X_2, \dots, X_n)$ un vettore aleatorio $\mathbf{X}: \Omega \rightarrow \mathbb{R}^n$. La funzione

$$\begin{aligned} F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) &= P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) \\ &= P(\{\omega \in \Omega: X_1(\omega) \leq x_1, X_2(\omega) \leq x_2, \dots, X_n(\omega) \leq x_n\}), \end{aligned} \quad (3.33)$$

con $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$, prende il nome di *funzione di distribuzione del vettore aleatorio \mathbf{X}* , o *funzione di distribuzione congiunta delle variabili aleatorie X_1, X_2, \dots, X_n* .

Si osservi che, come nella (3.32), la notazione “ $X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n$ ” nella (3.33) è un’abbreviazione della scrittura $\{\omega \in \Omega: X_1(\omega) \leq x_1\} \cap \{\omega \in \Omega: X_2(\omega) \leq x_2\} \cap \dots \cap \{\omega \in \Omega: X_n(\omega) \leq x_n\}$. La funzione di distribuzione congiunta gode delle proprietà di cui al seguente teorema, che peraltro ci limitiamo ad enunciare.

Teorema 3.6 La funzione di distribuzione $F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$ di un vettore aleatorio $\mathbf{X} = (X_1, X_2, \dots, X_n): \Omega \rightarrow \mathbb{R}^n$ gode delle seguenti proprietà:

- (i) $F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$ è non decrescente;³
- (ii) $F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$ è continua a destra rispetto a ciascun argomento, ossia per ogni k ($k = 1, 2, \dots, n$) si ha:

$$\lim_{\varepsilon \downarrow 0} F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_k + \varepsilon, \dots, x_n) = F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n),$$

con x_1, x_2, \dots, x_n reali arbitrari;

- (iii) per ogni k ($k = 1, 2, \dots, n$) si ha:

$$\lim_{x_k \rightarrow -\infty} F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_k, \dots, x_n) = 0,$$

con $x_1, x_2, \dots, x_{k-1}, x_{k+1}, \dots, x_n$ reali arbitrari;

- (iv) $\lim_{\substack{x_1 \rightarrow +\infty \\ x_n \rightarrow +\infty}} F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = 1.$

Si osservi che le proprietà (iii) e (iv) trovano una semplice giustificazione. Infatti, se fissato un indice k ($k = 1, 2, \dots, n$) si fa tendere x_k a $-\infty$, il corrispondente insieme $\{\omega \in \Omega: X_k(\omega) \leq x_k\}$ tende all’evento impossibile ed a questo tende, quindi, anche l’intersezione al secondo membro della (3.32), restando così verificata la (iii). Invece, se per ogni k ($k = 1, 2, \dots, n$), si fa tendere x_k a $+\infty$, i corrispondenti eventi $\{\omega \in \Omega: X_k(\omega) \leq x_k\}$

³Una funzione $F(x_1, x_2, \dots, x_n)$ si dice non decrescente se comunque si scelgano n reali positivi $\vartheta_1, \vartheta_2, \dots, \vartheta_n$, per ogni n -upla $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ risulta $\Delta_{x_1}^{\vartheta_1} \Delta_{x_2}^{\vartheta_2} \dots \Delta_{x_n}^{\vartheta_n} F \geq 0$, dove $\Delta_{x_k}^{\vartheta_k} F := F(x_1, \dots, x_{k-1}, x_k + \vartheta_k, x_{k+1}, \dots, x_n) - F(x_1, x_2, \dots, x_n)$.

convergono tutti all'evento certo ed a questo tende, quindi, anche l'intersezione al secondo membro della (3.32), verificando così la (iv).

Si noti, inoltre, che nel caso di un vettore aleatorio bidimensionale la proprietà (i) equivale ad affermare che per ogni coppia $(x_1, x_2) \in \mathbb{R}^2$ si ha $\Delta_{x_1}^{\vartheta_1} \Delta_{x_2}^{\vartheta_2} F_{X_1, X_2}(x_1, x_2) \geq 0$ comunque si scelgano $\vartheta_1 > 0$ e $\vartheta_2 > 0$. Infatti, risulta:

$$\begin{aligned} \Delta_{x_1}^{\vartheta_1} \Delta_{x_2}^{\vartheta_2} F_{X_1, X_2}(x_1, x_2) &= \Delta_{x_1}^{\vartheta_1} \left[F_{X_1, X_2}(x_1, x_2 + \vartheta_2) - F_{X_1, X_2}(x_1, x_2) \right] \\ &= \left[F_{X_1, X_2}(x_1 + \vartheta_1, x_2 + \vartheta_2) - F_{X_1, X_2}(x_1 + \vartheta_1, x_2) \right] \\ &\quad - \left[F_{X_1, X_2}(x_1, x_2 + \vartheta_2) - F_{X_1, X_2}(x_1, x_2) \right] \\ &= P(X_1 \leq x_1 + \vartheta_1, x_2 < X_2 \leq x_2 + \vartheta_2) - P(X_1 \leq x_1, x_2 < X_2 \leq x_2 + \vartheta_2) \\ &= P(x_1 < X_1 \leq x_1 + \vartheta_1, x_2 < X_2 \leq x_2 + \vartheta_2) \geq 0. \end{aligned}$$

Esempio 3.15 Si consideri la funzione

$$F(x, y) = \begin{cases} 1, & x + y \geq 0 \\ 0, & \text{altrimenti.} \end{cases}$$

È agevole verificare che $F(x, y)$ soddisfa le condizioni (ii), (iii) e (iv). Ciò nonostante, sebbene $F(x, y)$ sia non decrescente sia rispetto a x sia rispetto a y , essa non può riguardarsi come una funzione di distribuzione congiunta in quanto non soddisfa la condizione (i), avendosi ad esempio

$$\Delta_{x_1}^2 \Delta_{x_2}^2 F_{X_1, X_2}(-1, -1) = F(1, 1) - F(1, -1) - F(-1, 1) + F(-1, -1) = -1 < 0.$$

◇

Proposizione 3.4 Sia $\mathbf{X} = (X_1, X_2, \dots, X_n)$ un vettore aleatorio e sia $F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$ la sua funzione di distribuzione. La funzione di distribuzione congiunta delle variabili aleatorie X_1, X_2, \dots, X_m ($m < n$), detta funzione di distribuzione marginale, è ottenibile al seguente modo:

$$F_{X_1, X_2, \dots, X_m}(x_1, x_2, \dots, x_m) = \lim_{\substack{x_{m+1} \rightarrow +\infty \\ \vdots \\ x_n \rightarrow +\infty}} F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_m, x_{m+1}, \dots, x_n). \quad (3.34)$$

Dimostrazione Se $x_k \rightarrow +\infty$ per ogni k ($k = m+1, m+2, \dots, n$), i corrispondenti eventi $\{\omega \in \Omega: X_k(\omega) \leq x_k\}$ tendono tutti all'evento certo. Quindi, per ogni m -upla $(x_1, x_2, \dots, x_m) \in \mathbb{R}^m$ segue che $\bigcap_{i=1}^n \{\omega \in \Omega: X_i(\omega) \leq x_i\}$ tende a $\bigcap_{i=1}^m \{\omega \in \Omega: X_i(\omega) \leq x_i\}$, che per la (3.32) coincide con $\{\omega \in \Omega: X_1(\omega) \leq x_1, X_2(\omega) \leq x_2, \dots, X_m(\omega) \leq x_m\}$. Ricordando la (3.33) segue immediatamente la (3.34). □

La Proposizione 3.4 è suscettibile di immediata estensione. Invero, se nella (3.33) si fanno tendere $n-m$ delle variabili x_1, x_2, \dots, x_n a $+\infty$, si ottiene la funzione di distribuzione marginale del vettore aleatorio m -dimensionale costituito dalle rimanenti m variabili aleatorie.

Se ne trae che la conoscenza della funzione di distribuzione $F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$ permette di ricavare le funzioni di distribuzione delle singole variabili aleatorie oltre che di tutte le loro coppie, triple, ecc. Il viceversa in generale non sussiste: con l'eccezione di casi particolari, la funzione di distribuzione congiunta non può determinarsi a partire dalle funzioni di distribuzione marginali.

3.7 Classificazione dei vettori aleatori

Analogamente al caso delle variabili aleatorie unidimensionali, è conveniente considerare separatamente i vettori aleatori discreti e quelli assolutamente continui. Per semplicità di trattazione, e per non appesantire sin dall'inizio eccessivamente le notazioni, fisseremo in prima istanza l'attenzione su vettori bidimensionali, destinando il Paragrafo 3.7.3 al caso generale. A tal fine, osserviamo anzitutto che se (X, Y) è una variabile aleatoria bidimensionale, la (3.33) assume la forma:

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y) \quad \forall (x, y) \in \mathbb{R}^2. \quad (3.35)$$

Dalla (iii) e (iv) del Teorema 3.6 risulta poi:

$$\lim_{x \rightarrow -\infty} F_{X,Y}(x, y) = \lim_{y \rightarrow -\infty} F_{X,Y}(x, y) = 0, \quad \lim_{\substack{x \rightarrow +\infty \\ y \rightarrow +\infty}} F_{X,Y}(x, y) = 1. \quad (3.36)$$

Inoltre, dalla (3.34) si ha:

$$F_X(x) = \lim_{y \rightarrow +\infty} F_{X,Y}(x, y) \quad \forall x \in \mathbb{R}, \quad F_Y(y) = \lim_{x \rightarrow +\infty} F_{X,Y}(x, y) \quad \forall y \in \mathbb{R}. \quad (3.37)$$

3.7.1 Vettori aleatori bidimensionali discreti

Definizione 3.10 Un vettore aleatorio (X, Y) è detto discreto se esiste un insieme $S = S_1 \times S_2$ al più numerabile di coppie distinte di \mathbb{R}^2 , con $S_1 = \{x_1, x_2, \dots\}$ e $S_2 = \{y_1, y_2, \dots\}$, tale che $P[(X, Y) \in S] = 1$.

Poiché

$$\{\omega \in \Omega: (X(\omega), Y(\omega)) \in S\} = \bigcup_{\{i: x_i \in S_1\}} \bigcup_{\{j: y_j \in S_2\}} \{\omega \in \Omega: X_1(\omega) = x_i, Y(\omega) = y_j\},$$

per l'incompatibilità degli eventi dei quali si effettuano queste unioni, segue:

$$\sum_{\{i: x_i \in S_1\}} \sum_{\{j: y_j \in S_2\}} P(X = x_i, Y = y_j) = P\{(X, Y) \in S\} = 1.$$

Un vettore aleatorio discreto (X, Y) è quindi completamente individuato dall'insieme $\{(x_i, y_j), p_{i,j}; i, j = 1, 2, \dots\}$, dove (x_i, y_j) sono le coppie di valori assunti dalle variabili aleatorie X, Y e dove si è posto $p_{i,j} = P(X = x_i, Y = y_j)$.

Al vettore aleatorio (X, Y) è possibile associare una funzione di probabilità $p_{X,Y}: \mathbb{R}^2 \rightarrow [0, 1]$ così definita:

$$p_{X,Y}(x, y) = \begin{cases} p_{i,j}, & x = x_i, y = y_j \quad (i, j = 1, 2, \dots) \\ 0, & \text{altrimenti.} \end{cases} \quad (3.38)$$

Ad essa si dà il nome di *funzione di probabilità congiunta*.

Per variabili aleatorie discrete la conoscenza della funzione di probabilità congiunta consente di calcolare probabilità del tipo

$$P\{(X, Y) \in D\} = \sum_{(x,y) \in D} p_{X,Y}(x, y) \quad (3.39)$$

per ogni assegnata regione D di \mathbb{R}^2 . Dalla (3.39) segue che la funzione di distribuzione congiunta del vettore aleatorio (X, Y) può essere così espressa:

$$F_{X,Y}(x, y) = \sum_{\{i: x_i \leq x\}} \sum_{\{j: y_j \leq y\}} p_{i,j} \quad \forall (x, y) \in \mathbb{R}^2. \quad (3.40)$$

Nel caso di un vettore aleatorio (X, Y) discreto è possibile determinare la funzione di probabilità di una qualsiasi delle due variabili aleatorie componenti a partire dalla funzione di probabilità congiunta. Infatti, essendo

$$\{X = x\} = \bigcup_{\{j: y_j \in S_2\}} \{X = x, Y = y_j\}, \quad \{Y = y\} = \bigcup_{\{i: x_i \in S_1\}} \{X = x_i, Y = y\},$$

con gli eventi delle singole unioni incompatibili, si ricava:

$$p_X(x) = \sum_{\{j: y_j \in S_2\}} p_{X,Y}(x, y_j) \quad \forall x \in \mathbb{R}, \quad p_Y(y) = \sum_{\{i: x_i \in S_1\}} p_{X,Y}(x_i, y) \quad \forall y \in \mathbb{R}. \quad (3.41)$$

Queste funzioni prendono il nome di *funzioni di probabilità marginali*.

La (3.41) mostra che la conoscenza della funzione di probabilità $p_{X,Y}(x, y)$ permette di determinare le funzioni di probabilità marginali relative alle singole variabili aleatorie X e Y . Occorre sottolineare che il viceversa non sussiste, nel senso che la funzione di probabilità congiunta non può essere determinata a partire dalle funzioni di probabilità marginali.

Esempio 3.16 Riconsideriamo l'esperimento di cui all'Esempio 3.14 consistente nel lanciare due volte una moneta e denotiamo con X la variabile aleatoria che rappresenta il numero di volte in cui esce testa nei due lanci e con Y la variabile aleatoria che indica il numero di variazioni del risultato nei due lanci. (Con riferimento alla notazione dell'Esempio 3.14 è stato ora posto $X = X_1$ e $Y = X_2$). La funzione di probabilità del vettore aleatorio discreto (X, Y) è:

$$p_{X,Y}(x, y) = \begin{cases} 1/4, & (x = 0, y = 0) \text{ oppure } (x = 2, y = 0) \\ 1/2, & x = 1, y = 1 \\ 0, & \text{altrimenti.} \end{cases}$$

Facendo uso della relazione (3.41) è possibile ottenere le funzioni di probabilità marginali relative alle singole variabili aleatorie X e Y :

$$p_X(x) = \begin{cases} 1/4, & x = 0 \\ 1/2, & x = 1 \\ 1/4, & x = 2 \\ 0, & \text{altrimenti;} \end{cases} \quad p_Y(y) = \begin{cases} 1/2, & y = 0 \\ 1/2, & y = 1 \\ 0, & \text{altrimenti.} \end{cases}$$

Inoltre, dalla (3.40) si può determinare la funzione di distribuzione congiunta

$$F_{X,Y}(x, y) = \begin{cases} 0, & (x < 0, y \in \mathbb{R}) \text{ oppure } (x \in \mathbb{R}, y < 0) \\ 1/4, & (0 \leq x < 1, y \geq 0) \text{ oppure } (1 \leq x < 2, 0 \leq y < 1) \\ 3/4, & 1 \leq x < 2, y \geq 1 \\ 1/2, & x \geq 2, 0 \leq y < 1 \\ 1, & x \geq 2, y \geq 1, \end{cases}$$

da cui, facendo uso della (3.37), si ricava:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1/4, & 0 \leq x < 1 \\ 3/4, & 1 \leq x < 2 \\ 1, & x \geq 2, \end{cases} \quad F_Y(y) = \begin{cases} 0, & y < 0 \\ 1/2, & 0 \leq y < 1 \\ 1, & y \geq 2. \end{cases}$$

◇

3.7.2 Vettori aleatori bidimensionali assolutamente continui

Definizione 3.11 Un vettore aleatorio (X, Y) si dice assolutamente continuo se esiste una funzione non negativa $f_{X,Y}(x, y)$ tale che per ogni coppia $(x, y) \in \mathbb{R}^2$ la sua funzione di distribuzione $F_{X,Y}(x, y)$ può esprimersi nella seguente forma:

$$F_{X,Y}(x, y) = \int_{-\infty}^x du \int_{-\infty}^y f_{X,Y}(u, v) dv. \quad (3.42)$$

La funzione $f_{X,Y}(x, y)$ prende il nome *densità di probabilità congiunta* delle variabili aleatorie X e Y .

Per ogni coppia di valori (x, y) in cui la funzione di distribuzione ammette derivate seconde continue, dalla (3.42) segue:

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y). \quad (3.43)$$

La densità di probabilità congiunta di una variabile aleatoria bidimensionale (X, Y) soddisfa le proprietà:

$$f_{X,Y}(x, y) \geq 0, \quad \forall (x, y) \in \mathbb{R}^2, \quad \int_{-\infty}^{+\infty} du \int_{-\infty}^{+\infty} f_{X,Y}(u, v) dv = 1. \quad (3.44)$$

Si noti che per variabili aleatorie bidimensionali assolutamente continue la conoscenza della densità di probabilità congiunta consente di calcolare probabilità del tipo

$$P\{(X, Y) \in D\} = \int_D f_{X,Y}(u, v) du dv \quad (3.45)$$

per ogni assegnata regione D di \mathbb{R}^2 . Le difficoltà nell'ottenere tali probabilità sono evidentemente legate al calcolo effettivo degli integrali coinvolti e vanno affrontate di volta in volta con procedimenti ed accorgimenti particolari.

Nel caso di un vettore aleatorio bidimensionale (X, Y) assolutamente continuo, è possibile determinare la densità di probabilità di ognuna delle variabili aleatorie componenti. In virtù delle (3.37) e della (3.42) si ha infatti:

$$F_X(x) = \int_{-\infty}^x \left[\int_{-\infty}^{+\infty} f_{X,Y}(u, v) dv \right] du \quad \forall x \in \mathbb{R} \quad (3.46)$$

$$F_Y(y) = \int_{-\infty}^y \left[\int_{-\infty}^{+\infty} f_{X,Y}(u, v) du \right] dv \quad \forall y \in \mathbb{R}.$$

La (3.46) mostra che $F_X(x)$ e $F_Y(y)$ sono esprimibili nella forma integrale (3.42); pertanto X e Y sono entrambe assolutamente continue con rispettive densità di probabilità

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, v) dv \quad \forall x \in \mathbb{R} \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(u, y) du \quad \forall y \in \mathbb{R}. \quad (3.47)$$

Le funzioni (3.47) prendono il nome di *densità di probabilità marginali*.

La (3.47) mostra che la conoscenza della densità di probabilità congiunta $f_{X,Y}(x, y)$ permette di determinare le densità di probabilità marginali relative alle singole variabili aleatorie, mentre il viceversa non è in generale vero: la densità di probabilità congiunta non può essere determinata a partire dalle densità di probabilità marginali.

Esempio 3.17 Sia (X, Y) un vettore aleatorio assolutamente continuo di densità di probabilità congiunta:

$$f_{X,Y}(x, y) = \begin{cases} 2, & 0 < x \leq y < 1 \\ 0, & \text{altrimenti.} \end{cases}$$

Questa soddisfa le proprietà (3.44). Infatti, risulta $f_{X,Y}(x, y) \geq 0$ per ogni $(x, y) \in \mathbb{R}^2$ ed inoltre si ha:

$$\int_{-\infty}^{+\infty} du \int_{-\infty}^{+\infty} f_{X,Y}(u, v) dv = 2 \int_0^1 du \int_u^1 dv = 2 \int_0^1 (1 - u) du = 1.$$

In Figura 3.3 è rappresentata la funzione $f_{X,Y}(x, y)$. Dalla (3.47) segue:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, v) dv = \begin{cases} 2(1 - x), & 0 < x < 1 \\ 0, & \text{altrimenti;} \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(u, y) du = \begin{cases} 2y, & 0 < y < 1 \\ 0, & \text{altrimenti.} \end{cases}$$

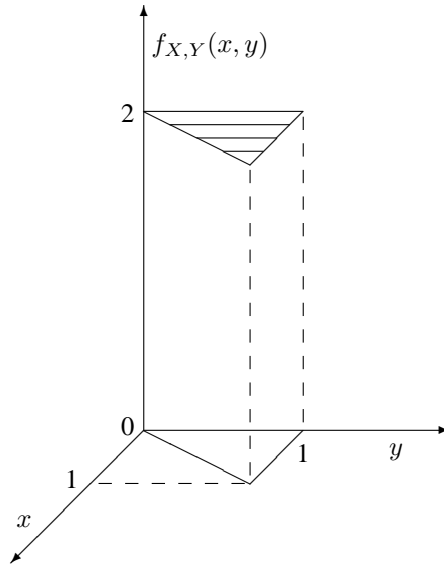


Figura 3.3 – La densità di probabilità congiunta di cui all'Esempio 3.17.

◇

Il vettore aleatorio (X, Y) di componenti assolutamente continue non è necessariamente assolutamente continuo come evidenziato nel seguente esempio.

Esempio 3.18 Sia $X: \Omega \rightarrow \mathbb{R}$ una variabile aleatoria assolutamente continua di densità di probabilità $f_X(x)$ e sia $Y = X$, ossia $Y(\omega) = X(\omega)$ per ogni $\omega \in \Omega$. Ovviamente anche Y è una variabile aleatoria assolutamente continua. Notiamo in primo luogo che se si pone $D = \{(x, y) \in \mathbb{R}^2: x = y\}$ si deve necessariamente avere $P\{(X, Y) \in D\} = 1$. Vogliamo mostrare che il vettore aleatorio (X, Y) non è assolutamente continuo. Ragioniamo per assurdo e supponiamo che (X, Y) sia assolutamente continuo con densità di probabilità $f_{X,Y}(x, y)$. Dovrebbe allora risultare:

$$1 = P\{(X, Y) \in D\} = \int_D f_{X,Y}(u, v) du dv.$$

Poiché D ha area nulla, l'integrale è nullo, il che conduce ad una contraddizione. Se ne conclude che (X, Y) non è un vettore aleatorio assolutamente continuo. ◇

3.7.3 Vettori aleatori multidimensionali

Esaminiamo come si generalizzano le relazioni determinate nei Paragrafi 3.7.1 e 3.7.2 nel caso $n > 2$.

Definizione 3.12 Un vettore aleatorio $\mathbf{X} = (X_1, X_2, \dots, X_n)$ è detto discreto se esiste un insieme $S = S_1 \times S_2 \times \dots \times S_n$ al più numerabile di n -uple distinte di \mathbb{R}^n , con $S_k = \{x_1^{(k)}, x_2^{(k)}, \dots\}$ per $k = 1, 2, \dots, n$, tale che $P(\mathbf{X} \in S) = 1$.

Dalla Definizione 3.12 segue:

$$\sum_{\{r_1: x_{r_1}^{(1)} \in S_1\}} \dots \sum_{\{r_n: x_{r_n}^{(n)} \in S_n\}} P(X_1 = x_{r_1}^{(1)}, \dots, X_n = x_{r_n}^{(n)}) = P(\mathbf{X} \in S) = 1.$$

Un vettore aleatorio discreto $\mathbf{X} = (X_1, X_2, \dots, X_n)$ è quindi completamente individuato dall'insieme $\{(x_{r_1}^{(1)}, x_{r_2}^{(2)}, \dots, x_{r_n}^{(n)}), p_{r_1, r_2, \dots, r_n}\}$, dove $(x_{r_1}^{(1)}, x_{r_2}^{(2)}, \dots, x_{r_n}^{(n)})$ sono le n -uple di valori assunti dalle variabili aleatorie X_1, X_2, \dots, X_n e dove si è posto $p_{r_1, r_2, \dots, r_n} = P(X_1 = x_{r_1}^{(1)}, X_2 = x_{r_2}^{(2)}, \dots, X_n = x_{r_n}^{(n)})$.

Al vettore aleatorio $\mathbf{X} = (X_1, X_2, \dots, X_n)$ è possibile associare la funzione di probabilità $p_{X_1, X_2, \dots, X_n}: \mathbb{R}^n \rightarrow [0, 1]$ così definita:

$$p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = \begin{cases} p_{r_1, r_2, \dots, r_n}, & x_1 = x_{r_1}^{(1)}, x_2 = x_{r_2}^{(2)}, \dots, x_n = x_{r_n}^{(n)} \\ & (r_1, r_2, \dots, r_n = 1, 2, \dots) \\ 0, & \text{altrimenti.} \end{cases} \quad (3.48)$$

Ad essa si dà il nome di *funzione di probabilità congiunta*.

Per variabili aleatorie discrete la conoscenza della funzione di probabilità consente di calcolare, almeno in linea di principio, probabilità del tipo

$$P\{(X_1, X_2, \dots, X_n) \in D\} = \sum_{(x_1, x_2, \dots, x_n) \in D} p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) \quad (3.49)$$

per ogni assegnata regione D di \mathbb{R}^n . Dalla (3.49) segue che per ogni $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ la funzione di distribuzione congiunta di X_1, X_2, \dots, X_n può essere così espressa:

$$F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = \sum_{\{r_1: x_{r_1}^{(1)} \leq x_1\}} \sum_{\{r_2: x_{r_2}^{(2)} \leq x_2\}} \dots \sum_{\{r_n: x_{r_n}^{(n)} \leq x_n\}} p_{r_1, r_2, \dots, r_n}. \quad (3.50)$$

Nel caso di una variabile aleatoria n -dimensionale (X_1, X_2, \dots, X_n) discreta, è possibile esprimere la funzione di probabilità di un qualsiasi numero m ($m < n$) di variabili aleatorie. Ad esempio, se si considerano le prime m , risulta:

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2, \dots, X_m = x_m) \\ = \sum_{\{x_{m+1} \in S_{m+1}\}} \dots \sum_{\{x_n \in S_n\}} P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n). \end{aligned} \quad (3.51)$$

A questa si dà il nome di *funzione di probabilità marginale*.

Nel caso di vettori aleatori discreti la (3.51) mostra che la conoscenza della funzione di probabilità congiunta $p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$ permette di determinare tutte le funzioni di probabilità marginali relative a coppie, triple, $(n-1)$ -uple di variabili aleatorie scelte tra le n variabili aleatorie X_1, X_2, \dots, X_n , oltre che le funzioni di probabilità delle singole variabili aleatorie.

Definizione 3.13 Un vettore aleatorio $\mathbf{X} = (X_1, X_2, \dots, X_n)$ si dice assolutamente continuo se esiste una funzione non negativa $f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$ tale che per ogni $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ la sua funzione di distribuzione $F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$ possa esprimersi nella forma:

$$F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} dz_1 \int_{-\infty}^{x_2} dz_2 \dots \int_{-\infty}^{x_n} f_{X_1, X_2, \dots, X_n}(z_1, z_2, \dots, z_n) dz_n. \quad (3.52)$$

La funzione $f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$ prende il nome *densità di probabilità congiunta* delle variabili aleatorie X_1, X_2, \dots, X_n .

Per ogni n -upla di valori (x_1, x_2, \dots, x_n) in cui la funzione di distribuzione ammette derivate di ordine n continue, per la (3.52) si ha:

$$f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \dots \partial x_n} F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n). \quad (3.53)$$

La densità di probabilità congiunta di una variabile aleatoria n -dimensionale $\mathbf{X} = (X_1, X_2, \dots, X_n)$ soddisfa le proprietà:

$$f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) \geq 0, \quad \forall \mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n \quad (3.54)$$

$$\int_{-\infty}^{+\infty} dz_1 \int_{-\infty}^{+\infty} dz_2 \dots \int_{-\infty}^{+\infty} f_{X_1, X_2, \dots, X_n}(z_1, z_2, \dots, z_n) dz_n = 1.$$

Si noti che per variabili aleatorie assolutamente continue la conoscenza della densità di probabilità consente, in linea di principio, di calcolare probabilità del tipo

$$P\{(X_1, X_2, \dots, X_n) \in D\} = \int_D f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n$$

per ogni assegnata regione D di \mathbb{R}^n .

Nel caso di un vettore aleatorio $\mathbf{X} = (X_1, X_2, \dots, X_n)$ assolutamente continuo, è possibile esprimere la densità di probabilità di un qualsiasi numero m ($m < n$) di variabili aleatorie. Ad esempio, se si considerano le prime m , in virtù della Proposizione 3.4 e della (3.52) si ha:

$$F_{X_1, X_2, \dots, X_m}(x_1, x_2, \dots, x_m) = \int_{-\infty}^{x_1} dz_1 \dots \int_{-\infty}^{x_m} \left[\int_{-\infty}^{+\infty} dz_{m+1} \dots \int_{-\infty}^{+\infty} f_{X_1, \dots, X_n}(z_1, \dots, z_n) dz_n \right] dz_m \quad (3.55)$$

e quindi, dalla (3.53) segue:

$$f_{X_1, X_2, \dots, X_m}(x_1, x_2, \dots, x_m) = \int_{-\infty}^{+\infty} dx_{m+1} \dots \int_{-\infty}^{+\infty} f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) dx_n. \quad (3.56)$$

La funzione $f_{X_1, X_2, \dots, X_m}(x_1, x_2, \dots, x_m)$ ($m = 1, 2, \dots, n-1$) prende il nome di *densità di probabilità marginale*.

Nel caso di vettori aleatori assolutamente continui, la (3.56) mostra che la conoscenza della densità di probabilità congiunta $f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$ permette di determinare tutte le densità di probabilità marginali relative a coppie, triple, $(n-1)$ -uple di variabili aleatorie scelte tra le n variabili aleatorie X_1, X_2, \dots, X_n , oltre che le densità di probabilità delle variabili aleatorie singolarmente considerate.

3.8 Indipendenza di variabili aleatorie

Tra le possibili proprietà delle variabili aleatorie particolare rilievo riveste la nozione di indipendenza.

Definizione 3.14 Siano X_1, X_2, \dots, X_n variabili aleatorie tutte definite in uno stesso spazio di probabilità (Ω, \mathcal{F}, P) ed a valori in \mathbb{R} . Esse si dicono indipendenti se e solo se per ogni n -upla arbitraria di sottoinsiemi di Borel B_1, B_2, \dots, B_n di \mathbb{R} si ha:

$$P(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n) = P(X_1 \in B_1) P(X_2 \in B_2) \cdots P(X_n \in B_n). \quad (3.57)$$

In generale, se X_1, X_2, \dots è una successione di variabili aleatorie definite in uno stesso spazio di probabilità ed a valori in \mathbb{R} , esse si dicono indipendenti se e solo se per ogni insieme finito di indici distinti i_1, i_2, \dots, i_k , le variabili aleatorie $X_{i_1}, X_{i_2}, \dots, X_{i_k}$ sono indipendenti.

Dalla Definizione 3.14 segue che se X_1, X_2, \dots, X_n sono indipendenti allora per ogni $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ si ha

$$\begin{aligned} F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) &= P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) \\ &= P(X_1 \leq x_1) P(X_2 \leq x_2) \cdots P(X_n \leq x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \cdots F_{X_n}(x_n), \end{aligned} \quad (3.58)$$

ossia la funzione di distribuzione congiunta di X_1, X_2, \dots, X_n si fattorizza nel prodotto delle funzioni di distribuzione delle singole variabili aleatorie. Si noti che è possibile dimostrare che vale anche il viceversa, nel senso che se sussiste la (3.58) per ogni $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$, allora X_1, X_2, \dots, X_n sono indipendenti.

Proposizione 3.5 Se le variabili aleatorie X_1, X_2, \dots, X_n sono indipendenti, tali risultano anche essere m di esse ($m < n$), comunque scelte.

Dimostrazione Senza perdita di generalità, dimostriamo che se X_1, X_2, \dots, X_n sono variabili aleatorie indipendenti, allora anche le variabili aleatorie X_1, X_2, \dots, X_m ($m < n$) sono indipendenti. Essendo le variabili X_1, X_2, \dots, X_n indipendenti, per la Definizione 3.14 si ha infatti:

$$\begin{aligned} P(X_1 \in B_1, X_2 \in B_2, \dots, X_m \in B_m) &= P(X_1 \in B_1, \dots, X_m \in B_m, X_{m+1} \in \mathbb{R}, \dots, X_n \in \mathbb{R}) \\ &= P(X_1 \in B_1) P(X_2 \in B_2) \cdots P(X_m \in B_m) P(X_{m+1} \in \mathbb{R}) \cdots P(X_n \in \mathbb{R}) \\ &= P(X_1 \in B_1) P(X_2 \in B_2) \cdots P(X_m \in B_m). \end{aligned}$$

□

Proposizione 3.6 Siano X_1, X_2, \dots, X_n variabili aleatorie discrete definite in uno stesso spazio di probabilità (Ω, \mathcal{F}, P) ed a valori in \mathbb{R} , con rispettive funzioni di probabilità $p_{X_1}(x_1), p_{X_2}(x_2), \dots, p_{X_n}(x_n)$. Allora X_1, X_2, \dots, X_n sono indipendenti se e solo se per ogni n -upla $\mathbf{x} = (x_1, x_2, \dots, x_n)$ di valori assunti da $\mathbf{X} = (X_1, X_2, \dots, X_n)$ risulta:

$$p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = p_{X_1}(x_1)p_{X_2}(x_2) \cdots p_{X_n}(x_n). \quad (3.59)$$

Dimostrazione Se X_1, X_2, \dots, X_n sono indipendenti, dalla Definizione 3.14 segue immediatamente la (3.59). Viceversa, se vale la (3.59), comunque si scelgano i sottoinsiemi di Borel B_1, B_2, \dots, B_n di \mathbb{R} e posto $D = \{(x_1, x_2, \dots, x_n) \in B_1 \times B_2 \times \cdots \times B_n\}$, si ha:

$$\begin{aligned} P(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n) &= \sum_D p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) \\ &= \sum_{x_1 \in B_1} p_{X_1}(x_1) \sum_{x_2 \in B_2} p_{X_2}(x_2) \cdots \sum_{x_n \in B_n} p_{X_n}(x_n) \\ &= P(X_1 \in B_1) P(X_2 \in B_2) \cdots P(X_n \in B_n). \end{aligned}$$

Dalla Definizione 3.14 segue quindi che le variabili aleatorie X_1, X_2, \dots, X_n sono indipendenti. \square

Si noti che se X_1, X_2, \dots, X_n sono variabili aleatorie discrete indipendenti, allora il vettore aleatorio $\mathbf{X} = (X_1, X_2, \dots, X_n)$ è discreto. La Proposizione 3.6 mostra in particolare che se X_1, X_2, \dots, X_n sono variabili aleatorie discrete indipendenti, la funzione di probabilità congiunta può essere determinata a partire dalle funzioni di probabilità delle singole variabili aleatorie.

Esempio 3.19 Consideriamo una variabile aleatoria bidimensionale discreta (X, Y) di funzione di probabilità:

$$p_{X,Y}(x, y) = \begin{cases} a, & (x = 1, y = 1) \text{ oppure } (x = 2, y = 2) \\ \frac{1}{2} - a, & (x = 1, y = 2) \text{ oppure } (x = 2, y = 1) \\ 0, & \text{altrimenti} \end{cases}$$

con $0 \leq a \leq 1/2$. Essendo $P(X = 1) = P(X = 2) = 1/2$ e $P(Y = 1) = P(Y = 2) = 1/2$, le funzioni di probabilità marginali di X e Y sono rispettivamente:

$$p_X(x) = \begin{cases} 1/2, & x = 1 \\ 1/2, & x = 2 \\ 0, & \text{altrimenti,} \end{cases} \quad p_Y(y) = \begin{cases} 1/2, & y = 1 \\ 1/2, & y = 2 \\ 0, & \text{altrimenti.} \end{cases}$$

Le variabili aleatorie X e Y sono indipendenti se e solo se si sceglie $a = 1/4$; infatti, soltanto in tal caso risulta $p_{X,Y}(x, y) = p_X(x)p_Y(y)$ per ogni $(x, y) \in \mathbb{R}^2$. \diamond

Proposizione 3.7 Siano X_1, X_2, \dots, X_n variabili aleatorie assolutamente continue definite in uno stesso spazio di probabilità ed a valori in \mathbb{R} , di rispettive densità di probabilità

$f_{X_1}(x_1), f_{X_2}(x_2), \dots, f_{X_n}(x_n)$. Allora X_1, X_2, \dots, X_n sono indipendenti se e solo se per ogni n -upla $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ si ha

$$f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n). \quad (3.60)$$

Dimostrazione Se X_1, X_2, \dots, X_n sono indipendenti, dalla (3.58) segue:

$$\begin{aligned} F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) &= F_{X_1}(x_1) F_{X_2}(x_2) \cdots F_{X_n}(x_n) \\ &= \int_{-\infty}^{x_1} f_{X_1}(z_1) dz_1 \cdots \int_{-\infty}^{x_n} f_{X_n}(z_n) dz_n \\ &= \int_{-\infty}^{x_1} dz_1 \cdots \int_{-\infty}^{x_n} [f_{X_1}(z_1) \cdots f_{X_n}(z_n)] dz_n, \end{aligned}$$

che mostra che $F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$ è suscettibile di rappresentazione nella forma integrale (3.52). Pertanto, il vettore aleatorio (X_1, X_2, \dots, X_n) è assolutamente continuo con densità di probabilità congiunta (3.60). Viceversa, se vale la (3.60), allora comunque si scelgano i sottoinsiemi di Borel B_1, B_2, \dots, B_n di \mathbb{R} , posto $D = \{(x_1, x_2, \dots, x_n) \in B_1 \times B_2 \times \cdots \times B_n\}$, risulta:

$$\begin{aligned} P(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n) &= \int_D f_{X_1, X_2, \dots, X_n}(z_1, z_2, \dots, z_n) dz_1 dz_2 \cdots dz_n \\ &= \int_{B_1} f_{X_1}(z_1) dz_1 \int_{B_2} f_{X_2}(z_2) dz_2 \cdots \int_{B_n} f_{X_n}(z_n) dz_n \\ &= P(X_1 \in B_1) P(X_2 \in B_2) \cdots P(X_n \in B_n) \end{aligned}$$

così che, per la Definizione 3.14, X_1, X_2, \dots, X_n sono indipendenti. \square

Se X_1, X_2, \dots, X_n sono indipendenti, il vettore aleatorio $\mathbf{X} = (X_1, X_2, \dots, X_n)$ è assolutamente continuo. La Proposizione 3.7 mostra in particolare che se X_1, X_2, \dots, X_n sono variabili aleatorie assolutamente continue indipendenti, la densità di probabilità congiunta può essere espressa mediante le densità di probabilità delle singole variabili aleatorie.

Si noti che dalla Proposizione 3.7 discende che per variabili aleatorie assolutamente continue la condizione di indipendenza si esprime equivalentemente nelle forme (3.58) o (3.60).

Esempio 3.20 Si consideri il vettore aleatorio assolutamente continuo (X, Y) di densità di probabilità congiunta:

$$f_{X,Y}(x, y) = \begin{cases} 2, & 0 < x < 1, 0 < y < 1/2 \\ 0, & \text{altrimenti.} \end{cases}$$

È facile rendersi conto che tale funzione soddisfa le proprietà (3.44). Invero, $f_{X,Y}(x, y) \geq 0$ per ogni $(x, y) \in \mathbb{R}^2$ ed inoltre

$$\int_{-\infty}^{+\infty} du \int_{-\infty}^{+\infty} f_{X,Y}(u, v) dv = 2 \int_0^1 du \int_0^{1/2} dv = 1.$$

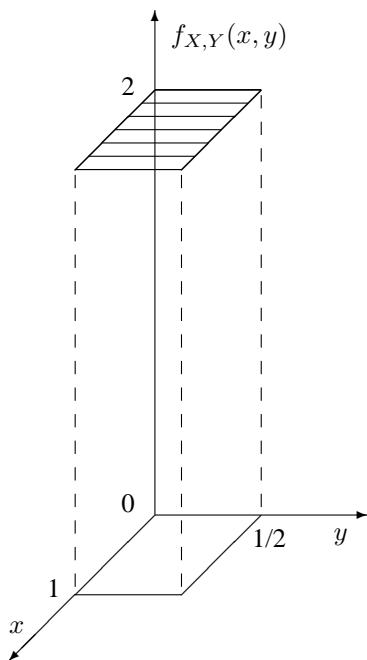


Figura 3.4 – La densità di probabilità congiunta di cui all'Esempio 3.20.

La Figura 3.4 mostra il grafico di $f_{X,Y}(x,y)$. Dalla (3.47) segue:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,v) dv = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{altrimenti;} \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(u,y) du = \begin{cases} 2, & 0 < y < 1/2 \\ 0, & \text{altrimenti.} \end{cases}$$

Poiché risulta $f_{X,Y}(x,y) = f_X(x) f_Y(y)$ per ogni coppia $(x,y) \in \mathbb{R}^2$, si trae che le variabili aleatorie X e Y sono indipendenti. Al contrario, le variabili aleatorie X e Y considerate nell'Esempio 3.17 non sono indipendenti in quanto l'uguaglianza $f_{X,Y}(x,y) = f_X(x) f_Y(y)$ non è soddisfatta per ogni coppia $(x,y) \in \mathbb{R}^2$. \diamond

Va infine menzionato che variabili aleatorie che descrivono risultati ottenuti in differenti prove indipendenti di un esperimento casuale sono solitamente supposte indipendenti. In altri termini, l'indipendenza delle prove si suole ragionevolmente assumere come implicante l'indipendenza delle corrispondenti variabili aleatorie.

Esempio 3.21 Un gioco tra due individui A e B si svolge con le seguenti modalità: A lancia un dado e acquisisce come punteggio il numero uscito; indipendentemente B lancia un altro

dado per due volte e gli si attribuisce come punteggio il maggiore dei numeri usciti: A vince se il suo punteggio non è inferiore di quello di B, altrimenti perde e vince B. Si intende calcolare la probabilità dell'evento $E = \{\text{il giocatore A vince il gioco}\}$.

A tal fine, si denoti con R_1 la variabile aleatoria che descrive il punteggio del giocatore A. Indicando poi con Y_k la variabile aleatoria che descrive il punteggio di B nel lancio k -esimo ($k = 1, 2$), sia $R_2 = \max(Y_1, Y_2)$ la variabile aleatoria che descrive il punteggio conseguito da B a seguito dei suoi due lanci. L'evento E si verifica se e solo se risulta $R_1 \geq R_2$. Per l'indipendenza dei due lanci effettuati da B si ha:

$$P(R_2 \leq j) = P(Y_1 \leq j, Y_2 \leq j) = P(Y_1 \leq j) P(Y_2 \leq j) = \frac{j^2}{36} \quad (j = 1, 2, \dots, 6),$$

$$P(R_2 = j) = P(R_2 \leq j) - P(R_2 \leq j-1) = \frac{2j-1}{36} \quad (j = 1, 2, \dots, 6).$$

Inoltre, per l'indipendenza delle prove che descrivono i lanci effettuati dai due giocatori, le variabili aleatorie R_1 ed R_2 possono ragionevolmente ritenersi indipendenti, così da aversi:

$$\begin{aligned} P(E) &= P(R_1 \geq R_2) = \sum_{i=1}^6 \sum_{j=1}^i P(R_1 = i, R_2 = j) = \sum_{i=1}^6 P(R_1 = i) \sum_{j=1}^i P(R_2 = j) \\ &= \sum_{i=1}^6 P(R_1 = i) P(R_2 \leq i) = \sum_{i=1}^6 \frac{1}{6} \frac{i^2}{36} = \frac{1}{6^3} \sum_{i=1}^6 i^2 = \frac{91}{216}. \end{aligned}$$

La probabilità che B vinca al gioco è dunque $P(\overline{E}) = 1 - P(E) = 125/216$ che, come del resto c'era da attendersi, è maggiore della probabilità di vincita di A. \diamond

Teorema 3.7 Siano X_1, X_2, \dots, X_n variabili aleatorie indipendenti tali che $X_i: \Omega \rightarrow \mathbb{R}$ ($i = 1, 2, \dots, n$) ed inoltre siano $g_i: \mathbb{R} \rightarrow \mathbb{R}$ ($i = 1, 2, \dots, n$) delle funzioni misurabili. Allora le variabili aleatorie $Y_1 = g_1(X_1), Y_2 = g_2(X_2), \dots, Y_n = g_n(X_n)$ sono indipendenti.

Dimostrazione Per tutti i sottoinsiemi di Borel B_1, B_2, \dots, B_n di \mathbb{R} si ha:

$$\begin{aligned} P(Y_1 \in B_1, Y_2 \in B_2, \dots, Y_n \in B_n) &= P\{g_1(X_1) \in B_1, g_2(X_2) \in B_2, \dots, g_n(X_n) \in B_n\} \\ &= P\{X_1 \in g_1^{-1}(B_1), X_2 \in g_2^{-1}(B_2), \dots, X_n \in g_n^{-1}(B_n)\}. \end{aligned} \quad (3.61)$$

Poiché per ipotesi le variabili aleatorie X_1, X_2, \dots, X_n sono indipendenti e poiché le funzioni $g_i: \mathbb{R} \rightarrow \mathbb{R}$ ($i = 1, 2, \dots, n$) sono misurabili, dalla (3.61) segue:

$$\begin{aligned} P(Y_1 \in B_1, Y_2 \in B_2, \dots, Y_n \in B_n) &= P\{X_1 \in g_1^{-1}(B_1)\} P\{X_2 \in g_2^{-1}(B_2)\} \cdots P\{X_n \in g_n^{-1}(B_n)\} \\ &= P(Y_1 \in B_1) P(Y_2 \in B_2) \cdots P(Y_n \in B_n) \end{aligned}$$

che implica l'indipendenza delle variabili aleatorie Y_1, Y_2, \dots, Y_n . \square

Esempio 3.22 Sia (X, Y) un vettore aleatorio assolutamente continuo di densità di probabilità congiunta

$$f_{X,Y}(x, y) = \begin{cases} 4x(1-y), & 0 < x < 1, 0 < y < 1 \\ 0, & \text{altrimenti.} \end{cases}$$

Determiniamo la densità di probabilità congiunta del vettore aleatorio (Z, W) , dove $Z = X^2$ e $W = (1 - Y)^2$.

A tal fine notiamo in primo luogo che X e Y sono indipendenti e che risulta:

$$f_X(x) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & \text{altrimenti} \end{cases} \quad f_Y(y) = \begin{cases} 2(1-y), & 0 < y < 1 \\ 0, & \text{altrimenti.} \end{cases}$$

Pertanto, dal Teorema 3.7 si trae che Z e W sono indipendenti, così che $f_{Z,W}(z, w) = f_Z(z) f_W(w)$ per ogni $(z, w) \in \mathbb{R}^2$. Essendo

$$F_Z(z) = P(X^2 \leq z) = \begin{cases} 0, & z < 0 \\ P(X \leq \sqrt{z}) = 2 \int_0^{\sqrt{z}} x \, dx = z, & 0 \leq z < 1 \\ 1, & z \geq 1 \end{cases}$$

e

$$F_W(w) = P[(1-Y)^2 \leq w] = \begin{cases} 0, & w < 0 \\ P(Y \geq 1 - \sqrt{w}) = 2 \int_{1-\sqrt{w}}^1 (1-y) \, dy = w, & 0 \leq w < 1 \\ 1, & w \geq 1, \end{cases}$$

segue:

$$f_{Z,W}(z, w) = f_Z(z) f_W(w) = \begin{cases} 1, & 0 < z < 1, 0 < w < 1 \\ 0, & \text{altrimenti.} \end{cases}$$

Si noti che Z e W sono identicamente distribuite. ◇

3.9 Trasformazioni di vettori aleatori

Nel Paragrafo 3.4 abbiamo esaminato le trasformazioni di variabili aleatorie. Vogliamo ora affrontare lo stesso problema nel caso di vettori aleatori. In primo luogo notiamo che la Proposizione 3.3 può essere così generalizzata:

Proposizione 3.8 Sia $\mathbf{X}: \Omega \rightarrow \mathbb{R}^n$ una variabile aleatoria n -dimensionale e sia $g(\mathbf{x})$ con $g: \mathbb{R}^n \rightarrow \mathbb{R}^m$ una funzione Borel-misurabile⁴. La funzione $\mathbf{Y} = g(\mathbf{X})$ è una variabile aleatoria m -dimensionale funzione di \mathbf{X} .

Poiché \mathbf{Y} è funzione di \mathbf{X} , per ogni insieme di Borel B in \mathcal{B} su \mathbb{R}^m risulta:

$$P(\mathbf{Y} \in B) = P(g(\mathbf{X}) \in B) = P(\mathbf{X} \in g^{-1}(B)). \quad (3.62)$$

La formula (3.62) permette di esprimere $P(\mathbf{Y} \in B)$ in termini di probabilità inerenti il vettore aleatorio \mathbf{X} .

⁴Una funzione $g: \mathbb{R}^n \rightarrow \mathbb{R}^m$ è Borel-misurabile se per ogni sottoinsieme di Borel $B \in \mathcal{B}$ di \mathbb{R}^m si ha che $g^{-1}(B) = \{\mathbf{x} \in \mathbb{R}^n : g(\mathbf{x}) \in B\}$ è un sottoinsieme di Borel di \mathbb{R}^n .

Esempio 3.23 Siano X_1, X_2, \dots, X_n variabili aleatorie indipendenti, tutte definite nello stesso spazio di probabilità. Determiniamo la funzione di distribuzione delle variabili aleatorie

$$U = \max\{X_1, X_2, \dots, X_n\}, \quad V = \min\{X_1, X_2, \dots, X_n\}$$

e la funzione di distribuzione del vettore (U, V) .

È anzitutto immediato mostrare che, in virtù dell'indipendenza di X_1, X_2, \dots, X_n , risulta:

$$\begin{aligned} F_U(u) &= P(\max\{X_1, X_2, \dots, X_n\} \leq u) = P(X_1 \leq u, X_2 \leq u, \dots, X_n \leq u) \\ &= P(X_1 \leq u) P(X_2 \leq u) \cdots P(X_n \leq u) = \prod_{i=1}^n F_{X_i}(u), \quad \forall u \in \mathbb{R}, \end{aligned} \quad (3.63)$$

$$\begin{aligned} F_V(v) &= P(\min\{X_1, X_2, \dots, X_n\} \leq v) = 1 - P(\min\{X_1, X_2, \dots, X_n\} > v) \\ &= 1 - P(X_1 > v, X_2 > v, \dots, X_n > v) = 1 - \prod_{i=1}^n [1 - F_{X_i}(v)], \quad \forall v \in \mathbb{R}. \end{aligned} \quad (3.64)$$

Poiché risulta $P(U \leq u) = P(U \leq u, V \leq v) + P(U \leq u, V > v)$, la funzione di distribuzione di (U, V) è:

$$\begin{aligned} F_{U,V}(u, v) &= P(U \leq u, V \leq v) = P(U \leq u) - P(U \leq u, V > v) \\ &= F_U(u) - P(v < X_1 \leq u, v < X_2 \leq u, \dots, v < X_n \leq u) \\ &= \begin{cases} \prod_{i=1}^n F_{X_i}(u), & u < v \\ \prod_{i=1}^n F_{X_i}(u) - \prod_{i=1}^n [F_{X_i}(u) - F_{X_i}(v)], & u \geq v. \end{cases} \end{aligned} \quad (3.65)$$

Si noti che la conoscenza della funzione di distribuzione delle variabili aleatorie indipendenti X_1, X_2, \dots, X_n permette di determinare le funzioni di distribuzione (3.63)-(3.65). È facile rendersi conto che se si procede al limite per $u \rightarrow +\infty$ nella (3.65) si ottiene la (3.64), mentre al limite per $v \rightarrow +\infty$ si ricava la (3.63). Infine, dalle (3.63)-(3.65) si vede facilmente che in generale U e V non sono indipendenti.

Lo studio delle variabili aleatorie U e V riveste ampio interesse in problemi applicativi. Invero, se ad esempio le variabili aleatorie indipendenti X_1, X_2, \dots, X_n descrivono i tempi di corretto funzionamento di n componenti di un dispositivo, allora U rappresenta il tempo di corretto funzionamento del dispositivo quando le n componenti sono collegate in parallelo, mentre V rappresenta il tempo di corretto funzionamento del dispositivo quando le n componenti sono collegate in serie. \diamond

Trasformazioni che frequentemente si incontrano nelle applicazioni sono del tipo $Y = g(X_1, X_2, \dots, X_n)$, con $\mathbf{X} = (X_1, X_2, \dots, X_n)$ vettore aleatorio e con $g: \mathbb{R}^n \rightarrow \mathbb{R}$ funzione Borel-misurabile. Per determinare la funzione di distribuzione di Y è opportuno distinguere il caso in cui \mathbf{X} è discreto da quello in cui \mathbf{X} è assolutamente continuo. Sia $D = \{(x_1, x_2, \dots, x_n) \in \mathbb{R}^n: g(x_1, x_2, \dots, x_n) \leq y\}$. Per ogni $y \in \mathbb{R}$, se \mathbf{X} è un vettore

aleatorio discreto si ha:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P\{g(X_1, X_2, \dots, X_n) \leq y\} \\ &= \sum_{(x_1, x_2, \dots, x_n) \in D} p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n), \end{aligned} \quad (3.66)$$

mentre se \mathbf{X} è un vettore aleatorio assolutamente continuo risulta:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P\{g(X_1, X_2, \dots, X_n) \leq y\} \\ &= \int_D f_{X_1, X_2, \dots, X_n}(z_1, z_2, \dots, z_n) dz_1 dz_2 \dots dz_n. \end{aligned} \quad (3.67)$$

Esempio 3.24 Sia (X, Y) un vettore aleatorio assolutamente continuo di densità di probabilità congiunta (v. Esempio 3.17)

$$f_{X,Y}(x, y) = \begin{cases} 2, & 0 < x \leq y < 1 \\ 0, & \text{altrimenti.} \end{cases}$$

Si desidera calcolare le densità di probabilità delle variabili aleatorie $Z = Y - X$ e $W = X + Y$.

Determiniamo in primo luogo la funzione di distribuzione di Z . Tale variabile aleatoria assume valori nell'intervallo $(0, 1)$; pertanto $F_Z(z) = 0$ se $z < 0$ e $F_Z(z) = 1$ se $z \geq 1$. Invece per $0 \leq z < 1$ dalla (3.67) segue:

$$F_Z(z) = P(Y - X \leq z) = \int_D f_{X,Y}(x, y) dx dy = 2 \int_D dx dy$$

dove $D = \{(x, y): 0 < x \leq y < 1, y - x \leq z\}$.

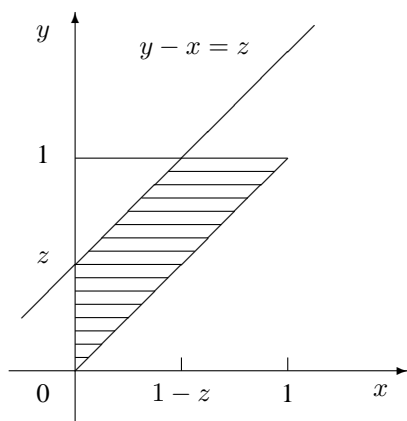


Figura 3.5– La regione tratteggiata individua l'insieme $D = \{(x, y): 0 < x \leq y < 1, y - x \leq z\}$ di cui all'Esempio 3.24.

Come si evidenzia dalla Figura 3.5, per determinare $F_Z(z)$ occorre moltiplicare l'area della regione tratteggiata per 2. Si giunge così al seguente risultato:

$$F_Z(z) = \begin{cases} 0, & z < 0 \\ 2 \left[\frac{1}{2} - \frac{(1-z)^2}{2} \right] = z(2-z), & 0 \leq z < 1 \\ 1, & z \geq 1, \end{cases}$$

da cui si trae:

$$f_Z(z) = \begin{cases} 2(1-z), & 0 < z < 1 \\ 0, & \text{altrimenti.} \end{cases}$$

Determiniamo ora la funzione di distribuzione di W . Tale variabile aleatoria assume valori nell'intervallo $(0, 2)$, così che si ha $F_W(w) = 0$ se $w < 0$ e $F_W(w) = 1$ se $w \geq 2$. Per $0 \leq w < 2$ dalla (3.67) si ricava:

$$F_W(w) = P(X + Y \leq w) = \int_D f_{X,Y}(x, y) dx dy = 2 \int_D dx dy,$$

dove $D = \{(x, y): 0 < x \leq y < 1, x + y \leq w\}$. Come si evince dalla Figura 3.6, occorre distinguere i seguenti casi: (a) $0 \leq w < 1$ e (b) $1 \leq w < 2$.

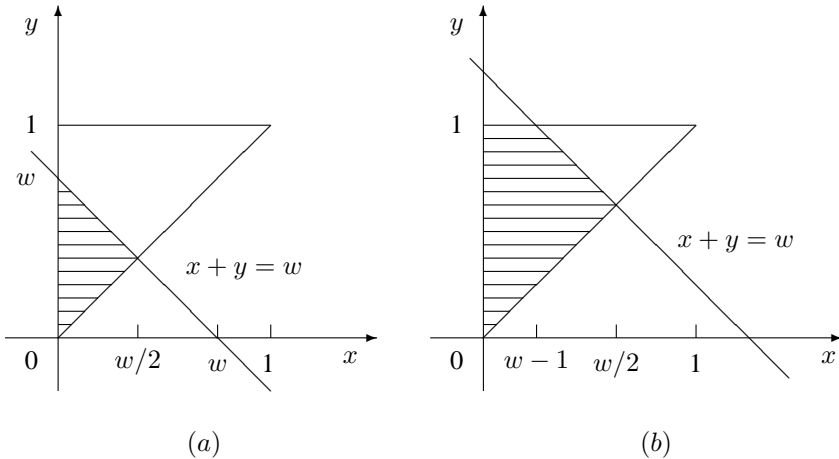


Figura 3.6– La regione tratteggiata individua l'insieme $D = \{(x, y): 0 < x \leq y < 1, x + y \leq w\}$ nelle situazioni (a) e (b) considerate nell'Esempio 3.24.

Per determinare $F_Z(z)$ occorre semplicemente moltiplicare per 2 le aree delle regioni tratte-

giate in (a) e (b). Si ottiene pertanto

$$F_W(w) = \begin{cases} 0, & w < 0 \\ 2 \frac{w(w/2)}{2} = \frac{w^2}{2}, & 0 \leq w < 1 \\ 2 \left[\frac{1}{2} - \frac{(2-w)(1-w/2)}{2} \right] = 1 - \frac{(2-w)^2}{2}, & 1 \leq w < 2 \\ 1, & w \geq 2, \end{cases}$$

da cui si ricava:

$$f_W(w) = \begin{cases} w, & 0 < w < 1 \\ 2 - w, & 1 \leq w < 2 \\ 0, & \text{altrimenti.} \end{cases}$$

◇

Esempio 3.25 Sia (X, Y) un vettore aleatorio assolutamente continuo di densità di probabilità congiunta

$$f_{X,Y}(x, y) = \begin{cases} \frac{3x}{2}, & 0 < x < 1, -x < y < x \\ 0, & \text{altrimenti.} \end{cases}$$

Si desidera calcolare le densità di probabilità delle variabili aleatorie $Z = XY$ e $W = Y/X$.

Determiniamo in primo luogo la funzione di distribuzione di Z . Tale variabile aleatoria assume valori nell'intervallo $(-1, 1)$; pertanto risulta che $F_Z(z) = 0$ se $z < -1$ e $F_Z(z) = 1$ se $z \geq 1$. Invece se $-1 \leq z < 1$ dalla (3.67) segue:

$$F_Z(z) = P(XY \leq z) = \int_D f_{X,Y}(x, y) dx dy = \frac{3}{2} \int_D x dx dy,$$

dove $D = \{(x, y): 0 < x < 1, -x < y < x, xy \leq z\}$.

Come mostrato in Figura 3.7, per determinare $F_Z(z)$ occorre distinguere i seguenti casi: (a) $-1 \leq z < 0$ e (b) $0 \leq z < 1$. Se $-1 \leq z < 0$ si ha:

$$F_Z(z) = \frac{3}{2} \int_{\sqrt{-z}}^1 x dx \int_{-x}^{z/x} dy = \frac{3}{2} \int_{\sqrt{-z}}^1 x \left(\frac{z}{x} + x \right) dx = \frac{1 + 3z - 2z\sqrt{-z}}{2},$$

mentre se $0 \leq z < 1$ risulta:

$$F_Z(z) = \frac{3}{2} \int_0^{\sqrt{z}} x dx \int_{-x}^x dy + \frac{3}{2} \int_{\sqrt{z}}^1 x dx \int_{-x}^{z/x} dy = \frac{1 + 3z - 2z\sqrt{z}}{2}.$$

Si giunge così al seguente risultato:

$$F_Z(z) = \begin{cases} 0, & z < -1 \\ \frac{1 + 3z - 2z\sqrt{|z|}}{2}, & -1 \leq z < 1 \\ 1, & z \geq 1, \end{cases}$$

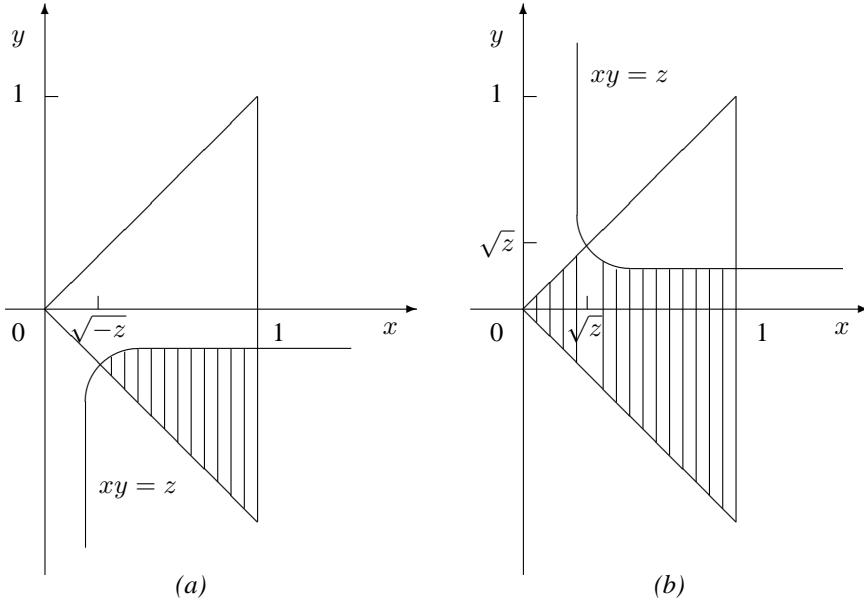


Figura 3.7– La regione tratteggiata individua l'insieme $D = \{(x, y): 0 < x < 1, -x < y < x, xy \leq z\}$ nei due casi (a) e (b) considerati nell'Esempio 3.25.

così che

$$f_Z(z) = \begin{cases} \frac{3}{2} (1 - \sqrt{|z|}), & -1 < z < 1 \\ 0, & \text{altrimenti.} \end{cases}$$

Determiniamo ora la funzione di distribuzione di W . Tale variabile aleatoria assume valori nell'intervallo $(-1, 1)$ e quindi si ha $F_W(w) = 0$ se $w < 0$ e $F_W(w) = 1$ se $w \geq 1$. Se $-1 \leq w < 1$ dalla (3.67) si ricava:

$$F_W(w) = P\left(\frac{Y}{X} \leq w\right) = \int_D f_{X,Y}(x, y) dx dy = \frac{3}{2} \int_D x dx dy,$$

dove $D = \{(x, y): 0 < x < 1, -x < y < x, y \leq wx\}$.

Come si evince dalla Figura 3.8 risulta:

$$F_W(w) = \begin{cases} 0, & w < -1 \\ \frac{3}{2} \int_0^1 x dx \int_{-x}^{wx} dy = \frac{1+w}{2}, & -1 \leq w < 1 \\ 1, & w \geq 1, \end{cases}$$

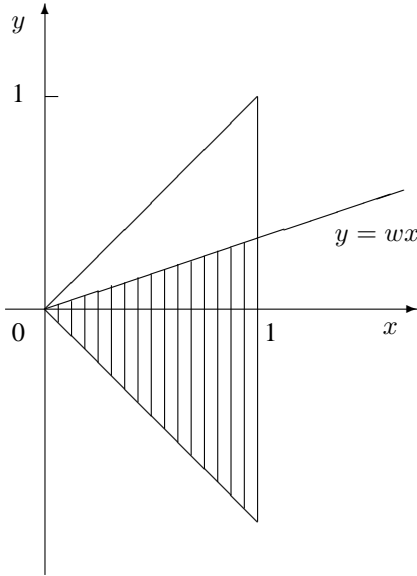


Figura 3.8– La regione tratteggiata individua l'insieme $D = \{(x, y): 0 < x < 1, -x < y < x, y \leq wx\}$ di cui all'Esempio 3.25.

da cui segue:

$$f_W(w) = \begin{cases} \frac{1}{2}, & -1 < w < 1 \\ 0, & \text{altrimenti.} \end{cases}$$

◇

Per vettori aleatori assolutamente continui sussiste il seguente teorema, che costituisce un'estensione del Teorema 3.4 relativo al caso unidimensionale. Per brevità ne omettiamo la dimostrazione.

Teorema 3.8 Sia (X_1, X_2, \dots, X_n) un vettore aleatorio assolutamente continuo di densità di probabilità $f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$. Siano $g_i: \mathbb{R}^n \rightarrow \mathbb{R}$ ($i = 1, 2, \dots, n$) funzioni tali che la trasformazione definita da $y_i = g_i(x_1, x_2, \dots, x_n)$ ($i = 1, 2, \dots, n$) sia biunivoca. Denotata con $x_i = h_i(y_1, y_2, \dots, y_n)$ ($i = 1, 2, \dots, n$) la trasformazione inversa, indichiamo con D il suo dominio. Supponiamo inoltre che le derivate parziali prime di $x_i = h_i(y_1, y_2, \dots, y_n)$ ($i = 1, 2, \dots, n$) rispetto a tutte le variabili esistono e sono

continue, e che lo Jacobiano

$$J_h(\mathbf{y}) = \frac{\partial(h_1, h_2, \dots, h_n)}{\partial(y_1, y_2, \dots, y_n)} = \begin{vmatrix} \frac{\partial h_1}{\partial y_1} & \frac{\partial h_1}{\partial y_2} & \dots & \frac{\partial h_1}{\partial y_n} \\ \frac{\partial h_2}{\partial y_1} & \frac{\partial h_2}{\partial y_2} & \dots & \frac{\partial h_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial y_1} & \frac{\partial h_n}{\partial y_2} & \dots & \frac{\partial h_n}{\partial y_n} \end{vmatrix}$$

è non nullo per ogni $\mathbf{y} \in D$. Allora, il vettore aleatorio (Y_1, Y_2, \dots, Y_n) , definito tramite la trasformazione $Y_i = g_i(X_1, X_2, \dots, X_n)$ ($i = 1, 2, \dots, n$), è assolutamente continuo con densità di probabilità

$$f_{Y_1, Y_2, \dots, Y_n}(y_1, y_2, \dots, y_n) = \begin{cases} f_{X_1, X_2, \dots, X_n}(h_1(\mathbf{y}), h_2(\mathbf{y}), \dots, h_n(\mathbf{y})) |J_h(\mathbf{y})|, & \mathbf{y} \in D \\ 0, & \text{altrimenti,} \end{cases}$$

dove si è posto $\mathbf{y} = (y_1, y_2, \dots, y_n)$.

Esempio 3.26 Sia (X, Y) un vettore aleatorio assolutamente continuo di densità di probabilità congiunta (v. Esempio 3.20)

$$f_{X,Y}(x, y) = \begin{cases} 2, & 0 < x < 1, 0 < y < 1/2 \\ 0, & \text{altrimenti.} \end{cases}$$

Determiniamo la densità di probabilità congiunta del vettore aleatorio (Z, W) ottenuto tramite le trasformazioni $Z = X - Y$ e $W = X + Y$.

Il sistema

$$\begin{cases} z = x - y \\ w = x + y \end{cases}$$

ammette come unica soluzione $x = h_1(z, w) = (z + w)/2$ e $y = h_2(z, w) = (w - z)/2$. Lo Jacobiano è:

$$J_h(z, w) = \frac{\partial(h_1, h_2)}{\partial(z, w)} = \begin{vmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{vmatrix} = \frac{1}{2}.$$

Essendo soddisfatte le ipotesi del Teorema 3.8, si ha:

$$f_{Z,W}(z, w) = \begin{cases} \frac{1}{2} f_{X,Y}\left(\frac{z+w}{2}, \frac{w-z}{2}\right) = 1, & 0 < z + w < 2, 0 < w - z < 1 \\ 0, & \text{altrimenti.} \end{cases}$$

La conoscenza della densità congiunta permette di ottenere le densità marginali $f_Z(z)$ e $f_W(w)$.

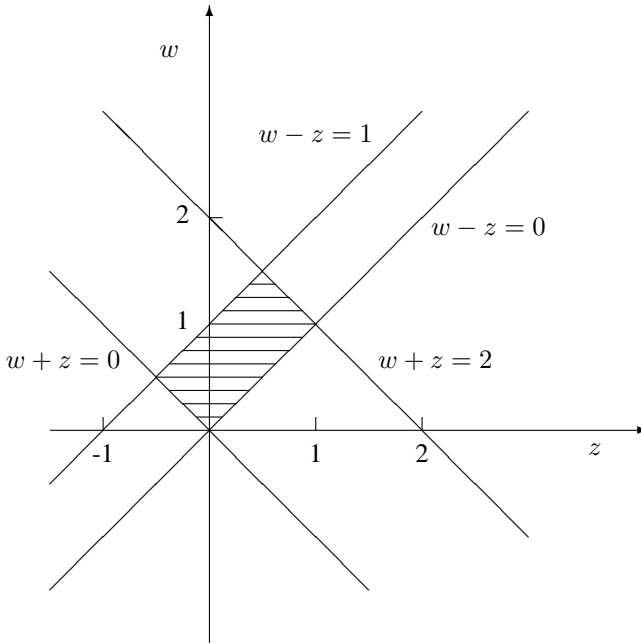


Figura 3.9– La regione tratteggiata indica l'insieme delle coppie (z, w) tali che $f_{Z,W}(z, w) > 0$ (v. Esempio 3.26).

Con riferimento alla Figura 3.9 si ricava:

$$f_Z(z) = \int_{-\infty}^{+\infty} f_{Z,W}(z, w) dw = \begin{cases} \int_{-z}^{1+z} dw = 1 + 2z, & -1/2 < z < 0 \\ \int_z^{1+z} dw = 1, & 0 < z < 1/2 \\ \int_z^{2-z} dw = 2(1 - z), & 1/2 < z < 1 \\ 0, & \text{altrimenti} \end{cases}$$

$$f_W(w) = \int_{-\infty}^{+\infty} f_{Z,W}(z, w) dz = \begin{cases} \int_{-w}^w dz = 2w, & 0 < w < 1/2 \\ \int_{w-1}^w dz = 1, & 1/2 < w < 1 \\ \int_{w-1}^{2-w} dz = 3 - 2w, & 1 < w < 3/2 \\ 0, & \text{altrimenti} \end{cases}$$

Si noti che Z e W non sono indipendenti, pur essendo indipendenti le variabili aleatorie X e Y . Inoltre, come mostra la Figura 3.10, la densità di probabilità di W è traslata di $1/2$ rispetto

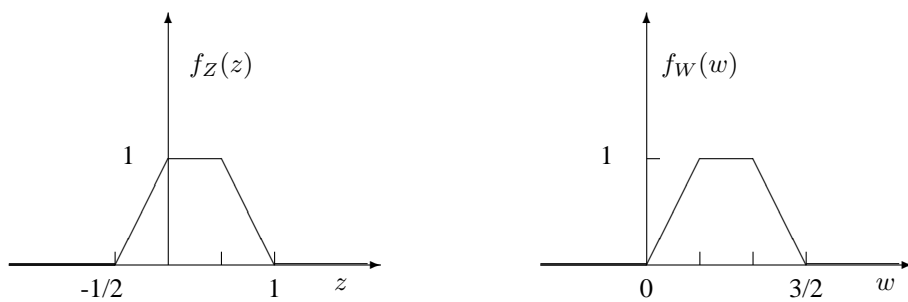


Figura 3.10 – Densità di probabilità marginali $f_Z(z)$ e $f_W(w)$ relative all'Esempio 3.26.

a quella di Z .



Capitolo 4

Distribuzioni notevoli

4.1 Famiglie parametriche di distribuzioni

Nel Capitolo 3 sono stati introdotti i concetti di variabile aleatoria e di funzione di distribuzione e si è giunti alla conclusione che una variabile aleatoria è completamente specificata dalla sua funzione di distribuzione. Le funzioni di distribuzione possono essere classificate in famiglie parametriche in base alla loro forma funzionale.

Una *famiglia parametrica* di funzioni di distribuzione è una collezione di funzioni di distribuzione indicizzate da parametri reali. Ad esempio, sia $F(x) = 0$ per $x < 0$ e $F(x) = 1 - e^{-\lambda x}$ per $x \geq 0$, con $\lambda > 0$; chiaramente $F(x)$ è una funzione di distribuzione di parametro λ . La collezione delle $F(x)$, ottenuta al variare di λ nell'insieme dei reali positivi, individua dunque una particolare famiglia parametrica di funzioni di distribuzione indicizzate appunto dal parametro λ .

Per la corrispondenza tra funzione di distribuzione e funzione di probabilità nel caso di variabili aleatorie discrete, e tra funzione di distribuzione e densità di probabilità se le variabili aleatorie sono assolutamente continue, il concetto di famiglia parametrica di distribuzioni si estende in modo naturale anche alle funzioni di probabilità ed alle densità di probabilità.

Nel seguito si studieranno alcune famiglie parametriche di funzioni di distribuzione che rivestono particolare importanza in problemi applicativi. A tal fine si esamineranno separatamente istanze relative a variabili aleatorie discrete ed a variabili aleatorie assolutamente continue.

4.2 Variabili aleatorie discrete

Nel presente paragrafo passeremo in rassegna alcune distribuzioni di probabilità unidimensionali di particolare rilevanza.

4.2.1 Distribuzione uniforme

Definizione 4.1 Una variabile aleatoria X che assume i valori $1, 2, \dots, n$, ciascuno con probabilità $1/n$, è detta avere distribuzione uniforme di parametro n .¹ Essa è caratterizzata dalla funzione di probabilità

$$p_X(x) = \begin{cases} \frac{1}{n}, & x = 1, 2, \dots, n \\ 0, & \text{altrimenti.} \end{cases} \quad (4.1)$$

La notazione $X \sim \mathcal{U}_d(n)$ verrà utilizzata per indicare che X ha distribuzione uniforme di parametro n . Dalla (4.1) è immediato verificare che risulta $p_X(x) \geq 0$ per ogni $x \in \mathbb{R}$, $\sum_{r=1}^n p_X(r) = 1$ e che

$$F_X(x) = P(X \leq x) = \begin{cases} 0, & x < 1 \\ \frac{k}{n}, & k \leq x < k+1 \quad (k = 1, 2, \dots, n-1) \\ 1, & x \geq n. \end{cases} \quad (4.2)$$

è la funzione di distribuzione di X . Nel caso particolare $n = 1$, la distribuzione uniforme si riduce ad una distribuzione degenere, avendosi $X = 1$ quasi certamente.

Esempio 4.1 Se X è la variabile aleatoria descrivente l'uscita nel lancio di un dado, risulta $X \sim \mathcal{U}_d(6)$. Nell'Esempio 3.6 è stata determinata e graficata la funzione di distribuzione di X . \diamond

Esempio 4.2 Nell'esperimento consistente nell'estrarre n biglie da un'urna che ne contiene N numerate da 1 a N , calcoliamo la probabilità che il massimo dei numeri estratti sia pari a k ($k = 1, 2, \dots, N$) considerando separatamente i casi di estrazioni effettuate con reinserimento e senza reinserimento.

Sia dunque X_i la variabile aleatoria che descrive il numero della biglia estratta nella i -esima estrazione ($i = 1, 2, \dots, n$) e sia $X = \max\{X_1, X_2, \dots, X_n\}$. Distinguiamo i due casi.

(i) Se l'estrazione è effettuata con reinserimento, X_1, X_2, \dots, X_n sono variabili aleatorie indipendenti e identicamente distribuite, con $X_i \sim \mathcal{U}_d(N)$ ($i = 1, 2, \dots, n$). Come mostrato in (3.63), per ogni intero positivo n , facendo uso della (4.2), risulta:

$$F_X(k) = P(X \leq k) = [F_{X_i}(k)]^n = \left(\frac{k}{N}\right)^n \quad (k = 1, 2, \dots, N),$$

da cui segue:

$$P(X = k) = F_X(k) - F_X(k-1) = \frac{k^n - (k-1)^n}{N^n} \quad (k = 1, 2, \dots, N).$$

¹Più in generale, nel discreto si può definire uniforme ogni variabile aleatoria che assume n valori distinti, ciascuno con probabilità $1/n$.

(ii) Se l'estrazione è effettuata senza reinserimento, le variabili aleatorie X_1, X_2, \dots, X_n non sono indipendenti. Nel caso in esame risulta $X = k$ se e solo se nel campione estratto compaiono la biglia contraddistinta dal numero k e $n - 1$ biglie recanti numeri distinti dell'insieme $\{1, 2, \dots, k - 1\}$. Pertanto risulta:

$$P(X = k) = \frac{\binom{k-1}{n-1}}{\binom{N}{n}} \quad (k = n, n+1, \dots, N).$$

◇

4.2.2 Distribuzione di Bernoulli

Una prova di Bernoulli è un esperimento casuale caratterizzato da due soli possibili risultati, interpretabili l'uno come *successo* e l'altro come *insuccesso*, che si verificano rispettivamente con probabilità p e $1 - p$, con $0 < p < 1$. La variabile aleatoria X che descrive il risultato di una prova di Bernoulli assume soltanto due valori: 1 (indicante il successo) con probabilità p e 0 (indicante l'insuccesso) con probabilità $1 - p$.

Definizione 4.2 Una variabile aleatoria X di funzione di probabilità

$$p_X(x) = \begin{cases} 1 - p, & x = 0 \\ p, & x = 1 \\ 0, & \text{altrimenti,} \end{cases} \quad (4.3)$$

con $0 < p < 1$, è detta avere distribuzione di Bernoulli di parametro p .

La funzione di distribuzione di X è pertanto

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - p, & 0 \leq x < 1 \\ 1, & x \geq 1. \end{cases} \quad (4.4)$$

Con la notazione $X \sim \mathcal{B}(1, p)$ intenderemo che X è una variabile aleatoria avente distribuzione di Bernoulli di parametro p , che chiameremo anche *variabile di Bernoulli*.

Esempio 4.3 Si consideri l'esperimento casuale consistente nel lancio di una moneta (v. Esempio 3.1) e sia X la variabile aleatoria definita dalle posizioni $X(C) = 0$ e $X(T) = 1$. Risulta $X \sim \mathcal{B}(1, 1/2)$. ◇

Più in generale, una prova di Bernoulli può essere costruita a partire da un qualsiasi spazio campione Ω considerando un evento A tale che $0 < P(A) < 1$ e identificando il verificarsi di A con il successo ed il verificarsi di \bar{A} con l'insuccesso. È allora utile definire la variabile aleatoria

$$X(\omega) = I_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \in \bar{A}, \end{cases}$$

dove $I_A(\omega)$ indica la funzione indicatrice di A . Se si pone $p = P(A)$, allora $X \sim \mathcal{B}(1, p)$.

4.2.3 Distribuzione binomiale

Consideriamo l'esperimento consistente in n prove di Bernoulli indipendenti ed effettuate tutte in condizioni identiche, ed assumiamo che in ogni prova i risultati di interesse siano sintetizzabili nel verificarsi dei seguenti due eventi necessari ed incompatibili: A (interpretabile come successo) e \bar{A} (interpretabile come insuccesso), con $P(A) = p$ ($0 < p < 1$). Un siffatto esperimento si dice costituito da n prove ripetute di Bernoulli. I possibili risultati possono essere rappresentati mediante un insieme di 2^n sequenze, ossia mediante l'insieme di tutte le possibili n -uple aventi A oppure \bar{A} in ognuna delle n posizioni. In virtù dell'indipendenza delle prove si può facilmente associare a ciascun risultato di detto esperimento una probabilità che, evidentemente, dipende da p , da n e dal numero di volte in cui l'evento A si è verificato nelle n prove. Ad esempio, la probabilità del seguente risultato

$$\underbrace{(A, \dots, A)}_{k \text{ volte}}, \underbrace{(\bar{A}, \dots, \bar{A})}_{n-k \text{ volte}}$$

esprime la circostanza che A si verifica nelle prime k prove e \bar{A} si verifica nelle rimanenti $n - k$ prove, è $p^k (1 - p)^{n-k}$. Questa è, tuttavia, anche la probabilità di qualsiasi n -upla caratterizzata dal verificarsi di A in k prove e di \bar{A} nelle rimanenti $n - k$ prove, a prescindere dall'ordine con cui si succedono i risultati nelle singole prove; tali n -uple sono in numero di $\binom{n}{k}$, tante quante sono le combinazioni di n oggetti a gruppi di k . Pertanto, l'evento

$$E_k = \{A \text{ si verifica in } k \text{ prove e } \bar{A} \text{ si verifica nelle rimanenti } n - k \text{ prove}\}$$

ha probabilità:

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

Sia X la variabile aleatoria che rappresenta il numero di volte in cui si verifica l'evento A nelle n prove; è evidente che $P(X = k) = P(E_k)$ per $k = 0, 1, \dots, n$.

Definizione 4.3 Una variabile aleatoria X di funzione di probabilità

$$p_X(x) = \begin{cases} \binom{n}{x} p^x (1 - p)^{n-x}, & x = 0, 1, \dots, n \\ 0, & \text{altrimenti,} \end{cases} \quad (4.5)$$

con $0 < p < 1$ e n intero positivo, è detta avere distribuzione binomiale di parametri n e p .

I valori ammissibili per p sono i reali dell'intervallo $(0, 1)$; quando p tende a 0 oppure ad 1 la variabile aleatoria X diventa degenerare. Il parametro n può invece assumere qualunque valore intero positivo. Si osservi che è $p_X(x) \geq 0$ per ogni $x \in \mathbb{R}$ e che è soddisfatta la condizione di normalizzazione:

$$\sum_{r=0}^n p_X(r) = \sum_{r=0}^n \binom{n}{r} p^r (1 - p)^{n-r} = [p + (1 - p)]^n = 1,$$

dove si è fatto uso della formula del binomio di Newton. La locuzione “binomiale” deriva appunto dal ruolo qui giocato dallo sviluppo della potenza n -esima del binomio. La funzione di distribuzione di X è poi immediatamente ottenibile:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \sum_{i=0}^k \binom{n}{i} p^i (1-p)^{n-i}, & k \leq x < k+1 \quad (k = 0, 1, \dots, n-1) \\ 1, & x \geq n. \end{cases} \quad (4.6)$$

Nel caso particolare $n = 1$, essa si riduce alla funzione di distribuzione di Bernoulli di parametro p .

Nel seguito con la notazione $X \sim \mathcal{B}(n, p)$ intenderemo che X è una variabile aleatoria con distribuzione binomiale di parametri n e p , che chiameremo anche *variabile binomiale*.

Esempio 4.4 Sia $X \sim \mathcal{B}(n, p)$. Posto $Y = n - X$, risulta $Y \sim \mathcal{B}(n, 1 - p)$. Infatti, poiché X assume i valori $0, 1, \dots, n$, i possibili valori di Y sono ancora $0, 1, \dots, n$. Pertanto, per $k = 0, 1, \dots, n$ si ha:

$$p_Y(k) = P(Y = k) = P(X = n - k) = p_X(n - k) = \binom{n}{k} p^{n-k} (1-p)^k,$$

ossia $Y \sim \mathcal{B}(n, 1 - p)$. ◇

Dalla (4.5) si ricava:

$$\frac{p_X(r)}{p_X(r-1)} = \frac{p}{1-p} \frac{n-r+1}{r}, \quad r = 1, 2, \dots, n, \quad (4.7)$$

da cui segue che le probabilità binomiali (4.5) sono calcolabili ricorsivamente al seguente modo:

$$p_X(0) = (1-p)^n, \quad p_X(r) = \frac{p}{1-p} \frac{n-r+1}{r} p_X(r-1), \quad r = 1, 2, \dots, n.$$

Inoltre, dalla (4.7) si trae la seguente

Osservazione 4.1 Sia $X \sim \mathcal{B}(n, p)$. Allora, per $r = 1, 2, \dots, n$ si ha:

- (i) $p_X(r) > p_X(r-1)$ per $r < (n+1)p$;
- (ii) $p_X(r) = p_X(r-1)$ per $r = (n+1)p$;
- (iii) $p_X(r) < p_X(r-1)$ per $r > (n+1)p$.

Da questa osservazione, ricordando la Definizione 4.3, è possibile dedurre quanto segue:

- (a) se $(n+1)p < 1$, allora $p_X(r)$ è strettamente decrescente per $r = 0, 1, \dots, n$;
- (b) se $(n+1)p > n$, allora $p_X(r)$ è strettamente crescente per $r = 0, 1, \dots, n$;
- (c) se $(n+1)p$ è un intero ed inoltre $1 \leq (n+1)p \leq n$, allora $p_X(r)$ presenta due massimi rispettivamente in $r = (n+1)p - 1$ e $r = (n+1)p$;

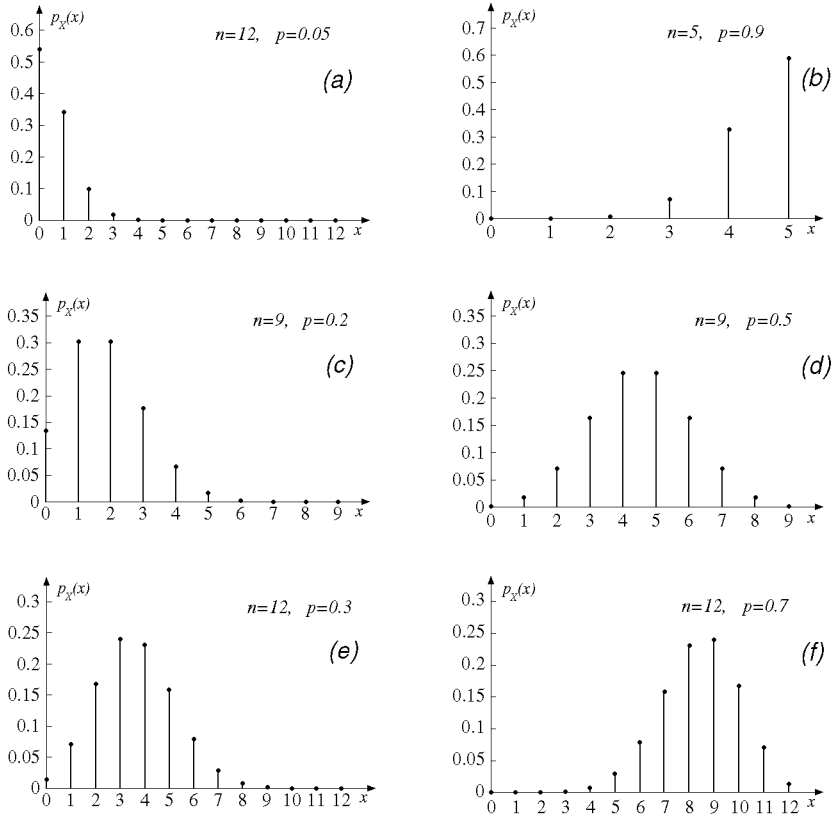


Figura 4.1 – Rappresentazione della funzione di probabilità (4.5) per vari valori di n e p .

(d) se $(n+1)p$ non è un intero ed inoltre $1 < (n+1)p < n$, allora $p_X(r)$ presenta un unico massimo in $r = \lfloor (n+1)p \rfloor$, dove con la notazione $\lfloor x \rfloor$ si denota il massimo intero non maggiore di x .

In Figura 4.1 sono mostrati alcuni grafici della funzione $p_X(x)$ per $x = 0, 1, \dots, n$ e per vari valori di p . In (a) e (b) $p_X(x)$ è strettamente monotona in x : decrescente nel primo caso, crescente nel secondo; in (c) e (d) $p_X(x)$ è massima in due valori di x . Si noti che per $n = 9$ e $p = 0.5$, come si vede in (d), la funzione $p_X(x)$ è simmetrica. Invero, se $X \sim \mathcal{B}(n, 1/2)$, risulta $p_X(x) = p_X(n-x)$ per $x = 0, 1, \dots, n$. Un'ulteriore proprietà di simmetria della distribuzione binomiale, illustrata in (e) ed (f), segue dall'osservazione che se $X \sim \mathcal{B}(n, p)$ e $Y \sim \mathcal{B}(n, 1-p)$, allora $p_X(x) = p_Y(n-x)$ per $x = 0, 1, \dots, n$.

Esempio 4.5 Si supponga che un'azienda realizzi dei componenti elettronici e che ciascuno di essi, indipendentemente dagli altri, sia affetto da imperfezioni con probabilità 5%. Calcoliamo la probabilità che in un lotto di 75 componenti almeno 2 siano imperfetti.

Sia X la variabile aleatoria che descrive il numero di componenti imperfetti del lotto;

evidentemente, $X \sim \mathcal{B}(75, 1/20)$. Pertanto la probabilità richiesta è la seguente:

$$P(X \geq 2) = \sum_{r=2}^{75} p_X(r) = \sum_{r=2}^{75} \binom{75}{r} \left(\frac{1}{20}\right)^r \left(1 - \frac{1}{20}\right)^{75-r} = 0.8944.$$

Si noti che il risultato desiderato può essere ottenuto molto più speditamente come segue:

$$\begin{aligned} P(X \geq 2) &= 1 - P(X < 2) = 1 - p_X(0) - p_X(1) \\ &= 1 - \binom{75}{0} \left(\frac{1}{20}\right)^0 \left(1 - \frac{1}{20}\right)^{75-0} - \binom{75}{1} \left(\frac{1}{20}\right)^1 \left(1 - \frac{1}{20}\right)^{75-1} \\ &= 1 - \left(\frac{19}{20}\right)^{75} - \frac{75}{20} \left(\frac{19}{20}\right)^{74} = 1 - 0.0213 - 0.0843 = 0.8944. \end{aligned}$$

◇

Esempio 4.6 Si supponga di estrarre n biglie con reinserimento da un'urna contenente N biglie, delle quali m sono azzurre e $N - m$ bianche. Calcoliamo la probabilità che nel campione estratto vi siano k biglie azzurre.

Si osservi a tal fine che la probabilità che in una singola estrazione si ottenga una biglia azzurra è m/N . Pertanto, indicata con X la variabile aleatoria che denota il numero di biglie azzurre presenti nel campione estratto, in virtù dell'indipendenza delle estrazioni risulta $X \sim \mathcal{B}(n, m/N)$, così che

$$p_X(k) = \binom{n}{k} \left(\frac{m}{N}\right)^k \left(1 - \frac{m}{N}\right)^{n-k}, \quad k = 0, 1, \dots, n$$

◇

Proposizione 4.1 Se X_1, X_2, \dots, X_n sono variabili aleatorie indipendenti e identicamente distribuite con $X_i \sim \mathcal{B}(1, p)$ per $i = 1, 2, \dots, n$, posto $X = X_1 + X_2 + \dots + X_n$ risulta $X \sim \mathcal{B}(n, p)$.

Dimostrazione Poiché ciascuna delle X_i può assumere i valori 0 o 1, i possibili valori di X sono $0, 1, \dots, n$. Inoltre, per $k = 0, 1, \dots, n$, si ha $X = k$ se e solo se esattamente k delle suddette variabili aleatorie assumono il valore 1 e le rimanenti $n - k$ il valore 0. Poiché X_1, X_2, \dots, X_n sono indipendenti e identicamente distribuite, X ha distribuzione binomiale di parametri n e p . □

Proposizione 4.2 Se Y_1, Y_2, \dots, Y_k sono variabili aleatorie indipendenti con $Y_1 \sim \mathcal{B}(n_1, p)$, $Y_2 \sim \mathcal{B}(n_2, p)$, \dots , $Y_k \sim \mathcal{B}(n_k, p)$, posto $X = Y_1 + Y_2 + \dots + Y_k$ risulta $X \sim \mathcal{B}(n_1 + n_2 + \dots + n_k, p)$.

Dimostrazione I possibili valori di X sono $0, 1, \dots, n_1 + n_2 + \dots + n_k$. Poiché, poi, ognuna delle Y_i può essere espressa come somma di n_i variabili aleatorie indipendenti aventi distribuzione di Bernoulli di parametro p , dall'indipendenza delle Y_i segue che la loro somma, X , è interpretabile come somma di $n_1 + n_2 + \dots + n_k$ variabili aleatorie indipendenti con identica distribuzione di Bernoulli, così che essa, in virtù della Proposizione 4.1, ha distribuzione binomiale di parametri $n_1 + n_2 + \dots + n_k$ e p . □

4.2.4 Distribuzione ipergeometrica

La distribuzione ipergeometrica interviene specificamente nella descrizione di estrazioni senza reinserimento oppure di estrazioni in blocco.

Si consideri l'esperimento che consiste nell'estrarre n biglie senza reinserimento da un'urna contenente N biglie, di cui m sono azzurre e $N - m$ sono bianche ($0 \leq n \leq N$, $0 \leq m \leq N$) e si consideri l'evento

$$E_k = \{k \text{ delle } n \text{ biglie estratte sono azzurre}\} \quad (k = 0, 1, \dots, n).$$

Facendo ricorso alla definizione classica di probabilità, si ha:

$$P(E_k) = \frac{\binom{m}{k} \binom{N-m}{n-k}}{\binom{N}{n}}, \quad (4.8)$$

dove il numeratore fornisce il numero di modi in cui si possono estrarre k delle m biglie azzurre e $n - k$ delle $N - m$ biglie bianche presenti nell'urna, mentre il denominatore dà il numero di modi in cui si possono estrarre n delle N biglie contenute nell'urna. Evidentemente deve essere $0 \leq k \leq m$ e $0 \leq n - k \leq N - m$, ossia $\max\{0, n + m - N\} \leq k \leq \min\{n, m\}$.

Se si indica con X la variabile aleatoria che descrive il numero di biglie azzurre estratte senza reinserimento, risulta $X = k$ se e solo se si verifica l'evento E_k ; pertanto risulta $P(X = k) = P(E_k)$ per $\max\{0, n + m - N\} \leq k \leq \min\{n, m\}$.

Definizione 4.4 Una variabile aleatoria X di funzione di probabilità

$$p_X(x) = \begin{cases} \frac{\binom{m}{x} \binom{N-m}{n-x}}{\binom{N}{n}}, & \max\{0, n + m - N\} \leq x \leq \min\{n, m\} \\ 0, & \text{altrimenti,} \end{cases} \quad (4.9)$$

con n, m e N interi tali che $0 \leq n \leq N$ e $0 \leq m \leq N$, è detta avere distribuzione ipergeometrica di parametri $n, m, N - m$.

Evidentemente, risulta $p_X(x) \geq 0$ per ogni $x \in \mathbb{R}$. Facendo uso della formula di Vandermonde

$$\sum_{r=\max\{0, n+m-N\}}^{\min\{n, m\}} \binom{m}{r} \binom{N-m}{n-r} = \binom{N}{n}$$

è poi possibile dimostrare che si ha:

$$\sum_{r=\max\{0, n+m-N\}}^{\min\{n, m\}} p_X(r) = \frac{1}{\binom{N}{n}} \sum_{r=\max\{0, n+m-N\}}^{\min\{n, m\}} \binom{m}{r} \binom{N-m}{n-r} = 1.$$

Nel seguito con la notazione $X \sim \mathcal{I}(n, m, N - m)$ intenderemo che X è una variabile aleatoria avente distribuzione ipergeometrica di parametri $n, m, N - m$; X sarà anche detta *variabile ipergeometrica*.

Si noti che, in virtù della (4.9), le variabili aleatorie $X \sim \mathcal{I}(n, m, N - m)$ e $Y \sim \mathcal{I}(m, n, N - n)$ sono identicamente distribuite. Sempre dalla (4.9) si ottiene poi:

$$\frac{p_X(r)}{p_X(r-1)} = \frac{m-r+1}{r} \frac{n-r+1}{N-n-m+r}, \quad \max\{0, n+m-N\} < r \leq \min\{n, m\},$$

da cui è possibile dedurre quanto espresso dall'osservazione seguente.

Osservazione 4.2 Sia $X \sim \mathcal{I}(n, m, N - m)$. Allora per $r = \max\{0, n+m-N\} + 1, \dots, \min\{n, m\}$ si ha:

- (i) $p_X(r) > p_X(r-1)$ per $r < (n+1)(m+1)/(N+2)$;
- (ii) $p_X(r) = p_X(r-1)$ per $r = (n+1)(m+1)/(N+2)$;
- (iii) $p_X(r) < p_X(r-1)$ per $r > (n+1)(m+1)/(N+2)$.

Esempio 4.7 In una classe di n maschi e n femmine si sceglie a caso un gruppo di n individui. Calcoliamo la probabilità che nel gruppo siano presenti k maschi ($k = 0, 1, \dots, n$).

Se si denota con X la variabile aleatoria che descrive il numero di maschi presenti nel gruppo scelto, dalla (4.9) si ottiene:

$$p_X(k) = \frac{\binom{n}{k} \binom{2n-n}{n-k}}{\binom{2n}{n}} = \frac{\left[\binom{n}{k}\right]^2}{\binom{2n}{n}} \quad (k = 0, 1, \dots, n). \quad (4.10)$$

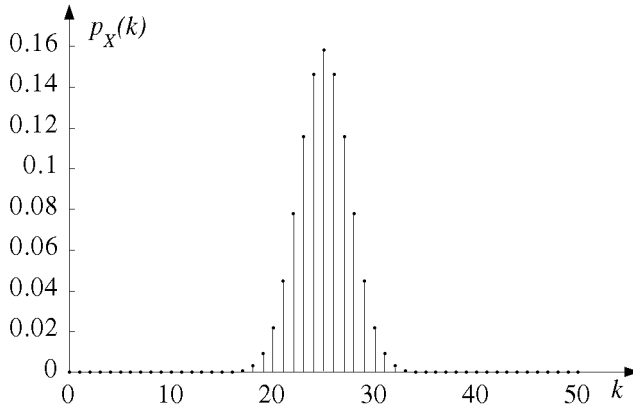


Figura 4.2 – Rappresentazione delle probabilità (4.10) per $n = 50$.

Poiché la (4.10) mostra che $p_X(k) = p_X(n-k)$ per $k = 0, 1, \dots, n$, facendo anche uso dell'Osservazione 4.2, segue che se n è pari $p_X(k)$ presenta un unico massimo in $n/2$ mentre

se n è dispari $p_X(k)$ presenta due massimi, rispettivamente in $(n-1)/2$ e in $(n+1)/2$. In Figura 4.2 è evidenziato il caso $n = 50$. \diamond

Esempio 4.8 Una fabbrica produce in un dato giorno 200 manufatti 60 dei quali sono imperfetti. Si seleziona a caso un lotto di 10 manufatti da sottoporre a controllo. Calcoliamo la probabilità che nel lotto selezionato siano presenti k manufatti imperfetti, con $k = 0, 1, \dots, 10$.

