

Aniello Buonocore
Antonio Di Crescenzo
Luigi Maria Ricciardi

519.2
BUS
1A

Appunti di probabilità

SYS: 963073

COD: 519.2



UNIVERSITÀ DI NAPOLI FEDERICO II
BIBLIOTECA DI AREA SCIENZE
INV. INT. 2016/185

Liguori Editore

Indice

Questa opera è protetta dalla Legge sul diritto d'autore
(<http://www.liguori.it/areadownload/LeggeDirittoAutore.pdf>).

Tutti i diritti, in particolare quelli relativi alla traduzione, alla citazione, alla riproduzione in qualsiasi forma, all'uso delle illustrazioni, delle tabelle e dei materiali software e codice, alla trasmissione radiofonica o televisiva, alla registrazione analogica o digitale, alla pubblicazione e diffusione attraverso la rete Internet sono riservati. La riproduzione di questa opera, anche se parziale o in copia digitale, fatte salve le eccezioni di legge, è vietata senza l'autorizzazione scritta dell'Editore.

Liguori Editore
Via Posillipo 394 - I 80123 Napoli NA
<http://www.liguori.it/>

© 2011 by Liguori Editore, S.r.l.
Tutti i diritti sono riservati
Prima edizione italiana Luglio 2011
Stampato in Italia da Liguori Editore, Napoli

Buonocore, Aniello :
Appunti di probabilità Aniello Buonocore, Antonio Di Crescenzo, Luigi Maria Ricciardi
Napoli : Liguori, 2011
ISBN-13 978 - 88 - 207 - 5400 - 6

I. Distribuzioni probabilistiche II. Variabili aleatorie I. Titolo II. Collana III. Serie

Ristampe:

20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0

La carta utilizzata per la stampa di questo volume è inalterabile, priva di acidi, a pH neutro, conforme alle norme UNI EN Iso 9706 ∞ , realizzata con materie prime fibrose vergini provenienti da piantagioni rinnovabili e prodotti ausiliari assolutamente naturali, non inquinanti e totalmente biodegradabili (FSC, PEFC, ISO 14001, Paper Profile, EMAS).

Premessa

xi

1 Concetti Introduttivi

1.1 Le origini	1
1.2 Alcuni esempi	6
1.2.1 Problema del tappeto	7
1.2.2 Passeggiate aleatorie	8
1.2.3 Moto browniano	9
1.2.4 Lancio della moneta	11
1.3 Gli eventi	11

2 Prime Definizioni di Probabilità

2.1 La legge empirica del caso	17
2.2 Frequenza empirica e probabilità	17
2.3 Probabilità a priori	20
2.4 Probabilità geometriche	21
2.5 Definizione soggettiva	24

3 La Costruzione Assiomatica

3.1 La struttura degli eventi	27
3.2 Successioni e loro limiti	35
3.3 Altre strutture algebriche di interesse probabilistico	40
3.4 La probabilità	43
3.5 Il problema dell'estensione	49
3.6 Il teorema di equivalenza	55
3.7 Eventi quasi impossibili ed eventi quasi certi	59

4 Indipendenza e Condizionamento

4.1 Indipendenza di eventi	63
4.2 Legge 0-1	68

- 4.3 Probabilità condizionate	71
- 4.4 Insiemi di alternative	76
- 4.5 Teorema di Bayes <small>FONDAZIONE ADEGUATIVA TEOREMA DI MONTY-HALL PROBLEMA DELLE CONCONIANZE</small>	78
5 Variabili Casuali <small>(CASUALI → AZIONI)</small>	81
- 5.1 Il concetto di variabile casuale	81
- 5.2 La funzione di distribuzione	83
- 5.3 Variabili casuali discrete	87
5.3.1 Esempi di funzioni di distribuzione	91
- 5.4 Variabili casuali assolutamente continue	94
5.4.1 La mediana	97
5.5 Alcune proprietà generali	99
5.6 Funzioni di variabili casuali	100
5.7 Variabili casuali multidimensionali	106
5.7.1 Variabili casuali bidimensionali	109
5.8 Indipendenza di variabili casuali	110
5.9 Somme, prodotti e rapporti di variabili casuali	112
5.9.1 Somma di due variabili casuali	112
5.9.2 Prodotto di due variabili casuali	114
5.9.3 Rapporto di due variabili casuali	115
6 Momenti di Variabili Casuali	117
6.1 Momenti di variabili unidimensionali	117
6.1.1 La variabile gaussiana	127
6.1.2 La variabile esponenziale	129
6.2 Momenti di vettori casuali	130
6.3 Alcune proprietà dei momenti	134
6.4 Il caso delle variabili casuali indipendenti	138
6.5 Coefficiente di correlazione	142
7 Funzioni Ausiliarie	149
7.1 Funzioni generatrici	149
7.2 Il processo di Poisson	157
7.2.1 Utilizzazione della f.g.p.	159
8 La Disuguaglianza di Chebyshev	161
8.1 Alcune proprietà delle variabili casuali normali	161
8.2 La disuguaglianza di Chebyshev	164
8.2.1 Un caso particolare	166
8.3 Il teorema di Bernoulli	167

8.4 Altre disuguaglianze notevoli	168
8.5 Funzione di distribuzione della normale standard	171
9 Funzioni Caratteristiche	173
9.1 La funzione caratteristica	173
9.2 Funzione caratteristica e momenti	179
9.3 Formula di inversione	183
9.4 Esempi di funzioni caratteristiche	188
9.4.1 Funzioni caratteristiche di vv.cc. discrete	188
9.4.2 Variabili casuali assolutamente continue	190
10 Teoremi Asintotici	195
10.1 Limite normale della distribuzione binomiale	195
10.2 Convergenza in legge	198
10.3 Legge debole dei grandi numeri	206
10.4 Il teorema centrale del limite	208
10.5 Applicazione alle passeggiate aleatorie	209
10.6 Passeggiata aleatoria in presenza di barriere assorbenti	212
11 Condizionamento nel Continuo	229
11.1 Variabili casuali unidimensionali	229
11.2 Momenti condizionati	235
11.3 Vettori casuali	238
11.3.1 Una particolare classe di processi	240
12 Cenno ai Processi Stocastici	243
12.1 Proprietà di carattere generale	243
12.2 Processi di Markov	245
12.3 Equazione di Smolukowski	247
12.4 Equazioni di diffusione	253
12.5 Densità stazionaria	256
12.6 Il processo di Wiener senza drift	258
12.7 Il processo di Wiener con drift	262
12.8 Il processo di Ornstein-Uhlenbeck	265
A Principali Variabili Casuali	269
A.1 Variabili casuali discrete	269
A.1.1 Distribuzione di Bernoulli	269
A.1.2 Distribuzione binomiale	270
A.1.3 Distribuzione geometrica	271

A.1.4 Distribuzione binomiale negativa	272
A.1.5 Distribuzione di Poisson	274
A.1.6 Distribuzione ipergeometrica	275
A.2 Variabili casuali assolutamente continue	276
A.2.1 Densità uniforme	276
A.2.2 Densità esponenziale	278
A.2.3 Densità normale	279
A.2.4 Densità gamma	280
A.2.5 Densità beta	283
B Complementi	289
B.1 La funzione delta di Dirac	289
B.2 Integrale di Stieltjes	292
Indice analitico	299

Premessa

Questi Appunti contengono una sintesi di alcuni argomenti del Calcolo delle Probabilità e della Teoria dei Processi Stocastici svolti nell'ambito di corsi tenuti da uno degli autori sin dagli inizi degli anni 80 e diretti a studenti della Facoltà di Scienze dell'Università di Napoli Federico II. La lettura del materiale offerto è agevole e, con l'eccezione di alcuni paragrafi, non offre particolari difficoltà a chi possiede una buona conoscenza degli argomenti di matematica di base quali quelli impartiti nei corsi delle lauree triennali delle Facoltà di Scienze, di Economia e nel biennio d'Ingegneria.

A complemento di quanto contenuto nel volume "Un Primo Corso in Probabilità" (V. Giorno, A. Di Crescenzo, A.G. Nobile, L.M. Ricciardi, 2009, per i tipi di Liguori Editore), la costruzione assiomatica della Probabilità viene qui alquanto approfondita facendo ricorso alla nozione di successioni di eventi e dei relativi limiti. Ci si sofferma così sulla discussione dei teoremi di estensione e di equivalenza e, successivamente, si dimostra la cosiddetta "Legge 0-1" dopo aver introdotto il Lemma di Borel-Cantelli.

Un intero capitolo (Cap. 9) è dedicato alle funzioni caratteristiche, qui specificamente utilizzate a differenza del metodo di cui al sopra-citato volume, per la dimostrazione della legge debole dei grandi numeri e del teorema centrale del limite. Il capitolo si conclude con un'applicazione del teorema centrale del limite alla descrizione qualitativa dell'evoluzione di una passeggiata aleatoria il cui comportamento asintotico viene studiato al divergere del numero di passi compiuti. La stessa passeggiata aleatoria viene poi dettagliatamente analizzata quando ristretta tra due barriere assorbenti, e viene ricavata in forma esplicita la probabilità di occupazione degli stati accessibili ad ogni prefissato numero di passi; ciò ricorrendo all'uso della funzione generatrice di probabilità la cui efficacia, normalmente non sufficientemente sottolineata, emerge in questo contesto con chiarezza sia pure al prezzo del necessario studio di minuziose proprietà algebriche delle varie funzioni coinvolte.

L'ultimo capitolo espone, infine, gli elementi della teoria dei processi stocastici spazialmente e temporalmente continui, con particolare riferimento ai processi di diffusione. Questi

vengono descritti mediante sviluppi differenziali che, nei casi d'interesse, si riducono ad equazioni differenziali alle derivate parziali. Un cenno ai processi di Wiener e di Ornstein-Uhlenbeck conclude questa carrellata su argomenti che, se trattati con completezza, richiederebbero non un solo breve capitolo ma un intero poderoso volume.

Due appendici elencano rispettivamente le principali proprietà di alcune variabili casuali discrete e continue e riassumono le definizioni essenziali della funzione delta di Dirac e dell'integrale di Stieltjes.

Napoli, 10 novembre 2010

Gli Autori

CAPITOLO 1

Concetti Introduttivi

1.1 Le origini

A differenza di altri settori 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 annerovati 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, è dello stesso periodo (1657) ed è dovuto a Christiaan Huyghens (*De Raziociniis 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 e 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 Friederich 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. Lo sviluppo moderno del calcolo delle probabilità deve comunque riguardarsi come una grande conquista del XX secolo.

Un fondamentale elemento di progresso nella trattazione analitica della teoria della probabilità è stato offerto dalla comparsa dell'integrazione secondo Lebesgue tra gli strumenti

matematici a disposizione. Dal punto di vista pratico lo sviluppo della teoria della probabilità si è poi ulteriormente accentuato anche grazie ai progressi fatti registrare dalle scienze naturali, in particolare dalla fisica e dalla chimica, che hanno allargato il campo delle sue possibili applicazioni prima confinate quasi esclusivamente ai giochi d'azzardo o a situazioni a questi riconducibili.

Il calcolo delle probabilità nasce dall'osservazione dei fenomeni casuali sistematicamente riscontrabili nell'osservazione della natura, prefissandosi 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"; essa 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. Si riconduce quindi, di fatto, al problema stesso dell'inferenza induttiva avendo a che fare con la natura ed il significato delle dimostrazioni scientifiche, con la credibilità di proposizioni astratte sulla base delle osservazioni empiriche disponibili e, in generale, con il modo di utilizzare dati empirici per trarre indicazioni o suggerimenti su eventuali esperimenti critici da effettuarsi. La seconda area, che privilegia l'aspetto sperimentale, è incentrata sul problema della individuazione di relazioni che permettano di collegare i possibili risultati di esperimenti casuali con grandezze intrinseche. In particolare, essa è rivolta allo studio di eventi ripetitivi caratterizzati dalla validità della cosiddetta *legge empirica del caso*, ossia dalla proprietà che le rispettive frequenze empiriche di occorrenza in un gran numero di prove ripetute in condizioni sempre macroscopicamente identiche appaiono stabilizzarsi intorno a dei valori che assumiamo come misure delle rispettive probabilità di occorrenza.

La teoria matematica delle probabilità (punto (ii)) è invece fondata sulla formulazione di un insieme di assiomi dai quali deriva 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 fissati 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, di modo che proprio attraverso questi essa va a posteriori valutata; la necessità di sostituzioni o di ritocchi del sistema assiomatico diventerebbe invero ineludibile qualora si riscontrassero contraddizioni o incongruenze con fatti sperimentali.

1.1 Le origini

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 Andrej Nikolaevič 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; questa ha per oggetto l'analisi di dati tratti da esperimenti effettivi, costituendo così un naturale collegamento tra la teoria matematica della probabilità e la realtà. Va tuttavia subito menzionato che in Statistica si valutano probabilità di eventi; queste, ovviamente, non forniscono certezze, e non è quindi escluso che dal punto di vista teorico possano talvolta prevedersi risultati anche assai discordanti da quelli poi sperimentalmente osservati. Occorre quindi sempre prestare gran cura proprio alla corretta registrazione ed interpretazione dei dati sperimentali.

È opportuno sottolineare che fluttuazioni o discordanze, sempre rilevabili quando si studiano sistemi di natura intrinsecamente aleatoria, non sono imputabili ad errori di misura, come avviene per i fenomeni deterministici, ma proprio alla circostanza che i risultati della teoria concernono delle probabilità. Ad esempio, nel caso di un esperimento consistente nel lanciare ripetutamente una moneta non truccata ci si aspetta che testa e croce si verifichino un numero pressoché uguale di volte (così come previsto dal valore teorico $p = 1/2$ delle rispettive probabilità); niente tuttavia esclude che in un esperimento effettivo, consistente ad esempio in 100 lanci di una moneta non truccata, esca sempre testa.

È interessante una considerazione di Max Born,¹ che puntualizza un aspetto della rilevanza della teoria della probabilità nella Scienza: "... Il concetto di causalità interviene già nei primi passi dell'attività scientifica in virtù del fatto che nessuna misura è corretta in maniera assoluta. Io ritengo che la causalità 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 causalità ...".

Verrà ora effettuato un breve excursus di carattere storico.

Va anzitutto osservato l'assenza di una nozione quantitativa di probabilità nel mondo pregalileiano sia 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

¹Max Born. *Natural Philosophy of Cause and Chance*, Oxford University Press, 1949. Max Born nacque nel 1882 a Breslau, in Germania, attualmente la polacca Wroclaw. Morì a Gottinga il 5 Gennaio 1970. Nel 1954 gli fu conferito il Premio Nobel per la Fisica "per le sue fondamentali ricerche in meccanica quantistica, specialmente per la sua interpretazione statistica della funzione d'onda".

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 — medico, matematico e filosofo italiano —, risalente forse al 1526 ma pubblicata postuma nel 1663. Precedentemente il concetto di probabilità era stato oggetto d'interesse del matematico Luca Pacioli e, successivamente, dello stesso Galileo. Ma fu soltanto con Pascal che il calcolo delle probabilità venne alla luce soprattutto per merito del cavalier Antoine Gombaud de Méré, 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 Fermat che si estese successivamente a numerosi problemi di probabilità combinatoria. Seguirono l'opera dell'olandese Huyghens, della quale si è già fatta menzione, ed i lavori dei Bernoulli, matematici svizzeri. A Jacob Bernoulli, autore della *Ars Conjectandi*, che apparve postuma nel 1713, sono dovuti un celebre teorema ed un famoso modello matematico che porta il suo nome. Daniel Bernoulli si preoccupò invece di applicare il calcolo delle probabilità a svariati 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 la pena qui 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 tassa di iscrizione, diciamo di s rubli. Si suppone che il giocatore riceve 2 rubli se si verifica testa al primo lancio, 4 rubli se testa si verifica per la prima volta al secondo lancio, 8 rubli se testa si verifica per la prima volta al terzo lancio, e in generale 2^n rubli 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 essere infinita. 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. Ruolo importante per lo sviluppo della teoria della probabilità fu svolto dal matematico inglese De Moivre che con la sua *Doctrine of Chance* del 1718 gettò le basi della convergenza alla legge normale della distribuzione binomiale, 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 formulò il cosiddetto “problema della probabilità delle cause”, fornendone una prima soluzione attraverso una nota

1.1 Le origini

formula che porta il suo nome. Al matematico francese Antoine Deparcieux 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 Laplace, francese, il cui trattato sulla teoria analitica della probabilità è già stato richiamato, ed il tedesco Gauss 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 a lungo fatto bella mostra di sé sulle banconote tedesche da 10 marchi.²

Notevoli ulteriori contributi alla teoria della probabilità furono apportati nel Secolo XIX dalle matematiche francesi e inglese. Basti ricordare le opere di Poisson al cui nome è legata una fondamentale legge della teoria, e di George Boole la cui intuizione doveva poi condurre 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 degli Accademici sovietici Pafnutij L'vovič Chebyshev e dei suoi allievi Andrei Andreyevich Markov e Aleksandr Mikhailovich Lyapunov ai quali va, tra l'altro, il merito di aver introdotto e sistematicamente utilizzato il concetto di variabile casuale. Essi molto contribuirono a dar lustro alla celebre scuola matematica di San Pietroburgo alla quale sono ascrivibili fondamento logico e impostazione astratta della teoria della probabilità, nonché la nascita 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 in molti Paesi tra cui, ma non soltanto, U.S.A., Francia, Unione Sovietica, Italia, Gran Bretagna, Ungheria, Giappone, Polonia. Va comunque detto che ruolo centrale ha svolto la Scuola sovietica, particolarmente con Serge Bernstein, Kolmogorov e Books by Aleksandr Iakovlevich Khinchin. Fu però durante la prima decade del '900 che Émil Borel 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 a opera di Khinchin, Kolmogorov, Eugene Slutsky, Paul Lévy, Anton Lomnitsky e molti altri, con risultati davvero fecondi per lo sviluppo della teoria. Ulteriori importanti sviluppi si ebbero anche con J.W. Lindberg, Sergei Natanovich Bernstein, William Feller ed innumerevoli altri studiosi, tra cui l'eclettico Norbert Wiener. 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, Adriaan Daniël Fokker e Max Planck, ha trovato fondamento matematico rigoroso

²Il marco era la moneta corrente nella Germania prima della messa in circolazione della nuova moneta europea avvenuta il primo gennaio 2002.

agli inizi degli anni trenta ad opera di Kinchin e di Kolmogorov. Proprio a quest'ultimo, come si è già detto, è dovuta la formulazione assiomatica della teoria della probabilità.

1.2 Alcuni esempi

Nello studio di fenomeni naturali spesso accade che esperimenti ed osservazioni siano suscettibili di essere ripetuti più volte in condizioni che possono ritenersi identiche. Solitamente l'interesse dello sperimentatore è rivolto al risultato ottenuto, che è di norma espresso da uno o più numeri o che a uno o più numeri è riconducibile. Quando accade che i risultati delle osservazioni sono collegati in maniera univoca alle condizioni iniziali e alle forze che causano il fenomeno — come ad esempio accade in meccanica classica ove è possibile prevedere posizioni e velocità di gravi in movimento — il metodo probabilistico è generalmente inopportuno. Esistono tuttavia delle situazioni, almeno in linea di principio descrivibili in termini deterministic, nell'ambito delle quali un approccio di tipo probabilistico può risultare utile e fecondo. Si pensi, ad esempio, all'evoluzione dinamica di qualche mole di gas.³ Pur nella più semplice schematizzazione in cui questo insieme di tante molecole possa essere descritto a mezzo della sola meccanica classica, ci si trova immediatamente a confronto con insormontabili difficoltà pratiche, se non altro per l'impossibilità di conoscere, ad esempio, le condizioni iniziali relative a ciascuna molecola. Ma anche ammesso che ciò fosse possibile, che fossero note con esattezza le forze di interazione molecolare e che si riuscisse a risolvere il numero astronomico di equazioni descriventi l'evoluzione dinamica del gas, non si otterrebbe nulla di particolarmente significativo. Invero, per sistemi siffatti le informazioni utili sono quelle di tipo globale più che quelle legate alla conoscenza di posizioni e velocità di ciascuna molecola. È ad esempio utile conoscere la percentuale di molecole aventi energia maggiore di un preassegnato valore, oppure essere in grado di prevedere la pressione del gas noti il volume da esso occupato e la sua temperatura. Informazioni di questa natura possono ottenersi, come insegnava la Meccanica Statistica, partendo da una descrizione probabilistica ad un livello microscopico (molecolare, nel nostro caso) e considerando come significative delle opportune "medie" che sono poi suscettibili di essere poste in relazione con le grandezze macroscopiche in cui si è interessati (pressione, volume, temperatura, ecc.). Si comprende, così, come la teoria della probabilità possa essere utilizzata non soltanto in riferimento a situazioni di carattere intrinsecamente aleatorio, quali ad esempio quelle assimilabili a questi riconducibili, ma anche per studiare fenomeni che appaiono deterministic quando considerati ad un livello macroscopico, ma che in realtà sono la manifestazione di altri fenomeni che, a livello microscopico, sono suscettibili solo di una trattazione di tipo probabilistico.

Alla stregua della geometria, anche la probabilità può riguardarsi come uno strumento atto a descrivere la natura. Quando si è, ad esempio, interessati a fenomeni coinvolgenti un

³Si ricordi che una mole consta di $6.02252 \cdot 10^{23}$ molecole. Ciò segue dall'aver convenuto che una mole dell'isotopo C¹² del carbonio abbia una massa di 12 grammi.

1.2 Alcuni esempi

enorme numero di particelle, l'approccio probabilistico mira a trascurarne il comportamento dettagliato a favore della ricerca di qualche tipo di regolarità globale. Ad esempio, ci si può chiedere se si possa determinare una legge che esprima l'addensamento di particelle (supposte puntiformi) in una fissata regione di spazio. Una delle leggi più semplici spesso applicabili in problemi di questo tipo è la cosiddetta *distribuzione di Poisson* che, nella fattispecie, indicata con α una costante positiva e con k un intero non negativo,⁴ consente di determinare la probabilità di trovare k particelle in una regione di estensione A :

$$p_k = \frac{(\alpha A)^k}{k!} e^{-\alpha A}, \quad k \in \mathbb{N}_0. \quad (1.1)$$

Si noti che si ha $0 < p_k < 1$ per ogni $k \in \mathbb{N}_0$ e che risulta $\sum_{k=0}^{\infty} p_k = 1$. Queste due proprietà esprimono la circostanza che nella regione considerata vi è un numero non negativo di particelle e che all'interno della regione considerata è presente con certezza un qualche numero k (eventualmente nullo) di particelle. Leggi di questo tipo prendono il nome di *leggi di punto* e i processi dipendenti da queste leggi si dicono *processi stocastici di punto*.⁵

Accenneremo ora a qualche ulteriore esempio di tipici problemi che vanno affrontati mediante l'uso degli strumenti della teoria della probabilità. Questi esempi, se da un lato sottolineano le difficoltà risolutive cui si può andare incontro, dall'altro indicano come talora i risultati ottenuti non solo non siano intuitivi, e quindi a priori facilmente immaginabili o prevedibili, ma possano a volte persino apparire contrari alla più immediata intuizione. Proprio in questi casi con particolare evidenza emerge il ruolo essenziale svolto dalla teoria della probabilità per la comprensione e per l'analisi quantitativa dei fenomeni aleatori.

1.2.1 Problema del tappeto

Un tale problema può enunciarsi nel seguente modo:⁶ dato un tappeto di dimensioni assegnate, calcolare la probabilità che lanciando su di esso un gran numero di monete identiche non se ne trovino mai due sovrapposte. Alla semplicità di impostazione di questo problema fa riscontro la circostanza che esso è tuttora irrisolto. Occorre in realtà precisare che questa è la formulazione semplificata di un problema di estremo interesse in relazione a questioni di fisica della materia. La soluzione di tale problema fornirebbe ad esempio un'interpretazione all'esistenza di un punto critico nei liquidi e nei gas. Le complicazioni principali che non ne permettono la risoluzione sono in parte dovute a situazioni di frontiera e a problemi di vincoli (ad esempio la distanza tra le monete non deve essere inferiore ad un certo preassegnato valore). Un ruolo essenziale gioca poi anche il numero di dimensioni. Infatti è invece risolubile il problema analogo in una dimensione, che consiste nel calcolare la probabilità

⁴Nel seguito, denoteremo con \mathbb{N}_0 l'insieme dei numeri interi positivi incluso lo zero.

⁵Questi processi vengono detti "di punto" in quanto le grandezze cui ci si riferisce, ad esempio le dimensioni delle particelle nel precedente esempio, sono da considerarsi di misura nulla.

⁶K. Baclawski, M. Cerasoli, G.C. Rota "Introduzione alla Probabilità". Unione Matematica Italiana (1990)

che non vi sia sovrapposizione alcuna di n aghi identici lasciati cadere su di un segmento di fissata lunghezza.

1.2.2 Passeggiate aleatorie

Considerato un reticolo costituito da linee parallele e quidistanti, scegliamo un sistema di assi cartesiani ortogonali con gli assi coincidenti con due di queste rette e unità di misura uguale alla distanza tra due rette adiacenti. Riferiamoci poi all'esperimento consistente nel generare il moto di una particella che partendo dall'origine possa, ad ogni istante di un'arbitraria scala discreta di tempi, raggiungere con probabilità $1/4$ uno qualsiasi dei 4 vertici più vicini e che poi continui a muoversi, ad ogni successivo istante, con la stessa regola a partire dalla posizione raggiunta all'istante precedente. Si dice che la particella compie una *passeggiata aleatoria*. Questo esperimento potrebbe ad esempio schematizzare il moto di un ubriaco che, muovendosi a caso, cerchi di raggiungere la propria abitazione; un problema che si può porre consiste nel calcolo del tempo necessario affinché egli prima o poi vi pervenga. Questo problema evidentemente non può ammettere come risposta un numero, ma una *distribuzione di probabilità*, ossia una successione di numeri esprimenti le probabilità che l'ubriaco per venga alla propria abitazione in $1, 2, \dots$ passi. Comunque, si intuisce che debba esistere una soluzione quando sia nota la posizione e la forma dell'abitazione. Un altro problema, apparentemente semplice ma la cui soluzione invece non è ancora stata ottenuta, è il seguente: calcolare la probabilità che fino all' n -esimo passo ($n = 1, 2, \dots$) l'ubriaco non sia mai ripassato per uno stesso punto.

Si può anche formulare un tipo simile di problema al continuo; invero un diverso modello matematico per descrivere il cammino dell'ubriaco può essere costruito supponendo che i suoi percorsi siano rappresentabili mediante traiettorie continue che si dipartono dall'origine.

Riferendoci ad un ulteriore esempio, supponiamo che al tempo $t = 0$ una massa unitaria di materia venga situata in un particolare punto di un liquido e venga poi lasciata libera di diffondere. Ogni particella di materia percorrerà allora una certa traiettoria. Si ha così un processo di diffusione che interessa l'intero ambiente. Questo fenomeno va evidentemente descritto in termini tridimensionali. Talora è tuttavia lecito ritenere che esso si sviluppi in maniera indipendente lungo i tre assi di un sistema cartesiano ortogonale di riferimento; ci si può allora ad esempio chiedere di determinare la probabilità $p(1)$ che una particella si trovi all'interno di un fissato intervallo (a, b) al tempo $t = 1$, ossia dopo un'unità di tempo. Questo tipo di problema rientra in un quadro molto più generale, di estremo interesse teorico e applicativo, noto come teoria del moto browniano, sul quale val la pena dire qui subito qualcosa.

1.2 Alcuni esempi

1.2.3 Moto browniano

Tra gli innumerevoli fenomeni naturali che appaiono evolvere in maniera casuale, e che pertanto possono essere descritti solo ricorrendo alla teoria della probabilità, spicca per notorietà e rilevanza il moto browniano. Tutto ebbe inizio nell'anno 1827 allorché il botanico inglese Robert Brown scoprì che particelle di polline sospese in soluzione acquosa sono soggette a movimenti erratici, di direzione e di ampiezza imprevedibili, non attribuibili ad alcuna visibile causa.⁷ Successivamente egli constatò che ogni volta che si pongono in sospensione liquida particelle molto minute di materia, organica o inorganica, queste appaiono soggette — quando osservate al microscopio — ad un simile moto irregolare e caratteristico che fu denominato appunto "moto browniano". Questo, proprio in virtù dell'assenza di cause visibili, o all'epoca immaginabili, atte a giustificare l'origine, venne da taluni invocato quale inconfondibile riprova della teoria animistica della materia. Invero, proprio grazie alle osservazioni di Brown, era stato possibile concludere definitivamente che questo tipo di moto è comune a qualunque aggregato di materia sospeso in un liquido, purché di dimensioni lineari di qualche ordine di grandezza maggiori di quelle delle molecole del mezzo circostante. La ragione di tale moto apparentemente spontaneo rimase a lungo oscura, finché una spiegazione non venne fornita facendo ricorso alla teoria cinetica dei gas. In realtà si capì che il moto delle particelle in sospensione è sostanzialmente causato dalle collisioni che queste subiscono da parte delle molecole del mezzo circostante. Queste collisioni generano forze, dovute dunque all'agitazione termica della materia, caratterizzate da una risultante avente direzione e modulo casuali che fa compiere ad ognuna delle particelle un movimento bizzarro ed imprevedibile.

La teoria della probabilità fornisce gli strumenti atti a consentire una formulazione matematica rigorosa del moto browniano, permettendone pertanto uno studio sistematico. Alcuni dati numerici, con riferimento al caso più frequente in cui si osservano particelle in sospensione aquosa caratterizzate da moto browniano, sono i seguenti:

- una particella in sospensione subisce collisioni da parte delle molecole d'acqua circostanti con frequenza dell'ordine di 10^{20} s^{-1} ;
- le dimensioni lineari delle molecole d'acqua variano da 10^{-8} cm a 10^{-7} cm ;
- le particelle in sospensione devono avere dimensioni lineari di circa 10^{-4} cm ;
- il cammino libero medio delle particelle in sospensione è di circa $1.5 \cdot 10^{-4} \text{ cm}$.⁸

⁷R. Brown (1828). A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *Philos. Mag. Ann. of Philos. New ser.* Vol. 4, 161-178.

⁸Il cammino libero medio di una particella è la distanza che essa percorre in media nell'intervallo di tempo intercorrente tra due successive collisioni.

Ad Albert Einstein va il merito di aver per primo formulato una teoria quantitativa del moto browniano. Egli concluse che, sotto plausibili ipotesi, considerata la proiezione del moto su di un asse, ad esempio l'asse X , la densità $u(x, t)$ di particelle all'istante t nel punto x soddisfa l'equazione di diffusione

$$\frac{\partial}{\partial t} u(x, t) = \frac{D}{2} \frac{\partial^2}{\partial x^2} u(x, t), \quad x \in \mathbb{R}, t \geq 0, \quad (1.2)$$

dove

$$D = \frac{RT}{Nf}, \quad (1.3)$$

con R costante universale dei gas, T temperatura assoluta, N numero di Avogadro e f coefficiente di viscosità del liquido nel quale avviene il moto. Come è facile verificare, la funzione

$$u(x, t) = \frac{1}{\sqrt{2\pi D t}} \exp\left\{-\frac{(x - x_0)^2}{2Dt}\right\} \quad (1.4)$$

soddisfa l'equazione (1.2). Si può inoltre dimostrare che si ha

$$\lim_{t \rightarrow 0} u(x, t) = \delta(x - x_0), \quad (1.5)$$

dove $\delta(\cdot)$ denota la funzione delta di Dirac (cf. Appendice I). Si noti che la condizione limite (1.5) esprime l'ipotesi che all'istante $t = 0$ una massa unitaria di materia viene collocata nel punto di ascissa x_0 (così che ivi la densità è infinita all'istante iniziale) per poi venire lasciata libera di diffondere con modalità descritte dall'equazione (1.2).

Val la pena menzionare che nel 1926 fu conferito al fisico francese Jean Perrin il premio Nobel per la Fisica per essere egli riuscito a fornire un'eccellente valutazione del numero di Avogadro N a mezzo della formula (1.3) grazie alla teoria di Einstein del moto browniano, fornendo così al contempo una convincente riprova indiretta della validità dell'ipotesi di agitazione termica della materia.

Ritorniamo ora al calcolo della probabilità $p(1)$ di cui si è detto alla fine del § 1.2.2. Questa può ragionevolmente identificarsi con il rapporto tra la massa di materia che si ritrova al tempo 1 nell'intervallo (a, b) e la massa iniziale di materia, che si è supposta unitaria. Facendo uso della (1.4) si ha allora:

$$p(1) \equiv \int_a^b u(x, 1) dx = \frac{1}{\sqrt{2\pi D}} \int_a^b \exp\left\{-\frac{(x - x_0)^2}{2D}\right\} dx.$$

Se, in particolare, assumiamo $D = 1$ e $x_0 = 0$, per $t = 1$ la (1.4) diventa:

$$u(x, 1) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}. \quad (1.6)$$

È questa la celebre *funzione di Gauss*, o *gaussiana*, che avremo modo di incontrare frequentemente nel seguito.

1.3 Gli eventi

1.2.4 LANCIO DELLA MONETA

Consideriamo l'esperimento casuale (rappresentabile ancora tramite una passeggiata aleatoria) così definito: "ad ogni lancio di una moneta equa si vince una somma s se si verifica testa, si perde una somma s se il risultato è croce". Partendo da un certo capitale iniziale k , ci si può chiedere quando si ritorni al valore k , ossia quando ci si ritrovi alla pari. Si dimostra che gli attraversamenti della retta $x = k$ sono eventi rari; in altri termini è molto più probabile che cominciando a vincere si continui a vincere per tempi piuttosto lunghi, oppure che, una volta cominciato a perdere, si continui a perdere per tempi piuttosto lunghi. Questo è un risultato molto interessante che non è affatto scontato a priori, ma che anzi sembrerebbe contrastare con l'intuizione dato che (essendo la moneta equa) ci si attenderebbe che il capitale si attestasse intorno al valore iniziale k .

1.3 Gli eventi

Convenzionalmente con la locuzione "esperimento casuale" si indica ogni atto o processo, spontaneamente avente luogo 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 e 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. Tra i fenomeni del primo tipo, al di là dei familiari casi relativi a lanci di monete o di dadi, estrazioni di carte da mazzi ben mescolati, di biglie da urne, o in generale riguardanti giochi d'azzardo, ricordiamo l'esempio del § 1.2 concernente il moto delle molecole di un gas. In linea di principio è possibile scrivere le equazioni del moto delle singole molecole, ma tali equazioni sono in realtà improduttive; ciò, come si è già menzionato, non solo per il loro numero astronomico, ma anche perché 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.

Si consideri un esperimento casuale; si definisce *spazio campione*, e lo si indica tradizionalmente con Ω , l'insieme dei possibili risultati dell'esperimento.⁹ I possibili risultati dell'esperimento casuale vengono detti *punti campione*.¹⁰ Nel seguito il generico punto cam-

⁹ Talvolta per spazio campione si intende un qualsiasi insieme i cui elementi siano in corrispondenza biunivoca con i possibili risultati dell'esperimento casuale considerato.

¹⁰ Taluni autori li indicano, invece, come *eventi elementari*. Qui, se preferisce non utilizzare tale locuzione, ad

pione sarà denotato con ω . Lo spazio campione Ω indica dunque la totalità dei punti campione. Si noti che esperimenti casuali distinti possono dar luogo allo stesso spazio campione. 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.

L'associare ad ogni esperimento casuale uno spazio campione permette di introdurre il concetto di *evento*: eventi sono taluni sottoinsiemi di Ω . Si noti che con tale definizione un punto campione non è un evento, mentre evento potrebbe essere un sottoinsieme di Ω costituito da un solo punto campione. 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 o il numero 2, oppure il numero 4, oppure il numero 6, diciamo che l'evento E si è verificato.

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 \bar{E} di E ; se, allorché, 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 sotto le operazioni di unione finita, complementazione 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 all'unione numerabile. Ciò non risulta giustificabile su base intuitiva e va, per il momento, interpretato come un requisitoatto a pervenire ad una più ricca teoria matematica della probabilità.

È importante sottolineare che non sempre tutti i sottoinsiemi dello spazio campione Ω possono essere considerati eventi. Il concetto di evento è, invero, insindibilmente legato a quello 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 occorrenza 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 costituito da una successione di 4 elementi, ognuno di questi essendo T (testa) o

evitare ambiguità che potrebbero nascere in relazione alla definizione di evento che verrà data tra breve.

C (croce). Sia E il sottoinsieme di Ω contenente i punti campioni aventi almeno tre teste:

$$E = \{(TTT), (CTTT), (TCCT), (TTCT), (TTTC)\}.$$

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, così che l'osservatore è nell'impossibilità di stabilire, sulla base dei dati in suo possesso, se il risultato dell'esperimento casuale appartiene, oppure no, ad E . In tal caso E non è osservabile, e pertanto non può essere riguardato come evento. D'altra parte, si comprende come questa non osservabilità 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 di occorrenza.

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 coerenza matematica a considerare come eventi la classe degli insiemi di Borel in luogo della collezione di tutti i sottoinsiemi di Ω . L'identificazione dell'insieme delle parti di Ω con la classe degli eventi è di solito priva di significato in quanto la conseguente estensione della misura perde il necessario significato probabilistico.

Tenendo conto delle definizioni date è immediato rendersi conto che esiste un'analogia tra il linguaggio usato nel calcolo delle probabilità e quello della teoria degli insiemi. Forniamo qui di seguito alcuni esempi concernenti i termini maggiormente usati:

- al termine *elementi* di un insieme Ω corrispondono i *punti campione* $\omega \in \Omega$;
- al termine *sottoinsieme* di Ω corrisponde il termine *evento*;
- all'insieme Ω corrisponde l'*evento certo* ("certo" perché qualunque sia il punto campione ω che si verifica, risulta $\omega \in \Omega$);
- all'insieme vuoto \emptyset corrisponde l'*evento impossibile* (perché \emptyset non contiene alcun punto campione).

Le operazioni tra insiemi trovano espressione nel linguaggio degli eventi nel modo seguente:

- $E = A \cup B$: l'evento E consiste nel verificarsi di almeno uno degli eventi A e B ;
- $E = A \cap B$: l'evento E consiste nel verificarsi sia dell'evento A che dell'evento B ;
- $A \cap B = \emptyset$ indica che gli eventi A e B non possono entrambi verificarsi, ossia che A e B sono *incompatibili*, ovvero *mutuamente esclusivi* (se $\omega \in A$ allora $\omega \notin B$ e se $\omega \in B$ allora $\omega \notin A$);

- $A \cup B = \Omega$ esprime la circostanza che gli eventi A e B costituiscono un *sistema esaustivo* o, come si suole anche dire, che A e B sono *eventi necessari*, nel senso che uno di essi deve necessariamente verificarsi;
- \bar{A} (il complementare di A rispetto a Ω) è l'*evento contrario* di A : dunque \bar{A} consiste degli eventi elementari che non sono in A , così che \bar{A} si verifica se e solo se non si verifica A ;
- $B \subset A$ si esprime nella seguente forma: l'evento B implica l'evento A , ovvero A si verifica ogni volta che B si verifica (se $\omega \in B$ allora $\omega \in A$);
- $A = B$ significa che ogni punto campione in A è anche in B e ogni punto campione in B è anche in A . Quindi $A = B$ se e solo se $A \subset B$ e $B \subset A$.

Per snellire la notazione, nelle operazioni di insiemi che coinvolgono eventi si preferisce talora sostituire al simbolo di intersezione il simbolo di prodotto:

$$A \cdot B = AB = A \cap B.$$

Se $AB = \emptyset$ gli eventi A e B sono incompatibili e in tal caso al simbolo di unione si sostituisce talora quello di somma. La scrittura $A + B$ implica dunque che A e B sono incompatibili, così che l'evento $A + B$ si verifica se almeno uno, ed al più uno, degli eventi A e B si verifica.

Nel seguito le dizioni “eventi” o “insiemi” verranno talora utilizzate in maniera intercambiabile. Più in generale, dato l’isomorfismo sussistente tra algebre di insiemi e algebre di eventi (le cui definizioni saranno richiamate nel § 3.1), faremo sovente uso di proprietà tipiche della teoria degli insiemi estendendole, con le necessarie varianti terminologiche, al caso di famiglie di eventi. Ad esempio le relazioni di De Morgan e le proprietà distributive delle operazioni di unione e intersezione troveranno diretta utilizzazione nell’ambito di operazioni tra eventi. Così, in particolare, se A e B sono sottoinsiemi di un insieme Ω , com’è noto risulta:

$$A = (A \cap B) \cup (A \cap \bar{B}) \quad (1.7)$$

avendosi:

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

e sussistendo la proprietà distributiva dell’intersezione rispetto all’unione.¹¹ Lo stesso risultato (1.7) sussiste dunque anche se Ω viene interpretato come spazio campione associato ad un esperimento casuale, e se A e B vengono riguardati come eventi. Ciò è d’altronde giustificabile con la considerazione che se l’evento A si verifica, esso necessariamente deve verificarsi congiuntamente all’evento B o al suo complementare \bar{B} .

1.3 Gli eventi

La (1.7), nota anche quale *formula della partizione con il certo*, troverà frequente utilizzazione nel seguito. Altrettanto utile risulterà la proprietà insiemistica esprimente l’unione di due insiemi qualsiasi come unione di due insiemi disgiunti,

$$A \cup B = A \cup (\bar{A} \cap B), \quad (1.8)$$

che nel contesto degli eventi si interpreta affermando che l’occorrenza di almeno uno degli eventi A e B è equivalente all’occorrenza di A o all’occorrenza di B senza che A si sia verificato.

¹¹ La (1.7) può semplicemente leggersi dicendo che A è l’unione della parte di A comune a B con la parte di A che non è comune a B (ossia che è comune al complemento di B).

CAPITOLO 2

Prime Definizioni di Probabilità

2.1 La legge empirica del caso

Introdotti nei paragrafi precedenti i necessari elementi preliminari, ci rivolgiamo ora alla costruzione di una teoria assiomatica della probabilità. Va a tal proposito sottolineato che se è vero che ogni teoria assiomatica coerente è intrinsecamente valida, essa risulta assolutamente inutile se, messa alla prova con la realtà, si manifesta inidonea a rendere conto dei risultati osservati. Anche nel formulare gli assiomi della teoria della probabilità occorre tenere conto di questa considerazione.

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 ν il numero di volte in cui l'evento si verifica in n prove ripetute tra loro indipendenti, la *frequenza relativa* (o, più brevemente, *la frequenza*) di occorrenza dell'evento, ossia il rapporto ν/n , al crescere di n appare stabilizzarsi intorno ad un valore costante. Questa osservazione, di natura esclusivamente empirica, per la sua fondamentale importanza viene designata quale *legge empirica del caso*.

2.2 Frequenza empirica e probabilità

Se si accetta come valida la legge empirica del caso, è naturale richiedere che gli assiomi da individuare per la formulazione di una teoria della probabilità siano tali che la probabilità che si va a definire esibisca proprietà analoghe a quelle di cui gode la frequenza empirica. È dunque utile cominciare col determinare le proprietà caratteristiche delle frequenze in un generico schema di n prove ripetute; si richiederà poi che tali proprietà vengano soddisfatte anche dalle probabilità nella edificanda teoria.

Con riferimento ad un esperimento casuale e ad uno spazio campione Ω ad esso associato, e denotato con $E \subset \Omega$ un generico evento, si supponga di ripetere l'esperimento casuale n

volte. Indicato con ν il numero di *successi*, ossia il numero di volte in cui E si è verificato, la frequenza f di occorrenza dell'evento E nelle n prove è per definizione:¹

$$f = \frac{\nu}{n}. \quad (2.1)$$

Notiamo innanzitutto che si ha:

$$0 \leq f \leq 1. \quad (2.2)$$

Inoltre, se $E = \Omega$ risulta:

$$f = \frac{n}{n} = 1, \quad (2.3)$$

mentre se $E = \emptyset$ si ha:

$$f = \frac{0}{n} = 0. \quad (2.4)$$

È quindi ragionevole richiedere che analoghe proprietà valgano anche per la probabilità $\mathbb{P}(E)$ teorica di occorrenza del generico evento E , ossia che si abbia:

$$0 \leq \mathbb{P}(E) \leq 1, \quad \mathbb{P}(\Omega) = 1, \quad \mathbb{P}(\emptyset) = 0. \quad (2.5)$$

Quella che abbiamo fin qui usato è la frequenza relativa f ; proprietà analoghe, però, si ottengono anche se ci si riferisce alla *frequenza assoluta*, ossia al numero ν di volte in cui l'evento E si è verificato nelle n prove. In questo caso le relazioni (2.2)-(2.4) dovrebbero sostituirsi nell'ordine con le seguenti:

$$0 \leq \nu \leq n, \quad \nu = n, \quad \nu = 0.$$

L'utilizzazione della frequenza assoluta in luogo della frequenza relativa è tuttavia meno significativa in quanto i valori che essa può assumere appartengono ad un insieme che dipende dal numero n delle prove; l'uso della frequenza relativa riconduce, invece, sempre all'intervallo $[0, 1]$ qualunque sia l'esperimento casuale e qualunque sia il numero n delle prove.

Sempre con riferimento allo schema precedente, indichiamo rispettivamente con 1 e 0 il verificarsi e il non verificarsi di un evento, e siano A e B eventi qualsiasi. Notiamo che in ogni singola prova si presenta una ed una sola delle seguenti eventualità:

- (i) A si verifica e B non si verifica;
- (ii) A non si verifica e B si verifica;
- (iii) A e B si verificano;
- (iv) né A né B si verificano.

¹In linea di principio dovremmo indicare tale frequenza con f_n ; ma le considerazioni che seguono, come vedremo, sono indipendenti da n cosicché, per semplicità di scrittura, ometteremo la specificazione del suffisso n .

2.2 Frequenza empirica e probabilità

Denotiamo rispettivamente con n_1, n_2, n_3, n_4 i numeri di volte in cui nelle n prove si presentano le situazioni (i), (ii), (iii), (iv). In relazione alle n prove si può allora costruire la Tabella 2.1. Si noti che qualunque siano i valori degli n_i ($i = 1, 2, 3, 4$) osservati deve sempre risultare $\sum_{i=1}^4 n_i = n$. Dalla tabella si ricava che le frequenze di occorrenza degli eventi A e B sono rispettivamente:

$$f(A) = \frac{n_1 + n_3}{n}, \quad f(B) = \frac{n_2 + n_3}{n}.$$

Inoltre, per la frequenza di occorrenza di uno o di entrambi gli eventi A, B si ha:

$$f(A \cup B) = \frac{n_1 + n_2 + n_3}{n},$$

mentre la frequenza di occorrenza simultanea di A e B è data da:

$$f(A \cap B) = \frac{n_3}{n}.$$

Tabella 2.1 – Prove ripetute con due possibili eventi.

A	B	n
1	0	n_1
0	1	n_2
1	1	n_3
0	0	n_4

Si può anche fissare l'attenzione su un sottoinsieme di possibili risultati. Ad esempio, fissati i casi in cui si verifica B si può determinare la frequenza $f(A|B)$ che esprime la frequenza con cui si verifica A "condizionatamente" a B , ossia una volta che si sia verificato B . Dalla Tabella 2.1 segue:

$$f(A|B) = \frac{n_3}{n_2 + n_3}.$$

Analogamente, si trova:

$$f(B|A) = \frac{n_3}{n_1 + n_3}.$$

Le frequenze così determinate possono poi mettersi in mutua relazione in vario modo. Ad esempio si ha:

$$\begin{aligned} f(A \cup B) &= \frac{n_1 + n_2 + n_3}{n} = \frac{n_1 + n_3}{n} + \frac{n_2 + n_3}{n} - \frac{n_3}{n} \\ &= f(A) + f(B) - f(A \cap B). \end{aligned}$$

Se, in particolare, è $n_3 = 0$, ossia se risulta $f(A \cap B) = 0$ (evento composto impossibile ossia A e B incompatibili), sussiste la proprietà additiva:

$$f(A \cup B) = f(A) + f(B).$$

Ancora, risulta:

$$f(A \cap B) = \frac{n_3}{n} = \frac{n_1 + n_3}{n} \cdot \frac{n_3}{n_1 + n_3} = f(A) \cdot f(B|A), \quad (2.6)$$

oppure:

$$f(A \cap B) = \frac{n_3}{n} = \frac{n_2 + n_3}{n} \cdot \frac{n_3}{n_2 + n_3} = f(B) \cdot f(A|B). \quad (2.7)$$

Le proprietà sopra elencate per la frequenza sono valide qualunque sia il numero n di prove e qualunque siano gli eventi A e B , senza necessità di ulteriori ipotesi. È quindi ancora una volta naturale postulare che le stesse proprietà debbano essere verificate dalla probabilità $\mathbb{P}(\cdot)$ nella teoria assiomatica. Ad esempio, ci attendiamo che risulti:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B), \quad (2.8)$$

dalla quale, se $A \cap B = \emptyset$, si ottiene la proprietà di additività finita della probabilità:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B). \quad (2.9)$$

Inoltre dovrà aversi:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B|A) = \mathbb{P}(B) \mathbb{P}(A|B). \quad (2.10)$$

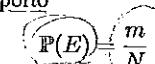
In realtà c'è da attendersi che la teoria matematica della probabilità si basi su assiomi che riflettano solo talune di queste proprietà; le rimanenti potranno infatti costituire oggetto di teoremi dimostrabili a partire dagli assiomi.

2.3 Probabilità a priori

La definizione di probabilità come "limite di frequenze empiriche" costituisce il collegamento tra la teoria matematica della probabilità ed il mondo reale; essa è quindi, in particolare, alla base della statistica. Tra la definizione frequentistica e quella assiomatica, di cui si dirà tra breve, si inserisce la cosiddetta definizione "aprioristica" di probabilità, la cui genesi risale essenzialmente allo studio dei giochi d'azzardo, ma che risulta ancora di notevole utilità in problemi coinvolgenti spazi campione finiti. Precisamente, dato uno spazio campione Ω finito, se $E \subseteq \Omega$ è un evento, detti m il numero di "casi favorevoli" all'occorrenza di E (ossia

2.4 Probabilità geometriche

il numero di punti campione in E) ed N il numero dei "casi possibili" (la cardinalità di Ω), tutti considerati equiprobabili, il rapporto



$$\mathbb{P}(E) = \frac{m}{N} \quad (2.11)$$

viene detto *probabilità a priori* dell'evento E . Si noti che si ha sempre $0 \leq \mathbb{P}(E) \leq 1$ e che l'evento è certo se e solo se $\mathbb{P}(E) = 1$, impossibile se e solo se $\mathbb{P}(E) = 0$. La probabilità a priori dovrebbe poi fornire una previsione della frequenza empirica in un esperimento di prove ripetute un elevato numero di volte.

Se la definizione frequentistica di probabilità può apparire per taluni aspetti insoddisfacente, in quanto basata su di un postulato empirico (la legge empirica del caso), la definizione aprioristica non lo è da meno in quanto manifestamente dà luogo ad un circolo vizioso poiché in tale definizione i casi possibili vengono supposti "equiprobabili". Il carattere intrinsecamente tautologico di questa definizione in realtà sfuma quando la probabilità a priori venga riferita ad un universo di oggetti ideali, caratterizzati da perfezione geometrica, isotropia, omogeneità, ecc., cosicché non vi sia ad esempio da ritenere che una faccia di un dado sia privilegiata rispetto alle altre. Pur tuttavia, per poter pervenire ad una previsione sulla frequenza di occorrenza di un evento a partire dalla sua probabilità a priori, è comunque necessario un postulato che affermi la tendenza della frequenza empirica alla probabilità a priori al crescere del numero delle prove. L'esigenza di un postulato empirico atto a collegare la teoria matematica della probabilità con il mondo reale appare dunque inevitabile, anche se non è ovvio che cosa debba intendersi, ad esempio, per rigorosa verifica sperimentale di una legge di probabilità.

2.4 Probabilità geometriche

La definizione di probabilità a priori, che richiede la finitezza della cardinalità dello spazio campione, trova un'analogia di carattere geometrico in cui lo spazio campione può essere di natura più generale. Ad esempio, sia AB un segmento di lunghezza L , Ω l'insieme dei punti di AB ed Γ il sottoinsieme di Ω costituito dai punti del segmento MN di lunghezza l (v. Figura 2.1). Si consideri poi l'esperimento casuale consistente nello scegliere a caso un punto di AB , ossia un elemento di Ω . Ci si chiede quale sia la probabilità P che il punto scelto appartenga a MN , ossia che tale punto sia un elemento di Γ . Ci si trova qui in presenza di uno spazio campione continuo, di modo che la formula (2.11) non è direttamente utilizzabile. È però ragionevole, in analogia con la definizione di probabilità a priori, identificare la richiesta probabilità con un rapporto di lunghezze, al seguente modo:²

$$P = \frac{\overline{MN}}{\overline{AB}} = \frac{l}{L}. \quad (2.12)$$

²Ciò corrisponde ad assumere che la probabilità che il punto cada in un prefissato segmento sia proporzionale alla lunghezza di questo.

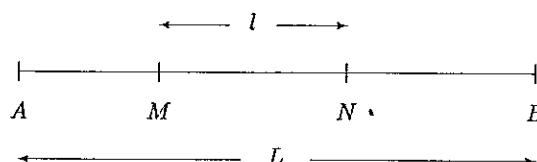


Figura 2.1 – Ad illustrazione di probabilità geometriche in 1 dimensione.

Si noti che P dipende solo dalla distanza dei punti M e N e non dalle loro posizioni.

Come secondo esempio si consideri un quadrato Q di lato d , sia Ω l'insieme dei punti di Q e A il sottoinsieme di Ω costituito dai punti del cerchio inscritto in Q (v. Figura 2.2). Se un esperimento casuale consiste nello scegliere a caso un punto in Ω , si calcoli la probabilità $P(A)$ che il punto scelto cada in A . Di nuovo ci si trova in presenza di uno spazio campione di cardinalità infinita; è tuttavia ragionevole procedere come nel caso precedente definendo questa volta $P(A)$ come rapporto di aree al seguente modo:

$$P(A) = \frac{\text{area di } A}{\text{area di } Q} = \frac{\pi d^2 / 4}{d^2} = \frac{\pi}{4} \approx 0.7853. \quad (2.13)$$

Si noti che poiché $P(A)$ non dipende da d , si può scegliere $d = 1$.

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 lunghezza di tale corda sia minore della lunghezza del lato del triangolo equilatero inscritto nella circonferenza. A seconda del criterio usato per individuare casi favorevoli e casi totali si ottengono risultati diversi. Invero, possiamo anzitutto ragionare al seguente modo. Detto r il raggio della circonferenza (v. Figura 2.3), ricordiamo che il lato del triangolo equilatero in essa inscritto ha lunghezza $\sqrt{3}r$ e che l'altezza di tale triangolo è $3r/2$. La distanza di ciascun lato del triangolo dal centro del cerchio è dunque $r/2$. Tracciata ora a caso una corda, la distanza d di questa dal centro del cerchio è un numero compreso tra 0 e r . La lunghezza l della corda è allora inferiore a $\sqrt{3}r$ se risulta $d > r/2$. Se quindi la probabilità $P(l < \sqrt{3}r)$ richiesta viene calcolata come rapporto tra la “lunghezza favorevole” $r/2$ e la “lunghezza possibile” r , si ottiene $1/2$. D’altro canto è lecito procedere anche in modo diverso. Precisamente, sempre con riferimento alla Figura 2.3, osserviamo che l’angolo al centro ϑ sotteso dalla corda è compreso tra 0 e π e che si ha $\vartheta = \alpha = 2\pi/3$ quando la lunghezza di questa è pari alla lunghezza del lato del triangolo. La lunghezza l della corda è allora inferiore a $\sqrt{3}r$ se risulta $\vartheta < 2\pi/3$. Calcolando la probabilità richiesta

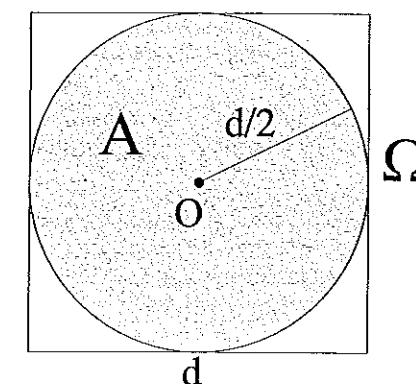


Figura 2.2 – Ad illustrazione di probabilità geometriche nel piano.

come rapporto tra “l’angolo al centro favorevole” $2\pi/3$ e “l’angolo al centro possibile” π si ottiene $P(l < \sqrt{3}r) = 2/3$, che è un risultato diverso da quello ottenuto prima.

Questo paradosso ha comunque una spiegazione abbastanza semplice: le diverse soluzioni ottenute sono in realtà soluzioni di *problemi diversi*. Ciò è dovuto alla circostanza che l’enunciato del problema non definisce in modo univoco cosa debba intendersi per “tracciare a caso una corda”.

La formula (2.13) suggerisce che la teoria della probabilità può essere utilizzata anche per effettuare calcoli relativi a problemi non connessi con fenomeni aleatori. Infatti, se è possibile determinare la frequenza $f(A) = \nu/n$ per n molto grande, dove n denota il numero di punti scelti a caso nel quadrato e ν il numero di quelli che giacciono anche all’interno del cerchio, si ha $\nu/n \approx P(A)$. Dalla (2.13) si ricava così una stima del numero π avendosi $\pi \approx 4\nu/n$; tale stima c’è da attendersi sia tanto migliore quanto più elevato è il numero dei punti scelti a caso, ossia quanto più grande è il numero delle “prove” effettuato. Simili procedimenti di stima, che rientrano nella categoria dei cosiddetti *metodi di Monte Carlo*, possono essere utilizzati per effettuare stime numeriche in numerosi contesti non probabilistici. Si voglia, ad esempio, calcolare l’integrale definito di una funzione di più variabili di forma molto complicata oppure di ignota espressione analitica. È noto che il calcolo a mezzo di tecniche numeriche è possibile con buona approssimazione fino a un numero di dimensioni non superiore a 5. Invece, qualunque sia il numero di dimensioni, per il calcolo di un tale integrale — che equivale al calcolo della misura di una certa regione A — ci si può ricondurre ad uno schema simile a quello usato nel caso del cerchio inscritto nel quadrato,

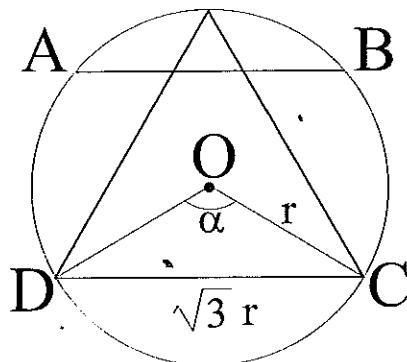


Figura 2.3 – Ad illustrazione del Paradosso di Bertrand.

considerando A come un sottoinsieme di un opportuno ipercubo \mathcal{I} e calcolando la frequenza data dal rapporto tra il numero di punti generati a caso che cadono in A e il numero totale di punti generati in \mathcal{I} .

Metodi di tipo Monte Carlo trovano attualmente crescente utilizzazione in problemi di grande interesse richiedenti, ad esempio, la risoluzione di complicate equazioni differenziali o di sistemi di equazioni algebriche di grandi dimensioni.

2.5 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 dei punti campione, 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 appare ragionevole ritenere che i tre possibili esiti (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, ecc.).

Proprio da considerazioni di questo tipo nasce l'impostazione "soggettiva" (o "personale") della probabilità. Questa, già accennata da Pascal ma più propriamente attribuibile a D. Bernoulli, è stata ripresa e sviluppata in tempi recenti soprattutto da Bruno de Finetti

2.5 Definizione soggettiva

e da Leonard Jimmie Savage. 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 2.1 (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. \triangleleft

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

La Definizione 2.1 trova un'immediata interpretazione quando ci si riferisce 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 vittoria 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 vittoria 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 vittoria 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) $sP(A)$ 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 $P(A) \leq 1$. 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, così che deve aversi $P(A) = 0$.

Dalla condizione di coerenza è possibile inoltre far discendere la legge di additività finita (2.9) delle probabilità. Siano invero A_1, A_2, \dots, A_n eventi necessari ed incompatibili, e 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 $\sum_{i=1}^n \mathbb{P}(A_i) < 1$ e negativo se $\sum_{i=1}^n \mathbb{P}(A_i) > 1$, 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, $\mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(\overline{A_1 \cup A_2}) = 1$. D'altra parte anche gli eventi $A_1 \cup A_2$ e $\overline{A_1 \cup A_2}$ sono necessari e incompatibili; quindi $\mathbb{P}(A_1 \cup A_2) + \mathbb{P}(\overline{A_1 \cup A_2}) = 1$. Queste due ultime uguaglianze implicano che $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2)$.

Nemmeno la definizione soggettiva è esente da critiche: solitamente l'individuo non è indifferente di fronte al rischio, tanto che in alcuni casi lo ricerca (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 funzione dell'individuo di riferimento: la probabilità di un dato evento può invece risultare notevolmente diversa per persone diverse, riflettendo le differenti informazioni, esperienze e atteggiamenti su cui esse basano le proprie convinzioni.

CAPITOLO 3

La Costruzione Assiomatica

3.1 La struttura degli eventi

Il grande sviluppo che il calcolo delle probabilità ha avuto, e che a tutt'oggi ancora conosce con l'utilizzazione in nuovi filoni di ricerca sia applicativi che teorici, è dovuto alla rigorosa formulazione datane nel 1933 da Kolmogorov. In questo capitolo presenteremo i suoi elementi essenziali: collezione degli eventi e misura delle loro probabilità. Seguirà anche l'esposizione di una serie di risultati che rappresentano il patrimonio di conoscenze su cui fondare i successivi sviluppi ed approfondimenti.

Un primo passo verso la teoria assiomatica è già stato presentato nel § 1.3 nel quale la nozione di *evento* ha trovato una rappresentazione matematica come sottoinsieme dello spazio campione Ω che è l'insieme dei possibili risultati di un esperimento casuale \mathcal{E} . Ovviamente ciò non esclude la possibilità di contesti che sfuggono da questa visualizzazione; né che si è affermato qualcosa riguardante la natura degli eventi. Ciò nondimeno una tale rappresentazione è applicabile nella stragrande maggioranza delle situazioni d'interesse concreto; inoltre, l'aspetto negativo consistente nel venire così meno considerazioni legate all'intuizione, viene largamente compensato dalla possibilità di utilizzare la teoria degli insiemi e di tutte le costruzioni matematiche ottenute da essa, con particolare riferimento alle teorie della misura e dell'integrazione.

Sempre nel § 1.3 è stata proposta una trasposizione in termini insiemistici della logica degli eventi. Ad esempio, si è già sottolineato che, assegnato un evento A , potrebbe risultare d'interesse anche l'evento che si verifica quando A non si verifica, ossia la negazione di A . Rappresentando l'evento A come insieme possiamo quindi affermare che siamo interessati anche ad \overline{A} (il complemento di A rispetto a Ω). In tale ottica, come si è visto, l'unione di due insiemi rappresentanti rispettivamente gli eventi A e B è la trasposizione dell'evento che si verifica se si verifica almeno uno degli eventi A e B . Allo stesso modo l'intersezione rappresenta l'evento che si verifica se entrambi gli eventi A e B si verificano. Dalla logica degli eventi scaturisce quindi la necessità di riguardare \overline{A} , $A \cup B$ e $A \cap B$ come eventi. È

utile, a questo punto, richiamare la struttura algebrica che caratterizza proprietà di un siffatto tipo.

Definizione 3.1. Una famiglia \mathcal{A} di parti di un insieme Ω tale che

- (i) $\Omega \in \mathcal{A}$,
- (ii) $A \in \mathcal{A} \Rightarrow \bar{A} \in \mathcal{A}$,
- (iii) $A \in \mathcal{A}, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$.

costituisce un'algebra (o un campo) su Ω . \triangleleft

Algebra su Ω sono ad esempio le seguenti:

$$\mathcal{A}_1 = \{\emptyset, \Omega\}, \quad \mathcal{A}_2 = \{\emptyset, A, \bar{A}, \Omega\}, \quad \mathcal{A}_3 = \mathcal{P}(\Omega),$$

dove A è un sottoinsieme di Ω e $\mathcal{P}(\Omega)$ indica l'insieme delle parti di Ω . Se \mathcal{A} è una qualunque algebra su Ω si ha poi evidentemente:

$$\mathcal{A}_1 = \{\emptyset, \Omega\} \subseteq \mathcal{A} \subseteq \mathcal{A}_3 = \mathcal{P}(\Omega).$$

Dalla Definizione 3.1 scaturiscono alcune immediate conseguenze: dalle (i) e (ii) segue in primo luogo che anche $\emptyset \in \mathcal{A}$; inoltre, dalla (iii) e dalla proprietà associativa dell'unione, segue che anche l'unione di un qualsiasi numero finito di elementi di \mathcal{A} appartiene a \mathcal{A} ; infine, da quest'ultima, dalla (iii) e dalla formula di De Morgan segue che anche l'intersezione di un qualsiasi numero finito di elementi di \mathcal{A} appartiene a \mathcal{A} .

Giova osservare esplicitamente che la chiusura di una classe rispetto ad un'operazione non sempre resta soddisfatta con il passaggio dal finito al numerabile. Per rendersene conto si faccia ad esempio riferimento alla famiglia \mathcal{J} costituita dagli intervalli del tipo $I_n = (1/n, 1]$ per ogni $n \in \mathbb{N}$.¹ Risulta di immediata verifica che l'unione di un numero finito di elementi di \mathcal{J} è uguale all'intervallo che possiede il pedice massimo e pertanto appartiene a \mathcal{J} . D'altra parte, indicata con $S = \{y_n\}_{n \in \mathbb{N}}$ una qualsiasi successione di interi positivi, risulta $\bigcup_{y_n \in S} I_{y_n} = (0, 1] \notin \mathcal{J}$.

Per far sì che la proprietà di chiusura sussista anche nel caso di unioni numerabili, si rende necessario il richiederlo esplicitamente. Ciò conduce alla seguente

Definizione 3.2. Una famiglia \mathcal{F} di parti di un insieme Ω tale che

- (i) $\Omega \in \mathcal{F}$;
- (ii) $A \in \mathcal{F} \Rightarrow \bar{A} \in \mathcal{F}$;

¹Con \mathbb{N} denotiamo l'insieme dei numeri interi positivi.

3.1 La struttura degli eventi

$$(iii) \forall n \in \mathbb{N}, A_n \in \mathcal{F} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F},$$

costituisce una σ -algebra (o un σ -campo) su Ω . \triangleleft

Dalla Definizione 3.2 segue che una σ -algebra è anche un'algebra, mentre vale anche il viceversa solo se Ω è finito. Inoltre, l'intersezione di σ -algebra è essa stessa una σ -algebra. Le dimostrazioni, semplici, di queste affermazioni sono lasciate al lettore.

D'ora innanzi chiameremo *eventi* di Ω soltanto gli elementi di una σ -algebra \mathcal{F} per essere certi che, operando su di essi mediante il complemento e l'unione finita o numerabile, ci si mantenga nella classe degli eventi. Inoltre, la σ -algebra \mathcal{F} sarà costruita a partire da opportuni sottoinsiemi di Ω (ai quali risulti possibile associare, in maniera diretta, una qualche "misura" che in seguito sarà interpretata come probabilità di tali eventi). Più precisamente introduciamo la seguente definizione.

Definizione 3.3. Una σ -algebra \mathcal{F} su Ω si dice generata da una famiglia \mathcal{G} di sottoinsiemi di Ω , e si designa con il simbolo $\sigma(\mathcal{G})$, se ogni σ -algebra su Ω contenente \mathcal{G} contiene \mathcal{F} . Si può anche dire che $\sigma(\mathcal{G})$ è la minima σ -algebra su Ω contenente \mathcal{G} . Gli elementi di \mathcal{G} sono detti eventi generatori. \triangleleft

Teorema 3.1. Data una qualsiasi famiglia \mathcal{G} di sottoinsiemi di Ω , esiste ed è unica la σ -algebra da essa generata.

Dimostrazione. Ricordiamo che l'intersezione di σ -algebra su Ω è una σ -algebra su Ω ; inoltre, la famiglia di tutte le σ -algebra su Ω è non vuota in quanto ad essa appartiene certamente $\mathcal{P}(\Omega)$. La σ -algebra $\sigma(\mathcal{G})$ coincide allora con la σ -algebra che si ottiene mediante l'intersezione di tutti gli elementi di tale famiglia. Questo dimostra la parte relativa all'esistenza. L'unicità è una conseguenza immediata della definizione di σ -algebra generata. Infatti, se esistessero due σ -algebra generate da \mathcal{G} , diciamo \mathcal{F}_1 e \mathcal{F}_2 , entrambe dovrebbero essere contenute in tutte le σ -algebra contenenti \mathcal{G} . In particolare, dovrebbe allora risultare $\mathcal{F}_1 \supseteq \mathcal{F}_2$ e $\mathcal{F}_2 \supseteq \mathcal{F}_1$ da cui scaturisce $\mathcal{F}_1 = \mathcal{F}_2$. \square

Osservazione 3.1. Quanto detto a proposito di σ -algebra generata da una famiglia \mathcal{G} di sottoinsiemi di Ω si trasferisce in maniera immediata al concetto di algebre generata da \mathcal{G} e che, nel seguito sarà indicata con il simbolo $a(\mathcal{G})$. \square

Nella definizione seguente vengono riassunti i fatti essenziali fin qui esposti.

Definizione 3.4. Si dice spazio probabilizzabile ogni coppia (Ω, \mathcal{F}) , dove Ω è uno spazio campione e \mathcal{F} è una σ -algebra generata da una famiglia \mathcal{G} di sottoinsiemi di Ω . Gli elementi di \mathcal{F} sono chiamati eventi; in particolare Ω è l'evento certo e \emptyset è l'evento impossibile. \square

Seguiranno ora alcuni esempi utili ad illustrare le definizioni date sopra e quelle introdotte nel precedente § 1.3; essi forniscono al contempo spunti per ulteriori approfondimenti essenziali per il successivo sviluppo della trattazione. Ad essi risulta conveniente premettere la definizione degli eventi elementari.

Definizione 3.5. Sia (Ω, \mathcal{F}) uno spazio probabilizzabile e sia E un evento diverso dall'evento impossibile. E è detto essere un evento elementare se la sua intersezione con un altro qualsiasi evento o lo lascia inalterato oppure fornisce l'evento impossibile. \triangleleft

In altri termini un evento elementare è “indivisibile” nel senso che esso non può essere ottenuto come unione di altri eventi di Ω .

Esempio 3.1. Sia $\Omega = \{1, 2, 3, 4, 5, 6\}$ l'insieme dei risultati ottenibili mediante l'esperimento del lancio di un dado. Si supponga di avere interesse a considerare come evento ogni singolo possibile esito dell'esperimento. In termini insiemistici, la famiglia di tali eventi generatori è allora:

$$\mathcal{G}_1 = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}.$$

Applicando le operazioni di complemento, unione e intersezione agli elementi di \mathcal{G}_1 si verifica (si lascia al lettore) che la σ -algebra \mathcal{F}_1 così generata coincide con $\mathcal{P}(\Omega)$, ossia con l'insieme delle $2^6 = 64$ parti di Ω . Due osservazioni concludono questo esempio.

1. Un qualsiasi evento appartenente a \mathcal{G}_1 è un evento elementare. Quindi, nella situazione in esame gli eventi generatori coincidono con gli eventi elementari.
2. La famiglia di tutti gli eventi elementari forma una partizione (finita) di Ω ; per ottenerne \mathcal{F}_1 basta unire, in tutti i modi possibili, tali eventi, senza dover ricorrere ad operazioni di intersezione e complemento, con l'eccezione dell'evento impossibile che si ottiene come complemento di Ω . \triangleleft

Esempio 3.2. Sia Ω come nell'Esempio 3.1 e si supponga di avere interesse a considerare come eventi le seguenti due proposizioni: “il risultato è un numero dispari” e “il risultato è un numero minore di 3”. In termini insiemistici, la famiglia di eventi generatori siffatti, è allora:

$$\mathcal{G}_2 = \{\{1, 2\}, \{1, 3, 5\}\}.$$

Indichiamo con A e B gli eventi $\{1, 3, 5\}$ e $\{1, 2\}$, rispettivamente. Per costruire la σ -algebra generata da \mathcal{G}_2 bisogna in primo luogo aggiungere agli elementi di \mathcal{G}_2 gli insiemi $A \cup B$, $A \cap B$, \bar{A} e \bar{B} . Abbiamo in tal modo ottenuto una classe di sottoinsiemi di Ω

$$\mathcal{C} = \{\{1, 2\}, \{1, 3, 5\}, \{1, 2, 3, 5\}, \{1\}, \{3, 4, 5, 6\}, \{2, 4, 6\}\}$$

che non costituisce un'algebra (ad esempio $\{1, 2, 3, 5\} \cap \{2, 4, 6\} = \{2\} \notin \mathcal{C}$). Bisogna infatti aggiungere le unioni che mancano, poi le negazioni e così via. Il procedimento avrà certamente termine dal momento che la potenza dell'insieme delle parti dello spazio campione è

finita. Procedendo in questa maniera si ottiene una classe \mathcal{F}_2 costituita da 16 eventi (si lascia al lettore) chiusa rispetto all'unione, all'intersezione ed al complemento, e quindi un'algebra su Ω che, data la finitezza di Ω , risulta essere la σ -algebra richiesta. Possiamo quindi operare su questa classe, invece che sulla classe di tutti i sottoinsiemi di Ω , ottenendo una notevole economia avendo a che fare con 16 eventi invece che con 64. \triangleleft

Esempio 3.3. Proseguiamo con l'Esempio 3.2 e indichiamo, rispettivamente, con G_1 e G_2 i due eventi di \mathcal{G}_2 . Il lettore che ha costruito la classe \mathcal{F}_2 si sarà certamente reso conto delle difficoltà operative insite in tale procedimento. Ricordando la seconda osservazione dell'Esempio 3.1, la σ -algebra \mathcal{F}_2 generata da \mathcal{G}_2 , può, invece, essere ottenuta più rapidamente se si passa attraverso gli eventi elementari generati da \mathcal{G}_2 . Questi si ottengono eseguendo tutte le intersezioni del tipo $G_1^* \cap G_2^*$ dove, per $k = 1$ e $k = 2$, con G_k^* si è indicato l'evento G_k oppure il suo complemento:

$$\begin{aligned} E_1 &= G_1 \cap G_2 = \{1\}; & E_2 &= G_1 \cap \bar{G}_2 = \{2\}; \\ E_3 &= \bar{G}_1 \cap G_2 = \{3, 5\}; & E_4 &= \bar{G}_1 \cap \bar{G}_2 = \{4, 6\}. \end{aligned}$$

La famiglia di tutti gli eventi elementari così determinata forma una partizione (finita) di Ω ; per ottenerne \mathcal{F}_2 basta unire, in tutti i modi possibili, tali eventi. Come prima, l'evento impossibile si ottiene come complemento di Ω . Va sottolineato che in questo specifico contesto la famiglia degli eventi generatori \mathcal{G}_2 è diversa dalla famiglia degli eventi elementari $\mathcal{E} = \{E_1, E_2, E_3, E_4\}$ e che, ad esempio, $\{3\}$ non è un evento in quanto non è ottenibile mediante unioni di elementi di \mathcal{E} . \triangleleft

Osservazione 3.2. Il procedimento per ottenere la famiglia \mathcal{E} degli eventi elementari può essere generalizzato a qualsiasi cardinalità n finita della famiglia $\mathcal{G} = \{G_k : k = 1, 2, \dots, n\}$ di generatori. Il generico evento elementare E è del tipo $G_1^* \cap G_2^* \dots \cap G_n^*$ dove, per $i = 1, 2, \dots, n$, con G_k^* si è indicato l'evento G_k oppure il suo complemento. Inoltre, dal momento che nella formula per la costruzione del generico evento elementare E compaiono n operandi, ognuno dei quali può essere scelto in due modi diversi, la cardinalità della famiglia \mathcal{E} risulta essere 2^n . Nulla vieta che uno o più di tali eventi possa essere l'evento impossibile. \triangleleft

La giustificazione generale e rigorosa dell'equivalenza delle procedure indicate negli Esempi 3.2 e 3.3 per la costruzione di $\sigma(\mathcal{G}_2)$ (la seconda delle quali, come si è sottolineato, è molto meno laboriosa) poggia sul seguente

Teorema 3.2. Assegnata una famiglia $\mathcal{G} = \{G_k : k = 1, 2, \dots, n\}$ a cardinalità finita di eventi generatori, si denoti con $\mathcal{E} = \{E_k : k = 1, 2, \dots, 2^n\}$ la famiglia degli eventi elementari da essa ottenibile con il procedimento esposto nella Osservazione 3.2. Si ha $\sigma(\mathcal{G}) = \sigma(\mathcal{E})$.

Dimostrazione. Un qualsiasi evento elementare è stato ottenuto dagli eventi generatori mediante operazioni di intersezione e di complementazione; pertanto $\sigma(\mathcal{G}) \supseteq \mathcal{E}$. Da quest'ultima, e dalla Definizione 3.3 di σ -algebra generata da una famiglia, segue immediatamente che $\sigma(\mathcal{G}) \supseteq \sigma(\mathcal{E})$. Viceversa, considerato un qualsiasi evento generatore G_j , applicando $n - 1$ volte la formula della partizione con l'evento certo possiamo scrivere:

$$G_j = \bigcup_{i \neq j, i=1}^n (G_k \cup \bar{G}_k) \cap G_j.$$

Utilizzando in quest'ultima la proprietà distributiva dell'intersezione rispetto all'unione, segue che un qualsiasi evento generatore è ottenibile mediante unioni di eventi elementari; pertanto $\sigma(\mathcal{E}) \supseteq \mathcal{G}$ e $\sigma(\mathcal{E}) \supseteq \sigma(\mathcal{G})$. \square

Esempio 3.4 (Spazio campione di Bernoulli). Consideriamo l'esperimento casuale che, per semplicità di rappresentazione, assumeremo modellizzato dal lancio indefinitamente ripetuto di una moneta "equa", ossia non truccata. A questo esperimento si dà il nome di *processo di Bernoulli*. Lo spazio campione Ω è costituito da tutte le successioni di teste e croci (T e C nel seguito); esso prende il nome di *spazio di Bernoulli*. In tale spazio siamo interessati a considerare come evento, per ogni $n \in \mathbb{N}$, la proposizione T_n : "al lancio n -esimo esce testa" che assumeremo costituire collettivamente la famiglia \mathcal{G}_1 degli eventi generatori. Per ogni $n \in \mathbb{N}$, l'evento T_n ha come negazione l'evento C_n : "all' n -esimo lancio esce croce". In termini insiemistici, T_n è l'insieme di tutte le successioni di T e C aventi T nella posizione n -esima, mentre C_n è l'insieme di tutte le successioni di T e C aventi C nella posizione n -esima. Si noti che T_n e C_n non sono elementi di Ω , ma sottoinsiemi di Ω . Inoltre, mentre risulta immediato affermare che la negazione della proposizione T_n è la proposizione C_n , risulta più difficoltoso verificare in termini insiemistici che $\bar{T}_n = C_n$ (in quanto i due sottoinsiemi considerati sono entrambi infiniti). Lo stesso vale per la negazione della proposizione C_n che è la proposizione T_n .

Da quanto detto in precedenza, si possono considerare come *eventi dello spazio di Bernoulli* tutti i sottoinsiemi di Ω che si ottengono a partire da tali eventi generatori per complemento e per unioni e intersezioni finite o numerabili. Tali eventi costituiscono la σ -algebra \mathcal{F}_1 generata da \mathcal{G}_1 . In particolare, si dimostra agevolmente che ogni singoletto di Ω è un evento.

A titolo di esempio, mostriamo come può esprimersi mediante gli eventi generatori la seguente proposizione A : "due teste consecutive si presentano prima di due croci consecutive". Si noti come sia la natura stessa dell'evento A a richiedere la non limitatezza del numero delle ripetizioni del lancio della moneta: si continuerà a lanciare la moneta fino a quando si verificano consecutivamente due teste oppure due croci. Inoltre, dal momento che l'evento A può realizzarsi solo se la sequenza TT si presenta prima della sequenza CC , risulta conveniente introdurre, per ogni $n \in \mathbb{N}$ e maggiore di 1, l'evento A_n : " TT si verifica

3.1 La struttura degli eventi

per la prima volta all' n -esimo lancio e CC non si è ancora verificata". Il generico evento A_n è esprimibile in termini di eventi generatori. Ad esempio, per i primi cinque di essi, si ha:

$$\begin{aligned} A_2 &= T_1 \cap T_2; & A_3 &= C_1 \cap T_2 \cap T_3; \\ A_4 &= T_1 \cap C_2 \cap T_3 \cap T_4; & A_5 &= C_1 \cap T_2 \cap C_3 \cap T_4 \cap T_5; \\ A_6 &= T_1 \cap C_2 \cap T_3 \cap C_4 \cap T_5 \cap T_6. \end{aligned}$$

L'evento A si verifica quando si verifica uno qualsiasi di essi, così che risulta:

$$A = \bigcup_{n=2}^{\infty} A_n. \quad (3.1)$$

L'evento A è quindi stato rappresentato mediante unione numerabile di eventi (che risultano anche a due a due incompatibili) ciascuno dei quali è espresso mediante intersezione di eventi generatori. \triangleleft

Esempio 3.5. Continuiamo a considerare l'esperimento casuale del processo di Bernoulli definito nell'Esempio 3.4. In esso, per scopi applicativi facilmente intuibili, rivestono particolare importanza, per ogni $n \in \mathbb{N}$, gli eventi E_n : "testa si verifica per la prima volta all' n -esimo lancio" che rappresentano il *ritardo* o l'*attesa* di T . Essi sono a due a due incompatibili ma non sono necessari. Infatti, in essi rientrano tutti i punti campione di Ω (successioni di T e C) che contengono almeno una T , mentre resta fuori la successione costituita da sole C . Indicato allora con E_0 l'evento che contiene solo tale punto campione, si ha che la famiglia $\mathcal{E}_2 = \{E_n : n \in \mathbb{N}_0\}$ costituisce una partizione numerabile di Ω ; in base alla Definizione 3.5 gli elementi di \mathcal{E}_2 sono eventi elementari e pertanto la σ -algebra \mathcal{F}_2 generata da \mathcal{E}_2 è costituita da tutte le unioni finite o numerabili di elementi di \mathcal{E}_2 .

Ricordando il significato del simbolo T_n di cui all'Esempio 3.4, risulta di immediata verifica che

$$E_n = \bar{T}_1 \cap \bar{T}_2 \cap \cdots \cap \bar{T}_{n-1} \cap T_n, \quad (n = 1, 2, \dots).$$

Quindi $\mathcal{F}_2 \subseteq \mathcal{F}_1$. Non sussiste invece il viceversa. Consideriamo, infatti, il sottoinsieme T_2 di Ω costituito da tutte le successioni di T e C aventi T al secondo posto. Si tratta di un evento dello spazio di Bernoulli (in altri termini $T_2 \in \mathcal{F}_1$) in quanto esso è uno degli elementi generatori di \mathcal{F}_1 . Non è invece un elemento di \mathcal{F}_2 in quanto esso è dato dall'unione di E_2 con quella parte propria di E_1 costituita da tutte le successioni di T e C aventi T nel primo e nel secondo posto. Essendo evento elementare, E_1 non può essere decomponibile in termini di eventi, e quindi T_2 non può essere espresso in termini di eventi elementari della famiglia \mathcal{E}_2 . In definitiva risulta $\mathcal{F}_2 \subset \mathcal{F}_1$. \triangleleft

Esempio 3.6 (σ -algebra di Borel di \mathbb{B}). Consideriamo come spazio campione l'insieme \mathbb{R} dei numeri reali e sia \mathcal{G}_1 la famiglia degli aperti della topologia naturale di \mathbb{R} . La σ -algebra generata da \mathcal{G}_1 , $\sigma(\mathcal{G}_1)$, che solitamente si indica con $\mathcal{B}_{\mathbb{R}} \equiv \mathcal{B}$, prende il nome di *σ -algebra*

di Borel di \mathbb{R} , oppure \mathbb{R}; i suoi elementi sono detti . La σ -algebra \mathcal{B} si ottiene componendo gli aperti mediante le operazioni di complemento e di unione finita o numerabile. Si costruiscono in tal modo tutti gli insiemi che si considerano usualmente nell'analisi matematica e, in particolare, tutti gli intervalli aperti, chiusi, semiaperti a destra o a sinistra, finiti o infiniti, e tutti i numeri reali.

Indicata con \mathcal{G}_2 la famiglia delle semirette sinistre chiuse, faremo ora vedere che risulta $\sigma(\mathcal{G}_2) = \sigma(\mathcal{G}_1) \equiv \mathcal{B}$.

Cominciamo col dimostrare in primo luogo che $\sigma(\mathcal{G}_2) \subseteq \sigma(\mathcal{G}_1)$. A tale scopo osserviamo che una qualsiasi semiretta sinistra chiusa appartiene a $\sigma(\mathcal{G}_1)$ in quanto complementare di un aperto. Allora $\mathcal{G}_2 \subseteq \sigma(\mathcal{G}_1)$ e ciò, per la Definizione 3.3 di σ -algebra generata da una famiglia di insiemi, comporta anche che $\sigma(\mathcal{G}_2) \subseteq \sigma(\mathcal{G}_1)$.

Dimostriamo ora il viceversa; ossia che $\sigma(\mathcal{G}_1) \subseteq \sigma(\mathcal{G}_2)$. Ragionando in maniera analoga, basterà mostrare che un qualsiasi insieme aperto appartiene a $\sigma(\mathcal{G}_2)$, ossia che $\mathcal{G}_1 \subseteq \sigma(\mathcal{G}_2)$. Indichiamo all'uopo con G un aperto qualsiasi della topologia naturale di \mathbb{R} . Esso è costituito solo da punti interni. Per il generico punto x di G esiste, pertanto, un numero reale $\delta > 0$ tale che l'intorno circolare $|x - \delta, x + \delta| \subseteq G$. Per la proprietà di densità dei numeri razionali in \mathbb{R} è possibile individuare due numeri razionali r ed s , con $r < s$, tali che $x \in]r, s] \subseteq G$. L'insieme \mathbb{Q} dei razionali è numerabile, così come $\mathbb{Q} \times \mathbb{Q}$; quindi la famiglia degli intervalli $]r, s]$ contenuti in G è essa stessa numerabile. È lecito, pertanto, rappresentare tale famiglia con la successione $\{]r_n, s_n]\}_{n \in \mathbb{N}}$ e poiché, come già osservato, ogni $x \in G$ appartiene ad almeno uno di questi intervalli, si ha:

$$G = \bigcup_{n=1}^{\infty}]r_n, s_n] = \bigcup_{n=1}^{\infty} (-\infty, s_n] \cap]r_n, +\infty[.$$

In definitiva G è stato espresso come unione numerabile di insiemi che appartengono a $\sigma(\mathcal{G}_2)$ (intersezione di una semiretta sinistra chiusa con il complemento di un'altra semiretta sinistra chiusa), cosicché appartiene esso stesso a $\sigma(\mathcal{G}_2)$, il che completa la dimostrazione. \square

Si lascia al lettore la verifica che \mathcal{B} è generata anche dalle seguenti famiglie di sottoinsiemi di \mathbb{R} :

$$\mathcal{G}_3 = \{]r, s] : r \in \mathbb{R}, s \in \mathbb{R}\}; \quad \mathcal{G}_4 = \{[r, s] : r \in \mathbb{R}, s \in \mathbb{R}\}; \quad \mathcal{G}_5 = \{(-\infty, r] : r \in \mathbb{R}\};$$

$$\mathcal{G}_6 = \{]r, s] : r \in \mathbb{R}, s \in \mathbb{R}\}; \quad \mathcal{G}_7 = \{[r, s] : r \in \mathbb{R}, s \in \mathbb{R}\}; \quad \mathcal{G}_8 = \{[s, +\infty[: s \in \mathbb{R}\};$$

$$\mathcal{G}_9 = \{[s, +\infty[: s \in \mathbb{R}\}.$$

Nella definizione delle famiglie da \mathcal{G}_2 a \mathcal{G}_9 si può anche permettere agli estremi degli intervalli considerati di essere elementi di \mathbb{Q} ; ad esempio alla famiglia \mathcal{G}_3 si può sostituire la famiglia $\mathcal{G}'_3 = \{]r, s] : r \in \mathbb{Q}, s \in \mathbb{Q}\}$. \square

3.2 Successioni e loro limiti

Esempio 3.7 (σ -algebra di Borel di $[a, b]$). Siano a e b numeri reali con $a < b$. La σ -algebra generata dagli aperti della topologia naturale dell'intervallo $[a, b]$ prende il nome *σ -algebra di Borel di $[a, b]$* e si indica con il simbolo $\mathcal{B}_{[a, b]}$. Come nell'esempio precedente, si può dimostrare che essa è generata anche dalla famiglia degli intervalli del tipo $]r, s]$ con $a \leq r < s \leq b$. \square

Osservazione 3.3. Sia Ω un insieme e \mathcal{A} una sua σ -algebra. Assegnato un sottoinsieme A di Ω si designi con $\mathcal{A} \cap A$ la famiglia di parti di Ω che si ottiene intersecando ogni elemento di \mathcal{A} con A . Si lascia al lettore verificare che $\mathcal{A} \cap A$ è una σ -algebra su A : essa prende il nome di *traccia di \mathcal{A} su A* e si designa anche con il simbolo $\text{tr}(\mathcal{A}, A)$. Risultando, con $a \leq r < s \leq b$, $]r, s] \cap [a, b] =]r, s]$, relativamente alla σ -algebra di Borel dell'intervallo $[a, b]$ possiamo allora affermare che essa è la traccia della σ -algebra di Borel di \mathbb{R} su tale intervallo: $\mathcal{B}_{[a, b]} = \text{tr}(\mathcal{B}, [a, b])$. \square

Prima di concludere questo paragrafo è utile soffermarsi l'attenzione su alcune importanti considerazioni.

a) Nella Definizione 3.1 è presente la proprietà (i). Questa potrebbe omettersi a patto di aggiungere che \mathcal{A} è una famiglia *non vuota* di parti di Ω . Infatti se \mathcal{A} è non vuota, detto \bar{A} un suo elemento, dalla proprietà (ii) segue che $\bar{A} \in \mathcal{A}$ e dalla proprietà (iii) segue che $\bar{A} \cup \bar{A} = \Omega \in \mathcal{A}$.

b) Analoga considerazione vale per la Definizione 3.2: la proprietà (i) si potrebbe eliminare a patto di specificare che \mathcal{F} è una famiglia *non vuota* di parti di Ω .

c) In talune trattazioni della teoria della probabilità nella Definizione 3.2 compare la seguente quarta proprietà:

(iv) se $A_n \in \mathcal{F}$ per $n = 1, 2, \dots, k$ allora $A_1 \cup A_2 \cup \dots \cup A_k \in \mathcal{F}$,

esprimente la chiusura di \mathcal{F} rispetto all'unione finita. In realtà la proprietà (iv) discende dalla proprietà (iii) della Definizione 3.2 al seguente modo. Dati k elementi di \mathcal{F} , che denoteremo con A_1, A_2, \dots, A_k , si consideri la successione di insiemi $\{B_n\}_{n \in \mathbb{N}}$, dove $B_n = A_n$ ($n = 1, 2, \dots, k$) mentre $B_{k+1} = B_{k+2} = \dots = \emptyset$. Si osservi che per ogni $n \in \mathbb{N}$ risulta $B_n \in \mathcal{F}$. Infatti, per ipotesi $A_n \in \mathcal{F}$ per $(n = 1, 2, \dots, n)$; inoltre $\emptyset \in \mathcal{F}$ come già fatto osservare in precedenza. Dalla proprietà (iii) si trae infine $\bigcup_{k=1}^n A_k = \bigcup_{n=1}^{\infty} B_n \in \mathcal{F}$.

