

**Nicola Cufaro Petroni**

**Lezioni di**

**CALCOLO  
delle PROBABILITÀ**

**Edizioni dal Sud**

© 1996 Edizioni dal Sud  
S.S. 98 km 81,100 - 70026 MODUGNO (Bari)  
Tel/Fax (080) 5353705

Il testo è stato composto dall'autore in  $\text{T}_{\text{E}}\text{X}$

Opera stampata con il contributo dell'Università di Bari e del M.U.R.S.T.

Finito di stampare nel giugno 1996  
dalle Arti Grafiche Ariete - Modugno

*“ ... Es gibt keinen Zufall;  
Und was uns blindes Ohngefähr nur dünkt,  
Gerade das steigt aus den tiefsten Quellen.”*

F. Schiller, *Wallensteins Tod* (1799);  
zweiter Aufzug, dritter Auftritt.\*

---

\* *“ ... Il caso non esiste; e quel che a noi appare cieco accidente è proprio ciò che scaturisce dalle sorgenti più profonde.”* (F. Schiller: *La morte di Wallenstein*; 1799; atto secondo, scena terza).



## I N D I C E

PREFAZIONE .....	pag. 9
------------------	--------

### I. Teoria elementare delle Probabilità

I.1 Introduzione: Fortuna e Rovina del Cavalier de Méré .....	pag. 13
I.2 Modelli finiti di Probabilità .....	19
I.3 Distribuzione Binomiale .....	29
I.4 Condizionamento ed Indipendenza .....	37
I.5 Variabili aleatorie .....	49
I.6 Valore d'attesa e Varianza .....	61
I.7 Modello di Bernoulli e Legge dei grandi Numeri .....	71
I.8 Modello di Bernoulli e Teoremi limite .....	77
I.9 Valore d'attesa condizionato .....	93
I.10 Probabilità di Rovina .....	101
I.11 Esercizi svolti .....	111

### II. Basi matematiche del Calcolo delle Probabilità

II.1 Modelli di probabilità. Assiomi .....	125
II.2 Spazi probabilizzabili .....	135
II.3 Misure di Probabilità (I Parte) .....	153
II.4 Misure di Probabilità (II Parte) .....	173
II.5 Variabili aleatorie (I Parte) .....	187
II.6 Valore d'attesa e integrazione .....	211
II.7 Condizionamento .....	239
II.8 Variabili aleatorie (II Parte) .....	259
II.9 Convergenza. Spazi vettoriali di variabili aleatorie .....	277
II.10 Funzioni caratteristiche .....	293
II.11 Sistemi gaussiani .....	309
II.12 Esercizi svolti .....	317

### III. Successioni di Variabili Aleatorie

III.1 Teoremi Limite e Leggi dei Grandi Numeri .....	333
III.2 Martingale .....	351
III.3 Catene di Markov .....	363
III.4 Esercizi svolti .....	389

INDICE ANALITICO .....	397
------------------------	-----



## P R E F A Z I O N E

Queste Lezioni non costituiscono un *manuale*: di questo manca loro l'ambizione alla completezza e alla sistematicità. Si tratta piuttosto di un percorso guidato attraverso le principali idee del Calcolo delle Probabilità con lo scopo di dare agli studenti del Corso di Laurea in Matematica i concetti e gli strumenti necessari per affrontare sia eventuali ulteriori approfondimenti (Tesi di Laurea, lavoro di ricerca) sia le domande, non banali, poste dall'insegnamento nella Scuola Secondaria. È ben chiaro all'autore che tale obiettivo è solo molto imperfettamente raggiunto dalla presente edizione, ma gli è anche evidente che un reale miglioramento non potrebbe essere ottenuto solo cedendo alla tentazione di un indiscriminato completamento degli argomenti: se non altro, lo vieterebbero le limitazioni imposte dall'angustia di un corso universitario. Sarebbe necessario piuttosto un diverso bilancio fra alcune parti e, in taluni casi, il sacrificio di qualche sezione in favore di problemi (ingiustamente) rimasti fuori della trattazione.

Non è questo il luogo adatto per enumerare le debolezze e per proporre le correzioni del caso: le future edizioni, rese agevoli dai moderni mezzi di scrittura, si incaricheranno di evidenziare le nuove scelte e i nuovi equilibri. Qui ci limiteremo invece a poche altre osservazioni. Evitando di riprodurre, come spesso accade, tutte le dimostrazioni dei teoremi di un certo rilievo, eventualmente private dei dettagli tecnici di alcuni passaggi lasciati al lettore ("...è facile provare che..."), si è preferito sviluppare alcune prove anche nei loro particolari rinviando altre, corredate solo di alcuni commenti ed esempî, a testi ormai classici. L'autore ritiene che le dimostrazioni troppo sintetiche inducano piuttosto gli allievi a trascurare i dettagli ritenuti ovvî privandole così della loro doppia funzione di essere laboratori di *trucchi* del mestiere e strumenti per una comprensione più profonda degli enunciati. Per questo motivo le dimostrazioni qui riportate sono in generale svolte nei loro particolari anche se, per evitare i rischi di dispersione causati dalla loro lunghezza, sono spesso sezionate nei loro passaggi essenziali. L'autonomia e l'iniziativa degli studenti potranno poi essere sviluppate con piccole dimostrazioni o esercizi da svolgere parallelamente al corso.

Le lezioni si articolano in tre parti. Nella prima vengono presentati, in forma elementare e per spazi finiti, tutti i concetti rilevanti della probabilità, mostrando con vari esempî in che modo si costruiscono modelli probabilistici dei fenomeni reali. Uno degli scopi non secondari di questa prima esposizione è quello di mettere in evidenza la necessità di una costruzione probabilistica rigorosa in spazi, in genere, non finiti: tale costruzione viene poi realizzata nella seconda parte. È stata anche preoccupazione costante dell'autore in questa sezione quella di indicare le connessioni degli argomenti esposti con altri settori della matematica e della fisica: Teoria della misura, Analisi funzionale, Statistica, Teoria dei processi stocastici, Meccanica statistica. L'esposizione della potenza e della flessibilità degli strumenti probabilistici è stata infatti ritenuta essenziale per accendere l'interesse degli studenti. I collegamenti sono stati anche corredate, ovunque possibile, con un certo

numero di opportuni riferimenti bibliografici. La terza parte costituisce un primo saggio di applicazione delle tecniche e delle idee esposte alle successioni di variabili aleatorie con l'intenzione di alludere alla dinamica dei fenomeni aleatori e alla teoria dei processi stocastici pur rimanendo, a causa delle dimensioni limitate di un corso, nel dominio dei tempi discreti. Non si è trascurato, infine, di accennare alle vicende biografiche degli studiosi i cui nomi sono rimasti associati a idee, teoremi o modelli probabilistici con il doppio scopo di inquadrare cronologicamente lo sviluppo del Calcolo delle Probabilità e di mostrare come spesso le costruzioni intellettuali più ardite ed i loro autori non fossero estranei alla vita culturale e politica che si svolgeva attorno a loro.

Non sarebbe giusto chiudere queste osservazioni preliminari senza citare i testi classici dai quali è stato attinto, modificandolo, il materiale per queste lezioni, ma la redazione di una vera bibliografia costituisce un impegno che esula dalle intenzioni di questo scritto. Rinvierebbe pertanto il lettore alle citazioni richiamate di volta in volta nelle pagine del testo limitandoci per il rimanente a ricordare che i volumi verso i quali l'autore si sente più pesantemente debitore sono i manuali di A.N. Shiriyayev e di M. Métivier nelle edizioni più innanzi citate. Infine un sincero ringraziamento va a tutti i colleghi che con i loro consigli e le loro osservazioni hanno contribuito alla realizzazione di quest'opera.

*L'Autore*

Bari, Aprile 1996



I  
TEORIA ELEMENTARE  
delle PROBABILITÀ



# I.1 Introduzione:

## Fortuna e Rovina del Cavalier de Méré

La nascita del Calcolo delle Probabilità (collocata generalmente attorno alla metà del XVII secolo) è strettamente legata ai problemi del *gioco d'azzardo*: fin dall'inizio, infatti, esso fu considerato come un sistema che consentiva di stimare la verosimiglianza di eventi, e quindi come un utile strumento per dare veste quantitativa alle previsioni dei giocatori. Le probabilità di eventi complicati potevano essere ricostruite partendo da eventi più elementari cui facilmente potevano essere attribuiti dei *pesi* sulla base di osservazioni o di opportune ipotesi. Nel seguito (e fino ad avviso contrario) considereremo unicamente esperimenti (giochi) che diano luogo solo ad un *numero finito di possibili risultati*.

**I.1.1 Esempio:** Il caso più semplice, ma che comunque consente di rendere un po' più concrete le osservazioni precedenti, è quello del lancio di una moneta nel quale si osserva il verificarsi di uno dei due eventi elementari possibili: la moneta cade mostrando la faccia con la *testa* ( $T$ ); oppure la moneta cade mostrando la faccia con la *croce* ( $C$ ). Dire che la moneta è **equa** vuol dire supporre che, essendo essa stata prodotta nella maniera più *simmetrica* possibile (cioè essendo *non truccata*), è possibile attribuire ai due eventi elementari suddetti le medesime possibilità di verificarsi; in tal caso diremo anche che i due eventi elementari  $T$  e  $C$  sono **equiprobabili**. Per dare una veste quantitativa a queste considerazioni si usa attribuire ad ogni evento elementare una **probabilità** intesa come frazione dell'unità, sicché nel nostro caso avremo:

$$p = \mathbf{P}(T) = \frac{1}{2}; \quad q = \mathbf{P}(C) = \frac{1}{2}.$$

Osserviamo che  $p + q = 1$ , il che riflette il fatto che *con certezza* (ossia con probabilità eguale ad 1) uno dei due casi,  $T$  oppure  $C$ , si verifica e che non vi sono altre possibilità.  $\diamond$

**I.1.2 Esempio:** Considerazioni analoghe a quelle dell'Eempio I.1.1 applicate al caso di un dado *equo* conducono alla seguente attribuzione di probabilità per le sei facce che qui indicheremo con le cifre romane  $I, II, \dots, VI$ :

$$p_1 = \mathbf{P}(I) = \frac{1}{6}; \quad \dots \quad p_6 = \mathbf{P}(VI) = \frac{1}{6}.$$

Osserviamo che anche in questo caso si ha  $p_1 + \dots + p_6 = 1$ .  $\diamond$

Dagli esempi precedenti si ricava anche l'idea che, almeno per casi semplici, si possano attribuire delle probabilità mediante un'*enumerazione*. Questo conduce alla cosiddetta **definizione classica** della probabilità in base alla quale l'attribuzione di un peso probabilistico ad un evento  $A$  (in generale *non* elementare, cioè non ridotto ad un solo risultato) può essere eseguita enumerando sia gli eventi elementari

*possibili* (e ritenuti, in base a qualche ipotesi, equiprobabili) che quelli *favorevoli* all'evento in considerazione (quelli, cioè, che darebbero luogo al verificarsi di  $A$ ) ed attribuendo all'evento  $A$  la probabilità:

$$\mathbf{P}(A) = \frac{\text{numero dei casi favorevoli}}{\text{numero dei casi possibili}}.$$

Notiamo che anche in questo caso la probabilità assegnata ad  $A$  è un numero positivo e compreso fra 0 ed 1.

**I.1.3 Esempio:** Nel lancio di un dado equo (come quello dell'Esempio I.1.2) consideriamo gli eventi (*non* elementari)  $A =$  “appare una faccia contrassegnata da un numero pari”,  $B =$  “appare una faccia contrassegnata da un multiplo di tre”,  $C =$  “appare una faccia diversa da  $VI$ ”. Una semplice enumerazione in base alla definizione classica ci porta a concludere che, essendo 6 i casi possibili, e rispettivamente 3, 2 e 5 i casi favorevoli ad  $A, B$  e  $C$ , si avrà:

$$\mathbf{P}(A) = \frac{1}{2}; \quad \mathbf{P}(B) = \frac{1}{3}; \quad \mathbf{P}(C) = \frac{5}{6}.$$

È questo un primo esempio di *calcolo* delle probabilità inteso nel senso della valutazione della probabilità di eventi *complicati* a partire dalle probabilità di eventi *elementari*. Osserviamo che, sebbene uno degli eventi dati si verifica sempre, la somma delle tre probabilità *non* è eguale all'unità: questo è dovuto al fatto che gli eventi  $A, B$  e  $C$  non si escludono a vicenda.  $\diamond$

**I.1.4 Esempio:** Consideriamo ora un lancio di *due* dadi non truccati. È facile verificare che i risultati elementari possibili sono ora 36, cioè quante sono le coppie *ordinate*  $(n, m)$  dove  $n$  ed  $m$  possono assumere i 6 valori  $I, \dots, VI$ . L'ipotesi che i dadi siano equi vuol dunque dire ora che i 36 eventi elementari  $(I, I); (I, II); \dots; (VI, VI)$  sono tutti equiprobabili e pertanto si ha che:

$$\mathbf{P}(I, I) = \frac{1}{36}; \quad \mathbf{P}(I, II) = \frac{1}{36}; \quad \dots; \quad \mathbf{P}(VI, VI) = \frac{1}{36}.$$

Sempre per enumerazione si può verificare anche che all'evento  $A =$  “non appare la coppia  $(VI, VI)$ ” si può attribuire una probabilità

$$\mathbf{P}(A) = \frac{35}{36}. \quad \diamond$$

Dalla discussione precedente possiamo dunque trarre la prima conclusione che la probabilità di un evento può essere definita come un numero compreso tra 0 ed 1 che misura la possibilità del verificarsi del dato evento: il valore 1 indica la *certezza* del verificarsi ed il valore 0 la sua *impossibilità*; i valori intermedi rappresentano tutti gli altri casi. Queste probabilità possono essere calcolate (in alcuni casi)

mediante enumerazione di eventi elementari equiprobabili, ma va detto subito che questo metodo diviene rapidamente *inutilizzabile* a causa del gran numero dei casi elementari da esplorare.

**I.1.5 Esempio:** Consideriamo un mazzo di carte *francesi* (52 carte) ed enumeriamo le possibili *successioni* (*senza ripetizioni*) delle 52 carte. Si verifica subito che il loro numero è enorme:

$$52 \cdot 51 \cdot 50 \cdot \dots \cdot 2 \cdot 1 = 52! \simeq 8 \cdot 10^{67}$$

e che quindi qualunque speranza di risolvere problemi di calcolo delle probabilità *per enumerazione* è vana.  $\diamond$

Bisogna inoltre tenere presente che, anche nei casi in cui la definizione classica è applicabile, la costruzione del *modello* a cui essa va applicata deve essere fatta con estrema cura per evitare di commettere errori. La seguente discussione del celebre **problema del Cavalier de Méré**<sup>1</sup> ne è un eloquente esempio.

Un gioco in voga in Francia nel XVII secolo (da ora in poi detto *primo gioco*) consisteva nel lanciare per quattro volte un dado (supposto equo) e nello scommettere sull'evento  $A =$  “appare VI almeno una volta”. Da una lunga serie di osservazioni de Méré aveva tratto la conclusione che la probabilità di vincere scommettendo sul verificarsi di  $A$  era leggermente superiore alla probabilità di perdere. Sempre alla ricerca di regole complicate che gli consentissero di sfruttare un *vantaggio nascosto* a spese dei suoi avversari, il nostro cavaliere pensò di rendere più intricato il primo gioco modificandolo nel modo seguente (da ora in poi ci riferiremo a questa seconda versione come al *secondo gioco*): si lanciano 24 volte due dadi e si scommette sul verificarsi dell'evento  $B =$  “appare (VI, VI) almeno una volta”. Il suo ragionamento deve essere stato più o meno il seguente: *il lancio di un dado ammette 6 risultati possibili; quello di due dadi 36, cioè 6 volte di più. VI è uno dei 6 possibili risultati del primo gioco mentre (VI, VI) è uno dei 36 possibili risultati del secondo gioco. Pertanto, se nel secondo gioco si lanciano i dadi 6 volte di più che nel primo gioco (cioè 24 volte!), la probabilità di vincere scommettendo su (VI, VI) deve essere eguale alla probabilità di vincere nel primo gioco scommettendo su VI.*

<sup>1</sup> Il Cavalier de Méré, letterato e filosofo, visse in Francia tra il 1607 ed il 1684. Personaggio notevole della corte di Luigi XIV, acquistò fama di essere anche un giocatore impenitente, benché la sua corrispondenza con i più importanti pensatori della sua epoca (Pascal, Fermat, Roberval,...) lo faccia apparire più come uno spirito curioso che come un *habitué* del gioco d'azzardo. Non è chiaro quanto negli aneddoti che si raccontano di lui sia pura leggenda e quanto sia realtà. Resta il fatto che egli seppe proporre a Pascal e Fermat alcuni celebri problemi a proposito dei quali si conserva un lunga discussione nella corrispondenza scambiata fra questi due matematici nell'estate del 1654. Qui discuteremo solo uno di questi problemi, quello della cosiddetta *partita a dadi*, mentre quello della *suddivisione della posta in gioco* sarà affrontato in I.10 (vedi **P. Deheuvels**: *La Probabilité, le Hasard et la Certitude*; P.U.F., Paris, 1990; **R. von Mises**: *Probability, Statistics and Truth*; Dover, New York, 1981).

La leggenda vuole che egli abbia fatto la sua fortuna giocando al primo gioco, ma che si sia poi rovinato con il secondo! Infatti il ragionamento per analogia qui utilizzato non poteva che condurlo a delle previsioni errate come dimostrò poi Fermat<sup>2</sup> con un calcolo che è generalmente considerato come l'atto di nascita del calcolo delle probabilità.

Che il calcolo precedente si basi su premesse confuse può essere mostrato facilmente. Ad esempio esso potrebbe fondarsi sull'idea che se  $p_1 = \frac{1}{6}$  è la probabilità che esca  $VI$  in uno qualunque dei 4 lanci del primo gioco la probabilità di vincere deve essere  $\mathbf{P}(A) = 4p_1 = \frac{2}{3}$  e che per il secondo gioco l'analogo calcolo con  $p_2 = \frac{1}{36}$  conduce al medesimo risultato:  $\mathbf{P}(B) = 24p_2 = \frac{2}{3}$ . Ma che questi calcoli siano assurdi è presto visto<sup>3</sup>: infatti secondo questo ragionamento all'evento  $C =$  "appare  $VI$  almeno una volta su 7 lanci di un dado" bisognerebbe attribuire la probabilità  $\mathbf{P}(C) = 7p_1 = \frac{7}{6} > 1$  !!

Sarà istruttivo, a questo punto, esaminare in dettaglio la soluzione di questo problema. Cominciamo con il prendere in considerazione il primo gioco: se esso consistesse in un solo lancio del dado la probabilità di ottenere  $VI$  come risultato sarebbe banalmente  $\frac{1}{6} \simeq 0.167$ . Se invece lanciamo il dado per due volte, posto

$$p_1 = \mathbf{P}(\text{appare } VI \text{ al primo lancio}) = \frac{1}{6},$$

$$p_2 = \mathbf{P}(\text{appare } VI \text{ al secondo lancio}) = \frac{1}{6},$$

dobbiamo subito osservare che

$$p = \mathbf{P}(\text{appare } VI \text{ in almeno un lancio}) \neq p_1 + p_2 = \frac{1}{3}.$$

Infatti i due eventi  $A_1 =$  "appare  $VI$  al primo lancio" e  $A_2 =$  "appare  $VI$  al secondo lancio" non sono mutuamente esclusivi (il caso elementare in cui appare

---

<sup>2</sup> Pierre de Fermat (1608 - 1665), consigliere del Parlamento provinciale di Tolosa, fu attratto verso la teoria dei numeri dalla lettura di una traduzione delle opere di Diofanto. Le sue numerose scoperte in questo campo furono affidate per la maggior parte al suo epistolario con alcuni matematici suoi contemporanei o furono annotate (come nel caso del suo famoso teorema) in margine ai libri che leggeva. I suoi interessi, comunque, non si limitarono alla teoria dei numeri: ad esempio alcuni problemi di gioco d'azzardo (come quello discusso in queste pagine o quello della divisione della posta fra due giocatori che siano costretti ad interrompere il gioco prima della sua fine) propostigli dal Cavalier de Méré nel 1654 divennero oggetto di una serie di lettere scambiate con Pascal e che sono tuttora considerate fondamentali nello sviluppo del concetto di probabilità.

<sup>3</sup> In questa discussione viene per la prima volta adombrata la relazione esistente fra la disgiunzione, o alternativa logica non esclusiva, *oppure* e la *somma* delle probabilità. Infatti l'evento  $A =$  " $VI$  appare almeno una volta su 4 lanci" coincide con l'evento  $A =$  " $VI$  appare al primo lancio *oppure* al secondo *oppure* al terzo *oppure* al quarto". Vedremo in seguito, però, che la probabilità di una disgiunzione di eventi è pari alla somma delle probabilità dei singoli eventi solo se questi eventi sono mutuamente esclusivi, caso che non si verifica nella presente discussione in quanto il risultato  $VI$  in un lancio non esclude il medesimo risultato in un altro lancio.

$VI$  al primo e al secondo lancio è un caso favorevole ad ambedue questi eventi) per cui sommare le loro rispettive probabilità  $p_1$  e  $p_2$  significherebbe contare due volte alcuni eventi elementari. Più precisamente, considerati i 36 possibili (ed equiprobabili) risultati di due lanci di un dado:

$$\begin{pmatrix} (I, I) & (II, I) & \dots & (VI, I) \\ (I, II) & (II, II) & \dots & (VI, II) \\ \vdots & \vdots & \ddots & \vdots \\ (I, VI) & (II, VI) & \dots & (VI, VI) \end{pmatrix}$$

gli eventi elementari dell'ultima colonna sono *favorevoli* ad  $A_1$ , mentre quelli dell'ultima riga sono *favorevoli* ad  $A_2$ : ognuno degli eventi elementari porta con sé  $\frac{1}{36}$  di probabilità, ma bisogna fare attenzione a non contare due volte il caso  $(VI, VI)$ . Pertanto risulta

$$\begin{aligned} \mathbf{P}(VI \text{ in almeno un lancio}) &= \frac{1}{6} + \frac{1}{6} - \frac{1}{36} \\ &= \frac{6}{36} + \frac{5}{36} = \frac{11}{36} \simeq 0.305 < \frac{1}{3}. \end{aligned}$$

In maniera analoga sui 216 possibili risultati di 3 lanci si ha

$$\mathbf{P}(VI \text{ in almeno un lancio}) = \frac{6 \cdot 6 + 6 \cdot 5 + 5 \cdot 5}{6^3} \simeq 0.421,$$

e sui 1.296 possibili risultati di 4 lanci risulta

$$\mathbf{P}(VI \text{ in almeno un lancio}) = \frac{6 \cdot 6 \cdot 6 + 6 \cdot 6 \cdot 5 + 6 \cdot 5 \cdot 5 + 5 \cdot 5 \cdot 5}{6^4} \simeq 0.518,$$

risultato che rappresenta il corretto valore della probabilità di vincita al primo gioco: come de Méré aveva osservato, il gioco risulta effettivamente favorevole, anche se di poco, a chi scommette sull'uscita del  $VI$ .