Denotando con X la variabile aleatoria che descrive il numero di manufatti imperfetti presenti nel lotto selezionato, risulta che $X \sim \mathcal{I}(10, 60, 200)$. La probabilità che k di quelli sottoposti a controllo siano difettosi è pertanto:

$$p_X(k) = \frac{\binom{60}{k} \binom{140}{10-k}}{\binom{200}{10}} \quad (k = 0, 1, \dots, 10). \quad (4.11)$$

Tabella 4.1 – Le probabilità $P(X = k)$ e $P(X \leq k)$ per $k = 0, 1, \dots, 10$ ottenute dalla (4.11).

k	$P(X = k)$	$P(X \leq k)$
0	0.025552	0.025552
1	0.117030	0.142582
2	0.235390	0.377972
3	0.273737	0.651709
4	0.203771	0.855480
5	0.101433	0.956912
6	0.034184	0.991096
7	0.007699	0.998796
8	0.001109	0.999904
9	0.000092	0.999997
10	0.000003	1

Si noti che la percentuale di manufatti imperfetti tra i 200 prodotti è del 30%. Affinché anche la percentuale di manufatti imperfetti tra i 10 selezionati sia del 30% deve risultare $k = 3$; dalla Tabella 4.1 si vede che ciò si realizza con probabilità 0.273737. Si osservi inoltre che la probabilità che nessuno dei 10 manufatti sottoposti a controllo risulti imperfetto è $P(X = 0) = 0.025552$. \diamond

Vogliamo ora mostrare che sotto opportune ipotesi la distribuzione ipergeometrica tende alla distribuzione binomiale.

Proposizione 4.3 Sia $X \sim \mathcal{I}(n, m, N - m)$; se n è fissato, e se N e m divergono in modo tale che il loro rapporto converga ad un valore $p \in (0, 1)$, allora

$$\lim_{\substack{N \rightarrow +\infty, \quad m \rightarrow +\infty \\ m/N \rightarrow p}} p_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (k = 0, 1, \dots, n). \quad (4.12)$$

Dimostrazione Dalla (4.9) segue:

$$p_X(k) = \binom{n}{k} \frac{m(m-1) \cdots (m-k+1)}{N(N-1) \cdots (N-k+1)} \times \frac{(N-m)(N-m-1) \cdots (N-m-n+k+1)}{(N-k)(N-k-1) \cdots (N-n+1)}. \quad (4.13)$$

Poiché

$$\begin{aligned} \lim_{\substack{N \rightarrow +\infty, m \rightarrow +\infty \\ m/N \rightarrow p}} \frac{m(m-1) \cdots (m-k+1)}{N(N-1) \cdots (N-k+1)} \\ = \lim_{\substack{N \rightarrow +\infty, m \rightarrow +\infty \\ m/N \rightarrow p}} \frac{\frac{m}{N} \left(\frac{m}{N} - \frac{1}{N} \right) \cdots \left(\frac{m}{N} - \frac{k-1}{N} \right)}{1 \left(1 - \frac{1}{N} \right) \cdots \left(1 - \frac{k-1}{N} \right)} = p^k \end{aligned}$$

e

$$\begin{aligned} \lim_{\substack{N \rightarrow +\infty, m \rightarrow +\infty \\ m/N \rightarrow p}} \frac{(N-m)(N-m-1) \cdots (N-m-n+k+1)}{(N-k)(N-k-1) \cdots (N-n+1)} \\ = \lim_{\substack{N \rightarrow +\infty, m \rightarrow +\infty \\ m/N \rightarrow p}} \frac{\left(1 - \frac{m}{N} \right) \left(1 - \frac{m}{N} - \frac{1}{N} \right) \cdots \left(1 - \frac{m}{N} - \frac{n-k-1}{N} \right)}{\left(1 - \frac{k}{N} \right) \left(1 - \frac{k+1}{N} \right) \cdots \left(1 - \frac{n-1}{N} \right)} = (1-p)^{n-k}, \end{aligned}$$

dalla (4.13) segue la (4.12). □

Con riferimento allo schema di estrazione che ha condotto alla formula (4.8), la Proposizione 4.3 comporta che se il numero m delle biglie azzurre e il numero N di biglie presenti nell'urna sono entrambi sufficientemente elevati in modo tale che il loro rapporto sia una costante p , allora la probabilità che k delle n biglie estratte senza reinserimento siano azzurre è approssimabile con la medesima probabilità relativa al caso di estrazioni con reinserimento, essendo

$$\frac{\binom{m}{k} \binom{N-m}{n-k}}{\binom{N}{n}} \simeq \binom{n}{k} \left(\frac{m}{N} \right)^k \left(1 - \frac{m}{N} \right)^{n-k} \quad (k = 0, 1, \dots, n). \quad (4.14)$$

Nella Tabella 4.2 si confrontano le probabilità presenti al primo ed al secondo membro della (4.14) per $n = 5$ e varie scelte di m ed N tali che $m/N = 0.4$. Si noti che l'approssimazione della distribuzione ipergeometrica con quella binomiale, già buona per $m = 40$ e $N = 100$, tende a migliorare ulteriormente al crescere di N .

Tabella 4.2 – Confronto tra $P(X = k)$ e $P(Y = k)$ con $X \sim \mathcal{I}(5, m, N - m)$ e $Y \sim \mathcal{B}(5, 0.4)$.

k	$P(X = k)$ $m = 40$ $N = 100$	$P(X = k)$ $m = 400$ $N = 1000$	$P(X = k)$ $m = 4000$ $N = 10000$	$P(X = k)$ $m = 40000$ $N = 100000$	$P(Y = k)$
0	0.0725421	0.0772413	0.0777082	0.0777548	0.07776
1	0.259079	0.259199	0.2592	0.2592	0.2592
2	0.354529	0.346467	0.345686	0.345609	0.3456
3	0.232278	0.230592	0.230419	0.230402	0.2304
4	0.0728328	0.0764147	0.0767616	0.0767962	0.0768
5	0.00873993	0.0100867	0.0102246	0.0102385	0.01024

4.2.5 Distribuzione geometrica

Si consideri l'esperimento consistente in una successione di prove ripetute di Bernoulli di parametro $p \in (0, 1)$, il cui spazio campionario è quello di Bernoulli (v. Esempio 2.1). Si supponga di essere interessati all'evento

$$E_r = \{\text{il primo successo si verifica alla prova } r\text{-esima}\} \quad (r = 1, 2, \dots).$$

Dall'ipotesi di indipendenza delle prove si ricava che $P(E_r) = (1 - p)^{r-1} p$. Sia X la variabile aleatoria che descrive il numero di prove necessarie per ottenere il primo successo; è evidente che $P(X = r) = P(E_r)$ per $r = 1, 2, \dots$.

Definizione 4.5 Una variabile aleatoria X di funzione di probabilità

$$p_X(x) = \begin{cases} p(1 - p)^{x-1}, & x = 1, 2, \dots \\ 0, & \text{altrimenti,} \end{cases} \quad (4.15)$$

con $0 < p < 1$ si dice avere distribuzione geometrica di parametro p .

Per ogni $p \in (0, 1)$ la (4.15) individua una funzione di probabilità; infatti risulta $p_X(x) \geq 0$ per ogni $x \in \mathbb{R}$ ed inoltre è soddisfatta la condizione di normalizzazione:

$$\sum_{r=1}^{+\infty} p_X(r) = \sum_{r=1}^{+\infty} p(1 - p)^{r-1} = p \sum_{s=0}^{+\infty} (1 - p)^s = \frac{p}{1 - (1 - p)} = 1.$$

La locuzione “geometrica” deriva appunto dal ruolo qui giocato dallo sviluppo della serie geometrica di ragione $1 - p$. Poiché

$$\sum_{r=1}^k p_X(r) = \sum_{r=1}^k p(1 - p)^{r-1} = p \sum_{s=0}^{k-1} (1 - p)^s = p \frac{1 - (1 - p)^k}{1 - (1 - p)} = 1 - (1 - p)^k,$$

la funzione di distribuzione di X è la seguente:

$$F_X(x) = \begin{cases} 0, & x < 1 \\ 1 - (1 - p)^k, & k \leq x < k + 1 \end{cases} \quad (k = 1, 2, \dots). \quad (4.16)$$

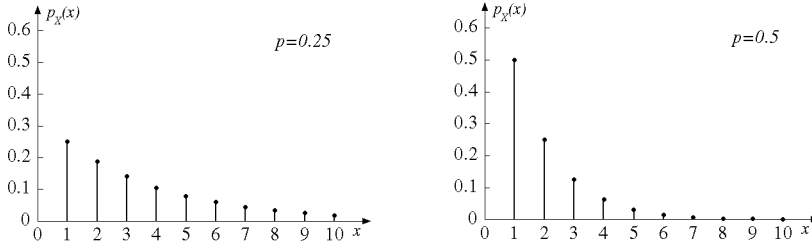


Figura 4.3 – Rappresentazione della funzione di probabilità (4.15) per due diverse scelte di p .

Dalla (4.15) segue immediatamente che $p_X(r)$ è strettamente decrescente in $r = 1, 2, \dots$, come è anche evidenziato nella Figura 4.3.

Un'interessante proprietà della distribuzione geometrica è la seguente.

Proposizione 4.4 *Se X è una variabile aleatoria con distribuzione geometrica, risulta:*

$$P(X > r + n \mid X > r) = P(X > n), \quad (4.17)$$

con r e n interi non negativi.

Dimostrazione Ricordando la definizione di probabilità condizionata, dalla (4.16) si ha:

$$\begin{aligned} P(X > r + n \mid X > r) &= \frac{P(X > r + n, X > r)}{P(X > r)} = \frac{P(X > r + n)}{P(X > r)} \\ &= \frac{1 - F_X(r + n)}{1 - F_X(r)} = \frac{(1 - p)^{r+n}}{(1 - p)^r} = (1 - p)^n = P(X > n), \end{aligned}$$

ossia la tesi. \square

Per illustrare il significato della (4.17) si consideri una successione di prove ripetute di Bernoulli. Se nelle prime r prove non si è avuto nessun successo, la probabilità che non si verifichi alcun successo fino alla prova $r + n$ non dipende da r , ossia da quanto si è atteso, ma solo dal numero n di prove ancora da effettuarsi. La (4.17) esprime dunque la proprietà di *assenza di memoria* della distribuzione geometrica.

Esempio 4.9 Calcoliamo la probabilità che nel gioco del lotto il numero 20 non esca in n successive estrazioni su una fissata ruota.

Sia $B_k = \{\text{nell'estrazione } k\text{-esima esce il numero 20 sulla fissata ruota}\}$ ($k = 1, 2, \dots$). Chiaramente gli eventi B_k sono indipendenti in quanto si riferiscono a distinte estrazioni. Dalla definizione classica di probabilità risulta:

$$P(B_k) = \frac{\binom{89}{4}}{\binom{90}{5}} = \frac{1}{18} \quad (k = 1, 2, \dots).$$

Sia ora X la variabile aleatoria che descrive il numero di estrazioni necessarie affinché si presenti per la prima volta il numero 20 sulla fissata ruota. Essa ha distribuzione geometrica di parametro $1/18$. Se si pone $A_n = \{\text{il numero 20 non esce nelle prossime } n \text{ estrazioni sulla fissata ruota}\}$, allora la probabilità richiesta è

$$P(A_n) = P(X > n) = \left(1 - \frac{1}{18}\right)^n = \left(\frac{17}{18}\right)^n \quad (n = 1, 2, \dots),$$

avendo fatto uso della (4.16). Si noti che la probabilità che il numero 20 non esca in n estrazioni successive è strettamente decrescente in n ; ciò mostra che ritardi via via più lunghi diventano sempre meno probabili. Infine, dalla (4.17) segue $P(A_{r+n}|A_r) = P(A_n)$; pertanto, la probabilità che il numero 20 non esca nelle successive $r+n$ estrazioni, avendo osservato che esso non è uscito nelle prime r estrazioni, non dipende da r , ossia dal cosiddetto “ritardo” dell’evento. \diamond

Esempio 4.10 Si consideri l’esperimento consistente nel lanciare ripetutamente una coppia di monete supponendo che ad ogni lancio le occorrenze di Testa avvengano con probabilità p_1 e p_2 , rispettivamente. Calcoliamo la probabilità che per la prima volta escano due teste nello stesso lancio.

Se indichiamo con X_i la variabile aleatoria che descrive il numero di lanci necessari per avere testa per la prima volta nel lancio della moneta i -esima ($i = 1, 2$), X_1 e X_2 sono indipendenti ed hanno distribuzione geometrica rispettivamente di parametri p_1 e p_2 . La probabilità desiderata è pertanto esprimibile come $P(X_1 = X_2)$. In virtù dell’indipendenza di X_1 e X_2 , dalla (4.15) risulta:

$$\begin{aligned} P(X_1 = X_2) &= \sum_{k=1}^{+\infty} P(X_1 = k, X_2 = k) = \sum_{k=1}^{+\infty} P(X_1 = k) P(X_2 = k) \\ &= p_1 p_2 \sum_{k=1}^{+\infty} [(1 - p_1)(1 - p_2)]^{k-1} = \frac{p_1 p_2}{p_1 + p_2 - p_1 p_2}. \end{aligned}$$

Si noti che se l’uscita di testa ad ogni lancio avviene per entrambe le monete con la stessa probabilità $p = p_1 = p_2$, risulta:

$$P(X_1 = X_2) = \frac{p}{2 - p}.$$

Se ne trae che $P(X_1 = X_2)$ è strettamente crescente in $p \in (0, 1)$. In particolare, nel caso di monete eque ($p = 1/2$), risulta $P(X_1 = X_2) = 1/3$. \diamond

4.2.6 Distribuzione binomiale negativa

Si consideri nuovamente l’esperimento consistente in una successione di prove ripetute di Bernoulli di parametro p e si supponga di essere interessati all’evento

$$E_{n,r} = \{\text{il successo } n\text{-esimo si verifica alla prova } r\text{-esima}\} \quad (r = n, n + 1, \dots).$$

Evidentemente $E_{n,r}$ si verifica allorché alla r -esima prova si ha un successo, il che avviene con probabilità p , mentre nelle precedenti $r - 1$ prove si sono avuti in un ordine qualsiasi $n - 1$ successi e $r - n$ insuccessi, il che si verifica con probabilità

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

Pertanto, in virtù dell'indipendenza delle prove, si ha:

$$P(E_{n,r}) = p \binom{r-1}{n-1} p^{n-1} (1-p)^{r-n} = \binom{r-1}{n-1} p^n (1-p)^{r-n} \quad (r = n, n+1, \dots).$$

Se X_n è la variabile aleatoria che descrive il numero di prove necessarie per ottenere il successo n -esimo, si ha $P(X_n = r) = P(E_{n,r})$ per $r = n, n+1, \dots$.

Definizione 4.6 Una variabile aleatoria X_n di funzione di probabilità

$$p_{X_n}(x) = \begin{cases} \binom{x-1}{n-1} p^n (1-p)^{x-n}, & x = n, n+1, \dots \\ 0, & \text{altrimenti,} \end{cases} \quad (4.18)$$

con $0 < p < 1$ e n intero positivo, si dice avere distribuzione binomiale negativa di parametri n e p .

Si noti che si ha $p_{X_n}(x) \geq 0$ per ogni $x \in \mathbb{R}$, e che inoltre risulta:

$$\begin{aligned} \sum_{r=n}^{+\infty} p_{X_n}(r) &= \sum_{r=n}^{+\infty} \binom{r-1}{n-1} p^n (1-p)^{r-n} = \sum_{j=0}^{+\infty} \binom{j+n-1}{n-1} p^n (1-p)^j \\ &= p^n \sum_{j=0}^{+\infty} \binom{j+n-1}{n-1} (1-p)^j = \frac{p^n}{[1 - (1-p)]^n} = 1, \end{aligned}$$

avendo fatto uso dell'identità:

$$(1-\alpha)^{-n} = \sum_{j=0}^{+\infty} \binom{j+n-1}{n-1} \alpha^j, \quad |\alpha| < 1.$$

Come è da attendersi, dal confronto della (4.15) con la (4.18) si verifica che il caso $n = 1$ conduce alla distribuzione geometrica. Nel seguito si userà la notazione $X \sim \mathcal{BN}(n, p)$ per indicare che X ha distribuzione binomiale negativa di parametri n e p ; X sarà anche detta *variabile binomiale negativa*. Conseguentemente $X \sim \mathcal{BN}(1, p)$ indicherà che X ha distribuzione geometrica di parametro p ; diremo anche che X è una *variabile geometrica*.

Notiamo che dalla (4.18) si ricava:

$$\frac{p_X(r)}{p_X(r-1)} = \frac{\binom{r-1}{n-1} p^n (1-p)^{r-n}}{\binom{r-2}{n-1} p^n (1-p)^{r-n-1}} = \frac{r-1}{r-n} (1-p) \quad (r = n+1, n+2, \dots), \quad (4.19)$$

che mostra che le probabilità binomiali negative (4.18) sono calcolabili in modo ricorsivo come segue:

$$p_X(n) = p^n, \quad p_X(r) = (1-p) \frac{r-1}{r-n} p_X(r-1) \quad (r = n+1, n+2, \dots).$$

Inoltre dalla (4.19) si trae la seguente

Osservazione 4.3 Se $X \sim \mathcal{BN}(n, p)$, per $r = n+1, n+2, \dots$ si ha:

- (i) $p_X(r) > p_X(r-1)$ per $r < 1 + (n-1)/p$;
- (ii) $p_X(r) = p_X(r-1)$ per $r = 1 + (n-1)/p$;
- (iii) $p_X(r) < p_X(r-1)$ per $r > 1 + (n-1)/p$.

Da questa osservazione, ricordando la Definizione 4.6, segue:

- (a) se $(n-1)/p < n$, allora $p_X(r)$ è strettamente decrescente;
- (b) se $(n-1)/p$ è un intero ed inoltre $(n-1)/p \geq n$, allora $p_X(r)$ presenta due massimi in $r = (n-1)/p$ ed in $r = 1 + (n-1)/p$;
- (c) se $(n-1)/p$ non è un intero ed inoltre $(n+1)/p > n$, allora $p_X(r)$ presenta un unico massimo in $r = \lfloor 1 + (n-1)/p \rfloor$.

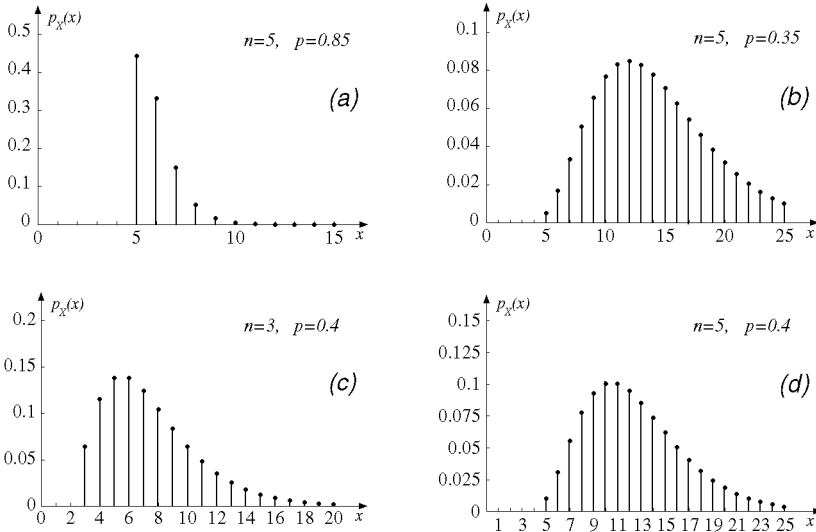


Figura 4.4 – Rappresentazione della funzione di probabilità (4.18) per vari valori di n e p .

In Figura 4.4 sono mostrati alcuni grafici della funzione $p_X(x)$ per $x = n, n+1, n+2, \dots$ e per varie scelte di n e p . In (a) $p_X(x)$ è strettamente decrescente, in (b) presenta un unico massimo e nei rimanenti casi $p_X(x)$ è massima in due valori di x .

Esempio 4.11 Ad un torneo partecipano due squadre, A e B. Il torneo è articolato in 5 partite in ognuna delle quali, indipendentemente da ciò che accade nelle altre, la squadra A ha probabilità p di vincere, mentre è $1 - p$ la probabilità che vinca la squadra B, restando così esclusa la possibilità del pareggio. Il torneo termina allorché una delle due squadre vince tre partite. Calcoliamo le probabilità di vincita del torneo da parte di ciascuna squadra.

Affinché una squadra vinca il torneo è dunque necessario che essa vinca tre partite (il che automaticamente esclude che l'altra squadra possa vincere, essendo 5 il numero complessivo di partite del torneo). Se X denota la variabile aleatoria che rappresenta il numero di partite da giocare fino a che la squadra A vince per la terza volta, risulta $X \sim \mathcal{BN}(3, p)$. Pertanto, considerando l'evento $E_1 = \{\text{la squadra A vince il torneo}\}$ si ha:

$$\begin{aligned} P(E_1) &= P(X = 3) + P(X = 4) + P(X = 5) = \sum_{r=3}^5 \binom{r-1}{2} p^3 (1-p)^{r-3} \\ &= p^3 + 3p^3(1-p) + 6p^3(1-p)^2. \end{aligned}$$

Analogamente, se Y è la variabile aleatoria che rappresenta il numero di partite necessarie perché la squadra B vinca per la terza volta, si ha $Y \sim \mathcal{BN}(3, 1-p)$; pertanto, posto $E_2 = \{\text{la squadra B vince il torneo}\}$, risulta:

$$\begin{aligned} P(E_2) &= P(Y = 3) + P(Y = 4) + P(Y = 5) = \sum_{r=3}^5 \binom{r-1}{2} (1-p)^3 p^{r-3} \\ &= (1-p)^3 + 3(1-p)^3 p + 6(1-p)^3 p^2 = 1 - P(E_1). \end{aligned}$$

Si noti che la probabilità che il gioco termini è unitaria, essendo $P(E_1 \cup E_2) = P(E_1) + P(E_2) = 1$. \diamond

Proposizione 4.5 Siano $X_i \sim \mathcal{BN}(1, p)$ con $i = 1, 2, \dots, n$ variabili aleatorie indipendenti. Posto $X = X_1 + X_2 + \dots + X_n$, risulta $X \sim \mathcal{BN}(n, p)$.

Dimostrazione Avendo distribuzione geometrica, X_1 può riguardarsi come il tempo di attesa per il primo successo in una successione di prove ripetute di Bernoulli, ossia come il numero di prove necessarie per ottenere il primo successo. Inoltre, in virtù dell'assenza di memoria della distribuzione geometrica, si può riguardare X_r ($r = 2, 3, \dots, n$) come il tempo di attesa del primo successo a partire dalla prova in cui si è verificato il successo $(r-1)$ -esimo, ossia come il numero di prove intercorrenti tra il successo $(r-1)$ -esimo ed il successo r -esimo. Per l'indipendenza delle prove, le variabili aleatorie X_1, X_2, \dots, X_n sono indipendenti e identicamente distribuite. Pertanto X descrive il tempo di attesa per il successo n -esimo che, per quanto visto, ha distribuzione binomiale negativa di parametri n e p . \square

Proposizione 4.6 Siano Y_1, Y_2, \dots, Y_k variabili aleatorie indipendenti con $Y_1 \sim \mathcal{BN}(n_1, p)$, $Y_2 \sim \mathcal{BN}(n_2, p)$, \dots , $Y_k \sim \mathcal{BN}(n_k, p)$, e sia $X = Y_1 + Y_2 + \dots + Y_k$. Allora risulta $X \sim \mathcal{BN}(n_1 + n_2 + \dots + n_k, p)$.

Dimostrazione I possibili valori di X sono $n, n+1, n+2, \dots$, avendo posto $n = n_1 + n_2 + \dots + n_k$. In virtù della Proposizione 4.5, ognuna delle Y_i può essere espressa come somma di n_i variabili aleatorie geometriche indipendenti di parametro p . Dall'indipendenza delle Y_i segue che la loro somma è interpretabile come somma di n variabili aleatorie geometriche indipendenti ed identicamente distribuite che, pertanto, ha distribuzione binomiale negativa di parametri $n_1 + n_2 + \dots + n_k$ e p . \square

4.2.7 Distribuzione di Poisson

La distribuzione di Poisson gioca un ruolo molto importante nel Calcolo delle Probabilità. Essa interviene spesso nella descrizione di alcuni fenomeni coinvolgenti qualche tipo di conteggio, quali il numero di chiamate telefoniche ricevute da un centralino in un fissato intervallo di tempo, il numero di particelle radioattive emesse per unità di tempo, il numero di microorganismi per unità di volume in un fluido, il numero di imperfezioni per unità di lunghezza di un cavo.

Definizione 4.7 Una variabile aleatoria X avente funzione di probabilità

$$p_X(x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda}, & x = 0, 1, \dots \quad (\lambda > 0), \\ 0, & \text{altrimenti} \end{cases} \quad (4.20)$$

è detta di distribuzione di Poisson di parametro λ .

Si noti che risulta $p_X(x) \geq 0$ per ogni $x \in \mathbb{R}$ e che è soddisfatta la condizione di normalizzazione:

$$\sum_{k=0}^{+\infty} p_X(k) = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

Nel seguito con la notazione $X \sim \mathcal{P}(\lambda)$ si indicherà che X è una variabile aleatoria avente distribuzione di Poisson di parametro λ , o più semplicemente che X è una *variabile di Poisson* o *poissoniana*. Dalla (4.20) si ricava:

$$\frac{p_X(r)}{p_X(r-1)} = \frac{\lambda}{r} \quad (r = 1, 2, \dots), \quad (4.21)$$

così che le probabilità di Poisson (4.20) sono calcolabili in modo ricorsivo:

$$p_X(0) = e^{-\lambda}, \quad p_X(r) = \frac{\lambda}{r} p_X(r-1) \quad (r = 1, 2, \dots).$$

Osservazione 4.4 Sia $X \sim \mathcal{P}(\lambda)$. Allora, per $r = 1, 2, \dots$ si ha:

- (i) $p_X(r) > p_X(r-1)$ se $\lambda < r$;
- (ii) $p_X(r) = p_X(r-1)$ se $\lambda = r$;
- (iii) $p_X(r) < p_X(r-1)$ se $\lambda > r$.

Da ciò e dalla Definizione 4.7 discende che

- (a) se $\lambda < 1$, $p_X(r)$ è strettamente decrescente per $r = 0, 1, \dots$;

- (b) se λ è intero positivo, $p_X(r)$ ha due massimi rispettivamente in $r = \lambda - 1$ ed in $r = \lambda$;
- (c) se λ non è intero ed è maggiore di 1, $p_X(r)$ ha un unico massimo in $r = \lfloor \lambda \rfloor$.

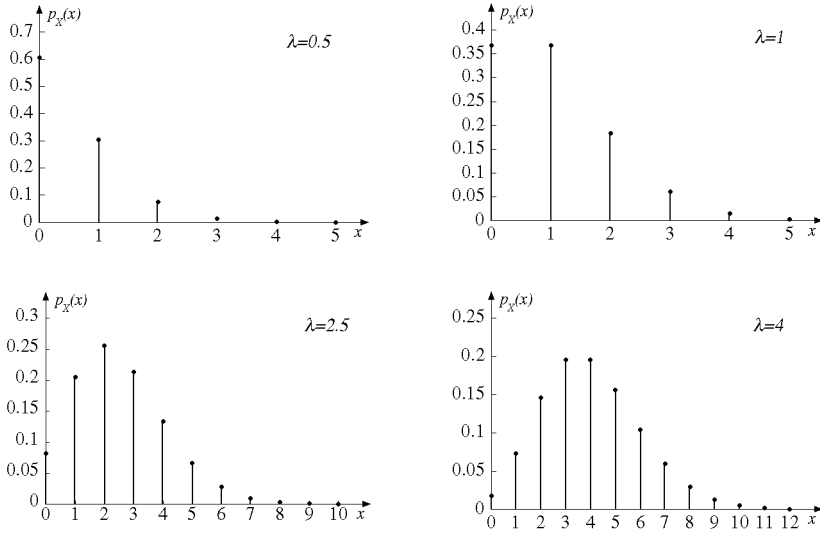


Figura 4.5 – Rappresentazione della funzione di probabilità (4.20) per vari valori di λ .

In Figura 4.5 sono riportati alcuni grafici di $p_X(x)$ con $x = 0, 1, 2, \dots$ ad illustrazione di tali proprietà. Nel primo caso ($\lambda = 0.5$), $p_X(x)$ è strettamente decrescente, mentre per $\lambda = 2.5$ essa presenta un unico massimo in $x = 2$; nei rimanenti casi, essendo λ intero, $p_X(r)$ presenta due massimi le cui ascisse aumentano all'aumentare di λ , mentre le ordinate diminuiscono.

La distribuzione di Poisson è spesso detta *degli eventi rari* o *dei piccoli numeri* poiché, come si vedrà qui di seguito, essa si rivela utile per trattare i cosiddetti eventi rari di schemi binomiali in cui la probabilità di successo in ogni singola prova è molto piccola mentre il numero di prove è molto grande, come spesso accade in numerosi fenomeni biologici (colonie di batteri, mutazioni genetiche), assicurativi (incidenti aerei, incendi), industriali (controllo statistico della qualità di prodotti), ecc.

Proposizione 4.7 Sia X_1, X_2, \dots una successione di variabili aleatorie con $X_n \sim \mathcal{B}(n, p_n)$. Se al divergere di n , p_n tende a zero in modo tale che $np_n \rightarrow \lambda$, con $\lambda > 0$, allora

$$\lim_{\substack{n \rightarrow +\infty, \\ np_n \rightarrow \lambda}} p_{X_n}(k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (k = 0, 1, \dots). \quad (4.22)$$

Dimostrazione Poiché per ipotesi $X_n \sim \mathcal{B}(n, p_n)$, dalla (4.5) si ricava:

$$p_{X_n}(k) = \binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{n(n-1) \cdots (n-k+1)}{n^k k!} (np_n)^k (1 - p_n)^{n-k}$$

$$= \frac{\left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)}{k!} (n p_n)^k \frac{\left(1 - \frac{n p_n}{n}\right)^n}{(1 - p_n)^k}. \quad (4.23)$$

Procediamo ora al limite per $n \rightarrow +\infty$, $p_n \rightarrow 0$ e $n p_n \rightarrow \lambda$ nella (4.23). Poiché²

$$\lim_{\substack{n \rightarrow +\infty, p_n \rightarrow 0 \\ n p_n \rightarrow \lambda}} \left(1 - \frac{n p_n}{n}\right)^n = e^{-\lambda},$$

la (4.22) segue immediatamente dalla (4.23). \square

Un'immediata conseguenza della Proposizione 4.7 è che se il numero n di prove ripetute di Bernoulli è elevato e se la probabilità di successo p in ogni prova è piccola, allora è possibile utilizzare la seguente approssimazione:

$$\binom{n}{k} p^k (1-p)^{n-k} \simeq \frac{(n p)^k}{k!} e^{-n p} \quad (k = 0, 1, \dots). \quad (4.24)$$

È opportuno menzionare che la distribuzione di Poisson costituisce una buona approssimazione della distribuzione binomiale quando nella (4.24) si ha $n \geq 20$ e $p \leq 0.05$; per $n \geq 100$ e $n p \leq 10$ l'approssimazione diviene poi eccellente.

Tabella 4.3 – Confronto tra $P(X = k)$ e $P(Y = k)$ con $X \sim \mathcal{B}(n, p)$ e $Y \sim \mathcal{P}(1)$ per alcune scelte di k .

k	$P(X = k)$ $n = 100$ $p = 0.01$	$P(X = k)$ $n = 1000$ $p = 0.001$	$P(X = k)$ $n = 10000$ $p = 0.0001$	$P(Y = k)$ $\lambda = 1$
0	0.366032	0.367695	0.367861	0.367879
1	0.369730	0.368063	0.367898	0.367879
2	0.184865	0.184032	0.183949	0.183940
3	0.060999	0.061282	0.061310	0.061313
4	0.014942	0.015290	0.015324	0.015328
5	0.002898	0.003049	0.003064	0.003066
6	0.000463	0.000506	0.000510	0.000511
7	0.000063	0.000072	0.000073	0.000073
8	0.000007	0.000009	0.000009	0.000009

In Tabella 4.3 sono riportati i valori delle probabilità binomiali per $n = 100, 1000, 10000$, scegliendo ogni volta il parametro p in modo che risulti $n p = 1$ (seconda, terza e quarta colonna). L'ultima colonna mostra i valori della probabilità di Poisson con parametro $\lambda = 1$. È evidente che l'approssimazione della distribuzione binomiale con la distribuzione di Poisson, già buona per $n = 100$ e $p = 0.01$ come si evince confrontando la seconda e la quinta colonna della Tabella 4.3, migliora al crescere di n e al decrescere di p , come emerge confrontando la terza e quarta colonna con la quinta colonna della Tabella 4.3.

²Qui si fa uso del seguente limite notevole $\lim_{n \rightarrow +\infty} \left(1 - \frac{\alpha_n}{n}\right)^n = e^{-\alpha}$, dove $\alpha_1, \alpha_2, \dots$ è una successione di reali tali da aversi $\lim_{n \rightarrow +\infty} \alpha_n = \alpha$.

Esempio 4.12 Si supponga che la probabilità che un autoveicolo si guasti all'interno di un certo tunnel sia 4×10^{-4} . Si desidera calcolare la probabilità che di 1000 autoveicoli che attraversano il tunnel se ne guasti al più uno.

Sia X una variabile aleatoria descrivente il numero di autoveicoli che si guastano all'interno del tunnel. Evidentemente $X \sim \mathcal{B}(1000, 4 \times 10^{-4})$, ossia X ha distribuzione binomiale di parametri $n = 1000$ e $p = 4 \times 10^{-4}$. Pertanto, considerato l'evento $A = \{\text{si guasta al più un autoveicolo ogni mille che attraversano il tunnel}\}$, si ha:

$$\begin{aligned} P(A) &= p_X(0) + p_X(1) = \binom{1000}{0} (4 \times 10^{-4})^0 (1 - 4 \times 10^{-4})^{1000} \\ &\quad + \binom{1000}{1} (4 \times 10^{-4})^1 (1 - 4 \times 10^{-4})^{999} \\ &= (0.9996)^{1000} + 1000 \cdot 4 \times 10^{-4} (0.9994)^{999} = 0.9385. \end{aligned}$$

Nel caso in esame, essendo $n = 1000$ e $np = 1000 \cdot 4 \times 10^{-4} = 0.4$, l'approssimazione di Poisson della distribuzione binomiale è ottima; infatti, ponendo $\lambda = 0.4$ si ha:

$$P(A) \simeq \frac{\lambda^0}{0!} e^{-\lambda} + \frac{\lambda^1}{1!} e^{-\lambda} = 1.4 \cdot e^{-0.4} = 0.9384$$

che differisce dal valore precedente sulla quarta cifra decimale. \diamond

Esempio 4.13 Siano X e Y variabili aleatorie indipendenti rispettivamente di distribuzioni $\mathcal{P}(\lambda)$ e $\mathcal{P}(\mu)$. Determiniamo la distribuzione della variabile aleatoria $Z = X + Y$.

A tal fine, si noti che per ogni $n \geq 0$ risulta:

$$P(Z = n) = P(X + Y = n) = \sum_{k=0}^n P(X = k, Y = n - k) = \sum_{k=0}^n P(X = k) P(Y = n - k),$$

dove l'ultima uguaglianza è conseguenza dell'ipotesi di indipendenza tra X e Y . Per l'ipotesi fatta sulle distribuzioni di X e Y , si ha infine:

$$\begin{aligned} P(Z = n) &= \sum_{k=0}^n \frac{\lambda^k}{k!} e^{-\lambda} \frac{\mu^{n-k}}{(n-k)!} e^{-\mu} = e^{-(\lambda+\mu)} \frac{1}{n!} \sum_{k=0}^n \binom{n}{k} \lambda^k \mu^{n-k} \\ &= \frac{(\lambda + \mu)^n}{n!} e^{-(\lambda+\mu)} \quad (n = 0, 1, \dots). \end{aligned}$$

Pertanto, $Z \sim \mathcal{P}(\lambda + \mu)$, ossia Z ha distribuzione di Poisson di parametro $\lambda + \mu$. \diamond

Dall'Esempio 4.13 si deduce, per iterazione, che date r variabili aleatorie indipendenti X_i di distribuzioni $\mathcal{P}(\lambda_i)$ ($i = 1, 2, \dots, r$), la variabile aleatoria somma $\sum_{i=1}^r X_i$ ha distribuzione di Poisson di parametro $\sum_{i=1}^r \lambda_i$.

4.3 Vettori aleatori discreti

Finora sono state studiate distribuzioni di variabili aleatorie discrete unidimensionali; introdurremo ora alcune distribuzioni di probabilità multivariate, relative quindi a vettori aleatori.

4.3.1 Distribuzione multinomiale

Come visto nel Paragrafo 4.2.2, una prova di Bernoulli è un esperimento i cui risultati definiscono due eventi incompatibili e necessari, interpretabili genericamente come successo ed insuccesso. A partire da una prova di Bernoulli è possibile definire una variabile aleatoria, coincidente con la funzione indicatrice, avente distribuzione di Bernoulli il cui parametro rappresenta la probabilità di successo. Inoltre, se si considerano n prove ripetute di Bernoulli, si può definire una variabile aleatoria indicante il numero di successi registrati nelle n prove; questa risulta avere distribuzione binomiale di parametri n e p , dove p è la probabilità di successo in ciascuna prova. Allo scopo di pervenire ad una distribuzione multivariata analoga alla distribuzione binomiale è necessario estendere il concetto di prova di Bernoulli. A tal fine riferiamoci ad un esperimento costituito da n prove indipendenti ed assumiamo che in ciascuna prova i risultati di interesse siano $k > 2$ eventi necessari ed incompatibili, che denoteremo con B_1, B_2, \dots, B_k , caratterizzati rispettivamente da probabilità $p_i = P(B_i) > 0$ ($i = 1, 2, \dots, k$). Si noti che le ipotesi fatte comportano $\sum_{i=1}^k p_i = 1$. Un esperimento del tipo considerato si dice essere costituito da n prove generalizzate di Bernoulli. I possibili risultati sono le k^n n -uple distinte ognuna delle quali è costituita da n degli eventi B_i ($i = 1, 2, \dots, k$). A ciascun risultato dell'esperimento è così associata una probabilità che in generale dipende da n , da p_i ($i = 1, 2, \dots, k$) e dal numero di volte in cui ciascun evento B_i si è verificato. Infatti, la probabilità del generico risultato

$$\left(\underbrace{B_1, \dots, B_1}_{n_1 \text{ volte}}, \underbrace{B_2, \dots, B_2}_{n_2 \text{ volte}}, \dots, \underbrace{B_k, \dots, B_k}_{n_k \text{ volte}} \right),$$

esprime che B_i si è verificato n_i volte con $i = 1, 2, \dots, k$ e con $\sum_{i=1}^k n_i = n$, per l'ipotesi di indipendenza delle prove è $p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}$. Questa è anche la probabilità di ciascun risultato in cui B_i si è verificato n_i volte ($i = 1, 2, \dots, k$) indipendentemente da quanto si è verificato nel corso delle n prove. Per calcolare il numero di tali risultati consideriamo una sequenza di lunghezza n e su di essa scegliamo n_1 posti in cui sistemare i simboli B_1 ; ciò è evidentemente possibile in $\binom{n}{n_1}$ modi distinti. Successivamente, nelle $n - n_1$ posizioni ancora libere sistemiamo n_2 dei simboli B_2 , il che può essere effettuato in $\binom{n-n_1}{n_2}$ modi distinti. Così procedendo, si comprende che nel momento in cui devono essere sistemati i simboli B_i ($1 < i \leq k$) sono ancora disponibili $n - n_1 - \dots - n_{i-1}$ posti nei quali gli n_i simboli B_i potranno essere collocati in $\binom{n-n_1-\dots-n_{i-1}}{n_i}$ modi distinti. In conclusione, il numero di sequenze in cui B_i si è verificato n_i volte ($i = 1, 2, \dots, k$) indipendentemente dall'ordine è

$$\binom{n}{n_1} \binom{n-n_1}{n_2} \binom{n-n_1-n_2}{n_3} \dots \binom{n-n_1-n_2-\dots-n_{k-2}}{n_{k-1}} \binom{n_k}{n_k} = \frac{n!}{n_1! n_2! \dots n_k!}.$$

Pertanto

$$P(E) = \frac{n!}{n_1! n_2! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}$$

è la probabilità dell'evento

$$E = \{B_i \text{ si verifica } n_i \text{ volte } (i = 1, 2, \dots, k) \text{ nelle } n \text{ prove}\}.$$

Sia $\mathbf{X} = (X_1, X_2, \dots, X_k)$ il vettore aleatorio la cui generica componente X_i ($i = 1, 2, \dots, k$) rappresenta il numero di volte in cui si è verificato l'evento B_i nelle n prove. Evidentemente, per ogni scelta di interi non negativi n_1, n_2, \dots, n_k tali che $\sum_{i=1}^k n_i = n$, risulta $P(X_1 = n_1, X_2 = n_2, \dots, X_k = n_k) = P(E)$.

Definizione 4.8 Un vettore aleatorio $\mathbf{X} = (X_1, X_2, \dots, X_k)$ di funzione di probabilità congiunta

$$p_{\mathbf{X}}(\mathbf{x}) = \begin{cases} \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}, & x_i = 0, 1, \dots, n \quad (i = 1, 2, \dots, k), \\ & x_1 + x_2 + \dots + x_k = n \\ 0, & \text{altrimenti,} \end{cases} \quad (4.25)$$

con $0 < p_i < 1$ ($i = 1, 2, \dots, k$) tali che $p_1 + p_2 + \dots + p_k = 1$, è detto avere distribuzione multinomiale di parametri n, p_1, p_2, \dots, p_k .

Si noti che $p_{\mathbf{X}}(\mathbf{x}) \geq 0$ per ogni $\mathbf{x} \in \mathbb{R}^k$ e che inoltre risulta:

$$\begin{aligned} \sum_{\substack{n_1=1, \dots, n_k=1 \\ n_1 + \dots + n_k = n}}^n p_{\mathbf{X}}(n_1, \dots, n_k) &= \sum_{\substack{n_1=1, \dots, n_k=1 \\ n_1 + \dots + n_k = n}}^n \frac{n!}{n_1! n_2! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k} \\ &= (p_1 + p_2 + \dots + p_k)^n = 1, \end{aligned}$$

dove la somma è estesa a tutti i valori di n_1, n_2, \dots, n_k soddisfacenti le condizioni $n_1 + n_2 + \dots + n_k = n$, $n_i = 0, 1, \dots, n$ ($i = 1, 2, \dots, k$). La locuzione “multinomiale” deriva appunto dal ruolo qui giocato dallo sviluppo multinomiale di $(p_1 + p_2 + \dots + p_k)^n$.

Si noti che \mathbf{X} è in effetti un vettore aleatorio a $k - 1$ dimensioni in quanto una qualsiasi delle componenti è univocamente determinata dalle rimanenti $k - 1$ in virtù del vincolo $X_1 + X_2 + \dots + X_k = n$. Nel caso particolare $k = 2$, il secondo membro della (4.25) si identifica con quello corrispondente della (4.5), pur rimanendo ben distinti i significati di $p_{\mathbf{X}}(\mathbf{x})$ e di $p_X(x)$, riferendosi la prima al vettore \mathbf{X} e la seconda alla variabile X .

Nel seguito con la notazione $\mathbf{X} \sim \mathcal{M}(n, p_1, p_2, \dots, p_k)$ intenderemo che \mathbf{X} è un vettore aleatorio di distribuzione multinomiale di parametri n, p_1, p_2, \dots, p_k ; \mathbf{X} sarà anche detto *vettore multinomiale*.

Esempio 4.14 Si consideri l'esperimento consistente nel lanciare un dado quattro volte. Il risultato è una quadrupla costituita dalle quattro uscite dei singoli lanci. Calcoliamo la probabilità che nei quattro lanci il numero 1 compaia due volte ed il numero 2 una volta.

Definiamo gli eventi $B_1 = \{\text{in un singolo lancio si verifica } 1\}$, $B_2 = \{\text{in un singolo lancio si verifica } 2\}$, $B_3 = \{\text{il risultato di un singolo lancio è uno qualsiasi dei numeri } 3, 4, 5, 6\}$, e sia \mathbf{X} il vettore aleatorio che descrive il numero di volte in cui B_1 , B_2 e B_3 si sono verificati nelle 4 prove. Evidentemente, $\mathbf{X} \sim \mathcal{M}(4, p_1, p_2, p_3)$ con $p_1 = 1/6$, $p_2 = 1/6$ e $p_3 = 2/3$. Pertanto, la probabilità desiderata è:

$$p(2, 1, 1, 0) = \frac{4!}{2! 1! 1! 0!} \left(\frac{1}{6}\right)^2 \left(\frac{2}{3}\right)^1 \left(\frac{1}{6}\right)^1 = \frac{1}{27}.$$

◇

Esempio 4.15 Si supponga che un'urna contiene biglie contrassegnate con i numeri 1, 2, 3. Il numero n_1 di biglie recanti il numero 1 è il doppio del numero n_2 di biglie con il numero 2, che a sua volta è uguale al numero n_3 di biglie contrassegnate con il numero 3. Si consideri l'esperimento consistente nell'estrarre, con reinserimento, tre biglie dall'urna. Per $k = 1, 2, 3$, calcoliamo la probabilità dell'evento $E_k = \{\text{le tre biglie estratte sono contrassegnate da } k \text{ numeri diversi}\}$.

Sia $B_{i,j} = \{\text{alla } j\text{-esima estrazione si è avuta una biglia contrassegnata con il numero } i\}$ ($i, j = 1, 2, 3$). In virtù dell'ipotesi fatta sulla composizione dell'urna, si ha $n_1 = 2n_2$, $n_2 = n_3$, con $n = n_1 + n_2 + n_3$, dove con n si è indicato il numero complessivo di biglie presenti nell'urna. Quindi, $n_1 = n/2$ e $n_2 = n_3 = n/4$, così che $P(B_{1,j}) = 1/2$ e $P(B_{2,j}) = P(B_{3,j}) = 1/4$ ($j = 1, 2, 3$). Sia $\mathbf{X} = (X_1, X_2, X_3)$ il vettore aleatorio che descrive il numero di biglie di ciascun tipo presenti nel campione estratto; è evidente che risulta $\mathbf{X} \sim \mathcal{M}(3, \frac{1}{2}, \frac{1}{4}, \frac{1}{4})$. La probabilità che le tre biglie estratte siano contrassegnate da differenti numeri può essere immediatamente calcolata applicando la (4.25):

$$P(E_3) = p_{\mathbf{X}}(1, 1, 1) = \frac{3!}{1! 1! 1!} \frac{1}{2} \left(\frac{1}{4}\right)^2 = \frac{3}{16}.$$

Per calcolare la probabilità che le tre biglie estratte siano dello stesso tipo, si osservi che E_1 si può esprimere come unione di 3 eventi incompatibili:

$$E_1 = \bigcup_{i=1}^3 (B_{i,1} \cap B_{i,2} \cap B_{i,3}).$$

Si ha quindi:

$$\begin{aligned} P(E_1) &= \sum_{i=1}^3 P(B_{i,1} \cap B_{i,2} \cap B_{i,3}) = p_{\mathbf{X}}(3, 0, 0) + p_{\mathbf{X}}(0, 3, 0) + p_{\mathbf{X}}(0, 0, 3) \\ &= \frac{3!}{3! 0! 0!} \left[\left(\frac{1}{2}\right)^3 + 2 \left(\frac{1}{4}\right)^3 \right] = \frac{5}{32}. \end{aligned}$$

Calcoliamo, infine, la probabilità che le tre biglie siano contrassegnate da due numeri diversi. A tal fine notiamo che l'evento E_2 può realizzarsi in corrispondenza delle seguenti situazioni incompatibili:

- (a) il campione estratto contiene due biglie contrassegnate dal numero 2 e una biglia contrassegnata dal numero 3, oppure il campione contiene una biglia contrassegnata dal numero 2 e due biglie contrassegnate dal numero 3;
- (b) nel campione estratto sono presenti una biglia contrassegnata dal numero 1 e due biglie che possono essere o entrambe contrassegnate dal numero 2 o entrambe contrassegnate dal numero 3;
- (c) il campione contiene due biglie contrassegnate dal numero 1, mentre la terza biglia è contrassegnata dal numero 2 o dal numero 3. Pertanto:

$$\begin{aligned} P(E_2) &= [p_{\mathbf{X}}(0, 2, 1) + p_{\mathbf{X}}(0, 1, 2)] + [p_{\mathbf{X}}(1, 2, 0) + p_{\mathbf{X}}(1, 0, 2)] \\ &\quad + [p_{\mathbf{X}}(2, 1, 0) + p_{\mathbf{X}}(2, 0, 1)] \\ &= 2 \left[\frac{3!}{0! 2! 1!} \left(\frac{1}{4}\right)^3 + \frac{3!}{1! 2! 0!} \frac{1}{2} \left(\frac{1}{4}\right)^2 + \frac{3!}{2! 1! 0!} \left(\frac{1}{2}\right) \frac{1}{4} \right] = \frac{21}{32}. \end{aligned}$$

Si verifica infine che risulta $P(E_1) + P(E_2) + P(E_3) = 1$. \diamond

Proposizione 4.8 Se $\mathbf{X} = (X_1, X_2, \dots, X_k) \sim \mathcal{M}(n, p_1, p_2, \dots, p_k)$, allora $X_i \sim \mathcal{B}(n, p_i)$ ($i = 1, 2, \dots, k$), ossia le distribuzioni marginali di ciascuna componente X_i sono binomiali di parametri n e p_i .

Dimostrazione Senza perdita di generalità possiamo riferirci ad X_1 . Si ha pertanto:

$$\begin{aligned} p_{X_1}(n_1) &= \sum_{\substack{n_2=1, \dots, n_k=1 \\ n_2+\dots+n_k=n-n_1}}^n p_{\mathbf{X}}(n_1, n_2, \dots, n_k) = \sum_{\substack{n_2=1, \dots, n_k=1 \\ n_2+\dots+n_k=n-n_1}}^n \frac{n!}{n_1! n_2! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k} \\ &= \frac{n! p_1^{n_1}}{n_1! (n - n_1)!} \sum_{\substack{n_2=1, \dots, n_k=1 \\ n_2+\dots+n_k=n-n_1}}^n \frac{(n - n_1)!}{n_2! \dots n_k!} p_2^{n_2} \dots p_k^{n_k} \\ &= \binom{n}{n_1} p_1^{n_1} (p_2 + \dots + p_k)^{n-n_1}, \end{aligned}$$

da cui, ricordando che $\sum_{i=1}^k p_i = 1$, si trae:

$$p_{X_1}(n_1) = \binom{n}{n_1} p_1^{n_1} (1 - p_1)^{n-n_1} \quad (n_1 = 0, 1, \dots, n).$$

□

Proposizione 4.9 Se $\mathbf{X} = (X_1, X_2, \dots, X_k) \sim \mathcal{M}(n, p_1, p_2, \dots, p_k)$, allora risulta $(X_i, X_j) \sim \mathcal{M}(n, p_i, p_j, 1 - p_i - p_j)$ ($i, j = 1, 2, \dots, k; i \neq j$), ossia le distribuzioni marginali delle coppie (X_i, X_j) sono multinomiali di parametri $n, p_i, p_j, 1 - p_i - p_j$.

Dimostrazione Senza perdita di generalità, riferiamoci alla coppia (X_1, X_2) . Si ha:

$$\begin{aligned} p_{X_1, X_2}(n_1, n_2) &= \sum_{\substack{n_3=1, \dots, n_k=1 \\ n_3+\dots+n_k=n-n_1-n_2}}^n \frac{n!}{n_1! n_2! n_3! \dots n_k!} p_1^{n_1} p_2^{n_2} p_3^{n_3} \dots p_k^{n_k} \\ &= \frac{n! p_1^{n_1} p_2^{n_2}}{n_1! n_2! (n - n_1 - n_2)!} \sum_{\substack{n_3=1, \dots, n_k=1 \\ n_3+\dots+n_k=n-n_1-n_2}}^n \frac{(n - n_1 - n_2)!}{n_3! \dots n_k!} p_3^{n_3} \dots p_k^{n_k} \\ &= \frac{n! p_1^{n_1} p_2^{n_2}}{n_1! n_2! (n - n_1 - n_2)!} (p_3 + \dots + p_k)^{n-n_1-n_2} \\ &= \frac{n!}{n_1! n_2! (n - n_1 - n_2)!} p_1^{n_1} p_2^{n_2} (1 - p_1 - p_2)^{n-n_1-n_2}. \end{aligned}$$

□

4.3.2 Distribuzione ipergeometrica multivariata

Così come la distribuzione ipergeometrica, la distribuzione ipergeometrica multivariata interviene nella descrizione di estrazioni senza reinserimento oppure di estrazioni in blocco.

Si consideri l'esperimento che consiste nell'estrarre a caso n biglie senza reinserimento da un'urna che ne contiene N , di cui N_1 di colore C_1 , N_2 di colore C_2 , ..., N_k di colore C_k , essendo $0 \leq N_i \leq N$ ($i = 1, 2, \dots, k$) e $N_1 + N_2 + \dots + N_k = N$. Sia poi l'evento

$$E_{n_1, n_2, \dots, n_k} = \{ \text{nel campione estratto di } n \text{ biglie } n_1 \text{ sono di colore } C_1, \\ n_2 \text{ di colore } C_2, \dots, n_k \text{ di colore } C_k \},$$

con $0 \leq n_i \leq N_i$ ($i = 1, 2, \dots, k$) e $n_1 + n_2 + \dots + n_k = n$. Facendo ricorso alla definizione classica di probabilità, si ha:

$$P(E_{n_1, n_2, \dots, n_k}) = \frac{\binom{N_1}{n_1} \binom{N_2}{n_2} \dots \binom{N_k}{n_k}}{\binom{N}{n}},$$

dove il numeratore fornisce il numero di modi in cui si possono estrarre n_1 biglie di colore C_1 dalle N_1 disponibili, n_2 biglie di colore C_2 dalle N_2 disponibili e così via, fino a n_k biglie di colore C_k dalle N_k presenti nell'urna, mentre il denominatore fornisce il numero di modi in cui si possono estrarre n biglie tra le N biglie contenute nell'urna. Affinché sia $P(E_{n_1, n_2, \dots, n_k}) > 0$ occorre che sia $0 \leq n_i \leq N_i$ ($i = 1, 2, \dots, k$) con $n_1 + n_2 + \dots + n_k = n$ e che inoltre risulti $0 \leq N_i \leq N$ ($i = 1, 2, \dots, k$) con $N_1 + N_2 + \dots + N_k = N$.

Se $\mathbf{X} = (X_1, X_2, \dots, X_k)$ è il vettore aleatorio che descrive il numero di biglie di colori C_1, C_2, \dots, C_k estratte senza reinserimento, risulta $P(X_1 = n_1, X_2 = n_2, \dots, X_k = n_k) = P(E_{n_1, n_2, \dots, n_k})$.

Definizione 4.9 Un vettore aleatorio $\mathbf{X} = (X_1, X_2, \dots, X_k)$ di funzione di probabilità

$$p_{\mathbf{X}}(\mathbf{x}) = \begin{cases} \frac{\binom{N_1}{x_1} \binom{N_2}{x_2} \dots \binom{N_k}{x_k}}{\binom{N}{n}}, & 0 \leq x_i \leq N_i \quad (i = 1, 2, \dots, k), \\ & x_1 + x_2 + \dots + x_k = n \\ 0, & \text{altrimenti,} \end{cases} \quad (4.26)$$

con n, N_1, N_2, \dots, N_k interi positivi e $N = N_1 + N_2 + \dots + N_k$, si dice avere distribuzione ipergeometrica multivariata di parametri n, N_1, N_2, \dots, N_k .

Si noti che è $p_{\mathbf{X}}(\mathbf{x}) \geq 0$ per ogni $\mathbf{x} \in \mathbb{R}^k$ e che, facendo uso della seguente formula:

$$\sum_{\substack{n_1, \dots, n_k \\ n_1 + \dots + n_k = n}} \binom{N_1}{n_1} \binom{N_2}{n_2} \dots \binom{N_k}{n_k} = \binom{N}{n}, \quad (4.27)$$

si ha:

$$\sum_{\substack{n_1, \dots, n_k \\ n_1 + \dots + n_k = n}} p_{\mathbf{X}}(n_1, n_2, \dots, n_k) = \sum_{\substack{n_1, \dots, n_k \\ n_1 + \dots + n_k = n}} \frac{\binom{N_1}{n_1} \binom{N_2}{n_2} \dots \binom{N_k}{n_k}}{\binom{N}{n}} = 1,$$

con le somme estese a tutti gli interi n_1, n_2, \dots, n_k soddisfacenti le condizioni $0 \leq n_i \leq N_i$ ($i = 1, 2, \dots, k$) con $n_1 + n_2 + \dots + n_k = n$.

Come nel caso multinomiale, a causa del vincolo $X_1 + X_2 + \dots + X_k = n$ il vettore aleatorio \mathbf{X} è in effetti a $k - 1$ dimensioni. Per $k = 2$ il secondo membro della (4.26) si riduce ad una distribuzione di probabilità ipergeometrica unidimensionale.

Nel seguito con la notazione $\mathbf{X} \sim \mathcal{I}(n, N_1, N_2, \dots, N_k)$ intenderemo che \mathbf{X} è un vettore aleatorio di distribuzione ipergeometrica di parametri n, N_1, N_2, \dots, N_k , detto anche *vettore ipergeometrico*.

Esempio 4.16 In un gruppo costituito da 4 maschi, di cui metà biondi e metà bruni, e da 4 femmine, di cui metà bionde e metà brune, si sceglie a caso un gruppo di 4 individui. Calcoliamo la probabilità dell'evento $E = \{\text{nel gruppo scelto sono presenti entrambi i sessi ed in ugual numero individui biondi e bruni}\}$.

Sia $\mathbf{X} = (X_1, X_2, X_3, X_4)$ il vettore aleatorio che descrive il numero X_1 di maschi bruni, il numero X_2 di maschi biondi, il numero X_3 di femmine brune ed il numero X_4 di femmine bionde presenti nel gruppo scelto. Evidentemente $\mathbf{X} \sim \mathcal{I}(4, 2, 2, 2, 2)$. Pertanto

$$\begin{aligned} P(E) &= [p_{\mathbf{X}}(0, 1, 2, 1) + p_{\mathbf{X}}(0, 2, 2, 0)] \\ &\quad + [p_{\mathbf{X}}(1, 0, 1, 2) + p_{\mathbf{X}}(1, 1, 1, 1) + p_{\mathbf{X}}(1, 2, 1, 0)] \\ &\quad + [p_{\mathbf{X}}(2, 0, 0, 2) + p_{\mathbf{X}}(2, 1, 0, 1)] \\ &= \left[\binom{8}{4} \right]^{-1} \left\{ \left[\binom{2}{0} \binom{2}{1} \binom{2}{2} \binom{2}{1} + \binom{2}{0} \binom{2}{2} \binom{2}{2} \binom{2}{0} \right] \right. \\ &\quad + \left[\binom{2}{1} \binom{2}{0} \binom{2}{1} \binom{2}{2} + \binom{2}{1} \binom{2}{1} \binom{2}{1} \binom{2}{1} + \binom{2}{1} \binom{2}{2} \binom{2}{1} \binom{2}{0} \right] \\ &\quad \left. + \left[\binom{2}{2} \binom{2}{0} \binom{2}{0} \binom{2}{2} + \binom{2}{2} \binom{2}{1} \binom{2}{0} \binom{2}{1} \right] \right\} = \frac{17}{35}. \end{aligned}$$

◇

Proposizione 4.10 Se $\mathbf{X} = (X_1, X_2, \dots, X_k) \sim \mathcal{I}(n, N_1, N_2, \dots, N_k)$, allora $X_i \sim \mathcal{I}(n, N_i, N - N_i)$ ($i = 1, 2, \dots, k$), ossia le distribuzioni marginali di ciascuna componente X_i sono ipergeometriche unidimensionali di parametri $n, N_i, N - N_i$.

Dimostrazione Con riferimento ad X_1 , si ha ad esempio:

$$\begin{aligned} p_{X_1}(n_1) &= \sum_{\substack{n_2, \dots, n_k \\ n_2 + \dots + n_k = n - n_1}} p_{\mathbf{X}}(n_1, \dots, n_k) = \sum_{\substack{n_2, \dots, n_k \\ n_2 + \dots + n_k = n - n_1}} \frac{\binom{N_1}{n_1} \binom{N_2}{n_2} \dots \binom{N_k}{n_k}}{\binom{N}{n}} \\ &= \frac{\binom{N_1}{n_1}}{\binom{N}{n}} \sum_{\substack{n_2, \dots, n_k \\ n_2 + \dots + n_k = n - n_1}} \binom{N_2}{n_2} \dots \binom{N_k}{n_k} = \frac{\binom{N_1}{n_1} \binom{N - N_1}{n - n_1}}{\binom{N}{n}}, \end{aligned}$$

dove l'ultima uguaglianza segue dall'utilizzazione della formula (4.27).

□

Proposizione 4.11 Se $\mathbf{X} = (X_1, X_2, \dots, X_k) \sim \mathcal{I}(n, N_1, N_2, \dots, N_k)$, allora si ha $(X_i, X_j) \sim \mathcal{I}(n, N_i, N_j, N - N_i - N_j)$ ($i, j = 1, 2, \dots, k; i \neq j$), ossia le distribuzioni marginali delle coppie (X_i, X_j) sono ipergeometriche multivariate di parametri $n, N_i, N_j, N - N_i - N_j$.

Dimostrazione Ad esempio, con riferimento alla coppia (X_1, X_2) , si ha:

$$\begin{aligned} p_{X_1, X_2}(n_1, n_2) &= \sum_{\substack{n_3, \dots, n_k \\ n_3 + \dots + n_k = n - n_1 - n_2}} p_{\mathbf{X}}(n_1, \dots, n_k) \\ &= \frac{\binom{N_1}{n_1} \binom{N_2}{n_2}}{\binom{N}{n}} \sum_{\substack{n_3, \dots, n_k \\ n_3 + \dots + n_k = n - n_1 - n_2}} \binom{N_3}{n_3} \dots \binom{N_k}{n_k} = \frac{\binom{N_1}{n_1} \binom{N_2}{n_2} \binom{N - N_1 - N_2}{n - n_1 - n_2}}{\binom{N}{n}}, \end{aligned}$$

dove di nuovo si è fatto uso della formula (4.27). \square

4.4 Variabili aleatorie assolutamente continue

In questo paragrafo passeremo in rassegna alcune distribuzioni di probabilità per variabili aleatorie assolutamente continue di particolare rilevanza.

4.4.1 Distribuzione uniforme

Quando lo spazio campione è costituito da un numero finito di elementi è talora ragionevole assumere che gli eventi costituiti dai singleton dei suoi elementi siano equiprobabili. Ciò conduce alla distribuzione di probabilità uniforme discreta discussa nel Paragrafo 4.2.1. Se, invece, lo spazio campione contiene un'infinità numerabile di punti non è possibile introdurre un simile concetto di equiprobabilità che implicherebbe che ad ogni singleton sia associata una probabilità nulla.

Nel caso in cui lo spazio campione è costituito da un'infinità non numerabile di punti, nuovamente l'equiprobabilità implicherebbe che ogni elemento è caratterizzato da probabilità nulla e, per l'assioma di additività completa, si deduce che la probabilità di scegliere un punto appartenente ad un insieme al più numerabile è ancora nulla. Invece, la probabilità di scegliere un punto appartenente ad un insieme non numerabile di punti, ad esempio un intervallo dell'asse reale, potrebbe non essere nulla poiché in tal caso non è più applicabile l'assioma di additività della probabilità. Un modo per tradurre l'esigenza intuitiva di assegnare uguale probabilità a tutti i punti di un intervallo dell'asse reale consiste nel richiedere che ad intervalli di ampiezza uguale siano associate uguali probabilità.

Sia (a, b) un intervallo dell'asse reale e sia X una variabile aleatoria assolutamente continua che assume valori in (a, b) con densità di probabilità $f_X(x)$. Inoltre, sia I un sottointervallo di (a, b) di misura ℓ . Per tradurre nel continuo il concetto di equiprobabilità che è alla base della distribuzione uniforme discreta, supponiamo che la probabilità che X appartenga ad I dipenda da I solo attraverso la sua misura ℓ al seguente modo:

$$P(X \in I) = \int_I f_X(x) dx = \frac{\ell}{b - a}.$$

Definizione 4.10 Siano a e b numeri reali tali che $a < b$. Una variabile aleatoria X di funzione di distribuzione

$$F_X(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b \end{cases} \quad (4.28)$$

e corrispondente densità di probabilità

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{altrimenti} \end{cases} \quad (4.29)$$

si dice uniformemente distribuita o equidistribuita nell'intervallo (a, b) .

Si noti che la (4.29) coincide con $dF_X(x)/dx$ laddove $F_X(x)$ è derivabile ed è posta arbitrariamente uguale a zero per $x = a$ e $x = b$.

La densità di probabilità uniforme traduce nel continuo il concetto di equiprobabilità che è alla base della distribuzione uniforme discreta. Infatti, anche se la probabilità $P(X = x)$ è nulla per ciascun punto x , la densità di probabilità uniforme assegna valori uguali a tutti gli intervalli di uguale ampiezza che vengono scelti in (a, b) .

Nel seguito, con la notazione $X \sim \mathcal{U}(a, b)$ intenderemo che X è una variabile aleatoria avente distribuzione uniforme nell'intervallo (a, b) ; X sarà anche detta *variabile uniforme*. La funzione di distribuzione e la densità di probabilità di una variabile $X \sim \mathcal{U}(-1, 1)$ sono rappresentate in Figura 4.6.

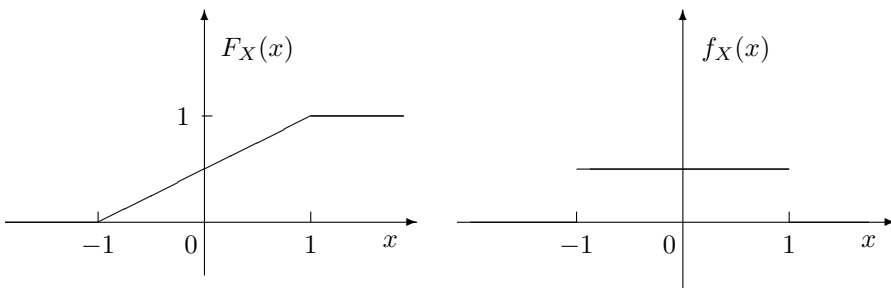


Figura 4.6 – Funzione di distribuzione e densità di probabilità di $X \sim \mathcal{U}(-1, 1)$.

Esempio 4.17 Con riferimento alla Figura 4.7, sul segmento AB di lunghezza a e punto medio C si scelga un punto D in accordo con la distribuzione uniforme. Si desidera calcolare la probabilità dell'evento $E = \{i \text{ tre segmenti AD, DB, AC possono riguardarsi come lati di un triangolo}\}$.

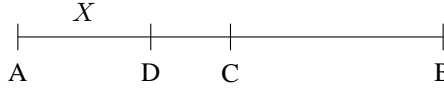


Figura 4.7 – Ad illustrazione dell'Esempio 4.17.

In base a quanto specificato, $X \sim \mathcal{U}(0, a)$ è la variabile aleatoria che indica la distanza del punto D da A. Affinché i tre segmenti in questione possano riguardarsi come lati di un triangolo è necessario e sufficiente che la lunghezza di ogni segmento sia minore della somma delle lunghezze degli altri due. Pertanto, essendo $\overline{AD} = X$, $\overline{DB} = a - X$ e $\overline{AC} = a/2$, si può concludere che l'evento E si realizza se e solo se risulta

$$X < (a - X) + \frac{a}{2}, \quad (a - X) < X + \frac{a}{2}, \quad \frac{a}{2} < X + (a - X),$$

ossia se e solo se $a/4 < X < 3a/4$. Quindi si ha:

$$P(E) = P\left(\frac{a}{4} < X < \frac{3a}{4}\right) = \frac{1}{a} \left(\frac{3a}{4} - \frac{a}{4}\right) = \frac{1}{2}.$$

◇

Proposizione 4.12 Sia $U \sim \mathcal{U}(0, 1)$ e sia X la variabile aleatoria assolutamente continua definita tramite la trasformazione

$$X = G^{-1}(U), \quad (4.30)$$

dove $G(x)$ denota un'arbitraria funzione di distribuzione invertibile. Allora

$$F_X(x) = \begin{cases} 0, & x < G^{-1}(0) \\ G(x), & G^{-1}(0) \leq x < G^{-1}(1) \\ 1, & x \geq G^{-1}(1), \end{cases} \quad (4.31)$$

è la funzione di distribuzione di X .

Dimostrazione Essendo $G(x)$ per ipotesi una funzione di distribuzione, essa è non decrescente; quindi, dalla (4.30) si ricava:

$$F_X(x) = P(X \leq x) = P[G^{-1}(U) \leq x] = P[U \leq G(x)] = \begin{cases} 0, & G(x) < 0 \\ G(x), & 0 \leq G(x) < 1 \\ 1, & G(x) \geq 1, \end{cases} \quad (4.32)$$

da cui segue la (4.31). □

La Proposizione 4.12 mostra che è possibile ottenere una variabile aleatoria X assolutamente continua caratterizzata da funzione di distribuzione $G(x)$ invertibile a partire da una variabile

aleatoria U uniformemente distribuita in $(0, 1)$ tramite la trasformazione $X = G^{-1}(U)$. Ciò si rivela particolarmente utile per la costruzione di algoritmi di simulazione probabilistica che richiedono la generazione di numeri che obbediscano approssimativamente a preassegnate leggi di probabilità.

Esempio 4.18 Si desidera determinare una trasformazione tra variabili aleatorie che permetta di ottenere $X \sim \mathcal{U}(a, b)$ a partire da $U \sim \mathcal{U}(0, 1)$.

Confrontando le (4.28) e (4.32), dalla Proposizione 4.12 segue

$$U = G(X) = \frac{X - a}{b - a},$$

da cui si ricava immediatamente:

$$X = a + (b - a) U.$$

Tale formula viene utilizzata ad esempio per simulare una variabile uniforme in (a, b) a partire da una variabile uniforme in $(0, 1)$. \diamond

4.4.2 Distribuzione esponenziale

Nel caso discreto abbiamo mostrato che una variabile aleatoria caratterizzata da funzione di probabilità geometrica può utilizzarsi per descrivere il tempo di attesa per l'occorrenza del primo successo in una successione di prove ripetute di Bernoulli. Introduciamo ora la densità di probabilità esponenziale; questa può interpretarsi come l'analogo nel continuo della funzione di probabilità geometrica nel senso che una variabile aleatoria caratterizzata da densità di probabilità esponenziale può immaginarsi idonea a descrivere anch'essa un tempo di attesa che, però, in questo caso viene riguardato come variabile nel continuo.

Per introdurre la densità di probabilità esponenziale consideriamo il seguente esempio.

Esempio 4.19 Si supponga che il numero di telefonate che pervengono ad un centralino telefonico nell'intervallo di tempo $(0, t)$, con t reale positivo, sia una variabile aleatoria N_t di distribuzione di Poisson di parametro λt . Sia poi T la variabile aleatoria rappresentante l'istante in cui perviene la prima telefonata. Se è $t < 0$, deve evidentemente assumersi $P(T \leq t) = 0$; per $t \geq 0$ risulta:

$$P(T \leq t) = 1 - P(N_t = 0) = 1 - e^{-\lambda t}.$$

La probabilità $P(T \leq t)$, con t reale, può interpretarsi come funzione di distribuzione di una variabile aleatoria assolutamente continua. \diamond

Definizione 4.11 Sia $\lambda > 0$. Una variabile aleatoria X di funzione di distribuzione

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0 \end{cases} \quad (4.33)$$

e corrispondente densità di probabilità

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{altrimenti} \end{cases} \quad (4.34)$$

si dice *esponenzialmente distribuita con parametro λ* .

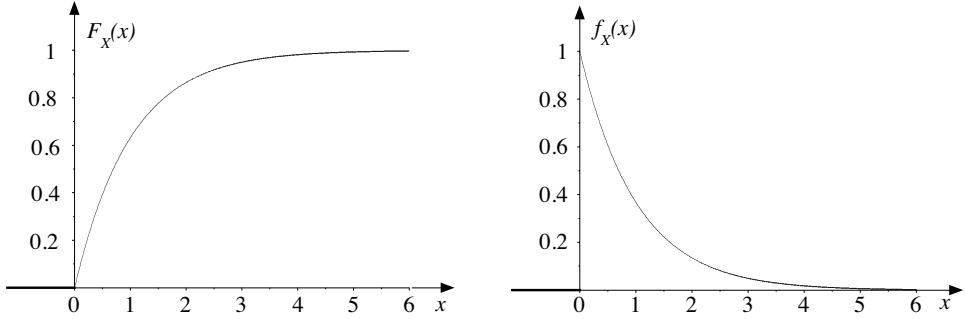


Figura 4.8 – Funzione di distribuzione e densità di probabilità di $X \sim \mathcal{E}(1, 1)$.

Nel seguito la notazione $X \sim \mathcal{E}(1, \lambda)$ verrà utilizzata per indicare che X ha distribuzione esponenziale di parametro λ ; X sarà anche detta *variabile esponenziale*. La funzione di distribuzione e la densità di probabilità di una variabile $X \sim \mathcal{E}(1, 1)$ sono rappresentate in Figura 4.8.

Esempio 4.20 Determiniamo una trasformazione che permetta di ottenere una variabile $X \sim \mathcal{E}(1, \lambda)$ a partire dalla variabile $U \sim \mathcal{U}(0, 1)$.

Confrontando le (4.33) e (4.32), per la Proposizione 4.12 segue:

$$U = G(X) = 1 - e^{-\lambda X},$$

da cui si ricava immediatamente:

$$X = -\frac{1}{\lambda} \ln(1 - U).$$

Essendo U e $1 - U$ entrambe uniformemente distribuite in $(0, 1)$, segue anche che $X = -\ln U/\lambda$ ha distribuzione esponenziale di parametro $\lambda > 0$. È questo il punto di partenza solitamente utilizzato per la simulazione di una variabile aleatoria esponenziale. \diamond

Proposizione 4.13 Sia X una variabile aleatoria di densità (4.34). Per ogni s, t reali positivi risulta:

$$P(X > s + t | X > s) = P(X > t). \quad (4.35)$$

Dimostrazione Dalla definizione di probabilità condizionata (2.19) e dalla (4.33) per $s \geq 0$ e $t \geq 0$ segue:

$$\begin{aligned} P(X > s + t | X > s) &= \frac{P(X > s + t, X > s)}{P(X > s)} = \frac{P(X > s + t)}{P(X > s)} \\ &= \frac{1 - F_X(s + t)}{1 - F_X(s)} = e^{-\lambda t} = P(X > t). \end{aligned}$$

□

Se si interpreta X come un tempo di attesa, la (4.35) mostra che la probabilità condizionata che il tempo di attesa X sia maggiore di $t + s$ dato che essa è maggiore di s non dipende da quanto si è già atteso, ossia da s . Pertanto, la distribuzione esponenziale, così come la distribuzione geometrica, gode della proprietà di “assenza di memoria”.

La funzione di distribuzione esponenziale riveste notevole importanza sia teorica che applicativa. Essa, ad esempio, interviene spesso quando si studiano sistemi di servizio in cui è ragionevole assumere che i tempi di interarrivo dei clienti oppure i tempi di espletamento dei servizi siano distribuiti proprio esponenzialmente, o allorché si considera la durata di funzionamento, realisticamente supposta aleatoria, di componenti elettronici o di dispositivi di varia natura.

Proposizione 4.14 *Sia X una variabile aleatoria di densità (4.34) e sia $Z = X - \tau$, con $\tau > 0$. Per ogni $z > 0$ risulta:*

$$P(Z \leq z | X > \tau) = P(X \leq z). \quad (4.36)$$

Dimostrazione Facendo uso della definizione di probabilità condizionata e della (4.33), per $z > 0$ si ottiene:

$$\begin{aligned} P(Z \leq z | X > \tau) &= P(X - \tau \leq z | X > \tau) = P(X \leq z + \tau | X > \tau) \\ &= \frac{P(X \leq z + \tau, X > \tau)}{P(X > \tau)} = \frac{P(\tau < X \leq z + \tau)}{P(X > \tau)} \\ &= \frac{e^{-\lambda\tau} - e^{-\lambda(z+\tau)}}{e^{-\lambda\tau}} = 1 - e^{-\lambda z} = P(X \leq z). \end{aligned}$$

□

Nella Proposizione 4.14 interpretiamo ora X come durata di funzionamento di un componente elettronico, meccanico o di altra natura (ossia l’intervallo di tempo in cui esso funziona perfettamente) e, conseguentemente, $Z = X - \tau$ come durata della sua vita residua sapendo che esso ha già funzionato per una durata τ . La (4.36) mostra che la durata di vita residua ha la stessa distribuzione della durata di vita del componente considerato. Questa proprietà è un’ovvia conseguenza dell’assenza di memoria della funzione di distribuzione esponenziale. Ciò costituisce evidente circostanza che variabili aleatorie esponenzialmente distribuite non sono idonee a descrivere la durata di vita di dispositivi soggetti ad usura se non entro prefissati limiti di approssimazione, ad esempio nel caso di semplici dispositivi non soggetti a rapidi deterioramenti significativi, ma che possono subire danni per casi accidentali quali, ad esempio, corti circuiti, fulmini, imprevedibili sollecitazioni meccaniche, ecc.

Esempio 4.21 Siano X_1, X_2, \dots, X_n variabili aleatorie indipendenti, con $X_i \sim \mathcal{E}(1, \lambda_i)$. Calcoliamo la funzione di distribuzione di $V = \min\{X_1, X_2, \dots, X_n\}$.

Dalla (3.64) segue:

$$F_V(v) = 1 - \prod_{i=1}^n [1 - F_{X_i}(v)] = \begin{cases} 0, & v < 0 \\ 1 - \prod_{i=1}^n e^{-\lambda_i v} = 1 - e^{-(\lambda_1 + \lambda_2 + \dots + \lambda_n)v}, & v \geq 0. \end{cases}$$

La variabile V è dunque esponenzialmente distribuita con parametro $\lambda = \lambda_1 + \lambda_2 + \dots + \lambda_n$.

È interessante osservare che se le variabili X_1, X_2, \dots, X_n descrivono i tempi di corretto funzionamento di n componenti di un dispositivo collegati in serie, allora V , che rappresenta il tempo di corretto funzionamento dell'intero dispositivo, è distribuita esponenzialmente con parametro dato dalla somma dei parametri delle distribuzioni esponenziali individuali. \diamond

Esempio 4.22 Si consideri (v. Figura 4.9) un sistema di servizio costituito da un'unica fila di attesa e da n addetti al servizio come illustrato in figura. Si supponga che i tempi di servizio

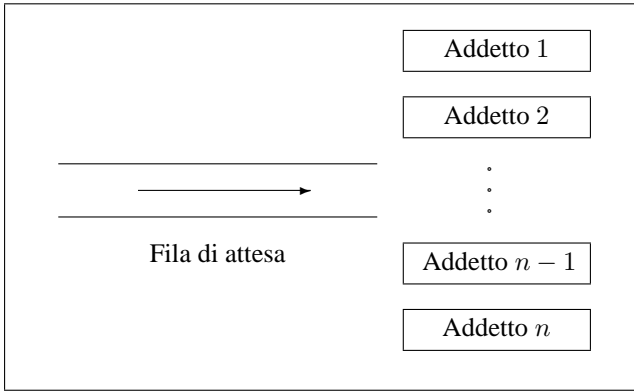


Figura 4.9 – Ad illustrazione dell'Esempio 4.22.

degli addetti $1, 2, \dots, n$ siano rispettivamente descritti dalle variabili aleatorie S_1, S_2, \dots, S_n indipendenti e identicamente distribuite con $S_i \sim \mathcal{E}(1, \mu)$ ($i = 1, 2, \dots, n$). Calcoliamo la funzione di distribuzione $F_T(t)$ della variabile aleatoria T rappresentante il tempo di permanenza in fila di attesa di un utente che al suo arrivo trova la fila di attesa vuota e tutti gli addetti occupati.

Per accedere al servizio l'utente in arrivo deve attendere un tempo pari al minimo dei tempi residui degli utenti già in corso di servizio. In virtù della proprietà di assenza di memoria della distribuzione esponenziale, i tempi residui di servizio degli n utenti già presenti nel sistema sono distribuiti esponenzialmente con parametro μ . Pertanto T ha la stessa distribuzione di $V = \min(S_1, S_2, \dots, S_n)$. Dall'Esempio 4.21 segue:

$$F_T(t) = \begin{cases} 0, & t < 0 \\ 1 - e^{-n\mu t}, & t \geq 0, \end{cases}$$

ossia $T \sim \mathcal{E}(1, n\mu)$. \diamond

Esempio 4.23 Un dispositivo è costituito da due componenti elettronici collegati in serie. Assumiamo che la durata di vita del componente r -esimo sia descritto da una variabile aleatoria X_r di distribuzione esponenziale di parametro λ_r ($r = 1, 2$). Nell'ipotesi di indipendenza di X_1 e X_2 , si vuole determinare la probabilità che sia stato il secondo componente la causa del guasto del dispositivo.

È evidente che l'evento a cui siamo interessati è esprimibile in termini di X_1 e X_2 . Infatti, la probabilità richiesta è, in definitiva, $P(X_1 > X_2)$, il cui calcolo può essere effettuato facendo ricorso alla (3.45) e all'indipendenza di X_1 e X_2 , avendosi pertanto:

$$\begin{aligned} P(X_1 > X_2) &= \int_0^{+\infty} \int_0^{x_1} f_{X_1, X_2}(x_1, x_2) dx_1 dx_2 = \int_0^{+\infty} \int_0^{x_1} f_{X_1}(x_1) f_{X_2}(x_2) dx_1 dx_2 \\ &= \int_0^{+\infty} \lambda_1 e^{-\lambda_1 x_1} dx_1 \int_0^{x_1} \lambda_2 e^{-\lambda_2 x_2} dx_2 = \int_0^{+\infty} \lambda_1 e^{-\lambda_1 x_1} [1 - e^{-\lambda_2 x_1}] dx_1 \\ &= 1 - \frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{\lambda_2}{\lambda_1 + \lambda_2}. \end{aligned}$$

È opportuno menzionare che il risultato ottenuto è suscettibile di generalizzazione. Infatti, se si considera un dispositivo costituito da n componenti elettronici collegati in serie, ognuno con durata di vita esponenziale di parametro λ_i , e si suppone che le durate di vita dei componenti siano tra loro indipendenti, allora la probabilità che la rottura del dispositivo sia dovuta al componente j -esimo è $\lambda_j / \sum_{k=1}^n \lambda_k$. \diamond

In precedenza, sono stati forniti alcuni esempi di dispositivi costituiti da componenti soggetti a guasti i cui tempi di vita sono indipendenti. Nel seguente esempio si vuole analizzare una situazione in cui ciò non si verifica.

Esempio 4.24 Un sistema costituito da due componenti elettronici è soggetto a guasti provocati da tre possibili cause contrassegnate rispettivamente con 1, 2 e 3 (v. Figura 4.10).

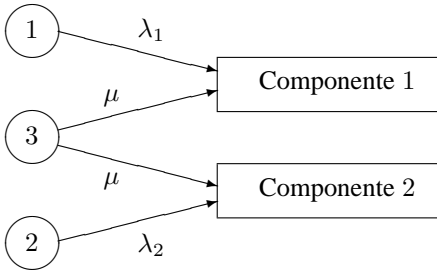


Figura 4.10 – Tre possibili cause di guasti per due componenti.

Si assuma che ognuna delle cause provochi un danno indipendentemente dalle altre, che un guasto provocato dalla causa k ($k = 1, 2$) arrechi un danno irreparabile al componente k e che l'istante di occorrenza del guasto sia descritto da una variabile aleatoria U_k esponenziale di parametro λ_k . Inoltre si assuma che un guasto prodotto dalla causa 3 danneggi entrambi i componenti e che l'istante di occorrenza di tale guasto sia descritto da una variabile aleatoria esponenziale Z di parametro μ . Si assuma infine che U_1 , U_2 e Z siano indipendenti. Denotati con X e Y i tempi di vita dei due componenti, si vuole determinare la funzione di distribuzione congiunta $F_{X,Y}(x, y)$ e le funzioni di distribuzione marginali $F_X(x)$ e $F_Y(y)$.

Osserviamo in primo luogo che, da quanto detto, risulta:

$$X = \min(U_1, Z), \quad Y = \min(U_2, Z),$$

così che, come mostrato nell'Esempio 4.21, si ha:

$$F_X(x) = \begin{cases} 0, & x \leq 0 \\ 1 - e^{-(\lambda_1 + \mu)x}, & x > 0 \end{cases} \quad F_Y(y) = \begin{cases} 0, & y \leq 0 \\ 1 - e^{-(\lambda_2 + \mu)y}, & y > 0. \end{cases}$$

Per determinare la distribuzione congiunta di X e Y osserviamo in primo luogo che per tutti gli x, y reali sussistono le seguenti relazioni:

$$\begin{aligned} F_{X,Y}(x, y) + P(X > x, Y > y) + P(X \leq x, Y > y) + P(X > x, Y \leq y) &= 1, \\ P(X \leq x, Y > y) &= F_X(x) - F_{X,Y}(x, y), \\ P(X > x, Y \leq y) &= F_Y(y) - F_{X,Y}(x, y), \end{aligned}$$

da cui segue:

$$F_{X,Y}(x, y) = P(X > x, Y > y) + F_X(x) + F_Y(y) - 1. \quad (4.37)$$

Ricordando che i tempi di guasto U_1, U_2, Z sono per ipotesi indipendenti ed esponenzialmente distribuiti con rispettivi parametri $\lambda_1, \lambda_2, \mu$, si ricava:

$$\begin{aligned} P(X > x, Y > y) &= P\{\min(U_1, Z) > x, \min(U_2, Z) > y\} \\ &= P\{U_1 > x, U_2 > y, Z > \max(x, y)\} \\ &= \exp\{-\lambda_1 x - \lambda_2 y - \mu \max(x, y)\} \quad x \geq 0, y \geq 0, \end{aligned}$$

così che dalla (4.37) risulta:

$$F_{X,Y}(x, y) = \begin{cases} 1 + e^{-\lambda_1 x - \lambda_2 y - \mu \max(x, y)} - e^{-(\lambda_1 + \mu)x} - e^{-(\lambda_2 + \mu)y}, & x \geq 0, y \geq 0 \\ 0, & \text{altrimenti,} \end{cases}$$

che si identifica con la funzione di distribuzione del vettore aleatorio (X, Y) non assolutamente continuo. Essendo $F_{X,Y}(x, y) \neq F_X(x) F_Y(y)$, si conclude che X e Y , ossia i tempi di vita dei due componenti elettronici, non sono indipendenti. \diamond

4.4.3 Distribuzioni di Erlang e gamma

Si è visto che una variabile aleatoria caratterizzata da funzione di probabilità binomiale negativa è idonea a descrivere il tempo di attesa per l'occorrenza dell' n -esimo successo in una successione di prove ripetute di Bernoulli. Introduciamo ora la densità di probabilità di Erlang, interpretabile come l'analogo di questa nel continuo, nel senso che una variabile aleatoria caratterizzata da densità di Erlang può anch'essa utilizzarsi per descrivere il tempo di attesa per l'occorrenza dell' n -esimo successo, questa volta nel continuo.

Cominciamo col riferirci al seguente esempio.

Esempio 4.25 Si supponga che il numero di telefonate che pervengono ad un centralino telefonico nell'intervallo di tempo $(0, t)$, con t reale positivo, sia una variabile aleatoria N_t di distribuzione di Poisson di parametro λt . Inoltre, sia T_n la variabile aleatoria descrivente l'istante in cui perviene l' n -esima telefonata. Abbastanza realisticamente, riguarderemo gli eventi

$$A_k = \{\text{nell'intervallo } (0, t) \text{ pervengono } k \text{ telefonate}\} \quad (k = 0, 1, \dots)$$

come incompatibili ed aventi distribuzione di Poisson:

$$P(A_k) = P(N_t = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad (k = 0, 1, \dots).$$

Pertanto deve aversi $P(T_n \leq t) = 0$ per $t < 0$. Per $t \geq 0$ si ha invece:

$$P(T_n \leq t) = 1 - P(T_n > t) = 1 - P\left(\bigcup_{k=0}^{n-1} A_k\right) = 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t},$$

che può riguardarsi come funzione di distribuzione di una variabile aleatoria assolutamente continua. \diamond

Definizione 4.12 Siano $\lambda > 0$ e n un intero positivo. Una variabile aleatoria X di funzione di distribuzione

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!} e^{-\lambda x}, & x \geq 0 \end{cases} \quad (4.38)$$

e corrispondente densità di probabilità

$$f_X(x) = \begin{cases} \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x}, & x > 0 \\ 0, & \text{altrimenti} \end{cases} \quad (4.39)$$

si dice avere distribuzione di Erlang di parametri n e λ .

Si noti che la (4.39) si ottiene dalla (4.38) osservando che:

$$\frac{d}{dx} \left\{ e^{-\lambda x} \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!} \right\} = - \frac{(\lambda x)^{n-1}}{(n-1)!} \lambda e^{-\lambda x}.$$

Evidentemente, per ogni $\lambda > 0$ si ha:

$$f_X(x) \geq 0 \quad \forall x \in \mathbb{R}, \quad \int_{-\infty}^{+\infty} f_X(x) dx = \int_0^{+\infty} \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x} dx = 1,$$

dove l'ultima uguaglianza segue osservando che risulta:

$$\int_0^{+\infty} x^{n-1} e^{-x} dx = (n-1)!,$$

come si evince mediante ripetute integrazioni per parti. In particolare, se nella (4.39) si pone $n = 1$, si ritrova la densità di probabilità esponenziale.

Nel seguito con la notazione $X \sim \mathcal{E}(n, \lambda)$ intenderemo che X è una variabile aleatoria avente distribuzione di Erlang di parametri n e λ , o più semplicemente *variabile di Erlang*. In Figura 4.11 è rappresentata la densità di probabilità di una variabile $X \sim \mathcal{E}(n, 1)$ per $n = 2, 3, 4$.

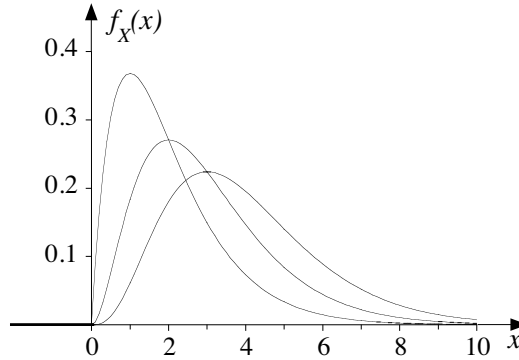


Figura 4.11 – Densità (4.39) con $\lambda = 1$ e $n = 2, 3, 4$ dall'alto verso il basso in prossimità dell'origine.

Proposizione 4.15 Siano $X \sim \mathcal{E}(1, \lambda)$ e $Y \sim \mathcal{E}(n, \lambda)$ variabili aleatorie indipendenti. Posto $Z = X + Y$, si ha $Z \sim \mathcal{E}(n + 1, \lambda)$, ossia Z ha densità di Erlang di parametri $n + 1$ e λ .

Dimostrazione Si osservi in primo luogo che deve aversi $F_Z(z) = 0$ per ogni $z < 0$, essendo X e Y positive quasi certamente. Inoltre, per ogni $z \geq 0$ risulta:

$$F_Z(z) = P(X + Y \leq z) = \iint_A f_{X,Y}(x, y) dx dy,$$

dove $A = \{(x, y) \in \mathbb{R}^2: x \geq 0, y \geq 0, x + y \leq z\}$. Facendo uso dell'indipendenza di X e Y e delle rispettive densità, per $z \geq 0$ si ha:

$$\begin{aligned} F_Z(z) &= \int_0^z \frac{\lambda^n y^{n-1}}{(\nu-1)!} e^{-\lambda y} dy \int_0^{z-y} \lambda e^{-\lambda x} dx = \int_0^z \frac{\lambda^n y^{n-1}}{(\nu-1)!} e^{-\lambda y} [1 - e^{-\lambda(z-y)}] dy \\ &= \int_0^z \frac{\lambda^n y^{n-1}}{(\nu-1)!} e^{-\lambda y} dy - e^{-\lambda z} \int_0^z \frac{\lambda^n y^{n-1}}{(\nu-1)!} dy = F_Y(z) - \frac{(\lambda z)^n}{n!} e^{-\lambda z}. \end{aligned}$$

Pertanto, per $n = 1, 2, \dots$ la densità di probabilità di Z è la seguente:

$$f_Z(z) = \begin{cases} \frac{\lambda^{n+1} z^n}{n!} e^{-\lambda z}, & z > 0 \\ 0, & \text{altrimenti,} \end{cases}$$

così che Z è di distribuzione di Erlang di parametri $n + 1$ e λ . □

Da quanto esposto nella Proposizione 4.15 si trae che una variabile aleatoria avente distribuzione di Erlang di parametri n e λ può essere riguardata come somma di n variabili indipendenti esponenziali di parametro λ .

Proposizione 4.16 Se Y_1, Y_2, \dots, Y_n sono variabili aleatorie indipendenti con $Y_i \sim \mathcal{E}(k_i, \lambda)$ ($i = 1, 2, \dots, n$), allora $Z = Z_1 + Z_2 + \dots + Z_n$ ha distribuzione di Erlang di parametri $k_1 + k_2 + \dots + k_n$ e λ .

Dimostrazione Poiché ognuna delle Y_i può essere espressa come somma di k_i variabili aleatorie indipendenti ed esponenziali di parametro λ , per l'indipendenza delle Y_i la loro somma è interpretabile come somma di $k_1 + k_2 + \dots + k_n$ variabili aleatorie indipendenti esponenziali, così che Z ha distribuzione di Erlang di parametri $k_1 + k_2 + \dots + k_n$ e λ . \square

Esempio 4.26 Consideriamo un centro di servizio ad ingresso unico che consiste di k stazioni di lavoro identiche ed indipendenti, come schematizzato nella Figura 4.12. Supponiamo

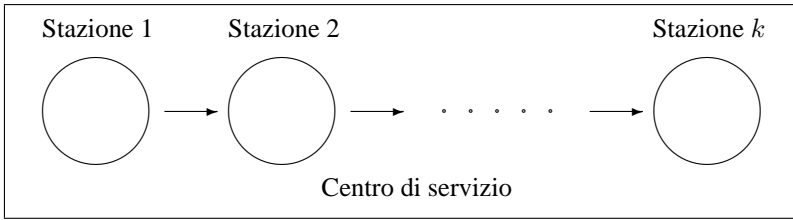


Figura 4.12 – Ad illustrazione dell'Esempio 4.26.

che la durata di servizio nella generica stazione j (ossia il tempo misurato dall'istante in cui l'utente entra nella stazione j -esima fino a quando ne esce) sia descritto da una variabile aleatoria $S_j \sim \mathcal{E}(1, \mu)$ ($j = 1, 2, \dots, k$). Una situazione reale tipica è, ad esempio, quella di un centro di assistenza automobilistico che prevede varie operazioni successive sulle auto in ingresso (controllo pneumatici, lavaggio e ingrassaggio, ecc.). Sia S la variabile aleatoria che rappresenta la durata di servizio di un utente (ossia il tempo misurato dall'istante in cui l'utente entra nella prima stazione fino a quando esce dalla k -esima stazione). Evidentemente, $S = S_1 + S_2 + \dots + S_k$ è la somma di k variabili aleatorie indipendenti ed esponenzialmente distribuite di parametro μ . Pertanto, in virtù della Proposizione 4.16, $S \sim \mathcal{E}(k, \mu)$, ossia S è caratterizzata da densità di Erlang di parametri k e μ . \diamond

La densità di probabilità (4.39) è un caso particolare della densità di probabilità gamma che passiamo a definire.

Definizione 4.13 Una variabile aleatoria X di densità di probabilità

$$f_X(x) = \begin{cases} \frac{\lambda^\nu x^{\nu-1}}{\Gamma(\nu)} e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0, \end{cases} \quad (4.40)$$

con λ, ν reali positivi e dove

$$\Gamma(\nu) = \int_0^{+\infty} x^{\nu-1} e^{-x} dx, \quad \nu > 0, \quad (4.41)$$

è la ben nota funzione gamma di Eulero, si dice di distribuzione gamma di parametri ν e λ .

Poiché per $\nu > 1$ sussiste la seguente proprietà di fattorizzazione:

$$\Gamma(\nu) = (\nu - 1) \Gamma(\nu - 1) \quad (\nu > 1), \quad (4.42)$$

se ν è un intero positivo, usando iterativamente la (4.42), si ottiene:

$$\Gamma(\nu) = (\nu - 1)! \quad (\nu = 1, 2, \dots),$$

avendo fatto uso della proprietà $\Gamma(1) = 1$ direttamente ricavata dalla (4.41).

Si noti che per $\nu = n$ la densità (4.40) coincide con la densità di Erlang di parametri n e λ , definita nella (4.39).

Nel seguito con $X \sim \mathcal{G}(\nu, \lambda)$ intenderemo che X ha distribuzione gamma di parametri ν e λ ; X sarà anche detta *variabile gamma*.

4.4.4 Distribuzione iperesponenziale

La funzione di distribuzione iperesponenziale è una combinazione lineare di funzioni di distribuzione di tipo esponenziale; essa interviene nella costruzione di modelli di fenomeni aleatori, quali tempi di attesa o di servizio.

Definizione 4.14 Siano a_1, a_2, \dots, a_n reali con $a_i \geq 0$ ($i = 1, 2, \dots, n$) e $\sum_{i=1}^n a_i = 1$. Una variabile aleatoria X di densità di probabilità

$$f_X(x) = \begin{cases} \sum_{i=1}^n a_i \lambda_i e^{-\lambda_i x}, & x > 0 \\ 0, & \text{altrimenti,} \end{cases} \quad (4.43)$$

con $\lambda_i > 0$ ($i = 1, 2, \dots, n$), si dice di distribuzione iperesponenziale di parametri $a_1, a_2, \dots, a_n, \lambda_1, \lambda_2, \dots, \lambda_n$.

È immediato verificare che risulta:

$$\int_{-\infty}^{+\infty} f_X(x) dx = \sum_{i=1}^n a_i \int_0^{+\infty} \lambda_i e^{-\lambda_i x} dx = \sum_{i=1}^n a_i = 1.$$

La funzione di distribuzione di X è immediatamente ottenibile dalla (4.43) per integrazione, avendosi:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \sum_{i=1}^n a_i [1 - e^{-\lambda_i x}] = 1 - \sum_{i=1}^n a_i e^{-\lambda_i x}, & x \geq 0. \end{cases} \quad (4.44)$$

Per $n = 1$, oppure se $\lambda_1 = \lambda_2 = \dots = \lambda_n$, la (4.43) coincide con una densità esponenziale di parametro λ_1 .

In Figura 4.13 è mostrata la densità di probabilità iperesponenziale (4.43) per $n = 2$, $a_1 = 0.4$, $a_2 = 0.6$, $\lambda_1 = 0.5 \lambda$, $\lambda_2 = 3 \lambda$ con $\lambda = 1, 0.5, 0.25$ (dall'alto verso il basso in prossimità dell'origine).