3.2 Successioni e loro limiti

Nei prossimi paragrafi presenteremo due risultati tra i più importanti nell'impianto di base del Calcolo delle Probabilità: il Teorema di Equivalenza ed il Teorema di Estensione. Entrambi i teoremi, o già nell'enunciato, o nel corso della dimostrazione che sarà qui proposta,

fanno uso del concetto di successione di eventi e del relativo concetto di limite. Riteniamo pertanto utile richiamare definizioni e risultati concernenti le successioni di sottoinsiemi di un generico spazio Ω e parallelamente riguardare Ω come lo spazio campione di un opportuno esperimento casuale e i sottoinsiemi costituenti la successione appartenenti ad una σ -algebra \mathcal{F} di parti di Ω , ed illustrare le relative interpretazioni nella logica degli eventi.

Definizione 3.6. Consideriamo una successione $\{C_n\}_{n \in \mathbb{N}}$ di parti di Ω ; essa si dice crescente se, $\forall n \in \mathbb{N}$, risulta $C_n \subseteq C_{n+1}$. Per una tale successione, l'insieme, che si designa con il simbolo $\lim_{n \rightarrow \infty} C_n$, costituito dai punti di Ω che appartengono ad almeno uno dei suoi elementi, viene detto esserne il limite. \square

Ricordando la definizione dell'unione insiemistica possiamo anche scrivere:

$$\lim_{n \rightarrow \infty} C_n = \bigcup_{n=1}^{\infty} C_n. \quad (3.2)$$

In virtù di quest'ultima è abbastanza agevole rendersi conto che il limite di una successione crescente di eventi è quell'evento che si verifica quando si verifica almeno uno degli eventi che la costituiscono.

Definizione 3.7. Consideriamo una successione $\{D_n\}_{n \in \mathbb{N}}$ di parti di Ω ; essa si dice decrescente sse, $\forall n \in \mathbb{N}$, risulta $D_n \supseteq D_{n+1}$. Per una tale successione, l'insieme, che si designa con il simbolo $\lim_{n \rightarrow \infty} D_n$, costituito dai punti di Ω che appartengono a tutti i suoi elementi, viene detto esserne il limite. \square

Ricordando la definizione dell'intersezione insiemistica possiamo anche scrivere:

$$\lim_{n \rightarrow \infty} D_n = \bigcap_{n=1}^{\infty} D_n. \quad (3.3)$$

Tramite quest'ultima è abbastanza agevole rendersi conto che il limite di una successione decrescente di eventi è quell'evento che si verifica quando si verificano tutti gli eventi che la costituiscono.

Definizione 3.8. Una successione di parti di Ω crescente o decrescente viene detta monotona. \square

Esempio 3.8. Nello spazio $\Omega = [-1, 1]$ consideriamo la successione $\{C_n\}_{n \in \mathbb{N}}$ con termine generale C_n coincidente con l'intervallo $(-1 + 1/n, 1 - 1/n)$. In primo luogo, è del tutto evidente che essa è crescente. Inoltre, applicando la (3.2), si ha:

$$\lim_{n \rightarrow \infty} C_n = \bigcup_{n=1}^{\infty} \left(-1 + \frac{1}{n}, 1 - \frac{1}{n} \right) = (-1, 1).$$

Invece, per la successione di termine generale $D_n = (-1/n, 1/n)$, che è immediato verificare essere decrescente, la (3.3) fa ottenere:

$$\lim_{n \rightarrow \infty} D_n = \bigcap_{n=1}^{\infty} \left(-\frac{1}{n}, \frac{1}{n} \right) = \{0\}. \quad \square$$

Definizione 3.9. Consideriamo una generica successione $\{A_n\}_{n \in \mathbb{N}}$ di parti di Ω . Si definisce limite inferiore di tale successione l'insieme, che si designa con $\liminf_{n \rightarrow \infty} A_n$, costituito dai punti di Ω che appartengono definitivamente alla successione. In simboli:

$$\liminf_{n \rightarrow \infty} A_n = \{\omega \in \Omega : \text{esiste un } n \in \mathbb{N} \text{ per il quale } \omega \in A_k \text{ per ogni } k \geq n\}. \quad (3.4)$$

A partire dalla successione di termine generale A_n possiamo definire una successione crescente ponendo:

$$C_n = \bigcap_{k=n}^{\infty} A_k, \quad \forall n \in \mathbb{N}.$$

Infatti, $C_n = A_n \cap C_{n+1}$, e quindi $C_n \subseteq C_{n+1}$. Osservando che un punto di Ω appartiene a C_n se esso appartiene a A_k per ogni $k \geq n$, e ricordando la (3.2) valida per le successioni crescenti, possiamo affermare che il limite inferiore della successione di termine generale A_n può essere espresso in uno dei seguenti modi:

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} C_n = \lim_{n \rightarrow \infty} C_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \lim_{n \rightarrow \infty} \bigcap_{k=n}^{\infty} A_k. \quad (3.5)$$

Alla luce di quest'ultima possiamo quindi interpretare il limite inferiore di una successione di eventi come quell'evento che si verifica se si verificano tutti gli eventi della successione fatta al più eccezione per un numero finito di essi.

Definizione 3.10. Consideriamo una generica successione $\{A_n\}_{n \in \mathbb{N}}$ di parti di Ω . Si definisce limite superiore di tale successione l'insieme, che si designa con $\limsup_{n \rightarrow \infty} A_n$, costituito dai punti di Ω che appartengono a infiniti elementi della successione. In simboli:

$$\limsup_{n \rightarrow \infty} A_n = \{\omega \in \Omega : \text{per ogni } n \in \mathbb{N} \text{ esiste un } k \geq n \text{ per il quale } \omega \in A_k\}. \quad (3.6)$$

A partire dalla successione di termine generale A_n possiamo definire una successione decrescente ponendo:

$$D_n = \bigcup_{k=n}^{\infty} A_k, \quad \forall n \in \mathbb{N}.$$

Infatti, $D_n = A_n \cup D_{n+1}$, e quindi $D_n \supseteq D_{n+1}$. Osservando che un punto di Ω appartiene a D_n se esso appartiene a A_k per almeno un $k \geq n$, e ricordando la (3.3) valida per le successioni decrescenti, possiamo affermare che il limite superiore della successione di termine generale A_n può essere espresso in uno dei seguenti modi:

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} D_n = \lim_{n \rightarrow \infty} D_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \lim_{n \rightarrow \infty} \bigcup_{k=n}^{\infty} A_k. \quad (3.7)$$

In base all'ultima uguaglianza della (3.7) possiamo quindi interpretare il limite superiore di una successione di eventi come quell'evento che si verifica se si verificano infiniti eventi della successione.

✓ **Proposizione 3.1.** *Sia assegnata una generica successione $\{A_n\}_{n \in \mathbb{N}}$ di parti di Ω . Risulta:*

$$\liminf_{n \rightarrow \infty} A_n \subseteq \limsup_{n \rightarrow \infty} A_n. \quad (3.8)$$

Dimostrazione. Se un punto ω dello spazio Ω appartiene a $\liminf_{n \rightarrow \infty} A_n$ significa che esso appartiene agli elementi della successione $\{A_n\}_{n \in \mathbb{N}}$ a partire da un certo indice in poi. Quindi ω appartiene ad infiniti elementi della successione $\{A_n\}_{n \in \mathbb{N}}$, così esso appartiene anche a $\limsup_{n \rightarrow \infty} A_n$. \square

In termine di eventi, possiamo quindi affermare che, se si verifica l'evento limite inferiore di una successione si verifica anche l'evento limite superiore della stessa successione.

✓ **Definizione 3.11.** Consideriamo una generica successione $\{A_n\}_{n \in \mathbb{N}}$ di parti di Ω . Se accade che

$$\liminf_{n \rightarrow \infty} A_n \supseteq \limsup_{n \rightarrow \infty} A_n$$

allora si dice che $\{A_n\}_{n \in \mathbb{N}}$ è dotata di limite. In tal caso, sussistendo anche la (3.8), si pone per definizione:

$$\lim_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n = \liminf_{n \rightarrow \infty} A_n.$$

\square

Esempio 3.9. Nello spazio $\Omega = [-1, 1]$ consideriamo la successione $\{A_n\}_{n \in \mathbb{N}}$ (non monotona) di sottoinsiemi il cui termine generale è costituito dall'intervallo

$$A_n = \begin{cases} (0, 1), & \text{se } n \text{ è pari,} \\ \left(0, \frac{1}{n}\right), & \text{se } n \text{ è dispari.} \end{cases}$$

3.2 Successioni e loro limiti

Dall'essere

$$C_n = \bigcap_{k=n}^{\infty} A_k = \begin{cases} (0, 1) \cap \left(0, \frac{1}{n+1}\right) \cap (0, 1) \cap \left(0, \frac{1}{n+3}\right) \cap \dots, & \text{per } n \text{ pari,} \\ \left(0, \frac{1}{n}\right) \cap (0, 1) \cap \left(0, \frac{1}{n+2}\right) \cap (0, 1) \cap \dots, & \text{per } n \text{ dispari,} \end{cases}$$

segue che $C_n = \emptyset$ per ogni $n \in \mathbb{N}$, e quindi

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} C_n = \bigcup_{n=1}^{\infty} \emptyset = \emptyset.$$

Dal momento che risulta $D_n = \bigcup_{k=n}^{\infty} A_k = (0, 1)$, per il limite superiore si ottiene invece:

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} D_n = \bigcap_{n=1}^{\infty} (0, 1) = (0, 1).$$

Dal momento che il limite inferiore differisce dal limite superiore, la successione $\{A_n\}_{n \in \mathbb{N}}$ non ammette limite. \square

✓ **Esempio 3.10.** Nello stesso spazio Ω dell'esempio precedente, si consideri la successione $\{B_n\}_{n \in \mathbb{N}}$ (non monotona) il cui termine generale è costituito dall'intervallo

$$B_n = \begin{cases} \emptyset, & \text{se } n \text{ è pari,} \\ \left(0, \frac{1}{n}\right), & \text{se } n \text{ è dispari.} \end{cases}$$

Dall'essere

$$C_n = \bigcap_{k=n}^{\infty} B_k = \begin{cases} \emptyset \cap \left(0, \frac{1}{n+1}\right) \cap \emptyset \cap \left(0, \frac{1}{n+3}\right) \cap \dots, & \text{per } n \text{ pari,} \\ \left(0, \frac{1}{n}\right) \cap \emptyset \cap \left(0, \frac{1}{n+2}\right) \cap \emptyset \cap \dots, & \text{per } n \text{ dispari,} \end{cases}$$

segue che, per ogni $n \in \mathbb{N}$, $C_n = \emptyset$, e quindi

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} C_n = \bigcup_{n=1}^{\infty} \emptyset = \emptyset.$$

Inoltre, risulta

$$D_n = \bigcup_{k=n}^{\infty} B_k = \begin{cases} \emptyset \cup \left(0, \frac{1}{n+1}\right) \cup \emptyset \cup \left(0, \frac{1}{n+3}\right) \cup \dots = \left(0, \frac{1}{n+1}\right), & \text{per } n \text{ pari,} \\ \left(0, \frac{1}{n}\right) \cup \emptyset \cup \left(0, \frac{1}{n+2}\right) \cup \emptyset \cup \dots = \left(0, \frac{1}{n}\right), & \text{per } n \text{ dispari,} \end{cases}$$

dalla quale si ricava immediatamente

$$\limsup_{n \rightarrow \infty} B_n = \bigcap_{n=1}^{\infty} D_n = (0, 1) \cap \left(0, \frac{1}{3}\right) \cap \left(0, \frac{1}{5}\right) \cap \dots \cap \left(0, \frac{1}{2n+1}\right) \cap \dots = \emptyset.$$

Il limite inferiore ed il limite superiore della successione $\{B_n\}_{n \in \mathbb{N}}$ coincidono, essendo entrambi uguali a \emptyset ; quindi tale successione è dotata di limite che risulta essere l'insieme vuoto. \square

3.3 Altre strutture algebriche di interesse probabilistico

Nel paragrafo iniziale di questo capitolo abbiamo mostrato che il requisito della stabilità rispetto al complemento e alle unioni finite scaturisce da considerazioni intuitive nella logica degli eventi. Da esse segue immediatamente il fatto che gli eventi formano un'algebra sullo spazio campionare Ω associato ad un esperimento casuale \mathcal{E} . Più riposta è, invece, la richiesta della stabilità della struttura degli eventi rispetto all'unione numerabile che di solito viene giustificata dicendo che essa è una condizione di comodo, suggerita dalla opportunità di disporre di una più ampia gamma di operatività sugli insiemi che rappresentano gli eventi. Per parte nostra, abbiamo già mostrato con un esempio che vi sono situazioni nelle quali gli eventi di interesse possono essere rappresentati solo in termini di unioni numerabili di eventi esprimibili attraverso gli eventi generatori.

In questo paragrafo, oltre a presentare un'altra struttura algebrica che riveste particolare importanza nella trattazione assiomatica del Calcolo delle Probabilità, affronteremo il problema di determinare quale ulteriore proprietà è necessario richiedere ad un'algebra affinché essa sia anche una σ -algebra, ovvero qual è il requisito necessario per passare dalla stabilità rispetto all'unione finita a quella rispetto all'unione numerabile.

Definizione 3.12. Sia \mathcal{M} una famiglia di parti di Ω . Se per ogni successione $\{A_n\}_{n \in \mathbb{N}}$ monotona e costituita da elementi di \mathcal{M} accade che anche $\lim_{n \rightarrow \infty} A_n$ appartiene a \mathcal{M} allora \mathcal{M} dicesi classe monotona su Ω . \square

In altre parole \mathcal{M} è una classe monotona se contiene i limiti di tutte le successioni monotone costituite con suoi elementi.

3.3 Altre strutture algebriche di interesse probabilistico

Teorema 3.3. Sia \mathcal{A} un'algebra su Ω . Le due affermazioni seguenti sono equivalenti:

- a) \mathcal{A} è una σ -algebra su Ω ;
- b) \mathcal{A} è una classe monotona su Ω .

Dimostrazione. Dimostriamo che $a) \rightarrow b)$. A tale scopo consideriamo una qualsiasi successione $\{A_n\}_{n \in \mathbb{N}}$ monotona e costituita da elementi di \mathcal{A} . Sia essa crescente. Dall'ipotesi che \mathcal{A} è una σ -algebra, applicando la (3.2) si trae:

$$\lim_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}.$$

Quindi \mathcal{A} contiene il limite delle sue successioni crescenti. Lo stesso accade per le successioni decrescenti, come si può facilmente vedere applicando la (3.3). Se ne conclude che \mathcal{A} è una classe monotona su Ω .

Dimostriamo che $b) \rightarrow a)$. Consideriamo una qualsiasi successione $\{A_n\}_{n \in \mathbb{N}}$ costituita da elementi di \mathcal{A} . Posto $C_n = \bigcup_{k=1}^n A_k$ per ogni $n \in \mathbb{N}$, risulta $C_n \in \mathcal{A}$ in quanto \mathcal{A} è un'algebra (ipotesi iniziale dell'enunciato del teorema). Inoltre, la successione $\{C_n\}_{n \in \mathbb{N}}$ è, per costruzione, crescente e quindi, per ogni $n \in \mathbb{N}$ risulta $\bigcup_{k=1}^n C_k = C_n = \bigcup_{k=1}^n A_k$. Da quest'ultima e dall'ipotesi che \mathcal{A} è una classe monotona, ricordando anche la (3.2), segue:

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} C_n = \lim_{n \rightarrow \infty} C_n \in \mathcal{A}.$$

Tanto basta per affermare che \mathcal{A} è una σ -algebra su Ω . \square

Siamo ora in grado di rispondere al quesito inizialmente posto: una σ -algebra su Ω è un'algebra su Ω che contiene i limiti delle proprie successioni monotone.

Passiamo ora a definire la struttura algebrica cui si è fatto riferimento all'inizio di questo paragrafo.

Definizione 3.13. Una famiglia \mathcal{S} di parti di uno spazio Ω costituisce una semi-algebra su Ω se verifica le seguenti proprietà:

- (i) $\Omega \in \mathcal{S}$;
- (ii) Se $A \in \mathcal{S}$ allora \bar{A} può essere rappresentato come unione di un numero finito di elementi di \mathcal{S} a due a due disgiunti; in altri termini esistono un intero n e una n -upla di elementi di \mathcal{S} a due a due disgiunti, diciamo $\{A_1, A_2, \dots, A_n\}$, tali che: $\bar{A} = \bigcup_{k=1}^n A_k$ ²;
- (iii) $A \in \mathcal{S}, B \in \mathcal{S} \Rightarrow A \cap B \in \mathcal{S}$. \square

² Il simbolo dell'unione con il puntino centrale è utilizzato per evidenziare che si stanno unendo insiemi a due a due disgiunti. Nel seguito si utilizzerà tale simbolo e, ove non ritenuto strettamente necessario, si ometterà la specificazione che gli insiemi sono a due a due disgiunti nel testo.

Da questa definizione discende che $\emptyset \in \mathcal{S}$. Infatti la (i) assicura che $\Omega \in \mathcal{S}$ e la (ii) afferma che $\emptyset = \overline{\Omega}$ si deve poter rappresentare come unione finita di elementi di \mathcal{S} . L'unica possibilità per tale rappresentazione è che $\emptyset \in \mathcal{S}$.

Ci proponiamo, adesso, di caratterizzare la forma degli elementi di un'algebra su Ω generata da una semi-algebra su Ω . A tal fine è utile premettere il seguente

Lemma 3.1. *Sia \mathcal{S} una semi-algebra sullo spazio Ω e sia \mathcal{C} la classe delle unioni finite di elementi di \mathcal{S} a due a due disgiunti. Si ha:*

$$C \in \mathcal{C}, D \in \mathcal{C} \Rightarrow C \cap D \in \mathcal{C}.$$

Dimostrazione. Per definizione $C \in \mathcal{C} \Rightarrow C = \bigcup_{i=1}^n C_i$ e $D \in \mathcal{C} \Rightarrow D = \bigcup_{j=1}^m D_j$ dove n e m sono opportuni interi e $\{C_1, C_2, \dots, C_n\}$ e $\{D_1, D_2, \dots, D_m\}$ insiemi costituiti da elementi di \mathcal{S} a due a due disgiunti. Quindi

$$C \cap D = \bigcup_{i=1}^n C_i \cap \bigcup_{j=1}^m D_j = \bigcup_{i=1}^n \bigcup_{j=1}^m C_i \cap D_j.$$

I sottoinsiemi $C_i \cap D_j$, appartenono a \mathcal{S} per $i = 1, 2, \dots, n$ e $j = 1, 2, \dots, m$, in virtù della (iii) della Definizione 3.13. L'intersezione di C e D , potendosi scrivere come unione finita di elementi \mathcal{S} a due a due disgiunti, appartiene per definizione a \mathcal{C} . \square

Osservazione 3.4. Il risultato appena dimostrato può essere esteso, per induzione, all'intersezione di un numero finito di sottoinsiemi di \mathcal{C} . \square

Osservazione 3.5. Risulta utile far notare esplicitamente che \mathcal{C} , oltre a contenere tutti gli elementi di \mathcal{S} (in particolare Ω e \emptyset), contiene anche i complementi di tutti gli elementi di \mathcal{S} . Infatti, se $A \in \mathcal{S}$, allora per la (ii) della Definizione 3.13 esistono un intero n e una n -upla di elementi di \mathcal{S} a due a due disgiunti, $\{A_1, A_2, \dots, A_n\}$, tali che $\overline{A} = \bigcup_{i=1}^n A_i$. Ma questo è proprio il requisito affinché \overline{A} appartenga a \mathcal{C} . \square

Siamo adesso in grado di dimostrare il teorema che fornisce la caratterizzazione ricercata.

Teorema 3.4. *Sia \mathcal{S} una semi-algebra sullo spazio Ω . L'algebra $a(\mathcal{S})$ generata dalla semi-algebra \mathcal{S} è equivalente alla classe \mathcal{C} delle unioni finite di elementi di \mathcal{S} a due a due disgiunti.*

Dimostrazione. In primo luogo, dalla definizione stessa di algebra discende immediatamente che $\mathcal{C} \subseteq a(\mathcal{S})$. La dimostrazione sarà, quindi, completata se mostreremo che \mathcal{C} è un'algebra, ossia che \mathcal{C} verifica le proprietà (i), (ii) e (iii) della Definizione 3.1. Abbiamo già visto $\Omega \in \mathcal{C}$. Dimostriamo che \mathcal{C} è chiusa rispetto al complemento. Sia C un elemento di \mathcal{C} ; esistono allora un intero n e una n -upla di elementi di \mathcal{S} a due a due disgiunti, $\{C_1, C_2, \dots, C_n\}$, tali che

3.4 La probabilità

$C = \bigcup_{i=1}^n C_i$. Passando al complemento, le formule di De Morgan consentono di scrivere

$$\overline{C} = \overline{\bigcup_{i=1}^n C_i} = \bigcap_{i=1}^n \overline{C_i}$$

da cui, ricordando le Osservazioni 3.4 e 3.5, discende che $\overline{C} \in \mathcal{C}$. Terminiamo la dimostrazione mostrando che \mathcal{C} è chiusa rispetto all'unione di due suoi elementi. Infatti, siano C e D due elementi di \mathcal{C} ; applicando le leggi di De Morgan si ottiene:

$$C \cup D = \overline{\overline{C} \cap \overline{D}} \in \mathcal{C}$$

avendo utilizzato la appena dimostrata stabilità \mathcal{C} rispetto al complemento e il Lemma 3.1. \square

3.4 La probabilità

Nel precedente paragrafo è stata elucidata la struttura algebrica atta a definire gli eventi associabili ad un esperimento casuale. È stato altresì già menzionato che agli elementi della σ -algebra \mathcal{F} considerata deve potersi associare una qualche "misura". In questo paragrafo intendiamo affrontare tale questione introducendo, lungo la linea dovuta a Kolmogorov, la nozione di probabilità proprio come opportuna misura operante sulla σ -algebra degli eventi. A tal fine formuliamo un sistema assiomatico, atto a rappresentare in maniera rigorosa e coerente l'esperimento casuale. Come apparirà chiaro, questo sistema assiomatico costituisce la sintesi, sia pure con le opportune estensioni matematiche, dell'insieme di proprietà condivise dalle varie definizioni di probabilità in passato discusse (definizione aprioristica, frequentista, soggettiva), al di là delle diversità di impostazioni e dei limiti di applicazione di ognuna di esse.

✓ **Definizione 3.14.** *Sia assegnato uno spazio probabilizzabile (Ω, \mathcal{F}) . Una funzione di insieme $P: \mathcal{F} \rightarrow \mathbb{R}$ che gode delle proprietà*

$$(i) \forall A \in \mathcal{F}, P(A) \geq 0,$$

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

(iii) per ogni successione $\{A_n\}_{n \in \mathbb{N}}$ di eventi a due a due incompatibili, ossia tali che $A_i \cap A_j = \emptyset$ per $i \neq j$, si ha

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n), \quad (3.9)$$

viene detta (misura di) probabilità su (Ω, \mathcal{F}) . La (iii) esprime la proprietà di additività numerabile o completa della probabilità. \triangleleft

La probabilità è, quindi, una funzione definita sulla classe degli eventi di uno spazio campione Ω a valori non negativi (in particolare assegna 1 all'evento certo) numerabilmente additiva. Quindi la probabilità è una misura (di massa) finita.

La definizione seguente, oltre a racchiudere in sé i concetti basilari finora illustrati, è il punto di partenza per lo sviluppo della teoria assiomatica.

Definizione 3.15. Sia \mathfrak{D} un esperimento casuale e sia \mathcal{G} la famiglia degli eventi generatori. Dicesi spazio di probabilità la tripla $(\Omega, \mathcal{F}, \mathbb{P})$ dove Ω è lo spazio campione associato ad \mathfrak{D} , \mathcal{F} è la σ -algebra generata dalla famiglia \mathcal{G} e \mathbb{P} è una misura di probabilità su (Ω, \mathcal{F}) . \triangleleft

Quindi, nella descrizione dell'esperimento casuale \mathfrak{D} si assume che la classe degli eventi è una σ -algebra e che la funzione che assegna ad ogni evento la sua probabilità di occorrenza è una misura finita. In altri termini le proprietà che definiscono una σ -algebra e quelle che definiscono una misura di probabilità vengono scelte come assiomi.

Per quanto riguarda l'assioma dell'additività numerabile si può dire che di esso non se ne ha una diretta interpretazione intuitiva, né che esso era stato messo in luce dalle definizioni di probabilità incontrate nel capitolo precedente. D'altra parte, nel caso della definizione aprioristica, il problema non si pone, perché in tale contesto non venivano prese in considerazione collezioni infinite di eventi incompatibili. Lo stesso si può dire della definizione frequentista. Discorso a parte merita la definizione soggettiva. L'additività finita deriva dalla condizione di coerenza, per la quale, non deve essere possibile con un insieme di scommesse, ottenere una vincita certa (o una perdita certa). Riprendendo la dimostrazione data nel § 2.5, ci si convince facilmente che essa può essere facilmente estesa al caso di una infinità numerabile di eventi incompatibili ottenendo in tal modo anche la numerabile additività. La scelta dell'additività numerabile, anche se contestata da taluni, non è esclusa dalla varie definizioni. Inoltre, come vedremo successivamente con il già citato Teorema di Equivalenza, essa conduce ad un'importante proprietà che è peraltro suggerita anche da considerazioni di natura intuitiva: se una successione di eventi ammette limite, allora anche la successione delle probabilità dei suoi eventi ammette limite, e la probabilità del limite coincide con il limite delle probabilità. In altri termini la probabilità deve essere una funzione continua. Ad ogni modo, è facilmente intuibile che dall'assioma dell'additività numerabile discende la proprietà di additività finita (o semplice) per la probabilità. Vediamo come si perviene a tale risultato.

Lemma 3.2. Si ha $\mathbb{P}(\emptyset) = 0$.

Dimostrazione. Utilizzando la proprietà di idempotenza, possiamo esprimere l'evento impossibile \emptyset come unione numerabile di eventi impossibili: $\emptyset = \emptyset \cup \emptyset \cup \dots$. La numerabile additività di \mathbb{P} applicata a tale successione fornisce:

$$\mathbb{P}(\emptyset) = \mathbb{P}(\emptyset) + \mathbb{P}(\emptyset) + \dots$$

3.4 La probabilità

Ma $\mathbb{P}(\emptyset)$ è un numero reale finito, così che la precedente uguaglianza comporta $\mathbb{P}(\emptyset) = 0$. \square

✓ **Teorema 3.5.** La funzione \mathbb{P} è finitamente additiva ossia, se A_1, A_2, \dots, A_n sono eventi a due a due incompatibili, allora si ha:

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

Dimostrazione. Osserviamo che partendo dagli eventi A_1, A_2, \dots, A_n possiamo costruire una successione $\{B_n\}_{n \in \mathbb{N}}$ di eventi ponendo $B_k = A_k$, per $k = 1, 2, \dots, n$, e $B_{n+1} = B_{n+2} = \dots = \emptyset$. Dall'ipotesi di incompatibilità a due a due degli eventi A_1, A_2, \dots, A_n discende immediatamente che così sono anche gli eventi della successione $\{B_n\}_{n \in \mathbb{N}}$. Allora, dalla numerabile additività di \mathbb{P} e dal Lemma 3.2, si ottiene:

$$\begin{aligned} \mathbb{P}\left(\bigcup_{k=1}^n A_k\right) &= \mathbb{P}\left[\left(\bigcup_{k=1}^n A_k\right) \cup \emptyset \cup \emptyset \cup \dots\right] = \mathbb{P}\left(\bigcup_{k=1}^{\infty} B_k\right) \\ &= \sum_{k=1}^{\infty} \mathbb{P}(B_k) = \sum_{k=1}^n \mathbb{P}(A_k) + \sum_{k=n+1}^{\infty} \mathbb{P}(\emptyset) = \sum_{k=1}^n \mathbb{P}(A_k). \end{aligned} \quad \square$$

Dal Teorema 3.5 discendono facilmente alcuni risultati che rivestono, però, carattere particolarmente significativo dal punto di vista probabilistico.

✓ **Corollario 3.1.** Per ogni evento A risulta:

$$\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A). \quad (3.10)$$

Dimostrazione. Considerato un evento A qualsiasi, lo spazio campione Ω si può esprimere come unione degli eventi incompatibili A e \bar{A} . La proprietà (ii) della Definizione 3.14 e il Teorema 3.5 permettono di scrivere

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

Da qui segue la (3.10). \square

✓ **Corollario 3.2.** Per ogni evento A si ha $\mathbb{P}(A) \in [0, 1]$.

Dimostrazione. Sia $A \in \mathcal{F}$. In primo luogo, la (i) della Definizione 3.14 fornisce $\mathbb{P}(A) \geq 0$. Analogamente $\mathbb{P}(\bar{A}) \geq 0$ mentre il Corollario 3.1 fornisce $\mathbb{P}(A) = 1 - \mathbb{P}(\bar{A})$. Quindi, in definitiva, risulta:

$$0 \leq \mathbb{P}(A) \leq 1. \quad \square$$

✓ **Corollario 3.3** (Proprietà di monotonia). *Se A_1 e A_2 sono eventi tali che $A_1 \subseteq A_2$, si ha $\mathbb{P}(A_1) \leq \mathbb{P}(A_2)$.*

Dimostrazione. Anzitutto osserviamo che dall'essere $A_1 \cup A_2 = A_2$, in virtù della (1.8) è possibile esprimere A_2 come unione degli eventi incompatibili A_1 e $\overline{A}_1 \cap A_2$:

$$A_2 = A_1 \cup (\overline{A}_1 \cap A_2).$$

Applicando la finita additività della funzione \mathbb{P} possiamo allora scrivere:

$$\mathbb{P}(A_2) = \mathbb{P}(A_1) + \mathbb{P}(\overline{A}_1 \cap A_2)$$

e la tesi segue dall'essere $\mathbb{P}(\overline{A}_1 \cap A_2) \geq 0$ (per la (i) della Definizione 3.14). \square

Termineremo questo paragrafo presentando alcuni risultati che riguardano collezioni di eventi non incompatibili.

✓ **Teorema 3.6.** *Siano A_1 e A_2 eventi. Si ha:*

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2). \quad (3.11)$$

Dimostrazione. Applicando la (1.8) agli eventi A_1 e A_2 , si ottiene:

$$A_1 \cup A_2 = A_1 \cup (\overline{A}_1 \cap A_2), \quad (3.12)$$

mentre, la formula della partizione con il certo (1.7) applicata all'evento A_2 , fornisce:

$$A_2 = (A_1 \cap A_2) \cup (\overline{A}_1 \cap A_2). \quad (3.13)$$

Osserviamo esplicitamente che risulta

$$A_1 \cap (\overline{A}_1 \cap A_2) = \emptyset \quad \text{e} \quad (A_1 \cap A_2) \cap (\overline{A}_1 \cap A_2) = \emptyset \quad (3.14)$$

così che gli eventi A_2 e $A_1 \cup A_2$ sono stati entrambi espressi come unioni di eventi incompatibili. Il Teorema 3.5 applicato, rispettivamente, alla (3.12) ed alla (3.13) fornisce allora:

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(\overline{A}_1 \cap A_2) \quad (3.15)$$

$$\mathbb{P}(A_2) = \mathbb{P}(A_1 \cap A_2) + \mathbb{P}(\overline{A}_1 \cap A_2). \quad (3.16)$$

Da quest'ultima segue:

$$\mathbb{P}(\overline{A}_1 \cap A_2) = \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2). \quad (3.17)$$

Sostituendo il secondo membro della (3.17) nella (3.15) si ha infine la relazione (3.11). \square

Se A_1 e A_2 sono incompatibili si ha $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(\emptyset) = 0$, cosicché la (3.11) si riduce alla finita additività della funzione di probabilità \mathbb{P} . Il Teorema 3.6 è suscettibile della seguente generalizzazione nota quale *formula di inclusione-esclusione*.

✓ **Teorema 3.7** (Formula di inclusione-esclusione). *Sia $\{A_1, \dots, A_n\} \subseteq \mathcal{F}$. Risulta:*

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{i < j} \mathbb{P}(A_i \cap A_j) + \sum_{i < j < k} \mathbb{P}(A_i \cap A_j \cap A_k) + \\ &\quad + \cdots + (-1)^{n+1} \mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n). \end{aligned} \quad (3.18)$$

Dimostrazione. Procediamo per induzione. Osserviamo a tal fine che, in virtù del Teorema 3.6, la (3.18) sussiste per $n = 2$. Dimostriamo che se essa vale per $n = r$ allora essa vale anche per $n = r + 1$. A tal fine, posto

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

in virtù del Teorema 3.6 segue:

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

Poiché la (3.18) è stata assunta vera per $n = r$, si ha anche:

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^{r+1} A_i\right) &= \mathbb{P}\left(\bigcup_{i=1}^r A_i\right) + \mathbb{P}(A_{r+1}) - \mathbb{P}(B_r \cap A_{r+1}) \\ &= \left[\sum_{i=1}^r \mathbb{P}(A_i) - \sum_{i < j} \mathbb{P}(A_i \cap A_j) + \sum_{i < j < k} \mathbb{P}(A_i \cap A_j \cap A_k) + \right. \\ &\quad \left. + \cdots + (-1)^{r+1} \mathbb{P}(A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_r) \right] + \\ &\quad + \mathbb{P}(A_{r+1}) - \mathbb{P}\left[\bigcup_{i=1}^r (A_i \cap A_{r+1})\right]. \end{aligned} \quad (3.19)$$

Osserviamo ora che applicando di nuovo la formula (3.18) risulta:

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

Facendo uso della proprietà associativa dell'intersezione ed osservando che l'intersezione di un insieme con sé stesso dà l'insieme stesso, si ha poi:

$$\begin{aligned} \mathbb{P}\left[\bigcup_{i=1}^r (A_i \cap A_{r+1})\right] &= \sum_{i=1}^r \mathbb{P}(A_i \cap A_{r+1}) - \sum_{i < j} \mathbb{P}(A_i \cap A_j \cap A_{r+1}) \\ &\quad + \cdots + (-1)^{r+1} \mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_{r+1}). \end{aligned} \quad (3.20)$$

Sostituendo la (3.20) nell'ultima delle (3.19) si ricava la (3.18) scritta per $n = r + 1$. \square

Riprendiamo in esame la formula (3.11). Da essa, trascurando l'addendo non negativo $\mathbb{P}(A_1 \cap A_2)$, in particolare si trae:

$$\mathbb{P}(A_1 \cup A_2) \leq \mathbb{P}(A_1) + \mathbb{P}(A_2). \quad (3.21)$$

La formula (3.21) fornisce un maggiorante per la probabilità dell'unione di due eventi. Questo può essere utile per ottenere un'approssimazione della probabilità $\mathbb{P}(A_1 \cup A_2)$ quando essa non sia calcolabile in modo esatto e nell'ipotesi che $\mathbb{P}(A_1 \cap A_2)$ possa ritenersi trascurabile rispetto a $\mathbb{P}(A_1) + \mathbb{P}(A_2)$ (ovviamente, affinché essa abbia un significato, deve risultare $\mathbb{P}(A_1) + \mathbb{P}(A_2) \leq 1$). Si osservi che la (3.21) è suscettibile di immediata generalizzazione al caso di un numero arbitrario, ma finito, di eventi.

Teorema 3.8. Per una qualsiasi successione di eventi, $\{A_n\}_{n \in \mathbb{N}}$, risulta:

$$\forall n \in \mathbb{N}, \quad \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n \mathbb{P}(A_i). \quad (3.22)$$

Dimostrazione. Procediamo applicando il principio di induzione completa. La tesi è vera per $n = 2$ in virtù della (3.21). Supponiamo la tesi vera per $n = k > 2$. Applicando, di seguito, la (3.21) e la (3.22) scritta per $n = k$, si ottiene

$$\mathbb{P}\left(\bigcup_{i=1}^{k+1} A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^k A_i \cup A_{k+1}\right) \leq \mathbb{P}\left(\bigcup_{i=1}^k A_i\right) + \mathbb{P}(A_{k+1}) \leq \sum_{i=1}^k \mathbb{P}(A_i) + \mathbb{P}(A_{k+1})$$

che è proprio la (3.22) scritta per $n = k + 1$. \square

3.5 Il problema dell'estensione

La generalizzazione al caso infinito della (3.22) è altresì vera ed è nota come *disugualanza di Boole*; essa è espressa nel Teorema che segue.

Teorema 3.9 (Disugualanza di Boole). Per una qualsiasi successione di eventi, $\{A_n\}_{n \in \mathbb{N}}$, risulta:

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n). \quad (3.23)$$

Dimostrazione. La tesi è banale se la serie a secondo membro della (3.23) risulta divergente. Nel caso contrario, esprimiamo l'evento $\bigcup_{n=1}^{\infty} A_n$ come unione di eventi a due a due incompatibili al seguente modo:

$$\bigcup_{n=1}^{\infty} A_n = A_1 \cup (\bar{A}_1 \cap A_2) \cup (\bar{A}_1 \cap \bar{A}_2 \cap A_3) \cup \cdots \cup (\bar{A}_1 \cap \bar{A}_2 \cap \cdots \cap \bar{A}_{n-1} \cap A_n) \cup \cdots$$

In altri termini, $\bigcup_{n=1}^{\infty} A_n$ è espresso come A_1 unito con la parte di A_2 non comune ad A_1 , unito con la parte di A_3 non comune ad A_1 e ad A_2 , ecc. Dall'assioma di additività numerabile si ha allora:

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbb{P}(A_1) + \mathbb{P}(\bar{A}_1 \cap A_2) + \mathbb{P}(\bar{A}_1 \cap \bar{A}_2 \cap A_3) + \cdots$$

La (3.23) segue allora da quest'ultima maggiorando, $\forall n \in \mathbb{N}$, l'addendo di posto n con $\mathbb{P}(A_n)$ in virtù del Corollario 3.3. \square

3.5 Il problema dell'estensione



Affronteremo ora il problema della costruzione di una funzione d'insieme avente i requisiti per essere una misura di probabilità. A tale scopo prenderemo come punto di partenza uno spazio probabilizzabile (Ω, \mathcal{F}) essendo Ω uno spazio campione associato ad un esperimento casuale e \mathcal{F} una σ -algebra su Ω generata da una famiglia \mathcal{G} di sottoinsiemi di Ω .

Un caso particolarmente importante che si presenta in molte situazioni di interesse si ha quando risulta possibile determinare la famiglia \mathcal{E} degli eventi elementari avendo inoltre questa cardinalità finita: $\mathcal{E} = \{E_1, E_2, \dots, E_n\}$. In tal caso, un qualsiasi evento $A \in \mathcal{F}$ è espresso come unione (finita) di eventi elementari: $A = E_{i_1} \cup E_{i_2} \cup \cdots \cup E_{i_k}$. Assegnata, allora, una n -upla di numeri reali $\{\mathbb{P}(E_1), \mathbb{P}(E_2), \dots, \mathbb{P}(E_n)\}$ tale che:

$$\mathbb{P}(E_i) \geq 0, \quad (i = 1, 2, \dots, n), \quad (3.24)$$

$$\sum_{i=1}^n \mathbb{P}(E_i) = 1 \quad (3.25)$$

e posto per ogni $A \in \mathcal{F}$

$$\mathbb{P}(A) = \mathbb{P}(E_{i_1}) + \mathbb{P}(E_{i_2}) + \cdots + \mathbb{P}(E_{i_k}), \quad (3.26)$$

la funzione d'insieme così definita risulta essere una misura di probabilità su (Ω, \mathcal{F}) , ossia essa verifica gli assiomi della Definizione 3.14. Infatti la condizione (3.24) e la posizione (3.26) assicurano la sussistenza dell'assioma (i) mentre la condizione (3.25), dall'essere $\Omega = E_1 \cup E_2 \cup \cdots \cup E_n$, garantisce l'assioma (ii). Infine siano $A = E_{i_1} \cup E_{i_2} \cup \cdots \cup E_{i_k}$ e $B = E_{j_1} \cup E_{j_2} \cup \cdots \cup E_{j_s}$ eventi. Se essi sono incompatibili allora tutti gli eventi elementari che compongono A sono diversi da tutti gli eventi elementari che compongono B , così che

$$A \cup B = (E_{i_1} \cup E_{i_2} \cup \cdots \cup E_{i_k}) \cup (E_{j_1} \cup E_{j_2} \cup \cdots \cup E_{j_s}).$$

La posizione (3.26) conferisce, quindi, alla funzione \mathbb{P} la proprietà di finita additività che nella situazione in esame equivale all'assioma (iii).

Esempio 3.11. Riprendiamo in considerazione l'Esempio 3.1 dove $\Omega = \{1, 2, 3, 4, 5, 6\}$. In tale contesto la famiglia degli eventi generatori coincide con quella degli eventi elementari e questi sono

$$E_1 = \{1\}, E_2 = \{2\}, \dots, E_6 = \{6\}.$$

Pertanto $\mathcal{F} = \mathbb{P}(\Omega)$. Supposto allora il dado non truccato, evidenti considerazioni di simmetria suggeriscono la seguente assegnazione di probabilità agli eventi elementari:

$$\mathbb{P}(E_i) = \frac{1}{6}, \quad (i = 1, 2, \dots, n)$$

che soddisfa le (3.24) e (3.25). Con tale posizione la (3.26) fornisce la probabilità di un qualsiasi sottoinsieme A di Ω ; essa, come si verifica facilmente, equivale al rapporto tra la cardinalità di A e la cardinalità di Ω che è pari a 6.

Mantenendo l'ipotesi di regolarità del dado, riferiamoci ora al contesto illustrato nell'Esempio 3.3. Ricordiamo, in primo luogo, che in esso gli eventi elementari sono i quattro sottoinsiemi

$$E_1 = \{1\}, E_2 = \{3, 5\}, E_3 = \{2\}, E_4 = \{4, 6\};$$

quindi procediamo ad assegnare ad essi le seguenti probabilità:

$$\mathbb{P}(E_1) = \frac{1}{6}, \quad \mathbb{P}(E_2) = \frac{2}{6}, \quad \mathbb{P}(E_3) = \frac{1}{6}, \quad \mathbb{P}(E_4) = \frac{2}{6}.$$

Tale assegnazione soddisfa le (3.24) e (3.25). La probabilità di un qualsiasi evento si ottiene tramite la posizione (3.26). Ad esempio per l'evento $A = \{3, 4, 5, 6\}$ si ottiene $\mathbb{P}(A) = \mathbb{P}(E_2) + \mathbb{P}(E_4) = 4/6$, mentre non è definita la probabilità del sottoinsieme $\{1, 5\}$ che non si ottiene dall'unione di eventi elementari e che quindi non è un evento. \triangleleft

3.5 Il problema dell'estensione

Osservazione 3.6. Le considerazioni precedenti si estendono agevolmente al caso di una famiglia numerabile di eventi elementari, nel qual caso nella condizione (3.25) al posto della sommatoria compare il simbolo di serie. Lo stesso accade nella posizione 3.26 se l'evento A è unione di infiniti eventi elementari. \triangleleft

Esempio 3.12. Riprendiamo in considerazione lo spazio campione di Bernoulli Ω illustrato nell'Esempio 3.4 e supponiamo che i lanci si riferiscano ad una moneta regolare. Evidenti ragioni di simmetria suggeriscono la considerazione che tutti i singoletti di Ω (che abbiamo già osservato essere eventi) hanno la stessa probabilità p di occorrenza che risulta nulla. Infatti, se così non fosse si potrebbero avere eventi, costituiti da un numero sufficientemente grande di distinte successioni di T e C , con probabilità maggiore di 1. Da questa stessa considerazione si ricava anche che la cardinalità di Ω è maggiore di quella del numerabile, altrimenti per l'additività numerabile si avrebbe $\mathbb{P}(\Omega) = 0$. Inoltre, dal momento che, per ogni n intero, vi sono 2^n possibili n -disposizioni con ripetizione di T e C che le stesse ragioni di simmetria inducono a ritenere equiprobabili, allora ogni evento che si ottiene specificando i primi n posti deve avere probabilità 2^{-n} . Da questa osservazione possiamo ricavare la probabilità dell'evento A : “due teste consecutive si presentano prima di due croci consecutive”: Infatti, facendo uso della (3.1) si ottiene

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{n=2}^{\infty} A_n\right) = \sum_{n=2}^{\infty} \mathbb{P}(A_n) = \sum_{n=2}^{\infty} 2^{-n} = \frac{1}{4} \sum_{n=0}^{\infty} 2^{-n} = \frac{1}{2}.$$

Il risultato ottenuto è senza dubbio conseguenza della equità della moneta.

Allo stesso modo risulta agevole assegnare la probabilità agli eventi elementari che costituiscono la famiglia che abbiamo designato con \mathcal{E}_2 e che genera la σ -algebra \mathcal{F}_2 . In primo luogo, essendo E_0 un singoletto, si può porre $\mathbb{P}(E_0) = 0$. Inoltre, dato che E_n , per ogni n intero, è costituito da tutte le successioni di T e C che hanno ai primi n posti la sequenza $CC\dots CT$, si può porre $\mathbb{P}(E_n) = 2^{-n}$. Con tale assegnazione la (3.24) è immediatamente verificata, mentre per la (3.25) (estesa al caso numerabile) si ha:

$$\sum_{n=0}^{\infty} \mathbb{P}(E_n) = 0 + \sum_{n=1}^{\infty} \frac{1}{2^n} = 2 - 1 = 1.$$

Come applicazione della posizione (3.26) consideriamo il seguente evento B : “ T si verifica per la prima volta in un lancio di posto pari”. Gli eventi elementari che tramite l'operazione di unione formano B sono tutti quelli con indice pari e quindi:

$$\mathbb{P}(B) = \sum_{n=1}^{\infty} \mathbb{P}(E_{2n}) = \sum_{n=1}^{\infty} \frac{1}{2^{2n}} = \sum_{n=1}^{\infty} \frac{1}{4^n} = \frac{4}{3} - 1 = \frac{1}{3}.$$

Il risultato ottenuto esplicita, in maniera quantitativa, l'intuizione che l'evento B è meno probabile del suo negato “ T si verifica per la prima volta in un lancio di posto dispari”. \triangleleft

Un'altra situazione che pone un problema teorico generale di grande interesse si ha quando la famiglia \mathcal{G} di eventi generatori della σ -algebra \mathcal{F} è una semi-algebra su Ω . Supponiamo che su \mathcal{G} sia definita una funzione non negativa $\mathbb{P}_{\mathcal{G}}$ numerabilmente additiva e tale che $\mathbb{P}_{\mathcal{G}}(\Omega) = 1$.³ Ci si chiede se sia possibile prolungare, in modo univoco e conservando l'additività numerabile, tale funzione alla σ -algebra \mathcal{F} generata da \mathcal{G} ottenendo in tal modo una misura di probabilità \mathbb{P} . La risposta a questo quesito è affermativa come illustrato dai seguenti risultati.

Teorema 3.10 (Primo teorema di estensione). *Supponiamo che la famiglia di eventi generatori \mathcal{G} sia una semi-algebra su Ω e indichiamo con \mathcal{A} l'algebra da essa generata. Sia $\mathbb{P}_{\mathcal{G}}$ una funzione non negativa definita su \mathcal{G} numerabilmente additiva e tale che $\mathbb{P}_{\mathcal{G}}(\Omega) = 1$. Allora esiste ed è unico un prolungamento non negativo $\mathbb{P}_{\mathcal{A}}$ di $\mathbb{P}_{\mathcal{G}}$ da \mathcal{G} a \mathcal{A} che conserva la proprietà di additività numerabile.*

Dimostrazione. Si fornirà una dimostrazione suddivisa in quattro passi.

a) *Definizione del prolungamento*

Sia A un elemento qualsiasi di \mathcal{A} . Il Teorema 3.4 garantisce l'esistenza di un intero n e di una n -upla di elementi di \mathcal{G} a due a due disgiunti, diciamo $\{G_1, G_2, \dots, G_n\}$, per il quale si ha $A = \bigcup_{i=1}^n G_i$. Di conseguenza possiamo porre:

$$\mathbb{P}_{\mathcal{A}}(A) = \sum_{i=1}^n \mathbb{P}_{\mathcal{G}}(G_i). \quad (3.27)$$

La definizione è ben posta in quanto $\mathbb{P}_{\mathcal{A}}(A)$ non dipende dalla partizione $\{G_1, G_2, \dots, G_n\}$ di A . Infatti, se $\{G'_1, G'_2, \dots, G'_m\}$ è un'altra partizione finita di A costituita da m elementi \mathcal{G} , si ha:

$$\begin{aligned} \mathbb{P}_{\mathcal{A}}(A) &= \sum_{j=1}^m \mathbb{P}_{\mathcal{G}}(G'_j) = \sum_{j=1}^m \mathbb{P}_{\mathcal{G}}(G'_j \cap A) = \sum_{j=1}^m \mathbb{P}_{\mathcal{G}}\left(G'_j \cap \bigcup_{i=1}^n G_i\right) \\ &= \sum_{j=1}^m \mathbb{P}_{\mathcal{G}}\left(\bigcup_{i=1}^n G'_j \cap G_i\right) = \sum_{j=1}^m \sum_{i=1}^n \mathbb{P}_{\mathcal{G}}(G'_j \cap G_i) = \sum_{i=1}^n \sum_{j=1}^m \mathbb{P}_{\mathcal{G}}(G_i \cap G'_j) \\ &= \sum_{i=1}^n \mathbb{P}_{\mathcal{G}}\left(\bigcup_{j=1}^m G_i \cap G'_j\right) = \sum_{i=1}^n \mathbb{P}_{\mathcal{G}}\left(G_i \cap \bigcup_{j=1}^m G'_j\right) = \sum_{i=1}^n \mathbb{P}_{\mathcal{G}}(G_i \cap A) \\ &= \sum_{i=1}^n \mathbb{P}_{\mathcal{G}}(G_i). \end{aligned}$$

³Si osservi che richiedere ad una funzione di insiemi di essere numerabilmente additiva quando il suo dominio \mathcal{D} non è stabile rispetto all'unione numerabile significa limitarsi a considerare solo le successioni la cui unione appartiene a \mathcal{D} .

3.5 Il problema dell'estensione

b) *Non negatività del prolungamento*

Per ogni $A \in \mathcal{A}$ il valore della funzione $\mathbb{P}_{\mathcal{A}}(A)$ definito nella (3.27) è non negativo in quanto somma di quantità non negative. Questo dimostra la non negatività del prolungamento $\mathbb{P}_{\mathcal{A}}$.

c) *Numerabile additività del prolungamento*

Consideriamo una successione $\{A_j\}_{j \in \mathbb{N}}$ costituita da elementi di \mathcal{A} a due a due disgiunti. Posto $A = \bigcup_{j=1}^{\infty} A_j$, si supponga che $A \in \mathcal{A}$. Allora esiste una partizione finita di A costituita da elementi di \mathcal{G} : $A = \bigcup_{k=1}^n G_k$. Lo stesso vale per ognuno degli elementi A_j : $A_j = \bigcup_{i=1}^{n_j} G_{ji}$. Ricordando l'ipotesi che $\mathbb{P}_{\mathcal{G}}$ è numerabilmente additiva su \mathcal{G} e che \mathcal{G} è una semi-algebra cosicché le intersezioni $G_k \cap G_{ji}$ appartengono a \mathcal{G} , si ha:

$$\begin{aligned} \mathbb{P}_{\mathcal{A}}\left(\bigcup_{j=1}^{\infty} A_j\right) &= \mathbb{P}_{\mathcal{A}}(A) = \sum_{k=1}^n \mathbb{P}_{\mathcal{G}}(G_k) = \sum_{k=1}^n \mathbb{P}_{\mathcal{G}}(A \cap G_k) = \sum_{k=1}^n \mathbb{P}_{\mathcal{G}}\left(\bigcup_{j=1}^{\infty} A_j \cap G_k\right) \\ &= \sum_{k=1}^n \mathbb{P}_{\mathcal{G}}\left(\bigcup_{j=1}^{\infty} \bigcup_{i=1}^{n_j} (G_{ji} \cap G_k)\right) = \sum_{k=1}^n \sum_{j=1}^{\infty} \sum_{i=1}^{n_j} \mathbb{P}_{\mathcal{G}}(G_{ji} \cap G_k) \\ &= \sum_{j=1}^{\infty} \sum_{k=1}^n \sum_{i=1}^{n_j} \mathbb{P}_{\mathcal{G}}(G_{ji} \cap G_k) = \sum_{j=1}^{\infty} \sum_{k=1}^n \mathbb{P}_{\mathcal{G}}\left(\bigcup_{i=1}^{n_j} G_{ji} \cap G_k\right) \\ &= \sum_{j=1}^{\infty} \sum_{k=1}^n \mathbb{P}_{\mathcal{G}}(A_j \cap G_k) = \sum_{j=1}^{\infty} \mathbb{P}_{\mathcal{A}}(A_j \cap A) = \sum_{j=1}^{\infty} \mathbb{P}_{\mathcal{A}}(A_j). \end{aligned}$$

d) *Unicità del prolungamento*

Supponiamo che esiste un altro prolungamento $\mathbb{P}'_{\mathcal{A}}$ non negativo e numerabilmente additivo di $\mathbb{P}_{\mathcal{G}}$. Questo significa, in particolare, che $\mathbb{P}'_{\mathcal{A}}(G) = \mathbb{P}_{\mathcal{G}}(G) = \mathbb{P}_{\mathcal{A}}(G)$ per ogni $G \in \mathcal{G}$. Sia $A \in \mathcal{A}$ e sia $\{G_1, G_2, \dots, G_n\}$ una partizione finita di A effettuata con elementi di \mathcal{G} . Si ha

$$\mathbb{P}'_{\mathcal{A}}(A) = \mathbb{P}'_{\mathcal{A}}\left(\bigcup_{i=1}^n G_i\right) = \sum_{i=1}^n \mathbb{P}'_{\mathcal{A}}(G_i) = \sum_{i=1}^n \mathbb{P}_{\mathcal{A}}(G_i) = \mathbb{P}_{\mathcal{A}}\left(\bigcup_{i=1}^n G_i\right) = \mathbb{P}_{\mathcal{A}}(A). \quad \square$$

Il seguente teorema (ottenuto da Carathéodory in una formulazione più generale) completa la ricerca del prolungamento di una funzione numerabilmente additiva da una semi-algebra alla σ -algebra da essa generata⁴.

Teorema 3.11 (Teorema di estensione). *Supponiamo che la famiglia di eventi generatori \mathcal{G} sia una semi-algebra su Ω e indichiamo con \mathcal{F} la σ -algebra da essa generata. Sia $\mathbb{P}_{\mathcal{G}}$ una funzione non negativa definita su \mathcal{G} numerabilmente additiva e tale che $\mathbb{P}_{\mathcal{G}}(\Omega) = 1$.*

⁴La dimostrazione di questo risultato si trova, ad esempio, nel libro di L.E. Clarke dal titolo *Random variables* pubblicato da Longman Mathematical Texts nel 1975.

Allora esiste ed è unico un prolungamento non negativo \mathbb{P}_F di \mathbb{P}_G da \mathcal{G} a \mathcal{F} che conserva la proprietà di additività numerabile.

Terminiamo questo paragrafo con due esempi che elucidano l'importanza dei teoremi di estensione.

Esempio 3.13 (La misura di Lebesgue sull'intervallo $[a, b]$). Siano a e b numeri reali con $a < b$. Consideriamo l'esperimento casuale \mathfrak{D} consistente nella scelta di un punto a caso nell'intervallo $[a, b]$ e nel quale si vogliono considerare come eventi tutti gli intervalli in esso contenuti. Lo spazio campione e la σ -algebra associati a \mathfrak{D} coincidono, rispettivamente, con l'intervallo $[a, b]$ e con $\mathcal{B}_{[a, b]}$ che abbiamo introdotto nell'Esempio 3.7.

Per quanto riguarda la misura di probabilità la locuzione "scelta di un punto a caso" garantisce la presenza di simmetria al continuo e pertanto la probabilità di un qualsiasi intervallo I contenuto in $[a, b]$ è data dal rapporto della sua lunghezza rispetto a quella dell'intervallo $[a, b]$. Nel seguito indicheremo con $l(I)$ la lunghezza di un qualsiasi intervallo I .

Nell'esempio citato si è affermato che la famiglia $\mathcal{G} = \{[r, s] : a \leq r < s \leq b\}$ genera la σ -algebra $\mathcal{B}([a, b])$; d'altra parte \mathcal{G} è una semi-algebra su $[a, b]$ (si lascia al lettore) e la funzione

$$\mathbb{P}_G([r, s]) = \frac{l([r, s])}{l([a, b])} = \frac{s - r}{b - a}$$

risulta essere non negativa e numerabilmente additiva (si lascia al lettore).

L'algebra $\mathcal{A} = a(\mathcal{G})$ coincide con la famiglia dei sottoinsiemi dell'intervallo $[a, b]$ che sono unioni finite di intervalli del tipo $[r, s] \subseteq [a, b]$; in altri termini:

$$\forall A \in \mathcal{A}, \exists n \in \mathbb{N} \text{ e } \exists \{[r_1, s_1], \dots, [r_n, s_n]\} \subseteq \mathcal{G} : A = \bigcup_{i=1}^n [r_i, s_i] \quad (3.28)$$

Il primo teorema di estensione fornisce la funzione non negativa e numerabilmente additiva definita su \mathcal{A} :

$$\mathbb{P}_A(A) = \sum_{i=1}^n \frac{l([r_i, s_i])}{l([a, b])} = \sum_{i=1}^n \frac{s_i - r_i}{b - a},$$

nella quale si è fatto uso della rappresentazione (3.28) del generico $A \in \mathcal{A}$.

Il Teorema di estensione garantisce inoltre l'esistenza di un unico prolungamento non negativo e numerabilmente additivo sulla σ -algebra $\mathcal{B}_{[a, b]} = \sigma(\mathcal{A}) \equiv \sigma(\mathcal{G})$. Dal fatto che $\mathbb{P}_G([a, b]) = 1$ tale prolungamento risulta essere una misura di probabilità che viene chiamata *misura di Lebesgue sull'intervallo $[a, b]$* e denotata con il simbolo $\lambda_{[a, b]}$. \triangleleft

Esempio 3.14 (La misura di Lebesgue-Stieltjes). Consideriamo lo spazio probabilizzabile $(\mathbb{R}, \mathcal{B})$ e sia $F: \mathbb{R} \rightarrow [0, 1]$ una funzione non decrescente, continua a destra e tale che

3.6 Il teorema di equivalenza

$\lim_{x \rightarrow -\infty} F(x) = 0$ e $\lim_{x \rightarrow +\infty} F(x) = 1$. Indichiamo, inoltre, con \mathcal{G} la famiglia di sottoinsiemi di \mathbb{R} costituita da \mathbb{R} stesso, da tutte le semirette sinistre chiuse e destre aperte nonché da tutti gli intervalli limitati aperti a sinistra e chiusi a destra e poniamo

$$\forall x \in \mathbb{R}, \quad \mathbb{P}_G((-\infty, x]) = F(x), \quad (3.29)$$

$$\forall y \in \mathbb{R}, \quad \mathbb{P}_G([y, +\infty)) = 1 - F(y), \quad (3.30)$$

$$\forall (x, y) \in \mathbb{R}^2 : x < y, \quad \mathbb{P}_G([x, y]) = F(y) - F(x). \quad (3.31)$$

La famiglia \mathcal{G} , che è immediato riconoscere essere una semi-algebra su \mathbb{R} , contiene la famiglia \mathcal{G}_3 dell'Esempio 3.6 e pertanto $\sigma(\mathcal{G}) = \mathcal{B}$. Si lascia al lettore verificare che la funzione \mathbb{P}_G definita nelle (3.29)-(3.31) è non negativa e numerabilmente additiva. Applicando allora i teoremi di estensione possiamo affermare che esiste un unico prolungamento numerabilmente additivo di \mathbb{P}_G alla σ -algebra \mathcal{B} . Risultando, inoltre,

$$\begin{aligned} \forall x \in \mathbb{R}, \quad \mathbb{P}_G(\mathbb{R}) &= \mathbb{P}_G((-\infty, x] \cup [x, +\infty)) = \mathbb{P}_G((-\infty, x]) + \mathbb{P}([x, +\infty)) \\ &= F(x) + 1 - F(x) = 1, \end{aligned}$$

tal prolungamento risulta essere una misura di probabilità su $(\mathbb{R}, \mathcal{B})$ che viene chiamata *misura di Lebesgue-Stieltjes*; essa si designa con il simbolo λ_S . \triangleleft

3.6 Il teorema di equivalenza

Con riferimento ad uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, passiamo ad esaminare un'importante questione, più volte menzionata in precedenza, riguardante la proprietà di additività. Data una funzione \mathbb{P} definita su \mathcal{F} , ricordiamo che per essa sono definibili due proprietà di additività: la finita e la numerabile. Come abbiamo precedentemente dimostrato (cfr. il Teorema 3.5) l'additività numerabile implica l'additività finita. Non è invece vero il viceversa. Introduciamo ora un'ulteriore proprietà per una funzione definita in \mathcal{F} e a valori in \mathbb{R}

Definizione 3.16. Consideriamo una funzione $\mathbb{T}: \mathcal{F} \rightarrow \mathbb{R}$. Se per ogni successione di eventi $\{A_n\}_{n \in \mathbb{N}}$ di eventi dotata di limite, ossia tale che esiste un $A \in \mathcal{F}$ per il quale si ha $A = \lim_{n \rightarrow \infty} A_n = \liminf_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n$, risulta

$$\lim_{n \rightarrow \infty} \mathbb{T}(A_n) = \mathbb{T}\left(\lim_{n \rightarrow \infty} A_n\right) = \mathbb{T}(A),$$

allora si dice che \mathbb{T} verifica la proprietà di continuità. \triangleleft

Sussiste il seguente teorema.

Teorema 3.12 (Teorema di equivalenza). *Sia (Ω, \mathcal{F}) uno spazio probabilizzabile. Una funzione di insieme $\mathbb{P}: \mathcal{F} \rightarrow \mathbb{R}$ soddisfacente gli assiomi (i) e (ii) della Definizione 3.14 verifica la proprietà di additività numerabile se e solo se essa verifica sia la proprietà di additività finita che la proprietà di continuità.*

Dimostrazione. Sia \mathbb{P} numerabilmente additiva. Per il Teorema 3.5 essa è anche finitamente finita.

a) Sia $\{B_n\}_{n \in \mathbb{N}}$ una successione di eventi decrescente all'evento impossibile. Per ogni $n \in \mathbb{N}$, l'evento B_n può scriversi come

$$B_n = \bigcup_{k=n}^{\infty} (B_k \cap \bar{B}_{k+1}). \quad (3.32)$$

Se, infatti, un elemento ω appartiene a B_n , poiché la successione decrescente $\{B_n\}_{n \in \mathbb{N}}$ converge all'evento impossibile \emptyset esiste un evento B_{k+1} , con $k \geq n$, che non contiene ω . Allora ω appartiene sia a B_k che al complementare di B_{k+1} , così che risulta $\omega \in B_k \cap \bar{B}_{k+1}$. Ogni elemento ω di B_n è dunque contenuto nell'unione a secondo membro della (3.32). Viceversa, se un elemento ω appartiene a tale unione, esso certamente appartiene ad uno degli eventi B_k ($k \geq n$) e quindi anche a B_n per la decrescenza della successione $\{B_n\}_{n \in \mathbb{N}}$. La (3.32) è dunque vera. Dalla numerabile additività di \mathbb{P} e dalla (3.32) scritta per $n = 1$ si ottiene

$$\mathbb{P}(B_1) = \mathbb{P}\left[\bigcup_{k=1}^{\infty} (B_k \cap \bar{B}_{k+1})\right] = \sum_{k=1}^{\infty} \mathbb{P}(B_k \cap \bar{B}_{k+1}) < \infty.$$

Ne segue:

$$\lim_{n \rightarrow \infty} \mathbb{P}(B_n) = \lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} \mathbb{P}(B_k \cap \bar{B}_{k+1}) = 0, \quad (3.33)$$

in quanto $\sum_{k=n}^{\infty} \mathbb{P}(B_k \cap \bar{B}_{k+1})$ è il resto n -esimo di una serie convergente.

b) Sia $\{D_n\}_{n \in \mathbb{N}}$ una successione di eventi decrescente e sia D il suo limite. Indicato, $\forall n \in \mathbb{N}$, con $B_n = D_n \cap \bar{D}$ l'evento D_n si scrive come unione di due eventi incompatibili nel seguente modo:

$$D_n = (D_n \cap D) \cup (D_n \cap \bar{D}) = D \cup B_n.$$

Passando alle probabilità e al limite per $n \rightarrow \infty$, dalla finita additività di \mathbb{P} e dalla (3.33) si ottiene:

$$\lim_{n \rightarrow \infty} \mathbb{P}(D_n) = \mathbb{P}(D) + \lim_{n \rightarrow \infty} \mathbb{P}(B_n) = \mathbb{P}(D). \quad (3.34)$$

c) Sia $\{C_n\}_{n \in \mathbb{N}}$ una successione di eventi decrescente e sia C il suo limite. Dalle relazioni di De Morgan discende che $\{\bar{C}_n\}_{n \in \mathbb{N}}$ decresce a \bar{C} (si lascia al lettore). Quindi, per la (3.34) possiamo affermare che

$$\lim_{n \rightarrow \infty} \mathbb{P}(\bar{C}_n) = \mathbb{P}(\bar{C}),$$

3.6 Il teorema di equivalenza

da cui segue:

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(C_n) &= \lim_{n \rightarrow \infty} [1 - \mathbb{P}(\bar{C}_n)] = 1 - \lim_{n \rightarrow \infty} \mathbb{P}(\bar{C}_n) = 1 - \mathbb{P}(\bar{C}) \\ &= \mathbb{P}(C). \end{aligned} \quad (3.35)$$

d) Supponiamo, infine, che $\{A_n\}_{n \in \mathbb{N}}$ sia una qualsiasi successione di eventi dotata di limite:

$$\liminf_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n = A \Leftrightarrow \lim_{n \rightarrow \infty} \bigcap_{k=n}^{\infty} A_k = \lim_{n \rightarrow \infty} \bigcup_{k=n}^{\infty} A_k = A.$$

Applicando la (3.34) e la (3.35), rispettivamente, alle successioni $\{\cup_{k=n}^{\infty} A_k\}_{n \in \mathbb{N}}$ e $\{\cap_{k=n}^{\infty} A_k\}_{n \in \mathbb{N}}$, si può affermare che

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k\right) = \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcup_{k=n}^{\infty} A_k\right) = \mathbb{P}(A). \quad (3.36)$$

Inoltre, dall'essere $\bigcap_{k=n}^{\infty} A_k \subseteq A_n \subseteq \bigcup_{k=n}^{\infty} A_k$, la proprietà di monotonia di \mathbb{P} (cfr. Corollario 3.3) ci permette di scrivere anche che

$$\mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k\right) \leq \mathbb{P}(A_n) \leq \mathbb{P}\left(\bigcup_{k=n}^{\infty} A_k\right). \quad (3.37)$$

Infine, facendo uso del teorema del confronto per successioni numeriche dalle (3.36) e (3.37) si ricava che

$$\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \mathbb{P}(A). \quad (3.38)$$

Quest'ultima ci consente di concludere che se \mathbb{P} è numerabilmente additiva allora essa verifica anche la proprietà di continuità.

Dimostriamo ora l'implicazione inversa del teorema, ossia che se la funzione \mathbb{P} verifica le proprietà di additività finita e di continuità allora essa è anche numerabilmente additiva. Consideriamo, a tal fine, una successione $\{A_n\}_{n \in \mathbb{N}}$ di eventi a due a due incompatibili. Osserviamo che, per ogni intero k , si può scrivere:

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

in cui gli eventi in parentesi tonde a secondo membro sono evidentemente incompatibili.

Facendo uso della proprietà di additività finita si ottiene allora:

$$\begin{aligned} \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) &= \mathbb{P}\left(\bigcup_{n=1}^k A_n\right) + \mathbb{P}\left(\bigcup_{n=k+1}^{\infty} A_n\right) \\ &= \sum_{n=1}^k \mathbb{P}(A_n) + \mathbb{P}\left(\bigcup_{n=k+1}^{\infty} A_n\right). \end{aligned} \quad (3.39)$$

Si osservi ora che se $\omega \in \bigcup_{n=k+1}^{\infty} A_n$ esiste un evento A_r , con $r > k$, tale che $\omega \in A_r$, così che ω non appartiene ad alcuno degli eventi $\bigcup_{n=j+1}^{\infty} A_n$, per $j \geq r$. La successione $\{\bigcup_{n=k+1}^{\infty} A_n\}_{k \in \mathbb{N}}$ esclude dunque via via tutti gli elementi di Ω , e quindi decresce all'evento impossibile. Dalla proprietà di continuità di \mathbb{P} si ha allora:

$$\lim_{k \rightarrow \infty} \mathbb{P}\left(\bigcup_{n=k+1}^{\infty} A_n\right) = \mathbb{P}(\emptyset) = 0,$$

da cui, insieme alla (3.39), segue la numerabile additività di \mathbb{P} :

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

Gli esempi che seguono illustrano, esclusivamente a scopo didattico, la modalità di applicazione della proprietà di continuità.

Esempio 3.15. Riferiamoci di nuovo allo spazio di Bernoulli Ω costituito dalle successioni di T e C e alla σ -algebra \mathcal{F}_1 dei suoi eventi definiti nell'Esempio 3.4. Si è già mostrato che la proposizione A : "due teste consecutive si presentano prima di due croci consecutive" è un elemento di \mathcal{F}_1 e quindi è un evento. Allo stesso modo si dimostra essere un evento la proposizione B : "due croci consecutive si presentano prima di due teste consecutive". Vogliamo adesso fare uso della proprietà di continuità per dimostrare che è nulla la probabilità dell'evento $\overline{A \cup B}$. A tal fine introduciamo le proposizioni U : "testa si presenta nei posti pari e croce nei posti dispari" e la proposizione V : "croce si presenta nei posti pari e testa nei posti dispari". Sia U che V sono rappresentate insiemisticamente da un singoletto di Ω e quindi sono eventi. Di più si ha che se non si presenta né A e né B allora vuol dire che si è presentato U oppure V . In termini insiemistici $\overline{A \cup B} = U \cup V$, di modo che la proprietà di finita additività fornisce $\mathbb{P}(\overline{A \cup B}) = \mathbb{P}(U \cup V) = \mathbb{P}(U) + \mathbb{P}(V)$.

Consideriamo, ora, la successione $\{U_n\}_{n \in \mathbb{N}}$ il cui primo elemento è $U_1 = C_1$, ed i cui elementi di indice maggiore di 1 sono così definiti:

$$U_n = \begin{cases} U_{n-1} \cap C_n, & \text{per } n \text{ dispari,} \\ U_{n-1} \cap T_n, & \text{per } n \text{ pari.} \end{cases}$$

3.7 Eventi quasi impossibili ed eventi quasi certi

Notiamo che $\{U_n\}_{n \in \mathbb{N}}$ è decrescente e che il suo limite è proprio l'evento U sopra introdotto, ossia $\lim_{n \rightarrow \infty} U_n = \bigcap_{n=1}^{\infty} U_n = U$. Per quanto già detto nell'Esempio 3.12 e, osservando che l'evento U_n ha i suoi primi n simboli fissati, risulta $\mathbb{P}(U_n) = 2^{-n}$. La proprietà di continuità della probabilità fornisce allora:

$$\mathbb{P}(U) = \lim_{n \rightarrow \infty} \mathbb{P}(U_n) = \lim_{n \rightarrow \infty} 2^{-n} = 0. \quad (3.40)$$

Ragionando in maniera analoga si dimostra poi che risulta $\mathbb{P}(V) = 0$. In definitiva, si ottiene $\mathbb{P}(\overline{A \cup B}) = \mathbb{P}(U \cup V) = \mathbb{P}(U) + \mathbb{P}(V) = 0$.

Prima di concludere questo esempio risulta utile osservare che il risultato $\mathbb{P}(U) = 0$ è coerente con l'affermazione che tutti i singoletti di Ω hanno probabilità nulla, fatta all'inizio dell'Esempio 3.12. \square

Esempio 3.16. Riprendiamo la situazione illustrata nell'Esempio 3.12. In esso consideriamo la successione $\{A_n\}_{n \in \mathbb{N}}$ di eventi definita dalla posizione:

$$\forall n \in \mathbb{N}, \quad A_n = \overline{\bigcup_{k=1}^n E_k}.$$

In primo luogo tale successione risulta essere decrescente e quindi dotata di limite. Inoltre, applicando una delle relazioni di De Morgan è agevole rendersi conto che A_n è quell'evento che presenta il simbolo C ai primi n posti, così che la successione $\{A_n\}_{n \in \mathbb{N}}$ decresce all'evento costituito dall'unica sequenza infinita di T e C che presenta C in ogni posto, che abbiamo già denotato con E_0 . D'altra parte, applicando il Teorema 3.5 ed il suo Corollario 3.1, si ottiene

$$\mathbb{P}(A_n) = 1 - \sum_{k=1}^n \frac{1}{2^k} = 1 - \left(1 - \frac{1}{2^n}\right) = \frac{1}{2^n}$$

dalla quale segue

$$\mathbb{P}(E_0) = \mathbb{P}\left(\lim_{n \rightarrow \infty} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \lim_{n \rightarrow \infty} \frac{1}{2^n} = 0.$$

Si osservi che tale risultato è coerente con l'assegnazione iniziale della probabilità di E_0 di cui all'Esempio 3.12. \square

3.7 Eventi quasi impossibili ed eventi quasi certi

Il Lemma 3.2 esprime la circostanza che l'evento impossibile ha probabilità di occorrenza nulla. È bene però precisare che non è detto che un evento a probabilità zero coincida

con l'evento impossibile. Abbiamo già incontrato eventi di questo tipo. Precisamente, nell'Esempio 3.12 è nulla la probabilità di E_0 (e di tutti gli eventi rappresentati da un singololetto dello spazio di Bernoulli) e nell'Esempio 3.15 è nulla la probabilità dell'evento $\overline{A \cup B}$. Analogamente, un evento avente probabilità 1 non coincide necessariamente con l'evento certo Ω ; a tale scopo basta pensare all'evento $A \cup B$ complementare del secondo evento richiamato sopra.

Sia $(\Omega, \mathcal{F}, \mathbb{P})$ uno spazio di probabilità. Cominciamo col dare le seguenti definizioni.

Definizione 3.17. Un evento I si dice quasi impossibile se risulta $\mathbb{P}(I) = 0$. \square

Definizione 3.18. Un evento C si dice quasi certo se risulta $\mathbb{P}(C) = 1$. \square

Osservazione 3.7. Si noti che se per ogni $\omega \in \Omega$ risulta definita e positiva $\mathbb{P}(\{\omega\})$, l'essere $\mathbb{P}(C) = 1$ implica $C = \Omega$, mentre l'essere $\mathbb{P}(I) = 0$ implica $I = \emptyset$. \square

Osservazione 3.8. Il Corollario 3.1 comporta che il complementare di un evento quasi certo è un evento quasi impossibile, e viceversa. \square

Presenteremo adesso alcuni risultati che trasferiscono in probabilità agli eventi quasi impossibili e quasi certi le proprietà di assorbimento e di unità godute da \emptyset e da Ω .

Proposizione 3.2. Sia I un evento quasi impossibile. Per ogni $A \in \mathcal{F}$ si ha $\mathbb{P}(A \cap I) = 0$.

Dimostrazione. Dal fatto che $A \cap I \supseteq I$, la proprietà di monotonia di \mathbb{P} garantisce che $\mathbb{P}(A \cap I) \leq \mathbb{P}(I) = 0$; la tesi discende dal Corollario 3.2. \square

Proposizione 3.3. Sia C un evento quasi certo. Per ogni $A \in \mathcal{F}$ si ha $\mathbb{P}(A \cup C) = 1$.

Dimostrazione. Dal fatto che $A \cup C \supseteq C$, la proprietà di monotonia di \mathbb{P} garantisce che $\mathbb{P}(A \cup C) \geq \mathbb{P}(C) = 1$; la tesi discende dal Corollario 3.2. \square

Proposizione 3.4. Sia I un evento quasi impossibile. Per ogni $A \in \mathcal{F}$ si ha $\mathbb{P}(A \cup I) = \mathbb{P}(A)$.

Dimostrazione. Utilizziamo la relazione $A \cup I = A \cup (I \cap \overline{A})$ che esprime l'unione di A e I in termini di eventi incompatibili. Dalla finita additività di \mathbb{P} si trae $\mathbb{P}(A \cup I) = \mathbb{P}(A) + \mathbb{P}(I \cap \overline{A})$. La tesi segue dalla Proposizione 3.2 in virtù della quale sappiamo che $\mathbb{P}(I \cap \overline{A}) = 0$. \square

Proposizione 3.5. Sia C un evento quasi certo. Per ogni $A \in \mathcal{F}$ si ha $\mathbb{P}(A \cap C) = \mathbb{P}(A)$.

Dimostrazione. Utilizziamo la formula di partizione con l'evento certo per scrivere $A = (A \cap C) \cup (A \cap \overline{C})$. Dalla finita additività di \mathbb{P} , si trae $\mathbb{P}(A) = \mathbb{P}(A \cap C) + \mathbb{P}(A \cap \overline{C})$. La tesi segue dall'Osservazione 3.8 e dalla Proposizione 3.3 in virtù della quale sappiamo che risulta $\mathbb{P}(A \cap \overline{C}) = 0$. \square

Osservazione 3.9. In questa osservazione indichiamo con un X insieme e con \mathcal{A} una σ -algebra su di esso. Nella *Teoria della misura*, la coppia (X, \mathcal{A}) si dice uno spazio misurabile e gli elementi di \mathcal{A} si dicono misurabili. Inoltre, una funzione non negativa μ definita su \mathcal{A} , numerabilmente additiva e che assegna il valore zero all'insieme vuoto si dice che è una *misura su* (X, \mathcal{A}) . La terna (X, \mathcal{A}, μ) è detta *spazio di misura*, nel senso che, ad ogni $A \in \mathcal{A}$ il valore $\mu(A)$ ne rappresenta allora la sua misura, in particolare, i misurabili aventi misura nulla vengono detti *insiemi nulli*: essi sono i corrispettivi degli eventi quasi impossibili che abbiamo definito nello spazio di probabilità. Dal momento che nulla vieta di avere $\mu(X) = +\infty$ non è possibile avere, in generale, i corrispettivi degli eventi quasi certi. Pertanto nello spazio di misura (X, \mathcal{A}, μ) , in alternativa, si dice che una proprietà Π è vera quasi ovunque se l'insieme $\{x \in X : \Pi(x) \text{ non è vera}\}$ è contenuto in un insieme nullo. Ad esempio, se \mathcal{A} contiene una topologia di X , una funzione reale definita in X può essere continua quasi ovunque in X ; un'altra funzione potrebbe essere quasi ovunque positiva in X . Nel seguito, quando utilizzeremo tale locuzione, ci riferiremo esclusivamente allo spazio di misura $(\mathbb{R}, \mathcal{B}, \lambda)$ dove λ è la misura di Lebesgue su \mathbb{R} , ossia l'unico prolungamento numerabilmente additivo della lunghezza degli intervalli. Ovviamente λ non è una misura di probabilità dal momento che $\lambda(\mathbb{R}) = +\infty$. \square

CAPITOLO 4

Indipendenza e Condizionamento

4.1 Indipendenza di eventi

Fondamentale nel Calcolo delle Probabilità è il concetto di indipendenza di eventi, che ne costituisce una caratterizzazione non riscontrabile specificatamente, ad esempio, nella teoria della misura sulla quale in gran parte il modello calcolo delle probabilità è fondato. Vedremo che il concetto di indipendenza è strettamente collegato a quello di probabilità condizionata; da entrambi questi concetti traggono origine alcuni tra i più significativi e secondi campi di indagine ed applicazioni di questa disciplina tra i quali rientrano le teorie dei campioni casuali, delle martingale e dei processi markoviani. Al solito, in quanto segue intenderemo fissato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e per evento intenderemo sempre un elemento di \mathcal{F} .

Definizione 4.1. Siano A e B due eventi; se risulta

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

allora essi si dicono indipendenti. \triangleleft

Da questa definizione scaturisce immediatamente che l'evento impossibile e l'evento certo sono indipendenti da un qualsiasi altro evento. La proposizione seguente fornisce un altro esempio a carattere generale di coppie di eventi indipendenti.

Proposizione 4.1. *Ogni evento A è indipendente da un qualsiasi evento quasi impossibile e da un qualsiasi evento quasi certo.*

Dimostrazione. Sia I un evento quasi impossibile. Dall'essere $\mathbb{P}(I) = 0$, discende immediatamente $\mathbb{P}(A)\mathbb{P}(I) = 0$. D'altra parte la Proposizione 3.2 fornisce $\mathbb{P}(A \cap I) = 0$. Risulta in definitiva $\mathbb{P}(A \cap I) = 0 = \mathbb{P}(A)\mathbb{P}(I)$ ossia l'indipendenza di A e I in base alla Definizione 4.1.

Sia C un evento quasi certo. Dall'essere $\mathbb{P}(C) = 1$, discende immediatamente $\mathbb{P}(A)\mathbb{P}(C) = \mathbb{P}(A)$. D'altra parte la Proposizione 3.5 fornisce $\mathbb{P}(A \cap C) = \mathbb{P}(A)$. Risulta in definitiva $\mathbb{P}(A) = \mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C)$ ossia l'indipendenza di A e C in base alla Definizione 4.1. \square

Il risultato seguente mostra come la relazione di indipendenza di due eventi si conserva con il passaggio agli eventi complementari.

Teorema 4.1. Se A e B sono eventi indipendenti,

- (i) gli eventi A e \bar{B} sono indipendenti;
- (ii) gli eventi B e \bar{A} sono indipendenti;
- (iii) gli eventi \bar{A} e \bar{B} sono indipendenti.

Dimostrazione. Dimostriamo, in primo luogo la (i). La formula di partizione di A con l'evento certo $\Omega = B \cup \bar{B}$ fornisce $A = (A \cap \bar{B}) \cup (A \cap B)$. Dalla proprietà di finita additività della probabilità \mathbb{P} e dall'ipotesi di indipendenza di A e B si ha allora:

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

Ricordando il Corollario 3.1, da quest'ultima si trae:

$$\mathbb{P}(A \cap \bar{B}) = \mathbb{P}(A)[1 - \mathbb{P}(B)] = \mathbb{P}(A)\mathbb{P}(\bar{B}).$$

Gli eventi A e \bar{B} sono dunque indipendenti. Per dimostrare la (ii) basta applicare scambiare i ruoli di A e B nelle dimostrazione precedente. Infine, la (iii) si ottiene applicando la (i) agli eventi indipendenti \bar{A} e B . \square

Osservazione 4.1. La definizione di indipendenza data, essendo basata su una relazione tra probabilità, non dipende dalla struttura degli eventi ma solo dalle loro probabilità. Assegnate diversamente queste probabilità l'indipendenza può pertanto venire a mancare. Da notare la sostanziale differenza con la nozione di incompatibilità, che è invece una proprietà intrinseca degli eventi. \triangleleft

Intuitivamente, l'indipendenza spesso si realizza o si verifica non attraverso l'imposizione di una relazione di probabilità, ma in virtù delle caratteristiche esibite dall'esperimento casuale. Per esempio nel caso dell'esperimento casuale consistente nel lancio di una coppia di dadi, l'uno dopo l'altro, è ragionevole ritenere che gli eventi relativi al lancio del primo dado

4.1 Indipendenza di eventi

siano indipendenti dagli eventi relativi al lancio del secondo perché non sussistono legami tra i risultati dei lanci dei due dadi nell'effettuazione dell'esperimento. Si parte, cioè, dalla constatazione dell'indipendenza delle due sottoprove (lancio del primo dado e lancio del secondo dado) in cui può pensarsi decomposto l'esperimento, e da ciò si trae la convinzione dell'indipendenza degli eventi. La traduzione in termini matematici di questo concetto intuitivo di indipendenza di eventi si realizza attraverso la Definizione 4.1 la cui giustificazione non può risiedere in considerazioni intuitive ma deve scaturire dalle implicazioni che da essa discendono nell'ambito della teoria delle probabilità.

Va notato che l'indipendenza delle sottoprove può essere utilizzata come punto di partenza per la definizione di indipendenza. Vi sono comunque eventi indipendenti che non sono associabili a sottoprove diverse. È ad esempio facile dimostrare che sono indipendenti gli eventi $\{5, 6\}$ e $\{2, 4, 6\}$ nel contesto illustrato nel corso dell'Esempio 3.11, pur non essendo possibile considerare il lancio scisso in due sottoprove alle quali appartengono, rispettivamente, i due eventi.

La nozione di indipendenza si può generalizzare al caso di una infinità numerabile di eventi.

Definizione 4.2. Sia $\{A_n\}_{n \in \mathbb{N}}$ una successione di eventi; se comunque si scelgono un intero k e una k -upla $A_{i_1}, A_{i_2}, \dots, A_{i_k}$ di elementi di $\{A_n\}_{n \in \mathbb{N}}$ risulta

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_k}).$$

allora si dice che la successione $\{A_n\}_{n \in \mathbb{N}}$ è costituita da eventi collettivamente indipendenti. \triangleleft

Osservazione 4.2. La Definizione 4.2 può essere applicata anche al caso di un numero finito di eventi. Infatti, ricordando che lo spazio campione Ω è l'elemento unità dell'intersezione insiemistica e che ad esso è assegnata probabilità unitaria, basterà completare la successione con una infinità di eventi certi. \triangleleft

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

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B), \quad \mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C), \quad \mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C) \quad (4.1)$$

e dalla condizione

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C). \quad (4.2)$$

Le condizioni (4.1) e (4.2) devono essere tutte soddisfatte affinché si possa affermare che i tre eventi sono collettivamente indipendenti. Invero, per eventi che non sono collettivamente indipendenti è possibile che sia soddisfatta la (4.2), ma che una tra le (4.1) sia violata. È anche possibile che avvenga il contrario, ossia che tre eventi a due a due indipendenti non siano collettivamente indipendenti.

Esempio 4.1. Consideriamo un esperimento casuale caratterizzato dai risultati ω_i con $i = 1, 2, 3, 4$, così che $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$. Inoltre, sia $\mathcal{F} = \mathbb{P}(\Omega)$; ciò significa che la famiglia degli eventi elementari è costituita dai sottoinsiemi $E_i = \{\omega_i\}$ per $i = 1, 2, 3, 4$. Assegnato ad ognuno di essi la probabilità $1/4$, sappiamo come estendere la probabilità ad un qualsiasi elemento di \mathcal{F} . Consideriamo i seguenti eventi:

$$A = \{\omega_1, \omega_2\} \quad B = \{\omega_1, \omega_3\} \quad C = \{\omega_1, \omega_4\}. \quad (4.3)$$

Chiaramente risulta

$$A \cap B = A \cap C = B \cap C = A \cap B \cap C = E_1$$

mentre, con riguardo alla probabilità, si ha:

$$\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}$$

e

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

Le condizioni (4.1) sono dunque soddisfatte, ma la (4.2) non lo è in quanto risulta

$$\mathbb{P}(A \cap B \cap C) \equiv \mathbb{P}(E_1) = \frac{1}{4} \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = \frac{1}{8}.$$

In definitiva, gli eventi (4.3) pur essendo indipendenti a coppie, non lo sono se vengono considerati collettivamente. \square

Teorema 4.2. Siano A_1, A_2, \dots, A_n eventi collettivamente indipendenti. Sostituendo ad uno di tali eventi il proprio complemento si ottiene ancora una n -upla di eventi collettivamente indipendenti.

Dimostrazione. Basta ricalcare la linea della dimostrazione della parte (i) del Teorema 4.1. \square

Osservazione 4.3. Applicando più volte il Teorema 4.2 possiamo dire che esso continua a valere anche se più di un evento della n -upla viene sostituito dal rispettivo complemento. \square

Esempio 4.2. Riferiamoci ancora una volta allo spazio di Bernoulli Ω costituito dalle successioni di T e C e alla σ -algebra \mathcal{F}_1 dei suoi eventi definiti nell'Esempio 3.4 e ricordiamo che in esso la famiglia \mathcal{G} degli eventi generatori è costituita dalla successione $\{T_n\}_{n \in \mathbb{N}}$. Il termine generale T_n rappresenta il sottoinsieme dello spazio di Bernoulli costituito da tutte le successioni di T e C aventi T nel posto n -esimo. Come abbiamo già avuto modo di osservare risulta $C_n = \overline{T_n}$.

4.1 Indipendenza di eventi

Supponiamo ora di non sapere se la moneta è regolare o meno. Posto $\mathbb{P}(T) = p$, la differenza $p - 1/2$ viene detta *trucco* della moneta. L'ipotesi, abbastanza ragionevole nel caso di lanci che non sottopongono la moneta a deformazioni macroscopiche, che ad ogni lancio ci si trovi nelle medesime condizioni conduce in primo luogo ad avere $\mathbb{P}(T_n) = p$; ovviamente, per l'evento complementare possiamo scrivere $\mathbb{P}(C_n) = 1 - p \equiv q$. Inoltre, la stessa ipotesi comporta anche l'indipendenza delle sottoprove nelle quali può essere scissa la prova complessiva, conducendo a ritenere indipendenti gli eventi che si riferiscono a sottoprove diverse. È quindi lecito ritenere che gli eventi generatori siano tra loro collettivamente indipendenti. Indicato allora, $\forall n \in \mathbb{N}$, con G_n^* l'evento T_n oppure il suo complementare C_n , per $n \geq 2$ risulta:

$$\mathbb{P}(G_1^* \cap G_2^* \cap \dots \cap G_n^*) = \mathbb{P}(G_1^*)\mathbb{P}(G_2^*) \dots \mathbb{P}(G_n^*). \quad (4.4)$$

Dopo di ciò, ci proponiamo di calcolare la probabilità dell'evento “due teste consecutive si presentano prima di due croci consecutive”, già introdotto nell'esempio citato e ivi denotato con A . Questo, come mostrato nella (3.1), può essere rappresentato come unione numerabile di eventi a due a due incompatibili ed esprimibili in termini di eventi generatori:

$$\begin{aligned} A = & (T_1 \cap T_2) \cup (C_1 \cap T_2 \cap T_3) \cup (T_1 \cap C_2 \cap T_3 \cap T_4) \cup \\ & \cup (C_1 \cap T_2 \cap C_3 \cap T_4 \cap T_5) \cup (T_1 \cap C_2 \cap T_3 \cap C_4 \cap T_5 \cap T_6 \cap T_7) \cup \\ & \cup (C_1 \cap T_2 \cap C_3 \cap T_4 \cap C_5 \cap T_6 \cap T_7) \cup \dots \end{aligned}$$

La proprietà di additività completa della probabilità fa ottenere:

$$\begin{aligned} \mathbb{P}(A) = & \mathbb{P}(T_1 \cap T_2) + \mathbb{P}(C_1 \cap T_2 \cap T_3) + \mathbb{P}(T_1 \cap C_2 \cap T_3 \cap T_4) + \\ & + \mathbb{P}(C_1 \cap T_2 \cap C_3 \cap T_4 \cap T_5) + \mathbb{P}(T_1 \cap C_2 \cap T_3 \cap C_4 \cap T_5 \cap T_6) + \\ & + \mathbb{P}(C_1 \cap T_2 \cap C_3 \cap T_4 \cap C_5 \cap T_6 \cap T_7) + \dots \end{aligned}$$

In virtù della (4.4), ed essendo $pq < 1$, si ha poi:

$$\begin{aligned} \mathbb{P}(A) &= p^2 + p^2q + p^2(pq) + p^2q(pq) + p^2(pq)^2 + p^2q(pq)^2 + \dots \\ &= p^2[1 + pq + (pq)^2 + \dots] + p^2q[1 + pq + (pq)^2 + \dots] \\ &= p^2(1 + q)[1 + pq + (pq)^2 + \dots] = \frac{p^2(1 + q)}{1 - pq}, \end{aligned}$$

ovvero:

$$\mathbb{P}(A) = \frac{p^2(2 - p)}{1 - p + p^2}. \quad (4.5)$$

Terminiamo l'esempio con la seguente osservazione. Nel caso di una moneta regolare, ossia a trucco nullo, risulta $p = q = 1/2$ che sostituto nella (4.5) fa ottenere:

$$\mathbb{P}(A) = \frac{1}{4} \cdot \frac{3}{2} \left(\frac{1}{2} + \frac{1}{4} \right)^{-1} = \frac{1}{2}.$$

Questo risultato era già stato ottenuto nell'Esempio 3.12. Esso può essere riottenuto anche con il seguente ragionamento. Nel caso della moneta non truccata, possiamo affermare, in primo luogo, che l'evento A è perfettamente simmetrico all'evento incompatibile B : "due croci consecutive si presentano prima di due teste consecutive" (basta scambiare il ruolo di T con quello di C). Tali due eventi devono quindi avere uguali probabilità di occorrenza. Inoltre, nell'Esempio 3.15, relativo ad una moneta regolare, è stato mostrato che $\mathbb{P}(\overline{A \cup B}) = 0$. Allora l'evento $A \cup B$ è quasi certo. Da queste due affermazioni segue immediatamente che $\mathbb{P}(A) = \mathbb{P}(B) = 1/2$. \square

4.2 Legge 0-1

Esamineremo tra breve alcuni importanti risultati riassunti nella cosiddetta *legge 0-1*. Essa costituisce un importante strumento per la determinazione di ulteriori risultati riguardanti il comportamento asintotico di notevoli successioni di eventi. Consideriamo innanzitutto il seguente celebre teorema noto col nome di *Lemma di Borel-Cantelli*.