Va osservato subito, comunque, che il metodo dell'enumerazione dei casi possibili e di quelli favorevoli, che già nella discussione del primo gioco si è rivelato un po' laborioso, diviene assolutamente impraticabile nell'analisi del secondo gioco. Pertanto sarà utile tentare di trovare dei metodi diversi per ottenere il medesimo risultato. Sempre limitandoci per ora alla discussione del primo gioco, osserviamo che, siccome un evento  $A$  e la sua negazione  $\bar{A}$  sono ovviamente mutuamente esclusivi e siccome l'evento " $A$  oppure  $\bar{A}$ " è ovviamente l'evento certo, avremo anche che  $\mathbf{P}(A) + \mathbf{P}(\bar{A}) = 1$ . Nel nostro caso potremo quindi dire che

$$\mathbf{P}(VI \text{ almeno una volta su 4 lanci}) = 1 - \mathbf{P}(VI \text{ non appare mai su 4 lanci}).$$

Inoltre l'evento nel quale  $VI$  non appare mai può essere considerato come la congiunzione ( $e$ ) dei quattro eventi nei quali  $VI$  non appare in ciascuno dei singoli

lanci: “ $VI$  non appare al primo lancio e ... e non appare al quarto”. Siccome infine il risultato di ogni lancio non è influenzato dai risultati degli altri<sup>4</sup>, e siccome la probabilità che  $VI$  non esca in ciascun singolo lancio è  $\frac{5}{6}$ , si perviene alla conclusione che

$$\mathbf{P}(VI \text{ almeno una volta su 4 lanci}) = 1 - \left(\frac{5}{6}\right)^4 \simeq 0.518.$$

Ciò mostra che questo secondo metodo *indiretto* di calcolo è molto vantaggioso in quanto consente di evitare lunghe enumerazioni.

Passiamo infine ad applicare questo metodo indiretto al calcolo della probabilità di vincita al secondo gioco: tenendo conto delle osservazioni precedenti e del fatto che la probabilità che  $(VI, VI)$  non esca in un singolo lancio è  $\frac{35}{36}$ , otteniamo facilmente che

$$\mathbf{P}((VI, VI) \text{ almeno una volta su 24 lanci}) = 1 - \left(\frac{35}{36}\right)^{24} \simeq 0.491.$$

Ciò mostra che, contrariamente alle speranze di de Méré e diversamente dal primo gioco, il secondo gioco è, sia pur di poco, sfavorevole e spiega quindi come mai il nostro cavaliere si sia rovinato. Ma un'ultima osservazione è d'obbligo: se il secondo gioco consistesse in 25 (invece che in 24) lanci di due dadi il gioco tornerebbe ad essere favorevole:

$$\mathbf{P}((VI, VI) \text{ almeno una volta su 25 lanci}) = 1 - \left(\frac{35}{36}\right)^{25} \simeq 0.505.$$

---

<sup>4</sup> In tutta questa parte della discussione vengono introdotti in maniera intuitiva dei concetti particolarmente importanti che saranno ripresi a lungo nel seguito. In particolare la relazione fra la congiunzione *e* ed il prodotto delle probabilità degli eventi congiunti nel caso in cui tali eventi possano essere considerati *indipendenti*, il caso, cioè, in cui il verificarsi o meno di un evento non influenza il valore della probabilità di un altro evento.



## I.2 Modelli finiti di probabilità

Come abbiamo visto nel precedente capitolo, un ruolo importante nella costruzione degli schemi probabilistici è giocato dai cosiddetti *eventi elementari* cioè dai possibili risultati dell'esperimento che stiamo esaminando. Questa semplice osservazione rende chiaro il motivo per introdurre la seguente definizione<sup>1</sup>

**I.2.1 Definizione:** Chiameremo **spazio dei campioni** o **spazio degli eventi elementari** l'insieme finito

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_N\}$$

composto da tutti i possibili risultati  $\omega_k$  del nostro esperimento. △

**I.2.2 Esempio:** Richiamando quanto esposto negli esempi del paragrafo precedente è facile verificare che nel caso di un solo lancio di una moneta lo spazio dei campioni è composto di soli due elementi:

$$\Omega = \{T, C\}; \quad N = 2,$$

mentre nel caso di un solo lancio di un dado si ha

$$\Omega = \{I, II, \dots, VI\}; \quad N = 6.$$

Se invece l'esperimento consistesse in due lanci di una moneta si avrebbe:

$$\Omega = \{TT, TC, CT, CC\}; \quad N = 4.$$

Naturalmente si avrebbe

$$\Omega = \{\omega = (a_1, \dots, a_n) : a_i = T \text{ oppure } C\}; \quad N = 2^n$$

nel caso di  $n$  lanci di una moneta. ◇

**I.2.3 Esempio (Campionamento con rimessa):** Consideriamo una scatola contenente  $M$  palline numerate (e quindi distinguibili) ed estraiamo successivamente  $n$  palline rimettendo nella scatola dopo ogni estrazione la pallina estratta. Registriamo quindi i numeri segnati sulle palline estratte in modo che un risultato possibile dell'esperimento sarà  $\omega = (a_1, a_2, \dots, a_n)$  con  $a_i = 1, 2, \dots, M$  ed  $i = 1, 2, \dots, n$ . Naturalmente, siccome rimettiamo a posto le palline dopo averle

---

<sup>1</sup> In questa prima parte del testo verranno fornite delle definizioni *provvisorie* per la maggior parte dei concetti fondamentali del calcolo delle probabilità. Tali definizioni saranno riprese con maggior rigore nella seconda parte di queste lezioni.

osservate, i valori delle  $a_i$  possono coincidere (ci sono delle ripetizioni). Lo spazio dei campioni sarà ora l'insieme delle  $n$ -ple di risultati delle estrazioni:

$$\Omega = \{\omega = (a_1, \dots, a_n) : a_i = 1, \dots, M ; i = 1, \dots, n\};$$

In realtà è possibile costruire due diversi spazi dei campioni:

- a) Spazio dei **campioni ordinati**  $(a_1, \dots, a_n)$ : si tratta dei campioni che possono differire anche solo per l'ordine in cui si presentano le *etichette* delle palline estratte e che prendono il nome di *disposizioni*; in questo caso, ad esempio per  $n = 4$  estrazioni, il campione  $(4, 1, 2, 1)$  è considerato diverso dal campione  $(1, 1, 2, 4)$ . In queste condizioni è molto facile verificare che il numero di elementi di  $\Omega$  è  $N_d = M^n$ .
- b) Spazio dei **campioni non ordinati**  $[a_1, \dots, a_n]$ : in questo caso i campioni  $(4, 1, 2, 1)$  e  $(1, 1, 2, 4)$  vengono considerati identici, sicché il numero degli elementi di  $\Omega$ , che prendono il nome di *ripartizioni*, è inferiore a quello del caso a). Più precisamente dimostreremo ora che

$$N_r = \binom{M+n-1}{n} = \frac{(M+n-1)!}{n!(M-1)!}.$$

Per dimostrare la relazione precedente indichiamo con  $N(m, n)$  il numero delle  $n$ -ple non ordinate estratte da una scatola contenente  $m$  palline numerate ( $m = 1, \dots, M$ ) ed osserviamo che, essendo  $N_r = N(M, n)$ , per dimostrare il nostro risultato sarà sufficiente provare che

$$N(m, n) = \binom{m+n-1}{n}; \quad m = 1, \dots, M.$$

Dato che per  $n = 1$  la precedente relazione è banalmente vera:

$$N(m, 1) = m = \binom{m}{1} = \binom{m+1-1}{1}; \quad m = 1, \dots, M,$$

non ci resta che usare il metodo di induzione completa e provare che, se essa è vera per un dato  $n$ , risulta anche vera per  $n + 1$ , cioè

$$N(m, n+1) = \binom{m+n}{n+1}; \quad m = 1, \dots, M.$$

Osserviamo allora che, se  $[a_1, \dots, a_{n+1}]$  è un campione non ordinato, potremo supporre, per comodità, che  $a_1 \leq a_2 \leq \dots \leq a_{n+1}$ , sicché il numero delle possibili  $(n+1)$ -ple  $[a_1, \dots, a_{n+1}]$  si calcola notando che:

- il numero di campioni con  $a_1 = 1$  è  $N(m, n)$ ;
- il numero di campioni con  $a_1 = 2$  è  $N(m-1, n)$ ;
- .....
- il numero di campioni con  $a_1 = m$  è  $N(1, n)$ ;

in quanto, fissato il valore di  $a_1$ , i possibili valori delle successive  $a_i$  devono soddisfare la relazione  $a_i \geq a_1$ . Ne segue che

$$\begin{aligned} N(m, n+1) &= N(m, n) + \dots + N(1, n) \\ &= \binom{m+n-1}{n} + \binom{m+n-2}{n} + \dots + \binom{n+1}{n} + \binom{n}{n} \end{aligned}$$

e siccome è facile mostrare con un calcolo diretto che

$$\binom{k-1}{l} = \binom{k}{l+1} - \binom{k-1}{l+1},$$

otteniamo che

$$\begin{aligned} N(m, n+1) &= \left[ \binom{m+n}{n+1} - \binom{m+n-1}{n+1} \right] + \left[ \binom{m+n-1}{n+1} - \binom{m+n-2}{n+1} \right] + \\ &\quad \dots + \left[ \binom{n+2}{n+1} - \binom{n+1}{n+1} \right] + \binom{n}{n} = \binom{m+n}{n+1} \end{aligned}$$

cioè il risultato richiesto.  $\diamond$

**I.2.4 Esempio (Campionamento senza rimessa):** Si esegue una estrazione di palline numerate da una scatola esattamente come nell'Esempio I.2.3, ma senza rimettere a posto le palline dopo averle estratte. In tal caso è evidente che i campioni ottenuti  $(a_1, \dots, a_n)$  sono composti di elementi tutti diversi (non ci sono ripetizioni), e che  $n \leq M$  visto che non possiamo estrarre un numero di palline superiore a quello contenuto nella scatola. Anche in questo caso ci sono due possibili spazi dei campioni:

- a) Spazio dei **campioni ordinati**  $(a_1, \dots, a_n)$ : si tratta delle cosiddette *permutazioni* di  $M$  oggetti su  $n$  posti. Il numero di tali permutazioni (cioè il numero di elementi del nostro spazio  $\Omega$ ) è ora

$$N_p = (M)_n = \frac{M!}{(M-n)!} = M(M-1)\dots(M-n+1),$$

dato che, ogni volta che estraiamo una pallina, lasciamo una possibilità in meno per le successive estrazioni. Osserviamo che, nel caso in cui  $n = M$ , risulta  $N = M!$ , che è il numero delle permutazioni di  $M$  oggetti su  $M$  posti.

- b) Spazio dei **campioni non ordinati**  $[a_1, \dots, a_n]$ : si tratta questa volta delle cosiddette *combinazioni* di  $M$  oggetti su  $n$  posti. Il numero di queste combinazioni è

$$N_c = \binom{M}{n}.$$

Infatti ogni campione non ordinato  $[a_1, \dots, a_n]$  ammette  $n!$  permutazioni distinte dall'ordine dei suoi simboli (vedi il precedente punto a). Ne segue che  $N_c \cdot n! = (M)_n = M!/(M-n)!$  da cui discende la relazione richiesta.  $\diamond$

**I.2.5 Definizione:** Chiameremo **evento** ogni sottinsieme  $A \subseteq \Omega$  per il quale, nelle condizioni dell'esperimento eseguito, sia possibile dire se un risultato  $\omega \in \Omega$  è o meno un elemento di  $A$ .  $\triangle$

**I.2.6 Esempio:** Nel caso di tre lanci di una moneta lo spazio dei campioni è composto di  $N = 2^3 = 8$  elementi:

$$\Omega = \{TTT, TTC, \dots, CCC\}.$$

Il sottinsieme

$$A = \{TTT, TTC, TCT, CTT\} \subseteq \Omega$$

rappresenterà l'evento *T appare almeno due volte su tre lanci* se il nostro esperimento consiste nella osservazione dei tre lanci della moneta. Se invece ci è possibile osservare solo il risultato del primo lancio, il sottinsieme  $A$  non potrà essere considerato come un evento visto che non potremo rispondere alla domanda se un particolare risultato  $\omega \in \Omega$  appartiene ad  $A$  (cioè: non possiamo dire se  $T$  appare almeno due volte su tre lanci osservando solo uno di questi lanci).  $\diamond$

Gli eventi così definiti possono essere considerati come rappresentazioni di *proposizioni logiche* e le corrispondenti operazioni tra eventi (intese come operazioni tra insiemi) possono essere considerate come un modello per i *connettivi logici* che uniscono delle proposizioni. Così, ad esempio, i connettivi *oppure* (OR) ed *e* (AND) sono rappresentati rispettivamente dalle operazioni di *unione* ed *intersezione*:

$$\begin{aligned} A \cup B &= \{\omega : \omega \in A \text{ oppure } \omega \in B\} \\ A \cap B &= AB = \{\omega : \omega \in A \text{ e } \omega \in B\}. \end{aligned}$$

Il significato logico delle seguenti operazioni è facilmente deducibile<sup>2</sup> anche tenendo presenti i diagrammi della Fig. I.2.1:

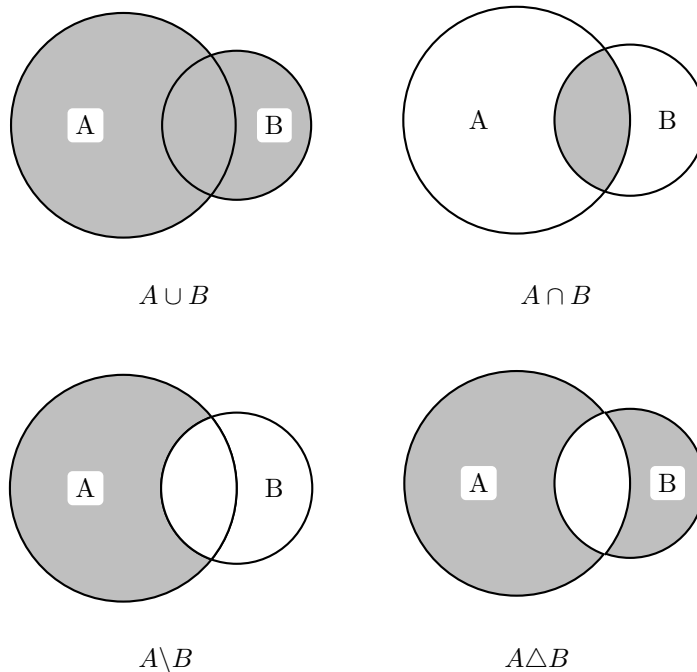
$$\begin{aligned} \overline{A} &= \{\omega : \omega \notin A\}; \\ A \setminus B &= A \cap \overline{B} = \{\omega : \omega \in A \text{ ma } \omega \notin B\}; \\ A \triangle B &= (A \setminus B) \cup (B \setminus A) \quad (\text{differenza simmetrica}). \end{aligned}$$

---

<sup>2</sup> Per dimostrare in modo elementare la validità delle leggi generali che governano le operazioni fra insiemi è consigliabile fare uso dei cosiddetti *diagrammi di Venn* (L. Koulikov: *Algèbre et Théorie des Nombres*; MIR, Moscou, 1982) qualche esempio dei quali è riportato in Fig. I.2.1. Il contributo dell'inglese John Venn (1834 - 1923) nel campo della logica è stato principalmente didattico, sistematizzatore ed orientato all'automatismo del calcolo logico. I suoi studi si inseriscono nell'ambito della scuola inglese della seconda metà del XIX secolo che faceva capo a G. Boole e J. Stuart Mill. Tra i suoi lavori ricordiamo *Logic of chance* (1866) e *Symbolic logic* (1881). L'aspetto più noto della sua opera è legato alla rappresentazione diagrammatica delle proposizioni e del ragionamento, ma va ricordato che prima di lui L. Euler e G.W. Leibniz avevano già utilizzato queste tecniche soprattutto per la loro efficacia didattica. Dopo Venn un altro significativo impiego dei diagrammi in logica lo si trova in opere (1896) di C.L. Dodgson, autore meglio noto con lo pseudonimo di Lewis Carroll.

Converrà anche notare che  $\Omega$  rappresenta l'evento *certo* (nel senso che l'evento  $\omega \in \Omega$  si verifica sempre),  $\emptyset$  rappresenta l'evento *impossibile* (dato che  $\omega \in \emptyset$  non si verifica mai) e che ovviamente, dalle definizioni precedenti, risulta  $\bar{A} = \Omega \setminus A$ . Diremo inoltre che i due eventi  $A$  e  $B$  sono **disgiunti** (o anche **incompatibili**) quando  $A \cap B = \emptyset$  (cioè quando un risultato  $\omega$  non può mai verificare contemporaneamente gli eventi  $A$  e  $B$ ). Naturalmente le proprietà delle operazioni fra insiemi riproducono anche le proprietà delle operazioni logiche, come ad esempio nel caso delle seguenti identità note anche come *leggi di de Morgan*<sup>3</sup>:

$$\overline{A \cap B} = \bar{A} \cup \bar{B}; \quad \overline{A \cup B} = \bar{A} \cap \bar{B}.$$



**Fig. I.2.1** Esempi di Diagrammi di Venn.

<sup>3</sup> Augustus de Morgan (1806 - 1871), logico e matematico inglese, diede degli importanti contributi allo sviluppo della teoria delle relazioni e, assieme a G. Boole, alla nascita della moderna logica formale. Egli fu uno dei primi a riconoscere la natura puramente simbolica dell'algebra e, quindi, a discutere la possibilità di algebre diverse da quella ordinaria: la sua interpretazione geometrica delle proprietà dei numeri complessi (in *Trigonometry and double algebra*, 1849) è, ad esempio, considerata come uno degli spunti che ha suggerito l'idea dei quaternioni. Le leggi logiche che portano il suo nome erano verbalmente note fin dai tempi di Ockham (XIV secolo), ma furono completamente esaminate e poste in forma matematica solo con la sua opera.

**I.2.7 Esempio:** Nel caso di due lanci di una moneta, a partire dagli eventi

$$A = \{TT, TC, CT\} = \text{appare } T \text{ almeno una volta}$$

$$B = \{TC, CT, CC\} = \text{appare } C \text{ almeno una volta}$$

è possibile formare i seguenti altri eventi:

$$A \cup B = \{TT, TC, CT, CC\} = \Omega,$$

$$A \cap B = \{TC, CT\},$$

$$A \setminus B = \{TT\}.$$

È consigliabile che il lettore costruisca qualche altro esempio per abituarsi a questo linguaggio.  $\diamond$

Dato un insieme finito  $\Omega$  ed una famiglia di suoi sottinsiemi (eventi) che contenga  $\Omega$  stesso, supponiamo di formare tutti i possibili eventi ottenibili mediante le operazioni  $\cup, \cap, \setminus$ : la famiglia (finita)  $\mathcal{A}$  così ottenuta è ovviamente *chiusa* per tutte le operazioni menzionate, nel senso che il risultato di una delle operazioni  $\cup, \cap, \setminus$  su due elementi di  $\mathcal{A}$  produce ancora un elemento di  $\mathcal{A}$ . Famiglie di eventi di questo tipo sono particolarmente importanti per il calcolo delle probabilità perché in esse le operazioni logiche risultano sempre possibili nel senso che producono risultati che appartengono alla stessa famiglia.

**I.2.8 Definizione:** Chiameremo **algebra** una famiglia  $\mathcal{A}$  di parti di  $\Omega$  tale che

- a.  $\Omega \in \mathcal{A}$ ;
- b.  $A \cup B, A \cap B, A \setminus B$  sono tutti elementi di  $\mathcal{A}$  se  $A$  e  $B$  lo sono.  $\triangle$

**I.2.9 Esempio:** Dato un insieme  $\Omega$  le seguenti famiglie di insiemi

$$\begin{aligned} \mathcal{A}_* &= \{\Omega, \emptyset\}, && (\text{algebra banale}) \\ \mathcal{A}_A &= \{A, \overline{A}, \Omega, \emptyset\}, && (\text{algebra generata da } A \subseteq \Omega) \\ \mathcal{A}^* &= \wp(\Omega), && (\text{algebra di tutte le parti di } \Omega) \end{aligned}$$

sono tutte algebre di parti di  $\Omega$ .  $\diamond$

**I.2.10 Definizione:** Diremo che una famiglia di sottinsiemi  $\mathcal{D} = \{D_1, \dots, D_n\}$  è una **decomposizione** di  $\Omega$  in **atomi**  $D_k$ , se le  $D_k$  sono parti di  $\Omega$  non vuote, disgiunte e tali che  $\bigcup_k D_k = \Omega$ .  $\triangle$

È abbastanza evidente che una decomposizione *non* è un'algebra: una decomposizione, ad esempio, non contiene le unioni dei suoi atomi e pertanto non è chiusa rispetto all'operazione di unione. Ciononostante, a partire da una data decomposizione  $\mathcal{D}$  è sempre possibile costruire un'algebra  $\mathcal{A}$  formando tutte le possibili unioni di elementi di  $\mathcal{D}$ ; tale algebra  $\mathcal{A} = \alpha(\mathcal{D})$  si chiama **algebra generata da**

$\mathcal{D}$ . Si verifica facilmente, nel caso di insiemi finiti  $\Omega$ , che data una generica algebra  $\mathcal{A}$  di parti di  $\Omega$  esiste sempre un'unica decomposizione  $\mathcal{D}$  che genera  $\mathcal{A}$ :

$$\mathcal{D} = \{D \in \mathcal{A} : D \cap A = \emptyset \text{ oppure } D \cap A = D, \forall A \in \mathcal{A}\}.$$

**I.2.11 Esempio:** Un dato insieme  $\Omega$  ammette diverse decomposizioni; così ad esempio  $\Omega = \{1, 2, 3\}$  ammette, tra le altre,

$$\begin{aligned} D_1 &= \{1\}, & D_2 &= \{2, 3\}; \\ D_1 &= \{1\}, & D_2 &= \{2\}, & D_3 &= \{3\}. \end{aligned}$$

Viceversa, data un'algebra di parti di un insieme  $\Omega$  di  $N$  elementi, la decomposizione che la genera è unica; le seguenti:

$$\begin{aligned} D_1 &= \Omega; \\ D_1 &= A, & D_2 &= \overline{A}; \\ D_k &= \{\omega_k\}, & k &= 1, \dots, N. \end{aligned}$$

sono le decomposizioni che generano le algebre dell'Esempio I.2.9. ◇

**I.2.12 Definizione:** Date due decomposizioni  $\mathcal{D}_1$  e  $\mathcal{D}_2$  diremo che  $\mathcal{D}_2$  è **più fine** di  $\mathcal{D}_1$  (e scriveremo  $\mathcal{D}_1 \preceq \mathcal{D}_2$ ) quando  $\alpha(\mathcal{D}_1) \subseteq \alpha(\mathcal{D}_2)$ . △

**I.2.13 Osservazione:** Nel caso di spazi finiti  $\Omega$  con  $N$  elementi si usa in generale assumere come algebra degli eventi l'algebra più ricca  $\mathcal{A} = \mathcal{A}^* = \wp(\Omega)$ . In questo caso si mostra che l'algebra degli eventi è composta di  $2^N$  elementi. Per provarlo basterà osservare che ogni evento  $A = \{\omega_{i_1}, \dots, \omega_{i_k}\}$  di  $\mathcal{A}$  può essere rappresentato con una  $N$ -pla di cifre eguali a 0 oppure 1:

$$\begin{pmatrix} 0, & 0, & \dots, & 1, & \dots, & 0, & \dots, & 1, & \dots, & 0 \end{pmatrix}$$

$$\begin{matrix} 1 & 2 & \dots & i_1 & \dots & i & \dots & i_k & \dots & N \end{matrix}$$

che indicano la presenza (o l'assenza) dell'elemento  $\omega_i$  nell'evento  $A$ . Il numero di eventi distinti con  $k$  elementi coincide quindi con il numero delle combinazioni di  $k$  oggetti su  $N$  posti<sup>4</sup> (vedi Esempio I.2.4). Pertanto risulta che il numero degli eventi di  $\mathcal{A}^*$  è

$$\sum_{k=0}^N \binom{N}{k} = (1+1)^N = 2^N. \quad \bigcirc$$

---

<sup>4</sup> Ogni evento di  $k$  elementi può essere pensato come l'estrazione (senza rimessa) di  $k$  palline numerate da una scatola contenente  $N$  palline con successiva registrazione degli indici estratti mediante le cifre 1 nei posti corrispondenti della  $N$ -pla che rappresenta l'evento. Ovviamente l'ordine dell'estrazione è irrilevante.