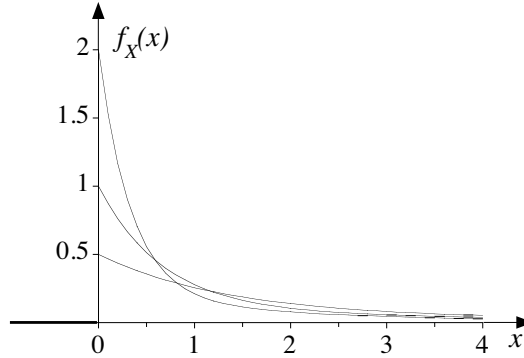


Figura 4.13 – La densità di probabilità (4.43) per alcune scelte dei parametri.

Esempio 4.27 Consideriamo un centro di assistenza consistente in un'unica stazione di lavoro che fornisce k diversi tipi di servizi. Supponiamo che la probabilità che l'utente richieda un servizio di tipo i sia p_i , con $p_i \geq 0$ ($i = 1, 2, \dots, k$) e $\sum_{i=1}^k p_i = 1$. Una situazione tipica è quella di uno sportello informativo in grado di erogare k tipi di informazioni. Supponiamo, inoltre, che la durata di un servizio di tipo i sia descritta dalla variabile aleatoria $S_i \sim \mathcal{E}(1, \lambda_i)$ per $i = 1, 2, \dots, k$. Determiniamo la distribuzione della variabile aleatoria S descrivente il cosiddetto tempo effettivo di servizio, ossia il tempo necessario per soddisfare un qualsiasi tipo di richiesta dell'utente.

Se si pone

$$\begin{aligned} A_t &= \{\text{il tempo di servizio non supera } t\} \\ B_r &= \{\text{l'utente ha richiesto il servizio di tipo } r\} \quad (r = 1, 2, \dots, k), \end{aligned}$$

risulta:

$$A_t = \bigcup_{i=1}^k (B_i \cap A_t).$$

Infatti l'evento A_t si realizza se e solo se si verifica uno qualunque dei k eventi incompatibili B_1, B_2, \dots, B_k . Pertanto, $P(S \leq t) = 0$ per $t < 0$, mentre per $t \geq 0$ si ha:

$$\begin{aligned} P(S \leq t) &= P(A_t) = \sum_{i=1}^k P(B_i \cap A_t) = \sum_{i=1}^k P(B_i) P(A_t | B_i) \\ &= \sum_{i=1}^k p_i P(S_i < t) = \sum_{i=1}^k p_i [1 - e^{-\mu_i t}]. \end{aligned}$$

Pertanto S ha dunque distribuzione iperesponenziale di parametri $p_1, p_2, \dots, p_k, \lambda_1, \lambda_2, \dots, \lambda_k$. \diamond

4.4.5 Distribuzione di Weibull

La funzione di distribuzione di Weibull trova applicazioni nella cosiddetta teoria dell'affidabilità che descrive fenomeni aleatori di rottura, decadimento, crollo, ecc.

Definizione 4.15 Una variabile aleatoria X di funzione di distribuzione

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - \exp\{-\lambda x^\alpha\}, & x \geq 0 \end{cases} \quad (\alpha > 0, \lambda > 0) \quad (4.45)$$

e corrispondente densità di probabilità

$$f_X(x) = \begin{cases} \alpha \lambda x^{\alpha-1} \exp\{-\lambda x^\alpha\}, & x > 0 \\ 0, & \text{altrimenti,} \end{cases} \quad (4.46)$$

con $\alpha > 0$ e $\lambda > 0$, si dice avere distribuzione di Weibull di parametri λ e α .

Si noti che si ha:

$$\int_{-\infty}^{+\infty} f_X(x) dx = \alpha \lambda \int_0^{+\infty} x^{\alpha-1} \exp\{-\lambda x^\alpha\} dx = \int_0^{+\infty} e^{-y} dy = 1.$$

Per $\alpha = 1$ la (4.46) coincide con la densità esponenziale di parametro λ .

In Figura 4.14 sono mostrate le densità di probabilità di Weibull ottenute scegliendo $\lambda = 1$ e $\alpha = 0.5, 1, 1.5$ (dall'alto verso il basso in prossimità dell'origine).

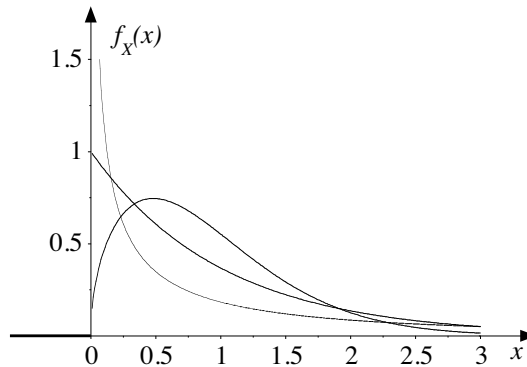


Figura 4.14 – Densità di probabilità (4.46) con $\alpha = 0.5, 1, 1.5$ e con $\lambda = 1$.

Proposizione 4.17 Siano X_1, X_2, \dots variabili aleatorie non-negative indipendenti e identicamente distribuite di funzione di distribuzione $F(x)$ tale da aversi:

$$\lim_{x \downarrow 0} \frac{F(x)}{\lambda x^\alpha} = 1, \quad (4.47)$$

con $\alpha > 0$ e $\lambda > 0$. Posto

$$Y_n = n^{1/\alpha} \min\{X_1, X_2, \dots, X_n\} \quad (n = 1, 2, \dots),$$

si ha che $\lim_{n \rightarrow +\infty} F_{Y_n}(x)$ è una funzione di distribuzione di Weibull di parametri λ e α .

Dimostrazione Dalla (3.64) segue:

$$\begin{aligned} F_{Y_n}(x) &= P(Y_n \leq x) = P\left(\min\{X_1, X_2, \dots, X_n\} \leq \frac{x}{n^{1/\alpha}}\right) \\ &= \begin{cases} 0 & x < 0 \\ 1 - \left[1 - F_X\left(\frac{x}{n^{1/\alpha}}\right)\right]^n, & x \geq 0. \end{cases} \end{aligned}$$

Dall'ipotesi (4.47) segue:

$$\begin{aligned} \lim_{n \rightarrow +\infty} \left[1 - F_X\left(\frac{x}{n^{1/\alpha}}\right)\right]^n &= \lim_{n \rightarrow +\infty} \left[1 - \lambda \left(\frac{x}{n^{1/\alpha}}\right)^\alpha\right]^n \\ &= \lim_{n \rightarrow +\infty} \left(1 - \frac{\lambda x^\alpha}{n}\right)^n = \exp\{-\lambda x^\alpha\} \end{aligned}$$

così che dalla (4.48) per $x > 0$ si trae:

$$\lim_{n \rightarrow +\infty} F_{Y_n}(x) = 1 - \exp\{-\lambda x^\alpha\}.$$

□

La Proposizione 4.17 mostra che se la funzione di distribuzione $F(x)$ delle variabili aleatorie indipendenti e identicamente distribuite X_1, X_2, \dots soddisfa la (4.47), per n grande sussiste la seguente approssimazione:

$$P(\min\{X_1, X_2, \dots, X_n\} \leq x) \simeq 1 - e^{-\lambda n x^\alpha}, \quad x > 0. \quad (4.48)$$

Si noti che se X_1, X_2, \dots, X_n descrivono i tempi di corretto funzionamento di n componenti di un dispositivo collegati in serie, e se $F(x)$ soddisfa la (4.47), allora per n grande la funzione di distribuzione di $V = \min\{X_1, X_2, \dots, X_n\}$, che rappresenta il tempo di corretto funzionamento dell'intero dispositivo, è approssimabile con una funzione di distribuzione di Weibull di parametri λ e α .

Esempio 4.28 Siano X_1, X_2, \dots, X_n variabili aleatorie indipendenti con $X_i \sim \mathcal{E}(k, \mu)$ ($i = 1, 2, \dots, n$). Calcoliamo la funzione di distribuzione $F_V(x)$ di $V = \min\{X_1, X_2, \dots, X_n\}$.

Dalla (3.64) segue:

$$F_V(x) = 1 - [1 - F(x)]^n = \begin{cases} 0, & x < 0 \\ 1 - e^{-n\mu x} \left[\sum_{j=0}^{k-1} \frac{(\mu x)^j}{j!} \right]^n, & x \geq 0, \end{cases} \quad (4.49)$$

dove $F(x)$ denota la comune funzione di distribuzione delle X_i .

Facendo uso della Proposizione 4.17 si può ottenere un'approssimazione per $F_V(x)$ valida per n grande. Infatti, dalla Definizione 4.12 segue che X_i ha densità di probabilità:

$$f(x) = \begin{cases} \frac{\mu^k x^{k-1}}{(k-1)!} e^{-\mu x}, & x > 0 \\ 0, & \text{altrimenti.} \end{cases}$$

Inoltre, risulta:

$$\lim_{x \downarrow 0} \frac{F(x)}{\lambda x^\alpha} = \lim_{x \downarrow 0} \frac{f(x)}{\lambda \alpha x^{\alpha-1}} = \begin{cases} 0, & k > \alpha \\ \frac{\mu^k}{\lambda k!}, & k = \alpha \\ +\infty, & k < \alpha. \end{cases}$$

Tabella 4.4—I valori delle funzioni $F_V(x)$ e $G(x)$ date dalle (4.49) e (4.50), approssimati alla sesta cifra significativa, sono riportati per alcuni valori di x avendo fissato $k = 3$, $\mu = 0.01$ e $n = 10^6$.

x	$F_V(x)$	$G(x)$
0.1	0.00447982	0.00448989
0.2	0.0352038	0.0353597
0.3	0.113685	0.114409
0.4	0.248302	0.250238
0.5	0.426622	0.430217
0.6	0.616713	0.621674
0.7	0.781156	0.786368
0.8	0.895954	0.900141
0.9	0.959836	0.962391
1.0	0.987722	0.988891

Pertanto la (4.47) è soddisfatta se e solo se si sceglie $\alpha = k$ e $\lambda = \mu^k/k!$. Facendo uso della (4.48), per n sufficientemente grande segue:

$$F_V(x) \simeq G(x) = 1 - \exp \left\{ -n \frac{(\mu x)^k}{k!} \right\}, \quad x > 0. \quad (4.50)$$

Nella Tabella 4.4 per $k = 3$, $\mu = 0.01$ e $n = 10^6$ sono riportati i valori delle funzioni di distribuzione $F_V(x)$ e $G(x)$ in corrispondenza di alcune scelte di x . \diamond

4.4.6 Distribuzione normale

La funzione di distribuzione normale, detta anche di Gauss o gaussiana, riveste estrema importanza nel calcolo delle probabilità e nella statistica anche in quanto essa costituisce una distribuzione limite alla quale tendono varie altre funzioni di distribuzioni sotto opportune ipotesi.

Definizione 4.16 Una variabile aleatoria X di densità di probabilità

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad x \in \mathbb{R} \quad (\mu \in \mathbb{R}, \sigma > 0), \quad (4.51)$$

si dice avere distribuzione normale di parametri μ e σ .

La notazione $X \sim \mathcal{N}(\mu, \sigma)$ verrà utilizzata nel seguito per indicare che X ha distribuzione normale di parametri μ e σ , o più semplicemente che è una *variabile normale*.

Dalla (4.51) è evidente che $f_X(x) \geq 0$ per ogni $x \in \mathbb{R}$. Meno immediato è rendersi conto che risulta $\int_{-\infty}^{+\infty} f_X(x) dx = 1$. Per dimostrarlo, poniamo $A = \int_{-\infty}^{+\infty} f_X(x) dx$, così che

$$\begin{aligned} A^2 &= \frac{1}{2\pi\sigma^2} \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} \exp\left\{-\frac{(x-\mu)^2 + (y-\mu)^2}{2\sigma^2}\right\} dy \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} du \int_{-\infty}^{+\infty} \exp\left\{-\frac{u^2 + v^2}{2}\right\} dv \end{aligned}$$

dove l'ultima uguaglianza è stata ottenuta effettuando il cambiamento di variabili $u = (x - \mu)/\sigma$, $v = (y - \mu)/\sigma$. Passando a coordinate polari, ossia ponendo $u = \varrho \cos \vartheta$ e $v = \varrho \sin \vartheta$, e ricordando che lo Jacobiano di tale trasformazione è ϱ , si ha:

$$A^2 = \frac{1}{2\pi} \int_0^{+\infty} \varrho \exp\left\{-\frac{\varrho^2}{2}\right\} d\varrho \int_0^{2\pi} d\vartheta = 1,$$

da cui segue $A = 1$, il che dimostra l'asserto.

Dalla (4.51) si evince che per ogni $x \in \mathbb{R}$ risulta $f_X(\mu - x) = f_X(\mu + x)$; pertanto la densità normale è simmetrica rispetto all'asse $x = \mu$. Si ha inoltre per ogni $x \in \mathbb{R}$:

$$\begin{aligned} \frac{df_X(x)}{dx} &= -\frac{x-\mu}{\sigma^3\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} = -\frac{x-\mu}{\sigma^2} f_X(x), \\ \frac{d^2f_X(x)}{dx^2} &= -\frac{(x-\mu)^2 - \sigma^2}{\sigma^5\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} = -\frac{(x-\mu)^2 - \sigma^2}{\sigma^4} f_X(x). \end{aligned}$$

Quindi $f_X(x)$ presenta il massimo $(\sigma\sqrt{2\pi})^{-1}$ nel punto di ascissa $x = \mu$ e due flessi nei punti di ascisse $\mu - \sigma$ e $\mu + \sigma$. Il grafico di $f_X(x)$ esibisce una caratteristica forma a campana, simmetrica rispetto a $x = \mu$, come indicato nella Figura 4.15. Variazioni del parametro μ comportano traslazioni della curva lungo l'asse delle ascisse; infatti, al crescere del parametro μ la curva si sposta lungo l'asse delle ascisse senza cambiare forma, come illustrato nella Figura 4.16 per $\mu = -2, -1, 0, 1, 2$ e $\sigma = 1$. Il parametro σ , pari alla semiampiezza tra i due punti di flesso, caratterizza la larghezza della funzione. Poiché l'ordinata massima è inversamente proporzionale a σ , al crescere di σ questa decresce, mentre l'area sottesa dalla densità deve rimanere unitaria. Quindi al crescere di σ la curva diventa sempre più piatta, mentre al decrescere di σ essa si allunga verso l'alto restringendosi contemporaneamente ai lati. Nella Figura 4.17 è riportato il grafico della densità di probabilità normale per $\mu = 0$ e $\sigma = 1/\sqrt{2}, 1, \sqrt{2}$ (dall'alto verso il basso in prossimità dell'origine).

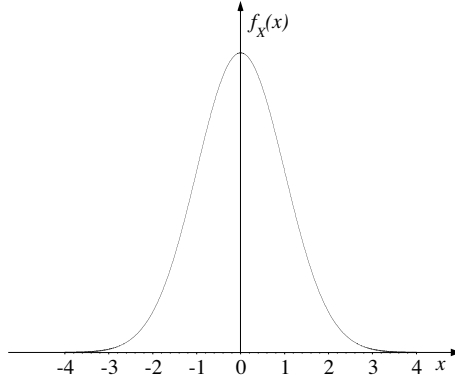


Figura 4.15 – Densità di probabilità di $X \sim \mathcal{N}(0, 1)$.

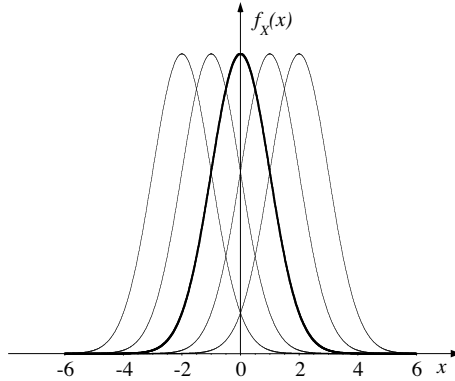


Figura 4.16 – Densità di probabilità di $X \sim \mathcal{N}(\mu, 1)$ con $\mu = -2, -1, 0, 1, 2$.

La funzione di distribuzione di una variabile aleatoria X normale di parametri μ e σ è:

$$\begin{aligned} F_X(x) &= \int_{-\infty}^x f_X(y) dy = \frac{1}{\sqrt{2\pi}\sigma^2} \int_{-\infty}^x \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\} dy \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-\mu)/\sigma} \exp\left\{-\frac{z^2}{2}\right\} dz. \end{aligned} \quad (4.52)$$

Particolare importanza riveste il caso $\mu = 0$, $\sigma = 1$. La variabile aleatoria $Z \sim \mathcal{N}(0, 1)$ è detta *normale standard*. Dalla (4.51) seguono le corrispondenti densità di probabilità

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\}, \quad x \in \mathbb{R}, \quad (4.53)$$

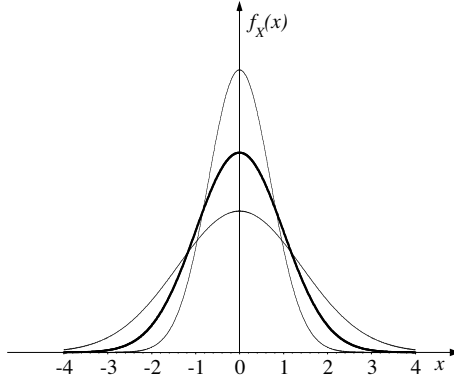


Figura 4.17 – Densità di probabilità di $X \sim \mathcal{N}(0, \sigma)$ con $\sigma = 1/\sqrt{2}, 1, \sqrt{2}$.

e funzione di distribuzione:

$$\Phi(z) = F_Z(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp\left\{-\frac{y^2}{2}\right\} dy, \quad z \in \mathbb{R}. \quad (4.54)$$

Come ogni funzione di distribuzione, $\Phi(z)$ gode delle seguenti proprietà:

- (a) $\lim_{z \rightarrow -\infty} \Phi(z) = 0$;
- (b) $\lim_{z \rightarrow +\infty} \Phi(z) = 1$;
- (c) $\Phi(z)$ è una funzione non decrescente.

Inoltre, si ha:

$$\Phi(-x) = 1 - \Phi(x), \quad x \in \mathbb{R} \quad (4.55)$$

conseguenza della simmetria della densità normale standard:

$$\Phi(-x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-x} \exp\left\{-\frac{y^2}{2}\right\} dy = \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} \exp\left\{-\frac{y^2}{2}\right\} dy = 1 - \Phi(x).$$

La (4.54) può essere valutata per via numerica. Facendo uso della proprietà (4.55) spesso ci si limita a tabellare $\Phi(z)$ soltanto per valori positivi dell'argomento.

In virtù delle (4.54), la funzione di distribuzione di $X \sim \mathcal{N}(\mu, \sigma)$ può essere espressa in termini della funzione di distribuzione normale standard $\Phi(x)$:

$$F_X(x) = \Phi\left(\frac{x - \mu}{\sigma}\right). \quad (4.56)$$

Pertanto, se $X \sim \mathcal{N}(\mu, \sigma)$ si ha:

$$P(a < X < b) = F_X(b) - F_X(a) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right). \quad (4.57)$$

Per $a = \mu - \varepsilon$ e $b = \mu + \varepsilon$, con $\varepsilon > 0$, da questa segue:

$$P(\mu - \varepsilon < X < \mu + \varepsilon) = \Phi\left(\frac{\varepsilon}{\sigma}\right) - \Phi\left(-\frac{\varepsilon}{\sigma}\right) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right) - 1. \quad (4.58)$$

Quindi,

$$\begin{aligned} P(\mu - \sigma < X < \mu + \sigma) &= 2\Phi(1) - 1 \simeq 2 \cdot 0.8413 - 1 = 0.6826, \\ P(\mu - 2\sigma < X < \mu + 2\sigma) &= 2\Phi(2) - 1 \simeq 2 \cdot 0.9772 - 1 = 0.9544, \\ P(\mu - 3\sigma < X < \mu + 3\sigma) &= 2\Phi(3) - 1 \simeq 2 \cdot 0.9987 - 1 = 0.9974, \end{aligned} \quad (4.59)$$

dove i valori numerici riportati seguono dalla tabella della distribuzione normale standard (v. Appendice C). Essi mostrano che la probabilità che una variabile aleatoria $X \sim \mathcal{N}(\mu, \sigma)$ assuma valori in un intervallo avente come centro μ e semiampiezza 3σ è prossima all'unità. Questa proprietà delle variabili aleatorie normali è nota come *regola del 3σ* .

Esempio 4.29 Sia $X \sim \mathcal{N}(\mu, \sigma)$, con $\mu \in \mathbb{R}$ e $\sigma > 0$. Determinare la distribuzione di $Y = aX + b$, con $a, b \in \mathbb{R}$ e $a \neq 0$.

Ricordando la (3.27), dalla (4.51) segue che la densità di probabilità di Y è:

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right) = \frac{1}{\sqrt{2\pi\sigma^2 a^2}} \exp\left\{-\frac{(y-a\mu-b)^2}{2\sigma^2 a^2}\right\}, \quad x \in \mathbb{R},$$

che mostra che $X \sim \mathcal{N}(a\mu + b, |a|\sigma)$

Esempio 4.30 Supponiamo che $X \sim \mathcal{N}(\mu, \sigma)$, con $\mu = 0$ e $\sigma = 0.01$, descriva l'errore di misura nel valutare una certa distanza. Determiniamo la probabilità p che il valore assoluto dell'errore di misura sia minore di $\beta = 0.02$.

Facendo uso delle (4.55), (4.57) e (4.59) si ottiene:

$$\begin{aligned} p &= P(|X| < \beta) = P(-\beta < X < \beta) = \Phi\left(\frac{\beta}{\sigma}\right) - \Phi\left(-\frac{\beta}{\sigma}\right) \\ &= 2\Phi\left(\frac{\beta}{\sigma}\right) - 1 = 2\Phi(2) - 1 \simeq 0.9544, \end{aligned}$$

dove il valore di $\Phi(2)$ è ottenuto dalla Tabella C.1 della distribuzione normale dell'Appendice. \diamond

Esempio 4.31 Sia $X \sim \mathcal{N}(\mu, \sigma)$, con $\mu \in \mathbb{R}$ e $\sigma > 0$. Determinare il reale ε tale che $P(|X - \mu| \geq \varepsilon) = 1/2$.

In virtù della (4.58) si ha:

$$P(|X - \mu| \geq \varepsilon) = 1 - P(|X - \mu| < \varepsilon) = 1 - P(\mu - \varepsilon < X < \mu + \varepsilon) = 2 - 2\Phi\left(\frac{\varepsilon}{\sigma}\right),$$

da cui, imponendo che $P(|X - \mu| \geq \varepsilon) = 1/2$, si ricava $\Phi(\varepsilon/\sigma) = 3/4 = 0.75$. Facendo uso della Tabella C.1 della distribuzione normale dell'Appendice, si può determinare il valore x per il quale $\Phi(x) = 0.75$; così facendo si ottiene $x \simeq 0.68$. In conclusione, affinché $P(|X - \mu| \geq \varepsilon) = 1/2$, deve risultare $\varepsilon \simeq 0.68\sigma$. \diamond

Capitolo 5

Valore medio, momenti e funzioni generatrici

5.1 Introduzione

La distribuzione di una variabile aleatoria X può essere definita tramite la funzione di distribuzione, tramite la funzione di probabilità se X è discreta, oppure, nel caso di variabile aleatoria assolutamente continua, tramite la densità di probabilità. Conoscendo la distribuzione di X si è in grado, in linea di principio, di calcolare le probabilità di qualsiasi evento connesso ad X . Tuttavia, spesso l'attenzione non viene rivolta alla distribuzione di probabilità, vuoi perché essa non è ottenibile in forma esplicita, vuoi semplicemente perché si desidera confrontare soltanto caratteristiche particolari di variabili aleatorie. Tra queste caratteristiche grande interesse rivestono il *valore medio* (detto anche *valore di aspettazione* oppure *valore atteso*), la *varianza*, la *moda* e la *mediana*.

5.2 Valore medio

Il concetto di valore medio nasce come generalizzazione del concetto di media aritmetica, come è chiarito dai seguenti esempi.

Esempio 5.1 I voti riportati da N studenti sono x_1, x_2, \dots, x_n . Per ogni i , sia N_i la frequenza assoluta del voto x_i ; questa indica il numero di studenti che ha riportato il voto x_i . Chiaramente risulta $N_1 + N_2 + \dots + N_n = N$. La media aritmetica dei voti è dunque:

$$\bar{x} = x_1 \frac{N_1}{N} + x_2 \frac{N_2}{N} + \dots + x_n \frac{N_n}{N} = \sum_{i=1}^n x_i \frac{N_i}{N},$$

dove il rapporto N_i/N esprime la frequenza relativa del voto x_i ($i = 1, \dots, n$). La media \bar{x} fornisce una sorta di sintesi dei dati che può essere utilizzata per indicare concisamente il rendimento della classe. Si consideri ora l'esperimento consistente nell'estrazione di una biglia da un'urna contenente N biglie contrassegnate dai voti degli studenti e sia X la variabile aleatoria rappresentante il numero presente sulla biglia estratta, ossia il voto riportato

da uno studente scelto a caso. Essa assume i valori x_1, x_2, \dots, x_n con rispettive probabilità $N_1/N, N_2/N, \dots, N_n/N$. Definiamo il valore medio di X al seguente modo:

$$E(X) = \sum_{i=1}^n x_i \frac{N_i}{N}.$$

Pertanto, il valore medio $E(X)$ testé definito costituisce un'estensione in termini probabilistici del concetto statistico di media aritmetica. \diamond

Esempio 5.2 Si consideri l'esperimento consistente nel lanciare ripetutamente un dado e si supponga che in n lanci il numero i sia uscito n_i volte ($n_1 + n_2 + \dots + n_6 = n$). La media aritmetica dei numeri ottenuti è data da

$$\bar{x} = \frac{1n_1 + 2n_2 + \dots + 6n_6}{n} = 1 \frac{n_1}{n} + 2 \frac{n_2}{n} + \dots + 6 \frac{n_6}{n} = 1f_n(1) + 2f_n(2) + \dots + 6f_n(6),$$

dove $f_n(1), f_n(2), \dots, f_n(6)$ sono le frequenze relative con cui sono comparsi i numeri $1, 2, \dots, 6$ negli n lanci, soddisfacenti ovviamente la relazione $f_n(1) + f_n(2) + \dots + f_n(6) = 1$. Al crescere del numero delle prove, per la definizione frequentista di probabilità tali frequenze tendono alle corrispondenti probabilità con cui si verificano i numeri nel lancio del dado. Supponiamo ora che un giocatore paghi inizialmente una certa somma per essere ammesso al gioco e che in ogni lancio del dado guadagni un importo uguale al numero uscito nel lancio. Consideriamo poi la variabile aleatoria X che descrive il guadagno che il giocatore si aspetta di conseguire in un lancio del dado. Questa assume i valori $1, 2, \dots, 6$ con probabilità $p_X(1), p_X(2), \dots, p_X(6)$, rispettivamente. Appare allora naturale definire il valore medio della variabile aleatoria X al seguente modo:

$$E(X) = \sum_{i=1}^6 i p_X(i),$$

esprimente il guadagno medio che un giocatore si aspetta di conseguire in un lancio del dado. \diamond

Gli esempi considerati suggeriscono la seguente definizione di valore medio per una variabile aleatoria discreta non negativa oppure non positiva.

Definizione 5.1 Sia X una variabile aleatoria discreta.

(a) Se X è non negativa (ossia se $P(X \geq 0) = 1$) ed assume valori in un insieme S al più numerabile di reali distinti non negativi x_1, x_2, \dots con rispettive probabilità $p_X(x_1), p_X(x_2), \dots$, il valore medio di X è

$$E(X) = \sum_{r: x_r \in S} x_r p_X(x_r). \quad (5.1)$$

(b) Se X è non positiva (ossia se $P(X \leq 0) = 1$) ed assume valori in un insieme S al più numerabile di reali distinti non positivi x_1, x_2, \dots con rispettive probabilità $p_X(x_1), p_X(x_2), \dots$, il valore medio di X è di nuovo dato dalla (5.1).

Seguendo la tradizione, è stata finora utilizzata la lettera E in quanto iniziale della parola inglese *EXPECTATION* (valore atteso), storicamente legata ai giochi d'azzardo.

Proposizione 5.1 *Sia c un numero reale. Risulta $E(c) = c$.*

Dimostrazione È sufficiente definire una variabile aleatoria X degenera che assume il valore c con probabilità 1. Poiché X è discreta, dalla (5.1) segue immediatamente $E(X) = E(c) = c$. \square

Proposizione 5.2 *Sia X una variabile aleatoria discreta. Se X assume i valori $1, 2, \dots, n$ si ha:*

$$E(X) = \sum_{k=1}^n P(X \geq k), \quad (5.2)$$

mentre se X assume i valori $1, 2, \dots$ risulta:

$$E(X) = \sum_{k=1}^{+\infty} P(X \geq k). \quad (5.3)$$

Dimostrazione Se X assume i valori $1, 2, \dots, n$, dalla Definizione 5.1 segue:

$$E(X) = \sum_{j=1}^n j P(X = j) = \sum_{j=1}^n P(X = j) \sum_{k=1}^j 1 = \sum_{k=1}^n \sum_{j=k}^n P(X = j) = \sum_{k=1}^n P(X \geq k),$$

ossia la (5.2). Se invece X assume i valori $1, 2, \dots$ procedendo al limite per $n \rightarrow +\infty$ nella (5.2) si ottiene immediatamente la (5.3). \square

Esempio 5.3 Sia $X \sim \mathcal{U}_d(n)$. Poiché X assume i valori $1, 2, \dots, n$, ciascuno con probabilità $1/n$, facendo uso della (5.1) si ha:

$$E(X) = \sum_{r=1}^n r \frac{1}{n} = \frac{1}{n} \sum_{r=1}^n r = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}. \quad (5.4)$$

Allo stesso risultato si giunge utilizzando la Proposizione 5.2. Infatti, poiché $P(X \geq k) = (n - k + 1)/n$, dalla (5.2) si ottiene:

$$E(X) = \sum_{k=1}^n P(X \geq k) = \sum_{k=1}^n \frac{n - k + 1}{n} = \sum_{r=1}^n \frac{r}{n} = \frac{n+1}{2}.$$

Si noti che se n è un numero pari, $E(X)$ non coincide con nessuno dei possibili valori di X ; ad esempio, per $n = 6$ dalla (5.4) segue $E(X) = 3.5$. \diamond

Esempio 5.4 Sia X una variabile aleatoria di distribuzione geometrica di parametro p . Poiché $P(X \geq k) = (1 - p)^{k-1}$ per $k = 1, 2, \dots$, dalla (5.3) segue:

$$E(X) = \sum_{k=1}^{+\infty} P(X \geq k) = \sum_{k=1}^{+\infty} (1 - p)^{k-1} = \sum_{r=0}^{+\infty} (1 - p)^r = \frac{1}{p}. \quad (5.5)$$

\diamond

La Definizione 5.1 può essere di spunto per dare un'idea di come operare nel caso di variabili aleatorie assolutamente continue non negative oppure non positive. L'esempio che segue è finalizzato a chiarire come procedere.

Esempio 5.5 Due individui, A e B, decidono di incontrarsi ad una data ora. Poiché B è solito arrivare in ritardo agli appuntamenti, diversamente da A, è ragionevole identificare l'istante di arrivo di A come origine di una scala temporale e riguardare il ritardo con cui B arriva all'appuntamento, ossia il tempo di attesa di A, come una variabile aleatoria T assolutamente continua che qui assumeremo misurata in minuti primi ed avente densità di probabilità:

$$f_T(t) = \begin{cases} \frac{4}{90} \left[1 - \left(\frac{t}{30} \right)^3 \right], & 0 < t < 30 \\ 0, & \text{altrimenti.} \end{cases} \quad (5.6)$$

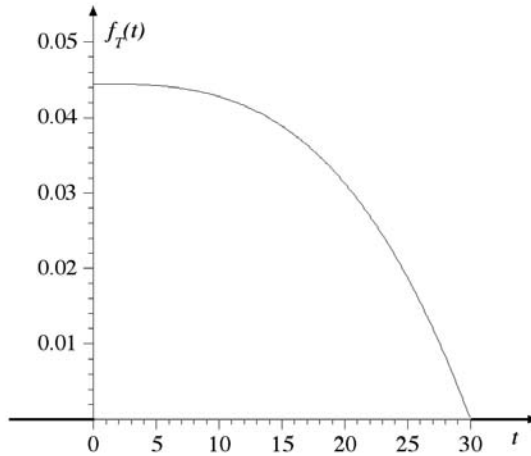


Figura 5.1 – Rappresentazione della densità di probabilità $f_T(t)$ data in (5.6).

La Figura 5.1 mostra il grafico di $f_T(t)$. Per calcolare il ritardo con cui mediamente B si presenta all'appuntamento procediamo nel seguente modo. Dividiamo l'intervallo $0 \leq t \leq 30$ in n sottointervalli di ampiezza $\Delta t = 30/n$ e denotiamo con t_1, t_2, \dots, t_n i rispettivi punti medi, ossia poniamo $t_i = (i - 1/2) \Delta t$ ($i = 1, 2, \dots, n$). Un'approssimazione al tempo medio di attesa di A è pertanto:

$$E_n = \sum_{i=1}^n t_i f_T(t_i) \Delta t = \frac{40}{n^2} \sum_{i=1}^n \left(i - \frac{1}{2} \right) \left[1 - \left(\frac{1}{n} \right)^3 \left(i - \frac{1}{2} \right)^3 \right] \quad (n = 1, 2, \dots). \quad (5.7)$$

Nella Tabella 5.1 sono riportati i valori di E_n per alcuni valori di n . Si noti che E_n si avvicina al valore 12 al crescere di n . Il limite di (5.7) per $n \rightarrow +\infty$ fornisce il tempo medio di attesa di A:

$$E(T) = \int_0^{30} t f_T(t) dt = \frac{4}{90} \int_0^{30} t \left[1 - \left(\frac{t}{30} \right)^3 \right] dt = 12.$$

Tabella 5.1 – I valori di E_n ottenuti dalla (5.7) sono riportati per alcune scelte di n .

n	E_n	n	E_n
1	17.500000	150	12.000296
10	12.066550	200	12.000167
50	12.002667	250	12.000107
100	12.000667	300	12.000074

Pertanto A uscendo da casa per recarsi in orario all'appuntamento con B si aspetta di rimanere in attesa presumibilmente per 12 minuti. \diamond

L'esempio precedente suggerisce che se X è una variabile aleatoria assolutamente continua non negativa (non positiva) di densità di probabilità $f_X(x)$, si può pensare di suddividere la semiretta positiva (negativa) in sottointervalli del tipo $[x_i, x_i + \Delta x_i)$. Ricordando che $P(x_i \leq X < x_i + \Delta x_i) \sim f_X(x_i) \Delta x_i$, si giunge ad un'approssimazione per l'aspettazione $E(X)$ del tipo $\sum_i x_i f_X(x_i) \Delta x_i$, che tende a migliorare al diminuire delle ampiezze Δx_i dei sottointervalli. Ciò conduce alla definizione che segue.

Definizione 5.2 Sia X una variabile aleatoria assolutamente continua.

(a) Se X è non negativa (ossia se $P(X \geq 0) = 1$) di densità di probabilità $f_X(x)$, il valore medio di X è

$$E(X) = \int_0^{+\infty} x f_X(x) dx. \quad (5.8)$$

(b) Se X è non positiva (ossia se $P(X \leq 0) = 1$) di densità di probabilità $f_X(x)$, il valore medio di X è

$$E(X) = \int_{-\infty}^0 x f_X(x) dx. \quad (5.9)$$

Esempio 5.6 Sia X una variabile aleatoria assolutamente continua avente funzione di distribuzione

$$F_X(x) = \begin{cases} 0, & x \leq 1 \\ 1 - \frac{1}{x}, & x > 1. \end{cases}$$

Calcoliamo $E(X)$.

A tal fine osserviamo che la densità di probabilità di X è

$$f_X(x) = \begin{cases} \frac{1}{x^2}, & x > 1 \\ 0, & \text{altrimenti} \end{cases}$$

(v. Figura 5.2). Poiché X è non negativa, il suo valore medio è:

$$E(X) = \int_1^{+\infty} x \frac{1}{x^2} dx = +\infty.$$

\diamond

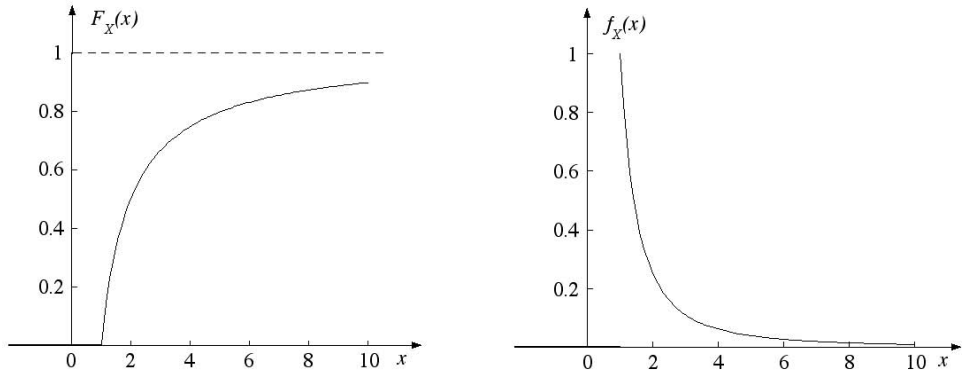


Figura 5.2 – Funzione di distribuzione e densità di probabilità dell'Esempio 5.6.

Finora abbiamo considerato variabili aleatorie discrete ed assolutamente continue non negative oppure non positive. Per esse il valore medio esiste sempre e può essere finito oppure infinito. Quando invece ci si riferisce a una variabile aleatoria che con probabilità non nulla assume valori sia positivi che negativi, per motivi che saranno chiariti nel seguito è necessario assicurarsi che le somme o gli integrali coinvolti nel calcolo del valore medio siano ben definiti.

Allo scopo di fornire una procedura di semplice utilizzazione per verificare l'esistenza del valore medio e per individuarne alcune semplici proprietà osserviamo che data una variabile aleatoria X è possibile associare ad essa le due variabili aleatorie X^+ e X^- così definite:

$$X^+ = \max(0, X), \quad X^- = \max(-X, 0), \quad (5.10)$$

rappresentanti rispettivamente la parte positiva e la parte negativa di X . Per ogni fissato risultato $\omega \in \Omega$ dell'esperimento casuale si ha:

$$X^+(\omega) = \begin{cases} 0, & X(\omega) < 0 \\ X(\omega), & X(\omega) \geq 0, \end{cases} \quad X^-(\omega) = \begin{cases} -X(\omega), & X(\omega) \leq 0 \\ 0, & X(\omega) > 0. \end{cases}$$

Pertanto entrambe X^+ e X^- assumono valori non negativi.

Dalla (5.10) risulta che X è esprimibile come differenza delle due variabili non negative:

$$X(\omega) = X^+(\omega) - X^-(\omega), \quad (5.11)$$

per ciascuna delle quali il valore medio è definito. Dalla (5.10) segue anche $|X(\omega)| = X^+(\omega) + X^-(\omega)$ (v. Figura 5.3).

Definizione 5.3 Si definisce *valore medio* di una variabile aleatoria X la differenza

$$E(X) = E(X^+) - E(X^-) \quad (5.12)$$

se almeno uno dei due valori medi $E(X^+)$ e $E(X^-)$ è finito. Se invece $E(X^+)$ e $E(X^-)$ sono entrambi divergenti si dice che il valore medio di X non esiste.

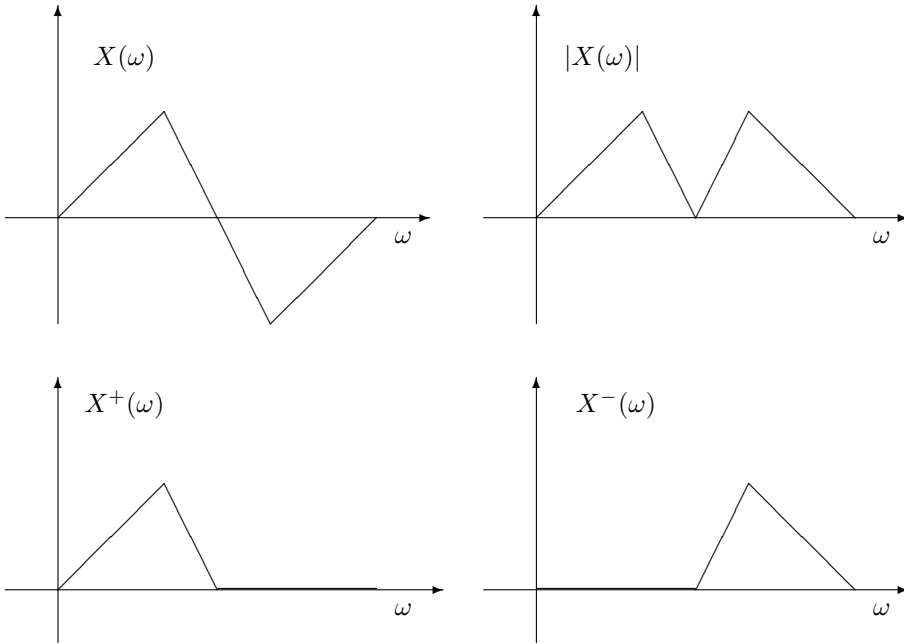


Figura 5.3 – Rappresentazione delle funzioni $X(\omega)$, $|X(\omega)|$, $X^+(\omega)$ e $X^-(\omega)$.

Quanto detto conduce alle seguenti semplici proposizioni di immediata dimostrazione:

Proposizione 5.3 *Se la variabile aleatoria X è non negativa oppure non positiva, allora il suo valore medio $E(X)$ esiste sempre, finito o infinito.*

Dimostrazione Se X è non negativa, per la Proposizione 5.1 risulta $E(X^-) = 0$ e quindi dalla (5.12) segue $E(X) = E(X^+)$. Analogamente, se X è non positiva, per la Proposizione 5.1 risulta $E(X^+) = 0$ e quindi dalla (5.12) si ottiene $E(X) = -E(X^-)$. \square

Proposizione 5.4 *Il valore medio $E(X)$ è finito se e solo se $E(X^+)$ e $E(X^-)$ sono entrambi finiti. Inoltre, $E(X)$ è finito se e solo se $E(|X|)$ è finito.*

Dimostrazione Dalla Definizione 5.3 discende che $E(X^+)$ e $E(X^-)$ sono entrambi finiti se $E(X)$ è finito, il che è soddisfatto se e solo se $E(|X|) = E(X^+) + E(X^-)$ è finito. \square

Proposizione 5.5 *Sia X una variabile aleatoria e si supponga che il suo valore medio $E(X)$ esista.*

(a) *Se X è discreta ed assume valori in un insieme S al più numerabile di reali distinti x_1, x_2, \dots con probabilità $p_X(x_1), p_X(x_2), \dots$ rispettivamente, si ha:*

$$E(X) = \sum_{r: x_r \in S} x_r p_X(x_r). \quad (5.13)$$

(b) Se X è assolutamente continua con densità di probabilità $f_X(x)$, risulta:

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx. \quad (5.14)$$

Dimostrazione Poiché X è una variabile aleatoria di valore medio $E(X)$, almeno uno dei due valori medi $E(X^+)$ e $E(X^-)$ è finito. Se X una variabile aleatoria discreta, facendo uso della Definizione 5.1 si ha:

$$E(X) = E(X^+) - E(X^-) = \sum_{\substack{r: x_r \in S \\ x_r \geq 0}} x_r p_X(x_r) - \sum_{\substack{r: x_r \in S \\ x_r \leq 0}} (-x_r) p_X(x_r) = \sum_{r: x_r \in S} x_r p_X(x_r),$$

che dimostra la (5.13). Analogamente, se X è assolutamente continua, utilizzando la Definizione 5.2 segue:

$$E(X) = E(X^+) - E(X^-) = \int_0^{+\infty} x f_X(x) dx - \int_{-\infty}^0 (-x) f_X(x) dx = \int_{-\infty}^{+\infty} x f_X(x) dx,$$

ossia la (5.14). \square

Dalla Definizione 5.3 e dalla Proposizione 5.5 scaturisce quanto segue: (i) se X è discreta, $E(X)$ esiste se X è non negativa, o non positiva, oppure se la serie al secondo membro della (5.13) è assolutamente convergente¹; se X è assolutamente continua, $E(X)$ esiste se X è non negativa, o non positiva, oppure se l'integrale al secondo membro della (5.14) è assolutamente convergente².

Esempio 5.7 Sia X una variabile aleatoria discreta di funzione di probabilità

$$p_X(x) = \begin{cases} \frac{1}{2^n}, & x = (-1)^{n+1} \frac{2^n}{n} \quad (n = 1, 2, \dots) \\ 0, & \text{altrimenti.} \end{cases} \quad (5.15)$$

Si noti che X assume con probabilità non nulle valori sia positivi che negativi, ossia i valori $x_n = 2^n/n$ quando $n = 1, 3, 5, \dots$ ed i valori $x_n = -2^n/n$ quando $n = 2, 4, 6, \dots$. Poiché risulta

$$E(X^+) = \sum_{n=0}^{+\infty} x_{2n+1} p_X(x_{2n+1}) = \sum_{n=0}^{+\infty} \frac{2^{2n+1}}{2n+1} 2^{-(2n+1)} = \sum_{n=0}^{+\infty} \frac{1}{2n+1} = +\infty,$$

$$E(X^-) = - \sum_{n=1}^{+\infty} x_{2n} p_X(x_{2n}) = \sum_{n=1}^{+\infty} \frac{2^{2n}}{2n} 2^{-2n} = \sum_{n=0}^{+\infty} \frac{1}{2n} = +\infty,$$

dalla Definizione 5.3 si conclude che il valore medio $E(X)$ non esiste. \diamond

¹La serie $\sum_{r: x_r \in S} x_r p(x_r)$ converge assolutamente se e solo se $\sum_{r: x_r \in S} |x_r| p(x_r)$ è finita.

²L'integrale $\int_{-\infty}^{+\infty} x f(x) dx$ converge assolutamente se e solo se $\int_{-\infty}^{+\infty} |x| f(x) dx$ è finito.

Esempio 5.8 Sia X una variabile aleatoria assolutamente continua di funzione di distribuzione

$$F_X(x) = \begin{cases} -\frac{1}{2x}, & x \leq -1 \\ \frac{1}{2}, & -1 < x \leq 1 \\ 1 - \frac{1}{2x}, & x > 1 \end{cases}$$

e quindi di densità di probabilità (v. Figura 5.4):

$$f_X(x) = \begin{cases} \frac{1}{2x^2}, & x < -1 \text{ oppure } x > 1 \\ 0, & \text{altrimenti.} \end{cases}$$

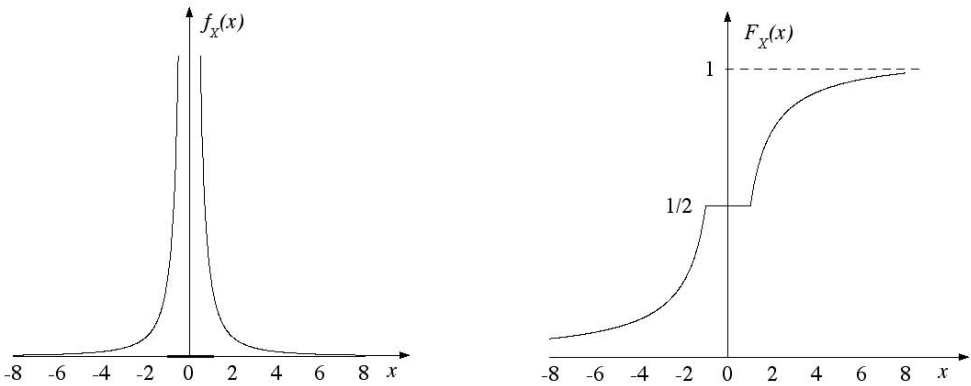


Figura 5.4 – Densità di probabilità e funzione di distribuzione dell'Esempio 5.8.

Poiché risulta

$$\begin{aligned} E(X^+) &= \int_1^{+\infty} x f_X(x) dx = \int_1^{+\infty} x \frac{1}{2x^2} dx = \frac{1}{2} \int_1^{+\infty} \frac{1}{x} dx = +\infty, \\ E(X^-) &= \int_{-\infty}^{-1} (-x) f_X(x) dx = \int_{-\infty}^{-1} (-x) \frac{1}{2x^2} dx = -\frac{1}{2} \int_{-\infty}^{-1} \frac{1}{x} dx = +\infty, \end{aligned}$$

si conclude che X non possiede valore medio. \diamond

Esempio 5.9 Sia $X \sim \mathcal{U}(a, b)$. Facendo uso della (5.14) si ha:

$$E(X) = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \int_a^b x dx = \frac{b^2 - a^2}{b-a} = \frac{a+b}{2}, \quad (5.16)$$

ossia il valore medio è il punto medio dell'intervallo (a, b) . \diamond

Proposizione 5.6 Se X una variabile aleatoria assolutamente continua la cui densità $f_X(x)$ è una funzione pari³, si ha $E(X^+) = E(X^-)$. Pertanto, il valore medio $E(X)$ esiste se e

³Ricordiamo che una funzione $f: \mathbb{R} \rightarrow \mathbb{R}$ si dice pari se per ogni $x \in \mathbb{R}$ risulta $f(x) = f(-x)$.

solo se $E(X^+)$ e $E(X^-)$ sono finiti, ed in tal caso risulta $E(X) = 0$.

Dimostrazione Poiché

$$E(X^-) = \int_{-\infty}^0 (-x) f_X(x) dx = \int_0^{+\infty} y f_X(-y) dy = \int_0^{+\infty} y f_X(y) dy,$$

risulta $E(X^+) = E(X^-)$. Pertanto, dalla (5.12) segue che $E(X)$ esiste se e solo se $E(X^+)$ e $E(X^-)$ sono finiti. In tal caso $E(X) = E(X^+) - E(X^-) = 0$. \square

Esempio 5.10 Sia $X \sim \mathcal{N}(0, 1)$. Poiché $f_X(x) = f_X(-x)$ per ogni $x \in \mathbb{R}$, ed inoltre

$$\begin{aligned} E(X^+) &= \int_0^{+\infty} x f_X(x) dx = \int_0^{+\infty} \frac{x}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} dx \\ &= -\frac{1}{\sqrt{2\pi}} \int_0^{+\infty} \frac{d}{dx} \left[\exp\left\{-\frac{x^2}{2}\right\} \right] dx = \frac{1}{\sqrt{2\pi}} \end{aligned}$$

risulta finito, dalla Proposizione 5.6 segue $E(X) = 0$. \diamond

Esempio 5.11 Sia X una variabile aleatoria assolutamente continua avente densità di probabilità di Cauchy, ossia:

$$f_X(x) = \frac{1}{\pi(1+x^2)} \quad (x \in \mathbb{R}) \quad (5.17)$$

il cui grafico è mostrato in Figura 5.5.

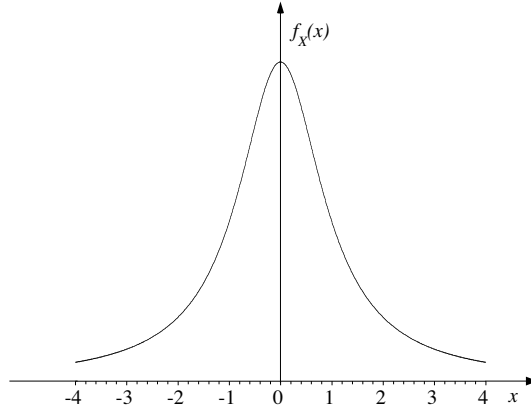


Figura 5.5 – Densità di probabilità di Cauchy.

Poiché $f_X(x)$ è una funzione pari e risulta

$$E(X^+) = \int_0^{+\infty} x f_X(x) dx = \int_0^{+\infty} \frac{x}{\pi(1+x^2)} dx = \frac{1}{2\pi} \int_0^{+\infty} \frac{d}{dx} \ln(1+x^2) dx = +\infty,$$

per la Proposizione 5.6 si conclude che il valore medio $E(X)$ non esiste. \diamond

La definizione che segue fornisce una caratterizzazione del valore medio di una variabile aleatoria in termini della sua funzione di distribuzione, così che risulta possibile unificare le varie definizioni di valore medio ed estendere il concetto di valore medio a variabili aleatorie di qualsiasi tipo. La Definizione 5.3 può essere così riformulata.

Definizione 5.4 Il valore medio di una variabile aleatoria X di funzione di distribuzione $F_X(x)$ è

$$E(X) = \int_0^{+\infty} [1 - F_X(x)] dx - \int_{-\infty}^0 F_X(x) dx, \quad (5.18)$$

purché almeno uno dei due integrali sia finito. Se invece gli integrali sono entrambi divergenti si conviene che la variabile aleatoria X non possiede valore medio.

Questa definizione ha carattere generale in quanto non richiede la specificazione del tipo di variabile aleatoria, ed è particolarmente utile quando si considerano variabili aleatorie che non sono né discrete né assolutamente continue. La (5.18) si presta ad un'interessante interpretazione geometrica. Infatti, graficando la funzione $F_X(x)$ si nota che il valore medio di X è, in definitiva, ottenibile come differenza tra due aree: l'area delimitata dall'asse y , dalla retta $y = 1$ e dalla curva $y = F_X(x)$ nell'intervallo $(0, +\infty)$, e l'area delimitata dall'asse x , dalla curva $y = F_X(x)$ e dall'asse y nell'intervallo $(-\infty, 0)$. Nella Figura 5.6 tali aree sono tratteggiate ed è indicato il segno da attribuire a ciascuna di esse quando se ne effettua la somma algebrica per ottenere $E(X)$.

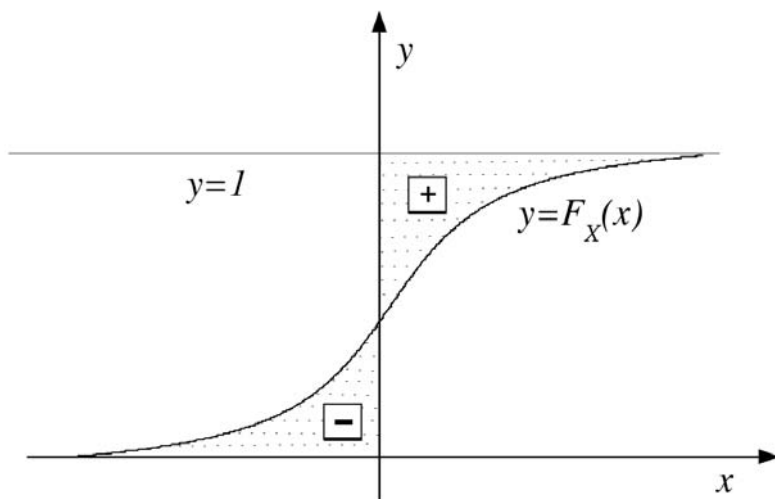


Figura 5.6 – Rappresentazione grafica per il calcolo del valore medio (5.18).

Il valore medio è suscettibile anche di un'interpretazione di carattere meccanico. Infatti nel caso di una variabile aleatoria X assolutamente continua l'asse delle ascisse può riguardarsi come un ideale filo rigido senza peso lungo alcuni dei cui tratti una massa unitaria è distribuita con continuità. Se la densità di massa nel generico punto x del filo coincide con la densità di probabilità $f_X(x)$, il centro di massa del filo è proprio $E(X)$; quindi il filo resta in

equilibrio se viene sospeso nel punto di ascissa $E(X)$. Analogamente, nel caso di una variabile aleatoria X discreta l'asse delle ascisse può ancora riguardarsi come un filo rigido lungo cui la massa unitaria è concentrata nei punti x_i ($i = 1, 2, \dots$) a ciascuno dei quali compete una massa $p_X(x_i)$ scelta in accordo con la funzione di probabilità di X . Anche in questo caso il filo rimane in equilibrio se è sospeso nel punto di ascissa $E(X)$ che, pertanto, assume ancora il significato di centro di massa del filo. Si noti che $E(X)$ può non coincidere con nessuno dei punti x_i .

Proposizione 5.7 *Sia X una variabile aleatoria dotata di valore medio. La (5.18) coincide con la (5.13) se X è discreta e con la (5.14) se X è assolutamente continua.*

Dimostrazione Sia X una variabile aleatoria discreta che assume valori in un insieme S al più numerabile di reali distinti x_1, x_2, \dots con rispettive probabilità $p_X(x_1), p_X(x_2), \dots$. Posto per $r = 1, 2, \dots$ e per ogni $x \in \mathbb{R}$

$$A_r(x) = \begin{cases} 1, & 0 \leq x < x_r \\ 0, & \text{altrimenti,} \end{cases} \quad B_r(x) = \begin{cases} 1, & x_r \leq x \leq 0 \\ 0, & \text{altrimenti,} \end{cases}$$

risulta:

$$\begin{aligned} \int_0^{+\infty} [1 - F_X(x)] dx &= \int_0^{+\infty} \sum_{r: x_r > x} p_X(x_r) dx = \int_0^{+\infty} \sum_{r: x_r \in S} p_X(x_r) A_r(x) dx \\ &= \sum_{r: x_r \in S} p_X(x_r) \int_0^{+\infty} A_r(x) dx = \sum_{\substack{r: x_r \in S \\ x_r \geq 0}} x_r p_X(x_r), \\ \int_{-\infty}^0 F_X(x) dx &= \int_{-\infty}^0 \sum_{r: x_r \leq x} p_X(x_r) dx = \int_{-\infty}^0 \sum_{r: x_r \in S} p_X(x_r) B_r(x) dx \\ &= \sum_{r: x_r \in S} p_X(x_r) \int_{-\infty}^0 B_r(x) dx = \sum_{\substack{r: x_r \in S \\ x_r \leq 0}} (-x_r) p_X(x_r), \end{aligned} \quad (5.19)$$

da cui segue che se X è discreta la (5.13) è equivalente alla (5.18). D'altra parte se X è una variabile aleatoria assolutamente continua con densità di probabilità $f_X(x)$ si ha:

$$\begin{aligned} \int_0^{+\infty} [1 - F_X(x)] dx &= \int_0^{+\infty} dx \int_x^{+\infty} f_X(y) dy = \int_0^{+\infty} dy f_X(y) \int_0^y dx = \int_0^{+\infty} y f_X(y) dy, \\ \int_{-\infty}^0 F_X(x) dx &= \int_{-\infty}^0 dx \int_{-\infty}^x f_X(y) dy = \int_{-\infty}^0 dy f_X(y) \int_y^0 dx = \int_{-\infty}^0 (-y) f_X(y) dy, \end{aligned} \quad (5.20)$$

da cui segue che se X è assolutamente continua la (5.18) è equivalente alla (5.14). \square

Esempio 5.12 Si supponga che X ha la seguente funzione di distribuzione (v. Figura 5.7):

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

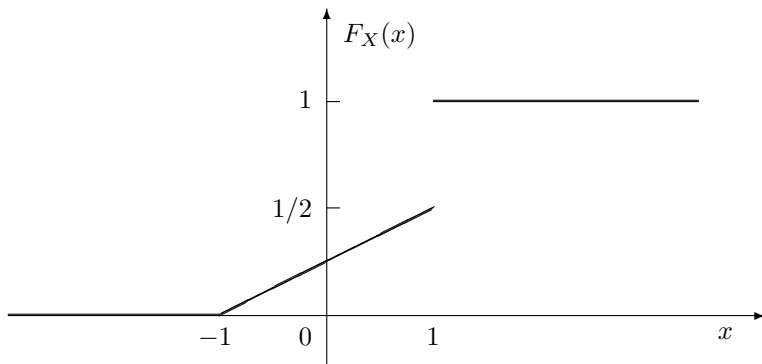


Figura 5.7 – La funzione di distribuzione di cui all'Esempio 5.12.

Si noti che $F_X(x)$ può essere riguardata come combinazione lineare di due funzioni di distribuzione, la prima uniforme in $(-1, 1)$ e l'altra degenere in 1, entrambe pesate con coefficienti $1/2$. Pertanto X non è né discreta né assolutamente continua, essendo $F_X(x)$ discontinua nel punto $x = 1$ con un salto di ampiezza $1/2$. Poiché risulta:

$$\begin{aligned} \int_0^{+\infty} [1 - F_X(x)] dx &= \int_0^1 \frac{3-x}{4} dx = \frac{5}{8}, \\ \int_{-\infty}^0 F_X(x) dx &= \int_{-1}^0 \frac{x+1}{4} dx = \frac{1}{8}, \end{aligned}$$

dalla (5.18) segue che $E(X) = 1/2$. ◇

Proposizione 5.8 *Sia X una variabile aleatoria tale che sia $X \geq 0$ oppure $X \leq 0$. Allora $E(X) = 0$ se e solo se X è degenere in 0.*

Dimostrazione Se X è degenere in 0, il suo valore medio è sicuramente nullo per la Proposizione 5.1. Si assuma ora $X \geq 0$. Dalla (5.18) si ha:

$$E(X) = \int_0^{+\infty} [1 - F_X(x)] dx.$$

Affinché $E(X)$ sia nullo è necessario che tale sia la funzione integranda, il che può verificarsi solo se $F_X(x) = 1$ per ogni punto dell'intervallo $(0, +\infty)$. Essendo $X \geq 0$ per ipotesi, e dovendo essere $F_X(x) = 1$ per ogni $x \in (0, +\infty)$, si conclude che X è una variabile aleatoria degenere in 0. Analogamente, se è $X \leq 0$, dalla (5.18) si ha:

$$E(X) = - \int_{-\infty}^0 F_X(x) dx.$$

Perché sia $E(X) = 0$ è necessario che sia identicamente nulla la funzione integranda, ossia che risulti $F_X(x) = 0$ in ogni punto dell'intervallo $(-\infty, 0)$. Essendo per ipotesi $X \leq 0$ ed avendosi $F_X(x) = 0$ per ogni $x \in (-\infty, 0)$, X deve risultare degenere nel punto 0. □

5.3 Valore medio di una funzione di variabile aleatoria

In molti problemi concreti di probabilità e di statistica oltre che il valore medio di una variabile aleatoria X riveste interesse anche il valore medio di funzioni di X . In particolare, detta X una variabile aleatoria e $g: \mathbb{R} \rightarrow \mathbb{R}$ una funzione Borel-misurabile, si consideri la nuova variabile $Y = g(X)$. Come dimostrato nella Proposizione 3.3, Y è una variabile aleatoria, così che è possibile calcolarne il valore medio, se esistente. Denotando con Y^+ e Y^- rispettivamente le parti positiva e negativa di Y , da quanto esposto nel Paragrafo 5.2 segue che se almeno uno tra $E(Y^+)$ e $E(Y^-)$ è finito, il valore medio di Y esiste e si ha:

$$E(Y) = E[g(X)] = E(Y^+) - E(Y^-) = E\{[g(X)]^+\} - E\{[g(X)]^-\}. \quad (5.21)$$

Se invece $E(Y^+)$ e $E(Y^-)$ sono entrambi infiniti, il valore medio di Y non esiste. Limitandosi ai casi in cui $Y = g(X)$ è discreta o assolutamente continua è possibile mostrare che per calcolare $E(Y)$ non è necessario determinare preliminarmente la distribuzione di probabilità di Y ma è sufficiente la conoscenza della distribuzione di X , come affermato dal teorema che segue.

Teorema 5.1 *Sia X una variabile aleatoria e sia $Y = g(X)$, con $g: \mathbb{R} \rightarrow \mathbb{R}$ funzione Borel-misurabile, tale che $E(Y)$ esista.*

(a) *Se X è discreta ed assume valori in un insieme S al più numerabile di reali distinti x_1, x_2, \dots con rispettive probabilità $p_X(x_1), p_X(x_2), \dots$ si ha:*

$$E[g(X)] = \sum_{r: x_r \in S} g(x_r) p_X(x_r). \quad (5.22)$$

(b) *Se X è assolutamente continua con densità di probabilità $f_X(x)$ risulta:*

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x) dx. \quad (5.23)$$

Dimostrazione Supponiamo per semplicità che il dominio di g si possa rappresentare come unione di intervalli disgiunti I_1, I_2, \dots, I_n e che g sia strettamente monotona in I_j ($j = 1, 2, \dots, n$).

(a) Dal Teorema 3.3 segue che se si denota con h_j la funzione inversa di g in I_j , allora $Y = g(X)$ è una variabile aleatoria discreta. In virtù della (3.21), essa assume valori in un insieme S_Y al più numerabile di reali distinti y_1, y_2, \dots con

$$p_Y(y_k) = P(Y = y_k) = \sum_{j=1}^n P[X = h_j(y_k)],$$

dove il termine j -esimo della somma è nullo se y_k non appartiene al dominio di h_j . Pertanto dalla (5.13) segue che, se esistente, il valore medio $E(Y)$ si può esprimere come

$$E(Y) = \sum_{k: y_k \in S_Y} y_k P(Y = y_k) = \sum_{j=1}^n \sum_{k: y_k \in D_j} y_k P[X = h_j(y_k)],$$

dove D_j denota il dominio di h_j . Da ciò, ricordando che $y_k = g(x_r)$, si ha:

$$\begin{aligned} E(Y) &= \sum_{j=1}^n \sum_{r: x_r \in I_j} g(x_r) P\{X = h_j[g(x_r)]\} = \sum_{j=1}^n \sum_{r: x_r \in I_j} g(x_r) P(X = x_r) \\ &= \sum_{r: x_r \in S} g(x_r) p_X(x_r) \end{aligned}$$

che coincide con la (5.22).

(b) Assumiamo che g sia strettamente monotona in ogni intervallo I_1, I_2, \dots, I_n con derivata prima continua (tranne, al più, negli estremi degli intervalli). Dal Teorema 3.5 segue che se si denota con $h_j(y)$ la funzione inversa di g in I_j e con D_j il suo dominio, $Y = g(X)$ è assolutamente continua. In virtù della (3.28) si ha:

$$f_Y(y) = \sum_{j=1}^n f_X[h_j(y)] \left| \frac{dh_j(y)}{dy} \right|$$

se y appartiene ad almeno uno dei domini D_j ; se invece y è esterno all'unione dei D_j , risulta $f_Y(y) = 0$. Pertanto dalla (3.28) segue che, se esistente, $E(Y)$ può esprimersi come

$$E(Y) = \int_{-\infty}^{+\infty} y f_Y(y) dy = \sum_{j=1}^n \int_{D_j} y f_X[h_j(y)] \left| \frac{dh_j(y)}{dy} \right| dy. \quad (5.24)$$

Effettuiamo il cambiamento di variabile $x = h_j(y)$ ed osserviamo che se g è strettamente crescente in I_j si ha:

$$\int_{D_j} y f_X[h_j(y)] \left| \frac{dh_j(y)}{dy} \right| dy = \int_{D_j} y f_X[h_j(y)] \frac{dh_j(y)}{dy} dy = \int_{I_j} g(x) f_X\{h_j[g(x)]\} dx,$$

mentre se g è strettamente decrescente risulta:

$$\int_{D_j} y f_X[h_j(y)] \left| \frac{dh_j(y)}{dy} \right| dy = - \int_{D_j} y f_X[h_j(y)] \frac{dh_j(y)}{dy} dy = \int_{I_j} g(x) f_X\{h_j[g(x)]\} dx.$$

Dalla (5.24) segue allora immediatamente

$$E(Y) = \sum_{j=1}^n \int_{I_j} g(x) f_X\{h_j[g(x)]\} dx = \int_{-\infty}^{+\infty} g(x) f_X(x) dx,$$

che coincide con la (5.23). □

Esempio 5.13 Sia X una variabile aleatoria assolutamente continua di densità di probabilità

$$f_X(x) = \begin{cases} \frac{3}{2} x^2, & -1 < x < 1 \\ 0, & \text{altrimenti} \end{cases}$$

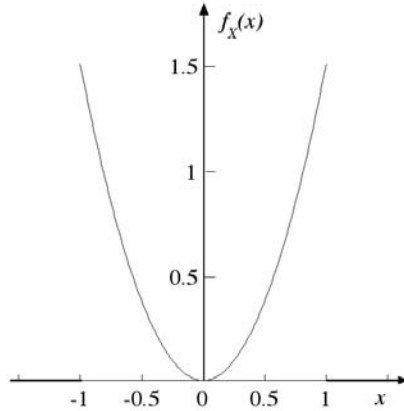


Figura 5.8 – Rappresentazione della densità di probabilità dell'Esempio 5.13.

e sia $Y = \sqrt{|X|}$. Calcoliamo $E(Y)$.

Dalla (5.23), osservando che $f_X(x)$ è simmetrica rispetto all'asse $x = 0$ (v. Figura 5.8), si ha:

$$E(Y) = \int_{-1}^1 \sqrt{|x|} \frac{3}{2} x^2 dx = \frac{3}{2} \left[\int_{-1}^0 x^2 \sqrt{-x} dx + \int_0^1 x^2 \sqrt{x} dx \right] = 3 \int_0^1 x^{5/2} dx = \frac{6}{7}.$$

Mostriamo come lo stesso risultato può essere ottenuto in base alla Proposizione 5.5 calcolando anzitutto la densità di probabilità di Y . A tal fine, osserviamo che la funzione di distribuzione di Y è esprimibile in termini della funzione di distribuzione di X avendosi:

$$F_Y(y) = P(\sqrt{|X|} \leq y) = \begin{cases} 0, & y < 0 \\ P(-y^2 \leq X \leq y^2) = F_X(y^2) - F_X(-y^2), & 0 \leq y < 1 \\ 1, & y \geq 1. \end{cases}$$

Pertanto, poiché $f_X(x)$ è una funzione pari, la densità di probabilità di Y risulta essere la seguente:

$$f_Y(y) = \begin{cases} 2y [f_X(y^2) + f_X(-y^2)] = 4y f_X(y^2) = 6y^5, & 0 < y < 1 \\ 0, & \text{altrimenti.} \end{cases}$$

Dalla (5.14) si ricava così:

$$E(Y) = \int_0^1 y f_Y(y) dy = 6 \int_0^1 y^6 dy = \frac{6}{7}.$$

◇

Esempio 5.14 Sia $X \sim \mathcal{U}(0, \pi)$ e siano $Y = \sin X$ e $Z = \cos X$. Calcoliamo $E(Y)$ e $E(Z)$.

Dalla (5.23), si ricava immediatamente:

$$E(Y) = \int_0^\pi \sin x \frac{1}{\pi} dx = \frac{2}{\pi}, \quad E(Z) = \int_0^\pi \cos x \frac{1}{\pi} dx = 0.$$

Mostriamo come, in base alla Proposizione 5.5, è possibile pervenire allo stesso risultato calcolando anzitutto le funzioni di distribuzione di Y e di Z . Notiamo a tal fine che $F_Y(y) = 0$ per $y < 0$, mentre $F_Y(y) = 1$ per $y \geq 1$; invece, per $0 \leq y < 1$ si ha:

$$\begin{aligned} F_Y(y) &= P(\sin X \leq y) = P(0 \leq X \leq \arcsin y) + P(\pi - \arcsin y \leq X \leq \pi) \\ &= \frac{1}{\pi} \left[\int_0^{\arcsin y} dx + \int_{\pi - \arcsin y}^{\pi} dx \right] = \frac{2 \arcsin y}{\pi}. \end{aligned}$$

In conclusione,

$$F_Y(y) = \begin{cases} 0, & y < 0 \\ \frac{2 \arcsin y}{\pi}, & 0 \leq y < 1 \\ 1, & y \geq 1, \end{cases} \quad (5.25)$$

da cui segue

$$f_Y(y) = \begin{cases} \frac{2}{\pi \sqrt{1-y^2}}, & 0 \leq y < 1 \\ 0, & \text{altrimenti,} \end{cases} \quad (5.26)$$

e quindi:

$$E(Y) = \frac{2}{\pi} \int_0^1 \frac{y}{\sqrt{1-y^2}} dy = -\frac{2}{\pi} \int_0^1 \left[\frac{d}{dy} \sqrt{1-y^2} \right] dy = \frac{2}{\pi}.$$

Procedendo in modo analogo, $F_Z(z) = 0$ se $z < -1$, mentre $F_Z(z) = 1$ se $z \geq 1$; invece, se $-1 \leq z < 1$ si ha:

$$F_Z(z) = P(\cos X \leq z) = P(\arccos z \leq X \leq \pi) = \frac{1}{\pi} \int_{\arccos z}^{\pi} dz = \frac{\pi - \arccos z}{\pi}.$$

Si ottiene quindi:

$$F_Z(z) = \begin{cases} 0, & z < -1 \\ \frac{\pi - \arccos z}{\pi}, & -1 \leq z < 1 \\ 1, & z \geq 1 \end{cases} \quad (5.27)$$

da cui segue:

$$f_Z(z) = \begin{cases} \frac{1}{\pi \sqrt{1-z^2}}, & -1 \leq z < 1 \\ 0, & \text{altrimenti,} \end{cases} \quad (5.28)$$

che comporta:

$$E(Z) = \frac{1}{\pi} \int_{-1}^1 \frac{z}{\sqrt{1-z^2}} dz = -\frac{1}{\pi} \int_{-1}^1 \left[\frac{d}{dz} \sqrt{1-z^2} \right] dz = 0.$$

Si noti che se si è interessati soltanto al calcolo dei valori medi, il primo metodo risulta evidentemente più rapido ed elegante.

È qui utile notare che le (5.26) e (5.28) possono anche ottenersi direttamente, senza passare attraverso le funzioni di distribuzione (5.25) e (5.27), utilizzando le relazioni (3.22) e (3.28) e tenendo conto della non monotonicità di $\sin x$ nell'intervallo $(0, \pi)$ e della monotonicità di $\cos x$ nello stesso intervallo. \diamond

Teorema 5.2 *Sia X una variabile aleatoria, $g: \mathbb{R} \rightarrow \mathbb{R}$ una funzione Borel-misurabile e a, b costanti reali. Se esiste $E[g(X)]$, si ha:*

$$E[a g(X) + b] = a E[g(X)] + b. \quad (5.29)$$

Dimostrazione Dal Teorema 5.1 segue che se X è discreta ed assume valori in un insieme S al più numerabile di reali distinti x_1, x_2, \dots con rispettive probabilità $p_X(x_1), p_X(x_2), \dots$ si ha:

$$E[a g(X) + b] = \sum_{r: x_r \in S} [a g(x_r) + b] p_X(x_r) = a \sum_{r: x_r \in S} g(x_r) p_X(x_r) + b = a E[g(X)] + b,$$

mentre se X è assolutamente continua di densità $f_X(x)$ risulta:

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

Ciò completa la dimostrazione. \square

Il Teorema 5.2 mostra che il valore medio di una variabile aleatoria è un operatore lineare.

Teorema 5.3 *Sia X una variabile aleatoria e siano $g_k: \mathbb{R} \rightarrow \mathbb{R}$ ($k = 1, 2, \dots, n$) funzioni Borel-misurabili tutte dotate di valori medi $E[g_k(X)]$ ($k = 1, 2, \dots, n$). Si ha:*

$$E\left[\sum_{k=1}^n g_k(X)\right] = \sum_{k=1}^n E[g_k(X)] \quad (5.30)$$

sempre che a secondo membro non siano presenti simultaneamente nella somma infiniti positivi e negativi.

Dimostrazione Dal Teorema 5.1 segue che se X è discreta ed assume valori in un insieme S al più numerabile di reali distinti x_1, x_2, \dots con rispettive probabilità $p_X(x_1), p_X(x_2), \dots$ risulta:

$$E\left[\sum_{k=1}^n g_k(X)\right] = \sum_{r: x_r \in S} \left[\sum_{k=1}^n g_k(x_r)\right] p_X(x_r) = \sum_{k=1}^n \left[\sum_{r: x_r \in S} g_k(x_r) p_X(x_r)\right] = \sum_{k=1}^n E[g_k(X)],$$

mentre se X è assolutamente continua di densità $f_X(x)$ si ha:

$$E\left[\sum_{k=1}^n g_k(X)\right] = \int_{-\infty}^{+\infty} \left[\sum_{k=1}^n g_k(x)\right] f_X(x) dx = \sum_{k=1}^n \left[\int_{-\infty}^{+\infty} g_k(x) f_X(x) dx\right] = \sum_{k=1}^n E[g_k(X)].$$

Ciò completa la dimostrazione. \square

Esempio 5.15 Sia X una variabile aleatoria discreta di funzione di probabilità

$$p_X(x) = \begin{cases} 1/3, & x = -2 \\ 1/6, & x = 1 \\ 1/2, & x = 2 \\ 0, & \text{altrimenti} \end{cases}$$

e sia $Y = X + X^2$. Calcoliamo $E(Y)$.

Dalla (5.30) si ha $E(Y) = E(X) + E(X^2)$. Utilizzando la (5.22) segue:

$$E(X) = (-2) \frac{1}{3} + 1 \frac{1}{6} + 2 \frac{1}{2} = \frac{1}{2}, \quad E(X^2) = 4 \frac{1}{3} + 1 \frac{1}{6} + 4 \frac{1}{2} = \frac{7}{2},$$

da cui risulta $E(Y) = 4$. È possibile giungere allo stesso risultato in base alla Proposizione 5.5, il che richiede la conoscenza della funzione di probabilità di Y che può essere calcolata ragionando come segue. Poiché X assume con probabilità non nulla i valori $-2, 1, 2$, la variabile $Y = X + X^2$ assumerà con probabilità non nulla il valore 2 (in corrispondenza di $x = -2$ e $x = 1$) ed il valore 6 (in corrispondenza di $x = 2$); pertanto risulta:

$$p_Y(y) = \begin{cases} 1/2, & y = 2 \\ 1/2, & y = 6 \\ 0, & \text{altrimenti.} \end{cases}$$

Dalla (5.13) segue allora:

$$E(Y) = 2 \frac{1}{2} + 6 \frac{1}{2} = 4,$$

che coincide con il risultato precedentemente ottenuto. \diamond

Concludiamo questo paragrafo osservando che se il valore medio di una variabile aleatoria assolutamente continua di densità di probabilità $f_X(x)$ fosse stato definito come

$$E(X) = \lim_{a \rightarrow +\infty} \int_{-a}^a x f_X(x) dx \quad (5.31)$$

con a reale positivo, allora non sarebbe stato necessariamente vero che

$$E(X + c) = E(X) + c \quad (5.32)$$

per ogni costante reale c . Invero, assumiamo che X sia assolutamente continua di densità $f_X(x)$ pari. Dalla (5.31) si ha $E(X) = 0$ visto che $\int_{-a}^a x f_X(x) dx = 0$ per ogni a reale positivo. D'altra parte, ponendo $Y = X + c$, con c costante reale positiva, dal Teorema 3.4 segue $f_Y(y) = f_X(y - c)$. Quindi, facendo uso della (5.31), risulta:

$$E(Y) = E(X + c) = \lim_{a \rightarrow +\infty} \int_{-a}^a y f_X(y - c) dy. \quad (5.33)$$

Sia ad esempio $c > 0$. Ricorrendo al cambiamento di variabile $u = y - c$, si ottiene:

$$\begin{aligned} \int_{-a}^a y f_X(y - c) dy &= \int_{-a-c}^{a-c} (u + c) f_X(u) du \\ &= \int_{-a-c}^{a+c} u f_X(u) du - \int_{a-c}^{a+c} u f_X(u) du + c \int_{-a-c}^{a-c} f_X(u) du. \end{aligned} \quad (5.34)$$

Per $a \rightarrow +\infty$ il primo integrale nell'ultima uguaglianza si annulla, mentre il terzo integrale tende ad 1. Quindi se al divergere di a anche il secondo integrale si annullasse per ogni densità di probabilità simmetrica intorno a $x = 0$ e per ogni costante c , allora $E(Y) = E(X + c)$ sarebbe uguale a $E(X) + c = c$. Pertanto, l'individuare una densità di probabilità simmetrica intorno a $x = 0$ ed una costante $c > 0$ tali che $\int_{a-c}^{a+c} u f_X(u) du$ non si annulli al divergere di a è sufficiente a dimostrare che la (5.31) è in contrasto con la (5.32) in quanto si avrebbe $E(Y) \neq E(X) + c$. Un esempio di una densità di probabilità che soddisfa tali requisiti è il seguente (v. Figura 5.9):

$$f_X(x) = \begin{cases} \frac{3}{\pi^2} \frac{1}{k^2} (1 - k^2 + |x|), & -1 \leq |x| - k^2 \leq 0, \quad k = 1, 2, \dots \\ \frac{3}{\pi^2} \frac{1}{k^2} (1 + k^2 - |x|), & 0 \leq |x| - k^2 \leq 1, \quad k = 1, 2, \dots \\ 0, & \text{altrimenti.} \end{cases} \quad (5.35)$$

In altre parole, $f_X(x)$ si annulla in tutti i punti x tranne che in quelli che distano meno di una unità da un punto che in valore assoluto è un quadrato perfetto. È evidente che $f_X(x)$ è una funzione simmetrica intorno a $x = 0$; inoltre è possibile mostrare che essa è una densità di probabilità osservando che risulta

$$\begin{aligned} \int_{-\infty}^{+\infty} f_X(x) dx &= \frac{6}{\pi^2} \sum_{k=1}^{+\infty} \frac{1}{k^2} \left[\int_{k^2-1}^{k^2} (1 - k^2 + x) dx + \int_{k^2}^{k^2+1} (1 + k^2 - x) dx \right] \\ &= \frac{12}{\pi^2} \sum_{k=1}^{+\infty} \frac{1}{k^2} \int_0^1 y dy = \frac{6}{\pi^2} \sum_{k=1}^{+\infty} \frac{1}{k^2} = \frac{6}{\pi^2} \frac{\pi^2}{6} = 1. \end{aligned}$$

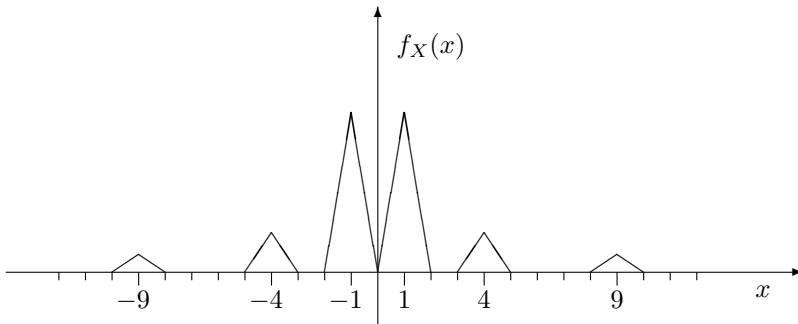


Figura 5.9 – La densità di probabilità (5.35).

Infine, per convincersi che qualunque sia $c > 0$ il secondo integrale che compare al secondo membro della (5.34) non si annulla al divergere di a , si noti che scegliendo $c = 1$ per $k = 2, 3, \dots$ si ha:

$$\int_{k^2-1}^{k^2+1} u f_X(u) du \geq (k^2 - 1) \int_{k^2-1}^{k^2+1} f_X(u) du = \frac{3}{\pi^2} \frac{k^2 - 1}{k^2} \geq \frac{9}{4\pi^2}.$$

5.4 Momenti di una variabile aleatoria

Tra i valori medi di funzioni di una variabile aleatoria ruolo particolarmente importante rivestono quelle ottenute scegliendo come funzione $g: \mathbb{R} \rightarrow \mathbb{R}$ le seguenti: $g(X) = X^n$ e $g(X) = [X - E(X)]^n$ ($n = 0, 1, \dots$).

Definizione 5.5 Il momento di ordine n ($n = 0, 1, \dots$) di una variabile aleatoria X è definito da

$$\mu_n = E(X^n), \quad (5.36)$$

sempre che il valore medio al secondo membro della (5.36) esista. Inoltre, se $E(X)$ esiste, il momento centrale di ordine n di X ($n = 0, 1, \dots$) è definito al seguente modo

$$\bar{\mu}_n = E[(X - E(X))^n], \quad (5.37)$$

sempre che il valore medio al secondo membro della (5.37) esista.

La seguente proposizione è di immediata dimostrazione.

Proposizione 5.9 Per ogni variabile aleatoria X risulta $\mu_0 = 1$ e, se $E(X)$ esiste, $\bar{\mu}_1 = 0$.

Dimostrazione Infatti μ_0 per definizione è il valore medio della variabile aleatoria X^0 che risulta degenerare nel punto 1; pertanto dalla Proposizione 5.1, scegliendo $c = 1$, segue $\mu_0 = E(1) = 1$. Inoltre, dal Teorema 5.2, scegliendo $a = 1$ e $b = E(X)$, si ricava $\bar{\mu}_1 = E[X - E(X)] = E(X) - E(X) = 0$. \square

Dalla Definizione 5.5 e dal Teorema 5.1 scaturisce quindi che se X è una variabile aleatoria discreta $\{x_r, p_X(x_r); r = 1, 2, \dots\}$ risulta:

$$\mu_n = \sum_{r: x_r \in S} x_r^n p_X(x_r), \quad \bar{\mu}_n = \sum_{r: x_r \in S} (x_r - \mu_1)^n p_X(x_r) \quad (n = 0, 1, \dots) \quad (5.38)$$

mentre se X è una variabile aleatoria assolutamente continua di densità $f_X(x)$ si ha:

$$\mu_n = \int_{-\infty}^{+\infty} x^n f_X(x) dx, \quad \bar{\mu}_n = \int_{-\infty}^{+\infty} (x - \mu_1)^n f_X(x) dx \quad (n = 0, 1, \dots). \quad (5.39)$$

Teorema 5.4 Siano a e b costanti reali e sia X una variabile aleatoria per la quale esistono i momenti $E(X^i)$ ($i = 1, 2, \dots, n$). Risulta:

$$E[(aX + b)^n] = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k} E(X^k), \quad (5.40)$$

sempre che nella somma al secondo membro non siano presenti simultaneamente infiniti positivi e negativi.

Dimostrazione Sia $g(X) = (aX + b)^n$. Usando la formula del binomio di Newton, se X è discreta $\{x_r, p_X(x_r); r = 1, 2, \dots\}$ risulta:

$$\begin{aligned} E[(aX + b)^n] &= \sum_{r: x_r \in S} (ax_r + b)^n p_X(x_r) = \sum_{r: x_r \in S} \left[\sum_{k=0}^n \binom{n}{k} (ax_r)^k b^{n-k} \right] p_X(x_r) \\ &= \sum_{k=0}^n \binom{n}{k} a^k b^{n-k} \sum_{r: x_r \in S} x_r^k p_X(x_r) = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k} E(X^k), \end{aligned}$$

dove nell'ultima uguaglianza si è fatto uso della (5.38). Analogamente, se X è assolutamente continua di densità $f_X(x)$ si ha:

$$\begin{aligned} E[(aX + b)^n] &= \int_{-\infty}^{+\infty} (ax + b)^n f_X(x) dx = \int_{-\infty}^{+\infty} \left[\sum_{k=0}^n \binom{n}{k} (ax)^k b^{n-k} \right] f_X(x) dx \\ &= \sum_{k=0}^n \binom{n}{k} a^k b^{n-k} \int_{-\infty}^{+\infty} x^k f_X(x) dx = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k} E(X^k) \end{aligned}$$

dove l'ultima uguaglianza segue in virtù della (5.39). \square

Il Teorema 5.4 permette di determinare una semplice relazione tra i momenti ed i momenti centrali. Infatti dalla (5.40), per $a = 1$ e $b = -E(X) = -\mu_1$, si ha:

$$\bar{\mu}_n = E[(X - \mu_1)^n] = \sum_{k=0}^n \binom{n}{k} (-\mu_1)^{n-k} E(X^k) = \sum_{k=0}^n \binom{n}{k} (-\mu_1)^{n-k} \mu_k. \quad (5.41)$$

Un'estensione della Definizione 5.4 fornisce il modo per calcolare i momenti di una variabile aleatoria in termini della funzione di distribuzione.

Teorema 5.5 Sia X una variabile aleatoria di funzione di distribuzione $F_X(x)$. Se $E(X^n)$ esiste, risulta:

$$E(X^n) = \int_0^{+\infty} n x^{n-1} [1 - F_X(x)] dx - \int_{-\infty}^0 n x^{n-1} F_X(x) dx \quad (n = 2, 3, \dots). \quad (5.42)$$

Dimostrazione Sia $Y = X^n$. Dalla (5.18) risulta:

$$E(Y) = \int_0^{+\infty} [1 - F_Y(y)] dy - \int_{-\infty}^0 F_Y(y) dy; \quad (5.43)$$

è quindi necessario calcolare la funzione di distribuzione di Y . A tal fine occorre distinguere due casi: (a) n dispari e (b) n pari.

(a) Se $n = 3, 5, \dots$ risulta:

$$F_Y(y) = P(X^n \leq y) = P(X \leq \sqrt[n]{y}) = F_X(\sqrt[n]{y}),$$

e pertanto dalla (5.43), effettuando il cambiamento di variabile di integrazione $y = z^n$, segue:

$$\begin{aligned} E(Y) &= \int_0^{+\infty} [1 - F_X(\sqrt[n]{y})] dy - \int_{-\infty}^0 F_X(\sqrt[n]{y}) dy \\ &= \int_0^{+\infty} [1 - F_X(z)] n z^{n-1} dz - \int_{-\infty}^0 F_X(z) n z^{n-1} dz, \end{aligned}$$

ossia la (5.42).

(b) Se $n = 2, 4, \dots$ risulta $F_Y(y) = 0$ se $y < 0$, così che il secondo integrale che compare nella (5.43) è nullo. Se è invece $y \geq 0$, si ha:

$$F_Y(y) = P(X^n \leq y) = P(-\sqrt[n]{y} \leq X \leq \sqrt[n]{y}).$$

Pertanto dalla (5.43), effettuando il cambiamento di variabile di integrazione $y = z^n$, segue:

$$\begin{aligned} E(Y) &= \int_0^{+\infty} [P(X < -\sqrt[n]{y}) + P(X > \sqrt[n]{y})] dy \\ &= \int_0^{+\infty} [P(X < -z) + P(X > z)] n z^{n-1} dz \\ &= \int_0^{+\infty} [1 - F_X(z)] n z^{n-1} dz + \int_0^{+\infty} P(X < -z) n z^{n-1} dz. \quad (5.44) \end{aligned}$$

Effettuando un cambiamento della variabile di integrazione ed un passaggio al limite sotto il segno di integrale, per n pari si ha:

$$\begin{aligned} \int_0^{+\infty} P(X < -z) n z^{n-1} dz &= \lim_{\varepsilon \rightarrow 0} \int_0^{+\infty} F_X(-z - \varepsilon) n z^{n-1} dz \\ &= \lim_{\varepsilon \rightarrow 0} \int_{-\infty}^{-\varepsilon} F_X(y) n (-y - \varepsilon)^{n-1} dy = - \lim_{\varepsilon \rightarrow 0} \int_{-\infty}^{-\varepsilon} F_X(y) n (y + \varepsilon)^{n-1} dy \\ &= - \int_{-\infty}^0 F_X(y) n y^{n-1} dy, \end{aligned}$$

così che dalla (5.44) segue immediatamente la (5.42). \square

Proposizione 5.10 *Se X è una variabile aleatoria con distribuzione simmetrica rispetto all'asse $x = 0$, ossia tale che $P(X \leq x) = P(X \geq -x)$ per ogni $x \in \mathbb{R}$, i suoi momenti di ordine dispari, se esistenti, sono tutti nulli.*

Dimostrazione Osserviamo in primo luogo che X è una variabile aleatoria avente distribuzione simmetrica rispetto all'asse $x = 0$ se e solo se $Y = -X$ ha la stessa distribuzione di X . Inoltre, se n è dispari risulta:

$$P(X^n \leq y) = P(X \leq \sqrt[n]{y}) = P(X \geq -\sqrt[n]{y}) = P(X^n \geq -y),$$

implicante che se X ha distribuzione simmetrica rispetto all'asse $x = 0$ ciò sussiste anche per X^n con n dispari. Pertanto, se n è dispari, X^n e $Y = -X^n$ hanno la stessa funzione di distribuzione e quindi lo stesso valore medio. Per n dispari si ha dunque $E(X^n) = E(-X^n) = -E(X^n)$, così che $E(X^n) = 0$. \square

Si osservi che condizione necessaria e sufficiente affinché una variabile aleatoria discreta X abbia distribuzione simmetrica rispetto all'asse $x = 0$ è che sia $p_X(x) = p_X(-x)$ per ogni $x \in \mathbb{R}$, ossia che la funzione di probabilità di X sia pari; analogamente se X è assolutamente continua, condizione necessaria e sufficiente affinché abbia distribuzione simmetrica rispetto all'asse $x = 0$ è che $f_X(x)$ sia una funzione pari. In entrambi i casi, dalla Proposizione 5.10, si deduce che i momenti di ordine dispari, se esistenti, sono nulli.

Esempio 5.16 Sia X una variabile aleatoria discreta di funzione di probabilità

$$p_X(x) = \begin{cases} \frac{1}{2} p (1-p)^{|x|-1}, & x = \pm 1, \pm 2, \dots \\ 0, & \text{altrimenti,} \end{cases} \quad (5.45)$$

con $0 < p < 1$. Poiché X ha distribuzione simmetrica rispetto all'asse $x = 0$, dalla Proposizione 5.10 segue che i suoi momenti di ordine dispari sono nulli. \diamond

Esempio 5.17 Sia X una variabile aleatoria assolutamente continua di densità di probabilità

$$f_X(x) = \frac{\lambda}{2} \exp\{-\lambda|x|\} \quad (x \in \mathbb{R}),$$

con $\lambda > 0$. Poiché X ha distribuzione simmetrica rispetto all'asse $x = 0$, in virtù della Proposizione 5.10 si ricava che i suoi momenti di ordine dispari sono nulli. \diamond

I momenti ed i momenti centrali sono molto importanti in quanto costituiscono delle caratteristiche numeriche sia della variabile aleatoria sia della sua distribuzione di probabilità. Il valore medio sintetizza solo un aspetto della variabile aleatoria; è infatti immediato costruire variabili aleatorie che, pur avendo lo stesso valore medio, sono tra loro profondamente diverse. È desiderabile quindi introdurre accanto al valore medio ulteriori caratteristiche numeriche; tra queste particolare interesse rivestono quelle che permettono di individuare delle *misure di dispersione* della variabile aleatoria rispetto al suo valore medio. La misura di dispersione più comune è legata alla cosiddetta *varianza*.

Definizione 5.6 La varianza σ_X^2 , o $\text{Var}(X)$, di una variabile aleatoria X è definita al seguente modo:

$$\sigma_X^2 = \text{Var}(X) = E[(X - E(X))^2], \quad (5.46)$$

sempre che il valore medio di X esista e sia finito.

Quindi $\text{Var}(X)$ è il momento centrale di ordine 2 di X .

Proposizione 5.11 Se X è una variabile aleatoria dotata di valore medio finito, allora $\text{Var}(X)$ esiste (finita o infinita) ed è non negativa. Inoltre, $\text{Var}(X) = 0$ se e solo se X è degenera nel punto $E(X)$.

Dimostrazione $\text{Var}(X)$ è il valore medio di una variabile aleatoria non negativa. Pertanto se $E(X)$ è finito, dalla Proposizione 5.3 segue che $\text{Var}(X)$ esiste sempre e risulta $\text{Var}(X) \geq 0$. Inoltre, ricordando la (5.46), dalla Proposizione 5.8 si ha $\text{Var}(X) = 0$ se e solo se X è una variabile aleatoria degenera nel punto $E(X)$. \square

Proposizione 5.12 *Sia X una variabile aleatoria dotata di valore medio finito. Risulta:*

$$\text{Var}(X) = E(X^2) - [E(X)]^2. \quad (5.47)$$

Dimostrazione Dalla (5.41) si ha

$$\begin{aligned} \text{Var}(X) &= \sum_{k=0}^2 \binom{2}{k} [-E(X)]^{2-k} E(X^k) = E(X^2) - 2 E(X) E(X) + [E(X)]^2 \\ &= E(X^2) - [E(X)]^2 \end{aligned}$$

ossia la (5.47). □

Dalle Proposizioni 5.11 e 5.12 segue una tanto immediata quanto utile disuguaglianza sussistente tra il momento del secondo ordine ed il quadrato del valore medio, come indicato dalla seguente proposizione.

Proposizione 5.13 *Sia X una variabile aleatoria dotata di valore medio finito. Risulta*

$$E(X^2) \geq [E(X)]^2. \quad (5.48)$$

Dimostrazione Dalla (5.47) e dalla Proposizione 5.11 si ha:

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

che conduce banalmente alla (5.48). □

Esempio 5.18 Con riferimento all'Esempio 2.10 prendiamo in esame la variabile aleatoria discreta X che descrive il numero di concordanze che si verificano nell'esperimento consistente nell'estrarre n biglie, numerate da 1 a n , l'una dopo l'altra fino ad esaurimento. Vogliamo calcolare il valore medio e la varianza di X .

Ricordiamo anzitutto che la funzione di probabilità di X , determinata in (2.14), è

$$p_X(r) = \frac{1}{r!} \sum_{i=0}^{n-r} (-1)^i \frac{1}{i!} \quad (r = 0, 1, \dots, n). \quad (5.49)$$

Essendo X discreta, dalla (5.49) segue:

$$E(X) = \sum_{r=0}^n r p_X(r) = \sum_{r=0}^n r \frac{1}{r!} \sum_{i=0}^{n-r} \frac{(-1)^i}{i!} = \sum_{r=1}^n \frac{1}{(r-1)!} \sum_{j=r}^n \frac{(-1)^{j-r}}{(j-r)!}, \quad (5.50)$$

dove l'ultima uguaglianza si è ricavata ponendo $j = i + r$. Invertendo l'ordine delle somme nell'ultimo termine della (5.50) e facendo ricorso alla formula del binomio di Newton si ottiene:

$$E(X) = \sum_{j=1}^n \frac{1}{(j-1)!} \sum_{r=1}^j \binom{j-1}{r-1} (-1)^{j-r},$$

da cui, ponendo $j - 1 = k$ e $r - 1 = m$, si trae:

$$E(X) = \sum_{k=0}^{n-1} \frac{1}{k!} \sum_{m=0}^k \binom{k}{m} (-1)^{k-m} = \sum_{k=0}^{n-1} \frac{1}{k!} (1-1)^k = 1.$$

Pertanto, qualsiasi sia il numero n di biglie contenute nell'urna, il numero medio di concordanze è unitario.⁴ Per ricavare la varianza distinguiamo il caso in cui $n = 1$ dai casi in cui $n = 2, 3, \dots$. Se l'urna contiene una sola biglia, è evidente che $\text{Var}(X) = 0$, essendo X degenerare in 1. Se invece $n = 2, 3, \dots$ procediamo in modo analogo a quanto appena visto per il valore medio, osservando che risulta:

$$\begin{aligned} E[X(X-1)] &= \sum_{r=0}^n r(r-1) p_X(r) = \sum_{r=0}^n r(r-1) \frac{1}{r!} \sum_{i=0}^{n-r} \frac{(-1)^i}{i!} \\ &= \sum_{r=2}^n \frac{1}{(r-2)!} \sum_{j=r}^n \frac{(-1)^{j-r}}{(j-r)!} = \sum_{j=2}^n \frac{1}{(j-2)!} \sum_{r=2}^j \binom{j-2}{r-2} (-1)^{j-r} \\ &= \sum_{k=0}^{n-2} \frac{1}{k!} \sum_{m=0}^k \binom{k}{m} (-1)^{k-m} = \sum_{k=0}^{n-2} \frac{1}{k!} (1-1)^k = 1. \end{aligned}$$

Pertanto, ricordando che $E(X) = 1$, si ricava:

$$\text{Var}(X) = E[X(X-1)] + E(X) - [E(X)]^2 = 1.$$

Quindi, qualsiasi sia il numero $n = 2, 3, \dots$ di biglie contenute nell'urna, la media e la varianza del numero di concordanze sono entrambe unitarie. \diamond

Nel Teorema 5.2 è stato mostrato che il valore medio di una variabile aleatoria è un operatore lineare; ciò, come mostra il seguente teorema, non vale per la varianza.