Teorema 4.3 (Borel-Cantelli). *Sia $\{A_n\}_{n \in \mathbb{N}}$ è una successione di eventi.*

$$\text{Se } \sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty \quad \text{allora si ha} \quad \mathbb{P}\left(\limsup_{n \rightarrow \infty} A_n\right) = 0.$$

Dimostrazione. Applicando di seguito la proprietà di continuità della probabilità e la disugualanza di Boole, dalla (3.7) si ottiene:

$$\begin{aligned} \mathbb{P}\left(\limsup_{n \rightarrow \infty} A_n\right) &= \mathbb{P}\left(\lim_{n \rightarrow \infty} \bigcup_{k=n}^{\infty} A_k\right) = \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcup_{k=n}^{\infty} A_k\right) \\ &\leq \lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} \mathbb{P}(A_k) = 0. \end{aligned}$$

L'ultimo passaggio è giustificato dal fatto che, per ipotesi, la serie $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$ converge e quindi il suo resto n -esimo $\sum_{k=n}^{\infty} \mathbb{P}(A_k)$ è infinitesimo al divergere di n . La tesi segue dalla non negatività della probabilità. \square

Ricordando la (3.6) che rappresenta in simboli la definizione del limite superiore di $\{A_n\}_{n \in \mathbb{N}}$, il Lemma di Borel-Cantelli afferma che se la serie $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$ converge, allora è nulla la probabilità che per ogni n esista un intero $k \geq n$ tale che A_k si verifichi. In altri termini, se la serie $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$ converge, il verificarsi di infiniti eventi della successione $\{A_n\}$ è un evento quasi impossibile.

Teorema 4.4 (Legge 0-1). *Sia $\{A_n\}_{n \in \mathbb{N}}$ una successione costituita da eventi collettivamente indipendenti. Si ha:*

4.2 Legge 0-1

- (i) $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty \quad \text{se e solo se} \quad \mathbb{P}\left(\limsup_{n \rightarrow \infty} A_n\right) = 0;$
- (ii) $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty \quad \text{se e solo se} \quad \mathbb{P}\left(\limsup_{n \rightarrow \infty} A_n\right) = 1.$

Dimostrazione. La necessarietà della (i) coincide con il lemma di Borel-Cantelli; per dimostrare la necessarietà della (ii) supponiamo che la serie $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$ diverge. Applicando la proprietà di continuità della probabilità e facendo uso di una delle leggi di De Morgan, dalla (3.7) si ottiene:

$$\begin{aligned} \mathbb{P}\left(\limsup_{n \rightarrow \infty} A_n\right) &= \mathbb{P}\left(\lim_{n \rightarrow \infty} \bigcup_{k=n}^{\infty} A_k\right) = \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcup_{k=n}^{\infty} A_k\right) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}\left(\overline{\bigcap_{k=n}^{\infty} \overline{A}_k}\right) = 1 - \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcap_{k=n}^{\infty} \overline{A}_k\right). \end{aligned} \quad (4.6)$$

D'altra parte dall'ipotesi di collettiva indipendenza degli eventi della successione $\{A_n\}_{n \in \mathbb{N}}$ si ottiene:

$$\begin{aligned} \mathbb{P}\left(\bigcap_{k=n}^{\infty} \overline{A}_k\right) &= \mathbb{P}\left(\lim_{r \rightarrow \infty} \bigcap_{k=n}^r \overline{A}_k\right) = \lim_{r \rightarrow \infty} \mathbb{P}\left(\bigcap_{k=n}^r \overline{A}_k\right) \\ &= \lim_{r \rightarrow \infty} \prod_{k=n}^r \mathbb{P}(\overline{A}_k) = \prod_{k=n}^{\infty} \mathbb{P}(\overline{A}_k). \end{aligned} \quad (4.7)$$

La (4.6) e la (4.7) consentono allora di scrivere

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} A_n\right) = 1 - \lim_{n \rightarrow \infty} \prod_{k=n}^{\infty} \mathbb{P}(\overline{A}_k) = 1 - \lim_{n \rightarrow \infty} \prod_{k=n}^{\infty} [1 - \mathbb{P}(A_k)].$$

Da quest'ultima facendo uso della diseguaglianza $1 - x \leq e^{-x}$, valida per ogni $x \in \mathbb{R}$, si ottiene poi la seguente maggiorazione:

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} A_n\right) \geq 1 - \lim_{n \rightarrow \infty} \prod_{k=n}^{\infty} e^{-\mathbb{P}(A_k)} = 1 - \lim_{n \rightarrow \infty} \exp\left\{-\sum_{k=n}^{\infty} \mathbb{P}(A_k)\right\} = 1. \quad (4.8)$$

L'ultimo passaggio è giustificato dal fatto che la serie a termini positivi $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$ diverge (per ipotesi) e così si comporta anche il suo resto n -esimo $\sum_{k=n}^{\infty} \mathbb{P}(A_k)$ al divergere di n . La tesi segue dal fatto che la probabilità di un qualsiasi evento non può eccedere l'unità.

Le implicazioni inverse della (i) e della (ii) possono essere facilmente dimostrate per assurdo. Infatti, se risultasse $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ nel caso in cui $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 0$, come conseguenza della (ii) dedurremmo che $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 1$, giungendo così ad una contraddizione. Analogamente, se risultasse $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$ nel caso in cui $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 1$, dalla (i) dedurremmo che $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 0$, giungendo così nuovamente ad una contraddizione. \square

In virtù dell'interpretazione data dell'evento $\limsup_{n \rightarrow \infty} A_n$, la Legge 0-1 afferma che, nell'ipotesi in cui gli eventi A_n sono indipendenti, la probabilità che per ogni n esista un $k \geq n$ tale che A_k si verifica (ovvero che infiniti tra gli eventi A_n si verificano) è nulla se e solo se la serie $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$ converge. Tale probabilità è invece unitaria se e solo se la serie $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$ diverge. E bene osservare che nel caso (ii) l'ipotesi di indipendenza è fondamentale. Infatti, se gli eventi non sono indipendenti può accadere che $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ e che $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 0$, come è illustrato nell'esempio che segue.

Esempio 4.3. Consideriamo l'esperimento casuale consistente nella scelta di un punto a caso nell'intervallo $[0, 1]$. Sappiamo che esso è descritto dallo spazio di probabilità $([0, 1], \mathcal{B}_{[0,1]}, \lambda_{[0,1]})$ (cfr. Esempio 3.13). Consideriamo in tale spazio la successione di eventi $\{[0, 1/n]\}_{n \in \mathbb{N}}$. Gli eventi della successione non sono indipendenti a due a due (e quindi neanche collettivamente). Infatti, ad esempio, per gli eventi $[0, 1/2]$ e $[0, 1/3]$ risulta $\lambda_{[0,1]}([0, 1/2] \cap [0, 1/3]) = \lambda_{[0,1]}([0, 1/3]) = 1/3$ mentre $\lambda_{[0,1]}([0, 1/2]) \cdot \lambda_{[0,1]}([0, 1/3]) = 1/2 \cdot 1/3 = 1/6$. Inoltre, dall'essere

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} [0, 1/n] = \bigcap_{n=1}^{\infty} [0, 1/n] = \emptyset,$$

segue $\lambda_{[0,1]}(\limsup_{n \rightarrow \infty} A_n) = \lambda_{[0,1]}(\emptyset) = 0$. D'altra parte si ha:

$$\sum_{n=1}^{\infty} \lambda_{[0,1]}([0, 1/n]) = \sum_{n=1}^{\infty} \frac{1}{n} = \infty.$$

In definitiva, il limite superiore della successione considerata è un evento quasi impossibile pur risultando divergente la serie avente per termine generale la probabilità del termine generale della successione considerata. \square

Un altro semplice esempio che mostra quanto l'ipotesi di indipendenza sia essenziale nel caso (ii) è il seguente.

Esempio 4.4. Sia A un evento per il quale risulti $0 < \mathbb{P}(A) < 1$. Se si pone $A_1 = A_2 = \dots = A$, gli eventi della successione non sono indipendenti dato che $\mathbb{P}(A \cap A) \neq \mathbb{P}(A)\mathbb{P}(A)$. In questo caso si ha $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$. D'altra parte, risulta $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = \mathbb{P}(A) < 1$ essendo $\limsup_{n \rightarrow \infty} A_n = A$. \square

4.3 Probabilità condizionate

Giova osservare che se gli eventi della successione $\{A_n\}_{n \in \mathbb{N}}$ sono collettivamente indipendenti ed equiprobabili, con $\mathbb{P}(A_n) > 0$, la serie $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$ diverge, e quindi deve risultare che il limite superiore della successione $\{A_n\}_{n \in \mathbb{N}}$ è un evento quasi certo. L'esempio che segue applica proprio tale tipo di considerazione per mostrare un risultato notevole e a priori inaspettato.

Esempio 4.5. Riprendiamo il contesto applicativo illustrato nell'Esempio 4.2 ed in esso prendiamo in considerazione le successioni $\{T_n\}_{n \in \mathbb{N}}$ e $\{C_n\}_{n \in \mathbb{N}}$. Risulta:

$$\sum_{n=1}^{\infty} \mathbb{P}(T_n) = \sum_{n=1}^{\infty} p = \infty \quad \text{e} \quad \sum_{n=1}^{\infty} \mathbb{P}(C_n) = \sum_{n=1}^{\infty} (1-p) = \infty.$$

Per la divergenza di queste serie e per la collettiva indipendenza degli eventi delle successioni $\{T_n\}_{n \in \mathbb{N}}$ e $\{C_n\}_{n \in \mathbb{N}}$, dalla Legge 0-1 si deduce:

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} T_n\right) = \mathbb{P}\left(\limsup_{n \rightarrow \infty} C_n\right) = 1.$$

La probabilità che si verifichino infiniti degli eventi T_n o degli eventi C_n è dunque unitaria. In altri termini, comunque fissato un intero k grande a piacere, quasi certamente prima o poi si osserverà l'occorrenza di k teste e di altrettante croci.

4.3 Probabilità condizionate

Prima di introdurre l'importante concetto di probabilità condizionata, illustriamone le motivazioni con alcune semplici considerazioni. Riferiamoci all'esperimento che consiste nel lancio di due dadi e calcoliamo la probabilità del seguente evento A :

A : "la somma dei risultati è un multiplo di 4."

In termini insiemistici l'evento A coincide con l'insieme delle seguenti coppie ordinate:

$$A = \{(1, 3), (2, 2), (2, 6), (3, 1), (3, 5), (4, 4), (5, 3), (6, 2), (6, 6)\}.$$

Supposto che i dadi non siano truccati, gli eventi semplici costituiti dalle sopraelencate coppie di risultati sono tutti equiprobabili, così che la probabilità $\mathbb{P}(A)$ si può calcolare come rapporto tra numero di casi favorevoli e numero di casi possibili:

$$\mathbb{P}(A) = \frac{9}{36} = \frac{1}{4}.$$

Supponiamo ora che i due dadi non vengano lanciati contemporaneamente, ma uno dopo l'altro, e calcoliamo la probabilità dell'evento A sapendo che nel primo lancio si è verificato il seguente evento B :

B : "uscita del numero 3."

Calcoliamo, in altri termini, la probabilità di A condizionata dall'evento B . In questo caso, risultando $A = \{(3, 1), (3, 5)\}$, sempre nella ragionevole ipotesi di equiprobabilità, si ha:

$$\mathbb{P}_B(A) = \frac{2}{6} = \frac{1}{3}.$$

Infatti, dei sei possibili risultati del lancio del secondo dado soltanto due (uscita dei numeri 1 e 5) fanno sì che la somma sia multiplo di 4. Si è quindi ottenuto un risultato diverso dal precedente.

Con riferimento allo stesso esperimento inizialmente considerato in questo paragrafo, calcoliamo ora la probabilità del seguente evento C :

C : "la somma dei risultati è un multiplo di 3."

L'evento C è costituito dalle coppie

$$C = \{(1, 2), (1, 5), (2, 1), (2, 4), (3, 3), (3, 6), (4, 2), (4, 5), (5, 1), (5, 4), (6, 3), (6, 6)\}$$

e quindi

$$\mathbb{P}(C) = \frac{12}{36} = \frac{1}{3}.$$

Se poi calcoliamo la probabilità di C condizionata dall'evento B (che, come si è detto, consiste nell'uscita del numero 3), ragionando come prima ottieniamo:

$$\mathbb{P}_B(C) \equiv \mathbb{P}(\{(3, 3), (3, 6)\}) = \frac{2}{6} = \frac{1}{3}.$$

A differenza della situazione precedentemente esaminata, in questo caso la conoscenza del risultato del primo lancio dell'esperimento non influisce sulla probabilità dell'evento.

Daremo ora una definizione di probabilità condizionata che ci si attende debba in qualche modo rispecchiare i risultati sopra illustrati. In quello che segue supporremo che sia stato fissato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$.

Definizione 4.3. Sia $B \in \mathcal{F}$ un sottoinsieme di Ω tale che $\mathbb{P}(B) > 0$ e sia $A \in \mathcal{F}$ un evento fissato. Il rapporto, denotato indifferentemente con $\mathbb{P}_B(A)$ o con $\mathbb{P}(A|B)$,

$$\mathbb{P}_B(A) \equiv \mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad (4.9)$$

definisce la probabilità di A condizionata da B (detta anche probabilità di A dato B). \triangleleft

1
2
3

4.3 Probabilità condizionate

Si noti l'analogia tra la (4.9) e la relazione (2.7) tra frequenze ricavate nel § 2.2.

Proposizione 4.2. La posizione 4.9 considerata per ogni $A \in \mathcal{F}$ definisce una misura di probabilità su (Ω, \mathcal{F}) .

Dimostrazione. Bisogna far vedere che \mathbb{P}_B verifica i tre requisiti della Definizione 3.14 che per comodità del lettore qui riscriviamo:

(i) $\forall A \in \mathcal{F}$, $\mathbb{P}_B(A) \geq 0$;

(ii) $\mathbb{P}_B(\Omega) = 1$;

(iii) se $\{A_n\}_{n \in \mathbb{N}}$ è una successione di eventi a due a due incompatibili, allora

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

Dal momento che la (4.9) definisce la probabilità condizionata come rapporto di un numero non negativo, $\mathbb{P}(A \cap B)$, e di un numero positivo, $\mathbb{P}(B)$, la (i) è banalmente verificata. Supposto ora $A = \Omega$, la (4.9) assume la forma

$$\mathbb{P}_B(\Omega) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1,$$

così che anche la (ii) risulta verificata. Mostriamo, infine, che \mathbb{P}_B è numerabilmente additiva. Consideriamo all'uopo una successione $\{A_n\}_{n \in \mathbb{N}}$ di eventi a due a due incompatibili. Dalla Definizione (4.9) applicata con $A = \bigcup_{n=1}^{\infty} A_n$ si trae:

$$\begin{aligned} \mathbb{P}_B\left(\bigcup_{n=1}^{\infty} A_n\right) &\equiv \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n | B\right) = \frac{\mathbb{P}\left[\left(\bigcup_{n=1}^{\infty} A_n\right) \cap B\right]}{\mathbb{P}(B)} = \frac{\mathbb{P}\left[\bigcup_{n=1}^{\infty} (A_n \cap B)\right]}{\mathbb{P}(B)} \\ &= \frac{\sum_{n=1}^{\infty} \mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} = \sum_{n=1}^{\infty} \frac{\mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} = \sum_{n=1}^{\infty} \mathbb{P}(A_n | B) \\ &\equiv \sum_{n=1}^{\infty} \mathbb{P}_B(A_n). \end{aligned}$$

Resta così dimostrata anche la (iii). La funzione \mathbb{P}_B definita nella 4.9 è dunque una misura di probabilità su (Ω, \mathcal{F}) . \square

Osservazione 4.4. È stata proposta da diversi studiosi un'assiomatizzazione del Calcolo delle Probabilità che prende le mosse proprio dal concetto di probabilità condizionata. Una delle motivazioni è il superamento dell'impossibilità di definire la probabilità condizionata nel caso di un evento condizionante con probabilità nulla di occorrenza, onde soddisfare talune aspettative di carattere intuitivo. Un esempio è offerto dall'esperimento consistente nella scelta di un punto a caso in un fissato intervallo; in tale contesto, consideriamo gli eventi A_1 e A_2 che si verificano quando viene scelto il punto di ascissa x_1 oppure il punto di il punto di ascissa x_2 , rispettivamente. In tali condizioni è abbastanza intuitivo richiedere che sia $\mathbb{P}(A_1|A_1 \cup A_2) = \mathbb{P}(A_2|A_1 \cup A_2) = 1/2$; ma tale probabilità non può essere presa in considerazione nell'ambito della Definizione 4.3 dal momento che è ovvio riguardare come quasi impossibile l'evento $A_2 \cup A_2$. \triangleleft

Notiamo che se risulta $\mathbb{P}(A) > 0$, scambiando i ruoli di A e B nella (4.9) si può definire la probabilità condizionata di B dato A tramite la posizione:

$$\mathbb{P}_A(B) \equiv \mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}, \quad (4.10)$$

reminiscente della relazione (2.6) tra frequenze.

Tenendo conto degli esempi discussi all'inizio di questo paragrafo, ci si deve attendere che risulti $\mathbb{P}_B(A) = \mathbb{P}(A)$ se A e B sono eventi indipendenti. Ciò è effettivamente vero; si può anzi fornire una caratterizzazione degli eventi indipendenti tramite tre proprietà equivalenti (cf. anche la Definizione 4.1), come qui di seguito indicato.

Proposizione 4.3. Siano $A, B \in \mathcal{F}$ sottoinsiemi di Ω , con $\mathbb{P}(A) > 0$ e $\mathbb{P}(B) > 0$. Le seguenti affermazioni sono equivalenti:

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

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:

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

Dimostriamo che $(ii) \Rightarrow (iii)$. Dalla 4.9 si ricava $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$, cosicché:

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

4.3 Probabilità condizionate

Dimostriamo, infine, che $(iii) \Rightarrow (i)$. A tale scopo notiamo che dalla (4.10) e dall'ipotesi segue:

$$\mathbb{P}(A \cap B) = \mathbb{P}(B|A)\mathbb{P}(A) = \mathbb{P}(B)\mathbb{P}(A). \quad \square$$

La definizione di probabilità condizionata è suscettibile di immediata estensione. Siano, infatti, $B \in \mathcal{F}$ e $C \in \mathcal{F}$ tali che $\mathbb{P}(B) > 0$ e $\mathbb{P}(B \cap C) > 0$ (il che implica $\mathbb{P}_B(C) > 0$). Sappiamo che \mathbb{P}_B definisce una misura di probabilità su (Ω, \mathcal{F}) ; facendo uso della Definizione 4.3 con riferimento alla misura di probabilità \mathbb{P}_B invece che alla misura di probabilità \mathbb{P} , $\forall A \in \mathcal{F}$ si ha:

$$\begin{aligned} (\mathbb{P}_B)_C(A) &= \frac{\mathbb{P}_B(A \cap C)}{\mathbb{P}_B(C)} = \frac{\mathbb{P}[(A \cap C) \cap B]}{\mathbb{P}(B)} \frac{\mathbb{P}(B)}{\mathbb{P}(B \cap C)} \\ &= \frac{\mathbb{P}[A \cap (B \cap C)]}{\mathbb{P}(B \cap C)} = \mathbb{P}(A|B \cap C). \end{aligned} \quad (4.11)$$

La (4.11) definisce la probabilità dell'evento A condizionata dall'evento intersezione $B \cap C$. In modo analogo si definisce la probabilità di un evento condizionata dall'intersezione di un numero finito di eventi. Invero se $n \in \mathbb{N}$ e $B_1, \dots, B_n \in \mathcal{F}$ sono eventi tali che $\mathbb{P}(B_1 \cap \dots \cap B_n) > 0$, si pone:

$$\mathbb{P}(A|B_1 \cap \dots \cap B_n) = \frac{\mathbb{P}(A \cap B_1 \cap \dots \cap B_n)}{\mathbb{P}(B_1 \cap \dots \cap B_n)}. \quad (4.12)$$

Da quest'ultima discende un risultato molto importante ovvero la legge delle probabilità congiunte, o composte, che permette di calcolare la probabilità dell'intersezione di un numero finito di eventi in termini di probabilità condizionate.¹ Precisamente, sussiste il seguente teorema.

Teorema 4.5 (delle probabilità congiunte). *La probabilità dell'intersezione, supposta non quasi impossibile, di $n \geq 2$ eventi A_1, A_2, \dots, A_n è esprimibile come*

$$\mathbb{P}(A_1 A_2 \dots A_n) = \mathbb{P}(A_n) \mathbb{P}(A_{n-1}|A_n) \dots \mathbb{P}(A_1|A_2 \dots A_n), \quad (4.13)$$

oppure come

$$\mathbb{P}(A_1 A_2 \dots A_n) = \mathbb{P}(A_1) \mathbb{P}(A_2|A_1) \dots \mathbb{P}(A_n|A_{n-1} \dots A_1). \quad (4.14)$$

Dimostrazione. Dall'ipotesi $\mathbb{P}(A_1 A_2 \dots A_n) > 0$ si ha anche che

$$\mathbb{P}(A_n) > 0, \mathbb{P}(A_{n-1} A_n) > 0, \dots, \mathbb{P}(A_2 \dots A_n) > 0,$$

¹Si ricorda che, come il segno di moltiplicazione tra numeri reali, si può omettere il segno di intersezione tra insiemi.

per cui è possibile scrivere:

$$\mathbb{P}(A_1 A_2 \cdots A_n) = \mathbb{P}(A_n) \frac{\mathbb{P}(A_{n-1} A_n)}{\mathbb{P}(A_n)} \cdots \frac{\mathbb{P}(A_1 A_2 \cdots A_n)}{\mathbb{P}(A_2 \cdots A_n)}.$$

La (4.13) segue da quest'ultima e dalla (4.12). La (4.14) si ottiene in maniera analoga dopo aver scritto l'intersezione degli n eventi come $A_n \cdots A_2 A_1$. \square

Alla $\mathbb{P}(A_1 A_2 \cdots A_n)$ si dà il nome di *probabilità congiunta* degli eventi A_1, A_2, \dots, A_n . Si fa osservare che le (4.13) e (4.14) sono state utilizzate, nel caso $n = 2$ di immediata derivazione, durante la dimostrazione della Proposizione 4.3.

4.4 Insiemi di alternative

La probabilità condizionata gode di proprietà interessanti ed utili anche sotto il profilo delle applicazioni. Prima di illustrarne alcune, introduciamo una terminologia particolare che rispecchia le applicazioni del concetto di probabilità condizionata ai cosiddetti "test di ipotesi", ben noti in Statistica.

Dato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, fissiamo in Ω un evento H , ossia un elemento di \mathcal{F} , tale che $\mathbb{P}(H) > 0$. Chiameremo tale evento *ipotesi*, e l'evento complementare \bar{H} *ipotesi alternativa*. Le probabilità condizionate $\mathbb{P}(A|H)$ e $\mathbb{P}(A|\bar{H})$ possono allora essere riguardate come le probabilità di occorrenza dell'evento A sotto l'ipotesi H oppure sotto l'ipotesi \bar{H} . Si può però pensare anche ad un insieme più ampio di alternative. Diamo a tale scopo la seguente definizione.

Definizione 4.4. Sia $\{A_i\}_{i \in \mathcal{J}}$ una famiglia di eventi finita o numerabile. Se $\{A_i\}_{i \in \mathcal{J}}$ verifica le condizioni

- (i) $\bigcup_{i \in \mathcal{J}} H_i = \Omega$,
- (ii) per $i \neq j$, $H_i \cap H_j = \emptyset$,
- (iii) $\forall i \in \mathcal{J}$, $\mathbb{P}(H_i) > 0$.

allora si dice che essa costituisce un insieme completo di alternative per lo spazio campione Ω . \square

In altri termini la famiglia $\{A_i\}_{i \in \mathcal{J}}$ costituisce un insieme completo di alternative per Ω se essa è una partizione di Ω costituita da un numero finito o numerabile di eventi non quasi impossibili. Sussiste il seguente teorema.

4.4 Insiemi di alternative

Teorema 4.6 (Teorema delle alternative). Siano $\{A_i\}_{i \in \mathcal{J}}$ un insieme completo di alternative di Ω e A un evento. Risulta:

$$\mathbb{P}(A) = \sum_{i \in \mathcal{J}} \mathbb{P}(A|H_i) \mathbb{P}(H_i). \quad (4.15)$$

Dimostrazione. Tenendo conto della (4.9) e del fatto che, per $i \neq j$, risulta $(A \cap H_i) \cap (A \cap H_j) = \emptyset$ si ricava:

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \cap \Omega) = \mathbb{P}\left[A \cap \left(\bigcup_{i \in \mathcal{J}} H_i\right)\right] = \mathbb{P}\left(\bigcup_{i \in \mathcal{J}} A \cap H_i\right) \\ &= \sum_{i \in \mathcal{J}} \mathbb{P}(A \cap H_i) = \sum_{i \in \mathcal{J}} \frac{\mathbb{P}(A \cap H_i)}{\mathbb{P}(H_i)} \mathbb{P}(H_i) = \sum_{i \in \mathcal{J}} \mathbb{P}(A|H_i) \mathbb{P}(H_i). \quad \square \end{aligned}$$

Il teorema delle alternative permette di calcolare la probabilità di A conoscendo le probabilità $\mathbb{P}(H_i)$ delle alternative e le probabilità condizionate $\mathbb{P}(A|H_i)$.

Osserviamo che nella Definizione 4.4 abbiamo considerato una ricopertura di Ω ; alla nozione di alternative si può però attribuire un senso anche nel caso di una famiglia $\{A_i\}_{i \in \mathcal{J}}$ costituita da un numero finito o numerabile di eventi la cui unione sia un evento H contenuto propriamente nell'evento certo, tale, quindi, da risultare:

$$\bigcup_{i \in \mathcal{J}} H_i = H \subset \Omega.$$

In tal caso la famiglia $\{A_i\}_{i \in \mathcal{J}}$ si dice costituire un insieme di alternative per l'evento H e per essa vale la cosiddetta *legge delle alternative* che è espressa nel seguente teorema.

Teorema 4.7 (Legge delle alternative). Se $\{A_i\}_{i \in \mathcal{J}}$ è un insieme di alternative per H e se $A \subseteq \Omega$ è un evento, la probabilità di A condizionata da H è esprimibile al seguente modo:

$$\mathbb{P}(A|H) = \sum_{i \in \mathcal{J}} \mathbb{P}(A|H_i) \mathbb{P}(H_i|H). \quad (4.16)$$

Dimostrazione. In primo luogo risulta $\mathbb{P}(H) > 0$ in quanto H è l'unione di alternative non quasi impossibili. Poiché, per $i \neq j$, risulta $(A \cap H_i) \cap (A \cap H_j) = \emptyset$, ed inoltre $H_i \cap H = H_i$

per ogni $i \in \mathcal{J}$, si ha:

$$\begin{aligned}\mathbb{P}(A|H) &= \frac{\mathbb{P}(A \cap H)}{\mathbb{P}(H)} = \frac{\mathbb{P}\left[A \cap \bigcup_{i \in \mathcal{J}} H_i\right]}{\mathbb{P}(H)} = \frac{\mathbb{P}\left[\bigcup_{i \in \mathcal{J}} (A \cap H_i)\right]}{\mathbb{P}(H)} \\ &= \sum_{i \in \mathcal{J}} \frac{\mathbb{P}(A \cap H_i)}{\mathbb{P}(H_i)} \cdot \frac{\mathbb{P}(H_i)}{\mathbb{P}(H)} = \sum_{i \in \mathcal{J}} \frac{\mathbb{P}(A \cap H_i)}{\mathbb{P}(H_i)} \cdot \frac{\mathbb{P}(H_i \cap H)}{\mathbb{P}(H)} \\ &= \sum_{i \in \mathcal{J}} \mathbb{P}(A|H_i) \mathbb{P}(H_i|H).\end{aligned}\quad \square$$

Osserviamo, per inciso, che la (4.16) è la generalizzazione della (4.15), a cui si riduce se $H = \Omega$.

4.5 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 produrre — ed ha invero spesso generato — equivoci e risultati paradossali. Si tratta del celebre *teorema o legge di Bayes*, che trovò la sua prima chiara enunciazione ad opera appunto di Bayes in un saggio del 1763 (*Towards Solving a Problem nell'opera Doctrine of Chance*).

Teorema 4.8 (Legge di Bayes). Sia $\{A_i\}_{i \in \mathcal{J}}$ un insieme completo di alternative e sia B un evento non quasi impossibile. Per ogni $i \in \mathcal{J}$ si ha:

$$\mathbb{P}(H_i|B) = \frac{\mathbb{P}(B|H_i)\mathbb{P}(H_i)}{\sum_{j \in \mathcal{J}} \mathbb{P}(B|H_j)\mathbb{P}(H_j)}. \quad (4.17)$$

Dimostrazione. Fissiamo $i \in \mathcal{J}$. Scrivendo le (4.9) e (4.10) con $A = H_i$, e ricavando poi da entrambe $\mathbb{P}(B \cap H_i)$, si ottiene:

$$\mathbb{P}(H_i|B) = \frac{\mathbb{P}(B|H_i)\mathbb{P}(H_i)}{\mathbb{P}(B)}. \quad (4.18)$$

La tesi allora segue da quest'ultima e dal fatto che, per il teorema delle alternative, si può scrivere

$$\mathbb{P}(B) = \sum_{j \in \mathcal{J}} \mathbb{P}(B|H_j)\mathbb{P}(H_j). \quad \square$$

4.5 Teorema di Bayes

Per $i \in \mathcal{J}$, le quantità $\mathbb{P}(H_i)$, $\mathbb{P}(H_i|B)$ e $\mathbb{P}(B|H_i)$ vengono chiamate, rispettivamente, probabilità a priori, probabilità a posteriori e verosimiglianze. Nel caso particolare di un numero finito di alternative equiprobabili, $\{H_1, \dots, H_n\}$, la (4.17) diventa:

$$\mathbb{P}(H_i|B) = \frac{\mathbb{P}(B|H_i)}{\sum_{j \in \mathcal{J}} \mathbb{P}(B|H_j)}.$$

Quindi, in questo caso, le probabilità a priori non giocano alcun ruolo e si mantiene costante il rapporto tra una qualsiasi verosimiglianza e la corrispondente probabilità a posteriori.

La legge di Bayes assume un significato particolare se si attribuisce all'insieme completo di alternative $\{A_i\}_{i \in \mathcal{J}}$ e all'evento B i ruoli, rispettivamente, di “cause” ed “effetto”. Infatti, essa permette allora di determinare la probabilità a posteriori di una fissata causa supponendo note la relativa verosimiglianza e le probabilità a priori di tutte le cause. Di conseguenza la (4.18) fornisce una misura di quale tra le alternative è stata con maggiore probabilità la causa che ha provocato l'effetto B , supposto che B si sia verificato. Tale interpretazione può generare errori ed equivoci se usata ad esempio nel contesto di fenomeni deterministicici e non con specifico riferimento ad uno spazio di probabilità.

È a tal proposito opportuno sottolineare che, in generale, 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.

Osservazione 4.5. Si può dimostrare che il teorema di Bayes continua a valere anche quando la famiglia $\{A_i\}_{i \in \mathcal{J}}$ costituisce un sistema di alternative (ossia $\cup H_i \subset \Omega$) con B evento arbitrario purché non quasi impossibile e soddisfacente la condizione $\cup H_i \supseteq B$. Ciò si può facilmente controllare calcolando la probabilità $\mathbb{P}(B)$ con lo stesso procedimento utilizzato per il teorema delle alternative.

CAPITOLO 5

Variabili Casuali

5.1 Il concetto di variabile casuale^{ALEATORIA}

Fondamentale nella teoria della probabilità è il concetto di variabile casuale. Prima di darne la definizione illustriamolo con alcuni esempi. Consideriamo in primo luogo l'esperimento casuale che consiste nel lancio di un dado e chiediamoci quanti lanci vadano effettuati prima che esca un numero pari. Evidentemente tale quesito non è suscettibile di una risposta costituita da un unico numero. Riferiamoci ora al moto di una particella su di un asse coordinato e supponiamo che essa compia una passeggiata aleatoria (cfr. il § 1.2.2). Inizialmente la particella occupa dunque una prefissata posizione, ad esempio l'origine; si effettua allora il lancio di una moneta e la particella si sposta di un tratto di lunghezza unitaria nel verso positivo o nel verso negativo a seconda che il risultato del lancio sia rispettivamente testa o croce. A questo punto viene effettuato un nuovo lancio e la particella viene spostata di un'unità secondo il precedente schema. Questa procedura viene reiterata indefinitamente, cosicché la particella si muove in maniera erratica lungo l'asse coordinato, compiendo appunto una passeggiata aleatoria. Ci chiediamo quanti punti distinti siano stati "visitati" dalla particella dopo un prefissato numero di spostamenti. Ancora una volta la risposta non è univoca. Un ulteriore esempio può essere offerto dal numero di chiamate che pervengono ad un centralino telefonico in un generico intervallo di tempo di durata fissata; questo numero non è unico in quanto varia in dipendenza di fattori che possiamo riguardare come casuali.

Questi tre esempi sono caratterizzati da un elemento comune: la variabile che esprime il numero richiesto non assume un unico valore, ma più valori che dipendono dai risultati degli esperimenti casuali sottostanti considerati. Si comprende dunque la necessità di introdurre un nuovo concetto di variabile che coinvolga intrinsecamente la struttura degli eventi e la misura

di probabilità definita su di essa. A tale scopo è opportuno richiamare, a questo punto, il concetto di misurabilità riferito ad un'applicazione.

Definizione 5.1. Sia H un insieme, \mathcal{H} una σ -algebra su H e g un'applicazione di H in \mathbb{R} . Si dice che g è \mathcal{H} -misurabile se la controimmagine tramite g di ogni insieme di Borel di \mathbb{R} è un elemento di \mathcal{H} . \square

Quindi g è \mathcal{H} -misurabile se, $\forall B \in \mathcal{B}$, $g^{-1}(B) \in \mathcal{H}$. In particolare se $H = \mathbb{R}$ e $\mathcal{H} = \mathcal{B}$ allora la funzione g si dice \mathcal{B} -misurabile, ovvero Borel-misurabile.

Nelle considerazioni seguenti faremo sempre riferimento ad uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$.

Definizione 5.2. Una funzione $X : \Omega \rightarrow \mathbb{R}$ si dice variabile casuale (v.c.), o variabile aleatoria, se essa è \mathcal{F} -misurabile. \square

Riferendo la definizione al contesto probabilistico nel quale stiamo operando possiamo allora dire che la controimmagine tramite una v.c. X di un qualsiasi boreliano di \mathbb{R} è un evento di \mathcal{F} . Nel caso in cui $\mathcal{F} = \mathcal{P}(\Omega)$ ogni applicazione di Ω in \mathcal{F} è una v.c.

Si noti che il termine "variabile casuale" non è molto felice; infatti una v.c. non è né una "variabile" nel senso consueto¹ né una quantità in alcun modo "casuale": essa è semplicemente una funzione caratterizzata dall'essere \mathcal{F} -misurabile.

Risulta opportuno ricordare che alla σ -algebra \mathcal{B} appartengono tutti gli intervalli di \mathbb{R} . Nel seguito useremo, per brevità, la notazione $\{X \leq x\}$ per indicare l'evento $\{\omega \in \Omega : X(\omega) \leq x\} \equiv X^{-1}((-\infty, x])$. Allo stesso modo $\{X > x\}$ rappresenta l'evento $\{\omega \in \Omega : X(\omega) > x\} \equiv X^{-1}([x, +\infty[)$, e così via. In generale, per ogni boreliano B porremo:

$$\{X \in B\} := \{\omega \in \Omega : X(\omega) \in B\} \equiv X^{-1}(B),$$

e ometteremo le parentesi graffe quando l'evento in questione è argomento della funzione \mathbb{P} ; ad esempio scriveremo $\mathbb{P}(X \in B)$ in luogo di $\mathbb{P}(\{X \in B\})$.

Teorema 5.1. Sia $X : \Omega \rightarrow \mathbb{R}$ una v.c. La posizione

$$\forall B \in \mathcal{B}, \quad \mathbb{P}_X(B) := \mathbb{P}(X \in B) \tag{5.1}$$

definisce una misura di probabilità su $(\mathbb{R}, \mathcal{B})$.

Dimostrazione. Bisogna far vedere che \mathbb{P}_X verifica i tre requisiti della Definizione 3.14.

(i) Si ha, $\forall B \in \mathcal{B}$, $\mathbb{P}_X(B) \equiv \mathbb{P}(X \in B) \geq 0$.

¹Una variabile è un simbolo che varia in un certo insieme: ad es. x che varia nell'insieme dei reali, n che varia nell'insieme dei numeri interi positivi, ecc.

5.2 La funzione di distribuzione

(ii) Risulta $\mathbb{P}_X(\mathbb{R}) \equiv \mathbb{P}(X \in \mathbb{R}) = \mathbb{P}(\Omega) = 1$.

(iii) Sia $\{B_n\}_{n \in \mathbb{N}}$ una successione di boreliani di \mathbb{R} a due a due incompatibili. Si ha

$$\begin{aligned} \mathbb{P}_X\left(\bigcup_{n=1}^{\infty} B_n\right) &= \mathbb{P}\left(X \in \bigcup_{n=1}^{\infty} B_n\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} (X \in B_n)\right) = \sum_{n=1}^{\infty} \mathbb{P}(X \in B_n) \\ &= \sum_{n=1}^{\infty} \mathbb{P}_X(B_n). \end{aligned} \quad \square$$

La misura di probabilità \mathbb{P}_X è chiamata la *distribuzione (o legge) di probabilità di X* .

Osservazione 5.1. Due variabili casuali (vv.cc.) differenti, X e Y , definite sul medesimo spazio di probabilità possono avere la medesima legge, nel senso che

$$\forall B \in \mathcal{B}, \quad \mathbb{P}_X(B) = \mathbb{P}_Y(B).$$

In tal caso si dice che, indifferentemente, le vv.cc. X e Y sono *somiglianti* oppure *identicamente distribuite* oppure *equidistribuite*. Ad esempio, nell'esperimento casuale consistente nel lancio di un dado onesto dove ogni singoletto è considerato come evento elementare, la funzione X che trasforma il risultato in se stesso e la funzione Y che trasforma il risultato nel suo complemento a 7 sono identicamente distribuite (si lascia al lettore). La somiglianza si estende naturalmente al caso di più di due variabili casuali. \square

Per evidenziare che \mathbb{P}_X si ottiene dalla misura di probabilità \mathbb{P} si dice che essa è indotta dalla v.c. X sullo spazio probabilizzabile $(\mathbb{R}, \mathcal{B})$. Con essa si completa la costruzione di un nuovo spazio di probabilità $(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$ al quale ci riferiremo ogni qual volta vogliamo specificare la classe delle vv.cc. tra loro somiglianti. In altri termini la conoscenza della legge di una v.c. X consente allora di determinare le probabilità degli eventi del tipo $\{X \in B\}$, con $B \in \mathcal{B}$. La difficoltà nell'ottenimento di misure di probabilità (che ricordiamo è una funzione avente per dominio una σ -algebra) è stata più volte messa in luce nel precedente Capitolo 3; nel prossimo paragrafo definiremo una restrizione della distribuzione di probabilità di una v.c. che ha il notevole vantaggio di essere una funzione di variabile reale.

5.2 La funzione di distribuzione

Iniziamo questo paragrafo con la seguente definizione.

Definizione 5.3. Data una v.c. X , la funzione $F_X : \mathbb{R} \rightarrow [0, 1]$ definita dalla posizione

$$\forall x \in \mathbb{R}, \quad F_X(x) := \mathbb{P}_X((-\infty, x]) \equiv \mathbb{P}(X \leq x), \tag{5.2}$$

prende il nome di *funzione di distribuzione* della v.c. X . \square

La funzione di distribuzione F_X di una v.c. X gode di alcune proprietà che la caratterizzano nel senso che se una funzione reale di variabile reale gode di queste stesse proprietà allora essa è la funzione di distribuzione di una qualche variabile casuale. Inoltre, la funzione di distribuzione F_X di una v.c. X pur fornendo esclusivamente la probabilità delle semirette sinistre chiuse consente, attraverso opportune operazioni, la ricostruzione della distribuzione di probabilità di X .

Quanto qui affermato viene formalizzato e dimostrato nei risultati che seguono.

Teorema 5.2. La funzione di distribuzione F_X di una v.c. X gode delle seguenti proprietà:

- (i) F_X è non decrescente in \mathbb{R} : $\forall x_1, x_2 \in \mathbb{R}$ tali che $x_1 < x_2 \Rightarrow F_X(x_1) \leq F_X(x_2)$;
- (ii) F_X è continua a destra in \mathbb{R} : $\forall x \in \mathbb{R}, \lim_{\epsilon \rightarrow 0^+} F_X(x + \epsilon) = F_X(x)$;
- (iii) $\lim_{x \rightarrow -\infty} F_X(x) = 0$ e $\lim_{x \rightarrow +\infty} F_X(x) = 1$.

Dimostrazione. Si ricordi che $F_X(x) = \mathbb{P}(X \leq x)$.

(i) L'asserto segue dalla monotonia di \mathbb{P} . Infatti l'evento $\{X \leq x_1\}$ è contenuto nell'evento $\{X \leq x_2\}$, cosicché $\mathbb{P}(X \leq x_1) \leq \mathbb{P}(X \leq x_2) \Leftrightarrow F_X(x_1) \leq F_X(x_2)$.

(ii) Dall'essere F_X una funzione monotona sarà sufficiente dimostrare che per ogni $x \in \mathbb{R}$ si ha:

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

A tale scopo posto, per ogni $n \in \mathbb{N}$, $A_n = \{X \in]x, x + 1/n]\}$ risulta:

$$\begin{aligned} \forall n \in \mathbb{N}, \quad F_X\left(x + \frac{1}{n}\right) &= \mathbb{P}\left(X \leq x + \frac{1}{n}\right) = \mathbb{P}(X \leq x) + \mathbb{P}\left(x < X \leq x + \frac{1}{n}\right) \\ &= F_X(x) + \mathbb{P}(A_n). \end{aligned}$$

Osservando che la successione $\{A_n\}_{n \in \mathbb{N}}$ decresce all'evento impossibile, da quest'ultima e dalla continuità di \mathbb{P} si ottiene:

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

(iii) Dall'essere F_X una funzione monotona sarà sufficiente dimostrare che risulta:

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

5.2 La funzione di distribuzione

A tal fine poniamo, per ogni $n \in \mathbb{N}$, $B_n^- = \{X \leq -n\}$ e $B_n^+ = \{X \leq n\}$ e osserviamo che $\{B_n^-\}_{n \in \mathbb{N}}$ è una successione decrescente all'evento impossibile mentre $\{B_n^+\}_{n \in \mathbb{N}}$ cresce all'evento certo. Dalla continuità di \mathbb{P} segue

$$\lim_{n \rightarrow +\infty} F_X(-n) = \mathbb{P}(X \leq -n) = \lim_{n \rightarrow +\infty} \mathbb{P}(B_n^-) = \mathbb{P}(\emptyset) = 0,$$

e

$$\lim_{n \rightarrow +\infty} F_X(n) = \mathbb{P}(X \leq n) = \lim_{n \rightarrow +\infty} \mathbb{P}(B_n^+) = \mathbb{P}(\Omega) = 1. \quad \square$$

Le proprietà (i), (ii) e (iii) caratterizzano la funzione di distribuzione di una variabile casuale nel senso specificato dal prossimo teorema.

Teorema 5.3. Sia assegnata una funzione $F: \mathbb{R} \rightarrow [0, 1]$ tale che

- (i) F è non decrescente in \mathbb{R} ;
- (ii) $F(x)$ è continua a destra in ogni $x \in \mathbb{R}$;
- (iii) $\lim_{x \rightarrow -\infty} F(x) = 0$ e $\lim_{x \rightarrow +\infty} F(x) = 1$.

Esistono allora uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e una v.c. X definita su Ω per i quali si ha, per ogni $x \in \mathbb{R}$, $F_X(x) = F(x)$.

Dimostrazione. Consideriamo lo spazio probabilizzabile $(\mathbb{R}, \mathcal{B})$ e sia λ_S la misura di Lebesgue-Stieltjes indotta da F su di esso (cfr. l'Esempio 3.14). Lo spazio di probabilità e la v.c. ricercati coincidono, rispettivamente, con $(\mathbb{R}, \mathcal{B}, \lambda_S)$ e con la funzione identità I di \mathbb{R} in \mathbb{R} . Infatti, per la definizione stessa di λ_S si ha

$$\forall x \in \mathbb{R}, \quad F_I(x) = \lambda_S([-\infty, x]) = F(x). \quad \square$$

La Proposizione che segue mostra come sia possibile esprimere in termini di funzione di distribuzione la probabilità degli eventi che hanno come immagini in \mathbb{R} i sottoinsiemi elencati nell'Esempio 3.6 e i singoletti di \mathbb{R} .

Proposizione 5.1. Sia X una v.c. avente funzione di distribuzione $F_X(x)$ e siano x_1, x_2 e x numeri reali con $x_1 < x_2$. Posto $F_X(x^-) = \lim_{\epsilon \rightarrow 0^+} F_X(x - \epsilon)$, si ha:

- a) $\mathbb{P}(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)$,
- b) $\mathbb{P}(X = x) = F_X(x) - F_X(x^-)$,
- c) $\mathbb{P}(x_1 \leq X \leq x_2) = F_X(x_2) - F_X(x_1^-)$,
- d) $\mathbb{P}(x_1 < X < x_2) = F_X(x_2^-) - F_X(x_1)$,
- e) $\mathbb{P}(x_1 \leq X < x_2) = F_X(x_2^-) - F_X(x_1^-)$,
- f) $\mathbb{P}(X > x) = 1 - F_X(x)$,
- g) $\mathbb{P}(X < x) = F_X(x^-)$,
- h) $\mathbb{P}(X \geq x) = 1 - F_X(x^-)$.

Dimostrazione. Si ricordi che $F_X(x) = \mathbb{P}(X \leq x)$.

a) Sussistendo la decomposizione

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

possiamo scrivere

$$F_X(x_2) = F_X(x_1) + \mathbb{P}(x_1 < X \leq x_2).$$

Da quest'ultima segue immediatamente la tesi.

b) Dalla relazione

$$\forall n \in \mathbb{N}, \quad \{X \leq x\} = \{X \leq x - 1/n\} \cup \{x - 1/n < X \leq x\}$$

discende

$$\forall n \in \mathbb{N}, \quad F_X(x) = F_X(x - 1/n) + \mathbb{P}(x - 1/n < X \leq x).$$

Osservando che la successione $\{x - 1/n < X \leq x\}_{n \in \mathbb{N}}$ decresce all'evento $\{x\}$, passando al limite per $n \rightarrow \infty$ nella precedente relazione, per la continuità della misura di probabilità \mathbb{P} si ottiene

$$F_X(x) = \lim_{n \rightarrow \infty} F_X(x - 1/n) + \mathbb{P}(X = x).$$

La tesi discende dal fatto che, per la non decrescenza di \mathbb{P} , risulta

$$F_X(x^-) = \lim_{\epsilon \rightarrow 0^+} F_X(x - \epsilon) = \lim_{n \rightarrow \infty} F_X(x - 1/n).$$

c) Sussistendo la decomposizione

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

possiamo scrivere

$$\mathbb{P}(x_1 \leq X \leq x_2) = \mathbb{P}(X = x_1) + \mathbb{P}(x_1 < X \leq x_2).$$

La tesi allora discende dalla a) e dalla b) applicata per $x = x_1$.

d) Sussistendo la decomposizione

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

possiamo scrivere

$$\mathbb{P}(x_1 < X < x_2) = \mathbb{P}(x_1 < X \leq x_2) - \mathbb{P}(X = x_1).$$

5.3 Variabili casuali discrete

La tesi allora discende dalla a) e dalla b) applicata per $x = x_2$.

e) Sussistendo la decomposizione

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

possiamo scrivere

$$\mathbb{P}(x_1 \leq X < x_2) = \mathbb{P}(X = x_1) + \mathbb{P}(x_1 < X < x_2).$$

La tesi allora discende dalla d) e dalla b) applicata per $x = x_1$.

f) La tesi segue direttamente dal fatto che $\{X > x\} = \overline{\{X \leq x\}}$

g) Sussistendo la decomposizione

$$\{X \leq x\} = \{X < x\} \cup \{X = x\},$$

possiamo scrivere

$$\mathbb{P}(X < x) = F_X(x) - \mathbb{P}(X = x).$$

La tesi allora discende dalla b).

h) La tesi segue dal fatto che $\{X \geq x\} = \overline{\{X < x\}}$ e dalla g).

Osservazione 5.2. Più generalmente, la teoria dell'integrazione di Stieltjes (v. l'Appendice B.2) consente di scrivere:

$$\forall B \in \mathcal{B}, \quad \mathbb{P}_X(B) \equiv \mathbb{P}(X \in B) = \int_B dF_X(x). \quad (5.3)$$

Dalla (5.3) si ricava anche che due o più vv.cc. sono somiglianti se e solo se esse hanno la stessa funzione di distribuzione. \square

5.3 Variabili casuali discrete

Quanto esposto nel § 5.2 ha carattere generale, ossia sussiste per ogni variabile casuale. Considereremo ora in particolare il caso delle cosiddette variabili casuali discrete.

Definizione 5.4. Una variabile casuale X si dice discreta se esiste un sottoinsieme S_X di \mathbb{R} finito o numerabile per il quale:²

(i) l'evento $\{X \in S_X\}$ è quasi certo

²Il fatto che S_X è al più numerabile ci garantisce che esso appartiene a \mathcal{B} .

(ii) $\forall x \in S_X, \mathbb{P}_X(\{x\}) \equiv \mathbb{P}(X = x) > 0$.

L'insieme S_X è detto *spettro* della v.c. X . \triangleleft

Per una v.c. X con spettro S_X ha senso allora considerare la restrizione $p_X : \mathbb{R} \rightarrow [0, 1]$ della distribuzione di probabilità \mathbb{P}_X ai singoletti $\{x\}$ di \mathbb{R} :

$$\forall x \in \mathbb{R}, \quad p_X(x) := \mathbb{P}_X(\{x\}) = \mathbb{P}(X = x). \quad (5.4)$$

La funzione p_X prende il nome di *funzione di probabilità* (o *funzione massa*) della v.c. discreta X ; essa è nulla quasi ovunque in quanto assume un valore positivo esclusivamente sugli elementi dello spettro S_X . Sussiste il seguente teorema.

Teorema 5.4. *Sia X una v.c. discreta il cui spettro S_X sia rappresentato da una successione $\{x_n\}_{n \in \mathbb{N}}$ crescente. La distribuzione di probabilità \mathbb{P}_X della v.c. X è univocamente individuata da tale successione e dalla successione di numeri reali positivi $\{p_X(x_n)\}_{n \in \mathbb{N}}$.*

Dimostrazione. In primo luogo, si osservi che

$$\sum_{n=1}^{\infty} p_X(x_n) = \mathbb{P}_X(S_X) = 1. \quad (5.5)$$

Inoltre, riconoscendo che l'evento $\{X \in (B \cap \bar{S}_X)\}$ è contenuto nell'evento quasi impossibile $\{X \in \bar{S}_X\}$ e sussistendo la decomposizione

$$\forall B \in \mathcal{B}, \quad \{X \in B\} = \{X \in (B \cap S_X)\} \cup \{X \in (B \cap \bar{S}_X)\}$$

risulta³

$$\forall B \in \mathcal{B}, \quad \mathbb{P}_X(B) = \mathbb{P}(X \in B) = \mathbb{P}(X \in B \cap S_X) = \sum_{\{n: x_n \in B\}} p_X(x_n). \quad (5.6) \quad \square$$

Osservazione 5.3. Il risultato precedente vale per ogni v.c. discreta X . Infatti, nel caso che lo spettro S_X ha cardinalità finita m la distribuzione di probabilità di X è univocamente individuata dalla m -upla ordinata (x_1, \dots, x_m) degli elementi dello spettro e dalla m -upla $(p_X(x_1), \dots, p_X(x_m))$ delle rispettive probabilità. Tutti gli ulteriori risultati che saranno presentati in questo paragrafo che si riferiscono ad uno spettro con cardinalità numerabile valgono, con gli opportuni adattamenti, anche nel caso di uno spettro finito. \triangleleft

³Con abuso di notazione si intende dire che la sommatoria $\sum_{\{n: \Pi(a_n)\}} a_n$ è estesa a tutti gli interi n per i quali il corrispondente a_n verifica la condizione Π .

In particolare, applicando la relazione (5.6) alle semirette sinistre chiuse si ottiene la funzione di distribuzione F_X della v.c. X discreta:

$$\forall x \in \mathbb{R}, \quad F_X(x) = \mathbb{P}(X \leq x) = \sum_{\{n: x_n \leq x\}} p_X(x_n). \quad (5.7)$$

Osservazione 5.4. La parte g) della Proposizione 5.1 e la relazione (5.6) applicata alle semirette sinistre aperte consentono di ricavare il valore del limite sinistro della funzione di distribuzione:

$$\forall x \in \mathbb{R}, \quad F_X(x^-) = \mathbb{P}(X < x) = \sum_{\{n: x_n < x\}} p_X(x_n). \quad (5.8)$$

Adesso siamo in condizione di valutare $\Delta F_X(x) = F_X(x) - F_X(x^-)$. Se $x \notin S_X$ allora $\Delta F_X(x) = 0$: in tal caso F_X è continua in x . Se, invece, esiste un intero k per il quale $x = x_k \in S_X$ allora $\Delta F_X(x) = p_X(x_k)$: la funzione di distribuzione F_X ha una discontinuità di prima specie in x_k il cui salto rappresenta la probabilità che X assuma il valore x_k . In entrambi i casi si è in accordo con la parte b) della Proposizione 5.1. \triangleleft

La funzione di distribuzione di una v.c. discreta possiede una ulteriore e, in alcune circostanze, estremamente utile rappresentazione che è basata sulla cosiddetta funzione di Heaviside, o funzione *gradino unitario*. Tale funzione assume il valore 1 quando il suo argomento è non negativo:

$$H(x) = \begin{cases} 0, & \text{per } x < 0, \\ 1, & \text{per } x \geq 0. \end{cases} \quad (5.9)$$

Infatti, dalla (5.7) si evince che al valore di F_X nell'argomento x contribuiscono i termini della successione $\{p_X(x_n)\}_{n \in \mathbb{N}}$ i cui corrispondenti valori dello spettro verificano la condizione $x_n \leq x$. Lo stesso risultato si ottiene considerando ogni termine della successione moltiplicato per una quantità che vale 1 solo se il corrispondente valore dello spettro è minore di x :

$$F(x) = \sum_{n=1}^{\infty} p_X(x_n) H(x - x_n). \quad (5.10)$$

Si noti che la funzione $H(x)$ può essa stessa riguardarsi come la funzione di distribuzione di una v.c. che assume con probabilità 1 il valore 0, mentre è nulla la probabilità di occorrenza di qualsiasi altro valore; una v.c. siffatta, che in qualche maniera generalizza il significato delle costanti reali, si dice *degenera*. Per tale ragione $H(x)$ è anche detta funzione di distribuzione *causale*.

Il risultato stabilito nel teorema seguente è di fondamentale importanza nella costruzione di variabili casuali discrete.

Teorema 5.5. *Siano assegnate due successioni di numeri reali. La prima di esse, $\{x_n\}_{n \in \mathbb{N}}$, è una successione crescente e la seconda, $\{p_n\}_{n \in \mathbb{N}}$, che verifica le seguenti condizioni:*

(i) $\forall n \in \mathbb{N}, p_n > 0$;

(ii) (condizione di normalizzazione) $\sum_{n=1}^{\infty} p_n = 1$.

Esistono allora uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e una v.c. X definita su Ω discreta con spettro $S_X = \{x_n\}_{n \in \mathbb{N}}$ per i quali si ha, per ogni $n \in \mathbb{N}$, $p_X(x_n) = p_n$.

Dimostrazione. Si consideri la funzione $F: \mathbb{R} \rightarrow [0, 1]$ definita nella posizione

$$\forall x \in \mathbb{R}, F(x) := \sum_{\{n : x_n \leq x\}} p_n.$$

Faremo vedere che F è (a) non decrescente, (b) continua a destra, (c) infinitesima per $x \rightarrow -\infty$ e tendente a 1 per $x \rightarrow +\infty$.

(a) Siano $x < y$ reali qualsiasi; risulta

$$F(y) - F(x) = \sum_{\{n : x < x_n \leq y\}} p_n \geq 0.$$

(b) Per ogni fissato reale x , sia k l'intero per il quale $x_k \leq x < x_{k+1}$. Posto $\delta = x_{k+1} - x_k$ si ha che

$$\forall \epsilon < \delta, F(x + \epsilon) = \sum_{\{n : x_n \leq x + \epsilon\}} p_n = \sum_{n=1}^k p_n = \sum_{\{n : x_n \leq x\}} p_n = F(x),$$

dalla quale discende immediatamente che $\lim_{\epsilon \rightarrow 0^+} F(x + \epsilon) = F(x)$.

(c) Per definizione, $F(x) = 0$ per ogni $x < x_1$ da cui $\lim_{x \rightarrow -\infty} F(x) = 0$; inoltre

$$\lim_{x \rightarrow +\infty} F(x) = \sum_{\{n : x_n < +\infty\}} p_n = \sum_{n=1}^{\infty} p_n = 1.$$

La tesi segue dal Teorema 5.3. \square

Osservazione 5.5. Si fa osservare che entrambe le successioni, quella dei valori $\{x_n\}_{n \in \mathbb{N}}$ e quella delle relative probabilità $\{p_n\}_{n \in \mathbb{N}}$ sono necessarie per individuare in maniera univoca la distribuzione di probabilità. Ad esempio, nell'ambito del processo di Bernoulli che abbiamo descritto nell'Esempio 3.4, si possono immaginare le seguenti due situazioni diverse: una prima situazione in cui all'uscita di testa è associato una vincita unitaria e all'uscita di croce una perdita nulla; una seconda situazione nella quale, permanendo la vincita unitaria all'uscita di testa, all'uscita di croce è associata una perdita unitaria. In altri termini,

$\forall k \in \mathbb{N}$, al lancio k -esimo della moneta stiamo considerando due vv.cc., X_k e Y_k , che assumono entrambe il valore 1 se il risultato del lancio è testa; all'uscita della croce, invece, la c.c. X_k assume il valore 0 mentre la v.c. Y_k assume il valore -1. Ovviamente Pur possedendo la stessa distribuzione di probabilità i due schemi sono fondamentalmente diversi: nell'ambito del primo si può invero solo vincere, mentre nell'ambito del secondo si può vincere o perdere. \triangleleft

Per giustificare l'utilizzazione della denominazione "funzione massa" precedentemente introdotta, è suggestivo fare riferimento ad una massa unitaria di una qualche sostanza. Immaginiamo tale massa distribuita in più modi su di un prefissato intervallo, con l'unica condizione che non si alteri la massa totale di sostanza presente: tutta in un punto, in più punti in parti uguali, con continuità sull'intero intervallo, ecc. Supponiamo, ad esempio, che questa massa unitaria sia concentrata nei punti dell'asse reale a coordinate intere, e denotiamo con $p(n)$ la massa presente nel punto n ; evidentemente risulta:⁴

$$\sum_{n \in \mathbb{Z}} p(n) = 1. \quad (5.11)$$

Inoltre, la massa complessivamente presente nell'intervallo $]-\infty, x]$ è data da

$$\sum_{\{n \leq x\}} p(n). \quad (5.12)$$

5.3.1 Esempi di funzioni di distribuzione

L'introduzione del concetto di v.c. permette di esprimere in modo quantitativo eventi a carattere anche qualitativo tramite, appunto, una funzione a valori reali.

Esempio 5.1. Ritorniamo all'esperimento casuale consistente in un unico lancio di una moneta, e sia $\Omega = \{T, C\}$ lo spazio campione. Associamo a questo esperimento lo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ con

$$\mathcal{F} = \{\emptyset, \{T\}, \{C\}, \Omega\}$$

e con \mathbb{P} definita al seguente modo:

$$\mathbb{P}(\emptyset) = 0, \quad \mathbb{P}(\{T\}) = p, \quad \mathbb{P}(\{C\}) = q \equiv 1 - p, \quad \mathbb{P}(\Omega) = 1,$$

dove p e q denotano rispettivamente le probabilità di occorrenza di testa e croce.⁵ Introduciamo ora una funzione $X: \Omega \rightarrow \mathbb{R}$ tramite le posizioni $X(T) = 1$ e $X(C) = 0$.

⁴Con \mathbb{Z} denotiamo l'insieme dei numeri interi positivi e negativi, zero incluso.

⁵La probabilità p si dice *trucco* della moneta; se è $p = 1/2$ la moneta è detta *equa*.

Dimostriamo anzitutto che X è una variabile casuale.⁶ A tal fine occorre mostrare che l'applicazione $X: \Omega \rightarrow \mathbb{R}$ è \mathcal{F} -misurabile, ossia che per ogni $x \in \mathbb{R}$ risulta $\{\omega \in \Omega: X(\omega) \leq x\} \in \mathcal{F}$. Ciò è peraltro immediato, avendosi

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

In virtù della (5.2) la funzione di distribuzione $F(x)$ di X è così specificata:

$$F(x) \equiv \mathbb{P}\{X \leq x\} = \begin{cases} \mathbb{P}(\emptyset) = 0, & \text{per } x < 0, \\ \mathbb{P}(\{C\}) = q, & \text{per } 0 \leq x < 1, \\ \mathbb{P}(\Omega) = 1, & \text{per } x \geq 1. \end{cases} \quad (5.13)$$

Utilizzando la funzione di Heaviside la funzione F si scrive molto più facilmente come

$$F(x) = q H(x) + p H(x - 1). \quad (5.14)$$

La funzione F è dunque costante in ognuno dei tre intervalli sopra specificati così che la probabilità di ottenere un qualunque valore diverso da 0 o da 1 è nulla. Nei punti $x = 0$ e $x = 1$ le probabilità $\mathbb{P}(X = 0)$ e $\mathbb{P}(X = 1)$ equivalgono, rispettivamente, alle ampiezze delle discontinuità in 0 e 1 che quindi valgono q in 0 e $1 - q \equiv p$ in 1. Da quanto abbiamo detto segue che la funzione di probabilità $p(x)$ assume i valori $p(0) = q$, $p(1) = p$; l'intera massa unitaria è così concentrata nei due soli punti 0 e 1. \triangleleft

Esempio 5.2. Un altro esempio di v.c. discreta (definita in questo caso su un insieme numerabile) è fornito dalla cosiddetta distribuzione di Poisson, già incontrata nel § 1.1. Si dice che X è di Poisson con parametro α ($\alpha > 0$) la v.c. – la cui esistenza è garantita dal Teorema 5.5 – avente come successione dei valori assumibili l'insieme \mathbb{N}_0 e come successione dei valori di probabilità quella di termine generale:

$$p_n = \frac{\alpha^n}{n!} e^{-\alpha}.$$

Infatti, si noti che risulta

$$\forall n \in \mathbb{N}_0, \quad p_n > 0$$

e

$$\sum_{n \in \mathbb{N}_0} p(n) = e^{-\alpha} \sum_{n=0}^{+\infty} \frac{\alpha^n}{n!} = e^{-\alpha} e^\alpha = 1.$$

⁶La dimostrazione è fornita solo allo scopo di illustrare la definizione di variabile casuale; infatti il fatto che la funzione X in questione è una v.c. è del tutto ovvio dal momento che \mathcal{F} coincide con l'insieme delle parti di Ω .

Inoltre, la funzione di distribuzione della v.c. X di Poisson con parametro λ di cui $p(x)$ è la seguente:

$$F(x) = \sum_{\{n \in \mathbb{N}_0 : n \leq x\}} p_n = e^{-\alpha} \sum_{n=0}^{\lceil x \rceil} \frac{\alpha^n}{n!},$$

dove $\lceil x \rceil$ denota il massimo intero non maggiore di x . In termini della funzione di Heaviside essa diviene:

$$F(x) = \sum_{n=0}^{\infty} p_n H(x - n) = e^{-\alpha} \sum_{n=0}^{\infty} \frac{\alpha^n}{n!} H(x - n).$$

In questo caso si possiede dunque un'espressione analitica della funzione di distribuzione. Ciò, tuttavia, non avviene sempre; la funzione di distribuzione potrebbe ad esempio essere la soluzione di un'equazione che non è risolubile in forma chiusa. \triangleleft

Esempio 5.3. Un'ulteriore v.c. discreta è illustrata dalle considerazioni che seguono. Riferiamoci allo spazio di Bernoulli associato ad un esperimento consistente nel lancio ripetuto di una moneta e sia $\Omega = \{0, 1\}$ lo spazio campione relativo alla singola prova, avendo associato 1 a testa (T) e 0 a croce (C). Denotiamo, poi, con $\Omega^{(\infty)} = \Omega \times \Omega \times \dots$ lo spazio campione i cui elementi $\omega^{(\infty)} = \{\omega_1, \omega_2, \dots\}$ sono le sequenze infinite di zeri e uni. Come σ -algebra consideriamo $\mathcal{F}_1 = \sigma(\mathcal{G}_1)$ dove \mathcal{G}_1 contiene tutti gli eventi T_n del tipo "al lancio n -esimo esce testa". Poniamo

$$X(\omega) = \min\{i \geq 1: \omega_i = 1\}.$$

È questa un'applicazione che ad ogni elemento $\omega^{(\infty)} \in \Omega^{(\infty)}$ associa l'intero positivo denotante il numero d'ordine del lancio della moneta effettuato il cui esito è stato per la prima volta T . L'applicazione $X: \Omega^{(\infty)} \rightarrow \mathbb{R}$ è una variabile casuale se è \mathcal{F}_1 -misurabile. Per dimostrare che ciò accade, mostriamo che per ogni $x \in \mathbb{R}$ risulta $A_X = \{\omega: X(\omega) \leq x\} \in \mathcal{F}_1$. Per $x < 1$ si ha $A_X = \emptyset$. Sia ora $i \leq x < i + 1$, con $i \in \mathbb{N}$. Con la notazione introdotta nel §3.4 si ha:

$$A_X = T_1 \cup (C_1 \cap T_2) \cup (C_1 \cap C_2 \cap T_3) \cup \dots \cup (C_1 \cap C_2 \cap \dots \cap C_{i-1} \cap T_i).$$

Per ogni x reale A_X è dunque esprimibile in termini di unioni e intersezioni di eventi generatori, che sono elementi di \mathcal{F}_1 ; pertanto risulta $A_X \in \mathcal{F}_1$. È infine immediato calcolare la distribuzione di probabilità $\mathbb{P}\{X = n\}$ di X :

$$\begin{aligned} \mathbb{P}\{X = n\} &= \mathbb{P}\{\omega^{(\infty)} \in \Omega^{(\infty)}: X(\omega) = n\} \\ &= \mathbb{P}\{(C_1 \cap C_2 \cap \dots \cap C_{n-1} \cap T_n\} \\ &= q^{n-1} p \quad (n = 1, 2, \dots), \end{aligned}$$

dove si è fatto uso dell'ipotesi di indipendenza tra eventi generatori e dove si è posto $\mathbb{P}\{T_n\} = p$ e $\mathbb{P}\{C_n\} = q \equiv 1 - p$ per ogni $n \in \mathbb{N}$. \triangleleft

5.4 Variabili casuali assolutamente continue

Come abbiamo visto nel § 5.3, ogni v.c. discreta possiede la distribuzione di probabilità. Quest'ultima è sufficiente per una completa caratterizzazione probabilistica della v.c.; mediante essa è infatti esprimibile la funzione di distribuzione.

Nel caso di vv.cc. con funzioni di distribuzione continue non è significativo riferirsi alle probabilità dei singoli valori in quanto, come sappiamo, tali probabilità sono tutte nulle. Così, nell'esperimento casuale consistente nella scelta a caso di un punto dell'intervallo $[0, 1]$ dell'asse reale, qualunque sia il punto $x \in [0, 1]$ l'evento "il punto scelto è x " ha probabilità nulla. Non è invece nulla la probabilità, ad esempio, dell'evento "il punto scelto è interno all'intervallo $[0, 1/2]$ ".

Per vv.cc. con funzioni di distribuzione continue conviene allora fare riferimento esclusivamente alla probabilità che la v.c. assuma valori in intervalli comunque preassegnati. In particolare, ha senso calcolare la probabilità che la v.c. assuma un valore interno all'intervallo $[x, x + \Delta x]$, con $x \in \mathbb{R}$. Tale intervallo, ovviamente, non individua il valore x ; possiamo però affermare che quanto più piccolo è Δx tanto migliore è l'individuazione del punto x . Nel passaggio dal discreto al continuo al concetto di probabilità di singoli valori va dunque sostituito il concetto di probabilità di valori in intervalli. Di conseguenza in questo caso, come subito vedremo, in luogo della distribuzione di probabilità va introdotta una funzione il cui integrale esteso ad un qualunque prefissato intervallo rappresenta una probabilità.

Definizione 5.5. Sia X una v.c. e sia $F(x)$ la sua funzione di distribuzione. Si dice che X è una v.c. assolutamente continua, e $F(x)$ viene detta assolutamente continua, se esiste una funzione $f(x)$ non negativa tale che $F(x)$ può esprimersi nella forma:

$$F(x) = \int_{-\infty}^x f(t) dt, \quad \forall x \in \mathbb{R}. \quad (5.15)$$

□

La funzione $f(x)$ prende il nome di *densità di probabilità* della v.c. X . Dalle proprietà della funzione di distribuzione $F(x)$ (cfr. la seconda della (iii) del Teorema 5.2) segue la condizione di normalizzazione:

$$\int_{-\infty}^{+\infty} f(t) dt = 1. \quad (5.16)$$

In ogni punto x di continuità della funzione f , in virtù teorema fondamentale del calcolo integrale si ha poi:

$$f(x) = \frac{d}{dx} F(x) = \frac{d}{dx} \int_{-\infty}^x f(t) dt. \quad (5.17)$$

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

5.4 Variabili casuali assolutamente continue

Giova ribadire che, a differenza della distribuzione di probabilità del caso discreto, la densità di probabilità non ha il significato di probabilità; è invece una probabilità il suo integrale esteso ad un qualunque intervallo, così come rappresenta una probabilità la scrittura $f(x) dx$ se $f(x)$ è continua in x . Per rendersene conto nel caso in cui, ad esempio, $f(x)$ è una funzione continua in x , si osservi che dalla (5.17) segue:

$$f(x) = \lim_{\Delta x \rightarrow 0^+} \frac{F(x + \Delta x) - F(x)}{\Delta x}. \quad (5.18)$$

Poiché $F(x)$ è la funzione di distribuzione di X si ha poi:

$$\begin{aligned} F(x + \Delta x) - F(x) &= \mathbb{P}(X \leq x + \Delta x) - \mathbb{P}(X \leq x) \\ &= \mathbb{P}(x < X \leq x + \Delta x) \\ &= \int_x^{x + \Delta x} f(t) dt. \end{aligned} \quad (5.19)$$

Per l'ipotesi di continuità di $f(x)$ la funzione di distribuzione $F(x)$ è derivabile, e quindi anche continua. L'evento $\{X = x\}$ ha allora probabilità nulla, così che nella (5.19) l'intervallo semiaperto può sostituirsi con l'intervallo chiuso:

$$F(x + \Delta x) - F(x) = \mathbb{P}(x \leq X \leq x + \Delta x).$$

Dalla (5.18) segue pertanto:

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

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

È opportuno menzionare che talora si fa uso della scrittura:

$$f(x) dx \approx \mathbb{P}(x < X \leq x + dx).$$

Osserviamo ancora che per vv.cc. assolutamente continue la conoscenza della funzione di distribuzione è equivalente alla conoscenza della densità di probabilità. Ad esempio si ha:

$$\begin{aligned} \int_a^b f(x) dx &= \int_{-\infty}^b f(x) dx - \int_{-\infty}^a f(x) dx \\ &= F(b) - F(a) = \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) \\ &= \mathbb{P}(a < X < b) = \mathbb{P}(a \leq X \leq b). \end{aligned} \quad (5.20)$$

L'integrale della funzione $f(x)$ esteso ad un qualunque intervallo (a, b) esprime dunque la probabilità che la v.c. X assume valori in tale intervallo.

Osservazione 5.6. Con B insieme di Borel di \mathbb{R} la (5.3) consente di ottenere più generalmente:

$$\mathbb{P}(X \in B) = \int_B f_X(x) dx. \quad (5.21)$$

□

Esempio 5.4. Tipici esempi di vv.cc. continue sono costituiti dall'intervallo di tempo intercorrente tra due successive emissioni di particelle da parte di una sostanza radioattiva oppure dall'istante di emissione del primo elettrone da parte di una valvola termoionica. Per fissare le idee, facciamo riferimento a una sostanza radioattiva. Com'è noto, questa emette particelle in modo casuale nel senso che l'istante di emissione di ogni particella non è prevedibile secondo una legge deterministica. Sia T la v.c. che rappresenta l'intervallo di tempo intercorrente tra due successive emissioni di particelle. Evidentemente, T assume valori in $(0, +\infty)$. Supponiamo che la densità di probabilità associata a T , che indicheremo con $f(t)$, sia di tipo esponenziale (in realtà ciò si può dimostrare partendo da ipotesi molto plausibili sul fenomeno fisico sottostante), ossia supponiamo che si abbia:

$$f(t) = \begin{cases} \lambda e^{-\lambda t}, & \text{per } t > 0, \\ 0, & \text{altròve,} \end{cases} \quad (5.22)$$

dove $\lambda > 0$ è un parametro dipendente dalla sostanza radioattiva considerata. Per $t > 0$ la funzione di distribuzione $F(t)$ di T è allora

$$F(t) = \int_{-\infty}^t f(\tau) d\tau = \int_0^t \lambda e^{-\lambda \tau} d\tau = 1 - e^{-\lambda t},$$

mentre essa è nulla per $t < 0$. La descrizione del fenomeno, di conseguenza, può essere effettuata sia attraverso la densità di probabilità che attraverso la funzione di distribuzione. Calcoliamo le probabilità di alcuni eventi.

(a) Sia $I = [\alpha, \alpha + k]$ un intervallo con $\alpha \geq 0$ e $k > 0$. Calcoliamo la probabilità $\mathbb{P}(\alpha \leq T \leq \alpha + k)$. Dalle (5.20) e (5.22) si ha evidentemente:

$$\begin{aligned} \mathbb{P}(\alpha \leq T \leq \alpha + k) &= \int_{\alpha}^{\alpha+k} f(t) dt = \int_{\alpha}^{\alpha+k} \lambda e^{-\lambda t} dt \\ &= e^{-\alpha \lambda} [1 - e^{-k \lambda}] \equiv F(\alpha + k) - F(\alpha). \end{aligned}$$

(b) Fissato $k > 0$, calcoliamo le seguenti quantità:

- (i) $\max_{\alpha} \mathbb{P}(\alpha \leq T \leq \alpha + k)$,
- (ii) $\min_{\alpha} \mathbb{P}(\alpha \leq T \leq \alpha + k)$,

5.4 Variabili casuali assolutamente continue

(iii) il valore di α tale che risulti $\mathbb{P}(\alpha \leq T \leq \alpha + k) = 1/2$.

A tal fine poniamo:

$$\pi(\alpha) = \mathbb{P}(\alpha \leq T \leq \alpha + k) = e^{-\alpha \lambda} (1 - e^{-k \lambda}). \quad (5.23)$$

Essendo

$$\pi'(\alpha) = -\lambda e^{-\alpha \lambda} (1 - e^{-k \lambda}) < 0,$$

$\pi(\alpha)$ è decrescente in α . Pertanto $\mathbb{P}(\alpha \leq T \leq \alpha + k)$ assume il suo valore massimo per $\alpha = 0$, e tale massimo è dato da $1 - e^{-k \lambda}$. Ciò d'altronde segue anche da un esame diretto della (5.23) osservando che è $e^{-\alpha \lambda} \leq 1$ e che $e^{-\alpha \lambda} = 1$ se e solo se risulta $\alpha = 0$. In maniera simile si risponde al quesito di cui al punto (ii), concludendo che $\pi(\alpha)$ non ammette minimo e che, come si evince dalla (5.23), $\mathbb{P}(\alpha \leq T \leq \alpha + k) \rightarrow 0$ per $\alpha \rightarrow +\infty$. Per rispondere infine alla (iii), occorre risolvere la seguente equazione in α :

$$e^{-\alpha \lambda} (1 - e^{-k \lambda}) = \frac{1}{2}.$$

Si ottiene così:

$$\alpha = \lambda^{-1} \ln [2(1 - e^{-k \lambda})].$$

(c) Determinare per quale valore di β si ha $\mathbb{P}(0 \leq T \leq \beta) = 1/2$. Risulta ovviamente:

$$\mathbb{P}(0 \leq T \leq \beta) = \int_0^{\beta} \lambda e^{-\lambda \tau} d\tau = 1 - e^{-\beta \lambda}.$$

Per rispondere al quesito si deve quindi risolvere l'equazione in β

$$1 - e^{-\beta \lambda} = \frac{1}{2}.$$

Si ricava così:

$$\beta = \frac{1}{\lambda} \ln 2. \quad (5.24)$$

□

5.4.1 La mediana

Il valore β dato dalla (5.24) dell'Esempio 5.4 si dice *mediana* della v.c. T considerata. In realtà si può parlare di mediana per una qualunque v.c., a prescindere dall'esempio dato. Precisamente, si dà la seguente definizione.

Definizione 5.6. Si definisce mediana di una v.c. X di funzione di distribuzione $F(x)$ ogni reale M tale che

$$F(M^-) \leq \frac{1}{2} \leq F(M)$$

ossia tale da aversi simultaneamente

$$\mathbb{P}(X < M) \leq \frac{1}{2}, \quad \mathbb{P}(X > M) \equiv 1 - F(M) \leq \frac{1}{2}.$$

□

Se in particolare X è assolutamente continua, la mediana M è tale che

$$\mathbb{P}(X \leq M) = \mathbb{P}(X \geq M) = \frac{1}{2}.$$

Si noti che in generale M può non essere unico.

In particolare, se X è assolutamente continua e se inoltre la sua densità di probabilità $f(x)$ possiede un punto Q di simmetria,⁷ questo è una mediana.

Della mediana si può dare la seguente visualizzazione geometrica. Si consideri il grafico della funzione di distribuzione $F(x)$, riguardando ogni salto verticale come parte del grafico, così da avere una singola curva connessa e non decrescente. Tale curva possiede almeno un punto di intersezione con la retta $y = 1/2$. Se il punto di intersezione M è unico, questo è l'unica mediana della v.c. X . Può tuttavia avvenire che la curva e la retta abbiano un intero intervallo chiuso in comune. In tal caso l'ascissa di ogni punto di tale intervallo soddisfa la condizione $F(x) = 1/2$, ed è pertanto da considerarsi una mediana.

Esempio 5.5. Sia X una v.c. di Bernoulli di parametro $1/2$. Ogni punto $x \in [0, 1]$ è una mediana, avendosi

$$\mathbb{P}(X < x) \leq \frac{1}{2} \quad \text{e} \quad \mathbb{P}(X > x) \leq \frac{1}{2}.$$

□

Altre quantità utili, soprattutto nelle applicazioni statistiche, sono i cosiddetti *percentili*.

Definizione 5.7. Dicesi k -esimo percentile della variabile casuale X di funzione di distribuzione $F(x)$ ogni reale t_k tale da aversi

$$F(t_k^-) \leq \frac{k}{100} \leq F(t_k).$$

□

Si noti che il percentile t_{50} fornisce la mediana. Per evidenti ragioni, i percentili t_{25} , t_{50} e t_{75} vengono anche chiamati *quartili* perché suddividono la funzione di probabilità o la densità di probabilità in quattro parti equiprobabili. Alla differenza $t_{75} - t_{25}$ si dà poi il nome di *intervallo interquartile*; questo viene talora utilizzato come misura di variabilità.

⁷Il punto Q è di simmetria per $f(x)$ se per ogni x risulta $f(Q - x) = f(Q + x)$.

5.5 Alcune proprietà generali

5.5 Alcune proprietà generali

Ogni funzione di distribuzione $F(x)$ gode di talune altre significative proprietà che ci limitiamo qui ad enunciare attraverso il teorema che segue. Le dimostrazioni delle proposizioni in esso presenti possono reperirsi in trattati sulla teoria delle funzioni di variabile reale.

Teorema 5.6. Una funzione di distribuzione $F(x)$ possiede al più un numero finito di punti in cui il salto è maggiore o uguale ad un reale positivo, e conseguentemente al più un'infinità numerabile di punti di discontinuità. La derivata $F'(x)$ esiste quasi ovunque. La funzione $F(x)$ può sempre rappresentarsi come somma di tre componenti

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

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

(i) $F_1(x)$ è la funzione di distribuzione di una v.c. assolutamente continua, ossia tale da aversi

$$F_1(x) = \int_{-\infty}^x f_1(\xi) d\xi \quad x \in \mathbb{R},$$

dove f_1 denota la densità di probabilità;

(ii) $F_2(x)$ è la funzione di distribuzione di una v.c. discreta; per ogni x $F_2(x)$ è uguale alla somma dei salti $p(x_n) = F(x_n) - F(x_n^-)$ di $F(x)$ in tutti i punti x_n di discontinuità con $x_n \leq x$:

$$F_2(x) = \sum_{\{n: x_n \leq x\}} p(x_n) \quad x \in \mathbb{R};$$

(iii) $F_3(x)$ è la componente singolare di F ; è una funzione ovunque continua a derivata quasi ovunque nulla. Essa caratterizza variabili casuali continue ma non assolutamente continue.

Le tre componenti $a_1 F_1, a_2 F_2, a_3 F_3$ sono univocamente determinate da $F(x)$.

Le vv.cc. che sono continue ma non assolutamente continue non rivestono nel presente contesto particolare rilevanza. Pertanto d'ora innanzi assumeremo sempre $a_3 = 0$, e quindi esprimeremo la funzione di distribuzione F nella forma

$$F(x) = a_1 F_1(x) + a_2 F_2(x), \tag{5.25}$$

con a_1, a_2 reali non negativi tali da aversi $a_1 + a_2 = 1$. Di conseguenza, considereremo ogni v.c. come composta soltanto da due parti, l'una assolutamente continua e l'altra discreta. In

particolare, se la v.c. è assolutamente continua si ha $a_1 = 1$ e $a_2 = 0$ mentre se la v.c. è discreta risulta $a_1 = 0$ e $a_2 = 1$.

Le proprietà della funzione di distribuzione espresse dal Teorema 5.6 suggeriscono l'opportunità di sviluppare una trattazione unitaria per lo studio delle vv.cc. senza dover dunque esaminare separatamente i casi di vv.cc. discrete e vv.cc. continue.

Esempio 5.6. Consideriamo una v.c. X di funzione di distribuzione

$$F_X(t) = \begin{cases} 0, & \text{per } t < 0, \\ 1 - 0.95 e^{-\lambda t}, & \text{per } t \geq 0, \end{cases}$$

con λ parametro positivo. Non è difficile rendersi conto che $F_X(t)$ è suscettibile della seguente rappresentazione:

$$F_X(t) = 0.95 G(t) + 0.05 H(t), \quad (5.26)$$

dove $H(t)$ è la funzione di Heaviside definita dalla (5.9) e $G(t)$ è la funzione di distribuzione di una v.c. assolutamente continua con densità di probabilità $g(t)$ esponenziale data dalla (5.22) (cfr. il § A.2.2). La (5.26) è un caso particolare della (5.25) in cui si ponga $a_1 = 0.95$ e $a_2 = 0.05$. La funzione di distribuzione (5.26) è suscettibile, ad esempio, di una semplice interpretazione. Si consideri l'esperimento consistente nello scegliere a caso una lampadina da un deposito contenente il 5% di lampadine fulminate. Nella ipotesi che la durata di funzionamento di una lampadina integra sia una v.c. a densità di probabilità esponenziale di parametro λ , la (5.26) è la funzione di distribuzione della v.c. rappresentante la durata di funzionamento di una lampadina scelta a caso dal suddetto deposito. \triangleleft

5.6 Funzioni di variabili casuali

In un problema di n prove ripetute sia E un evento e sia p la sua probabilità di occorrenza in ogni prova. L'evento E si può verificare da 0 a n volte con probabilità p_0, p_1, \dots, p_n esprimibili in funzione di p (cfr. il § A.1.2). La funzione di distribuzione $F_X(x)$ della v.c. X rappresentante il numero di successi nelle n prove è dunque:

$$F_X(x) = \sum_{\{k \leq x\}} p_k.$$

Consideriamo ora la v.c. $Y = X/n$ ($k = 0, 1, \dots, n$) che fornisce la frequenza con cui l'evento si verifica nelle n prove. Tale v.c. è funzione di X , ma le probabilità con cui essa assume i valori $0, 1/n, \dots, 1$ sono ancora p_0, p_1, \dots, p_n ; la distribuzione di probabilità di Y coincide dunque con la distribuzione di probabilità di X .

Questo semplice esempio è stato introdotto in quanto indicativo di due tipi di problemi:

(a) data una v.c. X e una funzione $g: \mathbb{R} \rightarrow \mathbb{R}$, ha senso definire una nuova v.c. Y tramite la posizione $Y = g(X)$?

(b) se $Y = g(X)$ è una v.c., quale legame sussiste tra le funzioni di distribuzione di X e Y ?

La proposizione che segue fornisce una risposta affermativa alla domanda (a) quando la funzione g soddisfa un requisito non molto restrittivo.

Proposizione 5.2. Sia $X: \Omega \rightarrow \mathbb{R}$ una variabile casuale con funzione di distribuzione $F_X(x)$ e sia $g: \mathbb{R} \rightarrow \mathbb{R}$ una funzione Borel-misurabile. La funzione $Y = g(X): \Omega \rightarrow \mathbb{R}$, definita tramite la posizione $Y(\omega) = g(X(\omega))$ per $\omega \in \Omega$, è allora una v.c. che si dice funzione di X .

Dimostrazione. La dimostrazione è immediata. Infatti per ogni $B \in \mathcal{B}$ si ha $g^{-1}(B) \in \mathcal{B}$ per la Borel-misurabilità di g e $X^{-1}(B) \in \mathcal{F}$ per la \mathcal{F} -misurabilità di X . Quindi per ogni $B \in \mathcal{B}$ si ha $Y^{-1}(B) = X^{-1}[g^{-1}(B)] \in \mathcal{F}$ ossia la \mathcal{F} -misurabilità di Y . \square

Affrontiamo adesso la questione posta nella domanda (b). Dal momento che Y è funzione di X risulta abbastanza agevole intuire che la probabilità di un qualsiasi evento relativo a Y si possa ottenere dal corrispondente evento relativo a X senza passare attraverso lo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$. In realtà è sempre a tale spazio che si fa implicitamente riferimento risultando, per ogni boreiano B , $\mathbb{P}(Y \in B) = \mathbb{P}(Y^{-1}(B))$ con $(Y^{-1}(B))$ evento di \mathcal{F} . Inoltre

$$\mathbb{P}(Y \in B) = \mathbb{P}(g(X) \in B) = \mathbb{P}(X \in g^{-1}([-\infty, y])). \quad (5.27)$$

In particolare per la funzione di distribuzione si ha:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \in g^{-1}([-\infty, y])). \quad (5.28)$$

La formula (5.28) è molto importante in quanto permette di ottenere la funzione di distribuzione di Y a partire dalla funzione di distribuzione di X . Il punto centrale di questo passaggio si ha nella seconda delle uguaglianze, che sposta l'attenzione dalla v.c. Y alla v.c. X , risolvendo in questo modo il problema probabilistico posto nella domanda (b). Ovviamente la natura della v.c. X e le caratteristiche della funzione trasformatrice g renderanno più o meno agevole il problema analitico (calcolo di un integrale o di una serie) posto nella (5.28). Nel seguito sono analizzate alcune delle situazioni che ricorrono sovente nelle applicazioni.

Teorema 5.7. Se X è una v.c. discreta con funzione massa $p_X(x)$ non nulla nei valori della successione $\{x_n\}_n$ e se $g: \mathbb{R} \rightarrow \mathbb{R}$ è una funzione Borel-misurabile, la funzione massa e la funzione di distribuzione della v.c. $Y = g(X)$ sono rispettivamente:

$$p_Y(y) = \sum_i p_X(x_i), \quad \forall x_i \in g^{-1}(y) \quad (5.29)$$

e

$$F_Y(y) = \sum_i p_X(x_i), \quad \forall x_i \in g^{-1}([-\infty, y]). \quad (5.30)$$

Dimostrazione. Sia y un qualunque numero reale e sia B_y il boreiano controimmagine di y tramite g : $B_y = g^{-1}(y)$. Tenendo conto che:

$$\mathbb{P}(X \in B_y) = \mathbb{P}(X \in B_y \cap \{x_n\}_n),$$

la (5.29) si ottiene immediatamente dal teorema delle probabilità totali e dalla (5.27). La (5.30) si ottiene dalla (5.28) con analogo ragionamento. \square

Nel teorema precedente, per non appesantire eccessivamente la notazione, è stato sottinteso che sia la funzione massa che la funzione di distribuzione si annullano in tutti gli argomenti per i quali nessuna x_i appartiene alla corrispondente controimmagine.

Corollario 5.1. *Se X è una v.c. discreta con funzione massa p_X non nulla nei valori della successione $\{x_n\}_{n \in \mathbb{N}}$ e se $g: \mathbb{R} \rightarrow \mathbb{R}$ è una funzione strettamente monotona, la funzione massa della v.c. $Y = g(X)$ è data da:*

$$p_Y(y) = \begin{cases} p_X(x_i), & \text{se } \exists x_i \in \{x_n\}_n \text{ tale che } y = g(x_i), \\ 0, & \text{altrimenti.} \end{cases} \quad (5.31)$$

Dimostrazione. Dalla proprietà di stretta monotonia discende la Borel-misurabilità della funzione g . Inoltre, per ogni y numero reale risulta $g^{-1}(y) \cap \{x_n\}_n = \emptyset$ oppure $g^{-1}(y) \cap \{x_n\}_n = \{x_i\}$ per un certo indice i . La tesi segue dalla (5.29). \square

Corollario 5.2. *Se X è una v.c. discreta con funzione massa p_X non nulla nei valori della successione $\{x_n\}_n$ e se $g: \mathbb{R} \rightarrow \mathbb{R}$ è una funzione strettamente monotona, la funzione di distribuzione della v.c. $Y = g(X)$ è data da:*

$$F_Y(y) = \begin{cases} \mathbb{P}(X \leq g^{-1}(y)) = F_X(g^{-1}(y)), & \text{se } g \text{ è strettamente crescente,} \\ \mathbb{P}(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y)^-), & \text{se } g \text{ è strettamente decrescente.} \end{cases} \quad (5.32)$$

Dimostrazione. Dalla proprietà di stretta monotonia discende la Borel-misurabilità della funzione g . Indicato con y un qualsiasi numero reale si ha:

$$g^{-1}(-\infty, y] = \begin{cases}]-\infty, g^{-1}(y)], & \text{se } g \text{ è strettamente crescente,} \\ [g^{-1}(y), +\infty[, & \text{se } g \text{ è strettamente decrescente.} \end{cases}$$

Il risultato segue immediatamente dalla (5.30) ricordando anche la posizione (5.7). \square

Se la v.c. X è assolutamente continua, rimanendo sotto l'ipotesi di monotonia di g segue che anche Y è assolutamente continua. In questo caso, però, le densità di probabilità non sono legate da una semplice relazione, quale la (5.31) dianzi dimostrata per le distribuzioni di probabilità. Si ha invece il seguente teorema.