Una volta definito uno spazio (finito) dei campioni ed un'algebra degli eventi è anche possibile assegnare una **probabilità** ad ogni evento elementare  $\omega_k$ , ossia un insieme di *pesi*  $p(\omega_k)$  tali che

$$\begin{aligned} a) \quad & 0 \leq p(\omega_k) \leq 1; \quad k = 1, \dots, N \\ b) \quad & \sum_{k=1}^N p(\omega_k) = 1. \end{aligned}$$

Inoltre si associa ad ogni evento (non elementare)  $A \in \mathcal{A}$  una probabilità definita da:

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

Sulla base di queste notazioni possiamo ora dare la seguente definizione:

**I.2.14 Definizione:** Chiameremo **spazio (o modello) finito di probabilità** la terna  $(\Omega, \mathcal{A}, \mathbf{P})$  dove  $\Omega$  è un insieme (finito),  $\mathcal{A}$  è un'algebra di eventi di  $\Omega$  e  $\mathbf{P} : \mathcal{A} \rightarrow [0, 1]$  è una funzione definita secondo lo schema precedente.  $\triangle$

**I.2.15 Osservazione:** È facile provare che le seguenti proprietà:

$$\begin{aligned} \mathbf{P}(\emptyset) &= 0 \\ \mathbf{P}(\Omega) &= 1 \\ \mathbf{P}(A \cup B) &= \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B) \\ \mathbf{P}(A \cup B) &= \mathbf{P}(A) + \mathbf{P}(B), \quad \text{se } A \cap B = \emptyset \quad (\text{additività}) \\ \mathbf{P}\left(\bigcup_{k=1}^n A_k\right) &\leq \sum_{k=1}^n \mathbf{P}(A_k), \quad \text{se } A_k \in \mathcal{A}, \quad k = 1, \dots, n \quad (\text{sub-additività}) \\ \mathbf{P}(\overline{A}) &= 1 - \mathbf{P}(A) \\ \mathbf{P}(A \Delta B) &= \mathbf{P}(A) + \mathbf{P}(B) - 2\mathbf{P}(A \cap B) \\ \mathbf{P}(A) &\leq \mathbf{P}(B), \quad \text{se } A \subseteq B \end{aligned}$$

sono sempre vere in uno spazio di probabilità. Ad esempio l'ultima proprietà si prova osservando che, se  $A \subseteq B$ , risulta  $B = A \cup (B \cap \overline{A})$  dove  $A \cap (B \cap \overline{A}) = \emptyset$  sicché dalla proprietà di additività si ha che  $\mathbf{P}(B) = \mathbf{P}(A) + \mathbf{P}(B \cap \overline{A}) \geq \mathbf{P}(A)$ . Va invece osservato che in generale risulta

$$\mathbf{P}(A \cap B) \neq \mathbf{P}(A)\mathbf{P}(B).$$

Avremo modo di discutere a lungo nel seguito le condizioni (*indipendenza* di  $A$  e  $B$ ) sotto le quali la relazione  $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$  è verificata. Sarà bene, infine, prendere familiarità con questo tipo di proprietà tentando di elaborarne delle semplici dimostrazioni.  $\circ$

**I.2.16 Osservazione:** L'assegnazione dei pesi probabilistici  $p(\omega_k)$  agli eventi elementari non è sempre un'impresa facile se si vogliono costruire dei modelli realistici.



Eppure si tratta del primo passo da compiere se il problema del calcolo delle probabilità è quello di calcolare le probabilità di eventi complicati dalla conoscenza delle probabilità degli eventi elementari. Le tecniche per la ricostruzione delle probabilità a partire dalle osservazioni empiriche sono argomento della **statistica** e noi non potremo che far cenno solo di sfuggita a questi problemi. Per il seguito, invece, adotteremo un altro modo di determinare le  $p(\omega_k)$  basato sulla **definizione classica**<sup>5</sup> della probabilità che è stata brevemente discussa all'inizio del capitolo I.1. Il metodo consiste nel ricondursi ad uno spazio dei campioni i cui elementi (per ragioni di simmetria) possano essere considerati **equiprobabili**. In tal caso ad ogni evento elementare si associa la probabilità  $p(\omega_k) = 1/N, k = 1, \dots, N$  e ad ogni evento  $A \in \mathcal{A}$  la probabilità

$$\mathbf{P}(A) = \frac{N(A)}{N}$$

dove  $N(A)$  indica il numero di  $\omega_k$  contenuti nell' evento  $A$ . ○

**I.2.17 Esempio (Problema delle coincidenze):** Supponiamo di estrarre *con rimessa* da una scatola contenente  $M$  palline numerate una successione di  $n$  palline e di registrare i numeri estratti tenendo conto dell'ordine di estrazione. Sappiamo, dall'Esempio I.2.3 che il nostro spazio  $\Omega$  sarà formato da  $N = M^n$  eventi elementari  $\omega = (a_1, \dots, a_n)$  che considereremo equiprobabili. Consideriamo ora l'evento:

$$A = \{\omega : \text{i valori delle } a_k \text{ sono tutti diversi}\}$$

e calcoliamone la probabilità secondo la definizione classica. È facile accorgersi, per enumerazione, che il numero  $N(A)$  degli eventi elementari contenuti in  $A$  è

$$N(A) = M(M-1) \dots (M-n+1) = (M)_n = \frac{M!}{(M-n)!}$$

e che pertanto la probabilità richiesta è

$$\mathbf{P}(A) = \frac{(M)_n}{M^n} = \left(1 - \frac{1}{M}\right) \left(1 - \frac{2}{M}\right) \dots \left(1 - \frac{n-1}{M}\right).$$

Questo risultato permette di discutere il cosiddetto **problema dei compleanni**: date  $n$  persone quale è la probabilità  $P_n$  che almeno due di esse celebrino il compleanno nello stesso giorno? Si tratta chiaramente di un problema di conteggi di campioni ordinati dato che una diversa disposizione delle stesse date sulle  $n$  persone deve essere ritenuta un'eventualità distinta. Il modello discusso in questo esempio ci permette pertanto di dare una risposta se si pone  $M = 365$ ; in tal

---

<sup>5</sup> Vedi a questo proposito il celebre testo di **P.-S. Laplace**: *Théorie analytique des probabilités* (1812), attualmente riedito da J. Gabay, Paris, 1995.

caso infatti, essendo  $\mathbf{P}(A)$  la probabilità che *tutti* i compleanni cadano in giorni differenti, si ha che

$$P_n = 1 - \mathbf{P}(A) = 1 - \frac{(365)_n}{365^n}.$$

In particolare si ottengono i seguenti sorprendenti risultati:

$$\begin{array}{cccccccc} n & = & 4 & & 16 & & 22 & & 23 & & 40 & & 64 & & \dots \\ P_n & = & 0,016 & & 0,284 & & 0,476 & & 0,507 & & 0,891 & & 0,997 & & \dots \end{array}$$

È notevole infatti che già con  $n = 23$  la probabilità di almeno due compleanni coincidenti supera  $1/2$  e che con solo 64 persone tale probabilità sfiora la certezza. Naturalmente se  $n \geq 366$  si ha  $P_n = 1$  e  $\mathbf{P}(A) = 0$  dato che nel prodotto comparirà un fattore nullo. Osserviamo inoltre che molta della sorpresa per questi risultati sparisce se si comprende che i risultati sarebbero diversi se la domanda posta fosse stata la seguente: supponendo che io sia una delle  $n$  persone considerate nel problema precedente, quale è la probabilità  $P'_n$  che almeno una celebri il suo compleanno nello stesso giorno in cui lo celebri io? Infatti in questo caso dobbiamo ragionare come nel problema del Cavalier de Méré discusso nel capitolo precedente: dato che per ognuna delle altre  $n - 1$  persone la probabilità di *non* celebrare il suo compleanno nello stesso giorno mio è  $364/365$  (e supposto che le date dei compleanni delle singole persone sono indipendenti) la probabilità cercata è

$$P'_n = 1 - \left(\frac{364}{365}\right)^{n-1}$$

sicché questa volta avremmo

$$\begin{array}{cccccccc} n & = & 4 & & 16 & & 22 & & 23 & & 40 & & 64 & & 101 & & \dots \\ P'_n & = & 0,011 & & 0,040 & & 0,056 & & 0,059 & & 0,101 & & 0,159 & & 0,240 & & \dots \end{array}$$

Inoltre in questo caso  $P'_n$  è sempre diversa da 1 (anche se  $n \geq 366$ ) in quanto con qualunque numero di persone può sempre capitare che nessuno celebri il suo compleanno nello stesso giorno in cui lo celebri io. Quest'ultima osservazione mette bene in luce la differenza fra i due problemi e sottolinea che, nell'eventualità di una scommessa, è sempre bene capire chiaramente su quale eventualità si sta puntando il proprio denaro.  $\diamond$

## I.3 Distribuzione binomiale

In questo capitolo introdurremo un modello che, nonostante la sua semplicità, riveste una considerevole importanza nella Teoria delle Probabilità. Consideriamo l'esperimento consistente in  $n$  lanci successivi di una moneta (in generale *non equa*) o equivalentemente in  $n$  successivi tentativi di verifica di un particolare evento aleatorio e poniamoci il problema di studiare gli eventi del tipo *su  $n$  lanci di moneta l'evento  $T$  (testa) si verifica  $k$  volte* (più in generale, *su  $n$  tentativi un particolare evento aleatorio si verifica  $k$  volte*, o anche *su  $n$  tentativi si registrano  $k$  successi*). Per impostare correttamente questo problema non potremo però limitarci a prendere in esame lo spazio dei campioni corrispondente alla singola verifica del nostro evento (o al singolo lancio di moneta), ma dovremo costruire uno spazio che descriva le  $n$ -ple di tentativi. A questo fine converremo, per comodità, di codificare i risultati di ogni singolo tentativo con un numero: il numero 1 indicherà che l'evento  $T$ , o più in generale l'evento aleatorio in esame, si è verificato (*successo*), mentre il numero 0 indicherà l'eventualità opposta (*fallimento*). In questo modo il risultato di ogni  $n$ -pla di tentativi si presenterà come una  $n$ -pla ordinata di numeri tutti eguali a 0 oppure 1:  $\omega = (a_1, \dots, a_n)$  e quindi lo spazio dei campioni sarà

$$\Omega = \{\omega = (a_1 \dots a_n) : a_i = 0, 1; i = 1, \dots, n\}.$$

Considereremo inoltre come famiglia degli eventi l'algebra più ricca  $\mathcal{A}^* = \wp(\Omega)$  (vedi Esempio I.2.9). Notando ora che  $\sum_{i=1}^n a_i = k$  risulta essere proprio il numero di volte in cui si verifica l'evento  $T$  su  $n$  lanci di moneta, definiremo la probabilità di ogni evento elementare  $\omega = (a_1, \dots, a_n)$  nel modo seguente:

$$\mathbf{P}\{\omega\} = p(\omega) = p^k q^{n-k},$$

dove  $p$  e  $q$  sono due numeri reali e positivi tali che  $p + q = 1$ .

**I.3.1 Osservazione:** Il vero significato di questa assegnazione di probabilità sarà chiaro più oltre. Per il momento ci contenteremo di notare che, se (come vedremo tra breve)  $p$  è interpretato come la probabilità che l'evento  $T$  si verifichi in un solo lancio di moneta (e corrispondentemente  $q$  è la probabilità che si verifichi  $C$ ), il numero  $p(\omega)$  consiste nel prodotto di tante  $p$  e  $q$  quante volte si verificano rispettivamente  $T$  (cioè  $k$ ) e  $C$  (cioè  $n-k$ ). Vedremo in seguito che questa maniera di calcolare le probabilità di eventi composti a partire dalle probabilità di eventi elementari è tipica del caso in cui gli eventi elementari sono *indipendenti*, ma non è applicabile in generale (vedi anche Osservazione I.2.15). Pertanto l'assegnazione dei pesi statistici da noi effettuata va considerata come una scelta arbitraria fra le tante possibili motivata, soltanto *a posteriori*, dall'intenzione di costruire un modello che descriva un esperimento consistente in  $n$  ripetizioni *indipendenti* di una verifica elementare (come un lancio di moneta).  $\circ$

La definizione delle probabilità elementari ci consente poi di definire anche la probabilità di tutti gli altri eventi  $A \in \mathcal{A}^*$  come

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

in modo che  $(\Omega, \mathcal{A}^*, \mathbf{P})$  resti definito come un possibile modello finito di probabilità per il nostro esperimento.

**I.3.2 Osservazione:** Va ricordato che l'assegnazione della probabilità  $\mathbf{P}(\cdot)$  non può essere completamente arbitraria: affinché  $(\Omega, \mathcal{A}^*, \mathbf{P})$  sia realmente uno spazio di probabilità  $\mathbf{P}(\cdot)$  deve soddisfare alcune proprietà elencate in Definizione I.2.14 ed in Osservazione I.2.15. Noi qui ci limiteremo solo a verificare che  $\mathbf{P}(\Omega) = 1$ , lasciando come esercizio la verifica (immediata) delle altre. Per far questo osserviamo che lo spazio  $\Omega$  può essere decomposto negli  $n+1$  atomi  $D_k$ ,  $k = 0, \dots, n$  di una decomposizione  $\mathcal{D}$  (vedi Definizione I.2.10) così definiti:

$$D_k = \left\{ \omega \in \Omega : \sum_{i=1}^n a_i = k \right\}.$$

È chiaro da questa definizione che  $D_k$  corrisponde all'evento *su  $n$  tentativi si registrano  $k$  successi*. Siccome inoltre i  $k$  successi possono collocarsi in vario modo fra gli  $n$  tentativi effettuati, ogni  $D_k$  sarà costituito da un certo numero  $n_k$  di eventi elementari a ciascuno dei quali la nostra definizione di  $\mathbf{P}(\cdot)$  assegna la medesima probabilità  $p^k q^{n-k}$ . Pertanto risulterà

$$\mathbf{P}(D_k) = \sum_{\omega \in D_k} p(\omega) = n_k p^k q^{n-k}.$$

Essendo  $\mathcal{D}$  una decomposizione di  $\Omega$  si ha anche che i  $D_k$  sono tutti disgiunti e  $\bigcup_{k=0}^n D_k = \Omega$ . Segue allora dalla proprietà di additività di  $\mathbf{P}(\cdot)$  (vedi Osservazione I.2.15) che

$$\mathbf{P}(\Omega) = \mathbf{P}\left(\bigcup_{k=0}^n D_k\right) = \sum_{k=0}^n \mathbf{P}(D_k) = \sum_{k=0}^n n_k p^k q^{n-k}.$$

Per terminare la nostra verifica dobbiamo quindi solo determinare il valore delle  $n_k$ . Notiamo a questo scopo che ogni risultato  $\omega = (a_1 \dots a_n)$  con  $k = \sum_{i=1}^n a_i$  fissato è univocamente individuato tramite un insieme di *numeri d'occupazione*  $[b_1, \dots, b_k]$  che indicano la posizione dei  $k$  numeri "1" sugli  $n$  posti di  $\omega$ ; ad esempio (con  $n = 7$ ):

$$\omega = (0, 1, 1, 0, 0, 1, 0) \leftrightarrow [2, 3, 6].$$

È inoltre chiaro che per i nostri calcoli l'ordine delle  $b_j$  in  $[b_1, \dots, b_k]$  è perfettamente irrilevante ( $[2, 3, 6]$  e  $[3, 6, 2]$  indicheranno nel nostro esempio la medesima

7-pla con il numero “1” al secondo, terzo e sesto posto) e che i valori delle  $b_j$  sono tutti diversi (dato che non avrebbe senso indicare due volte che lo stesso posto  $b_j$  è occupato da un “1”). In conclusione potremo dire che il numero  $n_k$  di  $\omega$  appartenenti a  $D_k$  coincide con il numero delle possibili  $k$ -ple non ordinate e senza ripetizione  $[b_1, \dots, b_k]$  dove ogni  $b_j$  può assumere gli  $n$  valori  $1, 2, \dots, n$ . Come discusso nell’Esempio I.2.4 tale numero è  $\binom{n}{k}$ , sicché risulterà (tenendo anche conto della formula del binomio di Newton):

$$\mathbf{P}(\Omega) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = (p + q)^n = 1,$$

che è quanto volevamo provare.  $\circ$

**I.3.3 Esempio:** In particolare, nel caso in cui  $n = 1$ , cioè quando si prende in considerazione un solo tentativo di verifica (un solo lancio di moneta), si ha  $\Omega = \{0, 1\}$  mentre l’algebra degli eventi si riduce a

$$\mathcal{A}^* = \{\{1\}, \{0\}, \{1, 0\}, \emptyset\},$$

e l’assegnazione di probabilità è  $\mathbf{P}\{1\} = p(1) = p$ ,  $\mathbf{P}\{0\} = p(0) = q$ . Da queste osservazioni si ricava subito che  $p$  assume il significato di probabilità di successo in un tentativo isolato come annunciato nell’Osservazione I.3.1.  $\diamond$

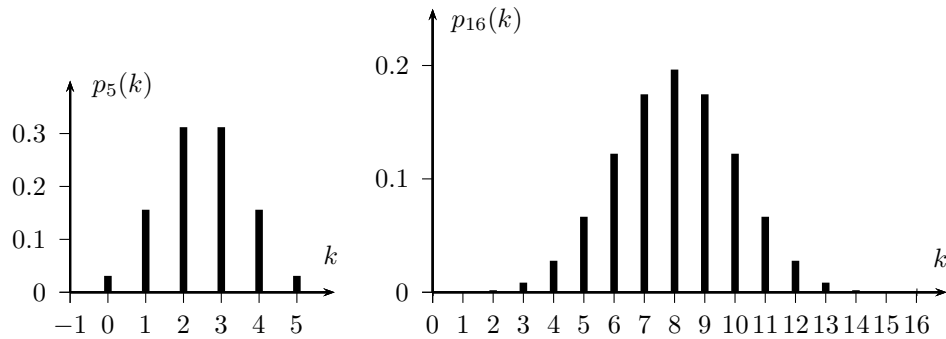


Fig. I.3.1 Esempî di distribuzioni binomiali ( $p = q = 1/2$ ).

Riprendendo la notazione dell’Osservazione I.3.2, notiamo ora che l’insieme dei numeri  $p_n(k)$ ,  $k = 0, 1, \dots, n$ , con

$$p_n(k) = \mathbf{P}(D_k) = \binom{n}{k} p^k q^{n-k}$$

rappresenta l’insieme delle probabilità che nel nostro modello competono agli eventi della decomposizione  $\mathcal{D}$ , cioè agli eventi del tipo *su  $n$  tentativi si registrano  $k$  successi*. Tale insieme di numeri  $\{p_n(k)\}_{k=0, \dots, n}$  prende il nome di **Distribuzione**

**Binomiale** (per  $n$  tentativi indipendenti di verifica di un dato evento). In Figura I.3.1 riportiamo l'andamento grafico di tale distribuzione nel caso  $p = q = \frac{1}{2}$  (moneta *equa*) per diversi valori di  $n$ . Nota anche che il valore massimo di  $p_n(k)$  si ottiene in  $k = n/2$  se  $n$  è pari, e nei due punti  $k = (n + 1)/2$  e  $k = (n - 1)/2$  se  $n$  è dispari.