Teorema 5.6 *Sia X una variabile aleatoria e sia $g: \mathbb{R} \rightarrow \mathbb{R}$ una funzione Borel-misurabile; inoltre siano a, b costanti reali. Se $E[g(X)]$ esiste finito, si ha:*

$$\text{Var}[a g(X) + b] = a^2 \text{Var}[g(X)]. \quad (5.51)$$

Dimostrazione Ponendo $Y = g(X)$, per la proprietà di linearità del valore medio risulta:

$$\begin{aligned} \text{Var}[a g(X) + b] &= E\{[aY + b - E(aY + b)]^2\} = E\{[aY + b - a E(Y) - b]^2\} \\ &= E\{a^2 [Y - E(Y)]^2\} = a^2 E[(Y - E(Y))^2] = a^2 \text{Var}(Y) = \text{Var}[g(X)], \end{aligned}$$

ossia la (5.51). \square

Immediata conseguenza dei Teoremi 5.2 e 5.6 è la seguente

⁴È questo un esempio paradigmatico che mostra come l'intuizione possa alle volte essere fuorviante

Proposizione 5.14 Siano a, b costanti positive e sia X una variabile aleatoria dotata di valore medio finito. Risulta:

$$E(aX + b) = aE(X) + b, \quad \text{Var}(aX + b) = a^2 \text{Var}(X). \quad (5.52)$$

Dimostrazione Il risultato segue immediatamente dalle (5.29) e (5.51) scegliendo per g la funzione identica. \square

Dalla Proposizione 5.14 appare evidente che il valore medio di una variabile aleatoria conserva le stesse dimensioni della variabile aleatoria mentre la varianza ha dimensioni quadratiche rispetto a quelle della variabile aleatoria. Per ricondursi alle stesse dimensioni si considera la radice quadrata con segno positivo della varianza che viene chiamata *deviazione standard*.

Definizione 5.7 La deviazione standard σ_X di una variabile aleatoria X dotata di valore medio finito è:

$$\sigma_X = \sqrt{\text{Var}(X)} = \sqrt{E[(X - E(X))^2]}. \quad (5.53)$$

Definizione 5.8 Il coefficiente di variazione di una variabile aleatoria X , dotata di varianza finita e di valore medio finito e non nullo, è il rapporto

$$C_X = \frac{\sqrt{\text{Var}(X)}}{E(X)}. \quad (5.54)$$

Il coefficiente di variazione, a differenza del valore medio e della deviazione standard, è una quantità sempre adimensionale.

Per confrontare variabili aleatorie differenti caratterizzate da valori medi diversi oppure da differenti varianze è conveniente ricorrere alle cosiddette variabili standardizzate, come indicato nella definizione che segue.

Definizione 5.9 Sia X una variabile aleatoria dotata di valore medio finito e di varianza finita e non nulla. La variabile aleatoria

$$Y = \frac{X - E(X)}{\sqrt{\text{Var}(X)}} \quad (5.55)$$

è detta *standardizzata* di X .

I valori che assume una variabile aleatoria standardizzata sono adimensionali.

Proposizione 5.15 Una variabile aleatoria standardizzata è caratterizzata da valore medio nullo e varianza unitaria.

Dimostrazione Facendo uso della (5.52), dalla Definizione 5.9 segue:

$$\begin{aligned} E(Y) &= E\left[\frac{X - E(X)}{\sigma_X}\right] = \frac{1}{\sigma_X} [E(X) - E(X)] = 0, \\ \text{Var}(Y) &= \text{Var}\left[\frac{X - E(X)}{\sigma_X}\right] = \frac{1}{\sigma_X^2} \text{Var}(X) = 1. \end{aligned}$$

Ciò completa la dimostrazione. \square

Proposizione 5.16 *Sia X una variabile aleatoria dotata di valore medio finito e avente distribuzione simmetrica rispetto all'asse $x = E(X)$, ossia tale che $P\{X \leq E(X) + x\} = P\{X \geq E(X) - x\}$ per ogni $x \in \mathbb{R}$. Allora i momenti centrali di ordine dispari, se esistenti, sono tutti nulli.*

Dimostrazione Consideriamo la variabile aleatoria $Y = X - E(X)$; poich  per ogni $y \in \mathbb{R}$ risulta

$$\begin{aligned} P(Y \leq y) &= P\{X - E(X) \leq y\} = P\{X \leq E(X) + y\} = P\{X \geq E(X) - y\} \\ &= P\{X - E(X) \geq x\} = P(Y \geq y), \end{aligned}$$

si trae che Y ha distribuzione simmetrica intorno all'asse $y = 0$. Dalla Proposizione 5.10 si deduce quindi che i momenti di ordine dispari di Y , che coincidono con i momenti centrali di ordine dispari di X , se esistenti, sono nulli. \square

Si osservi che, sulla base della Proposizione 5.16, condizione necessaria e sufficiente affinch  una variabile aleatoria discreta X sia a distribuzione simmetrica rispetto all'asse $x = E(X) = \mu_1$   che risulti $p_X(\mu_1 + x) = p_X(\mu_1 - x)$ per ogni $x \in \mathbb{R}$, ossia che la funzione di probabilit  di X sia simmetrica intorno a $x = \mu_1$; analogamente, se X   assolutamente continua condizione necessaria e sufficiente affinch  sia a distribuzione simmetrica rispetto all'asse $x = E(X) = \mu_1$   che si abbia $f_X(\mu_1 + x) = f_X(\mu_1 - x)$ per ogni $x \in \mathbb{R}$, ossia che la densit  di probabilit  di X sia simmetrica intorno a $x = \mu_1$. In entrambi i casi, dalla Proposizione 5.16, si deduce che i momenti centrali di ordine dispari, se esistenti, sono nulli.

Cos  come valore medio e varianza descrivono alcune caratteristiche della distribuzione di una variabile aleatoria, i momenti successivi ne descrivono altre. Nel seguito ci limiteremo ad analizzare i momenti del terzo e del quarto ordine ai quali sono associate delle cosiddette *costanti di forma*. In particolare, il momento centrale del terzo ordine   un parametro utile per descrivere il grado di asimmetria della funzione di probabilit  nel caso discreto e della densit  di probabilit  nel caso assolutamente continuo, mentre il momento centrale del quarto ordine   utilizzato per descriverne il grado di piccatezza.

Definizione 5.10 *Sia X una variabile aleatoria dotata di valore medio finito ed avente varianza positiva. Si definisce coefficiente di simmetria, o "skewness", di X il rapporto*

$$\alpha_3 = \frac{E[(X - E(X))^3]}{\{E[(X - E(X))^2]\}^{3/2}} = \frac{\bar{\mu}_3}{\sigma_X^3}. \quad (5.56)$$

Dalla Proposizione 5.16 si trae che per ogni variabile aleatoria X avente distribuzione simmetrica rispetto all'asse $x = E(X)$ e dotata di valore medio finito, il momento centrale del terzo ordine, se esiste,   nullo, cos  che $\alpha_3 = 0$. Se invece risulta $\alpha_3 > 0$ ($\alpha_3 < 0$) la distribuzione di X presenta asimmetria positiva (asimmetria negativa) rispetto all'asse $x = E(X)$.

Definizione 5.11 *Sia X una variabile aleatoria dotata di valore medio finito ed avente varianza positiva. Si definisce curtosi di X il rapporto*

$$\alpha_4 = \frac{E[(X - E(X))^4]}{\{E[(X - E(X))^2]\}^2} = \frac{\bar{\mu}_4}{\sigma_X^4}. \quad (5.57)$$

La curtosi è tanto maggiore quanto maggiore è la piccatezza intorno al valore medio della funzione di probabilità nel caso discreto, oppure della densità di probabilità nel caso assolutamente continuo.

Esempio 5.19 Sia $X \sim \mathcal{B}(1, p)$ e si consideri la variabile aleatoria

$$Y = \frac{X^a - (1 - X)^a}{a},$$

con a reale diverso da zero. Si nota che Y è una variabile aleatoria discreta che assume il valore $-1/a$ con probabilità $1 - p$ ed il valore $1/a$ con probabilità p . Pertanto, risulta:

$$E(Y^n) = \frac{(-1)^n}{a^n} (1 - p) + \frac{1}{a^n} p = \frac{1}{a^n} [p + (-1)^n (1 - p)] \quad (n = 1, 2, \dots),$$

che mostra che i momenti di ordine dispari sono nulli, mentre per i momenti di ordine pari si ha $E(Y^{2k}) = a^{-2k}$ ($k = 1, 2, \dots$). Essendo nullo il valore medio di Y , i momenti centrali coincidono con i momenti. Si ricava quindi immediatamente che la skewness è nulla e che la curtosi è unitaria. \diamond

Utili per la descrizione di ulteriori caratteristiche delle variabili aleatorie sono la *mediana* e la *moda*.

Definizione 5.12 Si definisce *mediana* di una variabile aleatoria X di funzione di distribuzione $F_X(x)$, ogni reale x_m tale che entrambe le seguenti disuguaglianze sussistono:

$$P(X \leq x_m) \geq \frac{1}{2}, \quad P(X \geq x_m) \geq \frac{1}{2}, \quad (5.58)$$

o equivalentemente

$$F_X(x_m^-) \leq \frac{1}{2} \leq F_X(x_m). \quad (5.59)$$

In generale si possono avere più mediane appartenenti ad un intervallo chiuso che si può anche ridurre ad un solo punto. Inoltre se la variabile aleatoria è assolutamente continua, dalla (5.59) segue che ogni valore x_m tale che $F_X(x_m) = 1/2$ è una mediana.

Esempio 5.20 Sia $X \sim \mathcal{B}(1, p)$. Si ha:

$$F_X(x^-) = \begin{cases} 0, & x \leq 0 \\ 1 - p, & 0 < x \leq 1 \\ 1, & x > 1 \end{cases} \quad F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - p, & 0 \leq x < 1 \\ 1, & x \geq 1. \end{cases}$$

Ricordando la (5.59) ed osservando la Figura 5.10 si evince che se $0 < p < 1/2$ esiste un'unica mediana $x_m = 0$, se $p = 1/2$ ogni reale x_m tale che $0 \leq x_m \leq 1$ è una mediana e se $1/2 < p < 1$ esiste ancora un'unica mediana $x_m = 1$. \diamond

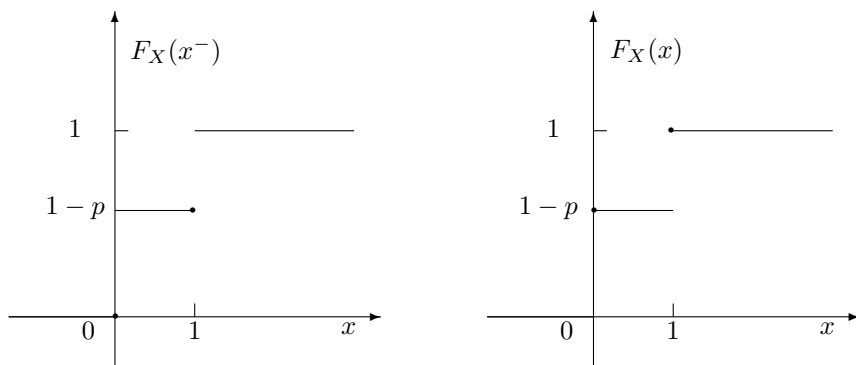


Figura 5.10 – Le funzioni $F_X(x^-)$ e $F_X(x)$ della variabile aleatoria di Bernoulli.

Esempio 5.21 Sia X una variabile aleatoria assolutamente continua di funzione di distribuzione

$$F_X(x) = \begin{cases} 0, & x \leq -1 \\ \frac{x+1}{2}, & -1 < x \leq 0 \\ 1 - \frac{1}{2}e^{-x}, & x > 0. \end{cases}$$

Come si evince dalla Figura 5.11 esiste un'unica mediana $x_m = 0$.

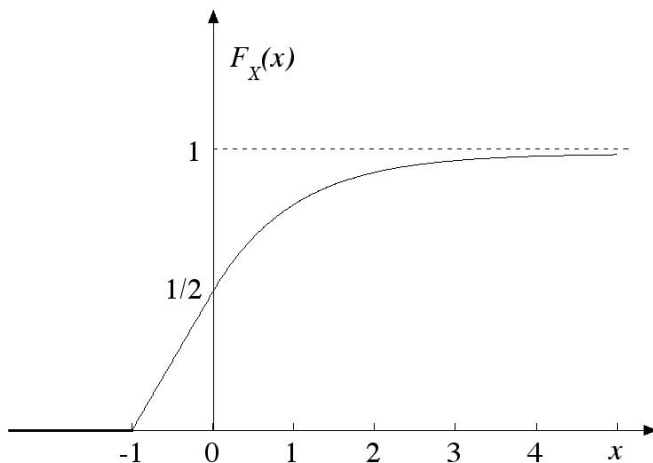


Figura 5.11 – Grafico della funzione di distribuzione dell'Esempio 5.21.

Invece, se si considera la variabile aleatoria assolutamente continua X di funzione di

distribuzione (v. Esempio 5.8)

$$F_X(x) = \begin{cases} -\frac{1}{2x}, & x \leq -1 \\ \frac{1}{2}, & -1 < x \leq 1 \\ 1 - \frac{1}{2x}, & x > 1 \end{cases}$$

ogni reale x_m tale che $-1 \leq x_m \leq 1$ è una mediana. \diamond

Definizione 5.13 Data una variabile aleatoria X di funzione di distribuzione $F_X(x)$, si definisce *moda* di X ogni punto x_M di massimo relativo della funzione di probabilità nel caso discreto oppure della densità di probabilità nel caso assolutamente continuo.

Si noti che la moda non è necessariamente unica.

Esempio 5.22 Sia $X \sim \mathcal{U}_d(n)$. Nell'Esempio 5.3 abbiamo mostrato che $\mu_1 = E(X) = (n+1)/2$. Vogliamo ora calcolare altre caratteristiche numeriche di X .

Poiché X assume i valori $1, 2, \dots, n$, ciascuno con probabilità $1/n$, facendo uso della (5.38) per i generici momenti di ordine k si ha:

$$\mu_k = \sum_{r=1}^n r^k \frac{1}{n} = \frac{1}{n} \sum_{r=1}^n r^k, \quad \bar{\mu}_k = \sum_{r=1}^n (r - \mu_1)^k \frac{1}{n} = \frac{1}{n} \sum_{r=1}^n (r - \mu_1)^k \quad (k = 1, 2, \dots).$$

Procedendo per induzione su n è possibile dimostrare che risulta:

$$\begin{aligned} \sum_{r=1}^n r &= \frac{n(n+1)}{2}, & \sum_{r=1}^n r^2 &= \frac{n(n+1)(2n+1)}{6}, \\ \sum_{r=1}^n r^3 &= \left[\frac{n(n+1)}{2} \right]^2, & \sum_{r=1}^n r^4 &= \frac{n(n+1)(2n+1)(3n^2+3n-1)}{30}, \end{aligned}$$

da cui segue:

$$\mu_2 = \frac{(n+1)(2n+1)}{6}, \quad \mu_3 = \frac{n(n+1)^2}{4}, \quad \mu_4 = \frac{(n+1)(2n+1)(3n^2+3n-1)}{30}.$$

Inoltre, dalla (5.47) si ottiene:

$$\text{Var}(X) = \bar{\mu}_2 = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{2} = \frac{n^2-1}{12}.$$

Essendo $p_X(\mu_1 + x) = p_X(\mu_1 - x)$ per ogni $x \in \mathbb{R}$, X ha distribuzione simmetrica rispetto all'asse $x = \mu_1$; dalla Proposizione 5.16 si deduce quindi che tutti i momenti centrali di ordine dispari sono nulli, ossia $\bar{\mu}_{2k+1} = 0$ ($k = 0, 1, \dots$), così che in particolare la skewness α_3 è nulla. Facendo uso della (5.41), per il momento centrale del quarto ordine si ha:

$$\bar{\mu}_4 = \sum_{i=0}^4 \binom{4}{i} (-\mu_1)^{4-i} \mu_i = -3\mu_1^2 + 6\mu_1^2\mu_2 - 4\mu_1\mu_3 + \mu_4 = \frac{(n^2-1)(3n^2-7)}{16 \cdot 15}.$$

Pertanto quando è $n > 1$, ossia nel caso in cui la variabile aleatoria è non degenera, la curtosi è data da

$$\alpha_4 = \frac{\bar{\mu}_4}{\sigma_X^4} = \frac{3}{5} \frac{3n^2 - 7}{n^2 - 1}.$$

Si noti che per $n > 1$ la curtosi è decrescente in n e tende a 1.8 quando $n \rightarrow +\infty$. Inoltre, ricordando la (4.2), segue che se n è dispari esiste un'unica mediana $x_m = (n+1)/2$ coincidente con il valore medio, mentre se n è pari ogni reale x_m tale che $n/2 \leq x_m \leq 1 + n/2$ è una mediana. Infine le mode sono tutti gli interi appartenenti all'intervallo $[1, n]$, ossia tutti i valori che può assumere X . \diamond

Esempio 5.23 Sia $X \sim \mathcal{U}(a, b)$. Nell'Esempio 5.9 abbiamo visto che $\mu_1 = E(X) = (a+b)/2$. Inoltre, dalla (5.39) segue:

$$\mu_n = E(X^n) = \frac{1}{b-a} \int_a^b x^n dx = \frac{b^{n+1} - a^{n+1}}{(n+1)(b-a)}.$$

Poiché

$$\mu_2 = E(X^2) = \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + b^2 + ab}{3},$$

dalla (5.47) si ottiene:

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{a^2 + b^2 + ab}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}.$$

Inoltre, essendo $f_X(\mu_1+x) = f_X(\mu_1-x)$ per ogni $x \in \mathbb{R}$, la distribuzione di X è simmetrica intorno all'asse $x = \mu_1 = (a+b)/2$; dalla Proposizione 5.16 si deduce quindi che tutti i momenti centrali di ordine dispari sono nulli, ossia $\bar{\mu}_{2k+1} = 0$ ($k = 0, 1, \dots$), da cui, in particolare, segue $\alpha_3 = 0$. I momenti centrali di ordine pari invece sono:

$$\begin{aligned} \bar{\mu}_{2k} &= E[(X - \mu_1)^{2k}] = \frac{1}{b-a} \int_a^b (x - \mu_1)^{2k} dx = \frac{1}{b-a} \int_{a-\mu_1}^{b-\mu_1} y^{2k} dy \\ &= \frac{(b-\mu_1)^{2k+1} - (a-\mu_1)^{2k+1}}{(b-a)(2k+1)} = \frac{1}{2k+1} \left(\frac{b-a}{2} \right)^{2k} \quad (k = 1, 2, \dots). \end{aligned}$$

Quindi per la curtosi si ha:

$$\alpha_4 = \frac{\bar{\mu}_4}{\sigma_X^4} = \frac{1}{5} \left(\frac{b-a}{2} \right)^4 \frac{144}{(b-a)^4} = \frac{9}{5} = 1.8.$$

In conclusione, il valore medio, coincidente con la mediana, è il punto medio dell'intervallo, mentre la varianza è uguale ad un dodicesimo del quadrato dell'ampiezza dell'intervallo. Inoltre, essendo la densità di probabilità simmetrica intorno al valore medio, la skewness è nulla. Infine, le mode sono tutti i punti dell'intervallo $[a, b]$. \diamond

Esempio 5.24 Sia $X \sim \mathcal{E}(1, \lambda)$. Poiché X è assolutamente continua, dalla (5.39), mediante ripetute integrazioni per parti, si ottiene:

$$\mu_n = E(X^n) = \int_0^{+\infty} x^n \lambda e^{-\lambda x} dx = \frac{n!}{\lambda^n},$$

da cui segue

$$E(X) = \frac{1}{\lambda}, \quad E(X^2) = \frac{2}{\lambda^2}, \quad \text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{1}{\lambda^2}.$$

Inoltre, facendo uso di (5.41), risulta:

$$\begin{aligned} \bar{\mu}_n &= E[(X - \mu_1)^n] = \sum_{i=0}^n \binom{n}{i} (-\mu_1)^{n-i} \mu_i = \sum_{i=0}^n \binom{n}{i} \left(-\frac{1}{\lambda}\right)^{n-i} \frac{i!}{\lambda^i} \\ &= n! \left(\frac{1}{\lambda}\right)^n \sum_{i=0}^n \frac{(-1)^{n-i}}{(n-i)!} = n! \left(\frac{1}{\lambda}\right)^n \sum_{j=0}^n \frac{(-1)^j}{j!} \quad (n = 0, 1, \dots), \end{aligned}$$

da cui è possibile ricavare la skewness e la curtosi:

$$\alpha_3 = \frac{\bar{\mu}_3}{\sigma_X^3} = 3! \left(\frac{1}{2} - \frac{1}{6}\right) = 2, \quad \alpha_4 = \frac{\bar{\mu}_4}{\sigma_X^4} = 4! \left(\frac{1}{2} - \frac{1}{6} + \frac{1}{24}\right) = 9.$$

Si noti che il valore medio e la deviazione standard coincidono con l'inverso del parametro λ . Infine, dalle Definizioni 5.12 e 5.13 risulta che la mediana è $x_m = \ln 2/\lambda$ e che la moda è $x_M = 0$. \diamond

Esempio 5.25 Sia $X \sim \mathcal{N}(\mu, \sigma^2)$. Essendo X assolutamente continua, dalla (5.39) risulta:

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} \frac{x}{\sigma \sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma y + \mu) \exp\left\{-\frac{y^2}{2}\right\} dy \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y \exp\left\{-\frac{y^2}{2}\right\} dy + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left\{-\frac{y^2}{2}\right\} dy = \mu, \end{aligned}$$

dove l'ultima uguaglianza segue poiché il primo integrale è nullo (essendo la funzione integranda $y e^{-y^2/2}$ dispari) mentre il secondo integrale fornisce $\sqrt{2\pi}$ (ricordando che la densità normale standard è integrata ad uno in \mathbb{R}). Inoltre, si ha:

$$\begin{aligned} \text{Var}(X) &= E[(X - \mu)^2] = \int_{-\infty}^{+\infty} \frac{(x-\mu)^2}{\sigma \sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y^2 \exp\left\{-\frac{y^2}{2}\right\} dy = -\frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y \left[\frac{d}{dy} \exp\left\{-\frac{y^2}{2}\right\}\right] dy \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left\{-\frac{y^2}{2}\right\} dy = \sigma^2. \end{aligned}$$

I parametri μ e σ^2 che compaiono nella densità normale hanno quindi proprio il significato rispettivamente di valore medio e di varianza. Inoltre, poiché μ risulta essere il centro di simmetria, la mediana di X è proprio μ . Quindi il parametro μ è sia moda (l'ascissa del massimo della densità), sia mediana, sia valore medio. Inoltre, essendo X a distribuzione simmetrica rispetto a $x = \mu$, i momenti centrali di ordine dispari sono nulli, ossia

$\bar{\mu}_{2k+1} = 0$ ($k = 0, 1, \dots$), di modo che in particolare è nulla la skewness. Infine, integrando ripetutamente per parti, si può dimostrare che risulta:

$$\bar{\mu}_{2k} = \frac{(2k)!}{k!} \left(\frac{\sigma^2}{2} \right)^k \quad (k = 1, 2, \dots).$$

Pertanto per la curtosi si ha:

$$\alpha_4 = \frac{\bar{\mu}_4}{\sigma_X^4} = \frac{(4)!}{2!} \left(\frac{1}{2} \right)^2 = 3.$$

Si noti che la curtosi della variabile uniforme ($\alpha_4 = 1.8$) è minore della curtosi della variabile normale ($\alpha_4 = 3$) che a sua volta è minore della curtosi della variabile esponenziale ($\alpha_4 = 9$). \diamond

5.5 Valore medio e momenti di una funzione di vettore aleatorio

Sia $\mathbf{X} = (X_1, X_2, \dots, X_n)$ un vettore aleatorio n -dimensionale e sia $g: \mathbb{R}^n \rightarrow \mathbb{R}$ una funzione Borel-misurabile. In taluni problemi si è interessati al calcolo dei momenti della variabile aleatoria $Y = g(\mathbf{X}) = g(X_1, X_2, \dots, X_n)$. In questi casi è possibile estendere i risultati del Paragrafo 5.3. Infatti Y è una variabile aleatoria unidimensionale, così che per essa si può definire la parte positiva Y^+ e la parte negativa Y^- . Se almeno uno tra $E(Y^+)$ e $E(Y^-)$ è finito, allora $E(Y)$ esiste e risulta:

$$E(Y) = E[g(\mathbf{X})] = E(Y^+) - E(Y^-) = E\{[g(\mathbf{X})]^+\} - E\{[g(\mathbf{X})]^-\}. \quad (5.60)$$

Nel caso in cui $E(Y^+) = E(Y^-) = +\infty$, il valore medio della variabile aleatoria Y non esiste. Il Teorema 5.1 può essere esteso al caso in esame. Infatti sussiste il seguente

Teorema 5.7 *Sia $\mathbf{X} = (X_1, X_2, \dots, X_n)$ un vettore aleatorio e sia $Y = g(\mathbf{X})$, con $g: \mathbb{R}^n \rightarrow \mathbb{R}$ una funzione Borel-misurabile, tale che $E(Y)$ esiste.*

(a) *Se \mathbf{X} è discreto e se $S = S_1 \times S_2 \times \dots \times S_n$ è l'insieme al più numerabile delle n -uple distinte \mathbf{x} di valori assunti dalle variabili aleatorie X_1, X_2, \dots, X_n con probabilità $p(\mathbf{x})$, si ha:*

$$\begin{aligned} E[g(\mathbf{X})] &= \sum_{\mathbf{x} \in S} g(\mathbf{x}) p_{\mathbf{X}}(\mathbf{x}) \\ &= \sum_{x_1 \in S_1} \dots \sum_{x_n \in S_n} g(x_1, \dots, x_n) P(X_1 = x_1, \dots, X_n = x_n). \end{aligned} \quad (5.61)$$

(b) *Se \mathbf{X} è assolutamente continuo con densità di probabilità $f_{\mathbf{X}}(\mathbf{x})$ risulta:*

$$\begin{aligned} E[g(\mathbf{X})] &= \int_{\mathbb{R}^n} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\ &= \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} g(x_1, \dots, x_n) f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \dots dx_n. \end{aligned} \quad (5.62)$$

Il seguente teorema mostra che il valore medio gode della *proprietà di linearità*, ossia che il valore medio di una combinazione lineare di variabili aleatorie è uguale alla combinazione lineare dei valori medi delle singole variabili.

Teorema 5.8 *Se $\mathbf{X} = (X_1, X_2, \dots, X_n)$ è un vettore aleatorio e c_1, c_2, \dots, c_n sono costanti reali, risulta:*

$$E(c_1 X_1 + c_2 X_2 + \dots + c_n X_n) = c_1 E(X_1) + c_2 E(X_2) + \dots + c_n E(X_n), \quad (5.63)$$

sempre che al secondo membro i valori medi esistano e che nella somma non siano presenti simultaneamente infiniti positivi e negativi.

Dimostrazione È possibile dimostrare la (5.63) facendo uso del Teorema 5.7 con $g(\mathbf{x}) = c_1 x_1 + c_2 x_2 + \dots + c_n x_n$. Infatti, se \mathbf{X} è un vettore aleatorio discreto, dalla (5.61), ricordando le proprietà delle funzioni di probabilità congiunte, si ha:

$$\begin{aligned} E(c_1 X_1 + \dots + c_n X_n) &= \sum_{x_1 \in S_1} \dots \sum_{x_n \in S_n} [c_1 x_1 + \dots + c_n x_n] P(X_1 = x_1, \dots, X_n = x_n) \\ &= c_1 \sum_{x_1 \in S_1} x_1 \left[\sum_{x_2 \in S_2} \dots \sum_{x_n \in S_n} P(X_1 = x_1, \dots, X_n = x_n) \right] + \dots \\ &\quad \dots + c_n \sum_{x_n \in S_n} x_n \left[\sum_{x_1 \in S_1} \dots \sum_{x_{n-1} \in S_{n-1}} P(X_1 = x_1, \dots, X_n = x_n) \right] \\ &= c_1 \sum_{x_1 \in S_1} x_1 P(X_1 = x_1) + \dots + c_n \sum_{x_n \in S_n} x_n P(X_n = x_n) \\ &= c_1 E(X_1) + \dots + c_n E(X_n). \end{aligned}$$

Similmente, se \mathbf{X} è un vettore aleatorio assolutamente continuo, facendo uso delle proprietà delle funzioni densità di probabilità congiunte, dalla (5.62) si trae:

$$\begin{aligned} E(c_1 X_1 + \dots + c_n X_n) &= \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} [c_1 x_1 + \dots + c_n x_n] f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \dots dx_n \\ &= c_1 \int_{-\infty}^{+\infty} x_1 dx_1 \left[\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_2 \dots dx_n \right] + \dots \\ &\quad \dots + c_n \int_{-\infty}^{+\infty} x_n dx_n \left[\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \dots dx_{n-1} \right] \\ &= c_1 \int_{-\infty}^{+\infty} x_1 f_{X_1}(x_1) dx_1 + \dots + c_n \int_{-\infty}^{+\infty} x_n f_{X_n}(x_n) dx_n \\ &= c_1 E(X_1) + \dots + c_n E(X_n). \end{aligned}$$

Ciò completa la dimostrazione. □

È opportuno osservare che la proprietà di linearità espressa dal Teorema 5.8 non sussiste sempre allorché si considera un'infinità numerabile di variabili aleatorie.

Esempio 5.26 Sia $\mathbf{X} = (X_1, X_2, \dots, X_n)$ un vettore aleatorio con componenti positive, indipendenti e identicamente distribuite. Calcoliamo il valore medio delle variabili aleatorie

$$Z_r = \frac{X_r}{X_1 + X_2 + \dots + X_n} \quad (r = 1, 2, \dots, n).$$

Per l'ipotesi di positività di X_1, X_2, \dots, X_n , ciascuna delle Z_r assume valori nell'intervallo $(0, 1)$, e pertanto $E(Z_r)$ esiste per $r = 1, 2, \dots, n$. Inoltre, è evidente che essendo X_1, X_2, \dots, X_n indipendenti e identicamente distribuite, le Z_1, Z_2, \dots, Z_n sono anch'esse identicamente distribuite, ma non necessariamente indipendenti. Pertanto, risulta $E(Z_1) = E(Z_2) = \dots = E(Z_n)$. Essendo $Z_1 + Z_2 + \dots + Z_n = 1$, dalla proprietà di linearità del valore medio risulta $E(Z_1) + E(Z_2) + \dots + E(Z_n) = 1$, da cui segue:

$$E(Z_r) = \frac{1}{n} \quad (r = 1, 2, \dots, n).$$

Si noti che in questo caso sussiste l'identità

$$E(Z_r) = \frac{E(X_r)}{E(X_1) + E(X_2) + \dots + E(X_n)} \quad (r = 1, 2, \dots, n).$$

◇

Alcuni importanti risultati possono essere dimostrati sotto l'ipotesi che le componenti del vettore aleatorio siano variabili aleatorie indipendenti, come qui di seguito indicato.

Teorema 5.9 *Se le componenti di $\mathbf{X} = (X_1, X_2, \dots, X_n)$ sono variabili aleatorie indipendenti, risulta:*

$$E(X_1 X_2 \cdots X_n) = E(X_1) E(X_2) \cdots E(X_n), \quad (5.64)$$

sempre che i valori medi al secondo membro esistano e che non siano presenti simultaneamente fattori nulli ed infiniti.

Dimostrazione Per dimostrare la validità della (5.64) è sufficiente far uso del Teorema 5.7 con $g(\mathbf{x}) = x_1 x_2 \cdots x_n$. Infatti, nel caso in cui le variabili aleatorie X_1, X_2, \dots, X_n sono discrete, dalla (5.61) e dalle proprietà delle funzioni di probabilità congiunte per variabili aleatorie indipendenti, segue:

$$\begin{aligned} E(X_1 X_2 \cdots X_n) &= \sum_{x_1 \in S_1} \cdots \sum_{x_n \in S_n} x_1 \cdots x_n P(X_1 = x_1, \dots, X_n = x_n) \\ &= \sum_{x_1 \in S_1} \cdots \sum_{x_n \in S_n} x_1 \cdots x_n P(X_1 = x_1) \cdots P(X_n = x_n) \\ &= \sum_{x_1 \in S_1} x_1 P(X_1 = x_1) \cdots \sum_{x_n \in S_n} x_n P(X_n = x_n) \\ &= E(X_1) E(X_2) \cdots E(X_n); \end{aligned}$$

se invece le variabili X_1, X_2, \dots, X_n sono assolutamente continue, dalla (5.62) e dalle proprietà delle funzioni densità di probabilità congiunte per variabili aleatorie indipendenti,

segue:

$$\begin{aligned}
 E(X_1 X_2 \cdots X_n) &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} x_1 \cdots x_n f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \cdots dx_n \\
 &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} x_1 \cdots x_n f_{X_1}(x_1) \cdots f_{X_n}(x_n) dx_1 \cdots dx_n \\
 &= \int_{-\infty}^{+\infty} x_1 f_{X_1}(x_1) dx_1 \cdots \int_{-\infty}^{+\infty} x_n f_{X_n}(x_n) dx_n \\
 &= E(X_1) E(X_2) \cdots E(X_n),
 \end{aligned}$$

il che completa la dimostrazione. \square

Si noti che l'affermazione del Teorema 5.9 non è invertibile: assegnate n variabili aleatorie X_1, X_2, \dots, X_n , la validità della (5.64) non comporta che esse siano indipendenti.

Proposizione 5.17 *Se le componenti di $\mathbf{X} = (X_1, X_2, \dots, X_n)$ sono variabili aleatorie indipendenti tali che $E(X_r^2)$ è finito per $r = 1, 2, \dots, n$, si ha:*

$$\text{Var}(X_1 X_2 \cdots X_n) = E(X_1^2) E(X_2^2) \cdots E(X_n^2) - [E(X_1)]^2 [E(X_2)]^2 \cdots [E(X_n)]^2. \quad (5.65)$$

Dimostrazione Essendo $E(X_r^2)$ finito per $r = 1, 2, \dots, n$, dalla (5.48) segue che $E(X_r)$ è finito per $r = 1, 2, \dots, n$. Si ponga ora $Z = X_1 X_2 \cdots X_n$; dalla (5.47) segue:

$$\text{Var}(X_1 X_2 \cdots X_n) = \text{Var}(Z) = E(Z^2) - [E(Z)]^2 = E(X_1^2 X_2^2 \cdots X_n^2) - [E(X_1 X_2 \cdots X_n)]^2.$$

Essendo X_1, X_2, \dots, X_n indipendenti, tali sono anche $X_1^2, X_2^2, \dots, X_n^2$; pertanto dal Teorema 5.9 segue immediatamente la (5.65). \square

Nel caso di un vettore aleatorio bidimensionale (X, Y) con componenti indipendenti e tali che $E(X^2)$ e $E(Y^2)$ sono entrambi finiti, ricordando che $E(X^2) = \text{Var}(X) + [E(X)]^2$ e $E(Y^2) = \text{Var}(Y) + [E(Y)]^2$, la (5.65) può essere così riscritta:

$$\text{Var}(XY) = \text{Var}(X) \text{Var}(Y) + [E(X)]^2 \text{Var}(Y) + [E(Y)]^2 \text{Var}(X). \quad (5.66)$$

5.6 Covarianza e coefficiente di correlazione

I momenti possono anche essere definiti per vettori aleatori. Nel seguito, per semplicità, consideriamo il caso di vettori aleatori bidimensionali sviluppando alcune considerazioni sui *momenti misti* e sui *momenti misti centrali*.

Definizione 5.14 *Il momento misto di ordine (i, j) ($i, j = 0, 1, \dots$) del vettore aleatorio (X, Y) è definito al seguente modo:*

$$\mu_{i,j} = E(X^i Y^j), \quad (5.67)$$

sempre che il valore medio al secondo membro della (5.67) esista. Inoltre, se esistono $E(X)$ e $E(Y)$, si definisce il momento misto centrale di ordine (i, j) ($i, j = 0, 1, \dots$):

$$\bar{\mu}_{i,j} = E[(X - E(X))^i (Y - E(Y))^j], \quad (5.68)$$

sempre che il valore medio al secondo membro della (5.68) esista.

In particolare, dalle (5.67) e (5.68) per $i = 0$ risulta

$$\mu_{0,j} = E(Y^j), \quad \bar{\mu}_{0,j} = E[(Y - E(Y))^j],$$

mentre per $j = 0$ si ottiene:

$$\mu_{i,0} = E(X^i), \quad \bar{\mu}_{i,0} = E[(X - E(X))^i].$$

Nella trattazione di problemi statistici particolare importanza riveste il momento misto $\bar{\mu}_{1,1}$, detto anche *covarianza*.

Definizione 5.15 Sia (X, Y) un vettore aleatorio. La covarianza di X e Y si definisce come

$$\text{Cov}(X, Y) = \bar{\mu}_{1,1} = E[(X - E(X))(Y - E(Y))], \quad (5.69)$$

sempre che il valore medio al secondo membro esista.

Il seguente teorema fornisce alcune proprietà della covarianza.

Teorema 5.10 Sia (X, Y) un vettore aleatorio. Risulta:

$$\text{Cov}(X, Y) = \text{Cov}(Y, X), \quad (5.70)$$

$$\text{Cov}(aX + b, cY + d) = ac \text{Cov}(X, Y). \quad (5.71)$$

Dimostrazione La (5.70) segue direttamente dalla Definizione 5.15 e dalla proprietà commutativa del prodotto. Inoltre, facendo nuovamente uso della Definizione 5.15 ed osservando che risulta

$$\begin{aligned} \text{Cov}(aX + b, cY + d) &= E\{[aX + b - E(aX + b)][cY + d - E(cY + d)]\} \\ &= E\{[aX - aE(X)][cY - cE(Y)]\} \\ &= ac E[(X - E(X))(Y - E(Y))], \end{aligned}$$

segue la (5.71). □

La (5.70) esprime la proprietà riflessiva della covarianza, mentre la (5.71) mostra che la covarianza resta invariata se ad una o ad entrambe le variabili si aggiunge una costante; la covarianza, inoltre, resta moltiplicata per una costante se una delle variabili è moltiplicata per tale costante. Se $\text{Cov}(X, Y) = 0$ si dice che X e Y sono non correlate o scorrelate; altrimenti esse si dicono correlate positivamente o negativamente a seconda del segno della (5.69). Le variabili sono correlate positivamente quando nel valore medio $E[(X - \mu_X)(Y - \mu_Y)]$ prevale la parte positiva sulla parte negativa, ossia quando agli effetti della media sono

prevalenti le differenze $X - E(X)$ e $Y - E(Y)$ dello stesso segno. Quando invece prevale la parte negativa, le variabili sono negativamente correlate. La covarianza rappresenta una *misura della dipendenza lineare tra X e Y* o, più precisamente, una misura della dipendenza lineare tra $X - E(X)$ e $Y - E(Y)$.

La seguente proposizione fornisce un modo alternativo per esprimere la covarianza.

Proposizione 5.18 *Sia (X, Y) un vettore aleatorio. Si ha:*

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y). \quad (5.72)$$

Dimostrazione Dalla Definizione 5.15, facendo uso del Teorema 5.8 segue:

$$\begin{aligned} \text{Cov}(X, Y) &= E[XY - E(Y)X - E(X)Y + E(X)E(Y)] \\ &= E(XY) - E(Y)E(X) - E(X)E(Y) + E(X)E(Y), \end{aligned}$$

da cui si trae immediatamente la (5.72). \square

Esempio 5.27 Sia (X, Y) un vettore aleatorio assolutamente continuo di densità congiunta

$$f_{X,Y}(x, y) = \begin{cases} 1 + \theta(x - y), & 0 < x < 1, 0 < y < 1 \\ 0, & \text{altrimenti,} \end{cases} \quad (5.73)$$

con $-1 \leq \theta \leq 1$. È facile rendersi conto che questa è una densità di probabilità essendo $f_{X,Y}(x, y) \geq 0$ e

$$\int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy = \int_0^1 dx \int_0^1 [1 + \theta(x - y)] dy = 1 + \theta \int_0^1 \left(x - \frac{1}{2}\right) dx = 1.$$

Facendo uso della relazione (3.47) è possibile ottenere le densità di probabilità marginali relative alle singole variabili aleatorie X e Y :

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy = \begin{cases} 1 + \theta \left(x - \frac{1}{2}\right), & 0 < x < 1 \\ 0, & \text{altrimenti,} \end{cases} \\ f_Y(y) &= \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx = \begin{cases} 1 + \theta \left(\frac{1}{2} - y\right), & 0 < y < 1 \\ 0, & \text{altrimenti.} \end{cases} \end{aligned}$$

Le variabili X e Y sono indipendenti se e solo se si sceglie $\theta = 0$; infatti, soltanto in tal caso risulta $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ per ogni $(x, y) \in \mathbb{R}^2$. Per determinare la covarianza di X e Y calcoliamo i seguenti valori medi:

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x f_X(x) dx = \int_0^1 x \left[1 + \theta \left(x - \frac{1}{2}\right)\right] dx = \frac{1}{2} + \frac{\theta}{12}, \\ E(Y) &= \int_{-\infty}^{+\infty} y f_Y(y) dy = \int_0^1 y \left[1 + \theta \left(\frac{1}{2} - y\right)\right] dy = \frac{1}{2} - \frac{\theta}{12}, \\ E(XY) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f_{X,Y}(x, y) dx dy = \int_0^1 \int_0^1 xy [1 + \theta(x - y)] dx dy = \frac{1}{4}, \end{aligned}$$

da cui segue:

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = \left(\frac{\theta}{12}\right)^2.$$

Pertanto X e Y sono correlate positivamente per $-1 \leq \theta < 0$ oppure $0 < \theta \leq 1$, mentre sono non correlate per $\theta = 0$. \diamond

Teorema 5.11 *Se le componenti del vettore aleatorio (X, Y) sono indipendenti, risulta $\text{Cov}(X, Y) = 0$.*

Dimostrazione Dal Teorema 5.9 segue che se X e Y sono indipendenti si ha $E(XY) = E(X)E(Y)$, così che dalla (5.72) segue la tesi. \square

Si noti che in generale non sussiste il viceversa nel senso che la covarianza può annullarsi anche se le variabili aleatorie coinvolte non sono indipendenti. I seguenti esempi sono finalizzati a chiarire quanto detto.

Esempio 5.28 Sia (X, Y) un vettore aleatorio discreto di funzione di probabilità

$$p_{X,Y}(x, y) = \begin{cases} 1/3, & (x = -1, y = 0) \text{ oppure } (x = 0, y = 1) \text{ oppure } (x = 1, y = 0) \\ 0, & \text{altrimenti.} \end{cases}$$

Facendo uso della relazione (3.41) si ottengono le funzioni di probabilità marginali relative alle singole variabili aleatorie X e Y :

$$p_X(x) = \begin{cases} 1/3, & x = -1 \\ 1/3, & x = 0 \\ 1/3, & x = 1 \\ 0, & \text{altrimenti;} \end{cases} \quad p_Y(y) = \begin{cases} 2/3, & y = 0 \\ 1/3, & y = 1 \\ 0, & \text{altrimenti.} \end{cases}$$

Le variabili aleatorie X e Y non sono indipendenti in quanto l'uguaglianza $p_{X,Y}(x, y) = p_X(x)p_Y(y)$ non è soddisfatta per ogni coppia $(x, y) \in \mathbb{R}^2$. Nonostante ciò esse sono non correlate. Infatti risulta $E(XY) = 0$, $E(X) = 0$, $E(Y) = 1/3$, da cui segue immediatamente $\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0$. \diamond

Esempio 5.29 Sia (X, Y) un vettore aleatorio, con X assolutamente continua a distribuzione simmetrica intorno all'asse $x = 0$ e sia $Y = X^2$. Si supponga che il momento del secondo ordine di X sia finito e che i momenti di ordine dispari esistano. Dalla Proposizione 5.10 segue che i momenti di ordine dispari di X sono tutti nulli. Pertanto, dalla (5.72) si ha:

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = E(X^3) - E(X)E(X^2) = 0.$$

Quest'ultima relazione mostra che X e Y sono non correlate. D'altra parte è facile convincersi della non indipendenza di X e Y , essendo Y una funzione di X . Più rigorosamente, la non indipendenza di X e Y può essere dimostrata mostrando che l'uguaglianza $F_{X,Y}(x, y) = F_X(x)F_Y(y)$ non è soddisfatta per ogni coppia $(x, y) \in \mathbb{R}^2$. \diamond

La seguente proposizione, nota come *disuguaglianza di Schwarz*, fornisce una relazione tra il momento misto del primo ordine e i momenti del secondo ordine di X e Y .

Proposizione 5.19 Sia (X, Y) un vettore aleatorio. Se $E(X^2)$ e $E(Y^2)$ sono finiti, allora anche $E(XY)$ è finito e risulta:

$$|E(XY)| \leq \sqrt{E(X^2) E(Y^2)}. \quad (5.74)$$

Dimostrazione Se almeno una delle variabili X e Y è degenera in zero, la (5.74) sussiste banalmente con il segno di uguaglianza. Assumiamo pertanto che le variabili aleatorie X e Y siano non degeneri così che $E(X^2) > 0$ e $E(Y^2) > 0$. Per ogni reale z sia

$$h(z) = E[(zX - Y)^2] = z^2 E(X^2) - 2z E(XY) + E(Y^2).$$

Poiché $h(z)$ è il valore medio della variabile aleatoria non negativa $(zX - Y)^2$, risulta $h(z) \geq 0$ per ogni reale z . D'altra parte $h(z)$ è un polinomio di secondo grado in z ; avendo supposto che $E(X^2) > 0$, risulta $h(z) \geq 0$ per ogni reale z se e solo se il discriminante dell'equazione $h(z) = 0$ è non positivo, ossia se e solo se $[E(XY)]^2 - E(X^2) E(Y^2) \leq 0$. Per le ipotesi di finitezza dei momenti del secondo ordine si ha quindi:

$$[E(XY)]^2 \leq E(X^2) E(Y^2) < +\infty.$$

Da questa segue che $E(XY)$ è finito ed inoltre che è verificata la (5.74). \square

Definizione 5.16 Sia (X, Y) un vettore aleatorio. Dicesi coefficiente di correlazione di X e Y il rapporto

$$\varrho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}, \quad (5.75)$$

sempre che $E(X^2)$ e $E(Y^2)$ sono finiti ed inoltre $\sigma_X = \sqrt{\text{Var}(X)} > 0$ e $\sigma_Y = \sqrt{\text{Var}(Y)} > 0$.

Si noti che le ipotesi $E(X^2) < +\infty$ e $E(Y^2) < +\infty$ presenti nella Definizione 5.16 implicano, per la (5.48), che $E(X)$ e $E(Y)$ sono finiti di modo che tali sono anche $\text{Var}(X)$ e $\text{Var}(Y)$; inoltre, se $E(X^2)$ e $E(Y^2)$ sono finiti, dalla (5.74) si deduce che $E(XY)$ è finito; pertanto anche la covarianza di X e Y è finita.

Se le componenti del vettore aleatorio (X, Y) sono indipendenti, dal Teorema 5.11 segue $\varrho(X, Y) = 0$. È inoltre opportuno sottolineare che, così come accade per la covarianza, il coefficiente di correlazione può essere nullo anche se X e Y non sono indipendenti. Inoltre, il coefficiente di correlazione, a differenza della covarianza, è sempre adimensionale, essendo la covarianza di X e Y delle stesse dimensioni del prodotto $\sigma_X \sigma_Y$.

Teorema 5.12 Se per il vettore aleatorio (X, Y) risulta che $E(X^2)$ e $E(Y^2)$ sono finiti e se inoltre $\sigma_X > 0$ e $\sigma_Y > 0$, allora

$$|\varrho(X, Y)| \leq 1. \quad (5.76)$$

Dimostrazione Si consideri il vettore aleatorio (U, V) , con $U = X - E(X)$ e $V = Y - E(Y)$. Dalla disuguaglianza di Schwarz (5.74), se $E(U^2)$ e $E(V^2)$ sono finiti risulta:

$$|E(UV)| \leq \sqrt{E(U^2) E(V^2)}. \quad (5.77)$$

Poiché

$$E(U^2) = E[(X - E(X))^2] = \text{Var}(X), \quad E(V^2) = E[(Y - E(Y))^2] = \text{Var}(Y),$$

$$E(UV) = E[(X - E(X))(Y - E(Y))] = \text{Cov}(X, Y),$$

dalla (5.77) segue:

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X) \text{Var}(Y)},$$

che, congiuntamente con la (5.75), conduce immediatamente alla (5.76). \square

Il coefficiente di correlazione fornisce una misura della dipendenza lineare tra X e Y o, più precisamente, una misura della dipendenza lineare tra $X - E(X)$ e $Y - E(Y)$, come mostrato dai seguenti teoremi.

Teorema 5.13 Sia (X, Y) un vettore aleatorio e sia $Y - E(Y) = a[X - E(X)]$, con a reale non nullo. Se $\text{Var}(X)$ è finita e non nulla, risulta $\varrho(X, Y) = 1$ se $a > 0$ e $\varrho(X, Y) = -1$ se $a < 0$.

Dimostrazione Poiché per ipotesi si ha $Y - E(Y) = a[X - E(X)]$ con $a \neq 0$, segue $\text{Var}(Y) = a^2 \text{Var}(X)$. Inoltre, risulta:

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = a E[(X - E(X))^2] = a \text{Var}(X).$$

Dalla Definizione 5.16 segue infine:

$$\varrho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}} = \frac{a \text{Var}(X)}{\sqrt{\text{Var}(X)}\sqrt{a^2 \text{Var}(X)}} = \frac{a}{|a|},$$

che completa la dimostrazione. \square

Teorema 5.14 Sia (X, Y) un vettore aleatorio e siano $E(X^2)$ e $E(Y^2)$ finiti; siano inoltre $\sigma_X > 0$ e $\sigma_Y > 0$. Se $\varrho(X, Y) = -1$ allora con probabilità unitaria risulta:

$$Y - E(Y) = -\frac{\sigma_X}{\sigma_Y} [X - E(X)], \quad (5.78)$$

mentre se $\varrho(X, Y) = 1$ allora con probabilità unitaria si ha:

$$Y - E(Y) = \frac{\sigma_X}{\sigma_Y} [X - E(X)]. \quad (5.79)$$

Dimostrazione Dal Teorema 5.8 seguono le seguenti identità:

$$E\left[\left(\frac{X - E(X)}{\sigma_X} + \frac{Y - E(Y)}{\sigma_Y}\right)^2\right] = \frac{\text{Var}(X)}{\sigma_X^2} + \frac{\text{Var}(Y)}{\sigma_Y^2} + 2 \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

$$= 2[1 + \varrho(X, Y)], \quad (5.80)$$

$$E\left[\left(\frac{X - E(X)}{\sigma_X} - \frac{Y - E(Y)}{\sigma_Y}\right)^2\right] = \frac{\text{Var}(X)}{\sigma_X^2} + \frac{\text{Var}(Y)}{\sigma_Y^2} - 2 \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

$$= 2[1 - \varrho(X, Y)]. \quad (5.81)$$

Poiché nella (5.80) compare il valore medio di una variabile aleatoria non negativa, si ha $\varrho(X, Y) \geq -1$. Dalla Proposizione 5.8 segue che la variabile aleatoria $[X - E(X)]/\sigma_X + [Y - E(Y)]/\sigma_Y$ è degenera in zero se $\varrho(X, Y) = -1$, ossia se con probabilità unitaria sussiste la (5.78). Analogamente, poiché la (5.81) esprime il valore medio di una variabile aleatoria non negativa, si ha $\varrho(X, Y) \leq 1$. Dalla Proposizione 5.8 segue che $[X - E(X)]/\sigma_X - [Y - E(Y)]/\sigma_Y$ è una variabile degenera in zero se $\varrho(X, Y) = 1$, ossia se la (5.79) sussiste con probabilità unitaria. \square

Il Teorema 5.13 mostra che X e Y sono in relazione lineare decrescente se $\varrho(X, Y) = -1$, mentre sono in relazione lineare crescente se $\varrho(X, Y) = 1$.

Esempio 5.30 Si consideri il vettore aleatorio discreto (X, Y) di funzione di probabilità congiunta

$$p_{X,Y}(x, y) = \begin{cases} \theta, & (x = 0, y = 1) \text{ oppure } (x = 1, y = 0), \\ \frac{1}{2} - \theta, & (x = 0, y = 0) \text{ oppure } (x = 1, y = 1) \\ 0, & \text{altrimenti,} \end{cases}$$

con $0 \leq \theta \leq 1/2$. È facile rendersi conto che le funzioni di probabilità marginali di X e Y sono:

$$p_X(x) = \begin{cases} 1/2, & x = 0 \\ 1/2, & x = 1 \\ 0, & \text{altrimenti,} \end{cases} \quad p_Y(y) = \begin{cases} 1/2, & y = 0 \\ 1/2, & y = 1 \\ 0, & \text{altrimenti;} \end{cases}$$

X e Y hanno dunque entrambe distribuzione di Bernoulli di parametro $1/2$. Si noti che X e Y sono indipendenti se e solo se si sceglie $\theta = 1/4$; ciò comporta che il coefficiente di correlazione è nullo. Inoltre, poiché

$$E(X) = E(Y) = \frac{1}{2}, \quad \text{Var}(X) = \text{Var}(Y) = \frac{1}{4}, \quad E(XY) = \frac{1}{2} - \theta,$$

risulta:

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = \frac{1}{4} - \theta, \quad \varrho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = 1 - 4\theta.$$

Si noti che per $\theta = 1/2$ risulta $\varrho(X, Y) = -1$ e che per la (5.78) ciò implica $Y = 1 - X$ con probabilità unitaria. Analogamente, se $\theta = 0$ risulta $\varrho(X, Y) = 1$ che, per la (5.79), implica $Y = X$ con probabilità unitaria. \diamond

Esempio 5.31 Sia (X, Y) un vettore aleatorio assolutamente continuo di densità congiunta (5.73). Calcoliamo il coefficiente di correlazione. Come mostrato nell'Esempio 5.27, risulta:

$$E(X) = \frac{1}{2} + \frac{\theta}{12}, \quad E(Y) = \frac{1}{2} - \frac{\theta}{12}, \quad \text{Cov}(X, Y) = \left(\frac{\theta}{12}\right)^2.$$

Poiché

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f_X(x) dx = \int_0^1 x^2 \left[1 + \theta \left(x - \frac{1}{2} \right) \right] dx = \frac{1}{3} + \frac{\theta}{12},$$

$$E(Y^2) = \int_{-\infty}^{+\infty} y^2 f_Y(y) dy = \int_0^1 y^2 \left[1 + \theta \left(\frac{1}{2} - y \right) \right] dy = \frac{1}{3} - \frac{\theta}{12},$$

dalla (5.47) segue:

$$\text{Var}(X) = \text{Var}(Y) = \frac{1}{12} - \left(\frac{\theta}{12} \right)^2.$$

Pertanto, facendo uso della (5.75), si trae:

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}} = \frac{\theta^2}{12 - \theta^2}.$$

Dall'ipotesi $-1 \leq \theta \leq 1$, segue $0 \leq \rho(X, Y) < 1/11$, così che X e Y non sono legate da una relazione lineare. \diamond

5.7 Momenti centrali di somme di variabili aleatorie

Nel Teorema 5.8 è stato mostrato che il valore medio gode della proprietà di linearità; ciò, come mostra il seguente teorema, non vale per la varianza.

Teorema 5.15 *Sia $\mathbf{X} = (X_1, X_2, \dots, X_n)$ un vettore aleatorio le cui componenti possiedono ciascuna momento del secondo ordine finito e siano a_1, a_2, \dots, a_n numeri reali. Risulta:*

$$\text{Var}(a_1 X_1 + a_2 X_2 + \dots + a_n X_n) = \sum_{i=1}^n a_i^2 \text{Var}(X_i) + 2 \sum_{\substack{i,j=1 \\ i < j}}^n a_i a_j \text{Cov}(X_i, X_j). \quad (5.82)$$

Dimostrazione Poiché per ipotesi $E(X_r^2)$ è finito per $r = 1, 2, \dots, n$, dalla Proposizione 5.19 segue che $E(X_i X_j)$ è finito per $i, j = 1, 2, \dots, n$ con $i \neq j$. Dalla Definizione 5.6 e dal Teorema 5.8 si ricava:

$$\begin{aligned} \text{Var}\left(\sum_{i=1}^n a_i X_i\right) &= E\left[\left(\sum_{i=1}^n a_i X_i - \sum_{i=1}^n a_i E(X_i)\right)^2\right] = E\left[\left(\sum_{i=1}^n a_i [X_i - E(X_i)]\right)^2\right] \\ &= \sum_{i=1}^n \sum_{j=1}^n a_i a_j E[(X_i - E(X_i))(X_j - E(X_j))]. \end{aligned} \quad (5.83)$$

Gli addendi presenti nel lato destro della (5.83) sono del tipo $E[(X_i - E(X_i))^2]$ per $i = j$ e pertanto coincidono con $\text{Var}(X_i)$, mentre per $i \neq j$ dalla Definizione 5.69 si ha:

$$E[(X_i - E(X_i))(X_j - E(X_j))] = \text{Cov}(X_i, X_j).$$

Pertanto, la (5.83) fornisce

$$\text{Var}(a_1X_1 + a_2X_2 + \dots + a_nX_n) = \sum_{i=1}^n a_i^2 \text{Var}(X_i) + \sum_{\substack{i,j=1 \\ i \neq j}}^n a_i a_j \text{Cov}(X_i, X_j),$$

che, per la proprietà di simmetria della covarianza, coincide con la (5.82). \square

Dal Teorema 5.15 si vede che la varianza di una combinazione lineare di variabili aleatorie non è in generale uguale alla combinazione lineare delle varianze delle singole variabili. Aggiungendo l'ipotesi di indipendenza delle variabili aleatorie si ottiene invece il seguente risultato.

Proposizione 5.20 *Se le componenti di $\mathbf{X} = (X_1, X_2, \dots, X_n)$ sono variabili aleatorie indipendenti, ognuna avente momento del secondo ordine finito, e se a_1, a_2, \dots, a_n sono numeri reali, si ha:*

$$\text{Var}(a_1X_1 + a_2X_2 + \dots + a_nX_n) = \sum_{i=1}^n a_i^2 \text{Var}(X_i). \quad (5.84)$$

Dimostrazione Se X_1, X_2, \dots, X_n sono indipendenti, risulta $\text{Cov}(X_i, X_j) = 0$ per $i, j = 1, 2, \dots, n$ con $i \neq j$, così che dalla (5.82) segue la (5.84). \square

Si noti che per $a_1 = a_2 = \dots = a_n = 1$ la Proposizione 5.20 afferma che la varianza di una somma di variabili aleatorie indipendenti è uguale alla somma delle varianze delle singole variabili.

Esempio 5.32 Si consideri un esperimento consistente in n prove ripetute e si assuma che nella prova i -esima i risultati di interesse siano sintetizzabili nel verificarsi di due eventi necessari ed incompatibili: A_i (interpretabile come successo) e \bar{A}_i (interpretabile come insuccesso). Si assuma che la probabilità del verificarsi di A_i , ossia di un successo alla prova i -esima, sia uguale a p_i ($i = 1, 2, \dots, n$). Se X denota la variabile aleatoria che rappresenta il numero di successi nelle n prove, risulta:

$$X = X_1 + X_2 + \dots + X_n,$$

dove ciascuna delle X_r ($r = 1, 2, \dots, n$) è una variabile discreta che può assumere due soli valori: 0 (interpretabile come insuccesso) e 1 (interpretabile come successo). È evidente che essendo $X_r \sim \mathcal{B}(1, p_r)$, risulta $E(X_r) = p_r$ e $\text{Var}(X_r) = E(X_r^2) - [E(X_r)]^2 = p_r(1 - p_r)$ per $r = 1, 2, \dots, n$. Dal Teorema 5.8 segue immediatamente:

$$E(X) = E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n) = \sum_{r=1}^n p_r, \quad (5.85)$$

sia che le prove siano indipendenti sia che non lo siano. Per determinare la varianza occorre invece analizzare separatamente questi due casi. Infatti, se si suppone che le n prove sono

indipendenti, dalla Proposizione 5.20 si ricava:

$$\text{Var}(X) = \text{Var}(X_1 + X_2 + \dots + X_n) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n) = \sum_{r=1}^n p_r (1 - p_r). \quad (5.86)$$

In particolare, se le prove sono effettuate tutte in condizioni identiche (ossia se $p_1 = p_2 = \dots = p_n = p$) e se inoltre esse sono indipendenti, per la Proposizione 4.1 risulta $X \sim \mathcal{B}(n, p)$. In tal caso dalle (5.85) e (5.86) si ottiene immediatamente:

$$E(X) = np, \quad \text{Var}(X) = np(1 - p).$$

Supponiamo ora che le prove non siano indipendenti. In questo caso non è sufficiente conoscere le probabilità p_1, p_2, \dots, p_n con le quali gli eventi A_1, A_2, \dots, A_n si verificano, ma è necessario conoscere la funzione di probabilità congiunta delle variabili X_1, X_2, \dots, X_n . Per calcolare la varianza di X è comunque sufficiente determinare le probabilità dell'evento $A_i \cap A_j$, ossia del verificarsi simultaneo di un successo nella i -esima e nella j -esima prova, per $i, j = 1, 2, \dots, n$, con $i \neq j$. Infatti, posto $q_{i,j} = P(X_i = 1, X_j = 1)$, risulta:

$$E(X_i X_j) = \sum_{i=0}^1 \sum_{j=0}^1 i j P(X_i = i, X_j = j) = P(X_i = 1, X_j = 1) = q_{i,j}$$

e quindi:

$$\text{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i) E(X_j) = q_{i,j} - p_i p_j$$

per $i, j = 1, 2, \dots, n$, con $i \neq j$. Facendo uso di tale risultato nella (5.82) si ottiene:

$$\text{Var}(X) = \text{Var}(X_1 + X_2 + \dots + X_n) = \sum_{i=1}^n p_i (1 - p_i) + 2 \sum_{\substack{i,j=1 \\ i < j}}^n (q_{i,j} - p_i p_j). \quad (5.87)$$

Se, in particolare, le prove sono effettuate tutte in condizioni identiche, ossia se si suppone $p_1 = p_2 = \dots = p_n = p$ e inoltre che $q_{i,j} = q$ per $i, j = 1, 2, \dots, n$ con $i \neq j$, la formula (5.87) diventa:

$$\text{Var}(X) = np(1 - p) + n(n - 1)(q - p^2), \quad (5.88)$$

essendo q la probabilità del verificarsi di un successo in una coppia qualsiasi di differenti prove. \diamond

Esempio 5.33 Consideriamo l'esperimento che consiste nell'estrarre n biglie senza reinserimento da un'urna contenente N biglie, m delle quali sono azzurre e $N - m$ bianche ($0 \leq n \leq N$, $0 \leq m \leq N$), e sia X la variabile aleatoria che descrive il numero di biglie azzurre estratte. Calcoliamo valore medio e varianza di X .

Come mostrato in (4.9), X ha distribuzione ipergeometrica di parametri $n, m, N - m$. Sia X_r ($r = 1, 2, \dots, n$) la variabile aleatoria che descrive l'uscita della biglia azzurra alla r -esima estrazione; X_r può assumere i due soli valori, 0 (interpretabile come uscita della

biglia bianca alla r -esima estrazione) e 1 (interpretabile come uscita della biglia azzurra alla r -esima estrazione). Chiaramente risulta:

$$X = X_1 + X_2 + \dots + X_n.$$

È inoltre evidente che si ha:

$$E(X_r) = \frac{m}{N}, \quad \text{Var}(X_r) = \frac{m}{N} \left(1 - \frac{m}{N}\right) \quad (r = 1, 2, \dots, n)$$

ed inoltre:

$$E(X_i X_j) = P(X_i = 1, X_j = 1) = \frac{m(m-1)}{N(N-1)} \quad (i, j = 1, 2, \dots, n; i \neq j).$$

Pertanto, X_i e X_j sono negativamente correlate, avendosi:

$$\text{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j) = \frac{m(m-1)}{N(N-1)} - \frac{m^2}{N^2} = -\frac{m}{N(N-1)} \left(1 - \frac{m}{N}\right).$$

Per la proprietà di linearità del valore medio si ha poi:

$$E(X) = \sum_{r=1}^n E(X_r) = n \frac{m}{N}.$$

Inoltre, ponendo $p = m/N$ e $q = m(m-1)/[N(N-1)]$ nella (5.88), si ottiene:

$$\begin{aligned} \text{Var}(X) &= n \frac{m}{N} \left(1 - \frac{m}{N}\right) - n(n-1) \frac{m}{N(N-1)} \left(1 - \frac{m}{N}\right) \\ &= n \frac{m(N-n)}{N(N-1)} \left(1 - \frac{m}{N}\right). \end{aligned}$$

◇

In alcune situazioni coinvolgenti somme di variabili aleatorie indipendenti è richiesto il calcolo dei momenti centrali di ordine superiore al secondo. Nel seguito, per $k = 3, 4, \dots$, denoteremo con $\bar{\mu}_k(X_1 + X_2 + \dots + X_n)$ il momento centrale di ordine k di $X_1 + X_2 + \dots + X_n$ e con $\bar{\mu}_k(X_r)$ il momento centrale di ordine k di X_r ($k = 3, 4, \dots$)

Proposizione 5.21 *Sia (X_1, X_2, \dots, X_n) un vettore aleatorio a componenti indipendenti, ognuna con momento del primo ordine finito. Risulta:*

$$\bar{\mu}_3(X_1 + X_2 + \dots + X_n) = \sum_{r=1}^n \bar{\mu}_3(X_r), \quad (5.89)$$

$$\bar{\mu}_4(X_1 + X_2 + \dots + X_n) = \sum_{r=1}^n \bar{\mu}_4(X_r) + 6 \sum_{\substack{i,j=1 \\ i < j}}^n \text{Var}(X_i) \text{Var}(X_j), \quad (5.90)$$

sempre che i momenti centrali al secondo membro siano finiti.

Dimostrazione Dalla (5.37) e dal Teorema 5.8, per $k = 3, 4, \dots$ si ha:

$$\begin{aligned} \bar{\mu}_k(X_1 + X_2 + \dots + X_n) &= E \left[\left(\sum_{i=1}^n X_i - \sum_{i=1}^n E(X_i) \right)^k \right] = E \left[\left(\sum_{i=1}^n [X_i - E(X_i)] \right)^k \right] \\ &= \sum_{i_1=1}^n \sum_{i_2=1}^n \dots \sum_{i_k=1}^n E \left[(X_{i_1} - E(X_{i_1})) (X_{i_2} - E(X_{i_2})) \dots (X_{i_k} - E(X_{i_k})) \right]. \end{aligned} \quad (5.91)$$

Sia $k = 3$. Per il calcolo del valore medio

$$E \{ [X_{i_1} - E(X_{i_1})] [X_{i_2} - E(X_{i_2})] [X_{i_3} - E(X_{i_3})] \} \quad (5.92)$$

occorre distinguere tre casi: (a) gli indici i_1, i_2, i_3 sono uguali, (b) solo due indici sono uguali, (c) i tre indici sono tutti diversi. Nel caso (a), posto $i_1 = i_2 = i_3 = i$, la (5.92) si identifica con $\bar{\mu}_3(X_i)$. Nel caso (b), denotando i due indici uguali con i e quello diverso con j , la (5.92) coincide con $\bar{\mu}_2(X_i) \bar{\mu}_1(X_j) = 0$. Infine, nel caso (c) si ottiene il prodotto di tre momenti centrali del primo ordine, così che la (5.92) è nulla. Quindi se $k = 3$, nell'ultima uguaglianza della (5.91) risultano presenti soltanto i termini con i tre indici uguali, risultando così verificata la (5.89).

Sia ora $k = 4$. Per il calcolo del valore medio

$$E \{ [X_{i_1} - E(X_{i_1})] [X_{i_2} - E(X_{i_2})] [X_{i_3} - E(X_{i_3})] [X_{i_4} - E(X_{i_4})] \} \quad (5.93)$$

occorre distinguere cinque casi: (a) gli indici i_1, i_2, i_3, i_4 sono uguali, (b) solo tre indici sono uguali, (c) due indici sono uguali ed i rimanenti sono diversi tra loro e dai due considerati, (d) due indici sono uguali ed i rimanenti due sono uguali tra loro e differenti dagli altri, (e) i quattro indici sono tutti diversi. Nel caso (a), posto $i_1 = i_2 = i_3 = i_4 = i$, la (5.93) si identifica con $\bar{\mu}_4(X_i)$. Nel caso (b), denotando i tre indici uguali con i ed il rimanente con j , la (5.93) coincide con $\bar{\mu}_3(X_i) \bar{\mu}_1(X_j) = 0$. Nel caso (c), denotando i due indici uguali con i e i restanti due indici con j e k , la (5.93) conduce a $\bar{\mu}_2(X_i) \bar{\mu}_1(X_j) \bar{\mu}_1(X_k) = 0$. Inoltre nel caso (d), denotando con i e j le due coppie di indici uguali, la (5.93) diventa $\bar{\mu}_2(X_i) \bar{\mu}_2(X_j)$. Infine, nel caso (e) si ottiene il prodotto dei quattro momenti centrali del primo ordine, così che la (5.93) è nulla. Quindi se $k = 4$, nell'ultima uguaglianza di (5.91) risultano presenti soltanto i termini con i quattro indici uguali e quelli con due differenti coppie di indici uguali, risultando così verificata la (5.90). \square

Esempio 5.34 Sia $X \sim \mathcal{B}(n, p)$. Si vuole calcolare skewness e curtosi di X .

Nella Proposizione 4.1 abbiamo mostrato che $X = X_1 + X_2 + \dots + X_n$, con X_1, X_2, \dots, X_n variabili indipendenti e identicamente distribuite tali che $X_i \sim \mathcal{B}(1, p)$ per $i = 1, 2, \dots, n$. Essendo $E(X_i) = p$, risulta:

$$\begin{aligned} \bar{\mu}_3(X_i) &= \sum_{k=0}^1 (k-p)^3 P(X_i = k) = (-p)^3 (1-p) + (1-p)^3 p = p(1-p)(1-2p), \\ \bar{\mu}_4(X_i) &= \sum_{k=0}^1 (k-p)^4 P(X_i = k) = (-p)^4 (1-p) + (1-p)^4 p \\ &= p(1-p)[1-3p(1-p)]. \end{aligned}$$

Per l'indipendenza e l'identica distribuzione di X_1, X_2, \dots, X_n dalla (5.89) si ha:

$$\begin{aligned}\bar{\mu}_3(X) &= \bar{\mu}_3(X_1 + X_2 + \dots + X_n) = n p (1 - p) (1 - 2p) \\ \bar{\mu}_4(X) &= \bar{\mu}_3(X_1 + X_2 + \dots + X_n) = n p (1 - p) [1 - 3 p (1 - p)] \\ &\quad + 3 n (n - 1) p^2 (1 - p)^2 = n p (1 - p) [1 + 3 p (1 - p) (n - 2)].\end{aligned}$$

Quindi, ricordando che $\text{Var}(X) = n p (1 - p)$ e facendo uso delle (5.56) e (5.57) si ottengono skewness e curtosi:

$$\alpha_3(X) = \frac{\bar{\mu}_3(X)}{\sigma_X^3} = \frac{1 - 2p}{\sqrt{n p (1 - p)}}, \quad \alpha_4(X) = \frac{\bar{\mu}_4(X)}{\sigma_X^4} = \frac{1 + 3 p (1 - p) (n - 2)}{n p (1 - p)}.$$

Si noti che per $p = 1/2$ la skewness è nulla. Inoltre, al crescere di n la skewness tende a 0 e la curtosi tende a 3, valori che coincidono con skewness e curtosi di una variabile aleatoria normale. \diamond

La proprietà (5.71) della covarianza può essere estesa al calcolo della covarianza di combinazioni lineari di variabili aleatorie, come mostrato nel seguente teorema.

Teorema 5.16 *Sia (X, Y) un vettore aleatorio e siano X_1, X_2, \dots, X_m e Y_1, Y_2, \dots, Y_n variabili aleatorie tali che*

$$X = a_1 X_1 + a_2 X_2 + \dots + a_m X_m, \quad Y = b_1 Y_1 + b_2 Y_2 + \dots + b_n Y_n,$$

con a_1, a_2, \dots, a_m e b_1, b_2, \dots, b_n costanti reali. Allora, se $\text{Cov}(X, Y)$ esiste, risulta:

$$\text{Cov}(X, Y) = \text{Cov} \left[\sum_{r=1}^m a_r X_r, \sum_{s=1}^n b_s Y_s \right] = \sum_{r=1}^m \sum_{s=1}^n a_r b_s \text{Cov}(X_r, Y_s). \quad (5.94)$$

Dimostrazione Dalla Definizione 5.15 e dalla (5.37) si ha:

$$\begin{aligned}\text{Cov}(X, Y) &= E \left[\left(\sum_{r=1}^m a_r X_r - \sum_{r=1}^m a_r E(X_r) \right) \left(\sum_{s=1}^n b_s Y_s - \sum_{s=1}^n b_s E(Y_s) \right) \right] \\ &= \sum_{r=1}^m \sum_{s=1}^n a_r b_s E[(X_r - E(X_r)) (Y_s - E(Y_s))],\end{aligned}$$

ossia la (5.94). \square

Esempio 5.35 Consideriamo un esperimento costituito da n prove ripetute indipendenti ed assumiamo che in ciascuna prova i risultati di interesse siano $k > 2$ eventi necessari ed incompatibili, che denoteremo con B_1, B_2, \dots, B_k , caratterizzati rispettivamente da probabilità $p_i = P(B_i) > 0$ ($i = 1, 2, \dots, k$). Sia $\mathbf{X} = (X_1, X_2, \dots, X_k)$ il vettore aleatorio la cui generica componente X_i ($i = 1, 2, \dots, k$) rappresenta il numero di volte in cui si è verificato l'evento B_i nelle n prove. Come mostrato in (4.25), risulta $\mathbf{X} \sim \mathcal{M}(n, p_1, p_2, \dots, p_k)$. Calcoliamo sia il valore medio e la varianza del numero di volte in cui si è verificato l'evento B_i ($i = 1, 2, \dots, k$) nelle n prove, sia la $\text{Cov}(X_i, X_j)$ ($i, j = 1, 2, \dots, k, i \neq j$).

Nella Proposizione 4.8 abbiamo mostrato che se $\mathbf{X} \sim \mathcal{M}(n, p_1, p_2, \dots, p_k)$, allora $X_i \sim \mathcal{B}(n, p_i)$ ($i = 1, 2, \dots, k$), ossia le distribuzioni marginali di ciascuna componente X_i sono binomiali di parametri n e p_i . Pertanto si ha:

$$E(X_i) = n p_i, \quad \text{Var}(X_i) = n p_i (1 - p_i) \quad (i = 1, 2, \dots, k).$$

Inoltre nella Proposizione 4.9 abbiamo mostrato che se $\mathbf{X} \sim \mathcal{M}(n, p_1, p_2, \dots, p_k)$, allora risulta $(X_i, X_j) \sim \mathcal{M}(n, p_i, p_j, 1 - p_i - p_j)$ ($i, j = 1, 2, \dots, k; i \neq j$), ossia le distribuzioni marginali delle coppie (X_i, X_j) sono multinomiali di parametri $n, p_i, p_j, 1 - p_i - p_j$. Denotiamo ora con X_{ir} una variabile aleatoria discreta che può assumere due soli valori: 0 (interpretabile come insuccesso dovuto al verificarsi di un evento diverso da B_i alla prova r -esima) e 1 (interpretabile come successo dovuto al verificarsi dell'evento B_i alla prova r -esima). È evidente che

$$X_i = X_{i1} + X_{i2} + \dots + X_{in} \quad (i = 1, 2, \dots, k),$$

con $X_{ir} \sim \mathcal{B}(1, p_i)$. Dal Teorema 5.16 si ricava:

$$\begin{aligned} \text{Cov}(X_i, X_j) &= \text{Cov}\left[\sum_{r=1}^n X_{ir}, \sum_{s=1}^n X_{js}\right] = \sum_{r=1}^n \sum_{s=1}^n \text{Cov}(X_{ir}, X_{js}) \\ &\quad (i, j = 1, 2, \dots, k; i \neq j). \end{aligned}$$

Se $r \neq s$ risulta $\text{Cov}(X_{ir}, X_{js}) = 0$; infatti, se $r \neq s$, X_{ir} e X_{js} sono indipendenti poiché corrispondenti a prove differenti (supposte indipendenti per ipotesi). Risulta quindi:

$$\text{Cov}(X_i, X_j) = \sum_{r=1}^n \text{Cov}(X_{ir}, X_{jr}) \quad (i, j = 1, 2, \dots, k; i \neq j).$$

Osserviamo ora che $E(X_{ir} X_{jr}) = 0$ poiché almeno una tra X_{ir} e X_{jr} è nulla (nella stessa prova non si possono verificare simultaneamente due eventi distinti); pertanto:

$$\text{Cov}(X_{ir}, X_{jr}) = E(X_{ir} X_{jr}) - E(X_{ir}) E(X_{jr}) = -E(X_{ir}) E(X_{jr}) = -p_i p_j.$$

Si giunge quindi alla relazione

$$\text{Cov}(X_i, X_j) = -n p_i p_j \quad (i, j = 1, 2, \dots, k; i \neq j),$$

che mostra che la covarianza è sempre negativa. Infine, osserviamo che si ha:

$$\varrho(X_i, X_j) = \frac{\text{Cov}(X_i, X_j)}{\sqrt{\text{Var}(X_i)} \sqrt{\text{Var}(X_j)}} = -\sqrt{\frac{p_i p_j}{(1 - p_i)(1 - p_j)}} \quad (i, j = 1, 2, \dots, k; i \neq j).$$

◇

5.8 Funzioni generatrici

Per lo studio di proprietà delle leggi di probabilità di una variabile aleatoria e, in particolare, dei momenti, è utile far ricorso ad opportune funzioni che permettono di ottenere notevoli semplificazioni. Tra queste, particolare importanza rivestono la *funzione generatrice dei momenti* e la *funzione generatrice di probabilità*. Introduciamo in primo luogo la funzione generatrice dei momenti.

5.8.1 Funzione generatrice dei momenti

Definizione 5.17 Data una variabile aleatoria X , la funzione

$$M_X(s) = E(e^{sX}) \quad (5.95)$$

è detta *funzione generatrice dei momenti* se il valore medio al secondo membro è finito almeno in un intorno dell'origine, $|s| < s_0$, con $s_0 > 0$.

Poiché e^{sX} è una variabile aleatoria non negativa per ogni s reale, dalla Proposizione 5.3 segue che $E(e^{sX})$ esiste sempre, anche se può risultare divergente per certi valori di s . Quindi la funzione generatrice dei momenti esiste per ogni s reale. Per l'effettiva utilizzazione di questa funzione occorre richiedere che detto valore medio sia finito almeno in un intorno dell'origine. Per alcune variabili aleatorie tale valore medio è finito per tutti gli s reali, per altre è finito in un fissato intorno dell'origine e per altre diverge per ogni reale s .

Se X è una variabile aleatoria discreta con funzione di probabilità $p_X(x)$ la (5.95) diventa:

$$M_X(s) = E(e^{sX}) = \sum_{r: x_r \in S} e^{s x_r} p_X(x_r), \quad (5.96)$$

dove S denota l'insieme dei valori assunti da X con probabilità non nulla. Se X è assolutamente continua con densità di probabilità $f_X(x)$, la (5.95) si scrive:

$$M_X(s) = E(e^{sX}) = \int_{-\infty}^{+\infty} e^{sx} f_X(x) dx. \quad (5.97)$$

Teorema 5.17 Sia X una variabile aleatoria con funzione generatrice dei momenti $M_X(s)$ per $|s| \leq s_0$ e siano a, b costanti reali. Risulta:

$$M_{aX+b}(s) = e^{bs} M_X(as). \quad (5.98)$$

Dimostrazione Dalla Definizione 5.17 e dal Teorema 5.2 segue immediatamente la (5.98) avendosi:

$$M_{aX+b}(s) = E[e^{(aX+b)s}] = e^{bs} E[e^{(as)X}] = e^{bs} M_X(as).$$

□

Teorema 5.18 Sia $M_X(s)$ la funzione generatrice dei momenti di una variabile aleatoria X . Se $M_X(s) < +\infty$ per $|s| \leq s_0$, allora tutti i momenti di X esistono finiti e risulta:

$$M_X(s) = \sum_{n=0}^{+\infty} \mu_n \frac{s^n}{n!}, \quad |s| \leq s_0 \quad (5.99)$$

dove

$$\mu_n = E(X^n) = \left. \frac{d^n}{ds^n} M_X(s) \right|_{s=0} \quad (n = 0, 1, \dots). \quad (5.100)$$

Dimostrazione Per ogni u reale risulta $e^u > 0$, così che $e^{|u|} \leq e^{|u|} + e^{-|u|} = e^u + e^{-u}$. Per ogni $|s| \leq s_0$ si ha pertanto:

$$E[e^{|sX|}] \leq E[e^{sX}] + E[e^{-sX}] = M_X(s) + M_X(-s) < +\infty. \quad (5.101)$$

Inoltre, poiché per ogni u reale risulta:

$$\frac{|u|^n}{n!} \leq \sum_{r=0}^{+\infty} \frac{|u|^r}{r!} = e^{|u|} \quad (n = 1, 2, \dots),$$

facendo uso della (5.101) segue:

$$\frac{|s|^n}{n!} E(|X|^n) = E\left[\frac{|sX|^n}{n!}\right] \leq E[e^{|sX|}] < +\infty \quad (n = 1, 2, \dots),$$

così che $E(|X|^n) < +\infty$ ($n = 1, 2, \dots$). Pertanto tutti i momenti della variabile aleatoria $|X|$ esistono finiti e, conseguentemente, per la Proposizione 5.4 esistono finiti tutti i momenti di X . Per dimostrare la (5.99) si noti che per ogni u reale risulta:

$$\left| \sum_{n=0}^k \frac{u^n}{n!} \right| \leq \sum_{n=0}^k \frac{|u|^n}{n!} \leq e^{|u|} \quad (k = 1, 2, \dots).$$

Pertanto, ricordando (5.101), segue che per ogni $|s| < s_0$ è possibile applicare il teorema della convergenza dominata. Applicando tale teorema si ha che se X è discreta con funzione di probabilità $p_X(x)$ risulta:

$$\begin{aligned} M_X(s) &= \sum_{r: x_r \in S} e^{sx_r} p_X(x_r) = \sum_{r: x_r \in S} \left[\lim_{k \rightarrow +\infty} \sum_{n=0}^k \frac{(sx_r)^n}{n!} \right] p_X(x_r) \\ &= \lim_{k \rightarrow +\infty} \sum_{r: x_r \in S} \left[\sum_{n=0}^k \frac{(sx_r)^n}{n!} \right] p_X(x_r) \\ &= \lim_{k \rightarrow +\infty} \sum_{n=0}^k \frac{s^n}{n!} \sum_{r: x_r \in S} x_r^n p_X(x_r) = \sum_{n=0}^{+\infty} \mu_n \frac{s^n}{n!}, \end{aligned}$$

dove con S si è denotato l'insieme dei valori per i quali si ha $p_X(x) > 0$; mentre se X è assolutamente continua con densità $f_X(x)$ si ha:

$$\begin{aligned} M_X(s) &= \int_{-\infty}^{+\infty} e^{sx} f_X(x) dx = \int_{-\infty}^{+\infty} \left[\lim_{k \rightarrow +\infty} \sum_{n=0}^k \frac{(sx)^n}{n!} \right] f_X(x) dx \\ &= \lim_{k \rightarrow +\infty} \int_{-\infty}^{+\infty} \left[\sum_{n=0}^k \frac{(sx)^n}{n!} \right] f_X(x) dx \\ &= \lim_{k \rightarrow +\infty} \sum_{n=0}^k \frac{s^n}{n!} \int_{-\infty}^{+\infty} x^n f_X(x) dx = \sum_{n=0}^{+\infty} \mu_n \frac{s^n}{n!}. \end{aligned}$$

La validità della (5.99) è così dimostrata. Infine, la (5.100) segue direttamente ricordando che lo sviluppo di una funzione in serie di potenze è unico. \square

Il nome attribuito alla funzione $M_X(s)$ trae origine dalla relazione (5.100) che mostra che il coefficiente di $s^n/n!$ è proprio il momento $E(X^n)$.

Si noti che dal Teorema 5.18 si deduce che esistono variabili aleatorie la cui funzione generatrice dei momenti non è finita in un intorno dell'origine e in tal caso i momenti non possono essere ottenuti tramite la (5.100). È inoltre possibile dimostrare che se in un intorno dell'origine due variabili aleatorie X e Y hanno la stessa funzione generatrice dei momenti finita, allora esse possiedono la stessa funzione di distribuzione. Quando la funzione generatrice dei momenti è finita in un intorno dell'origine, essa individua univocamente la funzione di distribuzione.

Un'importante proprietà della funzione generatrice dei momenti è espressa dal seguente

Teorema 5.19 *Siano X_1, X_2, \dots, X_n variabili aleatorie indipendenti aventi funzioni generatrici dei momenti finite in un intorno dell'origine e sia $X = X_1 + X_2 + \dots + X_n$. Risulta:*

$$M_X(s) = M_{X_1+X_2+\dots+X_n}(s) = M_{X_1}(s) M_{X_2}(s) \cdots M_{X_n}(s). \quad (5.102)$$

Dimostrazione Dall'ipotesi di indipendenza delle variabili X_1, X_2, \dots, X_n segue che anche le variabili aleatorie e^{sX_i} per $i = 1, 2, \dots, n$ sono indipendenti (v. Teorema 3.7). Pertanto, facendo uso del Teorema 5.9, si ha:

$$\begin{aligned} M_X(s) &= M_{X_1+X_2+\dots+X_n}(s) = E\left[e^{s(X_1+X_2+\dots+X_n)}\right] \\ &= E(e^{sX_1}) E(e^{sX_2}) \cdots E(e^{sX_n}), \end{aligned}$$

che conduce alla (5.102). □

Occorre sottolineare che $M_X(s) = E(e^{sX})$ è definita soltanto se il valore medio è finito almeno in un intorno dell'origine. Quando ciò non si verifica si rende necessario definire un'altra funzione, chiamata *funzione caratteristica*, ossia $H_X(u) = E(e^{iuX})$, avendo indicato con i l'unità immaginaria. La funzione caratteristica è quindi una funzione complessa di variabile reale; essa gode della proprietà di essere finita per ogni $u \in \mathbb{R}$. Si può dimostrare, ma la questione eccede i limiti di questa trattazione, che la funzione caratteristica e la funzione di distribuzione sono in corrispondenza biunivoca, nel senso che ad ogni funzione di distribuzione corrisponde una funzione caratteristica ed ogni funzione caratteristica proviene da una sola funzione di distribuzione. Quindi, la funzione caratteristica svolge un ruolo più generale, anche se simile a quello della funzione generatrice dei momenti quando quest'ultima è finita almeno in un intorno dell'origine.

5.8.2 Funzioni generatrici di probabilità

Introduciamo ora la cosiddetta funzione generatrice di probabilità.

Definizione 5.18 *Sia X una variabile aleatoria discreta che assume solo valori interi non negativi in numero finito o numerabile con funzione di probabilità $\{r, p_X(r); r = 0, 1, \dots\}$. La funzione*

$$G_X(z) = E(z^X) = \sum_{r=0}^{+\infty} z^r p_X(r) \quad (5.103)$$

è detta *funzione generatrice di probabilità* quando il valore medio al secondo membro esiste ed è finito almeno in un intorno dell'unità, ossia per $|z - 1| < \delta$, dove δ denota un qualche reale positivo.

Il nome *funzione generatrice di probabilità* attribuito a $G_X(z)$ deriva dal fatto che per ogni intero r il coefficiente di z^r è la probabilità $p_X(r) = P(X = r)$. Si noti che se la variabile aleatoria assume un numero finito di valori, la somma in (5.103) ha sempre significato. Se invece X assume un'infinità numerabile di valori, la serie in (5.103) converge almeno per $-1 < z \leq 1$. Infatti, per $|z| < 1$ si ha:

$$|G_X(z)| = \left| \sum_{r=0}^{+\infty} z^r p_X(r) \right| \leq \sum_{r=0}^{+\infty} |z|^r p_X(r) < \sum_{r=0}^{+\infty} p_X(r) = 1,$$

mentre per $z = 1$ risulta:

$$G_X(1) = \sum_{r=0}^{+\infty} p_X(r) = 1.$$

Per $|z| < 1$, per le proprietà delle serie di potenze, segue che la serie che definisce la funzione generatrice di probabilità può essere derivata termine a termine.

Teorema 5.20 Per $|z| < 1$ si ha:

$$\frac{d^k G_X(z)}{dz^k} = \sum_{r=k}^{+\infty} r(r-1) \dots (r-k+1) z^{r-k} p_X(r). \quad (5.104)$$

Dimostrazione Dimostriamo la (5.104) per induzione su k . Per $k = 1$ dalla (5.103) segue:

$$\frac{dG_X(z)}{dz} = \sum_{r=1}^{+\infty} r z^{r-1} p_X(r),$$

che coincide con la (5.104) per $k = 1$. Supponiamo che la (5.104) valga per k e dimostriamo che sussiste anche per $k + 1$. Infatti:

$$\frac{d^{k+1} G_X(z)}{dz^{k+1}} = \frac{d}{dz} \left[\frac{d^k G_X(z)}{dz^k} \right] = \sum_{r=k+1}^{+\infty} r(r-1) \dots (r-k+1)(r-k) z^{r-k-1} p_X(r),$$

che coincide con la (5.104) se si sostituisce k con $k + 1$. □

Se $G_X(z)$ è sviluppabile in serie in un intorno dell'unità, allora ivi sussiste la (5.104) da cui segue:

$$G_X^{(k)}(1) = \left. \frac{d^k G_X(z)}{dz^k} \right|_{z=1} = \sum_{r=k}^{+\infty} r(r-1) \dots (r-k+1) p_X(r), \quad (5.105)$$

che mostra che la derivata di ordine k della funzione generatrice di probabilità calcolata per $z = 1$ si può esprimere come combinazione lineare dei momenti della variabile aleatoria

X fino all'ordine k , sempre che $+\infty$ e $-\infty$ non appaiano simultaneamente nella somma al secondo membro. In particolare, si ha:

$$\begin{aligned} E(X) &= G_X^{(1)}(1), \\ E(X^2) &= G_X^{(2)}(1) + G_X^{(1)}(1), \\ \text{Var}(X) &= G_X^{(2)}(1) + G_X^{(1)}(1) - [G_X^{(1)}(1)]^2. \end{aligned}$$

La funzione generatrice di probabilità è spesso utile non solo per la valutazione dei momenti di una variabile aleatoria discreta ma anche per la determinazione della sua funzione di probabilità. Infatti è possibile dimostrare che se in un intorno dell'unità due variabili aleatorie X e Y discrete che assumono valori interi non negativi hanno la stessa funzione generatrice di probabilità finita, allora esse sono caratterizzate dalla stessa funzione di probabilità. Quindi, quando la funzione generatrice di probabilità è finita in un intorno dell'unità, essa individua univocamente la funzione di probabilità. Tale funzione può essere ottenuta espandendo la funzione generatrice di probabilità in serie di potenze.

Un'importante proprietà della funzione generatrice di probabilità è espressa dal seguente

Teorema 5.21 *Se X_1, X_2, \dots, X_n sono variabili aleatorie discrete indipendenti che assumono valori interi non negativi in numero finito o numerabile e che possiedono funzione generatrice di probabilità in un intorno dell'unità, allora anche la somma $X = X_1 + X_2 + \dots + X_n$ possiede funzione generatrice di probabilità e risulta:*

$$G_X(z) = G_{X_1+X_2+\dots+X_n}(z) = G_{X_1}(z) G_{X_2}(z) \cdots G_{X_n}(z). \quad (5.106)$$

Dimostrazione Dalla (5.103), utilizzando la proprietà moltiplicativa della media di un prodotto di variabili aleatorie indipendenti, si ottiene:

$$\begin{aligned} G_X(z) &= G_{X_1+X_2+\dots+X_n}(z) = E(z^{X_1+X_2+\dots+X_n}) = E(z^{X_1} z^{X_2} \dots z^{X_n}) \\ &= E(z^{X_1}) E(z^{X_2}) \dots E(z^{X_n}), \end{aligned}$$

ossia la (5.106). □

Se X una variabile aleatoria discreta che assume valori interi non negativi in numero finito o numerabile, la funzione generatrice dei momenti e la funzione generatrice di probabilità sono tra loro connesse. Sussiste, infatti, la seguente proposizione.

Proposizione 5.22 *Sia X è una variabile aleatoria discreta che assume valori interi non negativi in numero finito o numerabile.*

(a) *Se X possiede funzione generatrice di probabilità finita in un intorno dell'unità, allora la funzione generatrice dei momenti è finita in un intorno dell'origine e risulta $M_X(s) = G_X(e^s)$.*

(b) *Se X possiede funzione generatrice dei momenti finita in un intorno dell'origine, allora la funzione generatrice di probabilità esiste ed è finita in un intorno dell'unità e risulta $G_X(z) = M_X(\ln z)$.*

5.8.3 Esempi di funzioni generatrici

Ci proponiamo di calcolare le funzioni generatrici ed i momenti di alcune variabili aleatorie discrete ed assolutamente continue introdotte nel Capitolo 4.

Distribuzione uniforme discreta Sia $X \sim \mathcal{U}_d(n)$. La sua funzione di probabilità è:

$$p_X(x) = \begin{cases} \frac{1}{n}, & x = 1, 2, \dots, n \\ 0, & \text{altrimenti.} \end{cases}$$

La funzione generatrice dei momenti è pertanto:

$$M_X(s) = \frac{1}{n} \sum_{r=1}^n e^{sr} = \frac{1}{n} \left[\sum_{r=0}^n e^{sr} - 1 \right] = \frac{e^s (1 - e^{sn})}{n (1 - e^s)} \quad (s \in \mathbb{R}).$$

Poiché X è discreta ed assume valori interi non negativi in numero finito, dalla Proposizione 5.22 si ricava la funzione generatrice di probabilità:

$$G_X(z) = M_X(\ln z) = \frac{z (1 - z^n)}{n (1 - z)} \quad (z \in \mathbb{R}).$$

Come mostrato negli Esempi 5.3 e 5.22, risulta:

$$E(X) = \frac{n+1}{2}, \quad E(X^2) = \frac{(n+1)(2n+1)}{6}, \quad \text{Var}(X) = \frac{n^2-1}{12}.$$

Facendo uso della (5.54), si ottiene:

$$C_X = \frac{\sqrt{\text{Var}(X)}}{E(X)} = \sqrt{\frac{n-1}{3(n+1)}},$$

che mostra che il coefficiente di variazione soddisfa la limitazione $0 \leq C_X < 1/\sqrt{3}$.

Distribuzione di Bernoulli Sia $X \sim \mathcal{B}(1, p)$, così che la sua funzione di probabilità è:

$$p_X(x) = \begin{cases} 1-p, & x = 0 \\ p, & x = 1 \\ 0, & \text{altrimenti,} \end{cases}$$

con $0 < p < 1$. Risulta:

$$E(X) = p, \quad E(X^2) = p, \quad \text{Var}(X) = p(1-p)$$

così che il coefficiente di variazione è dato da

$$C_X = \sqrt{\frac{1-p}{p}}.$$

La funzione generatrice dei momenti è:

$$M_X(s) = 1 - p + p e^s \quad (s \in \mathbb{R});$$

dalla Proposizione 5.22 si ricava la funzione generatrice di probabilità:

$$G_X(z) = M_X(\ln z) = 1 - p + pz \quad (z \in \mathbb{R}).$$

Distribuzione binomiale Se $X \sim \mathcal{B}(n, p)$, la funzione di probabilità è la seguente:

$$p_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x = 0, 1, \dots, n \\ 0, & \text{altrimenti,} \end{cases}$$

con $0 < p < 1$ e n intero positivo. Nella Proposizione 4.1 abbiamo mostrato che se X_1, X_2, \dots, X_n sono variabili indipendenti con $X_i \sim \mathcal{B}(1, p)$ per $i = 1, 2, \dots, n$, allora $X = X_1 + X_2 + \dots + X_n$ ha distribuzione binomiale di parametri n e p . Pertanto, dal Teorema 5.19 si ricava:

$$M_X(s) = M_{X_1}(s) M_{X_2}(s) \cdots M_{X_n}(s) = [1 - p + p e^s]^n,$$

mentre per la Proposizione 5.22 risulta:

$$G_X(z) = M_X(\ln z) = [1 - p + p z]^n.$$

Si ottiene così immediatamente:

$$E(X) = np, \quad E(X^2) = np(1-p+np), \quad \text{Var}(X) = np(1-p),$$

da cui segue

$$C_X = \sqrt{\frac{1-p}{np}}.$$

Distribuzione ipergeometrica Sia $X \sim \mathcal{I}(n, m, N - m)$. La sua funzione di probabilità è:

$$p_X(x) = \begin{cases} \frac{\binom{m}{x} \binom{N-m}{n-x}}{\binom{N}{n}}, & \max\{0, n+m-N\} \leq x \leq \min\{n, m\} \\ 0, & \text{altrimenti,} \end{cases}$$

con n, m e N interi tali che $0 \leq n \leq N$ e $0 \leq m \leq N$. Nell'Esempio 5.33 abbiamo mostrato che

$$E(X) = n \frac{m}{N}, \quad \text{Var}(X) = \frac{nm(N-n)(N-m)}{N^2(N-1)}.$$

In questo caso la valutazione delle funzioni generatrici non risulta di particolare utilità pratica.