Teorema 5.8. *Se X è una v.c. assolutamente continua di densità di probabilità f_X e $g: \mathbb{R} \rightarrow (a, b) \subseteq \mathbb{R}$ è una funzione strettamente monotona, la densità di probabilità f_Y della v.c. $Y = g(X)$ è data da:*

$$f_Y(y) = \begin{cases} \frac{f_X[g^{-1}(y)]}{\left| \frac{dg}{dx} \right|_{x=g^{-1}(y)}}, & \text{per } y \in (a, b), \\ 0, & \text{altrove.} \end{cases} \quad (5.33)$$

Dimostrazione. Distinguiamo i casi di g crescente e di g decrescente.

(i) Supponiamo g strettamente crescente, così che si abbia $g'(x) > 0$. Per ogni $y \in \mathbb{R}$ risulta allora:

$$\mathbb{P}(Y \leq y) = \mathbb{P}\{X \leq g^{-1}(y)\}$$

e quindi:

$$F_Y(y) = \begin{cases} F_X[g^{-1}(y)], & \text{per } y \in [a, b], \\ 0, & \text{per } y < a, \\ 1, & \text{per } y \geq b. \end{cases}$$

Derivando ambo i membri rispetto a y segue così:

$$\begin{aligned} f_Y(y) &\equiv \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X[g^{-1}(y)] \\ &= f_X[g^{-1}(y)] \frac{d}{dy} g^{-1}(y) \\ &= \frac{f_X[g^{-1}(y)]}{\left(\frac{dg}{dx} \right)_{x=g^{-1}(y)}} \quad \text{per } y \in (a, b), \end{aligned} \quad (5.34)$$

mentre risulta $f_Y(y) = 0$ per $y \notin (a, b)$.

(ii) Sia g strettamente decrescente; è allora $g'(x) < 0$. Per ogni $y \in \mathbb{R}$ si ha dunque:

$$\begin{aligned} \mathbb{P}(Y \leq y) &= \mathbb{P}\{X \geq g^{-1}(y)\} = 1 - \mathbb{P}\{X < g^{-1}(y)\} \\ &= 1 - \mathbb{P}\{X \leq g^{-1}(y)\} \end{aligned}$$

e quindi:

$$F_Y(y) = \begin{cases} 1 - F_X[g^{-1}(y)], & \text{per } y \in [a, b], \\ 0, & \text{per } y < a, \\ 1, & \text{per } y \geq b. \end{cases}$$

Derivando ambo i membri rispetto a y si ha poi:

$$\begin{aligned} f_Y(y) &= -f_X[g^{-1}(y)] \frac{d}{dy} g^{-1}(y) \\ &= -\frac{f_X[g^{-1}(y)]}{\left(\frac{dg}{dx}\right)_{x=g^{-1}(y)}}, \quad y \in (a, b) \end{aligned} \tag{5.35}$$

mentre risulta $f_Y(y) = 0$ per $y \notin (a, b)$. Dalle (5.34) e (5.35) segue infine la (5.33).

□

Si noti la diversità tra le formule (5.31) e (5.33): la prima esprime una legge di trasformazione di probabilità mentre la seconda riguarda densità di probabilità. Così, mentre per trasformazioni monotone crescenti di v.v.c. nel caso discreto le distribuzioni di probabilità non cambiano, nel caso continuo le densità di probabilità sono diverse a meno che non sia $g'(x) = 1$ e quindi $Y = X + \text{costante}$. In ogni caso il divario tra le due densità di probabilità dipende dalla funzione $|dg/dx|$; solo se $|dg/dx| \approx 1$, le due densità di probabilità sono prossime l'una all'altra.

Esempio 5.7. Sia X una v.c. *normale standard*, ossia di densità di probabilità

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad (x \in \mathbb{R}) \tag{5.36}$$

e sia $g(x) = e^x$. Considerata la v.c. $Y = e^X$, dalla (5.33) segue:

$$f_Y(y) = \frac{f_X(\ln y)}{y} = \begin{cases} \frac{1}{y\sqrt{2\pi}} e^{-(\ln y)^2/2}, & \text{per } y > 0, \\ 0, & \text{per } y \leq 0. \end{cases} \tag{5.37}$$

Si noti che $f_Y(y)$ è definibile per continuità nel punto $y = 0$, avendosi $\lim_{y \downarrow 0} f_Y(y) = 0$. È evidente che le densità di probabilità (5.36) e (5.37) esibiscono comportamenti molto diversi.

Ad esempio, $f_Y(y)$ è nulla per $y < 0$ e non possiede la proprietà di simmetria intorno al suo massimo, che invece caratterizza $f_X(x)$. La (5.37) è nota quale densità di probabilità *lognormale*. □

Se g non è invertibile bisognerà cercare di risolvere caso per caso, come già si è detto, il problema analitico posto nella (5.28). Ad illustrazione di questo punto viene proposto il seguente esempio.

Esempio 5.8. Sia $g(x) = x^2$. Data una v.c. assolutamente continua X , definiamo la v.c. $Y = g(X) = X^2$. In questo caso, per ogni y non negativo si ha:

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

Quindi, in definitiva, per y positivo risulta:

$$F_Y(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y}),$$

laddove, per le densità di probabilità, si ha:

$$f_Y(y) = \frac{f_X(\sqrt{y})}{2\sqrt{y}} - \frac{f_X(-\sqrt{y})}{2\sqrt{y}} = \frac{f_X(\sqrt{y}) + f_X(-\sqrt{y})}{2\sqrt{y}}.$$

Infine sia $F_Y(y)$ che $f_Y(y)$ sono nulle per $y \leq 0$. □

Analogamente si procede in casi più complessi.

Concludiamo questo paragrafo facendo esplicitamente osservare che mentre la stretta monotonia della funzione g non altera la natura della v.c. X , (nel senso che ad una v.c. discreta corrisponde una v.c. discreta ed a una v.c. assolutamente continua corrisponde una v.c. assolutamente continua) la stessa cosa non accade se g non è strettamente monotona. Precisamente, se X è una v.c. assolutamente continua la v.c. $Y = g(X)$ non è necessariamente tale. Supponiamo ad esempio che sia $g(x) = k$, con k costante, per $x \in [a, b]$ e supponiamo che risulti $\mathbb{P}(a \leq X \leq b) > 0$. In questo caso si ha:

$$\mathbb{P}(Y = k) = \mathbb{P}(a \leq X \leq b) > 0.$$

Quindi la funzione di distribuzione di Y non può essere continua dal momento che risulta $\mathbb{P}(Y = k) > 0$ e non $\mathbb{P}(Y = k) = 0$.

5.7 Variabili casuali multidimensionali

In questo paragrafo, detta X una variabile casuale, indicheremo l'evento $\{\omega \in \Omega : X(\omega) \leq x\}$ semplicemente con $\{X \leq x\}$ e l'evento $\{\omega \in \Omega : X(\omega) = x\}$ con $\{X = x\}$.

Definizione 5.8. Sia n un intero maggiore di 1 e siano X_1, \dots, X_n n variabili casuali definite in uno stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$.⁸ La n -upla $X = (X_1, \dots, X_n)$ viene detta variabile casuale n -dimensionale, o vettore casuale. \triangleleft

Per pervenire alla definizione della funzione di distribuzione di X , cominciamo con l'osservare che per ogni n -upla $x = (x_1, \dots, x_n)$ di numeri reali risulta:

$$\{X_1 \leq x_1; \dots; X_n \leq x_n\} = \{X_1 \leq x_1\} \cap \dots \cap \{X_n \leq x_n\}. \quad (5.38)$$

L'insieme a primo membro della (5.38) è un elemento di \mathcal{F} , ossia un evento, in quanto intersezione di elementi di \mathcal{F} .

Definizione 5.9. La funzione $F_X \equiv F_{X_1 \dots X_n} : \mathbb{R}^n \rightarrow [0, 1]$ così definita

$$F_X(x) \equiv F_{X_1 \dots X_n}(x_1, \dots, x_n) = \mathbb{P}(X_1 \leq x_1; \dots; X_n \leq x_n) \quad (5.39)$$

viene chiamata funzione di distribuzione congiunta delle variabili casuali X_1, \dots, X_n o funzione di distribuzione della v.c. n -dimensionale X . \triangleleft

Si noti poi che, se k è un intero compreso tra 1 e n , il $\lim_{x_k \rightarrow +\infty} F_X(x)$ fornisce la probabilità dell'evento:

$$\{X_1 \leq x_1; \dots; X_{k-1} \leq x_{k-1}; \Omega; X_{k+1} \leq x_{k+1}; \dots; X_n \leq x_n\}.$$

Si può allora concludere che la funzione di distribuzione della v.c. $(n - 1)$ -dimensionale

$$(X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_n)$$

si ottiene da quella di X eseguendone il limite per x_k tendente a $+\infty$. Procedendo allo stesso modo si può determinare la funzione di distribuzione congiunta di m vv.cc. X_{i_1}, \dots, X_{i_m} scelte tra (X_1, \dots, X_n) .

Definizione 5.10. Si considerino m ($m < n$) delle n vv.cc. X_1, \dots, X_n e le si riordinino in modo da indicarle con X_1, \dots, X_m . La funzione di distribuzione marginale di X_1, \dots, X_m è così definita:

$$F_{X_1 \dots X_m}(x_1, \dots, x_m) = \lim_{x_{m+1} \rightarrow +\infty} \dots \lim_{x_n \rightarrow +\infty} F_{X_1 \dots X_n}(x_1, \dots, x_m, x_{m+1}, \dots, x_n). \quad (5.40)$$

\triangleleft

⁸In tal caso le vv.cc. X_1, \dots, X_n si dicono congiuntamente distribuite.

5.7 Variabili casuali multidimensionali

Determiniamo le proprietà della funzione di distribuzione congiunta. Dalle (5.38) e (5.39) segue che $F_{X_1 \dots X_n}$ è non decrescente in ciascuno dei suoi argomenti.⁹ Inoltre, se x_k ($k = 1, 2, \dots, n$) tende a $-\infty$ il corrispondente evento $\{X_k \leq x_k\}$ tende all'evento impossibile ed a questo tende, quindi, anche l'intersezione a secondo membro della (5.38). Si ha dunque:

$$\lim_{x_k \rightarrow -\infty} F_{X_1 \dots X_n}(x_1, \dots, x_n) = 0 \quad (k = 1, \dots, n),$$

con $x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n$ reali arbitrari. Dalla (5.39) si trae anche:

$$\lim_{x_1 \rightarrow +\infty} \dots \lim_{x_n \rightarrow +\infty} F_{X_1 \dots X_n}(x_1, \dots, x_n) = 1.$$

Analogamente al caso unidimensionale si può definire la densità di probabilità di una v.c. n -dimensionale.

Definizione 5.11. Una v.c. n -dimensionale $X = (X_1, X_2, \dots, X_n)$ avente funzione di distribuzione $F_{X_1 \dots X_n}$, è assolutamente continua se esiste una funzione non negativa

$$f_X = f_{X_1 \dots X_n} : \mathbb{R}^n \rightarrow \mathbb{R}$$

tale che per ogni n -upla (x_1, \dots, x_n) di reali risulti

$$F_{X_1 \dots X_n}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} d\xi_1 \dots \int_{-\infty}^{x_n} f_{X_1 \dots X_n}(\xi_1, \dots, \xi_n) d\xi_n. \quad (5.41)$$

La funzione f_X è detta funzione densità di probabilità congiunta della v.c. n -dimensionale (X_1, \dots, X_n) . \triangleleft

In ogni n -upla di valori (x_1, \dots, x_n) in cui la funzione di distribuzione ammette derivate di ordine n continue, per la (5.41) si ha:

$$f_{X_1 \dots X_n}(x_1, \dots, x_n) = \frac{\partial^n}{\partial x_1 \dots \partial x_n} F_{X_1 \dots X_n}(x_1, \dots, x_n). \quad (5.42)$$

Nel caso in cui la v.c. n -dimensionale (X_1, \dots, X_n) è assolutamente continua, facendo uso delle (5.40), (5.41) e (5.42) è possibile esprimere la densità di probabilità marginale di un qualsiasi numero m ($m < n$) di variabili casuali. Ad esempio, per la densità di probabilità congiunta delle prime m variabili casuali si ha:

$$f_{X_1 \dots X_m}(x_1, \dots, x_m) = \int_{-\infty}^{+\infty} dx_{m+1} \dots \int_{-\infty}^{+\infty} f_{X_1 \dots X_n}(x_1, \dots, x_n) dx_n. \quad (5.43)$$

⁹Una funzione $F(x_1, x_2, \dots, x_n)$ si dice non decrescente in ciascuno dei suoi argomenti se risulta $\Delta_{x_1} \Delta_{x_2} \dots \Delta_{x_n} F \geq 0$, dove $\Delta_{x_i} F := F(x_1, \dots, x_{i-1}, x_i + \Delta x_i, x_{i+1}, \dots, x_n) - F(x_1, \dots, x_n)$.

Si noti che per vv.cc. assolutamente continue la conoscenza della densità di probabilità consente di calcolare $\mathbb{P}(X \in B)$ per ogni assegnato $B \in \mathcal{B}^n$.¹⁰

$$\mathbb{P}(X \in B) = \mathbb{P}((X_1, \dots, X_n) \in B) = \int_B f_{X_1 \dots X_n}(x_1, \dots, x_n) dx_1 \dots dx_n. \quad (5.44)$$

Definizione 5.12. Una v.c. n -dimensionale $X = (X_1, X_2, \dots, X_n)$ è detta discreta se esistono una funzione

$$p_X = p_{X_1 \dots X_n} : \mathbb{R}^n \rightarrow [0, 1]$$

ed n successioni $\{x_i^{(k)}\}_{i \in \mathbb{N}}$ ($k = 1, 2, \dots, n$) di reali tali che si abbia:

$$F_X(x) = F_{X_1 \dots X_n}(x_1, \dots, x_n) = \sum_{\{r_1 : x_{r_1}^{(1)} \leq x_1\}} \dots \sum_{\{r_n : x_{r_n}^{(n)} \leq x_n\}} p_X(x_{r_1}^{(1)}, \dots, x_{r_n}^{(n)}), \quad (5.45)$$

dove

$$p_X(x_1, \dots, x_n) = \mathbb{P}(X_1 = x_1; \dots; X_n = x_n). \quad (5.46)$$

La funzione $p_{X_1 \dots X_n}(x_1, \dots, x_n)$ prende il nome di distribuzione di probabilità congiunta. \triangleleft

Nel dimostrare proprietà relative a vv.cc. multidimensionali considereremo sempre, a meno di esplicito avviso contrario, il caso di vv.cc. assolutamente continue. Le considerazioni svolte nel § 5.6 possono immediatamente estendersi al caso multidimensionale. In particolare è agevole dimostrare che un vettore ottenuto mediante una trasformazione Borel-misurabile di un vettore casuale risulta essere esso stesso un vettore casuale. A titolo di esempio si riporta l'enunciato della versione vettoriale del Teorema 5.8.

Teorema 5.9. Sia $X = (X_1, X_2, \dots, X_n)$ un vettore casuale a n -dimensioni assolutamente continuo di densità di probabilità congiunta f_X e sia $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ una funzione Borel-misurabile con jacobiano J sempre diverso da zero. Posto

$$Y = g(X) = (g_1(X), g_2(X), \dots, g_m(X)) = (Y_1, Y_2, \dots, Y_n)$$

si ha:

$$f_Y(y_1, y_2, \dots, y_m) = f_X(x_1, x_2, \dots, x_n) \Big|_{x_i = g_i^{-1}(y_1, y_2, \dots, y_m)} \cdot |J|. \quad (5.47)$$

¹⁰Come nel caso unidimensionale, \mathcal{B}^n rappresenta la σ -algebra di Borel di \mathbb{R}^n , ossia la minima sigma algebra su \mathbb{R}^n contenente tutti gli insiemi del tipo $]-\infty, x_1] \times]-\infty, x_2] \times \dots \times]-\infty, x_n]$.

5.7.1 Variabili casuali bidimensionali

Esaminiamo come si specializzano le relazioni determinate finora quando ci si ponga nel caso particolare $n = 2$. Date due vv.cc. X e Y congiuntamente distribuite, la (5.39) assume la forma:

$$F_{XY}(x, y) = \mathbb{P}(X \leq x; Y \leq y).$$

Risulta poi:

$$\lim_{x \rightarrow -\infty} F_{XY}(x, y) = \lim_{y \rightarrow -\infty} F_{XY}(x, y) = 0, \quad \lim_{x \rightarrow +\infty} \lim_{y \rightarrow +\infty} F_{XY}(x, y) = 1.$$

La v.c. bidimensionale (X, Y) è assolutamente continua se esiste una funzione $f_{XY}(x, y)$ tale da aversi:

$$F_{XY}(x, y) = \int_{-\infty}^x d\xi \int_{-\infty}^y f_{XY}(\xi, \eta) d\eta, \quad \forall x, y \in \mathbb{R}^2. \quad (5.48)$$

Se $F_{XY}(x, y)$ ammette derivate seconde continue si ha poi [Dimostrazione]:

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y}. \quad (5.49)$$

Applicando la (5.43) si ottengono le densità marginali di X e di Y :

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dy, \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dx.$$

Nel caso di vv.cc. discrete congiuntamente distribuite la (5.45) assume la forma:

$$F_{XY}(x, y) = \sum_{\xi \leq x} \sum_{\eta \leq y} p_{XY}(\xi, \eta),$$

dove

$$p_{XY}(x, y) \equiv \mathbb{P}(X = x, Y = y).$$

Le distribuzioni marginali in questo caso diventano:

$$p_X(x) = \sum_y p_{XY}(x, y)$$

$$p_Y(y) = \sum_x p_{XY}(x, y).$$

5.8 Indipendenza di variabili casuali

Siano X, Y vv.cc. congiuntamente distribuite su uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e sia $F_{XY}(x, y)$ la loro funzione di distribuzione congiunta. È ragionevole dire che X, Y sono *indipendenti* se per ogni $x, y \in \mathbb{R}$ gli eventi $\{X \leq x\}$ e $\{Y \leq y\}$ sono indipendenti, ossia se risulta:

$$\mathbb{P}(X \leq x; Y \leq y) = \mathbb{P}(X \leq x) \cdot \mathbb{P}(Y \leq y).$$

Invero, si dà la seguente definizione:

Definizione 5.13. Le vv.cc. X, Y si dicono indipendenti, o stocasticamente indipendenti, se la loro funzione di distribuzione congiunta si fattorizza nel prodotto delle funzioni di distribuzione delle vv.cc. singolarmente prese:

$$F_{XY}(x, y) = F_X(x)F_Y(y) \quad \forall x, y \in \mathbb{R}. \quad (5.50)$$

□

Tenendo conto della (5.42), se le vv.cc. X e Y indipendenti sono assolutamente continue, per la (5.50) la densità di probabilità congiunta $f_{XY}(x, y)$ verifica la relazione

$$f_{XY}(x, y) \equiv \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y} = \frac{dF_X(x)}{dx} \frac{dF_Y(y)}{dy} = f_X(x)f_Y(y). \quad (5.51)$$

La (5.51) afferma che la densità di probabilità congiunta di due vv.cc. indipendenti si fattorizza nel prodotto delle densità di probabilità delle singole variabili casuali.

La nozione di indipendenza espressa dalla Definizione 5.13 nel caso di due vv.cc. si estende al caso di un arbitrario numero di variabili.

Definizione 5.14. Siano X_1, \dots, X_n vv.cc. congiuntamente distribuite su uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ fissato e sia $F_{X_1 \dots X_n}(x_1, \dots, x_n)$ la loro funzione di distribuzione congiunta; esse si dicono indipendenti, o stocasticamente indipendenti, se risulta:

$$F_{X_1 \dots X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n), \quad \forall x_1, x_2, \dots, x_n \in \mathbb{R}. \quad (5.52)$$

□

Proposizione 5.3. Se le vv.cc. X_1, \dots, X_n sono indipendenti, tali risultano anche k di esse ($k < n$), comunque scelte.

Proof. Siano $X_{j_1}, X_{j_2}, \dots, X_{j_k}$ k delle n variabili casuali. Si ha:

$$\begin{aligned} F_{X_1 \dots X_n}(x_1, \dots, x_n) &= F_{Y_1 \dots Y_n}(y_1, \dots, y_n) \\ &= F_{Y_1}(y_1)F_{Y_2}(y_2) \cdots F_{Y_n}(y_n), \end{aligned}$$

5.8 Indipendenza di variabili casuali

dove

$$Y_i = \begin{cases} X_i, & \text{se } X_i \neq X_{j_1}, X_{j_2}, \dots, X_{j_k}, \\ X_{j_r}, & \text{se } X_i = X_{j_r} \quad (r = 1, 2, \dots, k), \end{cases}$$

e dove si è fatto uso della (5.52). L'asserto segue allora ponendo

$$y_i = \begin{cases} +\infty, & \text{se } X_i \neq X_{j_1}, X_{j_2}, \dots, X_{j_k}, \\ x_i, & \text{altrimenti.} \end{cases}$$

Dalla (5.52) si trae infatti:

$$F_{X_{j_1} \dots X_{j_k}}(x_{j_1}, \dots, x_{j_k}) = F_{X_{j_1}}(x_{j_1}) \cdots F_{X_{j_k}}(x_{j_k}). \quad \square$$

Nel caso di n vv.cc. assolutamente continue indipendenti sussiste una relazione analoga alla (5.51) per la densità di probabilità congiunta:

$$f_{X_1 \dots X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n). \quad (5.53)$$

Definizione 5.15. Siano (X_1, \dots, X_n) e (Y_1, \dots, Y_m) vv.cc., rispettivamente n -dimensionale e m -dimensionale, assolutamente continue di densità di probabilità $f_{X_1 \dots X_n}$ e $f_{Y_1 \dots Y_m}$; esse si dicono indipendenti se e solo se risulta:

$$f_{X_1 \dots X_n, Y_1 \dots Y_m}(x_1, \dots, x_n, y_1, \dots, y_m) = f_{X_1 \dots X_n}(x_1, \dots, x_n) f_{Y_1 \dots Y_m}(y_1, \dots, y_m). \quad (5.54)$$

□

Proposizione 5.4. Siano X e Y vv.cc. indipendenti e siano h e g funzioni monotone a valori reali. Le vv.cc. $X_1 = h(X)$ e $Y_1 = g(Y)$ sono indipendenti.

Proof. Per brevità effettuiamo la dimostrazione supponendo che h e g siano entrambe crescenti. Si ha:

$$\begin{aligned} F_{X_1, Y_1}(x_1, y_1) &= \mathbb{P}(X_1 \leq x_1; Y_1 \leq y_1) = \mathbb{P}\{h(X) \leq x_1; g(Y) \leq y_1\} \\ &= \mathbb{P}\{X \leq h^{-1}(x_1); Y \leq g^{-1}(y_1)\} = \mathbb{P}\{X \leq h^{-1}(x_1)\} \cdot \mathbb{P}\{Y \leq g^{-1}(y_1)\} \\ &= \mathbb{P}\{h(X) \leq x_1\} \cdot \mathbb{P}\{g(Y) \leq y_1\} = \mathbb{P}(X_1 \leq x_1) \cdot \mathbb{P}(Y_1 \leq y_1) \\ &= F_{X_1}(x_1) F_{Y_1}(y_1), \end{aligned}$$

che mostra l'indipendenza di X_1 e Y_1 . Analogamente si procede negli altri casi. □

L'indipendenza delle vv.cc. $h(X)$ e $g(Y)$ di cui alla Proposizione 5.4 sussiste più in generale nell'ipotesi in cui h e g sono funzioni di Borel, ossia funzioni invertibili che trasformano insiemi di Borel in insiemi di Borel.

5.9 Somme, prodotti e rapporti di variabili casuali

Lo studio di vv.cc. n -dimensionali fornisce uno strumento utile anche per evidenziare proprietà relative a particolari vv.cc. unidimensionali. Ad esempio, date n vv.cc. X_1, X_2, \dots, X_n possiamo porci il problema di determinare la funzione di distribuzione della v.c. somma, ossia della v.c.

$$Y = X_1 + \dots + X_n.$$

Per definizione si ha:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X_1 + \dots + X_n \leq y);$$

$F_Y(y)$ è dunque la probabilità che la v.c. $(X_1 + \dots + X_n)$ assuma valore $(x_1 + \dots + x_n)$ e che al contempo il punto $x = (x_1, \dots, x_n)$ appartenga al semispazio

$$\{x \in \mathbb{R}^n : x_1 + \dots + x_n \leq y\}.$$

Supponiamo ora che X_1, \dots, X_n siano vv.cc. assolutamente continue congiuntamente distribuite. Indicata allora con $f_{X_1 \dots X_n}(x_1, \dots, x_n)$ la loro densità di probabilità congiunta, facendo uso della (5.44) si ottiene:

$$F_Y(y) = \int_D f_{X_1 \dots X_n}(x_1, \dots, x_n) dx_1 \dots dx_n, \quad (5.55)$$

dove $D = \{x \in \mathbb{R}^n : x_1 + \dots + x_n \leq y\}$.

5.9.1 Somma di due variabili casuali

Per semplicità di calcolo consideriamo in particolare il caso $n = 2$; i risultati possono essere poi estesi al caso di n intero qualsiasi. Posto $Y = X_1 + X_2$, dalla (5.55) segue:

$$\begin{aligned} F_Y(y) &= \int \int_{x_1+x_2 \leq y} f_{X_1 X_2}(x_1, x_2) dx_1 dx_2 \\ &= \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{y-x_1} f_{X_1 X_2}(x_1, x_2) dx_2 \\ &= \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^y f_{X_1 X_2}(x_1, z - x_1) dz, \end{aligned} \quad (5.56)$$

dove si è posto $x_2 = z - x_1$.

5.9 Somme, prodotti e rapporti di variabili casuali

Esempio 5.9. Diamo un esempio di calcolo della densità di probabilità della somma $Y = X_1 + X_2$ di due vv.cc. assolutamente continue X_1 e X_2 . Supponiamo che la loro funzione di distribuzione congiunta sia

$$f_{X_1 X_2}(x_1, x_2) = \begin{cases} (1 - e^{-\lambda x_1})(1 - e^{-\lambda x_2}), & \text{per } x_1, x_2 > 0, \\ 0, & \text{altrove,} \end{cases} \quad (5.57)$$

dove λ è un parametro reale positivo. Usando la (5.49) si ottiene la densità di probabilità congiunta:

$$f_{X_1 X_2}(x_1, x_2) = \begin{cases} \lambda^2 e^{-\lambda(x_1+x_2)}, & \text{per } x_1, x_2 > 0, \\ 0, & \text{altrove.} \end{cases}$$

Come nella (5.55) si ha poi:

$$F_Y(y) = \int_{\{(x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 \leq y\}} \lambda^2 e^{-\lambda(x_1+x_2)} dx_1 dx_2.$$

L'integrale a secondo membro è nullo se risulta $y \leq 0$, mentre per $y > 0$ si ha:

$$\begin{aligned} F_Y(y) &= \int_0^y dx_1 \int_0^{y-x_1} dx_2 \lambda^2 e^{-\lambda(x_1+x_2)} = \lambda \int_0^y dx_1 (e^{-\lambda x_1} - e^{-\lambda y}) \\ &= 1 - e^{-\lambda y} - \lambda y e^{-\lambda y}. \end{aligned} \quad (5.58)$$

Derivando la (5.58) rispetto a y si ottiene infine la densità di probabilità della variabile casuale Y :

$$f_Y(y) = \begin{cases} \lambda^2 y e^{-\lambda y}, & \text{per } y > 0, \\ 0, & \text{altrove.} \end{cases} \quad (5.59)$$

Si noti che la (5.57) è la funzione di distribuzione congiunta di due variabili casuali indipendenti ed esponenziali di parametro λ . Come mostra la (A.39), la (5.59) è una densità gamma di parametri $\alpha = 2$ e $\beta = \lambda^{-1}$. \triangleleft

La densità di probabilità di $Y = X_1 + X_2$ si può ottenere direttamente dalla formula (5.56) calcolando la derivata rispetto a y di ambo i membri. Infatti, applicando la formula di derivazione sotto il segno di integrale, si ottiene:

$$\begin{aligned} f_Y(y) &= \frac{dF_Y(y)}{dy} = \frac{d}{dy} \left[\int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{y-x_1} f_{X_1 X_2}(x_1, x_2) dx_2 \right] \\ &= \int_{-\infty}^{+\infty} f_{X_1 X_2}(x_1, y - x_1) dx_1. \end{aligned} \quad (5.60)$$

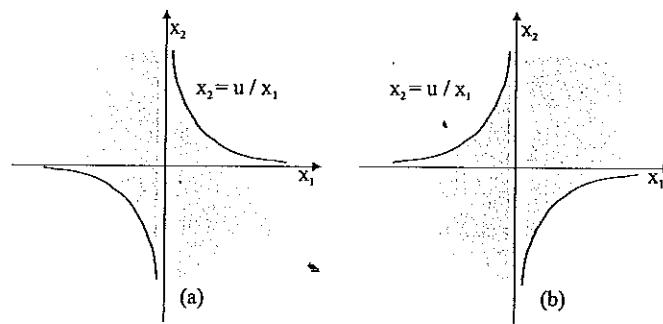


Figura 5.1 – Regioni di integrazione per il prodotto di due variabili casuali.

La formula (5.60) esprime l'assoluta continuità della v.c. Y . Nel caso di vv.cc. discrete procedendo in modo analogo si ottiene:

$$\mathbb{P}_Y(y) = \sum_{x_1} \mathbb{P}_{X_1 X_2}(x_1, y - x_1)$$

dove $\mathbb{P}_{X_1 X_2}(x_1, x_2)$ denota la distribuzione di probabilità congiunta della coppia di vv.cc. (X_1, X_2) .

Inoltre, nel caso di indipendenza tra le due variabili casuali X_1 e X_2 (nel qual caso la funzione di densità di probabilità congiunta si fattorizza nel prodotto delle funzioni di densità di probabilità unidimensionali) la (5.60) diventa la formula di convoluzione:

$$f_Y(y) = f_{X_1 + X_2}(y) = \int_{-\infty}^{+\infty} f_{X_1}(x_1) f_{X_2}(y - x_1) dx_1. \quad (5.61)$$

5.9.2 Prodotto di due variabili casuali

Passiamo ora a considerare la variabile casuale U prodotto delle variabili casuali X_1 e X_2 . Ricordando la (5.44), applicata al caso $n = 2$, si ottiene:

$$F_U(u) = \mathbb{P}(U \leq u) = \mathbb{P}(X_1 \cdot X_2 \leq u) = \int_{x_1 x_2 \leq u} f_{X_1 X_2}(x_1, x_2) dx_1 dx_2.$$

Se è $u > 0$, la curva di equazione $u = x_1 x_2$ rappresenta l'iperbole equilatera $x_2 = u/x_1$ i cui rami giacciono nel primo e nel terzo quadrante; decomponendo pertanto la regione

5.9 Somme, prodotti e rapporti di variabili casuali

d'integrazione (ombreggiata nella parte (a) della Figura 5.1) nelle due zone $x_1 < 0$ e $x_1 > 0$ e integrando per verticali si ha:

$$F_U(u) = \int_{-\infty}^0 dx_1 \int_{u/x_1}^{+\infty} f_{X_1 X_2}(x_1, x_2) dx_2 + \int_0^{+\infty} dx_1 \int_{-\infty}^{u/x_1} f_{X_1 X_2}(x_1, x_2) dx_2.$$

Derivando rispetto ad u entrambi i membri di quest'ultima tramite la formula di derivazione sotto il segno di integrale si ottiene infine:

$$\begin{aligned} f_U(u) &= f_{X_1 \cdot X_2}(u) = \int_{-\infty}^0 -\frac{1}{x_1} f_{X_1 X_2}(x_1, \frac{u}{x_1}) dx_1 + \int_0^{+\infty} \frac{1}{x_1} f_{X_1 X_2}(x_1, \frac{u}{x_1}) dx_1 \\ &= \int_{-\infty}^{+\infty} \frac{1}{|x_1|} f_{X_1 X_2}(x_1, \frac{u}{x_1}) dx_1. \end{aligned} \quad (5.62)$$

Nel caso $u < 0$, facendo riferimento alla regione d'integrazione ombreggiata nella parte (b) della Figura 5.1 e riapplicando il precedente ragionamento si perviene allo stesso risultato espresso nella (5.62) che quindi fornisce la funzione di densità di probabilità della variabile casuale prodotto.

5.9.3 Rapporto di due variabili casuali

Passiamo ora a considerare la variabile casuale W rapporto delle vv.cc. X_1 e X_2 . Ricordando la (5.44), applicata al caso $n = 2$, si ottiene:

$$F_W(w) = \mathbb{P}(W \leq w) = \mathbb{P}\left(\frac{X_1}{X_2} \leq w\right) = \int_{x_1/x_2 \leq w} f_{X_1 X_2}(x_1, x_2) dx_1 dx_2.$$

La curva di equazione $w = x_1/x_2$ rappresenta la retta $x_1 = w x_2$ giacente nel primo e nel terzo quadrante se $w > 0$, nel secondo e nel quarto quadrante se $w < 0$. Il grafico di tali rette è mostrato, rispettivamente, nelle parti (a) e (b) della Figura 5.2. Decomponendo la regione d'integrazione (ombreggiata nella Figura 5.2) nelle due zone $x_2 < 0$ e $x_2 > 0$ e integrando per orizzontali si ha:

$$F_W(w) = \int_{-\infty}^0 dx_2 \int_{w x_2}^{+\infty} f_{X_1 X_2}(x_1, x_2) dx_1 + \int_0^{+\infty} dx_2 \int_{-\infty}^{w x_2} f_{X_1 X_2}(x_1, x_2) dx_1.$$

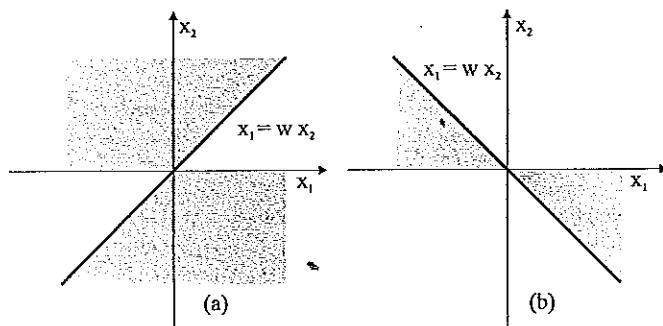


Figura 5.2 – Regioni di integrazione per il rapporto di due variabili casuali.

Derivando rispetto a w entrambi i membri di quest'ultima tramite la formula di derivazione sotto il segno di integrale si ottiene infine:

$$\begin{aligned} f_W(w) &= f_{\frac{x_1}{x_2}}(w) = \int_{-\infty}^0 -x_2 f_{X_1 X_2}(wx_2, x_2) dx_2 + \int_0^{+\infty} x_2 f_{X_1 X_2}(wx_2, x_2) dx_2 \\ &= \int_{-\infty}^{+\infty} |x_2| f_{X_1 X_2}(wx_2, x_2) dx_2. \end{aligned} \quad (5.63)$$

La (5.63) fornisce la funzione di densità di probabilità della variabile casuale quoziente.

CAPITOLO 6

Momenti di Variabili Casuali

6.1 Momenti di variabili unidimensionali

Come abbiamo visto, una v.c. è caratterizzata dalla sua funzione di distribuzione. Tuttavia quest'ultima non è sempre determinabile esplicitamente; ad esempio, come già menzionato in passato, essa potrebbe essere soluzione di un'equazione che non si riesce a risolvere. In questi casi può risultare utile la conoscenza di alcune costanti che vengono associate alla v.c. per ottenerne una sia pur approssimativa descrizione. Tra queste particolare importanza rivestono il *valore medio*, la *varianza* e, in generale, i *momenti*.

Definizione 6.1. Sia X una v.c. di funzione di distribuzione F . Se non sono entrambi infiniti i limiti

$$\lim_{b \rightarrow +\infty} \int_0^b x dF(x) \quad \text{e} \quad \lim_{a \rightarrow -\infty} \int_a^0 x dF(x), \quad (6.1)$$

l'integrale

$$E(X) := \int_{-\infty}^{+\infty} x dF(x) \equiv \lim_{b \rightarrow +\infty} \int_0^b x dF(x) + \lim_{a \rightarrow -\infty} \int_a^0 x dF(x) \quad (6.2)$$

prende il nome di *valore medio* di X . □

Spesso il valore medio viene chiamato indifferentemente *media*, *valore di aspettazione* (*expectation* in Inglese, da cui ha origine la notazione $E(\cdot)$), *speranza matematica* o *valore atteso*, e denotato anche con il simbolo $\langle X \rangle$.

Si noti che può avversi $E(X) = +\infty$ oppure $E(X) = -\infty$. Affinché il valore medio di X sia finito occorre che i limiti nella (6.1) siano entrambi finiti, il che comporta l'assoluta convergenza dell'integrale che definisce $E(X)$. Infine, se i limiti nella (6.1) risultano valere rispettivamente $+\infty$ e $-\infty$, si dice che la v.c. X non è dotata di valore medio.

Dalle proprietà dell'integrale di Stieltjes (cfr. il § B.2) segue che se X è una v.c. assolutamente continua, risulta

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx, \quad (6.3)$$

dove $f = F'$ è la densità di probabilità di X ; se invece X è discreta, si ha:

$$E(X) = \sum_r x_r p(x_r), \quad (6.4)$$

dove $\{p(x_r)\}$ è la distribuzione di probabilità di X .

È poi appena il caso di menzionare esplicitamente che una v.c. ad un numero finito di valori, oppure a valori tutti non negativi o tutti non positivi, è sempre dotata di media. Noi, per gli scopi di questo testo, da questo punto in poi tralasseremo il caso delle vv.cc. aventi media infinita e, pertanto, quando diremo che una v.c. è dotata di media intenderemo che l'integrale che la definisce è assolutamente convergente.

Alla nozione di valore medio si può dare una suggestiva interpretazione geometrica. Invero, il valore medio definito dalla (6.2) esiste finito se e solo se i seguenti integrali

$$\int_0^{+\infty} [1 - F(x)] dx, \quad \int_{-\infty}^0 F(x) dx$$

sono finiti, e in tal caso si ha:

$$E(X) = - \int_{-\infty}^0 F(x) dx + \int_0^{+\infty} [1 - F(x)] dx. \quad (6.5)$$

Accenniamo alla dimostrazione riferendoci, per semplicità, ad una v.c. assolutamente continua di densità di probabilità $f = F'$, così che

$$E(X) = \int_{-\infty}^0 y f(y) dy + \int_0^{+\infty} y f(y) dy. \quad (6.6)$$

Osserviamo che risulta:

$$\begin{aligned} \int_{-\infty}^0 y f(y) dy &= \int_{-\infty}^0 dy f(y) \int_0^y dx \\ &= \int_{-\infty}^0 dx \int_x^{-\infty} dy f(y) dy = - \int_{-\infty}^0 F(x) dx \end{aligned}$$

e

$$\begin{aligned} \int_0^{+\infty} y f(y) dy &= \int_0^{+\infty} dy f(y) \int_0^y dx \\ &= \int_0^{+\infty} dx \int_x^{+\infty} dy f(y) dy = \int_0^{+\infty} [1 - F(x)] dx, \end{aligned}$$

6.1 Momenti di variabili unidimensionali

di modo che la (6.6) si identifica con la (6.5). Quest'ultima mostra che il valore medio rappresenta la differenza tra le seguenti due aree: (i) l'area delimitata dall'asse y , dalla retta $y = 1$ e dalla curva $y = F(x)$ nell'intervallo $(0, +\infty)$; (ii) l'area delimitata dall'asse x , dall'asse y e dalla curva $y = F(x)$ nell'intervallo $(-\infty, 0)$. In Figura 6.1 queste aree sono state indicate a mezzo di tratteggio insieme con i segni con cui vanno algebricamente sommate.

L'esempio seguente mostra il caso di una v.c. che non possiede valor medio.

Esempio 6.1. Consideriamo la funzione

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad \forall x \in \mathbb{R} \quad (6.7)$$

e mostriamo che essa rappresenta una funzione di densità di probabilità. Invero, per ogni $x \in \mathbb{R}$ risulta $f(x) > 0$; inoltre

$$\frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{1}{(1+x^2)} dx = \frac{1}{\pi} \left(\lim_{a \rightarrow -\infty} \arctan x|_a^0 + \lim_{b \rightarrow +\infty} \arctan x|_0^b \right) = \frac{1}{\pi} \left(\frac{\pi}{2} + \frac{\pi}{2} \right) = 1,$$

così che anche la condizione di normalizzazione è soddisfatta. Poiché:

$$f'(x) = -\frac{2x}{\pi(1+x^2)^2}, \quad f''(x) = \frac{2}{\pi} \frac{3x^2 - 1}{(1+x^2)^3},$$

il grafico della funzione f è simmetrico rispetto all'asse y e possiede due punti di flesso in $x = \pm 1/\sqrt{3}$. La v.c. X che ammette la f definita nella (6.7) come funzione di densità di probabilità è detta *variabile di Cauchy*.

Verifichiamo che X non è dotata di valore medio. Invero risulta:

$$\lim_{b \rightarrow +\infty} \int_0^b \frac{x}{1+x^2} dx = \frac{1}{2} \lim_{b \rightarrow +\infty} [\ln(1+x^2)]_0^b = +\infty$$

e

$$\lim_{a \rightarrow -\infty} \int_a^0 \frac{x}{1+x^2} dx = \frac{1}{2} \lim_{a \rightarrow -\infty} [\ln(1+x^2)]_a^0 = -\infty.$$

Notiamo, incidentalmente, che si ha:

$$\lim_{p \rightarrow +\infty} \frac{1}{\pi} \int_{-p}^{+p} \frac{x}{(1+x^2)} dx = 0;$$

esiste quindi l'integrale a valor principale secondo Cauchy della densità (6.7). Ciò, tuttavia, non comporta l'esistenza del valore medio secondo la Definizione 6.1. \square

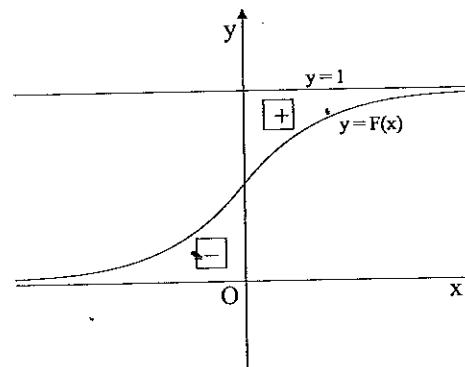


Figura 6.1 – Valore medio e funzione di distribuzione.

Oltre alla precedente interpretazione geometrica come differenze di aree, il valore medio di una v.c. è suscettibile di un'altra suggestiva interpretazione. Nel caso di una v.c. X assolutamente continua l'asse delle ascisse può infatti essere riguardato come un ideale filo rigido senza peso lungo alcuni dei suoi 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 di X , il centro di massa del filo risulta essere proprio $E(X)$; quindi il filo resta in equilibrio se viene sospeso nel punto di ascissa $E(X)$. Analogamente, nel caso in cui X è discreta l'asse delle ascisse può essere considerato ancora come un filo rigido senza peso lungo cui la massa unitaria è concentrata nei punti x_i ($i = 1, 2, \dots$) a ciascuno dei quali compete la massa $p(x_i)$, dove $\{p(x_i)\}$ è la distribuzione di probabilità di X . Anche in questo caso il filo rimane in equilibrio se sospeso nel punto di ascissa $E(X)$, che assume ancora il significato di centro di massa del filo.

L'analogia meccanica tra valore medio e centro di massa aiuta ad intendere, su base intuitiva, come possa accadere che a variazioni anche minime della distribuzione di probabilità (o della densità di probabilità) possano corrispondere notevoli variazioni del valore medio.

Per comprendere l'origine della denominazione "valore medio" della quantità sopra introdotta prendiamo in esame il seguente esempio. Consideriamo l'esperimento che consiste nel lanciare ripetutamente un dado, e supponiamo che in n lanci escano i numeri da 1 a 6 rispettivamente n_1, n_2, \dots, n_6 volte, con $n_1 + n_2 + \dots + n_6 = n$. Il rapporto

$$M = \frac{n_1 \cdot 1 + n_2 \cdot 2 + \dots + n_6 \cdot 6}{n} \quad (6.8)$$

6.1 Momenti di variabili unidimensionali

esprime allora la media aritmetica dei risultati ottenuti. D'altra parte si ha:

$$M = \frac{n_1}{n} \cdot 1 + \frac{n_2}{n} \cdot 2 + \dots + \frac{n_6}{n} \cdot 6 \approx p_1 \cdot 1 + p_2 \cdot 2 + \dots + p_6 \cdot 6$$

in quanto se n è molto grande per la legge empirica del caso la frequenza n_i/n dà approssimativamente la probabilità p_i di uscita del numero i ($i = 1, \dots, 6$) in ogni prova singolarmente considerata. La (6.8) esprime quindi il *valore di aspettazione* della v.c. discreta associata all'esperimento. La Definizione 6.1 si presenta dunque come un'estensione, nella teoria della probabilità, del concetto di media aritmetica. La media aritmetica possiede una proprietà importante che risulterà poi essere condivisa dal valore medio. Invero sussiste la seguente proposizione:

Proposizione 6.1. *Siano dati n reali x_1, x_2, \dots, x_n , e sia*

$$f(c) = \sum_{i=1}^n (x_i - c)^2, \quad (6.9)$$

con $c \in \mathbb{R}$. Il valore \bar{c} che minimizza $f(c)$ è la media aritmetica degli x_i .

Dimostrazione. Per ottenere i punti stazionari occorre risolvere l'equazione:

$$f'(c) = -2 \sum_{i=1}^n (x_i - c) = -2 \left(\sum_{i=1}^n x_i - nc \right) = 0.$$

Si ottiene così:

$$\bar{c} = \frac{1}{n} \sum_{i=1}^n x_i.$$

D'altra parte è $f''(c) = 2n > 0$ per ogni $c \in \mathbb{R}$; pertanto \bar{c} è un punto di minimo. Ciò, d'altronde, segue direttamente dal fatto che f è una somma di quadrati. \square

Si noti che se i reali x_1, x_2, \dots, x_n di cui alla Proposizione 6.1 non sono tutti distinti, la (6.9) assume la seguente forma:

$$f(c) = \sum_{r=1}^s k_r (y_r - c)^2, \quad (6.10)$$

dove con y_1, y_2, \dots, y_s si sono indicati i reali distinti e con k_1, k_2, \dots, k_s le rispettive frequenze assolute. Il valore \bar{c} che minimizza la (6.10) è allora esprimibile come

$$\bar{c} = \frac{\sum_{r=1}^s k_r y_r}{\sum_{r=1}^s k_r},$$

ovvero:

$$\bar{c} = \sum_{r=1}^s f_r y_r,$$

dove f_r ($r = 1, 2, \dots, s$) denota la frequenza di occorrenza del reale y_r nell'insieme degli n reali x_1, x_2, \dots, x_n considerati:

$$f_r = \frac{k_r}{\sum_{r=1}^s k_r} \equiv \frac{k_r}{n} \quad (r = 1, 2, \dots, s).$$

Va subito detto che il valore medio non è sempre significativo. È intuitivo che affinché esso sia in qualche modo rappresentativo della v.c. non devono avversi grandi discrepanze tra i valori assunti da questa con alta probabilità e il suo valore medio. È anche utile rilevare esplicitamente che il valore medio non è necessariamente uno dei valori assunti dalla variabile casuale.

Consideriamo ora il caso di una v.c. Y funzione di una v.c. X .

Definizione 6.2. Siano X una v.c. di funzione di distribuzione $F_X(x)$ e $g(x): \mathbb{R} \rightarrow \mathbb{R}$ una funzione Borel-misurabile. Se assolutamente convergente, l'integrale

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x) dF_X(x) \quad (6.11)$$

prende il nome di valore medio della v.c. $Y = g(X)$. \square

In base alla Definizione 6.2 per poter calcolare il valore medio di una funzione di una v.c. basta conoscere la funzione di distribuzione di quest'ultima. In effetti la (6.11) si può dimostrare, come ora vedremo limitatamente a due casi particolarmente semplici.

(a) Supponiamo che g sia strettamente monotona e che X sia discreta, di modo che anche Y è discreta. Risulta allora:

$$\begin{aligned} E[g(X)] &\equiv E(Y) = \int_{-\infty}^{+\infty} y dF_Y(y) = \sum_y y p_Y(y) = \sum_y y \mathbb{P}(g(X) = y) \\ &= \sum_y y \mathbb{P}(X = g^{-1}(y)) = \sum_{x=g^{-1}(y)} g(x) p_X(x) \\ &= \int_{-\infty}^{+\infty} g(x) dF_X(x). \end{aligned}$$

(b) Supponiamo che $g: \mathbb{R} \rightarrow (a, b) \subseteq \mathbb{R}$ sia strettamente monotona e che X sia assolutamente continua. Ricordando il Teorema 5.8 si ha:

$$\int_{-\infty}^{+\infty} y dF_Y(y) = \int_{-\infty}^{+\infty} y f_Y(y) dy = \int_a^b y \frac{f_X(g^{-1}(y))}{\left| \frac{dg}{dx} \right|_{x=g^{-1}(y)}} dy.$$

6.1 Momenti di variabili unidimensionali

Effettuando la sostituzione $y = g(x)$ si ricava allora

$$\begin{aligned} E[g(X)] &\equiv E(Y) = \int_{-\infty}^{+\infty} y dF_Y(y) = \int_{-\infty}^{+\infty} g(x) f_X(x) dx \\ &= \int_{-\infty}^{+\infty} g(x) dF_X(x), \end{aligned}$$

e quindi anche in questo caso si è ottenuta la (6.11).

Esaminiamo alcune proprietà del valore medio di una variabile casuale.

Proposizione 6.2. Siano X una v.c. e g una funzione Borel-misurabile. Sussiste quanto segue:

- (i) se X coincide con una costante c , si ha $E(c) = c$;
- (ii) $E[cg(X)] = c E[g(X)]$ per ogni costante c ;
- (iii) $E(aX + b) = a E(X) + b$ per ogni coppia a e b di costanti.

Dimostrazione. Le (i) e (ii) si ottengono immediatamente dalle Definizioni 7 e 8. Infatti risulta:

$$E(c) = \int_{-\infty}^{+\infty} c dF_X(x) = c \int_{-\infty}^{+\infty} dF_X(x) = c ;$$

$$E[cg(X)] = \int_{-\infty}^{+\infty} c g(x) dF_X(x) = c E[g(X)].$$

La (iii) segue infine dalle (i) e (ii) e dalla proprietà additiva dell'integrale di Stieltjes. \square

Proposizione 6.3. Siano X una v.c. e $a, b \in \mathbb{R}$ costanti. Risulta:

$$E[(aX + b)^n] = \sum_{i=0}^n \binom{n}{i} b^i a^{n-i} E(X^{n-i}). \quad (6.12)$$

Dimostrazione. Utilizzando la formula di Newton per lo sviluppo della potenza n -esima del binomio e facendo uso della Proposizione 6.2 si ottiene:

$$E[(aX + b)^n] = E \left[\sum_{i=0}^n \binom{n}{i} b^i a^{n-i} X^{n-i} \right] = \sum_{i=0}^n \binom{n}{i} b^i a^{n-i} E[X^{n-i}]. \quad \square$$

Due costanti di particolare rilievo associate ad una v.c. X sono quelle di cui alle definizioni che seguono.

Definizione 6.3. Data una v.c. X di funzione di distribuzione F e dotata di valore medio finito, se assolutamente convergente l'integrale

$$D^2(X) := E\{[X - E(X)]^2\} = \int_{-\infty}^{+\infty} [x - E(X)]^2 dF(x), \quad (6.13)$$

prende anche il nome di varianza di X . \triangleleft

Definizione 6.4. Si definisce deviazione standard di X la radice quadrata positiva $D(X)$ della varianza. \triangleleft

Notiamo che per la varianza non sussiste una proprietà di linearità quale quella espressa dalla (iii) della Proposizione 6.2 per il valor medio. Calcolando infatti la varianza della v.c. $Y = aX + b$, con $a, b \in \mathbb{R}$, si ha:

$$\begin{aligned} D^2(Y) &= E\{[Y - E(Y)]^2\} = E\{[aX + b - aE(X) - b]^2\} \\ &= a^2 E\{[X - E(X)]^2\} = a^2 D^2(X) \end{aligned}$$

e quindi anche:

$$D(Y) = |a| D(X).$$

Altre quantità che possono associarsi ad una v.c. X sono i cosiddetti *momenti*. Essi si ottengono considerando le vv.cc. $Y_r = X^r$ ($r = 1, 2, \dots$). I valori medi delle vv.cc. così ottenute al variare di r rappresentano i *momenti intorno all'origine* di X .

Definizione 6.5. Sia X una v.c. di funzione di distribuzione $F_X(x)$.

(i) Se assolutamente convergente, l'integrale

$$\mu'_r \equiv E(X^r) = \int_{-\infty}^{+\infty} x^r dF_X(x) \quad (r = 1, 2, \dots) \quad (6.14)$$

prende il nome di momento di ordine r intorno all'origine.

(ii) Si definisce momento centrale di ordine r , o momento intorno alla media di ordine r , della v.c. X dotata di valore medio finito l'integrale

$$\mu_r \equiv E\{[X - E(X)]^r\} = \int_{-\infty}^{+\infty} [x - E(X)]^r dF_X(x) \quad (r = 0, 1, \dots), \quad (6.15)$$

se assolutamente convergente. \triangleleft

D'ora innanzi, quando conveniente, in luogo di "momento intorno all'origine" e di "momenti intorno all'origine" parleremo più semplicemente di "momento" e di "momenti", mentre manteremo inalterate le locuzioni "momento centrale" e "momenti centrali".

Si noti che nel definire i momenti intorno all'origine il valore $r = 0$ non è stato preso in considerazione in quanto il "momento di ordine 0" esprimerebbe semplicemente la condizione di normalizzazione. Osserviamo anche che per $r = 1$ si ottiene il valore medio:

$$\mu'_1 = E(X).$$

Se, com'è consuetudine, si pone $\mu'_1 = \mu$, la (6.15) può anche riscriversi nella forma

$$\mu_r = E[(X - \mu)^r].$$

Per quanto riguarda poi i momenti centrali, si ha in particolare $\mu_0 = 1$ in quanto la (6.15) per $r = 0$ dà la condizione di normalizzazione. Per $r = 1$ dalla (6.15) segue:

$$\mu_1 = E(X - \mu) = E(X) - \mu = 0;$$

per $r = 2$ si ha invece:

$$\mu_2 = D^2(X).$$

Ribadiamo che l'esistenza dei momenti sia centrali che intorno all'origine richiede l'assoluta convergenza degli integrali coinvolti.

Proposizione 6.4. *Data una v.c. X dotata di momenti intorno all'origine finiti, si ha:*

$$\mu_r = \mu'_r - \binom{r}{1} \mu'_{r-1} \mu + \cdots + (-1)^k \binom{r}{k} \mu'_{r-k} \mu^k + \cdots + (-1)^{r-1} \mu^r (r-1). \quad (6.16)$$

Dimostrazione. Basta porre $a = 1$, $b = -\mu$, $n = r$ nella (6.12) per avere:

$$\mu_r \equiv E[(X - \mu)^r] = \sum_{k=0}^r \binom{r}{k} (-\mu)^k E(X^{r-k}) = \sum_{k=0}^r (-1)^k \binom{r}{k} \mu^k \mu'_{r-k}.$$

D'altra parte risulta:

$$\begin{aligned} (-1)^{r-1} \binom{r}{r-1} \mu^{r-1} \mu'_1 + (-1)^r \binom{r}{r} \mu^r \mu'_0 &\equiv (-1)^{r-1} r \mu^r + (-1)^r \mu^r \\ &= (-1)^{r-1} (r-1) \mu^r, \end{aligned}$$

in quanto $\mu'_1 = \mu$ e $\mu'_0 = 1$. Segue pertanto la (6.16). \square

In particolare dalla Proposizione 6.4 per $r = 2$ si trae:

$$\mu_2 = \mu'_2 - \mu^2,$$

ossia:

$$D^2(X) = E(X^2) - [E(X)]^2. \quad (6.17)$$

I momenti centrali rivestono grande rilevanza in quanto forniscono utili informazioni sulla variabile casuale. In particolare, alla varianza è legata una misura della dispersione intorno alla media dei valori assunti dalla v.c. Sussiste in particolare il seguente teorema:

Teorema 6.1. *Una variabile casuale X a valor medio μ e varianza nulla assume il valore μ con probabilità 1, ossia risulta:*

$$\mathbb{P}(|X - \mu| > \epsilon) = 0, \quad \forall \epsilon > 0.$$

Dimostrazione. Supponiamo per assurdo che la tesi non sia vera, ossia che esista un $\epsilon > 0$ tale da aversi $\mathbb{P}(|X - \mu| > \epsilon) > 0$. Dovrebbe allora risultare:

$$\begin{aligned} D^2(X) &= \int_{-\infty}^{+\infty} (x - \mu)^2 dF_X(x) \\ &= \int_{|x-\mu|>\epsilon} (x - \mu)^2 dF_X(x) + \int_{|x-\mu|\leq\epsilon} (x - \mu)^2 dF_X(x) \\ &\geq \int_{|x-\mu|>\epsilon} (x - \mu)^2 dF_X(x) > \epsilon^2 \int_{|x-\mu|>\epsilon} dF_X(x) \\ &= \epsilon^2 \mathbb{P}(|X - \mu| > \epsilon) > 0, \end{aligned}$$

contro l'ipotesi $D^2(X) = 0$. Deve dunque essere $\mathbb{P}(|X - \mu| > \epsilon) = 0$ per ogni $\epsilon > 0$; ciò, per l'arbitrarietà di ϵ , implica che X assume il valore μ con probabilità 1. \square

Osserviamo che nel caso di una v.c. discreta X che rappresenti i risultati di un esperimento casuale consistente in n prove ripetute, la varianza empirica per n molto grande è approssimativamente esprimibile al seguente modo:

$$D^2(X) \equiv \sum_{i=1}^s (x_i - \mu)^2 p(x_i) \approx \sum_{i=1}^s (x_i - \mu)^2 f_i = \frac{1}{n} \sum_{i=1}^s n_i (x_i - \mu)^2 \quad (6.18)$$

dove x_1, x_2, \dots, x_s sono i possibili risultati in ogni prova e dove $f_i = n_i/n$ è la frequenza di occorrenza del valore x_i nelle n prove. La (6.18) è analoga alla (6.10). Si può allora affermare che la media aritmetica μ minimizza la varianza empirica; è questo uno dei motivi per i quali si preferisce dare la definizione di valore medio generalizzando proprio il concetto di media aritmetica e non quello di altri tipi di medie.

I momenti sono grandezze significative per la caratterizzazione di una variabile casuale. Media e varianza assumono particolare importanza anche nel confronto di coppie di variabili casuali. Per un confronto esauriente dovrebbero essere prese in esame le funzioni di

distribuzione delle due variabili; se, però, queste non sono note, si possono spesso ottenere significative informazioni ponendo a confronto i relativi momenti. Tuttavia, in generale, le v.v.c. possono essere di natura diversa, cosicché esse vanno preliminarmente standardizzate. Una variabile casuale può essere standardizzata se sono note e finite la sua media e la sua varianza, quest'ultima essendo inoltre non nulla. Precisamente, si dà la seguente definizione:

Definizione 6.6. Se X è una v.c. di valore medio $E(X)$ finito e deviazione standard $D(X)$ finita e non nulla, la v.c.

$$Y = \frac{X - E(X)}{D(X)}$$

si dice essere la standardizzata, o normalizzata, di X . \square

L'origine di questa definizione è legata alle evidenti proprietà

$$E(Y) = 0, \quad D(Y) = 1$$

che caratterizzano una tale variabile casuale. Due v.v.c. X_1 e X_2 di cui siano note e finite medie e varianze, essendo inoltre queste non nulle, possono dunque sempre standardizzarsi; si ottengono così due v.v.c. Y_1 e Y_2 entrambe a valor medio nullo e varianza unitaria le quali possono essere messe a confronto. In generale non si può dire che Y_1 e Y_2 hanno la stessa distribuzione di probabilità: ciò in quanto media e varianza di una v.c. di solito non bastano per individuare univocamente la funzione di distribuzione della variabile casuale. Un'importante eccezione è costituita dalle variabili casuali gaussiane, cui ora accenneremo.

6.1.1 La variabile gaussiana

Definizione 6.7. Si definisce normale, o gaussiana, una v.c. X assolutamente continua di densità di probabilità

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right], \quad \forall x \in \mathbb{R}, \quad (6.19)$$

dove $m \in \mathbb{R}$ e $\sigma > 0$ sono costanti arbitrarie. \square

La densità di probabilità f definita nella (6.19), ha per supporto l'asse reale ed è simmetrica rispetto alla retta $x = m$. Ciò implica che m è il valore medio della variabile casuale.¹ Il valore medio può peraltro essere calcolato direttamente dalla Definizione 6.3. Sia per il calcolo del valore medio che della varianza occorre fare uso della nota formula

$$\int_{-\infty}^{+\infty} e^{-z^2} dz = \sqrt{\pi}, \quad (6.20)$$

¹Si noti che per la forma (6.19) della densità di probabilità il valor medio esiste.

della quale diamo qui una rapida dimostrazione. A tal fine notiamo che si ha:

$$\begin{aligned} \left(\int_{-\infty}^{+\infty} e^{-z^2} dz \right)^2 &= \int_{-\infty}^{+\infty} e^{-x^2} dx \int_{-\infty}^{+\infty} e^{-y^2} dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-(x^2+y^2)} dxdy. \end{aligned}$$

Effettuando la trasformazione di coordinate $x = \rho \cos \vartheta$, $y = \rho \sin \vartheta$ e ricordando che lo jacobiano di tale trasformazione vale ρ , si ottiene:

$$\left(\int_{-\infty}^{+\infty} e^{-z^2} dz \right)^2 = \int_0^{2\pi} d\vartheta \int_0^{+\infty} \rho e^{-\rho^2} d\rho = \int_0^{2\pi} d\vartheta \left[-\frac{e^{-\rho^2}}{2} \right]_0^{+\infty} = \pi,$$

il che dimostra la (6.20). Calcoliamo ora, ricorrendo alle rispettive definizioni, media e varianza della v.c. gaussiana. Si ha:

$$E(X) \equiv \int_{-\infty}^{+\infty} x f(x) dx = \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(x-m)^2}{2\sigma^2} \right\} dx.$$

Posto $z = (x-m)/(\sqrt{2}\sigma)$, segue $x = \sqrt{2}\sigma z + m$ e $dx = \sqrt{2}\sigma dz$. Quindi,

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} \frac{\sqrt{2}\sigma}{\sqrt{2\pi}\sigma} (\sqrt{2}\sigma z + m) e^{-z^2} dz \\ &= \frac{\sqrt{2}\sigma}{\sqrt{\pi}} \int_{-\infty}^{+\infty} ze^{-z^2} dz + \frac{m}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-z^2} dz = m, \end{aligned}$$

dove si è fatto uso della (6.20) e della circostanza che ze^{-z^2} è una funzione dispari, cosicché il suo integrale esteso ad \mathbb{R} è nullo. Si ha poi:

$$D^2(X) = \int_{-\infty}^{+\infty} (x-m)^2 \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(x-m)^2}{2\sigma^2} \right\} dx. \quad (6.21)$$

Effettuando lo stesso cambiamento di coordinate utilizzato per il calcolo del valore medio, dalla (6.21) si trae:

$$\begin{aligned} D^2(X) &\equiv \int_{-\infty}^{+\infty} (x-m)^2 f(x) dx \\ &= \int_{-\infty}^{+\infty} 2\sigma^2 z^2 \frac{\sqrt{2}\sigma}{\sqrt{2\pi}\sigma} e^{-z^2} dz = \frac{\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{+\infty} z \left(2ze^{-z^2} \right) dz \\ &= \frac{\sigma^2}{\sqrt{\pi}} \left\{ \left[-ze^{-z^2} \right]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-z^2} dz \right\} \\ &= \sigma^2. \end{aligned}$$

Se la v.c. X è gaussiana, una volta assegnate le costanti m e σ^2 resta univocamente specificata la sua densità di probabilità, ed è quindi anche ricavabile la sua funzione di distribuzione. Ma poiché m e σ^2 sono media e varianza di X , si conclude che media e varianza sono sufficienti a determinare univocamente la funzione di distribuzione di una v.c. normale.

Uno studio del grafico della densità di probabilità permette di trarre altre conclusioni interessanti sulla v.c. gaussiana. Mostriamo anzitutto che tale densità possiede due punti di flesso. Dalla (6.19) segue infatti:

$$\begin{aligned} f'(x) &= -\frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(x-m)^2}{2\sigma^2} \right\} \frac{x-m}{\sigma^2} \\ &= -\frac{x-m}{\sqrt{2\pi}\sigma^3} \exp \left\{ -\frac{(x-m)^2}{2\sigma^2} \right\}, \end{aligned}$$

avendosi inoltre:

$$\begin{aligned} f''(x) &= \frac{1}{\sqrt{2\pi}\sigma^3} \exp \left\{ -\frac{(x-m)^2}{2\sigma^2} \right\} + \frac{(x-m)^2}{\sqrt{2\pi}\sigma^5} \exp \left\{ -\frac{(x-m)^2}{2\sigma^2} \right\} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma^5} \exp \left\{ -\frac{(x-m)^2}{2\sigma^2} \right\} \right) [(x-m)^2 - \sigma^2]. \end{aligned}$$

Quindi f possiede due punti di flesso di ascisse $x_1 = m - \sigma$ e $x_2 = m + \sigma$. La distanza di tali punti è dunque 2σ , così che possiamo affermare che la deviazione standard $\sigma = D(X)$ può riguardarsi come una misura della "larghezza" della curva. Il massimo esibito dal grafico della densità gaussiana è tanto più pronunciato quanto più piccolo è σ . Dato che l'area sottesa alla curva vale sempre 1, si comprende come che nel caso di valori piccoli di σ il valore medio è significativo, mentre per valori grandi di σ il valore medio non lo è particolarmente.

La v.c. gaussiana a media nulla e varianza unitaria viene detta *gaussiana standard*. È interessante osservare, come è possibile dimostrare utilizzando la formula (5.63), che la densità di probabilità della v.c. definita come rapporto di due gaussiane standard è la variabile di Cauchy cui si è accennato nell'Esempio 6.7.

6.1.2 La variabile esponenziale

Come si è detto, la conoscenza di media e varianza consente di determinare univocamente la funzione di distribuzione di una v.c. gaussiana. Vi sono però altri tipi di vv.cc. individuate dal solo valore medio. Consideriamo, ad esempio, una v.c. X esponenziale (cfr. il § A.2.2), ossia una v.c. di densità di probabilità

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{per } x > 0, \\ 0, & \text{altrove,} \end{cases} \quad (6.22)$$

con λ costante positiva. Si osservi che la funzione f così definita è effettivamente una densità di probabilità; si ha infatti $f(x) \geq 0$ per ogni $x \in \mathbb{R}$ ed inoltre risulta:

$$\int_{-\infty}^{+\infty} f(x) dx = \int_0^{+\infty} \lambda e^{-\lambda x} dx = 1.$$

Calcoliamo il valore medio di X :

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \int_0^{+\infty} \lambda x e^{-\lambda x} dx = \frac{1}{\lambda}.$$

Per individuare univocamente la densità di probabilità di X basta dunque assegnare il parametro λ che è l'inverso del valore medio di X . La conoscenza del solo valore medio consente in questo caso di specificare univocamente la densità di probabilità di una v.c. di tipo esponenziale.

Concludiamo questo paragrafo menzionando che è possibile estendere la definizione di valore medio di una funzione di variabile casuale al caso di funzioni complesse di variabile reale. Si dà infatti la seguente definizione:

Definizione 6.8. Siano X una v.c. di funzione di distribuzione F_X e g una funzione complessa di variabile reale con parte reale e parte immaginaria funzioni Borel-misurabili. Si definisce valore medio di $g(X)$ il numero complesso avente per parte reale e per parte immaginaria, rispettivamente, la media della parte reale reale e della parte immaginaria di g :

$$\begin{aligned} E[g(X)] &= \int_{-\infty}^{+\infty} \Re g(x) dF_X(x) + i \int_{-\infty}^{+\infty} \Im g(x) dF_X(x) \\ &= E[\Re g(X)] + i E[\Im g(X)]. \end{aligned} \quad (6.23)$$

△

Diremo poi che $E[g(X)]$ esiste finito se è finito $E[|g(X)|]$, ossia il valore medio del modulo di X . È infine possibile mostrare che la Definizione 6.8 estende alle funzioni complesse di variabile reale le usuali proprietà dell'operazione di media.

6.2 Momenti di vettori casuali

Siano X_1, \dots, X_n vv.cc. congiuntamente distribuite, definite su uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$. Come sappiamo, resta così definito un vettore casuale n -dimensionale X le cui componenti sono le vv.cc. X_1, \dots, X_n . Nelle considerazioni che seguono supporremo che queste siano assolutamente continue,² con densità di probabilità congiunta $f(x_1, \dots, x_n)$. In

6.2 Momenti di vettori casuali

tal caso il vettore X si dice assolutamente continuo. Sia inoltre $g: \mathbb{R}^n \rightarrow \mathbb{R}$ una funzione Borel-misurabile e sia $Y = g(X_1, \dots, X_n)$ una v.c. funzione di X_1, \dots, X_n .

Passiamo a dare delle definizioni alcune delle quali generalizzano le corrispondenti precedentemente date per variabili casuali unidimensionali.

Definizione 6.9. Si definisce:

(i) valore medio (o media, valore di aspettazione, valore atteso, speranza matematica) della v.c. X_k l'integrale

$$E(X_k) = \int_{-\infty}^{+\infty} dx_1 \dots \int_{-\infty}^{+\infty} dx_n x_k f(x_1, \dots, x_n), \quad (6.24)$$

sempre che esso sia assolutamente convergente;

(ii) valore medio (o media, valore di aspettazione, valore atteso, speranza matematica) della v.c. $Y = g(X_1, \dots, X_n)$ l'integrale

$$\begin{aligned} E(Y) &= \int_{-\infty}^{+\infty} y f_Y(y) dy \\ &= \int_{-\infty}^{+\infty} dx_1 \dots \int_{-\infty}^{+\infty} dx_n g(x_1, \dots, x_n) f(x_1, \dots, x_n), \end{aligned} \quad (6.25)$$

se assolutamente convergente. □

Ricordiamo che la densità di probabilità di una v.c. unidimensionale X si può ottenere come densità marginale di una v.c. n -dimensionale. Posto, ad esempio, $X_1 = X$, dalla (5.40) si ha:

$$f_X(x) = \int_{-\infty}^{+\infty} dx_2 \int_{-\infty}^{+\infty} dx_3 \dots \int_{-\infty}^{+\infty} dx_n f(x_1, \dots, x_n).$$

Dalla (6.24) si ricava allora:

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx,$$

che coincide con la (6.3).

Facendo uso della (6.25) è possibile dimostrare facilmente la seguente proposizione:

Proposizione 6.5. L'operatore E è lineare.

Dimostrazione. Nel § 6.1 si è già visto che risulta $E(cX) = cE(X)$ per ogni $c \in \mathbb{R}$ (cfr. la Proposizione 6.2); resta da mostrare che si ha:

$$E(X_1 + X_2) = E(X_1) + E(X_2)$$

²Ad analoghi risultati si perviene anche nel caso discreto. Per semplicità di notazione qui ci limiteremo però ad esaminare il solo caso di vv.cc. assolutamente continue.

per ogni coppia di vv.cc. X_1, X_2 dotate di valore medio. A tal fine, supponendo che X_1 e X_2 siano assolutamente continue con densità di probabilità congiunta $f(x_1, x_2)$ e posto $Z = X_1 + X_2$, dalle (6.24) e (6.25) segue:

$$\begin{aligned} E(Z) &\equiv E(X_1 + X_2) = \int_{-\infty}^{+\infty} z f_Z(z) dz \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x_1 + x_2) f(x_1, x_2) dx_1 dx_2 \\ &= \int_{-\infty}^{+\infty} x_1 \left[\int_{-\infty}^{+\infty} f(x_1, x_2) dx_2 \right] dx_1 + \int_{-\infty}^{+\infty} x_2 \left[\int_{-\infty}^{+\infty} f(x_1, x_2) dx_1 \right] dx_2 \\ &= E(X_1) + E(X_2). \end{aligned}$$

Nel caso di vv.cc. discrete si perviene ad un risultato analogo usando la distribuzione di probabilità congiunta. \square

Definizione 6.10. Sia r un intero positivo:

(i) se assolutamente convergente, l'integrale

$$E[(X_k)^r] = \int_{-\infty}^{+\infty} dx_1 \dots \int_{-\infty}^{+\infty} dx_n x_k^r f(x_1, \dots, x_n) \quad (6.26)$$

è detto momento intorno all'origine di ordine r della variabile casuale X_k ;

(ii) se X_k possiede valore medio finito, l'integrale

$$E[(X_k - E(X_k))^r] = \int_{-\infty}^{+\infty} dx_1 \dots \int_{-\infty}^{+\infty} dx_n [x_k - E(X_k)]^r f(x_1, \dots, x_n), \quad (6.27)$$

se assolutamente convergente, è detto momento centrale di ordine r della variabile casuale X_k ;

(iii) si dice momento assoluto intorno all'origine di ordine r della v.c. X_k il momento intorno all'origine di ordine r della variabile $Y_k = |X_k|$, se esistente:

$$E(|X_k|^r) = \int_{-\infty}^{+\infty} dx_1 \dots \int_{-\infty}^{+\infty} dx_n |x_k|^r f(x_1, \dots, x_n). \quad (6.28)$$

\square

Dalle (6.26) e (6.28) discende che il momento intorno all'origine di ordine r di una v.c. esiste finito se e solo se esiste finito il suo momento assoluto di ordine r . Con lo stesso tipo di ragionamento usato in precedenza si può poi mostrare che le definizioni (6.14) e (6.15) per vv.cc. unidimensionali coincidono rispettivamente con le (6.26) e (6.27).

6.2 Momenti di vettori casuali

Il momento assoluto intorno all'origine di ordine r di una v.c. X verrà nel seguito denotato al seguente modo:

$$\beta_r \equiv E(|X|^r) = \int_{-\infty}^{+\infty} |x|^r dF(x). \quad (6.29)$$

Definizione 6.11. Fissata una coppia di interi positivi r ed s ,

(i) se è assolutamente convergente, l'integrale

$$E[(X_k)^r (X_l)^s] = \int_{-\infty}^{+\infty} dx_1 \dots \int_{-\infty}^{+\infty} dx_n x_k^r x_l^s f(x_1, \dots, x_n) \quad (6.30)$$

dicesi momento prodotto di ordine r, s delle vv.cc. X_k e X_l ;

(ii) se le variabili casuali X_k e X_l sono dotate di valori medi finiti, dicesi momento centrale prodotto di ordine r, s l'integrale

$$\begin{aligned} E\{[(X_k - E(X_k))^r [X_l - E(X_l)]^s]\} &= \\ &= \int_{-\infty}^{+\infty} dx_1 \dots \int_{-\infty}^{+\infty} dx_n [x_k - E(X_k)]^r [x_l - E(X_l)]^s f(x_1, \dots, x_n), \end{aligned} \quad (6.31)$$

se assolutamente convergente. \square

Particolare rilievo riveste il caso in cui nella (6.31) si sceglie $r = 1, s = 1$. Si dà infatti la seguente definizione:

Definizione 6.12. La quantità

$$\text{cov}\{X_k, X_l\} = E\{[(X_k - E(X_k))(X_l - E(X_l))]\} \quad (6.32)$$

si dice covarianza delle vv.cc. X_k, X_l . \square

Per la linearità dell'operatore E , dalla (6.32) si trae:

$$\text{cov}\{X_k, X_l\} = E(X_k X_l) - E(X_k) E(X_l). \quad (6.33)$$

Infatti risulta:

$$\begin{aligned} \text{cov}\{X_k, X_l\} &= E[X_k X_l - E(X_k) X_l - X_k E(X_l) + E(X_k) E(X_l)] \\ &= E(X_k X_l) - E(X_k) E(X_l) - E(X_k) E(X_l) + E(X_k) E(X_l) \\ &= E(X_k X_l) - E(X_k) E(X_l), \end{aligned}$$

ossia l'asserto. In particolare, per $k = l$ si ha:

$$\text{cov}\{X_k, X_k\} = E(X_k^2) - [E(X_k)]^2 \equiv D^2(X_k). \quad (6.34)$$

Definizione 6.13. Si dice coefficiente di correlazione di due vv.cc. X_k e X_l non degeneri³ il rapporto

$$\rho(X_k, X_l) = \frac{\text{cov}\{X_k, X_l\}}{D(X_k) D(X_l)}. \quad (6.35)$$

□

Si noti che, a differenza della covarianza, il coefficiente di correlazione è sempre adimensionale.

6.3 Alcune proprietà dei momenti

In questo paragrafo esporremo alcune ulteriori proprietà dei momenti di una v.c. sottolineando talune utili interrelazioni.

Proposizione 6.6. Se esiste un intero positivo n tale che il momento assoluto β_n di ordine n è finito, allora esistono i momenti sia assoluti che intorno all'origine di ogni ordine r con $r \leq n$.

Dimostrazione. Per $r \leq n$ segue:

$$\begin{aligned} \beta_r &\equiv \int_{-\infty}^{+\infty} |x|^r dF(x) = \int_{|x| \leq 1} |x|^r dF(x) + \int_{|x| > 1} |x|^r dF(x) \\ &\leq \int_{|x| \leq 1} dF(x) + \int_{|x| > 1} |x|^n dF(x) \\ &\leq 1 + \beta_n. \end{aligned} \quad (6.36)$$

Poiché β_n è per ipotesi finito, dalla (6.36) discende che tale risulta anche β_r per ogni $r \leq n$. Di conseguenza, per la loro stessa definizione, sono finiti anche tutti i momenti intorno all'origine per ogni $r \leq n$. □

La Proposizione 6.6 può anche enunciarsi nella seguente forma:

Proposizione 6.7. Se una v.c. possiede finito un momento di un certo ordine, essa possiede finiti anche tutti i momenti di ordine inferiore.

Per poter determinare altre relazioni atte ad assicurare l'esistenza di momenti di un certo ordine, è conveniente fare uso di alcune diseguaglianze notevoli.

Proposizione 6.8. Se X e Y sono vv.cc. congiuntamente distribuite e se $E(X^2), E(Y^2)$ e $E(|XY|)$ esistono finiti, risulta:

$$[E(XY)]^2 \leq E(X^2)E(Y^2). \quad (6.37)$$

³Per non degenero si intende ogni variabile casuale a varianza non nulla.

6.3 Alcune proprietà dei momenti

Dimostrazione. Anzitutto osserviamo che, dalla linearità dell'operatore E , per ogni $a \in \mathbb{R}$ si ha:

$$E\{(a|X| - |Y|)^2\} \equiv a^2 E(X^2) - 2a E(|XY|) + E(Y^2) \geq 0. \quad (6.38)$$

La (6.38) può risciversi al seguente modo:

$$E(X^2) \left\{ \left[a - \frac{E|XY|}{E(X^2)} \right]^2 - \frac{\Delta}{E^2(X^2)} \right\} \geq 0, \quad (6.39)$$

dove Δ è il discriminante:

$$\Delta = [E(|XY|)]^2 - E(X^2)E(Y^2).$$

Perché la disequazione (6.39) sussista per ogni reale a , tale discriminante deve essere non positivo, ossia deve aversi:

$$\Delta = [E(|XY|)]^2 - E(X^2)E(Y^2) \leq 0;$$

di qui segue la (6.37). □

La (6.37) è nota come *diseguaglianza di Schwarz* per l'operatore E . Si noti come dalla supposta validità della (6.37) segue che se esistono finiti i momenti intorno all'origine del secondo ordine di X e Y , esiste finita anche $\text{cov}\{X, Y\}$.

Osservazione 6.1. Notiamo che per $r = 2$ la Proposizione 6.6 può essere dimostrata anche applicando la diseguaglianza di Schwarz alle vv.cc. $|X|$ e $Y \equiv 1$. Dire, infatti, che β_2 esiste finito equivale a dire che esiste finito il momento intorno all'origine del secondo ordine di X , così che dalla (6.37) segue:

$$[E(|X|)]^2 \leq E(|X|^2) = E(X^2).$$

Estraendo la radice quadrata di ambo i membri si trae infine:

$$E(X) \leq E(|X|) \leq \sqrt{E(X^2)}.$$

Pertanto esistono finiti sia il momento intorno all'origine $E(X)$ che il momento assoluto del primo ordine $E(|X|)$. □

Proposizione 6.9. L'operazione di media soddisfa la seguente diseguaglianza triangolare:

$$E(|X + Y|) \leq E(|X| + |Y|) = E(|X|) + E(|Y|). \quad (6.40)$$

Dimostrazione. Segue immediatamente usando la diseguaglianza

$$|X + Y| \leq |X| + |Y|$$

e la proprietà di linearità di E . □

Appena più riposta è la dimostrazione della seguente

Proposizione 6.10. *Se X e Y sono vvcc. dotate di momenti assoluti del secondo ordine risulta*

$$E(|X + Y|^2) \leq 2E(|X|^2) + 2E(|Y|^2). \quad (6.41)$$

Dimostrazione. Si ha:

$$\begin{aligned} |X + Y|^2 &= |X^2 + Y^2 + 2XY| \leq |X|^2 + |Y|^2 + 2|XY| \\ &= X^2 + Y^2 + 2|XY| - 2X^2 - 2Y^2 + 2X^2 + 2Y^2 \\ &= 2X^2 + 2Y^2 - (X^2 + Y^2 - 2|XY|) \\ &= 2X^2 + 2Y^2 - (|X| - |Y|)^2 \\ &\leq 2|X|^2 + 2|Y|^2. \end{aligned}$$

Passando ai valori di aspettazione si ottiene infine la (6.41). \square

Osserviamo che le (6.40) e (6.41) si possono riguardare come casi particolari, ottenuti rispettivamente per $r = 1$ e $r = 2$, della seguente proposizione:

Proposizione 6.11. *Se X e Y sono vvcc. dotate di momenti assoluti di ordine r , si ha:*

$$E(|X + Y|^r) \leq 2^{r-1} E(|X|^r) + 2^{r-1} E(|Y|^r). \quad (6.42)$$

Dimostrazione. Basta dimostrare che sussiste la seguente diseguaglianza:

$$|X + Y|^r \leq 2^{r-1} (|X|^r + |Y|^r). \quad (6.43)$$

Cominciamo col dimostrare che si ha:

$$(|X|^r + |Y|^r)(|X| + |Y|) \leq 2 (|X|^{r+1} + |Y|^{r+1}). \quad (6.44)$$

La (6.44) è verificata se e solo se risulta

$$|X|^{r+1} + |Y|^{r+1} - |X|^r|Y| - |X||Y|^r \geq 0,$$

ossia se e solo se

$$(|X|^r - |Y|^r)(|X| - |Y|) \geq 0. \quad (6.45)$$

D'altro canto la (6.45) è vera per ogni scelta di X e Y poiché

$$|X| \leq |Y| \Leftrightarrow |X|^r \leq |Y|^r.$$

6.3 Alcune proprietà dei momenti

La (6.44) è dunque vera. Dimostrare ora la (6.43) per induzione. Essa è vera per $r = 1$,⁴ in quanto coincide con la diseguaglianza triangolare di cui alla Proposizione 6.9. Supponiamo ora che sia vera per $r = k$ e dimostriamo che è vera per $r = k + 1$. Si ha infatti:

$$\begin{aligned} |X + Y|^{k+1} &= |X + Y|^k |X + Y| \leq 2^{k-1} (|X|^k + |Y|^k) (|X + Y|) \\ &\leq 2^{k-1} (|X|^k + |Y|^k) (|X| + |Y|) \leq 2^k (|X|^{k+1} + |Y|^{k+1}). \end{aligned}$$

La (6.43) è dunque verificata. Passando ai valori di aspettazione si ottiene la (6.42). \square

Facendo uso delle diseguaglianze ricavate si possono immediatamente dimostrare le due proposizioni che seguono.

Proposizione 6.12. *Se X e Y sono vvcc. congiuntamente distribuite e dotate di momenti assoluti del secondo ordine finiti, allora:*

- (i) *esiste finito il momento assoluto del secondo ordine, $E(|X + Y|^2)$, della v.c. $X + Y$;*
- (ii) *il momento prodotto $E(|XY|)$ è finito.*

Dimostrazione. La (i) è conseguenza immediata della (6.41); la (ii) segue invece dalla diseguaglianza di Schwarz. \square

Proposizione 6.13. *Se X e Y sono vvcc. congiuntamente distribuite dotate di momenti assoluti di ordine $n > 1$ finiti, esiste finito anche il momento assoluto $E(|X + Y|^n)$ della v.c. $X + Y$.*

Dimostrazione. La dimostrazione è diretta conseguenza della (6.42). \square

Corollario 6.1. *Siano X e Y vvcc. dotate di momenti intorno all'origine del secondo ordine finiti. Allora esistono finiti i momenti intorno all'origine del primo ordine, la varianza e la covarianza di X e Y .*

Dimostrazione. La supposta finitezza dei momenti del secondo ordine di X e Y implica che esistono finiti anche i momenti assoluti del secondo ordine di X e Y . Pertanto:

(a) $E(X)$ e $E(Y)$ esistono finiti in virtù della Proposizione 6.6.

(b) Poiché, per definizione, risulta

$$D^2(X) = E(X^2) - [E(X)]^2 \quad D^2(Y) = E(Y^2) - [E(Y)]^2,$$

possiamo affermare che le varianze di X e Y esistono finite in conseguenza dell'ipotesi e per quanto dimostrato in (a).

⁴Si noti che essa è anche vera per $r = 2$, come è stato dimostrato nella Proposizione 6.10.

(c) Facciamo uso della rappresentazione (6.33) della covarianza delle variabili casuali X e Y :

$$\text{cov}\{X, Y\} = E(XY) - E(X)E(Y).$$

Osservando che $E(XY)$ è finito per la (ii) della Proposizione 6.12 e che $E(X)$ e $E(Y)$ sono finiti per quanto dimostrato in (a), segue che anche la covarianza è finita. \square

6.4 Il caso delle variabili casuali indipendenti

Le vv.cc. indipendenti godono di alcune proprietà notevoli, come espresso dalla proposizioni che seguono.

Proposizione 6.14. *Se X_1, \dots, X_n sono vv.cc. indipendenti a valori medi $E(X_i)$ ($i = 1, \dots, n$) finiti, il valore medio del loro prodotto esiste ed è uguale al prodotto dei valori medi:*

$$E(X) \equiv E\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n E(X_i). \quad (6.46)$$

Dimostrazione. Per semplicità supponiamo che X_1, \dots, X_n siano assolutamente continue e denotiamo con $f_{X_1 \dots X_n}(x_1, \dots, x_n)$ la loro densità di probabilità congiunta; dalla (5.53) segue:

$$f_{X_1 \dots X_n}(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i).$$

Pertanto,

$$\begin{aligned} E(X) &\equiv E\left(\prod_{i=1}^n X_i\right) = \int_{-\infty}^{+\infty} dx_1 \dots \int_{-\infty}^{+\infty} \left(\prod_{i=1}^n x_i\right) f_{X_1 \dots X_n}(x_1, \dots, x_n) dx_n \\ &= \int_{-\infty}^{+\infty} dx_1 \dots \int_{-\infty}^{+\infty} \prod_{i=1}^n x_i f_{X_i}(x_i) dx_n \\ &= \int_{-\infty}^{+\infty} x_1 f_{X_1}(x_1) dx_1 \dots \int_{-\infty}^{+\infty} x_n f_{X_n}(x_n) dx_n \\ &= \prod_{i=1}^n E(X_i). \end{aligned}$$

In modo analogo si può procedere se le vv.cc. sono discrete. \square

Dalla Proposizione 6.18 e dalla definizione stessa di covarianza segue immediatamente la seguente proposizione:

6.4 Il caso delle variabili casuali indipendenti

Proposizione 6.15. *Due vv.cc. X, Y congiuntamente distribuite e indipendenti hanno covarianza nulla.*

Dimostrazione. Ricordiamo che per la (6.33) la covarianza di X e Y può esprimersi al seguente modo:

$$\text{cov}\{X, Y\} = E(XY) - E(X)E(Y). \quad (6.47)$$

In virtù della supposta indipendenza di X e Y si ha poi:

$$E(XY) = E(X)E(Y)$$

che, sostituita nella (6.47), dimostra l'asserto. \square

Notiamo che il viceversa non è in generale vero: l'annullarsi della covarianza di due vv.cc. X, Y non implica l'indipendenza di queste.

Esempio 6.2. Siano X e Y variabili casuali congiuntamente distribuite al seguente modo:

$$\mathbb{P}(X = -1; Y = 1) = \mathbb{P}(X = 0; Y = 0) = \mathbb{P}(X = 1; Y = 1) = \frac{1}{3}. \quad (6.48)$$

Si ha allora

$$\mathbb{P}(X = -1) = \mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \frac{1}{3} \quad (6.49)$$

così che risulta

$$E(X) = 0.$$

Analogamente, si ottiene:

$$\mathbb{P}(Y = -1) = \frac{1}{3}, \quad \mathbb{P}(Y = 1) = \frac{2}{3}, \quad (6.50)$$

e pertanto

$$E(Y) = \frac{2}{3}.$$

Infine, è facile mostrare che risulta

$$E(XY) = 0.$$

In conclusione, risultando

$$\text{cov}\{X, Y\} = E(XY) - E(X)E(Y) = 0,$$

le variabili casuali X e Y hanno covarianza nulla. Tuttavia esse non sono indipendenti, come segue dalle (6.48), (6.49) e (6.50). \triangleleft

Esempio 6.3. Un ulteriore esempio è fornito dalla variabile casuale bidimensionale (X, Y) avente distribuzione uniforme sull'insieme

$$T = \{(x, y) \in \mathbb{R}^2 : -y < x < y; 0 < y < 1\},$$

ossia avente densità di probabilità

$$f_{XY}(x, y) = \begin{cases} 1, & \text{per } y \in (0, 1) \text{ e } x \in (-y, y), \\ 0, & \text{altrimenti.} \end{cases}$$

Da questa seguono le densità marginali:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dy = \begin{cases} 1 - |x|, & \text{per } -1 < x < 1, \\ 0, & \text{altrimenti,} \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dx = \begin{cases} 2y, & \text{per } 0 < y < 1, \\ 0, & \text{altrimenti.} \end{cases}$$

Si vede poi facilmente che X e Y hanno covarianza nulla, avendosi

$$E(XY) = \int_0^1 dy \int_{-y}^y xy dx = 0,$$

$$E(X) = \int_{-1}^1 x(1 - |x|) dx = 0.$$

In virtù della (6.47) segue che le vv.cc. X e Y hanno covarianza nulla, pur non essendo indipendenti in quanto risulta

$$f_{XY}(x, y) \neq f_X(x)f_Y(y).$$

□

Le considerazioni svolte suggeriscono di dare la seguente definizione:

Definizione 6.14. Due vv.cc. X, Y congiuntamente distribuite si dicono scorrelate se la loro covarianza esiste ed è nulla. □

Dalla Definizione 6.14 e dalla (6.35) discende che se due vv.cc. sono scorrelate il loro coefficiente di correlazione $\rho\{X, Y\}$ è nullo. Inoltre, dalla Proposizione 6.19 e dalla Definizione 6.14 discende la seguente proposizione di immediata dimostrazione:

Proposizione 6.16. Due vv.cc. congiuntamente distribuite, indipendenti e dotate di momenti assoluti del secondo ordine sono scorrelate.