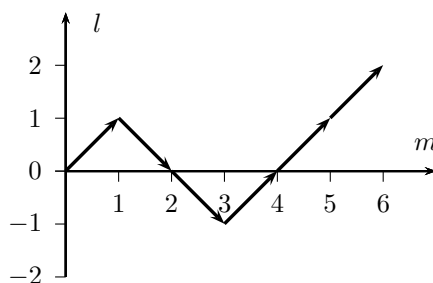


Fig. I.3.2 Esempio di *random walk*

**I.3.4 Esempio:** La distribuzione binomiale, nonostante la sua semplicità, è largamente usata in un gran numero di modelli, anche di interesse pratico. In questo esempio descriveremo brevemente alcuni aspetti di un caso estremamente semplificato di **passeggiata aleatoria** (*random walk*): un primo, elementare esempio di una vasta classe di importanti modelli probabilistici che vanno sotto il nome di *processi stocastici*. Supponiamo di prendere in considerazione una particella puntiforme che si muove lungo una retta partendo dall'origine ed eseguendo, dopo ogni unità di tempo, un passo di lunghezza unitaria in avanti o all'indietro (vedi Figura I.3.2). Se prendiamo in esame il moto della particella nell'intervallo di tempo  $[0, n]$ , gli istanti in cui essa effettua i suoi spostamenti sono  $i = 1, \dots, n$ , e ogni possibile traiettoria  $\omega \in \Omega$  sarà completamente specificata da  $n$  numeri  $b_i = \pm 1; i = 1, \dots, n$  che rappresentano lo spostamento in avanti (+1) o all'indietro (-1) del nostro punto materiale. Gli eventi elementari del nostro spazio  $\Omega$  saranno dunque le  $n$ -uple ordinate  $\omega = (b_1, \dots, b_n)$  di numeri che possono valere +1 o -1. Ci poniamo allora il problema di valutare la probabilità di eventi del tipo *dopo  $n$  istanti di tempo la posizione della particella è  $l$*  (dove  $l$  è un numero intero che varia fra  $-n$  ed  $n$ ). Per far questo osserviamo che per un dato  $\omega$  la posizione finale del punto si ottiene sommando algebricamente tutti gli spostamenti successivi, cioè calcolando il numero  $l = \sum_{i=1}^n b_i$  che ovviamente può assumere valori fra  $-n$  ed  $n$ . Inoltre, siccome i numeri  $a_i = \frac{1}{2}(b_i + 1)$  assumeranno valore +1 se  $b_i = +1$  e valore 0 se  $b_i = -1$ , il numero

$$k = \sum_{i=1}^n a_i = \sum_{i=1}^n \frac{b_i + 1}{2} = \frac{l + n}{2}$$

assumerà valori tra 0 ed  $n$  e rappresenterà il numero di “+1” presenti in  $\omega$ . Si verifica inoltre facilmente che  $l$  risulta pari se  $n$  è pari, mentre risulta dispari se  $n$  è dispari: ne consegue che  $l + n$  è sempre un numero pari e quindi che  $k$  è sempre

intero. Una traiettoria può quindi essere equivalentemente individuata tramite la  $n$ -pla ordinata  $(a_1, \dots, a_n)$  di numeri eguali a 0 oppure 1 come nel caso discusso all'inizio di questo capitolo. Assegnati allora due numeri positivi  $p$  e  $q$  ( $p + q = 1$ ), associeremo ad ogni traiettoria la probabilità

$$p(\omega) = p^k q^{n-k},$$

ricordando che (come nel caso della distribuzione binomiale, vedi Osservazione I.3.1) questa attribuzione è arbitraria e corrisponde all'ipotesi sottintesa che lungo la traiettoria i passi vengano effettuati in maniera indipendente uno dall'altro. L'evento *la posizione della particella è  $l$  dopo  $n$  istanti di tempo* è descritto dunque nel nostro modello dal seguente insieme di traiettorie:

$$B_l = \left\{ (b_1, \dots, b_n) : \sum_{i=1}^n b_i = l \right\},$$

o dal corrispondente insieme  $D_k$  con  $k = (l + n)/2$  di  $n$ -ple  $(a_1, \dots, a_n)$  definito in I.3.2; ad ogni traiettoria di tali insiemi viene attribuito il medesimo peso probabilistico

$$p^k q^{n-k} = p^{\frac{n+l}{2}} q^{\frac{n-l}{2}},$$

e pertanto, per valutare  $\mathbf{P}(B_l)$ , sarà sufficiente calcolare il numero di traiettorie  $\omega$  contenute in  $B_l$ . Ma il numero di queste traiettorie è pari al numero delle  $n$ -ple  $(a_1, \dots, a_n)$  in cui "1" compare  $k = \frac{1}{2}(n+l)$  volte, e queste sono (vedi Osservazione I.3.2)

$$\binom{n}{k} = \binom{n}{\frac{n+l}{2}}.$$

Pertanto risulta  $\mathbf{P}(B_l) = \mathbf{P}(D_k) = p_n(k)$  per  $k = (n+l)/2$  per cui dalle discussioni precedenti si ha:

$$\mathbf{P}(B_l) = p_n \left( \frac{n+l}{2} \right) = \binom{n}{\frac{n+l}{2}} p^{\frac{n+l}{2}} q^{\frac{n-l}{2}},$$

per  $l$  con la medesima parità di  $n$ , e  $\mathbf{P}(B_l) = 0$  per  $l$  con parità diversa da  $n$ . Inoltre i numeri  $p$  e  $q$  rappresentano le probabilità che in un singolo salto la particella si muova rispettivamente in avanti o all'indietro lungo la retta. In particolare, nel caso simmetrico  $p = q = \frac{1}{2}$ , risulta

$$\mathbf{P}(B_l) = p_n \left( \frac{n+l}{2} \right) = \binom{n}{\frac{n+l}{2}} 2^{-n}.$$

Inoltre, supponendo per semplicità  $n = 2m$ , abbiamo osservato nell'Esempio I.3.3 che  $p_n(k)$  assume il valore massimo in  $k = m$ , sicché, essendo anche  $k = (n+l)/2 = m+l/2$ , il valore massimo di  $\mathbf{P}(B_l)$  viene assunto per  $l = 0$ . Può essere interessante allora stimare tali probabilità al crescere di  $n$ ; ad esempio, per semplificare i conti,

potremmo stimare  $\mathbf{P}(B_0)$  (la probabilità che la particella resti nell'origine) nel caso in cui il numero di istanti trascorsi sia  $2n$  (con un numero dispari di passi la particella con certezza non è in 0). In tal caso

$$\mathbf{P}(B_0) = p_{2n}(n) = \binom{2n}{n} 2^{-2n},$$

che per grandi valori di  $n$  può essere stimata tramite la formula di Stirling<sup>1</sup> e vale

$$\mathbf{P}(B_0) \sim \frac{1}{\sqrt{\pi n}}, \quad (n \rightarrow \infty).$$

Sarà bene osservare che l'evoluzione di  $\mathbf{P}(B_k)$  al crescere di  $n$  costituisce un primo semplice esempio di fenomeno di *diffusione*.  $\diamond$

Una generalizzazione della distribuzione binomiale è costituita dalla cosiddetta **Distribuzione Multinomiale** che corrisponde ad un analogo problema di *campionamento con rimessa* (vedi Esempio I.2.3), cioè ad un problema in cui gli eventi elementari sono  $n$ -ple ordinate  $\omega = (a_1, \dots, a_n)$  di numeri che non sono tutti diversi fra loro. La generalizzazione consiste nel fatto che ora tali numeri non prendono solo due valori (0 e 1, oppure  $\pm 1$ ) ma gli  $r$  ( $\geq 2$ ) valori distinti  $b_1, \dots, b_r$  (ad esempio  $1, 2, \dots, r$ , ma tali valori non sono necessariamente numerici). Se indichiamo con  $n_i$  il numero di elementi di  $\omega$  che assumono il valore  $b_i$ , assegneremo a tale evento elementare la probabilità

$$\mathbf{P}\{\omega\} = p(\omega) = p_1^{n_1} \cdot \dots \cdot p_r^{n_r}; \quad (p_1 + \dots + p_r = 1).$$

Si può provare che, dati  $n_1, \dots, n_r$ , ci sono

$$\binom{n}{n_1, \dots, n_r} = \frac{n!}{n_1! \dots n_r!}$$

---

<sup>1</sup> La cosiddetta *Formula di Stirling* afferma che, per ogni numero intero  $n$ , è sempre possibile trovare un numero  $\theta_n \in (0, 1)$  tale che

$$n! = \sqrt{2\pi n} \left(\frac{n}{e}\right)^n e^{\theta_n/12n},$$

sicché si può anche affermare che  $n! \sim \sqrt{2\pi n} n^n e^{-n}$ , ( $n \rightarrow \infty$ ). Ricordiamo qui una volta per tutte che l'espressione  $f(n) \sim g(n)$ , ( $n \rightarrow \infty$ ), indica che

$$\lim_n \frac{f(n)}{g(n)} = 1.$$

James Stirling (1692 - 1770), detto il *veneziano*, fu un matematico scozzese che prese parte attiva al dibattito dei circoli scientifici della Londra della prima metà del '700 (la Royal Society, Newton, Taylor, de Moivre ecc.). Egli è accreditato della scoperta della formula che porta il suo nome, ma tutti i libri di storia della matematica (vedi ad esempio **D. E. Smith: A Source Book in Mathematics**; Dover, New York, 1959) ne attribuiscono la prima idea a de Moivre che ne parla in un suo celebre lavoro del 1733 sull'approssimazione Gaussiana della distribuzione binomiale.



possibili  $\omega = (a_1, \dots, a_n)$  con i dati valori di  $n_1, \dots, n_r$ . Pertanto, detto  $D_{n_1 \dots n_r}$  l'evento *il valore  $b_1$  è assunto  $n_1$  volte, ..., il valore  $b_r$  è assunto  $n_r$  volte*, si ha che

$$\mathbf{P}(D_{n_1 \dots n_r}) = \binom{n}{n_1, \dots, n_r} p_1^{n_1} \cdot \dots \cdot p_r^{n_r},$$

e l'insieme di tali probabilità prende il nome di distribuzione multinomiale.

**I.3.5 Esempio:** La distribuzione multinomiale viene usata per descrivere  $n$  ripetizioni (*indipendenti*) di un esperimento elementare il risultato del quale, però, non consiste solo nella verifica di un evento dato (*successo* o *fallimento*, 0 oppure 1), ma ha  $r$  valori possibili. Il problema che ci si pone è la valutazione della probabilità di eventi del tipo *il primo valore è stato trovato  $n_1$  volte, ..., l' $r$ -mo valore è stato trovato  $n_r$  volte*. La risposta a questo problema è appunto nella distribuzione multinomiale. Ad esempio supponiamo di voler misurare una data lunghezza e di essere interessati a sapere in quale dei seguenti tre intervalli (espressi in cm) cade il valore di tale misura:

$$[0, 10), \quad [10, 12), \quad [12, +\infty).$$

Se supponiamo di sapere che le probabilità che il risultato di una singola misura cada nei suddetti intervalli sono

$$p_1 = 0,25; \quad p_2 = 0,65; \quad p_3 = 0,10;$$

la probabilità che su  $n = 10$  risultati di misure ve ne siano  $n_1 = 5$  minori di 10 cm,  $n_2 = 3$  fra 10 e 12 cm ed  $n_3 = 2$  maggiori di 12 cm è data da

$$\frac{10!}{5!3!2!} (0,25)^5 (0,65)^3 (0,10)^2 = 0,00676.$$

Qui ovviamente  $r = 3$  mentre  $b_1, b_2, b_3$  corrispondono ai risultati (non numerici): *il risultato della misura è minore di 10 cm, ..., e così via.* ◇



## I.4 Condizionamento ed Indipendenza

Abbiamo già avuto modo, nei capitoli precedenti, di accennare all'indipendenza di eventi: è ora arrivato il momento di introdurre questo concetto in maniera un po' più sistematica anche se ancora molto elementare. Il primo problema da affrontare è quello del modo in cui una certa quantità di nuova informazione può essere amalgamata con l'informazione già immagazzinata in un dato modello di probabilità  $(\Omega, \mathcal{A}, \mathbf{P})$ . L'idea è che l'acquisizione di nuova informazione modifica le nostre conoscenze e quindi ci mette in condizione di valutare la probabilità degli eventi considerati in una maniera diversa da quella consentita dalle nostre informazioni iniziali. Un breve esempio ci permetterà di essere più chiari.

**I.4.1 Esempio:** Supponiamo di considerare una scatola contenente  $M$  palline delle quali  $m$  sono bianche ed  $M - m$  nere ed eseguiamo due estrazioni successive. Se le palline sono estratte tutte con la medesima probabilità, e se la prima estrazione è effettuata *con rimessa*, è facile convincersi del fatto che l'evento

$B =$  viene estratta una pallina bianca alla seconda estrazione

si verifica con una probabilità  $\mathbf{P}(B) = \frac{m}{M}$ . Diversa sarebbe, però, la nostra valutazione se la prima estrazione venisse effettuata *senza rimessa* e se sapessimo, ad esempio, che la prima pallina estratta è bianca. Infatti in questo caso l'eventualità che venga estratta una pallina bianca *sapendo che* in precedenza ne è stata estratta un'altra bianca si vedrebbe attribuire una probabilità pari ad  $\frac{m-1}{M-1}$ . Se invece in precedenza fosse stata estratta una pallina nera si avrebbe  $\mathbf{P}(B) = \frac{m}{M-1}$ .  $\diamond$

**I.4.2 Definizione:** Dato uno spazio di probabilità  $(\Omega, \mathcal{A}, \mathbf{P})$  e due eventi  $A, B \in \mathcal{A}$  tali che sia verificata la relazione  $\mathbf{P}(A) > 0$  ( $\neq 0$ ), chiameremo la quantità

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

**probabilità condizionata** di  $B$  rispetto ad  $A$  (*probabilità che si verifichi  $B$  sapendo che si è verificato  $A$* ). La quantità  $\mathbf{P}(A B)$  prende invece il nome di **probabilità congiunta** dei due eventi  $A$  e  $B$  (*probabilità che si verifichino contemporaneamente  $A$  e  $B$* ).  $\triangle$

Si controlla ora facilmente che la funzione  $\mathbf{P}(\cdot | A) : \mathcal{A} \rightarrow [0, 1]$  così definita gode delle proprietà tipiche di una probabilità (vedi Osservazione I.2.15), come ad esempio:

$$\begin{aligned} \mathbf{P}(\emptyset | A) &= 0, \\ \mathbf{P}(\Omega | A) &= \mathbf{P}(A | A) = 1, \\ \mathbf{P}(B_1 \cup B_2 | A) &= \mathbf{P}(B_1 | A) + \mathbf{P}(B_2 | A), \quad \text{se } B_1 \cap B_2 = \emptyset, \quad (\text{additività}) \\ \mathbf{P}(\overline{B} | A) &= 1 - \mathbf{P}(B | A), \end{aligned}$$

sicché  $(\Omega, \mathcal{A}, \mathbf{P}(\cdot | A))$  si comporta come un nuovo spazio di probabilità nel quale è ora incorporata l'informazione che l'evento  $A$  si verifica certamente ( $\mathbf{P}(A | A) = 1$ ).

**I.4.3 Osservazione:** Va detto subito che l'espressione  $\mathbf{P}(\cdot | \cdot)$  non è affatto simmetrica nei suoi due argomenti: non solo  $\mathbf{P}(B | A) \neq \mathbf{P}(A | B)$ , ma  $\mathbf{P}(B | \cdot)$ , a differenza di  $\mathbf{P}(\cdot | A)$ , non gode affatto delle proprietà tipiche di una probabilità; ad esempio risulta

$$\mathbf{P}(B | A) + \mathbf{P}(B | \bar{A}) \neq 1,$$

$$\mathbf{P}(B | A) + \mathbf{P}(\bar{B} | \bar{A}) \neq 1,$$

per cui  $\mathbf{P}(B | \cdot)$  non è additiva. Nota che anche la probabilità congiunta di due eventi permette di definire un'altra applicazione  $\mathbf{P}(\cdot | A) : \mathcal{A} \rightarrow [0, 1]$  ma che questa non è affatto una nuova probabilità: ad esempio in generale risulta  $\mathbf{P}(\Omega | A) = \mathbf{P}(A) \leq 1$ , cioè non vale sempre 1 come per una probabilità. Infine è importante osservare che, sebbene un condizionamento modifichi le attribuzioni di probabilità, esso lascia comunque intatte alcune particolari valutazioni. Ad esempio, se  $A$  è un evento tale che  $\mathbf{P}(A) = 0$ , quale che sia l'evento condizionante  $B$  risulterà sempre  $\mathbf{P}(A | B) = 0$ , dato che  $AB \subseteq A$  e quindi  $0 \leq \mathbf{P}(AB) \leq \mathbf{P}(A) = 0$ . Analogamente, se  $\mathbf{P}(A) = 1$ , risulterà sempre  $\mathbf{P}(A | B) = 1$  per qualunque scelta dell'evento condizionante  $B$ .  $\circ$

**I.4.4 Esempio:** Supponiamo di considerare una famiglia con due figli e calcoliamo la probabilità che ambedue siano maschi ( $M$ ) condizionata da uno dei due seguenti eventi:

- a) il più anziano è maschio;
- b) almeno uno dei due è maschio.

Lo spazio dei campioni è ora  $\Omega = \{MM, MF, FM, FF\}$ , cioè è l'insieme delle possibili coppie dei sessi dei due figli nelle quali conveniamo di indicare l'anzianità con l'ordine: il primo posto è quello del più anziano. Supporremo, in mancanza di altre informazioni, che i quattro eventi elementari considerati siano equiprobabili con probabilità  $\frac{1}{4}$ . Gli eventi che ci interessano sono ora:

$$A = \{MM, MF\} = \text{il più anziano è maschio,}$$

$$B = \{MM, FM\} = \text{il più giovane è maschio,}$$

$$A \cup B = \{MM, MF, FM\} = \text{almeno uno dei due è maschio,}$$

$$A \cap B = \{MM\} = \text{ambedue i figli sono maschi.}$$

Le rispettive probabilità di tali eventi sono ovviamente:

$$\mathbf{P}(A) = \frac{1}{2}, \quad \mathbf{P}(B) = \frac{1}{2}, \quad \mathbf{P}(A \cup B) = \frac{3}{4}, \quad \mathbf{P}(A \cap B) = \frac{1}{4}.$$

Ne segue che

$$\mathbf{P}(A \cap B | A) = \frac{\mathbf{P}((A \cap B) \cap A)}{\mathbf{P}(A)} = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A)} = \frac{1}{2},$$

$$\mathbf{P}(A \cap B | A \cup B) = \frac{\mathbf{P}((A \cap B) \cap (A \cup B))}{\mathbf{P}(A \cup B)} = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A \cup B)} = \frac{1}{3},$$

sono le probabilità richieste.  $\diamond$

**I.4.5 Proposizione (Formula della Probabilità Totale):** Dati in uno spazio finito di probabilità  $(\Omega, \mathcal{A}, \mathbf{P})$  un arbitrario evento  $A \in \mathcal{A}$  ed una decomposizione  $\mathcal{D} = \{D_1, \dots, D_n\}$  risulta sempre

$$\mathbf{P}(A) = \sum_{i=1}^n \mathbf{P}(A | D_i) \mathbf{P}(D_i),$$

se  $\mathbf{P}(D_i) > 0, i = 1, \dots, n$ .

**Dimostrazione:** Basterà osservare che

$$A = A \cap \Omega = A \cap \left( \bigcup_{i=1}^n D_i \right) = \bigcup_{i=1}^n (A \cap D_i),$$

e che gli eventi  $A \cap D_i$  sono tutti disgiunti per ottenere

$$\mathbf{P}(A) = \mathbf{P} \left( \bigcup_{i=1}^n (A \cap D_i) \right) = \sum_{i=1}^n \mathbf{P}(A \cap D_i) = \sum_{i=1}^n \mathbf{P}(A | D_i) \mathbf{P}(D_i),$$

dalla proprietà di additività.  $\square$

**I.4.6 Osservazione:** Osserviamo che in particolare, quando la decomposizione si riduce a  $\mathcal{D} = \{B, \bar{B}\}$ , la formula della Probabilità Totale diviene

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

espressione particolarmente facile da usare ed interpretare.  $\circ$

**I.4.7 Esempio:** Riprendiamo in considerazione la scatola di palline dell'esempio I.4.1, estraiamo in successione e senza rimessa due palline e, senza guardare la prima, chiediamoci quale è la probabilità che la seconda sia bianca. Definiamo, a questo scopo, gli eventi

$A$  = la prima pallina estratta è bianca,

$B$  = la seconda pallina estratta è bianca,

e notiamo che, in base all'Osservazione I.4.6, risulta

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

Si ha ovviamente che

$$\mathbf{P}(A) = \frac{m}{M}, \quad \mathbf{P}(\bar{A}) = \frac{M - m}{M}.$$

Inoltre, nel caso di due estrazioni successive, enumerando i casi possibili e i casi favorevoli agli eventi presi in considerazione, otteniamo facilmente che

$$\mathbf{P}(B | A) = \frac{m-1}{M-1}, \quad \mathbf{P}(B | \bar{A}) = \frac{m}{M-1},$$

per cui si ha, tenendo conto dell'Osservazione I.4.6,

$$\mathbf{P}(B) = \frac{m-1}{M-1} \frac{m}{M} + \frac{m}{M-1} \frac{M-m}{M} = \frac{m}{M}.$$

La valutazione di  $B$  è dunque diversa secondo le informazioni che si hanno: ad esempio, mentre  $\mathbf{P}(B | A)$  e  $\mathbf{P}(B | \bar{A})$  sono diversi da  $\mathbf{P}(A)$ , risulta  $\mathbf{P}(B) = \mathbf{P}(A)$ : quando il primo risultato è sconosciuto esso non influenza la probabilità del secondo.  $\diamond$

**I.4.8 Proposizione (Formula di Moltiplicazione):** Comunque assegnati  $A_1, \dots, A_n$  in uno spazio finito di probabilità  $(\Omega, \mathcal{A}, \mathbf{P})$  risulta sempre

$$\begin{aligned} \mathbf{P}(A_1 \dots A_n) \\ = \mathbf{P}(A_n | A_{n-1} \dots A_1) \mathbf{P}(A_{n-1} | A_{n-2} \dots A_1) \dots \mathbf{P}(A_2 | A_1) \mathbf{P}(A_1), \end{aligned}$$

se  $\mathbf{P}(A_1 \dots A_{n-1}) \neq 0$ .

(Useremo spesso la notazione  $AB \dots G$  per indicare  $A \cap B \cap \dots \cap G$ ).

**Dimostrazione:** Infatti dalla definizione di probabilità condizionate risulta

$$\begin{aligned} \mathbf{P}(A_n | A_{n-1} \dots A_1) \mathbf{P}(A_{n-1} | A_{n-2} \dots A_1) \dots \mathbf{P}(A_2 | A_1) \mathbf{P}(A_1) \\ = \frac{\mathbf{P}(A_1 \dots A_n)}{\mathbf{P}(A_1 \dots A_{n-1})} \frac{\mathbf{P}(A_1 \dots A_{n-1})}{\mathbf{P}(A_1 \dots A_{n-2})} \dots \frac{\mathbf{P}(A_1 A_2)}{\mathbf{P}(A_1)} \mathbf{P}(A_1) = \mathbf{P}(A_1 \dots A_n), \end{aligned}$$

come volevasi dimostrare.  $\square$

**I.4.9 Proposizione (Teorema di Bayes):** Dati due eventi  $A, B$  in uno spazio finito di probabilità  $(\Omega, \mathcal{A}, \mathbf{P})$ , se  $\mathbf{P}(A) > 0$ ,  $\mathbf{P}(B) > 0$ , risulta

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

inoltre risulta

$$\mathbf{P}(D_i | B) = \frac{\mathbf{P}(B | D_i) \mathbf{P}(D_i)}{\sum_{i=1}^n \mathbf{P}(B | D_i) \mathbf{P}(D_i)},$$

se  $\mathcal{D} = \{D_1, \dots, D_n\}$  è una decomposizione di  $\Omega$  con  $\mathbf{P}(D_i) > 0$ ,  $i = 1, \dots, n$ .

**Dimostrazione:** La dimostrazione della prima relazione (detta anche **Formula di Bayes**<sup>1</sup>) si basa sul fatto che per definizione di probabilità condizionata si ha

$$\mathbf{P}(B | A) \mathbf{P}(A) = \mathbf{P}(AB) = \mathbf{P}(A | B) \mathbf{P}(B);$$

la seconda relazione si ottiene poi dalla prima tramite il Teorema della Probabilità Totale.  $\square$

**I.4.10 Osservazione:** Nelle applicazioni statistiche gli eventi  $D_i$  del Teorema di Bayes sono spesso chiamati *ipotesi* e  $\mathbf{P}(D_i)$  *probabilità a priori* di tali ipotesi, mentre le probabilità condizionate  $\mathbf{P}(D_i | B)$  si chiamano *probabilità a posteriori*. Questa terminologia è basata, come vedremo nell'esempio successivo, sull'idea che la conoscenza del fatto che un certo evento  $B$  si è verificato modifica la probabilità che viene assegnata alle ipotesi del nostro problema.  $\circ$

**I.4.11 Esempio:** Supponiamo che in una scatola ci siano due monete, una equa (tale cioè che la probabilità di ottenere  $T$  oppure  $C$  è la stessa:  $\frac{1}{2}$ ) ed una truccata (per la quale supporremo che  $\mathbf{P}(T) = \frac{1}{3}$ ;  $\mathbf{P}(C) = \frac{2}{3}$ ). Supponiamo ora di estrarre una moneta a caso e di lanciarla: che influenza avrà la conoscenza del risultato del lancio sulla probabilità che io abbia estratto proprio la moneta equa? Per rispondere a questa domanda definiamo i seguenti eventi:

- $D_1$  = la moneta estratta è quella equa,
- $D_2$  = la moneta estratta è quella truccata,
- $A$  = il risultato del lancio è  $T$ ,
- $B$  = il risultato del lancio è  $C$ .