Distribuzione geometrica Sia X una variabile aleatoria di distribuzione geometrica di parametro p : $X \sim \mathcal{BN}(1, p)$. La sua funzione di probabilità è:

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

con $0 < p < 1$. La funzione generatrice di probabilità è pertanto:

$$G_X(z) = \sum_{r=1}^{+\infty} p(1-p)^{r-1} z^r = p z \sum_{n=0}^{+\infty} [(1-p)z]^n = \frac{p z}{1 - (1-p)z}$$

se $|z| < (1-p)^{-1}$. Dalla Proposizione 5.22 si ottiene la funzione generatrice dei momenti:

$$M_X(s) = G_X(e^s) = \frac{p e^s}{1 - (1-p)e^s} \quad (s < -\ln(1-p)).$$

Essendo $M_X(s)$ finita in un intorno dell'origine, facendo uso della (5.100), si ricava:

$$E(X) = \left. \frac{dM_X(s)}{ds} \right|_{s=0} = \left. \frac{p e^s}{[1 - (1-p)e^s]^2} \right|_{s=0} = \frac{1}{p},$$

$$E(X^2) = \left. \frac{d^2 M_X(s)}{ds^2} \right|_{s=0} = \left. \frac{p e^s [1 + (1-p)e^s]}{[1 - (1-p)e^s]^3} \right|_{s=0} = \frac{2-p}{p^2},$$

da cui segue:

$$\text{Var}(X) = \frac{1-p}{p^2}, \quad C_X = \sqrt{1-p}.$$

Distribuzione di Pascal Sia Y una variabile aleatoria caratterizzata da funzione di probabilità

$$p_Y(y) = \begin{cases} p(1-p)^y, & y = 0, 1, \dots \\ 0, & \text{altrimenti,} \end{cases}$$

con $0 < p < 1$. Si dice che Y ha *distribuzione di Pascal* di parametro p . Essendo $Y = X - 1$, con X di distribuzione geometrica di parametro p , dalla (5.98) si ha:

$$M_Y(s) = M_{X-1}(s) = e^{-s} M_X(s) = \frac{p}{1 - (1-p)e^s} \quad (s < -\ln(1-p)),$$

mentre dalla Proposizione 5.22 segue:

$$G_Y(z) = \frac{p}{1 - (1-p)z} \quad (|z| < (1-p)^{-1}).$$

Inoltre, risulta:

$$E(Y) = E(X - 1) = \frac{1}{p} - 1 = \frac{1-p}{p}, \quad \text{Var}(Y) = \text{Var}(X - 1) = \frac{1-p}{p^2},$$

da cui segue $C_X = (1-p)^{-1/2}$.

Distribuzione binomiale negativa Sia $X \sim \mathcal{BN}(n, p)$ di funzione di probabilità:

$$p_X(x) = \begin{cases} \binom{x-1}{n-1} p^n (1-p)^{x-n}, & x = n, n+1, \dots \\ 0, & \text{altrimenti,} \end{cases}$$

con $0 < p < 1$ e n intero positivo. Nella Proposizione 4.5 si è visto che se X_1, X_2, \dots, X_n sono variabili aleatorie indipendenti con $X_i \sim \mathcal{BN}(1, p)$ per $i = 1, 2, \dots, n$, allora $X = X_1 + X_2 + \dots + X_n$ ha distribuzione binomiale negativa di parametri n e p . Pertanto, dal Teorema 5.19 discende:

$$M_X(s) = M_{X_1}(s) M_{X_2}(s) \cdots M_{X_n}(s) = \left[\frac{p e^s}{1 - (1-p) e^s} \right]^n$$

se $s < -\ln(1-p)$ e dalla Proposizione 5.22:

$$G_X(z) = \left[\frac{p z}{1 - (1-p) z} \right]^n \quad (|z| < (1-p)^{-1}).$$

Si ottiene così immediatamente

$$E(X) = \frac{n}{p}, \quad \text{Var}(X) = \frac{n(1-p)}{p^2}, \quad C_X = \sqrt{\frac{1-p}{n}}.$$

Distribuzione di Poisson Sia $X \sim \mathcal{P}(\lambda)$. La funzione di probabilità è dunque:

$$p_X(x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda}, & x = 0, 1, \dots \\ 0, & \text{altrimenti,} \end{cases}$$

con $\lambda > 0$. Per la funzione generatrice di probabilità si ottiene:

$$G_X(z) = \sum_{r=0}^{+\infty} \frac{\lambda^r}{r!} e^{-\lambda} z^r = e^{-\lambda} \sum_{r=0}^{+\infty} \frac{(\lambda z)^r}{r!} = e^{\lambda(z-1)} \quad (z \in \mathbb{R})$$

e quindi

$$M_X(s) = G_X(e^s) = \exp\{\lambda(e^s - 1)\} \quad (s \in \mathbb{R}).$$

Essendo

$$E(X) = \left. \frac{dM_X(s)}{ds} \right|_{s=0} = \lambda e^s M_X(s) \Big|_{s=0} = \lambda,$$

$$E(X^2) = \left. \frac{d^2 M_X(s)}{ds^2} \right|_{s=0} = \lambda e^s (1 + \lambda e^s) M_X(s) \Big|_{s=0} = \lambda(1 + \lambda),$$

si ricava:

$$\text{Var}(X) = \lambda, \quad C_X = \frac{1}{\sqrt{\lambda}}.$$

Proposizione 5.23 Se X_1, X_2, \dots, X_n sono variabili indipendenti con $X_i \sim \mathcal{P}(\lambda_i)$ per $i = 1, 2, \dots, n$, allora $Y = X_1 + X_2 + \dots + X_n$ ha distribuzione di Poisson di parametro $\lambda = \lambda_1 + \lambda_2 + \dots + \lambda_n$.

Dimostrazione Tale risultato può essere agevolmente dimostrato osservando che dal Teorema 5.21 si ha:

$$\begin{aligned} G_Y(z) &= G_{X_1}(z) G_{X_2}(z) \dots G_{X_n}(z) = e^{\lambda_1(z-1)} e^{\lambda_2(z-1)} \dots e^{\lambda_n(z-1)} \\ &= e^{(\lambda_1 + \lambda_2 + \dots + \lambda_n)(z-1)}. \end{aligned}$$

Poiché $G_Y(z)$ è la funzione generatrice di probabilità di una variabile aleatoria con distribuzione di Poisson di parametro $\lambda = \lambda_1 + \lambda_2 + \dots + \lambda_n$, risulta $Y \sim \mathcal{P}(\lambda)$. \square

Distribuzione uniforme Sia $X \sim \mathcal{U}(a, b)$. La sua densità di probabilità è:

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{altrimenti} \end{cases}$$

con a e b numeri reali. La funzione generatrice dei momenti è:

$$M_X(s) = \frac{1}{b-a} \int_a^b e^{sx} dx = \frac{e^{sb} - e^{sa}}{s(b-a)} \quad (s \in \mathbb{R}).$$

Come mostrato negli Esempi 5.9 e 5.23 risulta:

$$E(X) = \frac{a+b}{2}, \quad E(X^2) = \frac{a^2 + ab + b^2}{3}, \quad \text{Var}(X) = \frac{(b-a)^2}{12}.$$

Quindi, il coefficiente di variazione è

$$C_X = \frac{b-a}{\sqrt{3}(a+b)}.$$

Distribuzione esponenziale Sia $X \sim \mathcal{E}(1, \lambda)$. Dalla densità di probabilità

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{altrimenti,} \end{cases}$$

con $\lambda > 0$, si ricava la funzione generatrice dei momenti:

$$M_X(s) = \lambda \int_0^{+\infty} e^{-x(\lambda-s)} dx = \frac{\lambda}{\lambda-s} \quad (s < \lambda).$$

Sviluppando questa in serie di potenze di s si ottiene:

$$M_X(s) = \frac{1}{1-s/\lambda} = \sum_{k=0}^{+\infty} \left(\frac{s}{\lambda}\right)^k = \sum_{k=0}^{+\infty} k! \left(\frac{1}{\lambda}\right)^k \frac{s^k}{k!},$$

da cui, ricordando la (5.99), segue:

$$E(X^k) = k! \left(\frac{1}{\lambda}\right)^k \quad (k = 0, 1, \dots).$$

In particolare, risulta:

$$E(X) = \frac{1}{\lambda}, \quad E(X^2) = 2 \left(\frac{1}{\lambda}\right)^2, \quad \text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{1}{\lambda^2}.$$

Si noti che $C_X = 1$, così che il coefficiente di variazione di una generica variabile aleatoria si può interpretare come una misura della sua deviazione dalla variabile esponenziale.

Distribuzione di Erlang Sia $X \sim \mathcal{E}(n, \lambda)$. La sua densità di probabilità è pertanto:

$$f_X(x) = \begin{cases} \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x}, & x > 0 \\ 0, & \text{altrimenti,} \end{cases}$$

con $\lambda > 0$ e n intero positivo. Come si evince dalla Proposizione 4.15, se X_1, X_2, \dots, X_n sono variabili aleatorie indipendenti con $X \sim \mathcal{E}(1, \lambda)$ per $i = 1, 2, \dots, n$, allora $X = X_1 + X_2 + \dots + X_n$ ha distribuzione di Erlang di parametri n e λ . Pertanto, dal Teorema 5.19 segue:

$$M_X(s) = M_{X_1}(s) M_{X_2}(s) \cdots M_{X_n}(s) = \left(\frac{\lambda}{\lambda - s}\right)^n \quad (s < \lambda).$$

Inoltre, risulta:

$$E(X) = \frac{n}{\lambda}, \quad \text{Var}(X) = \frac{n}{\lambda^2}, \quad C_X = \frac{1}{\sqrt{n}}.$$

Si noti che il coefficiente di variazione è una funzione decrescente in n che tende a zero al crescere di n . Inoltre, se è $n > 1$, risulta $0 < C_X < 1$.

Distribuzione gamma Sia $X \sim \mathcal{G}(\nu, \lambda)$ di densità di probabilità

$$f_X(x) = \begin{cases} \frac{\lambda^\nu x^{\nu-1}}{\Gamma(\nu)} e^{-\lambda x}, & x > 0 \\ 0, & \text{altrimenti,} \end{cases}$$

con λ, ν reali positivi e con $\Gamma(\nu) = \int_0^{+\infty} e^{-x} x^{\nu-1} dx$. La funzione generatrice dei momenti di X è:

$$\begin{aligned} M_X(s) &= \int_0^{+\infty} e^{sx} \frac{\lambda^\nu x^{\nu-1}}{\Gamma(\nu)} e^{-\lambda x} dx = \frac{\lambda^\nu}{\Gamma(\nu)} \int_0^{+\infty} x^{\nu-1} e^{(s-\lambda)x} dx \\ &= \frac{\lambda^\nu}{\Gamma(\nu)} (\lambda - s)^{-\nu} \int_0^{+\infty} e^{-y} y^{\nu-1} dy = \left(\frac{\lambda}{\lambda - s}\right)^\nu \quad (s < \lambda). \end{aligned}$$

Si noti che per $\nu = n$, $M_X(s)$ coincide con la funzione generatrice dei momenti di una variabile di Erlang di parametri n e λ . Più in generale, si ha:

$$E(X) = \frac{dM_X(s)}{ds} \Big|_{s=0} = \frac{\nu}{\lambda} \left(\frac{\lambda}{\lambda - s} \right)^{\nu+1} \Big|_{s=0} = \frac{\nu}{\lambda},$$

$$E(X^2) = \frac{d^2 M_X(s)}{ds^2} \Big|_{s=0} = M_X^{(2)}(0) = \frac{\nu(\nu+1)}{\lambda^2} \left(\frac{\lambda}{\lambda - s} \right)^{\nu+2} \Big|_{s=0} = \frac{\nu(\nu+1)}{\lambda^2},$$

da cui segue:

$$\text{Var}(X) = \frac{\nu}{\lambda^2}, \quad C_X = \frac{1}{\sqrt{\nu}}.$$

Il coefficiente di variazione di X è minore di 1 per $\nu > 1$, uguale a 1 per $\nu = 1$ e maggiore di 1 per $\nu < 1$.

Distribuzione iperesponenziale Sia X una variabile aleatoria di densità iperesponenziale di parametri $a_1, a_2, \dots, a_n, \lambda_1, \lambda_2, \dots, \lambda_n$, quindi di densità:

$$f_X(x) = \begin{cases} \sum_{i=1}^n a_i \lambda_i e^{-\lambda_i x}, & x > 0 \\ 0, & \text{altrimenti,} \end{cases}$$

con $\lambda_i > 0$ ($i = 1, 2, \dots, n$) e con a_1, a_2, \dots, a_n numeri reali tali che $a_i \geq 0$ ($i = 1, 2, \dots, n$) e $\sum_{i=1}^n a_i = 1$. La funzione generatrice dei momenti è:

$$M_X(s) = \int_0^{+\infty} e^{sx} \sum_{i=1}^n a_i \lambda_i e^{-\lambda_i x} dx = \sum_{i=1}^n a_i \int_0^{+\infty} e^{sx} \lambda_i e^{-\lambda_i x} dx = \sum_{i=1}^n \frac{a_i \lambda_i}{\lambda_i - s}$$

se $s < \min(\lambda_1, \lambda_2, \dots, \lambda_n)$. Inoltre, si ha:

$$E(X) = \frac{dM_X(s)}{ds} \Big|_{s=0} = \sum_{i=1}^n \frac{a_i \lambda_i}{(\lambda_i - s)^2} \Big|_{s=0} = \sum_{i=1}^n \frac{a_i}{\lambda_i},$$

$$E(X^2) = \frac{d^2 M_X(s)}{ds^2} \Big|_{s=0} = 2 \sum_{i=1}^n \frac{a_i \lambda_i}{(\lambda_i - s)^3} \Big|_{s=0} = 2 \sum_{i=1}^n \frac{a_i}{\lambda_i^2},$$

da cui segue:

$$\text{Var}(X) = 2 \sum_{i=1}^n \frac{a_i}{\lambda_i^2} - \left(\sum_{i=1}^n \frac{a_i}{\lambda_i} \right)^2.$$

Vogliamo ora mostrare che per ogni $n > 1$ risulta $C_X \geq 1$, ossia che $\text{Var}(X) \geq [E(X)]^2$. A tal fine utilizzeremo la disuguaglianza algebrica di Cauchy-Schwarz-Bunyakowsky:

$$\left(\sum_{i=1}^n x_i y_i \right)^2 \leq \left(\sum_{i=1}^n x_i^2 \right) \left(\sum_{i=1}^n y_i^2 \right),$$

con x_1, x_2, \dots, x_n e y_1, y_2, \dots, y_n arbitrari numeri reali. Ponendo $x_i = \sqrt{a_i}$ e $y_i = \sqrt{a_i}/\lambda_i$ per $i = 1, 2, \dots, n$, la disuguaglianza di Cauchy-Schwarz-Bunyakowsky diventa:

$$\left(\sum_{i=1}^n \frac{a_i}{\lambda_i} \right)^2 \leq \sum_{i=1}^n a_i \sum_{i=1}^n \frac{a_i}{\lambda_i^2} = \sum_{i=1}^n \frac{a_i}{\lambda_i^2}.$$

Pertanto,

$$\text{Var}(X) - [E(X)]^2 = 2 \left[\sum_{i=1}^n \frac{a_i}{\lambda_i^2} - \left(\sum_{j=1}^n \frac{a_j}{\lambda_j} \right)^2 \right] \geq 0,$$

e quindi $C_X^2 \geq 1$. Si noti che $C_X = 1$ se e solo se $\lambda_1 = \lambda_2 = \dots = \lambda_n$, ossia se la densità di probabilità di X è esponenziale.

Distribuzione di Weibull Sia X una variabile aleatoria di densità di probabilità di Weibull di parametri λ e α :

$$f_X(x) = \begin{cases} \alpha \lambda x^{\alpha-1} \exp\{-\lambda x^\alpha\}, & x > 0 \\ 0, & \text{altrimenti,} \end{cases}$$

con $\alpha > 0$ e $\lambda > 0$. La funzione generatrice dei momenti è data da:

$$\begin{aligned} M_X(s) &= \alpha \lambda \int_0^{+\infty} e^{sx} x^{\alpha-1} \exp\{-\lambda x^\alpha\} dx \\ &= \alpha \lambda \sum_{k=0}^{+\infty} \frac{s^k}{k!} \int_0^{+\infty} x^{\alpha+k-1} \exp\{-\lambda x^\alpha\} dx \\ &= \sum_{k=0}^{+\infty} \frac{s^k}{k!} \lambda^{-k/\alpha} \int_0^{+\infty} y^{k/\alpha} e^{-y} dy = \sum_{k=0}^{+\infty} \frac{s^k}{k!} \frac{\Gamma\left(1 + \frac{k}{\alpha}\right)}{\lambda^{k/\alpha}}, \end{aligned}$$

ottenuta ricorrendo prima all'espansione in serie di Taylor di e^{sx} , effettuando successivamente il cambiamento di variabile di integrazione $y = \lambda x^\alpha$ e ricordando, infine, la definizione della funzione gamma. Ricordando la (5.99), si ha:

$$E(X^n) = \frac{\Gamma\left(1 + \frac{n}{\alpha}\right)}{\lambda^{n/\alpha}} \quad (n = 0, 1, \dots).$$

Si noti che per $\alpha = 1$ la densità di probabilità Weibull diventa una densità esponenziale di parametro λ e che $E(X^n) = n!/\lambda^n$.

Distribuzione normale Sia $X \sim \mathcal{N}(\mu, \sigma)$ di densità di probabilità

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

con $\mu \in \mathbb{R}$ e $\sigma > 0$. La funzione generatrice dei momenti è data da:

$$\begin{aligned} M_X(s) &= \int_{-\infty}^{+\infty} e^{sx} \frac{1}{\sigma \sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{(\sigma y + \mu)s} \exp\left\{-\frac{y^2}{2}\right\} dy \\ &= \frac{1}{\sqrt{2\pi}} \exp\left\{\mu s + \frac{\sigma^2}{2} s^2\right\} \int_{-\infty}^{+\infty} \exp\left\{-\frac{(y-\sigma s)^2}{2}\right\} dy \\ &= \exp\left\{\mu s + \frac{\sigma^2}{2} s^2\right\}, \end{aligned}$$

ottenuta effettuando prima il cambiamento di variabile di integrazione $y = (x - \mu)/\sigma$ ed osservando poi che l'integrale su \mathbb{R} di una densità normale è unitario. Inoltre, si ottiene:

$$\begin{aligned} E(X) &= \left. \frac{dM_X(s)}{ds} \right|_{s=0} = (\mu + \sigma^2 s) M_X(s) \Big|_{s=0} = \mu, \\ E(X^2) &= \left. \frac{d^2 M_X(s)}{ds^2} \right|_{s=0} = \sigma^2 M_X(s) + (\mu + \sigma^2 s) \left. \frac{dM_X(s)}{ds} \right|_{s=0} = \sigma^2 + \mu^2, \end{aligned}$$

così che

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

Per $\mu \neq 0$, il coefficiente di correlazione è $C_X = \sigma/\mu$; inoltre si ha $C_X < 1$ per $\mu > \sigma$, $C_X = 1$ per $\mu = \sigma$ ed, infine, $C_X > 1$ per $\mu < \sigma$.

Da quanto appena dimostrato scaturiscono i seguenti importanti risultati.

Proposizione 5.24 *Se X_1, X_2, \dots, X_n sono variabili indipendenti con $X_i \sim \mathcal{N}(\mu_i, \sigma_i)$ per $i = 1, 2, \dots, n$ e se a_1, a_2, \dots, a_n sono numeri reali, allora $Y = a_1 X_1 + a_2 X_2 + \dots + a_n X_n$ ha distribuzione normale di parametri μ e σ , dove $\mu = a_1 \mu_1 + a_2 \mu_2 + \dots + a_n \mu_n$ e $\sigma^2 = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \dots + a_n^2 \sigma_n^2$.*

Dimostrazione Tale risultato può essere agevolmente dimostrato osservando che dal Teorema 5.19 si trae:

$$\begin{aligned} M_{Y_n}(s) &= M_{X_1}(a_1 s) M_{X_2}(a_2 s) \cdots M_{X_n}(a_n s) \\ &= \exp\left\{a_1 \mu_1 s + \frac{a_1^2 \sigma_1^2}{2} s^2\right\} \exp\left\{a_2 \mu_2 s + \frac{a_2^2 \sigma_2^2}{2} s^2\right\} \cdots \exp\left\{a_n \mu_n s + \frac{a_n^2 \sigma_n^2}{2} s^2\right\} \\ &= \exp\left\{(a_1 \mu_1 + a_2 \mu_2 + \dots + a_n \mu_n) s + \frac{1}{2} (a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \dots + a_n^2 \sigma_n^2) s^2\right\}. \end{aligned}$$

Essendo $M_Y(s)$ la funzione generatrice dei momenti di una variabile aleatoria con distribuzione normale di parametri μ e σ , dove $\mu = a_1 \mu_1 + a_2 \mu_2 + \dots + a_n \mu_n$ e $\sigma^2 = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \dots + a_n^2 \sigma_n^2$, si ha $Y \sim \mathcal{N}(\mu, \sigma)$. \square

In particolare, se X_1, X_2, \dots, X_n sono variabili aleatorie indipendenti ognuna con distribuzione normale di parametri μ e σ , allora la media campionaria

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

ha distribuzione normale di parametri μ e σ/\sqrt{n} . Da ciò segue che $E(\bar{X}) = \mu$ e $\text{Var}(\bar{X}) = \sigma^2/n$.

Distribuzione di Laplace Sia X una variabile aleatoria di densità di probabilità:

$$f_X(x) = \frac{1}{2\beta} \exp\left\{-\frac{|x - \alpha|}{\beta}\right\}, \quad x \in \mathbb{R},$$

con $\alpha \in \mathbb{R}$ e $\beta > 0$. In tal caso si dice che X ha *distribuzione di Laplace* di parametri α e β . La funzione generatrice dei momenti è:

$$\begin{aligned} M_X(s) &= \int_{-\infty}^{+\infty} e^{sx} \frac{1}{2\beta} \exp\left\{-\frac{|x - \alpha|}{\beta}\right\} dx \\ &= \frac{1}{2\beta} \int_{-\infty}^{\alpha} e^{sx} \exp\left\{-\frac{(\alpha - x)}{\beta}\right\} dx + \frac{1}{2\beta} \int_{\alpha}^{+\infty} e^{sx} \exp\left\{-\frac{(x - \alpha)}{\beta}\right\} dx \\ &= \frac{1}{2\beta} \int_0^{+\infty} e^{s(\alpha - y)} \exp\left\{-\frac{y}{\beta}\right\} dy + \frac{1}{2\beta} \int_0^{+\infty} e^{s(\alpha + y)} \exp\left\{-\frac{y}{\beta}\right\} dy \\ &= \frac{e^{\alpha s}}{2} \frac{1}{1 + \beta s} + \frac{e^{\alpha s}}{2} \frac{1}{1 - \beta s} = \frac{e^{\alpha s}}{1 - \beta^2 s^2} \quad \left(|s| < \frac{1}{\beta}\right). \end{aligned}$$

Inoltre, si ottiene:

$$E(X) = \left. \frac{d M_X(s)}{ds} \right|_{s=0} = \alpha, \quad E(X^2) = E(X^2) = \left. \frac{d^2 M_X(s)}{ds^2} \right|_{s=0} = \alpha^2 + 2\beta^2,$$

così che si ha:

$$\text{Var}(X) = 2\beta^2, \quad C_X = \frac{\beta\sqrt{2}}{\alpha}.$$

Capitolo 6

Distribuzioni e momenti condizionati

6.1 Introduzione

Lo studio delle relazioni sussistenti tra variabili aleatorie richiede l'introduzione del concetto di *distribuzione condizionata*. Nel Capitolo 2 si è parlato di probabilità condizionata per eventi che appartengono ad uno stesso spazio campione. In particolare si è visto (v. Definizione 2.11) che, dati gli eventi A e B , con $P(B) > 0$, la probabilità dell'evento A condizionata dal verificarsi dell'evento B è per definizione:

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

Qui si desidera estendere il concetto di probabilità condizionata al caso di variabili aleatorie; ciò viene effettuato attraverso la definizione delle distribuzioni condizionate. In seguito si parlerà anche di *momenti condizionati*, ossia dei momenti di una distribuzione condizionata.

6.2 Distribuzioni condizionate per variabili discrete

Sia (X, Y) un vettore aleatorio discreto di funzione di probabilità congiunta $p_{X,Y}(x, y)$ e siano $p_X(x)$ e $p_Y(y)$, rispettivamente, le funzioni di probabilità marginali di X e di Y . Denotiamo con D_X l'insieme, finito o numerabile, degli $x \in \mathbb{R}$ tali che $p_X(x) > 0$, e con D_Y l'insieme, finito o numerabile, degli $y \in \mathbb{R}$ tali che $p_Y(y) > 0$. Ricordando la Definizione 2.11, non è difficile pervenire alla funzione di probabilità condizionata nel caso di variabili aleatorie discrete purché ci si limiti a considerare quei valori della variabile condizionante Y caratterizzati da probabilità positiva.

Definizione 6.1 Se (X, Y) è un vettore aleatorio discreto di funzione di probabilità congiunta $p_{X,Y}(x, y)$, allora per ogni $y \in D_Y$ la funzione di probabilità della variabile aleatoria X condizionata da $Y = y$ è:

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x, y)}{p_Y(y)} \quad (x \in \mathbb{R}). \quad (6.1)$$

Per ogni $y \in D_Y$, $p_{X|Y}(\cdot|y)$ definisce una funzione di probabilità, così che risulta:

$$p_{X|Y}(x|y) \geq 0, \quad \sum_{\{r: x_r \in D_X\}} p_{X|Y}(x_r|y) = 1. \quad (6.2)$$

Infatti, essendo $p_{X,Y}(x, y)$ non negativa e $p_Y(y)$ positiva, dalla (6.1) segue che $p_{X|Y}(x|y) \geq 0$. Si noti inoltre che se $x \notin D_X$, risulta $p_{X,Y}(x, y) \leq p_X(x) = 0$, da cui si deduce in tal caso $p_{X|Y}(x|y) = 0$. Inoltre, ricordando che la funzione di probabilità marginale di Y viene ottenuta sommando la probabilità congiunta di X e Y su tutti i valori possibili di X , dalla (6.1) si ha:

$$\sum_{\{r: x_r \in D_X\}} p_{X|Y}(x_r|y) = \frac{1}{p_Y(y)} \sum_{\{r: x_r \in D_X\}} p_{X,Y}(x_r, y) = \frac{p_Y(y)}{p_Y(y)} = 1.$$

Per ogni $y \in D_Y$ è anche possibile definire la funzione di distribuzione condizionata di X dato $Y = y$, come qui di seguito specificato.

Definizione 6.2 Sia (X, Y) un vettore aleatorio discreto di funzione di probabilità congiunta $p_{X,Y}(x, y)$. Per ogni $y \in D_Y$ la funzione di distribuzione condizionata di X dato $Y = y$ è:

$$F_{X|Y}(x|y) = P(X \leq x|Y = y) \quad (x \in \mathbb{R}). \quad (6.3)$$

La funzione di distribuzione condizionata di X dato $Y = y$ può essere calcolata a partire dalla (6.3); infatti, per ogni $y \in D_Y$ si ha:

$$\begin{aligned} F_{X|Y}(x|y) &= \frac{P(X \leq x, Y = y)}{p_Y(y)} = \sum_{\{r: x_r \leq x, x_r \in D_X\}} \frac{p_{X,Y}(x_r, y)}{p_Y(y)} \\ &= \sum_{\{r: x_r \leq x, x_r \in D_X\}} p_{X|Y}(x_r|y) \quad (x \in \mathbb{R}). \end{aligned} \quad (6.4)$$

Si può poi facilmente dimostrare che $F_{X|Y}(x|y)$ è una funzione non decrescente in x , continua a destra per ogni x e che per ogni fissato $y \in D_Y$ risulta:

$$\lim_{x \rightarrow -\infty} F_{X|Y}(x|y) = 0, \quad \lim_{x \rightarrow +\infty} F_{X|Y}(x|y) = 1.$$

Analogamente alla (6.1), per ogni $x \in D_X$ si può definire la funzione di probabilità della variabile aleatoria Y condizionata da $X = x$ al seguente modo:

$$p_{Y|X}(y|x) = \frac{p_{X,Y}(x, y)}{p_X(x)} \quad (y \in \mathbb{R}). \quad (6.5)$$

Inoltre, similmente alla (6.4), per ogni $x \in D_X$ è possibile definire la funzione di distribuzione della variabile aleatoria Y condizionata da $X = x$:

$$F_{Y|X}(y|x) = P(Y \leq y|X = x) = \sum_{\{r: y_r \leq y, y_r \in D_Y\}} p_{Y|X}(y_r|x) \quad (y \in \mathbb{R}). \quad (6.6)$$

Esempio 6.1 Sia (X, Y) un vettore aleatorio discreto di funzione di probabilità congiunta:

$$p_{X,Y}(x, y) = \begin{cases} 1/4, & x = 1, y = 2 \\ 1/8, & (x = 2, y = 2) \text{ oppure } (x = 2, y = 4) \\ 1/2, & x = 1, y = 4 \\ 0, & \text{altrimenti.} \end{cases}$$

Determiniamo le funzioni di probabilità condizionate $p_{X|Y}(x|y)$ e $p_{Y|X}(y|x)$.

A tale scopo osserviamo che le funzioni di probabilità marginali delle variabili X e Y sono rispettivamente:

$$p_X(x) = \begin{cases} 3/4, & x = 1 \\ 1/4, & x = 2 \\ 0, & \text{altrimenti,} \end{cases} \quad p_Y(y) = \begin{cases} 3/8, & y = 2 \\ 5/8, & y = 4 \\ 0, & \text{altrimenti.} \end{cases}$$

Pertanto, dalla (6.1) segue la funzione di probabilità di X condizionata da $Y = 2$:

$$p_{X|Y}(x|2) = \frac{p_{X,Y}(x, 2)}{p_Y(2)} = \begin{cases} 2/3, & x = 1 \\ 1/3, & x = 2 \\ 0, & \text{altrimenti} \end{cases}$$

mentre la funzione di probabilità di X condizionata da $Y = 4$ risulta data da:

$$p_{X|Y}(x|4) = \frac{p_{X,Y}(x, 4)}{p_Y(4)} = \begin{cases} 4/5, & x = 1 \\ 1/5, & x = 2 \\ 0, & \text{altrimenti.} \end{cases}$$

Facendo uso della (6.4) è possibile determinare la funzione di distribuzione di X condizionata da $Y = 2$ e quella condizionata da $Y = 4$:

$$F_{X|Y}(x|2) = \begin{cases} 0, & x < 1 \\ 2/3, & 1 \leq x < 2 \\ 1, & x \geq 2, \end{cases} \quad F_{X|Y}(x|4) = \begin{cases} 0, & x < 1 \\ 4/5, & 1 \leq x < 2 \\ 1, & x \geq 2. \end{cases}$$

Similmente, dalla (6.5) si ha:

$$p_{Y|X}(y|1) = \frac{p_{X,Y}(1, y)}{p_X(1)} = \begin{cases} 1/3, & y = 2 \\ 2/3, & y = 4 \\ 0, & \text{altrimenti,} \end{cases}$$

$$p_{Y|X}(y|2) = \frac{p_{X,Y}(2, y)}{p_X(2)} = \begin{cases} 1/2, & y = 2 \\ 1/2, & y = 4 \\ 0, & \text{altrimenti.} \end{cases}$$

Pertanto, facendo uso della (6.6), è possibile determinare la funzione di distribuzione di Y condizionata da $X = 1$ e quella condizionata da $X = 2$:

$$F_{Y|X}(y|1) = \begin{cases} 0, & y < 2 \\ 1/3, & 2 \leq y < 4 \\ 1, & y \geq 4, \end{cases} \quad F_{Y|X}(y|2) = \begin{cases} 0, & y < 2 \\ 1/2, & 2 \leq y < 4 \\ 1, & y \geq 4. \end{cases}$$

◇

Esempio 6.2 Siano $X \sim \mathcal{B}(m, p)$ e $Y \sim \mathcal{B}(n, p)$ variabili aleatorie indipendenti e sia $Z = X + Y$. Come mostrato nella Proposizione 4.2 si ha $Z \sim \mathcal{B}(m + n, p)$. Qui si desidera determinare la funzione di probabilità condizionata di X conoscendo il valore assunto dalla somma Z .

A tal fine, facendo uso della (6.1) e dell'assunta indipendenza delle variabili X e Y , per ogni $z = 0, 1, \dots, m + n$ se $\max(0, z - n) \leq x \leq \min(z, m)$ si ha:

$$\begin{aligned} p_{X|Z}(x|z) &= \frac{p_{X,Z}(x, z)}{p_Z(z)} = \frac{P(X = x, X + Y = z)}{p_Z(z)} = \frac{P(X = x) P(Y = z - x)}{p_Z(z)} \\ &= \frac{\binom{m}{x} p^x (1 - p)^{m-x} \binom{n}{z-x} p^{z-x} (1 - p)^{n-z+x}}{\binom{m+n}{z} p^z (1 - p)^{m+n-z}} = \frac{\binom{m}{x} \binom{n}{z-x}}{\binom{m+n}{z}}, \end{aligned}$$

mentre $p_{X|Z}(x|z)$ è altrimenti nulla. Pertanto, per ogni fissato $z = 0, 1, \dots, m + n$, la variabile X condizionata da $Z = z$ ha distribuzione ipergeometrica di parametri z, m, n . \diamond

Esempio 6.3 Siano X_1, X_2, \dots, X_n variabili aleatorie indipendenti con $X_r \sim \mathcal{P}(\lambda_r)$ per $r = 1, 2, \dots, n$, e sia $Y = X_1 + X_2 + \dots + X_n$. Come mostrato nel Paragrafo 5.8.3, Y ha distribuzione di Poisson di parametro $\lambda = \lambda_1 + \lambda_2 + \dots + \lambda_n$. Qui si vuole determinare la funzione di probabilità condizionata di ciascuna delle variabili X_r conoscendo il valore assunto dalla somma Y .

A tale scopo, osserviamo che dall'assunta indipendenza delle variabili X_1, X_2, \dots, X_n , se $y = 0, 1, \dots$ e $x = 0, 1, \dots, y$ segue:

$$\begin{aligned} p_{X_r, Y}(x, y) &= P(X_r = x, X_1 + X_2 + \dots + X_n = y) = P\left(X_r = x, \sum_{\substack{i=1 \\ i \neq r}}^n X_i = y - x\right) \\ &= P(X_r = x) P\left(\sum_{\substack{i=1 \\ i \neq r}}^n X_i = y - x\right) = \frac{\lambda_r^x}{x!} e^{-\lambda_r} \frac{(\lambda - \lambda_r)^{y-x}}{(y-x)!} e^{-(\lambda - \lambda_r)}, \end{aligned}$$

mentre tale probabilità è altrimenti nulla. Pertanto, facendo uso della (6.1), per ogni $y = 0, 1, \dots$ si ottiene:

$$p_{X_r|Y}(x|y) = \frac{p_{X_r, Y}(x, y)}{p_Y(y)} = \frac{\frac{\lambda_r^x}{x!} e^{-\lambda_r} \frac{(\lambda - \lambda_r)^{y-x}}{(y-x)!} e^{-(\lambda - \lambda_r)}}{\frac{\lambda^y}{y!} e^{-\lambda}} = \binom{y}{x} \left(\frac{\lambda_r}{\lambda}\right)^x \left(1 - \frac{\lambda_r}{\lambda}\right)^{y-x}$$

se $x = 0, 1, \dots, y$, mentre tale probabilità è altrimenti nulla. È possibile quindi concludere che per ogni fissato $y = 0, 1, \dots$ la variabile X_r condizionata da $Y = y$ ha distribuzione binomiale di parametri y e λ_r/λ . \diamond

Le (6.1) e (6.5) mostrano anche che se (X, Y) è un vettore aleatorio discreto di funzione di probabilità congiunta $p_{X,Y}(x, y)$, sussistono le seguenti relazioni:

$$p_{X,Y}(x, y) = p_{X|Y}(x|y) p_Y(y) \quad (x \in \mathbb{R}, y \in D_Y) \quad (6.7)$$

$$p_X(x) = \sum_{\{r: y_r \in D_Y\}} p_{X|Y}(x|y_r) p_Y(y_r) \quad (x \in \mathbb{R}) \quad (6.8)$$

$$p_{Y|X}(y|x) = \frac{p_{X|Y}(x|y) p_Y(y)}{p_X(x)} \quad (x \in D_X, y \in D_Y), \quad (6.9)$$

avendosi inoltre:

$$p_{X,Y}(x, y) = p_{Y|X}(y|x) p_X(x) \quad (x \in D_X, y \in \mathbb{R}) \quad (6.10)$$

$$p_Y(y) = \sum_{\{r: x_r \in D_X\}} p_{Y|X}(y|x_r) p_X(x_r) \quad (y \in \mathbb{R}) \quad (6.11)$$

$$p_{X|Y}(x|y) = \frac{p_{Y|X}(y|x) p_X(x)}{p_Y(y)} \quad (x \in D_X, y \in D_Y). \quad (6.12)$$

Si noti che le (6.8) e (6.11) possono essere riguardate come estensioni alle variabili aleatorie discrete della legge delle alternative sussistente per gli eventi, e le (6.9) e (6.12) come estensioni alle variabili aleatorie discrete del teorema di Bayes.

Esempio 6.4 In un ufficio postale sono attivi gli sportelli A e B. Denotiamo con Y la variabile aleatoria che descrive il numero di clienti che perviene nell'unità di tempo all'ufficio e assumiamo che sia $Y \sim \mathcal{P}(\lambda)$. I clienti, indipendentemente gli uni dagli altri, arrivando scelgono di accodarsi allo sportello A con probabilità p e allo sportello B con probabilità $1 - p$. Valutiamo la distribuzione di probabilità della variabile X che rappresenta il numero di clienti pervenuti allo sportello A nell'unità di tempo.

A tale scopo, osserviamo che per ogni $n = 0, 1, \dots$ si ha:

$$p_{X|Y}(x|n) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x = 0, 1, \dots, n \\ 0, & \text{altrimenti.} \end{cases}$$

Infatti, se supponiamo che sia n il numero di clienti giunti all'ufficio nell'unità di tempo, $p_{X|Y}(x|n)$ è la probabilità che x degli n clienti hanno scelto lo sportello A. Ricordando quindi la (6.8) si ha:

$$p_X(x) = \sum_{n=0}^{+\infty} p_{X|Y}(x|n) p_Y(n) = \begin{cases} \sum_{n=x}^{+\infty} \binom{n}{x} p^x (1-p)^{n-x} \frac{\lambda^n e^{-\lambda}}{n!}, & x = 0, 1, \dots \\ 0, & \text{altrimenti.} \end{cases}$$

Infine, osservando che per $x = 0, 1, \dots$ risulta

$$\begin{aligned} \sum_{n=x}^{+\infty} \binom{n}{x} p^x (1-p)^{n-x} \frac{\lambda^n e^{-\lambda}}{n!} &= \frac{(\lambda p)^x e^{-\lambda}}{x!} \sum_{n=x}^{+\infty} \frac{[\lambda(1-p)]^{n-x}}{(n-x)!} \\ &= \frac{(\lambda p)^x e^{-\lambda}}{x!} e^{\lambda(1-p)} = \frac{(\lambda p)^x e^{-\lambda p}}{x!}, \end{aligned}$$

si trae:

$$p_X(x) = \begin{cases} \frac{(\lambda p)^x e^{-\lambda p}}{x!}, & x = 0, 1, \dots \\ 0, & \text{altrimenti.} \end{cases}$$

Nello schema considerato, la variabile X , rappresentante il numero di clienti che sceglie lo sportello A, è dunque distribuita alla Poisson con parametro λp . \diamond

Si noti che se X e Y sono variabili aleatorie indipendenti discrete, allora $p_{X,Y}(x, y) = p_X(x)p_Y(y)$ per ogni $x \in \mathbb{R}$ e $y \in \mathbb{R}$, così che dalle (6.1) e (6.4) segue:

$$p_{X|Y}(x|y) = p_X(x), \quad F_{X|Y}(x|y) = F_X(x) \quad (x \in \mathbb{R}, y \in D_Y). \quad (6.13)$$

Le relazioni (6.13) rivestono un ovvio significato: se X e Y sono indipendenti il sapere che Y ha assunto con probabilità positiva un certo valore y non influenza le probabilità relative a X . Analogamente, se X e Y sono variabili aleatorie indipendenti discrete, dalle (6.5) e (6.6) si ottiene:

$$p_{Y|X}(y|x) = p_Y(y), \quad F_{Y|X}(y|x) = F_Y(y) \quad (x \in D_X, y \in \mathbb{R}). \quad (6.14)$$

6.3 Distribuzioni condizionate per variabili continue

Sia (X, Y) un vettore aleatorio, che qui assumeremo essere assolutamente continuo, di densità di probabilità congiunta $f_{X,Y}(x, y)$ e siano $f_X(x)$ e $f_Y(y)$, rispettivamente, le densità di probabilità di X e di Y . Denotiamo ora con D_X l'insieme degli $x \in \mathbb{R}$ tali che $f_X(x) > 0$ e con D_Y l'insieme degli $y \in \mathbb{R}$ tali che $f_Y(y) > 0$. Poiché (X, Y) è assolutamente continuo, non è più possibile applicare direttamente la definizione di funzione di distribuzione condizionata prima data per variabili discrete in quanto la probabilità dell'evento condizionante $\{Y = y\}$ è nulla (v. (3.13)). Si giunge quindi alla seguente definizione:

Definizione 6.3 *Se (X, Y) è un vettore aleatorio assolutamente continuo di densità di probabilità congiunta $f_{X,Y}(x, y)$, per ogni $y \in D_Y$ la funzione di distribuzione condizionata di X dato $Y = y$ è:*

$$F_{X|Y}(x|y) = \lim_{h \downarrow 0} P(X \leq x \mid y < Y \leq y + h) \quad (x \in \mathbb{R}). \quad (6.15)$$

In tal caso la funzione di distribuzione condizionata di X dato $Y = y$ può essere calcolata a partire dalla (6.15) con il procedimento limite qui di seguito riportato. Osserviamo che per ogni $y \in D_Y$ e per ogni $x \in \mathbb{R}$ si ha:

$$P(X \leq x \mid y < Y \leq y + h) = \frac{P(X \leq x, y < Y \leq y + h)}{P(y < Y \leq y + h)} = \frac{\int_{-\infty}^x \left(\int_y^{y+h} f_{X,Y}(u, v) dv \right) du}{\int_y^{y+h} f_Y(v) dv}. \quad (6.16)$$

In accordo con la (6.15), per determinare $F_{X|Y}(x|y)$ si passa al limite per $h \downarrow 0$ nella (6.16) e si applica la regola di De L'Hospital per eliminare la forma indeterminata ottenuta. Quindi, per ogni $y \in D_Y$ si ottiene:

$$F_{X|Y}(x|y) = \lim_{h \downarrow 0} P(X \leq x | y < Y \leq y + h) = \frac{\int_{-\infty}^x f_{X,Y}(u, y) du}{f_Y(y)} = \int_{-\infty}^x \frac{f_{X,Y}(u, y)}{f_Y(y)} du. \quad (6.17)$$

Ricordando la Definizione 3.5, si nota che l'ultimo integrale nella (6.17) è chiaramente la funzione di distribuzione di una variabile aleatoria assolutamente continua; pertanto chiameremo densità di probabilità condizionata di X dato $Y = y$ la funzione integranda nell'ultimo termine della (6.17). Questo procedimento consente di dare la seguente

Definizione 6.4 Se (X, Y) è un vettore assolutamente continuo di densità di probabilità congiunta $f_{X,Y}(x, y)$, allora per ogni $y \in D_Y$ si definisce densità di probabilità condizionata di X dato $Y = y$ il rapporto

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} \quad (x \in \mathbb{R}). \quad (6.18)$$

Dicesi poi funzione di distribuzione condizionata di X dato $Y = y$ la seguente funzione:

$$F_{X|Y}(x|y) = \int_{-\infty}^x f_{X|Y}(u|y) du \quad (x \in \mathbb{R}). \quad (6.19)$$

Si noti che $f_{X|Y}(\cdot|y)$ definisce una densità di probabilità per ogni $y \in D_Y$ e quindi:

$$f_{X|Y}(x|y) \geq 0, \quad \int_{-\infty}^{+\infty} f_{X|Y}(x|y) dx = 1. \quad (6.20)$$

La disuguaglianza segue dalla (6.18) e dall'essere $f_{X,Y}(x, y)$ non negativa e $f_Y(y)$ positiva. Inoltre, sempre dalla (6.18) per ogni $y \in D_Y$ si ricava:

$$\int_{-\infty}^{+\infty} f_{X|Y}(x|y) dx = \frac{1}{f_Y(y)} \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx = \frac{f_Y(y)}{f_Y(y)} = 1.$$

Per ogni $x \in D_X$ si possono anche definire la densità di probabilità della variabile aleatoria Y condizionata da $X = x$ al seguente modo:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} \quad (y \in \mathbb{R}) \quad (6.21)$$

e la funzione di distribuzione condizionata di Y dato $X = x$:

$$F_{Y|X}(y|x) = \int_{-\infty}^y f_{Y|X}(v|x) dv \quad (y \in \mathbb{R}). \quad (6.22)$$

Esempio 6.5 Sia (X, Y) un vettore aleatorio assolutamente continuo di densità di probabilità congiunta

$$f_{X,Y}(x, y) = \begin{cases} 3x, & 0 < x < 1, 0 < y \leq x \\ 0, & \text{altrimenti.} \end{cases}$$

Per determinare le densità di probabilità condizionate di X dato $Y = y$ e di Y dato $X = x$ osserviamo che le densità di probabilità marginali di X e di Y sono rispettivamente:

$$f_X(x) = \int_0^x 3x \, dy = \begin{cases} 3x^2, & 0 < x < 1 \\ 0, & \text{altrimenti} \end{cases}$$

$$f_Y(y) = \int_y^1 3x \, dx = \begin{cases} 3(1 - y^2)/2, & 0 < y < 1 \\ 0, & \text{altrimenti.} \end{cases}$$

Pertanto dalla (6.18), per $0 < y < 1$, risulta:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \begin{cases} 2x/(1 - y^2), & y \leq x < 1 \\ 0, & \text{altrimenti,} \end{cases}$$

mentre dalla (6.21) per $0 < x < 1$ si ha:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \begin{cases} 1/x, & 0 < y \leq x \\ 0, & \text{altrimenti.} \end{cases}$$

Mediante le (6.19) e (6.22) è inoltre possibile determinare le funzioni di distribuzione condizionate di X dato $Y = y$ e di Y dato $X = x$. In particolare, per $0 < y < 1$ si ha:

$$F_{X|Y}(x|y) = \int_{-\infty}^x f_{X|Y}(u|y) \, du = \begin{cases} 0, & x < y \\ \frac{2}{(1 - y^2)} \int_y^x u \, du = \frac{x^2 - y^2}{1 - y^2}, & y \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

mentre per ogni $0 < x < 1$ risulta:

$$F_{Y|X}(y|x) = \int_{-\infty}^y f_{Y|X}(u|x) \, du = \begin{cases} 0, & y < 0 \\ \frac{1}{x} \int_0^y du = \frac{y}{x}, & 0 \leq y < x \\ 1, & y \geq x. \end{cases}$$

◇

Le (6.18) e (6.21) mostrano anche che se (X, Y) è un vettore aleatorio assolutamente continuo di densità di probabilità congiunta $f_{X,Y}(x, y)$, sussistono le seguenti relazioni:

$$f_{X,Y}(x, y) = f_{X|Y}(x|y) f_Y(y) \quad (x \in \mathbb{R}, y \in D_Y) \quad (6.23)$$

$$f_X(x) = \int_{D_Y} f_{X|Y}(x|y) f_Y(y) \, dy \quad (x \in \mathbb{R}) \quad (6.24)$$

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x|y) f_Y(y)}{f_X(x)} \quad (x \in D_X, y \in D_Y), \quad (6.25)$$

ed inoltre:

$$f_{X,Y}(x, y) = f_{Y|X}(y|x) f_X(x) \quad (x \in D_X, y \in \mathbb{R}) \quad (6.26)$$

$$f_Y(y) = \int_{D_X} f_{Y|X}(y|x) f_X(x) dx \quad (y \in \mathbb{R}) \quad (6.27)$$

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)} \quad (x \in D_X, y \in D_Y). \quad (6.28)$$

Analogamente al caso discreto, le (6.24) e (6.27) possono essere riguardate come estensioni alle variabili aleatorie assolutamente continue della legge delle alternative, e le (6.25) e (6.28) come estensioni alle variabili aleatorie assolutamente continue del teorema di Bayes.

Esempio 6.6 Sia (X, Y) un vettore aleatorio assolutamente continuo di densità congiunta

$$f_{X,Y}(x, y) = \begin{cases} 1/x, & 0 < x < 1, 0 < y \leq x \\ 0, & \text{altrimenti.} \end{cases}$$

Si ha evidentemente:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy = \begin{cases} \int_0^x \frac{1}{x} dy = 1, & 0 < x < 1 \\ 0, & \text{altrimenti,} \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx = \begin{cases} \int_y^1 \frac{1}{x} dx = -\ln y, & 0 < y < 1 \\ 0, & \text{altrimenti.} \end{cases}$$

Facendo uso della (6.21), per ogni $0 < x < 1$ si ottiene:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \begin{cases} 1/x, & 0 < y \leq x \\ 0, & \text{altrimenti,} \end{cases}$$

mentre dalla (6.18) per ogni $0 < y < 1$ segue:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \begin{cases} -\frac{1}{x \ln y} & y \leq x < 1 \\ 0, & \text{altrimenti.} \end{cases}$$

Le densità congiunte, condizionate e marginali sono quindi strettamente collegate tra loro tramite le (6.23)÷(6.28). \diamond

Si noti che se X e Y sono indipendenti e assolutamente continue, allora $f_{X,Y}(x, y) = f_X(x) f_Y(y)$ per ogni $x \in \mathbb{R}$ e $y \in \mathbb{R}$, così che dalle (6.18) e (6.19) per ogni $y \in D_Y$ si ha:

$$f_{X|Y}(x|y) = f_X(x), \quad F_{X|Y}(x|y) = F_X(x) \quad (x \in \mathbb{R}). \quad (6.29)$$

Similmente, dalle (6.21) e (6.22) per ogni $x \in D_X$ si ottiene:

$$f_{Y|X}(y|x) = f_Y(y), \quad F_{Y|X}(y|x) = F_Y(y) \quad (y \in \mathbb{R}). \quad (6.30)$$

6.4 Distribuzioni condizionate per vettori aleatori misti

La definizione di funzione di distribuzione condizionata può essere estesa anche ai casi in cui X è una variabile discreta mentre Y è assolutamente continua, e viceversa. Distinguiamo quindi due casi: (a) X è una variabile discreta e Y è assolutamente continua, (b) X è una variabile assolutamente continua e Y è discreta.

6.4.1 X discreta e Y assolutamente continua

Sia (X, Y) un vettore aleatorio misto, tale che X è discreta con funzione di probabilità $p_X(x)$ e Y è assolutamente continua con densità di probabilità $f_Y(y)$. Denotiamo con D_X l'insieme finito o numerabile degli $x \in \mathbb{R}$ tali che $p_X(x) > 0$, e con D_Y l'insieme degli $y \in \mathbb{R}$ tali che $f_Y(y) > 0$.

Definizione 6.5 Se (X, Y) è un vettore aleatorio misto, con X discreta e Y assolutamente continua, allora per ogni $x \in D_X$ la funzione di distribuzione condizionata di Y dato $X = x$ è:

$$F_{Y|X}(y|x) = P(Y \leq y | X = x) = \frac{P(Y \leq y, X = x)}{P(X = x)} \quad (y \in \mathbb{R}). \quad (6.31)$$

Inoltre, per ogni $x \in D_X$ la densità di probabilità condizionata di Y dato $X = x$ è:

$$f_{Y|X}(y|x) = \frac{d}{dy} F_{Y|X}(y|x) \quad (y \in \mathbb{R}), \quad (6.32)$$

se la derivata esiste ovunque, eccetto al più in un numero finito di punti.

È anche possibile definire la funzione di probabilità condizionata di X dato $Y = y$ come segue:

Definizione 6.6 Se (X, Y) è un vettore aleatorio misto, con X discreta e Y assolutamente continua, allora per ogni $y \in D_Y$ la funzione di probabilità condizionata di X dato $Y = y$ è data da

$$p_{X|Y}(x|y) = \lim_{h \downarrow 0} P(X = x | y < Y \leq y + h) \quad (x \in \mathbb{R}), \quad (6.33)$$

sempre che il limite esiste.

Si noti in primo luogo che se $x \notin D_X$ risulta $P(X = x, y < Y \leq y + h) \leq p_X(x) = 0$ e quindi $P(X = x | y < Y \leq y + h) = 0$, il che implica l'annullarsi del primo membro della (6.33). Se invece $x \in D_X$, la funzione di probabilità condizionata di X dato $Y = y$ può essere calcolata a partire dalla (6.33) con il procedimento limite qui di seguito riportato. Osserviamo anzitutto che per ogni $x \in D_X$ e $y \in D_Y$ si ha:

$$\begin{aligned} P(X = x | y < Y \leq y + h) &= \frac{P(X = x, y < Y \leq y + h)}{P(y < Y \leq y + h)} \\ &= \frac{P(X = x) P(y < Y \leq y + h | X = x)}{P(y < Y \leq y + h)} \\ &= \frac{p_X(x) [F_{Y|X}(y + h | x) - F_{Y|X}(y | x)]}{F_Y(y + h) - F_Y(y)}. \end{aligned} \quad (6.34)$$

In accordo con la (6.33), per determinare $p_{X|Y}(x|y)$ si passa poi al limite per $h \downarrow 0$ nella (6.34). Ricordando la (6.32) e l'ipotesi che Y è assolutamente continua, per ogni $x \in D_X$ e $y \in D_Y$ segue:

$$p_{X|Y}(x|y) = p_X(x) \lim_{h \downarrow 0} \frac{[F_{Y|X}(y+h|x) - F_{Y|X}(y|x)]/h}{[F_Y(y+h) - F_Y(y)]/h} = \frac{f_{Y|X}(y|x) p_X(x)}{f_Y(y)}. \quad (6.35)$$

Questo procedimento suggerisce la seguente

Definizione 6.7 Se (X, Y) è un vettore aleatorio misto, con X discreta e Y assolutamente continua, allora per ogni $y \in D_Y$ il rapporto

$$p_{X|Y}(x|y) = \frac{f_{Y|X}(y|x) p_X(x)}{f_Y(y)} \quad (x \in D_X) \quad (6.36)$$

dicesi *funzione di probabilità condizionata di X dato $Y = y$* . Tale funzione è invece nulla se $x \notin D_X$. Analogamente, per ogni $x \in D_X$ la densità di probabilità condizionata di Y dato $X = x$ è:

$$f_{Y|X}(y|x) = \frac{p_{X|Y}(x|y) f_Y(y)}{p_X(x)} \quad (y \in D_Y) \quad (6.37)$$

essendo invece tale funzione nulla se $y \notin D_Y$.

Per ogni $y \in D_Y$, $p_{X|Y}(\cdot|y)$ definisce una funzione di probabilità, avendosi:

$$p_{X|Y}(x|y) \geq 0, \quad \sum_{\{r: x_r \in D_X\}} p_{X|Y}(x_r|y) = 1. \quad (6.38)$$

Inoltre, per ogni $x \in D_X$, $f_{Y|X}(\cdot|x)$ definisce una densità di probabilità, così che:

$$f_{Y|X}(y|x) \geq 0, \quad \int_{D_Y} f_{Y|X}(y|x) dy = 1. \quad (6.39)$$

Come nei casi discreto e assolutamente continuo, sussistono le relazioni

$$p_X(x) = \int_{D_Y} p_{X|Y}(x|y) f_Y(y) dy \quad (x \in D_X), \quad (6.40)$$

$$f_Y(y) = \sum_{\{r: x_r \in D_X\}} f_{Y|X}(y|x_r) p_X(x_r) \quad (y \in D_Y). \quad (6.41)$$

Infatti, facendo uso della (6.39), se $x \in D_X$ si ha:

$$p_X(x) = p_X(x) \int_{D_Y} f_{Y|X}(y|x) dy = \int_{D_Y} \frac{f_{Y|X}(y|x) p_X(x)}{f_Y(y)} f_Y(y) dy,$$

da cui, ricordando la (6.36), segue la (6.40). Inoltre, poiché sussiste la (6.38), se $y \in D_Y$ si ha:

$$f_Y(y) = f_Y(y) \sum_{\{r: x_r \in D_X\}} p_{X|Y}(x_r|y) = \sum_{\{r: x_r \in D_X\}} \frac{p_{X|Y}(x_r|y) f_Y(y)}{p_X(x_r)} p_X(x_r)$$

da cui, tenendo conto della (6.37), si ricava la (6.40).

Si noti infine che se X e Y sono indipendenti, risulta $F_{Y|X}(y|x) = P(Y \leq y|X = x) = P(Y \leq y)$ così che, derivando rispetto a y , si ottiene $f_{Y|X}(y|x) = f_Y(y)$ per ogni $x \in D_X$. Inoltre, se X e Y sono indipendenti, facendo uso della (6.33), per ogni $y \in D_Y$ si ha $p_{X|Y}(x|y) = p_X(x)$.

Esempio 6.7 Sia (X, Y) un vettore aleatorio misto, con X discreta e $Y \sim \mathcal{U}(0, 1)$. Si supponga che per ogni $0 < y < 1$ risulti:

$$p_{X|Y}(x|y) = \begin{cases} 1 - y, & x = 0 \\ y, & x = 1 \\ 0, & \text{altrimenti,} \end{cases}$$

ossia che per ogni fissato $0 < y < 1$ la distribuzione condizionata di X dato che $Y = y$ sia di Bernoulli di parametro y . Dalla (6.40) si ricava:

$$\begin{aligned} p_X(0) &= \int_0^1 p_{X|Y}(0|y) f_Y(y) dy = \int_0^1 (1 - y) dy = \frac{1}{2}, \\ p_X(1) &= \int_0^1 p_{X|Y}(1|y) f_Y(y) dy = \int_0^1 y dy = \frac{1}{2}, \end{aligned}$$

da cui segue:

$$p_X(x) = \begin{cases} 1/2, & x = 0, 1 \\ 0, & \text{altrimenti.} \end{cases}$$

Quindi X ha distribuzione di Bernoulli di parametro $1/2$. La densità condizionata di Y dato che $X = x$ può essere calcolata mediante la (6.37). Infatti, per $x = 0, 1$ si ha:

$$f_{Y|X}(y|x) = \begin{cases} \frac{p_{X|Y}(x|y) f_Y(y)}{p_X(x)} = \frac{(1 - y)^{1-x} y^x}{1/2}, & 0 < y < 1 \\ 0, & \text{altrimenti.} \end{cases}$$

◇

Esempio 6.8 Sia (X, Y) un vettore aleatorio misto, con X discreta e $Y \sim \mathcal{U}(0, 1)$. Si supponga che per ogni $0 < y < 1$ risulti:

$$p_{X|Y}(x|y) = \begin{cases} y(1 - y)^{x-1}, & x = 1, 2, \dots \\ 0, & \text{altrimenti.} \end{cases}$$

Per ogni fissato $0 < y < 1$ la distribuzione condizionata di X dato che $Y = y$ è pertanto geometrica di parametro y . Dalla (6.40), per $x = 1, 2, \dots$ si ha:

$$p_X(x) = \int_0^1 p_{X|Y}(x|y) f_Y(y) dy = \int_0^1 y(1 - y)^{x-1} dy = \int_0^1 (1 - u) u^{x-1} du = \frac{1}{x(x + 1)},$$

da cui segue:

$$p_X(x) = \begin{cases} \frac{1}{x(x + 1)}, & x = 1, 2, \dots \\ 0, & \text{altrimenti.} \end{cases}$$

La densità condizionata di Y dato $X = x$ può essere calcolata mediante la (6.37). Infatti, per $x = 1, 2, \dots$ si ha:

$$f_{Y|X}(y|x) = \begin{cases} \frac{p_{X|Y}(x|y) f_Y(y)}{p_X(x)} = x(x+1)y(1-y)^{x-1}, & 0 < y < 1 \\ 0, & \text{altrimenti.} \end{cases}$$

◇

6.4.2 X assolutamente continua e Y discreta

Sia (X, Y) un vettore aleatorio misto, dove X è assolutamente continua con densità di probabilità $f_X(x)$ e Y è discreta con funzione di probabilità $p_Y(y)$. Denotiamo con D_X l'insieme degli $x \in \mathbb{R}$ tali che $f_X(x) > 0$, e con D_Y l'insieme finito o numerabile degli $y \in \mathbb{R}$ tali che $p_Y(y) > 0$. Si noti che scambiando i ruoli di X e di Y è possibile definire $F_{X|Y}(x|y)$, $f_{X|Y}(x|y)$ e $p_{Y|X}(y|x)$ rispettivamente, come nella (6.31), (6.32) e (6.33). Procedendo in analogia con quanto fatto nel Paragrafo 6.4.1, si giunge alla seguente definizione:

Definizione 6.8 Se (X, Y) è un vettore aleatorio misto, con X assolutamente continua e Y discreta, allora per ogni $x \in D_X$ la funzione di probabilità condizionata di Y dato $X = x$ è:

$$p_{Y|X}(y|x) = \frac{f_{X|Y}(x|y) p_Y(y)}{f_X(x)} \quad (y \in D_Y) \quad (6.42)$$

mentre essa è nulla se $y \notin D_Y$. Inoltre, per ogni $y \in D_Y$ la densità di probabilità condizionata di X dato $Y = y$ è:

$$f_{X|Y}(x|y) = \frac{p_{Y|X}(y|x) f_X(x)}{p_Y(y)} \quad (x \in D_X) \quad (6.43)$$

essendo invece nulla se $x \notin D_X$.

Per ogni $x \in D_X$, $p_{Y|X}(\cdot|x)$ definisce una funzione di probabilità; inoltre, per ogni $y \in D_Y$, la funzione $f_{X|Y}(\cdot|y)$ definisce una densità di probabilità. Evidentemente sussistono anche le relazioni:

$$p_Y(y) = \int_{D_X} p_{Y|X}(y|x) f_X(x) dx \quad (y \in D_Y), \quad (6.44)$$

$$f_X(x) = \sum_{\{r: y_r \in D_Y\}} f_{X|Y}(x|y_r) p_Y(y_r) \quad (x \in D_X). \quad (6.45)$$

Se X e Y sono indipendenti, per ogni $x \in D_X$ risulta $f_{Y|X}(y|x) = f_Y(y)$, ed inoltre per ogni $y \in D_Y$ si ha $p_{X|Y}(x|y) = p_X(x)$.

Si noti infine che se (X, Y) è un vettore aleatorio misto, la sua distribuzione viene spesso specificata assegnando la distribuzione di una delle componenti e la distribuzione condizionata dell'altra.

Esempio 6.9 Sia (X, Y) un vettore aleatorio misto, con X assolutamente continua e con Y discreta. Si supponga che

$$f_X(x) = \begin{cases} x e^{-x}, & x > 0 \\ 0, & \text{altrimenti,} \end{cases}$$

e che per ogni $x > 0$ risulti:

$$p_{Y|X}(y|x) = \begin{cases} \frac{x^y}{y!} e^{-x}, & y = 0, 1, \dots \\ 0, & \text{altrimenti.} \end{cases}$$

Per ogni fissato $x > 0$ la distribuzione condizionata di Y dato che $X = x$ è pertanto di Poisson di parametro x . Dalla (6.44), si ha poi:

$$\begin{aligned} p_Y(y) &= \int_0^{+\infty} p_{Y|X}(y|x) f_X(x) dx = \int_0^{+\infty} \frac{x^y}{y!} e^{-x} x e^{-x} dx = \frac{1}{y!} \int_0^{+\infty} x^{y+1} e^{-2x} dx \\ &= \frac{1}{2y!} \frac{(y+1)!}{2^{y+1}} \quad y = 0, 1, \dots \end{aligned}$$

dove l'ultima uguaglianza segue ricordando l'espressione dei momenti di una variabile aleatoria esponenziale (v. Paragrafo 5.8.3). Pertanto

$$p_Y(y) = \begin{cases} \frac{y+1}{2^{y+2}}, & y = 0, 1, \dots \\ 0, & \text{altrimenti.} \end{cases}$$

Determiniamo ora la densità di probabilità di X supponendo che la variabile Y abbia assunto il valore y . Questa segue immediatamente dalla (6.43) e per ogni $y = 0, 1, \dots$ si ottiene:

$$f_{X|Y}(x|y) = \begin{cases} \frac{p_{Y|X}(y|x) f_X(x)}{p_Y(y)} = \frac{2(2x)^{y+1} e^{-2x}}{(y+1)!}, & x > 0 \\ 0, & \text{altrimenti.} \end{cases}$$

◇

Esempio 6.10 Sia (X, Y) un vettore aleatorio misto, con X assolutamente continua e con $Y \sim \mathcal{P}(\lambda)$. Si supponga che per ogni $y = 0, 1, \dots$ si abbia:

$$f_{X|Y}(x|y) = \begin{cases} y e^{-y} x, & x > 0 \\ 0, & \text{altrimenti,} \end{cases}$$

ossia si assuma che per ogni fissato $y = 0, 1, \dots$ la distribuzione condizionata di X dato che $Y = y$ è esponenziale di parametro y . Dalla (6.45) si ottiene:

$$\begin{aligned} f_X(x) &= \sum_{y=0}^{+\infty} f_{X|Y}(x|y) p_Y(y) = \sum_{y=0}^{+\infty} y e^{-y} x \frac{\lambda^y}{y!} e^{-\lambda} = e^{-\lambda} (\lambda e^{-x}) \sum_{y=1}^{+\infty} \frac{(\lambda e^{-x})^{y-1}}{(y-1)!} \\ &= e^{-\lambda} (\lambda e^{-x}) \exp\{\lambda e^{-x}\} \quad (x > 0). \end{aligned}$$

Quindi, si ha:

$$f_X(x) = \begin{cases} \lambda \exp\{-(\lambda + x) + \lambda e^{-x}\}, & x > 0 \\ 0, & \text{altrimenti.} \end{cases}$$

Inoltre, facendo uso della (6.42), per ogni $x > 0$ si nota che $p_{Y|X}(0|x) = 0$ e che inoltre se $y = 1, 2, \dots$ risulta

$$p_{Y|X}(y|x) = \frac{f_{X|Y}(x|y) p_Y(y)}{f_X(x)} = \frac{\lambda^{y-1}}{(y-1)!} \exp\{-x(y-1) - \lambda e^{-x}\}.$$

Pertanto

$$p_{Y|X}(y|x) = \begin{cases} \frac{\lambda^{y-1}}{(y-1)!} \exp\{-x(y-1) - \lambda e^{-x}\}, & y = 1, 2, \dots \\ 0, & \text{altrimenti.} \end{cases}$$

◇

6.5 Probabilità condizionate

Nei precedenti paragrafi abbiamo definito la funzione di distribuzione condizionata per vettori aleatori (X, Y) discreti, assolutamente continui e misti. Vogliamo ora definire probabilità del tipo $P(X \in B | Y = y)$, dove B è un sottoinsieme della classe di Borel \mathcal{B} . Distinguiamo i seguenti due casi per la variabile Y condizionante: (a) Y è discreta, (b) Y è assolutamente continua.

6.5.1 Variabile condizionante discreta

Definizione 6.9 Se Y è discreta, denotando con D_Y l'insieme, finito o numerabile, degli $y \in \mathbb{R}$ tali che $p_Y(y) > 0$, per ogni $y \in D_Y$

$$P(X \in B | Y = y) = \frac{P(X \in B, Y = y)}{p_Y(y)} \quad (6.46)$$

dicesi probabilità condizionata dell'evento $\{X \in B\}$ dato che $Y = y$.

In particolare, se anche X è discreta e se B è un sottoinsieme finito o numerabile di \mathbb{R} , per ogni $y \in D_Y$ la (6.46) diventa:

$$P(X \in B | Y = y) = \sum_{\{r: x_r \in B \cap D_X\}} p_{X|Y}(x_r|y) \quad (6.47)$$

dove D_X denota l'insieme, finito o numerabile, degli $x \in \mathbb{R}$ tali che $p_X(x) > 0$, e dove $p_{X|Y}(x|y) = p_{X,Y}(x, y)/p_Y(y)$ per la definizione (6.1). Se, invece, X è assolutamente continua e B è un sottoinsieme di \mathbb{R} , per ogni $y \in D_Y$ la (6.46) diventa:

$$P(X \in B | Y = y) = \int_{B \cap D_X} f_{X|Y}(x|y) dx \quad (6.48)$$

dove D_X denota l'insieme, finito o numerabile, degli $x \in \mathbb{R}$ tali che $f_X(x) > 0$, e dove $f_{X|Y}(x|y) = p_{Y|X}(y|x) f_X(x)/p_Y(y)$ per la definizione (6.43).

6.5.2 Variabile condizionante assolutamente continua

Definizione 6.10 Se Y è assolutamente continua, denotando con D_Y l'insieme degli $y \in \mathbb{R}$ tali che $f_Y(y) > 0$, per ogni $y \in D_Y$ la probabilità condizionata di un evento $\{X \in B\}$ dato che $Y = y$ è:

$$\begin{aligned} P(X \in B \mid Y = y) &= \lim_{h \downarrow 0} P(X \in B \mid y < Y \leq y + h) \\ &= \lim_{h \downarrow 0} \frac{P(X \in B, y < Y \leq y + h)}{P(y < Y \leq y + h)} \end{aligned} \quad (6.49)$$

sempre che il limite esista.

In particolare, se X è discreta e B è un sottoinsieme finito o numerabile di \mathbb{R} , per ogni $y \in D_Y$ la (6.49) conduce nuovamente alla (6.47), con $p_{X|Y}(x|y) = f_{Y|X}(y|x) p_X(x) / f_Y(y)$ definita in (6.36). Se invece X è assolutamente continua e se B è un sottoinsieme di \mathbb{R} , per ogni $y \in D_Y$ la (6.49) conduce di nuovo alla (6.48), con $f_{X|Y}(x|y) = f_{X,Y}(x, y) / f_Y(y)$ definita in (6.18).

6.6 Legge delle alternative e teorema di Bayes per variabili aleatorie

Come visto nel precedente paragrafo, per un vettore aleatorio (X, Y) è possibile definire $P(X \in B \mid Y = y)$, dove B è un sottoinsieme della classe di Borel \mathcal{B} . Qui si vuole mostrare che tale probabilità gioca un ruolo rilevante poiché permette di estendere la legge delle alternative e la legge di Bayes per eventi, di cui ai Paragrafi 2.6 e 2.7, anche a situazioni coinvolgenti variabili aleatorie.

Teorema 6.1 Sia (X, Y) un vettore aleatorio e sia $B \in \mathcal{B}$.

(a) Se Y è discreta, denotando con D_Y l'insieme finito o numerabile degli $y \in \mathbb{R}$ tali che $p_Y(y) > 0$, si ha:

$$P(X \in B) = \sum_{\{r: y_r \in D_Y\}} P(X \in B \mid Y = y_r) p_Y(y_r), \quad B \in \mathcal{B}. \quad (6.50)$$

(b) Se Y è assolutamente continua, denotando con D_Y l'insieme degli $y \in \mathbb{R}$ tali che $f_Y(y) > 0$, risulta:

$$P(X \in B) = \int_{D_Y} P(X \in B \mid Y = y) f_Y(y) dy, \quad B \in \mathcal{B}. \quad (6.51)$$

Dimostrazione

(a) Se Y è discreta, essendo D_Y l'insieme finito o numerabile degli $y \in \mathbb{R}$ tali che $p_Y(y) > 0$, si ha:

$$\begin{aligned} P(X \in B) &= P(X \in B, Y \in D_Y) = \sum_{\{r: y_r \in D_Y\}} P(X \in B, Y = y_r) \\ &= \sum_{\{r: y_r \in D_Y\}} P(X \in B \mid Y = y_r) p_Y(y_r), \end{aligned}$$

dove l'ultima uguaglianza segue dalla (6.46). La (6.50) è così dimostrata.

(b) Se Y è assolutamente continua, il risultato non può essere ottenuto per via diretta essendo $P(Y = y) = 0$ per ogni $y \in \mathbb{R}$, e si rende così necessario un procedimento limite. A tal fine, rappresentiamo l'insieme D_Y come unione dei sottointervalli $(y_r, y_{r+1}]$ ottenuti tramite i punti y_r ($r = 0, \pm 1, \pm 2, \dots$), ed indichiamo con H_r l'evento $\{y_r < Y \leq y_{r+1}\}$. In virtù dell'assioma di additività completa si ha:

$$\begin{aligned} P(X \in B) &= \sum_r P(X \in B, Y \in H_r) = \sum_r P(X \in B \mid Y \in H_r) P(Y \in H_r) \\ &= \sum_r P(X \in B \mid y_r < Y \leq y_{r+1}) P(y_r < Y \leq y_{r+1}), \end{aligned} \quad (6.52)$$

dove le somme sono estese a tutti gli r tali che $(y_r, y_{r+1}] \subseteq D_Y$. Essendo Y assolutamente continua, il teorema della media conduce a scrivere:

$$P(y_r < Y \leq y_{r+1}) = \int_{y_r}^{y_{r+1}} f_Y(y) dy = f_Y(\xi_r) \Delta y_r, \quad (6.53)$$

con $y_r \leq \xi_r \leq y_{r+1}$ e con $\Delta y_r = y_{r+1} - y_r$. Utilizzando (6.53) nella (6.52) si ottiene:

$$P(X \in B) = \sum_r P(X \in B \mid y_r < Y \leq y_{r+1}) f_Y(\xi_r) \Delta y_r. \quad (6.54)$$

Scegliendo una partizione di D_Y progressivamente più fine in maniera da ottenere intervalli $(y_r, y_{r+1}]$ la cui ampiezza tende a zero, e ricordando la definizione (6.49) di probabilità condizionata, attraverso un procedimento limite dalla (6.54) si ottiene la (6.51). Ciò completa la dimostrazione. \square

Si noti che se (X, Y) è discreto e $B = \{x\}$, dalla (6.50) si ottiene la (6.8). Se, invece, X è discreta, Y assolutamente continua e $B = \{x\}$, la (6.51) si identifica con la (6.40). Inoltre, se X è assolutamente continua, Y discreta e se si sceglie B coincidente con l'intervallo $(x, x + h]$, la (6.50) conduce a scrivere:

$$P(x < X \leq x + h) = \sum_{\{r: y_r \in D_Y\}} P(x < X \leq x + h \mid Y = y_r) p_Y(y_r),$$

da cui, dividendo ambo i membri per h e procedendo al limite per $h \downarrow 0$, si riottiene la (6.45). Infine, se (X, Y) è assolutamente continuo e si sceglie B coincidente con l'intervallo $(x, x + h]$, dalla (6.51) si ricava:

$$P(x < X \leq x + h) = \int_{D_Y} P(x < X \leq x + h \mid Y = y) f_Y(y) dy$$

che, previa divisione di ambo i membri per h e passaggio al limite per $h \downarrow 0$, conduce alla (6.24).

Scegliendo $B \in \mathcal{B}$ coincidente con l'intervallo $(-\infty, x]$, dal Teorema 6.1 segue anche che è possibile esprimere la funzione di distribuzione di una variabile aleatoria X in termini

di funzione di distribuzione condizionata di X dato $Y = y$. Infatti, se Y è discreta e D_Y è l'insieme finito o numerabile degli $y \in \mathbb{R}$ tali che $p_Y(y) > 0$, dalla (6.50) si ricava:

$$F_X(x) = \sum_{\{r: y_r \in D_Y\}} F_{X|Y}(x|y_r) p_Y(y_r) \quad (x \in \mathbb{R}); \quad (6.55)$$

se invece Y è assolutamente continua e D_Y è l'insieme degli $y \in \mathbb{R}$ tali che $f_Y(y) > 0$, dalla (6.51) si ottiene:

$$F_X(x) = \int_{D_Y} F_{X|Y}(x|y) f_Y(y) dy \quad (x \in \mathbb{R}). \quad (6.56)$$

Le leggi delle alternative per vettori aleatori sono riassunte nella Tabella 6.1.

Tabella 6.1 – Leggi delle alternative per il vettore aleatorio (X, Y) .

<p><i>Y discreta</i></p> $P(X \in B) = \sum_{\{r: y_r \in D_Y\}} P(X \in B Y = y_r) p_Y(y_r)$ $F_X(x) = \sum_{\{r: y_r \in D_Y\}} F_{X Y}(x y_r) p_Y(y_r) \quad (x \in \mathbb{R})$	<p><i>Y assolutamente continua</i></p> $P(X \in B) = \int_{D_Y} P(X \in B Y = y) f_Y(y) dy$ $F_X(x) = \int_{D_Y} F_{X Y}(x y) f_Y(y) dy, \quad (x \in \mathbb{R})$
<p><i>(X, Y) discreto</i></p> $p_X(x) = \sum_{\{r: y_r \in D_Y\}} p_{X Y}(x y_r) p_Y(y_r) \quad (x \in \mathbb{R})$	<p><i>X discreta, Y assolutamente continua</i></p> $p_X(x) = \int_{D_Y} p_{X Y}(x y) f_Y(y) dy \quad (x \in \mathbb{R})$
<p><i>X assolutamente continua, Y discreta</i></p> $f_X(x) = \sum_{\{r: y_r \in D_Y\}} f_{X Y}(x y_r) p_Y(y_r) \quad (x \in \mathbb{R})$	<p><i>(X, Y) assolutamente continuo</i></p> $f_X(x) = \int_{D_Y} f_{X Y}(x y) f_Y(y) dy \quad (x \in \mathbb{R})$

Esempio 6.11 Sia X_0, X_1, \dots una successione di variabili aleatorie indipendenti e identicamente distribuite con funzione di distribuzione $F_X(x)$ e sia $N \sim \mathcal{P}(\lambda)$. Sotto l'ipotesi che N è indipendente da ognuna delle X_r , determiniamo la funzione di distribuzione di $U_N = \max(X_0, X_1, \dots, X_N)$ e di $V_N = \min(X_0, X_1, \dots, X_N)$.

Si osservi che U_N e V_N dipendono dai valori assunti dalla variabile N , così che le distribuzioni richieste non possono essere determinate direttamente. Se invece si suppone che sia $N = n$, il numero di variabili tra cui scegliere il massimo o il minimo non è più aleatorio. In altri termini, la variabile U_N condizionata da $N = n$ è il massimo tra n variabili aleatorie indipendenti e identicamente distribuite, così che la funzione di distribuzione di U_N condizionata da $N = n$ può essere determinata come già mostrato nel Paragrafo 3.9. Invero, facendo uso della (3.63), dell'indipendenza di N dalle X_i nonché dell'indipendenza di X_1, X_2, \dots ,

per $n = 0, 1, \dots$ risulta:

$$\begin{aligned} F_{U_N|N}(x|n) &= P(U_N \leq x \mid N = n) = \frac{P(U_N \leq x, N = n)}{p_N(n)} = \frac{P(U_n \leq x) p_N(n)}{p_N(n)} \\ &= P\{\max(X_0, X_1, \dots, X_n) \leq x\} = [F(x)]^{n+1} \quad (x \in \mathbb{R}). \end{aligned}$$

Per determinare la funzione di distribuzione di U_N è ora sufficiente applicare la (6.55). Infatti, ricordando la (4.20), per ogni $x \in \mathbb{R}$ si ha:

$$F_{U_N}(x) = \sum_{n=0}^{+\infty} F_{U_N|N}(x|n) p_N(n) = \sum_{n=0}^{+\infty} e^{-\lambda} \frac{\lambda^n}{n!} [F(x)]^{n+1} = F(x) \exp\{-\lambda[1-F(x)]\}.$$

Per la variabile V_N procediamo in modo analogo. In primo luogo valutiamo la funzione di distribuzione di V_N condizionata da $N = n$ che, facendo uso della (3.64) e delle ipotesi di indipendenza tra le variabili in questione, per ogni $n = 0, 1, \dots$ risulta essere:

$$\begin{aligned} F_{V_N|N}(x|n) &= P(V_N \leq x \mid N = n) = \frac{P(V_N \leq x, N = n)}{p_N(n)} = P(V_n \leq x) \\ &= P\{\min(X_0, X_1, \dots, X_n) \leq x\} = 1 - [1 - F(x)]^{n+1} \quad (x \in \mathbb{R}). \end{aligned}$$

Utilizzando nuovamente la (6.55), si ottiene poi:

$$\begin{aligned} F_{V_N}(x) &= \sum_{n=0}^{+\infty} F_{V_N|N}(x|n) p_N(n) = \sum_{n=0}^{+\infty} e^{-\lambda} \frac{\lambda^n}{n!} \{1 - [1 - F(x)]^{n+1}\} \\ &= 1 - [1 - F(x)] \exp\{-\lambda F(x)\} \quad (x \in \mathbb{R}). \end{aligned}$$

Sottolineiamo, infine, che questi risultati sono stati ottenuti sotto l'ipotesi di indipendenza e di identica distribuzione delle variabili X_i , ma non è stata necessaria alcuna ipotesi sulla natura delle variabili X_i . Se assumiamo che esse sono assolutamente continue, tali risultano U_N e V_N , e si ha:

$$\begin{aligned} f_{U_N}(x) &= \frac{dF_{U_N}(x)}{dx} = [1 + \lambda F(x)] f(x) \exp\{-\lambda[1 - F(x)]\} \quad (x \in \mathbb{R}), \\ f_{V_N}(x) &= \frac{dF_{V_N}(x)}{dx} = \{1 + \lambda[1 - F(x)]\} f(x) \exp\{-\lambda F(x)\} \quad (x \in \mathbb{R}), \end{aligned}$$

dove $f(x)$ denota la densità di ognuna delle variabili X_i . ◇

Così come accade nel contesto degli eventi, anche per le variabili aleatorie la formula delle alternative costituisce la base per la determinazione di distribuzioni “a posteriori”.

Il seguente teorema è la riformulazione della legge di Bayes per distribuzioni di variabili aleatorie.

Teorema 6.2 Sia (X, Y) un vettore aleatorio e sia $B \in \mathcal{B}$.

(a) Sia Y discreta e sia D_Y l'insieme, finito o numerabile, degli $y \in \mathbb{R}$ tali che $p_Y(y) > 0$. Se $P(X \in B) > 0$, per ogni $y \in D_Y$ si ha:

$$\begin{aligned} p_{Y|X}(y|B) &= P(Y = y \mid X \in B) = \frac{P(X \in B \mid Y = y) p_Y(y)}{P(X \in B)} \\ &= \frac{P(X \in B \mid Y = y) p_Y(y)}{\sum_{\{r: p_Y(r) > 0\}} P(X \in B \mid Y = r) p_Y(r)}. \end{aligned} \quad (6.57)$$

(b) Siano Y assolutamente continua e D_Y l'insieme degli $y \in \mathbb{R}$ tali che $f_Y(y) > 0$. Se $P(X \in B) > 0$, per ogni $y \in D_Y$ risulta:

$$\begin{aligned} f_{Y|X}(y|B) &= \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} P(y < Y \leq y + \varepsilon \mid X \in B) = \frac{P(X \in B \mid Y = y) f_Y(y)}{P(X \in B)} \\ &= \frac{P(X \in B \mid Y = y) f_Y(y)}{\int_{D_Y} P(X \in B \mid Y = v) f_Y(v) dv}. \end{aligned} \quad (6.58)$$

Dimostrazione Nel caso (a), essendo Y discreta e $P(X \in B) > 0$, si ha:

$$P(Y = y \mid X \in B) = \frac{P(X \in B, Y = y)}{P(X \in B)} = \frac{P(X \in B \mid Y = y) p_Y(y)}{P(X \in B)},$$

da cui, ricordando la (6.50), per ogni $y \in D_Y$ segue immediatamente la (6.57). Nel caso (b), essendo $P(X \in B) > 0$, per ogni $\varepsilon > 0$ si ha:

$$\begin{aligned} P(y < Y \leq y + \varepsilon \mid X \in B) &= \frac{P(y < Y \leq y + \varepsilon, X \in B)}{P(X \in B)} \\ &= \frac{P(X \in B \mid y < Y \leq y + \varepsilon) P(y < Y \leq y + \varepsilon)}{P(X \in B)}; \end{aligned}$$

di qui, dividendo ambo i membri per h , procedendo al limite per $\varepsilon \downarrow 0$ e utilizzando la (6.49), si ottiene la prima delle (6.57). L'ultima uguaglianza nella (6.58) segue poi dalla precedente tenendo conto della (6.51). \square

Si noti che se (X, Y) è discreto e $B = \{x\}$, dalla (6.57) si ottiene la (6.9). Se invece X è discreta, Y assolutamente continua e $B = \{x\}$, la (6.58) si identifica con la (6.37). Inoltre se X è assolutamente continua, Y discreta e si sceglie B coincidente con l'intervallo $(x, x + h]$, la (6.57) fornisce:

$$P(Y = y \mid x < X \leq x + h) = \frac{P(x < X \leq x + h \mid Y = y)/h}{P(x < X \leq x + h)/h} p_Y(y),$$

da cui, procedendo al limite per $h \downarrow 0$, si ottiene la (6.42). Infine, se (X, Y) è assolutamente continuo e se si sceglie di nuovo B coincidente con l'intervallo $(x, x + h]$, la (6.58) si può scrivere al seguente modo:

$$\lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} P(y < Y \leq y + \varepsilon \mid x < X \leq x + h) = \frac{P(x < X \leq x + h \mid Y = y)/h}{P(x < X \leq x + h)/h} f_Y(y),$$

così che, procedendo al limite per $h \downarrow 0$, si ricava la (6.25).

Come nel caso degli eventi, la legge di Bayes assume un significato particolarmente importante se alle variabili aleatorie coinvolte si assegnano i ruoli di *causa* ed *effetto*. Precisamente, chiamiamo Y la causa del verificarsi di $\{X \in B\}$, ossia dell'effetto di tale causa. Se Y è discreta, la probabilità condizionata $p_{Y|X}(y|B)$ è la probabilità di Y valutata sapendo che $\{X \in B\}$; essa è denominata *probabilità a posteriori*. La legge di Bayes (6.57) permette quindi di calcolare la probabilità a posteriori $p_{Y|X}(y|B)$ supponendo note la funzione di probabilità di X e le probabilità condizionate $P(X \in B | Y = y)$. Analogamente, se Y è assolutamente continua, la densità condizionata $f_{Y|X}(y|B)$ è la densità di Y valutata sapendo che $\{X \in B\}$; essa prende il nome di *densità a posteriori*. La legge di Bayes (6.58) consente quindi di calcolare la densità di probabilità a posteriori $f_{Y|X}(y|B)$ supponendo note la densità di probabilità di X e le probabilità condizionate $P(X \in B | Y = y)$.

Le leggi di Bayes per variabili aleatorie sono riassunte nella Tabella 6.2.

Tabella 6.2 – Leggi di Bayes per il vettore aleatorio (X, Y) .

$p_{Y X}(y B) = \frac{Y \text{ discreta}}{P(X \in B Y = y) p_Y(y)} \\ \frac{P(X \in B)}{P(X \in B)} \\ (y \in D_Y, P(X \in B) > 0)$	$f_{Y X}(y B) = \frac{Y \text{ assolutamente continua}}{P(X \in B Y = y) f_Y(y)} \\ \frac{P(X \in B)}{P(X \in B)} \\ (y \in D_Y, P(X \in B) > 0)$
$(X, Y) \text{ discreto} \\ p_{Y X}(y x) = \frac{p_{X Y}(x y) p_Y(y)}{p_X(x)} \\ (x \in D_X, y \in D_Y)$	$X \text{ discreta}, Y \text{ assolutamente continua} \\ f_{Y X}(y x) = \frac{p_{X Y}(x y) f_Y(y)}{p_X(x)} \\ (x \in D_X, y \in D_Y)$
$X \text{ assolutamente continua}, Y \text{ discreta} \\ p_{Y X}(y x) = \frac{f_{X Y}(x y) p_Y(y)}{f_X(x)} \\ (x \in D_X, y \in D_Y)$	$(X, Y) \text{ assolutamente continuo} \\ f_{Y X}(y x) = \frac{f_{X Y}(x y) f_Y(y)}{f_X(x)} \\ (x \in D_X, y \in D_Y)$

Esempio 6.12 Siano Y_1, Y_2, \dots, Y_n variabili aleatorie indipendenti e identicamente distribuite con funzione di distribuzione $F(x)$ e siano $U = \max(Y_1, Y_2, \dots, Y_n)$ e $V = \min(Y_1, Y_2, \dots, Y_n)$. Determiniamo la distribuzione di Y_1 condizionata da $U \leq u$ e la distribuzione di Y_1 condizionata da $V > v$. Distinguiamo due casi: (a) Y_1, Y_2, \dots, Y_n discrete e (b) Y_1, Y_2, \dots, Y_n assolutamente continue.

Nel caso (a), essendo Y_1, Y_2, \dots, Y_n discrete, tali sono anche U e V . Posto $B = \{U \leq u\}$, dalla (3.63) segue $P(U \leq u) = [F(u)]^n$, così che se $F(u) > 0$ dalla (6.57) si ottiene:

$$\begin{aligned}
 P(Y_1 = y | U \leq u) &= \frac{P(U \leq u | Y_1 = y) p_{Y_1}(y)}{P(U \leq u)} \\
 &= \begin{cases} \frac{[F(u)]^{n-1} p_{Y_1}(y)}{[F(u)]^n} = \frac{p_{Y_1}(y)}{F(u)}, & y < u \\ 0, & y \geq v, \end{cases}
 \end{aligned}$$

Similmente, se $B = \{V > v\}$ dalla (3.64) segue $P(V > v) = [1 - F(u)]^n$. Se $F(u) < 1$, dalla (6.57) si ricava quindi:

$$\begin{aligned} P(Y_1 = y \mid V > v) &= \frac{P(V > v \mid Y_1 = y) p_{Y_1}(y)}{P(V > v)} \\ &= \begin{cases} \frac{[1 - F(v)]^{n-1} p_{Y_1}(y)}{[1 - F(v)]^n} = \frac{p_{Y_1}(y)}{1 - F(v)}, & y > v \\ 0, & y \leq v. \end{cases} \end{aligned}$$

Nel caso (b) invece, essendo Y_1, Y_2, \dots, Y_n assolutamente continue, tali sono anche U e V ; pertanto, facendo uso della (6.58), si può procedere in modo analogo al caso discreto, così che se $F(u) > 0$ si ha:

$$\lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} P(y < Y \leq y + \varepsilon \mid U \leq u) = \frac{P(U \leq u \mid Y = y) f_{Y_1}(y)}{P(U \leq u)} = \begin{cases} \frac{f_{Y_1}(y)}{F(u)}, & y < u \\ 0, & y \geq v. \end{cases}$$

Se $F(v) < 1$ risulta invece:

$$\lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} P(y < Y \leq y + \varepsilon \mid V > v) = \frac{P(V > v \mid Y_1 = y) f_{Y_1}(y)}{P(V > v)} = \begin{cases} \frac{f_{Y_1}(y)}{1 - F(v)}, & y > v \\ 0, & y \leq v. \end{cases}$$

◇

6.7 Medie e momenti condizionati

Sia (X, Y) un vettore aleatorio con funzione di distribuzione $F_{X,Y}(x, y)$ e sia Z una variabile aleatoria funzione di X e Y :

$$Z = g(X, Y), \quad (6.59)$$

con $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ funzione Borel-misurabile. Ci proponiamo di definire la *media condizionata di Z dato $Y = y$* . Esaminiamo separatamente i seguenti quattro casi: (a) (X, Y) discreto, (b) (X, Y) assolutamente continuo, (c) (X, Y) misto con X discreta e Y assolutamente continua, (d) (X, Y) misto con X assolutamente continua e Y discreta.

(a) Sia (X, Y) discreto di funzione di probabilità congiunta $p_{X,Y}(x, y)$. Siano poi $p_X(x)$ e $p_Y(y)$ rispettivamente le funzioni di probabilità marginali di X e di Y . Denotiamo con D_X l'insieme finito o numerabile degli $x \in \mathbb{R}$ tali che $p_X(x) > 0$ e indichiamo con D_Y l'insieme finito o numerabile degli $y \in \mathbb{R}$ tali che $p_Y(y) > 0$. Per ogni $y \in D_Y$, la media condizionata di $Z = g(X, Y)$ dato $Y = y$ è:

$$E(Z \mid Y = y) = \sum_{\{r: x_r \in D_X\}} g(x_r, y) p_{X|Y}(x_r|y) = \sum_{\{r: x_r \in D_X\}} g(x_r, y) \frac{p_{X,Y}(x_r, y)}{p_Y(y)}, \quad (6.60)$$

sempre che tale media esista. Così come accade per il valore medio di una variabile discreta non condizionata, la media condizionata di Z dato $Y = y$ esiste sempre se Z è non negativa oppure se Z è non positiva, anche se non necessariamente finita. Se Z assume con probabilità

non nulla valori sia positivi che negativi, la media condizionata di Z dato $Y = y$ esiste se la serie al secondo membro della (6.60) è assolutamente convergente oppure se nel calcolo non compaiono simultaneamente infiniti positivi e negativi.

(b) Sia (X, Y) assolutamente continuo di densità di probabilità congiunta $f_{X,Y}(x, y)$. Siano inoltre $f_X(x)$ e $f_Y(y)$ rispettivamente le densità di probabilità di X e di Y . Denotati con D_X l'insieme degli $x \in \mathbb{R}$ tali che $f_X(x) > 0$ e con D_Y l'insieme degli $y \in \mathbb{R}$ tali che $f_Y(y) > 0$, per ogni $y \in D_Y$ la media condizionata di $Z = g(X, Y)$ dato $Y = y$ è:

$$E(Z | Y = y) = \int_{-\infty}^{+\infty} g(x, y) f_{X|Y}(x|y) dx = \int_{-\infty}^{+\infty} g(x, y) \frac{f_{X,Y}(x, y)}{f_Y(y)} dx, \quad (6.61)$$

sempre che essa esista. Così come accade per il valore medio di una variabile assolutamente continua non condizionata, la media condizionata di Z dato $Y = y$ esiste sempre se Z è non negativa oppure Z è non positiva, anche se non necessariamente finita. Se Z assume con probabilità non nulla valori sia positivi che negativi, la media condizionata di Z dato $Y = y$ esiste se l'integrale al secondo membro della (6.61) è assolutamente convergente oppure se nel corso del suo calcolo non compaiono simultaneamente infiniti positivi e negativi.

(c) Sia (X, Y) misto, con X discreta con funzione di probabilità $p_X(x)$ e Y assolutamente continua con densità di probabilità $f_Y(y)$. Siano D_X l'insieme finito o numerabile degli $x \in \mathbb{R}$ tali che $p_X(x) > 0$ e D_Y l'insieme degli $y \in \mathbb{R}$ tali che $f_Y(y) > 0$. Per ogni $y \in D_Y$, la media condizionata di $Z = g(X, Y)$ dato $Y = y$ è:

$$E(Z | Y = y) = \sum_{\{r: x_r \in D_X\}} g(x_r, y) p_{X|Y}(x_r|y) = \sum_{\{r: x_r \in D_X\}} g(x_r, y) \frac{f_{Y|X}(y|x_r) p_X(x_r)}{f_Y(y)} \quad (6.62)$$

sempre che essa esista. Le condizioni per l'esistenza della media condizionata sono le stesse del caso (a).

(d) Sia (X, Y) misto, con X assolutamente continua con densità di probabilità $f_X(x)$ e Y discreta con funzione di probabilità $p_Y(y)$. Siano D_X l'insieme degli $x \in \mathbb{R}$ tali che $f_X(x) > 0$ e D_Y l'insieme finito o numerabile degli $y \in \mathbb{R}$ per i quali risulta $p_Y(y) > 0$. Per ogni $y \in D_Y$, la media condizionata di $Z = g(X, Y)$ dato $Y = y$ è:

$$E(Z | Y = y) = \int_{D_X} g(x, y) f_{X|Y}(x|y) dx = \int_{D_X} g(x, y) \frac{p_{Y|X}(y|x) f_X(x)}{p_Y(y)} dx \quad (6.63)$$

se esistente. Le condizioni per l'esistenza della media condizionata sono le stesse del caso (b).

Se nelle (6.60), (6.61), (6.62) e (6.63) si sceglie $g(x, y) = x$, si ottiene $E(X | Y = y)$ cui si dà il nome di *media condizionata di X dato $Y = y$* . Se invece si pone $g(x, y) = x^n$ ($n = 2, 3, \dots$) si ottiene $E(X^n | Y = y)$, che è detto *momento condizionato di ordine n di X dato $Y = y$* . I momenti condizionati di ordine n di X dato $Y = y$ sono indicati nella Tabella 6.3. Inoltre, se $\mu(y) = E(X | Y = y)$ esiste finito e se nelle (6.60), (6.61), (6.62) e (6.63) si sceglie $g(x, y) = [x - \mu(y)]^n$ ($n = 2, 3, \dots$) si ottiene $E\{[X - \mu(y)]^n | Y = y\}$, detto *momento condizionato centrale di ordine n di X dato $Y = y$* . In particolare, così come accade in assenza di condizionamento, se $\mu(y)$ è finito, la *varianza condizionata di X dato*

Tabella 6.3 – Momenti condizionati di ordine n ($n = 1, 2, \dots$) di X dato $Y = y$.

(X, Y) discreto	
$E(X^n Y = y) = \sum_{\{r: x_r \in D_X\}} x_r^n p_{X Y}(x_r y) = \sum_{\{r: x_r \in D_X\}} x_r^n \frac{p_{X,Y}(x_r, y)}{p_Y(y)} \quad (y \in D_Y)$	
X discreta, Y assolutamente continua	
$E(X^n Y = y) = \sum_{\{r: x_r \in D_X\}} x_r^n p_{X Y}(x_r y) = \sum_{\{r: x_r \in D_X\}} x_r^n \frac{f_{Y X}(y x_r) p_X(x_r)}{f_Y(y)} \quad (y \in D_Y)$	
X assolutamente continua, Y discreta	
$E(X^n Y = y) = \int_{D_X} x^n f_{X Y}(x y) dx = \int_{D_X} x^n \frac{p_{Y X}(y x) f_X(x)}{p_Y(y)} dx \quad (y \in D_Y)$	
(X, Y) assolutamente continuo	
$E(X^n Y = y) = \int_{-\infty}^{+\infty} x^n f_{X Y}(x y) dx = \int_{-\infty}^{+\infty} x^n \frac{f_{X,Y}(x, y)}{f_Y(y)} dx \quad (y \in D_Y)$	

$Y = y$ è così definita:

$$\text{Var}(X | Y = y) = E\{[X - \mu(y)]^2 | Y = y\} = E(X^2 | Y = y) - [E(X | Y = y)]^2. \quad (6.64)$$

Esempio 6.13 Supponiamo che $X \sim \mathcal{P}(\lambda)$ descriva, in una qualche ragionevole approssimazione, il numero di decadimenti radioattivi prodotti da un certo materiale fisico e che ogni decadimento sia rilevato con probabilità p indipendentemente da ogni altro. Sia Y la variabile aleatoria descrivente il numero decadimenti rilevati. Ci proponiamo di determinare $E(X|Y = y)$ e $\text{Var}(X|Y = y)$, ossia il numero medio e la varianza dei decadimenti avvenuti sapendo che ne sono stati rilevati y .