Notiamo infine che la Definizione 5.14 di indipendenza è equivalente alla (5.53) (o alla corrispondente formula per distribuzioni di probabilità), ma non alla (6.46); in realtà la nozione di indipendenza si può fornire anche tramite la seguente proposizione, che per semplicità enunciamo nel caso di due sole variabili casuali.

Proposizione 6.17. Le vv.cc. X e Y sono indipendenti se e solo se è soddisfatta una delle seguenti condizioni tra esse equivalenti:

$$(i) F_{XY}(x, y) = F_X(x)F_Y(y) \quad \forall x, y \in \mathbb{R};$$

$$(ii) f_{XY}(x, y) = f_X(x)f_Y(y) \quad \forall x, y \in \mathbb{R};$$

$$(iii) E[h(X)g(Y)] = E[h(X)]E[g(Y)],$$

con h e g funzioni Borel-misurabili arbitrarie per le quali esistono i valori di aspettazione indicati.

Dimostrazione. Basterà dimostrare la catena di implicazioni

$$(i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (i).$$

In virtù delle (5.50) e (5.51) si ricava che la (i) implica la (ii). Dalla definizione di valore medio, per la supposta validità della (ii) si ha poi:

$$\begin{aligned} E[h(X)g(Y)] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x)g(y)f_{XY}(x, y) dx dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x)g(y)f_X(x)f_Y(y) dx dy \\ &= \int_{-\infty}^{+\infty} h(x)f_X(x) dx \int_{-\infty}^{+\infty} g(y)f_Y(y) dy \\ &= E[h(X)]E[g(Y)], \end{aligned}$$

così che la (ii) implica la (iii). Dimostriamo infine che la (iii) implica la (i). Si ha invece:

$$\begin{aligned} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x)g(y)f_{XY}(x, y) dx dy &= \int_{-\infty}^{+\infty} h(x)f_X(x) dx \int_{-\infty}^{+\infty} g(y)f_Y(y) dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x)g(y)f_X(x)f_Y(y) dx dy. \end{aligned}$$

Di qui si trae:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x)g(y)[f_{XY}(x,y) - f_X(x)f_Y(y)]dxdy = 0$$

per ogni scelta di f e g . Di conseguenza deve essere

$$f_{XY}(x,y) = f_X(x)f_Y(y), \quad (6.51)$$

che è proprio la (ii). Integrando ambo i membri della (6.51) segue anche:

$$\int_{-\infty}^{\xi} dx \int_{-\infty}^{\eta} dy f_{XY}(x,y) = \int_{-\infty}^{\xi} dx f_X(x) \int_{-\infty}^{\eta} dy f_Y(y),$$

e quindi:

$$F_{XY}(\xi, \eta) = F_X(\xi)F_Y(\eta) \quad \forall \xi, \eta \in \mathbb{R},$$

ossia la (i).

□

6.5 Coefficiente di correlazione

La covarianza di due vv.cc. X e Y descrive la loro dipendenza statistica nel senso che quando è non nulla, essa fornisce una misura di come le due variabili casuali tendono a variare insieme. Infatti, in termini intuitivi possiamo affermare che $\text{cov}(X, Y)$ risulterà positiva se quando X assume valori maggiori della sua media lo stesso accade ad Y , e se quando X assume valori minori della sua media lo stesso succede ad Y . Se $X - E(X)$ e $Y - E(Y)$ hanno segno opposto con alta probabilità, $\text{cov}(X, Y)$ risulterà negativa, mentre risulterà prossima a zero quando con pari frequenza $Y - E(Y)$ è negativa allorché $X - E(X)$ è positiva, e viceversa. In tal senso $\text{cov}(X, Y)$ misura la "covariabilità" di X e Y .

La covarianza, però, non solo non è adimensionale, ma in generale assume valori sull'intero asse reale. Per poter quindi porre in relazione vv.cc. arbitrarie conviene ricorrere ad un coefficiente adimensionale che assuma valori in un intervallo limitato e preassegnato ma che abbia un significato analogo a quello della covarianza. Proprio a tal fine si è già introdotto il coefficiente di correlazione del quale ci occuperemo qui in qualche dettaglio.

Siano X, Y vv.cc. dotate di momenti del secondo ordine finiti; come sappiamo, esistono allora valore medio e varianza di X e Y nonché il momento prodotto $E(XY)$.⁵ Di conseguenza è possibile definire la covarianza di X e Y . Ricordiamo che se X e Y sono non degeneri, si dice coefficiente di correlazione di X, Y (cfr. la (6.35)) la costante $\rho(X, Y)$ così definita:

$$\rho(X, Y) = \frac{\text{cov}\{X, Y\}}{D(X) D(Y)}. \quad (6.52)$$

⁵ $E(XY)$ è denominata *correlazione* di X e Y . Questa si identifica con la covarianza se almeno una delle vv.cc. ha media nulla.

Il coefficiente di correlazione gode di varie interessanti proprietà.

Proposizione 6.18. *Il coefficiente di correlazione $\rho(X, Y)$ soddisfa la condizione:*

$$-1 \leq \rho(X, Y) \leq 1. \quad (6.53)$$

Dimostrazione. Essa segue dalla diseguaglianza di Schwarz (6.37). Si ha invero:

$$\left\{ E\{[X - E(X)][Y - E(Y)]\} \right\}^2 \leq E\{[X - E(X)]^2\} E\{[Y - E(Y)]^2\},$$

ossia:

$$\text{cov}^2\{X, Y\} \leq D^2(X) D^2(Y). \quad (6.54)$$

Se X e Y sono non degeneri, dalla (6.54) segue:

$$\rho^2(X, Y) = \frac{\text{cov}^2\{X, Y\}}{D^2(X) D^2(Y)} \leq 1,$$

che, a sua volta, implica la (6.53). □

A titolo di esercizio effettuiamo la dimostrazione della Proposizione 6.18 facendo uso di un diverso procedimento. A tal fine, siano W e Z variabili casuali. Denotato con a un reale arbitrario, $(aW - Z)^2$ è una v.c. a valori non negativi, così che il suo valore medio è non negativo:

$$E[(aW - Z)^2] \geq 0.$$

Pertanto

$$E(a^2 W^2 - 2aWZ + Z^2) \geq 0,$$

e quindi per ogni a risulta:

$$a^2 E(W^2) - 2a E(WZ) + E(Z^2) \geq 0. \quad (6.55)$$

Scegliamo in particolare

$$a = \frac{E(WZ)}{E(W^2)};$$

con questa scelta di a la (6.55) fornisce:

$$-\frac{[E(WZ)]^2}{E(W^2)} + E(Z^2) \geq 0,$$

e quindi:

$$1 \geq \frac{[E(WZ)]^2}{E(W^2) E(Z^2)}. \quad (6.56)$$

Per $W = X - E(X)$ e $Z = Y - E(Y)$ la (6.56) dà così $1 \geq \rho^2$, ossia la tesi.⁶

Proposizione 6.19. *Se due vv.cc. X e Y non degeneri sono legate da una relazione lineare del tipo*

$$Y = aX + b, \quad (6.57)$$

con $a, b \in \mathbb{R}$ ($a \neq 0$), risulta:

$$\rho(X, Y) = \frac{a}{|a|} = \begin{cases} +1, & \text{per } a > 0, \\ -1, & \text{per } a < 0. \end{cases} \quad (6.58)$$

Dimostrazione. Dalla (6.57) e dalle proprietà del valore medio segue

$$E(Y) = aE(X) + b.$$

Si ha inoltre:

$$\begin{aligned} D^2(Y) &= E\left\{\left[aX + b - [aE(X) + b]\right]^2\right\} \\ &= E\left\{[aX + b - aE(X) - b]^2\right\} \\ &= a^2 E\left\{[X - E(X)]^2\right\} \\ &\equiv a^2 D^2(X). \end{aligned} \quad (6.59)$$

Ancora, risulta:

$$\begin{aligned} \text{cov}\{X, Y\} &= E\left\{[X - E(X)][Y - E(Y)]\right\} \\ &= E\left\{[X - E(X)][aX + b - aE(X) - b]\right\} \\ &= aE\left\{[X - E(X)]^2\right\} \\ &\equiv aD^2(X). \end{aligned} \quad (6.60)$$

Dalle relazioni (6.59) e (6.60) si trae allora:

$$\rho(X, Y) \equiv \frac{\text{cov}\{X, Y\}}{D(X)D(Y)} = \frac{aD^2(X)}{|a|D^2(X)} = \frac{a}{|a|},$$

da cui segue la (6.58). \square

⁶Si noti che si è assunta l'esistenza e la finitezza di medie, varianze e covarianze.

Val la pena notare che qualunque sia il segno di a risulta $|\rho(X, Y)| = 1$.

Proposizione 6.20. *Nell'ipotesi in cui risulta $|\rho(X, Y)| = 1$, con probabilità 1 si ha $Y = aX + b$, con*

$$a = \frac{\rho}{|\rho|} \frac{D(Y)}{D(X)}, \quad (6.61)$$

$$b = E(Y) - \frac{\rho}{|\rho|} \frac{D(Y)}{D(X)} E(X). \quad (6.62)$$

Dimostrazione. La dimostrazione si basa essenzialmente sull'osservazione che, posto

$$Z = aX + b - Y, \quad (6.63)$$

con le scelte (6.61) e (6.62) di a e b , si ottiene:

$$E(Z) = 0, \quad (6.64)$$

$$D^2(Z) = 0. \quad (6.65)$$

In virtù del Teorema 6.1 segue allora che Z assume il valore zero con probabilità 1 o, equivalentemente, che Y assume il valore $aX + b$ con probabilità unitaria.

Al fine di ottenere le (6.64) e (6.65) mostriamo che i coefficienti della (6.63) devono essere scelti proprio come indicato dalle (6.61) e (6.62). Si ha infatti:

$$E(Z) = E(aX + b - Y) = aE(X) + b - E(Y). \quad (6.66)$$

Imponendo che $E(Z)$ sia nullo, dalla (6.66) si ricava:

$$b = E(Y) - aE(X). \quad (6.67)$$

Calcoliamo la varianza di Z ; poiché $E(Z) = E(aX + b - Y) = 0$, si ha:

$$D^2(Z) = D^2(aX + b - Y) = E[(aX + b - Y)^2]$$

e quindi, in virtù della (6.67):

$$\begin{aligned} D^2(Z) &= E[(aX + E(Y) - aE(X) - Y)^2] \\ &= E\left\{[a(X - E(X)) - (Y - E(Y))]^2\right\} \\ &= a^2 E\left\{[X - E(X)]^2\right\} - 2a E\left\{[X - E(X)][Y - E(Y)]\right\} \\ &\quad + E\left\{(Y - E(Y))^2\right\} \\ &= a^2 D^2(X) - 2a \text{cov}\{X, Y\} + D^2(Y). \end{aligned} \quad (6.68)$$

Osserviamo che dall'ipotesi $|\rho| = 1$ segue

$$\left| \frac{\text{cov}\{X, Y\}}{D(X) D(Y)} \right| = 1,$$

ovvero:

$$\text{cov}\{X, Y\} = \begin{cases} +D(X) D(Y), & \text{per } \rho = +1, \\ -D(X) D(Y), & \text{per } \rho = -1. \end{cases}$$

Qualunque sia ρ , purché in modulo unitario, possiamo allora scrivere:

$$\text{cov}\{X, Y\} = \frac{\rho}{|\rho|} D(X) D(Y). \quad (6.69)$$

Imponendo che risulti $D^2(Z) = 0$ e facendo uso delle (6.68) e (6.69) si ottiene:

$$\begin{aligned} 0 &= a^2 D^2(X) - 2a \frac{\rho}{|\rho|} D(X) d(Y) + \left(\frac{\rho}{|\rho|} \right)^2 D^2(Y) \\ &= \left[a D(X) - \frac{\rho}{|\rho|} D(Y) \right]^2 \end{aligned} \quad (6.70)$$

che, risolta rispetto ad a , fornisce la (6.61). Sostituendo, infine, il valore di a così ricavato nella (6.67), segue la (6.62). \square

Proposizione 6.21. *Date due vv.cc. non degeneri X e Y dotate di varianza finita, sussiste la relazione seguente:*

$$D^2(X \pm Y) = D^2(X) + D^2(Y) \pm 2\rho(X, Y) D(X) D(Y). \quad (6.71)$$

Dimostrazione. Dalla definizione di varianza segue:

$$\begin{aligned} D^2(X \pm Y) &\equiv E\left\{ [X \pm Y - E(X \pm Y)]^2 \right\} \\ &= E\left\{ \left\{ [X - E(X)] \pm [Y - E(Y)] \right\}^2 \right\} = E\left\{ [(X - E(X))^2] \right\} \\ &\quad + E\left\{ [Y - E(Y)]^2 \right\} \pm 2E\left\{ [X - E(X)][Y - E(Y)] \right\} \\ &= D^2(X) + D^2(Y) \pm 2\text{cov}\{X, Y\}, \end{aligned}$$

che è equivalente alla (6.71). \square

La (6.71) è anche nota come *legge della varianza*. Notiamo che se X e Y sono scorrelate si ha $\rho(X, Y) = 0$ così che la (6.71) assume la forma più semplice:

$$D^2(X \pm Y) = D^2(X) + D^2(Y). \quad (6.72)$$

La (6.72) vale certamente se, in particolare, X e Y sono indipendenti.

Dalla (6.72) si conclude che effettuando o somma o differenza di due vv.cc. indipendenti, la varianza della v.c. risultante è sempre uguale alla *somma* delle varianze delle vv.cc. originarie. Questa proprietà additiva delle varianze non è condivisa dalle deviazioni standard, per le quali si ha invece:

$$D(X + Y) = \sqrt{D^2(X) + D^2(Y)}.$$

CAPITOLO 7

Funzioni Ausiliarie

7.1 Funzioni generatrici

Talora è conveniente effettuare la determinazione di distribuzioni di probabilità o di momenti di variabili casuali facendo ricorso a delle funzioni ausiliarie che spesso conducono a notevoli semplificazioni.

Definizione 7.1. Sia X una v.c. discreta a valori interi non negativi e sia $\{p_X(k)\}$ la sua distribuzione di probabilità. La funzione $G: \mathbb{R} \rightarrow \mathbb{R}$ così definita

$$G(s) = E(s^X) = \sum_k p_X(k) s^k \quad (|s| \leq 1), \quad (7.1)$$

prende il nome di funzione generatrice di probabilità (f.g.p.). ◀

Il valore medio nella (7.1) esiste sempre finito per $|s| \leq 1$ poiché

$$E(|s|^X) = \sum_k p_X(k) |s|^k \leq \sum_k p_X(k) = 1.$$

La denominazione data alla funzione G è conseguenza del fatto che se questa viene sviluppata in serie di potenze di s , il coefficiente di s^k coincide con la probabilità $p_X(k)$ per ogni intero non negativo k . In virtù della continuità di G e delle proprietà delle distribuzioni di probabilità, si ha poi:

$$\lim_{s \uparrow 1} G(s) = 1. \quad (7.2)$$

La (7.2) esprime, in termini di funzione generatrice, la condizione di normalizzazione della distribuzione di probabilità $\{p_X(k)\}$.

Proposizione 7.1. La f.g.p. $G(s)$ gode delle proprietà approssimamente elencate.

(i) Per ogni intero positivo n , la derivata di ordine n di $G(s)$, calcolata per $s = 1$, fornisce una combinazione lineare dei momenti della v.c. X fino all'ordine n , sempre che questi esistano. In particolare, sussistono le seguenti relazioni:

$$\frac{dG(s)}{ds} \Big|_{s=1} = E(X), \quad (7.3)$$

$$\frac{d^2G(s)}{ds^2} \Big|_{s=1} = E(X^2) - E(X), \quad (7.4)$$

$$\frac{d^3G(s)}{ds^3} \Big|_{s=1} = E(X^3) - 3E(X^2) + 2E(X). \quad (7.5)$$

(ii) La varianza di X è esprimibile mediante le derivate del primo e del secondo ordine di $G(s)$ valutate per $s = 1$,

$$D^2(X) = \left\{ \frac{d^2G(s)}{ds^2} + \frac{dG(s)}{ds} \left[1 - \frac{dG(s)}{ds} \right] \right\}_{s=1}. \quad (7.6)$$

Dimostrazione. L'affermazione (i) si dimostra immediatamente. Infatti dalla (7.1) si ha:

$$\frac{dG(s)}{ds} \Big|_{s=1} = \sum_k k p_X(k) = E(X)$$

$$\frac{d^2G(s)}{ds^2} \Big|_{s=1} = \sum_k k(k-1) p_X(k) = E(X^2) - E(X)$$

$$\begin{aligned} \frac{d^3G(s)}{ds^3} \Big|_{s=1} &= \sum_k k(k-1)(k-2) p_X(k) = \sum_k (k^3 - 3k^2 + 2k) p_X(k) \\ &= E(X^3) - 3E(X^2) + 2E(X), \end{aligned}$$

e così via. Per quanto riguarda l'affermazione (ii), la (7.6) segue direttamente dalle (7.3) e (7.4), avendosi:

$$\begin{aligned} D^2(X) &= E \left\{ [X - E(X)]^2 \right\} = E(X^2) - [E(X)]^2 \\ &= \left\{ \frac{d^2G(s)}{ds^2} + \frac{dG(s)}{ds} - \left[\frac{dG(s)}{ds} \right]^2 \right\}_{s=1} \\ &= \left\{ \frac{d^2G(s)}{ds^2} + \frac{dG(s)}{ds} \left[1 - \frac{dG(s)}{ds} \right] \right\}_{s=1}. \end{aligned} \quad \square$$

7.1 Funzioni generatrici

Si noti che la descrizione della v.c. X richiede la conoscenza della sua distribuzione di probabilità che può riguardarsi come un vettore in generale a infinite componenti p_0, p_1, \dots . Alternativamente, e più semplicemente, per la descrizione di X si può assegnare un'unica funzione, la f.g.p.: in tal modo l'intero vettore a infinite componenti viene, per così dire, condensato in un'unica funzione. In molti casi concreti la f.g.p. si rivela particolarmente utile anche in quanto può risultare più semplice determinare questa piuttosto che la distribuzione di probabilità; è possibile poi ottenere la distribuzione di probabilità sviluppando la f.g.p., se possibile, in serie di potenze secondo la (7.1). Un esempio non banale verrà fornito in seguito, quando discuteremo di passeggiate aleatorie in presenza di barriere assorbenti. Qui ci limitiamo a fornire un semplice esempio di calcolo di una funzione generatrice di probabilità.

Esempio 7.1. Sia X una v.c. avente distribuzione di probabilità binomiale:

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

Facendo uso della Definizione 7.1 si ha:

$$G(s) = \sum_{k=0}^n \binom{n}{k} (ps)^k (1-p)^{n-k} = (ps+1-p)^n = [1+p(s-1)]^n. \quad (7.7)$$

Mediane le (7.3) e (7.6) si ricavano valor medio e varianza. Si ottiene così:

$$E(X) = \frac{dG(s)}{ds} \Big|_{s=1} = \left\{ np [1+p(s-1)]^{n-1} \right\}_{s=1} = np, \quad (7.8)$$

ed inoltre:

$$\begin{aligned} D^2(X) &= \left\{ \frac{d^2G(s)}{ds^2} + \frac{dG(s)}{ds} \left[1 - \frac{dG(s)}{ds} \right] \right\}_{s=1} \\ &= \left\{ n(n-1)p^2 [1+p(s-1)] \right\}_{s=1}^{n-2} + np(1-np) \\ &= n^2p^2 - np^2 + np - n^2p^2 = np(1-p). \end{aligned} \quad (7.9)$$

Notiamo poi che risulta:

$$\lim_{s \uparrow 1} [1+p(s-1)] = 1. \quad (7.10)$$

La (7.10) esprime la condizione di normalizzazione della distribuzione binomiale. \triangleleft

L'estensione della nozione di f.g.p. al caso di una v.c. arbitraria X si ottiene considerando la funzione

$$G(s) = E(e^{-sx}) = \int_{-\infty}^{+\infty} e^{-sx} dF_X(x), \quad (7.11)$$

dove s è una variabile reale di comodo scelta in modo da rendere convergente, se possibile, l'integrale che compare nella (7.11).

Definizione 7.2. Si dice funzione generatrice dei momenti (f.g.m.) della v.c. X di funzione di distribuzione F_X la funzione

$$M(t) = E(e^{tX}) = \int_{-\infty}^{+\infty} e^{tx} dF_X(x), \quad (7.12)$$

con t reale, sempre che l'integrale nella (7.12) esista finito in un intorno $|t| \leq t_0$ dell'origine. \triangleleft

Proposizione 7.2. La f.g.m. gode delle seguenti proprietà:

(i) soddisfa la condizione di normalizzazione:

$$\lim_{t \rightarrow 0} M(t) = 1; \quad (7.13)$$

(ii) data la v.c. $Y = aX + b$ ($a, b \in \mathbb{R}$), e denotate con $M_X(t)$ e $M_Y(t)$ le f.f.g.m. rispettivamente di X e Y , per ogni t tale che $M_X(at)$ esiste si ha:

$$M_Y(t) = e^{bt} M_X(at). \quad (7.14)$$

Dimostrazione. La (7.13) segue immediatamente dalla (7.12) e dalla continuità di $M(t)$. Anche la (7.14) è facilmente dimostrata avendosi:

$$M_Y(t) = E(e^{tY}) = E\left[e^{t(aX+b)}\right] = e^{bt} E(e^{atX}) = e^{bt} M_X(at). \quad \square$$

Teorema 7.1. Se una v.c. X possiede la f.g.m. $M(t)$ e se questa è finita per $|t| \leq t_0$, i momenti $\mu'_k = E(X^k)$ sono tutti finiti e si ha:

$$M(t) = \sum_{k=0}^{\infty} \mu'_k \frac{t^k}{k!} \quad \text{per } |t| \leq t_0, \quad (7.15)$$

con

$$\mu'_n = \left[\frac{d^n}{dt^n} M(t) \right]_{t=0} \quad (n = 1, 2, \dots). \quad (7.16)$$

Dimostrazione. Poiché per ogni z reale risulta

$$e^{|z|} \leq e^{|z|} + e^{-|z|} = e^z + e^{-z},$$

per ogni $|t| \leq t_0$ si ha:

$$E\left(e^{|tX|}\right) = \int_{-\infty}^{+\infty} e^{|tx|} dF(x) \leq M(t) + M(-t) < +\infty. \quad (7.17)$$

7.1 Funzioni generatrici

Facciamo ora uso della relazione

$$\frac{|z|^n}{n!} \leq \sum_{k=0}^{\infty} \frac{|z|^k}{k!} \equiv e^{|z|} \quad (n = 1, 2, \dots) \quad (7.18)$$

valida per ogni $z \in \mathbb{R}$. Ponendo $z = tx$ nella (7.18), dove x denota il generico valore di X , otteniamo:

$$\frac{|t|^n}{n!} |x|^n \leq e^{|tx|}$$

e quindi:

$$\frac{|t|^n}{n!} \int_{-\infty}^{+\infty} |x|^n dF(x) \leq \int_{-\infty}^{+\infty} e^{|tx|} dF(x) < +\infty,$$

dove si è fatto ricorso alla (7.17). Pertanto tutti i momenti assoluti di X , e quindi anche tutti i suoi momenti intorno all'origine, esistono finiti. Inoltre si ha:

$$M(t) = \int_{-\infty}^{+\infty} e^{tx} dF(x) = \int_{-\infty}^{+\infty} \lim_{n \rightarrow +\infty} \sum_{k=0}^n \frac{t^k x^k}{k!} dF(x). \quad (7.19)$$

Notiamo che, essendo

$$\left| \sum_{k=0}^n \frac{t^k x^k}{k!} \right| \leq \sum_{k=0}^n \frac{|t^k x^k|}{k!} \leq e^{|tx|},$$

ed avendosi per la (7.17) $E(e^{|tX|}) < +\infty$ per $|t| \leq t_0$, si può applicare il teorema della convergenza dominata. Dalla (7.19) si ottiene così:

$$\begin{aligned} M(t) &= \lim_{n \rightarrow +\infty} \int_{-\infty}^{+\infty} \sum_{k=0}^n \frac{t^k x^k}{k!} dF(x) = \lim_{n \rightarrow +\infty} \sum_{k=0}^n \frac{t^k}{k!} \int_{-\infty}^{+\infty} x^k dF(x) \\ &= \sum_{k=0}^{\infty} \mu'_k \frac{t^k}{k!} \end{aligned}$$

che fornisce la (7.15). Ricordando, infine, che una serie di potenze può essere derivata termine a termine, dalla (7.15) segue:

$$\frac{d^n}{dt^n} M(t) = \sum_{k=n}^{\infty} \mu'_k \frac{t^{k-n}}{(k-n)!} \quad \text{per } |t| \leq t_0,$$

da cui si trae:

$$\left[\frac{d^n}{dt^n} M(t) \right]_{t=0} = \mu'_n \quad (n = 1, 2, \dots),$$

ossia la (7.16). \square

Si noti che nelle ipotesi del Teorema 7.1 sia la f.g.m. che la successione dei momenti individuano la distribuzione di X . Dunque due vv.cc. X e Y sono identicamente distribuite nei seguenti due casi:

- (i) le rispettive ff.gg.mf. sono tali da aversi $M_X(t) = M_Y(t)$ per $|t| \leq t_0$;
- (ii) X è dotata di f.g.m. ed inoltre risulta $E(X^k) = E(Y^k)$ per ogni k intero positivo.

Teorema 7.2. Siano X_1, X_2, \dots, X_n vv.cc. indipendenti, dotate di ff.gg.mm. $M_i(t)$ ($i = 1, \dots, n$) e sia Y la v.c. somma delle variabili X_i :

$$Y = X_1 + X_2 + \dots + X_n.$$

Se Y è dotata di f.g.m. $M(t)$, si ha:

$$M(t) = \prod_{i=1}^n M_i(t). \quad (7.20)$$

Dimostrazione. Facendo uso della (7.12), in virtù dell'ipotizzata indipendenza delle vv.cc. X_i , può scriversi:

$$M(t) = E(e^{tY}) = E\left[e^{t(X_1+\dots+X_n)}\right] = E(e^{tX_1})E(e^{tX_2})\cdots E(e^{tX_n})$$

da cui segue immediatamente la (7.20). \square

Notiamo che le funzioni generatrici G e M di una v.c. discreta sono legate dalle seguenti relazioni:

$$G(s) = M(\ln s), \quad M(t) = G(e^t). \quad (7.21)$$

Infatti si ha:

$$\begin{aligned} G(s) &= \sum_k p_X(k)s^k = \sum_k p_X(k)(e^{\ln s})^k = \sum_k p_X(k)e^{k\ln s} = \int_{-\infty}^{+\infty} e^{x\ln s} dF_X(x) \\ &\equiv M(\ln s). \end{aligned}$$

Risulta inoltre:

$$\begin{aligned} M(t) &= \int_{-\infty}^{+\infty} e^{tx} dF_X(x) = \sum_k e^{tk} p_X(k) = \sum_k (e^t)^k p_X(k) \\ &\equiv G(e^t). \end{aligned}$$

L'importanza della f.g.m. in parte risiede nella circostanza che spesso si ha a che fare con vv.cc. per le quali essa è facilmente determinabile, laddove non risulta tale la funzione

7.1 Funzioni generatrici

di distribuzione. In questi casi, anche se non è possibile risalire dalla f.g.m. alla funzione di distribuzione, è possibile comunque trarre informazioni sulla v.c. dato che i momenti sono ottenibili, almeno in linea di principio, a mezzo della (7.16).

Diamo ora degli esempi di calcolo di funzione generatrice dei momenti.

Esempio 7.2. Sia X una v.c. esponenziale, avente quindi densità di probabilità¹

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{per } x > 0, \\ 0, & \text{altrove.} \end{cases}$$

Usando la (7.12) si ricava la f.g.m.:

$$M(t) = \int_0^{+\infty} e^{tx} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - t} \quad (t < \lambda).$$

Possiamo quindi calcolare il valor medio al seguente modo:

$$E(X) = \frac{d}{dt} M(t) \Big|_{t=0} = \frac{\lambda}{(\lambda - t)^2} \Big|_{t=0} = \frac{1}{\lambda}.$$

Per la varianza si ottiene poi:

$$\begin{aligned} D^2(X) &= E(X^2) - [E(X)]^2 = \left\{ \frac{d^2}{dt^2} M(t) - \left[\frac{d}{dt} M(t) \right]^2 \right\}_{t=0} \\ &= \left\{ \frac{2\lambda}{(\lambda - t)^3} - \left[\frac{\lambda}{(\lambda - t)^2} \right]^2 \right\}_{t=0} \\ &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}. \end{aligned}$$

Esempio 7.3. Consideriamo ora il caso di una v.c. normale $X = N(m, \sigma^2)$ la cui densità di probabilità è pertanto

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

¹Cfr. la (6.22) del § 6.1.2.

e calcoliamone la f.g.m. $M(t)$. Si ha:

$$\begin{aligned} M(t) &= \int_{-\infty}^{+\infty} e^{xt} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\} dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2\sigma^2}[-2xt\sigma^2 + (x-m)^2]\right\} dx \\ &= e^{mt+t^2\sigma^2/2} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2\sigma^2}[x-(m+t\sigma^2)]^2\right\} dx. \end{aligned} \quad (7.22)$$

Poiché la funzione

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}[x-(m+t\sigma^2)]^2\right\}$$

rappresenta la densità di una variabile normale di valor medio $m+t\sigma^2$ e varianza σ^2 , per la condizione di normalizzazione risulta:

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2\sigma^2}[x-(m+t\sigma^2)]^2\right\} dx = 1. \quad (7.23)$$

Dalle (7.22) e (7.23) si ricava dunque:

$$M(t) = \exp\left(mt + \frac{1}{2}\sigma^2 t^2\right) \quad (t \in \mathbb{R}). \quad (7.24)$$

Calcolando le derivate prime e seconde della (7.24) è poi immediato ottenere media e varianza. Si ha infatti:

$$\begin{aligned} M'(t) &= (m + \sigma^2 t) M(t) \\ M''(t) &= [\sigma^2 + (m + \sigma^2 t)^2] M(t), \end{aligned}$$

da cui segue:

$$\begin{aligned} M'(0) &= m \\ M''(0) &= \sigma^2 + m^2. \end{aligned}$$

In conclusione si trae:

$$\begin{aligned} E(X) &\equiv M'(0) = m \\ D^2(X) &\equiv M''(0) - [M'(0)]^2 = \sigma^2. \end{aligned}$$

□

7.2 Il processo di Poisson

Consideriamo un sistema che riceve in successione nel tempo sollecitazioni di una qualche natura e assumiamo che ciascuna di queste abbia durata infinitesimale, in modo che si possa parlare di *istante di occorrenza* di ciascuna sollecitazione. Supponiamo, inoltre, che le sollecitazioni abbiano luogo in maniera indipendente le une dalle altre. Il quadro descritto rientra nella categoria dei cosiddetti *processi di punto*, caratterizzati, come il termine indica, da eventi di durate assimilabili a punti (ossia di durate infinitamente piccole). Si tratta evidentemente di un'astrazione, che peraltro risulta comoda ed utile in molte applicazioni.² Fissato un intervallo di tempo di durata t , ci proponiamo di calcolare la probabilità $p_n(t)$ che n eventi abbiano luogo in questo intervallo, sotto opportune ipotesi. Denotiamo a tal fine con $N(t)$ la v.c. rappresentante il numero di eventi che hanno luogo nell'intervallo di tempo $(0, t]$. Evidentemente $N(t)$ è monotona non decrescente in t . Poniamo poi:

$$\Delta N(t) = N(t + \Delta t) - N(t).$$

Supporremo ora che esista una costante $\alpha > 0$ tale da aversi:

$$(i) \mathbb{P}(\Delta N(t) = 1) = \alpha \Delta t + o(\Delta t),$$

$$(ii) \mathbb{P}(\Delta N(t) > 1) = o(\Delta t),$$

con $o(\Delta t)$ infinitesimi di ordine superiore rispetto a Δt . Dalle (i) e (ii) segue:

$$\mathbb{P}(\Delta N(t) = 0) = 1 - \alpha \Delta t + o(\Delta t). \quad (7.25)$$

Cominciamo col calcolare $p_0(t)$. Per l'indipendenza postulata e per la (7.25) si ha:

$$p_0(t + \Delta t) = p_0(t)[1 - \alpha \Delta t + o(\Delta t)],$$

ovvero:

$$\frac{p_0(t + \Delta t) - p_0(t)}{\Delta t} = -\alpha p_0(t), \quad (7.26)$$

dove si sono trascurati gli infinitesimi di ordine superiore rispetto a Δt . Passando al limite per $\Delta t \rightarrow 0$, dalla (7.26) si ottiene l'equazione differenziale

$$\frac{d}{dt} p_0(t) = -\alpha p_0(t), \quad (7.27)$$

che va integrata con la condizione iniziale $p_0(0) = 1$. Si ha così:

$$p_0(t) = e^{-\alpha t}. \quad (7.28)$$

²Cfr. anche il § 1.2.

In maniera simile si può calcolare $p_1(t)$. Questa volta si fa uso delle (i) e (ii) per scrivere la relazione:

$$p_1(t + \Delta t) = p_1(t)[1 - \alpha \Delta t + o(\Delta t)] + p_0(t)[\alpha \Delta t + o(\Delta t)].$$

Di qui, trascurando gli infinitesimi di ordine superiore rispetto a Δt , si ottiene:

$$\frac{p_1(t + \Delta t) - p_1(t)}{\Delta t} = -\alpha[p_1(t) - p_0(t)]. \quad (7.29)$$

Passando infine al limite per $\Delta t \rightarrow 0$, dalla (7.29) si trae:

$$\frac{d}{dt}p_1(t) = -\alpha[p_1(t) - p_0(t)]. \quad (7.30)$$

Tale equazione va integrata con la condizione iniziale $p_1(0) = 0$. Sostituendo la (7.28) nella (7.30) si ha l'equazione differenziale lineare non omogenea

$$\frac{d}{dt}p_1(t) = -\alpha p_1(t) + \alpha e^{-\alpha t}.$$

Il suo integrale generale si ottiene comandando all'integrale generale dell'equazione omogenea associata

$$\frac{d}{dt}p_1(t) = -\alpha p_1(t)$$

un integrale particolare dell'equazione completa. Poiché una soluzione particolare dell'equazione completa è $\alpha t e^{-\alpha t}$ e la soluzione generale dell'equazione omogenea associata è $C e^{-\alpha t}$, con C costante arbitraria, ricordando che è $p_1(0) = 0$ si ha in definitiva:

$$p_1(t) = \alpha t e^{-\alpha t}.$$

In maniera perfettamente analoga si ottiene l'equazione

$$\frac{d}{dt}p_n(t) = -\alpha[p_n(t) - p_{n-1}(t)] \quad (n = 1, 2, \dots), \quad (7.31)$$

che va risolta iterativamente utilizzando le condizioni iniziali $p_n(0) = 0$, con $n = 1, 2, \dots$. Non è difficile convincersi che risulta:

$$p_n(t) = \frac{(\alpha t)^n}{n!} e^{-\alpha t} \quad (n = 1, 2, \dots). \quad (7.32)$$

In conclusione, dalle (7.28) e (7.32) segue:

$$p_n(t) = \frac{(\alpha t)^n}{n!} e^{-\alpha t} \quad (n = 0, 1, \dots). \quad (7.33)$$

$N(t)$ è dunque una v.c. di Poisson di valore medio αt .

7.2.1 Utilizzazione della f.g.p.

Il procedimento indicato per il calcolo di $p_n(t)$ risulta piuttosto laborioso. Molto più semplicemente si può procedere nel seguente modo. Sia

$$G(s, t) = \sum_{n=0}^{\infty} p_n(t)s^n \quad (7.34)$$

la f.g.p. di $N(t)$. Dall'equazione (7.31) segue facilmente:

$$\frac{\partial}{\partial t} \left[\sum_{n=1}^{\infty} s^n p_n(t) \right] = -\alpha \left[\sum_{n=1}^{\infty} s^n p_n(t) - s \sum_{n=1}^{\infty} s^{n-1} p_{n-1}(t) \right],$$

ovvero, facendo uso della (7.34):

$$\frac{\partial}{\partial t}[G(s, t) - p_0(t)] = -\alpha[G(s, t) - p_0(t) - s G(s, t)]$$

che può riscriversi nel seguente modo:

$$\frac{\partial}{\partial t}G(s, t) = -\alpha G(s, t)(1 - s) + \frac{d}{dt}p_0(t) + \alpha p_0(t). \quad (7.35)$$

Facendo uso della (7.27), dalla (7.35) segue:

$$\frac{\partial}{\partial t}G(s, t) = -\alpha(1 - s)G(s, t).$$

Questa equazione va integrata con la condizione iniziale:

$$G(s, 0) \equiv \sum_{n=0}^{\infty} p_n(0)s^n = \sum_{n=0}^{\infty} \delta_{n,0}s^n = 1, \quad (7.36)$$

dove $\delta_{n,m}$ denota il simbolo di Kronecker:

$$\delta_{n,m} = \begin{cases} 1, & \text{per } n = m, \\ 0, & \text{per } n \neq m. \end{cases}$$

Si ha pertanto:

$$G(s, t) = \psi(s) e^{-\alpha(1-s)t}. \quad (7.37)$$

La funzione arbitraria $\psi(s)$ si riconosce poi valere 1. Infatti, dalle (7.36) e (7.37) si trae:

$$\psi(s) = G(s, 0) = 1. \quad (7.38)$$

In definitiva, sostituendo la (7.38) nella (7.37) si ha:

$$G(s, t) = e^{-\alpha(1-s)t}. \quad (7.39)$$

Per ottenere le incognite $p_n(t)$ basta sviluppare e^{st} in serie di potenze di punto iniziale $s = 0$. Dalla (7.39) segue così:

$$G(s, t) = e^{-\alpha t} \sum_{n=0}^{\infty} \frac{(\alpha t s)^n}{n!} = \sum_{n=0}^{\infty} \left[\frac{(\alpha t)^n}{n!} e^{-\alpha t} \right] s^n. \quad (7.40)$$

Confrontando la (7.40) con la (7.34) si ottiene immediatamente:

$$p_n(t) = \frac{(\alpha t)^n}{n!} e^{-\alpha t} \quad (n = 0, 1, \dots),$$

che coincide con la (7.33).

Mediane la funzione $G(s, t)$ sono poi calcolabili i momenti di $N(t)$, che però si possono ottenere più agevolmente facendo ricorso alla funzione generatrice dei momenti. Questa si ricava dalla f.g.p. come indicato nella (7.21):

$$M(\tau, t) \equiv G(e^\tau, t) = e^{-\alpha(1 - e^\tau)t}.$$

Segue così:

$$E[N(t)] = \frac{\partial M(\tau, t)}{\partial \tau} \Big|_{\tau=0} = \alpha t, \quad (7.41)$$

$$E[N^2(t)] = \frac{\partial^2 M(\tau, t)}{\partial \tau^2} \Big|_{\tau=0} = (\alpha t)^2 + \alpha t. \quad (7.42)$$

Dalle (7.41) e (7.42) si ricava in particolare la varianza di $N(t)$:

$$D^2[N(t)] \equiv E[N^2(t)] - \{E[N(t)]\}^2 = \alpha t.$$

Analogamente si procede per il calcolo dei momenti di ordine superiore.

CAPITOLO 8

La Disuguaglianza di Chebyshev

8.1 Alcune proprietà delle variabili casuali normali

Come sappiamo (cfr. § 6.19), si definisce *gaussiana* o *normale* una variabile casuale X assolutamente continua la cui densità di probabilità f è espressa dalla (6.19) che, per comodità, qui riscriviamo:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(x-m)^2}{2\sigma^2} \right], \quad \forall x \in \mathbb{R}. \quad (8.1)$$

Come abbiamo visto, si ha:

$$E(X) = m, \quad D^2(X) = \sigma^2.$$

Note, quindi, media e varianza, la v.c. normale risulta essere completamente individuata. Per tale motivo una v.c. normale di media m e varianza σ^2 viene convenzionalmente indicata con la scrittura:

$$X = N(m, \sigma^2).$$

Dai calcoli effettuati nel § 6.19 si è anche visto che la densità (8.1) possiede un punto stazionario, che è un massimo situato in $x = m$, e due punti di flesso in $x = m \pm \sigma$: inoltre, la funzione di distribuzione F di X è data da:

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x \exp \left[-\frac{(\xi-m)^2}{2\sigma^2} \right] d\xi, \quad \forall x \in \mathbb{R}. \quad (8.2)$$

Come è noto, l'integrale a secondo membro della (8.2) non è integrabile per via elementare e pertanto risulta necessario ricercare una rappresentazione di F in termini di funzioni speciali. A tale scopo, una volta effettuato nella (8.2) il cambiamento di variabile $z = (\xi - m)/\sqrt{2}\sigma$,

la funzione di distribuzione F di X si scrive

$$\begin{aligned} F(x) &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{(x-m)/\sqrt{2}\sigma} e^{-z^2} dz \\ &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^0 e^{-z^2} dz + \frac{1}{\sqrt{\pi}} \int_0^{(x-m)/\sqrt{2}\sigma} e^{-z^2} dz \\ &= \frac{1}{\sqrt{\pi}} \cdot \frac{1}{2} + \frac{1}{\sqrt{\pi}} \int_0^{(x-m)/\sqrt{2}\sigma} e^{-z^2} dz. \end{aligned}$$

Nell'ultimo passaggio si è fatto ricorso alla (6.20) ed al fatto che e^{-z^2} è una funzione pari. Da quest'ultima si evince che la funzione speciale a cui si è accennato in precedenza è la *funzione dell'errore o funzione di Gauss* definita mediante la posizione:

$$\text{Erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x e^{-z^2} dz. \quad (8.3)$$

Per quanto riguarda la funzione di distribuzione della v.c. normale X risulta in definitiva:

$$F(x) = \frac{1}{2} \left[1 + \text{Erf} \left(\frac{x-m}{\sqrt{2}\sigma} \right) \right]. \quad (8.4)$$

La funzione definita dalla (8.3) prende il nome di *funzione dell'errore o funzione di Gauss*. Si noti che questa è una funzione dispari, avendosi:

$$\text{Erf}(-x) = \frac{2}{\sqrt{\pi}} \int_0^{-x} e^{-z^2} dz = -\frac{2}{\sqrt{\pi}} \int_0^x e^{-z^2} dz = -\text{Erf}(x). \quad (8.5)$$

Inoltre, si lascia al lettore verificare che

$$\lim_{x \rightarrow +\infty} \text{Erf}(x) = 1. \quad (8.6)$$

L'utilità dell'espressione (8.4) consiste nel fatto che Erf è una funzione tabulata; ricorrendo a tavole numeriche si può quindi valutare F qualunque siano (compatibilmente con i valori numerici inclusi nelle tavole) media e varianza della v.c. normale assegnata.

Fissiamo ora un intervallo (a, b) . In virtù delle proprietà della funzione di distribuzione si ha:

$$\begin{aligned} \mathbb{P}(a < X < b) &= F(b) - F(a) \\ &= \frac{1}{2} \left[1 + \text{Erf} \left(\frac{b-m}{\sqrt{2}\sigma} \right) \right] - \frac{1}{2} \left[1 + \text{Erf} \left(\frac{a-m}{\sqrt{2}\sigma} \right) \right] \\ &= \frac{1}{2} \left[\text{Erf} \left(\frac{b-m}{\sqrt{2}\sigma} \right) - \text{Erf} \left(\frac{a-m}{\sqrt{2}\sigma} \right) \right]. \end{aligned} \quad (8.7)$$

Scelti, in particolare, $a = m - \epsilon$ e $b = m + \epsilon$, con $\epsilon > 0$ arbitrario, dalla (8.7) segue:

$$\begin{aligned} \mathbb{P}(m - \epsilon < X < m + \epsilon) &= \mathbb{P}(|X - m| < \epsilon) \\ &= \frac{1}{2} \left[\text{Erf} \left(\frac{\epsilon}{\sqrt{2}\sigma} \right) - \text{Erf} \left(\frac{-\epsilon}{\sqrt{2}\sigma} \right) \right] \end{aligned} \quad (8.8)$$

ovvero, facendo uso della (8.5):

$$\mathbb{P}(m - \epsilon < X < m + \epsilon) = \mathbb{P}(|X - m| < \epsilon) = \text{Erf} \left(\frac{\epsilon}{\sqrt{2}\sigma} \right). \quad (8.9)$$

Utilizziamo la (8.9) con le seguenti scelte di ϵ :

$$\epsilon = \sigma, \quad \epsilon = 2\sigma, \quad \epsilon = 3\sigma;$$

otteniamo così:

$$\mathbb{P}(m - \sigma < X < m + \sigma) = \mathbb{P}(|X - m| < \sigma) = \text{Erf} \left(\frac{1}{\sqrt{2}} \right), \quad (8.10)$$

$$\mathbb{P}(m - 2\sigma < X < m + 2\sigma) = \mathbb{P}(|X - m| < 2\sigma) = \text{Erf} \left(\frac{2}{\sqrt{2}} \right), \quad (8.11)$$

$$\mathbb{P}(m - 3\sigma < X < m + 3\sigma) = \mathbb{P}(|X - m| < 3\sigma) = \text{Erf} \left(\frac{3}{\sqrt{2}} \right). \quad (8.12)$$

Di solito, nei libri di Probabilità o di Statistica, al posto della tabella relativa ai valori della funzione dell'errore, viene fornita una tabella relativa ai valori assunti dalla funzione di distribuzione Φ della v.c. normale standard¹

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp \left(-\frac{\xi^2}{2} \right) d\xi, \quad \forall x \in \mathbb{R}. \quad (8.13)$$

per alcuni argomenti x iniziati da 0 ed equispaziati tra loro di un centesimo. Anche noi ci uniformeremo a tale consuetudine e forniremo una tale tabella alla fine di questo paragrafo. La ragione risiede nel fatto che la (8.4) scritta per $m = 0$ e $\sigma^2 = 1$ fornisce

$$\text{Erf} \left(\frac{x}{\sqrt{2}} \right) = 2\Phi(x) - 1, \quad (8.14)$$

in virtù della quale per le (8.10)-(8.12) otteniamo

$$\mathbb{P}(|X - m| < \sigma) = 2\Phi(1) - 1 \simeq 0.6826, \quad (8.15)$$

$$\mathbb{P}(|X - m| < 2\sigma) = 2\Phi(2) - 1 \simeq 0.9544, \quad (8.16)$$

$$\mathbb{P}(|X - m| < 3\sigma) = 2\Phi(3) - 1 \simeq 0.9974. \quad (8.17)$$

¹si ricorda che la v.c. normale standard è quella con media nulla e varianza unitaria.

Le approssimazioni numeriche riportate in queste ultime sono state tratte dalla tabella del § 8.5.

Osservando i risultati ottenuti si può concludere che vi è la pratica certezza che una v.c. normale assuma valori in un intervallo avente centro nel valore medio e semidimensone pari a 3σ . Questa proprietà delle vv.cc. normali è nota come *legge dei 3σ*.

Per una v.c. normale la (8.9) esprime la probabilità che il discostamento dal valore medio sia minore di un arbitrariamente preassegnato reale positivo ϵ ; da essa e dalle (8.6) si vede poi che $\mathbb{P}(|X - m| < \epsilon) \rightarrow 1$ allorché $\sigma \rightarrow 0$ per ogni fissato ϵ positivo. Possiamo così affermare che piccole varianze comportano grandi accentramenti intorno al valore medio dei valori che la v.c. assume, e quindi piccole dispersioni di questi intorno al valore medio stesso. Le (8.15) e (8.17) esprimono la quantificazione di queste dispersioni in taluni casi particolari.

Come si vedrà tra breve, la deviazione standard di una qualsiasi v.c., qualora finita, è suscettibile di una simile interpretazione (cfr. la Proposizione 8.2).

8.2 La disuguaglianza di Chebyshev

Sia X una v.c. di funzione di distribuzione F e sia $\psi: \mathbb{R} \rightarrow \mathbb{R}$ una funzione Borel-misurabile a valori non negativi che gode della seguente proprietà:

$$\exists a > 0: \mathbb{P}(\omega \in \Omega: \psi(X(\omega)) \geq a) > 0. \quad (8.18)$$

Sussiste la seguente

Proposizione 8.1. Se $Y = \psi(X)$ è dotata di valore medio finito, si ha:

$$\mathbb{P}(\psi(X) < a) \geq 1 - \frac{E[\psi(X)]}{a}. \quad (8.19)$$

Dimostrazione. Sia $A = \{\omega \in \Omega: \psi(X(\omega)) \geq a\}$. Dalla definizione di valore medio segue:

$$\begin{aligned} E[\psi(X)] &= \int_{-\infty}^{+\infty} \psi(x) dF(x) \\ &= \int_{\psi(x) \geq a} \psi(x) dF(x) + \int_{\psi(x) < a} \psi(x) dF(x) \\ &\geq \int_{\psi(x) \geq a} \psi(x) dF(x) \geq a \int_A dF(x) = a \mathbb{P}(X \in A), \end{aligned}$$

ossia:

$$\mathbb{P}(X \in A) \leq \frac{E[\psi(X)]}{a}.$$

8.2 La disuguaglianza di Chebyshev

Quindi:

$$\mathbb{P}(X \in \bar{A}) \geq 1 - \frac{E[\psi(X)]}{a},$$

dove \bar{A} denota l'insieme complementare di A rispetto a \mathbb{R} . D'altra parte, dalla definizione di A segue:

$$\mathbb{P}(X \in \bar{A}) = \mathbb{P}(\psi(X) < a),$$

così che in definitiva si perviene alla (8.19). \square

Il risultato espresso dalla (8.19) è valido qualunque sia la funzione ψ a valori non negativi soddisfacente la condizione (8.18). A seconda della forma della funzione ψ si possono ricavare varie disuguaglianze notevoli.

Proposizione 8.2 (Chebyshev). Se la v.c. X è non degenere e dotata di valore medio e varianza finiti, sussiste la relazione:

$$\mathbb{P}(|X - E(X)| < \epsilon) \geq 1 - \frac{D^2(X)}{\epsilon^2}. \quad (8.20)$$

Dimostrazione. Segue immediatamente dalla (8.19) ponendovi

$$\psi(X) = [X - E(X)]^2. \quad (8.21)$$

Ciò è lecito in quanto la funzione ψ è a valori non negativi; inoltre essa non può essere identicamente nulla in quanto per ipotesi X è non degenere. La funzione (8.21) soddisfa dunque la condizione (8.18). Posto allora $a = \epsilon^2$, dalla (8.19) si ricava:

$$\mathbb{P}([X - E(X)]^2 < \epsilon^2) \geq 1 - \frac{E\{[X - E(X)]^2\}}{\epsilon^2} \equiv 1 - \frac{D^2(X)}{\epsilon^2},$$

da cui segue la (8.20). \square

La Proposizione 8.2 esprime la celeberrima *disuguaglianza di Chebyshev* che riveste notevolissimo interesse sia teorico che applicativo. Consideriamone, ad esempio, l'utilizzazione per illustrare il significato della varianza di una v.c. A tal fine ricordiamo che nel caso della v.c. normale, come abbiamo già visto, alla varianza è legata una misura della sua dispersione intorno al valore medio. Ci si può allora chiedere se una proprietà analoga valga anche per vv.cc. non gaussiane. Ora, più in generale, ci poniamo il problema del significato della varianza di una v.c. Al fine di fornire una risposta a tale quesito, ed al contempo per delucidare il significato stesso della varianza di una v.c., osserviamo che dalla disuguaglianza di Chebyshev (8.20) per $\epsilon = h D(X)$, con $h > 0$, si ottiene:

$$\mathbb{P}(E(X) - h D(X) < X < E(X) + h D(X)) \geq 1 - \frac{1}{h^2}. \quad (8.22)$$

In particolare, per $h = 3$ dalla (8.22) si trae:

$$\mathbb{P}(E(X) - 3D(X) < X < E(X) + 3D(X)) \geq 1 - \frac{1}{9} = 0,8.$$

Con probabilità almeno pari all'89% ci si attende dunque che la v.c. assuma valori compresi in un intervallo avente centro nel valore medio e semidimensione pari al triplo della sua deviazione standard. Questa affermazione estende dunque la "legge del 3σ " a vv.cc. non gaussiane. Dalla disuguaglianza di Chebyshev segue anche che se $D(X) \rightarrow 0$ risulta $\mathbb{P}(|X - E(X)| < \epsilon) \rightarrow 1$ per ogni scelta di ϵ . La differenza con il caso gaussiano consiste sostanzialmente nella circostanza che nella (8.22) compare il segno di disuguaglianza.

Grazie alla disuguaglianza di Chebyshev, tramite la varianza è possibile determinare un limite inferiore alla probabilità della dispersione dei valori di un'arbitraria v.c. intorno al suo valor medio.

8.2.1 Un caso particolare

A titolo di esercizio ricaviamo direttamente la disuguaglianza di Chebyshev in un caso particolare. Siano X e Y vv.cc. discrete, definite al seguente modo:

$$\begin{aligned}\mathbb{P}(X = x_i) &= p_i \quad (i = 1, 2, \dots, n) \\ \mathbb{P}(Y = x_i - E(X)) &= p_i \quad (i = 1, 2, \dots, n).\end{aligned}$$

Il valore medio e la varianza di X sono allora:

$$\begin{aligned}E(X) &= \sum_{i=1}^n p_i x_i \\ D^2(X) &\equiv E\left\{\left[X - E(X)\right]^2\right\} = \sum_{i=1}^n p_i y_i^2,\end{aligned}$$

dove $y_i = x_i - E(X)$. A condizione di effettuare un cambiamento di indici è sempre possibile ottenere che le y_i siano ordinate in modo tale da essere crescenti in modulo:

$$|y_1| < |y_2| < \dots < |y_n|.$$

Per ogni fissato $\epsilon > 0$ consideriamo l'intervallo $I_\epsilon = [E(X) - \epsilon, E(X) + \epsilon]$ e sia $p = \mathbb{P}(X \in I_\epsilon)$. Sia inoltre $k \in \{0, 1, \dots, n\}$ quell'intero tale per il quale risulta:

$$\begin{aligned}0 < |y_i| < \epsilon, \quad &(i = 1, 2, \dots, k), \\ |y_i| \geq \epsilon, \quad &(i = k + 1, k + 2, \dots, n).\end{aligned}$$

8.3 Il teorema di Bernoulli

Si ha:

$$D^2(X) = \sum_{i=1}^n p_i y_i^2 = \sum_{i=1}^k p_i y_i^2 + \sum_{i=k+1}^n p_i y_i^2 \geq \sum_{i=k+1}^n p_i y_i^2 \geq \epsilon^2 \sum_{i=k+1}^n p_i.$$

Si ottiene così:

$$\sum_{i=k+1}^n p_i \leq \frac{D^2(X)}{\epsilon^2},$$

e quindi:

$$1 - \sum_{i=k+1}^n p_i \geq 1 - \frac{D^2(X)}{\epsilon^2}.$$

D'altra parte risulta:

$$\begin{aligned}1 - \sum_{i=k+1}^n p_i &= \mathbb{P}(|Y| < \epsilon) = \mathbb{P}(|X - E(X)| < \epsilon) \\ &= \mathbb{P}(E(X) - \epsilon < X < E(X) + \epsilon) \\ &= p.\end{aligned}$$

In definitiva abbiamo ottenuto:

$$\mathbb{P}(X \in I_\epsilon) \geq 1 - \frac{D^2(X)}{\epsilon^2},$$

che è proprio la disuguaglianza di Chebyshev per la v.c. X considerata.

8.3 Il teorema di Bernoulli

Un'importante conseguenza della disuguaglianza di Chebyshev è un celebre risultato dovuto a J. Bernoulli, pubblicato nel 1713 nella sua *Ars Conjectandi*.

Teorema 8.1 (Bernoulli). *Sia Y_n la frequenza con cui si verifica un evento A in n prove ripetute indipendenti e sia p la sua probabilità di occorrenza in ogni prova. Per ogni $\epsilon > 0$ si ha:*

$$\lim_{n \rightarrow \infty} \mathbb{P}(|Y_n - p| < \epsilon) = 1. \quad (8.23)$$

Dimostrazione. Sia X_n la v.c. che rappresenta il numero di successi nelle n prove indipendenti. Essendo X_n una v.c. binomiale, risulta:

$$\begin{aligned}E(X_n) &= np \\ D^2(X_n) &= np(1-p).\end{aligned}$$

La v.c. Y_n di cui all'enunciato del teorema è definita al seguente modo:

$$Y_n = \frac{1}{n} X_n.$$

Essa è dunque tale da aversi:

$$\mathbb{P}\left(Y_n = \frac{k}{n}\right) = \mathbb{P}(X_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (k = 0, 1, \dots, n).$$

Calcoliamo media e varianza di Y_n . Per la media otteniamo:

$$E(Y_n) = E\left(\frac{X_n}{n}\right) = \frac{1}{n} E(X_n) = \frac{1}{n} np = p, \quad (8.24)$$

e per la varianza:

$$\begin{aligned} D^2(Y_n) &= E\left\{\left[\frac{X_n}{n} - E\left(\frac{X_n}{n}\right)\right]^2\right\} = \frac{1}{n^2} E\left\{\left[X_n - E(X_n)\right]^2\right\} \\ &= \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}. \end{aligned} \quad (8.25)$$

La disuguaglianza di Chebyshev (8.20) fornisce allora:

$$\mathbb{P}(|Y_n - p| < \epsilon) \geq 1 - \frac{p(1-p)}{n\epsilon^2}. \quad (8.26)$$

Passando al limite per $n \rightarrow \infty$, dalla (8.26) per ogni $\epsilon > 0$ segue la (8.23). \square

L'evidente analogia tra il teorema di Bernoulli e la legge empirica del caso non induca a ritenerne che il primo costituisca la dimostrazione della seconda. Il teorema è, invece, un'affermazione valida solo nell'ambito della teoria; esso, peraltro, suggerisce che la frequenza empirica di occorrenza di un evento possa costituire, nel caso di prove identiche e indipendenti ripetute un gran numero di volte, una stima della probabilità di occorrenza dell'evento nella singola prova. La validità di questa indicazione è, tuttavia, una conclusione di natura esclusivamente empirica e di gran lunga cronologicamente precedente il teorema di Bernoulli.

8.4 Altre disuguaglianze notevoli

Effettuando alcune scelte particolari della funzione non negativa ψ soddisfacente (8.18), dalla Proposizione 8.1 si ricavano, oltre alla disuguaglianza di Chebyshev, anche altre notevoli disuguaglianze. Queste sono indicate nella seguente proposizione.

8.4 Altre disuguaglianze notevoli

Proposizione 8.3. *Sussistono i seguenti risultati:*

(i) *Sia X una v.c. dotata di momento del second'ordine finito. Per ogni arbitrariamente fissato reale positivo ϵ risulta*

$$\mathbb{P}(|X| \geq \epsilon) \leq \frac{E(X^2)}{\epsilon^2}; \quad (8.27)$$

(ii) *se X è una v.c. dotata di momento assoluto finito β_ν di un ordine $\nu > 0$, si ha:*

$$\mathbb{P}(|X| \geq k(\beta_\nu)^{1/\nu}) \leq \frac{1}{k^\nu}, \quad (8.28)$$

dove k è una costante positiva qualsiasi;

(iii) *siano X una v.c. e c una costante positiva. Se $E(e^{cX})$ esiste, risulta:*

$$\mathbb{P}(X \geq \lambda) \leq \frac{E(e^{cX})}{e^{c\lambda}}, \quad (8.29)$$

con $\lambda \in \mathbb{R}^+$.

Dimostrazione. Per dimostrare la (i) osserviamo che dalla (8.19) segue

$$\mathbb{P}(\psi(X) \geq a) \leq \frac{E[\psi(X)]}{a}. \quad (8.30)$$

Ponendovi $\psi(X) = X^2$ e $a = \epsilon^2$ si ottiene:

$$\mathbb{P}(X^2 \geq \epsilon^2) \leq \frac{E(X^2)}{\epsilon^2},$$

da cui segue la (8.27). Utilizziamo ora la (8.30) ponendovi $\psi(x) = |x|^\nu$. Anche tale funzione è a valori non negativi e soddisfa la condizione (8.18); posto allora $a = k^\nu \beta_\nu$, con $k > 0$ qualsiasi, dalla (8.30) segue

$$\mathbb{P}(|X|^\nu \geq k^\nu \beta_\nu) \leq \frac{E[|X|^\nu]}{k^\nu \beta_\nu} = \frac{\beta_\nu}{k^\nu \beta_\nu} = \frac{1}{k^\nu}$$

dalla quale discende la (8.28), così che la (ii) è dimostrata. Dimostriamo la (iii). In questo caso poniamo:

$$\psi(x) = e^{cx}.$$

Chiaramente si ha ancora $\psi(x) \geq 0$ e inoltre la (8.18) è soddisfatta. Posto $a = e^{c\lambda}$, con $\lambda \in \mathbb{R}$, dalla (8.30) si ottiene:

$$\mathbb{P}(e^{cX} \geq e^{c\lambda}) \leq \frac{E(e^{cX})}{e^{c\lambda}},$$

da cui segue la (8.29). \square

Osserviamo che, in particolare, per $c = 1$ la (8.29) assume la forma:

$$\mathbb{P}(X \geq \lambda) \leq \frac{E(e^X)}{e^\lambda}.$$

Va segnalato che alla (8.27), invece che alla (8.20), taluni autori attribuiscono la denominazione di disuguaglianza di Chebyshev. È peraltro evidente che la (8.27) discende dalla (8.20) quando quest'ultima venga scritta sostituendo la v.c. X con lo scarto $Y = X - E(X)$ di X dal suo valore medio.

Ricaveremo ora un altro tipo di disuguaglianza che conduce ad una interessante interpretazione del valore medio di una v.c. a valori positivi.

Proposizione 8.4 (Markov). *Sia X una v.c. positiva e dotata di valore medio $E(X) = m$ finito.² Risulta :*

$$\mathbb{P}(X > \lambda m) \leq \frac{1}{\lambda}, \quad (8.31)$$

dove λ è una costante positiva arbitraria.

Dimostrazione. Se $\lambda \in (0, 1]$ si ha $1/\lambda \geq 1$ e quindi la (8.31) è certamente verificata. Supponiamo ora che sia $\lambda > 1$. Dalla ipotesi di positività di X segue:

$$m = \int_0^{+\infty} x dF(x) > 0.$$

Possiamo allora scrivere:

$$\begin{aligned} m &= \int_0^{+\infty} x dF(x) \\ &= \int_0^{\lambda m} x dF(x) + \int_{\lambda m}^{+\infty} x dF(x) \geq \int_{\lambda m}^{+\infty} x dF(x) \\ &\geq \lambda m \int_{\lambda m}^{+\infty} dF(x) = \lambda m \mathbb{P}(X > \lambda m), \end{aligned}$$

da cui segue infine la (8.31). \square

La (8.31) è nota in letteratura quale *disuguaglianza di Markov*. Questa non è una disuguaglianza molto fine ma, tenendo conto della generalità delle ipotesi, è molto significativa in quanto fornisce un'interpretazione del valore medio. Invero, grazie alla disuguaglianza di Markov se una v.c. X è positiva risulta individuato un limite superiore alla probabilità di

²Una variabile casuale viene detta positiva quando assume valori positivi con probabilità 1, e quindi se ha misura nulla l'insieme dei punti in cui essa è non positiva.

discostamento dei suoi valori ad esempio da multipli del valore medio. La probabilità che X assuma valori maggiori di λ volte il suo valore medio al crescere di λ tende infatti a zero almeno come $1/\lambda$.

A differenza della disuguaglianza di Chebyshev, la disuguaglianza di Markov coinvolge un limite superiore, e non inferiore, alla probabilità considerata. Osserviamo, infine, che se la v.c. X è negativa si può ottenere una disuguaglianza analoga considerando la v.c. $-X$.

Notiamo che se X è una v.c. di media m e varianza $D^2(X)$, posto $\epsilon^2 = \lambda^2 m^2$, dalla disuguaglianza di Chebyshev (8.20)) segue:

$$\mathbb{P}(|X - m| < \lambda m) \geq 1 - \frac{D^2(X)}{\lambda^2 m^2}$$

e quindi:

$$\mathbb{P}(|X - m| > \lambda m) = \mathbb{P}(X < -(\lambda - 1)m) + \mathbb{P}(X > (\lambda + 1)m) \leq \frac{D^2(X)}{\lambda^2 m^2}$$

che, in un certo senso, è reminiscente della (8.31).

8.5 Funzione di distribuzione della normale standard

Nella tabella a doppia entrata che segue sono riportati valori della funzione di distribuzione normale standard. I valori di x si ottengono sommando quelli della prima colonna con quelli della prima riga. Ad esempio, 1.87 si ottiene sommando 1.8 della prima colonna con 0.07 della prima riga, e in corrispondenza si ottiene $\Phi(1.87) = 0.9693$. Per valori negativi dell'argomento Φ può essere calcolata tramite la relazione $\Phi(-x) = 1 - \Phi(x)$.

<i>x</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	.5474	.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

CAPITOLO 9

Funzioni Caratteristiche

9.1 La funzione caratteristica

Sia X una v.c. definita su uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e sia F la sua funzione di distribuzione.

Definizione 9.1. Si definisce funzione caratteristica (f.c.) della v.c. X la funzione $\varphi: \mathbb{R} \rightarrow \mathbb{C}$ definita per ogni $u \in \mathbb{R}$ al seguente modo:

$$\varphi(u) \equiv E(e^{iuX}) = \int_{-\infty}^{+\infty} e^{iux} dF(x). \quad (9.1)$$

□

Si noti che la f.c. esiste sempre, in quanto l'integrale nella (9.1) è finito avendosi:

$$\left| \int_{-\infty}^{+\infty} e^{iux} dF(x) \right| \leq \int_{-\infty}^{+\infty} |e^{iux}| dF(x) = \int_{-\infty}^{+\infty} dF(x) = 1.$$

In teoria della probabilità ed in molteplici applicazioni la funzione caratteristica gioca un ruolo fondamentale.

Nel caso in cui la v.c. X è assolutamente continua ed ha densità di probabilità f , la (9.1) assume la seguente forma:

$$\varphi(u) = \int_{-\infty}^{+\infty} e^{iux} f(x) dx; \quad (9.2)$$

la f.c. φ è dunque la trasformata di Fourier della densità di probabilità f .

Se X è discreta ed ha distribuzione di probabilità $\{p_k\}$, la (9.1) assume la forma di una serie convergente:

$$\varphi(u) = \sum_k e^{iu x_k} p_k. \quad (9.3)$$

Si noti che facendo uso della formula di Eulero

$$e^{iu x} = \cos ux + i \sin ux,$$

valida per u e x reali, la (9.1) assume la forma:

$$\begin{aligned} \varphi(u) &= \int_{-\infty}^{+\infty} \cos(ux) dF(x) + i \int_{-\infty}^{+\infty} \sin(ux) dF(x) \\ &= E[\cos(uX)] + iE[\sin(uX)] \equiv E(e^{iuX}) \end{aligned}$$

che risulta talora conveniente in quanto evita di dover effettuare integrazioni nel campo complesso.

Proposizione 9.1. La f.c. di una v.c. X gode delle seguenti proprietà:

- (i) $\varphi(0) = 1$;
 - (ii) $|\varphi(u)| \leq 1$;
 - (iii) $\varphi(u)$ è uniformemente continua;
 - (iv) $\varphi(u)$ è hermitiana;
 - (v) $\varphi_{aX+b}(u) = e^{ibu} \varphi_X(au)$ ($a, b \in \mathbb{R}$);
 - (vi) $\varphi(u)$ è definita positiva.
- (9.4)

(Ricordiamo che una funzione $\psi(z): \mathbb{R} \rightarrow \mathbb{C}$ si dice hermitiana se per ogni $z \in \mathbb{R}$ risulta $\psi(-z) = \overline{\psi(z)}$.)

Dimostrazione. La (i) è immediata, avendosi:

$$\varphi(0) = \int_{-\infty}^{+\infty} dF(x) = 1.$$

Anche la (ii) è una conseguenza immediata della definizione di f.c. Infatti abbiamo già visto che risulta:

$$\left| \int_{-\infty}^{+\infty} e^{iu x} dF(x) \right| \equiv |\varphi(u)| \leq 1.$$

9.1 La funzione caratteristica

Dimostriamo ora la (iii). Si deve mostrare che $\forall \epsilon > 0$ esiste una quantità positiva δ tale da aversi $|\varphi(u_1) - \varphi(u_2)| < \epsilon$ per ogni $u_1, u_2 \in \mathbb{R}$ con $|u_1 - u_2| < \delta$. Siano dunque $u_1, u_2 \in \mathbb{R}$ e sia a una costante positiva arbitraria. Si ha:

$$\begin{aligned} |\varphi(u_1) - \varphi(u_2)| &= \left| \int_{-\infty}^{+\infty} e^{iu_1 x} dF(x) - \int_{-\infty}^{+\infty} e^{iu_2 x} dF(x) \right| \\ &\leq \int_{-\infty}^{+\infty} |e^{iu_1 x} - e^{iu_2 x}| dF(x) \\ &= \int_{|x| \leq a} |e^{iu_1 x} - e^{iu_2 x}| dF(x) + \int_{|x| > a} |e^{iu_1 x} - e^{iu_2 x}| dF(x). \end{aligned}$$

Faremo ora uso delle seguenti diseguaglianze:

$$|e^{iu_1 x} - e^{iu_2 x}| \leq |x| \cdot |u_1 - u_2| \quad (9.5)$$

$$|e^{iu_1 x} - e^{iu_2 x}| \leq |e^{iu_1 x}| + |e^{iu_2 x}| = 2. \quad (9.6)$$

Notiamo che la (9.5) esprime la circostanza che la lunghezza della corda congiungente due punti della circonferenza di raggio unitario nel piano complesso è non maggiore della lunghezza dell'arco che li ha per estremi; la (9.6) è invece immediata conseguenza della proprietà triangolare. Facendo dunque uso delle (9.5) e (9.6) si trae:

$$|\varphi(u_1) - \varphi(u_2)| \leq |u_1 - u_2| \int_{|x| \leq a} |x| dF(x) + 2 \int_{|x| > a} dF(x). \quad (9.7)$$

Per ogni prefissato $\epsilon > 0$ scegliamo a in modo da aversi

$$\mathbb{P}(|X| > a) < \epsilon/2.$$

Poiché

$$\int_{|x| \leq a} |x| dF(x) \leq a \int_{|x| \leq a} dF(x) \leq a, \quad (9.8)$$

dalle (9.7) e (9.8) si ottiene:

$$\begin{aligned} |\varphi(u_1) - \varphi(u_2)| &\leq a |u_1 - u_2| + 2 \mathbb{P}(|X| > a) \\ &< a |u_1 - u_2| + \epsilon. \end{aligned} \quad (9.9)$$

Poniamo ora $\delta = \epsilon/a$ e fissiamo $u_1, u_2 \in \mathbb{R}$ con la condizione

$$|u_1 - u_2| < \delta = \frac{\epsilon}{a}. \quad (9.10)$$

Dalle (9.9) e (9.10) segue allora:

$$|\varphi(u_1) - \varphi(u_2)| < 2\epsilon$$

con ϵ reale positivo arbitrario e δ dipendente solo da ϵ . La funzione $\varphi(u)$ è dunque uniformemente continua.

Dimostriamo la (iv). Poiché $e^{-iuX} = \overline{e^{iuX}}$, si ottiene:

$$\varphi(-u) = \int_{-\infty}^{+\infty} e^{-iuX} dF(X) = \int_{-\infty}^{+\infty} \overline{e^{iuX}} dF(X) = \overline{\int_{-\infty}^{+\infty} e^{iuX} dF(X)} = \overline{\varphi(u)},$$

che esprime l'hermitianità della funzione caratteristica.

Siano ora a, b costanti reali arbitrarie; posto $Y = aX + b$ si ha:

$$\varphi_Y(u) = E(e^{iuY}) = E\left[e^{iu(aX+b)}\right] = e^{ibu}E\left(e^{iu(aX)}\right) = e^{ibu}\varphi_X(au),$$

che dimostra la (v).

Dimostriamo infine la (vi). A tale scopo ricordiamo che una funzione $\varphi(u)$ si dice definita positiva se per $(u_1, \dots, u_n) \in \mathbb{R}^n$ e $(z_1, \dots, z_n) \in \mathbb{C}^n$ risulta:

$$\sum_{k=1}^n \sum_{r=1}^n \varphi(u_k - u_r) z_k \bar{z}_r \begin{cases} = 0 & \text{se } z_1 = z_2 = \dots = z_n = 0 \\ > 0 & \text{altrimenti.} \end{cases}$$

Orbene, scelti arbitrariamente $u_1, u_2, \dots, u_n \in \mathbb{R}$ e $z_1, z_2, \dots, z_n \in \mathbb{C}$, si ha:

$$\begin{aligned} \sum_{k=1}^n \sum_{r=1}^n \varphi(u_k - u_r) z_k \bar{z}_r &= \sum_{k=1}^n \sum_{r=1}^n E\left[e^{i(u_k - u_r)X}\right] z_k \bar{z}_r \\ &= E\left[\left(\sum_{k=1}^n e^{iu_k X} z_k\right) \left(\sum_{r=1}^n e^{-iu_r X} \bar{z}_r\right)\right] \end{aligned}$$

dalla quale, in virtù di note proprietà dei numeri complessi, segue:

$$\sum_{k=1}^n \sum_{r=1}^n \varphi(u_k - u_r) z_k \bar{z}_r = E\left\{\left|\sum_{k=1}^n e^{iu_k X} z_k\right|^2\right\}. \quad (9.11)$$

Il valore medio che compare a secondo membro della (9.11) è non negativo, annullandosi se e solo se le costanti z_k sono tutte nulle. Ciò dimostra che $\varphi(u)$ è definita positiva. \square

9.1 La funzione caratteristica

La proprietà (v) risulta particolarmente utile in alcuni casi. Ad esempio, si consideri una v.c. X di valore medio m e varianza σ^2 , e sia $Y = (X - m)/\sigma$ la sua standardizzata, così che $E(Y) = 0$ e $D^2(Y) = 1$. Poiché risulta $Y = aX + b$ con $a = 1/\sigma$ e $b = -m/\sigma$, conoscendo la f.c. di X possiamo, utilizzando la (v), immediatamente risalire alla f.c. di Y , e viceversa ottenere la funzione caratteristica di X a partire da quella di Y .

Nel caso di vv.cc. indipendenti sussiste il seguente teorema:

Teorema 9.1. Se X_1, X_2, \dots, X_n sono vv.cc. indipendenti di funzioni caratteristiche $\varphi_{X_k}(u)$ ($k = 1, 2, \dots, n$), la f.c. della loro somma $Y = X_1 + X_2 + \dots + X_n$ è data dal prodotto delle singole funzioni caratteristiche $\varphi_{X_k}(u)$:

$$\varphi_Y(u) = \prod_{k=1}^n \varphi_{X_k}(u). \quad (9.12)$$

Dimostrazione. Applicando la Definizione 9.1 si ha:

$$\varphi_Y(u) = E(e^{iuY}) = E\left[\exp\left(iu \sum_{k=1}^n X_k\right)\right] = E\left(\prod_{k=1}^n e^{iuX_k}\right)$$

e quindi, per l'ipotesi di indipendenza, segue:

$$\varphi_Y(u) = \prod_{k=1}^n E(e^{iuX_k}) = \prod_{k=1}^n \varphi_{X_k}(u). \quad (9.13) \quad \square$$

Osserviamo che in generale se la f.c. di una v.c. X si può esprimere come prodotto di funzioni caratteristiche, non è detto che X sia somma di vv.cc. indipendenti.

Nel caso di somma di due vv.cc. indipendenti, sussistono le due seguenti Proposizioni, dovute rispettivamente a Cramér e a Raikov, che ci limitiamo ad enunciare.

Proposizione 9.2 (Cramér). Se $\varphi_X(u)$ e $\varphi_Y(u)$ sono funzioni caratteristiche e se risulta

$$\varphi_X(u) \cdot \varphi_Y(u) = e^{-u^2/2},$$

si ha:

$$\varphi_X(u) = e^{imu - \sigma^2 u^2/2}$$

$$\varphi_Y(u) = e^{-imu - (1-\sigma^2)u^2/2},$$

ossia:

$$X = N(m, \sigma^2)$$

$$Y = N(-m, 1 - \sigma^2),$$

con m reale arbitrario e $0 \leq \sigma^2 \leq 1$. \square

Osservazione 9.1. Dalla Proposizione 9.2 discende che se X e Y sono vv.cc. indipendenti e se $Z = X + Y$ è normale, allora X e Y sono individualmente normali. \triangleleft

Proposizione 9.3 (Raikov). Se la somma di due vv.cc. indipendenti ha distribuzione di Poisson, così che

$$\varphi_{X+Y}(u) \equiv \varphi_X(u) \cdot \varphi_Y(u) = \exp\{\lambda(e^{iu} - 1)\}$$

risulta anche

$$\begin{aligned}\varphi_X(u) &= \exp\{\mu(e^{iu} - 1)\} \\ \varphi_Y(u) &= \exp\{(\lambda - \mu)(e^{iu} - 1)\},\end{aligned}$$

ossia X e Y sono individualmente distribuite alla Poisson con parametri rispettivamente dati da μ e $\lambda - \mu$, dove $0 \leq \mu \leq \lambda$.

Osservazione 9.2. Se $Z = X + Y$ ha distribuzione di Poisson con X e Y indipendenti, X e Y risultano individualmente distribuite alla Poisson. \triangleleft

Il Teorema 9.1 risulta spesso utile. Ad esempio, se X_1, X_2, \dots, X_n vv.cc. indipendenti di densità di probabilità $f_{X_1}(x), f_{X_2}(x), \dots, f_{X_n}(x)$, si consideri il problema della determinazione della densità di probabilità della v.c.

$$Z = \sum_{k=1}^n X_k. \quad (9.14)$$

Fissati ad arbitrio i valori delle prime $n-1$ vv.cc., il valore della n -esima è vincolato dalla relazione lineare (9.14). La densità di probabilità $f_Z(z)$ di Z è allora:

$$f_Z(z) = \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2 \dots \int_{-\infty}^{+\infty} dx_{n-1} \prod_{k=1}^{n-1} f_{X_k}(x_k) f_{X_n} \left(z - \sum_{k=1}^{n-1} x_k \right), \quad (9.15)$$

ossia è espressa da un integrale a $n-1$ dimensioni di non semplice calcolo. Se sono, invece, note le funzioni caratteristiche delle vv.cc. X_k , la f.c. della somma (9.14) si può immediatamente esprimere a mezzo della 9.13.

Osserviamo ancora che dalla continuità della f.c. $\varphi(u)$ e dalla circostanza che risulta $\varphi(0) = 1 > 0$ segue che si può determinare un intorno I di $t = 0$ in cui è $\varphi(u) > 0$. Possiamo quindi dare la seguente definizione:

Definizione 9.2. La funzione $\psi: I \rightarrow \mathbb{C}$ definita tramite la posizione

$$\psi(u) = \ln \varphi(u) \quad (9.16)$$

prende il nome di seconda funzione caratteristica. \triangleleft

Dalla (9.16) segue anche:

$$\psi(u) = \ln |\varphi(u)| + i \arg[\varphi(u)].$$

La funzione $\psi(u)$ è anche detta *funzione generatrice dei cumulanti*. Su essa ritorneremo nel prossimo paragrafo.

9.2 Funzione caratteristica e momenti

Sia $(\Omega, \mathcal{F}, \mathbb{P})$ uno spazio di probabilità e sia X una v.c. di funzione di distribuzione F e funzione caratteristica φ . Sussiste il seguente teorema:

Teorema 9.2. Sia X dotata dei primi n momenti intorno all'origine

$$\mu'_r = E(X^r), \quad r = 1, 2, \dots, n.$$

La sua f.c. φ è allora derivabile n volte e si ha:

$$\varphi^{(r)}(0) = i^r \mu'_r, \quad r = 1, 2, \dots, n. \quad (9.17)$$

Dimostrazione. Per ipotesi esistono finiti gli integrali

$$\int_{-\infty}^{+\infty} |x|^r dF(x), \quad r = 1, 2, \dots, n.$$

In particolare, per $r = 1$ risulta:

$$\left| \int_{-\infty}^{+\infty} x e^{ixu} dF(x) \right| \leq \int_{-\infty}^{+\infty} |x| dF(x),$$

così che l'integrale a primo membro è uniformemente convergente. Pertanto si ha:¹

$$\varphi'(u) = \int_{-\infty}^{+\infty} i x e^{ixu} dF(x),$$

da cui si trae:

$$\varphi'(0) = i \int_{-\infty}^{+\infty} x dF(x) = i \mu'_1.$$

Se r è un intero qualsiasi minore o uguale ad n , con lo stesso ragionamento si conclude che l'integrale

$$\int_{-\infty}^{+\infty} x^r e^{ixu} dF(x)$$

¹Qui assumiamo euristicamente che sia consentito il passaggio al limite sotto il segno di integrale.

è uniformemente convergente, così che risulta:

$$\varphi^{(r)}(u) = i^r \int_{-\infty}^{+\infty} x^r e^{ixu} dF(x),$$

ossia:

$$\varphi^{(r)}(0) = i^r \mu'_r, \quad r = 1, 2, \dots, n.$$

□

Il Teorema 9.2 espone la circostanza che se la v.c. X possiede dei momenti, questi possono essere calcolati direttamente tramite la funzione caratteristica. Sussiste anche il seguente teorema che fornisce una condizione sufficiente per la sviluppabilità della f.c. in serie di potenze.

Teorema 9.3. Si supponga che tutti i momenti assoluti della v.c. X esistano e che siano uniformemente limitati:

$$\beta_n = E(|X|^n) = \int_{-\infty}^{+\infty} |x|^n dF(x) < M \quad \forall n \in \mathbb{N}. \quad (9.18)$$

Si ha allora:

$$\varphi(u) = \sum_{k=0}^{\infty} \frac{(iu)^k}{k!} \mu'_k, \quad (9.19)$$

dove $\mu'_0 = 1$ e $\mu'_k = E(X^k)$ ($k = 1, 2, \dots$).

Dimostrazione. Sviluppiamo la funzione $\exp(iux)$ con la formula di Taylor, per un fissato $n \in \mathbb{N}$:

$$e^{iux} = 1 + \frac{iux}{1!} + \frac{(iux)^2}{2!} + \dots + \frac{(iux)^{n-1}}{(n-1)!} + \frac{(iux)^n}{n!} e^{iux\vartheta}, \quad \vartheta \in (0, 1).$$

Si ha allora:

$$\begin{aligned} \varphi(u) &= \int_{-\infty}^{+\infty} e^{iux} dF(x) \\ &= \int_{-\infty}^{+\infty} \left[1 + \frac{iux}{1!} + \frac{(iux)^2}{2!} + \dots + \frac{(iux)^{n-1}}{(n-1)!} + \frac{(iux)^n}{n!} e^{iux\vartheta} \right] dF(x) \\ &= 1 + \frac{iu}{1!} \mu'_1 + \frac{(iu)^2}{2!} \mu'_2 + \dots + \frac{(iu)^{n-1}}{(n-1)!} \mu'_{n-1} + R_n, \end{aligned}$$

con

$$R_n = \frac{(iu)^n}{n!} \int_{-\infty}^{+\infty} e^{iux\vartheta} x^n dF(x).$$

9.2 Funzione caratteristica e momenti

D'altra parte,

$$\begin{aligned} |R_n| &= \frac{|u|^n}{n!} \left| \int_{-\infty}^{+\infty} e^{iux\vartheta} x^n dF(x) \right| \\ &\leq \frac{|u|^n}{n!} \int_{-\infty}^{+\infty} |e^{iux\vartheta}| |x|^n dF(x) \\ &\leq \frac{|u|^n}{n!} \beta_n < \frac{|u|^n}{n!} M, \end{aligned}$$

dove l'ultima diseguaglianza è conseguenza dell'ipotesi (9.18). Poiché si ha che $|u|^n/n! \rightarrow 0$ per $n \rightarrow \infty$, segue:

$$\lim_{n \rightarrow \infty} R_n = 0.$$

Quindi φ è sviluppabile in serie di potenze, ossia sussiste la (9.19). □

Le ipotesi del Teorema 9.3, che consentono di dimostrare la sviluppabilità in serie di potenze della f.c., sono certamente molto forti. Sussiste peraltro il seguente teorema, del quale verrà fatto uso nel seguito, che qui ci limitiamo ad enunciare.

Teorema 9.4. Se $E(X^n)$ esiste finito, si ha:

$$\varphi(u) = \sum_{k=0}^n \frac{(iu)^k}{k!} \mu'_k + h(u), \quad (9.20)$$

dove $\mu'_0 = 1$, $\mu'_k = E(X^k)$ per $k = 1, 2, \dots, n$ e dove $h(u)$ è una funzione tale da aversi:

$$\lim_{u \rightarrow 0} \frac{h(u)}{u^n} = 0.$$

In altri termini, se X possiede i momenti intorno all'origine fino a quelli di ordine n , la sua f.c. è la somma di un polinomio in u di grado n e di una funzione la cui forma non interessa specificare, ma che è del tipo $o(u^n)$, ossia è un infinitesimo di ordine superiore a u^n .

Anche la funzione ψ definita nella (9.16) può essere messa in relazione con i momenti di una v.c. X , come indicato dal teorema che segue.

Teorema 9.5. Se la v.c. X è dotata di momenti assoluti finiti di qualunque ordine, la funzione ψ è sviluppabile in una serie di potenze del tipo

$$\psi(u) = \sum_{j=1}^{\infty} \frac{(iu)^j}{j!} k_j. \quad (9.21)$$

Dimostrazione. In virtù della (9.19) si ha:

$$\psi(u) = \ln \varphi(u) = \ln \left[1 + \frac{iu}{1!} \mu'_1 + \frac{(iu)^2}{2!} \mu'_2 + \dots \right].$$

Utilizzando lo sviluppo

$$\ln(1+u) = \sum_{r=1}^{\infty} \frac{(-1)^{r-1} u^r}{r},$$

legittimo pur di scegliere $|u|$ sufficientemente piccolo, si ottiene:

$$\begin{aligned} \psi(u) &= \sum_{j=1}^{\infty} \frac{(iu)^j}{j!} \mu'_j - \frac{1}{2} \left[\sum_{j=1}^{\infty} \frac{(iu)^j}{j!} \mu'_j \right]^2 + \frac{1}{3} \left[\sum_{j=1}^{\infty} \frac{(iu)^j}{j!} \mu'_j \right]^3 + \\ &\quad - \frac{1}{4} \left[\sum_{j=1}^{\infty} \frac{(iu)^j}{j!} \mu'_j \right]^4 + \dots \\ &= \frac{iu}{1!} \mu'_1 + \frac{(iu)^2}{2!} [\mu'_2 - (\mu'_1)^2] + \frac{(iu)^3}{3!} [\mu'_3 - 3\mu'_1\mu'_2 + 2(\mu'_1)^3] + \\ &\quad + \frac{(iu)^4}{4!} [\mu'_4 - 4\mu'_1\mu'_3 - 3(\mu'_2)^2 + 12\mu'_2(\mu'_1)^2 - 6\mu'_1]^4 + \dots \\ &= \frac{iu}{1!} \mu'_1 + \frac{(iu)^2}{2!} D^2(X) + \frac{(iu)^3}{3!} \mu_3 + \frac{(iu)^4}{4!} (\mu_4 - 3\mu_2^2) + \dots, \end{aligned} \tag{9.22}$$

dove si è fatto uso della (6.16). In conclusione si è ottenuta la (9.21) con i coefficienti k_j specificati tramite la (9.22). \square

I coefficienti k_j sono detti *cumulanti* e sono collegati ai momenti di ordine da 1 a j della v.c. X .

È istruttivo osservare che è possibile effettuare la dimostrazione del Teorema 9.5 anche con un diverso procedimento. Precisamente, consideriamo la v.c. $Y = X - \mu'_1$, così che $X = Y + \mu'_1$. Per la (v) della Proposizione 9.1 si ottiene allora:

$$\varphi_X(u) = e^{iu\mu'_1} \varphi_Y(u).$$

D'altra parte i momenti di Y intorno all'origine coincidono con i momenti centrali di X , così che

$$\begin{aligned} \varphi_X(u) &= e^{iu\mu'_1} \left[1 + \frac{iu}{1!} \mu_1 + \frac{(iu)^2}{2!} \mu_2 + \dots \right] \\ &= e^{iu\mu'_1} \left[1 + \frac{(iu)^2}{2!} \mu_2 + \dots \right], \end{aligned}$$

9.3 Formula di inversione

in quanto $\mu_1 = 0$. Si ha allora:

$$\begin{aligned} \psi_X(u) &= \ln \varphi_X(u) \\ &= iu\mu'_1 + \ln \left[1 + \frac{(iu)^2}{2!} \mu_2 + \frac{(iu)^3}{3!} \mu_3 + \dots \right] \\ &= \frac{iu}{1!} \mu'_1 + \frac{(iu)^2}{2!} \mu_2 + \frac{(iu)^3}{3!} \mu_3 + \frac{(iu)^4}{4!} (\mu_4 - 3\mu_2^2) + \dots, \end{aligned}$$

che coincide con la (9.22).

Osserviamo che l'utilità dell'introduzione della funzione ψ risiede essenzialmente nel fatto che il passaggio ai logaritmi trasforma i prodotti in somme, e quindi semplifica notevolmente taluni calcoli. I primi tre cumulanti rivestono grande importanza per il tipo di informazioni che forniscono sulla relativa v.c. Infatti k_1 coincide con il valore medio che, come sappiamo, nell'analogia meccanica è il baricentro della distribuzione di probabilità [densità di probabilità] della v.c.; k_2 coincide con la varianza, a cui è legata una misura della dispersione intorno al valore medio dei valori assunti dalla v.c.; k_3 , infine, coincide con il momento centrale del terzo ordine, il quale misura il grado di asimmetria della distribuzione di probabilità [densità di probabilità].

Precisiamo meglio quest'ultimo concetto. Si ha per definizione

$$\mu_3 = \int_{-\infty}^{+\infty} [x - E(X)]^3 dF(x).$$

Poiché la funzione integranda è dispari, risulta $\mu_3 = 0$ se la densità di probabilità [distribuzione di probabilità] è simmetrica intorno al valore medio. Risulta invece $\mu_3 > 0$ [$\mu_3 < 0$] se la massa di probabilità è per lo più distribuita su valori maggiori [minori] del valore medio. Il momento del terz'ordine intorno all'origine, μ_3 , è dunque idoneo a misurare il grado di asimmetria della v.c. In realtà, nelle applicazioni si preferisce usare un coefficiente adimensionale definito al seguente modo:

$$\alpha = \frac{\mu_3}{D^3(X)}; \tag{9.23}$$

ad esso si dà il nome di *coefficiente di asimmetria* o *skewness*.

9.3 Formula di inversione

Nel § 9.1 si è visto che data la funzione di distribuzione di una v.c. è, almeno in linea di principio, sempre possibile ottenere la funzione caratteristica. Si può dimostrare anche il viceversa: la funzione di distribuzione di una v.c. X è completamente determinata dalla conoscenza della sua f.c. Cominciamo col dimostrare il seguente teorema.