Siccome inoltre sappiamo che  $\mathcal{D} = \{D_1, D_2\}$  è una decomposizione di  $\Omega$  e che

$$\begin{aligned} \mathbf{P}(D_1) &= \mathbf{P}(D_2) = \frac{1}{2}, \\ \mathbf{P}(A | D_1) &= \frac{1}{2}, \quad \mathbf{P}(B | D_1) = \frac{1}{2}, \\ \mathbf{P}(A | D_2) &= \frac{1}{3}, \quad \mathbf{P}(B | D_2) = \frac{2}{3}, \end{aligned}$$

dal Teorema di Bayes otteniamo:

$$\begin{aligned} \mathbf{P}(D_1 | A) &= \frac{\mathbf{P}(A | D_1) \mathbf{P}(D_1)}{\mathbf{P}(A | D_1) \mathbf{P}(D_1) + \mathbf{P}(A | D_2) \mathbf{P}(D_2)} = \frac{3}{5}, \\ \mathbf{P}(D_2 | A) &= \frac{2}{5}. \end{aligned}$$

---

<sup>1</sup> Thomas Bayes (1702 - 1761): teologo e matematico inglese, pioniere nell'uso induttivo della probabilità. I suoi lavori hanno gettato le basi matematiche della cosiddetta *stima Bayesiana*, un metodo di inferenza statistica usato per calcolare la probabilità di eventi futuri sulla base della frequenza con cui tali eventi si sono verificati nel passato. La proposizione qui riportata fu enunciata per la prima volta da Bayes verso la metà del Settecento e da allora è riportata quasi imm modificata in tutti i manuali di calcolo delle probabilità.

Pertanto, se il risultato del lancio della moneta è  $T$ , la probabilità di aver estratto la moneta equa (ipotesi) è aumentata dal suo valore a priori  $\frac{1}{2}$  al suo valore a posteriori (cioè dopo aver effettuato il lancio)  $\frac{3}{5}$ . Naturalmente la confidenza con cui riconosceremo il tipo di moneta estratto aumenterà con il numero di lanci effettuato.  $\diamond$

A questo punto è possibile introdurre il concetto di *eventi indipendenti*: è piuttosto naturale pensare che due eventi sono indipendenti quando il verificarsi di uno di essi non ha alcun effetto sul valore della probabilità che viene attribuita all'altro. Sulla base del concetto di probabilità condizionata introdotto prima diremo quindi che l'evento  $A$  è indipendente dall'evento  $B$  quando  $\mathbf{P}(A | B) = \mathbf{P}(A)$  e quindi se  $\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B)$  (osserva che questa espressione, a differenza di quella basata sulle probabilità condizionate, ha senso anche quando  $\mathbf{P}(B) = 0$ ). È facile inoltre, sulla base della simmetria di queste relazioni, convincersi del fatto che se  $A$  è indipendente da  $B$ , anche  $B$  è indipendente da  $A$ .

**I.4.12 Definizione:** Dato uno spazio di probabilità  $(\Omega, \mathcal{A}, \mathbf{P})$  diremo che  $A$  e  $B$  sono **eventi indipendenti** quando

$$\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B);$$

diremo inoltre che due algebre  $\mathcal{A}_1$  e  $\mathcal{A}_2$  di eventi di  $\Omega$  (più precisamente due sottoalgebre di  $\mathcal{A}$ ) sono due **algebre indipendenti** se ogni evento di  $\mathcal{A}_1$  è indipendente da ogni evento di  $\mathcal{A}_2$ .  $\triangle$

**I.4.13 Esempio:** È facile controllare che le due algebre

$$\begin{aligned}\mathcal{A}_1 &= \{A_1, \bar{A}_1, \Omega, \emptyset\}, \\ \mathcal{A}_2 &= \{A_2, \bar{A}_2, \Omega, \emptyset\},\end{aligned}$$

generate dagli eventi  $A_1$  e  $A_2$  sono indipendenti se e solo se sono indipendenti gli eventi  $A_1$  e  $A_2$ . Infatti se  $\mathcal{A}_1$  è indipendente da  $\mathcal{A}_2$ , anche  $A_1$  risulterà indipendente da  $A_2$  per definizione di indipendenza di due algebre. Viceversa, se  $A_1$  e  $A_2$  sono indipendenti basterà verificare che tutte le altre possibili coppie di eventi estratti da  $\mathcal{A}_1$  e  $\mathcal{A}_2$  sono indipendenti; ad esempio:

$$\begin{aligned}\mathbf{P}(A_1 \bar{A}_2) &= \mathbf{P}(A_1) - \mathbf{P}(A_1 A_2) = \mathbf{P}(A_1) - \mathbf{P}(A_1)\mathbf{P}(A_2) \\ &= \mathbf{P}(A_1)\mathbf{P}(\bar{A}_2).\end{aligned}$$

Lasciamo al lettore la verifica delle altre relazioni.  $\diamond$

Il concetto di indipendenza può essere esteso anche al caso in cui il numero di eventi è maggiore di due, ma bisogna fare molta attenzione al fatto che sarà ora possibile parlare di indipendenza *due a due*, nel senso di  $\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B)$ , di indipendenza *tre a tre*, nel senso di  $\mathbf{P}(ABC) = \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C)$ , e così via, e che tali *livelli di indipendenza* non si implicano affatto l'uno con l'altro, nel senso che



(come vedremo fra breve) ad esempio tre eventi possono essere indipendenti due a due senza esserlo tre a tre e viceversa.

**I.4.14 Definizione:** Dato uno spazio di probabilità  $(\Omega, \mathcal{A}, \mathbf{P})$  diremo che gli eventi  $A_1, \dots, A_n$  sono  $n$  **eventi indipendenti** quando  $\forall k = 1, \dots, n$  e  $\forall \{i_1, \dots, i_k\}$  con  $1 \leq i_1 < i_2 < \dots < i_k \leq n$  risulta

$$\mathbf{P}(A_{i_1} \dots A_{i_k}) = \mathbf{P}(A_{i_1}) \dots \mathbf{P}(A_{i_k});$$

cioè quando sono indipendenti due a due, tre a tre,  $\dots$ ,  $n$  a  $n$ . △

**I.4.15 Esempio:** Consideriamo uno spazio dei campioni composto di quattro soli elementi tutti equiprobabili:  $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$  con  $\mathbf{P}\{\omega_i\} = \frac{1}{4}$ ,  $i = 1, 2, 3, 4$ . Gli eventi

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

sono indipendenti due a due (come si controlla facilmente) ma non tre a tre in quanto  $\mathbf{P}(ABC) = \mathbf{P}\{\omega_1\} = \frac{1}{4}$  mentre  $\mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C) = (\frac{1}{2})^3 = \frac{1}{8}$ . Ciò mostra che l'indipendenza due a due non implica quella tre a tre. Per mostrare che neanche l'indipendenza tre a tre implica quella due a due consideriamo lo spazio dei campioni di due lanci di un dado equo  $\Omega = \{(i, j) : i, j = 1, \dots, 6\}$  (si tratta di 36 coppie ordinate a ciascuna delle quali viene attribuita una probabilità di  $\frac{1}{36}$ ) e consideriamo gli eventi

$$A = \{(i, j) : j = 1, 2, 5\}, \quad B = \{(i, j) : j = 4, 5, 6\}, \quad C = \{(i, j) : i + j = 9\}.$$

Si controlla facilmente ora che

$$\begin{aligned} \mathbf{P}(A) &= \frac{1}{2}, & \mathbf{P}(B) &= \frac{1}{2}, & \mathbf{P}(C) &= \frac{4}{36} = \frac{1}{9}, \\ \mathbf{P}(AB) &= \frac{1}{6}, & \mathbf{P}(AC) &= \frac{1}{36}, & \mathbf{P}(BC) &= \frac{3}{36} = \frac{1}{12}, \\ & & \mathbf{P}(ABC) &= \frac{1}{36} \end{aligned}$$

sicché risulta

$$\mathbf{P}(ABC) = \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C),$$

ma

$$\mathbf{P}(AB) \neq \mathbf{P}(A)\mathbf{P}(B), \quad \mathbf{P}(AC) \neq \mathbf{P}(A)\mathbf{P}(C), \quad \mathbf{P}(BC) \neq \mathbf{P}(B)\mathbf{P}(C),$$

cioè  $A, B, C$  sono indipendenti tre a tre ma non due a due. ◇

**I.4.16 Osservazione:** L'indipendenza di due, o più, eventi è determinata non solo dalla natura o dalla definizione di tali eventi (come essi sono dati nell'algebra  $\mathcal{A}$  di parti di  $\Omega$ ), ma anche, e soprattutto, dalla probabilità  $\mathbf{P}(\cdot)$  definita sullo spazio di probabilità dato: gli stessi eventi (identici dal punto di vista insiemistico in

$(\Omega, \mathcal{A})$ ) possono risultare indipendenti o dipendenti secondo il tipo di probabilità  $\mathbf{P}(\cdot)$  assegnata. Questo sarà particolarmente evidente nel concetto di *indipendenza condizionata* che ci permetterà di confrontare l'indipendenza di eventi rispetto a due diverse probabilità:  $\mathbf{P}(\cdot)$  e  $\mathbf{P}(\cdot | D)$ .  $\circ$

**I.4.17 Definizione:** Dato uno spazio di probabilità  $(\Omega, \mathcal{A}, \mathbf{P})$  diremo che  $A$  e  $B$  sono **eventi condizionatamente indipendenti** (rispetto a  $D$ ) quando

$$\mathbf{P}(AB | D) = \mathbf{P}(A | D) \mathbf{P}(B | D)$$

essendo  $D \in \mathcal{A}$  un evento tale che  $\mathbf{P}(D) > 0$ .  $\triangle$

**I.4.18 Esempio:** Sia  $\Omega = \{1, 2, \dots, 6\}$  lo spazio dei campioni di un dado equo: i due eventi  $A = \{1, 2\}$ ,  $B = \{2, 4, 6\}$  con  $\mathbf{P}(A) = \frac{1}{3}$  e  $\mathbf{P}(B) = \frac{1}{2}$  sono evidentemente indipendenti in quanto  $AB = \{2\}$  e quindi  $\mathbf{P}(AB) = \frac{1}{6} = \mathbf{P}(A)\mathbf{P}(B)$ . Se però consideriamo l'evento  $D = \{2, 3\}$  con  $\mathbf{P}(D) = \frac{1}{3}$ , si ha che  $\mathbf{P}(A | D) = \frac{1}{2}$ ,  $\mathbf{P}(B | D) = \frac{1}{2}$ , e quindi che  $\mathbf{P}(AB | D) = \frac{1}{2}$  è diverso da  $\mathbf{P}(A | D)\mathbf{P}(B | D) = \frac{1}{4}$ , sicché  $A$  e  $B$  non sono condizionatamente indipendenti rispetto a  $D$ . Viceversa, se avessimo considerato gli eventi  $A = \{1, 2, 3\}$ ,  $B = \{2, 4, 6\}$  con  $\mathbf{P}(A) = \mathbf{P}(B) = \frac{1}{2}$  li avremmo trovati non indipendenti in quanto  $\mathbf{P}(AB) = \frac{1}{6} \neq \mathbf{P}(A)\mathbf{P}(B)$ ; ma dato l'evento  $D = \{2, 3, 4, 5\}$  con  $\mathbf{P}(D) = \frac{2}{3}$  si controlla facilmente che  $\mathbf{P}(A | D) = \mathbf{P}(B | D) = \frac{1}{2}$  e che  $\mathbf{P}(AB | D) = \frac{1}{4}$ , sicché  $A$  e  $B$  risultano condizionatamente indipendenti rispetto a  $D$ .  $\diamond$

**I.4.19 Esempio:** Dato  $(\Omega, \mathcal{A}, \mathbf{P})$ , un evento  $A$  può essere indipendente da se stesso se e solo se  $\mathbf{P}(A) = 0$  oppure  $\mathbf{P}(A) = 1$ . Infatti deve risultare  $\mathbf{P}(AA) = \mathbf{P}(A) = [\mathbf{P}(A)]^2$  e quindi  $\mathbf{P}(A) = 0$  oppure  $\mathbf{P}(A) = 1$ . In tal caso si verifica anche che  $A$  risulta indipendente da qualunque altro evento  $B \in \mathcal{A}$ . Infatti, se  $\mathbf{P}(A) = 0$ , siccome  $AB \subseteq A$ , si ha (vedi Osservazione I.2.15)  $0 \leq \mathbf{P}(AB) \leq \mathbf{P}(A) = 0$  cioè  $\mathbf{P}(AB) = 0 = \mathbf{P}(A)\mathbf{P}(B)$ . Se invece  $\mathbf{P}(A) = 1$ , sarà anche  $\mathbf{P}(\bar{A}) = 0$  e quindi, come nel caso precedente,  $\mathbf{P}(\bar{A}B) = 0 = \mathbf{P}(\bar{A})\mathbf{P}(B)$ , cioè  $\mathbf{P}(B) - \mathbf{P}(AB) = [1 - \mathbf{P}(A)]\mathbf{P}(B)$  da cui  $\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B)$ . Sarà il caso di notare, a proposito di questo esempio, che un evento di probabilità nulla non è necessariamente coincidente con l'insieme vuoto  $\emptyset$ , e che un evento che avviene con probabilità 1 non è necessariamente coincidente con tutto  $\Omega$ : questa particolarità (che sarà ripresa più oltre, quando tratteremo spazi *non finiti* di probabilità) può essere meglio compresa se si pensa che, in generale, è possibile assegnare dei pesi probabilistici nulli ad alcuni degli eventi elementari  $\omega \in \Omega$ . Ovviamente sarebbe facile, almeno nel caso di spazi finiti di probabilità, modificare la struttura di  $\Omega$  eliminando gli eventi elementari che si verificano con probabilità nulla, in quanto ci attendiamo che non si verifichino mai; ma questa operazione (del tutto impossibile, come vedremo, nel caso più generale di spazi non finiti) si rivelerebbe spesso solo un'apparente razionalizzazione anche quando fosse possibile. Infatti, come nel caso dell'Esempio I.4.18, molte volte è importante poter confrontare due o più assegnazioni di pesi probabilistici (come  $\mathbf{P}(\cdot)$  e  $\mathbf{P}(\cdot | D)$  dell'esempio citato) e può capitare, ad esempio, che ad eventi elementari con probabilità non nulla secondo

un'assegnazione venga attribuita una probabilità nulla secondo un'altra. In questo caso sarebbe particolarmente inopportuno modificare ogni volta la struttura dello spazio  $\Omega$  e si preferisce piuttosto conservarne tutti gli elementi anche se a qualcuno può capitare di assegnare una probabilità nulla.  $\diamond$

**I.4.20 Esempio (Modello di Bernoulli):** Riprendiamo ora in considerazione il modello costruito nel capitolo I.3 per introdurre la distribuzione binomiale, e consideriamo gli  $n$  eventi

$$A_k = \{\omega \in \Omega : a_k = 1\}$$

= si registra un successo al  $k$ -mo tentativo.

Osserviamo innanzitutto che tali  $A_k$  sono eventi in  $\mathcal{A}^*$ , cioè sono insiemi di  $n$ -ple  $\omega = (a_1, \dots, a_n)$ , più precisamente l'insieme delle  $n$ -ple tali che  $a_k = 1$  mentre le  $a_j$  con  $j \neq k$  assumono indifferentemente valore 0 oppure 1. Inoltre, a differenza degli eventi  $D_k$  introdotti nell'osservazione I.3.2, gli eventi  $A_k$  non sono disgiunti in quanto l'aver registrato un successo al  $k$ -mo tentativo non esclude la possibilità che si possano registrare successi anche negli altri tentativi. Definiremo inoltre le algebre

$$\mathcal{A}_k = \{A_k, \bar{A}_k, \Omega, \emptyset\}.$$

Si verifica ora facilmente che

$$\mathbf{P}(A_k) = p, \quad \mathbf{P}(\bar{A}_k) = q = 1 - p;$$

infatti, dalle definizioni date in I.3, siccome  $l = a_1 + \dots + a_{k-1} + a_{k+1} + \dots + a_n$  può assumere valori da 0 ad  $n - 1$ , e calcolato come in Esempio I.2.4 il numero di modi in cui si possono disporre  $l$  simboli "1" su  $n - 1$  posti, si ha

$$\begin{aligned} \mathbf{P}(A_k) &= \sum_{\omega \in A_k} p(\omega) \\ &= \sum_{a_i; i \neq k} p^{a_1 + \dots + a_{k-1} + 1 + a_{k+1} + \dots + a_n} \cdot q^{n - (a_1 + \dots + a_{k-1} + 1 + a_{k+1} + \dots + a_n)} \\ &= p \sum_{l=0}^{n-1} \binom{n-1}{l} p^l q^{n-1-l} = p, \end{aligned}$$

ed analogamente per  $\bar{A}_k$ . Con un calcolo simile si prova anche che

$$\mathbf{P}(A_k A_l) = p^2, \quad \mathbf{P}(A_k \bar{A}_l) = pq, \quad \mathbf{P}(\bar{A}_k \bar{A}_l) = q^2, \quad (k \neq l)$$

e quindi che gli eventi  $A_k$  ed  $A_l$  e le algebre  $\mathcal{A}_k$  e  $\mathcal{A}_l$  sono indipendenti ( $k \neq l$ ) rispetto alla probabilità da noi assegnata al nostro modello in I.3; anzi, generalizzando il calcolo, si potrebbe dimostrare che le algebre  $\mathcal{A}_1, \dots, \mathcal{A}_n$  sono tutte indipendenti rispetto alla probabilità  $\mathbf{P}(\cdot)$  assegnata. Tutto ciò conferma quanto anticipato nell'Osservazione I.3.1 circa il fatto che lo spazio di probabilità da noi

definito costituisce *un modello per la descrizione di  $n$  tentativi indipendenti di verifica di un dato evento con probabilità di successo  $p$  in ogni singolo tentativo*. Un modello siffatto prende il nome di **Modello di Bernoulli**<sup>2</sup> e costituisce il primo esempio di uno spazio di probabilità costruito come **prodotto diretto** di un numero finito di altri spazi di probabilità: dati  $n$  spazi finiti di probabilità  $(\Omega_k, \mathcal{B}_k, \mathbf{P}_k)$  con  $k = 1, \dots, n$  si costruisce un nuovo spazio dei campioni

$$\Omega^{(n)} = \Omega_1 \times \dots \times \Omega_n$$

i cui elementi  $\omega \in \Omega^{(n)}$  sono  $n$ -ple di elementi degli spazi  $\Omega_k$  del tipo  $\omega = (b_1, \dots, b_n)$  dove  $b_k \in \Omega_k$  con  $k = 1, \dots, n$ ; si costruisce poi l'algebra

$$\mathcal{A}^{(n)} = \mathcal{B}_1 \otimes \dots \otimes \mathcal{B}_n$$

ottenuta eseguendo tutte le possibili (sono in numero finito, dato che i nostri sono spazi finiti di probabilità) unioni di parti di  $\Omega^{(n)}$ , date come prodotti cartesiani fra insiemi, del tipo  $A = B_1 \times \dots \times B_n$  con  $B_k \in \mathcal{B}_k$ ; infine si assegna ad ogni evento elementare  $\omega = (b_1, \dots, b_n)$  la probabilità  $p_1(b_1) \cdot \dots \cdot p_n(b_n)$  di ogni evento elementare (dove ovviamente  $p_k(b_k) = \mathbf{P}_k(\{b_k\})$ ), quella di eventi del tipo  $A = B_1 \times \dots \times B_n$ :

$$\mathbf{P}^{(n)}(A) = \sum_{b_k \in B_k; k=1 \dots n} p_1(b_1) \cdot \dots \cdot p_n(b_n),$$

e quella di generici eventi non rettangolari  $A \in \mathcal{A}^{(n)}$ :

$$\mathbf{P}^{(n)}(A) = \sum_{(b_1, \dots, b_n) \in A} p_1(b_1) \cdot \dots \cdot p_n(b_n).$$

Ciò fatto si verifica che  $\mathbf{P}^{(n)}(\cdot)$  è additiva e che  $\mathbf{P}^{(n)}(\Omega^{(n)}) = 1$  sicché lo spazio  $(\Omega^{(n)}, \mathcal{A}^{(n)}, \mathbf{P}^{(n)})$  è un nuovo spazio di probabilità detto *prodotto diretto* degli  $n$  spazi iniziali. Nel caso particolare del modello di Bernoulli gli spazi  $(\Omega_k, \mathcal{B}_k, \mathbf{P}_k)$  sono tutti uguali fra loro (sono  $n$  copie distinte di un unico spazio di probabilità)

---

<sup>2</sup> Primo di una numerosa famiglia di scienziati (tra i quali, oltre a Jacques, ricorderemo almeno suo fratello Jean e suo nipote Daniel), stabilitasi in Svizzera dalle Fiandre per sfuggire alle persecuzioni religiose del Duca d'Alba contro i protestanti, Jacques Bernoulli (1654 - 1705) fu professore di matematica all'università di Basilea e scrisse un gran numero di memorie scientifiche sui soggetti più svariati: serie, coniche, curve trascendenti, isoperimetria, per citarne solo alcune. Pubblicò il primo libro conosciuto dedicato interamente alla teoria delle probabilità: l'*Ars Coniectandi* (Basilea, 1713; postumo). Prima di lui infatti il calcolo delle probabilità era stato trattato da vari matematici (Cardano, Tartaglia, Fermat e Pascal), ma solo Christian Huyghens (1629 - 1695) aveva pubblicato qualcosa di simile ad un trattato: il *De ratiociniis in ludo aleae* (Leida, 1657). Tra le altre cose l'*Ars Coniectandi* contiene la parte essenziale della teoria delle combinazioni come è conosciuta oggi e per la prima volta vi appare, con l'attuale significato, la parola *permutazione*.

e risulta  $\Omega_k = \{0, 1\}$ ,  $\mathcal{B}_k = \{\{0\}, \{1\}, \Omega, \emptyset\}$ ,  $p_k(1) = p$  e  $p_k(0) = q$ . Va osservato ancora una volta che negli spazi *prodotto diretto* la probabilità  $\mathbf{P}^{(n)}(\cdot)$  è definita in modo tale che gli eventi di  $\mathcal{A}^{(n)}$  del tipo

$$A_k = \{\omega \in \Omega^{(n)} : b_k \in B_k\} = \Omega_1 \times \dots \times \Omega_{n-1} \times B_k \times \Omega_{n+1} \times \dots \times \Omega_n$$

con  $k = 1, \dots, n$  e  $B_k \in \mathcal{B}_k$ , risultino tutti indipendenti quale che sia la scelta degli eventi  $B_k \in \mathcal{B}_k$ . Ricordiamo infine che questa non è l'unica maniera in cui è possibile definire una probabilità su  $(\Omega^{(n)}, \mathcal{A}^{(n)})$ : si tratta solamente della più semplice.  $\diamond$



## I.5 Variabili Aleatorie

In un generico spazio di probabilità  $(\Omega, \mathcal{A}, \mathbf{P})$  (nel seguito e fino ad avviso contrario supporremo di aver a che fare solo con spazi finiti) l'insieme  $\Omega$  non è un insieme numerico come può essere desunto da molti degli esempi dei capitoli precedenti. D'altra parte nella realizzazione di esperimenti concreti si ha per lo più a che fare con numeri che rappresentano i risultati delle misure di quantità fisiche. È dunque importante che i nostri modelli matematici ci consentano di associare entità numeriche agli oggetti astratti che costituiscono il generico spazio dei campioni  $\Omega$ .