Per le ipotesi fatte, per $x = 0, 1, \dots$ si ha:

$$p_{Y|X}(y|x) = \begin{cases} \binom{x}{y} p^y (1-p)^{x-y}, & y = 0, 1, \dots, x \\ 0, & \text{altrimenti.} \end{cases}$$

Per determinare $E(X|Y = y)$ occorre conoscere la funzione di probabilità condizionata di X dato $Y = y$. A tal fine osserviamo che dalla (6.11) per $y = 0, 1, \dots$ segue:

$$\begin{aligned} p_Y(y) &= \sum_{n=y}^{+\infty} p_{Y|X}(y|n) p_X(n) = \sum_{n=y}^{+\infty} \binom{n}{y} p^y (1-p)^{n-y} \frac{\lambda^n}{n!} e^{-\lambda} \\ &= \frac{(p\lambda)^y}{y!} e^{-\lambda} \sum_{n=y}^{+\infty} \frac{[(1-p)\lambda]^{n-y}}{(n-y)!} = \frac{(p\lambda)^y}{y!} e^{-\lambda} e^{(1-p)\lambda} = \frac{(p\lambda)^y}{y!} e^{-p\lambda}. \end{aligned}$$

Pertanto si ha $Y \sim \mathcal{P}(p\lambda)$. Inoltre, tenendo conto della (6.12), per ogni $y = 0, 1, \dots$ risulta:

$$p_{X|Y}(x|y) = \frac{p_{Y|X}(y|x)p_X(x)}{p_Y(y)} = \begin{cases} \frac{[(1-p)\lambda]^{x-y}}{(x-y)!} e^{-(1-p)\lambda}, & x = y, y+1, \dots \\ 0, & \text{altrimenti.} \end{cases}$$

Facendo, quindi, uso della (6.60) e della (6.64), per $y = 0, 1, \dots$ si trae:

$$\begin{aligned} E(X|Y = y) &= \sum_{k=y}^{+\infty} k \frac{[(1-p)\lambda]^{k-y}}{(k-y)!} e^{-(1-p)\lambda} = \sum_{r=0}^{+\infty} (r+y) \frac{[(1-p)\lambda]^r}{r!} e^{-(1-p)\lambda} \\ &= y + (1-p)\lambda, \\ \text{Var}(X|Y = y) &= \sum_{k=y}^{+\infty} [k - y - (1-p)\lambda]^2 \frac{[(1-p)\lambda]^{k-y}}{(k-y)!} e^{-(1-p)\lambda} \\ &= \sum_{r=0}^{+\infty} [r - (1-p)\lambda]^2 \frac{[(1-p)\lambda]^r}{r!} e^{-(1-p)\lambda} = (1-p)\lambda. \end{aligned}$$

◇

Esempio 6.14 Sia (X, Y) assolutamente continuo con densità congiunta:

$$f_{X,Y}(x, y) = \begin{cases} e^{-y}, & y > 0, 0 < x \leq y \\ 0, & \text{altrimenti.} \end{cases}$$

Risulta:

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx = \begin{cases} \int_0^y e^{-y} dx = y e^{-y}, & y > 0 \\ 0, & \text{altrimenti.} \end{cases}$$

Inoltre, facendo uso della (6.18), per ogni $y > 0$ si ha:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \begin{cases} 1/y, & 0 < x \leq y \\ 0, & \text{altrimenti.} \end{cases}$$

È quindi possibile determinare i momenti condizionati e la varianza di X dato $Y = y$. Infatti, dalla (6.61) per ogni $y > 0$ si ha:

$$E(X^n | Y = y) = \int_{-\infty}^{+\infty} x^n f_{X|Y}(x|y) dx = \frac{1}{y} \int_0^y x^n dx = \frac{y^n}{n+1} \quad (n = 1, 2, \dots),$$

da cui, facendo uso della (6.64), si ottiene:

$$\text{Var}(X | Y = y) = \frac{y^2}{3} - \frac{y^2}{4} = \frac{y^2}{12}.$$

◇

Esempio 6.15 Sia (X, Y) misto con X discreta e $Y \sim \mathcal{U}(0, 1)$. Supponiamo che per ogni $0 < y < 1$ risulti:

$$p_{X|Y}(x|y) = \begin{cases} \binom{n}{x} y^x (1-y)^{n-x}, & x = 0, 1, \dots, n \\ 0, & \text{altrimenti,} \end{cases}$$

ossia assumiamo che la funzione di probabilità condizionata di X dato $Y = y$ sia binomiale di parametri n e y . Dalle (6.62) e (6.64) per ogni $0 < y < 1$ segue:

$$\begin{aligned} E(X | Y = y) &= \sum_{k=0}^n k \binom{n}{k} y^k (1-y)^{n-k} = n y \\ \text{Var}(X | Y = y) &= E[(X - n y)^2 | Y = y] = n y (1 - y). \end{aligned}$$

◇

Esempio 6.16 Sia (X, Y) misto con X assolutamente continua e Y discreta. Supponiamo che sia:

$$p_Y(y) = \begin{cases} \frac{1}{y(y+1)}, & y = 1, 2, \dots \\ 0, & \text{altrimenti} \end{cases}$$

e che per ogni $y = 1, 2, \dots$ risulti:

$$f_{X|Y}(x|y) = \begin{cases} y(y+1) x(1-x)^{y-1}, & 0 < x < 1 \\ 0, & \text{altrimenti.} \end{cases}$$

Dalla (6.63) per ogni $y = 1, 2, \dots$ si ottiene:

$$\begin{aligned} E(X | Y = y) &= \int_0^1 x f_{X|Y}(x|y) dx = y(y+1) \int_0^1 x^2 (1-x)^{y-1} dx \\ &= y(y+1) \int_0^1 (1-z)^2 z^{y-1} dz = \frac{2}{y+2}. \end{aligned}$$

◇

Il seguente teorema illustra la proprietà di linearità di cui godono le medie condizionate.

Teorema 6.3 Sia (X, Y) un vettore aleatorio. Si considerino le variabili $Z_1 = g_1(X, Y)$ e $Z_2 = g_2(X, Y)$ con $g_i: \mathbb{R}^2 \rightarrow \mathbb{R}$ funzione Borel-misurabile ($i = 1, 2$) e siano α_1 e α_2 reali arbitrari. Risulta:

$$E(\alpha_1 Z_1 + \alpha_2 Z_2 | Y = y) = \alpha_1 E(Z_1 | Y = y) + \alpha_2 E(Z_2 | Y = y), \quad (6.65)$$

sempre che al secondo membro le medie condizionate esistano e che nella somma non siano presenti simultaneamente infiniti di segno opposto.

Sussistono inoltre i seguenti teoremi.

Teorema 6.4 *Siano X e Y variabili aleatorie indipendenti e sia $T = r(X)$ con $r: \mathbb{R} \rightarrow \mathbb{R}$ funzione Borel-misurabile. Se $E(T)$ esiste, si ha:*

$$E(T \mid Y = y) = E(T) \quad (6.66)$$

per tutti gli $y \in \mathbb{R}$ per i quali la media condizionata di T dato $Y = y$ è definita.

Teorema 6.5 *Sia (X, Y) un vettore aleatorio. Si considerino le variabili aleatorie $T = r(Y)$ e $Z = g(X, Y)$ con $r: \mathbb{R} \rightarrow \mathbb{R}$ e $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ funzioni Borel-misurabili. Si ha:*

$$E(TZ \mid Y = y) = r(y) E(Z \mid Y = y) \quad (6.67)$$

sempre che al secondo membro esista la media condizionata.

Le dimostrazioni dei precedenti tre teoremi sono state omesse in quanto seguono direttamente dalle definizioni (6.60), (6.61), (6.62) e (6.63) sfruttando le proprietà di linearità di somme e di integrali.

6.8 Valori medi delle medie condizionate

Dato un vettore aleatorio (X, Y) e una variabile aleatoria $Z = g(X, Y)$, con $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ funzione Borel-misurabile, la media condizionata $E(Z \mid Y = y)$ è, per sua stessa definizione, una funzione dei valori y assunti dalla variabile aleatoria Y . Per ogni y appartenente all'insieme D_Y dei valori che Y assume con probabilità non nulla nel caso discreto, o con densità di probabilità non nulla nel caso continuo, denotiamo con

$$h(Y) = E(Z \mid Y) = E[g(X, Y) \mid Y] \quad (6.68)$$

la funzione ottenuta facendo variare y sull'insieme D_Y . Se $E(Z)$ esiste finito, si può dimostrare che $h(Y)$ è una funzione misurabile che può quindi essere interpretata come una variabile aleatoria ottenuta a partire da Y . È naturale chiedersi quali siano le caratteristiche probabilistiche di $h(Y)$. Mentre la distribuzione di $h(Y)$ dipende dalla distribuzione del vettore aleatorio (X, Y) di modo che la sua determinazione deve essere effettuata esaminando le varie situazioni che caso per caso si presentano, è invece possibile fornire dei risultati generali per il valore medio di $h(Y)$.

Teorema 6.6 *Sia (X, Y) un vettore aleatorio e sia $Z = g(X, Y)$. Se $E(Z)$ è finito, allora*

$$E(Z) = E[E(Z \mid Y)]. \quad (6.69)$$

Dimostrazione Poiché $\mu(Y) = E(Z \mid Y)$ è una variabile aleatoria ottenuta come funzione di Y , la sua media può essere calcolata facendo uso dei risultati ottenuti nel Paragrafo 5.3. Utilizzando notazione e risultati del Paragrafo 6.5, dimostriamo il teorema nei casi in cui (X, Y) è discreto e in cui (X, Y) è assolutamente continuo, anche se la (6.69) è valida in situazioni più generali.

Caso (a) Se (X, Y) è discreto, ricordando la (5.22), si ha:

$$E[E(Z|Y)] = \sum_{\{s: y_s \in D_Y\}} E(Z | Y = y) p_Y(y). \quad (6.70)$$

Facendo ora uso della (6.7) e della (6.60) si ottiene:

$$\begin{aligned} E(Z) &= \sum_{\{r: x_r \in D_X\}} \sum_{\{s: y_s \in D_Y\}} g(x_r, y_s) p_{X,Y}(x_r, y_s) \\ &= \sum_{\{s: y_s \in D_Y\}} \left[\sum_{\{r: x_r \in D_X\}} g(x_r, y_s) p_{X|Y}(x_r | y_s) \right] p_Y(y_s) = \sum_{\{s: y_s \in D_Y\}} E(Z | Y = y) p_Y(y), \end{aligned}$$

che per la (6.70) si identifica con $E[E(Z|Y)]$.

Caso (b) Se (X, Y) è assolutamente continuo, ricordando la (5.23), si ha:

$$E[E(Z|Y)] = \int_{D_Y} E(Z | Y = y) f_Y(y) dy. \quad (6.71)$$

Quindi, utilizzando la (6.23) e la (6.61), si ricava:

$$\begin{aligned} E(Z) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f_{X,Y}(x, y) dx dy = \int_{D_Y} \left[\int_{-\infty}^{+\infty} g(x, y) f_{X|Y}(x | y) dx \right] f_Y(y) dy \\ &= \int_{D_Y} E(Z | Y = y) f_Y(y) dy, \end{aligned}$$

che, tenendo conto della (6.71), coincide con $E[E(Z|Y)]$. □

Se nella (6.69) si sceglie $g(x, y) = x$, si ottiene $E(X) = E[E(X|Y)]$, mentre se si pone $g(x, y) = x^n$ ($n = 2, 3, \dots$) si ottiene $E(X^n) = E[E(X^n | Y)]$. Ciò significa che è possibile esprimere il valore medio di X come valore medio della variabile aleatoria $E(X|Y)$, ed il momento di ordine n di X come valore medio della variabile aleatoria $E(X^n|Y)$ ($n = 2, 3, \dots$). Inoltre, se $\mu(y)$ esiste finito per ogni $y \in D_Y$ e se si sceglie $g(x, y) = [x - \mu(y)]^n$, la (6.69) diventa:

$$E\{[X - \mu(Y)]^n\} = E[E\{[X - \mu(Y)]^n | Y\}] \quad (n = 1, 2, \dots). \quad (6.72)$$

Si noti che il primo membro nella (6.72) non corrisponde al momento centrale di ordine n di X poiché $\mu(Y) = E(X|Y)$ non è il valore medio della variabile X ma è una variabile aleatoria che per ogni $y \in D_Y$ assume i valori $E(X | Y = y)$.

Corollario 6.1 Sia (X, Y) un vettore aleatorio. Si considerino le variabili $T = r(Y)$ e $Z = g(X, Y)$ con $r: \mathbb{R} \rightarrow \mathbb{R}$ e $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ funzioni Borel-misurabili. Se $E(TZ)$ esiste finito, risulta:

$$E(TZ) = E[T E(Z|Y)]. \quad (6.73)$$

Dimostrazione Dalla (6.69) si ha $E(TZ) = E[E(TZ|Y)]$. Poiché, per il Teorema 6.5, $E(TZ | Y = y) = r(y) E(Z | Y = y)$ per ogni $y \in D_Y$, si ricava immediatamente la (6.73). □

Se si denota con $\text{Var}(X|Y)$ una variabile aleatoria che per ogni $y \in D_Y$ assume come valori $\text{Var}(X | Y = y) = E\{[X - \mu(y)]^2 | Y = y\}$, sussistono i risultati che seguono.

Teorema 6.7 Sia (X, Y) un vettore aleatorio. Se $\mu(y) = E(X | Y = y)$ esiste finito per ogni $y \in D_Y$ e X ha varianza finita, si ha:

$$E[\text{Var}(X|Y)] = E\{[X - \mu(Y)]^2\} = E(X^2) - E[\mu^2(Y)] \quad (6.74)$$

$$\text{Var}[E(X|Y)] = E\{[\mu(Y) - E(X)]^2\} = E[\mu^2(Y)] - [E(X)]^2 \quad (6.75)$$

$$\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}[E(X|Y)]. \quad (6.76)$$

Dimostrazione Scegliendo $n = 2$ nella (6.72), per la proprietà di linearità della media si ha:

$$E[\text{Var}(X|Y)] = E\{[X - \mu(Y)]^2\} = E(X^2) + E[\mu^2(Y)] - 2E[X\mu(Y)].$$

Poiché per la (6.73) risulta $E[X\mu(Y)] = E[\mu(Y)E(X|Y)] = E[\mu^2(Y)]$, si ricava immediatamente la (6.74). Inoltre, facendo uso della (6.69), si ottiene:

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

da cui segue la (6.75). Infine, sommando membro a membro le (6.74) e (6.75), segue la (6.76). \square

Il Teorema 6.7 mostra quindi che la varianza di X può essere ottenuta come somma del valore medio della variabile aleatoria $\text{Var}(X|Y)$ e della varianza della variabile aleatoria $E(X|Y)$.

Esempio 6.17 Dei clienti pervengono l'uno dopo l'altro ad un centro per richiedere un servizio. Se il servizio non è immediatamente disponibile essi rimangono in attesa formando una coda. Assumiamo che il tempo richiesto per espletare un singolo servizio sia descrivibile mediante una variabile aleatoria Y esponenzialmente distribuita con parametro μ . Denotiamo con X la variabile aleatoria rappresentante il numero di clienti che perviene al centro durante un intervallo di tempo uguale a quello necessario per espletare il singolo servizio. Per ogni $y > 0$ supponiamo che si abbia:

$$p_{X|Y}(x|y) = \begin{cases} \frac{(\lambda y)^x}{x!} e^{-\lambda y}, & x = 0, 1, \dots \\ 0, & \text{altrimenti.} \end{cases}$$

Calcoliamo quanti clienti pervengono in media al centro durante detto intervallo di tempo.

Osserviamo a tal fine che dalla (6.40) per $x = 0, 1, \dots$ si ha:

$$\begin{aligned} p_X(x) &= \int_0^{+\infty} p_{X|Y}(x|y) f_Y(y) dy = \int_0^{+\infty} \frac{(\lambda y)^x}{x!} e^{-\lambda y} \mu e^{-\mu y} dy \\ &= \mu \frac{\lambda^x}{x!} \int_0^{+\infty} y^x e^{-(\lambda+\mu)y} dy = \frac{\mu}{\lambda + \mu} \frac{\lambda^x}{x!} \frac{x!}{(\lambda + \mu)^x}, \end{aligned}$$

dove l'ultima uguaglianza segue ricordando l'espressione dei momenti di una variabile aleatoria esponenziale (v. Paragrafo 5.8.3). Pertanto,

$$p_X(x) = \begin{cases} \frac{\mu}{\lambda + \mu} \left(\frac{\lambda}{\lambda + \mu} \right)^x, & x = 0, 1, \dots \\ 0, & \text{altrimenti,} \end{cases}$$

ossia X ha distribuzione di Pascal di parametro $p = \mu/(\lambda + \mu)$. Concludiamo quindi che risulta:

$$E(X) = \frac{1-p}{p} = \frac{\lambda}{\mu}.$$

Si noti che lo stesso risultato si può ottenere più rapidamente senza dover determinare la distribuzione di X . Infatti, essendo $E(X | Y = y) = \lambda y$ per ogni $y > 0$, applicando la (6.69) si ottiene:

$$E(X) = \int_0^{+\infty} E(X|Y = y) f_Y(y) dy = \int_0^{+\infty} \lambda y \mu e^{-\mu y} dy = \frac{\lambda}{\mu}.$$

◇

Esempio 6.18 Consideriamo un elaboratore elettronico il cui carico di lavoro è suddiviso in n distinte classi di priorità numerate da 1 a n . Denotiamo con Y la variabile aleatoria descrivente tali classi e assumiamo che si abbia:

$$p_Y(y) = \begin{cases} a_y, & y = 1, 2, \dots, n \\ 0, & \text{altrimenti,} \end{cases}$$

con $a_i > 0$ per $i = 1, 2, \dots, n$ e $\sum_{i=1}^n a_i = 1$. Sia X la variabile aleatoria descrivente il tempo di utilizzazione della CPU da parte di un job, e assumiamo che la densità condizionata di X dato che il job in esame appartiene alla classe i -esima sia esponenziale di parametro λ_i ($i = 1, 2, \dots, n$). In altri termini, assumiamo che per $i = 1, 2, \dots, n$ risulta:

$$f_{X|Y}(x|i) = \begin{cases} \lambda_i e^{-\lambda_i x}, & x > 0 \\ 0, & \text{altrimenti.} \end{cases}$$

Vogliamo calcolare il tempo medio di utilizzazione della CPU.

Essendo $E(X | Y = i) = 1/\lambda_i$ per ogni $i = 1, 2, \dots, n$, dalla (6.69) si ha:

$$E(X) = \sum_{i=1}^n p_Y(i) E(X|Y = i) = \sum_{i=1}^n \frac{a_i}{\lambda_i}.$$

È opportuno osservare che nello schema proposto la variabile aleatoria X ha distribuzione iperesponenziale (v. Paragrafo 4.4.4). Per convincersene è sufficiente applicare la (6.45) da cui si ottiene:

$$f_X(x) = \begin{cases} \sum_{i=1}^n f_{X|Y}(x|i) p_Y(i) = \sum_{i=1}^n a_i \lambda_i e^{-\lambda_i x}, & x > 0 \\ 0, & \text{altrimenti.} \end{cases}$$

◇

Esempio 6.19 Sia Z_1, Z_2, \dots una sequenza di variabili aleatorie indipendenti e identicamente distribuite con valore medio μ e varianza σ^2 entrambi finiti. Denotiamo con N una variabile aleatoria, indipendente da ognuna delle Z_i , che assume valori interi positivi. Calcoliamo la media e la varianza di $X = Z_1 + Z_2 + \dots + Z_N$.

Facendo uso dell'indipendenza di N dalle Z_i nonché della mutua indipendenza di Z_1, Z_2, \dots , per $n = 0, 1, \dots$ si ha:

$$\mu(n) = E(X \mid N = n) = n\mu, \quad \text{Var}(X \mid N = n) = n\sigma^2.$$

Per l'indipendenza delle variabili coinvolte, dalla (6.69) si ha poi:

$$E(X) = \sum_{n=0}^{+\infty} E(X \mid N = n) p_X(n) = \sum_{n=0}^{+\infty} n\mu p_X(n) = \mu E(N).$$

Inoltre, utilizzando ancora l'indipendenza, dalle (6.69), (6.74) e (6.75) si ottiene:

$$\begin{aligned} E[\text{Var}(X|N)] &= \sum_{n=1}^{+\infty} \text{Var}(X \mid N = n) p_X(n) = \sigma^2 E(N), \\ \text{Var}[E(X|N)] &= E\{[\mu(N) - E(X)]^2\} = \sum_{n=1}^{+\infty} E\{[\mu(N) - E(X)]^2 \mid N = n\} p_X(n) \\ &= \sum_{n=1}^{+\infty} [n\mu - \mu E(N)]^2 p_X(n) = \mu^2 \text{Var}(N), \end{aligned}$$

da cui, per la (6.76), segue:

$$\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}[E(X|Y)] = \sigma^2 E(N) + \mu^2 \text{Var}(N).$$

◇

6.9 La densità normale bivariata

La densità di probabilità normale bivariata è una generalizzazione per vettori aleatori bidimensionali della densità di probabilità normale discussa nel Paragrafo 4.4.6.

Definizione 6.11 Un vettore aleatorio (X, Y) di densità di probabilità

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\varrho^2}} \exp\left\{-\frac{1}{2(1-\varrho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\varrho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)\right]\right\} \quad (x, y \in \mathbb{R}), \quad (6.77)$$

con $\mu_X \in \mathbb{R}, \mu_Y \in \mathbb{R}, \sigma_X > 0, \sigma_Y > 0$ e $-1 < \varrho < 1$, si dice di distribuzione normale di parametri $\mu_X, \mu_Y, \sigma_X, \sigma_Y, \varrho$.

La densità normale bivariata soddisfa alcune proprietà, qui di seguito esaminate, che la rendono particolarmente interessante nel contesto delle distribuzioni condizionate.

Proposizione 6.1 *Se (X, Y) è un vettore aleatorio di densità di probabilità (6.77), allora $X \sim \mathcal{N}(\mu_X, \sigma_X)$ e $Y \sim \mathcal{N}(\mu_Y, \sigma_Y)$.*

Dimostrazione Calcoliamo la densità di probabilità $f_X(x)$ della variabile aleatoria X . Dalla (3.47) risulta:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\varrho^2}} \exp\left\{-\frac{1}{2(1-\varrho^2)}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right\} \\ \times \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2(1-\varrho^2)}\left[-2\varrho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\} dy.$$

Completando il quadrato rispetto a x nell'esponente si ha:

$$f_X(x) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\varrho^2}} \exp\left\{-\frac{1}{2(1-\varrho^2)}\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \frac{\varrho^2}{2(1-\varrho^2)}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right\} \\ \times \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2(1-\varrho^2)}\left[\frac{y-\mu_Y}{\sigma_Y} - \varrho\frac{x-\mu_X}{\sigma_X}\right]^2\right\} dy \\ = \frac{1}{\sigma_X\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right\} \quad (x \in \mathbb{R}). \quad (6.78)$$

Quindi X ha distribuzione normale i cui parametri μ_X e σ_X^2 , come visto nell'Esempio 5.25, coincidono rispettivamente con media e varianza di X . Un procedimento del tutto analogo permette di dimostrare che risulta:

$$f_Y(y) = \frac{1}{\sigma_Y\sqrt{2\pi}} \exp\left\{-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}\right\} \quad (y \in \mathbb{R}), \quad (6.79)$$

e di qui concludere che Y ha distribuzione normale di parametri μ_Y , coincidente con la media e σ_Y^2 rappresentante la varianza. \square

In Figura 6.1 è rappresentata la densità normale bivariata per $\mu_X = \mu_Y = 0$, $\sigma_X = \sigma_Y = 1$ e $\varrho = 0.9, 0.5, 0, -0.9$.

Proposizione 6.2 *Sia (X, Y) un vettore aleatorio normale di densità di probabilità congiunta (6.77). Le densità di probabilità condizionate $f_{X|Y}(x|y)$ e $f_{Y|X}(y|x)$ sono normali e per $x, y \in \mathbb{R}$ risulta:*

$$f_{X|Y}(x|y) = \frac{1}{\sigma_X\sqrt{2\pi(1-\varrho^2)}} \exp\left\{-\frac{1}{2\sigma_X^2(1-\varrho^2)}\left[x - \mu_X - \varrho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y)\right]^2\right\} \quad (6.80)$$

$$f_{Y|X}(y|x) = \frac{1}{\sigma_Y\sqrt{2\pi(1-\varrho^2)}} \exp\left\{-\frac{1}{2\sigma_Y^2(1-\varrho^2)}\left[y - \mu_Y - \varrho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)\right]^2\right\}. \quad (6.81)$$

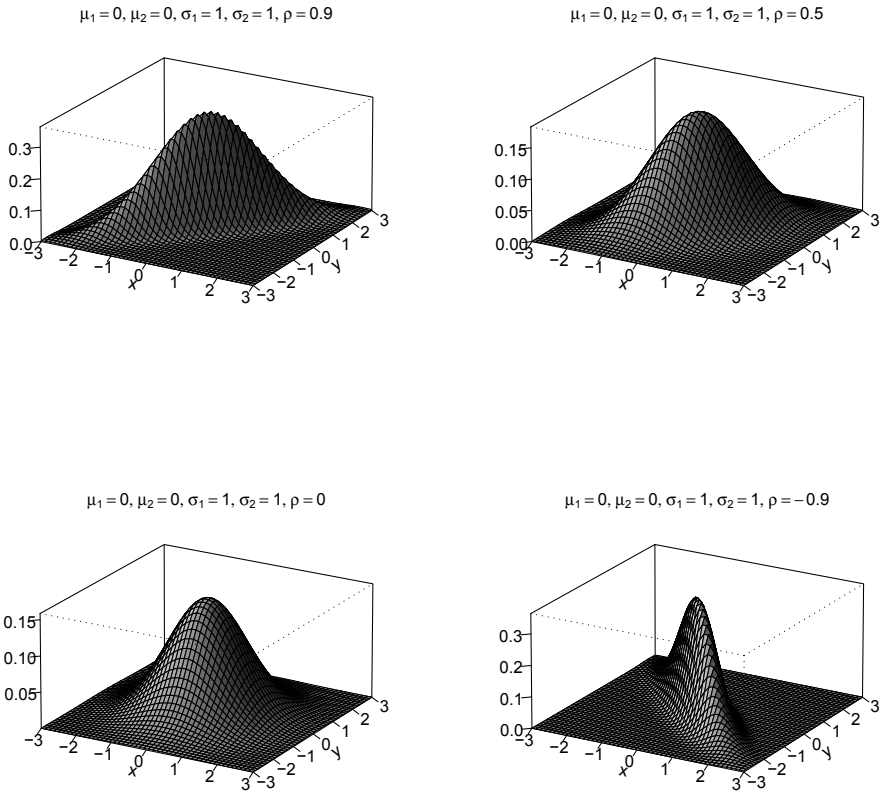


Figura 6.1 – Rappresentazioni grafiche di densità normali bivariate.

Inoltre, le medie e le varianze condizionate sono:

$$E(X|Y = y) = \mu_X + \varrho \frac{\sigma_X}{\sigma_Y} (y - \mu_Y), \quad \text{Var}(X|Y = y) = \sigma_X^2 (1 - \varrho^2), \quad (6.82)$$

$$E(Y|X = x) = \mu_Y + \varrho \frac{\sigma_Y}{\sigma_X} (x - \mu_X), \quad \text{Var}(Y|X = x) = \sigma_Y^2 (1 - \varrho^2). \quad (6.83)$$

Dimostrazione Ricordando la (6.18), dalle (6.77) e (6.79) segue la (6.80). Analogamente, facendo uso della (6.21) dalle (6.77) e (6.78) segue la (6.81). Pertanto, le densità di probabilità condizionate di X dato $Y = y$ e di Y dato $X = x$ nel caso in esame sono normali. In virtù delle proprietà della densità normale si conclude che la media e la varianza di X dato $Y = y$ sono date da (6.82), mentre per la variabile Y condizionata da $X = x$ sussistono le (6.83). \square

Nel caso di un vettore aleatorio bidimensionale normale, $E(X|Y = y)$ e $E(Y|X = x)$ sono funzioni lineari rispettivamente di y e di x , dette in Statistica *curve di regressione*.

Proposizione 6.3 Se (X, Y) è un vettore aleatorio di densità di probabilità congiunta (6.77) la covarianza ed il coefficiente di correlazione di X e Y sono rispettivamente:

$$\text{Cov}(X, Y) = \sigma_X \sigma_Y \varrho, \quad \varrho(X, Y) = \varrho. \quad (6.84)$$

Dimostrazione Calcoliamo la covarianza. Ricordando la (5.69), dalla (6.77) si ricava:

$$\begin{aligned} \text{Cov}(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \mu_X)(y - \mu_Y) f_{X,Y}(x, y) dx dy \\ &= \frac{[2\sigma_X^2(1 - \varrho^2)][2\sigma_Y^2(1 - \varrho^2)]}{2\pi\sigma_X\sigma_Y\sqrt{1 - \varrho^2}} \int_{-\infty}^{+\infty} dz_1 \int_{-\infty}^{+\infty} z_1 z_2 \exp\left\{-[z_1^2 - 2\varrho z_1 z_2 + z_2^2]\right\} dz_2 \end{aligned}$$

dove l'ultima uguaglianza segue effettuando il cambiamento di variabili di integrazione

$$z_1 = (x - \mu_X)/\sqrt{2\sigma_X^2(1 - \varrho^2)}, \quad z_2 = (y - \mu_Y)/\sqrt{2\sigma_Y^2(1 - \varrho^2)}.$$

Completando il quadrato rispetto a z_2 nell'esponente si ha:

$$\begin{aligned} \text{Cov}(X, Y) &= \frac{2\sigma_X\sigma_Y(1 - \varrho^2)^{3/2}}{\pi} \int_{-\infty}^{+\infty} dz_1 z_1 \exp\left\{-(1 - \varrho^2)z_1^2\right\} \\ &\quad \times \int_{-\infty}^{+\infty} z_2 \exp\left\{-(z_2 - \varrho z_1)^2\right\} dz_2 \\ &= \frac{2\sigma_X\sigma_Y(1 - \varrho^2)^{3/2}}{\pi} \int_{-\infty}^{+\infty} dz_1 z_1 \exp\left\{-(1 - \varrho^2)z_1^2\right\} \\ &\quad \times \int_{-\infty}^{+\infty} (y + \varrho z_1) e^{-y^2} dy \\ &= \frac{2\sigma_X\sigma_Y\varrho(1 - \varrho^2)^{3/2}}{\pi} \int_{-\infty}^{+\infty} z_1^2 \exp\left\{-(1 - \varrho^2)z_1^2\right\} dz_1 \\ &= \frac{2\sigma_X\sigma_Y\varrho}{\pi} \int_{-\infty}^{+\infty} x^2 e^{-x^2} dx = \sigma_X\sigma_Y\varrho. \end{aligned}$$

Da quest'ultima, in virtù della (5.75), segue immediatamente che il coefficiente di correlazione del vettore normale in esame è ϱ . \square

Il Teorema 5.11 mostra che se due variabili aleatorie sono indipendenti, esse sono anche non correlate e quindi hanno coefficiente di correlazione nullo. Nel caso di variabili normali, eccezionalmente (v. Paragrafo 5.6) vale anche il viceversa, ossia:

Proposizione 6.4 *Due variabili aleatorie X e Y di densità di probabilità congiunta (6.77) sono indipendenti se e solo se sono non correlate.*

Dimostrazione È sufficiente mostrare che se il coefficiente di correlazione, ossia ϱ , è nullo le variabili aleatorie X e Y sono indipendenti. In questo caso ponendo $\varrho = 0$ nella (6.77) si ottiene $f_{X,Y}(x, y) = f_X(x) f_Y(y)$, che mostra l'indipendenza delle variabili X e Y . \square

Capitolo 7

Densità di probabilità speciali e loro proprietà

7.1 Introduzione

In questo capitolo fisseremo l'attenzione su alcune densità di probabilità che giocano ruolo rilevante in contesti applicativi: le densità beta, chi-quadrato, di Fisher e di Student.

7.2 Statistiche ordinate e distribuzione beta

Siano X_1, X_2, \dots, X_n variabili aleatorie indipendenti e identicamente distribuite con funzione di distribuzione $F(x)$. Denotiamo con $Y_1 = \min(X_1, X_2, \dots, X_n)$ e con $Y_n = \max(X_1, X_2, \dots, X_n)$ le variabili aleatorie rappresentanti rispettivamente il minimo ed il massimo dei valori assunti dalle variabili X_1, X_2, \dots, X_n . Sia inoltre Y_k ($k = 2, \dots, n-1$) la variabile aleatoria denotante il valore di X_1, X_2, \dots, X_n che si colloca al k -esimo posto tra tutti i valori ordinati in maniera crescente tra i suddetti minimo e massimo. Le variabili aleatorie Y_1, Y_2, \dots, Y_n sono dette *variabili ordinate*; in statistica prendono il nome di *statistiche ordinate*. Essendo X_1, X_2, \dots, X_n indipendenti e identicamente distribuite, dalle (3.63) e (3.64) per ogni $y \in \mathbb{R}$ si ha:

$$F_{Y_1}(y) = 1 - [1 - F(y)]^n, \quad F_{Y_n}(y) = [F(y)]^n, \quad (7.1)$$

dove per $i = 1, 2, \dots, n$ si è posto $F(y) = F_{X_i}(y)$. Per determinare la funzione di distribuzione di Y_k osserviamo che l'evento $\{Y_k \leq y\}$ coincide con l'evento “*si verificano almeno k degli n eventi $\{X_1 \leq y\}, \{X_2 \leq y\}, \dots, \{X_n \leq y\}$ ”.*

Poiché questi sono indipendenti e si verificano ciascuno con probabilità $F(y)$, ricordando la distribuzione binomiale (v. Paragrafo 4.2.3), per ogni $y \in \mathbb{R}$ risulta:

$$F_{Y_k}(y) = P(Y_k \leq y) = \sum_{r=k}^n \binom{n}{r} [F(y)]^r [1 - F(y)]^{n-r} \quad (k = 1, 2, \dots, n). \quad (7.2)$$

Si noti che ponendo $k = 1$ e $k = n$ nella (7.2) si ritrovano rispettivamente le funzioni di distribuzione del minimo e del massimo di X_1, X_2, \dots, X_n , indicate in (7.1). Se poi le X_i

sono assolutamente continue con densità di probabilità $f(y)$, dalla (7.2) si ottiene:

$$f_{Y_k}(y) = \binom{n}{k} k f(y) [F(y)]^{k-1} [1 - F(y)]^{n-k} \quad (k = 1, 2, \dots, n). \quad (7.3)$$

Infatti, in tal caso dalla (7.2) si ha:

$$\begin{aligned} f_{Y_k}(y) &= \frac{d}{dy} F_{Y_k}(y) = \sum_{r=k}^n \binom{n}{r} r f(y) [F(y)]^{r-1} [1 - F(y)]^{n-r} \\ &\quad - \sum_{r=k}^{n-1} \binom{n}{r} [F(y)]^r (n-r) f(y) [1 - F(y)]^{n-r-1}. \end{aligned}$$

Isolando nella prima sommatoria il termine corrispondente a $r = k$ ed effettuando nella seconda sommatoria il cambiamento di indice $s = r + 1$ si ottiene:

$$\begin{aligned} f_{Y_k}(y) &= \binom{n}{k} k f(y) [F(y)]^{k-1} [1 - F(y)]^{n-k} \\ &\quad + \sum_{r=k+1}^n \left\{ r \binom{n}{r} - (n-r+1) \binom{n}{r-1} \right\} f(y) [F(y)]^{r-1} [1 - F(y)]^{n-r}, \end{aligned}$$

che conduce alla (7.3) essendo nulla la differenza in parentesi graffa.

Supponiamo ora che X_1, X_2, \dots, X_n siano indipendenti ed uniformemente distribuite nell'intervallo $(0, 1)$. Dalla (7.3) per $k = 1, 2, \dots, n$ si ricava allora:

$$f_{Y_k}(y) = \begin{cases} \binom{n}{k} k y^{k-1} (1-y)^{n-k}, & 0 < y < 1 \\ 0, & \text{altrimenti.} \end{cases} \quad (7.4)$$

La densità di probabilità (7.4) è un caso particolare della densità di probabilità beta che passiamo a definire.

Definizione 7.1 Una variabile aleatoria X di densità di probabilità

$$f_X(x) = \begin{cases} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, & 0 < x < 1 \\ 0, & \text{altrimenti,} \end{cases} \quad (7.5)$$

con α, β reali positivi e con $\Gamma(\nu)$ definita in (4.41), si dice avere distribuzione beta di parametri α e β .

La locuzione “beta” deriva dalla presenza in (7.5) della funzione beta di Eulero di parametri α e β così definita:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha) \Gamma(\beta)}{\Gamma(\alpha + \beta)} = \int_0^1 y^{\alpha-1} (1-y)^{\beta-1} dy \quad (\alpha > 0, \beta > 0). \quad (7.6)$$

Facendo uso della (7.6), si può immediatamente verificare che la (7.5) è effettivamente una densità di probabilità per ogni $\alpha > 0$ e $\beta > 0$. Nel seguito la scrittura $X \sim \mathcal{Be}(\alpha, \beta)$ indicherà che X ha distribuzione beta di parametri α e β .

Si noti che ponendo $\alpha = k$ e $\beta = n - k + 1$ nella (7.5) si ottiene la (7.4). Infatti, ricordando che $\Gamma(v) = (v-1)!$ per $v = 1, 2, \dots$, si ha:

$$\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)} = \frac{n!}{(k-1)!(n-k)!} = k \binom{n}{k}.$$

Se risulta invece $\alpha = \beta = 1$, la (7.5) diventa una densità di probabilità uniforme nell'intervallo $(0, 1)$.

La densità beta trova svariate utilizzazioni in probabilità e in statistica grazie alla flessibilità della sua forma al variare dei parametri. In particolare, dalla (7.5) segue:

$$\lim_{x \rightarrow 0} f_X(x) = \begin{cases} 0, & \alpha > 1 \\ \beta, & \alpha = 1 \\ +\infty, & 0 < \alpha < 1, \end{cases} \quad \lim_{x \rightarrow 1} f_X(x) = \begin{cases} 0, & \beta > 1 \\ \alpha, & \beta = 1 \\ +\infty, & 0 < \beta < 1, \end{cases}$$

ed inoltre, per ogni $0 < x < 1$, risulta:

$$\frac{d}{dx} f_X(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-2} (1-x)^{\beta-2} [\alpha - 1 - x(\alpha + \beta - 2)].$$

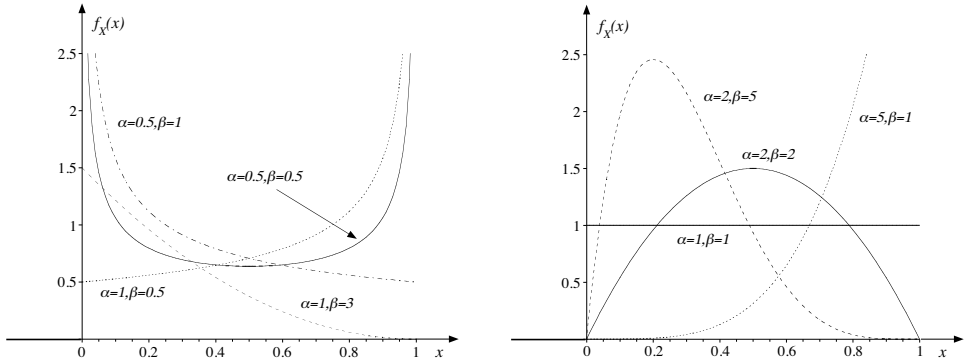
Ciò mostra che la densità (7.5) può esibire i differenti tipi di comportamenti indicati nella Tabella 7.1.

Tabella 7.1 – Comportamenti della densità di probabilità beta al variare dei parametri.

Condizioni	Comportamenti di $f_X(x)$ nell'intervallo $(0, 1)$
$0 < \alpha < 1, 0 < \beta < 1$	diverge positivamente sia quando $x \rightarrow 0$ sia quando $x \rightarrow 1$ e presenta un minimo nel punto $(1 - \alpha)/(2 - \alpha - \beta)$
$0 < \alpha < 1, \beta \geq 1$	diverge positivamente quando $x \rightarrow 0$ ed è monotona decrescente nell'intervallo $(0, 1)$
$\alpha \geq 1, 0 < \beta < 1$	è monotona crescente nell'intervallo $(0, 1)$ e diverge positivamente quando $x \rightarrow 1$
$\alpha = 1, \beta = 1$	è unitaria nell'intervallo $(0, 1)$
$\alpha = 1, \beta > 1$	è monotona decrescente nell'intervallo $(0, 1)$
$\alpha > 1, \beta = 1$	è monotona crescente nell'intervallo $(0, 1)$
$\alpha > 1, \beta > 1$	presenta un massimo nel punto $(\alpha - 1)/(\alpha + \beta - 2)$

In Figura 7.1 è rappresentata la densità di probabilità beta (7.5) per differenti scelte dei parametri α e β .

La funzione generatrice dei momenti della distribuzione beta non è suscettibile di una semplice rappresentazione. Si possono tuttavia calcolare i momenti in base alla loro stessa

Figura 7.1 – Densità di probabilità di $X \sim \mathcal{Be}(\alpha, \beta)$.

definizione:

$$\begin{aligned} E(X^n) &= \int_0^1 x^n f_X(x) dx = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} \int_0^1 x^{n+\alpha-1} (1-x)^{\beta-1} dx \\ &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} B(\alpha + n, \beta) = \frac{\Gamma(\alpha + \beta) \Gamma(\alpha + n)}{\Gamma(\alpha) \Gamma(\alpha + \beta + n)} \quad (n = 1, 2, \dots). \end{aligned} \quad (7.7)$$

In particolare, ricordando la (4.42), è possibile ricavare dalla (7.7) valore medio e varianza:

$$E(X) = \frac{\Gamma(\alpha + \beta) \Gamma(\alpha + 1)}{\Gamma(\alpha) \Gamma(\alpha + \beta + 1)} = \frac{\alpha}{\alpha + \beta}, \quad (7.8)$$

$$\begin{aligned} \text{Var}(X) &= E(X^2) - [E(X)]^2 = \frac{\Gamma(\alpha + \beta) \Gamma(\alpha + 2)}{\Gamma(\alpha) \Gamma(\alpha + \beta + 2)} - \left(\frac{\alpha}{\alpha + \beta} \right)^2 \\ &= \frac{\alpha(\alpha + 1)}{(\alpha + \beta)(\alpha + \beta + 1)} - \left(\frac{\alpha}{\alpha + \beta} \right)^2 = \frac{\alpha\beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}. \end{aligned} \quad (7.9)$$

Il seguente teorema indica talune relazioni esistenti tra variabili aleatorie di densità gamma e di densità beta.

Teorema 7.1 Se Z_1 e Z_2 sono variabili aleatorie indipendenti con $Z_1 \sim \mathcal{G}(\alpha, \lambda)$ e $Z_2 \sim \mathcal{G}(\beta, \lambda)$, allora le variabili aleatorie

$$X = \frac{Z_1}{Z_1 + Z_2}, \quad Y = Z_1 + Z_2 \quad (7.10)$$

sono indipendenti e risulta $X \sim \mathcal{Be}(\alpha, \beta)$ e $Y \sim \mathcal{G}(\alpha + \beta, \lambda)$.

Dimostrazione Determiniamo in primo luogo la densità di probabilità congiunta del vettore aleatorio

$$(X, Y) = \left(\frac{Z_1}{Z_1 + Z_2}, Z_1 + Z_2 \right)$$

e successivamente le densità di probabilità di ognuna delle sue componenti. Essendo Z_1 e Z_2 indipendenti, ricordando l'espressione (4.40) della densità gamma si ha:

$$f_{Z_1, Z_2}(z_1, z_2) = \begin{cases} \frac{\lambda^\alpha z_1^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda z_1} \frac{\lambda^\beta z_2^{\beta-1}}{\Gamma(\beta)} e^{-\lambda z_2}, & z_1 > 0, z_2 > 0 \\ 0, & \text{altrimenti.} \end{cases}$$

Per determinare la densità congiunta $f_{X,Y}(x, y)$ ricorriamo al Teorema 3.8. Consideriamo pertanto la trasformazione

$$x = g_1(z_1, z_2) = \frac{z_1}{z_1 + z_2}, \quad y = g_2(z_1, z_2) = z_1 + z_2.$$

Per ogni coppia (z_1, z_2) con $z_1 > 0$ e $z_2 > 0$ risulta $0 < x < 1$ e $y > 0$. Quindi la trasformazione inversa

$$z_1 = h_1(x, y) = xy, \quad z_2 = h_2(x, y) = (1-x)y$$

ha dominio $D = \{(x, y): 0 < x < 1, y > 0\}$. Lo Jacobiano della trasformazione

$$J_h(x, y) = \frac{\partial(h_1, h_2)}{\partial(x, y)} = \begin{vmatrix} y & x \\ -y & 1-x \end{vmatrix} = y$$

è non nullo per ogni $(x, y) \in D$. Essendo soddisfatte le ipotesi del Teorema 3.8, si ha:

$$f_{X,Y}(x, y) = \begin{cases} \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} y^{\alpha+\beta-1} e^{-\lambda y}, & 0 < x < 1, y > 0 \\ 0, & \text{altrimenti.} \end{cases}$$

La conoscenza della densità di (X, Y) permette di ricavare le densità di probabilità di X e di Y . Infatti, se $0 < x < 1$ si ha:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f_{XY}(x, y) dy = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \int_0^{+\infty} y^{\alpha+\beta-1} e^{-\lambda y} dy \\ &= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \int_0^{+\infty} u^{\alpha+\beta-1} e^{-u} du = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, \end{aligned}$$

dove l'ultima uguaglianza segue dalla (4.41). Facendo poi uso della (7.6), per $y > 0$ si ricava:

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{+\infty} f_{XY}(x, y) dx = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha+\beta-1} e^{-\lambda y} \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha+\beta-1} e^{-\lambda y} B(\alpha, \beta) = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)} y^{\alpha+\beta-1} e^{-\lambda y}. \end{aligned}$$

Possiamo quindi affermare che $X \sim \mathcal{B}e(\alpha, \beta)$ e che $Y \sim \mathcal{G}(\alpha + \beta, \lambda)$. In conclusione, essendo $f_{X,Y}(x, y) = f_X(x) f_Y(y)$ per ogni $(x, y) \in \mathbb{R}^2$, le variabili aleatorie X e Y sono indipendenti. \square

Si noti che l'indipendenza di X e Y , espressa dal Teorema 7.1, è conseguenza dell'ipotesi che Z_1 e Z_2 sono di distribuzioni gamma. Scegliendo altre distribuzioni di probabilità per le variabili indipendenti Z_1 e Z_2 , usualmente accade che X e Y non sono più indipendenti. Un ulteriore interessante risultato è il seguente:

$$E(X) = \frac{E(Z_1)}{E(Z_1) + E(Z_2)};$$

esso deriva dalla (7.8) e dalla circostanza che $E(Z_1) = \alpha/\lambda$ e $E(Z_2) = \beta/\lambda$ (v. Paragrafo 5.8.3).

7.3 Distribuzione chi-quadrato

Nel caso $\nu = n/2$ e $\lambda = 1/2$, con n intero positivo, la densità gamma specificata nella (4.40) prende il nome di *densità chi-quadrato* ed al parametro n si dà il nome di *numero di gradi di libertà*. Precisamente, si ha la seguente

Definizione 7.2 Una variabile aleatoria X di densità di probabilità

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{n/2} x^{n/2-1} e^{-x/2}, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (7.11)$$

con n intero positivo e con $\Gamma(\nu)$ definita in (4.41), si dice di *distribuzione chi-quadrato con n gradi di libertà*.

Nel seguito con $X \sim \chi^2(n)$ intenderemo che X ha distribuzione chi-quadrato con n gradi di libertà. In Figura 7.2 è rappresentata la densità di probabilità chi-quadrato (7.11) per $n = 1, 3, 5, 7$.

Il seguente teorema evidenzia il ruolo giocato dal parametro n .

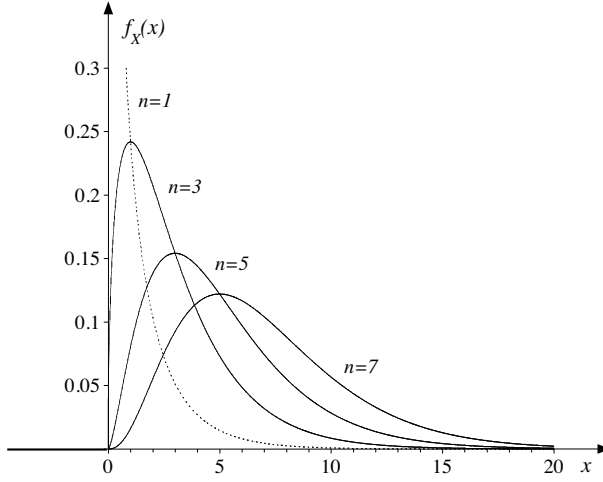
Teorema 7.2 Siano X_1, X_2, \dots, X_n variabili aleatorie indipendenti, con $X_i \sim \mathcal{N}(0, 1)$ per $i = 1, 2, \dots, n$. Allora, $Y_n = X_1^2 + X_2^2 + \dots + X_n^2$ ha distribuzione chi-quadrato con n gradi di libertà.

Dimostrazione Calcoliamo in primo luogo la funzione di distribuzione $F_Y(y)$ di $Y = X^2$, con $X \sim \mathcal{N}(0, 1)$. Per $y < 0$ si ha $F_Y(y) = 0$, mentre per $y \geq 0$ risulta:

$$F_Y(y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x) dx = 2\Phi(\sqrt{y}) - 1,$$

con $\Phi(z)$ definita in (4.54). La densità di probabilità di $Y = X^2$ è pertanto:

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi y}} e^{-y/2}, & y > 0 \\ 0, & y \leq 0. \end{cases} \quad (7.12)$$

Figura 7.2 – Densità di probabilità di $X \sim \chi^2(n)$.

Si noti che $Y \sim \mathcal{G}(1/2, 1/2)$. Ciò segue ponendo $\nu = 1/2$ e $\lambda = 1/2$ nella (4.40) e ricordando che sussiste l'identità:

$$\Gamma\left(\frac{1}{2}\right) = \int_0^{+\infty} x^{-1/2} e^{-x} dx = \sqrt{2} \int_0^{+\infty} e^{-y^2/2} dy = \frac{1}{\sqrt{2}} \int_{-\infty}^{+\infty} e^{-y^2/2} dy = \sqrt{\pi}, \quad (7.13)$$

ottenuta dalla (4.41) effettuando il cambiamento di variabile di integrazione $x = y^2/2$ e facendo poi ricorso alle proprietà della densità normale standard. Avendo Y distribuzione gamma di parametri $\nu = 1/2$ e $\lambda = 1/2$, la sua funzione generatrice dei momenti è (v. Paragrafo 5.8.3):

$$M_Y(s) = \left(\frac{\lambda}{\lambda - s}\right)^\nu = (1 - 2s)^{-1/2} \quad (s < 1/2). \quad (7.14)$$

Consideriamo ora le variabili aleatorie Z_1, Z_2, \dots, Z_n , con $Z_i = X_i^2$ per $i = 1, 2, \dots, n$. Per il Teorema 3.7 esse sono indipendenti e, per quanto poc'anzi dimostrato, risulta $Z_i \sim \mathcal{G}(1/2, 1/2)$ per $i = 1, 2, \dots, n$. Pertanto, si può scrivere:

$$Y_n = X_1^2 + X_2^2 + \dots + X_n^2 = Z_1 + Z_2 + \dots + Z_n. \quad (7.15)$$

Dal Teorema 5.19, facendo uso della (7.14), si ottiene poi:

$$M_{Y_n}(s) = M_{Z_1}(s) M_{Z_2}(s) \cdots M_{Z_n}(s) = (1 - 2s)^{-n/2} \quad (s < 1/2), \quad (7.16)$$

che si riconosce essere la funzione generatrice dei momenti di una variabile di densità gamma di parametri $\nu = n/2$ e $\lambda = 1/2$. Si può quindi in definitiva affermare che risulta $Y_n \sim \chi^2(n)$. \square

Il Teorema 7.2 afferma che la somma dei quadrati di variabili aleatorie normali standard indipendenti ha distribuzione chi-quadrato con un numero di gradi di libertà uguale al numero

degli addendi. Quindi, la denominazione “numero di gradi di libertà”, attribuita al parametro n , assume il significato di numero di addendi indipendenti presenti nella somma.

La densità chi-quadrato (7.11) contiene un solo parametro: il numero n di gradi di libertà. Inoltre, poiché per ogni $x > 0$ si ha

$$\frac{d}{dy} f_X(x) = \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{n/2+1} x^{n/2-2} e^{-x/2} [(n-2) - x],$$

la funzione densità chi-quadrato con n gradi di libertà è strettamente decrescente per $n = 1, 2$, mentre per $n > 2$ presenta un unico punto di massimo in $x = n - 2$.

Vogliamo ora calcolare i momenti di $X \sim \chi^2(n)$. Essendo X non negativa, essi esistono e dalla (5.39) sono dati da

$$\begin{aligned} E(X^k) &= \int_0^{+\infty} x^k f_X(x) dx = \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{n/2} \int_0^{+\infty} x^{n/2+k-1} e^{-x/2} dx \\ &= \frac{2^k}{\Gamma(n/2)} \int_0^{+\infty} z^{n/2+k-1} e^{-z} dz = \frac{2^k \Gamma(n/2 + k)}{\Gamma(n/2)} \quad (k = 1, 2, \dots), \end{aligned} \quad (7.17)$$

dove l'ultima uguaglianza segue per la (4.41). In particolare, dalla (7.17) si ottiene:

$$E(X) = \frac{2\Gamma(n/2 + 1)}{\Gamma(n/2)} = n, \quad E(X^2) = \frac{4\Gamma(n/2 + 2)}{\Gamma(n/2)} = n(n+2), \quad \text{Var}(X) = 2n.$$

Osservazione 7.1 Sia $X \sim \chi^2(n)$. Determiniamo la distribuzione e i momenti della variabile aleatoria $Y = 1/X$. Dal Teorema 3.4 segue:

$$f_Y(y) = f_X\left(\frac{1}{y}\right) \frac{1}{y^2} = \begin{cases} \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{n/2} \left(\frac{1}{y}\right)^{n/2+1} \exp\left\{-\frac{1}{2y}\right\}, & y > 0 \\ 0, & y \leq 0. \end{cases} \quad (7.18)$$

Inoltre, dal Teorema 5.1 si ha:

$$\begin{aligned} E(Y^k) &= E(X^{-k}) = \int_0^{+\infty} \frac{1}{x^k} f_X(x) dx = \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{n/2} \int_0^{+\infty} x^{n/2-1-k} e^{-x/2} dx \\ &= \frac{1}{2^k \Gamma(n/2)} \int_0^{+\infty} z^{n/2-k-1} e^{-z} dz. \end{aligned}$$

Facendo uso della (4.41), si ottiene così:

$$E(Y^k) = \begin{cases} \frac{\Gamma(n/2 - k)}{2^k \Gamma(n/2)}, & k < n/2 \\ +\infty, & k \geq n/2. \end{cases} \quad (7.19)$$

Si noti che se $X \sim \chi^2(n)$, i momenti $E(X^{-k})$ divergono quando $k \geq n/2$, a differenza dei momenti $E(X^k)$ che sono invece finiti per $k = 1, 2, \dots$ \diamond

Come mostrato dalla (7.16), la funzione generatrice dei momenti di una variabile aleatoria $X \sim \chi^2(n)$ è data da

$$M_X(s) = (1 - 2s)^{-n/2} \quad (s < 1/2). \quad (7.20)$$

Nei due esempi seguenti utilizziamo tale funzione per determinare la distribuzione di alcune variabili aleatorie.

Esempio 7.1 Siano X_1, X_2, \dots, X_n variabili aleatorie indipendenti, con $X_i \sim \chi^2(k_i)$ per $i = 1, 2, \dots, n$. Determiniamo la distribuzione di $Y = X_1 + X_2 + \dots + X_n$.

Osserviamo che dal Teorema 5.19 e dalla (7.20) segue:

$$\begin{aligned} M_Y(s) &= M_{X_1}(s) M_{X_2}(s) \cdots M_{X_n}(s) = (1 - 2s)^{-k_1/2} (1 - 2s)^{-k_2/2} \cdots (1 - 2s)^{-k_n/2} \\ &= (1 - 2s)^{-(k_1 + k_2 + \dots + k_n)/2} \quad (s < 1/2). \end{aligned}$$

Si è così ottenuta la funzione generatrice dei momenti di una variabile aleatoria chi-quadrato con $k = k_1 + k_2 + \dots + k_n$ gradi di libertà. Pertanto, $Y \sim \chi^2(k)$. \diamond

Esempio 7.2 Siano X_1 e X_2 variabili aleatorie indipendenti. Supponiamo che $X_1 \sim \chi^2(n_1)$ e che $Y = X_1 + X_2$ sia tale che $Y \sim \chi^2(n)$ con $n > n_1$. Per calcolare la distribuzione di X_2 osserviamo che dal Teorema 5.19 e dalla (7.20) si ricava $M_Y(s) = M_{X_1}(s) M_{X_2}(s)$, così che

$$M_{X_2}(s) = \frac{M_Y(s)}{M_{X_1}(s)} = \frac{(1 - 2s)^{-n/2}}{(1 - 2s)^{-n_1/2}} = (1 - 2s)^{-(n - n_1)/2} \quad (s < 1/2).$$

Questa si riconosce essere la funzione generatrice dei momenti di una variabile aleatoria a distribuzione chi-quadrato con $n - n_1$ gradi di libertà. Pertanto risulta $X_2 \sim \chi^2(n - n_1)$. \diamond

7.4 Distribuzione di Fisher

Vogliamo ora definire un'altra interessante densità di probabilità di notevole interesse in contesti applicativi.

Definizione 7.3 Una variabile aleatoria X di densità di probabilità

$$f_X(x) = \begin{cases} \frac{\Gamma\left(\frac{n_1 + n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right) \Gamma\left(\frac{n_2}{2}\right)} \left(\frac{n_1}{n_2}\right)^{n_1/2} x^{n_1/2 - 1} \left(1 + \frac{n_1}{n_2} x\right)^{-(n_1 + n_2)/2}, & x > 0 \\ 0, & \text{altrimenti} \end{cases} \quad (7.21)$$

con n_1, n_2 interi positivi e con $\Gamma(\nu)$ definita in (4.41), si dice di distribuzione di Fisher (o di distribuzione F di Fisher) con n_1 e n_2 gradi di libertà.

Nel seguito con la scrittura $X \sim \mathcal{F}(n_1, n_2)$ intenderemo che X ha distribuzione di Fisher con n_1 e n_2 gradi di libertà. In Figura 7.3 è rappresentata la densità di probabilità di Fisher (7.21) per alcune scelte dei parametri n_1 e n_2 .

Il seguente teorema mostra la connessione esistente tra una variabile di distribuzione di Fisher e due variabili aleatorie a distribuzioni chi-quadrato.

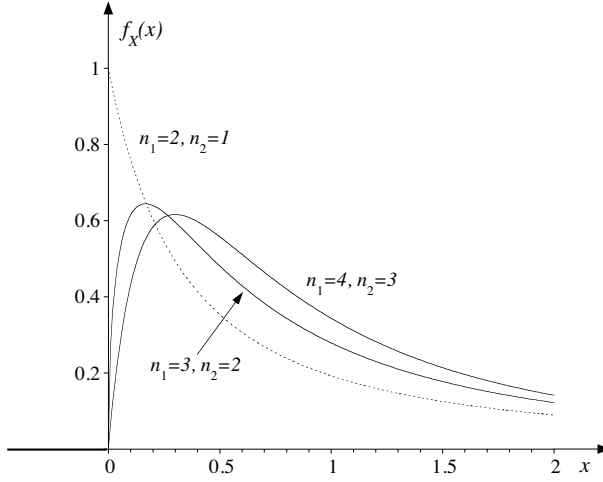


Figura 7.3 – Densità di probabilità di $X \sim \mathcal{F}(n_1, n_2)$.

Teorema 7.3 Se $U \sim \chi^2(n_1)$ e $V \sim \chi^2(n_2)$ sono indipendenti, la variabile aleatoria

$$X = \frac{U/n_1}{V/n_2} \quad (7.22)$$

ha distribuzione di Fisher con n_1 e n_2 gradi di libertà.

Dimostrazione Determiniamo dapprima la densità di probabilità congiunta della coppia

$$(X, Y) = \left(\frac{U/n_1}{V/n_2}, V \right) \quad (7.23)$$

e successivamente la densità di probabilità di X . Essendo U e V indipendenti, si ha:

$$f_{U,V}(u, v) = \begin{cases} \frac{1}{\Gamma\left(\frac{n_1}{2}\right)} \left(\frac{1}{2}\right)^{n_1/2} u^{n_1/2-1} e^{-u/2} \frac{1}{\Gamma\left(\frac{n_2}{2}\right)} \left(\frac{1}{2}\right)^{n_2/2} v^{n_2/2-1} e^{-v/2}, & u > 0, v > 0 \\ 0, & \text{altrimenti.} \end{cases}$$

Per determinare la densità congiunta $f_{X,Y}(x, y)$ facciamo appello al Teorema 3.8. A tal fine consideriamo la trasformazione

$$x = g_1(u, v) = \frac{u/n_1}{v/n_2}, \quad y = g_2(u, v) = v.$$

Conseguentemente, per ogni coppia (u, v) tale che $u > 0$ e $v > 0$ risulta $x > 0$ e $y > 0$. Quindi la trasformazione inversa

$$u = h_1(x, y) = \frac{n_1}{n_2} x y, \quad v = h_2(x, y) = y$$

ha come dominio $D = \{(x, y): x > 0, y > 0\}$. Lo Jacobiano della trasformazione

$$J_h(x, y) = \frac{\partial(h_1, h_2)}{\partial(x, y)} = \begin{vmatrix} \frac{n_1}{n_2} y & \frac{n_1}{n_2} x \\ 0 & 1 \end{vmatrix} = \frac{n_1}{n_2} y$$

è non nullo per ogni $(x, y) \in D$. Essendo soddisfatte le ipotesi del Teorema 3.8, per $x > 0$ e $y > 0$ si ha:

$$f_{X,Y}(x, y) = \frac{2^{-(n_1+n_2)/2}}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \left(\frac{n_1}{n_2} x y\right)^{n_1/2-1} y^{n_2/2-1} \exp\left\{-\frac{y}{2}\left(1 + \frac{n_1}{n_2} x\right)\right\} \frac{n_1}{n_2} y.$$

Quindi,

$$f_{X,Y}(x, y) = \begin{cases} \frac{2^{-(n_1+n_2)/2}}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \left(\frac{n_1 y}{n_2}\right)^{n_1/2} x^{n_1/2-1} y^{n_2/2-1} \exp\left\{-\frac{y}{2}\left(1 + \frac{n_1}{n_2} x\right)\right\}, & x > 0, y > 0 \\ 0, & \text{altrimenti.} \end{cases}$$

La conoscenza della densità del vettore (X, Y) permette di ricavare la densità di probabilità di X . Infatti, per ogni $x > 0$ risulta:

$$f_X(x) = \frac{2^{-(n_1+n_2)/2}}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \left(\frac{n_1}{n_2}\right)^{n_1/2} x^{n_1/2-1} \int_0^{+\infty} y^{(n_1+n_2)/2-1} \exp\left\{-\frac{y}{2}\left(1 + \frac{n_1}{n_2} x\right)\right\} dy.$$

Operando nell'integrale il cambiamento di variabile $z = (1 + n_1 x/n_2) y/2$, per ogni $x > 0$ si ottiene:

$$f_X(x) = \frac{1}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \left(\frac{n_1}{n_2}\right)^{n_1/2} x^{n_1/2-1} \left(1 + \frac{n_1}{n_2} x\right)^{-(n_1+n_2)/2} \int_0^{+\infty} z^{(n_1+n_2)/2-1} e^{-z} dz. \quad (7.24)$$

Ricordando quindi la (4.41), la (7.21) segue. In conclusione, $X \sim \mathcal{F}(n_1, n_2)$. \square

Ci proponiamo ora di effettuare il calcolo dei momenti di $X \sim \mathcal{F}(n_1, n_2)$, certo esistenti essendo X non negativa. Dal Teorema 7.3 segue:

$$E(X^k) = E\left[\left(\frac{U/n_1}{V/n_2}\right)^k\right] = \left(\frac{n_2}{n_1}\right)^k E(U^k) E\left(\frac{1}{V^k}\right),$$

con $U \sim \chi^2(n_1)$ e $V \sim \chi^2(n_2)$ variabili aleatorie indipendenti. Ricordando le (7.17) e (7.19) si ha poi:

$$E(X^k) = \begin{cases} \left(\frac{n_2}{n_1}\right)^k \frac{\Gamma(n_1/2 + k) \Gamma(n_2/2 - k)}{\Gamma(n_1/2) \Gamma(n_2/2)}, & k < n_2/2 \\ +\infty, & k \geq n_2/2. \end{cases} \quad (7.25)$$

In particolare,

$$E(X) = \begin{cases} \frac{n_2}{n_1} \frac{\Gamma(n_1/2 + 1) \Gamma(n_2/2 - 1)}{\Gamma(n_1/2) \Gamma(n_2/2)} = \frac{n_2}{n_2 - 2}, & n_2 > 2 \\ +\infty, & n_2 \leq 2, \end{cases} \quad (7.26)$$

che mostra che $E(X)$ dipende soltanto dal numero di gradi di libertà della variabile V che compare al denominatore della (7.22). Inoltre, poiché

$$E(X^2) = \begin{cases} \left(\frac{n_2}{n_1}\right)^2 \frac{\Gamma(n_1/2 + 2) \Gamma(n_2/2 - 2)}{\Gamma(n_1/2) \Gamma(n_2/2)} = \frac{n_2^2 (n_1 + 2)}{n_1 (n_2 - 2) (n_2 - 4)}, & n_2 > 4 \\ +\infty, & n_2 \leq 4, \end{cases}$$

per $n_2 > 4$ la varianza di $X \sim \mathcal{F}(n_1, n_2)$ è:

$$\text{Var}(X) = \frac{2 n_2^2 (n_1 + n_2 - 2)}{n_1 (n_2 - 2)^2 (n_2 - 4)}.$$

Osservazione 7.2 Sia $X \sim \mathcal{F}(n_1, n_2)$. Per determinare la distribuzione di $Y = 1/X$, osserviamo che risulta $f_Y(y) = 0$ per $y \leq 0$; utilizzando il Teorema 3.4, per $y > 0$ si ha invece:

$$f_Y(y) = f_X\left(\frac{1}{y}\right) \frac{1}{y^2} = \frac{\Gamma\left(\frac{n_1 + n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right) \Gamma\left(\frac{n_2}{2}\right)} \left(\frac{n_2}{n_1}\right)^{n_2/2} y^{n_2/2-1} \left(1 + \frac{n_2}{n_1} y\right)^{-(n_1+n_2)/2},$$

che mostra che $Y \sim \mathcal{F}(n_2, n_1)$. Possiamo quindi affermare che se X ha distribuzione di Fisher con n_1 e n_2 gradi di libertà, allora $1/X$ ha distribuzione di Fisher con n_2 e n_1 gradi di libertà. \diamond

L'osservazione seguente esplicita la connessione esistente tra una variabile a distribuzione di Fisher e una variabile a distribuzione beta.

Osservazione 7.3 Sia $X \sim \mathcal{F}(n_1, n_2)$. Per determinare la distribuzione di

$$Y = \frac{n_1 X/n_2}{1 + n_1 X/n_2}$$

osserviamo che risulta $f_Y(y) = 0$ per $y \notin (0, 1)$; in virtù del Teorema 3.4, per $0 < y < 1$ si ha:

$$f_Y(y) = f_X\left(\frac{n_2 y}{n_1(1-y)}\right) \frac{n_2}{n_1} \frac{1}{(1-y)^2} = \frac{\Gamma\left(\frac{n_1 + n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right) \Gamma\left(\frac{n_2}{2}\right)} y^{n_1/2-1} (1-y)^{n_2/2-1},$$

indicante che $Y \sim \mathcal{Be}(n_1, n_2)$. \diamond

7.5 Distribuzione di Student

Un'altra distribuzione di considerevole interesse applicativo è quella di Student¹ che passiamo a definire.

Definizione 7.4 Una variabile aleatoria X di densità di probabilità

$$f_X(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}, \quad x \in \mathbb{R} \quad (7.27)$$

con n intero positivo e con $\Gamma(\nu)$ definita in (4.41), si dice avere distribuzione di Student, o avere “distribuzione t di Student”, con n gradi di libertà.

Nel seguito con $X \sim \mathcal{T}(n)$ intenderemo che X ha distribuzione di Student con n gradi di libertà. Il teorema seguente mostra la stretta connessione esistente tra una variabile a distribuzione di Student (7.27) e variabili a distribuzione chi-quadrato e normale standard.

Teorema 7.4 Siano $U \sim \chi^2(n)$ e $Z \sim \mathcal{N}(0, 1)$ variabili aleatorie indipendenti. Allora

$$X = \frac{Z}{\sqrt{U/n}} \quad (7.28)$$

ha distribuzione di Student con n gradi di libertà.

Dimostrazione Procederemo determinando inizialmente la densità di probabilità congiunta di

$$(X, Y) = \left(\frac{Z}{\sqrt{U/n}}, U \right); \quad (7.29)$$

successivamente otterremo la densità di probabilità di X . Essendo U e Z indipendenti, si ha:

$$f_{U,Z}(u, z) = \begin{cases} \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{n/2} u^{n/2-1} e^{-u/2} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, & u > 0, z \in \mathbb{R} \\ 0, & u \leq 0, z \in \mathbb{R}. \end{cases}$$

Per determinare la densità congiunta $f_{X,Y}(x, y)$ utilizziamo il Teorema 3.8. A tal fine consideriamo la trasformazione

$$x = g_1(u, z) = \frac{z}{\sqrt{u/n}}, \quad y = g_2(u, z) = u.$$

Per ogni coppia (u, z) tale che $u > 0$ e $z \in \mathbb{R}$, risulta $x \in \mathbb{R}$ e $y > 0$. Quindi la trasformazione inversa

$$z = h_1(x, y) = x \sqrt{\frac{y}{n}}, \quad u = h_2(x, y) = y$$

¹ Student è lo pseudonimo con cui il matematico inglese W.S. Gosset pubblicava i suoi articoli.

ha dominio $D = \{(x, y): x \in \mathbb{R}, y > 0\}$. Lo Jacobiano

$$J_h(x, y) = \frac{\partial(h_1, h_2)}{\partial(x, y)} = \begin{vmatrix} \sqrt{\frac{y}{n}} & \frac{x}{2\sqrt{ny}} \\ 0 & 1 \end{vmatrix} = \sqrt{\frac{y}{n}}.$$

della trasformazione è non nullo per ogni $(x, y) \in D$. Essendo soddisfatte le ipotesi del Teorema 3.8, per $x \in \mathbb{R}$ e $y > 0$ si ha:

$$f_{X,Y}(x, y) = \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{n/2} y^{n/2-1} e^{-y/2} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2 y}{2n}\right\} \sqrt{\frac{y}{n}}.$$

Quindi,

$$f_{X,Y}(x, y) = \begin{cases} \frac{2^{-n/2-1/2}}{\sqrt{\pi n} \Gamma\left(\frac{n}{2}\right)} y^{n/2-1/2} \exp\left\{-\frac{y}{2}\left(1 + \frac{x^2}{n}\right)\right\}, & x \in \mathbb{R}, y > 0 \\ 0, & \text{altrimenti.} \end{cases}$$

La conoscenza della densità di (X, Y) permette di ricavare la densità di probabilità di X . Infatti, per ogni $x \in \mathbb{R}$ si ha:

$$f_X(x) = \int_0^{+\infty} f_{X,Y}(x, y) dy = \frac{2^{-n/2-1/2}}{\sqrt{\pi n} \Gamma\left(\frac{n}{2}\right)} \int_0^{+\infty} y^{n/2-1/2} \exp\left\{-\frac{y}{2}\left(1 + \frac{x^2}{n}\right)\right\} dy.$$

Operando il cambiamento di variabile $z = y(1 + x^2/n)/2$ segue poi:

$$f_X(x) = \frac{1}{\sqrt{\pi n} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2} \int_0^{+\infty} z^{(n+1)/2-1} e^{-z} dz$$

da cui, ricordando la (4.41), si ottiene la (7.27). In conclusione, $X \sim T(n)$. \square

Notiamo che la densità (7.27) è unimodale, simmetrica intorno all'asse $x = 0$ e dipendente dal solo parametro rappresentante il numero di gradi di libertà. Si noti che per $n = 1$ la (7.27) si identifica con la densità di Cauchy (5.17). Invece, al limite per $n \rightarrow +\infty$ essa converge alla densità normale standard. Infatti, facendo uso dell'approssimazione di Stirling

$$\Gamma(z+1) \sim \left(\frac{z}{e}\right)^z \sqrt{2\pi z} \quad (z \rightarrow +\infty)$$

e osservando che

$$\lim_{n \rightarrow +\infty} \left(1 + \frac{x^2}{n}\right)^{(n+1)/2} = \lim_{n \rightarrow +\infty} \left(1 + \frac{x^2}{n}\right)^{1/2} \left[\left(1 + \frac{x^2}{n}\right)^n\right]^{1/2} = \exp\left\{\frac{x^2}{2}\right\},$$

si ha

$$\lim_{n \rightarrow +\infty} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2} = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}, \quad x \in \mathbb{R}.$$

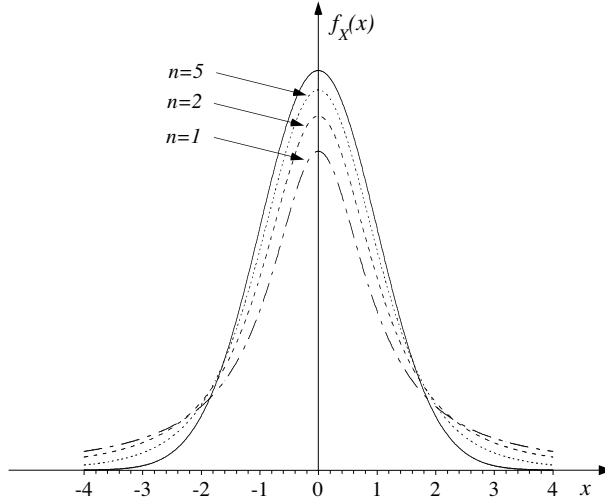


Figura 7.4– Densità di probabilità di $X \sim \mathcal{T}(n)$ per $n = 1, 2, 5$ e densità normale standard (curva con tratto continuo)

In Figura 7.4 è rappresentata la densità di probabilità di Student (7.27) per $n = 1, 2, 5$ e la densità normale standard (curva con tratto continuo).

La variabile $X \sim \mathcal{T}(n)$ non possiede funzione generatrice dei momenti. Per quanto attiene ai momenti $E(X^k)$ non sempre esistono e, quando esistono, non sempre sono finiti. Dalla Definizione 5.5 si ha infatti:

$$E(X^k) = \int_{-\infty}^{+\infty} x^k f_X(x) dx = \int_{-\infty}^0 x^k f_X(x) dx + \int_0^{+\infty} x^k f_X(x) dx. \quad (7.30)$$

Ricordando la (7.27) segue poi:

$$\begin{aligned} \int_0^{+\infty} x^k f_X(x) dx &= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right)} \int_0^{+\infty} x^k \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2} dx \\ &= \frac{n^{k/2} \Gamma\left(\frac{n+1}{2}\right)}{2\sqrt{\pi} \Gamma\left(\frac{n}{2}\right)} \int_0^1 z^{(n-k)/2-1} (1-z)^{(k+1)/2-1} dz, \end{aligned}$$

dove l'ultima uguaglianza è frutto del cambiamento di variabile di integrazione $z = (1 + x^2/n)^{-1}$. Pertanto, ricordando la (7.6) e la (7.13), si ottiene:

$$\int_0^{+\infty} x^k f_X(x) dx = \begin{cases} \frac{n^{k/2} B\left(\frac{k+1}{2}, \frac{n-k}{2}\right)}{2 B\left(\frac{1}{2}, \frac{n}{2}\right)}, & k < n \\ +\infty, & k \geq n. \end{cases} \quad (7.31)$$

Inoltre, poiché $f_X(x)$ è una funzione pari, risulta:

$$\int_{-\infty}^0 x^k f_X(x) dx = \int_0^{+\infty} (-z)^k f_X(z) dz = (-1)^k \int_0^{+\infty} z^k f_X(z) dz. \quad (7.32)$$

Le (7.30), (7.31) e (7.32) permettono quindi di affermare che i momenti di ordine dispari sono nulli se $k < n$ e non esistono se $k \geq n$; per i momenti di ordine pari, che esistono sempre, si ha invece:

$$E(X^k) = \begin{cases} \frac{n^{k/2} B\left(\frac{k+1}{2}, \frac{n-k}{2}\right)}{B\left(\frac{1}{2}, \frac{n}{2}\right)}, & k < n, k \text{ pari} \\ +\infty, & k \geq n, k \text{ pari}. \end{cases}$$

In particolare, il valore medio di $X \sim \mathcal{T}(n)$ non esiste se $n = 1$ (ossia quando la variabile aleatoria è di Cauchy) e risulta nullo se $n = 2, 3, \dots$. Inoltre la varianza non esiste se $n = 1$ e diverge se $n = 2$; per $n = 3, 4, \dots$ si ha poi:

$$\text{Var}(X) = E(X^2) = \frac{n B\left(\frac{3}{2}, \frac{n-2}{2}\right)}{B\left(\frac{1}{2}, \frac{n}{2}\right)} = n \frac{\Gamma\left(\frac{3}{2}\right) \Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{n}{2}\right)} = \frac{n}{n-2}.$$

Osservazione 7.4 Sia $X \sim \mathcal{T}(n)$. Per la distribuzione di $Y = X^2$, dalla (7.28) otteniamo:

$$Y = X^2 = \frac{Z^2}{U/n},$$

con U e Z variabili aleatorie indipendenti, con $U \sim \chi^2(n)$ e $Z \sim \mathcal{N}(0, 1)$. Dal Teorema 7.2 segue poi $Z^2 \sim \chi(1)$. Quindi Y si presenta come il rapporto di due variabili di distribuzione chi-quadrato entrambe divise per i rispettivi gradi di libertà. Pertanto, facendo uso del Teorema 7.3, si conclude che risulta $Y \sim \mathcal{F}(1, n)$. \diamond

Capitolo 8

Disuguaglianze notevoli

8.1 Introduzione

In questo capitolo ricaveremo talune utili disuguaglianze coinvolgenti probabilità e momenti. Alcune di esse consentono di ottenere limitazioni sulla probabilità delle “code” di una distribuzione e mostrano che tali limitazioni spesso dipendono esclusivamente dai soli primi momenti, ad esempio da media e varianza. Un’analisi di tali limitazioni può rivelarsi utile per individuare intervalli nei quali una variabile aleatoria assume valori con probabilità prossime all’unità, e può così condurre a semplificazioni di calcoli. Altre disuguaglianze si riferiscono invece a momenti di trasformazioni convesse o concave di variabili aleatorie, oppure a momenti di variabili aleatorie stocasticamente ordinate in distribuzione. Occorre infine sottolineare che, come si vedrà nel Capitolo 9, vi sono disuguaglianze che svolgono un ruolo centrale nello studio di leggi limite della probabilità.

8.2 Alcune semplici disuguaglianze

Una delle più semplici e importanti disuguaglianze è la seguente.

Teorema 8.1 (Disuguaglianza di Markov) *Sia X una variabile aleatoria non negativa. Per ogni $a > 0$ risulta:*

$$P(X \geq a) \leq \frac{E(X)}{a}. \quad (8.1)$$

Dimostrazione Consideriamo l’evento $A = \{\omega \in \Omega : X(\omega) \geq a\}$ e la corrispondente funzione indicatrice

$$I_A(\omega) = \begin{cases} 0, & \omega \in \bar{A} \\ 1, & \omega \in A. \end{cases}$$

Quest’ultima è una variabile aleatoria di Bernoulli che assume il valore 0 con probabilità $P(\bar{A}) = P(X < a)$ e il valore 1 con probabilità $P(A) = P(X \geq a)$. Quindi $E(I_A) = P(X \geq a)$. Essendo $X \geq a I_A$, sussiste la seguente disuguaglianza:

$$X \geq X I_A \geq a I_A. \quad (8.2)$$

Infatti, se $\omega \in \bar{A}$ si ha $I_A(\omega) = 0$ e la (8.2) è soddisfatta essendo $X(\omega) \geq 0 = X(\omega) I_A(\omega) = a I_A(\omega)$. Se invece $\omega \in A$ si ha $I_A(\omega) = 1$, così che la (8.2) è nuovamente verificata essendo $X(\omega) = X(\omega) I_A(\omega) \geq a I_A(\omega)$. Dalla (8.2) si ottiene poi $X - a I_A \geq 0$, da cui segue $E(X - a I_A) \geq 0$. Per la proprietà di linearità del valore medio si ha infine $E(X) - a E(I_A) \geq 0$, ossia

$$E(X) \geq a E(I_A) = a P(X \geq a),$$

che conduce alla (8.1). Si noti che se X è assolutamente continua e non negativa la (8.1) può essere dimostrata più semplicemente nel seguente modo:

$$E(X) = \int_0^{+\infty} x f_X(x) dx \geq \int_a^{+\infty} x f_X(x) dx \geq a \int_a^{+\infty} f_X(x) dx = a P(X \geq a).$$

□

Il Teorema 8.1 sussiste per qualsiasi variabile aleatoria non negativa, quale che sia la sua funzione di distribuzione. Proprio a tale grande generalità è dovuta la sua importanza. Si noti che grazie alla (8.1) la sola conoscenza del valore medio è sufficiente per avere una maggiorazione della funzione $P(X \geq a)$.

Va menzionato che la rilevanza della disuguaglianza di Markov risiede anche nel ruolo che essa gioca nel consentire la deduzione di importanti altre disuguaglianze, quali le disuguaglianze di Chebyshev e di Chernoff. Modesta appare invece in generale la possibilità di farne uso per ottenere significative approssimazioni miranti alla determinazione di utili maggioranti alle probabilità, come esemplificato negli Esempi 8.1 e 8.2. Si noti infine che nel caso $a \leq E(X)$, la (8.1) non è significativa.

Tabella 8.1— Per tre scelte dei parametri n e p , con $np = 1$, sono elencate le probabilità (8.3) e i maggioranti forniti dalla (8.4) in corrispondenza dei primi cinque interi x .

x	$P(X \geq x)$ $p = 0.05$ $n = 20$	$P(X \geq x)$ $p = 0.1$ $n = 10$	$P(X \geq x)$ $p = 0.2$ $n = 5$	Markov $(np)/x$
1	0.6415	0.6513	0.6723	1.0000
2	0.2642	0.2639	0.2627	0.5000
3	0.0755	0.0702	0.0579	0.3333
4	0.0159	0.0128	0.0067	0.2500
5	0.0026	0.0016	0.0003	0.2000

Esempio 8.1 Sia $X \sim \mathcal{B}(n, p)$. Si ha

$$P(X \geq x) = \sum_{k=\lceil x \rceil}^n \binom{n}{k} p^k (1-p)^{n-k} \quad (x \geq 0), \quad (8.3)$$

dove $\lceil x \rceil$ denota il minimo intero maggiore o uguale ad x . Essendo $E(X) = np$, dalla disuguaglianza di Markov per ogni x positivo si ricava:

$$P(X \geq x) \leq \frac{np}{x}. \quad (8.4)$$

Nella Tabella 8.1 confrontiamo le probabilità binomiali (8.3) con la limitazione $(np)/x$ fornita dalla disuguaglianza (8.4) per alcune scelte di n e p tali che $np = 1$. \diamond

Esempio 8.2 Si consideri l'esperimento consistente nel lanciare ripetutamente per n volte una moneta. Si è interessati alla probabilità che il numero di volte in cui si verifica testa sia non inferiore a $3n/4$. Denotando con X la variabile aleatoria che rappresenta il numero di volte in cui si realizza testa negli n lanci, si ha $X \sim \mathcal{B}(n, 1/2)$. Ponendo $p = 1/2$ e $x = 3n/4$ nella (8.3) si ottiene:

$$P\left(X \geq \frac{3n}{4}\right) = \left(\frac{1}{2}\right)^n \sum_{k=\lceil 3n/4 \rceil}^n \binom{n}{k}, \quad (8.5)$$

mentre, facendo uso della disuguaglianza (8.4), risulta:

$$P\left(X \geq \frac{3n}{4}\right) \leq \frac{2}{3}. \quad (8.6)$$

Tabella 8.2 – Le probabilità (8.5) per alcune scelte di n .

n	$P(X \geq 3n/4)$	n	$P(X \geq 3n/4)$
5	0.18750	30	0.00261
10	0.05469	35	0.00094
15	0.01758	40	0.00111
20	0.02069	45	0.00041
25	0.00732	50	0.00015

Si noti che la probabilità al primo membro della (8.5) tende a zero al crescere del numero n di lanci, il che si riflette nella Tabella 8.2, mentre la disuguaglianza (8.6) fornisce una limitazione a tale probabilità piuttosto debole che non dipende dal numero di lanci. \diamond

Una generalizzazione della disuguaglianza di Markov (8.1) è fornita nel seguente teorema:

Teorema 8.2 Sia X una variabile aleatoria e siano a e ν reali positivi. Risulta:

$$P(|X| \geq a) \leq \frac{E(|X|^\nu)}{a^\nu}. \quad (8.7)$$

Dimostrazione Essendo $|X|$ una variabile aleatoria non negativa, tale è anche $|X|^\nu$, così che $E(|X|^\nu)$ esiste (finito o infinito). Pertanto, per $a > 0$ e $\nu > 0$, in virtù della disuguaglianza di Markov si ottiene:

$$P(|X| \geq a) = P(|X|^\nu \geq a^\nu) \leq \frac{E(|X|^\nu)}{a^\nu}.$$

□

Poiché $P(|X| < a) = 1 - P(|X| \geq a)$, dalla (8.7) segue immediatamente:

$$P(|X| < a) \geq 1 - \frac{E[|X|^\nu]}{a^\nu} \quad (a > 0, \nu > 0). \quad (8.8)$$

Si noti che la conoscenza anche di un solo momento di $|X|$ consente di ottenere una maggiorazione per $P(|X| \geq a)$ attraverso la (8.7), nonché una minorazione per $P(|X| < a)$ mediante la (8.8), sempre che non risulti $a^\nu \leq E(|X|^\nu)$, nel qual caso né la (8.7) né la (8.8) sono significative.

Abbiamo visto nel Paragrafo 5.4 che la varianza può essere riguardata come una misura della dispersione di una variabile aleatoria intorno al suo valore medio. Questa interpretazione è rafforzata dalla disuguaglianza di cui al teorema che segue.

Teorema 8.3 (Disuguaglianza di Chebyshev) *Sia X una variabile aleatoria con $E(X^2)$ finito. Per ogni reale positivo ε si ha:*

$$P\{|X - E(X)| \geq \varepsilon\} \leq \frac{\text{Var}(X)}{\varepsilon^2}. \quad (8.9)$$

Dimostrazione Osserviamo anzitutto che essendo per ipotesi $E(X^2)$ finito, tali sono anche $E(X)$ e $\text{Var}(X)$. Ponendo nella (8.7) $|X - E(X)|$ al posto di X , $\nu = 2$ e $a = \varepsilon$, segue la (8.9) avendosi:

$$P\{|X - E(X)| \geq \varepsilon\} \leq \frac{E\{[X - E(X)]^2\}}{\varepsilon^2} = \frac{\text{Var}(X)}{\varepsilon^2}.$$

□

Essendo $P\{|X - E(X)| < \varepsilon\} = 1 - P\{|X - E(X)| \geq \varepsilon\}$, dalla (8.9) si ha poi:

$$P\{|X - E(X)| < \varepsilon\} \geq 1 - \frac{\text{Var}(X)}{\varepsilon^2}. \quad (8.10)$$

Osserviamo che mentre per calcolare la probabilità che una variabile aleatoria X assuma valori in un fissato insieme occorre in generale conoscerne la funzione di distribuzione, la disuguaglianza di Chebyshev mostra che la probabilità con cui una variabile aleatoria di valore medio μ e varianza σ^2 assuma valori nell'insieme $(-\infty, \mu - \varepsilon] \cup [\mu + \varepsilon, +\infty)$ è minore o uguale a σ^2/ε^2 , così che la probabilità che detta variabile assuma valori nell'intervallo $(\mu - \varepsilon, \mu + \varepsilon)$ è non minore di $1 - \sigma^2/\varepsilon^2$ pur non conoscendone la sua funzione di distribuzione.

Nel Capitolo 9 si vedrà come la disuguaglianza di Chebyshev si riveli particolarmente utile anche in svariati problemi di convergenza di successioni di variabili aleatorie.

Esempio 8.3 Siano $X \sim \mathcal{B}(n, p)$ e $Y = X/n$. Se si interpreta X come il numero di successi in n prove di Bernoulli indipendenti, allora Y rappresenta la frequenza relativa dei successi. Ricordando che $E(Y) = p$ e $\text{Var}(Y) = p(1-p)/n$, dalla disuguaglianza di Chebyshev (8.9) si ottiene:

$$P(|Y - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2}. \quad (8.11)$$

Poiché è sempre $p(1-p) \leq 1/4$, dalla disuguaglianza (8.11) si ricava:

$$P(|Y - p| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2}. \quad (8.12)$$

La probabilità che Y differisca da p per almeno ε è dunque non minore di $(4n\varepsilon^2)^{-1}$ qualsiasi sia p .

Tabella 8.3– Numero minimo di prove di Bernoulli indipendenti da effettuare affinché qualsiasi sia p risulti $P(|Y - p| \geq \varepsilon) \leq c$ per varie scelte di ε e c .

ε	$c = 0.1$	$c = 0.05$	$c = 0.025$	$c = 0.01$	$c = 0.005$
0.1	250	500	1000	2500	5000
0.2	63	125	250	625	1250
0.3	28	56	112	278	556
0.4	16	32	63	157	313
0.5	10	20	40	100	200

Nella Tabella 8.3 è indicato il numero minimo n_0 di prove di Bernoulli indipendenti da effettuare affinché qualsiasi sia p risulti $P(|Y - p| \geq \varepsilon) \leq c$ per varie scelte di ε e c . Tale numero è ottenuto dalla (8.12) mediante la relazione $n_0 = \min\{n : (4n\varepsilon^2)^{-1} \leq c\}$, ossia $n_0 = \lceil (4c\varepsilon^2)^{-1} \rceil$. Ad esempio, come si evince dalla Tabella 8.3, per $\varepsilon = 0.1$ e $c = 0.025$, risulta $n_0 = 1000$. \diamond

Esempio 8.4 Riprendiamo in considerazione l'Esempio 8.2 e sia $X \sim \mathcal{B}(n, 1/2)$. In tal caso $E(X) = n/2$ e $\text{Var}(X) = n/4$ e quindi si ha:

$$P\left(X \geq \frac{3n}{4}\right) = P\left(X - \frac{n}{2} \geq \frac{n}{4}\right) = P\left(X - E(X) \geq \frac{E(X)}{2}\right) \leq P\left(|X - E(X)| \geq \frac{E(X)}{2}\right),$$

da cui, applicando la disuguaglianza di Chebyshev con $\varepsilon = E(X)/2$, si ottiene:

$$P\left(X \geq \frac{3n}{4}\right) \leq \frac{4\text{Var}(X)}{[E(X)]^2} = \frac{4}{n}. \quad (8.13)$$

Per $n > 6$, la (8.13) fornisce una limitazione migliore di quella fornita dalla (8.6). In particolare, la (8.13) mostra esplicitamente che la probabilità (8.5) tende a zero al crescere del numero di lanci. \diamond

Teorema 8.4 Sia X una variabile aleatoria con $E(X^2)$ finito. Per ogni reale positivo r si ha:

$$P\left\{|X - E(X)| < r\sqrt{\text{Var}(X)}\right\} \geq 1 - \frac{1}{r^2}. \quad (8.14)$$

Dimostrazione Segue immediatamente dalla (8.10) ponendovi $\varepsilon = r\sqrt{\text{Var}(X)}$. \square

Questo teorema fornisce limitazioni in soli termini di valore medio e di deviazione standard. Esso è quindi particolarmente utile proprio quando la funzione di distribuzione non è nota. Ovviamente, quando tale funzione è conosciuta, mediante essa si ottengono limitazioni più accurate, come evidenziato nell'esempio seguente.

Esempio 8.5 Ci proponiamo di confrontare $P(|X - \mu| < r\sigma)$ con la limitazione superiore fornita dalla (8.14) per le variabili aleatorie uniformi, normale e di Laplace caratterizzate da stesso valore medio μ finito e stessa deviazione standard σ finita e positiva.

(a) Nel caso $X \sim \mathcal{U}(a, b)$, scegliamo $a = \mu - \sigma\sqrt{3}$ e $b = \mu + \sigma\sqrt{3}$, così che $E(X) = \mu$ e $\text{Var}(X) = \sigma^2$. Si ha:

$$P(|X - \mu| < r\sigma) = \begin{cases} \int_{\mu-r\sigma}^{\mu+r\sigma} \frac{1}{2\sigma\sqrt{3}} dx = \frac{r}{\sqrt{3}}, & 0 < r < \sqrt{3} \\ 1, & r \geq \sqrt{3}. \end{cases} \quad (8.15)$$

(b) Se $X \sim \mathcal{N}(\mu, \sigma)$, in virtù della (4.57) segue:

$$P(|X - \mu| < r\sigma) = P\left(-r < \frac{X - \mu}{\sigma} < r\right) = 2\Phi(r) - 1, \quad (8.16)$$

dove $\Phi(z)$ è la funzione di distribuzione della variabile aleatoria normale standard (v. Paragrafo 4.4.6).

(c) Sia X una variabile aleatoria di distribuzione di Laplace di parametri α e β , con $\alpha = \mu$ e $\beta = \sigma/\sqrt{2}$. Con tali scelte risulta $E(X) = \mu$ e $\text{Var}(X) = \sigma^2$ ed inoltre si ha:

$$\begin{aligned} P(|X - \mu| < r\sigma) &= \int_{\mu-r\sigma}^{\mu+r\sigma} \frac{1}{\sigma\sqrt{2}} \exp\left\{-\frac{\sqrt{2}|x - \mu|}{\sigma}\right\} dx \\ &= \frac{1}{\sigma\sqrt{2}} \int_{\mu-r\sigma}^{\mu} \exp\left\{-\frac{\sqrt{2}(\mu - x)}{\sigma}\right\} dx + \frac{1}{\sigma\sqrt{2}} \int_{\mu}^{\mu+r\sigma} \exp\left\{-\frac{\sqrt{2}(x - \mu)}{\sigma}\right\} dx \\ &= \frac{2}{\sqrt{2}} \int_0^r e^{-\sqrt{2}z} dz = 1 - e^{-\sqrt{2}r}. \end{aligned} \quad (8.17)$$

Nella Tabella 8.4 vengono confrontate le probabilità (8.15), (8.16) e (8.17) con la limitazione $1 - 1/r^2$ fornita dalla disuguaglianza (8.14) per alcune scelte di r .

Si noti che nel caso della distribuzione normale la disuguaglianza (8.14) fornisce un risultato più debole di quello espresso dalla legge del 3σ , illustrata nel Paragrafo 4.4.6. Infatti, come mostra la Tabella 8.4, con precisione sulla quarta cifra decimale, si ha $P(|X - \mu| < 3\sigma) \geq 0.8889$, mentre dalla legge del 3σ si ottiene $P(|X - \mu| < 3\sigma) \simeq 0.9974$. \diamond

Esempio 8.6 Ci proponiamo di confrontare $P\{|X - E(X)| < r\sqrt{\text{Var}(X)}\}$ con la limitazione superiore della (8.14) per le variabili aleatorie esponenziale e di Erlang di ordine 2, caratterizzate dallo stesso valore medio finito $1/\lambda$ ($\lambda > 0$).

(a) Sia $X \sim \mathcal{E}(1, \lambda)$, così che $E(X) = 1/\lambda$ e $\text{Var}(X) = 1/\lambda^2$. Si ha:

$$P\{|X - E(X)| < r\sqrt{\text{Var}(X)}\} = P\left(\frac{1-r}{\lambda} < X < \frac{1+r}{\lambda}\right)$$

Tabella 8.4– Confronto tra le probabilità (8.15), (8.16) e (8.17) e i maggioranti forniti dalla disuguaglianza (8.14) per alcune scelte di r .

r	Uniforme $P(X - \mu < r\sigma)$	Normale $P(X - \mu < r\sigma)$	Laplace $P(X - \mu < r\sigma)$	$1 - \frac{1}{r^2}$
1	0.5774	0.6826	0.7569	0.0000
1.5	0.8660	0.8664	0.8801	0.5556
2	1.0000	0.9544	0.9409	0.7500
2.5	1.0000	0.9876	0.9709	0.8400
3	1.0000	0.9974	0.9856	0.8889
3.5	1.0000	0.9995	0.9929	0.9184
4	1.0000	1.0000	0.9965	0.9375

$$= \begin{cases} \int_{(1-r)/\lambda}^{(1+r)/\lambda} \lambda e^{-\lambda x} dx = e^{-(1-r)} - e^{-(1+r)}, & 0 \leq r \leq 1 \\ \int_0^{(1+r)/\lambda} \lambda e^{-\lambda x} dx = 1 - e^{-(1+r)}, & r > 1. \end{cases} \quad (8.18)$$

(b) Sia $X \sim \mathcal{E}(2, 2\lambda)$, di modo che $E(X) = 1/\lambda$ e $\text{Var}(X) = 1/(2\lambda^2)$. In tal caso risulta:

$$P\left\{|X - E(X)| < r \sqrt{\text{Var}(X)}\right\} = P\left\{\frac{1}{\lambda}\left(1 - \frac{r}{\sqrt{2}}\right) < X < \frac{1}{\lambda}\left(1 + \frac{r}{\sqrt{2}}\right)\right\}$$

$$= \begin{cases} \int_{(1-r/\sqrt{2})/\lambda}^{(1+r/\sqrt{2})/\lambda} 4\lambda^2 x e^{-2\lambda x} dx = (3 - r\sqrt{2})e^{-2+r\sqrt{2}} - (3 + r\sqrt{2})e^{-2-r\sqrt{2}}, & 0 \leq r \leq \sqrt{2} \\ \int_0^{(1+r/\sqrt{2})/\lambda} 4\lambda^2 x e^{-2\lambda x} dx = 1 - (3 + r\sqrt{2})e^{-2-r\sqrt{2}}, & r > \sqrt{2}. \end{cases} \quad (8.19)$$

Nella Tabella 8.5 vengono riportate le probabilità (8.18) e (8.19) con la limitazione $1 - 1/r^2$ fornita dalla disuguaglianza (8.14) per alcune scelte di r . \diamond

Teorema 8.5 Sia X una variabile aleatoria con $E(X) \neq 0$ e $E(X^2)$ finito. Per ogni reale positivo δ si ha:

$$P\left\{\left|\frac{X - E(X)}{E(X)}\right| < \delta\right\} \geq 1 - \frac{(C_X)^2}{\delta^2}, \quad (8.20)$$

dove $C_X = \sqrt{\text{Var}(X)}/E(X)$ denota il coefficiente di variazione di X .