Teorema 9.6. Sia X una v.c. di funzione di distribuzione F e funzione caratteristica φ . Per ogni coppia $x_1, x_2 \in \mathbb{R}$, con $x_1 < x_2$, risulta:

$$\frac{F(x_2) + F(x_2^-)}{2} - \frac{F(x_1) + F(x_1^-)}{2} = \lim_{c \rightarrow +\infty} \frac{1}{2\pi} \int_{-c}^c \frac{e^{-iu x_1} - e^{-iu x_2}}{iu} \varphi(u) du. \quad (9.24)$$

Dimostrazione. Si ha:

$$\begin{aligned} I &\equiv \int_{-c}^c \frac{e^{-iu x_1} - e^{-iu x_2}}{iu} \varphi(u) du \\ &= \int_{-c}^c \frac{e^{-iu x_1} - e^{-iu x_2}}{iu} \left[\int_{-\infty}^{+\infty} e^{iu x} dF(x) \right] du \\ &= \int_{-\infty}^{+\infty} \left[\int_{-c}^c \frac{e^{iu(x-x_1)}}{iu} du \right] dF(x) - \int_{-\infty}^{+\infty} \left[\int_{-c}^c \frac{e^{iu(x-x_2)}}{iu} du \right] dF(x). \end{aligned} \quad (9.25)$$

D'altra parte risulta:

$$\begin{aligned} \int_{-c}^c \frac{e^{iu(x-x_1)}}{iu} du &= \int_{-c}^0 \frac{e^{iu(x-x_1)}}{iu} du + \int_0^c \frac{e^{iu(x-x_1)}}{iu} du \\ &= \int_0^c \frac{e^{iu(x-x_1)}}{iu} du - \int_0^c \frac{e^{-iu(x-x_1)}}{iu} du. \end{aligned}$$

Si ha pertanto:

$$\begin{aligned} \int_{-c}^c \frac{e^{iu(x-x_1)}}{iu} du &= \int_0^c \frac{e^{iu(x-x_1)} - e^{-iu(x-x_1)}}{iu} du \\ &= \int_0^c \frac{\cos[u(x-x_1)] + i \sin[u(x-x_1)]}{iu} du + \\ &\quad - \int_0^c \frac{\cos[u(x-x_1)] - i \sin[u(x-x_1)]}{iu} du \\ &= 2 \int_0^c \frac{\sin[u(x-x_1)]}{u} du. \end{aligned}$$

Effettuando la sostituzione $y = u(x-x_1)$ si ottiene poi:

$$\int_{-c}^c \frac{e^{iu(x-x_1)}}{iu} du = 2 \int_0^{c(x-x_1)} \frac{\sin y}{y} dy. \quad (9.26)$$

Com'è noto, la funzione $(\sin y)/y$ è oscillante ed assume il suo massimo nel punto $y = 0$ in cui è definibile per continuità e vale 1. Analogamente si trae:

$$\int_{-c}^c \frac{e^{iu(x-x_2)}}{iu} du = 2 \int_0^{c(x-x_2)} \frac{\sin y}{y} dy. \quad (9.27)$$

9.3 Formula di inversione

Facendo uso delle (9.26) e (9.27), la (9.25) diventa:

$$\begin{aligned} I &= 2 \int_{-\infty}^{+\infty} \left[\int_0^{c(x-x_1)} \frac{\sin y}{y} dy - \int_0^{c(x-x_2)} \frac{\sin y}{y} dy \right] dF(x) \\ &= 2 \int_{-\infty}^{+\infty} \left[\int_{c(x-x_2)}^{c(x-x_1)} \frac{\sin y}{y} dy \right] dF(x). \end{aligned}$$

Quindi,

$$\lim_{c \rightarrow +\infty} \frac{1}{2\pi} \int_{-c}^c \frac{e^{-iu x_1} - e^{-iu x_2}}{iu} \varphi(u) du = \int_{-\infty}^{+\infty} \left[\lim_{c \rightarrow +\infty} \frac{1}{\pi} \int_{c(x-x_2)}^{c(x-x_1)} \frac{\sin y}{y} dy \right] dF(x).$$

Facendo uso della nota formula integrale di Eulero

$$\lim_{a,b \rightarrow +\infty} \int_a^b \frac{\sin u}{u} du = \pi,$$

è possibile calcolare l'integrale

$$I^*(x) = \lim_{c \rightarrow +\infty} \frac{1}{\pi} \int_{c(x-x_2)}^{c(x-x_1)} \frac{\sin y}{y} dy.$$

Infatti si ha:

$$I^*(x) = \begin{cases} 1, & \text{per } x_1 < x < x_2 \\ \frac{1}{2}, & \text{per } x = x_1 \text{ oppure } x = x_2, \\ 0, & \text{per } x < x_1 \text{ oppure } x > x_2. \end{cases} \quad (9.28)$$

I primi due valori di I^* si ottengono dalla formula di Eulero; in particolare, il secondo scaturisce in virtù della parità della funzione $(\sin y)/y$, mentre l'ultimo si ottiene invece osservando che per $x < x_1$ [$x > x_2$] i limiti di integrazione tendono entrambi a $-\infty$ [$+\infty$] se $c \rightarrow +\infty$.

D'ora in avanti faremo uso della seguente convenzione: con

$$\int_a^b \eta(x) dF(x), \quad \int_{a^+}^b \eta(x) dF(x), \quad \int_a^{b^-} \eta(x) dF(x), \quad \int_{a^+}^{b^-} \eta(x) dF(x)$$

indicheremo gli integrali estesi rispettivamente agli intervalli

$$[a, b], \quad (a, b], \quad [a, b), \quad (a, b).$$

Si noti che è necessario procedere in questo modo in quanto la v.c. in considerazione è di natura arbitraria, ossia può essere assolutamente continua o discreta. Nel nostro caso si ha dunque:

$$\begin{aligned}\lim_{c \rightarrow +\infty} I &= \int_{-\infty}^{+\infty} I^*(x) dF(x) \\ &= \int_{-\infty}^{x_1} I^*(x) dF(x) + \int_{x_1^+}^{x_2^-} I^*(x) dF(x) + \int_{x_2}^{+\infty} I^*(x) dF(x).\end{aligned}$$

Ricordando che l'integrale di Stieltjes gode delle seguenti proprietà²

$$\begin{aligned}\int_a^b g(x) dF(x) &= \int_{a^+}^b g(x) dF(x) + g(a)[F(a) - F(a^-)] \\ \int_a^b g(x) dF(x) &= \int_a^{b^-} g(x) dF(x) + g(b)[F(b) - F(b^-)],\end{aligned}$$

facendo uso della (9.28) si ricava:

$$\begin{aligned}\lim_{c \rightarrow \infty} \frac{1}{2\pi} \int_{-c}^c \frac{e^{-iu x_1} - e^{-iu x_2}}{iu} \varphi(u) du &= \int_{-\infty}^{+\infty} I^*(x) dF(x) \\ &= \frac{1}{2} [F(x_1) - F(x_1^-)] + \int_{x_1^+}^{x_2^-} dF(x) + \frac{1}{2} [F(x_2) - F(x_2^-)] \\ &= \frac{1}{2} [F(x_1) - F(x_1^-)] + F(x_2^-) - F(x_1) + \frac{1}{2} [F(x_2) - F(x_2^-)] \\ &= \frac{1}{2} [F(x_2) + F(x_2^-)] - \frac{1}{2} [F(x_1) + F(x_1^-)].\end{aligned}$$

Resta così dimostra la validità della (9.24). \square

Un corollario di immediata dimostrazione del Teorema 9.6 è il seguente:

Corollario 9.1. Se x_1 e x_2 sono punti di continuità di F , risulta:

$$F(x_2) - F(x_1) = \lim_{c \rightarrow +\infty} \frac{1}{2\pi} \int_{-c}^c \frac{e^{-iu x_1} - e^{-iu x_2}}{iu} \varphi(u) du. \quad (9.29)$$

Si ha, inoltre, il seguente

²Cfr. § B.2 dell'Appendice.

Corollario 9.2. Data una v.c. X di funzione caratteristica φ , si ha:

$$F(x) = \lim_{\eta \rightarrow -\infty} \lim_{c \rightarrow \infty} \frac{1}{2\pi} \int_{-c}^c \frac{e^{-iu \eta} - e^{-iu x}}{iu} \varphi(u) du \quad (9.30)$$

in ogni punto di continuità della funzione di distribuzione F di X .

Dimostrazione. Poniamo $x_2 = x$ e $x_1 = \eta$ nella (9.24). Poiché

$$\lim_{\eta \rightarrow -\infty} \frac{F(\eta) + F(\eta^-)}{2} = 0,$$

e avendosi

$$\frac{F(x) + F(x^-)}{2} = F(x)$$

in virtù della supposta continuità di F in x segue la (9.30). \square

Si noti che la (9.30) permette di ottenere la funzione di distribuzione F di X in ogni punto di continuità. La funzione di distribuzione è peraltro univocamente determinata una volta nota la funzione caratteristica. Invero, nei punti di discontinuità si conviene di valutare il limite dalla sinistra di $F(x)$.

Corollario 9.3. Data una v.c. X assolutamente continua di f.c. φ assolutamente integrabile su \mathbb{R} , la densità di probabilità di X è l'antitrasformata di Fourier di φ :

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-iu x} \varphi(u) du. \quad (9.31)$$

Dimostrazione. Posto $x_2 = x + h$ e $x_1 = x - h$ ($h > 0$) nella (9.29), si ha:

$$\begin{aligned}\frac{F(x+h) - F(x-h)}{2h} &= \lim_{c \rightarrow \infty} \frac{1}{2\pi} \int_{-c}^c \frac{e^{-iu(x-h)} - e^{-iu(x+h)}}{2hiu} \varphi(u) du \\ &= \lim_{c \rightarrow \infty} \frac{1}{2\pi} \int_{-c}^c e^{-iu x} \varphi(u) \frac{e^{iu h} - e^{-iu h}}{2hiu} du \\ &= \lim_{c \rightarrow \infty} \frac{1}{2\pi} \int_{-c}^c e^{-iu x} \varphi(u) \frac{\sin(uh)}{uh} du.\end{aligned}$$

D'altra parte, risulta:

$$\begin{aligned}\left| \int_{-c}^c e^{-iu x} \varphi(u) \frac{\sin(uh)}{uh} du \right| &\leq \int_{-c}^c |e^{-iu x}| |\varphi(u)| \left| \frac{\sin(uh)}{uh} \right| du \\ &\leq \int_{-\infty}^{+\infty} |\varphi(u)| du.\end{aligned}$$

Poiché φ è assolutamente integrabile, l'integrale

$$\int_{-c}^c e^{-iu} \varphi(u) \frac{\sin(uh)}{uh} dt$$

è uniformemente convergente; si può quindi passare al limite sotto il segno di integrale ottenendo:

$$\lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h} = \lim_{h \rightarrow 0} \lim_{c \rightarrow \infty} \frac{1}{2\pi} \int_{-c}^c e^{-iu} \varphi(u) \frac{\sin uh}{uh} du,$$

ovvero:

$$f(x) = \lim_{c \rightarrow \infty} \frac{1}{2\pi} \int_{-c}^c e^{-iu} \varphi(u) du = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-iu} \varphi(u) du. \quad \square$$

Nel caso di una v.c. assolutamente continua, nota la funzione caratteristica φ la densità di probabilità, che è univocamente determinata dalla (9.31), può talora ottersi utilizzando le numerose esistenti tavole di trasformate di Fourier.

9.4 Esempi di funzioni caratteristiche

In questo paragrafo mostriamo come si ottengono le funzioni caratteristiche di alcune vv.cc. Esamineremo separatamente esempi relativi a vv.cc. discrete e a vv.cc. assolutamente continue.

9.4.1 Funzioni caratteristiche di vv.cc. discrete

Esempio 9.1 (Distribuzione binomiale). Come primo esempio consideriamo il caso di una v.c. X avente distribuzione binomiale in un problema di n prove ripetute. Riferiamoci, dunque, ad un evento che si verifica con probabilità p costante in ogni prova. Come sappiamo, la probabilità che esso si verifichi k volte nelle n prove è dunque:

$$b(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k} \quad (k = 0, 1, \dots, n). \quad (9.32)$$

Calcoliamo la f.c. utilizzando la definizione (9.1). Abbiamo:

$$\begin{aligned} \varphi(u) &= E(e^{iuX}) = \sum_{k=0}^n e^{iuk} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^n \binom{n}{k} (e^{iu} p)^k (1-p)^{n-k} \\ &= [e^{iu} p + (1-p)]^n. \end{aligned} \quad (9.33)$$

9.4 Esempi di funzioni caratteristiche

La (9.33) dà dunque la f.c. della v.c. binomiale nel problema considerato di n prove ripetute. Si noti che questa v.c. ammette certamente momenti finiti in quanto è una v.c. ad un numero finito di valori. Tali momenti possono essere calcolati tramite la funzione caratteristica φ , facendo uso del Teorema 9.2. Ad esempio si ha:

$$\varphi'(u) = n i p e^{iu} [e^{iu} p + (1-p)]^{n-1}.$$

Di qui segue:

$$\varphi'(0) = n i p,$$

ossia:

$$\mu'_1 \equiv \frac{\varphi'(0)}{i} = n p.$$

Analogamente si può procedere per il calcolo dei momenti di ordine superiore. \square

Esempio 9.2 (Distribuzione di Poisson). Ricordiamo che una v.c. X di Poisson è discreta ed assume i valori $0, 1, \dots$ con probabilità

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda} \quad (k = 0, 1, \dots), \quad (9.34)$$

dove λ è una costante positiva. Calcoliamo anche in questo caso la f.c. e dimostriamo che il valore medio di X è dato proprio dalla costante λ . Dalla (9.1) si trae:

$$\begin{aligned} \varphi(u) &= \sum_{k=0}^{\infty} \frac{e^{iu k} \lambda^k}{k!} e^{-\lambda} \\ &= \sum_{k=0}^{\infty} \frac{(e^{iu} \lambda)^k}{k!} e^{-\lambda} \\ &= \exp\{\lambda(e^{iu} - 1)\}. \end{aligned} \quad (9.35)$$

Si ha quindi:

$$\varphi'(u) = \exp\{\lambda(e^{iu} - 1)\} \lambda i e^{iu},$$

da cui segue:

$$\varphi'(0) = \lambda i.$$

In definitiva si ottiene:

$$\mu'_1 \equiv \frac{\varphi'(0)}{i} = \lambda. \quad \square$$

9.4.2 Variabili casuali assolutamente continue

Esempio 9.3 (Densità uniforme). Consideriamo il caso di una v.c. X di densità di probabilità uniforme nell'intervallo $(-a, a)$. Costruiamo la sua densità di probabilità f imponendo che questa sia costante nell'intervallo $(-a, a)$, nulla al di fuori di esso e normalizzata:

$$1 = \int_{-\infty}^{+\infty} f(x) dx = \int_{-a}^a k dx = 2a k.$$

Qui con k si è denotato il valore costante di f in $(-a, a)$. Quindi,

$$f(x) = \begin{cases} \frac{1}{2a}, & \text{per } x \in (-a, a), \\ 0, & \text{altrove.} \end{cases}$$

Dalla definizione di f.c. si trae allora:

$$\begin{aligned} \varphi(u) &= \int_{-a}^a \frac{1}{2a} e^{iux} dx = \frac{1}{2a i u} [e^{iua} - e^{-iua}] \\ &= \frac{1}{2a i u} [\cos(ua) + i \sin(ua) - \cos(ua) + i \sin(ua)] \\ &= \frac{\sin(ua)}{ua}. \end{aligned}$$

Si noti che in questo caso il calcolo del valore medio (e degli altri momenti) risulta più semplice se effettuato per via diretta anziché a mezzo della funzione caratteristica. \triangleleft

Esempio 9.4 (Densità normale). Consideriamo la variabile casuale normale $X = N(m, \sigma^2)$ e calcoliamone la f.c. Per semplicità conviene riferirsi alla v.c. standardizzata di X :

$$Y = N(0, 1) \equiv \frac{X - m}{\sigma}.$$

Dalla relazione

$$X = \sigma Y + m$$

e dalla (9.4) segue:

$$\varphi_X(u) = e^{im u} \varphi_Y(u \sigma). \quad (9.36)$$

Calcoliamo $\varphi_Y(u)$; poiché la densità di probabilità di Y è

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2},$$

si ha:

$$\varphi_Y(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{iuy} e^{-y^2/2} dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp \left\{ -\frac{y^2}{2} + iuy \right\} dy.$$

D'altra parte (trattando euristicamente, com'è consuetudine, l'unità immaginaria alla stregua di una costante) possiamo scrivere:

$$\frac{y^2}{2} - iuy = \frac{1}{2} (y - iu)^2 + \frac{u^2}{2},$$

e quindi:

$$\begin{aligned} \varphi_Y(u) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp \left\{ -\frac{1}{2} (y - iu)^2 - \frac{u^2}{2} \right\} dy \\ &= \frac{e^{-u^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp \left\{ -\frac{1}{2} (y - iu)^2 \right\} dy. \end{aligned}$$

Effettuando la sostituzione $z = y - iu$, otteniamo:

$$\begin{aligned} \varphi_Y(u) &= \frac{e^{-u^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-z^2/2} dz \\ &= e^{-u^2/2}. \end{aligned}$$

In definitiva, dalla (9.36) si ha:

$$\varphi_X(u) = \exp \left\{ imu - \frac{\sigma^2 u^2}{2} \right\}. \quad (9.37)$$

\triangleleft

Esempio 9.5 (Densità di Cauchy). Consideriamo una v.c. X assolutamente continua avente densità di probabilità f di Cauchy (cfr. la (6.7)):

$$f(x) = \frac{1}{\pi(1+x^2)}. \quad (9.38)$$

Dalla definizione di f.c. segue:

$$\varphi(u) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{e^{iux}}{1+x^2} dx. \quad (9.39)$$

L'integrale a secondo membro della (9.39) va calcolato con il metodo dei residui. A tal fine osserviamo preliminarmente che si ha:

$$\begin{aligned} \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{e^{ix}}{1+x^2} dx &= \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{\cos ux}{1+x^2} dx + i \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{\sin ux}{1+x^2} dx \\ &= \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{\cos ux}{1+x^2} dx = \frac{2}{\pi} \int_0^{+\infty} \frac{\cos(|u|x)}{1+x^2} dx \end{aligned} \quad (9.40)$$

in quanto $(\sin ux)/(1+x^2)$ è una funzione dispari, $(\cos ux)/(1+x^2)$ è una funzione pari e $\cos(|u|x) = \cos(-|u|x)$. La funzione

$$f(z) = \frac{\cos(|u|z)}{1+z^2}$$

è intera nel semipiano superiore del piano complesso con l'eccezione del punto $z = i$ che è un polo semplice cui corrisponde il residuo

$$a_{-1} = \frac{e^{-|u|}}{2i}.$$

Il teorema dei residui fornisce allora:

$$\frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{e^{ix}}{1+x^2} dx = \frac{2}{\pi} \int_0^{+\infty} \frac{\cos(|u|x)}{1+x^2} dx = \frac{2}{\pi} \pi i \frac{e^{-|u|}}{2i} = e^{-|u|}. \quad (9.41)$$

In definitiva dalle (9.39) e (9.41) si trae:

$$\varphi(u) = e^{-|u|}. \quad (9.42)$$

Si noti che $\varphi(u)$ è simmetrica rispetto all'asse y e che inoltre risulta:

$$\begin{aligned} \lim_{u \rightarrow 0^-} \varphi'(u) &= \lim_{u \rightarrow 0^-} e^u = 1, \\ \lim_{u \rightarrow 0^+} \varphi'(u) &= \lim_{u \rightarrow 0^+} -e^{-u} = -1. \end{aligned}$$

Quindi $\varphi(u)$ possiede un punto angoloso in $u = 0$ con tangenti sinistra e destra date rispettivamente dalle rette $y = x + 1$ e $y = -x + 1$, e non è quindi dotata di derivate in $u = 0$. Ciò non deve meravigliare in quanto, come sappiamo, la v.c. di Cauchy non possiede momenti finiti di alcun ordine. \square

Passiamo infine in rassegna alcuni esempi indicanti come sia utile talora prendere in esame funzioni caratteristiche di vv.cc. piuttosto che funzioni di distribuzione al fine di evidenziare talune proprietà notevoli.

9.4 Esempi di funzioni caratteristiche

Esempio 9.6. Siano X_1, X_2, \dots, X_n vv.cc. di Cauchy indipendenti e consideriamone la media aritmetica

$$Y = \frac{1}{n} \sum_{i=1}^n X_i$$

con l'intento di determinarne la densità di probabilità. Si noti che se a tal fine si ricorresse alle densità di probabilità delle vv.cc. X_1, \dots, X_n , si presenterebbe il calcolo di un integrale $(n-1)$ -dimensionale. Invece, facendo uso del Teorema 9.1, si ha:

$$\varphi_{(nY)}(u) = \prod_{j=1}^n \varphi_{X_j}(u) = \prod_{j=1}^n e^{-|u|} = e^{-n|u|},$$

e quindi:

$$\varphi_Y(u) = e^{-|u|}$$

che è proprio la f.c. della v.c. di Cauchy. Si trae così la conclusione che la media aritmetica di n variabili casuali di Cauchy mutuamente indipendenti è essa stessa una v.c. di Cauchy. \square

Esempio 9.7. Consideriamo n vv.cc. X_1, X_2, \dots, X_n indipendenti distribuite alla Poisson con parametri $\lambda_1, \lambda_2, \dots, \lambda_n$ e la v.c. somma

$$Y = \sum_{j=1}^n X_j.$$

Si ha:

$$\begin{aligned} \varphi_Y(u) &= \prod_{j=1}^n \varphi_{X_j}(u) = \prod_{j=1}^n \exp\{\lambda_j(e^{iu} - 1)\} \\ &= \exp\left\{\left(\sum_{j=1}^n \lambda_j\right)(e^{iu} - 1)\right\}. \end{aligned}$$

Possiamo quindi affermare che la somma di n variabili casuali indipendenti di Poisson di parametri $\lambda_1, \lambda_2, \dots, \lambda_n$ è una v.c. di Poisson di parametro $\lambda = \lambda_1 + \lambda_2 + \dots + \lambda_n$. \square

Esempio 9.8. Consideriamo il caso di n vv.cc. normali indipendenti di medie m_1, \dots, m_n e varianze $\sigma_1^2, \dots, \sigma_n^2$ rispettivamente. Posto

$$Y = \sum_{j=1}^n X_j$$

si ha:

$$\begin{aligned}\varphi_Y(u) &= \prod_{j=1}^n \varphi_{X_j}(u) = \prod_{j=1}^n \exp \left\{ i u m_j - \frac{u^2}{2} \sigma_j^2 \right\} \\ &= \exp \left\{ i u \sum_{j=1}^n m_j - \frac{u^2}{2} \sum_{j=1}^n \sigma_j^2 \right\}.\end{aligned}$$

Possiamo quindi affermare che la v.c. somma di n vv.cc. normali indipendenti è una normale di valore medio pari alla somma dei valori medi e di varianza pari alla somma delle varianze.

□

CAPITOLO 10

Teoremi Asintotici

10.1 Limite normale della distribuzione binomiale

Nella teoria della probabilità grande rilevanza, sia teorica che applicativa, rivestono alcuni teoremi concernenti le funzioni di distribuzione associate a successioni di vv.cc. ed ai rispettivi limiti. In via preliminare mostriamo come una distribuzione normale si possa riguardare come limite di una successione di distribuzioni binomiali. Sia pertanto X una v.c. a distribuzione binomiale

$$p_k(n) = \binom{n}{k} p^k (1-p)^{n-k} \quad (k = 0, 1, \dots, n), \quad (10.1)$$

di valore medio

$$E(X) = \mu \equiv np \quad (10.2)$$

e varianza

$$D^2(X) = \sigma^2 \equiv np(1-p). \quad (10.3)$$

Sia poi Z la v.c. standard associata ad X :

$$Z = \frac{X - \mu}{\sigma} = \frac{X - np}{\sqrt{np(1-p)}}. \quad (10.4)$$

Sussiste il seguente teorema:

Teorema 10.1. *Indicata con $M_Z(t)$ la funzione generatrice dei momenti della v.c. Z definita dalla (10.4), per ogni t reale si ha:*

$$\lim_{n \rightarrow \infty} M_Z(t) = e^{t^2/2}. \quad (10.5)$$

Al divergere di n la funzione generatrice dei momenti di Z tende dunque alla funzione generatrice dei momenti di una v.c. normale standard.

Dimostrazione. La funzione generatrice dei momenti $M_X(t)$ di una v.c. a distribuzione binomiale si ottiene immediatamente a partire dall'espressione della funzione generatrice di probabilità $G(s)$, ricavata nell'Esempio 7.1, facendo uso delle relazioni (7.21) che legano queste due funzioni. Si ha dunque:

$$M_X(t) = [1 + p(e^t - 1)]^n. \quad (10.6)$$

Dalla (iii) della Proposizione 7.2 segue poi:

$$\begin{aligned} M_Z(t) &= E(e^{tZ}) = E\left[\exp\left\{\frac{t(X-\mu)}{\sigma}\right\}\right] = \exp\left\{-\frac{\mu}{\sigma}t\right\} E\left[\exp\left\{\frac{t}{\sigma}X\right\}\right] \\ &= \exp\left\{-\frac{\mu}{\sigma}t\right\} M_X\left(\frac{t}{\sigma}\right) = \exp\left\{-\frac{\mu}{\sigma}t\right\} [1 + p(e^{t/\sigma} - 1)]^n. \end{aligned} \quad (10.7)$$

Passando ai logaritmi, dalla (10.7) si trae:

$$\begin{aligned} \ln M_Z(t) &= -\frac{\mu}{\sigma}t + n \ln [1 + p(e^{t/\sigma} - 1)] \\ &= -\frac{\mu}{\sigma}t + n \ln \left[1 + p\left(\frac{t}{\sigma} + \frac{1}{2!} \frac{t^2}{\sigma^2} + \frac{1}{3!} \frac{t^3}{\sigma^3} + \dots\right)\right], \end{aligned} \quad (10.8)$$

dove si è sviluppata la funzione $e^{t/\sigma}$ in serie di potenze di t . In virtù della (10.3), per n sufficientemente grande si può ricorrere allo sviluppo

$$\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \dots \quad (|x| < 1),$$

così che la (10.8) diviene:

$$\begin{aligned} \ln M_Z(t) &= -\frac{\mu}{\sigma}t + n \left\{ p\left(\frac{t}{\sigma} + \frac{1}{2!} \frac{t^2}{\sigma^2} + \frac{1}{3!} \frac{t^3}{\sigma^3} + \dots\right) \right. \\ &\quad - \frac{p^2}{2} \left(\frac{t}{\sigma} + \frac{1}{2!} \frac{t^2}{\sigma^2} + \frac{1}{3!} \frac{t^3}{\sigma^3} + \dots \right)^2 \\ &\quad \left. + \frac{p^3}{3} \left(\frac{t}{\sigma} + \frac{1}{2!} \frac{t^2}{\sigma^2} + \frac{1}{3!} \frac{t^3}{\sigma^3} + \dots \right)^3 - \dots \right\}. \end{aligned} \quad (10.9)$$

Raccogliendo le potenze uguali di t e ricordando la (10.3), dalla (10.9) si ottiene:

$$\begin{aligned} \ln M_Z(t) &= (np - np)\frac{t}{\sigma} + \left(\frac{np}{2} - \frac{np^2}{2}\right) \frac{t^2}{\sigma^2} + \left(\frac{np}{6} - \frac{np^2}{2} + \frac{np^3}{3}\right) \frac{t^3}{\sigma^3} + \dots \\ &= \frac{np(1-p)}{2\sigma^2} t^2 + \frac{n(p-3p^2+2p^3)}{6\sigma^3} t^3 + o\left(\frac{n}{\sigma^3}\right). \end{aligned}$$

10.1 Limite normale della distribuzione binomiale

Quindi,

$$\ln M_Z(t) = \frac{t^2}{2} + \frac{n}{n^{3/2}} \frac{p-3p^2+2p^3}{6[p(1-p)]^{3/2}} t^3 + \dots,$$

ossia:

$$\ln M_Z(t) = \frac{t^2}{2} + o\left(\frac{1}{\sqrt{n}}\right). \quad (10.10)$$

Passando al limite per $n \rightarrow \infty$, dalla (10.10) segue infine la (10.5). \square

Il Teorema 10.1 dà informazioni sul comportamento asintotico della v.c. standard. Il limite (10.5), invero, suggerisce la convergenza della v.c. Z ad una v.c. normale standard al crescere di n . Si vedrà in seguito come tale concetto, per ora ancora vago, sia suscettibile di precisazione.

Al fine di applicare il teorema in casi concreti è necessario stabilire per quali valori di n sia legittimo usare l'approssimazione normale per una v.c. binomiale. Si può affermare, su base empirica, che sussiste la seguente "regola del pollice": l'approssimazione normale è soddisfacente se n e p sono tali da aversi

$$np > 5, \quad n(1-p) > 5. \quad (10.11)$$

Notiamo che nell'approssimazione normale è ragionevole effettuare la seguente corrispondenza tra gli interi della binomiale e gli intervalli della gaussiana limite:

$$0 \rightarrow \left(-\frac{1}{2}, \frac{1}{2}\right), \quad 1 \rightarrow \left(\frac{1}{2}, \frac{3}{2}\right), \dots, \quad k \rightarrow \left(k - \frac{1}{2}, k + \frac{1}{2}\right), \quad \dots$$

Esempio 10.1. Valutiamo la probabilità che in 20 lanci di una moneta non truccata si verifichi 12 volte testa:

$$p_{12}(20) = \binom{20}{12} \left(\frac{1}{2}\right)^{20} = \frac{125970}{1048576} = 0.1201. \quad (10.12)$$

D'altra parte, poiché in questo caso risulta

$$np = n(1-p) = 10,$$

in base alla regola empirica (10.11) concludiamo che l'approssimazione gaussiana è applicabile. Utilizzando una densità normale di media $\mu \equiv np = 10$ e varianza $\sigma^2 \equiv np(1-p) = 5$ può dunque scriversi:

$$p_{12}(20) \simeq \int_{11.5}^{12.5} \frac{e^{-(x-10)^2/10}}{\sqrt{2\pi} \sqrt{5}} dx. \quad (10.13)$$

Con il cambiamento di variabile $y = (x - 10)/\sqrt{5}$ la (10.13) diviene:

$$p_{12}(20) \simeq \int_{\frac{1.5}{\sqrt{5}}}^{\frac{2.5}{\sqrt{5}}} \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy = \Phi\left(\frac{2.5}{\sqrt{5}}\right) - \Phi\left(\frac{1.5}{\sqrt{5}}\right),$$

dove Φ denota la funzione di distribuzione della v.c. normale standard. Consultando la tabella posta nel § 8.5 risulta infine:

$$\begin{aligned} \Phi\left(\frac{2.5}{\sqrt{5}}\right) - \Phi\left(\frac{1.5}{\sqrt{5}}\right) &\simeq \Phi(1.1180) - \Phi(0.6708) \\ &\simeq \Phi(1.12) - \Phi(0.67) \\ &\simeq 0.8686 - 0.7486 \\ &= 0.1200. \end{aligned} \quad (10.14)$$

Si noti l'eccellente accordo tra il valore esatto fornito nella (10.12) con il valore approssimato ottenuto nella (10.14). \square

Mediante un procedimento analogo a quello usato per dimostrare il Teorema 10.1, si potrebbe dimostrare il cosiddetto *teorema centrale del limite*. Per il ruolo fondamentale che questo teorema riveste nella teoria della probabilità preferiamo invece fornire una dimostrazione più rigorosa facendo uso di proprietà delle funzioni caratteristiche.

10.2 Convergenza in legge

Assumiamo che X, X_1, X_2, \dots siano vv.cc. definite in un medesimo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, e siano F, F_1, F_2, \dots le funzioni di distribuzione corrispondenti.

Definizione 10.1. La successione $\{X_n\}$ di vv.cc. si dice convergere in distribuzione, o in legge, ad una v.c. X se risulta:

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

per ogni reale x in cui F è continua. Tale convergenza si denota con la scrittura $X_n \xrightarrow{D} X$. \square

Esempio 10.2. Nel caso delle vv.cc. degeneri $X_n = 1/n$ ($n = 1, 2, \dots$) e $X = 0$, la successione $\{X_n\}$ converge a X qualunque sia il senso che si intende attribuire al termine "convergenza". Dalle espressioni delle corrispondenti funzioni di distribuzione

$$F_n(x) = \begin{cases} 0, & \text{per } x < \frac{1}{n}, \\ 1, & \text{per } x \geq \frac{1}{n}, \end{cases}$$

10.2 Convergenza in legge

e

$$F(x) = \begin{cases} 0, & \text{per } x < 0, \\ 1, & \text{per } x \geq 0, \end{cases}$$

segue poi che la convergenza sussiste nei punti di continuità di F , avendosi:

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \quad \text{per } x \neq 0.$$

Non si ha invece convergenza nel punto $x = 0$ ove risulta:

$$\lim_{n \rightarrow \infty} F_n(0) \neq F(0),$$

essendo ivi $F_n(0) = 0$ ($n = 1, 2, \dots$) e $F(0) = 1$. \square

Nei due teoremi che seguono si dimostra che esiste una corrispondenza biunivoca tra convergenza in distribuzione di successioni di vv.cc. e convergenza usuale di successioni di ff.cc. Ciò consente di poter adoperare uno strumento molto efficace, quale è quello costituito dalle ff.cc., per trattare questioni legate alla convergenza di vv.cc.

Teorema 10.2. Siano X, X_1, X_2, \dots vv.cc. diff.cc. $\varphi, \varphi_1, \varphi_2, \dots$ rispettivamente. Se la successione $\{X_n\}$ converge in distribuzione alla v.c. X , per ogni reale u si ha:

$$\lim_{n \rightarrow \infty} \varphi_n(u) = \varphi(u). \quad (10.15)$$

Dimostrazione. Cominciamo col mostrare che nell'ipotesi di convergenza in distribuzione di $\{X_n\}$ ad X , per ogni funzione $g: \mathbb{R} \rightarrow \mathbb{R}$ limitata e continua si ha:

$$\lim_{n \rightarrow \infty} E[g(X_n)] = E[g(X)] \quad (10.16)$$

dove, per semplicità, assumiamo che i valori d'aspettazione $E[g(X_n)]$ e $E[g(X)]$ siano finiti. A tal fine scegliamo tre reali ϵ, a e b , con $\epsilon > 0$ e $a < b$, tali che F sia continua in a e in b e che sussistano le diseguaglianze seguenti:

$$F(a) \leq \epsilon, \quad F(b) \geq 1 - \epsilon. \quad (10.17)$$

Si ha evidentemente:

$$\begin{aligned} E[g(X)] &= \int_{-\infty}^{+\infty} g(x)dF(x) \\ &= \int_{-\infty}^a g(x)dF(x) + \int_a^b g(x)dF(x) + \int_b^{+\infty} g(x)dF(x). \end{aligned} \quad (10.18)$$

Sia poi

$$M = \sup\{|g(x)| : x \in \mathbb{R}\}.$$

Da $g(x) \leq M$ e dalle (10.17) si trae:

$$\left| \int_{-\infty}^a g(x) dF(x) \right| \leq \int_{-\infty}^a |g(x)| dF(x) \leq M F(a) \leq M\epsilon, \quad (10.19)$$

$$\left| \int_b^{+\infty} g(x) dF(x) \right| \leq \int_b^{+\infty} |g(x)| dF(x) \leq M [1 - F(b)] \leq M\epsilon. \quad (10.20)$$

Facendo uso delle (10.19) e (10.20), dalla (10.18) si ottiene:

$$\begin{aligned} \left| E[g(X)] - \int_a^b g(x) dF(x) \right| &= \left| \int_{-\infty}^a g(x) dF(x) + \int_b^{+\infty} g(x) dF(x) \right| \\ &\leq 2M\epsilon. \end{aligned} \quad (10.21)$$

Poiché si è supposto che a e b sono punti di continuità per F e che $X_n \xrightarrow{\mathcal{D}} X$, dalla Definizione 10.1 segue la convergenza della successione $\{F_n\}$ in a e in b :

$$\lim_{n \rightarrow \infty} F_n(a) = F(a), \quad \lim_{n \rightarrow \infty} F_n(b) = F(b). \quad (10.22)$$

In virtù delle (10.22) e della (10.17) esiste un intero n_0 tale da aversi:

$$F_n(a) \leq \epsilon, \quad F_n(b) \geq 1 - \epsilon$$

per ogni $n \geq n_0$. Procedendo in maniera analoga si mostra che per $n \geq n_0$ risulta:

$$\left| E[g(X_n)] - \int_a^b g(x) dF_n(x) \right| \leq 2M\epsilon. \quad (10.23)$$

Facendo uso della diseguaglianza

$$|\alpha| - |\beta| \leq |\alpha - \beta| \leq |\alpha| + |\beta|,$$

valida per ogni coppia α, β di reali, in virtù della (10.21) e della (10.23) si ha poi:

$$\begin{aligned} \left| E[g(X)] - E[g(X_n)] \right| &= \left| \int_a^b g(x) dF(x) - \int_a^b g(x) dF_n(x) \right| \\ &\leq \left| \left\{ E[g(X)] - \int_a^b g(x) dF(x) \right\} - \left\{ E[g(X_n)] - \int_a^b g(x) dF_n(x) \right\} \right| \\ &\leq \left| E[g(X)] - \int_a^b g(x) dF(x) \right| + \left| E[g(X_n)] - \int_a^b g(x) dF_n(x) \right| \\ &\leq 4M\epsilon. \end{aligned} \quad (10.24)$$

10.2 Convergenza in legge

Per $n \geq n_0$, dalla (10.24) si ha quindi:

$$\left| E[g(X)] - E[g(X_n)] \right| \leq \left| \int_a^b g(x) dF(x) - \int_a^b g(x) dF_n(x) \right| + 4M\epsilon. \quad (10.25)$$

Dalla supposta continuità di g in $[a, b]$, per la proprietà di non decrescenza delle funzioni di distribuzione F, F_1, F_2, \dots e dall'ipotesi che $\{F_n\}$ converge a F , segue poi:

$$\lim_{n \rightarrow \infty} \int_a^b g(x) dF_n(x) = \int_a^b g(x) dF(x),$$

e quindi per $n \geq n_1$, dove n_1 è un intero sufficientemente grande:

$$\left| \int_a^b g(x) dF_n(x) - \int_a^b g(x) dF(x) \right| \leq \epsilon. \quad (10.26)$$

Dalle (10.24) e (10.26) per $n \geq \max(n_0, n_1)$ si trae:

$$\left| E[g(X)] - E[g(X_n)] \right| \leq \epsilon(1 + 4M),$$

da cui segue la (10.16). Facendo poi uso della (10.16) con $g(X_n) = \cos(uX_n)$ dapprima, e poi con $g(X_n) = \sin(uX_n)$ si ottiene:

$$\begin{aligned} \lim_{n \rightarrow \infty} E[e^{iuX_n}] &\equiv \lim_{n \rightarrow \infty} E[\cos(uX_n)] + i \lim_{n \rightarrow \infty} E[\sin(uX_n)] \\ &= E[\cos(uX)] + i E[\sin(uX)] = E[e^{iuX}], \end{aligned}$$

e quindi la tesi. \square

Lemma 10.1. Se $\varphi(u)$ è la f.c. di una v.c. X , per ogni reale positivo c si ha:

$$\mathbb{P}\left(|X| \geq \frac{2}{c}\right) \leq \frac{1}{c} \int_{-c}^c [1 - \varphi(u)] du. \quad (10.27)$$

Dimostrazione. Facciamo uso della relazione

$$\frac{1}{c} \int_{-c}^c [1 - \varphi(u)] du = \frac{1}{c} \int_{-c}^c E(1 - e^{iuX}) du = \frac{1}{c} E \left[\int_{-c}^c (1 - e^{iuX}) du \right], \quad (10.28)$$

sussistente in ipotesi molto generali atte a garantire l'esistenza del valore di aspettazione nell'ultima uguaglianza della (10.28) nonché la possibilità di invertire l'operazione di inte-

grazione con quella di media. Notiamo anzitutto che sussiste la seguente diseguaglianza:¹

$$1 - \frac{\sin cX}{cX} \geq \frac{1}{2} \quad \text{per } |X| \geq 2/c. \quad (10.29)$$

Posto allora

$$A = \left\{ \omega \in \Omega : |X(\omega)| \geq \frac{2}{c} \right\} \quad (10.30)$$

e denotata con I_A la sua funzione indicatrice, si ha:

$$\left(1 - \frac{\sin cX}{cX} \right) I_A(X) \geq \frac{1}{2} I_A(X). \quad (10.31)$$

Facendo uso delle (10.29) e (10.31), l'ultimo integrale che compare nella (10.28) per $X \neq 0$ fornisce:

$$\begin{aligned} \int_{-c}^c (1 - e^{iuX}) du &= 2c - \frac{e^{icX} - e^{-icX}}{iX} = 2c \left(1 - \frac{\sin cX}{cX} \right) \\ &\geq 2c \left(1 - \frac{\sin cX}{cX} \right) I_A(X) \geq c I_A(X). \end{aligned} \quad (10.32)$$

Dalla (10.28) e dalla diseguaglianza (10.32), che vale anche per $X = 0$, si trae infine:

$$\frac{1}{c} \int_{-c}^c [1 - \varphi(u)] du \geq \frac{1}{c} E[c I_A(X)] = \mathbb{P}(A),$$

ossia la (10.27). \square

Teorema 10.3. *Siano X_1, X_2, \dots v.v.c. rispettivamente di ff.cc. $\varphi_1, \varphi_2, \dots$. Se esiste una funzione $\varphi(u)$ continua in $u = 0$ tale che per ogni reale u si abbia*

$$\lim_{n \rightarrow \infty} \varphi_n(u) = \varphi(u), \quad (10.33)$$

allora φ è la f.c. di una v.c. X e la successione $\{X_n\}$ converge ad X in distribuzione.

Dimostrazione. Poiché φ è il limite di una successione di funzioni φ_n che in zero valgono 1, si ha $\varphi(0) = 1$. Da ciò e dalla continuità di $\varphi(u)$ in $u = 0$ segue che per ogni $\epsilon > 0$ esiste un reale positivo c tale che per $-c \leq u \leq c$ risulta:

$$|\varphi(u) - 1| < \frac{\epsilon}{2}. \quad (10.34)$$

¹Invero, essendo $\frac{\sin cX}{cX}$ funzione pari di X , basta dimostrare che la diseguaglianza $\frac{\sin cX}{cX} \leq \frac{1}{2}$, che è equivalente alla (10.29), sussiste per $X \geq 2/c$. Ma ciò è certamente vero, avendosi $\frac{1}{2} - \frac{\sin cX}{cX} = \frac{1}{cX} \left[\frac{cX}{2} - \sin cX \right] \geq \frac{1}{cX} [1 - \sin cX] \geq 0$.

10.2 Convergenza in legge

Poiché le ff.cc. $\varphi_n(u)$ sono limitate, la convergenza sussiste anche per i loro integrali:

$$\lim_{n \rightarrow \infty} \int_{-c}^c \varphi_n(u) du = \int_{-c}^c \varphi(u) du. \quad (10.35)$$

Facendo uso della (10.34) esprimente la continuità di $\varphi(u)$ in $u = 0$, si ottiene:

$$\left| \frac{1}{c} \int_{-c}^c [1 - \varphi(u)] du \right| \leq \frac{1}{c} \int_{-c}^c |1 - \varphi(u)| du < \epsilon,$$

mentre per la convergenza degli integrali (10.35) per n sufficientemente grande segue:

$$\left| \frac{1}{c} \int_{-c}^c [1 - \varphi_n(u)] du \right| < \epsilon. \quad (10.36)$$

Dalla (10.15) del Lemma 10.1 scritta per X_n e dalla (10.36) si trae allora:

$$\mathbb{P}\left(|X_n| \geq \frac{2}{c}\right) \equiv \mathbb{P}\left(-X_n \leq -\frac{2}{c} \leq X_n\right) < \epsilon. \quad (10.37)$$

Osserviamo che, denotata con F_n la funzione di distribuzione di X_n , per un n sufficientemente grande risulta:

$$\begin{aligned} \mathbb{P}\left(|X_n| \geq \frac{2}{c}\right) &\equiv \mathbb{P}\left(-X_n \geq \frac{2}{c}\right) + \mathbb{P}\left(X_n \geq \frac{2}{c}\right) \\ &= \mathbb{P}\left(X_n \leq -\frac{2}{c}\right) + 1 - \left[\mathbb{P}\left(X_n \leq \frac{2}{c}\right) - \mathbb{P}\left(X_n = \frac{2}{c}\right) \right] \\ &= F_n\left(-\frac{2}{c}\right) - F_n\left(\frac{2}{c}\right) + 1 - \mathbb{P}\left(X_n = \frac{2}{c}\right), \end{aligned}$$

e quindi:

$$\begin{aligned} F_n\left(\frac{2}{c}\right) - F_n\left(-\frac{2}{c}\right) &= 1 - \mathbb{P}\left(|X_n| \geq \frac{2}{c}\right) - \mathbb{P}\left(X_n = \frac{2}{c}\right) \\ &\geq 1 - \mathbb{P}\left(|X_n| \geq \frac{2}{c}\right) \geq 1 - \epsilon \end{aligned} \quad (10.38)$$

dove l'ultima diseguaglianza è conseguenza della (10.37). Faremo ora uso della (10.38) per dimostrare che esistono una sottosequenza $\{F_{n_j}\}$ di $\{F_n\}$ ed una funzione di distribuzione F tali che $\{F_{n_j}\}$ per $j \rightarrow \infty$ converge ad F in ogni punto di continuità. Invero, per ogni razionale r poniamo

$$G(r) = \lim_{i \rightarrow \infty} F_{n_i}(r) \quad (10.39)$$

e

$$F(x) = \inf_{r>x} \{G(r) : r \text{ razionale}\} \quad (10.40)$$

per ogni reale x . Osserviamo che la (10.40) è una funzione di distribuzione. Infatti si ha $0 \leq G(r) \leq 1$ per ogni razionale r , e $0 \leq F(x) \leq 1$ per ogni x reale. Per mostrare poi che F è continua a destra, osserviamo che dalla sua stessa definizione (10.40) segue che per ogni $\epsilon > 0$ e per ogni x reale esiste un razionale $r > x$ tale che:

$$G(r) < F(x) + \epsilon.$$

Poiché, inoltre, è $F(y) \leq G(r)$ per $y < r$, per $x \leq y < r$ si ricava:

$$F(x) \leq F(y) < F(x) + \epsilon,$$

che dimostra la continuità a destra di $F(x)$. D'altra parte si ha

$$F_n(s) - F_n(-s) > 1 - \epsilon$$

per $\epsilon > 0$, $s > 0$ e per n sufficientemente grande. Per j sufficientemente grande si ha allora:

$$F_{n_j}(r) - F_{n_j}(-s) > 1 - \epsilon.$$

Per ogni razionale $r > s$ risulta così:

$$F_{n_j}(r) \geq F_{n_j}(s) > 1 - \epsilon + F_{n_j}(-s) \geq 1 - \epsilon$$

e quindi:

$$G(r) \geq 1 - \epsilon.$$

Per $x > s$ ne segue:

$$1 \geq F(x) \geq 1 - \epsilon$$

e quindi:

$$\lim_{x \rightarrow +\infty} F(x) = 1.$$

Analogamente si dimostra che per $x < -s$ risulta:

$$0 \leq F(x) \leq \epsilon,$$

così che:

$$\lim_{x \rightarrow -\infty} F(x) = 0.$$

Resta dunque dimostrato che F è una funzione di distribuzione. Mostriamo ora che se x è un punto di continuità si ha:

$$\lim_{j \rightarrow \infty} F_{n_j}(x) = F(x). \quad (10.41)$$

10.2 Convergenza in legge

Se, infatti, r è un razionale e x un suo punto di continuità di F con $r > x$, risulta $F_{n_j}(x) \leq F_{n_j}(r)$ e quindi:

$$\max \lim_{j \rightarrow \infty} F_{n_j}(x) \leq \lim_{j \rightarrow \infty} F_{n_j}(r) = G(r) \leq F(x). \quad (10.42)$$

Se scegliamo un reale y tale che $y < r < x$, dalla disegualanza $F_{n_j}(r) \leq F_{n_j}(x)$ segue:

$$G(r) \leq \min \lim_{i \rightarrow \infty} F_{n_j}(x),$$

e quindi:

$$F(y) \leq \min \lim_{j \rightarrow \infty} F_{n_j}(x).$$

Poiché F è continua in x , si ha:

$$F(x) \leq \min \lim_{j \rightarrow \infty} F_{n_j}(x). \quad (10.43)$$

Dalle (10.42) e (10.43) si trae dunque la (10.41). Osserviamo ora che se φ^* è la f.c. della v.c. X avente funzione di distribuzione F , in virtù del Teorema 10.3 per ogni reale t si ha:

$$\lim_{j \rightarrow \infty} \varphi_{n_j}(t) = \varphi^*(t).$$

Dall'ipotesi (10.33) segue quindi $\varphi(u) = \varphi^*(u)$, così che φ è la funzione caratteristica di X . Rimane da dimostrare che la successione $\{X_n\}$ converge in distribuzione a X . A tal fine supponiamo per assurdo che b sia un reale tale che F sia ivi continua e che risulti

$$\lim_{n \rightarrow \infty} F_n(b) \neq F(b).$$

Poiché la successione $\{F_r(b)\}_r$ è limitata, essa contiene una sottosuccessione $\{F_{m_k}(b)\}_k$ che converge ad un limite diverso da $F(b)$. Ma da una siffatta sottosuccessione, come si è visto, si può estrarre un'ulteriore sottosuccessione $\{F_{m_{r_k}}(b)\}_k$ che converge ad una funzione di distribuzione G in ogni punto di continuità. Se ψ è la f.c. di G , si ha:

$$\psi(u) = \lim_{k \rightarrow \infty} \varphi_{m_{r_k}}(u) = \varphi(u).$$

Ne segue $G = F$; quindi $F_{m_{r_k}} \rightarrow F$ per $k \rightarrow \infty$. Poiché F è continua in b , in conclusione si perviene alla relazione:

$$F(b) = \lim_{k \rightarrow \infty} F_{m_{r_k}}(b) = \lim_{r \rightarrow \infty} F_{m_r}(b) \neq F(b),$$

ossia all'assurdo $F(b) \neq F(b)$. Se ne conclude che la successione delle F_n converge alla distribuzione F . Il teorema resta così dimostrato. \square

Corollario 10.1. Se X, X_1, X_2, \dots sono v.v.c. aventi f.f.c., rispettivamente, $\varphi, \varphi_1, \varphi_2, \dots$ si ha:

$$X_n \xrightarrow{\mathcal{D}} X \quad \text{per } n \rightarrow \infty$$

se e solo se per ogni reale u risulta:

$$\lim_{n \rightarrow \infty} \varphi_n(u) = \varphi(u).$$

Dimostrazione. Segue immediatamente dai Teoremi 10.2 e 10.3. \square

10.3 Legge debole dei grandi numeri

Esamineremo qui il comportamento asintotico della media aritmetica di variabili casuali indipendenti e identicamente distribuite nell'ipotesi in cui esse posseggono valore medio finito. Nulla verrà invece assunto sulla loro varianza, che potrà pertanto non esistere o essere infinita.

Teorema 10.4 (Legge debole dei grandi numeri). Se X_1, X_2, \dots sono v.v.c. indipendenti e identicamente distribuite di valore medio μ finito, per ogni $\epsilon > 0$ si ha:

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - \mu\right| \geq \epsilon\right) = 0.$$

Dimostrazione. Sia $\varphi(u)$ la f.c. delle v.v.c. X_i . Per l'ipotesi di indipendenza di queste, la v.c. $\bar{X}_n = (X_1 + \dots + X_n)/n$ ha f.c.

$$\varphi_{\bar{X}_n}(u) = E\left(e^{iu\bar{X}_n}\right) = E\left[\exp\left\{\frac{iu}{n} \sum_{k=1}^n X_k\right\}\right] = \left[\varphi\left(\frac{u}{n}\right)\right]^n.$$

Dallo sviluppo (9.20) di $\varphi(u)$, per $u \rightarrow 0$ segue:

$$\varphi(u) = 1 + iu[\mu + o(1)],$$

e quindi:

$$\left[\varphi\left(\frac{u}{n}\right)\right]^n = \left[1 + i\frac{u}{n}(\mu + \epsilon_n)\right]^n, \quad (10.44)$$

dove $\epsilon_n \rightarrow 0$ per $n \rightarrow \infty$. Poiché, com'è ben noto, $\lim_{n \rightarrow \infty} (1 + \alpha_n/n)^n = e^\alpha$ se $\lim_{n \rightarrow \infty} \alpha_n = \alpha$, dalla (10.44) si ottiene:

$$\lim_{n \rightarrow \infty} \varphi_{\bar{X}_n}(u) = \lim_{n \rightarrow \infty} \left[\varphi\left(\frac{u}{n}\right)\right]^n = e^{i\mu u}.$$

La f.c. di \bar{X}_n tende dunque alla funzione caratteristica di una v.c. Y che assume valore μ con probabilità 1. Dal Corollario 10.1 segue allora che \bar{X}_n converge in distribuzione alla

variabile casuale $Y \equiv \mu$. Se quindi indichiamo con F_n la funzione di distribuzione di \bar{X}_n e con

$$F_Y(x) = \begin{cases} 0, & \text{per } x < \mu, \\ 1, & \text{per } x \geq \mu, \end{cases} \quad (10.45)$$

la funzione di distribuzione di $Y \equiv \mu$, si ha:

$$\lim_{n \rightarrow \infty} F_n = F_Y.$$

Fissato $\epsilon > 0$ si ha dunque:

$$\begin{aligned} \mathbb{P}(|\bar{X}_n - \mu| \geq \epsilon) &= \mathbb{P}(\bar{X}_n \leq \mu - \epsilon) + \mathbb{P}(\bar{X}_n \geq \mu + \epsilon) \\ &\leq \mathbb{P}(\bar{X}_n \leq \mu - \epsilon) + \mathbb{P}\left(\bar{X}_n > \mu + \frac{1}{2}\epsilon\right) \\ &= F_n(\mu - \epsilon) + \left[1 - F_n\left(\mu + \frac{1}{2}\epsilon\right)\right], \end{aligned}$$

ovvero:

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\bar{X}_n - \mu| \geq \epsilon) \leq F_Y(\mu - \epsilon) + \left[1 - F_Y\left(\mu + \frac{1}{2}\epsilon\right)\right]. \quad (10.46)$$

In virtù della definizione (10.45), il secondo membro della (10.46) è nullo, così che per ogni $\epsilon > 0$ segue:

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\bar{X}_n - \mu| \geq \epsilon) = 0,$$

ossia la tesi. \square

È istruttivo notare come la legge debole dei grandi numeri fornisca come corollario il teorema di Bernoulli. Invero, il numero di successi Z_n in un esperimento di n prove ripetute con probabilità p di successo nella singola prova può esprimersi come somma di n v.v.c. X_i , dove X_i ($i = 1, 2, \dots, n$) denota il numero di successi nella i -esima prova:

$$Z_n = X_1 + X_2 + \dots + X_n.$$

Le X_i sono v.v.c. di Bernoulli, ossia tali da aversi:

$$\mathbb{P}(X_i = 1) = p, \quad \mathbb{P}(X_i = 0) = 1 - p, \quad E(X_i) = p \quad (i = 1, 2, \dots, n).$$

Poiché esse sono indipendenti, identicamente distribuite ed a valore medio p finito, per la legge debole dei grandi numeri risulta:

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - p\right| \geq \epsilon\right) = 0$$

e quindi:

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - p\right| < \epsilon\right) = 1,$$

che coincide con la (8.23).

10.4 Il teorema centrale del limite

Passiamo ora ad enunciare e a dimostrare un fondamentale teorema noto quale *teorema centrale del limite* o *teorema centrale di convergenza*.²

Teorema 10.5 (Teorema centrale del limite). *Se X_1, X_2, \dots sono v.v.c. indipendenti e identicamente distribuite dotate di valore medio μ e varianza $\sigma^2 > 0$ finiti, per ogni reale x si ha:*

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x),$$

dove

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

è la funzione di distribuzione della v.c. normale standard.

Dimostrazione. Siano F_n e φ_n , rispettivamente, la funzione di distribuzione e la f.c. della v.c.

$$Y_n = \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \quad (n = 1, 2, \dots). \quad (10.47)$$

Occorre dimostrare che per ogni reale x risulta:

$$\lim_{n \rightarrow \infty} F_n(x) = \Phi(x).$$

È sufficiente a tal fine mostrare che per $n \rightarrow \infty$ la f.c. $\varphi_n(u)$ tende a $\exp(-u^2/2)$, ossia alla f.c. di una v.c. normale standard. Poiché le v.v.c. $X_1 - \mu, X_2 - \mu, \dots$ sono indipendenti e identicamente distribuite, detta $\varphi(u)$ la loro comune f.c., $[\varphi(u)]^n$ è la f.c. della loro somma $X_1 + \dots + X_n - n\mu$. La funzione caratteristica $\varphi_n(u)$ della somma standardizzata Y_n è quindi $[\varphi(u/(\sigma\sqrt{n}))]^n$. Dallo sviluppo (9.20) di φ , osservando che questa è la f.c. di una v.c. di valore medio 0 e varianza σ^2 , per $u \rightarrow 0$ segue:

$$\varphi(u) = 1 - \frac{1}{2} \sigma^2 u^2 [1 - o(1)],$$

e quindi:

$$\varphi_n(u) = \left[\varphi\left(\frac{u}{\sigma\sqrt{n}}\right) \right]^n = \left[1 - \frac{u^2}{2n} (1 + \epsilon_n) \right]^n, \quad (10.48)$$

dove $\epsilon_n \rightarrow 0$ per $n \rightarrow \infty$. Passando al limite nella (10.48) per $n \rightarrow \infty$ si ottiene così:

$$\lim_{n \rightarrow \infty} \varphi_n(u) = \lim_{n \rightarrow \infty} \left[\varphi\left(\frac{u}{\sigma\sqrt{n}}\right) \right]^n = \exp\left(-\frac{1}{2} u^2\right). \quad \square$$

Il teorema centrale del limite esprime la convergenza in legge della successione delle (10.47) alla normale standard $N(0, 1)$. Una sua frequente utilizzazione è basata sull'osservazione che, dall'enunciato stesso del teorema, discende che per n "grande" e per ogni reale x risulta:

$$\mathbb{P}(X_1 + \dots + X_n \leq x) \simeq \Phi\left(\frac{x - n\mu}{\sigma\sqrt{n}}\right).$$

Tale approssimazione nella pratica risulta legittima per $n \geq 30$.

10.5 Applicazione alle passeggiate aleatorie

Un esempio che indica chiaramente l'importanza anche pratica del teorema centrale del limite è offerto dallo studio del comportamento asintotico, che passiamo a descrivere, della cosiddetta *passeggiata aleatoria semplice*.

Si consideri il moto di una particella puntiforme su di un asse coordinato, sul quale sia stata fissata un'origine, nelle seguenti ipotesi:

(i) all'istante $t = 0$ di una scala discreta di tempi $t = 0, 1, \dots$ la particella occupi la posizione $X_0 = 0$;

(ii) ad ognuno dei suddetti istanti la posizione X_n della particella sia data da

$$X_n = X_{n-1} + Z_n \quad (n = 1, 2, \dots), \quad (10.49)$$

dove Z_1, Z_2, \dots è una successione di v.v.c. indipendenti ed equidistribuite, con

$$\mathbb{P}(Z_n = 1) = p, \quad \mathbb{P}(Z_n = -1) = q, \quad \mathbb{P}(Z_n = 0) = 1 - p - q.$$

La descrizione della passeggiata aleatoria (10.49) richiede il calcolo delle seguenti probabilità:

$$p_{0k}^{(n)} = \mathbb{P}(X_n = k | X_0 = 0) \quad (n = 1, 2, \dots; k = 0, \pm 1, \pm 2, \dots).$$

Orbene, non è difficile dimostrare che risulta

$$p_{0k}^{(n)} = \sum \frac{n!}{k_1! k_2! k_3!} p^{k_1} q^{k_2} (1-p-q)^{k_3}, \quad (10.50)$$

²Spesso si incontra la meno opportuna denominazione di "teorema del limite centrale".

dove la somma va estesa a tutte le possibili scelte di k_1, k_2, k_3 soddisfacenti le seguenti condizioni:

$$k_1 + k_2 + k_3 = n, \quad k_1 - k_2 = k. \quad (10.51)$$

È tuttavia evidente che la quantità a secondo membro della (10.50) mal si presta ad essere calcolata per grandi valori di n .

Incidentalmente, notiamo che per ogni intero j si ha anche:

$$p_{0,k}^{(n)} \equiv \mathbb{P}(X_n = k | X_0 = 0) = \mathbb{P}(X_n = k + j | X_0 = j) \equiv p_{j,k+j}^{(n)}.$$

Di conseguenza la (10.50) fornisce anche la probabilità $p_{j,k+j}^{(n)}$ qualunque sia lo stato iniziale j . Si noti ancora che se risulta $p + q = 1^3$ si ha $p_{0,k}^{(n)} = 0$ quando k e n non sono contemporaneamente pari o dispari. Infatti in questo caso è $k_3 = 0$ così che le (10.51) si riducono a

$$k_1 + k_2 = n, \quad k_1 - k_2 = k,$$

che comportano

$$k_1 = \frac{n+k}{2}, \quad k_2 = \frac{n-k}{2},$$

di modo che k_1 e k_2 sono interi solo se k e n sono entrambi pari o dispari.

Dalla (10.49) per iterazione, ricordando che $X_0 = 0$, si ottiene:

$$X_n = \sum_{i=1}^n Z_i, \quad (10.52)$$

cosicché X_n è la somma di n vv.cc. indipendenti e identicamente distribuite. Posto

$$\begin{aligned} \mu &\equiv E(Z_n) = p - q \\ \sigma^2 &\equiv D^2(Z_n) = E(Z_n^2) - \mu^2 = p + q - \mu^2, \end{aligned}$$

dalla (10.52) si ricava:

$$E(X_n) = n\mu, \quad D^2(X_n) = n\sigma^2.$$

Ricorrendo al teorema centrale del limite concludiamo allora che la successione di vv.cc.

$$Y_n = \frac{X_n - n\mu}{\sqrt{n}\sigma}$$

converge in legge alla v.c. normale standard $N(0, 1)$; per ogni $\epsilon > 0$ esiste dunque un intero n_0 tale che per ogni $n \geq n_0$ risulta:

$$\left| \mathbb{P}(Y_n \leq x) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-z^2/2} dz \right| < \epsilon.$$

³In tal caso la particella ad ogni istante è soggetta a variazioni di posizione pari a ± 1 .

Per n grande possiamo quindi scrivere:

$$\mathbb{P}(Y_n \leq x) \simeq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-z^2/2} dz,$$

ovvero:

$$\begin{aligned} \mathbb{P}\left(\frac{X_n - n\mu}{\sqrt{n}\sigma} \leq x\right) &= \mathbb{P}(X_n \leq n\mu + x\sqrt{n}\sigma) \\ &\simeq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-z^2/2} dz. \end{aligned} \quad (10.53)$$

Posto

$$k = n\mu + x\sqrt{n}\sigma,$$

dalla (10.53) si ricava:

$$\mathbb{P}(X_n \leq k) \simeq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{k-n\mu}{\sqrt{n}\sigma}} e^{-z^2/2} dz. \quad (10.54)$$

La (10.54) consente di fornire immediatamente una risposta ad alcuni interessanti quesiti quali:

- (a) la determinazione della probabilità p_a che per n grande la particella occupi una posizione di ascissa maggiore di un reale a fissato ad arbitrio;
- (b) la determinazione della probabilità p_{ab} che per n grande la particella si trovi all'interno di un preassegnato intervallo $(-b, a)$.

Facendo uso della (10.54) abbiamo invero:

$$p_a = \mathbb{P}(X_n > a) \simeq \frac{1}{\sqrt{2\pi}} \int_{\frac{a-n\mu}{\sqrt{n}\sigma}}^{+\infty} e^{-z^2/2} dz.$$

Si ha quindi:

$$\lim_{n \rightarrow \infty} p_a = \begin{cases} 1, & \text{per } \mu > 0, \\ \frac{1}{2}, & \text{per } \mu = 0, \\ 0, & \text{per } \mu < 0. \end{cases}$$

In altri termini, al crescere di n se risulta $\mu > 0$ (ossia $p > q$), la particella occuperà con certezza posizioni al di sopra di un qualunque livello a preassegnato. Nel caso $\mu < 0$ (ossia $p < q$) la particella, pur di attendere un tempo sufficientemente lungo, verrà trovata con certezza al di sotto di un qualunque livello a assegnato, non importa quanto lontano

dalla sua posizione iniziale. Se, infine, è $\mu = 0$ (ossia $p = q$), asintoticamente la particella si troverà con uguale probabilità al di sopra o al di sotto di un qualsiasi livello a fissato. Queste considerazioni giustificano il termine di "deriva" o "drift" con cui si suole designare il parametro $\mu = p - q$. Osserviamo, infine, che si ha:

$$p_{ab} \equiv \mathbb{P}(-b < X_n < a) \simeq \frac{1}{\sqrt{2\pi}} \int_{\frac{-b-n\mu}{\sqrt{n}\sigma}}^{\frac{a-n\mu}{\sqrt{n}\sigma}} e^{-z^2/2} dz,$$

che chiaramente tende a zero per $n \rightarrow \infty$. In altri termini, pur di attendere un tempo sufficientemente lungo, risulta arbitrariamente piccola la probabilità che la particella resti confinata in un intervallo di ampiezza finita, per grande che questa sia.

L'esempio testé discusso indica l'utilità del teorema centrale del limite quando ci si accontenti di un'analisi qualitativa del comportamento di sistemi del tipo descritto, rinunciando a valutazioni quantitative che richiederebbero il ben poco pratico uso di formule quali (10.50). Va peraltro menzionato che considerare valori di n molto grandi è in realtà naturale e necessario in talune circostanze, come quando la passeggiata aleatoria venga usata per descrivere fenomeni, come il moto browniano, che sono caratterizzati da una "costante di tempo" estremamente piccola rispetto a quella delle misurazioni o delle osservazioni cosicché, ad esempio, tra due successive misure, per ravvicinate che siano, la particella in esame subisce un numero estremamente elevato di urti, e quindi di possibili variazioni della sua posizione.

10.6 Passeggiata aleatoria in presenza di barriere assorbenti

Si è visto in passato come attraverso la funzione generatrice di probabilità sia possibile risalire alla distribuzione di probabilità ed ai momenti mediante un'espansione in serie di potenze e mediante operazioni di derivazione. Tuttavia negli esempi considerati i medesimi risultati possono ottenersi anche con metodi alternativi, così che l'importanza della f.g.p. almeno in quei casi non emerge con particolare evidenza. In questo paragrafo ci riferiremo, invece, ad una situazione ben più complessa nella quale la f.g.p. appare particolarmente utile. Considereremo a tal fine nuovamente il problema della descrizione di una passeggiata aleatoria semplice, ma, a differenza di quanto prima fatto, supporremo che detta passeggiata aleatoria evolva in presenza di barriere che ne limitano lo spazio degli stati disponibile. Precisamente, supponiamo che la particella in moto occupi inizialmente lo stato 0 con $-b < 0 < a$, dove $a > 0$ e $-b < 0$ sono arbitrariamente fissati stati che chiamiamo *stati assorbenti* nel seguente senso: se la particella nel proprio moto raggiunge per la prima volta uno di questi stati, il suo moto termina. Pertanto, la particella continua a muoversi fintanto che essa non viene *assorbita* in uno di questi stati.

Indicheremo con $S_{-b,a}$ l'insieme degli interi relativi maggiori di $-b$ e minori di a ; analogamente $U_{-b,a}$ rappresenterà l'insieme degli interi relativi minori di $-b$ o maggiori di a .

10.6 Passeggiata aleatoria in presenza di barriere assorbenti

Risulta, pertanto, $\mathbb{Z} = S_{-b,a} \cup U_{-b,a} \cup \{a, b\}$. Con X , invece, indicheremo una v.c. discreta tale che $\mathbb{P}(X = -1) = q$, $\mathbb{P}(X = 0) = 1 - p - q$ e $\mathbb{P}(X = 1) = p$.⁴

Consideriamo ora una particella che al tempo iniziale delle osservazioni ($n = 0$) si trovi nello stato 0 e che ad ogni unità di tempo si muova di un passo aleatorio X .

In questo contesto si possono definire tre eventi relativamente al moto della particella:

A: "la particella, dopo un numero finito di passi, viene assorbita nello stato a ";

B: "la particella, dopo un numero finito di passi, viene assorbita nello stato $-b$ ";

C: "il moto non ha mai termine e, pertanto, la particella si muove indefinitamente mantenendo la sua posizione nell'insieme $S_{-b,a}$ ".

È immediato poi rendersi conto che risulta $\mathbb{P}(C) = 0$. Invero, indicata con X_n la posizione occupata dalla particella al tempo n quando essa si muove liberamente in \mathbb{Z} ,⁵ risulta

$$\mathbb{P}(C) \leq \lim_{n \rightarrow \infty} \mathbb{P}(-b \leq X_n \leq a)$$

e la tesi segue dal fatto a noi già noto che $\lim_{n \rightarrow \infty} \mathbb{P}(-b \leq X_n \leq a) = 0$.

Possiamo allora affermare che, quasi certamente, prima o poi la particella sarà assorbita o nello stato $-b$ o nello stato a . Per ottenerlo, ad esempio, la probabilità dell'evento A utilizzeremo l'artificio di considerare come stato iniziale del moto della particella uno qualsiasi degli stati di $S_{-b,a}$. Dopo di ciò, se $j \in S_{-b,a}$ e $n \in \mathbb{N}$, indichiamo con $f_{j,a}^{(n)}$ la probabilità che la particella venga assorbita nello stato a al tempo n sapendo che essa era partita al tempo 0 dallo stato j ,⁶ in simboli

$$f_{j,a}^{(n)} = \mathbb{P}(-b < X_1 < a, \dots, -b < X_{n-1} < a, X_n = a | X_0 = j), \quad j \in S_{-b,a} \text{ e } n \in \mathbb{N}.$$

Con ovvia interpretazione dei vari termini, sussiste la seguente equazione alle differenze:

$$f_{j,a}^{(n)} = pf_{j+1,a}^{(n-1)} + qf_{j-1,a}^{(n-1)} + (1-p-q)f_{j,a}^{(n-1)}, \quad j \in S_{-b,a} \text{ e } n \in \mathbb{N}. \quad (10.55)$$

L'equazione (10.55) è del primo ordine rispetto al tempo n e del secondo ordine rispetto allo stato iniziale j ;⁷ per la sua risoluzione abbiamo, pertanto bisogno, di una condizione iniziale

⁴Ovviamente p e q indicano numeri reali positivi, tali che $p + q \leq 1$.

⁵Ossia senza la presenza degli stati assorbenti.

⁶Si osservi che $f_{j,a}^{(n)}$ può essere anche interpretata come la probabilità che la particella in moto libero, partita al tempo 0 dallo stato j , raggiunga per la prima volta al tempo n lo stato a senza mai essere passata in precedenza per lo stato $-b$.

⁷Essa è classificata come *equazione a ritroso* proprio per il fatto che è lo stato iniziale della particella ad essere riguardato come variabile.

(ossia quando $n = 0$) e di due condizioni alla frontiera (ossia quando $j = -b$ o $j = a$). Tenendo conto del significato che abbiamo attribuito a $f_{j,a}^{(n)}$, imponiamo che si abbia

$$f_{j,a}^{(0)} = \delta_{j,a} = \begin{cases} 0 & \text{se } j \neq a; \\ 1 & \text{se } j = a \end{cases} \quad (10.56)$$

e

$$f_{-b,a}^{(n)} = f_{a,a}^{(n)} = 0, \quad n \in \mathbb{N}. \quad (10.57)$$

Fissato lo stato iniziale $j \in S_{-b,a} \cup \{-b, a\}$ possiamo considerare la v.a. discreta $T_{j,a}$ che con probabilità $f_{j,a}^{(n)}$ rappresenta il tempo nel quale si verifica l'assorbimento nello stato a .⁸ Mostriamo ora che per la sua f.g.p.

$$F_{j,a}(s) := E[s^{T_{j,a}}] = \sum_{n=0}^{\infty} s^n f_{j,a}^{(n)}, \quad -1 \leq s \leq 1, \quad (10.58)$$

sussiste il risultato seguente.

Proposizione 10.1. *Sia $j \in S_{-b,a}$. La f.g.p. $F_{j,a}(s)$ della v.a. $T_{j,a}$ è soluzione dell'equazione alle differenze del secondo ordine (rispetto allo stato iniziale)*

$$F_{j,a}(s) = s[pF_{j+1,a}(s) + qF_{j-1,a}(s) + (1-p-q)F_{j,a}(s)], \quad -1 \leq s \leq 1 \quad (10.59)$$

da risolvere imponendo le condizioni

$$F_{-b,a}(s) = 0 \quad \text{e} \quad F_{a,a}(s) = 1, \quad \forall s \in [-1, 1]. \quad (10.60)$$

Dimostrazione. In primo luogo si osservi che dalla definizione (10.58) e dalle condizioni (10.56) e (10.57) si ottiene immediatamente, per ogni $s \in [-1, 1]$, $F_{-b,a}(s) = 0$ e $F_{a,a}(s) = 1$.

Si moltiplichli ora ambo i membri della (10.55) per s^n e si sommi il risultato ottenuto su tutti gli interi; si ottiene in tal modo:

$$\sum_{n=1}^{\infty} s^n f_{j,a}^{(n)} = p \sum_{n=1}^{\infty} s^n f_{j+1,a}^{(n-1)} + q \sum_{n=1}^{\infty} s^n f_{j-1,a}^{(n-1)} + (1-p-q) \sum_{n=1}^{\infty} s^n f_{j,a}^{(n-1)}. \quad (10.61)$$

Tenendo conto che essendo $j \neq a$ risulta $f_{j,a}^{(0)} = \delta_{j,a} = 0$, si ha

$$\sum_{n=1}^{\infty} s^n f_{j,a}^{(n)} = \sum_{n=0}^{\infty} s^n f_{j,a}^{(n)} = F_{j,a}(s). \quad (10.62)$$

⁸Si osservi che, in generale, tale v.a non è onesta nel senso che la sua massa complessiva di probabilità non è unitaria: l'evento costituito da tutte le traiettorie della passeggiata aleatoria che sono assorbite nello stato $-b$ è quasi impossibile solo se $q = 0$.

Inoltre, con $k \in S_{-b,a} \cup \{-b, a\}$ e con $m = n - 1$, si ottiene

$$\sum_{n=1}^{\infty} s^n f_{k,a}^{(n-1)} = s \sum_{n=1}^{\infty} s^{n-1} f_{k,a}^{(n-1)} = s \sum_{m=0}^{\infty} s^m f_{k,a}^{(m)} = s F_{k,a}(s). \quad (10.63)$$

La (10.59) segue allora dalla (10.61) nella quale bisogna tenere conto della (10.62) nonché della (10.63) scritta per $k = j+1$, $k = j-1$ e $k = j$, rispettivamente. \square

Teorema 10.6. *Sia $j \in S_{-b,a}$. Posto*

$$\lambda_1(s) = \frac{1 - s(1 - p - q) + \sqrt{[1 - s(1 - p - q)]^2 - 4pq s^2}}{2ps} \quad (10.64)$$

$$\lambda_2(s) = \frac{1 - s(1 - p - q) - \sqrt{[1 - s(1 - p - q)]^2 - 4pq s^2}}{2ps}, \quad (10.65)$$

risulta:

$$F_{j,a}(s) = \frac{\lambda_1^{j+b}(s) - \lambda_2^{j+b}(s)}{\lambda_1^{a+b}(s) - \lambda_2^{a+b}(s)}, \quad -1 \leq s \leq 1. \quad (10.66)$$

Dimostrazione. Ricercando le soluzioni dell'equazione alle differenze (10.59) nella forma $\lambda^j(s)$, con $\lambda(s)$ funzione reale, si ottiene l'equazione di secondo grado in $\lambda(s)$

$$ps\lambda^2(s) - [1 - s(1 - p - q)]\lambda(s) + qs = 0 \quad (10.67)$$

le cui soluzioni sono date dalle (10.64) e (10.65). Affinché, queste ultime risultino funzioni reali bisogna però imporre che sia

$$[1 - s(1 - p - q)]^2 - 4pq s^2 > 0 \Leftrightarrow 1 + s^2(1 - p - q)^2 - 2s(1 - p - q) - 4pq s^2 > 0. \quad (10.68)$$

Questa, posto $\beta = (1 - p - q) > 0$ e $\eta = \beta^2 - 4pq = (\beta - 2\sqrt{pq})(\beta + 2\sqrt{pq})$, e ordinando secondo potenze decrescenti in s si scrive anche

$$\eta s^2 - 2\beta s + 1 > 0 \Leftrightarrow \mathbb{P}(s) > 0. \quad (10.69)$$

In riferimento alla disequazione (10.69) sussistono i seguenti risultati:

$$1. \frac{\Delta_P}{4} = \beta^2 - (\beta^2 - 4pq) = 4pq > 0 \Rightarrow \exists s_0 \in \mathbb{R} \text{ e } \exists s_{\pi} \in \overline{\mathbb{R}} : s_0 \neq s_{\pi}, \\ \lim_{s \rightarrow s_{\pi}} \mathbb{P}(s) = 0 \text{ e } \mathbb{P}(s_0) = 0;$$

$$2. s_0 = \frac{\beta - 2\sqrt{pq}}{(\beta - 2\sqrt{pq})(\beta + 2\sqrt{pq})} = \frac{1}{1 - p - q + 2\sqrt{pq}} = \frac{1}{1 - (\sqrt{p} - \sqrt{q})^2} > 0;$$

$$3. s_\pi = \frac{\beta + 2\sqrt{pq}}{(\beta - 2\sqrt{pq})(\beta + 2\sqrt{pq})} = \frac{1}{1 - p - q - 2\sqrt{pq}};$$

$$4. s_0 - 1 = \frac{1}{1 - (\sqrt{p} - \sqrt{q})^2} - 1 = \frac{(\sqrt{p} - \sqrt{q})^2}{1 - (\sqrt{p} - \sqrt{q})^2} = (\sqrt{p} - \sqrt{q})^2 s_0 \Rightarrow s_0 \geq 1;^9$$

5. $s_\pi \cdot s_0 = 1/\eta \Rightarrow s_\pi$ è concorde ad η ;

6. se $\eta < 0$ allora $s_0 + s_\pi = 2\beta/\eta \leq 0 \Rightarrow s_\pi \leq -s_0$;¹⁰

7. se $\eta = 0 \Leftrightarrow 2\sqrt{pq} = \beta$ allora $P(s) > 0 \Leftrightarrow s < \frac{1}{2\beta} = s_0$;

8. se $\eta > 0$ allora $s_0 < s_\pi$.

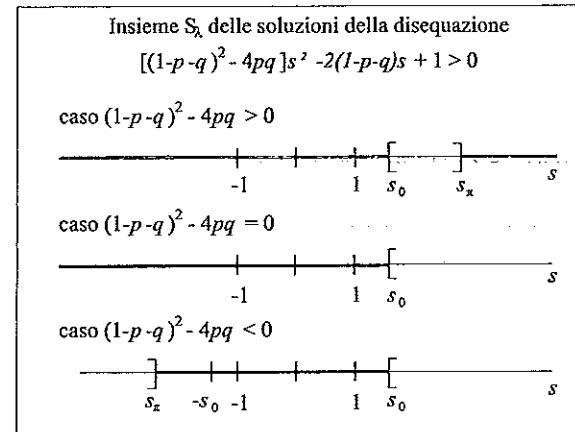


Figura 10.1 – La figura illustra in maniera grafica l'insieme S_λ delle soluzioni della disequazione (10.69) per tutti i possibili casi del coefficiente del suo termine di secondo grado.

La Figura 10.1 riassume graficamente quanto affermato; da essa si evince che per qualunque coppia (p, q) l'insieme S_λ delle soluzioni della disequazione (10.69) contiene l'intervallo $[-1, 1]$. Pertanto per $-1 \leq s \leq 1$ le funzioni $\lambda_1(s)$ e $\lambda_2(s)$ sono entrambe reali così che, in

⁹Risulta $s_0 = 1$ solo nel caso $p = q$.

¹⁰Risulta $s_\pi = -s_0$ solo nel caso $\beta = 0 \Leftrightarrow p + q = 1$.

tale intervallo, la soluzione generale dell'equazione alle differenze (10.59) si scrive come

$$F_{j,a}(s) = \gamma(s)\lambda_1^j(s) + \delta(s)\lambda_2^j(s)$$

con $\gamma(s)$ e $\delta(s)$ funzioni reali arbitrarie definite in $[-1, 1]$. Ricordiamo a questo punto che la soluzione particolare che stiamo ricercando verifica la condizioni sulla frontiera (10.60) che, per ogni $s \in [-1, 1]$, danno luogo al sistema lineare in $\gamma(s)$ e $\delta(s)$

$$\begin{cases} \gamma(s)\lambda_1^{-b}(s) + \delta(s)\lambda_2^{-b}(s) = 0 \\ \gamma(s)\lambda_1^a(s) + \delta(s)\lambda_2^a(s) = 1 \end{cases}$$

dalla cui risoluzione, ottenuta mediante agevoli passaggi algebrici, si ottiene la (10.66). \square

Osservazione 10.1. Risulta

$$\lim_{s \rightarrow 0} \lambda_2(s) = 0; \quad \lim_{s \rightarrow 0} \lambda_1(s) = -\infty; \quad \lim_{s \downarrow 0} \lambda_1(s) = +\infty.$$

Pertanto,

$$F_{j,a}(0) = \lim_{s \rightarrow 0} F_{j,a}(s) = 0.$$

\square

Ciò è d'altronde coerente con il fatto che $F_{j,a}(0) = f_{j,a}^{(0)}$, ossia $F_{j,a}(0)$ coincide con la probabilità che la particella, partendo da $j \neq a$, venga istantaneamente assorbita nello stato a che come abbiamo già visto vale $\delta_{j,a} = 0$.

Osservazione 10.2. Ricordato che S_λ è l'insieme delle soluzioni della disequazione (10.69), il risultato espresso nella (10.66) sussiste, più in generale, per $s \in S_\lambda$. \square

Ritorniamo ora al problema posto all'inizio e cioè alla determinazione della probabilità dell'evento A che la particella partendo dallo stato $j = 0$ venga prima o poi assorbita nello stato a . Per essa sussiste il seguente risultato.

Corollario 10.2. Si ha

$$P(A) = \begin{cases} \frac{1 - (q/p)^b}{1 - (q/p)^{a+b}} & \text{se } p \neq q; \\ \frac{b}{a+b} & \text{se } p = q. \end{cases} \quad (10.70)$$

Dimostrazione. Ponendo $j = 0$ nella (10.66) si ottiene

$$F_{0,a}(s) = \frac{\lambda_1^b(s) - \lambda_2^b(s)}{\lambda_1^{a+b}(s) - \lambda_2^{a+b}(s)}, \quad s \in S_\lambda \supseteq [-1, 1]. \quad (10.71)$$

Di conseguenza, riconoscendo che $f_{0,a}^{(n)}$ rappresenta la probabilità che la particella partendo dallo stato $j = 0$ venga assorbita al passo n nello stato a e ricordando la posizione (10.58) scritta per $j = 0$, si ha

$$\mathbb{P}(A) = \sum_{n=0}^{+\infty} f_{0,a}^{(n)} = F_{0,a}(1) = \frac{\lambda_1^b(1) - \lambda_2^b(1)}{\lambda_1^{a+b}(1) - \lambda_2^{a+b}(1)}. \quad (10.72)$$

Per determinare i due valori $\lambda_1(1)$ e $\lambda_2(1)$ si osservi che la (10.67) scritta per $s = 1$ stabilisce che il loro prodotto e loro somma valgono, rispettivamente, q/p e $1 + q/p$; di conseguenza uno di essi vale 1 e l'altro q/p .¹¹ Pertanto, se $p \neq q$ si ha

$$F_{0,a}(1) = \frac{\lambda_1^b(1) - \lambda_2^b(1)}{\lambda_1^{a+b}(1) - \lambda_2^{a+b}(1)} = \frac{\pm [1 - (q/p)^b]}{\pm [1 - (q/p)^{a+b}]} = \frac{1 - (q/p)^b}{1 - (q/p)^{a+b}},$$

inoltre

$$\lim_{q \rightarrow p} F_{0,a}(1) = \lim_{q \rightarrow p} \frac{1 - (q/p)^b}{1 - (q/p)^{a+b}} = \frac{(b/p)(q/p)^{b-1}}{[(a+b)/p](q/p)^{a+b-1}} = \frac{b}{a+b}$$

e questo completa la dimostrazione della (10.70). \square

Relativamente all'evento B che si verifica quando la particella partendo dallo stato $j = 0$ viene assorbita prima o poi nello stato $-b$ sussiste il seguente risultato.

Corollario 10.3. *Si ha*

$$\mathbb{P}(B) = \begin{cases} \frac{(q/p)^b - (q/p)^{a+b}}{1 - (q/p)^{a+b}} & \text{se } p \neq q; \\ \frac{a}{a+b} & \text{se } p = q. \end{cases} \quad (10.73)$$

Dimostrazione. La dimostrazione è del tutto immediata dal momento che quasi certamente la particella sarà assorbita o nello stato a o nello stato $-b$:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) = 1 \Rightarrow \mathbb{P}(B) = 1 - \mathbb{P}(A). \quad \square$$

Quanto finora visto già indica l'utilità della f.g.p. per fornire indicazioni quantitative sul considerato moto ristretto della passeggiata aleatoria. Tale ruolo verrà ulteriormente sottolineato dalle considerazioni che seguono, miranti alla determinazione esplicita della successione $\{f_{0,a}^{(n)}\}_{n \in \mathbb{N}}$. A tale scopo, dimostreremo che $F_{0,a}(s)$ è una funzione razionale fratta e per essa ricercheremo la forma ridotta ai minimi termini. Negli enunciati che seguono faremo riferimento al contesto e alle posizioni precedenti.

¹¹Dalle (10.64) e (10.65) scritte per $s = 1$ si ricava immediatamente che $\lambda_1(1) \geq \lambda_2(1)$ e questo comporta che $\lambda_1(1)$ vale 1 o q/p a seconda che risulti $p > q$ o $p < q$, rispettivamente. Nel caso $p = q$ si ha $\lambda_1(1) = \lambda_2(1) = 1$.

Proposizione 10.2. *La funzione $F_{0,a}(s)$ risulta essere una funzione razionale fratta con numeratore e denominatore aventi entrambi grado $v < a+b-1$; in particolare il numeratore possiede la radice $s = 0$ con molteplicità a . In simboli*

$$F_{0,a}(s) = (2ps)^a \frac{U(s)}{V(s)}, \quad s \in S_\lambda \supseteq [-1, 1] \quad (10.74)$$

dove $U(s)$ è un polinomio di grado al più $b-1$ e $V(s)$ è un polinomio di grado al più $a+b-1$.

Dimostrazione. Ricordando che abbiamo posto $\beta = 1 - p - q$, per $s \in S_\lambda$, definiamo ora, $\phi(s) = \sqrt{(1 - \beta s)^2 - 4pq s^2}$. Introducendo queste posizioni nelle (10.64) e (10.65) la (10.71) si riscrive come

$$F_{0,a}(s) = \frac{[(1 - \beta s) + \phi(s)]^b - [(1 - \beta s) - \phi(s)]^b}{[(1 - \beta s) + \phi(s)]^{a+b} - [(1 - \beta s) - \phi(s)]^{a+b}} \cdot \frac{(2ps)^{a+b}}{(2ps)^b}.$$

Quest'ultima, sviluppando tutte le potenze con la formula del binomio di Newton, diventa

$$F_{0,a}(s) = (2ps)^a \frac{\sum_{k=0}^b \binom{b}{k} \phi^k(s) (1 - \beta s)^{b-k} [1 - (-1)^k]}{\sum_{k=0}^{a+b} \binom{a+b}{k} \phi^k(s) (1 - \beta s)^{a+b-k} [1 - (-1)^k]}. \quad (10.75)$$

Si osservi ora che nelle sommatorie presenti al numeratore e al denominatore della (10.75), che indichiamo, rispettivamente, con $U(s)$ e $V(s)$, gli addendi di posto pari si annullano. Quindi, dopo aver posto $\bar{b} = \lfloor (b-1)/2 \rfloor$,¹² $\bar{a} = \lfloor (a+b-1)/2 \rfloor$ e semplificato il fattore comune $2\phi(s)$ presente in $U(s)$ e $V(s)$, la (10.76) diventa

$$F_{0,a}(s) = (2ps)^a \frac{U(s)}{V(s)} = (2ps)^a \frac{\sum_{k=0}^{\bar{b}} \binom{b}{2k} \phi^{2k}(s) (1 - \beta s)^{b-2k-1}}{\sum_{k=0}^{\bar{a}} \binom{a+b}{2k+1} \phi^{2k}(s) (1 - \beta s)^{a+b-2k-1}}. \quad (10.76)$$

L'asserto segue dal fatto che, essendo $\phi^2(s)$ un polinomio in s di grado 2, $U(s)$ e $V(s)$ risultano essere polinomi in s di grado $b-1$ e $a+b-1$, rispettivamente.

Nel caso in cui a e b sono entrambi pari sia $U(s)$ che $V(s)$ contengono il fattore comune $(1 - \beta s)$ che pertanto viene cancellato riducendo di un'unità sia il grado di $U(s)$ che quello di $V(s)$. Analogamente si ragiona nel caso di ulteriori radici comuni. \square

¹²Si ricorda che $\lfloor u \rfloor$ restituisce il minimo intero contenuto in $u \in \mathbb{R}$ e pertanto se $u > 0$ allora $\lfloor u \rfloor$ indica la parte intera di u .

Indicato con S l'insieme costituito dalle radici di $V(s)$ che non sono contestualmente anche radici di $U(s)$, nella prossima proposizione si determina un prolungamento della funzione $F_{0,a}(s)$ nell'insieme $\mathbb{R} - (S_\lambda \cup S)$.

Proposizione 10.3. Si ha

$$F_{0,a}(s) = \left(\frac{p}{q} \right)^{a/2} \frac{\sin(b\varphi)}{\sin[(a+b)\varphi]}, \quad s \in \mathbb{R} - (S_\lambda \cup S) \quad (10.77)$$

dove $\varphi \in [0, \pi]$ è collegato ad s tramite la relazione

$$s = \frac{1}{1 - p - q + 2\sqrt{pq}\cos(\varphi)}. \quad (10.78)$$

Dimostrazione. Sia $s \notin S_\lambda$. Dalla (10.68) si ricava che $[1 - s(1 - p - q)]^2 - 4pq s^2 \leq 0$ così che esiste un angolo $\varphi \in [0, \pi]$ tale che

$$[1 - s(1 - p - q)]^2 = 4pq s^2 \cos^2(\varphi) \Leftrightarrow 1 - s(1 - p - q) = 2\sqrt{pq}s \cos(\varphi) \quad (10.79)$$

$$\Leftrightarrow 1 = s[(1 - p - q) + 2\sqrt{pq}\cos(\varphi)].$$

La (10.78) segue immediatamente dall'ultima uguaglianza nella (10.79).¹³ Inoltre, per $s > 0$, dalla (10.64) e dalle varie uguaglianze presenti nella (10.79), segue

$$\begin{aligned} \lambda_1(s) &= \frac{1 - s(1 - p - q) + \sqrt{[1 - s(1 - p - q)]^2 - 4pq s^2}}{2ps} \\ &= \frac{2s\sqrt{pq}\cos(\varphi) + \sqrt{4pq s^2 \cos^2(\varphi) - 4pq s^2}}{2ps} \\ &= \frac{2s\sqrt{pq}\cos(\varphi) + \sqrt{4pq s^2 [\cos^2(\varphi) - 1]}}{2ps} \\ &= \frac{2s\sqrt{pq}\cos(\varphi) + 2s\sqrt{pq}\sqrt{\cos^2(\varphi) - 1}}{2ps} = \sqrt{\frac{q}{p}} \left[\cos(\varphi) + \sqrt{-\sin^2(\varphi)} \right] \end{aligned}$$

e quindi

$$\lambda_1(s) = \sqrt{\frac{q}{p}} [\cos(\varphi) + i\sin(\varphi)] = \left[\sqrt{\frac{q}{p}}, \varphi \right]. \quad (10.80)$$

¹³Come si può verificare a posteriori, con s e φ in relazione mediante la (10.78), risulta $[1 - s(1 - p - q)][s\cos(\varphi)] > 0$ e questo giustifica la prima equivalenza nella (10.79).