**I.5.1 Definizione:** una **variabile aleatoria (v.a.)** è una funzione a valori reali  $\xi = \xi(\omega)$  definita su uno spazio dei campioni  $\Omega$ .  $\triangle$

**I.5.2 Esempio:** Sullo spazio dei campioni relativo a due lanci di una moneta

$$\Omega = \{TT, TC, CT, CC\}$$

può essere definita la v.a.

$$\xi = \text{numero di } \textit{teste} \text{ su due lanci}$$

i cui valori sono ovviamente dati dalla seguente tabella

$$\begin{array}{cccc} \omega & = & TT & TC & CT & CC \\ \xi(\omega) & = & 2 & 1 & 1 & 0 \end{array} .$$

Naturalmente sul medesimo spazio  $\Omega$  possono essere definite diverse v.a.: il lettore è invitato a costruire qualche altro semplice esempio.  $\diamond$

**I.5.3 Esempio:** Dato un evento  $A \in \mathcal{A}$  è sempre possibile definire una v.a. particolarmente semplice, ma importante, che prende il nome di **indicatore** dell'evento  $A$  e che è definita come

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

Avremo modo di utilizzare ampiamente questa nozione nei seguenti capitoli.  $\diamond$

**I.5.4 Osservazione:** Se  $\Omega$  è finito, anche i possibili valori distinti di  $\xi$  saranno in numero finito e costituiranno l'insieme  $\mathcal{X} = \{x_1, \dots, x_m\} \subseteq \mathbf{R}$ . Per ragioni di comodità, nel seguito supporremo sempre (salvo avviso contrario) che i numeri  $x_1, \dots, x_m$  siano ordinati in modo crescente, sicché  $x_1 < \dots < x_m$ . Naturalmente, siccome è possibile che  $\xi(\omega') = \xi(\omega'')$  anche quando  $\omega' \neq \omega''$ , il numero di elementi di  $\mathcal{X}$  sarà minore o eguale al numero di elementi di  $\Omega$  e inoltre gli insiemi

$$\{\xi = x_k\} = \xi^{-1}(x_k) = \{\omega \in \Omega : \xi(\omega) = x_k\}$$

saranno parti di  $\Omega$ : noi faremo sempre l'ipotesi che l'algebra  $\mathcal{A}$  di parti di  $\Omega$  usata per la definizione del nostro spazio di probabilità sia abbastanza grande da

contenere tutti gli insiemi del tipo  $\{\xi = x_k\}$ ; supporremo cioè che  $\{\xi = x_k\} \in \mathcal{A}$  con  $k = 1, \dots, m$ . Ovviamente questa proprietà (detta **misurabilità** della funzione  $\xi : \Omega \rightarrow \mathbf{R}$ ) sarà sempre soddisfatta se useremo come algebra degli eventi l'algebra  $\mathcal{A}^*$  di tutte le parti di  $\Omega$ . L'importanza di questa ipotesi sarà subito chiarita dalla seguente costruzione: se definiamo come  $\mathcal{B} = \wp(\mathcal{X})$  la famiglia di tutte le parti di  $\mathcal{X}$ , la coppia  $(\mathcal{X}, \mathcal{B})$  avrà le stesse caratteristiche di  $(\Omega, \mathcal{A})$ , anzi l'ipotesi fatta ci consente di dire che

$$\{\xi \in B\} = \xi^{-1}(B) = \{\omega \in \Omega : \xi(\omega) \in B\} \in \mathcal{A}, \quad \forall B \in \mathcal{B}.$$

Potremo quindi pensare di dotare  $(\mathcal{X}, \mathcal{B})$  di una probabilità  $\mathbf{P}_\xi(\cdot)$  ottenuta a partire da  $\xi$  e da  $\mathbf{P}(\cdot)$ . Infatti, comunque scelto  $B \in \mathcal{B}$ , sarà sufficiente porre

$$\mathbf{P}_\xi(B) = \mathbf{P}\{\xi \in B\} = \mathbf{P}(\xi^{-1}(B)) = \sum_{\omega \in \xi^{-1}(B)} \mathbf{P}\{\omega\}$$

per ottenere la probabilità  $\mathbf{P}_\xi(\cdot)$  richiesta. È lasciata come esercizio per il lettore la verifica del fatto che  $\mathbf{P}_\xi(\cdot)$  soddisfi tutte le proprietà richieste ad una vera probabilità (vedi Osservazione I.2.15). In pratica, dato lo spazio  $(\Omega, \mathcal{A}, \mathbf{P})$  e una v.a.  $\xi$  definita su di esso, si può sempre definire un nuovo spazio di probabilità (questa volta numerico)  $(\mathcal{X}, \mathcal{B}, \mathbf{P}_\xi)$  mediante la costruzione qui descritta. In particolare  $\mathbf{P}_\xi\{x_k\} = \mathbf{P}\{\xi = x_k\}$ , ed è evidente che<sup>1</sup>  $\mathbf{P}_\xi(B) = \sum_{x_k \in B} \mathbf{P}_\xi\{x_k\}$ . Ciò mostra che la conoscenza dell'insieme di numeri  $(\mathbf{P}_\xi\{x_i\})_{i=1, \dots, k}$  è tutto ciò che è necessario per definire completamente la probabilità  $\mathbf{P}_\xi(\cdot)$ . Tale insieme di numeri prende il nome di **Distribuzione di Probabilità (DdP)** della v.a.  $\xi$ , ed è chiaro che essa dipende sia dalla probabilità  $\mathbf{P}(\cdot)$  definita su  $(\Omega, \mathcal{A})$  che dalla v.a.  $\xi(\omega)$  definita su  $\Omega$ .  $\circ$

**I.5.5 Esempio:** Riprendendo la v.a. definita nell'Esempio I.5.2, se supponiamo che la moneta sia equa, otteniamo immediatamente la seguente DdP:

$$\mathbf{P}_\xi\{0\} = \mathbf{P}\{CC\} = \frac{1}{4}, \quad \mathbf{P}_\xi\{1\} = \mathbf{P}\{TC, CT\} = \frac{1}{2}, \quad \mathbf{P}_\xi\{2\} = \mathbf{P}\{TT\} = \frac{1}{4};$$

mentre la DdP dell'indicatore dell'Esempio I.5.3 è

$$\mathbf{P}_{I_A}\{0\} = 1 - p, \quad \mathbf{P}_{I_A}\{1\} = p,$$

se  $p = \mathbf{P}(A)$ . Naturalmente il fatto che in questi esempi le nostre v.a. prendano solo valori interi è del tutto accidentale e dipende solo dal desiderio di semplificare le notazioni. Niente impedirebbe, ad esempio, di assegnare sullo spazio  $\Omega$

<sup>1</sup> Se si tiene presente che la funzione  $\xi(\cdot)$  è ad un sol valore, ci si rende subito conto del fatto che gli eventi  $\{\xi = x_j\}$  e  $\{\xi = x_k\}$  sono sempre disgiunti se  $j \neq k$  e quindi che, a causa dell'additività di  $\mathbf{P}(\cdot)$ , la probabilità  $\mathbf{P}_\xi(B)$  sarà somma di probabilità di eventi del tipo  $\{\xi = x_k\}$  dato che  $B$  è unione di tali eventi.

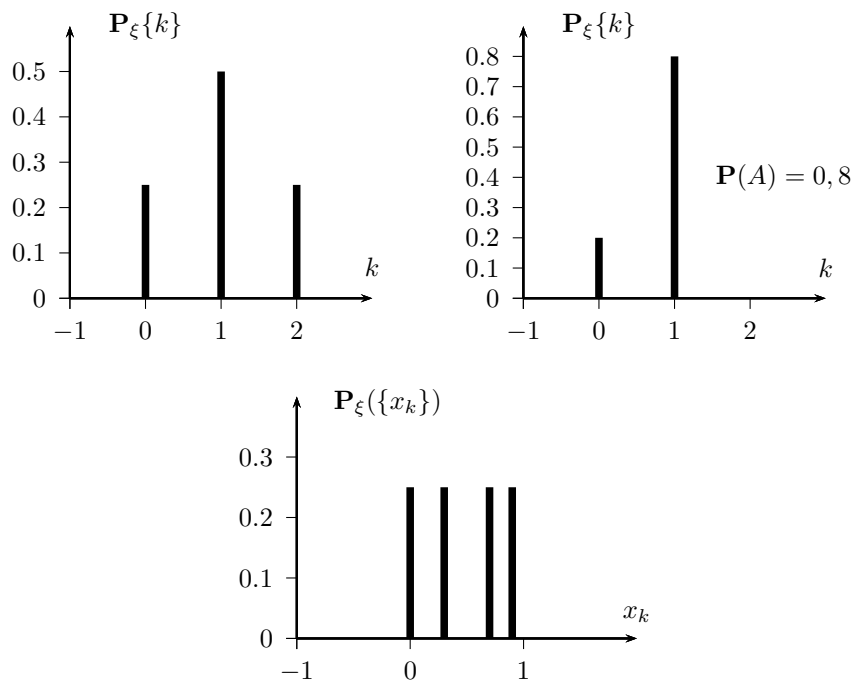


dell'Esempio I.5.2 una v.a.  $\eta$  che rappresenti (in centimetri) lo spostamento di un corpo puntiforme lungo una retta e che sia così definita:

$$\begin{aligned} \omega &= TT \quad TC \quad CT \quad CC \\ \eta(\omega) &= 0,9 \quad 0,7 \quad 0,3 \quad 0,0 \end{aligned}$$

in questo caso la nostra v.a. assumerebbe valori reali (non interi) e la sua DdP è facilmente calcolabile dato che i quattro valori possibili risulteranno tutti equiprobabili.  $\diamond$

**I.5.6 Osservazione:** A differenza di una probabilità  $\mathbf{P}$  su uno spazio astratto  $(\Omega, \mathcal{A})$ , una DdP  $(\mathbf{P}_\xi\{x_i\})_{i=1,\dots,k}$  definisce una associazione fra numeri visto che  $(\mathcal{X}, \mathcal{B})$  è uno spazio numerico. Ciò consente di dare per le DdP delle rappresentazioni grafiche che non sono in generale possibili per le probabilità  $\mathbf{P}$  su  $(\Omega, \mathcal{A})$ . Si veda la Figura I.5.1 per le rappresentazioni delle DdP descritte negli Esempi I.5.2, I.5.3 e I.5.5.  $\circ$



**Fig. I.5.1** Distribuzioni di probabilità per gli Esempi I.5.2, I.5.3 e I.5.5.

**I.5.7 Osservazione:** Gli indicatori  $I_A(\omega)$  con  $A \in \mathcal{A}$  sono v.a. che godono di un certo numero di semplici proprietà. In particolare le seguenti relazioni sono tutte

valide  $\forall \omega \in \Omega$ :

$$\begin{aligned} I_{\emptyset}(\omega) &= 0 \\ I_{\Omega}(\omega) &= 1 \\ I_A(\omega) + I_{\bar{A}}(\omega) &= 1 \\ I_{AB}(\omega) &= I_A(\omega) I_B(\omega) \\ I_{A \cup B}(\omega) &= I_A(\omega) + I_B(\omega) - I_A(\omega) I_B(\omega), \end{aligned}$$

e noi ne lasceremo la dimostrazione come esercizio per il lettore. ○

Abbiamo rilevato in Osservazione I.5.4 che lo spazio di probabilità  $(\mathcal{X}, \mathcal{B}, \mathbf{P}_{\xi})$  può essere definito assegnando una DdP  $(\mathbf{P}_{\xi}\{x_i\})_{i=1, \dots, k}$ : vedremo ora che la probabilità  $\mathbf{P}_{\xi}$  su  $(\mathcal{X}, \mathcal{B})$  può essere anche data mediante un altro strumento analitico che si rivelerà particolarmente importante nel seguito della nostra discussione.

**I.5.8 Definizione:** Assegnata su  $(\Omega, \mathcal{A}, \mathbf{P})$  una v.a.  $\xi$  a valori reali in  $(\mathcal{X}, \mathcal{B})$ , si chiama **Funzione di Distribuzione (FdD)** di  $\xi$  la funzione

$$F_{\xi}(x) = \mathbf{P}\{\omega \in \Omega : \xi(\omega) \leq x\} = \mathbf{P}\{\xi \leq x\} = \mathbf{P}\{\xi \in (-\infty, x]\} = \sum_{x_k \leq x} \mathbf{P}_{\xi}\{x_k\}$$

definita  $\forall x \in \mathbf{R}$ . △

**I.5.9 Osservazione:** Per come è definita, la FdD  $F_{\xi}$  di una v.a.  $\xi$  a valori in  $\mathcal{X} = \{x_1, \dots, x_m\}$  risulta essere una funzione monotona non decrescente<sup>2</sup> e limitata fra 0 ed 1 dato che:

$$\lim_{x \rightarrow -\infty} F_{\xi} = 0; \quad \lim_{x \rightarrow +\infty} F_{\xi} = 1.$$

Inoltre  $F_{\xi}$  è continua e costante su tutti gli intervalli del tipo  $(x_k, x_{k+1})$ , mentre presenta delle discontinuità di prima specie nei punti  $x_k$ ; più precisamente, siccome la definizione di FdD è data mediante eventi del tipo  $\{\xi \leq x\}$  e non del tipo  $\{\xi < x\}$ , la  $F_{\xi}$  risulta continua solo da destra nei punti  $x_k$ , cioè<sup>3</sup>

$$F_{\xi}(x_k^+) = F_{\xi}(x_k), \quad k = 1, \dots, m,$$

---

<sup>2</sup> Per convincersene basterà ricordare le proprietà generali di una probabilità (vedi Osservazione I.2.15) e notare che  $\{\xi \leq x'\} \subseteq \{\xi \leq x''\}$  quando  $x' \leq x''$ .

<sup>3</sup> Stabiliamo qui una volta per tutte le seguenti notazioni

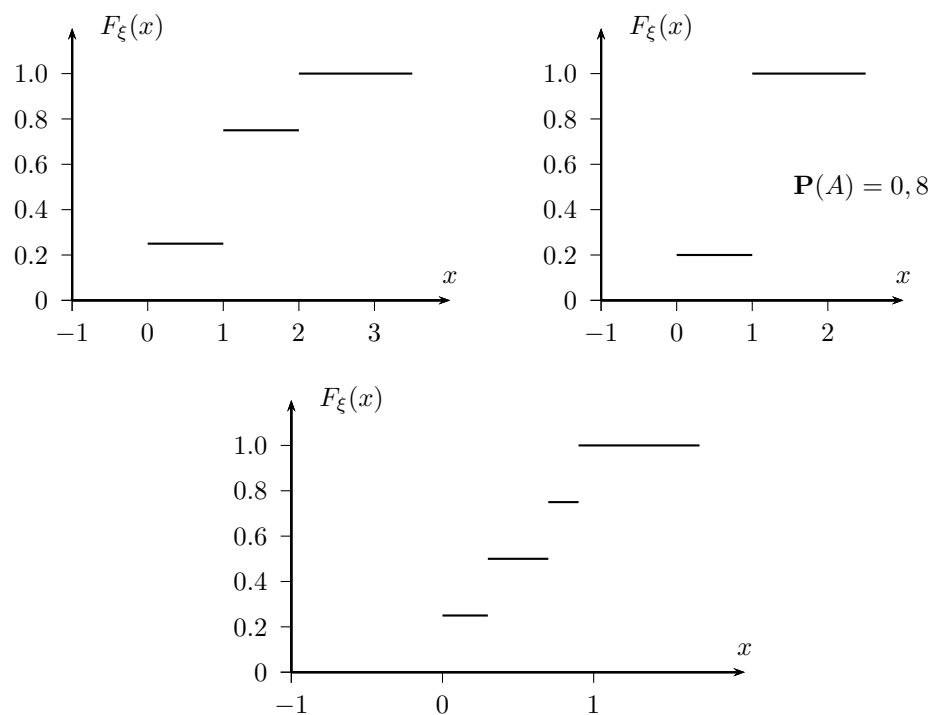
$$f(x^-) = \lim_{y \uparrow x} f(y) = \lim_{y \rightarrow x^-} f(y); \quad f(x^+) = \lim_{y \downarrow x} f(y) = \lim_{y \rightarrow x^+} f(y)$$

per i limiti rispettivamente da sinistra e da destra nel punto  $x$ .

mentre il limite da sinistra  $F_\xi(x_k^-)$  sicuramente esiste (dato che  $F_\xi$  è monotona e limitata) ma non coincide con  $F_\xi(x_k)$ . Infine è chiaro dalla definizione non solo che la conoscenza della DdP  $(\mathbf{P}_\xi\{x_i\})_{i=1,\dots,k}$  di  $\xi$  ci permette sempre di ricavare la FdD  $F_\xi$ , ma anche che, viceversa, la conoscenza della FdD  $F_\xi$  ci permette di ricostruire la DdP mediante la relazione

$$\mathbf{P}_\xi\{x_k\} = F_\xi(x_k^+) - F_\xi(x_k^-) = F_\xi(x_k) - F_\xi(x_k^-),$$

che può facilmente essere dedotta dalle osservazioni precedenti. In conclusione possiamo dire che il grafico della FdD  $F_\xi$  di una v.a. definita su uno spazio finito di probabilità (come vedremo, molte ma non tutte le proprietà delle FdD qui indicate potranno essere estese al caso più generale di spazi non finiti di probabilità) presenterà un caratteristico andamento *a scala* nell'intervallo fra  $x_1$  ed  $x_m$ ; sarà identicamente nulla per  $x < x_1$  ed identicamente eguale ad uno per  $x \geq x_m$ ; ed infine compirà dei *salto* nei punti  $x_k$  in modo tale che l'altezza di ogni *gradino* sia proprio eguale a  $\mathbf{P}_\xi\{x_k\}$ . Come esempio, sono riportati nella Figura I.5.2 i grafici delle FdD delle v.a. definite negli Esempi I.5.5, I.5.3 e I.5.5. ○



**Fig. I.5.2** Funzioni di distribuzione per gli Esempî I.5.2, I.5.3 e I.5.5.

Risulta naturale a questo punto pensare ad una prima generalizzazione dei concetti di v.a., di DdP e di FdD: è possibile infatti associare ad ogni  $\omega \in \Omega$  non un

solo numero reale, ma un intero *vettore* di numeri con due o più *componenti*. Le definizioni seguenti generalizzano le Definizioni I.5.1 e I.5.8 proprio in questa prospettiva.

**I.5.10 Definizione:** Chiameremo **Vettore Aleatorio (vett.a.)** ad  $r$  componenti una funzione  $\xi(\omega) = (\xi_1(\omega), \dots, \xi_r(\omega))$  che associ ad ogni  $\omega \in \Omega$  una  $r$ -pla di numeri reali  $x = (x_1, \dots, x_r)$ ; chiameremo inoltre **Distribuzione di Probabilità Congiunta** del vett.a.  $\xi$  l'insieme di numeri così definito:

$$\mathbf{P}_\xi\{x_1, \dots, x_r\} = \mathbf{P}\{\omega \in \Omega : \xi_1(\omega) = x_1, \dots, \xi_r(\omega) = x_r\},$$

mentre la funzione

$$F_\xi(x_1, \dots, x_r) = \mathbf{P}\{\omega \in \Omega : \xi_1(\omega) \leq x_1, \dots, \xi_r(\omega) \leq x_r\}$$

prenderà il nome di **Funzione di Distribuzione Congiunta** del dato vett.a.  $\Delta$

**I.5.11 Osservazione:** Se indichiamo con  $\mathcal{X}$  l'insieme delle  $r$ -ple  $x = (x_1, \dots, x_r)$  che sono possibili valori di un vett.a.  $\xi$ , è chiaro che ora  $\mathcal{X}$  risulta essere un insieme (finito) di punti di  $\mathbf{R}^r$ ; se poi indichiamo con  $\mathcal{B}$  un'algebra di parti di  $\mathcal{X}$  (quella di tutte le sue parti, se non diversamente precisato), potremo verificare che  $(\mathcal{X}, \mathcal{B}, \mathbf{P}_\xi)$  è ora un nuovo spazio di probabilità. Naturalmente, affinché tutta questa costruzione e la stessa Definizione I.5.10 abbiano senso, faremo anche qui (come in Osservazione I.5.4) l'ipotesi che l'algebra  $\mathcal{A}$  dello spazio  $(\Omega, \mathcal{A}, \mathbf{P})$  su cui il vett.a. è definito sia abbastanza grande da contenere tutti gli eventi del tipo  $\{\xi \in B\}$  con  $B \in \mathcal{B}$ . In queste condizioni è facile mostrare che ogni componente  $\xi_k$  di un vett.a.  $\xi$  è una v.a. che soddisfa le ipotesi di Osservazione I.5.4. Sarà bene infine notare che, se con  $\mathcal{X}_k$  indichiamo l'insieme dei possibili valori della componente  $\xi_k$  di un vett.a.  $\xi$ , e con  $\mathcal{B}_k$  la corrispondente algebra degli eventi, risulterà  $\mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_r$  e  $\mathcal{B} = \mathcal{B}_1 \otimes \dots \otimes \mathcal{B}_r$  secondo la notazione dell'Osservazione I.4.20. Con questa notazione, dato  $B = B_1 \times \dots \times B_r \in \mathcal{B}$ , l'evento

$$\{\xi \in B\} = \{\xi \in B_1 \times \dots \times B_r\} = \{\xi_1 \in B_1, \dots, \xi_r \in B_r\},$$

con  $B_k \in \mathcal{X}_k$ , coinciderà con l'intersezione  $\{\xi_1 \in B_1\} \cap \dots \cap \{\xi_r \in B_r\}$  di eventi  $\{\xi_k \in B_k\}$  che risultano tutti elementi della medesima algebra  $\mathcal{A}$  degli eventi di  $\Omega$ . ○

**I.5.12 Esempio:** Dato lo spazio dei campioni di due lanci di una moneta equa  $\Omega$  (vedi Esempi I.5.2 e I.5.5) definiamo i due eventi

$$A_1 = \text{appare } T \text{ almeno una volta} = \{TT, TC, CT\}$$

$$A_2 = \text{appare } C \text{ almeno una volta} = \{TC, CT, CC\}$$

e il vett.a. a due componenti  $\xi = (\xi_1, \xi_2)$ , con  $\xi_1 = I_{A_1}$  e  $\xi_2 = I_{A_2}$ . Siccome la moneta è equa,  $\mathbf{P}(A_1) = \mathbf{P}(A_2) = \frac{3}{4}$ , e quindi le DdP delle due componenti  $\xi_1$  e  $\xi_2$  sono identiche:

$$\mathbf{P}_{\xi_1}\{1\} = \mathbf{P}_{\xi_2}\{1\} = \frac{3}{4}, \quad \mathbf{P}_{\xi_1}\{0\} = \mathbf{P}_{\xi_2}\{0\} = \frac{1}{4}.$$

Per calcolare la DdP congiunta dobbiamo allora osservare che

$$\begin{aligned} \{\xi_1 = 1, \xi_2 = 1\} &= A_1 \cap A_2 = \{TC, CT\} \\ \{\xi_1 = 0, \xi_2 = 1\} &= \overline{A_1} \cap A_2 = \{CC\} \\ \{\xi_1 = 1, \xi_2 = 0\} &= A_1 \cap \overline{A_2} = \{TT\} \\ \{\xi_1 = 0, \xi_2 = 0\} &= \overline{A_1} \cap \overline{A_2} = \emptyset, \end{aligned}$$

per cui si ha:

$$\mathbf{P}_\xi\{1, 1\} = \frac{1}{2}, \quad \mathbf{P}_\xi\{1, 0\} = \mathbf{P}_\xi\{0, 1\} = \frac{1}{4}, \quad \mathbf{P}_\xi\{0, 0\} = 0.$$