Dimostrazione Segue immediatamente dalla (8.10) ponendo $\varepsilon = \delta |E(X)|$. \square

Tabella 8.5– Confronto tra le probabilità (8.18), (8.19) e i maggioranti forniti dalla disuguaglianza (8.14) per alcune scelte di r .

r	Esponenziale $P(X - \mu < r\sigma)$	Erlang $P(X - \mu < r\sigma)$	$1 - \frac{1}{r^2}$
1	0.8647	0.7375	0.0000
1.5	0.9179	0.9169	0.5556
2	0.9502	0.9534	0.7500
2.5	0.9698	0.9742	0.8400
3	0.9817	0.9859	0.8889
3.5	0.9889	0.9924	0.9184
4	0.9933	0.9959	0.9375
4.5	0.9959	0.9978	0.9506
5	0.9975	0.9988	0.9600

Il Teorema 8.5 mostra che la probabilità con cui una variabile aleatoria X di valore medio $\mu \neq 0$ e varianza σ^2 assume valori nell'intervallo $((1 - \delta)\mu, (1 + \delta)\mu)$ è non minore di $1 - (C_X/\delta)^2$. Inoltre, denotando con $Y = [X - E(X)]/E(X)$ l'errore relativo di X rispetto a $E(X)$, la (8.20) mostra anche che la probabilità con cui Y assume valori nell'intervallo $(-\delta, \delta)$ è vicina all'unità se $(C_X)^2$ è prossimo a zero. L'utilizzazione della disuguaglianza (8.20) necessita della conoscenza di valore medio e coefficiente di variazione. Al fine di consentire l'applicazione immediata della (8.20), nella Tabella 8.6 sono riportati valore medio e quadrato del coefficiente di variazione per alcune variabili aleatorie di uso frequente.

Come si è visto, la disuguaglianza di Chebyshev permette di ricavare delle limitazioni alle probabilità con cui una variabile aleatoria di valore medio μ e varianza σ^2 assume valori nell'insieme $(-\infty, \mu - \varepsilon] \cup [\mu + \varepsilon, +\infty)$. Talora, tuttavia, è desiderabile ottenere delle limitazioni alla probabilità con cui una variabile aleatoria assume valori in intervalli quali $[\mu + \varepsilon, +\infty)$, $(-\infty, \mu - \varepsilon]$. A tal proposito sussiste il seguente

Teorema 8.6 (Disuguaglianze di Chebyshev unilaterali) *Sia X una variabile aleatoria con $E(X^2)$ finito. Per ogni reale positivo ε si ha:*

$$P\{X \geq E(X) + \varepsilon\} \leq \frac{\text{Var}(X)}{\text{Var}(X) + \varepsilon^2}, \quad P\{X \leq E(X) - \varepsilon\} \leq \frac{\text{Var}(X)}{\text{Var}(X) + \varepsilon^2}. \quad (8.21)$$

Dimostrazione Essendo per ipotesi $E(X^2)$ finito, tali risultano anche $E(X)$ e $\text{Var}(X)$. Per ogni reale positivo b si ha poi:

$$\begin{aligned} P\{X \geq E(X) + \varepsilon\} &= P\{X - E(X) + b \geq \varepsilon + b\} \leq P\{[X - E(X) + b]^2 \geq (\varepsilon + b)^2\} \\ &\leq \frac{E\{[X - E(X) + b]^2\}}{(\varepsilon + b)^2} = \frac{\text{Var}(X) + b^2}{(\varepsilon + b)^2}, \end{aligned} \quad (8.22)$$

dove l'ultima maggiorazione segue dalla disuguaglianza di Markov. Il reale b_0 che minimizza il termine a destra della (8.22) è $\text{Var}(X)/\varepsilon$. Pertanto, con tale scelta, la (8.22) conduce alla

Tabella 8.6 – Per alcune variabili aleatorie X sono elencati $E(X)$ e C_X^2 .

Distribuzione	X	$E(X)$	C_X^2
Binomiale	$X \sim \mathcal{B}(n, p)$	np	$\frac{1-p}{np}$
Binomiale Negativa	$X \sim \mathcal{BN}(n, p)$	$\frac{n}{p}$	$\frac{1-p}{n}$
Poisson	$X \sim \mathcal{P}(\lambda)$	λ	$\frac{1}{\lambda}$
Uniforme	$X \sim \mathcal{U}(a, b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{3(a+b)^2}$
Erlang	$X \sim \mathcal{E}(n, \lambda)$	$\frac{\lambda}{n}$	$\frac{1}{n}$
Normale	$X \sim \mathcal{N}(\mu, \sigma)$	μ	$\frac{\sigma^2}{\mu^2}$

prima delle (8.21). Procedendo in modo analogo, per ogni reale positivo b risulta:

$$P\{X \leq E(X) - \varepsilon\} = P\{-X + E(X) + b \geq \varepsilon + b\} \leq \frac{\text{Var}(X) + b^2}{(\varepsilon + b)^2},$$

da cui, con la stessa scelta di cui sopra, si ottiene la seconda delle (8.21). \square

È anche possibile determinare una limitazione inferiore alla probabilità che una variabile aleatoria assuma valori positivi, come mostra il teorema seguente.

Teorema 8.7 *Se X è una variabile aleatoria non negativa, risulta:*

$$P(X > 0) = 1 - P(X = 0) \geq \frac{[E(X)]^2}{E(X^2)}. \quad (8.23)$$

Dimostrazione Dall'assunta non negatività di X , per ogni intero positivo k si ha:

$$\begin{aligned} E(X^k) &= E(X^k | X > 0) P(X > 0) + E(X^k | X = 0) P(X = 0) \\ &= E(X^k | X > 0) P(X > 0). \end{aligned} \quad (8.24)$$

Ponendo $k = 2$ nella (8.24) e ricordando che $E(X^2 | X > 0) \geq [E(X | X > 0)]^2$, si ottiene:

$$E(X^2) = E(X^2 | X > 0) P(X > 0) \geq [E(X | X > 0)]^2 P(X > 0). \quad (8.25)$$

Inoltre, dalla (8.24) per $k = 1$ si ha:

$$[E(X | X > 0)]^2 P(X > 0) = \frac{[E(X | X > 0) P(X > 0)]^2}{P(X > 0)} = \frac{[E(X)]^2}{P(X > 0)}. \quad (8.26)$$

Pertanto, utilizzando la (8.26) nella (8.25) risulta

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

da cui segue la (8.23). \square

Per variabili non negative, alla (8.23) si può anche dare la forma seguente:

$$P(X = 0) = 1 - P(X > 0) \leq \frac{\text{Var}(X)}{E(X^2)}. \quad (8.27)$$

Si noti che, se X è assolutamente continua e non negativa entrambe le disuguaglianze (8.23) e (8.27) perdono di significato. Infatti, essendo $P(X = 0) = 0$, la (8.27) esprime l'ovvia relazione $\text{Var}(X) \geq 0$; analogamente, essendo $P(X > 0) = 1$ per la non negatività di X , la (8.23) esprime l'ovvia relazione $[E(X)]^2 \leq E(X^2)$.

Esempio 8.7 Se $X \sim \mathcal{B}(n, p)$ si ha $P(X = 0) = (1 - p)^n$, $E(X) = np$ e $\text{Var}(X) = np(1 - p)$. La (8.27) conduce quindi alla disuguaglianza

$$(1 - p)^n \leq \frac{1 - p}{1 - p + np} \quad (n = 1, 2, \dots; 0 < p < 1).$$

Se invece $X \sim \mathcal{P}(\lambda)$ risulta $P(X = 0) = e^{-\lambda}$, $E(X) = \text{Var}(X) = \lambda$ e dalla (8.27) si ottiene la ben nota disuguaglianza frequentemente utilizzata in contesti matematici diversi

$$e^{-\lambda} \leq \frac{1}{1 + \lambda} \quad (\lambda > 0).$$

\diamond

Introdurremo ora la cosiddetta *disuguaglianza di Chernoff* che coinvolge la funzione generatrice dei momenti e quindi implicitamente fa intervenire tutti i momenti della variabile aleatoria e non soltanto i primi due come avviene per le disuguaglianze finora dimostrate.

Teorema 8.8 (Disuguaglianza di Chernoff) *Sia X una variabile aleatoria la cui funzione generatrice dei momenti $M_X(s) = E(e^{sX})$ è finita almeno in un intorno dell'origine. Per ogni reale positivo a , si ha*

$$P(X \geq a) \leq e^{-sa} M_X(s) \quad \text{per } s \geq 0 \quad (8.28)$$

$$P(X \leq a) \leq e^{-sa} M_X(s) \quad \text{per } s \leq 0. \quad (8.29)$$

Dimostrazione Per $s = 0$ le (8.28) e (8.29) sono banalmente soddisfatte. Per $s > 0$, utilizzando la disuguaglianza di Markov (8.1) si ottiene:

$$P(X \geq a) = P\{e^{sX} \geq e^{sa}\} \leq \frac{E[e^{sX}]}{e^{sa}} = e^{-sa} M_X(s),$$

e quindi la (8.28) è dimostrata. Se è invece $s < 0$, utilizzando nuovamente la (8.1) si ricava:

$$P(X \leq a) = P\{e^{sX} \geq e^{sa}\} \leq \frac{E[e^{sX}]}{e^{sa}} = e^{-sa} M_X(s),$$

ossia la (8.29). \square

La disuguaglianza di Chernoff sussiste per ogni reale s per il quale $M_X(s)$ è finita; quindi, per ottenere la migliore limitazione, basta scegliere nelle (8.28) e (8.29) il reale s_0 che minimizza $e^{-s x} M_X(s)$.

Esempio 8.8 Sia $X \sim \mathcal{B}(n, p)$. Essendo $E(X) = np$ e $\text{Var}(X) = np(1-p)$, se si pone $\varepsilon = x - np$ nella prima delle disuguaglianze di Chebyshev unilaterali (8.21) si ha:

$$P(X \geq x) \leq \frac{np(1-p)}{np(1-p) + (x - np)^2}. \quad (8.30)$$

Per ogni $s \geq 0$ dalla disuguaglianza di Chernoff (8.28) risulta:

$$P(X \geq x) \leq e^{-s x} M_X(s) = e^{-s x} [1 - p + p e^s]^n. \quad (8.31)$$

Poiché il reale s_0 che minimizza $e^{-s x} [1 - p + p e^s]^n$ se $np \leq x < n$ è tale che $e^{s_0} = [x(1-p)]/[p(n-x)]$, dalla (8.31) si ottiene la seguente disuguaglianza:

$$P(X \geq x) \leq e^{-s_0 x} M_X(s_0) = \left[\frac{p(n-x)}{x(1-p)} \right]^x \left[\frac{n(1-p)}{n-x} \right]^n \quad (np \leq x < n). \quad (8.32)$$

La Tabella 8.7 consente di confrontare le probabilità binomiali (8.3), la limitazione superiore fornita dalla (8.30) nelle colonne 3 e 6 e la limitazione superiore fornita dalla (8.32) nelle colonne 4 e 7 per alcune scelte di n e p tali che $np = 1$.

Tabella 8.7– Per alcune scelte di n e p tali che $np = 1$, sono riportate le probabilità binomiali (8.3) (colonne 2 e 5), il maggiorante di cui alla (8.30) (colonne 3 e 6), ed il maggiorante dato dalla (8.32) (colonne 4 e 7).

x	$P(X \geq x)$ $p = 0.05$ $n = 20$	Chebyshev $p = 0.05$ $n = 20$	Chernoff $p = 0.05$ $n = 20$	$P(X \geq x)$ $p = 0.1$ $n = 10$	Chebyshev $p = 0.1$ $n = 10$	Chernoff $p = 0.1$ $n = 10$
1	0.6415	1.0000	1.0000	0.6513	1.0000	1.0000
2	0.2642	0.4872	0.6616	0.2639	0.4737	0.6414
3	0.0755	0.1919	0.2454	0.0702	0.1837	0.2151
4	0.0159	0.0955	0.0611	0.0128	0.0909	0.0445
5	0.0026	0.0560	0.0111	0.0016	0.0533	0.0060

Una limitazione più debole può essere ricavata dalla (8.31) facendo uso della disuguaglianza¹

$$1 - p + p e^s = 1 + p(e^s - 1) \leq \exp\{p(e^s - 1)\},$$

ottenendo

$$P(X \geq x) \leq \exp\{-s x + np(e^s - 1)\}.$$

Il reale s_0 che minimizza $\exp\{-s x + np(e^s - 1)\}$ minimizza anche $-s x + np(e^s - 1)$, così che $s_0 = \ln[x/(np)]$ se $x \geq np$. Pertanto dalla (8.31) si ottiene anche la seguente più

¹La disuguaglianza segue immediatamente dallo sviluppo in serie di MacLaurin dell'esponenziale.

debole, ma spesso utilizzata, limitazione:

$$P(X \geq x) \leq \left(\frac{np}{x}\right)^x e^{x-np} \quad (np \leq x < n). \quad (8.33)$$

◇

Esempio 8.9 Sia $X \sim \mathcal{P}(\lambda)$. Si ha:

$$P(X \geq x) = \sum_{k=\lceil x \rceil}^{+\infty} \frac{\lambda^k}{k!} e^{-\lambda} \quad (x \geq 0), \quad (8.34)$$

con $E(X) = \text{Var}(X) = \lambda$. Ponendo $\varepsilon = x - \lambda$ nella prima delle disuguaglianze di Chebyshev unilaterali (8.21), si ottiene:

$$P(X \geq x) \leq \frac{\lambda}{\lambda + (x - \lambda)^2}. \quad (8.35)$$

Per ogni $s \geq 0$, dalla disuguaglianza di Chernoff si ricava:

$$P(X \geq x) \leq e^{-sx} M_X(s) = e^{-sx} \exp\{\lambda(e^s - 1)\} = \exp\{\lambda(e^s - 1) - sx\}.$$

Il reale s_0 che minimizza $\exp\{\lambda(e^s - 1) - sx\}$ è $\ln(x/\lambda)$ se $x \geq \lambda$. Pertanto, per $x \geq \lambda$ si ottiene:

$$P(X \geq x) \leq e^{-s_0 x} M_X(s_0) = \exp\left\{\lambda\left(\frac{x}{\lambda} - 1\right) - x \ln\left(\frac{x}{\lambda}\right)\right\},$$

ossia:

$$P(X \geq x) \leq \frac{e^{-\lambda} (\lambda e)^x}{x^x} \quad (x \geq \lambda). \quad (8.36)$$

La Tabella 8.8 riporta le probabilità di Poisson (8.34), i maggioranti forniti dalla (8.35) e quelli forniti dalla (8.36) per alcune scelte di x e di λ . ◇

Esempio 8.10 Siano X_1, X_2, \dots, X_n variabili aleatorie indipendenti con funzione di probabilità

$$P(X_r = x) = \begin{cases} 1/2, & x = -1 \\ 1/2, & x = 1 \\ 0, & \text{altrimenti} \end{cases} \quad (r = 1, 2, \dots, n)$$

e sia $Y = X_1 + X_2 + \dots + X_n$. Si ha:

$$P(Y = y) = \begin{cases} \binom{n}{(n-y)/2} \left(\frac{1}{2}\right)^n, & y = 0, \pm 1, \dots, \pm n; \ n - y \text{ intero pari} \\ 0, & \text{altrimenti.} \end{cases} \quad (8.37)$$

Essendo $E(Y) = 0$ e $\text{Var}(Y) = n$, se si pone $\varepsilon = y$ nella prima delle disuguaglianze di Chebyshev unilaterali (8.21) risulta:

$$P(Y \geq y) \leq \frac{n}{n + y^2}. \quad (8.38)$$

Tabella 8.8– Per alcune scelte di λ e x , sono riportate le probabilità di Poisson (8.34) nelle colonne 2 e 5, il maggiorante fornito dalla (8.35) nelle colonne 3 e 6, il maggiorante fornito dalla (8.36) nelle colonne 4 e 7.

x	$P(X \geq x)$	Chebyshev	Chernoff	$P(X \geq x)$	Chebyshev	Chernoff
	$\lambda = 0.5$	$\lambda = 0.5$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1$	$\lambda = 1$
1	0.3935	0.6667	0.8244	0.6321	1.0000	1.0000
2	0.0902	0.1818	0.2801	0.2642	0.5000	0.6796
3	0.0144	0.0741	0.0564	0.0803	0.2000	0.2737
4	0.0018	0.0392	0.0081	0.0190	0.1000	0.0785
5	0.0002	0.0241	0.0009	0.0037	0.0588	0.0175
	$\lambda = 1.5$	$\lambda = 1.5$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2$	$\lambda = 2$
1	0.7769	0.8571	-	0.8647	0.6667	-
2	0.4422	0.8571	0.9274	0.5940	1.0000	1.0000
3	0.1912	0.4000	0.5602	0.3233	0.6667	0.8054
4	0.0656	0.1935	0.2409	0.1429	0.3333	0.4618
5	0.0186	0.1091	0.0805	0.0527	0.1818	0.2057
6	0.0045	0.0690	0.0220	0.0166	0.1111	0.0749
7	0.0009	0.0472	0.0051	0.0045	0.0741	0.0231
8	0.0002	0.0343	0.0010	0.0011	0.0526	0.0062

Poiché la funzione generatrice di ognuna delle X_r è data da

$$M_X(s) = E(e^{sX_1}) = \frac{1}{2}e^{-s} + \frac{1}{2}e^s = \cosh s,$$

per le ipotesi di indipendenza e di identica distribuzione di X_1, X_2, \dots, X_n , dal Teorema 5.19 segue $M_Y(s) = [M_X(s)]^n$, così che dalla disuguaglianza di Chernoff per ogni $s \geq 0$ si ha:

$$P(Y \geq y) \leq e^{-sy} [\cosh s]^n. \quad (8.39)$$

Il reale s_0 che minimizza $e^{-sy} [\cosh s]^n$ è tale che $e^{2s_0} = (n+y)/(n-y)$ se $0 \leq y < n$; pertanto, dalla (8.39) si ottiene:

$$P(Y \geq y) \leq \left(\frac{n-y}{n+y}\right)^{y/2} \left(\frac{n}{\sqrt{(n-y)(n+y)}}\right)^n \quad (0 \leq y < n). \quad (8.40)$$

La Tabella 8.9 consente di confrontare le probabilità $P(Y \geq y)$ ottenute tramite la (8.37), il maggiorante fornito dalla (8.38) nelle colonne 3 e 6, il maggiorante (8.40) nelle colonne 4 e 7 per $n = 10, 15$ e per varie scelte di y .

Una limitazione più debole può essere ricavata dalla (8.39) facendo uso della disuguaglianza

$$\cosh s = \sum_{k=0}^{+\infty} \frac{s^{2k}}{(2k)!} = \sum_{k=0}^{+\infty} \frac{s^{2k}}{k!} \frac{k!}{(2k)!} \leq \sum_{k=0}^{+\infty} \frac{(s^2/2)^k}{k!} = \exp\left\{\frac{s^2}{2}\right\}, \quad (8.41)$$

Tabella 8.9– Per $n = 10, 15$ e per varie scelte di y sono elencate le probabilità $P(Y \geq y)$ ed i maggioranti forniti rispettivamente dalle (8.38) e (8.40).

y	$P(Y \geq y)$ $n = 10$	Chebyshev $n = 10$	Chernoff $n = 10$	$P(Y \geq y)$ $n = 15$	Chebyshev $n = 15$	Chernoff $n = 15$
1	0.3770	0.9091	0.9512	0.5000	0.9375	0.9672
2	0.3770	0.7143	0.8176	0.3036	0.7895	0.8748
3	0.1719	0.5263	0.6332	0.3036	0.6250	0.7393
4	0.1719	0.3846	0.4392	0.1509	0.4839	0.5828
5	0.0547	0.2857	0.2703	0.1509	0.3750	0.4276
6	0.0547	0.2174	0.1455	0.0592	0.2941	0.2911
7	0.0107	0.1694	0.0669	0.0592	0.2344	0.1830
8	0.0107	0.1351	0.0252	0.0176	0.1899	0.1056

ottenendo

$$P(Y \geq y) \leq e^{-s y + n s^2 / 2}.$$

Il reale che minimizza $e^{-s y + n s^2 / 2}$ è $s_0 = y/n$ se $y \geq 0$. Pertanto dalle (8.39) e (8.41) si ricava:

$$P(Y \geq y) \leq \exp\left\{-\frac{y^2}{2n}\right\} \quad (0 \leq y < n),$$

una limitazione certo più debole delle precedenti. \diamond

Come si evince dai risultati riportati nelle Tabelle 8.7, 8.8 e 8.9, per valori sufficientemente grandi della variabile la disuguaglianza di Chernoff fornisce limitazioni migliori rispetto a quelle ottenute dalla disuguaglianza di Chebyshev.

8.3 Disuguaglianze coinvolgenti i soli momenti

Vogliamo ora introdurre una disuguaglianza che si riferisce ai valori medi e non alle probabilità: la cosiddetta *disuguaglianza di Jensen*. Essa riguarda il valore medio di trasformazioni convesse di variabili aleatorie. Per illustrare il concetto di convessità che questa disuguaglianza coinvolge, ricordiamo che una funzione $g(x)$ continua in un intervallo aperto $I = (a, b)$ si dice convessa se comunque si scelgano due punti x_1 e x_2 in I , con $x_1 < x_2$, per ogni reale $\alpha \in (0, 1)$ risulta:

$$g[\alpha x_1 + (1 - \alpha) x_2] \leq \alpha g(x_1) + (1 - \alpha) g(x_2). \quad (8.42)$$

In altri termini, g è convessa se la corda tracciata tra due punti qualsiasi della curva non giace mai al di sotto della curva stessa.

Se invece g è concava, ossia se $-g$ è convessa, in luogo della (8.42) si ha:

$$g[\alpha x_1 + (1 - \alpha) x_2] \geq \alpha g(x_1) + (1 - \alpha) g(x_2).$$

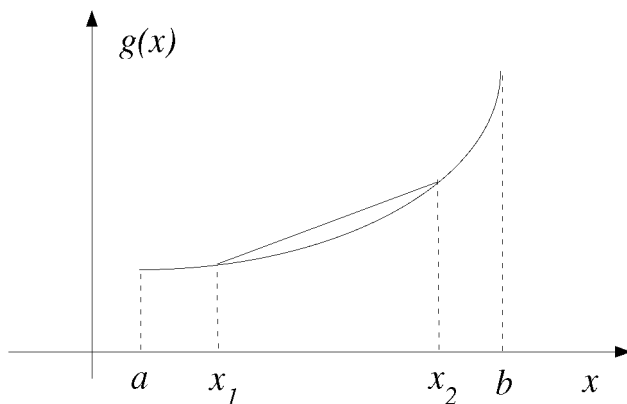


Figura 8.1 – Illustrazione della disuguaglianza (8.42).

Si noti che se g possiede derivate prima e seconda, la convessità in I equivale a richiedere che risulta $g''(x) \geq 0$ per $x \in I$, mentre la concavità in I equivale a richiedere $g''(x) \leq 0$ per $x \in I$.

Ciò premesso, prendiamo in considerazione in primo luogo una variabile aleatoria discreta X con funzione di probabilità

$$P(X = x) = \begin{cases} p, & x = x_1 \\ 1 - p, & x = x_2 \\ 0, & \text{altrimenti.} \end{cases}$$

Se $g(x)$ è una funzione convessa, in virtù della (8.42) risulta:

$$E[g(X)] = p g(x_1) + (1 - p) g(x_2) \geq g[p x_1 + (1 - p) x_2] = g[E(X)].$$

Questo risultato può essere esteso a una variabile aleatoria X qualsiasi per la quale esistono finiti sia $E(X)$ che $E[g(X)]$.

Teorema 8.9 (Disuguaglianza di Jensen) *Sia X una variabile aleatoria con valore medio finito e sia g una funzione convessa e continua in tutto l'intervallo contenente i valori assunti da X con probabilità non nulla. Se $E[g(X)]$ esiste finito risulta:*

$$E[g(X)] \geq g[E(X)], \quad (8.43)$$

il segno di uguaglianza sussistendo solo se X è degenere o se g è lineare.

Dimostrazione Dimostreremo il teorema sotto l'ipotesi che g sia rappresentabile mediante la formula di Taylor di punto iniziale $E(X) = \mu$ con termine complementare di Lagrange di ordine due, ossia se si ha:

$$g(x) = g(\mu) + g'(\mu) (x - \mu) + \frac{1}{2} g''(\xi) (x - \mu)^2$$

per qualche reale ξ interno all'intervallo di estremi x e μ . Per l'ipotizzata convessità di g risulta $g''(x) \geq 0$, così che per ogni valore assunto da X con probabilità non nulla si ha $g(x) \geq g(\mu) + g'(\mu)(x - \mu)$. Pertanto si ottiene la disuguaglianza

$$g(X) \geq g(\mu) + g'(\mu)(X - \mu).$$

Per la proprietà di linearità del valore medio si ha poi:

$$E[g(X)] \geq g(\mu) + g'(\mu) E(X - \mu) = g(\mu),$$

da cui segue la (8.43). □

In particolare, se $g(x) = x^\beta$ con $\beta > 1$, dalla (8.43) si ottiene $E(X^\beta) \geq [E(X)]^\beta$. Inoltre sussiste il seguente

Corollario 8.1 *Sia X una variabile aleatoria con funzione generatrice dei momenti $M_X(s)$ finita. Per ogni $s \in \mathbb{R}$ si ha:*

$$M_X(s) \geq e^{s E(X)}. \quad (8.44)$$

Dimostrazione Segue immediatamente dalla (8.43) ponendo $g(x) = e^{sx}$. □

Se g è concava, la disuguaglianza di Jensen (8.43) si inverte, ossia risulta $E[g(X)] \leq g[E(X)]$. In particolare, se $g(x) = x^\beta$ con $0 < \beta < 1$ si ottiene $E(X^\beta) \leq [E(X)]^\beta$. Inoltre, se $g(x) = \ln x$ ed X è una variabile aleatoria positiva, risulta $E(\ln X) \leq \ln[E(X)]$.

Vogliamo ora introdurre due importanti disuguaglianze riguardanti i valori medi, ossia la *disuguaglianza di Schwarz-Hölder* e la *disuguaglianza di Minkowsky*. Cominciamo, a tal fine, con l'enunciare il seguente

Lemma 8.1 (Disuguaglianza di Young) *Siano $y_1, y_2, \dots, y_n, \vartheta_1, \vartheta_2, \dots, \vartheta_n$ reali positivi tali che $\vartheta_1 + \vartheta_2 + \dots + \vartheta_n = 1$. Si ha:*

$$y_1 y_2 \cdots y_n \leq \sum_{r=1}^n \vartheta_r y_r^{1/\vartheta_r}. \quad (8.45)$$

Dimostrazione Osserviamo anzitutto che risulta:

$$y_1 y_2 \cdots y_n = e^{\ln y_1} e^{\ln y_2} \cdots e^{\ln y_n} = e^{\vartheta_1 z_1 + \vartheta_2 z_2 + \dots + \vartheta_n z_n},$$

dove l'ultima uguaglianza è stata ottenuta ponendo $\ln y_r = \vartheta_r z_r$ per $r = 1, 2, \dots, n$. Essendo $g(x) = e^x$ una funzione convessa, dalla (8.42) si ha:

$$|y_1 y_2 \cdots y_n| \leq \vartheta_1 e^{z_1} + \vartheta_2 e^{z_2} + \dots + \vartheta_n e^{z_n} = \sum_{r=1}^n \vartheta_r y_r^{1/\vartheta_r},$$

e quindi la (8.45). □

Teorema 8.10 (Disuguaglianza di Schwarz–Hölder) Se (X_1, X_2, \dots, X_n) è un vettore aleatorio e $\alpha_1, \alpha_2, \dots, \alpha_n$ sono reali positivi tali che $\alpha_1^{-1} + \alpha_2^{-1} + \dots + \alpha_n^{-1} = 1$, risulta:

$$E(|X_1 X_2 \cdots X_n|) \leq \prod_{r=1}^n [E(|X_r|^{\alpha_r})]^{1/\alpha_r}, \quad (8.46)$$

sempre che i valori medi al secondo membro siano finiti e diversi da zero.

Dimostrazione Se si pone

$$\vartheta_r = \frac{1}{\alpha_r}, \quad Y_r = \frac{|X_r|}{[E(|X_r|^{\alpha_r})]^{1/\alpha_r}} \quad (r = 1, 2, \dots, n),$$

utilizzando la disuguaglianza di Young (8.45) risulta:

$$\begin{aligned} \frac{E(|X_1 X_2 \cdots X_n|)}{\prod_{r=1}^n [E(|X_r|^{\alpha_r})]^{1/\alpha_r}} &= E(Y_1 Y_2 \cdots Y_n) \leq E\left[\sum_{r=1}^n \frac{1}{\alpha_r} Y_r^{\alpha_r}\right] \\ &= \sum_{r=1}^n \frac{1}{\alpha_r} E(Y_r^{\alpha_r}) = \sum_{r=1}^n \frac{1}{\alpha_r} = 1. \end{aligned} \quad (8.47)$$

Dall'essere dunque superiormente limitato dall'unità il rapporto al primo membro della (8.47), si riconosce immediatamente la (8.46). \square

Si noti che nel caso particolare $n = 2$, $\alpha_1 = \alpha_2 = 2$, $X_1 = X$, $X_2 = Y$, la disuguaglianza di Schwarz–Hölder diventa

$$[E(|XY|)]^2 \leq E(X^2) E(Y^2),$$

ossia si identifica con la disuguaglianza di Schwarz (5.74).

Teorema 8.11 (Disuguaglianza di Minkowsky) Sia (X_1, X_2, \dots, X_n) un vettore aleatorio e sia $\alpha > 1$. Risulta:

$$[E(|X_1 + X_2 + \dots + X_n|^\alpha)]^{1/\alpha} \leq \sum_{r=1}^n [E(|X_r|^\alpha)]^{1/\alpha}, \quad (8.48)$$

sempre che i valori medi al secondo membro siano finiti e diversi da zero.

Dimostrazione Sia $Y = X_1 + X_2 + \dots + X_n$. Essendo $\alpha > 1$, per la linearità del valore medio si ha

$$E(|Y|^\alpha) = E[|Y| |Y|^{\alpha-1}] \leq E\left[\sum_{r=1}^n |X_r| |Y|^{\alpha-1}\right] = \sum_{r=1}^n E(|X_r| |Y|^{\alpha-1}). \quad (8.49)$$

Dalla disuguaglianza di Schwarz–Hölder (8.46) si ricava poi:

$$E(|X_r| |Y|^{\alpha-1}) \leq [E(|X_r|^\alpha)]^{1/\alpha} [E(|Y|^{\beta(\alpha-1)})]^{1/\beta} \quad (r = 1, 2, \dots, n),$$

con $\beta = \alpha/(\alpha - 1)$. Facendo uso di questa nella (8.49), si ottiene infine:

$$E(|Y|^\alpha) \leq \sum_{r=1}^n [E(|X_r|^\alpha)]^{1/\alpha} [E(|Y|^\alpha)]^{(\alpha-1)/\alpha},$$

da cui segue immediatamente la (8.48). \square

Esempio 8.11 Siano X_1, X_2, \dots, X_n variabili aleatorie positive indipendenti e identicamente distribuite con valore medio finito μ e siano

$$H = \frac{n}{\frac{1}{X_1} + \frac{1}{X_2} + \dots + \frac{1}{X_n}}, \quad G = (X_1 X_2 \cdots X_n)^{1/n}, \quad \bar{X} = \frac{1}{n} \sum_{r=1}^n X_r, \quad (8.50)$$

rispettivamente le loro medie armonica, geometrica e campionaria. Vogliamo mostrare che per $n = 1, 2, \dots$ sussistono le disuguaglianze $E(H) \leq E(G) \leq E(\bar{X})$.

A tal fine, osserviamo in primo luogo che per la proprietà di linearità del valore medio e per l'ipotesi di identica distribuzione di X_1, X_2, \dots, X_n si ha $E(\bar{X}) = \mu$. Ponendo $g(x) = x^{1/n}$, dalla disuguaglianza di Jensen per funzioni concave scaturisce:

$$\begin{aligned} E(G) &= E[(X_1 X_2 \cdots X_n)^{1/n}] \leq [E(X_1 X_2 \cdots X_n)]^{1/n} \\ &= [E(X_1)E(X_2) \cdots E(X_n)]^{1/n} = \mu, \end{aligned}$$

avendo utilizzato nelle ultime due uguaglianze rispettivamente le ipotesi di indipendenza e di identica distribuzione di X_1, X_2, \dots, X_n . Risulta quindi $E(G) \leq E(\bar{X})$. Inoltre, essendo le variabili aleatorie positive, facendo uso della disuguaglianza di Young con $\vartheta_1 = \vartheta_2 = \dots = \vartheta_n = 1/n$ si ottiene:

$$\left(\frac{1}{X_1}\right)^{1/n} \left(\frac{1}{X_2}\right)^{1/n} \cdots \left(\frac{1}{X_n}\right)^{1/n} \leq \frac{1}{n} \left[\frac{1}{X_1} + \frac{1}{X_2} + \dots + \frac{1}{X_n}\right],$$

da cui segue $E(H) \leq E[(X_1 X_2 \cdots X_n)^{1/n}]$, ossia $E(H) \leq E(G)$. \diamond

8.4 Limitazioni per somme di variabili aleatorie indipendenti

In questo paragrafo vogliamo considerare alcune disuguaglianze riguardanti somme di variabili aleatorie indipendenti. A tal fine forniamo anzitutto una disuguaglianza riguardante la funzione generatrice dei momenti.

Lemma 8.2 Sia X una variabile aleatoria dotata di valore medio finito μ e tale che $P(a \leq X \leq b) = 1$ con a e b reali. Allora, per ogni $s \in \mathbb{R}$ risulta:

$$M_X(s) \leq \frac{b - \mu}{b - a} e^{sa} + \frac{\mu - a}{b - a} e^{sb}. \quad (8.51)$$

Dimostrazione Per la convessità della funzione $g(x) = e^x$ nell'intervallo $[a, b]$, dalla (8.42) ponendo $\alpha = (b - x)/(b - a)$, $x_1 = s a$, $x_2 = s b$, per ogni $x \in [a, b]$ si ha:

$$e^{sx} \leq \frac{b-x}{b-a} e^{sa} + \frac{x-a}{b-a} e^{sb}.$$

Essendo $P(a \leq X \leq b) = 1$, sussiste la disuguaglianza:

$$e^{sX} \leq \frac{b-X}{b-a} e^{sa} + \frac{X-a}{b-a} e^{sb}. \quad (8.52)$$

La (8.51) segue poi immediatamente passando ai valori medi nella (8.52). \square

In particolare, dal Lemma 8.2 scaturisce che se X è una variabile aleatoria a valore medio nullo e tale che $P(|X| \leq c) = 1$ con c reale positivo, allora per ogni $s \in \mathbb{R}$ risulta $M_X(s) \leq \cosh(sc)$.

Teorema 8.12 (Disuguaglianza di Hoeffding) *Siano X_1, X_2, \dots, X_n variabili aleatorie indipendenti ognuna dotata di valore medio finito $E(X_r) = \mu_r$ ($r = 1, 2, \dots, n$) e tali che per $r = 1, 2, \dots, n$ risulti $P\{a_r \leq X_r \leq b_r\} = 1$. Posto $Y = X_1 + X_2 + \dots + X_n$, per ogni $c > 0$ si ha:*

$$P\{Y - E(Y) \geq c\} \leq \exp\left\{-\frac{2c^2}{\sum_{r=1}^n (b_r - a_r)^2}\right\} \quad (8.53)$$

$$P\{Y - E(Y) \leq -c\} \leq \exp\left\{-\frac{2c^2}{\sum_{r=1}^n (b_r - a_r)^2}\right\}. \quad (8.54)$$

Dimostrazione Per ogni $s \geq 0$, utilizzando la disuguaglianza di Chernoff (8.28) si ricava:

$$P\{Y - E(Y) \geq c\} \leq e^{-sc} M_{Y-E(Y)}(s) = e^{-sc} e^{-sE(Y)} \prod_{r=1}^n M_{X_r}(s), \quad (8.55)$$

dove l'ultima uguaglianza segue per l'indipendenza di X_1, X_2, \dots, X_n . Facendo uso del Lemma 8.2, per ogni $s \geq 0$ si ha:

$$P\{Y - E(Y) \geq c\} \leq e^{-s[c+E(Y)]} \prod_{r=1}^n \left[\frac{b_r - \mu_r}{b_r - a_r} e^{sa_r} + \frac{\mu_r - a_r}{b_r - a_r} e^{sb_r} \right]. \quad (8.56)$$

Posto

$$a = \sum_{r=1}^n a_r, \quad b = \sum_{r=1}^n b_r, \quad p_r = \frac{\mu_r - a_r}{b_r - a_r}, \quad 1 - p_r = \frac{b_r - \mu_r}{b_r - a_r} \quad (8.57)$$

la (8.56) si può scrivere:

$$P\{Y - E(Y) \geq c\} \leq e^{-s[c+E(Y)-a]} \prod_{r=1}^n \left[1 - p_r + p_r e^{s(b_r - a_r)} \right]. \quad (8.58)$$

Poiché per ogni $0 \leq p \leq 1$ e $t \geq 0$ la funzione $e^{-pt}(1 - p + pe^t)$ è non crescente, si ha

$$e^{-pt}(1 - p + pe^t) \leq 1.$$

Sussiste quindi anche la disuguaglianza più debole

$$e^{-pt}(1 - p + pe^t) \leq e^{t^2/8},$$

che applicata alla (8.58) conduce per ogni $s \geq 0$ a scrivere

$$\begin{aligned} P\{Y - E(Y) \geq c\} &\leq e^{-s[c+E(Y)-a]} \prod_{r=1}^n \exp\left\{\frac{s^2(b_r - a_r)^2}{8}\right\} \exp\{s p_r (b_r - a_r)\} \\ &= \exp\left\{-s c + \frac{s^2}{8} \sum_{r=1}^n (b_r - a_r)^2\right\}. \end{aligned} \quad (8.59)$$

Il reale che minimizza $\exp\left\{-s c + (s^2/8) \sum_{r=1}^n (b_r - a_r)^2\right\}$ è $s_0 = 4c / \sum_{r=1}^n (b_r - a_r)^2$. Facendo uso di tale valore nella (8.59) si ottiene la (8.53).

Per dimostrare la (8.54) utilizziamo la disuguaglianza di Chernoff (8.29). Pertanto, per ogni $s \leq 0$, si ha:

$$P\{Y - E(Y) \leq -c\} \leq e^{s c} M_{Y-E(Y)}(s) = e^{s c} e^{-s E(Y)} \prod_{r=1}^n M_{X_r}(s), \quad (8.60)$$

dove l'uguaglianza scaturisce nuovamente dall'indipendenza di X_1, X_2, \dots, X_n . Facendo uso del Lemma 8.2 e delle posizioni (8.57), per ogni $s \leq 0$ si ottiene:

$$\begin{aligned} P\{Y - E(Y) \geq c\} &\leq e^{s[c-E(Y)+b]} \prod_{r=1}^n \left[p_r + (1 - p_r) e^{-s(b_r - a_r)} \right] \\ &\leq e^{s[c-E(Y)+b]} \prod_{r=1}^n \exp\left\{\frac{s^2(b_r - a_r)^2}{8}\right\} \exp\{-s(1 - p_r)(b_r - a_r)\} \\ &= \exp\left\{s c + \frac{s^2}{8} \sum_{r=1}^n (b_r - a_r)^2\right\}. \end{aligned} \quad (8.61)$$

Il reale che minimizza $\exp\left\{s c + (s^2/8) \sum_{r=1}^n (b_r - a_r)^2\right\}$ è $s_0 = -4c / \sum_{r=1}^n (b_r - a_r)^2$. Facendo uso di tale valore nella (8.61), si ottiene la (8.54). \square

Nel caso di somma di variabili aleatorie indipendenti di Bernoulli, la disuguaglianza di Hoeffding conduce al seguente

Corollario 8.2 *Siano X_1, X_2, \dots, X_n variabili aleatorie indipendenti con $X_r \sim \mathcal{B}(1, p)$ per $r = 1, 2, \dots, n$. Se $Y = X_1 + X_2 + \dots + X_n$, allora per ogni $c > 0$ si ha:*

$$P\{Y - np \geq c\} \leq e^{-2c^2/n}, \quad P\{Y - np \leq -c\} \leq e^{-2c^2/n}. \quad (8.62)$$

Tabella 8.10– Per $n = 10$, $p = 0.1$ e $p = 0.5$ sono riportate le probabilità binomiali in colonne 2 e 4, il maggiorante fornito dalla (8.63) in colonne 3 e 5, e il maggiorante dato dalla (8.64) in colonna 6.

c	$P(X - 1 \geq c)$ $p = 0.1$	Chebyshev $p = 0.1$	$P(X - 5 \geq c)$ $p = 0.5$	Chebyshev $p = 0.5$	Hoeffding
1	0.6126	0.9000	0.7539	2.5000	1.6375
2	0.0702	0.2250	0.3437	0.6250	0.8987
3	0.0128	0.1000	0.1093	0.2778	0.3306
4	0.0016	0.0563	0.0216	0.1563	0.0815
5	0.0001	0.0360	0.0020	0.1000	0.0135

Dimostrazione Segue immediatamente dalle (8.53) e (8.54) ponendo $a_r = 0$ e $b_r = 1$ per $r = 1, 2, \dots, n$. \square

Esempio 8.12 Sia $X \sim \mathcal{B}(n, p)$. Ponendo $E(X) = np$ e $\text{Var}(X) = np(1-p)$ nella disuguaglianza di Chebyshev (8.9), per ogni $c > 0$ si ottiene:

$$P(|X - np| \geq c) \leq \frac{np(1-p)}{c^2}, \quad (8.63)$$

mentre dalla disuguaglianza (8.62) per ogni $c > 0$ si ha:

$$P(|X - np| \geq c) = P(X - np \geq c) + P(X - np \leq -c) \leq 2e^{-c^2/n}. \quad (8.64)$$

Nella Tabella 8.10 confrontiamo le probabilità binomiali $P(|X - np| \geq c)$ con la limitazione superiore in (8.63) e con la limitazione superiore in (8.64) per $n = 10$ e per due diversi valori di p . \diamond

Esempio 8.13 Riprendiamo in considerazione gli Esempi 8.2 e 8.4 e sia $X \sim \mathcal{B}(n, 1/2)$. Ricordiamo che la disuguaglianza (8.4) fornisce $P\{X \geq 3n/4\} \leq 2/3$, mentre dalla disuguaglianza di Chebyshev si ricava $P\{X \geq 3n/4\} \leq 4/n$. Facendo invece uso della (8.62) si ottiene la disuguaglianza:

$$P\left(X \geq \frac{3n}{4}\right) = P\left(X - \frac{n}{2} \geq \frac{n}{4}\right) \leq e^{-n/8}, \quad (8.65)$$

che fornisce un maggiorante che tende a zero esponenzialmente con n . \diamond

8.5 Altre disuguaglianze rilevanti

Introduciamo due importanti disuguaglianze che svolgono un ruolo fondamentale nello studio di leggi della probabilità: la *disuguaglianza di Kolmogorov* e la *disuguaglianza di Lévy*.

Teorema 8.13 (Disuguaglianza di Kolmogorov) Siano X_1, X_2, \dots, X_n variabili aleatorie indipendenti ognuna a valore medio nullo e momento del secondo ordine finito. Inoltre, sia $Y_r = X_1 + X_2 + \dots + X_r$ per $r = 1, 2, \dots, n$. Per ogni $a > 0$ si ha:

$$P\left\{\max(|Y_1|, |Y_2|, \dots, |Y_n|) \geq a\right\} \leq \frac{E(Y_n^2)}{a^2}. \quad (8.66)$$

Dimostrazione Consideriamo i seguenti eventi necessari ed incompatibili:

$$\begin{aligned} A_0 &= \{\omega \in \Omega : |Y_1| < a, |Y_2| < a, \dots, |Y_n| < a\} \\ A_1 &= \{\omega \in \Omega : |Y_1| \geq a\} \\ A_k &= \{\omega \in \Omega : |Y_1| < a, |Y_2| < a, \dots, |Y_{k-1}| < a, |Y_k| \geq a\} \quad (k = 2, 3, \dots, n). \end{aligned} \quad (8.67)$$

Si noti che per $k = 1, 2, \dots, n$ l'evento A_k si realizza quando k è il primo indice per il quale si ha $|Y_k| \geq a$. È evidente che risulta

$$\left\{ \omega \in \Omega : \max(|Y_1|, |Y_2|, \dots, |Y_n|) \geq a \right\} = \bigcup_{r=1}^n A_r.$$

Se indichiamo con $I_{A_r}(\omega)$ la funzione indicatrice dell'evento A_r , allora per $r = 0, 1, \dots, n$ risulta $Y_r^2 I_{A_r} \geq a^2 I_{A_r}$. Quindi:

$$\begin{aligned} P\left\{ \max(|Y_1|, |Y_2|, \dots, |Y_n|) \geq a \right\} &= P\left(\bigcup_{k=1}^n A_k \right) = \sum_{r=1}^n P(A_r) = \sum_{r=1}^n E(I_{A_r}) \\ &\leq \sum_{r=1}^n \frac{E(Y_r^2 I_{A_r})}{a^2} = \frac{1}{a^2} \sum_{r=1}^n E(Y_r^2 I_{A_r}). \end{aligned} \quad (8.68)$$

Vogliamo ora mostrare che risulta $E(Y_r^2 I_{A_r}) \leq E(Y_n^2 I_{A_r})$ per $r = 1, 2, \dots, n$. A tal fine notiamo che essendo $Y_n^2 = [Y_r + (Y_n - Y_r)]^2$, si ha:

$$E(Y_n^2 I_{A_r}) = E(Y_r^2 I_{A_r}) + E[(Y_n - Y_r)^2 I_{A_r}] + 2E[Y_r (Y_n - Y_r) I_{A_r}]. \quad (8.69)$$

Nell'ultimo valore medio la variabile $Y_n - Y_r$ è funzione di $(X_{r+1}, X_{r+2}, \dots, X_n)$, mentre la variabile $Y_r I_{A_r}$ è funzione di (X_1, X_2, \dots, X_r) . Essendo X_1, X_2, \dots, X_n indipendenti, tali sono anche $Y_n - Y_r$ e $Y_r I_{A_r}$ così che

$$E[Y_r (Y_n - Y_r) I_{A_r}] = E(Y_r I_{A_r}) E(Y_n - Y_r) = E(Y_r I_{A_r}) E(X_{r+1} + X_{r+2} + \dots + X_n) = 0.$$

Poiché $(Y_n - Y_r)^2 I_{A_r} \geq 0$, dalla (8.69) si ricava:

$$E(Y_r^2 I_{A_r}) = E[Y_n^2 I_{A_r}] - E[(Y_n - Y_r)^2 I_{A_r}] \leq E[Y_n^2 I_{A_r}]. \quad (8.70)$$

Facendo uso della (8.70) nella (8.68), in conclusione si ottiene:

$$\begin{aligned} P\left\{ \max(|Y_1|, |Y_2|, \dots, |Y_n|) \geq a \right\} &\leq \frac{1}{a^2} \sum_{r=1}^n E(Y_n^2 I_{A_r}) \\ &= \frac{1}{a^2} E\left(Y_n^2 \sum_{r=1}^n I_{A_r}\right) \leq \frac{1}{a^2} E(Y_n^2), \end{aligned}$$

ossia la (8.66). □

Si noti che se $n = 1$ la disuguaglianza di Kolmogorov si riduce alla (8.7) con $\nu = 2$.

Dalla disuguaglianza di Kolmogorov scaturisce il seguente

Corollario 8.3 Siano $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$ variabili aleatorie indipendenti ognuna con momento del secondo ordine finito e sia $\tilde{Y}_r = \tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_r$ ($r = 1, 2, \dots, n$). Per ogni $a > 0$ si ha:

$$P\left\{\max_{1 \leq r \leq n} |\tilde{Y}_r - E(\tilde{Y}_r)| \geq a\right\} \leq \frac{1}{a^2} \sum_{r=1}^n \text{Var}(\tilde{X}_r). \quad (8.71)$$

Dimostrazione Ponendo $X_r = \tilde{X}_r - E(\tilde{X}_r)$, risulta $E(X_r) = 0$ ed inoltre $E(X_r^2) = \text{Var}(\tilde{X}_r)$ è finito. Per $r = 1, 2, \dots, n$ risulta poi $\tilde{Y}_r - E(\tilde{Y}_r) = X_1 + X_2 + \dots + X_n$. Essendo soddisfatte le ipotesi del Teorema 8.13, dalla (8.66) si ottiene:

$$P\left\{\max_{1 \leq r \leq n} |\tilde{Y}_r - E(\tilde{Y}_r)| \geq a\right\} \leq \frac{\text{Var}(\tilde{Y}_n)}{a^2} = \frac{1}{a^2} \sum_{r=1}^n \text{Var}(\tilde{X}_r), \quad (8.72)$$

avendo fatto ricorso nell'ultima uguaglianza all'indipendenza di $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$. \square

Esiste un'altra interessante disuguaglianza che si riferisce alla distribuzione del $\max_{1 \leq r \leq n} |Y_r|$, con $Y_r = X_1 + X_2 + \dots + X_r$, quando X_1, X_2, \dots, X_n sono indipendenti e possiedono una distribuzione simmetrica intorno allo zero. Questo risultato è noto come *disuguaglianza di Lévy*. Ricordiamo che una variabile aleatoria X è a distribuzione simmetrica rispetto all'asse $x = 0$ se e solo se $-X$ ha la stessa distribuzione di X , ossia se $P(X \leq x) = P(X \geq -x)$ per ogni $x \in \mathbb{R}$.

Teorema 8.14 (Disuguaglianza di Lévy) Siano X_1, X_2, \dots, X_n variabili aleatorie indipendenti le cui funzioni di distribuzione sono tutte simmetriche intorno allo zero. Inoltre, sia $Y_r = X_1 + X_2 + \dots + X_r$ ($r = 1, 2, \dots, n$). Per ogni $a > 0$ si ha:

$$P\left\{\max(Y_1, Y_2, \dots, Y_n) \geq a\right\} \leq 2 P(Y_n \geq a). \quad (8.73)$$

Dimostrazione Per semplicità di notazione, poniamo $M_n = \max(Y_1, Y_2, \dots, Y_n)$. Poiché

$$\begin{aligned} P(M_n \geq a) &= P(M_n \geq a, Y_n \geq a) + P(M_n \geq a, Y_n < a) \\ &= P(Y_n \geq a) + P(M_n \geq a, Y_n < a), \end{aligned} \quad (8.74)$$

per verificare la (8.73) è sufficiente mostrare che risulta $P(M_n \geq a, Y_n < a) \leq P(Y_n \geq a)$. Come già visto nel Teorema 8.13, si ha:

$$\{\omega \in \Omega : M_n \geq a\} = \bigcup_{r=1}^n \{\omega \in \Omega : Y_1 < a, Y_2 < a, \dots, Y_{r-1} < a, Y_r \geq a\}.$$

Quindi,

$$\begin{aligned} P(M_n \geq a, Y_n < a) &= \sum_{r=1}^n P(Y_1 < a, Y_2 < a, \dots, Y_{r-1} < a, Y_r \geq a, Y_n < a) \\ &\leq \sum_{r=1}^n P(Y_1 < a, Y_2 < a, \dots, Y_{r-1} < a, Y_r \geq a, Y_n - Y_r \leq 0) \\ &= \sum_{r=1}^n P(Y_1 < a, Y_2 < a, \dots, Y_{r-1} < a, Y_r \geq a) P(Y_n - Y_r \leq 0), \end{aligned} \quad (8.75)$$

dove l'ultima uguaglianza segue dall'indipendenza di X_1, X_2, \dots, X_n . Poiché per ipotesi ciascuna delle X_1, X_2, \dots, X_n ha distribuzione simmetrica intorno all'asse $x = 0$, si ha $P(Y_n - Y_r \leq 0) = P(Y_n - Y_r \geq 0)$, e quindi dalla (8.75) si ricava:

$$\begin{aligned} P(M_n \geq a, Y_n < a) &\leq \sum_{r=1}^n P(Y_1 < a, Y_2 < a, \dots, Y_{r-1} < a, Y_r \geq a, Y_n - Y_r \geq 0) \\ &\leq \sum_{r=1}^n P(Y_1 < a, Y_2 < a, \dots, Y_{r-1} < a, Y_r \geq a, Y_n \geq a) \\ &= P(M_n \geq a, Y_n \geq a) = P(Y_n \geq a), \end{aligned} \quad (8.76)$$

dove l'ultima uguaglianza segue osservando che l'evento $\{Y_n \geq a\}$ è incluso in $\{M_n \geq a\}$. Infine, utilizzando la (8.76) nella (8.74), si ottiene la (8.73). \square

Esempio 8.14 (*Passeggiata aleatoria semplice simmetrica*) Siano X_1, X_2, \dots, X_n variabili aleatorie indipendenti con funzione di probabilità

$$P(X_i = x) = \begin{cases} 1/2, & x = -1 \\ 1/2, & x = 1 \\ 0, & \text{altrimenti} \end{cases} \quad (i = 1, 2, \dots, n)$$

e sia $Y_r = X_1 + X_2 + \dots + X_r$ per $r = 1, 2, \dots, n$. Si noti che Y_r può descrivere la posizione occupata al tempo r da una particella che, a partire dalla posizione $Y_0 = 0$, si muove su una retta compiendo passi unitari verso destra o verso sinistra con uguali probabilità. La successione di somme parziali Y_0, Y_1, \dots , con $Y_0 = 0$, definisce allora una *passeggiata aleatoria semplice simmetrica*. Nella Figura 8.2 è rappresentata una realizzazione di una siffatta passeggiata aleatoria; sull'asse delle ascisse sono riportati i tempi $n = 1, 2, \dots$, mentre sull'asse delle ordinate sono indicate le corrispondenti posizioni occupate dalla particella.

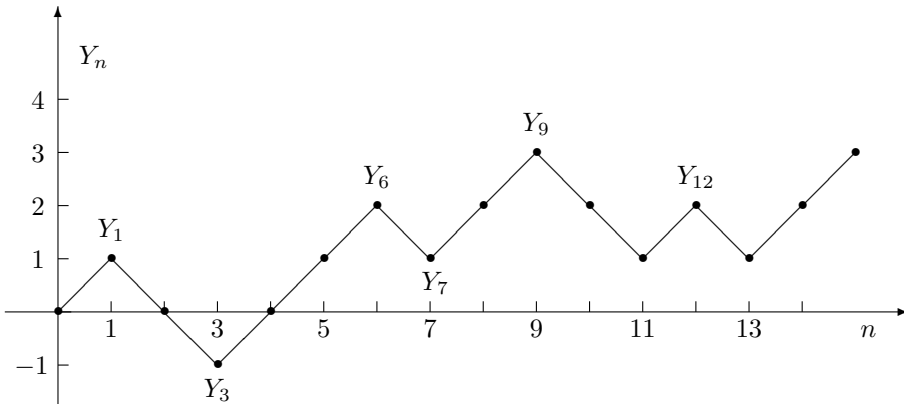


Figura 8.2 – Una particolare realizzazione di una passeggiata aleatoria semplice simmetrica.

Essendo ognuna delle X_1, X_2, \dots, X_n a distribuzione simmetrica intorno all'asse $x = 0$, è possibile applicare la disuguaglianza di Lévy (8.73). Si noti che, facendo uso della (8.37), per $k = 0, \pm 1, \dots, \pm n$ si ha:

$$P(Y_n = k) = \binom{n}{(n-k)/2} \left(\frac{1}{2}\right)^n$$

se $n - k$ è un intero pari e $P(Y_n = k) = 0$ altrimenti. Quindi per $k = 0, \pm 1, \dots, \pm n$ è possibile calcolare $2P(Y_n \geq k)$, ossia il secondo membro della (8.73). Inoltre, è anche possibile determinare l'espressione esatta della probabilità presente al primo membro della (8.73). Infatti, seguendo la linea di dimostrazione del Teorema 8.14, se si pone $M_n = \max(Y_1, Y_2, \dots, Y_n)$, si ha:

$$P(M_n \geq a) = P(Y_n \geq a) + P(M_n \geq a, Y_n < a), \quad (8.77)$$

essendo

$$P(M_n \geq k, Y_n < k) = \sum_{r=1}^n P(Y_1 < k, Y_2 < k, \dots, Y_{r-1} < k, Y_r \geq k, Y_n < k), \quad (8.78)$$

dove l'addendo corrispondente ad $r = n$ è nullo. Per calcolare la (8.78) osserviamo che in una passeggiata aleatoria semplice simmetrica l'evento

$$B_r = \{\omega \in \Omega : Y_1 < k, Y_2 < k, \dots, Y_{r-1} < k, Y_r \geq k, Y_n < k\}$$

si realizza se e solo se

$$\{Y_1 < k, Y_2 < k, \dots, Y_{r-1} < k, Y_r = k, Y_n < k\}.$$

Quindi, poiché le variabili X_1, X_2, \dots, X_n sono indipendenti e ognuna è a distribuzione simmetrica rispetto allo zero, per $r = 1, 2, \dots, n - 1$ sussistono le seguenti uguaglianze:

$$\begin{aligned} P(Y_1 < k, Y_2 < k, \dots, Y_{r-1} < k, Y_r \geq k, Y_n < k) \\ &= P(Y_1 < k, Y_2 < k, \dots, Y_{r-1} < k, Y_r = k, Y_n - Y_r < 0) \\ &= P(Y_1 < k, Y_2 < k, \dots, Y_{r-1} < k, Y_r = k) P(Y_n - Y_r < 0) \\ &= P(Y_1 < k, Y_2 < k, \dots, Y_{r-1} < k, Y_r = k) P(Y_n - Y_r > 0) \\ &= P(Y_1 < k, Y_2 < k, \dots, Y_{r-1} < k, Y_r = k, Y_n > k) \\ &= P(Y_1 < k, Y_2 < k, \dots, Y_{r-1} < k, Y_r \geq k, Y_n > k). \end{aligned}$$

Queste permettono di riscrivere la (8.78) nel seguente modo:

$$\begin{aligned} P(M_n \geq k, Y_n < k) &= \sum_{r=1}^n P(Y_1 < k, Y_2 < k, \dots, Y_{r-1} < k, Y_r \geq k, Y_n > k) \\ &= P(M_n \geq k, Y_n > k) = P(Y_n > k). \end{aligned}$$

Ricordando la (8.77), per $k = 1, 2, \dots, n$ si ha dunque:

$$P\{\max(Y_1, Y_2, \dots, Y_n) \geq k\} = P(Y_n \geq k) + P(Y_n > k) = 2P(Y_n \geq k) - P(Y_n = k), \quad (8.79)$$

che mostra che $P(M_n \geq k)$ si discosta di $P(Y_n = k)$ dal maggiorante presente nella disuguaglianza di Lévy (8.73). \diamond

8.6 Momenti di variabili aleatorie stocasticamente ordinate

Determineremo ora talune relazioni esistenti tra momenti di variabili aleatorie stocasticamente ordinate in distribuzione. Diamo anzitutto la seguente definizione.

Definizione 8.1 Siano X e Y variabili aleatorie con rispettive funzioni di distribuzione $F_X(x)$ e $F_Y(y)$. Si dice che X è maggiore di Y in distribuzione se e solo se per ogni $x \in \mathbb{R}$ risulta $F_X(x) \leq F_Y(x)$. Inoltre, X è minore di Y in distribuzione se e solo se per ogni $x \in \mathbb{R}$ risulta $F_X(x) \geq F_Y(x)$.

Teorema 8.15 Siano X e Y variabili aleatorie positive tali che per un fissato indice k , con $k = 1, 2, \dots$, i loro momenti di ordine k siano finiti. Se X è maggiore di Y in distribuzione, risulta $E(X^k) \geq E(Y^k)$. Inoltre se X è minore di Y in distribuzione, si ha $E(X^k) \leq E(Y^k)$.

Dimostrazione Se X e Y sono variabili aleatorie positive con momenti di ordine k finiti, allora dalla (5.42) si ricava:²

$$E(X^k) = \int_0^{+\infty} kx^{k-1}[1-F_X(x)] dx - \int_{-\infty}^0 kx^{k-1}F_X(x) dx = \int_0^{+\infty} kx^{k-1}[1-F_X(x)] dx.$$

Quindi, se X è maggiore di Y in distribuzione si ha $1 - F_X(x) \geq 1 - F_Y(x)$, così che dalla precedente relazione risulta:

$$E(X^k) \geq \int_0^{+\infty} kx^{k-1}[1-F_Y(x)] dx = E(Y^k),$$

mentre se X è minore di Y in distribuzione si ricava $E(X^k) \leq E(Y^k)$. □

Teorema 8.16 Siano X e Y variabili aleatorie tali che per un fissato intero dispari k , con $k = 1, 3, \dots$, i loro momenti di ordine k sono finiti. Se X è maggiore di Y in distribuzione, risulta $E(X^k) \geq E(Y^k)$. Inoltre se X è minore di Y in distribuzione, si ha $E(X^k) \leq E(Y^k)$.

Dimostrazione Essendo k un intero dispari, per ogni $x \in \mathbb{R}$ si ha $x^{k-1} \geq 0$. Quindi, se X è maggiore di Y in distribuzione dalla (5.42) per $k = 1, 3, \dots$ si ottiene:

$$\begin{aligned} E(X^k) &= \int_0^{+\infty} kx^{k-1}[1-F_X(x)] dx - \int_{-\infty}^0 kx^{k-1}F_X(x) dx \\ &\geq \int_0^{+\infty} kx^{k-1}[1-F_Y(x)] dx - \int_{-\infty}^0 kx^{k-1}F_Y(x) dx = E(Y^k), \end{aligned}$$

mentre se X è minore di Y in distribuzione si ottiene $E(X^k) \leq E(Y^k)$ per $k = 1, 3, \dots$ □

²Si ricordi che la (5.42) consente di esprimere i momenti di variabili aleatorie di natura arbitraria in termini delle corrispondenti funzioni di distribuzione

Esempio 8.15 Siano X_1, X_2, \dots, X_n variabili aleatorie indipendenti, tutte definite nello stesso spazio di probabilità, e siano $U = \max(X_1, X_2, \dots, X_n)$ e $V = \min(X_1, X_2, \dots, X_n)$. Mostriamo che risulta:

$$E(U) \geq \max_{1 \leq r \leq n} \{E(X_r)\}, \quad E(V) \leq \min_{1 \leq r \leq n} \{E(X_r)\}. \quad (8.80)$$

Invero, essendo X_1, X_2, \dots, X_n indipendenti, per $r = 1, 2, \dots, n$ dalla (3.63) segue:

$$F_U(u) = \prod_{i=1}^n F_{X_i}(u) \leq F_{X_r}(u), \quad \forall u \in \mathbb{R},$$

che implica, per la Definizione 8.1, che U è maggiore di X_r in distribuzione. Dal Teorema 8.16 segue quindi $E(U) \geq E(X_r)$ qualunque sia $r = 1, 2, \dots, n$, che implica la prima delle (8.80). Inoltre, ricordando nuovamente che X_1, X_2, \dots, X_n sono indipendenti, dalla (3.64) segue:

$$F_V(v) = 1 - \prod_{i=1}^n [1 - F_{X_i}(v)] \geq 1 - [1 - F_{X_r}(v)] = F_{X_r}(v), \quad \forall v \in \mathbb{R},$$

che comporta, per la Definizione 8.1, che V è minore di X_r in distribuzione. Dal Teorema 8.16 si deduce quindi che risulta $E(V) \leq E(X_r)$ qualunque sia $r = 1, 2, \dots, n$, da cui deriva la seconda delle (8.80). Utilizzando il Teorema 8.15 oppure il Teorema 8.16, le relazioni (8.80) possono essere estese anche a momenti successivi al primo.

Si noti che se ad esempio X_1, X_2, \dots, X_n rappresentano i tempi di corretto funzionamento di n componenti di un dispositivo collegati in parallelo, allora U rappresenta il tempo di corretto funzionamento dell'intero dispositivo. La prima delle (8.80) mostra quindi che il tempo medio di corretto funzionamento dell'intero dispositivo è non inferiore al tempo medio di funzionamento di ognuno dei dispositivi posti in parallelo. Analogamente, se gli n dispositivi sono collegati in serie, allora V rappresenta il tempo di corretto funzionamento dell'intero dispositivo. Quindi la seconda delle (8.80) mostra che il tempo medio di corretto funzionamento dell'intero dispositivo è inferiore del tempo di funzionamento di ognuno dei dispositivi posti in serie. \diamond

Capitolo 9

Teoremi asintotici

9.1 Successioni di variabili aleatorie

Nei precedenti capitoli abbiamo esaminato alcuni semplici problemi riguardanti distribuzioni limite ottenute al divergere di un parametro. Ad esempio, nel Capitolo 4 abbiamo visto in quali condizioni la distribuzione ipergeometrica converge a una distribuzione binomiale (v. Paragrafo 4.2.3) ed abbiamo esaminato in quali condizioni la distribuzione binomiale converge alla distribuzione di Poisson (v. Paragrafo 4.2.7), mentre nel Paragrafo 7.5 abbiamo mostrato che la distribuzione di Student converge a quella normale all'aumentare del numero di gradi di libertà.

In questo capitolo vogliamo analizzare alcuni dei principali problemi asintotici, di rilevante interesse e applicabilità in statistica, coinvolgenti collezioni di un numero infinitamente grande di variabili aleatorie, ordinate in maniera da costituire delle successioni. Ricordiamo che quando prendiamo in esame più variabili aleatorie, occorre supporre che esse siano definite in uno stesso spazio di probabilità (Ω, \mathcal{F}, P) . Una successione di variabili aleatorie va quindi riguardata come una successione $\{X_n(\omega); n = 1, 2, \dots\}$ di funzioni misurabili da (Ω, \mathcal{F}) in $(\mathbb{R}, \mathcal{B})$. Ricordiamo inoltre (v. Paragrafo 3.8) che una successione di variabili aleatorie X_1, X_2, \dots , tutte definite in uno stesso spazio di probabilità, si dicono indipendenti se e solo se per ogni insieme finito di indici distinti i_1, i_2, \dots, i_k le variabili $X_{i_1}, X_{i_2}, \dots, X_{i_k}$ sono indipendenti.

Discuteremo dapprima alcuni tipi di convergenza per successioni di variabili aleatorie e successivamente esporremo alcuni dei principali teoremi limite della probabilità.

9.2 Convergenza di variabili aleatorie

Ci proponiamo in questo paragrafo di analizzare il comportamento limite di successioni di variabili aleatorie mediante lo studio dei seguenti tipi di convergenza: *convergenza quasi certa*, *convergenza in probabilità* e *convergenza in distribuzione*.

Sia X_1, X_2, \dots una successione di variabili aleatorie, tutte definite nello stesso spazio di probabilità (Ω, \mathcal{F}, P) , e sia X una variabile aleatoria anch'essa ivi definita. Esistono

vari modi per introdurre il concetto di limite della successione, ossia per definire che cosa debba intendersi per convergenza della successione di variabili aleatorie X_1, X_2, \dots verso la variabile aleatoria X .

Un criterio naturale di convergenza che emerge dalle familiari nozioni dell'analisi matematica è la *convergenza su Ω* , così definita: per ogni $\varepsilon > 0$ e per ogni $\omega \in \Omega$ esiste un intero positivo $k = k(\varepsilon, \omega)$ tale da aversi

$$|X_n(\omega) - X(\omega)| < \varepsilon \quad \text{per ogni } n > k.$$

In altri termini, si richiede che la successione $X_n(\omega)$ converga ad $X(\omega)$ per ogni $\omega \in \Omega$.

In un contesto di tipo probabilistico risulta però eccessivo richiedere che la convergenza sussista per ogni $\omega \in \Omega$; è sufficiente che essa sussista per tutti i punti ω il cui insieme costituisce un evento di probabilità unitaria. Da ciò scaturisce quindi la seguente definizione:

Definizione 9.1 Una successione di variabili aleatorie X_1, X_2, \dots , tutte definite nello stesso spazio di probabilità (Ω, \mathcal{F}, P) , converge quasi certamente ad una variabile aleatoria X , definita nello stesso spazio di probabilità, se risulta:

$$\begin{aligned} P\left(\lim_{n \rightarrow +\infty} X_n = X\right) &= P\left(\left\{\omega \in \Omega : \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)\right\}\right) \\ &= P\left(\left\{\omega \in \Omega : \forall \varepsilon > 0 \quad \exists k : \forall n \geq k \quad |X_n(\omega) - X(\omega)| < \varepsilon\right\}\right) = 1. \end{aligned} \quad (9.1)$$

Tale convergenza, detta quasi certa, sarà denotata con $X_n \xrightarrow{q.c.} X$. Essa può essere così interpretata: la probabilità dell'evento $\left\{\omega \in \Omega : \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)\right\}$ è unitaria, ossia X_n al crescere di n tende a identificarsi con X con probabilità 1. È poi possibile dimostrare il seguente teorema:

Teorema 9.1 La successione di variabili aleatorie X_1, X_2, \dots converge quasi certamente a X se e solo se per ogni $\varepsilon > 0$ si ha

$$\lim_{n \rightarrow +\infty} P\left(\bigcap_{k=n}^{+\infty} \left\{\omega \in \Omega : |X_k(\omega) - X(\omega)| < \varepsilon\right\}\right) = 1. \quad (9.2)$$

Questo teorema afferma che risulta $X_n \xrightarrow{q.c.} X$ se e solo se tende a 1 la probabilità che tutte le variabili della successione da un certo indice in poi sono prossime a X quanto si vuole. Dalla (9.2) si evince che per mostrare che X_1, X_2, \dots converge quasi certamente a X occorre conoscere la distribuzione congiunta delle variabili aleatorie X, X_1, X_2, \dots .

Esempio 9.1 Data una variabile aleatoria X consideriamo la successione X_1, X_2, \dots , con $X_n = X + 1/n$ ($n = 1, 2, \dots$). Vogliamo dimostrare che risulta $X_n \xrightarrow{q.c.} X$.

Poiché per ogni $\varepsilon > 0$ si ha

$$P\left(\bigcap_{k=n}^{+\infty} \left\{\omega \in \Omega : |X_k(\omega) - X(\omega)| < \varepsilon\right\}\right) = P\left(\bigcap_{k=n}^{+\infty} \left\{\omega \in \Omega : \frac{1}{k} < \varepsilon\right\}\right) = \begin{cases} 0, & n \leq \frac{1}{\varepsilon} \\ 1, & n > \frac{1}{\varepsilon} \end{cases},$$

segue immediatamente che la (9.2) è soddisfatta; quindi $X_n \xrightarrow{q.c.} X$ in virtù del Teorema 9.1. \diamond

Introduciamo ora una ulteriore forma di convergenza, detta convergenza in probabilità, che sarà denotata con la scrittura $X_n \xrightarrow{P} X$.

Definizione 9.2 *Date una successione di variabili aleatorie X_1, X_2, \dots ed una variabile aleatoria X , tutte definite nello stesso spazio di probabilità (Ω, \mathcal{F}, P) , la successione X_1, X_2, \dots si dice convergere in probabilità a X se per ogni $\varepsilon > 0$ risulta:*

$$\lim_{n \rightarrow +\infty} P(|X_n - X| < \varepsilon) = \lim_{n \rightarrow +\infty} P(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \varepsilon\}) = 1. \quad (9.3)$$

La Definizione 9.2 comporta che la convergenza in probabilità dipende in maniera essenziale dalla distribuzione congiunta delle coppie (X_n, X) . Ciò è sottolineato dall'esempio che segue.

Esempio 9.2 Consideriamo la successione X_1, X_2, \dots e la variabile aleatoria X tali che la distribuzione congiunta di (X_n, X) per ogni intero positivo è la seguente:

$$\begin{aligned} P(X_n = 0, X = 1) &= P(X_n = 1, X = 0) = p_n, \\ P(X_n = 0, X = 0) &= P(X_n = 1, X = 1) = \frac{1}{2} - p_n, \end{aligned}$$

con $0 \leq p_n \leq 1/2$. Sia X che X_n hanno pertanto distribuzione di Bernoulli di parametro $1/2$. Osserviamo poi che si ha:

$$P(|X_n - X| < \varepsilon) = \begin{cases} 1 - 2p_n, & \text{se } 0 < \varepsilon \leq 1 \\ 1, & \text{se } \varepsilon > 1. \end{cases}$$

Da ciò discende che se risulta $\lim_{n \rightarrow +\infty} p_n = 0$, allora $X_n \xrightarrow{P} X$ in virtù della Definizione 9.2. Notiamo che se la successione delle p_n non tende a 0, la successione X_1, X_2, \dots non converge in probabilità ad X , pur essendo tali variabili tutte identicamente distribuite. \diamond

La convergenza quasi certa è più forte della convergenza in probabilità come affermato dal teorema che segue.

Teorema 9.2 *Date una successione di variabili aleatorie X_1, X_2, \dots ed una variabile aleatoria X , tutte definite nello stesso spazio di probabilità (Ω, \mathcal{F}, P) , si ha:*

$$X_n \xrightarrow{q.c.} X \implies X_n \xrightarrow{P} X.$$

Dimostrazione Per ogni $\varepsilon > 0$ risulta:

$$\bigcap_{k=n}^{+\infty} \{\omega \in \Omega : |X_k(\omega) - X(\omega)| < \varepsilon\} \subset \{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \varepsilon\}$$

e quindi

$$P\left(\bigcap_{k=n}^{+\infty} \{\omega \in \Omega : |X_k(\omega) - X(\omega)| < \varepsilon\}\right) \leq P(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \varepsilon\}).$$

Pertanto, dall'ipotesi $X_n \xrightarrow{q.c.} X$ e dalla (9.2), per ogni $\varepsilon > 0$ si ricava:

$$\lim_{n \rightarrow +\infty} P\left(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \varepsilon\}\right) \geq 1,$$

da cui, ricordando la Definizione 9.2, segue la tesi. \square

Un altro tipo di convergenza è la convergenza in distribuzione, che definiamo qui di seguito.

Definizione 9.3 *Date una successione di variabili aleatorie X_1, X_2, \dots con rispettive funzioni di distribuzione $F_{X_1}(x), F_{X_2}(x), \dots$ ed una variabile aleatoria X con funzione di distribuzione $F_X(x)$, la successione X_1, X_2, \dots converge in distribuzione a X se*

$$\lim_{n \rightarrow +\infty} F_{X_n}(x) = F_X(x) \quad (9.4)$$

in tutti i punti $x \in \mathbb{R}$ in cui $F_X(x)$ è continua.

La scrittura $X_n \xrightarrow{d} X$ denota la convergenza in distribuzione ad X della successione X_1, X_2, \dots .

È opportuno sottolineare che il limite di una successione di funzioni di distribuzione non è necessariamente una funzione di distribuzione. Per assicurare la convergenza in distribuzione della successione di variabili aleatorie X_1, X_2, \dots è necessario verificare che la successione $F_{X_1}(x), F_{X_2}(x), \dots$ delle funzioni di distribuzione converga ad una funzione limite ed inoltre occorre verificare che tale funzione limite sia effettivamente una funzione di distribuzione in ogni suo punto di continuità. Comunque, la conoscenza dei valori della funzione $F_X(x)$ nei suoi soli punti di continuità è sufficiente per la sua completa individuazione. Infatti, i valori di una funzione di distribuzione nei punti di discontinuità possono essere ricavati ricorrendo alla proprietà di continuità a destra delle funzioni di distribuzione.

È utile osservare che la convergenza in distribuzione è un tipo di convergenza piuttosto debole, la cui utilità principale consiste nel poter approssimare nei calcoli, per n sufficientemente grande, $F_{X_n}(x)$ con la funzione di distribuzione limite $F_X(x)$.

Esempio 9.3 Siano X_1, X_2, \dots variabili aleatorie indipendenti, ognuna a distribuzione geometrica di parametro p . Per $n = 1, 2, \dots$ sia poi $Y_n = \min(X_1, X_2, \dots, X_n)$. Desideriamo mostrare che $Y_n \xrightarrow{d} Y$, dove Y è una variabile aleatoria degenere che assume il valore 1 con probabilità unitaria.

Per la Definizione 9.3, occorre quindi dimostrare che per ogni $y \in \mathbb{R} - \{1\}$ risulta:

$$\lim_{n \rightarrow +\infty} F_{Y_n}(y) = \begin{cases} 0, & y < 1 \\ 1, & y > 1. \end{cases}$$

Ricordando la (4.16), in virtù della (3.64) si ha:

$$F_{Y_n}(y) = 1 - [1 - F_{X_1}(y)]^n = \begin{cases} 0, & y < 1 \\ 1 - (1 - p)^{kn}, & k \leq y < k + 1 \quad (k = 1, 2, \dots), \end{cases}$$

da cui si trae immediatamente che $\lim_{n \rightarrow +\infty} F_{Y_n}(y) = 0$ per $y < 1$ e $\lim_{n \rightarrow +\infty} F_{Y_n}(y) = 1$ per $y > 1$, così che in definitiva risulta $Y_n \xrightarrow{d} Y$. \diamond

Esempio 9.4 Siano X_1, X_2, \dots variabili aleatorie indipendenti, con $X_i \sim \mathcal{U}(0, 1)$. Consideriamo le successioni Y_1, Y_2, \dots e Z_1, Z_2, \dots con

$$Y_n = \min(X_1, X_2, \dots, X_n), \quad Z_n = n Y_n \quad (n = 1, 2, \dots).$$

Desideriamo mostrare che $Y_n \xrightarrow{d} Y$, dove Y è una variabile aleatoria degenere che assume il valore 0 con probabilità unitaria e inoltre che $Z_n \xrightarrow{d} Z$, con Z di distribuzione esponenziale di valore medio 1.

Per la Definizione 9.3, occorre quindi dimostrare che per ogni $y \in \mathbb{R} - \{0\}$ risulta:

$$\lim_{n \rightarrow +\infty} F_{Y_n}(y) = \begin{cases} 0, & y < 0 \\ 1, & y > 0, \end{cases}$$

e che per ogni $z \in \mathbb{R}$ si ha:

$$\lim_{n \rightarrow +\infty} F_{Z_n}(z) = \begin{cases} 0, & z < 0 \\ 1 - e^{-z}, & z \geq 0. \end{cases}$$

Ricordando la (4.28), in virtù della (3.64) si ha:

$$F_{Y_n}(y) = 1 - [1 - F_{X_1}(y)]^n = \begin{cases} 0, & y < 0 \\ 1 - (1 - y)^n, & 0 \leq y < 1 \\ 1, & y \geq 1, \end{cases}$$

implicante $\lim_{n \rightarrow +\infty} F_{Y_n}(y) = 0$ per $y < 0$ e $\lim_{n \rightarrow +\infty} F_{Y_n}(y) = 1$ per $y > 0$. Quindi, $Y_n \xrightarrow{d} Y$. Inoltre, essendo

$$F_{Z_n}(z) = P(n Y_n \leq z) = F_{Y_n}\left(\frac{z}{n}\right) = \begin{cases} 0, & z < 0 \\ 1 - \left(1 - \frac{z}{n}\right)^n, & 0 \leq z < n \\ 1, & z \geq n, \end{cases}$$

si trae che $\lim_{n \rightarrow +\infty} F_{Z_n}(z) = 0$ per $z < 0$ e $\lim_{n \rightarrow +\infty} F_{Z_n}(z) = 1 - e^{-z}$ per $z \geq 0$, il che comporta $Z_n \xrightarrow{d} Z$. \diamond

Esempio 9.5 Dato $X \sim B(1, p)$, per $n = 1, 2, \dots$ si ponga $Z_n = n X$. In virtù della (4.4) si ha:

$$F_{Z_n}(z) = F_X\left(\frac{z}{n}\right) = \begin{cases} 0, & z < 0 \\ 1 - p, & 0 \leq z < n \\ 1, & z \geq n, \end{cases}$$

il cui limite al crescere di n è 0 per $x < 0$ e $1 - p$ per $x > 0$. Poiché $\lim_{n \rightarrow +\infty} F_{Z_n}(x)$ non è una funzione di distribuzione, la successione Z_1, Z_2, \dots non converge in distribuzione. \diamond

Nell'Esempio 9.2 abbiamo visto che una successione X_1, X_2, \dots di variabili aleatorie identicamente distribuite può non convergere in probabilità. È invece ovvio che una siffatta successione converge in distribuzione. Questa osservazione indica che la convergenza in distribuzione è più debole della convergenza in probabilità, come puntualizzato dal seguente teorema:

Teorema 9.3 *Date una successione di variabili aleatorie X_1, X_2, \dots ed una variabile aleatoria X tutte definite nello stesso spazio di probabilità (Ω, \mathcal{F}, P) , sussiste la seguente implicazione:*

$$X_n \xrightarrow{P} X \implies X_n \xrightarrow{d} X.$$

Inoltre, se X è una variabile aleatoria degenera si ha:

$$X_n \xrightarrow{d} X \implies X_n \xrightarrow{P} X.$$

Dimostrazione Per $n = 1, 2, \dots$ sia $F_{X_n}(x)$ la funzione di distribuzione di X_n , e sia $F_X(x)$ la funzione di distribuzione di X . Allora, per ogni $\varepsilon > 0$ risulta:

$$\begin{aligned} F_{X_n}(x) &= P(X_n \leq x) = P(X_n \leq x, |X_n - X| < \varepsilon) + P(X_n \leq x, |X_n - X| \geq \varepsilon) \\ &= P(X_n \leq x, X_n - \varepsilon < X < X_n + \varepsilon) + P(X_n \leq x, |X_n - X| \geq \varepsilon) \\ &\leq P(X \leq x + \varepsilon) + P(|X_n - X| \geq \varepsilon) \\ &= F_X(x + \varepsilon) + P(|X_n - X| \geq \varepsilon), \end{aligned}$$

avendosi inoltre:

$$\begin{aligned} F_X(x - \varepsilon) &= P(X \leq x - \varepsilon) \\ &= P(X \leq x - \varepsilon, |X_n - X| < \varepsilon) + P(X \leq x - \varepsilon, |X_n - X| \geq \varepsilon) \\ &= P(X \leq x - \varepsilon, X - \varepsilon < X_n < X + \varepsilon) + P(X \leq x - \varepsilon, |X_n - X| \geq \varepsilon) \\ &\leq P(X_n \leq x) + P(|X_n - X| \geq \varepsilon) \\ &= F_{X_n}(x) + P(|X_n - X| \geq \varepsilon). \end{aligned}$$

Di qui, per ogni $\varepsilon > 0$ segue:

$$F_X(x - \varepsilon) - P(|X_n - X| \geq \varepsilon) \leq F_{X_n}(x) \leq F_X(x + \varepsilon) + P(|X_n - X| \geq \varepsilon). \quad (9.5)$$

Poiché per ipotesi $X_n \xrightarrow{P} X$, si ha che $P(|X_n - X| \geq \varepsilon)$ tende a zero per $n \rightarrow +\infty$, così che dalla (9.5) si trae:

$$F_X(x - \varepsilon) \leq \liminf_{n \rightarrow +\infty} F_{X_n}(x) \leq \limsup_{n \rightarrow +\infty} F_{X_n}(x) \leq F_X(x + \varepsilon). \quad (9.6)$$

Se $x \in \mathbb{R}$ è un punto di continuità di $F_X(x)$, per l'arbitrarietà di ε risulta:

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

che, insieme con la (9.6), comporta:

$$\lim_{n \rightarrow +\infty} F_{X_n}(x) = \liminf_{n \rightarrow +\infty} F_{X_n}(x) = \limsup_{n \rightarrow +\infty} F_{X_n}(x) = F_X(x).$$

Ciò dimostra che $X_n \xrightarrow{d} X$.

Supponiamo ora che $X_n \xrightarrow{d} X$, con X variabile aleatoria degenerare assumente un valore a con probabilità unitaria. Per ogni $x \in \mathbb{R} - \{a\}$ si ha allora:

$$\lim_{n \rightarrow +\infty} F_{X_n}(x) = \begin{cases} 0, & x < a \\ 1, & x > a. \end{cases}$$

Allo scopo di mostrare che $X_n \xrightarrow{P} X$, notiamo che per ogni $\varepsilon > 0$ risulta:

$$\begin{aligned} P(|X_n - X| < \varepsilon) &= P(|X_n - a| < \varepsilon) = P(a - \varepsilon < X_n < a + \varepsilon) \\ &\geq P\left(a - \varepsilon < X_n \leq a + \frac{\varepsilon}{2}\right) = F_{X_n}\left(a + \frac{\varepsilon}{2}\right) - F_{X_n}(a - \varepsilon). \end{aligned}$$

Pertanto per ogni $\varepsilon > 0$ si ottiene:

$$\lim_{n \rightarrow +\infty} P(|X_n - X| < \varepsilon) \geq \lim_{n \rightarrow +\infty} F_{X_n}\left(a + \frac{\varepsilon}{2}\right) - \lim_{n \rightarrow +\infty} F_{X_n}(a - \varepsilon) = 1.$$

Poiché la probabilità di un evento non può essere maggiore dell'unità, da qui segue che $X_n \xrightarrow{P} X$. \square

Esempio 9.6 Data una successione X_1, X_2, \dots di variabili aleatorie indipendenti, tali che $X_i \sim \mathcal{U}(0, \vartheta)$, si consideri la successione Y_1, Y_2, \dots con $Y_n = \max(X_1, X_2, \dots, X_n)$. Desideriamo mostrare che $Y_n \xrightarrow{d} Y$, con Y variabile aleatoria degenerare che assume il valore ϑ con probabilità unitaria. Per il Teorema 9.3 ciò è equivalente a richiedere che la successione Y_1, Y_2, \dots converga in distribuzione alla variabile aleatoria degenerare Y , ossia che per ogni $y \in \mathbb{R} - \{\vartheta\}$ risulti

$$\lim_{n \rightarrow +\infty} F_{Y_n}(y) = \begin{cases} 0, & y < \vartheta \\ 1, & y > \vartheta. \end{cases}$$

Dall'ipotesi di indipendenza e identica distribuzione delle variabili X_1, X_2, \dots, X_n , in virtù delle (3.63) e (4.31) si ha:

$$F_{Y_n}(y) = \begin{cases} 0, & y < 0 \\ (y/\vartheta)^n, & 0 \leq y < \vartheta \\ 1, & y \geq \vartheta, \end{cases}$$

da cui si ricava immediatamente $\lim_{n \rightarrow +\infty} F_{Y_n}(y) = 0$ per $y < \vartheta$, e $\lim_{n \rightarrow +\infty} F_{Y_n}(y) = 1$ per $y > \vartheta$. Pertanto $Y_n \xrightarrow{d} Y$ e, per il Teorema 9.3, $Y_n \xrightarrow{P} Y$. \diamond

Lo studio della convergenza in distribuzione di successioni di variabili aleatorie attraverso l'uso della Definizione 9.3 è spesso eccessivamente complicato richiedendo la conoscenza di tutte le funzioni di distribuzione coinvolte. Una valida alternativa consiste nel far ricorso alle funzioni generatrici dei momenti che, se finite in un intorno dell'origine, individuano univocamente le funzioni di distribuzione. Se la funzione generatrice dei momenti non è finita in un intorno dell'origine, si rende necessario ricorrere alla funzione caratteristica cui si è accennato nel Paragrafo 5.8.1.

Riportiamo ora gli enunciati di alcuni importanti teoremi la dimostrazione di alcuni dei quali eccede i limiti della presente trattazione.

Teorema 9.4 Se $X_n \xrightarrow{d} X$ e se si suppone che le funzioni generatrici dei momenti $M_{X_n}(s)$ ($n = 1, 2, \dots$) e $M_X(s)$ sono finite in un intorno $|s| < s_0$, con $s_0 > 0$, allora la successione $M_{X_1}(s), M_{X_2}(s), \dots$ converge a $M_X(s) = E(e^{sX})$ per ogni s appartenente a tale intorno.

Teorema 9.5 Se una successione $M_{X_1}(s), M_{X_2}(s), \dots$ di funzioni generatrici dei momenti converge a una funzione limite $M_0(s)$ continua in un intorno $|s| < s_0$, con $s_0 > 0$, allora $M_0(s)$ è essa stessa una funzione generatrice dei momenti e $X_n \xrightarrow{d} X$, dove X è una variabile aleatoria di cui $M_0(s)$ è la funzione generatrice dei momenti.

Esempio 9.7 Sia X_1, X_2, \dots una successione di variabili aleatorie e si supponga che per $n = 1, 2, \dots$ risulti:

$$p_{X_n}(x) = \begin{cases} \frac{1}{n}, & x = 0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n} \\ 0, & \text{altrimenti.} \end{cases}$$

In virtù della (5.96), possiamo calcolare la funzione generatrice dei momenti di X_n :

$$M_{X_n}(s) = \sum_{k=0}^{n-1} e^{sk/n} \frac{1}{n} = \frac{1}{n} \sum_{k=0}^{n-1} [e^{s/n}]^k = \frac{1}{n} \frac{1 - e^s}{1 - e^{s/n}} \quad (n = 1, 2, \dots),$$

da cui, utilizzando la regola di De L'Hospital, si ottiene:

$$M_0(s) = \lim_{n \rightarrow +\infty} M_{X_n}(s) = \frac{e^s - 1}{s}.$$

Come mostrato nel Paragrafo 5.8.3, $M_0(s)$ si riconosce essere la funzione generatrice dei momenti di una variabile aleatoria $X \sim U(0, 1)$. Dal Teorema 9.5 si trae quindi che $X_n \xrightarrow{d} X$. \diamond

Sussiste inoltre il seguente teorema, che ci limitiamo ad enunciare.

Teorema 9.6 Date una successione di variabili aleatorie X_1, X_2, \dots ed una variabile aleatoria X tutte definite nello stesso spazio di probabilità (Ω, \mathcal{F}, P) , risulta $X_n \xrightarrow{d} X$ se e solo se

$$\lim_{n \rightarrow +\infty} E[g(X_n)] = E[g(X)]$$

per ogni funzione $g(x)$ continua e limitata.

Il Teorema 9.6 prende in considerazione soltanto funzioni g limitate, così che quando questa condizione non è rispettata e $X_n \xrightarrow{d} X$, non è detto che $\lim_{n \rightarrow +\infty} E[g(X_n)]$ sia $E[g(X)]$, come mostra il seguente esempio.

Esempio 9.8 Sia X_1, X_2, \dots una successione di variabili aleatorie con

$$p_{X_n}(x) = \begin{cases} 1 - 1/n, & x = 0 \\ 1/n, & x = n \\ 0, & \text{altrimenti} \end{cases}$$

per $n = 1, 2, \dots$, così che X_n è caratterizzata dalla seguente funzione di distribuzione:

$$F_{X_n}(x) = \begin{cases} 0, & x < 0 \\ 1 - \frac{1}{n}, & 0 \leq x < n \\ 1, & x \geq n \end{cases} \quad (n = 1, 2, \dots).$$

Ne segue $X_n \xrightarrow{d} X$, dove X è una variabile aleatoria degenera che assume il valore 0 con probabilità 1. Consideriamo ora la funzione non limitata $g(x) = x$. Notiamo che $E(X_n) = 1$ mentre $E(X) = 0$. Ciò mostra che in generale $\lim_{n \rightarrow +\infty} E(X_n)$ e $E(X)$ non coincidono necessariamente, pur avendosi $X_n \xrightarrow{d} X$. Se invece consideriamo la funzione limitata $g(x) = e^{-|x|}$, allora risulta:

$$E[g(X_n)] = E[e^{-|X_n|}] = 1 - \frac{1}{n} + \frac{e^{-n}}{n},$$

che tende a 1 per $n \rightarrow +\infty$. Inoltre, avendo dimostrato che X è degenera in 0, $E[g(X)] = g(0) = 1$. Quindi, se $X_n \xrightarrow{d} X$ si ha $\lim_{n \rightarrow +\infty} E[e^{-|X_n|}] = E[e^{-|X|}]$. \diamond

Un importante risultato inerente trasformazioni continue di variabili aleatorie nel contesto della convergenza in distribuzione è espresso nel seguente teorema:

Teorema 9.7 Se $X_n \xrightarrow{d} X$ e se $h(x)$ è una funzione continua, allora $h(X_n) \xrightarrow{d} h(X)$.

Dimostrazione Sia $Y_n = h(X_n)$ e $Y = h(X)$, con h funzione continua. Si vuole dimostrare che $Y_n \xrightarrow{d} Y$. A tale scopo consideriamo un'arbitraria funzione $g(y)$ continua e limitata. La funzione $z(x) = g(h(x))$ è ancora continua e limitata, così che per il Teorema 9.6 si ha quindi:

$$\lim_{n \rightarrow +\infty} E[z(X_n)] = E[z(X)],$$

ossia

$$\lim_{n \rightarrow +\infty} E[g(Y_n)] = E[g(Y)]$$

per ogni funzione $g(y)$ continua e limitata. Ricorrendo nuovamente al Teorema 9.6 si ricava infine $Y_n \xrightarrow{d} Y$. \square

9.3 Teorema centrale di convergenza

Vogliamo ora introdurre uno dei più importanti risultati della teoria della probabilità, noto quale *teorema centrale di convergenza* o *teorema centrale del limite*, che fornisce una semplice ed utile approssimazione alla distribuzione della somma di variabili aleatorie indipendenti, evidenziando al contempo la grande importanza della distribuzione normale.

Teorema 9.8 (Teorema centrale di convergenza) Sia X_1, X_2, \dots una successione di variabili aleatorie, definite nello stesso spazio di probabilità, indipendenti e identicamente distribuite con valore medio μ finito e varianza σ^2 finita e positiva. Posto per ogni intero n

positivo $Y_n = X_1 + X_2 + \dots + X_n$, per ogni $x \in \mathbb{R}$ risulta:

$$\lim_{n \rightarrow +\infty} P\left(\frac{Y_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy = \Phi(x), \quad (9.7)$$

ossia la successione delle variabili aleatorie standardizzate

$$Z_n = \frac{Y_n - E(Y_n)}{\sqrt{\text{Var}(Y_n)}} = \frac{Y_n - n\mu}{\sqrt{n}\sigma}, \quad n = 1, 2, \dots$$

converge in distribuzione alla variabile aleatoria normale standard.

Dimostrazione Ci limitiamo a dimostrare il teorema nell'ipotesi aggiuntiva che le variabili aleatorie X_1, X_2, \dots siano dotate di funzione generatrice dei momenti $M(s)$ finita in un intorno $(-s_0, s_0)$ dell'origine. Essendo le variabili aleatorie X_1, X_2, \dots indipendenti e identicamente distribuite, per le note proprietà della funzione generatrice dei momenti si ha:

$$\begin{aligned} M_{Z_n}(s) &= E(e^{sZ_n}) = E\left[\exp\left\{s \sum_{i=1}^n \frac{X_i - \mu}{\sigma\sqrt{n}}\right\}\right] = E\left[\exp\left\{\frac{s}{\sigma\sqrt{n}} \left(\sum_{i=1}^n X_i - n\mu\right)\right\}\right] \\ &= \exp\left\{\frac{-\mu s \sqrt{n}}{\sigma}\right\} \left[M\left(\frac{s}{\sigma\sqrt{n}}\right)\right]^n, \end{aligned} \quad (9.8)$$

dove $M(s)$ denota la funzione generatrice dei momenti di ciascuna delle variabili aleatorie X_i della successione. Dalla (9.8) segue:

$$\ln M_{Z_n}(s) = n \psi\left(\frac{s}{\sigma\sqrt{n}}\right) - \frac{\mu s \sqrt{n}}{\sigma} \quad (9.9)$$

dove $\psi(s) = \ln M(s)$. Poiché

$$\psi'(s) = \frac{d \ln M(s)}{ds} = \frac{M'(s)}{M(s)}, \quad \psi''(s) = \frac{d^2 \ln M(s)}{ds^2} = \frac{M''(s)M(s) - [M'(s)]^2}{M^2(s)}$$

si ha $\psi'(0) = \mu$ e $\psi''(0) = \sigma^2$. Utilizzando la formula di Taylor di punto iniziale $s = 0$ con termine complementare di Lagrange di ordine tre, si ha:

$$\psi(s) = \psi(0) + \psi'(0)s + \psi''(0)\frac{s^2}{2} + \frac{1}{6}\psi'''(\xi)s^3 = \mu s + \sigma^2 \frac{s^2}{2} + \frac{1}{6}\psi'''(\xi)s^3$$

dove $\xi \in (-s_0, s_0)$. Pertanto la (9.9) si può riscrivere al seguente modo:

$$\ln M_{Z_n}(s) = \frac{s^2}{2} + \frac{1}{6}\psi'''(\xi)\frac{s^3}{\sigma^3\sqrt{n}}, \quad (9.10)$$

con $-s_0/(\sigma\sqrt{n}) < \xi < s_0/(\sigma\sqrt{n})$. Il secondo addendo nella (9.10) contiene il fattore $1/\sqrt{n}$ che tende a zero per $n \rightarrow +\infty$, così che $\ln M_{Z_n}(s)$ tende a $s^2/2$ per $n \rightarrow +\infty$. Quindi,

$$\lim_{n \rightarrow +\infty} M_{Z_n}(s) = \lim_{n \rightarrow +\infty} \exp[\ln M_{Z_n}(s)] = e^{-s^2/2}, \quad (9.11)$$

che si riconosce essere la funzione generatrice dei momenti della variabile aleatoria normale standard (v. Paragrafo 5.8.3). In virtù del Teorema 9.5 si conclude che Z_n converge in distribuzione alla variabile aleatoria normale standard. \square

Si noti che nel teorema centrale di convergenza è richiesto che le variabili aleatorie della successione X_1, X_2, \dots abbiano valore medio finito e varianza finita e positiva. Pertanto esistono successioni di variabili aleatorie per le quali la (9.7) non sussiste. Un esempio tipico è quello in cui le variabili della successione X_1, X_2, \dots hanno la distribuzione di Cauchy (5.17), nel qual caso $E(X_i)$ non esiste.

Il Teorema 9.8 mostra inoltre che sottraendo a $Y_n = X_1 + X_2 + \dots + X_n$ la sua media $n\mu$ e dividendo la differenza per la deviazione standard di Y_n , ossia per $\sigma\sqrt{n}$, si ottiene una variabile aleatoria standardizzata Z_n la cui funzione di distribuzione $P(Z_n \leq x)$ è per n sufficientemente grande approssimativamente normale standard:

$$P\left(\frac{Y_n - n\mu}{\sigma\sqrt{n}} \leq x\right) \simeq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy = \Phi(x) \quad (x \in \mathbb{R}). \quad (9.12)$$

Notiamo inoltre che per ogni coppia di reali α, β , con $\alpha < \beta$, dalla (9.7) segue anche

$$\lim_{n \rightarrow +\infty} P\left(\alpha < \frac{Y_n - n\mu}{\sigma\sqrt{n}} \leq \beta\right) = \Phi(\beta) - \Phi(\alpha), \quad (9.13)$$

così che, per n sufficientemente grande dalla (9.13) si ha:

$$P\left(\alpha < \frac{Y_n - n\mu}{\sigma\sqrt{n}} \leq \beta\right) \simeq \Phi(\beta) - \Phi(\alpha) \quad (\alpha, \beta \in \mathbb{R}, \alpha < \beta). \quad (9.14)$$

Va menzionato che la bontà delle approssimazioni (9.12) e (9.14) dipende da n , dal tipo di distribuzione delle variabili X_1, X_2, \dots, X_n e dai valori degli estremi x, α, β . Si mostra poi che per n grande l'approssimazione migliora al crescere di n . Nelle applicazioni spesso si verifica che essa è già soddisfacente per $n \geq 30$.