Nella (10.80) il simbolo i denota l'unità immaginaria, ossia $i = \sqrt{-1}$, e $[\rho, \theta]$ è la rappresentazione trigonometrica di un generico numero complesso $z = x + iy$.¹⁴ Procedendo allo stesso modo si ricava

$$\lambda_2(s) = \sqrt{\frac{q}{p}} [\cos(\varphi) - i\sin(\varphi)] = \left[\sqrt{\frac{q}{p}}, -\varphi \right]. \quad (10.81)$$

Ponendo le (10.80) e (10.81) nella (10.71), che evidentemente continua a valere nel campo complesso, si ottiene

$$\begin{aligned} F_{0,a}(s) &= \frac{\left[\sqrt{\frac{q}{p}}, \varphi \right]^b - \left[\sqrt{\frac{q}{p}}, -\varphi \right]^b}{\left[\sqrt{\frac{q}{p}}, \varphi \right]^{a+b} - \left[\sqrt{\frac{q}{p}}, -\varphi \right]^{a+b}} \\ &= \left[\left(\sqrt{\frac{q}{p}} \right)^b, b\varphi \right] - \left[\left(\sqrt{\frac{q}{p}} \right)^b, -b\varphi \right] \\ &= \left[\left(\sqrt{\frac{q}{p}} \right)^{a+b}, (a+b)\varphi \right] - \left[\left(\sqrt{\frac{q}{p}} \right)^{a+b}, -(a+b)\varphi \right] \\ &= \frac{\left(\sqrt{\frac{q}{p}} \right)^b [\cos(b\varphi) + i\sin(b\varphi) - \cos(-b\varphi) - i\sin(-b\varphi)]}{\left(\sqrt{\frac{q}{p}} \right)^{a+b} [\cos((a+b)\varphi) + i\sin((a+b)\varphi) - \cos(-(a+b)\varphi) - i\sin(-(a+b)\varphi)]} \\ &= \left(\frac{p}{q} \right)^{a/2} \frac{2i\sin(b\varphi)}{2i\sin((a+b)\varphi)} \end{aligned}$$

e con la cancellazione del fattore $2i$ la (10.77) è dimostrata. Nel caso di $s < 0$ si procede in maniera analoga. \square

Siamo ora in grado di determinare l'insieme S costituito dalle radici di $V(s)$ che non sono contemporaneamente radici di $U(s)$.

Proposizione 10.4. Posto

$$\varphi(V, k) = \frac{k}{a+b}\pi, \quad k \in \{1, 2, \dots, a+b-1\}, \quad (10.82)$$

$$\varphi(U, k) = \frac{k}{b}\pi, \quad k \in \{1, 2, \dots, b-1\} \quad (10.83)$$

¹⁴Si ricorda che risulta $\rho = \sqrt{x^2 + y^2}$ e $\theta = \arctan(y/x)$; viceversa $x = \rho \cos(\theta)$ e $y = \rho \sin(\theta)$. La rappresentazione trigonometrica dei numeri complessi risulta particolarmente vantaggiosa per il calcolo della potenza con esponente reale di un numero complesso: $z^u = [\rho, \theta]^u = [\rho^u, u\theta]$ per qualsiasi $u \in \mathbb{R}$.

e indicati con S_V e S_U l'insieme delle radici di $V(s)$ e di $U(s)$, rispettivamente, si ha

$$S_V = \left\{ s(V, k) = \frac{1}{1 - p - q + 2\sqrt{pq} \cos(\varphi(V, k))} : k \in \{1, 2, \dots, a+b-1\} \right\} \quad (10.84)$$

e

$$S_U = \left\{ s(U, k) = \frac{1}{1 - p - q + 2\sqrt{pq} \cos(\varphi(U, k))} : k \in \{1, 2, \dots, b-1\} \right\}. \quad (10.85)$$

Quindi

$$S = S_V \cup S_U. \quad (10.86)$$

Dimostrazione. Per $s \in \mathbb{R} - (S_\lambda \cup S)$ sussistono sia la (10.76) che la (10.77). Pertanto, imponendo l'uguaglianza dei secondi membri si ottiene

$$U(s) = (2ps)^{-a} \left(\frac{p}{q} \right)^{a/2} \kappa \sin(b\varphi) \quad (10.87)$$

e

$$V(s) = \kappa \sin[(a+b)\varphi] \quad (10.88)$$

dove con κ si è indicato una arbitraria costante reale non nulla. Di qui

$$V(s) = 0 \Leftrightarrow \sin[(a+b)\varphi] = 0 \Leftrightarrow (a+b)\varphi = k\pi$$

dalla quale discende immediatamente la (10.82). La (10.84) segue immediatamente dalla (10.78) e dalla (10.82). Pertanto le radici del polinomio $V(s)$ sono tutte reali, semplici e non appartenenti a S_λ . Da ciò si ricava che $|S_V| = a+b-1$ e al crescere di k da 1 ad $a+b-1$ il denominatore $1 - p - q + 2\sqrt{pq} \cos(\varphi(V, k))$ di $s(V, k)$ decresce strettamente.

In maniera analoga si procede per le radici di $U(s)$ e anche per esso si conclude che le sue radici sono tutte reali, semplici e non appartenenti a S_λ .

Si noti che $\varphi(V, 0) = 0 \Rightarrow s(V, 0) = s_0$ e $\varphi(V, a+b) = \pi \Rightarrow s(V, a+b) = s_\pi$. Pertanto sia s_0 che s_π risultano essere radici di $V(s)$ il che spiega il significato del pedice con il quale abbiamo denotato le radici del polinomio $P(s)$. D'altro canto, procedendo allo stesso modo si ricava che s_0 e s_π sono entrambe radici di $U(s)$. Dal momento che stiamo ricercando la forma ridotta ai minimi termini della funzione razionale fratta $F_{0,a}(s)$, si comprende perché i casi $k=0$ e $k=a+b$ per $V(s)$ e $k=0$ e $k=b$ per $U(s)$ sono state esclusi rispettivamente dalla (10.82) e dalla (10.83). \square

Osservazione 10.3. Come abbiamo già osservato in precedenza la (10.78) fa corrispondere a $\varphi = 0$ e a $\varphi = \pi$, rispettivamente, $s = s_0$ e $s = s_\pi$; inoltre, da $\cos(\varphi(V, 1)) =$

10.6 Passeggiata aleatoria in presenza di barriere assorbenti

$\cos(\pi/(a+b)) \geq 0$ segue che $s(V, 1) > s_0 > 0$. Esiste allora un $m \in \{1, 2, \dots, a+b-1\}$ per il quale

$$s_0 < s(V, 1) < s(V, 2) < \dots < s(V, m) \quad (10.89)$$

e

$$s(V, m+1) < s(V, m+2) < \dots < s(V, a+b-1) < s_\pi. \quad (10.90)$$

Ovviamente $m = a+b-1$ se $s(V, a+b-1) > 0$; tale condizione si verifica in due casi: (i) quando risulta $s_\pi > s_0$ oppure (ii) quando risulta $s_\pi \leq -s_0$ e $1 - p - q + 2\sqrt{pq} \cos(\varphi(V, a+b-1)) > 0$. \triangleleft

Relativamente alla localizzazione delle radici di $V(s)$ sussiste il seguente ulteriore risultato.

Corollario 10.4. *Si ha*

$$\frac{s(V, 1)}{|s(V, a+b-1)|} \leq 1. \quad (10.91)$$

Dimostrazione. La (10.91) è ovvia nei casi (i) e (ii) specificati nell'Osservazione 10.3. Nel caso in cui $s(V, a+b-1) < 0$, posto

$$c = \cos\left(\frac{1}{a+b}\pi\right) = -\cos\left(\frac{a+b-1}{a+b}\pi\right) > 0,$$

dalle (10.82) e (10.84) si ricava

$$\frac{s(V, 1)}{|s(V, a+b-1)|} = \frac{|1 - p - q - 2c\sqrt{pq}|}{|1 - p - q + 2c\sqrt{pq}|} = \frac{2c\sqrt{pq} - (1 - p - q)}{2c\sqrt{pq} + (1 - p - q)} \leq 1$$

che, pertanto, completa la dimostrazione. \square

Risulta evidente dalle (10.82), (10.83), (10.84) e (10.85) che $U(s)$ possiede radici comuni a $V(s)$ quando $a+b$ e b hanno almeno un divisore comune; in particolare quando $a+b$ e b sono entrambi pari, i polinomi $U(s)$ e $V(s)$ possiedono entrambi la radice $1/\beta$. Il risultato seguente, con i suoi immediati corollari, fornisce l'analisi completa per la determinazione di S .

Proposizione 10.5. *Posto $d = MCD(b, a+b)$, si ha*

$$|S_V \cap S_U| = d-1. \quad (10.92)$$

Dimostrazione. Indicato con k' un intero compreso tra 1 e $b-1$ e con k un altro intero compreso tra 1 e $a+b-1$, dalle (10.82), (10.83), (10.84) e (10.85) si evince che

$$\begin{aligned} s(V, k) = s(U, k') &\Leftrightarrow \varphi(V, k) = \varphi(U, k') \Leftrightarrow k = k' \frac{a+b}{b} \\ &\Leftrightarrow k = k' \frac{(a+b)/d}{b/d}. \end{aligned} \quad (10.93)$$

Dall'essere $k' < b$ e dal fatto che il secondo membro dell'ultima uguaglianza deve risultare inferiore, ne discende che i valori di k' che realizzano entrambi i vincoli sono

$$\frac{b}{d}, 2\frac{b}{d}, \dots, (d-1)\frac{b}{d} \quad (10.94)$$

e quindi in numero di $d-1$. \square

Ricordiamo che abbiamo indicato con ν il grado del polinomio denominatore della funzione razionale fratta $F_{0,a}(s)$ ridotta ai minimi termini.

Corollario 10.5. Si ha

$$\nu = a + b - d.$$

Dimostrazione. L'asserto segue immediatamente dalle (10.84), (10.86) e (10.92). \square

Corollario 10.6. I $d-1$ elementi dell'insieme S_V che non sono nell'insieme S sono quelli corrispondenti ai seguenti valori del pedice k :

$$-1 < \frac{a+b}{d}, 2\frac{a+b}{d}, \dots, (d-1)\frac{a+b}{d} < a+b-1. \quad (10.95)$$

Dimostrazione. L'individuazione degli indici segue immediatamente dalla (10.93) e dalla (10.94); la limitazione inferiore segue dal fatto che essendo d il massimo comune divisore tra b e $a+b-1$ risulta $(a+b)/d > 1$. Infine, supponiamo per assurdo che

$$(d-1)\frac{a+b}{d} = a+b-1 \Leftrightarrow \frac{a+b-1}{a+b} = \frac{d-1}{d}$$

per cui deve esistere un intero $m > 1$ per il quale risulta $a+b=md$ e

$$a+b-1 = m(d-1) \Leftrightarrow a+b-1 = md-m \Leftrightarrow a+b-1 = a+b-m \Leftrightarrow m=1.$$

L'assurdo al quale siamo pervenuti dimostra la limitazione superiore. \square

Poniamo, ora, nello stesso ordine con il quale si susseguono gli elementi di S_V ,

$$S = \{s_1, s_2, \dots, s_\nu\}$$

intendendo con questo che, per $i \in \{1, 2, \dots, \nu\}$, $s_i \in S_V$ e $s_i \notin S_U$. Per quanto dimostrato nel Corollario 10.6 risulta $s_1 = s(V, 1)$ e $s_\nu = s(V, a+b-1)$. La formula generale che stabilisce il pedice k dell'elemento di S_V che corrisponde all'elemento di pedice i dell'insieme S è fornita dalla seguente proposizione.

10.6 Passeggiata aleatoria in presenza di barriere assorbenti

Proposizione 10.6. Posto $d = MCD(b, a+b)$, sia $s_i \in S$. Allora $s_i = s[V, k(i)]$ con

$$k(i) = i + \left\lfloor \frac{i-1}{(a+b)/d-1} \right\rfloor, \quad i \in \{1, 2, \dots, \nu\}. \quad (10.96)$$

Dimostrazione. L'asserto segue dalla (10.95): dopo ogni gruppo di $(a+b)/d-1$ di elementi consecutivi di S_V che vengono messi in S ce n'è uno che viene scartato. \square

Possiamo a questo punto riscrivere la funzione razionale fratta $F_{0,a}(s)$ riducendo numeratore e denominatore ai minimi termini:

$$F_{0,a}(s) = (2ps)^a \frac{U(s)}{V(s)} = (2ps)^a \frac{\bar{U}(s)}{\bar{V}(s)} = (2ps)^a \frac{\bar{U}(s)}{c_\nu \prod_{i=1}^\nu (s-s_i)}, \quad s \in \mathbb{R} - S. \quad (10.97)$$

Nella (10.97) abbiamo denotato con la barra sulla stessa lettera i corrispondenti polinomi ottenuti in seguito alle cancellazioni e abbiamo indicato con c_ν il coefficiente di $\bar{V}(s)$ del termine di grado ν .

Proposizione 10.7. Tenendo conto della (10.96) e posto

$$\alpha_i = \frac{-\bar{U}(s_i)}{s_i \bar{V}'(s_i)}, \quad i \in \{1, 2, \dots, \nu\} \quad (10.98)$$

$$= \frac{(-1)^{k(i)+1} 2\sqrt{pq}}{(2p)^a (a+b) s_i^{a-1}} \left(\frac{p}{q}\right)^{a/2} \sin\left(\frac{bk(i)\pi}{a+b}\right) \sin\left(\frac{k(i)\pi}{a+b}\right), \quad i \in \{1, 2, \dots, \nu\} \quad (10.99)$$

si ha

$$F_{0,a}(s) = (2ps)^a \sum_{i=1}^\nu \frac{\alpha_i}{1-s/s_i}, \quad s \in \mathbb{R} - S. \quad (10.100)$$

Dimostrazione. La (10.100) segue immediatamente dal fatto che le radici di $\bar{V}(s)$ sono tutte reali e semplici. Per dimostrare la (10.98) osserviamo in primo luogo che, per qualsiasi $i \in \{1, 2, \dots, \nu\}$, risulta

$$\alpha_i = \lim_{s \rightarrow s_i} \left(1 - \frac{s}{s_i}\right) \sum_{j=1}^\nu \frac{\alpha_j}{1-s/s_j}$$

la quale, tenendo conto della (10.97) e della (10.100), consente di ottenere

$$\begin{aligned} \alpha_i &= \lim_{s \rightarrow s_i} \left(1 - \frac{s}{s_i}\right) \frac{\bar{U}(s)}{\bar{V}(s)} = \lim_{s \rightarrow s_i} \frac{(s_i - s) \bar{U}(s)}{s_i \bar{V}(s)} = \lim_{s \rightarrow s_i} \frac{-(s - s_i) \bar{U}(s)}{s_i c_\nu \prod_{j=1}^\nu (s - s_j)} \\ &= \lim_{s \rightarrow s_i} \frac{-\bar{U}(s)}{s_i c_\nu \prod_{j \neq i; j=1}^\nu (s - s_j)} = \lim_{s \rightarrow s_i} \frac{-\bar{U}(s)}{s_i \bar{V}'(s)} = \frac{-\bar{U}(s_i)}{s_i \bar{V}'(s_i)}. \end{aligned}$$

Per dimostrare la (10.99) osserviamo che per $i \in \{1, 2, \dots, \nu\}$ risulta $\bar{U}(s_i) = U(s_i)$ così che possiamo utilizzare la (10.87) insieme alla (10.82) per ottenere

$$\bar{U}(s_i) = (2ps_i)^{-a} \left(\frac{p}{q}\right)^{a/2} \kappa \sin\left(\frac{bk(i)\pi}{a+b}\right). \quad (10.101)$$

Allo stesso modo per $i \in \{1, 2, \dots, \nu\}$ risulta $\bar{V}(s_i) = V(s_i)$; possiamo quindi utilizzare la (10.88) per ottenere

$$\bar{V}'(s) = (a+b)\kappa \cos[(a+b)\varphi] \frac{d\varphi}{ds} \Leftrightarrow \bar{V}'(s) = (a+b)\kappa \cos[(a+b)\varphi] \frac{1}{ds/d\varphi}$$

e quindi, facendo uso della (10.78),

$$\bar{V}'(s) = \frac{(a+b)\kappa \cos[(a+b)\varphi]}{2\sqrt{pq} \sin(\varphi)s^2}.$$

Pertanto

$$\bar{V}'(s_i) = \frac{(a+b)\kappa \cos(k(i)\pi)}{2\sqrt{pq} \sin[k(i)\pi/(a+b)] s_i^2} = \frac{\kappa(a+b)(-1)^{k(i)}}{2\sqrt{pq} \sin[k(i)\pi/(a+b)] s_i^2}. \quad (10.102)$$

In definitiva, dalle (10.101) e (10.102) segue

$$\frac{-\bar{U}(s_i)}{s_i \bar{V}'(s_i)} = -\frac{1}{s_i} (2ps_i)^{-a} \left(\frac{p}{q}\right)^{a/2} \kappa \sin\left(\frac{bk(i)\pi}{a+b}\right) \frac{2\sqrt{pq} \sin[k(i)\pi/(a+b)] s_i^2}{\kappa(a+b)(-1)^{k(i)}}$$

dalla quale, riordinando e semplificando, si ottiene la (10.99). \square

Siamo ora in grado di determinare la successione $\{f_{0,a}^{(n)}\}_{n \in \mathbb{N}}$, il cui termine generale, si ricordi, rappresenta la ricercata probabilità che la particella partendo dallo stato 0 e in moto ristretto dagli stati $-b$ e a venga assorbita al tempo n nello stato a .

Teorema 10.7. Si ha

$$f_{0,a}^{(n)} = \begin{cases} 0 & \text{se } n < a; \\ (2p)^a \sum_{i=1}^{\nu} \frac{\alpha_i}{s_i^{n-a}} & \text{se } n \geq a \end{cases}$$

dove, per $i \in \{1, 2, \dots, \nu\}$, gli α_i si ottengono dalla (10.99) e gli $s_i = s(V, k(i))$ dalla (10.84) insieme alla (10.96).

10.6 Passeggiata aleatoria in presenza di barriere assorbenti

Dimostrazione. Sia $s \in \mathbb{R}$ tale che $s_0 < s < s(V, 1) = s_1$. Dalle (10.89), (10.90) e (10.91) segue

$$\frac{s}{|s(V, k)|} < 1, \quad k \in \{1, 2, \dots, a+b-1\}$$

e quindi anche

$$\frac{s}{|s_i|} < 1, \quad i \in \{1, 2, \dots, \nu\}.$$

Per un tale s si ha allora

$$\frac{1}{1-s/s_i} = \sum_{r=0}^{\infty} \left(\frac{s}{s_i}\right)^r, \quad i \in \{1, 2, \dots, \nu\}.$$

Dalla (10.100) scritta per $s_0 < s < s(V, 1) = s_1$, segue infine

$$F_{0,a}(s) = (2ps)^a \sum_{i=1}^{\nu} \frac{\alpha_i}{1-s/s_i} = (2ps)^a \sum_{i=1}^{\nu} \alpha_i \sum_{r=0}^{\infty} \left(\frac{s}{s_i}\right)^r = \sum_{r=0}^{\infty} \left[(2p)^a \sum_{i=1}^{\nu} \frac{\alpha_i}{s_i^r} \right] s^{a+r}.$$

La dimostrazione si completa ponendo $n = a+r$. \square

Osservazione 10.4. Dalla (10.99) si evince che la determinazione dei coefficienti α_i potrebbe anche essere direttamente ottenuta per tutti i valori di i compresi tra 1 e $a+b-1$, in quanto per un indice $1 \leq j \leq a+b-1$ per il quale $s(V, j) = s(U, j)$ si ha anche $\sin(bj\pi/(a+b)) = 0$. Sostanzialmente in corrispondenza delle radici comuni al numeratore e al denominatore di $F_{0,a}(s)$ si ottiene un valore nullo per il corrispondente addendo nella (10.100) scritta facendo variare i da 1 ad $a+b-1$. \triangleleft

In conclusione si è così pervenuto alla determinazione della distribuzione di probabilità $f_{0,a}^{(n)}$ attraverso lo sviluppo in serie di potenze della f.g.p. della v.c. $T_{0,a}$ che, per $n \geq a$ ha la seguente espressione

$$f_{0,a}^{(n)} = \frac{2\sqrt{pq}}{(a+b)} \left(\frac{p}{q}\right)^{\frac{n}{2}-a-b-1} \frac{(-1)^{k(i)+1} \sin\left(\frac{bk(i)\pi}{a+b}\right) \sin\left(\frac{k(i)\pi}{a+b}\right)}{\{1-p-q+2\sqrt{pq} \cos(\varphi[V, k(i)])\}^{1-n}} \quad (10.103)$$

con $k(i)$ definito nella Proposizione 10.6.

Ciò conclude il preannunciato esempio di utilizzazione della f.g.p. per determinare esplicitamente una distribuzione di probabilità in un caso caratterizzato da elevata complessità, risultando particolarmente evidenziata la rilevanza e l'efficacia dell'uso di tale funzione.

CAPITOLO 11

Condizionamento nel Continuo

11.1 Variabili casuali unidimensionali

Nel § 4.3 abbiamo definito la probabilità condizionata. Ricordiamo che fissato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e due eventi A, B con $\mathbb{P}(A) > 0$, si dice probabilità condizionata di B dato A , e si indica con $\mathbb{P}(B|A)$, il seguente rapporto:

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}. \quad (11.1)$$

Analogamente, se è $\mathbb{P}(B) > 0$ la probabilità condizionata di A dato B è per definizione:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \quad (11.2)$$

Se è $\mathbb{P}(A) = 0$ (oppure $\mathbb{P}(B) = 0$) non è possibile dare la precedente definizione di probabilità condizionata di B dato A (oppure di A dato B). In tal caso, tuttavia, gli eventi A, B sono “banalmente indipendenti”.

Sia ora X una v.c. con funzione di distribuzione F_X . In vari problemi ha interesse determinare la probabilità di un evento condizionato da eventi del tipo $\{X = x\}$. D'altra parte se X è assolutamente continua si ha $\mathbb{P}(X = x) = 0$, così che definizioni quali (11.1) e (11.2) perdono di significato. Al fine di superare questa difficoltà, procediamo al seguente modo. Dati due numeri reali h, k definiamo l'evento

$$A_{hk} = \{\omega : x - h < X(\omega) \leq x + k\}. \quad (11.3)$$

Scelto allora l'intervallo $(x-h, x+k]$ in modo tale che sia $\mathbb{P}(A_{hk}) > 0$, ha senso considerare la probabilità condizionata di un evento B rispetto ad A_{hk} . Applicando la (11.1) si ha:

$$\mathbb{P}(B|A_{hk}) = \frac{\mathbb{P}(B \cap A_{hk})}{\mathbb{P}(A_{hk})}. \quad (11.4)$$

Inoltre, dalla (11.3) risulta:

$$\mathbb{P}(A_{hk}) = F_X(x+k) - F_X(x-h),$$

così che la (11.4) diviene:

$$\mathbb{P}(B|A_{hk}) = \frac{\mathbb{P}(B \cap A_{hk})}{F_X(x+k) - F_X(x-h)}. \quad (11.5)$$

Definizione 11.1. Data una v.c. X di funzione di distribuzione F_X e scelto A_{hk} come in (11.3) con $\mathbb{P}(A_{hk}) > 0$, la probabilità dell'evento B condizionata dal valore x di X (o, più semplicemente, condizionata da x) è definita come

$$\mathbb{P}(B|x) := \mathbb{P}(B|X=x) = \lim_{h,k \rightarrow 0} \frac{\mathbb{P}(B \cap A_{hk})}{F_X(x+k) - F_X(x-h)}, \quad (11.6)$$

sempre che questo limite esista. \square

È possibile mostrare che in talune condizioni, solitamente soddisfatte, $\mathbb{P}(B|x)$ soddisfa gli assiomi di cui alla Definizione 3.14.

Consideriamo ora una seconda v.c. Y definita nello stesso spazio di probabilità di X e sia $F_{XY}(x,y)$ la funzione di distribuzione congiunta di X e Y :

$$F_{XY}(x,y) = \mathbb{P}(\omega : X(\omega) \leq x; Y(\omega) \leq y).$$

La Definizione 11.1 può essere applicata in particolare all'evento

$$B = \{\omega : Y(\omega) \leq y\}. \quad (11.7)$$

In questo caso risulta:

$$B \cap A_{hk} = \{\omega : Y(\omega) \leq y; x-h < X(\omega) \leq x+k\}$$

e quindi:

$$\mathbb{P}(B \cap A_{hk}) = F_{XY}(x+k; y) - F_{XY}(x-h; y).$$

Se esiste il limite, la (11.6) con la scelta (11.7) di B , può facilmente mostrarsi essere una funzione di distribuzione; a questa si dà pertanto il nome di funzione di distribuzione della v.c. Y condizionata dal valore x di X ; essa si indica con la scrittura $F_{Y|X}(y|x)$. Si ha dunque:

$$\begin{aligned} F_{Y|X}(y|x) &\equiv \mathbb{P}(Y \leq y|x) \\ &\equiv \lim_{h,k \rightarrow 0} \mathbb{P}(Y \leq y | x-h < X \leq x+k) \\ &= \lim_{h,k \rightarrow 0} \frac{\mathbb{P}(Y \leq y; x-h < X \leq x+k)}{\mathbb{P}(x-h < X \leq x+k)} \\ &= \lim_{h,k \rightarrow 0} \frac{F_{XY}(x+k; y) - F_{XY}(x-h; y)}{F_X(x+k) - F_X(x-h)}. \end{aligned} \quad (11.8)$$

11.1 Variabili casuali unidimensionali

Teorema 11.1. Se il limite (11.6) esiste e se la funzione $\mathbb{P}(B|x)$ è integrabile rispetto alla funzione di distribuzione F_X si ha:

$$\mathbb{P}(B) = \int_{-\infty}^{+\infty} \mathbb{P}(B|x) dF_X(x). \quad (11.9)$$

Dimostrazione. Dati $n-1$ reali $x_1 < x_2 < \dots < x_{n-1}$ e posto $x_0 = -\infty$, $x_n = +\infty$, consideriamo gli eventi

$$A_k = \{\omega : x_{k-1} < X(\omega) \leq x_k\} \quad (k = 1, 2, \dots, n).$$

Questi sono incompatibili ed esaustivi avendosi

$$A_i \cap A_j = \emptyset \quad (i \neq j), \quad \bigcup_{k=1}^n A_k = \Omega,$$

e quindi costituiscono un sistema completo di alternative. Applicando allora il Teorema 4.6 ed osservando che risulta

$$\mathbb{P}(A_k) = F_X(x_k) - F_X(x_{k-1}),$$

si ha:

$$\mathbb{P}(B) = \sum_{k=1}^n \mathbb{P}(B|A_k) \mathbb{P}(A_k) = \sum_{k=1}^n \mathbb{P}(B|A_k) [F_X(x_k) - F_X(x_{k-1})].$$

Passando al limite per $n \rightarrow \infty$, in virtù della definizione stessa di integrale di Stieltjes (cfr. § B.2) segue la (11.9). \square

Notiamo in particolare che ponendo $B = \{\omega : Y(\omega) \leq y\}$ risulta $\mathbb{P}(B) = F_Y(y)$, così che la (11.9) assume la seguente forma:

$$F_Y(y) = \int_{-\infty}^{+\infty} F_{Y|X}(y|x) dF_X(x). \quad (11.10)$$

Le (11.9) e (11.10) sono l'analogo, nel caso continuo, delle (4.15) e (4.16) relative al caso discreto (cfr. il Teorema 4.6 e il Teorema 4.7).

Supponiamo ora che il vettore casuale (X, Y) sia assolutamente continuo e dotato di densità di probabilità congiunta $f_{XY}(x, y)$ continua rispetto ad entrambe le variabili. Le v.v.c. X e Y sono allora assolutamente continue, dotate quindi di densità di probabilità f_X e f_Y . Calcoliamo in queste ipotesi la derivata della funzione di distribuzione condizionata

$F_{Y|X}(y|x)$. Per semplicità effettueremo una scelta particolare dei valori h e k ponendo $h = 0$ e $k = \epsilon$. Dalla (11.8) si ha allora:

$$\begin{aligned} F_{Y|X}(y|x) &= \lim_{\epsilon \rightarrow 0} \frac{F_{XY}(x + \epsilon, y) - F_{XY}(x, y)}{F_X(x + \epsilon) - F_X(x)} \\ &= \lim_{\epsilon \rightarrow 0} \frac{\epsilon}{F_X(x + \epsilon) - F_X(x)} \frac{F_{XY}(x + \epsilon, y) - F_{XY}(x, y)}{\epsilon} \\ &= \frac{1}{f_X(x)} \frac{\partial}{\partial x} F_{XY}(x, y) \end{aligned} \quad (11.11)$$

e quindi, tenendo conto della (5.42),

$$\frac{\partial}{\partial y} F_{Y|X}(y|x) = \frac{1}{f_X(x)} \frac{\partial^2}{\partial x \partial y} F_{XY}(x, y) = \frac{f_{XY}(x, y)}{f_X(x)}. \quad (11.12)$$

Possiamo quindi enunciare il seguente teorema:

Teorema 11.2. Siano X e Y vv.cc. assolutamente continue. Per ogni valore x di X per il quale sia possibile definire $F_{Y|X}(y|x)$ e tale che risulti $f_X(x) > 0$ esiste una funzione $f_{Y|X}(y|x) : \mathbb{R} \rightarrow \mathbb{R}$ tale da avere

$$f_{Y|X}(y|x) = \frac{\partial}{\partial y} F_{Y|X}(y|x). \quad (11.13)$$

La funzione $f_{Y|X}(y|x)$ prende il nome di densità di probabilità di Y condizionata dal valore x di X ; in virtù delle (11.12) e (11.13) essa è esprimibile al seguente modo:

$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_X(x)}. \quad (11.14)$$

Confrontando la (11.14) con la (11.1) si constata che sussiste una perfetta analogia tra il caso discreto e il caso continuo: quest'ultimo segue invero formalmente dal caso discreto pur di sostituire le probabilità di eventi con densità di probabilità. Ciò è d'altronde comprensibile osservando che $f_{XY}(x, y)\Delta x \Delta y$, $f_X(x)\Delta x$ e $f_{Y|X}(y|x)\Delta y$ sono delle probabilità per ogni scelta di Δx e Δy .

Notiamo che scambiando i ruoli di X e Y , in analogia con la (11.14) si ottiene:

$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)}. \quad (11.15)$$

Le (11.14) e (11.15) possono anche utilizzarsi per esprimere la densità di probabilità congiunta $f_{XY}(x, y)$ in funzione delle densità condizionate e delle densità marginali:

$$f_{XY}(x, y) = f_{X|Y}(x|y)f_Y(y) = f_{Y|X}(y|x)f_X(x). \quad (11.16)$$

Dalla (11.10), facendo uso delle (11.13) e (11.16), si ricava:

$$\begin{aligned} \frac{d}{dy} F_Y(y) &= \frac{\partial}{\partial y} \int_{-\infty}^{+\infty} F_{Y|X}(y|x) f_X(x) dx \\ &= \int_{-\infty}^{+\infty} f_{Y|X}(y|x) f_X(x) dx \\ &= \int_{-\infty}^{+\infty} f_{XY}(x, y) dx \\ &= f_Y(y). \end{aligned}$$

Si ritrova, così, la densità marginale, che può quindi esimersi al seguente modo:

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{Y|X}(y|x) f_X(x) dx. \quad (11.17)$$

Teorema 11.3. Se X e Y sono vv.cc. assolutamente continue, per ogni coppia (x, y) per la quale sia possibile definire le funzioni di distribuzione condizionate $F_{Y|X}(y|x)$ e $F_{X|Y}(x|y)$ si ha:

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{\int_{-\infty}^{+\infty} f_{Y|X}(y|x)f_X(x) dx}. \quad (11.18)$$

Dimostrazione. Segue immediatamente dalle (11.15) e (11.17). \square

Notiamo che la (11.18) è la trasposizione al caso continuo del teorema di Bayes, in termini, però, di densità di probabilità e non di probabilità.

Grazie alla (11.14) si possono immediatamente ricavare alcune proprietà della densità di probabilità condizionata.

Proposizione 11.1. Siano X e Y sono vv.cc. assolutamente continue. Nelle condizioni per le quali sussiste la (11.14), risulta:

$$f_{Y|X}(y|x) \geq 0, \quad (11.19)$$

con

$$\int_{-\infty}^{+\infty} f_{Y|X}(y|x) dy = 1. \quad (11.20)$$

Dimostrazione. La (11.19) segue immediatamente dalla (11.14) dal momento che f_{XY} e f_X sono densità di probabilità. Per quanto riguarda la condizione di normalizzazione (11.20), osserviamo che $f_{Y|X}(y|x) dy$ è la probabilità che dato il valore x di X si ottenga un valore di Y compreso tra y e $y + dy$; integrando allora su \mathbb{R} si ottiene la probabilità che, dato $X = x$,

la v.c. Y assuma un qualsiasi valore y , ossia la probabilità dell'evento certo che è unitaria. La dimostrazione analitica è semplice. Infatti, in virtù della (11.14), si ha:

$$\begin{aligned} \int_{-\infty}^{+\infty} f_{Y|X}(y|x) dy &= \int_{-\infty}^{+\infty} \frac{f_{XY}(x,y)}{f_X(x)} dy \\ &= \frac{1}{f_X(x)} \int_{-\infty}^{+\infty} f_{XY}(x,y) dy \\ &= \frac{f_X(x)}{f_X(x)} = 1. \end{aligned} \quad \square$$

Se le vv.cc. X e Y sono indipendenti, si ha per definizione:

$$F_{XY}(x,y) = F_X(x)F_Y(y), \quad (11.21)$$

e quindi dalla (11.11) si ottiene:

$$\begin{aligned} F_{Y|X}(y|x) &= \frac{1}{f_X(x)} \frac{\partial}{\partial x} [F_X(x)F_Y(y)] \\ &= \frac{f_X(x)F_Y(y)}{f_X(x)} = F_Y(y). \end{aligned}$$

Analogamente, scambiando i ruoli di X e Y e nell'ipotesi di indipendenza di queste, segue anche:

$$F_{X|Y}(x|y) = \frac{F_X(x)f_Y(y)}{f_Y(y)} = F_X(x).$$

Se poi X e Y oltre che indipendenti sono anche assolutamente continue, dalla (11.21) si trae:

$$f_{XY}(x,y) = f_X(x)f_Y(y).$$

In analogia con il caso discreto, il concetto di indipendenza di vv.cc. assolutamente continue è esprimibile in più modi equivalenti, come indica la Proposizione che segue.

Proposizione 11.2. *Sono equivalenti le seguenti relazioni:*

- (i) $f_{XY}(x,y) = f_X(x)f_Y(y);$
- (ii) $f_{X|Y}(x|y) = f_X(x);$
- (iii) $f_{Y|X}(y|x) = f_Y(y).$

Dimostrazione. Basta dimostrare che sussistono le implicazioni

$$(i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (i).$$

(a) Che $(i) \Rightarrow (ii)$ segue immediatamente dalla (11.15) e dalla (i):

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{f_X(x)f_Y(y)}{f_Y(y)} = f_X(x).$$

(b) Dimostriamo che $(ii) \Rightarrow (iii)$. Facendo uso delle (11.14), (11.16) e della (ii) si ottiene:

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{f_{X|Y}(x|y)f_Y(y)}{f_X(x)} = \frac{f_X(x)f_Y(y)}{f_X(x)} = f_Y(y).$$

(c) Dimostriamo infine che $(iii) \Rightarrow (i)$. Sempre dalla (11.16) e dalla (iii) segue:

$$f_{XY}(x,y) = f_{Y|X}(y|x)f_X(x) = f_X(x)f_Y(y). \quad \square$$

11.2 Momenti condizionati

Siano X e Y vv.cc. assolutamente continue definite in uno stesso spazio di probabilità (Ω, \mathcal{F}, P) .

Definizione 11.2. Si dice valore medio, o valore di attesa, di Y condizionato da $\{X(\omega) = x\}$, e si indica con $E\{Y|X(\omega) = x\}$, oppure con $E\{Y|x\}$ o anche con $\langle Y|x \rangle$, l'integrale

$$E\{Y|X(\omega) = x\} = \int_{-\infty}^{+\infty} y f_{Y|X}(y|x) dy, \quad (11.22)$$

se esso è assolutamente convergente.¹ \square

Notiamo che la funzione $f_{Y|X}(y|x)$ dipende da x ; di conseguenza $Z = g(X)$, con $g(x) = E(Y|x)$, per la (11.22) è una variabile casuale. Quando si vuole mettere in evidenza questo aspetto conviene usare la scrittura $E(Y|X)$.

Se $\mathbf{Y} := (Y_1, \dots, Y_n)$ è un vettore casuale n -dimensionale, si pone per definizione:

$$E\{\mathbf{Y}|X(\omega) = x\} := (E\{Y_1|x\}, \dots, E\{Y_n|x\}). \quad (11.23)$$

¹D'ora innanzi assumeremo tacitamente l'esistenza dei valori medi condizionati considerati.

Ritornando alla coppia di vv.cc. X e Y , tenendo conto dell'espressione della densità marginale e della (6.11) si ha:

$$\begin{aligned} f_Y(y) &\equiv \int_{-\infty}^{+\infty} f_{XY}(x,y) dx = \int_{-\infty}^{+\infty} f_{Y|X}(y|x) f_X(x) dx \\ &= E_X\{f_{Y|X}(y|X)\}. \end{aligned}$$

Analogamente al valore medio, possono definirsi i momenti condizionati di un qualunque ordine, ma non ci dilungheremo su questo. Vogliamo invece dimostrare un teorema che esprime importanti proprietà del valore medio condizionato.

Teorema 11.4. Siano X, Y, Z vv.cc. congiuntamente distribuite, c e d costanti, e $g: \mathbb{R} \rightarrow \mathbb{R}$ una funzione Borel-misurabile. Nell'ipotesi in cui $E(Y)$, $E(Z)$, $E[g(X)Y]$ esistono si ha:

- (i) $E(Y|X=x) = E(Y)$ se X e Y sono indipendenti;
- (ii) $E(Y) = E[E(Y|X)]$;
- (iii) $E[g(X)Y|X=x] = g(x)E(Y|x)$;
- (iv) $E[g(X)Y] = E[g(X)E(Y|X)]$;
- (v) $E[c|X=x] = c$;
- (vi) $E[g(X)|X=x] = g(x)$;
- (vii) $E[cY + dZ|X=x] = cE(Y|X=x) + dE(Z|X=x)$.

Dimostrazione. Dimostriamo la (i). Si ha:

$$E(Y|X=x) = \int_{-\infty}^{+\infty} y f_{Y|X}(y|x) dy = \int_{-\infty}^{+\infty} y f_Y(y) dy = E(Y) \quad (11.24)$$

per la supposta indipendenza di X e Y . Osserviamo che non vale il viceversa, ossia che la (11.24) può sussistere anche se X e Y non sono indipendenti. Dimostriamo ora la (ii).

Dalla definizione di valore medio di una funzione della v.c. X segue:

$$\begin{aligned} E[E(Y|X)] &= \int_{-\infty}^{+\infty} f_X(x) dx \int_{-\infty}^{+\infty} y f_{Y|X}(y|x) dy \\ &= \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} y f_{XY}(x,y) dy \\ &= \int_{-\infty}^{+\infty} y dy \int_{-\infty}^{+\infty} f_{XY}(x,y) dx \\ &= \int_{-\infty}^{+\infty} y f_Y(y) dy = E(Y). \end{aligned}$$

Dimostriamo la (iii). Sempre dalla definizione di valore medio si trae:

$$\begin{aligned} E[g(X)Y|X=x] &= \int_{-\infty}^{+\infty} g(x) y f_{Y|X}(y|x) dy \\ &= g(x) \int_{-\infty}^{+\infty} y f_{Y|X}(y|x) dy = g(x)E(Y|X=x). \end{aligned}$$

Per quanto riguarda la (iv), osserviamo che si ha:

$$\begin{aligned} E[g(X)E(Y|X)] &= \int_{-\infty}^{+\infty} g(x) f_X(x) dx \int_{-\infty}^{+\infty} y f_{Y|X}(y|x) dy \\ &= \int_{-\infty}^{+\infty} y dy \int_{-\infty}^{+\infty} g(x) f_{XY}(x,y) dx = E[g(X)Y]. \end{aligned}$$

Tenendo conto della condizione di normalizzazione (11.20) si ottiene:

$$E(c|X=x) = \int_{-\infty}^{+\infty} c f_{Y|X}(y|x) dy = c \int_{-\infty}^{+\infty} f_{Y|X}(y|x) dy = c,$$

così che la (v) è dimostrata. Sempre in base alla (11.20) segue poi la (vi):

$$\begin{aligned} E[g(X)|X=x] &= \int_{-\infty}^{+\infty} g(x) f_{Y|X}(y|x) dy \\ &= g(x) \int_{-\infty}^{+\infty} f_{Y|X}(y|x) dy = g(x). \end{aligned}$$

Infine, per quanto concerne la (vii) osserviamo che risulta:

$$\begin{aligned} E(cY + dZ|X=x) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (cy + dz) f_{YZ|X}(y,z|x) dy dz \\ &= c \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y f_{YZ|X}(y,z|x) dy dz + \\ &\quad + d \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} z f_{YZ|X}(y,z|x) dy dz \\ &= c \int_{-\infty}^{+\infty} y dy \int_{-\infty}^{+\infty} f_{YZ|X}(y,z|x) dz + \\ &\quad + d \int_{-\infty}^{+\infty} z dz \int_{-\infty}^{+\infty} f_{YZ|X}(y,z|x) dy \\ &= c \int_{-\infty}^{+\infty} y f_{Y|X}(y|x) dy + d \int_{-\infty}^{+\infty} z f_{Z|X}(z|x) dz \\ &= cE(Y|X=x) + dE(Z|X=x), \end{aligned}$$

avendo fatto uso delle relazioni

$$\begin{aligned} \int_{-\infty}^{+\infty} f_{YZ|X}(y, z|x) dy &= \int_{-\infty}^{+\infty} \frac{f_{XYZ}(x, y, z)}{f_X(x)} dy = \frac{f_{XZ}(x, z)}{f_X(x)} = f_{Z|X}(z|x) \\ \int_{-\infty}^{+\infty} f_{YZ|X}(y, z|x) dz &= f_{Y|X}(y|x). \end{aligned}$$

□

11.3 Vettori casuali

Siano $X = (X_1, \dots, X_n)$ e $Y = (Y_1, \dots, Y_m)$ vettori casuali definiti in uno stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e siano F_X , F_Y e F_{XY} le funzioni di distribuzione rispettivamente di X , di Y e di (X, Y) . Indichiamo poi con B l'evento

$$B = \{\omega : Y_1(\omega) \leq y_1; \dots; Y_m(\omega) \leq y_m\}.$$

Inoltre, scelte $2n$ costanti arbitrarie $h_1, \dots, h_n, k_1, \dots, k_n$ consideriamo l'evento

$$A := \{\omega : x_1 - h_1 < X_1(\omega) \leq x_1 + k_1; \dots; x_n - h_n < X_n(\omega) \leq x_n + k_n\} \quad (11.25)$$

e poniamo

$$x = (x_1, \dots, x_n), \quad y = (y_1, \dots, y_m), \quad h = (h_1, \dots, h_n), \quad k = (k_1, \dots, k_n).$$

Supposto di avere scelto le costanti in modo tale da aversi $\mathbb{P}(A) > 0$, consideriamo la probabilità di B condizionata da A :

$$\begin{aligned} \mathbb{P}(B|A) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \\ &= \frac{\mathbb{P}(\{Y \leq y\} \cap \{x - h < X \leq x + k\})}{\mathbb{P}(\{x - h < X \leq x + k\})}, \end{aligned}$$

dove, ad esempio, con la scrittura $U \leq u$ intendiamo che ogni componente U_i di U è minore o uguale della corrispondente componente u_i di u .

Definizione 11.3. Dati i vettori casuali X e Y e i vettori costanti h e k tali che risulta $\mathbb{P}(A) > 0$, con A definito dalla (11.25), il limite

$$F_{Y|X}(y|x) := \lim_{h, k \rightarrow 0} \frac{\mathbb{P}(\{Y \leq y\} \cap \{x - h < X \leq x + k\})}{\mathbb{P}(\{x - h < X \leq x + k\})}, \quad (11.26)$$

se esiste finito, prende il nome di funzione di distribuzione di Y condizionata da X . □

I risultati ottenuti nel § 11.1 si possono estendere al caso di vv.cc. multidimensionali.

Proposizione 11.3. Se il vettore casuale $(X, Y) = (X_1, \dots, X_n, Y_1, \dots, Y_m)$ è assolutamente continuo e la sua densità di probabilità congiunta $f_{XY}(x, y)$ è continua rispetto a tutte le variabili, si ha:

$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_X(x)} \quad (11.27)$$

$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)}. \quad (11.28)$$

Dimostrazione. Dimostriamo ad esempio la (11.27) ponendoci, per semplicità di scrittura, nell'ipotesi $h_1 = h_2 = \dots = h_n = 0$, ossia $h = 0$. Dalla (11.26) segue allora:

$$F_{Y|X}(y|x) = \lim_{k \rightarrow 0} \frac{\int_V dv \int_U du f_{XY}(u, v)}{\int_U du f_X(u)},$$

dove

$$U = \{u : x_1 < u_1 \leq x_1 + k_1; \dots; x_n < u_n \leq x_n + k_n\}$$

e

$$V = \{v : v_1 \leq y_1; \dots; v_m \leq y_m\}. \quad (11.29)$$

Dalla (11.29) segue anche:

$$F_{Y|X}(y|x) = \lim_{k \rightarrow 0} \left[\frac{\int_V dv \int_U du f_{XY}(u, v)}{k_1 \dots k_n} \times \frac{k_1 \dots k_n}{\int_U du f_X(u)} \right].$$

Poiché risulta

$$\lim_{k \rightarrow 0} \frac{\int_V dv \int_U du f_{XY}(u, v)}{k_1 \dots k_n} = \int_V dv f_{XY}(x, v)$$

ed inoltre

$$\lim_{k \rightarrow 0} \frac{\int_U du f_X(u)}{k_1 \dots k_n} = f_X(x),$$

si ricava:

$$F_{Y|X}(y|x) = \int_{-\infty}^{y_1} dv_1 \dots \int_{-\infty}^{y_m} dv_m \frac{f_{XY}(x, v)}{f_X(x)}.$$

Ciò, come espresso dalla (5.41), indica che $\hat{F}_{Y|X}(y|x)$ è la funzione di distribuzione di una v.c. assolutamente continua la cui densità di probabilità è data dalla (11.27). In maniera analoga si dimostra la (11.28).

Come conseguenza delle (11.27) e (11.28) si ha anche:

$$f_{XY}(x, y) = f_{Y|X}(y|x) f_X(x) = f_{X|Y}(x|y) f_Y(y). \quad (11.30)$$

Se poi i vettori casuali sono indipendenti, analogamente al caso unidimensionale risulta

$$f_{Y|X}(y|x) = f_Y(y) \quad (11.31)$$

e, scambiando i ruoli di X e Y ,

$$f_{X|Y}(x|y) = f_X(x). \quad (11.32)$$

La nozione di probabilità condizionata si può estendere al caso di n variabili casuali multidimensionali X_1, \dots, X_n . Queste si possono suddividere in due gruppi costituiti rispettivamente da r vv.cc. X_{i_1}, \dots, X_{i_r} e dalle rimanenti $n - r$ vv.cc. $X_{i_{r+1}}, \dots, X_{i_n}$, con $r = 1, 2, \dots, n - 1$. Si può quindi definire la probabilità condizionata di uno dei gruppi di vv.cc. rispetto all'altro in modo analogo a quanto espresso dalla (11.26).

11.3.1 Una particolare classe di processi

L'estensione del concetto di probabilità condizionata al caso di variabili casuali multidimensionali riveste notevole interesse specialmente per le sue applicazioni. Si consideri, ad esempio, un sistema dinamico retto da leggi stocastiche e sia $X_0, X_1, \dots, X_n, \dots$ una successione di vv.cc. — che per comodità supponiamo assolutamente continue — rappresentante la successione degli stati raggiungibili dal sistema agli istanti $0, 1, \dots, n, \dots$ di una fissata scala di tempi. In generale lo stato al "tempo presente" n dipende da tutti gli stati precedenti; se allora si vuole determinare lo stato del sistema al tempo n si devono esaminare le proprietà delle $n + 1$ vv.cc. X_0, X_1, \dots, X_n ; occorre quindi conoscere la densità di probabilità congiunta

$$f_{X_0 \dots X_n}(x_0, \dots, x_n) = \frac{\partial^n}{\partial x_0 \partial x_1 \dots \partial x_n} \mathbb{P}(X_0 \leq x_0; X_1 \leq x_1; \dots; X_n \leq x_n). \quad (11.33)$$

Mediante la (11.33) si possono poi determinare le densità marginali di qualunque ordine da 1 a $n - 1$. Se si desidera descrivere l'evoluzione del sistema nel tempo è dunque in generale necessario conoscere densità di probabilità di ordine via via sempre più elevato; il problema appare quindi irrisolvibile.

Esaminiamo ora come lo stato al tempo presente è influenzato dagli stati precedentemente attraversati. A tal fine occorre riferirsi ad una densità di probabilità condizionata del tipo

$$f_{X_n|X_{n-1} \dots X_0}(x_n|x_{n-1}, \dots, x_0) = \frac{f_{X_0 \dots X_n}(x_0, \dots, x_n)}{f_{X_{n-1} \dots X_0}(x_{n-1}, \dots, x_0)};$$

da questa, per iterazione, si ottiene poi facilmente:

$$f_{X_0 \dots X_n}(x_0, \dots, x_n) = f_{X_n|X_{n-1} \dots X_0}(x_n|x_{n-1}, \dots, x_0) \cdots f_{X_1|X_0}(x_1|x_0) f_{X_0}(x_0). \quad (11.34)$$

Orbene, esiste una classe particolare di processi, detti *processi di Markov*, caratterizzati dalla proprietà che la funzione di distribuzione dello stato del processo ad ogni istante dipende solo dallo stato occupato all'istante immediatamente precedente. Per tale classe di processi, nel caso assolutamente continuo si ha quindi:

$$f_{X_n|X_{n-1} \dots X_0}(x_n|x_{n-1}, \dots, x_0) = f_{X_n|X_{n-1}}(x_n|x_{n-1}). \quad (11.35)$$

La funzione definita dalla (11.35) si dice *densità di probabilità di transizione*. Se si riesce a determinare la funzione a secondo membro della (11.35) per ogni coppia (x_n, x_{n-1}) , e se inoltre è nota la densità di probabilità dello stato iniziale, tramite la (11.34) si può risalire alla densità di probabilità congiunta delle $n + 1$ vv.cc. X_0, \dots, X_n per ogni n , e quindi si può determinare l'evoluzione stocastica del sistema.

Sebbene tali processi costituiscano un'astrazione matematica, essi risultano utili ed efficaci per la modellizzazione di numerosi fenomeni. Su di essi ritorneremo nel Capitolo 12.

CAPITOLO 12

Cenno ai Processi Stocastici

12.1 Proprietà di carattere generale

Cominciamo col dare una definizione di ciò che debba intendersi allorché si parla di processi stocastici.¹

Definizione 12.1. Sia $(\Omega, \mathcal{F}, \mathbb{P})$ uno spazio di probabilità e $\mathcal{G} = \{X(t); t \geq t_0, t_0 \in \mathbb{R}\}$ una famiglia di funzioni i cui elementi sono in corrispondenza con gli elementi di Ω , nel senso che esiste un'applicazione $s: \Omega \rightarrow \mathcal{G}$ che ad ogni $\omega \in \Omega$ associa una funzione $X(t, \omega) \equiv X(t) \in \mathcal{G}$. La famiglia di funzioni \mathcal{G} è detta processo stocastico di parametro t . \triangleleft

L'elemento aleatorio connesso a tali funzioni è dovuto al fatto che gli elementi di Ω sono i risultati di un esperimento casuale le cui probabilità di occorrenza forniscono anche, per così dire, le "probabilità di occorrenza" delle funzioni di \mathcal{G} .

Esempio 12.1. Sia $\mathcal{G} = \{e^{t/2}, t^3; t \geq 0\}$, e sia $(\Omega, \mathcal{F}, \mathbb{P})$ uno spazio di probabilità associato all'esperimento consistente nel lancio di una moneta. Supponiamo, dunque, che sia $\Omega = \{T, C\}$. Consideriamo poi l'applicazione s così definita:

$$s(T) = e^{t/2}, \quad s(C) = t^3.$$

Possiamo allora dire che la funzione $e^{t/2}$ si presenta con probabilità uguale a quella dell'evento testa T , e la funzione t^3 con probabilità uguale a quella dell'evento croce C ; queste probabilità sono entrambe uguali a $1/2$ se la moneta è equa. \triangleleft

Possiamo riassumere le proprietà che caratterizzano un processo stocastico nel modo seguente:

- (a) per ogni fissato elemento $\omega^* \in \Omega$ la scrittura $X(t, \omega^*)$ rappresenta una funzione di t ;

¹Qui ci si riferisce esclusivamente al caso unidimensionale.

- (b) per ogni $\bar{t} \geq t_0$, $X(\bar{t}, \omega)$ è una variabile casuale;
- (c) per ogni fissata coppia $\bar{t} \geq t_0$, $\omega^* \in \Omega$, $X(\bar{t}, \omega^*)$ è un numero reale.

D'ora in avanti indicheremo in generale un processo stocastico con il simbolo $X(t)$, omettendo, quando non necessaria, l'indicazione di ω .

Definizione 12.2. Ogni elemento della famiglia \mathcal{G} si dice essere una realizzazione del processo stocastico $X(t)$. \triangleleft

Definizione 12.3. L'insieme dei valori assunti dal processo si dice costituire lo spazio degli stati. \triangleleft

Lo spazio degli stati può essere tanto discreto quanto continuo. Anche l'insieme dei valori del parametro t , che d'ora innanzi per comodità di linguaggio interpreteremo come variabile temporale, può essere discreto o continuo. I casi ai quali nel seguito ci riferiremo sono quelli in cui sia lo spazio degli stati che il dominio del parametro t sono continui.

Diremo che un processo stocastico $X(t)$ è noto se comunque fissata una n -upla ordinata di tempi $t_1 < t_2 < \dots < t_n$ è nota la funzione di distribuzione delle variabili casuali $X(t_1), X(t_2), \dots, X(t_n)$. Denoteremo questa funzione con la scrittura

$$F_n(x_1, t_1; \dots; x_n, t_n) = \mathbb{P}(X(t_1) \leq x_1, \dots, X(t_n) \leq x_n), \quad x_i \in \mathbb{R}, \quad (12.1)$$

dove la coppia (x_i, t_i) indica che x_i è il generico valore assunto dalla variabile casuale $X(t_i)$.

Esaminiamo ora in maggior dettaglio il caso di un processo stocastico continuo sia nello spazio che nel tempo nell'ipotesi che le variabili casuali $X(t_1), X(t_2), \dots, X(t_n)$ siano assolutamente continue. Il processo stocastico $X(t)$ è allora univocamente determinato se è nota la densità di probabilità congiunta

$$f_n(x_1, t_1; \dots; x_n, t_n) = \frac{\partial^n}{\partial x_1 \dots \partial x_n} F_n(x_1, t_1; \dots; x_n, t_n). \quad (12.2)$$

La descrizione di un processo stocastico richiede quindi la conoscenza di vettori casuali di dimensione comunque elevata; inoltre, a parità di n , al variare di $t_1 < t_2 < \dots < t_n$ la n -upla di variabili casuali da prendere in esame varia. Per studiare quindi l'evoluzione di un processo stocastico si deve determinare la densità di probabilità per ogni n e per ogni n -upla (t_1, \dots, t_n) con $t_1 < \dots < t_n$. Ricordando quanto indicato nel § 11.3.1, la densità f_n si può esprimere nella forma

$$f_n(x_1, t_1; \dots; x_n, t_n) = f_{n-1}(x_1, t_1; \dots; x_{n-1}, t_{n-1}) f(x_n, t_n | x_{n-1}, t_{n-1}; \dots; x_1, t_1). \quad (12.3)$$

Per determinare la densità n -dimensionale occorre quindi conoscere la densità di ordine $n-1$ e la densità condizionata che compare nella (12.3), che contiene informazioni sul

passato del processo. In generale, infatti, la densità di probabilità di $X(t_n)$ condizionata da $X(t_1), \dots, X(t_{n-1})$ dipende da tutte le variabili casuali $X(t_1), \dots, X(t_{n-1})$, ossia dipende dall'intera successione dei precedenti stati considerati. Nei casi in cui la densità condizionata è determinabile, se è nota $f_1(x_1, t_1)$ si può ottenere $f_n(x_1, t_1; \dots; x_n, t_n)$ tramite la (12.3) per ogni $n \geq 1$ e per ogni n -upla ordinata $t_1 < \dots < t_n$.

12.2 Processi di Markov

Come abbiamo visto nel § 12.1, lo studio dell'evoluzione di un processo stocastico pone in generale delle difficoltà apparentemente insormontabili in quanto per la sua descrizione sembrerebbe necessario dover conoscere densità di probabilità di ordine comunque elevato. Tali difficoltà possono essere tuttavia superate in casi particolari di dipendenza dello stato al tempo presente dagli stati precedenti, come accade nei cosiddetti processi di Markov, ai quali si è fatto cenno nel § 11.3.1, e dei quali passiamo ora ad occuparci.

Definizione 12.4. Un processo stocastico si dice di Markov, o markoviano, se per ogni intero n e comunque fissati n istanti ordinati $t_1 < \dots < t_n$ si ha:

$$f(x_n, t_n | x_{n-1}, t_{n-1}; \dots; x_1, t_1) = f(x_n, t_n | x_{n-1}, t_{n-1}). \quad (12.4)$$

\triangleleft

Per un processo di Markov la densità di probabilità condizionata non dipende, dunque, da tutte le variabili casuali condizionanti, ma solo dalla variabile casuale all'istante immediatamente precedente.

Se indichiamo con t_n il generico istante "futuro", con t_{n-1} l'istante "presente" e con t_i ($i < n-1$) gli istanti "passati", la proprietà (12.4) si può suggestivamente esprimere dicendo che per i processi markoviani il "futuro" $X(t_n)$ dipende solo dal "presente" $X(t_{n-1})$ e non dal "passato" $X(t_i)$ ($i < n-1$). I processi di Markov sono dunque caratterizzati da una limitata "memoria".

In base alla (12.4), la (12.3) si può scrivere nella forma

$$\begin{aligned} f_n(x_n, t_n; \dots; x_1, t_1) &= f_{n-1}(x_{n-1}, t_{n-1}; \dots; x_1, t_1) f(x_n, t_n | x_{n-1}, t_{n-1}) \\ &= f_{n-2}(x_{n-2}, t_{n-2}; \dots; x_1, t_1) f(x_n, t_n | x_{n-1}, t_{n-1}) \times \\ &\quad \times f(x_{n-1}, t_{n-1} | x_{n-2}, t_{n-2}) \dots \\ &= f(x_n, t_n | x_{n-1}, t_{n-1}) f(x_{n-1}, t_{n-1} | x_{n-2}, t_{n-2}) \times \\ &\quad \dots \times f(x_2, t_2 | x_1, t_1) f_1(x_1, t_1). \end{aligned} \quad (12.5)$$

Dalla (12.5) segue che se sono note la densità di probabilità del prim'ordine $f_1(x, t)$ e la funzione $f(x, t|y, \tau)$ per ogni $x, y \in \mathbb{R}$ e per ogni $\tau < t$, il processo di Markov è completamente determinato.

Definizione 12.5. La densità di probabilità condizionata $f(x, t|y, \tau)$ definita dalla (12.4) prende il nome di densità di transizione. \triangleleft

La funzione $f_1(x, t)$ deve essere riguardata come un dato del problema, e quindi considerata come assegnata, laddove la densità di transizione $f(x, t|y, \tau)$ va determinata. A tal fine occorre pervenire ad una qualche equazione di cui l'incognita densità di transizione sia soluzione.

Ricaveremo ora una prima relazione che coinvolge la densità di transizione. Siano dunque t_0, τ, t istanti arbitrari, con $t_0 < \tau < t$. Convenzionalmente t_0 prende il nome di *istante iniziale*. Siano inoltre x_0, y, x i generici valori assunti rispettivamente dalle variabili casuali $X(t_0), X(\tau), X(t)$. Va qui specificato che d'ora innanzi siano soddisfatte tutte le condizioni di derivabilità e di integrabilità atte a giustificare la correttezza degli sviluppi analitici che seguono.

Teorema 12.1. La densità di transizione di un processo di Markov soddisfa la seguente relazione:

$$f(x, t|x_0, t_0) = \int_{-\infty}^{+\infty} f(x, t|y, \tau) f(y, \tau|x_0, t_0) dy. \quad (12.6)$$

Dimostrazione. Osserviamo anzitutto che risulta

$$f(x, t|x_0, t_0) = \int_{-\infty}^{+\infty} f(x, t|y, \tau; x_0, t_0) f(y, \tau|x_0, t_0) dy. \quad (12.7)$$

Infatti la densità marginale $f_2(x, t; x_0, t_0)$ si può esprimere nella forma seguente:

$$f_2(x, t; x_0, t_0) = \int_{-\infty}^{+\infty} f_3(x, t; y, \tau; x_0, t_0) dy;$$

d'altra parte, dalla (12.3) segue:

$$f_2(x, t; x_0, t_0) = \int_{-\infty}^{+\infty} f(x, t|y, \tau; x_0, t_0) f_2(y, \tau; x_0, t_0) dy,$$

o anche:

$$f(x, t|x_0, t_0) f_1(x_0, t_0) = \int_{-\infty}^{+\infty} f(x, t|y, \tau; x_0, t_0) f(y, \tau|x_0, t_0) f_1(x_0, t_0) dy. \quad (12.8)$$

12.3 Equazione di Smolukowski

Poiché $f_1(x_0, t_0)$ non dipende dalla variabile di integrazione, dalla (12.8) segue la (12.7). Quest'ultima, che coinvolge solo densità condizionate, è valida qualunque sia il processo stocastico considerato; se, però, il processo stocastico è markoviano, risulta:

$$f(x, t|y, \tau; x_0, t_0) = f(x, t|y, \tau)$$

così che la (12.7) assume la forma (12.6). \square

Si osservi che la (12.6), nota quale *equazione di Smolukowski* oppure *equazione di Chapman-Kolmogorov*, è una identità che va riguardata come una relazione di compatibilità che deve essere soddisfatta dalla densità di transizione di ogni processo di Markov. Essa esprime la circostanza che per i processi di Markov la transizione dallo stato iniziale x_0 assunto al tempo t_0 allo stato generico x al tempo t si può decomporre in una transizione da (x_0, t_0) a (y, τ) seguita da una transizione da (y, τ) a (x, t) , dove $y = X(\tau)$ è un arbitrario generico stato e τ è un istante intermedio tra t_0 e t . Val la pena sottolineare che questo tipo di decomposizione non sussiste per qualunque processo stocastico, ma solo per i processi di Markov. Si può anzi dimostrare che nel caso di un processo stocastico continuo sia nello spazio che nel tempo una funzione f soddisfa l'equazione di Smolukowski se e solo se essa coincide con la densità di transizione di un processo di Markov.

12.3 Equazione di Smolukowski

Un ulteriore passo verso la determinazione della densità di transizione consiste nel trasformare l'equazione (12.6) in uno sviluppo differenziale, come indicato nella proposizione che segue.

Proposizione 12.1. L'equazione (12.6) può risciversi nella forma

$$\frac{\partial}{\partial t} f(x, t|x_0, t_0) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial x^n} [f(x, t|x_0, t_0) A_n(x, t)], \quad (12.9)$$

dove

$$A_n(y, t) = \lim_{\Delta t \rightarrow 0^+} \frac{1}{\Delta t} \int_{-\infty}^{+\infty} (x - y)^n f(x, t + \Delta t|y, t) dx, \quad (n = 1, 2, \dots). \quad (12.10)$$

Dimostrazione. Sostituendo nella (12.6) $t + \Delta t$ al posto di t , con $\Delta t > 0$, e t al posto di τ . Si noti che ciò è lecito in quanto è $t + \Delta t > t$. Dopo aver sottratto $f(x, t|x_0, t_0)$ ad ambo i

membri, si ha dunque:

$$\begin{aligned} f(x, t + \Delta t | x_0, t_0) - f(x, t | x_0, t_0) &= \int_{-\infty}^{+\infty} f(x, t + \Delta t | y, t) f(y, t | x_0, t_0) dy + \\ &\quad - f(x, t | x_0, t_0). \end{aligned} \quad (12.11)$$

Sia ora R una funzione arbitraria dotata di derivate di ordine arbitrariamente elevato, che tenda a zero con tutte le sue derivate per $y \rightarrow \pm\infty$. Dalla (12.11) si trae allora:

$$\begin{aligned} &\int_{-\infty}^{+\infty} \frac{f(x, t + \Delta t | x_0, t_0) - f(x, t | x_0, t_0)}{\Delta t} R(x) dx \\ &= \frac{1}{\Delta t} \int_{-\infty}^{+\infty} R(x) dx \int_{-\infty}^{+\infty} f(x, t + \Delta t | y, t) f(y, t | x_0, t_0) dy + \\ &\quad - \frac{1}{\Delta t} \int_{-\infty}^{+\infty} R(x) f(x, t | x_0, t_0) dx. \end{aligned} \quad (12.12)$$

Effettuiamo il limite per $\Delta t \rightarrow 0$ in entrambi i membri della (12.12) dopo aver sostituito nel primo termine del secondo membro a $R(x)$ il suo sviluppo in serie di Taylor di punto iniziale y :

$$R(x) = R(y) + \sum_{n=1}^{\infty} \frac{d^n R(y)}{dy^n} \frac{(x-y)^n}{n!}.$$

Si ottiene così:

$$\begin{aligned} &\int_{-\infty}^{+\infty} R(x) \frac{\partial}{\partial t} f(x, t | x_0, t_0) dx \\ &= \lim_{\Delta t \rightarrow 0^+} \frac{1}{\Delta t} \int_{-\infty}^{+\infty} R(y) f(y, t | x_0, t_0) dy \int_{-\infty}^{+\infty} f(x, t + \Delta t | y, t) dx + \\ &\quad + \lim_{\Delta t \rightarrow 0^+} \frac{1}{\Delta t} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \sum_{n=1}^{\infty} \frac{d^n R(y)}{dy^n} \frac{(x-y)^n}{n!} f(x, t + \Delta t | y, t) \times \\ &\quad \times f(y, t | x_0, t_0) dy dx - \lim_{\Delta t \rightarrow 0^+} \frac{1}{\Delta t} \int_{-\infty}^{+\infty} R(x) f(x, t | x_0, t_0) dx. \end{aligned}$$

In virtù della condizione di normalizzazione

$$\int_{-\infty}^{+\infty} f(x, t + \Delta t | y, t) dx = 1$$

12.3 Equazione di Smolukowski

si ha poi:

$$\begin{aligned} \int_{-\infty}^{+\infty} R(x) \frac{\partial}{\partial t} f(x, t | x_0, t_0) dx &= \lim_{\Delta t \rightarrow 0^+} \frac{1}{\Delta t} \int_{-\infty}^{+\infty} R(y) f(y, t | x_0, t_0) dy + \\ &\quad + \sum_{n=1}^{\infty} \frac{1}{n!} \int_{-\infty}^{+\infty} f(y, t | x_0, t_0) \frac{d^n R(y)}{dy^n} \left[\lim_{\Delta t \rightarrow 0^+} \frac{1}{\Delta t} \int_{-\infty}^{+\infty} (x-y)^n \times \right. \\ &\quad \left. \times f(x, t + \Delta t | y, t) dy \right] - \lim_{\Delta t \rightarrow 0^+} \frac{1}{\Delta t} \int_{-\infty}^{+\infty} R(x) f(x, t | x_0, t_0) dx. \end{aligned}$$

Di qui otteniamo:

$$\int_{-\infty}^{+\infty} R(x) \frac{\partial}{\partial t} f(x, t | x_0, t_0) dx = \sum_{n=1}^{\infty} \frac{1}{n!} \int_{-\infty}^{+\infty} f(y, t | x_0, t_0) A_n(y, t) \frac{d^n R(y)}{dy^n} dy, \quad (12.13)$$

dove si è fatto uso della definizione (12.10). Gli integrali a secondo membro della (12.13) possono essere calcolati per parti. Una prima integrazione dà

$$\begin{aligned} \int_{-\infty}^{+\infty} f(y, t | x_0, t_0) A_n(y, t) \frac{d^n R(y)}{dy^n} dy &= \left[f(y, t | x_0, t_0) A_n(y, t) \frac{d^{n-1} R(y)}{dy^{n-1}} \right]_{-\infty}^{+\infty} + \\ &\quad - \int_{-\infty}^{+\infty} \frac{\partial}{\partial y} \left[f(y, t | x_0, t_0) A_n(y, t) \right] \frac{d^{n-1} R(y)}{dy^{n-1}} dy \\ &= - \int_{-\infty}^{+\infty} \frac{\partial}{\partial y} \left[f(y, t | x_0, t_0) A_n(y, t) \right] \frac{d^{n-1} R(y)}{dy^{n-1}} dy, \end{aligned}$$

dove si è fatto uso delle ipotesi sopra formulate su R . Iterando il procedimento si ottiene:

$$\int_{-\infty}^{+\infty} R(x) \frac{\partial}{\partial t} f(x, t | x_0, t_0) dx = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \int_{-\infty}^{+\infty} \frac{\partial^n}{\partial y^n} [f(y, t | x_0, t_0) A_n(y, t)] R(y) dy,$$

e quindi:

$$\int_{-\infty}^{+\infty} R(x) \left\{ \frac{\partial}{\partial t} f(x, t | x_0, t_0) - \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial x^n} [f(x, t | x_0, t_0) A_n(x, t)] \right\} dx = 0. \quad (12.14)$$

Poiché la (12.14) deve sussistere qualunque sia R , segue la (12.9). \square

La (12.9) è il preannunciato sviluppo differenziale dell'equazione di Smolukowski; esso prende il nome di sviluppo *in avanti*, o *forward*, in quanto, fissati lo stato x_0 e il tempo

t_0 iniziali, la (12.9) esprime l'evoluzione della densità di transizione al variare dello stato presente x e del tempo presente t .

Come vedremo, sussiste anche uno sviluppo *all'indietro*, o *backward*; in questo caso, fissati stato e tempo presente, tale sviluppo mostra come sia possibile pervenire ad essi al variare dello stato iniziale x_0 e del tempo iniziale t_0 .

Le funzioni $A_n(x, t)$ definite dalla (12.10) prendono il nome di *momenti infinitesimali* del processo $X(t)$. Fissato un valore x di $X(t)$ consideriamo ora la variabile casuale

$$\Delta X := X(t + \Delta t) - x$$

che esprime l'incremento ΔX del processo in corrispondenza della transizione dal fissato stato $x = X(t)$ all'istante t al generico stato $X(t + \Delta t)$ all'istante $t + \Delta t$. I limiti dei rapporti tra i momenti di tale variabile casuale condizionati da $X(t) = x$ e l'intervallo di tempo Δt rappresentano proprio i coefficienti $A_n(x, t)$:

$$A_n(x, t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} E\{[X(t + \Delta t) - x]^n | X(t) = x\}. \quad (12.15)$$

Supporre che i coefficienti A_n siano finiti significa supporre che per piccoli valori di Δt tali momenti condizionati siano dell'ordine di Δt ; questa proprietà costituisce un'altra caratterizzazione dei processi di Markov, i quali possono essere definiti come quei processi stocastici i cui momenti infinitesimali di ordine n sono per ogni n proporzionali a Δt . Si ha infatti:

$$E\{[X(t + \Delta t) - x]^n | X(t) = x\} \approx A_n(x, t) \Delta t. \quad (12.16)$$

In particolare, per $n = 1$ si ha:

$$A_1(x, t) \approx \frac{1}{\Delta t} \int_{-\infty}^{+\infty} (y - x) f(y, t + \Delta t | x, t) dy. \quad (12.17)$$

Il rapporto a secondo membro è un momento per unità di tempo il cui limite per $\Delta t \rightarrow 0$ prende il nome di *deriva* o *drift* del processo. Per $n = 2$ risulta:

$$A_2(x, t) \approx \frac{1}{\Delta t} \int_{-\infty}^{+\infty} (y - x)^2 f(y, t + \Delta t | x, t) dy. \quad (12.18)$$

Calcoliamo la varianza $D^2[\Delta X | X(t) = x]$ dell'incremento del processo. Si ha:

$$D^2[\Delta X | X(t) = x] \approx A_2(x, t) \Delta t - A_1^2(x, t) (\Delta t)^2,$$

da cui segue:

$$\frac{D^2[\Delta X | X(t) = x]}{\Delta t} \approx A_2(x, t) - A_1^2(x, t) \Delta t.$$

Passando al limite per $\Delta t \rightarrow 0$ si trae infine:

$$A_2(x, t) = \lim_{\Delta t \rightarrow 0} \frac{D^2[\Delta X | X(t) = x]}{\Delta t}, \quad (12.19)$$

che giustifica la denominazione di *varianza infinitesimale* che si vuole dare al momento del secondo ordine $A_2(x, t)$.

I momenti infinitesimali A_1 e A_2 rivestono un'importanza fondamentale per la caratterizzazione di una sottoclasse dei processi di Markov, quella dei cosiddetti *processi di diffusione* sui quali ci soffermeremo nel seguito.

Il passo successivo da compiere per la determinazione della densità di transizione consiste nell'individuare condizioni che permettano di riscrivere lo sviluppo (12.9) in forma di equazione differenziale. Si tratta, cioè, di determinare in quali ipotesi esiste un intero $\nu > 0$ tale che si abbia $A_n(x, t) = 0$ per $n \geq \nu$.

Riprendiamo a tal fine in considerazione lo sviluppo differenziale (12.9) e supponiamo che il processo di Markov considerato sia temporalmente omogeneo, ossia che per ogni coppia di reali x, y e per ogni coppia di istanti t, τ ($t > \tau$) la densità di transizione dipenda solo dalla differenza $t - \tau$ e non da t e τ separatamente:

$$f(x, t | y, \tau) = f(x, t - \tau | y, 0) = f(x, t - \tau | y). \quad (12.20)$$

In questo caso i coefficienti $A_n(x, t)$ dipendono al più dalla variabile x avendosi:

$$\begin{aligned} A_n(x, t) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{+\infty} (y - x)^n f(y, t + \Delta t | x, t) dy \\ &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{+\infty} (y - x)^n f(y, \Delta t | x) dy \\ &= A_n(x). \end{aligned}$$

Si può in realtà facilmente dimostrare la seguente proposizione.

Proposizione 12.2. Condizione necessaria e sufficiente affinché un processo di Markov sia temporalmente omogeneo è che i coefficienti $A_n(x, t)$ nello sviluppo (12.9) siano indipendenti da t .

Nel caso in cui il processo è temporalmente omogeneo può accadere che $\lim_{t \rightarrow \infty} f(x, t - \tau | y)$ esista; in tal caso esso è funzione solo di x e non di y :

$$S(x) = \lim_{t \rightarrow \infty} f(x, t - \tau | y). \quad (12.21)$$

La funzione S definita dalla (12.21), detta *densità stazionaria* o *di equilibrio*, soddisfa lo sviluppo differenziale (12.9) nel quale venga posto

$$\partial S(x) / \partial t = 0.$$

Lo sviluppo (12.9) è, come abbiamo osservato, uno sviluppo in avanti, o forward, nel senso che esso descrive come evolve il sottoinsieme delle realizzazioni del processo che al tempo t_0 occupano la posizione x_0 (processo condizionato dal valore x_0). Un diverso modo di studiare il processo si ottiene fissando lo stato x all'istante presente t e determinando come evolvono gli stati iniziali x_0 assunti all'istante t_0 che conducono allo stato x al tempo t . Riguardiamo quindi t_0 e x_0 come variabili, e x e t come parametri. La densità di transizione è evidentemente la stessa del caso forward, solo che ora, come si è detto, x_0 e t_0 sono da considerarsi variabili, mentre x e t sono supposti fissi. Lo sviluppo differenziale che si ottiene in questo caso si dice *all'indietro*, o *backward*. Al fine di determinare tale sviluppo consideriamo ancora l'equazione di Smolukowski (12.6). Sostituendo $t_0 - \Delta t$ a t_0 e t_0 a τ si ha:

$$f(x, t|x_0, t_0 - \Delta t) = \int_{-\infty}^{+\infty} f(x, t|y, t_0) f(y, t_0|x_0, t_0 - \Delta t) dy. \quad (12.22)$$

Osserviamo poi che, in virtù della condizione di normalizzazione

$$\int_{-\infty}^{+\infty} f(y, t|x_0, t_0 - \Delta t) dy = 1,$$

la funzione $f(x, t|x_0, t_0)$ può esprimersi al seguente modo:

$$f(x, t|x_0, t_0) = \int_{-\infty}^{+\infty} f(x, t|x_0, t_0) f(y, t_0|x_0, t_0 - \Delta t) dy. \quad (12.23)$$

Pertanto, facendo uso delle (12.22) e (12.23), si può scrivere:

$$\begin{aligned} & f(x, t|x_0, t_0 - \Delta t) - f(x, t|x_0, t_0) \\ &= \int_{-\infty}^{+\infty} f(x, t|y, t_0) f(y, t_0|x_0, t_0 - \Delta t) dy + \\ & \quad - \int_{-\infty}^{+\infty} f(x, t|x_0, t_0) f(y, t_0|x_0, t_0 - \Delta t) dy \\ &= \int_{-\infty}^{+\infty} f(y, t_0|x_0, t_0 - \Delta t) [f(x, t|y, t_0) - f(x, t|x_0, t_0)] dy. \end{aligned} \quad (12.24)$$

Sviluppiamo ora $f(x, t|y, t_0)$ in serie di Taylor di punto iniziale x_0 :

$$f(x, t|y, t_0) - f(x, t|x_0, t_0) = \sum_{n=1}^{\infty} \frac{\partial^n f(x, t|x_0, t_0)}{\partial x_0^n} \frac{(y - x_0)^n}{n!}. \quad (12.25)$$

Sostituendo la (12.25) nella (12.24), dividendo per Δt e passando al limite per $\Delta t \rightarrow 0$, si ottiene:

$$\frac{\partial f(x, t|x_0, t_0)}{\partial t_0} = - \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \sum_{n=1}^{\infty} \frac{1}{n!} \int_{-\infty}^{+\infty} (y - x_0)^n \frac{\partial^n f(x, t|x_0, t_0)}{\partial x_0^n} f(y, t_0|x_0, t_0 - \Delta t) dy,$$

e quindi:

$$\frac{\partial f(x, t|x_0, t_0)}{\partial t_0} = - \sum_{n=1}^{\infty} \frac{1}{n!} \frac{\partial^n f(x, t|x_0, t_0)}{\partial x_0^n} A_n(x_0, t_0), \quad (12.26)$$

dove per $n = 1, 2, \dots$ si è posto:

$$A_n(x_0, t_0) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{+\infty} (y - x_0)^n f(y, t_0|x_0, t_0 - \Delta t) dy. \quad (12.27)$$

La (12.26), con i coefficienti (12.27), costituisce il desiderato sviluppo differenziale all'indietro dell'equazione di Smolukowski.

Si noti che le funzioni (12.27) coincidono con i momenti infinitesimali (12.10).

12.4 Equazioni di diffusione

Gli sviluppi differenziali (12.9) e (12.26) non permettono in generale di determinare la densità di transizione del processo in quanto in essi compaiono derivate di ordine comunque elevato. Tali sviluppi si riducono invece a equazioni differenziali alle derivate parziali se i momenti infinitesimali si annullano da un certo ordine n in poi. In tal caso è possibile ricercare le soluzioni di tali equazioni e quindi pervenire, a mezzo di tecniche analitiche o numeriche, alla densità di transizione del processo in forma esatta o approssimata.

Sussiste il seguente teorema.

Teorema 12.2 (Pawula). *Si consideri un processo di Markov avente finito il momento infinitesimale $A_n(x, t)$ per ogni $n \in \mathbb{N}$. Per un tale processo l'annullarsi di un momento infinitesimale di ordine pari implica $A_n(x, t) = 0$ per ogni $n \geq 3$.*

Dimostrazione. La dimostrazione consiste essenzialmente in una applicazione della disegualanza di Schwarz che qui, dopo aver indicato con $h(x)$ e $k(x)$ funzioni reali integrabili in \mathbb{R} , scriviamo al seguente modo:

$$\left(\int_{-\infty}^{+\infty} h(z)k(z) dz \right)^2 \leq \int_{-\infty}^{+\infty} h^2(z) dz \int_{-\infty}^{+\infty} k^2(z) dz \quad (12.28)$$

(i) Sia n dispari e $n \geq 3$.

Poniamo:

$$h(z) = \frac{1}{\sqrt{\Delta t}} (z-x)^{(n-1)/2} [f(z, t + \Delta t | x, t)]^{1/2} \quad (12.29)$$

$$k(z) = \frac{1}{\sqrt{\Delta t}} (z-x)^{(n+1)/2} [f(z, t + \Delta t | x, t)]^{1/2}. \quad (12.30)$$

Così che la (12.28), con le scelte (12.29) e (12.30), diviene:

$$\left[\frac{1}{\Delta t} \int_{-\infty}^{+\infty} (z-x)^n f(z, t + \Delta t | x, t) dz \right]^2 \leq \frac{1}{\Delta t} \int_{-\infty}^{+\infty} (z-x)^{(n-1)} f(z, t + \Delta t | x, t) dz \\ \times \frac{1}{\Delta t} \int_{-\infty}^{+\infty} (z-x)^{(n+1)} f(z, t + \Delta t | x, t) dz.$$

Di qui, passando al limite per $\Delta t \rightarrow 0^+$, si ottiene:

$$A_n^2(x, t) \leq A_{n-1}(x, t) A_{n+1}(x, t). \quad (12.31)$$

(ii) Sia n pari e $n \geq 4$.

Procedendo in maniera analoga al caso (i), si ottiene:

$$A_n^2(x, t) \leq A_{n-2}(x, t) A_{n+2}(x, t). \quad (12.32)$$

Indichiamo ora con r un intero pari. Ponendo $n = r - 1$ e $n = r + 1$ nella (12.31) otteniamo le seguenti diseguaglianze:

$$A_{r-1}^2(x, t) \leq A_{r-2}(x, t) A_r(x, t), \quad r \geq 4; \quad (12.33)$$

$$A_{r+1}^2(x, t) \leq A_r(x, t) A_{r+2}(x, t), \quad r \geq 2. \quad (12.34)$$

Inoltre per $n = r - 2$ e $n = r + 2$ nella (12.32) otteniamo le ulteriori seguenti diseguaglianze:

$$A_{r-2}^2(x, t) \leq A_{r-4}(x, t) A_r(x, t), \quad r \geq 6; \quad (12.35)$$

$$A_{r+2}^2(x, t) \leq A_r(x, t) A_{r+4}(x, t), \quad r \geq 2. \quad (12.36)$$

Ricordiamo ora che per ipotesi esiste finito il momento infinitesimale di ordine qualsiasi; supposto che risulti $A_r(x, t) = 0$ con r intero pari ed esaminiamo i seguenti tre casi esaustivi.

Caso $r \geq 6$. Valgono tutte le diseguaglianze (12.33)-(12.36) che comportano l'annullarsi di $A_{r-2}(x, t)$, $A_{r-1}(x, t)$, $A_{r+1}(x, t)$ e $A_{r+2}(x, t)$. In particolare, per induzione si ricava che $A_n(x, t) = 0$ per ogni $n \geq r$. Allo stesso modo, per iterazione all'indietro, si ottiene che $A_n(x, t) = 0$ per tutti i valori di n da r a 4. Dalla (12.33) segue poi che $n = 3$ è il minimo valore di n per il quale $A_n(x, t)$ si annulla.

Caso $r = 4$. Vale la diseguaglianza (12.36) dalla quale discende $A_6(x, t) = 0$ che ci riporta al caso $r \geq 6$.

Caso $r = 2$. Vale la diseguaglianza (12.36) dalla quale discende $A_4(x, t) = 0$ che ci riporta al caso $r = 4$. \square

Dal Teorema (12.2) segue immediatamente che se negli sviluppi differenziali (12.9) e (12.26) è presente un numero finito di addendi, tali sviluppi consistono di fatto in equazioni alle derivate parziali al più del secondo ordine.

Consideriamo i seguenti casi notevoli.

1. $A_n(x, t) = 0$, $n \geq 2$. Si ottengono allora le equazioni

$$\frac{\partial f}{\partial t} = -\frac{\partial}{\partial x} [A_1(x, t) f] \quad (\text{equazione forward}) \quad (12.37)$$

$$\text{e} \quad \frac{\partial f}{\partial t_0} + A_1(x_0, t_0) \frac{\partial f}{\partial x_0} = 0 \quad (\text{equazione backward}). \quad (12.38)$$

Queste sono equazioni differenziali a derivate parziali del primo ordine la cui soluzione (una volta assegnate le condizioni iniziali) è determinabile con metodi standard.

2. $A_n(x, t) = 0$, $n > 2$. In questo caso si ottengono le equazioni

$$\frac{\partial f}{\partial t} = -\frac{\partial}{\partial x} [A_1(x, t) f] + \frac{1}{2} \frac{\partial^2}{\partial x^2} [A_2(x, t) f] \quad (\text{equazione forward}) \quad (12.39)$$

$$\text{e} \quad \frac{\partial f}{\partial t_0} + A_1(x_0, t_0) \frac{\partial f}{\partial x_0} + \frac{1}{2} A_2(x_0, t_0) \frac{\partial^2 f}{\partial x_0^2} = 0 \quad (\text{equazione backward}). \quad (12.40)$$

Le (12.39) e (12.40) sono equazioni differenziali a derivate parziali del secondo ordine, dette *equazioni di diffusione*. La (12.39) è anche detta *equazione di Fokker-Planck*, mentre la (12.40) è anche nota come *equazione di Kolmogorov*.

La condizione iniziale cui deve soddisfare la densità di transizione f sia nel caso dell'equazione (12.39) che nel caso dell'equazione (12.40) è la seguente:

$$\lim_{t \rightarrow t_0} f(x, t | x_0, t_0) = \delta(x - x_0), \quad (12.41)$$

dove δ è la funzione delta di Dirac:

$$\delta(x) = \begin{cases} 0, & x \neq 0, \\ \infty, & x = 0, \end{cases}$$

che, com'è noto (cfr. il § B.1), gode della proprietà

$$\int_a^b \delta(x) dx = 1 \quad \forall a, b: a < 0 < b.$$

Nel caso dei processi temporalmente omogenei, per i quali i momenti infinitesimali non dipendono da t , le (12.39) e (12.40) assumono rispettivamente le forme

$$\frac{\partial f}{\partial t} = -\frac{\partial}{\partial x}[A_1(x)f] + \frac{1}{2} \frac{\partial^2}{\partial x^2}[A_2(x)f] \quad (12.42)$$

e

$$\frac{\partial f}{\partial t_0} + A_1(x_0) \frac{\partial f}{\partial x_0} + \frac{1}{2} A_2(x_0) \frac{\partial^2 f}{\partial x_0^2} = 0. \quad (12.43)$$

In particolare, per processi temporalmente omogenei, in virtù della (12.20), risulta

$$f(x, t|x_0, t_0) = f(x, t - t_0|x_0),$$

e quindi:

$$\frac{\partial f}{\partial t_0} = -\frac{\partial f}{\partial t}. \quad (12.44)$$

Infatti, posto $T = t - t_0$ si ha:

$$\frac{\partial f}{\partial t_0} = \frac{\partial f}{\partial T} \frac{\partial T}{\partial t_0} = -\frac{\partial f}{\partial T}, \quad \frac{\partial f}{\partial t} = \frac{\partial f}{\partial T} \frac{\partial T}{\partial t} = \frac{\partial f}{\partial T},$$

da cui segue la (12.44). Dalle (12.43) e (12.44) si trae allora:

$$\frac{\partial f}{\partial t} = A_1(x_0) \frac{\partial f}{\partial x_0} + \frac{1}{2} A_2(x_0) \frac{\partial^2 f}{\partial x_0^2}. \quad (12.45)$$

La (12.45) è un'equazione in cui compaiono variabili miste: tempo presente e stato iniziale.

12.5 Densità stazionaria

Supponiamo che il processo sia temporalmente omogeneo e che esista la densità stazionaria S definita dalla (12.21). La funzione S deve soddisfare l'equazione di Fokker-Planck (12.39), che quindi si riduce all'equazione differenziale ordinaria

$$-\frac{d}{dx}[A_1(x)S(x)] + \frac{1}{2} \frac{d^2}{dx^2}[A_2(x)S(x)] = 0. \quad (12.46)$$

Questa può risciversi nella forma

$$\frac{1}{2} \frac{d}{dx} \left\{ \frac{d}{dx}[A_2(x)S(x)] - 2A_1(x)S(x) \right\} = 0, \quad (12.47)$$

da cui segue:

$$\frac{d}{dx}[A_2(x)S(x)] - 2A_1(x)S(x) = \text{cost.} \quad (12.48)$$

Almeno per i processi di diffusione di cui ci occuperemo, ci si può poi convincere che la costante arbitraria a secondo membro della (12.48) deve essere nulla così che

$$\frac{d}{dx}[A_2(x)S(x)] - 2A_1(x)S(x) = 0. \quad (12.49)$$

La (12.49) è un'equazione lineare omogenea il cui integrale generale è il seguente:

$$S(x) = \frac{c}{A_2(x)} \exp \left\{ 2 \int_{x_1}^x \frac{A_1(y)}{A_2(y)} dy \right\} \quad (12.50)$$

con c e x_1 costanti arbitrarie.² Poiché S deve essere una densità di probabilità, la costante c nella (12.50) va determinata in modo che risulti:

$$\int_{-\infty}^{+\infty} S(x) dx = 1. \quad (12.51)$$

Dalle (12.50) e (12.51) segue:

$$c = \left\{ \int_{-\infty}^{+\infty} \frac{1}{A_2(x)} \exp \left[2 \int_{x_1}^x \frac{A_1(y)}{A_2(y)} dy \right] dx \right\}^{-1}. \quad (12.52)$$

Se è $c \neq 0$, la (12.50), con c data dalla (12.52), fornisce la densità di probabilità stazionaria del processo. Se quindi l'integrale a secondo membro della (12.52) è convergente, c è una costante non nulla e la densità stazionaria S esiste effettivamente; se invece l'integrale diverge, la costante c è nulla, e quindi il processo non possiede densità stazionaria.

Notiamo, infine, che la densità stazionaria è espressa in termini di integrali che coinvolgono solo la deriva A_1 e la varianza infinitesimale A_2 del processo che risultano evidentemente note una volta che siano assegnate le equazioni (12.39) o (12.40). Facendo allora uso della (12.50), la densità stazionaria può essere determinata in base alla sola conoscenza di A_1 e A_2 . Di conseguenza le proprietà stazionarie del processo possono essere analizzate, almeno in linea di principio, ossia compatibilmente con le connesse difficoltà computazionali, senza dover preventivamente risolvere le equazioni (12.39) e (12.40).

²Si noti che le costanti c e x_1 non sono indipendenti; variare una costante equivale a variare l'altra, così che una sola di esse è da considerarsi realmente arbitraria.