È facile vedere che anche la FdD congiunta può essere calcolata sulla base di questi dati.  $\diamond$

**I.5.13 Osservazione:** Innanzitutto va notato che v.a. diverse (nel senso che sono definite da differenti funzioni da  $\Omega$  in  $\mathbf{R}$ ) come  $\xi_1$  e  $\xi_2$  possono avere la medesima DdP: pur essendo diverse in quanto v.a., esse sono identiche sotto il profilo delle loro proprietà statistiche. In tal caso si dice che  $\xi_1$  e  $\xi_2$  sono v.a. **identicamente distribuite (i.d.)**. Le DdP  $\mathbf{P}_{\xi_k}$  delle singole componenti  $\xi_k$  di un vett.a.  $\xi$  vengono chiamate **Distribuzioni di Probabilità Marginali** e non vanno confuse con la DdP congiunta  $\mathbf{P}_\xi$  anche se tra loro vi è una relazione: si può mostrare, infatti, con qualche esempio (ma lo faremo più tardi in forma generale) che, nota la DdP congiunta  $\mathbf{P}_\xi$ , è sempre possibile da questa ricavare le DdP marginali delle singole componenti nel modo seguente:

$$\mathbf{P}_{\xi_k}\{x_k\} = \sum_{x_j \in \mathcal{X}_j; j \neq k} \mathbf{P}_\xi\{x_1, \dots, x_k, \dots, x_r\};$$

basterà cioè sommare la DdP congiunta su tutti i valori possibili delle componenti diverse da  $\xi_k$  per ottenere la DdP marginale di questa componente. Viceversa non è in generale possibile ricostruire in modo unico la DdP congiunta di un vett.a. a partire dalla semplice conoscenza delle DdP marginali delle sue componenti: vi è infatti più informazione nella DdP congiunta di quanta non ve ne sia nelle DdP marginali, la differenza essendo sostanzialmente data dalla informazione circa le correlazioni statistiche fra le varie componenti. Solo in un caso questa ricostruzione è possibile in modo unico (come vedremo tra breve): il caso in cui le componenti del vett.a. si comportano in maniera *statisticamente indipendente*. Un altro modo per esprimere la stessa idea consiste nel notare che, sebbene risulti  $\mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_r$  e  $\mathcal{B} = \mathcal{B}_1 \otimes \dots \otimes \mathcal{B}_r$  (vedi Osservazione I.5.11), non possiamo però affermare che lo spazio di probabilità  $(\mathcal{X}, \mathcal{B}, \mathbf{P}_\xi)$  sia uno spazio *prodotto diretto* nel senso dell'Osservazione I.4.20 perché la probabilità  $\mathbf{P}_\xi$  non è, in generale, ricavabile da prodotti delle DdP marginali (come appunto nel caso dell'Osservazione I.4.20 e più in generale, come vedremo presto, nel caso in cui le componenti siano indipendenti). Queste osservazioni possono, naturalmente, essere verificate nel caso

dell'Esempio I.5.12 dove si vede facilmente che la DdP congiunta non si ricava per prodotto dalle DdP marginali  $\mathbf{P}_{\xi_1}$  e  $\mathbf{P}_{\xi_2}$ , mentre le semplici relazioni

$$\mathbf{P}_{\xi_1}\{0\} = \mathbf{P}_{\xi}\{0,0\} + \mathbf{P}_{\xi}\{0,1\} = \frac{1}{4}, \quad \mathbf{P}_{\xi_1}\{1\} = \mathbf{P}_{\xi}\{1,0\} + \mathbf{P}_{\xi}\{1,1\} = \frac{3}{4}$$

consentono, ad esempio, di ricavare dalla DdP congiunta la DdP marginale di  $\xi_1$ ; un analogo calcolo consente di calcolare anche la DdP di  $\xi_2$ .  $\circ$

**I.5.14 Definizione:** Dato un vett.a.  $\xi = (\xi_1, \dots, \xi_r)$  ad  $r$  componenti che assume valori in un insieme  $\mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_r \subseteq \mathbf{R}^r$  e detta  $\mathcal{B} = \mathcal{B}_1 \otimes \dots \otimes \mathcal{B}_r$  la corrispondente algebra degli eventi, diremo che le componenti di  $\xi$  sono **indipendenti** se la relazione

$$\mathbf{P}_{\xi}(B_1 \times \dots \times B_r) = \mathbf{P}_{\xi_1}(B_1) \cdot \dots \cdot \mathbf{P}_{\xi_r}(B_r)$$

è verificata  $\forall B_k \in \mathcal{B}_k$  con  $k = 1, \dots, r$ .  $\triangle$

**I.5.15 Osservazione:** Innanzitutto va osservato che, a causa del fatto che le componenti di un vett.a. sono v.a., il concetto di indipendenza delle componenti di un vett.a. ad  $r$  componenti si traduce immediatamente nel concetto di indipendenza per  $r$  v.a. Inoltre la definizione data è suscettibile di altre formulazioni che sono ad essa perfettamente equivalenti (anche se noi ci asterremo, a questo punto della trattazione, dal verificare tale equivalenza). Ad esempio l'indipendenza delle componenti secondo la Definizione I.5.14 equivale alla richiesta che la relazione

$$\mathbf{P}_{\xi}\{x_1, \dots, x_r\} = \mathbf{P}_{\xi_1}\{x_1\} \cdot \dots \cdot \mathbf{P}_{\xi_r}\{x_r\}$$

sia verificata  $\forall (x_1, \dots, x_r) \in \mathcal{X}$ . Ancora in maniera del tutto equivalente si potrebbe richiedere che la seguente proprietà di fattorizzazione della FdD congiunta

$$F_{\xi}(x_1, \dots, x_r) = F_{\xi_1}(x_1) \cdot \dots \cdot F_{\xi_r}(x_r)$$

sia verificata  $\forall (x_1, \dots, x_r) \in \mathbf{R}^r$ . Siccome poi

$$\mathbf{P}_{\xi}(B_1 \times \dots \times B_r) = \mathbf{P}\{\xi_1 \in B_1, \dots, \xi_r \in B_r\}$$

e  $\mathbf{P}_{\xi_k}(B_k) = \mathbf{P}\{\xi_k \in B_k\}$ , va anche notato che il concetto di indipendenza delle componenti di un vett.a. coincide con il concetto di indipendenza di tutti gli eventi del tipo  $\{\xi_k \in B_k\}$  nel senso della Definizione I.4.14, visto che (come notato in Osservazione I.5.11) si ha anche

$$\{\xi_1 \in B_1, \dots, \xi_r \in B_r\} = \{\xi_1 \in B_1\} \cap \dots \cap \{\xi_r \in B_r\}.$$

Naturalmente, come già annunciato in Osservazione I.5.13, quando le componenti di un vett.a. sono indipendenti la DdP congiunta è sempre ricostruibile a partire dalle DdP marginali mediante dei prodotti: ciò risulta evidente se si considera la relazione della Definizione I.5.14 come un modo per definire la DdP congiunta

$\mathbf{P}_\xi$  mediante prodotti di DdP marginali. In questo caso lo spazio di probabilità  $(\mathcal{X}, \mathcal{B}, \mathbf{P}_\xi)$  risulta essere il **prodotto diretto** (nel senso dell'Osservazione I.4.20) degli  $r$  spazi di probabilità  $(\mathcal{X}_k, \mathcal{B}_k, \mathbf{P}_{\xi_k})$ . Infine è notevole il caso in cui le componenti di un vett.a. sono **indipendenti ed identicamente distribuite (i.i.d.)** in quanto esso, come vedremo in I.5.17, sarà il modello principale per la descrizione di ripetizioni indipendenti del medesimo esperimento.  $\circ$

Un ulteriore allargamento del concetto di v.a. si ottiene se si introducono dei metodi per combinare delle v.a. in modo da formare delle nuove v.a.; questo è reso possibile dalla introduzione del concetto di **funzione di v.a.** o, più in generale **funzione di un vett.a.**: dato un vett.a.  $\xi$  con  $r$  componenti e data una funzione reale di  $r$  variabili

$$f : \mathbf{R}^r \rightarrow \mathbf{R}$$

è possibile definire una nuova v.a.  $\eta$  mediante la posizione

$$\eta(\omega) = f[\xi_1(\omega), \dots, \xi_r(\omega)].$$

Naturalmente  $\eta$  è una nuova v.a. nel senso che essa definisce una nuova funzione su  $\Omega$  a valori reali.

**I.5.16 Esempio:** Date due v.a.  $\xi$  ed  $\eta$  (o, equivalentemente, il vett.a. di componenti  $\xi$  ed  $\eta$ ) e la funzione  $z = f(x, y) = x + y$ , si definisca la v.a.  $\zeta = f(\xi, \eta) = \xi + \eta$ ; se  $\mathcal{X} = \{x_1, \dots, x_m\}$  e  $\mathcal{Y} = \{y_1, \dots, y_l\}$  sono gli insiemi di possibili valori di  $\xi$  ed  $\eta$ , i possibili valori di  $\zeta$  sono gli elementi dell'insieme

$$\mathcal{Z} = \{z \in \mathbf{R} : z = x_i + y_j; \ x_i \in \mathcal{X}, \ y_j \in \mathcal{Y}\}.$$

La DdP di  $\zeta$  può essere calcolata nel modo seguente: se  $z \in \mathcal{Z}$

$$\mathbf{P}_\zeta\{z\} = \mathbf{P}\{\zeta = z\} = \mathbf{P}\{\xi + \eta = z\} = \sum_{(i,j):x_i+y_j=z} \mathbf{P}\{\xi = x_i, \eta = y_j\};$$

se poi  $\xi$  ed  $\eta$  sono indipendenti si ha che

$$\begin{aligned} \mathbf{P}_\zeta\{z\} &= \sum_{(i,j):x_i+y_j=z} \mathbf{P}\{\xi = x_i\} \mathbf{P}\{\eta = y_j\} \\ &= \sum_{i=1}^m \mathbf{P}\{\xi = x_i\} \mathbf{P}\{\eta = z - x_i\} = \sum_{i=1}^m \mathbf{P}_\xi\{x_i\} \mathbf{P}_\eta\{z - x_i\}. \end{aligned}$$

dove ovviamente  $z \in \mathcal{Z}$ .  $\diamond$

**I.5.17 Osservazione:** Applicheremo ora i concetti introdotti in questo capitolo al Modello di Bernoulli discusso nell'Osservazione I.4.20. La Distribuzione Binomiale discussa nel Capitolo I.3 è la DdP di una particolare v.a. detta v.a. Binomiale. Per costruirla consideriamo innanzitutto un esperimento con due soli possibili

risultati: *successo* e *fallimento* (ad esempio  $T$  e  $C$  nel lancio di una moneta), come quelli considerati per costruire la Distribuzione Binomiale (I.3) ed il Modello di Bernoulli (I.4.20), e supponiamo che siano  $p$  la probabilità di successo e  $q = 1 - p$  la probabilità di fallimento in un tentativo. Se conveniamo di indicare con  $T$  il *successo* e con  $C$  il *fallimento*, potremo definire la v.a.  $\xi$  nel modo seguente:

$$\xi(\omega) = \begin{cases} 1, & \text{se } \omega = T, \\ 0, & \text{se } \omega = C. \end{cases}$$

Come è evidente si tratta di un particolare indicatore: quello relativo all'evento  $\{T\}$ . In tal caso si ha  $\mathcal{X} = \{0, 1\}$ , e la DdP di  $\xi$  è semplicemente

$$\mathbf{P}_\xi\{0\} = q = 1 - p, \quad \mathbf{P}_\xi\{1\} = p.$$

Notiamo che, se indichiamo con  $k$  i possibili valori (0 e 1) di  $\xi$ , la DdP di  $\xi$  ha la forma di una distribuzione Binomiale con  $n = 1$ :

$$\mathbf{P}_\xi\{k\} = p_1(k) = p^k q^{1-k}; \quad k = 0, 1.$$

Consideriamo ora il Modello di Bernoulli (I.4.20) composto da  $n$  realizzazioni successive del nostro esperimento; in tal caso, come è già noto, un risultato  $\omega$  sarà una particolare successione di successi e fallimenti alla quale viene assegnata la probabilità  $\mathbf{P}\{\omega\} = p^k q^{n-k}$  se con  $k$  indichiamo il numero dei successi registrati in  $\omega$ . È anche noto che all'evento

$$D_k = \text{si registrano } k \text{ successi su } n \text{ tentativi}$$

il nostro modello assegna la probabilità

$$\mathbf{P}(D_k) = \binom{n}{k} p^k q^{n-k}.$$

Sullo spazio delle  $n$ -ple di tentativi  $\omega \in \Omega$  è possibile ora definire una nuova v.a.:

$$S^{(n)}(\omega) = \text{numero dei successi registrati in } \omega$$

i cui valori sono ovviamente gli  $n + 1$  numeri interi  $k = 0, 1, \dots, n$ , sicché  $\mathcal{X} = \{0, 1, \dots, n\}$ . È chiaro inoltre dalle definizioni che  $D_k = \{S^{(n)} = k\}$  per cui la DdP di  $S^{(n)}$  è binomiale:

$$\mathbf{P}_{S^{(n)}}\{k\} = \mathbf{P}\{S^{(n)} = k\} = \mathbf{P}(D_k) = p_n(k) = \binom{n}{k} p^k q^{n-k}.$$

Il caso relativo ad un solo tentativo trattato prima è ovviamente un caso particolare dal momento che evidentemente  $\xi = S^{(1)}$  e che le due DdP binomiali in questo caso coincidono. Le v.a. che, come  $S^{(n)}$  (che rappresenta il numero di successi in



$n$  tentativi di verifica di un dato evento), hanno come DdP una Distribuzione Binomiale si chiamano allora **Variabili Aleatorie Binomiali** (su  $n$  tentativi), mentre le v.a.  $\xi$  (relative ad un solo tentativo) si usano chiamare **Variabili Aleatorie di Bernoulli**. Inoltre, nel caso della v.a. Binomiale da noi costruita in questo esempio, lo spazio di probabilità  $(\mathcal{X}, \mathcal{B}, \mathbf{P}_{S^{(n)}})$  (dove ovviamente  $\mathcal{B}$  è l'algebra di tutte le parti di  $\mathcal{X} = \{0, 1, \dots, n\}$ ) mostra nella sua struttura la dipendenza dall'indice  $n$  che, come indicato dall'Osservazione I.4.21, dovrebbe essere esplicita in tutti gli spazi prodotto diretto.

Osserviamo infine che le  $n$  v.a. di Bernoulli  $\xi_k = I_{A_k}$ , dove gli eventi  $A_k$  sono stati definiti in I.4.20, sono le componenti indipendenti di un vett.a.  $\xi = (\xi_1, \dots, \xi_n)$ : infatti abbiamo già mostrato in I.4.20 che gli eventi  $A_k$  e  $\bar{A}_k$  sono tutti indipendenti fra loro, e pertanto anche le componenti di  $\xi$  risulteranno indipendenti. A questo proposito vogliamo ora mostrare che la definizione di una v.a. binomiale può anche essere data senza passare attraverso la specificazione dettagliata dello spazio di probabilità su cui essa è definita, cioè senza attribuire preventivamente (come fatto in maniera, a prima vista, un po' arbitraria in I.3.2) un peso probabilistico ai singoli risultati  $\omega \in \Omega$ . Supponiamo ad esempio di sapere soltanto che su  $(\Omega, \mathcal{A}, \mathbf{P})$  è possibile definire due v.a. di Bernoulli  $\xi$  ed  $\eta$  che siano i.i.d. (Osservazione I.5.15) con  $\mathbf{P}\{\xi = 1\} = \mathbf{P}\{\eta = 1\} = p$ , e definiamo poi la v.a.  $\zeta = \xi + \eta$ . L'insieme dei possibili valori di  $\zeta$  è ovviamente  $\mathcal{Z} = \{0, 1, 2\}$  e, data l'indipendenza di  $\xi$  ed  $\eta$ , mediante i risultati dell'Esempio I.5.16, posto  $q = 1 - p$ , se ne può facilmente calcolare la DdP:

$$\begin{aligned} \mathbf{P}_\zeta\{0\} &= \mathbf{P}_\xi\{0\} \mathbf{P}_\eta\{0\} = q^2 \\ \mathbf{P}_\zeta\{1\} &= \mathbf{P}_\xi\{0\} \mathbf{P}_\eta\{1\} + \mathbf{P}_\xi\{1\} \mathbf{P}_\eta\{0\} = 2pq \\ \mathbf{P}_\zeta\{2\} &= \mathbf{P}_\xi\{1\} \mathbf{P}_\eta\{1\} = p^2. \end{aligned}$$

Si vede subito che si tratta di una DdP Binomiale, nel senso che

$$\mathbf{P}_\zeta\{k\} = \binom{2}{k} p^k q^{2-k}.$$

Questo risultato può essere generalizzato alla somma di  $n$  v.a. e fornisce quindi un'altra maniera di definire una v.a. Binomiale: se su  $(\Omega, \mathcal{A}, \mathbf{P})$  è possibile definire  $n$  v.a. di Bernoulli  $\xi_k$  con  $k = 1, \dots, n$  che siano i.i.d. (con  $\mathbf{P}\{\xi_k = 1\} = p$ ), la loro somma  $S^{(n)} = \xi_1 + \dots + \xi_n$  ha come DdP

$$\mathbf{P}_{S^{(n)}}\{k\} = \binom{n}{k} p^k (1-p)^{n-k},$$

cioè è una v.a. Binomiale. Naturalmente le  $\xi_k$  sono ora proprio le  $I_{A_k}$  introdotte prima. D'ora in poi questo sarà il modo in cui definiremo v.a. binomiali in un modello di Bernoulli. ○



## I.6 Valore d'Attesa e Varianza

Consideriamo una v.a.  $\xi$  definita su uno spazio finito di probabilità  $(\Omega, \mathcal{A}, \mathbf{P})$  i cui valori reali distinti sono  $\mathcal{X} = \{x_1, \dots, x_k\}$ , e definiamo i seguenti eventi:

$$A_i = \{\xi = x_i\} = \{\omega \in \Omega : \xi(\omega) = x_i\}, \quad i = 1, \dots, k.$$

Come già notato alla fine dell'Osservazione I.5.4, tali eventi sono tutti disgiunti. Inoltre, siccome  $\mathcal{X}$  esaurisce tutti i possibili valori di  $\xi$ , è anche chiaro che  $\bigcup_{i=1}^k A_i = \Omega$ , sicché gli eventi  $A_i$  potranno essere considerati come atomi di una decomposizione  $\mathcal{D} = \{A_1, \dots, A_k\}$  indotta su  $\Omega$  da  $\xi$ . Su ognuno degli atomi di  $\mathcal{D}$  la v.a.  $\xi$  assume un valore costante (e diverso da quello di tutti gli altri atomi). Sarà dunque possibile rappresentare la v.a.  $\xi$  mediante una combinazione di indicatori degli atomi di  $\mathcal{D}$  nel modo seguente:

$$\xi(\omega) = \sum_{i=1}^k x_i I_{A_i}(\omega).$$

Le v.a. che possono essere rappresentate in questo modo (cioè come combinazione lineare *finita* di indicatori) prendono il nome di **Funzioni Semplici** o di **Variabili Aleatorie Semplici**; le osservazioni precedenti mostrano che ogni v.a. definita su uno spazio *finito* di probabilità è una v.a. semplice. Va subito notato, però, che questo non è più vero non appena lo spazio  $(\Omega, \mathcal{A}, \mathbf{P})$  non è più uno spazio finito. Inoltre, per brevità, useremo spesso nel seguito la notazione  $p_i = \mathbf{P}(A_i) = \mathbf{P}\{\xi = x_i\} = \mathbf{P}_\xi\{x_i\}$ .

**I.6.1 Definizione:** Data una v.a.  $\xi$  su uno spazio finito  $(\Omega, \mathcal{A}, \mathbf{P})$  diremo che il numero

$$\mathbf{E} \xi = \sum_{i=1}^k x_i p_i$$

è il **Valore d'Attesa (VdA)** della v.a.  $\xi$ . △

**I.6.2 Osservazione:** Si vede facilmente che sono verificate le seguenti relazioni:

$$\mathbf{E} \xi = \sum_{i=1}^k x_i \mathbf{P}\{\xi = x_i\} = \sum_{i=1}^k x_i \mathbf{P}_\xi\{x_i\},$$

e che, volendo definire  $\mathbf{E}(\cdot)$  direttamente su  $(\Omega, \mathcal{A}, \mathbf{P})$ ,

$$\mathbf{E} \xi = \sum_{\omega \in \Omega} \xi(\omega) \mathbf{P}\{\omega\}.$$

In altri termini: il VdA può essere calcolato sia mediante la probabilità  $\mathbf{P}$  definita su  $(\Omega, \mathcal{A}, \mathbf{P})$  che mediante la DdP  $(\mathbf{P}_\xi\{x_i\})_{i=1, \dots, k}$ , cioè mediante la probabilità

$\mathbf{P}_\xi$  definita su  $(\mathcal{X}, \mathcal{B}, \mathbf{P}_\xi)$ . Anzi si può far vedere che il VdA può essere calcolato anche mediante la FdD  $F_\xi$ : infatti ricordando che (vedi Osservazione I.5.9)

$$\Delta F_\xi(x) = F_\xi(x) - F_\xi(x^-) = \begin{cases} 0, & \forall x \notin \mathcal{X} \\ \mathbf{P}_\xi\{x_i\}, & \forall x = x_i \in \mathcal{X} \end{cases}$$

si ricava subito la relazione

$$\mathbf{E} \xi = \sum_{i=1}^k x_i \Delta F_\xi(x_i).$$

Tutte queste eguaglianze sono particolarmente suggestive per la loro somiglianza con le somme che definiscono un *integrale di Lebesgue*: avremo modo di vedere, nel seguito, che non si tratta affatto di un'analogia casuale e che c'è una profonda connessione fra il concetto di VdA e quello di integrale.  $\circ$

**I.6.3. Osservazione:** La rappresentazione di una v.a. in termini di funzioni semplici non è unica. È importante, pertanto, mostrare che il VdA della v.a. non dipende dalla particolare rappresentazione scelta. La rappresentazione di  $\xi$  come v.a. semplice data prima si basava su una particolare decomposizione  $\mathcal{D}$ : è chiaro però che sarà sempre possibile trovare delle decomposizioni  $\mathcal{D}' = \{B_1, \dots, B_l\}$  più fini di  $\mathcal{D}$  ( $\mathcal{D} \preceq \mathcal{D}'$ ; vedi I.2.12) sulla base delle quali dare una diversa rappresentazione di  $\xi$  come v.a. semplice. Infatti sarà sufficiente decomporre ogni atomo  $A_i$  di  $\mathcal{D}$  in un certo numero finito di eventi disgiunti  $B_j$  per raggiungere lo scopo. Naturalmente sugli atomi  $B_j$  della nuova decomposizione  $\xi$  assumerà sempre un valore costante, ma ora potrà anche assumere lo stesso valore su atomi distinti (in particolare: su tutti gli atomi in cui abbiamo decomposto un  $A_i$ ). Potremo allora dire che i valori assunti da  $\xi$  sono ora i numeri  $x'_j$  ( $j = 1, \dots, l$ ), ma con la precisazione che questi non sono più tutti distinti. Naturalmente ciò non impedisce affatto di dare una nuova rappresentazione di  $\xi$  come funzione semplice:

$$\xi(\omega) = \sum_{j=1}^l x'_j I_{B_j}(\omega),$$

che è ovviamente diversa dalla precedente. Ciononostante, siccome

$$\sum_{j: x'_j = x_i} x'_j \mathbf{P}(B_j) = x_i \sum_{j: x'_j = x_i} \mathbf{P}(B_j) = x_i \mathbf{P}\left(\bigcup_{j: x'_j = x_i} B_j\right) = x_i \mathbf{P}(A_i),$$

risulterà anche

$$\mathbf{E} \xi = \sum_{j=1}^l x'_j \mathbf{P}(B_j) = \sum_{i=1}^k x_i \mathbf{P}(A_i),$$

cioè il VdA di  $\xi$  non cambierà se modifichiamo la rappresentazione di  $\xi$ .  $\circ$

**I.6.4 Proposizione:** Data una v.a.  $\xi \geq 0$ ,  $\forall \omega \in \Omega$  la relazione

$$\mathbf{E} \xi \geq 0$$

risulta sempre verificata.