Ponendo $w = n\mu + x\sigma\sqrt{n}$ nella (9.12), $a = n\mu + \alpha\sigma\sqrt{n}$ e $b = n\mu + \beta\sigma\sqrt{n}$ nella (9.14) si ricava:

$$P(Y_n \leq w) \simeq \Phi\left(\frac{w - n\mu}{\sigma\sqrt{n}}\right) \quad (w \in \mathbb{R}), \quad (9.15)$$

$$P(a < Y_n \leq b) \simeq \Phi\left(\frac{b - n\mu}{\sigma\sqrt{n}}\right) - \Phi\left(\frac{a - n\mu}{\sigma\sqrt{n}}\right) \quad (a, b \in \mathbb{R}, a < b). \quad (9.16)$$

Per n grande la distribuzione di $Y_n = X_1 + X_2 + \dots + X_n$ è dunque approssimativamente normale con valore medio $n\mu$ e varianza $n\sigma^2$. Si noti che nelle (9.15) e (9.16) le approssimazioni dipendono soltanto dal valore medio μ e dalla varianza σ^2 di ciascuna delle variabili X_1, X_2, \dots, X_n .

Concludiamo ricordando che se X_1, X_2, \dots, X_n sono variabili aleatorie indipendenti, ognuna avente distribuzione normale di valore medio μ e varianza σ^2 , per ogni n la distribuzione di $Y_n = X_1 + X_2 + \dots + X_n$ è normale con valore medio $n\mu$ e varianza $n\sigma^2$ (v. Paragrafo 5.8.3), così che le (9.15) e (9.16) sussistono con il segno di uguaglianza.

Esempio 9.9 Supponiamo che i tempi, in fissate unità arbitrarie, di esecuzione di n programmi su di un elaboratore elettronico possano essere riguardati come variabili aleatorie X_1, X_2, \dots, X_n indipendenti e identicamente distribuite aventi distribuzione esponenziale di parametro $\lambda = 2$. Nell'ipotesi che $n = 40$, determiniamo un'approssimazione

della probabilità che il tempo totale di esecuzione degli n programmi, indicato con $Y_n = X_1 + X_2 + \dots + X_n$, sia compreso tra 15 e 30.

Per la Proposizione 4.16 risulta $Y_n \sim \mathcal{E}(n, \lambda)$, con $n = 40$ e $\lambda = 2$ di modo che, ricordando la (4.39), si ha:

$$P(15 < Y_{40} \leq 30) = \int_{15}^{30} \frac{2^{40}}{39!} x^{39} e^{-2x} dx = 0.9512. \quad (9.17)$$

Facendo ora uso della (9.16), e ricordando che per una distribuzione esponenziale di parametro $\lambda = 2$ risulta $\mu = \sigma = 1/\lambda = 0.5$, si ottiene:

$$\begin{aligned} P(15 < Y_{40} \leq 30) &\simeq \Phi\left(\frac{30 - 40 \cdot 0.5}{0.5 \sqrt{40}}\right) - \Phi\left(\frac{15 - 40 \cdot 0.5}{0.5 \sqrt{40}}\right) = \Phi(3.16) - \Phi(-1.58) \\ &= \Phi(3.16) - [1 - \Phi(1.58)] = 0.9992 - (1 - 0.9429) = 0.9421, \end{aligned}$$

dove si è fatto uso della relazione $\Phi(x) = 1 - \Phi(-x)$ e della Tabella C.1 dell'Appendice. Si noti che il valore approssimato 0.9421 appena ricavato si discosta per meno del 2% dal valore indicato nella (9.17). \diamond

Esempio 9.10 Supponiamo che X_1, X_2, \dots sia una successione di variabili aleatorie indipendenti con $X_i \sim \mathcal{U}(-1/2, 1/2)$ per $i = 1, 2, \dots$. Come mostrato nel Paragrafo 5.8.3 risulta $E(X_i) = 0$ e $\text{Var}(X_i) = 1/12$. Posto $Y_n = X_1 + X_2 + \dots + X_n$, dalla (9.12) segue:

$$P\left(\frac{Y_n}{\sqrt{n/12}} \leq x\right) \simeq \Phi(x) \quad (x \in \mathbb{R}).$$

Se, in particolare, si sceglie $n = 12$, ossia se si considera la somma di 12 variabili aleatorie indipendenti e uniformi in $(-1/2, 1/2)$, si ha:

$$P(Y_{12} \leq x) \simeq \Phi(x) \quad (x \in \mathbb{R}).$$

Il teorema centrale di convergenza mostra quindi che la funzione di distribuzione della somma di 12 variabili aleatorie indipendenti e uniformi in $(-1/2, 1/2)$ è approssimativamente normale standard. È qui opportuno notare che la simmetria della densità delle X_i concorre a far sì che con sole 12 variabili si ottenga una buona approssimazione alla normale standard. Questo risultato trova applicazione nella costruzione di un semplice algoritmo per la simulazione su elaboratori elettronici di numeri casuali di una variabile a distribuzione normale standard. A tal proposito, va menzionato che gli elaboratori posseggono routine atte a generare numeri casuali tratti da una variabile che con ottima approssimazione è $U \sim \mathcal{U}(0, 1)$. D'altronde, tramite la trasformazione $V = U - 1/2$ è possibile generare numeri tratti da una variabile $V \sim \mathcal{U}(-1/2, 1/2)$. Pertanto, se U_1, U_2, \dots, U_{12} sono indipendenti con $U_i \sim \mathcal{U}(0, 1)$, allora risulta $Y_{12} = U_1 + U_2 + \dots + U_{12} - 6 = V_1 + V_2 + \dots + V_{12}$, con $V_i \sim \mathcal{U}(-1/2, 1/2)$. Per il teorema centrale di convergenza segue che numeri generati da una variabile normale standard possono essere approssimativamente ottenuti generando numeri a partire dalla variabile Y_{12} . \diamond

Se le variabili X_1, X_2, \dots, X_n sono indipendenti e identicamente distribuite con lo stesso valore medio μ finito, possiamo interpretarle come descrittive i risultati di n diverse osservazioni di una variabile aleatoria X . In questo contesto è naturale definire la media campionaria delle n osservazioni:

$$\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n}. \quad (9.18)$$

Per la proprietà di linearità del valore medio e per l'ipotesi di identica distribuzione delle variabili si ha:

$$E(\bar{X}_n) = \frac{E(X_1) + E(X_2) + \dots + E(X_n)}{n} = \mu. \quad (9.19)$$

Inoltre, se si suppone che le X_1, X_2, \dots, X_n sono tutte dotate di varianza σ^2 finita e positiva, per le ipotesi di indipendenza e identica distribuzione delle variabili risulta:

$$\text{Var}(\bar{X}_n) = \frac{\text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n)}{n^2} = \frac{\sigma^2}{n}. \quad (9.20)$$

La (9.20) mostra che per $n = 2, 3, \dots$ la varianza di \bar{X}_n è minore della varianza delle variabili X_1, X_2, \dots, X_n . Ricordando che alla varianza è legata una misura della dispersione di una variabile aleatoria intorno al suo valore medio, segue che \bar{X}_n si discosta da μ meno di quanto avviene per ciascuna delle variabili X_1, X_2, \dots, X_n . Altre interessanti osservazioni sulla media campionaria scaturiscono dal teorema centrale di convergenza, come qui di seguito mostrato.

Corollario 9.1 *Se X_1, X_2, \dots è una successione di variabili aleatorie indipendenti e identicamente distribuite con valore medio μ finito e varianza σ^2 finita e positiva, si ha:*

$$\lim_{n \rightarrow +\infty} P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy = \Phi(x) \quad (x \in \mathbb{R}). \quad (9.21)$$

Dimostrazione Posto $Y_n = X_1 + X_2 + \dots + X_n$, risulta:

$$\frac{Y_n - n\mu}{\sigma\sqrt{n}} = \frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}.$$

Utilizzando poi tale identità nella (9.7) segue la (9.21). \square

La (9.21) mostra che sottraendo a \bar{X}_n la sua media μ e dividendo la differenza per la deviazione standard di \bar{X}_n , ossia per σ/\sqrt{n} , si ottiene una variabile aleatoria la cui funzione di distribuzione è approssimativamente normale standard per n grande:

$$P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq x\right) \simeq \Phi(x) \quad (x \in \mathbb{R}). \quad (9.22)$$

Anche in questo caso la bontà dell'approssimazione (9.22) dipende da n , dal tipo di distribuzione delle variabili X_1, X_2, \dots, X_n e da x . Posto poi $w = \mu + x\sigma/\sqrt{n}$ in (9.22), si ricava:

$$P(\bar{X}_n \leq w) \simeq \Phi\left(\frac{w - \mu}{\sigma/\sqrt{n}}\right) \quad (w \in \mathbb{R}). \quad (9.23)$$

Si noti infine che la (9.23) sussiste con il segno di uguaglianza nel caso di n variabili aleatorie a distribuzione normale, come evidenziato qui di seguito.

Proposizione 9.1 Se X_1, X_2, \dots, X_n sono variabili aleatorie indipendenti, con $X_i \sim \mathcal{N}(\mu, \sigma)$ per $i = 1, 2, \dots, n$, allora $\bar{X}_n \sim \mathcal{N}(\mu, \sigma/\sqrt{n})$.

Dimostrazione Come mostrato nel Paragrafo 5.8.3, la funzione generatrice dei momenti di X_i è:

$$M(s) = \exp\left\{\mu s + \frac{\sigma^2}{2} s^2\right\}.$$

Quindi,

$$M_{\bar{X}_n}(s) = E[e^{s\bar{X}_n}] = \left[M\left(\frac{s}{n}\right)\right]^n = \left[\exp\left\{\mu \frac{s}{n} + \frac{\sigma^2}{2} \frac{s^2}{n^2}\right\}\right]^n = \exp\left\{\mu s + \frac{\sigma^2}{2n} s^2\right\},$$

che si riconosce essere la funzione generatrice dei momenti di una variabile a distribuzione normale con valore medio μ e varianza σ^2/n . Pertanto, $\bar{X}_n \sim \mathcal{N}(\mu, \sigma/\sqrt{n})$. \square

9.4 Cenni alle leggi dei grandi numeri

Le leggi dei grandi numeri rivestono ruolo fondamentale nel calcolo delle probabilità e in statistica. Esamineremo in questo paragrafo innanzitutto tre diverse versioni della cosiddetta *legge debole* dei grandi numeri dovute rispettivamente a Markov, Chebyshev e Khintchin.

Teorema 9.9 (Legge debole dei grandi numeri di Markov) Sia X_1, X_2, \dots una successione di variabili aleatorie indipendenti aventi valori medi finiti e varianze $\sigma_1^2, \sigma_2^2, \dots$ finite, e sia $Y_n = Y_1 + Y_2 + \dots + X_n$ per $n = 1, 2, \dots$. Se risulta

$$\lim_{n \rightarrow +\infty} \frac{\sigma_1^2 + \dots + \sigma_n^2}{n^2} = 0, \quad (9.24)$$

allora per ogni $\varepsilon > 0$ si ha:

$$\lim_{n \rightarrow +\infty} P\left\{\left|\frac{Y_n - E(Y_n)}{n}\right| < \varepsilon\right\} = 1, \quad (9.25)$$

ovvero:

$$\frac{Y_n - E(Y_n)}{n} \xrightarrow{P} 0.$$

Dimostrazione Utilizzando la disuguaglianza di Chebyshev (8.9), per ogni $\varepsilon > 0$ risulta:

$$P\left\{\left|\frac{Y_n - E(Y_n)}{n}\right| \geq \varepsilon\right\} \leq \frac{\text{Var}(Y_n)}{n^2 \varepsilon^2} = \frac{1}{\varepsilon^2} \frac{\text{Var}(X_1) + \dots + \text{Var}(X_n)}{n^2},$$

dove l'uguaglianza segue dalla proprietà additiva della varianza di somme di variabili aleatorie indipendenti. Pertanto:

$$P\left\{\left|\frac{Y_n - E(Y_n)}{n}\right| < \varepsilon\right\} = 1 - P\left\{\left|\frac{Y_n - E(Y_n)}{n}\right| \geq \varepsilon\right\} \geq 1 - \frac{1}{\varepsilon^2} \frac{\sigma_1^2 + \dots + \sigma_n^2}{n^2},$$

che in virtù dell'ipotesi (9.24) conduce immediatamente alla (9.25). \square

È opportuno osservare che essenziale per la dimostrazione del Teorema 9.9 è la validità della relazione $\text{Var}(Y_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n)$, che è garantita dall'ipotesi di indipendenza delle variabili aleatorie X_1, X_2, \dots . Occorre sottolineare che tale uguaglianza sussiste anche nell'ipotesi più debole che le variabili aleatorie X_1, X_2, \dots siano non correlate, ossia che risulti $\text{Cov}(X_i, X_j) = 0$ per ogni $i \neq j$. Pertanto il Teorema 9.9 rimane valido anche se l'ipotesi di indipendenza delle X_1, X_2, \dots viene sostituita da quella di non correlazione delle stesse.

Dal Teorema 9.9 scaturisce il seguente

Teorema 9.10 (Legge debole dei grandi numeri di Chebyshev) *Sia X_1, X_2, \dots una successione di variabili aleatorie indipendenti aventi valori medi finiti e varianze $\sigma_1^2, \sigma_2^2, \dots$ uniformemente limitate e sia $Y_n = X_1 + X_2 + \dots + X_n$. Allora, per ogni $\varepsilon > 0$*

$$\frac{Y_n - E(Y_n)}{n} \xrightarrow{P} 0.$$

Dimostrazione Basta osservare che se le varianze $\sigma_1^2, \sigma_2^2, \dots$ sono uniformemente limitate, ossia se esiste un reale $C > 0$ tale che $\sigma_k^2 \leq C$ per $k = 1, 2, \dots$, allora risulta:

$$\frac{\sigma_1^2 + \dots + \sigma_n^2}{n^2} \leq \frac{C}{n},$$

così che la condizione (9.24) è sempre soddisfatta. \square

Si noti che se X_1, X_2, \dots è una successione di variabili aleatorie indipendenti e identicamente distribuite con valore medio μ finito e varianza σ^2 finita, la legge debole dei grandi numeri di Chebyshev afferma che la media campionaria converge in probabilità a μ , ossia che si ha $\overline{X}_n \xrightarrow{P} \mu$. In questo caso, è possibile ottenere un risultato più forte di quello fornito dalla legge debole dei grandi numeri di Chebyshev rilassando l'ipotesi della finitezza della varianza. Ciò è espresso nel seguente teorema, dovuto a Khintchin.

Teorema 9.11 (Legge debole dei grandi numeri di Khintchin) *Sia X_1, X_2, \dots una successione di variabili aleatorie indipendenti e identicamente distribuite, dotate di valore medio μ finito. Allora per ogni $\varepsilon > 0$ si ha:*

$$\lim_{n \rightarrow +\infty} P\{|\overline{X}_n - \mu| < \varepsilon\} = 1, \quad (9.26)$$

ossia $\overline{X}_n \xrightarrow{P} \mu$.

Dimostrazione La dimostrazione si basa su di una tecnica usata per la prima volta da Markov nel 1907 e successivamente chiamata “metodo del troncamento”. Sia $\alpha > 0$ fissato e per $k = 1, 2, \dots, n$ si ponga:

$$Y_k = \begin{cases} X_k, & |X_k| < \alpha n \\ 0, & |X_k| \geq \alpha n, \end{cases} \quad Z_k = \begin{cases} 0, & |X_k| < \alpha n \\ X_k, & |X_k| \geq \alpha n. \end{cases}$$

Ovviamente, per $k = 1, 2, \dots, n$, si ha $X_k = Y_k + Z_k$. Inoltre, le variabili Y_1, Y_2, \dots, Y_n (così come le variabili Z_1, Z_2, \dots, Z_n) sono indipendenti e identicamente distribuite. Indichiamo con \overline{Y}_n e \overline{Z}_n rispettivamente le medie campionarie di Y_1, Y_2, \dots, Y_n e di Z_1, Z_2, \dots ,

Z_n . Denotato con μ_n e σ_n^2 il valore medio e la varianza di ognuna delle Y_k , per la legge delle alternative si ha:

$$\begin{aligned}\mu_n &= E(Y_k) = E\{Y_k \mid |X_k| < \alpha n\} P(|X_k| < \alpha n) + E\{Y_k \mid |X_k| \geq \alpha n\} P(|X_k| \geq \alpha n) \\ &= E\{X_k \mid |X_k| < \alpha n\} P(|X_k| < \alpha n),\end{aligned}\quad (9.27)$$

$$\begin{aligned}\sigma_n^2 &= E(Y_k^2) - [E(Y_k)]^2 \leq E(Y_k^2) \\ &= E\{Y_k^2 \mid |X_k| < \alpha n\} P(|X_k| < \alpha n) + E\{Y_k^2 \mid |X_k| \geq \alpha n\} P(|X_k| \geq \alpha n) \\ &= E\{X_k^2 \mid |X_k| < \alpha n\} P(|X_k| < \alpha n) \\ &\leq \alpha n E(|X_k|) = \alpha n \gamma,\end{aligned}\quad (9.28)$$

dove si è posto $\gamma = E(|X_k|)$. Quando $n \rightarrow +\infty$, dalla (9.27) segue che $\lim_{n \rightarrow +\infty} \mu_n = E(X_k) = \mu$. Quindi, per ogni $\varepsilon > 0$ esiste un reale n_0 tale che per ogni $n \geq n_0$ risulta $|\mu_n - \mu| < \varepsilon/2$. Ciò implica che sussistono le seguenti disuguaglianze:

$$\mu - \varepsilon < \mu_n - \frac{\varepsilon}{2} < \mu < \mu_n + \frac{\varepsilon}{2} < \mu + \varepsilon.$$

Sfruttando queste, per $n \geq n_0$ segue:

$$\begin{aligned}P(|\bar{Y}_n - \mu| \geq \varepsilon) &= P(\bar{Y}_n \geq \mu + \varepsilon) + P(\bar{Y}_n \leq \mu - \varepsilon) \\ &\leq P\left(\bar{Y}_n \geq \mu_n + \frac{\varepsilon}{2}\right) + P\left(\bar{Y}_n \leq \mu_n - \frac{\varepsilon}{2}\right) = P\left(|\bar{Y}_n - \mu_n| \geq \frac{\varepsilon}{2}\right).\end{aligned}\quad (9.29)$$

Dalla disuguaglianza di Chebyshev (8.9), in virtù delle (9.20) e (9.28), per $n = 1, 2, \dots$ scaturisce poi:

$$P\left(|\bar{Y}_n - \mu_n| \geq \frac{\varepsilon}{2}\right) \leq \frac{4 \text{Var}(\bar{Y}_n)}{\varepsilon^2} = \frac{4 \sigma_n^2}{n \varepsilon^2} \leq \frac{4 \alpha \gamma}{\varepsilon^2}.$$

Quindi, dalla (9.29) per $n \geq n_0$ si ottiene:

$$P(|\bar{Y}_n - \mu| \geq \varepsilon) \leq \frac{\delta}{2}, \quad (9.30)$$

dove si è posto $\delta = 8 \alpha \gamma / \varepsilon^2$. Osserviamo ora che

$$\begin{aligned}P(|\bar{X}_n - \mu| \geq \varepsilon) &= P(|\bar{X}_n - \mu| \geq \varepsilon, \bar{Z}_n = 0) + P(|\bar{X}_n - \mu| \geq \varepsilon, \bar{Z}_n \neq 0) \\ &\leq P(|\bar{Y}_n - \mu| \geq \varepsilon) + P(\bar{Z}_n \neq 0).\end{aligned}\quad (9.31)$$

Una limitazione per il primo addendo nell'ultimo termine della (9.31) è data dalle (9.30). Determiniamo ora una limitazione per l'ultimo addendo. Osserviamo a tal fine che risulta:

$$\begin{aligned}P(\bar{Z}_n \neq 0) &= P\left(\sum_{k=1}^n Z_k \neq 0\right) \\ &\leq \sum_{k=1}^n P(Z_k \neq 0) = \sum_{k=1}^n P(|X_k| \geq \alpha n) = n P(|X_1| \geq \alpha n).\end{aligned}\quad (9.32)$$

Notiamo ora che ponendo

$$A = \{\omega \in \Omega : |X_1(\omega)| \geq \alpha n\}, \quad I_A(\omega) = \begin{cases} 0, & \omega \in \bar{A} \\ 1, & \omega \in A, \end{cases}$$

risulta $|X_1| I_A \geq \alpha n I_A$, da cui segue:

$$E[|X_1| I_A] \geq \alpha n E(I_A) = \alpha n P(|X_1| \geq \alpha n),$$

ossia

$$P(|X_1| \geq \alpha n) \leq \frac{1}{\alpha n} E[|X_1| I_A]. \quad (9.33)$$

Facendo uso della (9.33) nella (9.32), si ha poi:

$$P(\bar{Z}_n \neq 0) \leq \frac{1}{\alpha} E[|X_1| I_A].$$

Per l'ipotesi di finitezza del valore medio, $E[|X_1| I_A]$ tende a zero per $n \rightarrow +\infty$, da cui segue $\lim_{n \rightarrow +\infty} P(\bar{Z}_n \neq 0) = 0$. Quindi, per ogni $\delta > 0$ esiste un reale n_1 tale che per ogni $n \geq n_1$ risulta:

$$P(\bar{Z}_n \neq 0) \leq \frac{\delta}{2}. \quad (9.34)$$

Utilizzando le (9.30) e (9.34) nella (9.31), possiamo pertanto concludere che per ogni $\varepsilon > 0$ esiste un reale $n_2 = \max(n_0, n_1)$ tale che $P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \delta$ per ogni $n \geq n_2$, da cui segue la (9.26). La dimostrazione è così completa. \square

In sintesi il teorema di Khintchin afferma che se X_1, X_2, \dots sono variabili aleatorie indipendenti e identicamente distribuite con valore medio finito, allora la media campionaria \bar{X}_n converge in probabilità al valore medio μ .

Nel seguente esempio è indicata un'applicazione della legge debole dei grandi numeri di Khintchin che trova anche applicazione nella stima di integrali definiti utilizzando metodi di simulazione di tipo Monte Carlo.

Esempio 9.11 Siano X_1, X_2, \dots, X_n e Y_1, Y_2, \dots, Y_n variabili aleatorie indipendenti distribuite uniformemente in $(0, 1)$. Si definiscano poi le seguenti variabili aleatorie:

$$Z_i = \begin{cases} 1 & \text{se } Y_i \leq g(X_i) \\ 0 & \text{se } Y_i > g(X_i) \end{cases} \quad (i = 1, 2, \dots, n),$$

dove $g: (0, 1) \rightarrow (0, 1)$ è una funzione integrabile. Posto

$$\bar{Z}_n = \frac{1}{n} \sum_{i=1}^n Z_i \quad (n = 1, 2, \dots),$$

analizziamo la convergenza in probabilità della successione $\{\bar{Z}_n; n = 1, 2, \dots\}$. Notiamo che le variabili aleatorie Z_1, Z_2, \dots, Z_n hanno distribuzione di Bernoulli, così che risulta:

$$E(Z_i) = P(Z_i = 1) = P[Y_i \leq g(X_i)] = \int_{-\infty}^{+\infty} P\{Y_i \leq g(X_i) | X_i = x\} f_{X_i}(x) dx.$$

Dall'ipotesi che X_i e Y_i hanno distribuzione uniforme in $(0, 1)$, si ricava:

$$E(Z_i) = \int_0^1 P\{Y_i \leq g(x)\} dx = \int_0^1 g(x) dx \quad (i = 1, 2, \dots, n),$$

e quindi

$$E(\bar{Z}_n) = \frac{1}{n} E\left(\sum_{i=1}^n Z_i\right) = E(Z_i) = \int_0^1 g(x) dx \quad (n = 1, 2, \dots).$$

Poiché Z_1, Z_2, \dots, Z_n sono indipendenti e identicamente distribuite, in virtù del Teorema 9.11 si trae:

$$\bar{Z}_n \xrightarrow{P} \int_0^1 g(x) dx. \quad (9.35)$$

Tale risultato è alla base di un metodo di simulazione di tipo Monte Carlo per determinare una stima dell'integrale $\int_0^1 g(x) dx$ quando questo è difficile da calcolare per via analitica. A tal fine si generano n punti, le coordinate (X_i, Y_i) di ciascuno dei quali sono coppie indipendenti di variabili uniformi in $(0, 1) \times (0, 1)$. Risulta quindi che $X_i \sim \mathcal{U}(0, 1)$ e $Y_i \sim \mathcal{U}(0, 1)$ per $i = 1, 2, \dots, n$. Associamo a Z_i il valore 0 se il punto cade al di sopra del grafico della funzione $g(x)$ ed il valore 1 se il punto si presenta al di sotto o sul grafico stesso. La media campionaria \bar{Z}_n descrive quindi la frequenza relativa del numero di punti che cadono al di sotto del grafico o sul grafico. In virtù della (9.35), \bar{Z}_n fornisce un'approssimazione dell'integrale considerato che ci si attende migliori al crescere del numero di punti generati. \diamond

Abbiamo visto che nelle sue diverse formulazioni la legge debole dei grandi numeri coinvolge la convergenza in probabilità. È possibile esprimere analoghi risultati mediante il criterio più forte della convergenza quasi certa; in tal caso si parla di *legge forte dei grandi numeri*. Nel seguito di questo paragrafo ci limitiamo ad enunciarne due varianti omettendone peraltro le dimostrazioni.

Teorema 9.12 (Prima legge forte dei grandi numeri) *Sia X_1, X_2, \dots una successione di variabili aleatorie indipendenti aventi valori medi finiti e varianze $\sigma_n^2 = \text{Var}(X_n)$ finite, e sia $Y_n = X_1 + X_2 + \dots + X_n$. Se risulta*

$$\lim_{n \rightarrow +\infty} \frac{\sigma_1^2 + \dots + \sigma_n^2}{n^2} < +\infty,$$

allora si ha:

$$\frac{Y_n - E(Y_n)}{n} \xrightarrow{q.c.} 0. \quad (9.36)$$

Una conseguenza immediata del Teorema 9.12 è il seguente

Corollario 9.2 *Sia X_1, X_2, \dots una successione di variabili aleatorie indipendenti aventi valori medi finiti e varianze $\sigma_1^2, \sigma_2^2, \dots$ uniformemente limitate. Posto $Y_n = X_1 + X_2 + \dots + X_n$, sussiste la (9.36).*

Se X_1, X_2, \dots è una successione di variabili aleatorie indipendenti e identicamente distribuite con valore medio μ finito è possibile ottenere un risultato più forte. Ciò è espresso dal seguente

Teorema 9.13 (Seconda legge forte dei grandi numeri di Kolmogorov) *Sia X_1, X_2, \dots una successione di variabili aleatorie indipendenti e identicamente distribuite con valore medio μ . La finitezza di μ è condizione necessaria e sufficiente affinché risulti:*

$$\overline{X}_n \xrightarrow{q.c.} \mu. \quad (9.37)$$

Nel caso di una successione di variabili aleatorie indipendenti e identicamente distribuite con valore medio μ , la (9.37) mostra dunque che al crescere di n la media campionaria \overline{X}_n tende a μ con probabilità unitaria.

9.5 Convergenze di distribuzioni binomiali

Il teorema di convergenza e le leggi dei grandi numeri rivestono notevole interesse nel caso particolare in cui le variabili aleatorie X_1, X_2, \dots sono indipendenti con $X_i \sim \mathcal{B}(1, p)$ per $i = 1, 2, \dots$. Come mostrato nella Proposizione 4.1, $Y_n = X_1 + X_2 + \dots + X_n$ ha in tal caso distribuzione binomiale di parametri n e p , ossia $Y_n \sim \mathcal{B}(n, p)$. Pertanto, per $n = 0, 1, \dots$ si ha:

$$p_{Y_n}(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x = 0, 1, \dots, n \\ 0, & \text{altrimenti.} \end{cases} \quad (9.38)$$

Il calcolo delle probabilità (9.38) diviene rapidamente oneroso al crescere di n ; a maggior ragione tale risulta il calcolo diretto della funzione di distribuzione

$$F_{Y_n}(x) = \begin{cases} 0, & x < 0 \\ \sum_{i=0}^k \binom{n}{i} p^i (1-p)^{n-i}, & k \leq x < k+1 \quad (k = 0, 1, \dots, n-1), \\ 1, & x \geq n \end{cases} \quad (9.39)$$

la cui determinazione quantitativa al variare dei parametri coinvolti conduce a difficoltà numeriche non facilmente superabili. È quindi utile ricercare delle formule approssimate in grado di rendere agevole tale calcolo e, al contempo, accettabile l'errore derivante dall'approssimazione. Prenderemo in primo luogo in considerazione il *teorema di De Moivre-Laplace* e successivamente il *teorema di Poisson*.

Teorema 9.14 (Teorema di De Moivre-Laplace) *Sia X_1, X_2, \dots una successione di variabili aleatorie indipendenti distribuite alla Bernoulli con parametro p ($0 < p < 1$), e sia $Y_n = X_1 + X_2 + \dots + X_n$. Allora per ogni $x \in \mathbb{R}$ risulta:*

$$\lim_{n \rightarrow +\infty} P\left(\frac{Y_n - np}{\sqrt{np(1-p)}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy,$$

ossia:

$$Z_n = \frac{Y_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} Z,$$

con Z variabile aleatoria normale standard.

Dimostrazione Poiché risulta $E(X_i) = p$ e $\text{Var}(X_i) = p(1-p)$ per $i = 1, 2, \dots$, dal Teorema 9.8 segue immediatamente la tesi. \square

Il Teorema 9.14 mostra che sottraendo a Y_n la sua media np e dividendo la differenza per la deviazione standard $\sqrt{np(1-p)}$, si ottiene una variabile aleatoria standardizzata Z_n la cui funzione di distribuzione $P(Z_n \leq x)$ è per n grande approssimativamente normale standard:

$$P\left(\frac{Y_n - np}{\sqrt{np(1-p)}} \leq x\right) \simeq \Phi(x) \quad (x \in \mathbb{R}), \quad (9.40)$$

$$P\left(\alpha < \frac{Y_n - np}{\sqrt{np(1-p)}} \leq \beta\right) \simeq \Phi(\beta) - \Phi(\alpha) \quad (\alpha, \beta \in \mathbb{R}, \alpha < \beta) \quad (9.41)$$

La bontà delle approssimazioni (9.40) e (9.41) dipendono da n e da p e migliorano al tendere di p a $1/2$. In generale si suole assumere che l'approssimazione sia soddisfacente per $n > 10$ e per $5/n < p < 1 - 5/n$.

Dalle (9.40) e (9.41), per n grande si ricavano le seguenti approssimazioni:

$$P(Y_n \leq w) \simeq \Phi\left(\frac{w - np}{\sqrt{np(1-p)}}\right) \quad (w \in \mathbb{R}), \quad (9.42)$$

$$P(a < Y_n \leq b) \simeq \Phi\left(\frac{b - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - np}{\sqrt{np(1-p)}}\right) \quad (a, b \in \mathbb{R}, a < b). \quad (9.43)$$

Esempio 9.12 Un esperimento consiste nel ripetere 60 volte l'estrazione (con reinserimento) di una biglia da un'urna contenente 50 biglie, 15 delle quali sono bianche. Sia Y_{60} la variabile aleatoria che descrive il numero totale di biglie bianche estratte. Poiché essa ha distribuzione binomiale di parametri $n = 60$ e $p = 15/50 = 0.3$, la probabilità che il numero di biglie bianche estratte sia compreso tra 11 e 30 è:

$$P(10 < Y_{60} \leq 30) = \sum_{k=11}^{30} \binom{60}{k} (0.3)^k (0.7)^{n-k} = 0.985758.$$

Un'approssimazione di tale probabilità, in virtù della (9.43), è poi la seguente:

$$P(10 < Y_{60} \leq 30) \simeq \Phi\left(\frac{30 - 60 \cdot 0.3}{\sqrt{60 \cdot 0.3 \cdot 0.7}}\right) - \Phi\left(\frac{10 - 60 \cdot 0.3}{\sqrt{60 \cdot 0.3 \cdot 0.7}}\right) = 0.987532.$$

Si noti come il valore approssimato appena ricavato non si discosta molto dal valore esatto differendo da questo soltanto dello 0.18%. \diamond

Nel teorema di De Moivre-Laplace si è supposto che la successione Y_1, Y_2, \dots sia costituita da variabili aleatorie a distribuzione binomiale con lo stesso parametro p , ossia che $Y_n \sim \mathcal{B}(n, p)$ per $n = 0, 1, \dots$. Vogliamo ora modificare questa ipotesi supponendo che la successione Y_1, Y_2, \dots sia costituita da variabili aleatorie tali che $Y_n \sim \mathcal{B}(n, p_n)$ per $n = 0, 1, \dots$, dove la probabilità p_n è tale che $\lim_{n \rightarrow +\infty} n p_n = \lambda$, con $\lambda > 0$.

Teorema 9.15 *Sia Y_1, Y_2, \dots una successione di variabili aleatorie tali che per $n = 0, 1, \dots$ risulti $Y_n \sim \mathcal{B}(n, p_n)$. Se p_n , al divergere di n , tende a zero in modo tale che $n p_n \rightarrow \lambda$, con $\lambda > 0$, allora la successione Y_1, Y_2, \dots converge in distribuzione alla variabile aleatoria di Poisson di parametro λ :*

$$Y_n \xrightarrow{d} Y,$$

con $Y \sim \mathcal{P}(\lambda)$.

Dimostrazione Essendo $Y_n \sim \mathcal{B}(n, p_n)$, come mostrato nel Paragrafo 5.8.3 la sua funzione generatrice dei momenti è:

$$M_{Y_n}(s) = [1 - p_n + p_n e^s]^n = \left[1 + \frac{n p_n (e^s - 1)}{n}\right]^n.$$

Poiché per ipotesi $n p_n \rightarrow \lambda$, al limite per $n \rightarrow +\infty$ si ricava:

$$\lim_{\substack{n \rightarrow +\infty, \\ n p_n \rightarrow \lambda}} M_{Y_n}(s) = \exp\{\lambda(e^s - 1)\},$$

che si riconosce essere la funzione generatrice dei momenti della variabile $Y \sim \mathcal{P}(\lambda)$ (v. Paragrafo 5.8.3). Dal Teorema 9.5 segue quindi $Y_n \xrightarrow{d} Y$. \square

Si noti che abbiamo già incontrato in precedenza la convergenza della distribuzione binomiale a quella di Poisson. Invero, nella Proposizione 4.7 abbiamo mostrato che se Y_1, Y_2, \dots è una successione di variabili aleatorie tali che risulti $Y_n \sim \mathcal{B}(n, p_n)$ per $n = 0, 1, \dots$, e se p_n al divergere di n tende a zero in modo tale che $n p_n \rightarrow \lambda > 0$, allora

$$\lim_{\substack{n \rightarrow +\infty, \\ n p_n \rightarrow \lambda}} p_{Y_n}(k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (k = 0, 1, \dots).$$

Dal Teorema 9.15, per n grande e posto $\lambda = n p$, si ottiene la seguente approssimazione:

$$P(Y_n \leq x) \simeq P(Y \leq x),$$

dove $Y_n \sim \mathcal{B}(n, p)$ e $Y \sim \mathcal{P}(n p)$. È opportuno menzionare che la distribuzione di Poisson costituisce una buona approssimazione della distribuzione binomiale allorché $n \geq 20$ e $p \leq 0.05$; se poi è $n \geq 100$ e $n p \leq 10$ l'approssimazione diviene eccellente.

Le Tabelle 9.1 e 9.2 riportano la funzione di distribuzione binomiale $P(Y_n \leq x)$, l'approssimazione (9.42) e la distribuzione di Poisson $P(Y \leq x)$ per alcune scelte di x . Nella Tabella 9.1 si è scelto $n = 20$, $p = 0.05$ e $\lambda = n p = 1$. Si noti che essendo p piccolo e n grande, l'approssimazione di Poisson fornisce risultati migliori rispetto a quelli forniti dalla distribuzione normale.

Tabella 9.1– Confronto tra la funzione di distribuzione binomiale, l'approssimazione fornita dalla (9.42) e la distribuzione di Poisson per $x = 0, 1, \dots, 7$ e per $n = 20$, $p = 0.05$ e $\lambda = np = 1$.

x	Binomiale $P(Y_n \leq x)$ $p = 0.05, n = 20$	Normale $\Phi\left(\frac{x - np}{\sqrt{np(1-p)}}\right)$ $p = 0.05, n = 20$	Poisson $P(Y \leq x)$ $\lambda = 1$
0	0.3585	0.1515	0.3679
1	0.7359	0.5000	0.7358
2	0.9246	0.8485	0.9197
3	0.9842	0.9798	0.9810
4	0.9975	0.9990	0.9963
5	0.9997	1.0000	0.9994
6	1.0000	1.0000	0.9999
7	1.0000	1.0000	1.0000

Nella Tabella 9.2 si è invece scelto $n = 20$, $p = 0.5$ e $\lambda = np = 10$; a differenza del precedente caso l'approssimazione normale fornisce risultati migliori rispetto all'approssimazione di Poisson.

Osserviamo che se X_1, X_2, \dots sono indipendenti, con $X_i \sim \mathcal{B}(1, p)$ per $i = 1, 2, \dots$, allora

$$F_n = \frac{X_1 + X_2 + \dots + X_n}{n} \quad (n = 1, 2, \dots) \quad (9.44)$$

rappresenta la frequenza relativa dei successi in n prove indipendenti di Bernoulli. Il ben noto *Teorema di Jacques Bernoulli*, pubblicato nella sua “Ars conjectandi” nel 1713, fornisce una relazione tra la frequenza (9.44) e la probabilità p di successo in ogni singola prova, come specificato dal teorema che segue.

Teorema 9.16 (Teorema di Bernoulli) *Sia X_1, X_2, \dots una successione di variabili aleatorie indipendenti aventi distribuzione di Bernoulli di parametro p , e sia $F_n = (X_1 + X_2 + \dots + X_n)/n$. Allora per ogni $\varepsilon > 0$ si ha:*

$$\lim_{n \rightarrow +\infty} P\{|F_n - p| < \varepsilon\} = 1, \quad (9.45)$$

ossia $F_n \xrightarrow{P} p$.

Dimostrazione Essendo le variabili aleatorie X_1, X_2, \dots indipendenti con distribuzione di Bernoulli di parametro p , risulta $E(X_i) = p$ per $i = 1, 2, \dots$. In virtù del Teorema 9.11 per ogni $\varepsilon > 0$ segue:

$$\lim_{n \rightarrow +\infty} P\{|F_n - p| < \varepsilon\} = \lim_{n \rightarrow +\infty} P\left\{\left|\frac{X_1 + \dots + X_n}{n} - p\right| < \varepsilon\right\} = 1.$$

□

Tabella 9.2– Confronto tra la funzione di distribuzione binomiale, l'approssimazione (9.42) e la distribuzione di Poisson per $x = 10, 11, \dots, 18$ e per $n = 20$, $p = 0.5$ e $\lambda = np = 10$.

x	Binomiale $P(Y_n \leq x)$ $p = 0.5, n = 20$	Normale $\Phi\left(\frac{x - np}{\sqrt{np(1-p)}}\right)$ $p = 0.5, n = 20$	Poisson $P(Y \leq x)$ $\lambda = 10$
10	0.5881	0.5000	0.4580
11	0.7483	0.6736	0.5831
12	0.8684	0.8133	0.6968
13	0.9423	0.9099	0.7916
14	0.9793	0.9633	0.8645
15	0.9941	0.9875	0.9166
16	0.9987	0.9963	0.9513
17	0.9998	0.9991	0.9730
18	1.0000	1.0000	0.9858

Il Teorema di Bernoulli mostra che la frequenza relativa F_n dei successi quando il numero di prove tende all'infinito, converge in probabilità al valore p della probabilità di successo in ogni singola prova. Si noti che sussiste un'analogia tra il risultato del Teorema di Bernoulli e la legge empirica del caso, la quale afferma che in una successione di prove ripetute nelle stesse condizioni, la frequenza relativa di un evento si avvicina alla probabilità dell'evento stesso, con l'approssimazione che tende a migliorare all'aumentare del numero delle prove. Occorre però sottolineare che la legge empirica del caso è un postulato di natura sperimentale che trae origine da fenomeni reali osservati, mentre il risultato espresso dal Teorema 9.16 è una conseguenza del modello teorico ricavato a partire dall'assiomatizzazione della teoria della probabilità.

Teorema 9.17 Sia X_1, X_2, \dots una successione di variabili aleatorie indipendenti di distribuzione di Bernoulli di parametro p e sia $Y_n = X_1 + X_2 + \dots + X_n$. Allora per ogni $\varepsilon > 0$ si ha:

$$\lim_{n \rightarrow +\infty} P\{|Y_n - np| < \varepsilon\} = 0. \quad (9.46)$$

Dimostrazione Ragionando rapidamente, notiamo che facendo uso della (9.43), per ogni $\varepsilon > 0$ e per n grande si ricava:

$$\begin{aligned} P\{|Y_n - np| < \varepsilon\} &= P(np - \varepsilon < Y_n < np + \varepsilon) = P(np - \varepsilon < Y_n \leq \lceil np + \varepsilon \rceil - 1) \\ &\simeq \Phi\left(\frac{\lceil np + \varepsilon \rceil - 1 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{(np - \varepsilon) - np}{\sqrt{np(1-p)}}\right), \end{aligned}$$

dove $\lceil x \rceil$ denota il minimo intero maggiore o uguale a x . Per le proprietà della funzione $\Phi(x)$, quando $n \rightarrow +\infty$ si ottiene pertanto la (9.46). \square

Il Teorema 9.17 mostra che, per ogni arbitrario $\varepsilon > 0$, al crescere del numero delle prove tende a 0 la probabilità che il numero dei successi Y_n e il numero medio di successi np differiscano per meno di ε . Ad esempio, nell'esperimento consistente nel lanciare un numero elevato di volte una moneta non truccata, il numero totale di uscite di Testa non è necessariamente prossimo a $n/2$. Questo risultato è tanto più evidente quando si consideri un gioco d'azzardo consistente in una successione di partite tali che in ogni singola partita si lancia una moneta non truccata e si vince, ad esempio, 1 euro se esce Testa mentre si perde 1 euro se esce Croce. In base del Teorema 9.17, al crescere del numero delle partite ci si deve attendere una vincita o una perdita non necessariamente prossime a zero.

Nel Teorema di Bernoulli si è supposto implicitamente che le prove siano effettuate nelle medesime condizioni. Il seguente teorema considera il caso in cui le prove non sono effettuate nelle medesime condizioni e mostra che esiste una relazione tra la frequenza (9.44) e la media aritmetica delle probabilità di successo nelle singole prove.

Teorema 9.18 (Teorema di Poisson) *Sia X_1, X_2, \dots una successione di variabili aleatorie indipendenti tali che per $i = 1, 2, \dots$ risulti $X_i \sim \mathcal{B}(1, p_i)$ e sia $F_n = (X_1 + X_2 + \dots + X_n)/n$. Allora per ogni $\varepsilon > 0$ risulta:*

$$\lim_{n \rightarrow +\infty} P\left\{\left|F_n - \frac{p_1 + p_2 + \dots + p_n}{n}\right| < \varepsilon\right\} = 1, \quad (9.47)$$

ossia $F_n \xrightarrow{P} (p_1 + p_2 + \dots + p_n)/n$.

Dimostrazione Essendo X_1, X_2, \dots indipendenti e tali che $X_i \sim \mathcal{B}(1, p_i)$ per $i = 1, 2, \dots$, si ha:

$$\begin{aligned} E(F_n) &= E\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{p_1 + p_2 + \dots + p_n}{n} \\ \text{Var}(F_n) &= \text{Var}\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{1}{n^2} \sum_{i=1}^n p_i(1 - p_i). \end{aligned}$$

Poiché $p_i(1 - p_i) \leq 1/4$, segue che

$$\text{Var}(F_n) \leq \frac{1}{n^2} \frac{n}{4} = \frac{1}{4n} \leq \frac{1}{4}.$$

Essendo soddisfatte le ipotesi del Teorema 9.10 di Chebyshev, dalla (9.25) segue immediatamente la (9.47). \square

L'interesse di questo teorema risiede nella circostanza che, a differenza del sopra riportato teorema di Bernoulli, esso non richiede che ci si riferisca a prove ripetute con probabilità costante di successo in ciascuna di esse, essendo previsto che ciascuna prova possa essere caratterizzata da una propria specifica probabilità di successo.

Appendice A

Variabili aleatorie discrete

Qui sono riportate tabelle riassuntive relative a variabili aleatorie discrete. In particolare, nella Tabella A.1 sono elencate le funzioni di probabilità, nella Tabella A.2 sono riportati i valori medi, le varianze ed i coefficienti di variazione e nella Tabella A.3 sono indicate le funzioni generatrici di probabilità e dei momenti di tali variabili aleatorie. Infine, nella Tabella A.4 si considerano le funzioni di probabilità multinomiale e ipergeometrica multivariata.

Tabella A.1 – Funzioni di probabilità di variabili aleatorie discrete.

Distribuzione	Notazione	Funzione di probabilità
Uniforme discreta	$X \sim \mathcal{U}_d(n)$	$p_X(x) = \begin{cases} \frac{1}{n}, & x = 1, 2, \dots, n \\ 0, & \text{altrimenti} \end{cases} \quad (n = 1, 2, \dots)$
Bernoulli	$X \sim \mathcal{B}(1, p)$	$p_X(x) = \begin{cases} 1 - p, & x = 0 \\ p, & x = 1 \\ 0, & \text{altrimenti} \end{cases} \quad (0 < p < 1)$
Binomiale	$X \sim \mathcal{B}(n, p)$	$p_X(x) = \begin{cases} \binom{n}{x} p^x (1 - p)^{n-x}, & x = 0, 1, \dots, n \\ 0, & \text{altrimenti} \end{cases} \quad (n = 1, 2, \dots; 0 < p < 1)$
Ipergeometrica	$X \sim \mathcal{I}(n, m, N - m)$	$p_X(x) = \begin{cases} \frac{\binom{m}{x} \binom{N-m}{n-x}}{\binom{N}{n}}, & x \geq \max\{0, n + m - N\} \\ & x \leq \min\{n, m\} \\ 0, & \text{altrimenti} \end{cases} \quad (N = 1, 2, \dots; m = 0, 1, \dots, N; n = 0, 1, \dots, N)$
Geometrica	$X \sim \mathcal{BN}(1, p)$	$p_X(x) = \begin{cases} p(1 - p)^{x-1}, & x = 1, 2, \dots \\ 0, & \text{altrimenti} \end{cases} \quad (0 < p < 1)$
Binomiale negativa	$X \sim \mathcal{BN}(n, p)$	$p_X(x) = \begin{cases} \binom{x-1}{n-1} p^n (1 - p)^{x-n}, & x = n, n + 1, \dots \\ 0, & \text{altrimenti} \end{cases} \quad (n = 1, 2, \dots; 0 < p < 1)$
Poisson	$X \sim \mathcal{P}(\lambda)$	$p_X(x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda}, & x = 0, 1, \dots \\ 0, & \text{altrimenti} \end{cases} \quad (\lambda > 0)$

Tabella A.2 – Valori medi, varianze e coefficienti di variazione di variabili discrete.

Distribuzione	Valore medio $E(X)$	Varianza $\text{Var}(X)$	Coefficiente di variazione C_X
Uniforme discreta	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$\sqrt{\frac{n-1}{3(n+1)}}$
Bernoulli	p	$p(1-p)$	$\sqrt{\frac{1-p}{p}}$
Binomiale	np	$np(1-p)$	$\sqrt{\frac{1-p}{np}}$
Ipergeometrica	$\frac{nm}{N}$	$\frac{nm(N-m)(N-n)}{N^2(N-1)}$	$\sqrt{\frac{(N-m)(N-n)}{nm(N-1)}}$
Geometrica	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\sqrt{1-p}$
Binomiale negativa	$\frac{n}{p}$	$\frac{n(1-p)}{p^2}$	$\sqrt{\frac{1-p}{n}}$
Poisson	λ	λ	$\frac{1}{\sqrt{\lambda}}$

Tabella A.3 – Funzioni generatrici di variabili discrete.

Distribuzione	Funzione generatrice di probabilità $G_X(z)$	Funzione generatrice dei momenti $M_X(s)$
Uniforme discreta	$\frac{z(1-z^n)}{n(1-z)} \quad (z \in \mathbb{R})$	$\frac{e^s(1-e^{sn})}{n(1-e^s)} \quad (s \in \mathbb{R})$
Bernoulli	$1-p+pz \quad (z \in \mathbb{R})$	$1-p+pe^s \quad (s \in \mathbb{R})$
Binomiale	$(1-p+pz)^n \quad (z \in \mathbb{R})$	$(1-p+pe^s)^n \quad (s \in \mathbb{R})$
Ipergeometrica	Polinomio in z	Polinomio in e^s
Geometrica	$\frac{pz}{1-(1-p)z} \quad (z < (1-p)^{-1})$	$\frac{pe^s}{1-(1-p)e^s} \quad (s < -\ln(1-p))$
Binomiale negativa	$\left(\frac{pz}{1-(1-p)z}\right)^n \quad (z < (1-p)^{-1})$	$\left(\frac{pe^s}{1-(1-p)e^s}\right)^n \quad (s < -\ln(1-p))$
Poisson	$e^{\lambda(z-1)} \quad (z \in \mathbb{R})$	$e^{\lambda(e^s-1)} \quad (s \in \mathbb{R})$

Tabella A.4 – Funzioni di probabilità di alcuni vettori aleatori discreti.

Distribuzione	Funzione di probabilità di $\mathbf{X} = (X_1, X_2, \dots, X_k)$
Multinomiale	$p_{\mathbf{X}}(\mathbf{x}) = \begin{cases} \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}, & x_i = 0, 1, \dots, n \quad (i = 1, 2, \dots, k), \\ & x_1 + x_2 + \dots + x_k = n \\ 0, & \text{altrimenti,} \end{cases}$ <p style="text-align: center;">($0 < p_i < 1$ per $i = 1, 2, \dots, k$; $p_1 + p_2 + \dots + p_k = 1$)</p> <p style="text-align: center;">.....</p> $E(X_i) = n p_i, \quad \text{Var}(X_i) = n p_i (1 - p_i) \quad (i = 1, 2, \dots, k)$ $\text{Cov}(X_i, X_j) = -n p_i p_j \quad (i, j = 1, 2, \dots, k; i \neq j)$
Ipergeometrica multivariata	$p_{\mathbf{X}}(\mathbf{x}) = \begin{cases} \frac{\binom{N_1}{x_1} \binom{N_2}{x_2} \dots \binom{N_k}{x_k}}{\binom{N}{n}}, & 0 \leq x_i \leq N_i \quad (i = 1, 2, \dots, k), \\ & x_1 + x_2 + \dots + x_k = n \\ 0, & \text{altrimenti,} \end{cases}$ <p style="text-align: center;">(n, N_1, N_2, \dots, N_k interi positivi, $N = N_1 + N_2 + \dots + N_k$)</p>

Appendice B

Variabili aleatorie continue

Si riportano qui tabelle riassuntive relative a variabili aleatorie assolutamente continue. In particolare, nelle Tabelle B.1 e B.2 sono elencate le densità di probabilità, nella Tabella B.3 sono riportati i valori medi, le varianze ed i coefficienti di variazione e nella Tabella B.4 sono indicate le funzioni generatrici dei momenti di tali variabili aleatorie. Infine, la Tabella B.5 si riferisce alla densità di probabilità normale bivariata.

Tabella B.1 – Densità di probabilità di variabili aleatorie continue.

Distribuzione	Notazione	Funzione densità di probabilità
Uniforme	$X \sim \mathcal{U}(a, b)$	$f_X(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{altrimenti} \end{cases} \quad (a, b \in \mathbb{R})$
Esponenziale	$X \sim \mathcal{E}(1, \lambda)$	$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{altrimenti} \end{cases} \quad (\lambda > 0)$
Erlang	$X \sim \mathcal{E}(n, \lambda)$	$f_X(x) = \begin{cases} \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x}, & x > 0 \\ 0, & \text{altrimenti} \end{cases}$ <p style="text-align: center;">$(\lambda > 0, n = 1, 2, \dots)$</p>
Gamma	$X \sim \mathcal{G}(\nu, \lambda)$	$f_X(x) = \begin{cases} \frac{\lambda^\nu x^{\nu-1}}{\Gamma(\nu)} e^{-\lambda x}, & x > 0 \\ 0, & \text{altrimenti} \end{cases}$ <p style="text-align: center;">$(\lambda > 0, \nu > 0)$</p>
Iperesponenziale		$f_X(x) = \begin{cases} \sum_{i=1}^n a_i \lambda_i e^{-\lambda_i x}, & x > 0 \\ 0, & \text{altrimenti} \end{cases}$ <p style="text-align: center;">$(\lambda_i > 0, a_i \geq 0, a_1 + a_2 + \dots + a_n = 1)$</p>
Weibull		$f_X(x) = \begin{cases} \alpha \lambda x^{\alpha-1} \exp\{-\lambda x^\alpha\}, & x > 0 \\ 0, & \text{altrimenti} \end{cases}$ <p style="text-align: center;">$(\alpha > 0, \lambda > 0)$</p>

Tabella B.2 – Densità di probabilità di variabili aleatorie continue.

Distribuzione	Notazione	Funzione densità di probabilità
Normale	$X \sim \mathcal{N}(\mu, \sigma)$	$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad x \in \mathbb{R}$ $(\mu \in \mathbb{R}, \sigma > 0)$
Laplace		$f_X(x) = \frac{1}{2\beta} \exp\left\{-\frac{ x-\alpha }{\beta}\right\}, \quad x \in \mathbb{R}$ $(\alpha \in \mathbb{R}, \beta > 0)$
Chi-quadrato	$X \sim \chi^2(n)$	$f_X(x) = \begin{cases} \frac{1}{2^{n/2} \Gamma\left(\frac{n}{2}\right)} e^{-x/2} x^{n/2-1}, & x > 0 \\ 0, & \text{altrimenti} \end{cases}$ $(n = 1, 2, \dots)$
Cauchy	$X \sim \mathcal{T}(1)$	$f_X(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}$
Student	$X \sim \mathcal{T}(n)$	$f_X(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}, \quad x \in \mathbb{R}$ $(n = 1, 2, \dots)$
Beta	$X \sim \mathcal{Be}(\alpha, \beta)$	$f_X(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, & 0 < x < 1 \\ 0, & \text{altrimenti} \end{cases}$ $(\alpha > 0, \beta > 0)$

Tabella B.3 – Valori medi, varianze e coefficienti di variazione di variabili continue.

Distribuzione	Valore medio $E(X)$	Varianza $\text{Var}(X)$	Coefficiente di variazione C_X
Uniforme	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{b-a}{\sqrt{3}(a+b)}$
Esponenziale	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	1
Erlang	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$	$\frac{1}{\sqrt{n}}$
Gamma	$\frac{\nu}{\lambda}$	$\frac{\nu}{\lambda^2}$	$\frac{1}{\sqrt{\nu}}$
Iperesponenziale	$\sum_{j=1}^n \frac{a_j}{\lambda_j}$	$2 \sum_{j=1}^n \frac{a_j}{\lambda_j^2} - \left[\sum_{j=1}^n \frac{a_j}{\lambda_j} \right]^2$	
Weibull	$\lambda^{-1/\alpha} \Gamma\left(1 + \frac{1}{\alpha}\right)$	$\lambda^{-2/\alpha} \left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right) \right]$	
Normale	μ	σ^2	$\frac{\sigma}{\mu}$
Laplace	α	$2\beta^2$	$\frac{\beta\sqrt{2}}{\alpha}$
Chi-quadrato	n	$2n$	$\sqrt{\frac{2}{n}}$
Cauchy	non esiste	non esiste	
Student	0	$+\infty$ se $n = 2$ $\frac{n}{n-2}$ se $n = 3, 4, \dots$	
Beta	$\frac{\alpha}{\alpha + \beta}$	$\frac{\alpha\beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}$	$\sqrt{\frac{\beta(\alpha + \beta)}{\alpha(\alpha + \beta + 1)}}$

Tabella B.4 – Funzioni generatrici e momenti di variabili continue.

Distribuzione	Funzione generatrice dei momenti $M_X(s)$	Momenti $\mu_k = E(X^k)$ $\bar{\mu}_k = E[(X - \mu_1)^k]$
Uniforme	$\frac{e^{sb} - e^{sa}}{s(b-a)}$	$\mu_k = \frac{b^{k+1} - a^{k+1}}{(k+1)(b-a)}$
Esponenziale	$\frac{\lambda}{\lambda - s} \quad (s < \lambda)$	$\mu_k = \frac{k!}{\lambda^k}$
Erlang	$\left(\frac{\lambda}{\lambda - s}\right)^n \quad (s < \lambda)$	$\mu_k = \frac{n(n+1) \cdots (n+k-1)}{\lambda^k}$
Gamma	$\left(\frac{\lambda}{\lambda - s}\right)^\nu \quad (s < \lambda)$	$\mu_k = \frac{\Gamma(\nu + k)}{\lambda^k \Gamma(\nu)}$
Iperesponenziale	$\sum_{i=1}^n a_i \frac{\lambda_i}{\lambda_i - s}$ $(s < \min(\lambda_1, \lambda_2, \dots, \lambda_n))$	$\mu_k = k! \sum_{j=1}^n \frac{a_j}{\lambda_j^k}$
Weibull	$\sum_{k=0}^{+\infty} \frac{s^k}{k!} \lambda^{-k/\alpha} \Gamma\left(1 + \frac{k}{\alpha}\right)$	$\mu_k = \lambda^{-k/\alpha} \Gamma\left(1 + \frac{k}{\alpha}\right)$
Normale	$\exp\left\{\mu s + \frac{\sigma^2}{2} s^2\right\}$	$\bar{\mu}_k = \frac{k!}{(k/2)!} \frac{\sigma^2}{2^{k/2}} \quad \text{se } k \text{ è pari}$ $\bar{\mu}_k = 0 \quad \text{se } k \text{ è dispari}$
Laplace	$\frac{e^{\alpha s}}{1 - \beta^2 s^2} \quad (s < 1/\beta)$	$\bar{\mu}_k = k! \beta^k \quad \text{se } k \text{ è pari}$ $\bar{\mu}_k = 0 \quad \text{se } k \text{ è dispari}$
Chi-quadrato	$(1 - 2s)^{-n/2} \quad (s < 1/2)$	$\mu_k = \frac{2^k \Gamma(n/2 + k)}{\Gamma(n/2)}$
Cauchy	—	$\mu_k = +\infty \text{ se } k \text{ è pari}$ $\mu_k \text{ non esiste se } k \text{ è dispari}$
Student	—	$\mu_k = n^{k/2} \frac{B((k+1)/2, (n-k)/2)}{B(1/2, n/2)}$ $\text{se } n > k \text{ e } k \text{ pari}$ $\mu_k = +\infty \text{ se } n \leq k \text{ e } k \text{ pari}$ $\mu_k = 0 \text{ se } n > k \text{ e } k \text{ dispari}$ $\mu_k \text{ non esiste se } n \leq k \text{ e } k \text{ dispari}$
Beta	$\sum_{k=0}^{+\infty} \frac{s^k}{k!} \frac{B(\alpha + k, \beta)}{B(\alpha, \beta)}$	$\bar{\mu}_k = \frac{B(\alpha + k, \beta)}{B(\alpha, \beta)}$

Tabella B.5 – Vettore aleatorio bidimensionale normale.

Densità di probabilità normale bivariata

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\varrho^2}} \exp\left\{-\frac{1}{2(1-\varrho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\varrho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)\right]\right\}$$

$$(x,y \in \mathbb{R}; \mu_X \in \mathbb{R}, \mu_Y \in \mathbb{R}, \sigma_X > 0, \sigma_Y > 0, -1 < \varrho < 1)$$

Densità di probabilità marginali

$$f_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right\} \quad (x \in \mathbb{R})$$

$$f_Y(y) = \frac{1}{\sigma_Y\sqrt{2\pi}} \exp\left\{-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}\right\} \quad (y \in \mathbb{R})$$

Densità di probabilità condizionate

$$f_{X|Y}(x|y) = \frac{1}{\sigma_X\sqrt{2\pi(1-\varrho^2)}} \exp\left\{-\frac{1}{2\sigma_X^2(1-\varrho^2)}\left[x-\mu_X-\varrho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)\right]^2\right\}$$

$$f_{Y|X}(y|x) = \frac{1}{\sigma_Y\sqrt{2\pi(1-\varrho^2)}} \exp\left\{-\frac{1}{2\sigma_Y^2(1-\varrho^2)}\left[y-\mu_Y-\varrho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)\right]^2\right\}$$

Valori medi e varianze

$$E(X) = \mu_X, \quad \text{Var}(X) = \sigma_X^2, \quad E(Y) = \mu_Y, \quad \text{Var}(Y) = \sigma_Y^2$$

Medie e varianze condizionate

$$E(X|Y=y) = \mu_X + \varrho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y), \quad \text{Var}(X|Y=y) = \sigma_X^2(1-\varrho^2)$$

$$E(Y|X=x) = \mu_Y + \varrho\frac{\sigma_Y}{\sigma_X}(x-\mu_X), \quad \text{Var}(Y|X=x) = \sigma_Y^2(1-\varrho^2)$$

Covarianza e coefficiente di correlazione

$$\text{Cov}(X,Y) = \sigma_X\sigma_Y\varrho, \quad \varrho(X,Y) = \varrho$$

Appendice C

Distribuzione normale standard

Tabella C.1 – Valori della funzione di distribuzione normale standard $\Phi(x)$.

<i>x</i>	<i>Seconda cifra decimale</i>									
	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Nella Tabella C.1 a doppia entrata sono riportati i valori della funzione di distribuzione normale standard $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-z^2/2} dz$. I valori di x si ottengono sommando quelli della prima colonna con quelli della prima riga. Ad esempio, 0.68 si ottiene sommando 0.6 della prima colonna con 0.08 della prima riga, e in corrispondenza si ottiene $\Phi(0.68) = 0.7517$. Per valori negativi dell’argomento la funzione di distribuzione può essere calcolata tramite la relazione $\Phi(-x) = 1 - \Phi(x)$.

Appendice D

Fattoriali e coefficienti binomiali

$$\begin{aligned}
 0! &= 1, & n! &= n(n-1)! & (n = 1, 2, \dots) \\
 \binom{n}{k} &= \frac{n!}{k!(n-k)!} & (n = 0, 1, \dots; k = 0, 1, \dots, n)
 \end{aligned}$$

Tabella D.1 – Fattoriali.

<i>n</i>	<i>n</i> !	<i>n</i>	<i>n</i> !
0	1	8	40320
1	1	9	362880
2	2	10	3628800
3	6	11	39916800
4	24	12	479001600
5	120	13	6227020800
6	720	14	87178291200
7	5040	15	1307674368000

Tabella D.2 – Coefficienti Binomiali.

<i>n</i>	$\binom{n}{0}$	$\binom{n}{1}$	$\binom{n}{2}$	$\binom{n}{3}$	$\binom{n}{4}$	$\binom{n}{5}$	$\binom{n}{6}$	$\binom{n}{7}$	$\binom{n}{8}$	$\binom{n}{9}$	$\binom{n}{10}$
0	1										
1	1	1									
2	1	2	1								
3	1	3	3	1							
4	1	4	6	4	1						
5	1	5	10	10	5	1					
6	1	6	15	20	15	6	1				
7	1	7	21	35	35	21	7	1			
8	1	8	28	56	70	56	28	8	1		
9	1	9	36	84	126	126	84	36	9	1	
10	1	10	45	120	210	252	210	120	45	10	1
11	1	11	55	165	330	462	462	330	165	55	11
12	1	12	66	220	495	792	924	792	495	220	66
13	1	13	78	286	715	1287	1716	1716	1287	715	286
14	1	14	91	364	1001	2002	3003	3432	3003	2002	1001
15	1	15	105	455	1365	3003	5005	6435	6435	5005	3003
16	1	16	120	560	1820	4368	8008	11440	12870	11440	8008
17	1	17	136	680	2380	6188	12376	19448	24310	24310	19448
18	1	18	153	816	3060	8568	18564	31824	43758	48620	43758
19	1	19	171	969	3876	11628	27132	50388	75582	92378	92378
20	1	20	190	1140	4845	15504	38760	77520	125970	167960	184756

Appendice E

Formule notevoli

Questa appendice raccoglie alcune formule utilizzate nel testo riguardanti somme finite, serie, limiti notevoli e integrali definiti.

Tabella E.1 – Somme finite.

Binomio di Newton	$(a + b)^n = \sum_{r=0}^n \binom{n}{r} a^r b^{n-r} \quad (a, b \in \mathbb{R}; n = 0, 1, \dots)$ $\sum_{r=0}^n \binom{n}{r} = 2^n, \quad \sum_{r=0}^n \binom{n}{r} (-1)^r = 0 \quad (n = 0, 1, \dots)$
Formula di Vandermonde	$\sum_{r=\max\{0, n-m_2\}}^{\min\{n, m_1\}} \binom{m_1}{r} \binom{m_2}{n-r} = \binom{m_1+m_2}{n}$ $(n = 0, 1, \dots; m_1 = 0, 1, \dots; m_2 = 0, 1, \dots; 0 \leq n \leq m_1 + m_2)$ $\sum_{r=0}^n \binom{n}{r}^2 = \binom{2n}{n} \quad (n = 0, 1, \dots)$
Formula multinomiale	$\sum_{\substack{n_1=1, \dots, n_k=1 \\ n_1+\dots+n_k=n}}^n \frac{n!}{n_1! n_2! \dots n_k!} a_1^{n_1} a_2^{n_2} \dots a_k^{n_k} = (a_1 + a_2 + \dots + a_k)^n$ $(a_1, a_2, \dots, a_k \in \mathbb{R}; n = 1, 2, \dots)$
Formula generalizzata di Vandermonde	$\sum_{\substack{n_1, \dots, n_k \\ n_1+\dots+n_k=n}} \binom{N_1}{n_1} \binom{N_2}{n_2} \dots \binom{N_k}{n_k} = \binom{N}{n}$ $(n, N_1, N_2, \dots, N_k \text{ interi positivi}, N = N_1 + N_2 + \dots + N_k)$
Somme finite di potenze di interi	$\sum_{r=1}^n r = \frac{n(n+1)}{2}, \quad \sum_{r=1}^n r^2 = \frac{n(n+1)(2n+1)}{6} \quad (n = 1, 2, \dots)$ $\sum_{r=1}^n r^3 = \left[\frac{n(n+1)}{2} \right]^2, \quad \sum_{r=1}^n r^4 = \frac{n(n+1)(2n+1)(3n^2+3n-1)}{30} \quad (n = 1, 2, \dots)$
Formula geometrica	$\sum_{r=0}^n x^r = \frac{1-x^{n+1}}{1-x} \quad (n = 0, 1, \dots; x \neq 1)$
Formula aritmetica-geometrica	$\sum_{r=1}^n r x^r = \frac{[1-(n+1)x^n + nx^{n+1}]x}{(1-x)^2} \quad (n = 1, 2, \dots; x \neq 1)$

Tabella E.2 – Serie.

Serie geometrica	$\sum_{r=0}^{+\infty} x^r = \frac{1}{1-x} \quad (x < 1)$
Serie aritmetica- geometrica	$\sum_{r=0}^{+\infty} r x^r = \frac{x}{(1-x)^2} \quad (x < 1)$
Serie binomiale	$\sum_{r=0}^{+\infty} \binom{r+k}{k} x^r = (1-x)^{-k-1}, \quad (x < 1; k = 0, 1, \dots)$
Serie esponenziale	$\sum_{r=0}^{+\infty} \frac{x^r}{r!} = e^x \quad (x \in \mathbb{R})$
Serie iperboliche	$\sinh x = \frac{e^x - e^{-x}}{2} = \sum_{r=0}^{+\infty} \frac{x^{2r+1}}{(2r+1)!}, \quad \cosh x = \frac{e^x + e^{-x}}{2} = \sum_{r=0}^{+\infty} \frac{x^{2r}}{(2r)!} \quad (x \in \mathbb{R})$
Serie trigonometriche	$\sin x = \sum_{r=0}^{+\infty} (-1)^r \frac{x^{2r+1}}{(2r+1)!}, \quad \cos x = \sum_{r=0}^{+\infty} (-1)^r \frac{x^{2r}}{(2r)!} \quad (x \in \mathbb{R})$

Tabella E.3 – Limiti.

Limite di Eulero	
$\lim_{n \rightarrow +\infty} \left(1 - \frac{\alpha_n}{n}\right)^n = e^{-\alpha},$	dove $\alpha_1, \alpha_2, \dots$ è una successione di reali tali da aversi $\lim_{n \rightarrow +\infty} \alpha_n = \alpha$.
Caso particolare:	$\lim_{n \rightarrow +\infty} \left(1 - \frac{x}{n}\right)^n = e^{-x} \quad (x \in \mathbb{R})$
$\lim_{x \rightarrow 0} \frac{\sin(\alpha x)}{x} = \alpha,$	$\lim_{x \rightarrow 0} \frac{1 - \cos(\alpha x)}{x} = 0 \quad (\alpha \in \mathbb{R})$
$\lim_{x \rightarrow 0} \frac{1 - e^{-\alpha x}}{x} = \alpha$	$(\alpha \in \mathbb{R})$

Tabella E.4 – Integrali definiti.

$\int_a^b x^n dx = \frac{b^{n+1} - a^{n+1}}{n+1} \quad (a, b \in \mathbb{R}; a < b)$
$\int_0^{+\infty} x^n e^{-\lambda x} dx = \frac{n!}{\lambda^{n+1}} \quad (\lambda > 0; n = 0, 1, \dots)$
$\int_{-\infty}^{+\infty} e^{-z^2} dz = \sqrt{\pi}, \quad \int_0^{+\infty} e^{-z^2} dz = \frac{\sqrt{\pi}}{2}$
$\int_{-\infty}^{+\infty} e^{-\alpha z^2 + \beta z} dz = \sqrt{\frac{\pi}{\alpha}} \exp\left\{\frac{\beta^2}{4\alpha}\right\} \quad (\alpha > 0, \beta \in \mathbb{R})$
$\int_0^x \sin z dz = 1 - \cos x, \quad \int_0^x \cos z dz = \sin x$

Tabella E.5 – Funzioni gamma e beta.

Funzione gamma di Eulero	$\Gamma(\nu) = \int_0^{+\infty} x^{\nu-1} e^{-x} dx \quad (\nu > 0)$
Proprietà della funzione gamma	$\Gamma(\nu) = (\nu - 1) \Gamma(\nu - 1) \quad (\nu > 1)$ $\Gamma(n) = (n - 1)! \quad (n = 1, 2, \dots)$ $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ $\Gamma\left(n + \frac{1}{2}\right) = \frac{1 \cdot 3 \cdot 5 \cdots (2n - 1)}{2^n} \sqrt{\pi} \quad (n = 1, 2, \dots)$
Approssimazione di Stirling	$\Gamma(z + 1) \sim \left(\frac{z}{e}\right)^z \sqrt{2\pi z} \quad (z \rightarrow +\infty)$ $n! \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n} \quad (n \rightarrow +\infty)$
Funzione beta di Eulero	$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx \quad (\alpha > 0, \beta > 0)$
Proprietà della funzione beta	$B(\alpha, \beta) = \frac{\Gamma(\alpha) \Gamma(\beta)}{\Gamma(\alpha + \beta)} \quad (\alpha > 0, \beta > 0)$

Appendice F

Disuguaglianze

Questa appendice contiene una sintesi di alcune disuguaglianze analitiche e probabilistiche.

Tabella F.1 – Disuguaglianze analitiche.

Disuguaglianza di Cauchy-Schwarz-Bunyakowsky	$\left(\sum_{i=1}^n x_i y_i\right)^2 \leq \left(\sum_{i=1}^n x_i^2\right) \left(\sum_{i=1}^n y_i^2\right),$ $x_1, x_2, \dots, x_n \text{ e } y_1, y_2, \dots, y_n \text{ arbitrari numeri reali.}$
$g(x)$ convessa e continua in $I = (a, b)$	$g[\alpha x_1 + (1 - \alpha) x_2] \leq \alpha g(x_1) + (1 - \alpha) g(x_2),$ $0 < \alpha < 1; x_1, x_2 \in I, x_1 < x_2.$
$g(x)$ concava e continua in $I = (a, b)$	$g[\alpha x_1 + (1 - \alpha) x_2] \geq \alpha g(x_1) + (1 - \alpha) g(x_2),$ $0 < \alpha < 1; x_1, x_2 \in I, x_1 < x_2.$
Disuguaglianza di Young	$y_1 y_2 \cdots y_n \leq \sum_{r=1}^n \vartheta_r y_r^{1/\vartheta_r},$ $y_1, y_2, \dots, y_n, \vartheta_1, \vartheta_2, \dots, \vartheta_n \text{ reali positivi, } \vartheta_1 + \vartheta_2 + \dots + \vartheta_n = 1.$
	$1 - x \leq e^{-x} \leq \frac{1}{1 + x} \leq 1, \quad x \geq 0.$
	$p(1 - p) \leq \frac{1}{4}, \quad 0 < p < 1.$
	$(1 - p)^n \leq \frac{1 - p}{1 - p + np}, \quad 0 < p < 1; n = 0, 1, \dots$
	$1 - p + p e^x \leq e^{p x} \leq \exp\{p(e^x - 1)\}, \quad 0 \leq p \leq 1; x \geq 0.$
	$\cosh x \leq e^{x^2/2}, \quad x \in \mathbb{R}.$
	$e^{s x} \leq \frac{b - x}{b - a} e^{s a} + \frac{x - a}{b - a} e^{s b}, \quad s \in \mathbb{R}; a \leq x \leq b.$

Tabella F.2 – Disuguaglianze coinvolgenti probabilità.

Disuguaglianza di Boole	$P\left(\bigcup_{n=1}^{+\infty} A_n\right) \leq \sum_{n=1}^{+\infty} P(A_n), \quad A_1, A_2, \dots \text{eventi di } \mathcal{F}.$	
	$1 - \exp\left\{-\sum_{i=1}^n P(A_i)\right\} \leq \sum_{i=1}^n P(A_i), \quad A_1, A_2, \dots, A_n \text{eventi indipendenti di } \mathcal{F}.$	
Disuguaglianze di Markov	$P(X \geq a) \leq \frac{E(X)}{a}, \quad X \text{ non negativa, } a > 0.$	
	$P(X < a) \geq 1 - \frac{E(X)}{a}, \quad X \text{ non negativa, } a > 0.$	
Disuguaglianze di Markov generalizzata	$P(X \geq a) \leq \frac{E(X ^\nu)}{a^\nu}, \quad a, \nu \text{ reali positivi.}$	
	$P(X < a) \geq 1 - \frac{E[X ^\nu]}{a^\nu}, \quad a, \nu \text{ reali positivi.}$	
Disuguaglianze di Chebyshev	$P\{ X - E(X) \geq \varepsilon\} \leq \frac{\text{Var}(X)}{\varepsilon^2}, \quad \varepsilon > 0, E(X^2) \text{ finito.}$	
	$P\{ X - E(X) < \varepsilon\} \geq 1 - \frac{\text{Var}(X)}{\varepsilon^2}, \quad \varepsilon > 0, E(X^2) \text{ finito.}$	
	$P\{ X - E(X) < r \sqrt{\text{Var}(X)}\} \geq 1 - \frac{1}{r^2}, \quad r > 0, E(X^2) \text{ finito.}$	
	$P\left\{\left \frac{X - E(X)}{E(X)}\right < \delta\right\} \geq 1 - \frac{(CX)^2}{\delta^2}, \quad \delta > 0, E(X) \neq 0, E(X^2) \text{ finito.}$	
Disuguaglianze di Chebyshev unilaterali	$P\{X \geq E(X) + \varepsilon\} \leq \frac{\text{Var}(X)}{\text{Var}(X) + \varepsilon^2}, \quad \varepsilon > 0, E(X^2) \text{ finito.}$	
	$P\{X \leq E(X) - \varepsilon\} \leq \frac{\text{Var}(X)}{\text{Var}(X) + \varepsilon^2}, \quad \varepsilon > 0, E(X^2) \text{ finito.}$	
	$P(X > 0) \geq \frac{[E(X)]^2}{E(X^2)}, \quad X \text{ non negativa.}$	
	$P(X = 0) \leq \frac{\text{Var}(X)}{E(X^2)}, \quad X \text{ non negativa.}$	

Tabella F.3 – Disuguaglianze coinvolgenti funzioni generatrici dei momenti.

Disuguaglianze di Chernoff	$P(X \geq a) \leq e^{-s a} M_X(s), \quad s \geq 0, M_X(s) \text{ finita in un intorno dell'origine.}$
	$P(X \leq a) \leq e^{-s a} M_X(s), \quad s \leq 0, M_X(s) \text{ finita in un intorno dell'origine.}$
	$M_X(s) \geq e^{s E(X)}, \quad s \in \mathbb{R}, M_X(s) \text{ finita.}$
	$M_X(s) \leq \frac{b-\mu}{b-a} e^{s a} + \frac{\mu-a}{b-a} e^{s b}, \quad s, a, b \in \mathbb{R}, E(X) = \mu \text{ finito}, P(a \leq X \leq b) = 1.$
	$M_X(s) \leq \cosh(s c), \quad c > 0, s \in \mathbb{R}, E(X) = 0, P(X \leq c) = 1.$

Tabella F.4 – Disuguaglianze coinvolgenti momenti.

	$E(X^2) \geq [E(X)]^2, \quad E(X) \text{ finito.}$
Disuguaglianza di Schwarz	$ E(XY) \leq \sqrt{E(X^2) E(Y^2)}, \quad E(X^2), E(Y^2) \text{ finiti.}$
Disuguaglianze di Jensen	$E[g(X)] \geq g[E(X)], \quad E(X) \text{ finito, } g \text{ funzione convessa e continua.}$
	$E[g(X)] \leq g[E(X)], \quad E(X) \text{ finito, } g \text{ funzione concava e continua.}$
Disuguaglianza di Schwarz–Hölder	$E\left(X_1 X_2 \cdots X_n \right) \leq \prod_{r=1}^n \left[E\left(X_r ^{\alpha_r}\right)\right]^{1/\alpha_r},$ $\alpha_1, \alpha_2, \dots, \alpha_n \text{ reali positivi, } \alpha_1^{-1} + \alpha_2^{-1} + \dots + \alpha_n^{-1} = 1.$
Disuguaglianza di Minkowsky	$\left[E\left(X_1 + X_2 + \dots + X_n ^\alpha\right)\right]^{1/\alpha} \leq \sum_{r=1}^n \left[E\left(X_r ^\alpha\right)\right]^{1/\alpha}, \quad \alpha > 1.$

Tabella F.5 – Disuguaglianze coinvolgenti somme di variabili aleatorie.

Disuguaglianze di Hoeffding	$P\{Y - E(Y) \geq c\} \leq \exp\left\{-\frac{2c^2}{\sum_{r=1}^n (b_r - a_r)^2}\right\}, \quad c > 0, Y = X_1 + X_2 + \dots + X_n,$ $P\{Y - E(Y) \leq -c\} \leq \exp\left\{-\frac{2c^2}{\sum_{r=1}^n (b_r - a_r)^2}\right\}, \quad c > 0, Y = X_1 + X_2 + \dots + X_n,$ $X_1, X_2, \dots, X_n \text{ indipendenti, } E(X_r) \text{ finito, } P\{a_r \leq X_r \leq b_r\} = 1 \ (r = 1, 2, \dots, n).$
Disuguaglianza di Kolmogorov	$P\left\{\max\left(Y_1 , Y_2 , \dots, Y_n \right) \geq a\right\} \leq \frac{E(Y_n^2)}{a^2}, \quad a > 0, Y_r = X_1 + X_2 + \dots + X_r$ $(r = 1, 2, \dots, n), X_1, X_2, \dots, X_n \text{ indipendenti; } E(X_i) = 0, E(X_i^2) \text{ finito } (i = 1, 2, \dots, n).$
	$P\left\{\max_{1 \leq r \leq n} Y_r - E(Y_r) \geq a\right\} \leq \frac{1}{a^2} \sum_{r=1}^n \text{Var}(X_r), \quad a > 0, Y_r = X_1 + X_2 + \dots + X_r$ $(r = 1, 2, \dots, n), X_1, X_2, \dots, X_n \text{ indipendenti, } E(X_i^2) \text{ finito } (i = 1, 2, \dots, n).$
Disuguaglianza di Lévy	$P\left\{\max(Y_1, Y_2, \dots, Y_n) \geq a\right\} \leq 2P(Y_n \geq a), \quad a > 0, Y_r = X_1 + X_2 + \dots + X_r$ $(r = 1, 2, \dots, n), X_1, X_2, \dots, X_n \text{ indipendenti con funzioni di distribuzione simmetriche intorno allo zero.}$

Tabella F.6 – Disuguaglianze per alcune distribuzioni di probabilità.

$X \sim \mathcal{B}(n, p)$	$P(X \geq x) \leq \frac{np}{x}.$ $P(X \geq x) \leq \frac{np(1-p)}{np(1-p) + (x-np)^2}.$ $P(X \geq x) \leq \left[\frac{p(n-x)}{x(1-p)} \right]^x \left[\frac{n(1-p)}{n-x} \right]^n, \quad np \leq x < n.$ $P(X \geq x) \leq \left(\frac{np}{x} \right)^x e^{x-np}, \quad np \leq x < n.$ $P(X - np \geq c) \leq \frac{np(1-p)}{c^2}, \quad c > 0.$ $P\{X - np \geq c\} \leq e^{-2c^2/n}, \quad c > 0.$ $P\{X - np \leq -c\} \leq e^{-2c^2/n}, \quad c > 0.$ $P(X - np \geq c) \leq 2e^{-2c^2/n}, \quad c > 0.$
$Y = \frac{X}{n}$ con $X \sim \mathcal{B}(n, p)$	$P(Y - p \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2}, \quad \varepsilon > 0.$ $P(Y - p \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2}, \quad \varepsilon > 0.$
$X \sim \mathcal{P}(\lambda)$	$P(X \geq x) \leq \frac{\lambda}{\lambda + (x - \lambda)^2}.$ $P(X \geq x) \leq \frac{e^{-\lambda}(\lambda e)^x}{x^x}, \quad x \geq \lambda.$
$Y = X_1 + X_2 + \dots + X_n$, con $P(X_r = x) = \begin{cases} 1/2, & x = -1 \\ 1/2, & x = 1 \\ 0, & \text{altrimenti} \end{cases}$ per $r = 1, 2, \dots, n$	$P(Y \geq y) \leq \frac{n}{n + y^2}.$ $P(Y \geq y) \leq \left(\frac{n-y}{n+y} \right)^{y/2} \left(\frac{n}{\sqrt{(n-y)(n+y)}} \right)^n, \quad 0 \leq y < n.$ $P(Y \geq y) \leq \exp \left\{ -\frac{y^2}{2n} \right\}, \quad 0 \leq y < n.$

Alcuni personaggi

Bachelier, Louis (1870-1946)
Bayes, Thomas (1702-1761)
Bernoulli, Daniele (1700-1782)
Bernoulli, Jacques (1654-1705)
Bernstein, Sergei Natanovich (1880-1968)
Bertrand, Joseph Louis François (1822-1900)
Boole, George (1815-1864)
Borel, Émile (1871-1956)
Born, Max (1882-1970)
Bunyakovsky, Victor Jakowlewitseh (1804-1889)
Cardano, Gerolamo (1501-1576)
Carnap, Rudolf (1891-1970)
Cauchy, Augustin-Louis (1789-1857)
Chebyshev, Pafnuti Lvovich (1821-1894)
Chernoff, Herman (1923-)
de Fermat, Pierre (1601-1665)
de Finetti, Bruno (1906-1985)
de Méré, Cavaliere Antoine Gombaud (1607-1684)
de Moivre, Abraham (1667-1754)
De Morgan, Augustus (1806-1871)
Deparcieux, Antoine (1703-1768)
Erlang, Agner Krarup (1878 - 1929)
Eulero, Leonhard (1707-1783) Feller, William (1906-1970)
Fisher, Sir Ronald Aylmer (1890-1962)
Fokker, Adriaan Daniel (1887-1972)
Galilei, Galileo (1564-1642)
Gauss, Karl Friedrich (1777-1855)
Jensen, Johan Ludwig William Valdemar (1859-1925)
Hoeffding, Wassily (1914-1991)
Hölder, Otto Ludwig (1859-1937)
Huygens, Christian (1629-1695)
Khinchin, Aleksandr Yakovlevich (1894-1959)
Kolmogorov, Andrey Nikolaevich (1903-1987)
Lévy, Paul (1886-1971)
Laplace, Pierre-Simon (1749-1827)
Lindeberg, Jarl Waldemar (1876-1932)
Lyapunov, Aleksandr Mikhailovich (1857-1918)
Markov, Andrei Andreyevich (1856-1922)
Minkowsky, Hermann (1864-1909)
Newton, Isaac (1643-1727)
Pacioli, Luca (1445-1517)
Pascal, Blaise (1623-1662)
Planck, Max (1858-1947)
Poincaré, Jules-Henri (1854-1912)
Poisson, Simeon Denis (1781-1840)
Savage, Leonard Jimmie (1917-1971)
Schwarz, Hermann (1843-1921)
Slutsky, Evgeny Evgenievich (1880-1948)
Stirling, James (1692-1770)
Student, pseudonimo di William Sealy Gosset (1876-1937)
Vandermonde, Alexandre-Théophile (1736-1796)
Venn, John (1834-1923)
von Mises, Richard (1883-1953)
Weibull, Ernst Hjalmar Waloddi (1887-1979)
Wiener, Norbert (1894-1964)
Young, William Henry (1863-1942)

Indice analitico

- additività
 - completa della probabilità, 31
 - finita della probabilità, 11, 13, 15, 32
- affidabilità, 45
- algebra
 - σ -algebra, 26
 - σ -algebra generata, 27
- approssimazione normale, 309
 - della distribuzione binomiale, 318
- assenza di memoria, 117, 137
- assiomi della probabilità, 31

- binomio di Newton, 40, 109, 174, 177

- classe di Borel, 28, 60, 79
- coefficiente
 - di correlazione, 193
 - di variazione, 179
- combinazione
 - con ripetizione, 19
 - senza ripetizione, 19
- componenti
 - collegati in parallelo, 95, 297
 - collegati in serie, 95, 138, 147, 297
 - soggetti a guasti, 110, 139
- convergenza
 - in distribuzione, 302
 - in probabilità, 301
 - quasi certa, 300
- covarianza, 190, 201
- curtosi, 180

- densità di probabilità, 68
 - condizionata, 225, 228, 231
 - congiunta, 84, 88
 - marginale, 85, 88
- deviazione standard, 179

- diagrammi di Venn, 8
- disposizione
 - con ripetizione, 18
 - senza ripetizione, 18
- distribuzione
 - beta, 256
 - binomiale, 108, 200, 209
 - binomiale negativa, 119, 210
 - chi-quadrato, 260
 - degli eventi rari, 123
 - di Bernoulli, 107, 208
 - di Cauchy, 162, 268
 - di Erlang, 141, 213
 - di Fisher, 263
 - di Laplace, 217
 - di Pascal, 210
 - di Poisson, 122, 211
 - di Student, 267
 - di Weibull, 146, 215
 - esponenziale, 136, 184, 212
 - gamma, 144, 213
 - geometrica, 116, 209
 - iperesponenziale, 144, 214
 - ipergeometrica, 112, 198, 209
 - ipergeometrica multivariata, 130
 - multinomiale, 127, 201
 - normale, 149, 185, 215
 - normale bivariata, 249
 - normale standard, 150
 - simmetrica, 175, 180
 - uniforme, 133, 184, 212
 - uniforme discreta, 106, 183, 208
- distribuzioni condizionate, 219
 - per variabili continue, 224
 - per variabili discrete, 219
 - per vettori aleatori misti, 228

- disuguaglianza di
 - Boole, 34
 - Cauchy-Schwarz-Bunyakowsky, 214
 - Chebyshev, 274
 - unilaterale, 278
 - Chernoff, 280
 - Hoeffding, 289
 - Jensen, 285
 - Kolmogorov, 291
 - Lévy, 293
 - Markov, 271
 - generalizzazione, 273
 - Minkowsky, 287
 - Schwarz, 193, 287
 - Schwarz-Hölder, 287
 - Young, 286
- durata di
 - funzionamento, 137
 - servizio, 143, 145
 - vita, 137, 138
 - vita residua, 137
- esperimento casuale, 5
- eventi, 6, 27
 - evento certo, 7
 - evento elementare, 6
 - evento impossibile, 7
 - evento quasi certo, 31
 - evento quasi impossibile, 31
 - incompatibili, 7
 - indipendenti, 40
 - mutuamente esclusivi, 7
- expectation, 155
- famiglia parametrica di distribuzioni, 105
- formula di
 - De Morgan, 10
 - inclusione-esclusione, 36
 - Stirling, 268
 - Vandermonde, 112
- funzione
 - indicatrice, 67
 - beta di Eulero, 256
 - gamma di Eulero, 144
- funzione di distribuzione, 62
 - condizionata, 220, 224, 225
 - congiunta, 80
 - marginale, 81
- funzione di probabilità, 66
 - condizionata, 219, 228, 231
 - congiunta, 83, 87
 - marginale, 83, 87
- funzione di ripartizione, 62
- funzione generatrice
 - dei momenti, 203
 - di probabilità, 205
- indipendenza
 - di eventi, 40
 - di variabili aleatorie, 89
- insieme
 - di alternative, 53
 - potenza, 27
- legge
 - delle alternative, 54, 236
 - delle alternative condizionate, 55
 - delle probabilità composte, 51
 - di Bayes, 56, 239
- legge debole dei grandi numeri di
 - Khinchin, 313
 - Chebyshev, 313
 - Markov, 312
- legge empirica del caso, 2, 12, 321
- legge forte dei grandi numeri, 316, 317
- media
 - aritmetica, 153
 - armonica, 288
 - campionaria, 216, 288, 311, 313, 315
 - condizionata, 240, 241
 - geometrica, 288
- mediana, 181
- moda, 183
- momenti, 173, 174
 - centrali, 173, 199
 - condizionati, 241
 - misti, 189
 - misti centrali, 190
- paradosso di

- Bertrand, 16
- San Pietroburgo, 4
- passeggiata aleatoria, 294
- permutazione, 18
- probabilità
 - a posteriori, 57
 - a priori, 57
 - assiomatica, 31
 - classica, 10
 - condizionata, 47, 233, 234
 - frequentista, 12
 - geometrica, 15
 - soggettiva, 14
- problema
 - degli insiemi monocromatici, 35
 - dei compleanni, 24
 - del Cavaliere de Méré, 3, 44
 - delle concordanze, 38, 177
- prove
 - di Bernoulli, 107, 108
 - generalizzate di Bernoulli, 126
 - indipendenti, 198
 - non indipendenti, 198
- regola
 - del 3σ , 152
 - moltiplicativa, 51
- simulazione, 78, 135, 136, 310, 315, 316
- sistema di servizio, 137, 138, 143, 145, 247
- skewness, 180
- spazio
 - campione, 6
 - di Bernoulli, 25
 - di probabilità, 31
 - probabilizzabile, 30
 - statistiche ordinate, 255
- teorema
 - centrale di convergenza, 307
 - di Bernoulli, 320
 - di De Moivre-Laplace, 317
 - di Poisson, 322
- trasformazioni di
 - variabili aleatorie, 72
 - vettori aleatori, 94
- valore medio, 154, 158, 159, 163, 166, 186, 187
- variabili aleatorie, 59, 60
 - identicamente distribuite, 62
 - indipendenti, 89
 - stocasticamente ordinate, 296
 - variabile assolutamente continua, 68
 - variabile degenerare, 66
 - variabile discreta, 65
 - variabile mista, 71
 - variabile standardizzata, 179
- varianza, 176, 196
 - condizionata, 242
- vettore aleatorio, 78, 79
 - bidimensionale continuo, 84
 - bidimensionale discreto, 82
 - multidimensionale, 87

Indice delle abbreviazioni e dei simboli

Ω	spazio campione.
ω	$\omega \in \Omega$, elemento dello spazio campione.
\emptyset	evento impossibile.
\mathcal{F}	σ -algebra.
\mathbb{R}	insieme dei numeri reali.
\mathcal{B}	σ -algebra di Borel.
P	misura di probabilità.
(Ω, \mathcal{F}, P)	spazio di probabilità.
$P(A)$	probabilità dell'evento A .
$P(A B)$	probabilità dell'evento A condizionata dal verificarsi di B .
$F_X(x)$	funzione di distribuzione, o funzione di ripartizione, di X .
$p_X(x)$	funzione di probabilità di X discreta.
I_A	funzione indicatrice dell'evento A .
$f_X(x)$	densità di probabilità di X assolutamente continua.
$P(X \in B)$	probabilità che X assuma valori in B .
$F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$	funzione di distribuzione di $\mathbf{X} = (X_1, X_2, \dots, X_n)$.
$p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$	funzione di probabilità congiunta di $\mathbf{X} = (X_1, X_2, \dots, X_n)$.
$f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$	densità di probabilità congiunta di $\mathbf{X} = (X_1, X_2, \dots, X_n)$.
$P(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n)$	probabilità che $\mathbf{X} = (X_1, X_2, \dots, X_n)$ assuma valori in $B_1 \times B_2 \times \dots \times B_n$.
$E(X)$	valore medio di X .
$E(X^n)$	momento di ordine n di X .
μ_n	momento di ordine n .
$E[(X - E(X))^n]$	momento centrale di ordine n di X .
$\bar{\mu}_n$	momento centrale di ordine n .
μ_X	valore medio di X .
$\text{Var}(X)$	varianza di X .
σ_X	deviazione standard di X .
C_X	coefficiente di variazione di X .
α_3	coefficiente di simmetria, o skewness, di X .
α_4	curtosi di X .
$E(X^i Y^j)$	momento misto di ordine (i, j) di (X, Y) .
$\mu_{i,j}$	momento misto di ordine (i, j) .

$E[(X - E(X))^i(Y - E(Y))^j]$	momento misto centrale di ordine (i, j) di (X, Y) .
$\bar{\mu}_{i,j}$	momento misto centrale di ordine (i, j) .
$\text{Cov}(X, Y)$	covarianza di X e Y .
$\bar{\mu}_{1,1}$	covarianza.
$\varrho(X, Y)$	coefficiente di correlazione di X e Y .
$M_X(s)$	funzione generatrice dei momenti di X .
$G_X(z)$	funzione generatrice di probabilità di X .
$F_{X Y}(x y)$	funzione di distribuzione condizionata di X dato $Y = y$.
$p_{X Y}(x y)$	funzione di probabilità di X condizionata da $Y = y$.
$f_{X Y}(x y)$	densità di probabilità condizionata di X dato $Y = y$.
$P(X \in B Y = y)$	probabilità condizionata di $\{X \in B\}$ dato $Y = y$.
$E(X Y = y)$	media condizionata di X dato $Y = y$.
$E(X^n Y = y)$	momento condizionato di ordine n di X dato $Y = y$.
$\text{Var}(X Y = y)$	varianza condizionata di X dato $Y = y$.
\bar{X}	media campionaria.
$B(\alpha, \beta)$	funzione beta di Eulero.
$\Gamma(\nu)$	funzione gamma di Eulero.
$X \sim F$	X di distribuzione F .
$\mathcal{U}_d(n)$	distribuzione uniforme discreta di parametro n .
$\mathcal{B}(1, p)$	distribuzione di Bernoulli di parametro p .
$\mathcal{B}(n, p)$	distribuzione binomiale di parametri n e p .
$\mathcal{I}(n, m, N - m)$	distribuzione ipergeometrica di parametri $n, m, N - m$.
$\mathcal{BN}(1, p)$	distribuzione geometrica di parametro p .
$\mathcal{BN}(n, p)$	distribuzione binomiale negativa di parametri n e p .
$\mathcal{P}(\lambda)$	distribuzione di Poisson di parametro λ .
$\mathcal{M}(n, p_1, p_2, \dots, p_k)$	distribuzione multinomiale di parametri n, p_1, p_2, \dots, p_k .
$\mathcal{I}(n, N_1, N_2, \dots, N_k)$	distribuzione ipergeometrica di parametri n, N_1, N_2, \dots, N_k .
$\mathcal{U}(a, b)$	distribuzione uniforme in (a, b) .
$\mathcal{E}(1, \lambda)$	distribuzione esponenziale di parametro λ .
$\mathcal{E}(n, \lambda)$	distribuzione di Erlang di parametri n e λ .
$\mathcal{G}(\nu, \lambda)$	distribuzione gamma di parametri ν e λ .
$\mathcal{N}(\mu, \sigma)$	distribuzione normale di parametri μ e σ .
$\Phi(x)$	funzione di distribuzione normale standard.
$\mathcal{Be}(\alpha, \beta)$	distribuzione beta di parametri α e β .
$\chi^2(n)$	distribuzione chi-quadrato con n gradi di libertà.
$\mathcal{F}(n_1, n_2)$	distribuzione di Fisher con n_1 e n_2 gradi di libertà.
$T(n)$	distribuzione di Student con n gradi di libertà.
$X_n \xrightarrow{q.c.} X$	la successione X_1, X_2, \dots converge quasi certamente a X .
$X_n \xrightarrow{P} X$	la successione X_1, X_2, \dots converge in probabilità a X .
$X_n \xrightarrow{d} X$	la successione X_1, X_2, \dots converge in distribuzione a X .

Si espongono gli elementi del calcolo delle probabilità con la finalità di indicare come avvicinarsi alla descrizione quantitativa dei fenomeni aleatori: in breve, come guardare con correttezza matematica al "mondo dell'incerto". Il volume, rivolto particolarmente agli studenti dei nuovi ordinamenti dei corsi di studio afferenti alle Facoltà di Scienze, Ingegneria ed Economia, costituisce utile strumento di consultazione anche per operatori in aree disciplinari quali Sociologia e Medicina.

Antonio Di Crescenzo è professore associato di Probabilità e Statistica Matematica nella Facoltà di Scienze dell'Università di Salerno. Le sue ricerche sono rivolte alla teoria e alla simulazione di processi stocastici, alla modellistica in biomatematica ed ai metodi probabilistici e statistici in teoria dell'affidabilità.

Virginia Giorno è professore associato di Informatica nella Facoltà di Scienze dell'Università di Salerno. La sua attività di ricerca è incentrata sulla formulazione di modelli di sistemi in evoluzione stocastica, nonché su analisi e sviluppo di relativi algoritmi e metodi computazionali.

Amelia G. Nobile è professore ordinario di Informatica nella Facoltà di Scienze dell'Università di Salerno. I suoi interessi scientifici sono prevalentemente rivolti alla formulazione e all'analisi di modelli probabilistici per la descrizione di dinamiche neuronali, di evoluzioni di popolazioni e di sistemi con file di attesa.

Luigi Maria Ricciardi è professore ordinario di Calcolo delle Probabilità e Statistica Matematica nella Facoltà di Scienze dell'Università di Napoli Federico II. Ha svolto a lungo attività didattica e di ricerca anche in vari atenei ed istituzioni scientifiche di Stati Uniti e Giappone. I suoi volumi a carattere didattico includono gli "Esercizi di Calcolo delle Probabilità" (con S. Rinaldi) e gli "Elementi di Statistica" (con A. Di Crescenzo), entrambi per i tipi della Liguori Editore.

In copertina: rappresentazione schematica del moto aleatorio della proteina-motore Miosina II nella teoria di Toshio Yanagida.