12.6 Il processo di Wiener senza drift

In questo paragrafo fisseremo l'attenzione sul più semplice esempio di processo di diffusione, quello caratterizzato dai momenti infinitesimali

$$A_1(x, t) = 0$$

$$A_2(x, t) = \sigma^2,$$

con σ costante positiva. Esso prende il nome di *processo di Wiener senza drift di varianza infinitesimale σ^2* . Le equazioni di diffusione in questo caso assumono la forma seguente:

$$\frac{\partial f}{\partial t} = \frac{\sigma^2}{2} \frac{\partial^2 f}{\partial x^2}, \quad (\text{equazione forward}), \quad (12.53)$$

$$\frac{\partial f}{\partial t_0} = -\frac{\sigma^2}{2} \frac{\partial^2 f}{\partial x_0^2}, \quad (\text{equazione backward}). \quad (12.54)$$

Si noti che essendo A_1 e A_2 indipendenti da t , il processo di Wiener è temporalmente omogeneo. Per determinarne la densità di transizione risolveremo l'equazione (12.53) con la condizione iniziale

$$\lim_{t \rightarrow t_0} f(x, t|x_0, t_0) = \delta(x - x_0). \quad (12.55)$$

In modo analogo si può operare con l'equazione (12.54). Per la temporale omogeneità del processo, risulta:

$$f(x, t|x_0, t_0) = f(x, t - t_0|x_0); \quad (12.56)$$

sarà quindi sufficiente risolvere l'equazione

$$\frac{\partial}{\partial t} f(x, t|x_0) = \frac{\sigma^2}{2} \frac{\partial^2}{\partial x^2} f(x, t|x_0), \quad (12.57)$$

con la condizione iniziale

$$\lim_{t \rightarrow 0} f(x, t|x_0) = \delta(x - x_0), \quad (12.58)$$

e poi risalire alla funzione $f(x, t|x_0, t_0)$ facendo uso della (12.56).³ Indichiamo a tal fine con $\varphi(\lambda, t|x_0)$ la trasformata di Fourier (funzione caratteristica) della densità di transizione:

$$\varphi(\lambda, t|x_0) = \int_{-\infty}^{+\infty} e^{i\lambda x} f(x, t|x_0) dx. \quad (12.59)$$

³Per non appesantire eccessivamente le dimostrazioni, in questo e nei successivi paragrafi supporremo euristicamente leciti i necessari passaggi al limite sotto segno di integrale.

12.6 Il processo di Wiener senza drift

Proposizione 12.3. *La funzione $\varphi(\lambda, t|x_0)$ è soluzione dell'equazione*

$$\frac{\partial}{\partial t} \varphi(\lambda, t) = -\frac{\sigma^2}{2} \lambda^2 \varphi(\lambda, t) \quad (12.60)$$

e soddisfa la condizione iniziale

$$\lim_{t \rightarrow 0} \varphi(\lambda, t|x_0) = e^{i\lambda x_0}. \quad (12.61)$$

Dimostrazione. Denotiamo con \mathcal{F} l'operatore che, applicato ad una funzione g , ne fornisce la trasformata di Fourier $\mathcal{F}[g]$. Applicando \mathcal{F} ad entrambi i membri della (12.57), otteniamo:

$$\mathcal{F}\left[\frac{\partial f}{\partial t}\right] = \frac{\sigma^2}{2} \mathcal{F}\left[\frac{\partial^2 f}{\partial x^2}\right]. \quad (12.62)$$

Calcoliamo le trasformate che compaiono nella (12.62). Facendo uso della (12.59), si trae

$$\mathcal{F}\left[\frac{\partial f}{\partial t}\right] \equiv \int_{-\infty}^{+\infty} e^{i\lambda x} \frac{\partial f}{\partial t} dx = \frac{\partial}{\partial t} \varphi(\lambda, t). \quad (12.63)$$

Inoltre risulta:

$$\begin{aligned} \mathcal{F}\left[\frac{\partial^2 f}{\partial x^2}\right] &\equiv \int_{-\infty}^{+\infty} e^{i\lambda x} \frac{\partial^2 f}{\partial x^2} dx \\ &= \left[e^{i\lambda x} \frac{\partial f}{\partial x} \right]_{-\infty}^{+\infty} - i\lambda \int_{-\infty}^{+\infty} e^{i\lambda x} \frac{\partial f}{\partial x} dx \\ &= -i\lambda \left[e^{i\lambda x} f \right]_{-\infty}^{+\infty} + (i\lambda)^2 \int_{-\infty}^{+\infty} e^{i\lambda x} f dx \\ &= -\lambda^2 \int_{-\infty}^{+\infty} e^{i\lambda x} f dx \equiv -\lambda^2 \varphi(\lambda, t) \end{aligned} \quad (12.64)$$

in cui si è euristicamente assunto che la derivata rispetto a x della densità di transizione tende a zero per $x \rightarrow \pm\infty$. Sostituendo le (12.63) e (12.64) nella (12.62) si ottiene la (12.60). La (12.61) segue poi immediatamente dalle (12.58) e (12.59) avendosi

$$\lim_{t \rightarrow 0} \varphi(\lambda, t|x_0) = \int_{-\infty}^{+\infty} e^{i\lambda x} \delta(x - x_0) dx. \quad \square$$

Proposizione 12.4. *La funzione $\varphi(\lambda, t|x_0)$ è univocamente determinata come la soluzione dell'equazione (12.60) soddisfacente la condizione iniziale (12.61).*

Dimostrazione. L'integrale generale $\varphi(\lambda, t)$ della (12.60) è dato da

$$\varphi(\lambda, t) = k(\lambda) \exp\left(-\frac{\sigma^2 \lambda^2}{2} t\right), \quad (12.65)$$

dove $k(\lambda)$ denota una funzione arbitraria di λ . Tale funzione è peraltro univocamente determinata dalla (12.61). Invero, passando al limite per $t \rightarrow 0$ nella (12.65) e facendo uso della (12.61) si ottiene

$$k(\lambda) = e^{i\lambda x_0}. \quad (12.66)$$

Con $k(\lambda)$ dato dalla (12.66), la (12.65) fornisce infine l'integrale particolare

$$\varphi(\lambda, t|x_0) = \exp\left(-\frac{\sigma^2 t}{2} \lambda^2 + i x_0 \lambda\right) \quad (12.67)$$

che è dunque la soluzione della (12.60) che soddisfa la (12.61). \square

Proposizione 12.5. La densità di transizione $f(x, t|x_0)$, soluzione dell'equazione (12.57) soddisfacente la condizione iniziale (12.58), è data da

$$f(x, t|x_0) = \frac{1}{\sqrt{2\pi t} \sigma} \exp\left\{-\frac{(x - x_0)^2}{2\sigma^2 t}\right\} \quad (x \in \mathbb{R}). \quad (12.68)$$

Dimostrazione. La funzione $\varphi(\lambda, t|x_0)$ è stata definita come la trasformata di Fourier di $f(x, t|x_0)$. Quest'ultima si ottiene pertanto come antitrasformata di Fourier di $\varphi(\lambda, t|x_0)$:

$$f(x, t|x_0) \equiv \mathcal{F}^{-1}[\varphi(\lambda, t|x_0)] = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i\lambda x} \varphi(\lambda, t|x_0) d\lambda. \quad (12.69)$$

Dalle (12.67) e (12.69) segue:

$$\begin{aligned} f(x, t|x_0) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp\left\{-\left[\sigma^2 t \frac{\lambda^2}{2} + i\lambda(x - x_0)\right]\right\} d\lambda \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp\left\{-\frac{(x - x_0)^2}{2\sigma^2 t}\right\} \exp\left\{-\left[\sigma \sqrt{\frac{t}{2}} \lambda + i \frac{(x - x_0)}{2\sigma \sqrt{t/2}}\right]^2\right\} d\lambda \\ &= \frac{1}{2\pi} \exp\left\{-\frac{(x - x_0)^2}{2\sigma^2 t}\right\} \int_{-\infty}^{+\infty} \frac{e^{-z^2}}{\sigma \sqrt{t/2}} dz, \quad (x \in \mathbb{R}), \end{aligned}$$

e quindi la (12.68). Alternativamente, ricordando la (9.37), si può osservare che la (12.67) è la funzione caratteristica di una v.c. normale di media x_0 e varianza $\sigma^2 t$ (v. § 9.4); da ciò segue immediatamente la (12.68). \square

12.6 Il processo di Wiener senza drift

Si noti che l'ipotesi di annullamento di $\partial f / \partial x$ per $x \rightarrow \pm\infty$ risulta a posteriori giustificata in quanto di tale proprietà gode la densità di probabilità (12.68) cui si è pervenuti.

Dall'espressione (12.68) si vede che la massa probabilistica unitaria, inizialmente concentrata nel punto x_0 , ad ogni istante $t > 0$ è distribuita in maniera gaussiana con le seguenti caratteristiche:

- (a) la media resta costantemente uguale a x_0 (posizione iniziale);
- (b) la varianza cresce linearmente con il tempo.

Si hanno quindi densità gaussiane che, al trascorrere del tempo, progressivamente si allargano, rimanendo peraltro centrate nel punto x_0 .

Calcoliamo

$$P(s) = \lim_{t \rightarrow +\infty} \int_{-\infty}^s f(x, t|x_0) dx = \lim_{t \rightarrow +\infty} \mathbb{P}(W(t) \leq s)$$

ossia la probabilità che asintoticamente il processo $W(t)$ di Wiener considerato assuma valori non superiori ad s . Facendo uso della (12.68), dopo il cambiamento della variabile di integrazione $z = (x - x_0)/\sqrt{2t}\sigma$ si ottiene:

$$P(s) = \frac{1}{\sqrt{\pi}} \lim_{t \rightarrow \infty} \int_{-\infty}^{(s-x_0)/\sqrt{2t}\sigma} e^{-z^2} dz = \frac{1}{\sqrt{\pi}} \int_{-\infty}^0 e^{-z^2} dz = \frac{1}{2}.$$

Di conseguenza si ha anche:

$$\lim_{t \rightarrow \infty} \mathbb{P}(W(t) > s) = 1 - P(s) = \frac{1}{2}.$$

Si vede così che con uguali probabilità il processo di Wiener senza drift asintoticamente assume valori al di sopra o al di sotto di qualsiasi stato s preassegnato. In maniera analoga si può mostrare che, comunque fissata una coppia di stati u e v ($u < v$), asintoticamente è nulla la probabilità che il processo assuma valori compresi nell'intervallo (u, v) ; in altri termini è certa l'uscita asintotica di $W(t)$ da un qualsiasi intervallo preassegnato dello spazio degli stati.

Osserviamo anche che dalle (12.56) e (12.68) per ogni $t > t_0$ segue:

$$f(x, t|x_0, t_0) = \frac{1}{\sqrt{2\pi(t-t_0)\sigma}} \exp\left[-\frac{(x-x_0)^2}{2\sigma^2(t-t_0)}\right] \quad (x \in \mathbb{R}). \quad (12.70)$$

Va infine detto che nel caso in cui risulta $\sigma^2 = 1$, il processo $W(t)$ viene denominato *processo di Wiener standard*, o anche *moto browniano standard*.

12.7 Il processo di Wiener con drift

Una naturale generalizzazione del processo di Wiener senza drift è costituita dal processo $W_\mu(t)$ avente momenti infinitesimali

$$A_1(x, t) = \mu, \quad A_2(x, t) = \sigma^2,$$

con μ costante non nulla e σ costante positiva. La densità di transizione $f(x, t|x_0)$ di tale processo, anch'esso temporalmente omogeneo, può allora determinarsi come soluzione dell'equazione

$$\frac{\partial f}{\partial t} = -\mu \frac{\partial f}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 f}{\partial x^2}, \quad (12.71)$$

soddisfacente la condizione iniziale

$$\lim_{t \rightarrow 0} f(x, t|x_0) = \delta(x - x_0). \quad (12.72)$$

La determinazione di $f(x, t|x_0)$ si effettua in maniera analoga a quanto già visto per il processo di Wiener senza drift. A tal fine riferiamoci alla (12.59) in cui ora φ denota la trasformata di Fourier dell'incognita densità di transizione. Sussiste la seguente proposizione:

Proposizione 12.6. *La funzione $\varphi(\lambda, t|x_0)$ è soluzione dell'equazione*

$$\frac{\partial}{\partial t} \varphi(\lambda, t) = -\left(\frac{\sigma^2}{2}\lambda^2 - i\mu\lambda\right) \varphi(\lambda, t) \quad (12.73)$$

e soddisfa la condizione iniziale (12.61).

Dimostrazione. Passando alle trasformate di Fourier nella (12.71) si ottiene:

$$\mathcal{F}\left[\frac{\partial f}{\partial t}\right] = -\mu \mathcal{F}\left[\frac{\partial f}{\partial x}\right] + \frac{1}{2}\sigma^2 \mathcal{F}\left[\frac{\partial^2 f}{\partial x^2}\right], \quad (12.74)$$

dove $\mathcal{F}\left[\frac{\partial f}{\partial t}\right]$ e $\mathcal{F}\left[\frac{\partial^2 f}{\partial x^2}\right]$ sono date dalle (12.63) e (12.64). Resta da calcolare $\mathcal{F}\left[\frac{\partial f}{\partial x}\right]$. A tal fine osserviamo che risulta:

$$\begin{aligned} \mathcal{F}\left[\frac{\partial f}{\partial x}\right] &\equiv \int_{-\infty}^{+\infty} e^{i\lambda x} \frac{\partial f}{\partial x} dx \\ &= [e^{i\lambda x} f]_{-\infty}^{+\infty} - i\lambda \int_{-\infty}^{+\infty} e^{i\lambda x} f dx \\ &= -i\lambda \varphi(\lambda, t), \end{aligned} \quad (12.75)$$

avendo utilizzato l'annullarsi di f a $\pm\infty$. Sostituendo le (12.63), (12.64) e (12.75) nella (12.74) segue la (12.73). La dimostrazione, infine, che φ soddisfa la condizione (12.61) non differisce da quella relativa al caso del processo di Wiener senza drift. \square

12.7 Il processo di Wiener con drift

Proposizione 12.7. *La funzione $\varphi(\lambda, t|x_0)$ è univocamente determinata come la soluzione dell'equazione (12.73) soddisfacente la condizione (12.61).*

Dimostrazione. L'integrale generale della (12.73) è

$$\varphi(\lambda, t) = k(\lambda) \exp\left\{-\left(\frac{\sigma^2\lambda^2}{2} - \mu i\lambda\right)t\right\}, \quad (12.76)$$

dove $k(\lambda)$ è una funzione arbitraria. Questa, in virtù della (12.61), è ancora data dalla (12.66), di modo che dalla (12.76) si ricava:

$$\varphi(\lambda, t|x_0) = \exp\left\{-\frac{\sigma^2\lambda^2}{2}t + i\lambda(x_0 + \mu t)\right\}. \quad (12.77) \quad \square$$

La (12.77) è la funzione caratteristica di una variabile casuale normale di media $x_0 + \mu t$ e varianza $\sigma^2 t$. La densità di transizione del processo di Wiener con drift $W_\mu(t)$ è dunque:

$$f(x, t|x_0) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left\{-\frac{(x - x_0 - \mu t)^2}{2\sigma^2 t}\right\} \quad (x \in \mathbb{R}). \quad (12.78)$$

La presenza del drift μ fa sì che ora, come la varianza, anche la media di $W_\mu(t)$ varia linearmente nel tempo, crescendo per $\mu > 0$ e decrescendo per $\mu < 0$. In altri termini a causa del drift la massa di probabilità, pur rimanendo distribuita in maniera gaussiana sull'intero asse reale ad ogni t finito, subisce una deriva nel senso positivo se è $\mu > 0$ e nel senso negativo se è $\mu < 0$. Il comportamento asintotico del processo si analizza in maniera analoga a quanto prima fatto per il processo di Wiener senza drift, avendosi ora

$$P(s) = \lim_{t \rightarrow +\infty} \mathbb{P}(W_\mu(t) \leq s) = \begin{cases} 1, & \text{per } \mu < 0, \\ \frac{1}{2}, & \text{per } \mu = 0, \\ 0, & \text{per } \mu > 0. \end{cases}$$

Questo risultato può essere interpretato nel modo seguente:

- (a) se è $\mu < 0$ le realizzazioni del processo tendono a $-\infty$ con probabilità unitaria;
- (b) se è $\mu = 0$, fissato un qualunque reale s le realizzazioni del processo si trovano sia al di sopra che al di sotto di questo con uguali probabilità (caso del processo di Wiener senza drift);
- (c) se è $\mu > 0$ le realizzazioni del processo tendono a $+\infty$ con probabilità unitaria.

Ancora una volta si vede facilmente che è unitaria la probabilità di uscita asintotica da una qualsiasi striscia relativa ad un intervallo di stati (u, v) qualunque siano i reali u, v ($u < v$).

Poiché il processo di Wiener con drift è temporalmente omogeneo, dalle (12.56) e (12.78) segue poi:

$$f(x, t|x_0, t_0) = \frac{1}{\sqrt{2\pi(t-t_0)\sigma^2}} \exp\left\{-\frac{|x-x_0-\mu(t-t_0)|^2}{2\sigma^2(t-t_0)}\right\} \quad (x \in \mathbb{R}). \quad (12.79)$$

Osserviamo che, qualunque sia μ , il processo di Wiener non possiede densità stazionaria avendosi:

$$\lim_{t \rightarrow +\infty} f(x, t|x_0, t_0) = 0.$$

Ciò poteva peraltro stabilirsi a priori (ossia senza risolvere l'equazione differenziale) utilizzando la (12.52). Tenendo conto dei valori di $A_1(x, t)$ e $A_2(x, t)$ in questo caso si ha infatti:

$$\begin{aligned} c &= \left\{ \int_{-\infty}^{+\infty} \frac{1}{\sigma^2} \exp\left[2 \int_{x_1}^x \frac{\mu}{\sigma^2} dy\right] dx \right\}^{-1} \\ &= \sigma^2 \left\{ \int_{-\infty}^{+\infty} \exp\left[\frac{2\mu}{\sigma^2}(x-x_1)\right] dx \right\}^{-1}. \end{aligned} \quad (12.80)$$

La costante c risulta quindi essere nulla in quanto l'integrale nella (12.80) è divergente per ogni μ .

La forma (12.78) della densità di transizione suggerisce che il processo di Wiener $W_\mu(t)$ con drift μ può ricondursi al processo di Wiener senza drift $W(t)$ mediante una trasformazione che corrisponda ad osservare $W_\mu(t)$ da un sistema di ascisse in moto con velocità μ rispetto al sistema utilizzato per descrivere $W_\mu(t)$. Sussiste, invero, la seguente proposizione:

Proposizione 12.8. La trasformazione $x' = x - \mu t$, $x'_0 = x_0 - \mu t_0$, $t' = t$, $f'(x', t'|x'_0) = f(x, t|x_0)$ trasforma l'equazione (12.71) nell'equazione

$$\frac{\partial f'}{\partial t'} = \frac{\sigma^2}{2} \frac{\partial^2 f'}{\partial x'^2} \quad (12.81)$$

e la (12.58) nella condizione

$$\lim_{t' \rightarrow 0} f'(x', t'|x'_0) = \delta(x' - x'_0).$$

Dimostrazione. La trasformazione considerata implica:

$$\frac{\partial f}{\partial t} = \frac{\partial f'}{\partial t'} - \mu \frac{\partial f'}{\partial x'}, \quad \frac{\partial f}{\partial x} = \frac{\partial f'}{\partial x'}, \quad \frac{\partial^2 f}{\partial x^2} = \frac{\partial^2 f'}{\partial x'^2},$$

e quindi la (12.71) si trasforma nella (12.81). La stessa trasformazione applicata alla (12.58) dà infine:

$$\begin{aligned} \lim_{t' \rightarrow 0} f'(x' + \mu t', t'|x'_0 + \mu t') &\equiv \lim_{t' \rightarrow 0} f'(x', t', t'|x'_0) = \delta[x' + \mu t' - (x'_0 + \mu t')] \\ &= \delta(x' - x'_0). \end{aligned} \quad \square$$

La Proposizione 12.8 consente dunque di ottenere immediatamente la densità di transizione del processo $W_\mu(t)$ a partire da quella di $W(t)$. Basterà invero riscrivere la (12.68) con $x - \mu t$ al posto di x .

12.8 Il processo di Ornstein-Uhlenbeck

Consideriamo l'esempio più semplice in cui uno dei momenti infinitesimali è funzione della variabile di stato. Precisamente, riferiamoci al processo di diffusione $X(t)$ avente i seguenti momenti infinitesimali:

$$A_1(x, t) = -\alpha x \quad (12.82)$$

$$A_2(x, t) = \sigma^2, \quad (12.83)$$

con α e σ costanti positive; ad esso si dà il nome di *processo di Ornstein-Uhlenbeck*. L'equazione di Fokker-Planck in questo caso assume la forma

$$\frac{\partial f}{\partial t} = \alpha \frac{\partial}{\partial x}(x f) + \frac{1}{2} \sigma^2 \frac{\partial^2 f}{\partial x^2}. \quad (12.84)$$

Trattandosi di un processo temporalmente omogeneo, risolveremo la (12.84) con la solita condizione iniziale (12.55). Consideriamo anche in questo caso la trasformata (12.59) dell'incognita densità di transizione. Sussiste la seguente proposizione:

Proposizione 12.9. Nelle ipotesi in cui risulta

$$\lim_{x \rightarrow \pm\infty} \frac{\partial}{\partial x} f(x, t|x_0) = 0, \quad \lim_{x \rightarrow \pm\infty} [x f(x, t|x_0)] = 0,$$

la funzione $\varphi(\lambda, t|x_0)$ soddisfa l'equazione

$$\frac{\partial}{\partial t} \varphi(\lambda, t) + \alpha \lambda \frac{\partial}{\partial \lambda} \varphi(\lambda, t) = -\frac{1}{2} \sigma^2 \lambda^2 \varphi(\lambda, t). \quad (12.85)$$

Dimostrazione. Passando alle trasformate di Fourier, a partire dalla (12.84) si ottiene l'equazione

$$\mathcal{F}\left[\frac{\partial f}{\partial t}\right] = \alpha \mathcal{F}\left[\frac{\partial}{\partial x}(x f)\right] + \frac{1}{2} \sigma^2 \mathcal{F}\left[\frac{\partial^2 f}{\partial x^2}\right]. \quad (12.86)$$

Calcoliamo $\mathcal{F} \left[\frac{\partial}{\partial x} (x f) \right]$. Risulta evidentemente:

$$\begin{aligned}\mathcal{F} \left[\frac{\partial}{\partial x} (x f) \right] &= \int_{-\infty}^{+\infty} e^{i\lambda x} \frac{\partial}{\partial x} (x f) dx \\ &= [xe^{i\lambda x}]_{-\infty}^{+\infty} - i\lambda \int_{-\infty}^{+\infty} e^{i\lambda x} x f dx \\ &= -\lambda \frac{\partial}{\partial \lambda} \varphi(\lambda, t).\end{aligned}\quad (12.87)$$

La (12.85) segue sostituendo la (12.87) nella (12.86) e facendo uso delle (12.63) e (12.64). \square

Proposizione 12.10. La funzione $\varphi(\lambda, t|x_0)$ è univocamente determinata come la soluzione dell'equazione (12.85) che soddisfa la condizione iniziale (12.61).

Dimostrazione. La (12.85) è un'equazione a derivate parziali del primo ordine quasi lineare che può essere risolta in vari modi, ad esempio costruendo due integrali primi del sistema caratteristico.

$$\frac{dt}{1} = \frac{d\lambda}{\alpha\lambda} = \frac{d\varphi}{-\sigma^2\lambda^2\varphi/2}. \quad (12.88)$$

Dalla prima e dalla seconda delle (12.88) segue:

$$\lambda e^{-\alpha t} = c_1, \quad (12.89)$$

con c_1 costante arbitraria. Dalla seconda e dalla terza delle (12.88) si trae poi:

$$\varphi \exp \left\{ \frac{\lambda^2 \sigma^2}{4\alpha} \right\} = c_2, \quad (12.90)$$

con c_2 costante arbitraria. L'integrale generale della (12.85) è una funzione arbitraria delle costanti c_1 e c_2 oppure, equivalentemente, è esprimibile come

$$c_2 = H(c_1), \quad (12.91)$$

con H funzione arbitraria e c_1, c_2 date rispettivamente dalle (12.89) e (12.90). Si ha dunque:

$$\varphi(\lambda, t) \exp \left\{ \frac{\lambda^2 \sigma^2}{4\alpha} \right\} = H(\lambda e^{-\alpha t}),$$

ovvero

$$\varphi(\lambda, t) = \exp \left\{ -\frac{\lambda^2 \sigma^2}{4\alpha} \right\} H(\lambda e^{-\alpha t}). \quad (12.92)$$

Utilizzando la (12.61) segue poi:

$$H(\lambda) = \exp \left\{ \frac{\lambda^2 \sigma^2}{4\alpha} + i\lambda x_0 \right\}. \quad (12.93)$$

Facendo uso della (12.93) nella (12.92), si ha infine:

$$\varphi(\lambda, t|x_0) = \exp \left\{ -\frac{\lambda^2 \sigma^2}{4\alpha} (1 - e^{-2\alpha t}) + i\lambda e^{-\alpha t} x_0 \right\}. \quad (12.94) \quad \square$$

La (12.94) si riconosce essere la funzione caratteristica di una variabile casuale normale di media

$$M(t) = x_0 e^{-\alpha t} \quad (12.95)$$

e varianza

$$D^2(t) = \frac{\sigma^2}{2\alpha} (1 - e^{-2\alpha t}). \quad (12.96)$$

Si ha pertanto:

$$f(x, t|x_0) = \frac{1}{\sqrt{2\pi D(t)}} \exp \left\{ -\frac{[x - M(t)]^2}{2D^2(t)} \right\} \quad (x \in \mathbb{R}), \quad (12.97)$$

con $M(t)$ e $D(t)$ dati dalle (12.95) e (12.96). Risulta così dimostrata la seguente proposizione:

Proposizione 12.11. La densità di transizione del processo di Ornstein-Uhlenbeck di momenti infinitesimali (12.82) e (12.83) è univocamente determinata come soluzione dell'equazione (12.84) soddisfacente la condizione iniziale (12.58) ed è data dalla (12.97).

Per la omogeneità temporale del processo, si ha poi:

$$f(x, t|x_0, t_0) = \frac{1}{\sqrt{2\pi D(t-t_0)}} \exp \left\{ -\frac{|x - M(t-t_0)|^2}{2D^2(t-t_0)} \right\} \quad (x \in \mathbb{R}). \quad (12.98)$$

Si noti che essendo $\lim_{t \rightarrow +\infty} M(t) = 0$ e $\lim_{t \rightarrow +\infty} D^2(t) = \frac{\sigma^2}{2\alpha}$, in questo caso esiste la densità di probabilità stazionaria

$$S(x) = \frac{\sqrt{\alpha}}{\sqrt{\pi}\sigma} e^{-\alpha x^2/\sigma^2} \quad (12.99)$$

che è la densità di probabilità di una variabile casuale normale di media nulla e varianza $\sigma^2/(2\alpha)$. Alla stessa conclusione si perviene anche immediatamente utilizzando le (12.50) e (12.52).

Dalle considerazioni svolte si conclude che, diversamente da quanto accade per il processo di Wiener, al crescere del tempo il processo di Ornstein-Uhlenbeck raggiunge asintoticamente una situazione di equilibrio, detta anche "di regime", in cui la massa di probabilità è distribuita in modo gaussiano intorno all'origine, indipendentemente dallo stato iniziale x_0 . In tal senso è lecito affermare che lo stato zero costituisce il punto di equilibrio del processo di Ornstein-Uhlenbeck.

APPENDIX A

Principali Variabili Casuali

A.1 Variabili casuali discrete

Passeremo qui rapidamente in rassegna le principali variabili casuali discrete illustrandone il significato e fornendo le relative distribuzioni di probabilità, funzioni generatrici di probabilità, funzioni generatrici dei momenti e funzioni caratteristiche, nonché valori medi e varianze. Per semplicità di notazione denoteremo con p_X la distribuzione di probabilità della v.c. X , ossia porremo

$$p_X(x) = \mathbb{P}(X = x), \quad x \in S,$$

dove S indica l'insieme, finito o infinitamente numerabile, dei valori assunti da X con probabilità non nulla.

A.1.1 Distribuzione di Bernoulli

In un esperimento di prove ripetute, sia p la probabilità di successo (ossia la probabilità di occorrenza di un fissato evento E in ciascuna prova) e $q = 1-p$ la probabilità di insuccesso (la probabilità, quindi, dell'evento complementare \bar{E}). Il risultato della generica i -esima prova è dunque rappresentabile mediante una v.c. X che fornisce il numero di successi (1 oppure 0) in ciascuna prova. Ad essa si dà il nome di variabile di Bernoulli. La sua distribuzione di probabilità, detta *distribuzione di Bernoulli*, è dunque:

$$p_X(k) = p^k q^{1-k} \quad (k = 0, 1). \quad (\text{A.1})$$

è immediato calcolare media e varianza di X . Si trova così:

$$E(X) = p, \quad D^2(X) = p q. \quad (\text{A.2})$$

(i) *Funzione generatrice di probabilità*.

$$G_X(s) := E(s^X) = \sum_{k=0}^1 s^k p_X(k) = p s + q. \quad (\text{A.3})$$

(ii) *Funzione generatrice dei momenti.*

$$M_X(t) := E(e^{Xt}) = \sum_{k=0}^1 e^{kt} p_X(k) = p e^t + q.$$

(iii) *Funzione caratteristica.*

$$\varphi_X(u) := E(e^{iXu}) = \sum_{k=0}^1 e^{iku} p_X(k) = p e^{iu} + q. \quad (\text{A.4})$$

Si noti che la (A.1) dipende dal solo parametro p .

A.1.2 Distribuzione binomiale

Si consideri la somma Y_n di n variabili X_i di Bernoulli indipendenti:

$$Y_n = X_1 + X_2 + \dots + X_n. \quad (\text{A.5})$$

Essa rappresenta dunque il numero totale di successi nelle n ripetizioni dell'esperimento ed è chiamata variabile binomiale. Dalla (A.5) e dalla prima delle (A.2) segue:

$$E(Y_n) = \sum_{i=1}^n E(X_i) = np.$$

Per l'indipendenza delle X_i , dalla (A.5) e dalla seconda delle (A.2) si trae:

$$D^2(Y_n) = \sum_{i=1}^n D^2(X_i) = npq.$$

Sempre in virtù dell'indipendenza delle X_i e dalle (A.3)÷(A.4) seguono poi funzione generatrice di probabilità $G_{Y_n}(s)$, funzione generatrice dei momenti $M_{Y_n}(t)$ e funzione caratteristica $\varphi_{Y_n}(u)$. Ad esempio si ha:

$$G_{Y_n}(s) := E(s^{Y_n}) = E(s^{X_1+...+X_n}) = \prod_{i=1}^n E(s^{X_i}) = (ps + q)^n. \quad (\text{A.6})$$

Analogamente, si ottiene

$$M_{Y_n}(t) = (pe^t + q)^n \quad (\text{A.7})$$

$$\varphi_{Y_n}(u) = (pe^{iu} + q)^n. \quad (\text{A.8})$$

A.1 Variabili casuali discrete

La distribuzione di probabilità $p_{Y_n}(k)$ di Y_n , detta *distribuzione binomiale*, si ricava allora immediatamente. Ad esempio, dalla (A.6) segue:

$$G_{Y_n}(s) = (ps + q)^n$$

ovvero, sviluppando il secondo membro con la formula di Newton:

$$G_{Y_n}(s) = \sum_{k=0}^n \left[\binom{n}{k} p^k q^{n-k} \right] s^k.$$

Questa diviene

$$G_{Y_n}(s) = \sum_{k=0}^n p_{Y_n}(k) s^k$$

pur di porre

$$p_{Y_n}(k) = \binom{n}{k} p^k q^{n-k} \quad (k = 0, 1, \dots, n). \quad (\text{A.9})$$

La (A.9) è dunque la distribuzione di probabilità della variabile casuale binomiale. Al risultato (A.9) si sarebbe pervenuti utilizzando alternativamente la (A.7) o la (A.8). Si noti che la (A.9) dipende dai due parametri p ed n .

A.1.3 Distribuzione geometrica

Con riferimento ad un esperimento consistente nella ripetizione indefinita di prove in ciascuna delle quali un prefissato evento E si verifica con probabilità p e non si verifica con probabilità $q = 1 - p$, sia N la v.c. che denota il numero di prove dopo il quale E si verifica per la prima volta. Se le prove vengono effettuate ad istanti successivi e quiespaziati di una fissata scala di tempi — ad esempio agli istanti $\tau, 2\tau, 3\tau, \dots$, con $\tau \in \mathbb{R}^+$ fissata costante — il tempo di attesa per la prima occorrenza di E è $N\tau$. Scegliendo un'unità di misura del tempo uguale proprio a τ , la v.c. N si identifica con il suddetto tempo di attesa, misurato in unità di τ . La *distribuzione geometrica* è definita come la distribuzione di probabilità di N . Si ha pertanto:

$$p_N(k) = pq^{k-1} \quad (k = 1, 2, \dots). \quad (\text{A.10})$$

Invero, il secondo membro è la probabilità congiunta che l'evento in considerazione non si verifichi per $k-1$ prove consecutive e che si verifichi alla k -esima prova. Media e varianza di N si calcolano facilmente. Ad esempio, si ha:

$$E(N) = p \sum_{k=1}^{\infty} k q^{k-1} = p \frac{d}{dq} \sum_{k=1}^{\infty} q^k = p \frac{d}{dq} \left(\frac{1}{1-q} - 1 \right) = \frac{1}{p}. \quad (\text{A.11})$$

In maniera simile si può calcolare $E(N^2)$ e quindi $D^2(N)$. Il risultato è il seguente:

$$D^2(N) = \frac{q}{p^2}. \quad (\text{A.12})$$

La funzione generatrice di probabilità è data da:

$$G_N(s) = p \sum_{k=1}^{\infty} s^k q^{k-1} = \frac{ps}{1 - qs}, \quad 0 < s < \frac{1}{q}. \quad (\text{A.13})$$

La funzione generatrice dei momenti è allora:

$$M_N(t) = G_N(e^t) = \frac{pe^t}{1 - qe^t}, \quad t < \ln \frac{1}{q},$$

mentre per la funzione caratteristica si ha:

$$\varphi_N(u) = \frac{pe^{iu}}{1 - qe^{iu}}. \quad (\text{A.14})$$

Alle (A.11) e (A.12) si può evidentemente pervenire anche mediante uso delle (A.13)÷(A.14).

Concludiamo sottolineando che la distribuzione geometrica (A.10) dipende dal solo parametro p .

A.1.4 Distribuzione binomiale negativa

Describe il numero I di insuccessi che precedono il k -esimo successo in una successione di prove ripetute di Bernoulli in cui la probabilità di successo è p . Si noti che il numero di prove effettuate fino ad ottenere il k -esimo successo è dato dalla somma $i + k$ del numero k di successi e del numero i di insuccessi registrati fino ad ottenere il k -esimo successo. Si ha pertanto:

$$p_I(i) = \binom{k+i-1}{i} p^k q^i = \binom{k+i-1}{k-1} p^k q^i \quad (i = 0, 1, 2, \dots). \quad (\text{A.15})$$

Invero, la probabilità di avere i insuccessi prima dell' k -esimo successo è il prodotto delle probabilità p di avere un successo alla $(k+i)$ -esima prova per la probabilità di avere $k-1$ successi ed i insuccessi nelle precedenti $k+i-1$ prove. Il coefficiente binomiale dà il numero di modi distinti in cui i insuccessi si possono avere in $k+i-1$ prove.

Si noti che la distribuzione (A.15) contiene due parametri, $k \geq 1$ e p . Notiamo anche che la distribuzione binomiale negativa può equivalentemente interpretarsi come descrivente

il numero N di prove richieste nell'esperimento di prove ripetute fino al successo k -esimo. Invero, poiché N ed I sono legate dalla relazione $N = I + k$, dalla (A.15) si trae:

$$p_I(i) = p_N(i+k) = \binom{k+i-1}{i} p^k q^i \quad (i = 0, 1, 2, \dots),$$

da cui segue:

$$p_N(n) = \binom{n-1}{n-k} p^k q^{n-k} \quad (n = k, k+1, k+2, \dots),$$

ovvero:

$$p_N(n) = \binom{n-1}{k-1} p^k q^{n-k} \quad (n = k, k+1, k+2, \dots).$$

La funzione generatrice di probabilità di N si ottiene poi subito osservando che la v.c. N può esprimersi come somma di k variabili indipendenti e identicamente distribuite:

$$N = X_1 + X_2 + \cdots + X_k, \quad (\text{A.16})$$

dove X_i ($i = 1, 2, \dots, k$) denota il numero di prove intercorrenti tra l' $(i-1)$ -esimo e l' i -esimo successo. Ricordando la (A.13) segue pertanto:

$$g_N(s) = \prod_{i=1}^n g_{X_i}(s) = \left(\frac{ps}{1 - qs} \right)^k \quad 0 < s < \frac{1}{q}.$$

È poi immediato rendersi conto che la funzione generatrice dei momenti e la funzione caratteristica sono rispettivamente

$$M_N(t) = \left(\frac{pe^t}{1 - qe^t} \right)^k, \quad t < \ln \frac{1}{q}$$

e

$$\varphi_N(u) = \left(\frac{pe^{iu}}{1 - qe^{iu}} \right)^k.$$

Dalle (A.16) e dalla (A.11) segue poi:

$$E(N) = \sum_{i=1}^k E(X_i) = \frac{k}{p}.$$

Per l'indipendenza delle X_i , dalle (A.12) e (A.16) si trae infine:

$$D^2(N) = \sum_{i=1}^k D^2(X_i) = \frac{kq}{p^2}.$$

Concludiamo osservando che la (A.15) è suscettibile di una diversa scrittura. Posto inverò

$$\binom{-\alpha}{r} = \frac{-\alpha(-\alpha-1)(-\alpha-2)\cdots(-\alpha-r+1)}{r!},$$

per ogni α reale positivo arbitrario e per ogni r intero non negativo, si ha:

$$\begin{aligned}\binom{k+i-1}{i} &= \frac{(k+i-1)(k+i-2)(k+i-3)\cdots(k+1)k}{i!} \\ &= (-1)^i \binom{-k}{i},\end{aligned}$$

La (A.15) pertanto diviene:

$$p_I(i) = \binom{-k}{i} p^k (-q)^i \quad (i = 0, 1, 2, \dots).$$

Di qui la denominazione di distribuzione binomiale negativa. Essa viene talora denotata quale *distribuzione di Pascal*.

A.1.5 Distribuzione di Poisson

Si ottiene dalla distribuzione binomiale (A.9) al limite per $p \rightarrow 0$ e $n \rightarrow \infty$ con la condizione che il prodotto $p^n = \lambda$ sia costante. La condizione $p \rightarrow 0$ giustifica la denominazione di "eventi rari" attribuita agli eventi con distribuzione di Poisson. Effettuando il suddetto limite nella (A.9) si ottiene:

$$\begin{aligned}\lim_{\substack{n \rightarrow \infty \\ np = \lambda}} p_{Y_n}(k) &= \lim_{n \rightarrow \infty} \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \lim_{n \rightarrow \infty} \frac{\frac{n}{n} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)}{\left(1 - \frac{\lambda}{n}\right)^k} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \\ &= \frac{\lambda^k}{k!} e^{-\lambda},\end{aligned}$$

dove la condizione $k = 0, 1, 2, \dots, n$ nella (A.9) si traduce in $k \in \mathbb{N}$ per la distribuzione limite. In conclusione, una v.c. X di Poisson di parametro λ ha distribuzione di probabilità

$$p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (k = 0, 1, 2, \dots). \quad (\text{A.17})$$

A questa corrisponde la funzione generatrice di probabilità

$$G_X(s) = e^{\lambda(s-1)}, \quad s \in \mathbb{R}. \quad (\text{A.18})$$

A.1 Variabili casuali discrete

La funzione generatrice dei momenti e la funzione caratteristica sono rispettivamente

$$M_X(t) = e^{\lambda(e^t-1)}, \quad t \in \mathbb{R} \quad (\text{A.19})$$

e

$$\varphi_X(u) = e^{\lambda(e^{iu}-1)}, \quad u \in \mathbb{R}. \quad (\text{A.20})$$

Direttamente dalla (A.17), oppure tramite le (A.18), (A.19) o (A.20) si può facilmente mostrare che si ha:

$$E(X) = \lambda, \quad D^2(X) = \lambda.$$

Spesso la distribuzione di Poisson viene utilizzata come approssimazione della distribuzione binomiale allorché n è "piuttosto grande" e p "piuttosto piccolo". Un'accettabile regola del pollice suggerisce di fare uso di tale approssimazione quando risulta $n \geq 20$ e $p \leq 0.05$. Quando si ha $n \geq 100$, l'approssimazione è solitamente eccellente purché risulti $np \leq 10$.

A.1.6 Distribuzione ipergeometrica

Questa distribuzione nasce in problemi di estrazione senza rimpiazzamento. A titolo esemplificativo, riferiamoci ad un'urna contenente b biglie bianche ed n biglie nere, dalla quale si estraggono k biglie ($0 < k \leq b+n$), l'una dopo l'altra senza rimpiazzamento (nulla cambierebbe se le k biglie venissero estratte tutte simultaneamente o per gruppi senza rimpiazzamento). Lo spazio campione può identificarsi con l'insieme di tutte le k -uple ordinate di lettere scelte tra B (indicante biglia bianca) e N (indicante biglia nera) con la condizione che il numero di lettere B non superi b e il numero di lettere N non superi n . Denotiamo con X la v.c. rappresentante il numero di biglie bianche presenti tra le k biglie estratte (o, come si suol dire, nel campione di taglia k ottenuto). Evidentemente X assume valori interi non negativi; se, inoltre, è $k > n$, il minimo valore di X è $k-n$ poiché nel campione di taglia k devono necessariamente esserci almeno $k-n$ biglie bianche. Quindi, il minimo valore di X è $\max\{0, k-n\}$. Il massimo valore di X non può poi superare k e se il numero di biglie bianche è minore di k (ossia se è $b < k$) allora il massimo valore di X è proprio b . Quindi, il massimo valore di X è $\min\{k, b\}$. Calcoliamo ora la distribuzione di probabilità di X . A tal fine, per ogni intero s tale che $\max\{0, k-n\} \leq s \leq \min\{k, b\}$, calcoliamo $P(X=s)$ osservando che si possono scegliere s biglie bianche tra b biglie bianche in $\binom{b}{s}$ modi e che per ciascuno di tali modi vi sono $\binom{n}{k-s}$ modi di scegliere $k-s$ biglie nere tra le n biglie nere. Infine, vi sono $\binom{b+n}{k}$ modi di scegliere b biglie tra le $b+n$ biglie presenti nell'urna. In conclusione, la distribuzione di probabilità cercata è data dal rapporto tra il numero di casi

favorevoli alla presenza di s biglie bianche nel campione e il numero totale di casi:

$$\mathbb{P}(X = s) = \frac{\binom{b}{s} \binom{n}{k-s}}{\binom{b+n}{k}} \quad \max\{0, k-s\} \leq s \leq \min\{k, b\}.$$

Ad essa si dà il nome di distribuzione ipergeometrica. Si noti che, equivalentemente, la distribuzione ipergeometrica può definirsi al seguente modo:

$$\mathbb{P}(X = s) = \frac{\binom{Np}{s} \binom{Nq}{k-s}}{\binom{N}{k}} \quad s = 0, 1, \dots, k$$

con $N = 1, 2, \dots, k = 1, 2, \dots, N, p = 0, 1/N, 2/N, \dots, 1$ e dove si è posto $q = 1 - p$. È possibile mostrare che risulta

$$E(X) = kp, \quad D^2(X) = kpq \frac{N-k}{N-1}$$

ovvero, con la precedente notazione,

$$E(X) = \frac{kb}{b+n}, \quad D^2(X) = \frac{kbn(b+n-k)}{(b+n-1)(b+n)^2}.$$

Il calcolo delle funzioni generatrici e caratteristica, piuttosto laborioso, viene qui per brevità omesso.

A.2 Variabili casuali assolutamente continue

Passiamo ora ad una rassegna delle principali vv.cc. assolutamente continue, richiamando le definizioni delle corrispondenti densità di probabilità nonché le proprietà essenziali.

A.2.1 Densità uniforme

La v.c. X si dice uniformemente distribuita in (a, b) se ha densità di probabilità

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{per } a < x < b, \\ 0, & \text{altrove.} \end{cases}$$

La sua funzione generatrice dei momenti è allora:

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{+\infty} e^{tx} f_X(x) dx = \frac{1}{b-a} \int_a^b e^{tx} dx \\ &= \frac{e^{tb} - e^{ta}}{t(b-a)}. \end{aligned} \quad (\text{A.21})$$

Sviluppando gli esponenziali in serie di Mac Laurin, la (A.21) diviene:

$$M_X(t) = \frac{1}{t(b-a)} \sum_{k=1}^{\infty} \frac{t^k}{k!} (b^k - a^k) = \sum_{r=0}^{\infty} \frac{b^{r+1} - a^{r+1}}{(r+1)(b-a)} \frac{t^r}{r!}. \quad (\text{A.22})$$

Poiché X possiede momenti finiti di qualsiasi ordine, dallo sviluppo (A.22) della funzione generatrice dei momenti si trae:

$$E(X^r) = \frac{b^{r+1} - a^{r+1}}{(r+1)(b-a)} \quad (r = 1, 2, \dots).$$

In particolare, si ha:

$$E(X) = \frac{a+b}{2}, \quad E(X^2) = \frac{a^2 + ab + b^2}{3}, \quad (\text{A.23})$$

da cui segue:

$$D^2(X) = \frac{(b-a)^2}{12}. \quad (\text{A.24})$$

Ponendo $t = iu$ nella (A.21), si ottiene la funzione caratteristica:

$$\varphi_X(u) = \frac{e^{ibu} - e^{iau}}{iu(b-a)}.$$

Si noti che se X è uniforme nell'intervallo $(-\frac{1}{2}, \frac{1}{2})$, le (A.23) e (A.24) forniscono:

$$E(X) = 0, \quad D^2(X) = \frac{1}{12}.$$

Di qui segue che la somma di 12 vv.cc. indipendenti e uniformemente distribuite in $(-\frac{1}{2}, \frac{1}{2})$ ha media nulla e varianza unitaria. Su questa constatazione, e in virtù del Teorema centrale del limite, è basato un semplice metodo per la generazione di numeri pseudocasuali a distribuzione approssimativamente normale standard.

A.2.2 Densità esponenziale

Si dice che X è esponenzialmente distribuita con parametro λ se ha densità di probabilità

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{per } x > 0, \\ 0, & \text{per } x < 0. \end{cases} \quad (\text{A.25})$$

La funzione generatrice dei momenti $M_X(t)$ esiste per $t < \lambda$ ed è di calcolo immediato:

$$M_X(t) = \int_{-\infty}^{+\infty} e^{tx} f_X(x) dx = \lambda \int_0^{+\infty} e^{(t-\lambda)x} dx = \frac{\lambda}{\lambda - t} \quad (t < \lambda). \quad (\text{A.26})$$

La (A.26) è sviluppabile in serie di potenze di t . Invero, riscrivendo la (A.26) nella forma

$$M_X(t) = \frac{1}{1 - t/\lambda} \quad (t < \lambda),$$

sì riconosce che il secondo membro è la somma della serie geometrica di ragione t/λ . Pertanto si ha:

$$M_X(t) = \sum_{k=0}^{\infty} \left(\frac{t}{\lambda}\right)^k = \sum_{k=0}^{\infty} \frac{k!}{\lambda^k} \frac{t^k}{k!} \quad (t < \lambda). \quad (\text{A.27})$$

I momenti di X intorno all'origine, tutti esistenti, sono i coefficienti di $t^k/k!$ nello sviluppo (A.27):

$$E(X^k) = \frac{k!}{\lambda^k} \quad (k = 1, 2, \dots).$$

Si ha quindi in particolare:

$$E(X) = \frac{1}{\lambda}$$

$$D^2(X) = E(X^2) - [E(X)]^2 = \frac{1}{\lambda^2}.$$

è poi facile ottenere il coefficiente di asimmetria α di X :

$$\alpha \equiv \frac{E[X - E(X)]^3}{[D(X)]^3} = \frac{2/\lambda^3}{\lambda^3} = 2.$$

Infine, dalla (A.26) si ricava la funzione caratteristica:

$$\varphi_X(u) = \frac{\lambda}{\lambda - iu}.$$

Si noti che, a differenza di $M_X(t)$ che è definita solo per $t < \lambda$, la funzione caratteristica $\varphi_X(u)$ esiste per tutti i valori del suo argomento.

A.2.3 Densità normale

Si dice che X è normale di media μ e varianza σ^2 , e si scrive $X \sim N(\mu, \sigma^2)$, se essa ha densità di probabilità

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad -\infty < x < +\infty.$$

La funzione generatrice dei momenti è dunque:

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{+\infty} e^{tx} f_X(x) dx = \int_{-\infty}^{+\infty} e^{tx} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \\ &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} \exp\left[-y^2 + (\sigma\sqrt{2}y + \mu)t\right] dy \\ &= \frac{1}{\sqrt{\pi}} \exp\left(\frac{\sigma^2}{2}t^2 + \mu t\right) \int_{-\infty}^{+\infty} \exp\left[-\left(y - \frac{\sigma t}{2}\right)^2\right] dy \\ &= \frac{1}{\sqrt{\pi}} \exp\left(\frac{\sigma^2}{2}t^2 + \mu t\right) \int_{-\infty}^{+\infty} e^{-z^2} dz \\ &= \exp\left(\frac{\sigma^2}{2}t^2 + \mu t\right), \quad -\infty < t < +\infty. \end{aligned} \quad (\text{A.28})$$

I momenti intorno all'origine di X , tutti esistenti, si ottengono dalla (A.28):

$$E(X^k) = \frac{d^k}{dt^k} M_X(t)\Big|_{t=0} \quad (k = 1, 2, \dots). \quad (\text{A.29})$$

In particolare, dalla (A.29) si ricava:

$$E(X) = \frac{d}{dt} M_X(t)\Big|_{t=0} = [(\mu + \sigma^2 t) M_X(t)]_{t=0} = \mu.$$

Si ricava inoltre:

$$\begin{aligned} E(X^2) &= \frac{d^2}{dt^2} M_X(t)\Big|_{t=0} \\ &= [\sigma^2 M_X(t) + (\mu + \sigma^2 t) M_X(t)]_{t=0} \\ &= \mu^2 + \sigma^2, \end{aligned}$$

da cui segue:

$$D^2(X) = \sigma^2.$$

Nel caso di v.c. normale standard $Y \sim N(0, 1)$, la (A.28) diviene:

$$M_Y(t) = e^{t^2/2}.$$

Non è difficile mostrare ricorsivamente che in tal caso si ha:

$$E(X^n) = \begin{cases} (n-1)(n-3)(n-5)\cdots 3 \cdot 1, & \text{per } n \text{ pari,} \\ 0, & \text{per } n \text{ dispari.} \end{cases}$$

La funzione caratteristica $\varphi_X(u)$ si ottiene dalla (A.28) ponendovi $t = iu$:

$$\varphi_X(u) = \exp \left\{ -\frac{\sigma^2}{2} u^2 + i\mu u \right\}.$$

A.2.4 Densità gamma

È noto che l'integrale

$$\Gamma(p) = \int_0^{+\infty} x^{p-1} e^{-x} dx \quad (\text{A.30})$$

converge per ogni $p \in \mathbb{R}^+$. Esso è dunque una funzione definita per $p \in (0, +\infty)$ alla quale si dà il nome di funzione gamma. Dalla (A.30) per ogni $p > 0$ si ottiene poi:

$$\Gamma(p+1) = p\Gamma(p). \quad (\text{A.31})$$

Per $p = 1$ risulta:

$$\Gamma(1) = \int_0^{+\infty} e^{-x} dx = 1,$$

e quindi per ogni n intero positivo si ha:

$$\Gamma(n+1) = n!$$

Notevole è anche il seguente risultato:

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}. \quad (\text{A.32})$$

Dalla (A.30) si ha infatti:

$$\Gamma\left(\frac{1}{2}\right) = \int_0^{+\infty} x^{-1/2} e^{-x} dx = 2 \int_0^{+\infty} e^{-y^2} dy, \quad (\text{A.33})$$

A.2 Variabili casuali assolutamente continue

dove il secondo integrale segue dal primo con il cambiamento di variabile $y = \sqrt{x}$. Poiché $\Gamma(1/2)$ è una quantità positiva, essa può ottersi come la radice quadrata positiva di $[\Gamma(1/2)]^2$, dove per la (A.33) risulta:

$$\left[\Gamma\left(\frac{1}{2}\right) \right]^2 = 4 \int_0^{+\infty} e^{-x^2} dx \int_0^{+\infty} e^{-y^2} dy = 4 \int_0^{+\infty} \int_0^{+\infty} e^{-(x^2+y^2)} du dv.$$

Passiamo alle coordinate polari

$$x = \rho \cos \theta \quad y = \rho \sin \theta \quad (\text{A.34})$$

ed osserviamo che lo jacobiano della trasformazione è ρ . Poiché

$$\{(x, y): 0 < x < +\infty, 0 < y < +\infty\} = \{(\rho, \theta): 0 < \rho < +\infty, 0 < \theta < \pi/2\},$$

otteniamo:

$$\left[\Gamma\left(\frac{1}{2}\right) \right]^2 = 4 \int_0^{\pi/2} d\theta \int_0^{+\infty} e^{-\rho^2} \rho d\rho = \pi,$$

da cui segue la (A.32).

Definiamo ora la densità di probabilità gamma al seguente modo:

$$f_X(x) = \begin{cases} k x^\alpha e^{-x/\beta}, & \text{per } x > 0, \\ 0, & \text{per } x \leq 0, \end{cases} \quad (\text{A.35})$$

con $\alpha > -1$, $\beta > 0$ e k costante da determinarsi. Quest'ultima si ottiene imponendo che $f_X(x)$ soddisfi la condizione di normalizzazione

$$\int_{-\infty}^{+\infty} f_X(x) dx = 1. \quad (\text{A.36})$$

Dalle (A.35) e (A.36) segue allora:

$$k = \frac{1}{\int_0^{+\infty} x^\alpha e^{-x/\beta} dx}. \quad (\text{A.37})$$

Osserviamo ora che risulta:

$$\int_0^{+\infty} x^\alpha e^{-x/\beta} dx = \beta^{\alpha+1} \Gamma(\alpha+1), \quad (\text{A.38})$$

come si dimostra facendo uso del cambiamento di variabile $y = x/\beta$ e della definizione (A.30). Utilizzando le (A.37) e (A.38) si ha dunque:

$$k = \frac{1}{\beta^{\alpha+1} \Gamma(\alpha+1)},$$

così che la (A.35) diviene:

$$f_X(x) = \begin{cases} \frac{1}{\beta^{\alpha+1} \Gamma(\alpha+1)} x^\alpha e^{-x/\beta}, & \text{per } x > 0, \\ 0, & \text{per } x < 0. \end{cases} \quad (\text{A.39})$$

Se una v.c. X ha densità di probabilità (A.39) con $\alpha > -1$ e $\beta > 0$, essa si dice avere distribuzione gamma.

Calcoliamo la funzione generatrice dei momenti, esistente per $t < 1/\beta$. Si ha:

$$\begin{aligned} M_X(t) &= \frac{1}{\beta^{\alpha+1} \Gamma(\alpha+1)} \int_0^{+\infty} e^{tx} x^\alpha e^{-x/\beta} dx \\ &= \frac{(1/\beta - t)^{-\alpha-1}}{\beta^{\alpha+1} \Gamma(\alpha+1)} \int_0^{+\infty} y^\alpha e^{-y} dy = \left(\frac{1/\beta}{1/\beta - t} \right)^{\alpha+1} \\ &= (1 - \beta t)^{-\alpha-1} \quad (t < \frac{1}{\beta}). \end{aligned} \quad (\text{A.40})$$

Si noti che per $\alpha = 0$ e $\beta = 1/\lambda$ la (A.40) coincide con la funzione generatrice dei momenti (A.26) della v.c. esponenziale di parametro λ . Invero, con tale scelta la (A.39) si identifica con la (A.25). La distribuzione esponenziale è quindi un caso particolare della distribuzione gamma.

Dalla (A.40), per derivazione rispetto a t segue poi:

$$\begin{aligned} \frac{d}{dt} M_X(t) \Big|_{t=0} &= \beta(\alpha+1), \\ \frac{d^2}{dt^2} M_X(t) \Big|_{t=0} &= \beta^2(\alpha+1)(\alpha+2) \end{aligned}$$

da cui si ricava:

$$D^2(X) = \beta^2(\alpha+1).$$

Esempio A.1. Un esempio che conduce ad una v.c. con distribuzione gamma si ottiene ponendo

$$X = \sum_{i=1}^n X_i$$

A.2 Variabili casuali assolutamente continue

con X_1, X_2, \dots, X_n v.v.c. indipendenti distribuite esponenzialmente con parametro λ . Si ha:

$$M_X(t) = [M_{X_i}(t)]^n$$

ovvero, per la (A.26):

$$M_X(t) = \left(1 - \frac{t}{\lambda} \right)^{-n}. \quad (\text{A.41})$$

Paragonando la (A.41) con la (A.40), si conclude che X ha distribuzione gamma di parametri $\alpha = n-1$ e $\beta = 1/\lambda$. \square

Un'immmediata applicazione di quanto visto nell'Esempio A.1 nasce allorché si considera un processo di Poisson di Parametro λ . Infatti, sia $t_0 = 0$ l'istante iniziale, T_1 l'istante di occorrenza del primo evento, T_2 l'intervallo di tempo intercorrente tra il primo ed il secondo evento, T_3 l'intervallo di tempo intercorrente tra il secondo ed il terzo evento, e così via. Quindi $T_1 + T_2 + \dots + T_n$ è l'intervallo di tempo intercorrente tra l'istante iniziale e l'occorrenza dell' n -esimo evento, mentre T_k ($k = 1, 2, \dots, n$) denota il cosiddetto *tempo di attesa* del k -esimo evento. Indicato dunque genericamente con T il tempo di attesa dell' n -esimo evento, si ha:

$$T = T_1 + T_2 + \dots + T_n,$$

con le T_i ($i = 1, 2, \dots, n$) indipendenti ed esponenzialmente distribuite con parametro λ . Per quanto mostrato nell'Esempio A.1, concludiamo che T ha distribuzione gamma di parametri $\alpha = n-1$ e $\beta = 1/\lambda$, e che quindi ha densità di probabilità

$$f_T(x) = \begin{cases} \frac{\lambda^n}{(n-1)!} e^{-\lambda x} x^{n-1}, & \text{per } x > 0, \\ 0, & \text{per } x \leq 0. \end{cases}$$

A.2.5 Densità beta

La funzione beta è definita al seguente modo:

$$B(p, q) = \int_0^1 x^{p-1} (1-x)^{q-1} dx, \quad (\text{A.42})$$

con p e q reali positivi. Essa è intimamente legata alla funzione gamma, come mostra la seguente proposizione:

Proposizione A.1. Per $p > 0$ e $q > 0$ si ha:

$$B(p, q) = \frac{\Gamma(p) \Gamma(q)}{\Gamma(p+q)} = B(q, p).$$

Dimostrazione. Per $x = y^2$ la funzione gamma data dalla (A.30) diventa:

$$\Gamma(p) = 2 \int_0^{+\infty} y^{2p-1} e^{-y^2} dy.$$

Si può dunque scrivere:

$$\begin{aligned}\Gamma(p)\Gamma(q) &= 4 \int_0^{+\infty} x^{2p-1} e^{-x^2} dx \int_0^{+\infty} y^{2q-1} e^{-y^2} dy \\ &= 4 \int_0^{+\infty} \int_0^{+\infty} x^{2p-1} y^{2q-1} e^{-(x^2+y^2)} dx dy\end{aligned}$$

ovvero, passando a coordinate polari tramite la trasformazione (A.34):

$$\begin{aligned}\Gamma(p)\Gamma(q) &= 4 \int_0^{+\infty} d\rho \int_0^{\pi/2} (\rho \cos \theta)^{2p-1} (\rho \sin \theta)^{2q-1} e^{-\rho^2} \rho d\theta \\ &= 4 \int_0^{+\infty} \rho^{2p+2q-1} e^{-\rho^2} d\rho \int_0^{\pi/2} (\cos \theta)^{2p-1} (\sin \theta)^{2q-1} d\theta.\end{aligned}$$

Effettuando nel primo integrale a secondo membro il cambiamento di variabile $\rho^2 = x$, nel secondo integrale il cambiamento di variabile $\sin^2 \theta = y$ e ricordando la definizione (A.42), si ottiene:

$$\begin{aligned}\Gamma(p)\Gamma(q) &= \int_0^{+\infty} x^{p+q-1} e^{-x} dx \int_0^1 y^{q-1} (1-y)^{p-1} dy \\ &= \Gamma(p+q) B(q,p),\end{aligned}$$

ossia:

$$B(q,p) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}.$$

Scambiando i ruoli di p e q segue infine l'asserto. \square

Ciò premesso, una v.c. X si dice avere distribuzione beta se la sua densità di probabilità è così definita:

$$f_X(x) = \begin{cases} k x^{p-1} (1-x)^{q-1}, & \text{per } 0 < x < 1, \\ 0, & \text{altrove,} \end{cases} \quad (\text{A.43})$$

con $p > 0, q > 0$ e k costante tale da rendere f_X normalizzata. Imponendo che risulti:

$$\int_{-\infty}^{+\infty} f_X(x) dx = 1,$$

si ottiene:

$$k = \frac{1}{\int_0^1 x^{p-1} (1-x)^{q-1} dx} = \frac{1}{B(p,q)}.$$

In conclusione, se p e q sono reali positivi, la densità di probabilità beta di parametri p e q è la seguente:

$$f_X(x) = \begin{cases} \frac{1}{B(p,q)} x^{p-1} (1-x)^{q-1}, & \text{per } 0 < x < 1, \\ 0, & \text{altrove.} \end{cases} \quad (\text{A.44})$$

Calcoliamo la funzione generatrice dei momenti:¹

$$\begin{aligned}M_X(t) &= \frac{1}{B(p,q)} \int_0^1 e^{tx} x^{p-1} (1-x)^{q-1} dx \\ &= \frac{1}{B(p,q)} \sum_{k=0}^{+\infty} \frac{t^k}{k!} \int_0^1 x^{p+k-1} (1-x)^{q-1} dx \\ &= \sum_{k=0}^{\infty} \frac{B(p+k,q)}{B(p,q)} \frac{t^k}{k!}.\end{aligned} \quad (\text{A.45})$$

Di qui segue:

$$E(X^k) = \frac{B(p+k,q)}{B(p,q)} \quad (k = 1, 2, \dots). \quad (\text{A.46})$$

In particolare, dalla (A.46) si ricava:

$$\begin{aligned}E(X) &= \frac{B(p+1,q)}{B(p,q)} = \frac{\Gamma(p+1)\Gamma(q)/\Gamma(p+q+1)}{\Gamma(p)\Gamma(q)/\Gamma(p+q)} = \frac{p}{p+q} \\ E(X^2) &= \frac{B(p+2,q)}{B(p,q)} = \frac{\Gamma(p+2)\Gamma(q)/\Gamma(p+q+2)}{\Gamma(p)\Gamma(q)/\Gamma(p+q)} = \frac{p(p+1)}{(p+q)(p+q+1)},\end{aligned}$$

da cui segue:

$$D^2(X) = \frac{pq}{(p+q)^2(p+q+1)}.$$

Si noti che dalla (A.44) si trae:

$$\lim_{x \rightarrow 0} f_X(x) = \begin{cases} 0, & \text{per } p > 1, \\ q, & \text{per } p = 1, \\ +\infty, & \text{per } p < 1. \end{cases}$$

¹Non è difficile mostrare, usando ad esempio il criterio del rapporto, che la serie (A.45) ha raggio di convergenza infinito.

Ciò indica che la densità beta, come del resto anche la densità gamma, può risultare particolarmente utile in problemi di modellistica grazie alla sua flessibilità. Si noti anche che per $p = q = 1$ la densità beta si riduce alla densità uniforme in $(0, 1)$.

Esempio A.2. Un esempio di densità beta nasce allorché si considerano le statistiche d'ordine. Siano X_1, X_2, \dots, X_n v.v.c. assolutamente continue, indipendenti e identicamente distribuite con funzione di distribuzione $F_{X_i} = F$ e densità di probabilità $f_{X_i} = f$. Siano poi Y_1, Y_2, \dots, Y_n le corrispondenti statistiche d'ordine; per $(k = 1, 2, \dots, n)$ la v.c. Y_k denota dunque il k -esimo più piccolo valore di X_1, X_2, \dots, X_n . Osserviamo anzitutto che per $k = 1$ si ha:

$$\begin{aligned} F_{Y_1}(y) &= \mathbb{P}(Y_1 \leq y) = \mathbb{P}(\min\{X_1, X_2, \dots, X_n\} \leq y) \\ &= 1 - \mathbb{P}(\min\{X_1, X_2, \dots, X_n\} > y) \\ &= 1 - \mathbb{P}(X_1 > y; X_2 > y; \dots; X_n > y) \\ &= 1 - [1 - F(y)]^n. \end{aligned}$$

Nel caso $k = n$ si ha poi:

$$\begin{aligned} F_{Y_n}(y) &= \mathbb{P}(Y_n \leq y) = \mathbb{P}(\max\{X_1, X_2, \dots, X_n\} \leq y) \\ &= \mathbb{P}(X_1 \leq y; X_2 \leq y; \dots; X_n \leq y) \\ &= [F(y)]^n. \end{aligned}$$

Per un generico k , $\{Y_k \leq y\}$ coincide con l'evento si verificano almeno k degli n eventi $\{X_1 \leq y\}, \{X_2 \leq y\}, \dots, \{X_n \leq y\}$. Questi sono indipendenti e si verificano ciascuno con probabilità $F(y)$. Per $k = 1, 2, \dots, n$ si ha dunque:

$$F_{Y_k}(y) := \mathbb{P}(Y_k \leq y) = \sum_{r=k}^n \binom{n}{r} [F(y)]^r [1 - F(y)]^{n-r}. \quad (\text{A.47})$$

A.2 Variabili casuali assolutamente continue

Derivando ambo i membri della (A.47) otteniamo:

$$\begin{aligned} f_{Y_k}(y) &= \sum_{r=k}^n \binom{n}{r} r f(y) [F(y)]^{r-1} [1 - F(y)]^{n-r} + \\ &\quad - \sum_{r=k}^{n-1} \binom{n}{r} [F(y)]^r (n-r) f(y) [1 - F(y)]^{n-r-1} \\ &= \sum_{r=k}^n \binom{n}{r} r f(y) [F(y)]^{r-1} [1 - F(y)]^{n-r} + \\ &\quad - \sum_{s=k+1}^n \binom{n}{s-1} [F(y)]^{s-1} (n-s+1) f(y) [1 - F(y)]^{n-s} \\ &= \binom{n}{k} k f(y) [F(y)]^{k-1} [1 - F(y)]^{n-k} + \\ &\quad + \sum_{r=k+1}^n \left[r \binom{n}{r} - (n-r+1) \binom{n}{r-1} \right] f(y) [F(y)]^{r-1} [1 - F(y)]^{n-r} \\ &= \binom{n}{k} k f(y) [F(y)]^{k-1} [1 - F(y)]^{n-k} \quad (k = 1, 2, \dots, n), \end{aligned}$$

poiché l'espressione in parentesi quadra nell'ultima somma è nulla. Abbiamo così ottenuto:

$$f_{Y_k}(y) = \binom{n}{k} k f(y) [F(y)]^{k-1} [1 - F(y)]^{n-k} \quad (k = 1, 2, \dots, n). \quad (\text{A.48})$$

Supponiamo ora, in particolare, che le X_i ($i = 1, 2, \dots, n$) siano indipendenti e uniformemente distribuite in $(0, 1)$. Dalla (A.48) si ricava allora:

$$f_{Y_k}(y) = \begin{cases} \binom{n}{k} k y^{k-1} (1-y)^{n-k}, & \text{per } 0 < y < 1, \\ 0, & \text{altrove.} \end{cases} \quad (\text{A.49})$$

Pertanto la k -esima statistica d'ordine Y_k ha in questo caso distribuzione beta di parametri $p = k$ e $q = n - p + 1$; inverso, la (A.49) con questa scelta di p e q diviene:

$$f_{Y_k}(y) = \begin{cases} \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1} (1-y)^{q-1}, & \text{per } 0 < y < 1, \\ 0, & \text{altrove,} \end{cases}$$

avendo fatto uso dell'osservazione che per $p = k$ e $q = n - p + 1$ si ha:

$$\binom{n}{k} k = \frac{n!}{(k-1)! (n-k)!} = \frac{\Gamma(n+1)}{\Gamma(k) \Gamma(n-k+1)} = \frac{\Gamma(p+q)}{\Gamma(p) \Gamma(q)}.$$

△

APPENDIX B

Complementi

B.1 La funzione delta di Dirac

Riassumiamo le proprietà principali della funzione delta di Dirac cui si è spesso fatto riferimento.

Definizione B.1. Si dice funzione delta di Dirac, o funzione δ , la funzione definita tramite la posizione

$$\delta(x) = \begin{cases} 0, & \text{se } x \neq 0 \\ +\infty, & \text{se } x = 0, \end{cases} \quad (\text{B.1})$$

con la condizione

$$\int_{-\infty}^{+\infty} \delta(x) dx = 1. \quad (\text{B.2})$$

△

Si osservi che, qualunque sia la definizione di integrale usata, la Definizione B.1 non ha senso da un punto di vista usuale in quanto l'integrale di una funzione nulla ovunque tranne che in un punto è necessariamente zero. D'altra parte una giustificazione rigorosa di tale definizione è resa necessaria dalle sue molteplici applicazioni in numerosi settori disciplinari. Ciò si può ottenere ricorrendo alla cosiddetta teoria delle distribuzioni introdotta da L. Schwartz. Tale teoria è molto potente in quanto in essa è lecito scambiare operazioni di limiti, impossibili per funzioni ordinarie, usare serie che in un senso usuale sarebbero divergenti, e così via. Nell'ambito della teoria delle distribuzioni si vede che è in particolare perfettamente giustificata la definizione non solo della funzione δ ma anche di tutte le sue derivate. Una trattazione della teoria delle distribuzioni esula tuttavia dai nostri scopi; ci limitiamo perciò a menzionare che la funzione δ può essere usata come una funzione ordinaria e che le usuali tecniche di integrazione possono essere applicate agli integrali che la coinvolgono.

Passiamo quindi ad elencare alcune utili proprietà della funzione δ espresse dalla proposizione che segue della quale, per brevità, omettiamo la dimostrazione.

Proposizione B.1. Per ogni $x \in \mathbb{R}$ e per ogni funzione continua u risulta:

$$u(x) = \int_{-\infty}^{+\infty} \delta(t-x) u(t) dt. \quad (\text{B.3})$$

In particolare, si ha

$$u(0) = \int_{-\infty}^{+\infty} \delta(t) u(t) dt. \quad (\text{B.4})$$

Proposizione B.2. Se f è una funzione monotona che si annulla in x_0 e che è ivi derivabile, si ha:

$$\delta[f(x)] = \frac{\delta(x-x_0)}{|f'(x_0)|}. \quad (\text{B.5})$$

Dimostrazione. In virtù della Proposizione B.1, per ogni funzione u continua e per ogni suo punto di continuità x_0 risulta

$$\int_{-\infty}^{+\infty} \delta(t-x_0) u(t) dt = u(x_0).$$

Consideriamo ora l'integrale

$$\int_{-\infty}^{+\infty} \delta[f(x)] u(x) dx \quad (\text{B.6})$$

ed effettuiamovi la sostituzione $y = f(x)$, $V(y) = \frac{u(x)}{|f'(x)|}$. Notiamo che risulta $y = 0$ se e solo se è $x = x_0$ in quanto f è monotona. Si ha allora:

$$\int_{-\infty}^{+\infty} \delta[f(x)] u(x) dx = \int_{-\infty}^{+\infty} \delta(y) V(y) dy = V(0) = \frac{u(x_0)}{|f'(x_0)|}. \quad (\text{B.7})$$

Un confronto tra le (B.4) e (B.7) mostra che la (B.5) è valida. \square

Casi particolari della (B.7) sono evidentemente i seguenti:

$$\int_{-\infty}^{+\infty} \delta(ax-b) u(x) dx = |a|^{-1} u(ba^{-1}), \quad (a \in \mathbb{R}, b \in \mathbb{R}), \quad (\text{B.8})$$

$$\delta(x) = \delta(-x). \quad (\text{B.9})$$

La funzione u finora utilizzata, detta *funzione test*, deve essere continua. Per comodità, per dimostrare proprietà che coinvolgono le derivate della funzione δ , che ora definiremo, conviene restringere l'insieme delle funzioni test alle funzioni continue con tutte le loro derivate di ordine qualsiasi e nulle al di fuori di un intervallo finito all'interno del quale cade il punto x_0 in cui la funzione δ è non nulla. Indichiamo siffatto insieme con U .

Definizione B.2. La funzione δ' verificante l'uguaglianza

$$\int_{-\infty}^{+\infty} \delta'(t) u(t) dt = - \int_{-\infty}^{+\infty} \delta(t) u'(t) dt \quad (\text{B.10})$$

per ogni $u \in U$ si dice derivata prima della funzione δ . \triangleleft

Dalla (B.10) segue che la funzione δ' è ancora una funzione delta di Dirac. Possono così iterativamente definirsi le derivate successive. Ad esempio, si dice derivata seconda della funzione δ la derivata prima della funzione δ' , e così via. Sussiste la seguente proposizione:

Proposizione B.3. Per ogni $n \in \mathbb{N}$ e per ogni $u \in U$ risulta:

$$\int_{-\infty}^{+\infty} \delta^{(n)}(x) u(x) dx = (-1)^n \left[\frac{d^n}{dx^n} u(x) \right]_{x=0}, \quad (\text{B.11})$$

dove

$$\delta^{(n)}(x) = \frac{d^n}{dx^n} \delta(x).$$

Dimostrazione. La dimostrazione segue immediatamente dalla (B.10) e dalle proprietà integrali della funzione δ . \square

Proposizione B.4. La funzione δ è la derivata, nel senso della (B.10), della funzione di Heaviside (cfr. § 5.3):

$$H(x) = \begin{cases} 0, & \text{per } x < 0, \\ 1, & \text{per } x \geq 0. \end{cases} \quad (\text{B.12})$$

Dimostrazione. Per ogni $u \in U$ si ha infatti:

$$\begin{aligned} \int_{-\infty}^{+\infty} H'(t) u(t) dt &= \left[H(t) u(t) \right]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} H(t) u'(t) dt \\ &= - \int_0^{+\infty} u'(t) dt = -u(t) \Big|_0^{+\infty} = u(0). \end{aligned}$$

D'altra parte dalla (B.4) per ogni $u \in U$ risulta:

$$\int_{-\infty}^{+\infty} \delta(t) u(t) dt = u(0).$$

Si ha in definitiva:

$$H'(x) = \delta(x), \quad (\text{B.13})$$

e quindi l'asserto. \square

La Proposizione B.4 ammette un'evidente interpretazione probabilistica. Infatti la funzione di Heaviside può essere riguardata come la funzione di distribuzione di una variabile casuale X ; la funzione δ di Dirac, che come mostra la (B.13) ne è la derivata, rappresenta allora la densità di probabilità di X . La (B.2), infine, esprime la condizione di normalizzazione della variabile casuale.

B.2 Integrale di Stieltjes

Al fine di trattare in modo unitario variabili casuali discrete e variabili casuali assolutamente continue è conveniente richiamare la nozione di integrale di Stieltjes. Questo costituisce una generalizzazione dell'integrale di Riemann nel senso che nel calcolo delle somme parziali (il cui limite porta alla definizione di integrale) la partizione non viene effettuata sull'intervallo in cui è definita la funzione integranda, ma sui valori assunti da un'altra funzione preassegnata. Precisamente, sia g una funzione reale definita nell'intervallo $[a, b]$. Ricordiamo che se suddividiamo $[a, b]$ in intervallini parziali mediante i punti $a = x_0 < x_1 < x_2 < \dots < x_n = b$ e scegliamo un punto ξ_i in ognuno degli intervallini (x_{i-1}, x_i) , di ampiezza $\Delta x_i = x_i - x_{i-1}$ ($i = 1, 2, \dots, n$), il limite

$$\lim_{\max \Delta x_i \rightarrow 0} \sum_{i=1}^n g(\xi_i) (x_i - x_{i-1}),$$

se esiste finito, non dipende né dalla partizione effettuata né dalla scelta dei punti ξ_i ; esso prende il nome di integrale di Riemann di g e viene denotato con la scrittura

$$\int_a^b g(x) dx.$$

È intuitivo pensare che ove la suddivisione in intervallini parziali venga effettuata non sull'intervallo $[a, b]$ ma sull'insieme dei valori assunti da una funzione F non decrescente, a variazione limitata, definita in $[a, b]$ ed ivi continua a destra, si possa pervenire ad un risultato analogo. Invero, se g è una funzione continua in $[a, b]$ si può dimostrare che la somma

$$\sum_{i=1}^n g(\xi_i) [F(x_i) - F(x_{i-1})] \quad (\text{B.14})$$

ammette limite quando $n \rightarrow \infty$ e inoltre $\max \Delta x_i \rightarrow 0$. Questo limite prende il nome di integrale di Stieltjes della funzione g rispetto alla funzione F esteso all'intervallo $[a, b]$ e viene denotato con

$$\int_a^b g(x) dF(x), \quad (\text{B.15})$$

B.2 Integrale di Stieltjes

oppure con

$$\int_a^b g dF.$$

Vediamo ora come si definisce l'integrale improprio di Stieltjes, ossia quello in cui l'intervallo di integrazione è infinito. Precisamente, sia g una funzione reale definita e continua nell'intervallo $[a, +\infty)$ e sia F una funzione reale non decrescente, a variazione limitata, definita in $[a, +\infty)$ ed ivi continua a destra. L'integrale di Stieltjes della funzione g rispetto alla funzione F esteso all'intervallo infinito $[a, +\infty)$ è convergente ed è dato da

$$\int_a^{+\infty} g(x) dF(x) := \lim_{b \rightarrow +\infty} \int_a^b g(x) dF(x),$$

qualora il limite a secondo membro esista finito. In maniera analoga si definisce l'integrale improprio di Stieltjes quando l'intervallo infinito di integrazione è $(-\infty, a]$. Supponiamo infine che g sia una funzione reale definita e continua nell'intervallo in \mathbb{R} e che F sia una funzione reale non decrescente, a variazione limitata, anch'essa definita in \mathbb{R} ed ivi continua a destra. L'integrale di Stieltjes della funzione g rispetto alla funzione F esteso all'intervallo $(-\infty, +\infty)$ è convergente ed è dato da

$$\int_{-\infty}^{+\infty} g(x) dF(x) := \int_{-\infty}^a g(x) dF(x) + \int_a^{+\infty} g(x) dF(x),$$

qualora gli integrali impropri presenti a secondo membro siano convergenti.

Si può dimostrare che se la funzione g è continua e limitata, allora esiste il suo integrale di Stieltjes rispetto alla funzione F esteso sia a intervalli finiti che infiniti.

In taluni casi l'integrale di Stieltjes esiste anche se la funzione g è non limitata, il che è di particolare interesse nel calcolo delle probabilità come accade, ad esempio, allorché si definiscono i momenti di una variabile casuale (cfr. Capitolo 6).

L'integrale di Stieltjes evidentemente coincide con l'integrale di Riemann quando $F(x) = x$; esso, oltre a godere di tutte le proprietà fondamentali dell'integrale di Riemann, possiede una proprietà specifica che lo rende particolarmente utile nella trattazione di problemi del calcolo delle probabilità. Sussiste infatti la seguente proposizione:

Proposizione B.5. *Sia $F: [a, b] \rightarrow \mathbb{R}$ una funzione monotona non decrescente e sia g una funzione continua o a variazione limitata in $[a, b]$.*

(i) Se F ammette derivata continua F' , l'integrale di Stieltjes (B.15) si riduce all'integrale di Riemann ordinario:

$$\int_a^b g(x) dF(x) = \int_a^b g(x) F'(x) dx; \quad (\text{B.16})$$

(ii) se F è una funzione a gradini con discontinuità nei punti x_i ($i = 1, 2, \dots$) e corrispondenti salti di ampiezze F_i si ha:

$$\int_a^b g(x) dF(x) = \sum_i g(x_i) F_i; \quad (\text{B.17})$$

(iii) se F è in parte derivabile e in parte discontinua, con discontinuità (numerabili) nei punti x_i ($i = 1, 2, \dots$) e salti di ampiezze F_i , allora risulta:

$$\int_a^b g(x) dF(x) = \int_I g(x) F'(x) dx + \sum_i g(x_i) F_i, \quad (\text{B.18})$$

dove I rappresenta l'unione degli intervalli in cui F è derivabile e dove la somma è estesa a tutti i punti di discontinuità.

Dimostrazione. (i) Se F è derivabile, per il teorema della media si ha:

$$F(x_i) - F(x_{i-1}) = F'(\xi_i)(x_i - x_{i-1}), \quad (\text{B.19})$$

dove ξ_i è un punto opportuno dell'intervallo (x_{i-1}, x_i) . Moltiplicando ambo i membri della (B.19) per $g(\xi_i)$ e sommando su i da 1 a n si ottiene:

$$\sum_{i=1}^n g(\xi_i) [F(x_i) - F(x_{i-1})] = \sum_{i=1}^n g(\xi_i) F'(\xi_i) (x_i - x_{i-1}).$$

Passando al limite per $n \rightarrow \infty$, dalla definizione stessa di integrale di Stieltjes segue la (B.16).

(ii) Se F è una funzione a gradini, scelta una partizione di $[a, b]$ tale che in ogni intervallo parziale cada al più un punto di discontinuità, al crescere del numero n di intervalli parziali in ciascuno di essi continua a cadere al più un punto di discontinuità di F . Pertanto si ha $F(x_i) - F(x_{i-1}) = 0$ per ogni intero i tale che in $[x_{i-1}, x_i]$ la funzione F è continua (e quindi costante), così che risulta:

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n [F(x_i) - F(x_{i-1})] = \sum_i F_i$$

che si identifica con la (B.17) per $g \equiv 1$. Se g è una funzione qualsiasi analogamente segue la (B.17). (iii) La (B.20) si ottiene, infine, combinando le (i) e (ii). \square

Dalla Proposizione B.5 discende che se si sceglie come funzione F la funzione di distribuzione di una variabile casuale X (il che è lecito in quanto tale funzione è monotona

non decrescente), l'integrale di Stieltjes può essere sempre definito, quale che sia la variabile casuale X (assolutamente continua, discreta, o mista).

Abbiamo già detto che per l'integrale di Stieltjes valgono tutte le proprietà e regole che sussistono per l'integrale di Riemann; abbiamo inoltre visto che esso, così come l'integrale di Riemann, può essere esteso a intervalli illimitati. Se poi F è una funzione di distribuzione risulta:

$$\int_{-\infty}^{+\infty} dF(x) = 1, \quad (\text{B.20})$$

come è immediato dimostrare ricordando le proprietà delle funzioni di distribuzione. La (B.20) esprime in forma unitaria le formule (5.16) e (5.5) determinate separatamente per variabili casuali assolutamente continue e discrete. Infatti nel primo caso F è derivabile e risulta $F' = f$, con f funzione densità di probabilità; la (B.20) per la (B.16) dà allora:

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

Nel secondo caso F è costante a tratti e i salti F_i sono le probabilità dei valori della variabile casuale nei punti di discontinuità. Per la (B.17), la (B.20) assume allora la forma:

$$\sum_i p(x_i) = 1,$$

che equivale alla (5.5).

Occorre precisare che esiste una differenza rilevante tra l'integrale di Riemann e l'integrale di Stieltjes. Per quest'ultimo, infatti, nel definire l'intervallo di integrazione è indispensabile indicare se si includono o meno i punti estremi. Tale precisazione è necessaria in quanto dalla definizione di integrale di Stieltjes segue:

$$\begin{aligned} \int_a^b g(x) dF(x) &= \lim_{n \rightarrow \infty} \sum_{i=1}^n g(\xi_i) [F(x_i) - F(x_{i-1})] \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^{n-1} g(\xi_i) [F(x_i) - F(x_{i-1})] + \lim_{x_{n-1} \rightarrow x_n = b} g(\xi_n) [F(x_n) - F(x_{n-1})] \\ &= \int_a^{b^-} g(x) dF(x) + g(b) [F(b) - F(b^-)], \end{aligned}$$

dove il simbolo b^- indica che b è escluso dall'intervallo di integrazione. Analogamente, risulta:

$$\int_a^b g(x) dF(x) = \int_{a^+}^b g(x) dF(x) + g(a) [F(a^+) - F(a)],$$

dove a^+ indica che l'intervallo di integrazione è $(a, b]$.

Chiaramente, se b è un punto di continuità per F è indifferente che l'estremo b sia incluso o escluso dall'intervallo di integrazione (come accade per l'integrale di Riemann). Da quanto detto segue che l'integrale di Stieltjes esteso ad un intervallo che si riduce ad un punto può essere non nullo: ciò accade appunto quando F possiede una discontinuità nel punto in questione. D'ora in poi, a meno di diversa esplicita indicazione, assumeremo che l'estremo destro dell'intervallo è incluso nell'intervallo di integrazione, mentre quello sinistro è escluso. Tale convenzione consente di scrivere

$$\int_a^b dF(x) = F(b) - F(a).$$

Infatti, per definizione risulta:

$$\begin{aligned} \int_a^b dF(x) &= \lim_{n \rightarrow \infty} \sum_{i=1}^n [F(x_i) - F(x_{i-1})] \\ &= \lim_{n \rightarrow \infty} [F(x_n) - F(x_0)] \\ &= F(b) - F(a), \end{aligned}$$

avendo ricordato che F è continua a destra così che

$$\lim_{\epsilon \downarrow 0} F(a + \epsilon) = F(a).$$

In particolare si osservi che se F è la funzione di distribuzione di una variabile casuale X , si ha:

$$\int_a^b dF(x) = F(b) - F(a) = \mathbb{P}(a < X \leq b).$$

Elenchiamo qui di seguito, senza dimostrazione, alcune proprietà dell'integrale di Stieltjes.

(a) Se $g(x) = k$ (k costante) per $x \in (a, b]$, allora risulta:

$$\int_a^b g(x) dF(x) = k[F(b) - F(a)].$$

(b) Se $\int_a^b g(x) dF(x)$ esiste, si ha:

$$\int_a^b k g(x) dF(x) = k \int_a^b g(x) dF(x).$$

B.2 Integrale di Stieltjes

(c) Se $\int_a^b g_1(x) dF(x)$ e $\int_a^b g_2(x) dF(x)$ esistono, sussiste l'uguaglianza

$$\int_a^b [g_1(x) + g_2(x)] dF(x) = \int_a^b g_1(x) dF(x) + \int_a^b g_2(x) dF(x).$$

(d) Se $\int_a^b g_1(x) dF(x)$ e $\int_a^b g_2(x) dF(x)$ esistono e se $g_1(x) \leq g_2(x)$ per $x \in (a, b]$, sussiste la disegualanza

$$\int_a^b g_1(x) dF(x) \leq \int_a^b g_2(x) dF(x).$$

(e) Sia $g(x)$ limitata in $(a, c]$ e sia $a < b < c$; se $\int_a^c g(x) dF(x)$ esiste, allora $\int_a^b g(x) dF(x)$ e $\int_b^c g(x) dF(x)$ esistono e si ha:

$$\int_a^c g(x) dF(x) = \int_a^b g(x) dF(x) + \int_b^c g(x) dF(x). \quad (\text{B.21})$$

(f) Se $\int_a^b g(x) dF(x)$ e $\int_b^c g(x) dF(x)$ esistono e se g è continua in c , allora $\int_a^c g(x) dF(x)$ esiste e vale la (B.21).

(g) Se $g(x)$ è continua in $(a, b]$ si ha:

$$\left| \int_a^b g(x) dF(x) \right| \leq \int_a^b |g(x)| dF(x).$$

(h) Se $\int_a^b g(x) dF(x)$ esiste, esiste anche $\int_a^b F(x) dg(x)$ e risulta:

$$\int_a^b F(x) dg(x) = g(b)F(b) - g(a)F(a) - \int_a^b g(x) dF(x).$$

È anche possibile considerare integrali di funzioni complesse di variabile reale, ossia del tipo $g = h + ik$, dove h e k sono funzioni reali e continue in $[a, b]$. In tal caso si pone

$$\int_a^b g(x) dF(x) = \int_a^b h(x) dF(x) + i \int_a^b k(x) dF(x).$$

Tutte le proprietà sopra enunciate si estendono facilmente al caso di integrali di funzioni complesse.

Indice analitico

- algebra, 28
 - σ -algebra, 29
 - σ -algebra di Borel, 33
- barriere assorbenti, 212
- boreliani, 34
- coefficiente
 - di asimmetria, 183
 - di correlazione, 134, 142
- condizionamento nel continuo, 229
- convergenza
 - in distribuzione, 198
 - in legge, 198
- covarianza, 133
- cumulant, 182
- densità di probabilità, 94
 - beta, 284, 286
 - congiunta, 107
 - di Cauchy, 191
 - di transizione, 241, 246
 - esponenziale, 278
 - gamma, 281
 - gaussiana, 161
 - lognormale, 105
 - marginale, 233
 - normale, 190, 279
 - normale standard, 104
- drift, 250
- equazione
 - di Chapman-Kolmogorov, 247
 - di diffusione, 253
 - di Fokker-Planck, 255
 - di Kolmogorov, 255
 - di Smolukowski, 247
- stazionaria, 252, 257
- uniforme, 190, 276

esperimento casuale, 11
 evento, 11, 12
 impossibile, 29
 certo, 29
 elementare, 30
 quasi certo, 60
 quasi impossibile, 60
 formula di inclusione-esclusione, 47
 funzione
 beta, 283
 caratteristica, 173
 delta di Dirac, 10, 289
 di distribuzione, 83
 di distribuzione congiunta, 106
 di distribuzione marginale, 106
 di Gauss, 10
 di probabilità, 88
 gamma, 280
 generatrice dei cumulanti, 179
 generatrice dei momenti, 152
 generatrice di probabilità, 149
 massa, 88
 indipendenza, 63, 234
 di variabili casuali, 110
 insiemi di alternative, 76
 integrale di Stieltjes, 292
 intervallo interquartile, 98
 legge
 0-1, 68
 debole dei grandi numeri, 206
 dei 3σ , 164
 della varianza, 147
 delle alternative, 77
 empirica del caso, 17
 lemma di Borel-Cantelli, 68
 media aritmetica, 121

mediana, 97
 metodi di Monte Carlo, 24
 momenti
 assoluti, 133
 centrali, 125, 132
 centrali prodotto, 133
 condizionati, 235
 infinitesimali, 250
 intorno all'origine, 125, 132, 179
 prodotto, 133
 moto browniano, 9
 standard, 261
 paradosso di Bertrand, 22
 passeggiata aleatoria, 209, 212
 percentile, 98
 probabilità, 43
 a priori, 21
 condizionata, 71, 240
 congiunta, 76
 costruzione assiomatica, 27
 geometrica, 21
 soggettiva, 24
 totale, 46
 processo
 di Bernoulli, 32
 di diffusione, 251
 di Markov, 241, 245, 250
 di Ornstein-Uhlenbeck, 265
 di Poisson, 157, 283
 di Wiener, 258
 di Wiener con drift, 262
 di Wiener standard, 261
 stocastico, 243
 quartile, 98
 skewness, 183
 spazio
 campione, 11

campione di Bernoulli, 32
 degli stati, 244
 di probabilità, 44
 statistiche d'ordine, 286
 tempo di attesa, 283
 teorema
 centrale del limite, 208
 delle alternative, 77
 di Bayes, 78, 233
 di Bernoulli, 167
 di equivalenza, 55
 di estensione, 52, 53
 valore medio, 117, 131
 condizionato, 236

variabile aleatoria, 82
 variabile casuale, 81
 n-dimensionale, 106
 bidimensionale, 109
 di Cauchy, 119
 discreta, 87
 esponenziale, 130
 gaussiana, 127
 standardizzata, 127
 variabili casuali
 prodotto di, 114
 rapporto di, 115
 somma di, 112
 varianza, 124
 infinitesimale, 251
 vettore casuale, 106, 238