**Dimostrazione:** La dimostrazione (banale) è lasciata al lettore.  $\square$

**I.6.5 Proposizione:** Date due v.a.  $\xi$  ed  $\eta$  e due numeri  $a$  e  $b$  risulta sempre

$$\mathbf{E}(a\xi + b\eta) = a \mathbf{E} \xi + b \mathbf{E} \eta,$$

cioè  $\mathbf{E}(\cdot)$  è un funzionale lineare.

**Dimostrazione:** Supponiamo di indicare con  $x_i$  ed  $y_j$  i valori rispettivamente di  $\xi$  ed  $\eta$  e di porre  $A_i = \{\xi = x_i\}$  e  $B_j = \{\eta = y_j\}$ : siccome dalle proprietà degli indicatori (Osservazione I.5.8) risulta

$$\begin{aligned} \sum_j I_{A_i B_j} &= \sum_j I_{A_i} I_{B_j} = I_{A_i} \sum_j I_{B_j} = I_{A_i} I_{\Omega} = I_{A_i} \\ \sum_i I_{A_i B_j} &= I_{B_j} \end{aligned}$$

avremo anche che

$$\begin{aligned} \xi &= \sum_i x_i I_{A_i} = \sum_{i,j} x_i I_{A_i B_j} \\ \eta &= \sum_j y_j I_{B_j} = \sum_{i,j} y_j I_{A_i B_j} \end{aligned}$$

Ne segue che

$$a\xi + b\eta = \sum_{i,j} (ax_i + by_j) I_{A_i B_j}$$

e quindi

$$\begin{aligned} \mathbf{E}(a\xi + b\eta) &= \sum_{i,j} (ax_i + by_j) \mathbf{P}(A_i B_j) \\ &= a \sum_i x_i \sum_j \mathbf{P}(A_i B_j) + b \sum_j y_j \sum_i \mathbf{P}(A_i B_j) \\ &= a \sum_i x_i \mathbf{P}(A_i) + b \sum_j y_j \mathbf{P}(B_j) = a \mathbf{E} \xi + b \mathbf{E} \eta \end{aligned}$$

come volevasi dimostrare.  $\square$

**I.6.6 Proposizione:** Date due v.a.  $\xi$  ed  $\eta$  risulta sempre

$$\mathbf{E} \xi \geq \mathbf{E} \eta$$

se  $\xi \geq \eta, \forall \omega \in \Omega$ .

**Dimostrazione:** Infatti se  $\xi - \eta \geq 0$ , dalle Proposizioni I.6.4 e I.6.5 si ha che

$$\mathbf{E} (\xi - \eta) = \mathbf{E} \xi - \mathbf{E} \eta \geq 0$$

da cui segue la tesi. □

**I.6.7 Proposizione:** La relazione

$$|\mathbf{E} \xi| \leq \mathbf{E} |\xi|$$

è verificata da qualunque v.a.  $\xi$ .

**Dimostrazione:** Infatti si ha

$$|\mathbf{E} \xi| = \left| \sum_i x_i \mathbf{P}(A_i) \right| \leq \sum_i |x_i| \mathbf{P}(A_i) = \mathbf{E} |\xi|$$

dove  $A_i = \{\xi = x_i\}$ . □

**I.6.8 Proposizione:** Se  $\xi$  ed  $\eta$  sono v.a. indipendenti si ha

$$\mathbf{E} (\xi \eta) = \mathbf{E} \xi \mathbf{E} \eta.$$

Questa relazione si generalizza facilmente a qualunque numero finito di v.a.

**Dimostrazione:** Infatti se le nostre v.a. sono indipendenti (vedi I.5.15) si avrà

$$\begin{aligned} \mathbf{E} (\xi \eta) &= \mathbf{E} \left[ \left( \sum_i x_i I_{A_i} \right) \left( \sum_j y_j I_{B_j} \right) \right] = \mathbf{E} \left( \sum_{i,j} x_i y_j I_{A_i B_j} \right) \\ &= \sum_{i,j} x_i y_j \mathbf{P}(A_i B_j) = \sum_{i,j} x_i y_j \mathbf{P}(A_i) \mathbf{P}(B_j) \\ &= \sum_i x_i \mathbf{P}(A_i) \sum_j y_j \mathbf{P}(B_j) = \mathbf{E} \xi \mathbf{E} \eta \end{aligned}$$

dove gli eventi  $A_i$  e  $B_j$  sono definiti al solito modo. □

**I.6.9 Proposizione:** Dato un evento  $A \in \mathcal{A}$  risulta sempre

$$\mathbf{E} I_A = \mathbf{P}(A)$$

se  $I_A$  è l'indicatore di  $A$ .

**Dimostrazione:** Banale. □

**I.6.10 Esempio:** Se  $\xi$  è una v.a. di Bernoulli (vedi Osservazione I.5.17) con  $\mathbf{P}\{\xi = 1\} = p$  risulta che

$$\mathbf{E} \xi = 1 \cdot p + 0 \cdot (1 - p) = p.$$

Se poi  $\xi_1, \dots, \xi_n$  sono  $n$  v.a. di Bernoulli, anche non indipendenti, ma i.d. con  $\mathbf{P}\{\xi_i = 1\} = p$  ( $i = 1, \dots, n$ ), risulta

$$\mathbf{E} S^{(n)} = \sum_{i=1}^n \mathbf{E} \xi_i = n p$$

dove  $S^{(n)} = \xi_1 + \dots + \xi_n$ . ◇

**I.6.11 Osservazione:** Data una v.a.  $\xi$ , detti  $A_i = \{\xi = x_i\}$  con  $i = 1, \dots, n$  gli atomi della decomposizione  $\mathcal{D}_\xi$  in modo che

$$\xi = \sum_{i=1}^n x_i I_{A_i},$$

e data una funzione  $f(\cdot)$ , definita sui punti  $x_i$  con valori  $y_j$ , consideriamo la nuova v.a.  $\eta(\omega) = f[\xi(\omega)]$ . Se  $B_j = \{\eta = y_j\}$  con  $j = 1, \dots, k$  sono gli atomi della decomposizione  $\mathcal{D}_\eta$ , si verifica subito che  $\mathcal{D}_\eta \preceq \mathcal{D}_\xi$ . Infatti, siccome può risultare  $f(x_i) = f(x_l) = y_j$  anche se  $x_i \neq x_l$ , in generale gli eventi  $B_j$  saranno delle unioni di eventi  $A_i$ . Potremo dunque sempre scrivere

$$\eta = f(\xi) = \sum_{j=1}^k y_j I_{B_j} = \sum_{i=1}^n f(x_i) I_{A_i},$$

con la precisazione che mentre le  $y_j$  sono tutte diverse, alcune delle  $f(x_i)$  possono coincidere. Pertanto le seguenti espressioni per i VdA

$$\mathbf{E} \eta = \mathbf{E} f(\xi) = \begin{cases} \sum_{j=1}^k y_j \mathbf{P}(B_j) = \sum_{j=1}^k y_j \mathbf{P}_\eta\{y_j\} = \sum_{j=1}^k y_j \Delta F_\eta(y_j) \\ \sum_{i=1}^n f(x_i) \mathbf{P}(A_i) = \sum_{i=1}^n f(x_i) \mathbf{P}_\xi\{x_i\} = \sum_{i=1}^n f(x_i) \Delta F_\xi(x_i) \end{cases}$$

sono tutte equivalenti. ○

**I.6.12 Proposizione:** Se una v.a.  $\xi \geq 0$  ha VdA nullo ( $\mathbf{E}\xi = 0$ ), allora essa non può che essere la v.a. identicamente nulla:  $\xi(\omega) = 0, \forall \omega \in \Omega$ .

**Dimostrazione:** Infatti per ipotesi i valori di  $\xi$  sono tutti non negativi, cioè  $x_i \geq 0, i = 1, \dots, n$ ; pertanto, se

$$\mathbf{E}\xi = \sum_{i=1}^n x_i \mathbf{P}\{\xi = x_i\} = 0,$$

non può che risultare<sup>1</sup>  $x_i = 0, i = 1, \dots, n$ . □

**I.6.13 Proposizione (Diseguaglianza di Schwarz):** La diseguaglianza

$$(\mathbf{E}|\xi\eta|)^2 \leq \mathbf{E}\xi^2 \mathbf{E}\eta^2$$

è verificata da qualunque coppia di v.a.  $\xi$  ed  $\eta$ .

**Dimostrazione:** Consideriamo dapprima il caso in cui  $\mathbf{E}\xi^2 \neq 0$  e  $\mathbf{E}\eta^2 \neq 0$  e poniamo

$$\tilde{\xi} = \frac{\xi}{\sqrt{\mathbf{E}\xi^2}}, \quad \tilde{\eta} = \frac{\eta}{\sqrt{\mathbf{E}\eta^2}},$$

in modo che  $\mathbf{E}\tilde{\xi}^2 = 1$  e  $\mathbf{E}\tilde{\eta}^2 = 1$ . Siccome inoltre  $(|\tilde{\xi}| - |\tilde{\eta}|)^2 \geq 0$ , e quindi  $2|\tilde{\xi}\tilde{\eta}| \leq \tilde{\xi}^2 + \tilde{\eta}^2$ , abbiamo che  $2\mathbf{E}|\tilde{\xi}\tilde{\eta}| \leq \mathbf{E}\tilde{\xi}^2 + \mathbf{E}\tilde{\eta}^2 = 2$ , ossia  $\mathbf{E}|\tilde{\xi}\tilde{\eta}| \leq 1$  da cui si ricava immediatamente la tesi. Se invece almeno uno dei due VdA è nullo, poniamo, per fissare la idee,  $\mathbf{E}\xi^2 = 0$ , dalla Proposizione I.6.12 segue che l'unico valore che  $\xi$  può assumere con probabilità diversa da zero è  $x_i = 0$ . Ma allora è evidente che anche  $\mathbf{E}|\xi\eta| = 0$  e la tesi è verificata anche in questo caso<sup>2</sup>. □

<sup>1</sup> Naturalmente potrebbe risultare  $x_i \neq 0$  qualora fosse verificata la relazione  $\mathbf{P}\{\xi = x_i\} = 0$ . Ciò significa solo che l'enunciato della presente proposizione andrebbe modificato dicendo che  $\xi$  può assumere valori diversi da zero solo con probabilità nulla. A proposito dei valori di una v.a. assunti con probabilità nulla, vedi le analoghe considerazioni fatte nell'Esempio I.4.19.

<sup>2</sup> La dimostrazione di questa diseguaglianza è rimasta legata, per la maggior parte degli studiosi moderni, al nome di Hermann Amandus Schwarz (1843 - 1921), un allievo di Karl Weierstraß a Berlino che diede contributi essenziali alla sistemazione dell'analisi ed alle sue applicazioni alla geometria (vedi ad esempio **F. Riesz e B.Sz.-Nagy: Functional Analysis**; Dover, New York, 1990). Ciononostante è giusto ricordare che, in forme diverse, questo risultato era stato stabilito in precedenza da altri ricercatori. In particolare ricorderemo V. J. Bunyakovskii (1804 - 1889), fondatore della scuola russa di calcolo delle probabilità al cui nome la nostra diseguaglianza è legata per molti studiosi di origine orientale, ed A. L. Cauchy (1789 - 1857) del quale avremo modo di riparlare e che resta lo scopritore iniziale della nostra proposizione per gli studiosi francofoni.



**I.6.14 Osservazione:** È facile accorgersi dagli esempi fatti (vedi Esempio I.6.10) che il VdA di una v.a. può benissimo non essere uno dei suoi possibili valori: nel caso di una v.a. di Bernoulli (che assume valori 0 ed 1), ad esempio, il VdA è  $p$  che in generale non è né 0, né 1. Il VdA di una v.a., in effetti, è una misura (data mediante un solo numero) della collocazione del *baricentro* della DdP  $(\mathbf{P}_\xi\{x_i\})_{i=1,\dots,k}$  di  $\xi$ . Va pertanto fatta attenzione a non confondere il VdA con il valore di  $\xi$  in cui la DdP assume il suo massimo valore: il VdA non è il *valore più probabile* di  $\xi$ , non solo perché esso può non essere affatto uno dei possibili valori di  $\xi$ , ma anche perché la DdP può essere *non simmetrica* e quindi può avere il suo massimo ed il suo baricentro in collocazioni diverse. Naturalmente tutta l'informazione contenuta nella DdP non può essere concentrata in un solo numero (come il VdA): conoscere il VdA di una v.a. vuol dire conoscere un parametro importante della sua DdP, ma questo non esaurisce l'informazione in essa contenuta. Un altro parametro rilevante della v.a.  $\xi$  sarebbe una misura della *dispersione* con cui i valori della sua DdP si distribuiscono attorno al VdA: ovviamente, ancora una volta, tutti i dettagli di una DdP non potranno essere contenuti in due soli numeri, ma la conoscenza della dispersione sicuramente può giocare un ruolo fondamentale nella descrizione del comportamento dettagliato della v.a. data.  $\circ$

**I.6.15 Definizione:** Si dice **Varianza (Var)** della v.a.  $\xi$  il numero

$$\mathbf{V}\xi = \mathbf{E}(\xi - \mathbf{E}\xi)^2,$$

mentre la sua radice quadrata

$$\sigma_\xi = +\sqrt{\mathbf{V}\xi}$$

prende il nome di **Deviazione standard**. Una v.a. con Var nulla ( $\mathbf{V}\xi = 0$ ) viene detta anche v.a. **degenere**.  $\triangle$

**I.6.16 Osservazione:** La v.a.  $\xi - \mathbf{E}\xi$  indica la *deviazione* dei valori di  $\xi$  dal suo VdA  $\mathbf{E}\xi$ . Naturalmente tali deviazioni possono essere sia positive che negative, anzi si può vedere subito che esse, in media, sono nulle:

$$\mathbf{E}(\xi - \mathbf{E}\xi) = \mathbf{E}\xi - \mathbf{E}\xi = 0.$$

Questa proprietà (che dipende solo dalla definizione di VdA) è ovviamente verificata per qualunque v.a. e pertanto la v.a.  $\xi - \mathbf{E}\xi$  non può essere assunta come una buona misura della dispersione della DdP di  $\xi$ . Viceversa la v.a.  $(\xi - \mathbf{E}\xi)^2$  assume solo valori positivi e il suo VdA è, in generale, diverso da zero: per questo motivo tale VdA viene assunto come misura della dispersione della DdP di  $\xi$  attorno al suo VdA. Naturalmente sarebbero possibili altre scelte, come ad esempio  $\mathbf{E}|\xi - \mathbf{E}\xi|$ , ma sarà chiaro nel seguito che la scelta della Var quale indice di dispersione è particolarmente appropriata.  $\circ$

**I.6.17 Proposizione:** La relazione

$$\mathbf{V}\xi = \mathbf{E}\xi^2 - (\mathbf{E}\xi)^2$$

è verificata da ogni v.a.  $\xi$ .

**Dimostrazione:** Si ha infatti:

$$\begin{aligned}\mathbf{V}\xi &= \mathbf{E}(\xi - \mathbf{E}\xi)^2 = \mathbf{E}(\xi^2 - 2\xi\mathbf{E}\xi + (\mathbf{E}\xi)^2) \\ &= \mathbf{E}\xi^2 - 2(\mathbf{E}\xi)^2 + (\mathbf{E}\xi)^2 = \mathbf{E}\xi^2 - (\mathbf{E}\xi)^2,\end{aligned}$$

il che prova la tesi.  $\square$

**I.6.18 Proposizione:** Una v.a. degenera ( $\mathbf{V}\xi = 0$ ) è identicamente eguale al suo VdA, cioè

$$\xi(\omega) = \mathbf{E}\xi, \quad \forall \omega \in \Omega,$$

e si comporta quindi come un numero.

**Dimostrazione:** Infatti, siccome  $(\xi - \mathbf{E}\xi)^2 \geq 0$ , il risultato segue direttamente dalla Proposizione I.6.12.  $\square$

**I.6.19 Proposizione:** Data una v.a.  $\xi$ , le seguenti relazioni:

$$\begin{aligned}\mathbf{V}a &= 0 \\ \mathbf{V}(b\xi) &= b^2\mathbf{V}\xi \\ \mathbf{V}(a + b\xi) &= b^2\mathbf{V}\xi\end{aligned}$$

sono sempre verificate se  $a$  e  $b$  sono due numeri arbitrari.

**Dimostrazione:** Se  $a$  è un numero, esso ovviamente coincide con il suo VdA ( $\mathbf{E}a = a$ ) e quindi  $\mathbf{V}a = 0$ . Inoltre

$$\mathbf{V}(b\xi) = \mathbf{E}[b\xi - \mathbf{E}(b\xi)]^2 = b^2\mathbf{E}(\xi - \mathbf{E}\xi)^2 = b^2\mathbf{V}\xi.$$

La terza relazione segue facilmente dalle prime due.  $\square$

**I.6.20 Definizione:** Date due v.a.  $\xi$  ed  $\eta$  si chiama **Covarianza** il numero

$$\text{cov}(\xi, \eta) = \mathbf{E}[(\xi - \mathbf{E}\xi)(\eta - \mathbf{E}\eta)] = \mathbf{E}(\xi, \eta) - \mathbf{E}\xi\mathbf{E}\eta,$$

e (se  $\mathbf{V}\xi \neq 0$ ,  $\mathbf{V}\eta \neq 0$ ) **Coefficiente di Correlazione** il numero

$$\rho(\xi, \eta) = \frac{\mathbf{cov}(\xi, \eta)}{\sigma_\xi \sigma_\eta} = \frac{\mathbf{E}[(\xi - \mathbf{E}\xi)(\eta - \mathbf{E}\eta)]}{\sqrt{\mathbf{V}\xi} \sqrt{\mathbf{V}\eta}}.$$

Naturalmente risulta  $\mathbf{V}\xi = \mathbf{cov}(\xi, \xi)$ . △

**I.6.21 Proposizione:** Comunque assegnate due v.a.  $\xi$  e  $\eta$ , risulta sempre  $|\rho(\xi, \eta)| \leq 1$ . Inoltre  $|\rho(\xi, \eta)| = 1$  se e solo se esistono due numeri  $a \neq 0$  e  $b$  tali che  $\eta = a\xi + b$ ; in particolare risulta  $a > 0$  se  $\rho(\xi, \eta) = +1$  e  $a < 0$  se  $\rho(\xi, \eta) = -1$ .

**Dimostrazione:** La dimostrazione è lasciata per esercizio al lettore. □

**I.6.22 Proposizione:** Se  $\xi$  e  $\eta$  sono due v.a. indipendenti si ha che

$$\mathbf{cov}(\xi, \eta) = 0, \quad \mathbf{V}(\xi + \eta) = \mathbf{V}\xi + \mathbf{V}\eta.$$

Queste relazioni si generalizzano facilmente al caso di un numero arbitrario, ma finito, di v.a.

**Dimostrazione:** Si ha infatti

$$\begin{aligned} \mathbf{V}(\xi + \eta) &= \mathbf{E}[(\xi + \eta) - \mathbf{E}(\xi + \eta)]^2 = \mathbf{E}[(\xi - \mathbf{E}\xi) + (\eta - \mathbf{E}\eta)]^2 \\ &= \mathbf{V}\xi + \mathbf{V}\eta + 2\mathbf{E}[(\xi - \mathbf{E}\xi)(\eta - \mathbf{E}\eta)] = \mathbf{V}\xi + \mathbf{V}\eta + 2\mathbf{cov}(\xi, \eta); \end{aligned}$$

ma a causa dell'indipendenza di  $\xi$  ed  $\eta$  si ha anche

$$\mathbf{cov}(\xi, \eta) = \mathbf{E}[(\xi - \mathbf{E}\xi)(\eta - \mathbf{E}\eta)] = \mathbf{E}(\xi - \mathbf{E}\xi) \mathbf{E}(\eta - \mathbf{E}\eta) = 0,$$

da cui segue immediatamente che  $\mathbf{V}(\xi + \eta) = \mathbf{V}\xi + \mathbf{V}\eta$ . □

**I.6.23 Osservazione:** Quando risulta  $\mathbf{cov}(\xi, \eta) = 0$ , si dice che le v.a.  $\xi$  ed  $\eta$  sono **non correlate**. La Proposizione I.6.22 mostra che *se due v.a. sono indipendenti esse sono anche non correlate*. È importante però osservare che il viceversa non è vero: due v.a. possono non essere indipendenti pur essendo non correlate. Ne segue anche che l'ipotesi di indipendenza di  $\xi$  ed  $\eta$  non è strettamente necessaria per garantire che  $\mathbf{V}(\xi + \eta) = \mathbf{V}\xi + \mathbf{V}\eta$ : uno sguardo alla dimostrazione mostra infatti che è sufficiente fare l'ipotesi che  $\xi$  ed  $\eta$  siano non correlate. ○

**I.6.24 Esempio:** Sia  $\alpha$  una v.a. che assume tre soli valori:  $0, \frac{\pi}{2}, \pi$  con egual probabilità  $\frac{1}{3}$ , e siano definite le v.a.  $\xi = \sin \alpha$  e  $\eta = \cos \alpha$ . Si verifica subito che

$\xi$  ed  $\eta$  sono non correlate; infatti

$$\begin{aligned}\mathbf{E} \xi &= \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 0 = \frac{1}{3} \\ \mathbf{E} \eta &= \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot (-1) = 0 \\ \mathbf{E}(\xi \eta) &= \frac{1}{3} \cdot (1 \cdot 0) + \frac{1}{3} \cdot (0 \cdot 1) + \frac{1}{3} \cdot (-1 \cdot 0) = 0\end{aligned}$$

e quindi  $\mathbf{E} \xi \mathbf{E} \eta = 0 = \mathbf{E}(\xi \eta)$ . Ma si verifica altrettanto facilmente che esse non sono indipendenti dato che ad esempio

$$\mathbf{P}\{\xi = 1\} = \frac{1}{3}, \quad \mathbf{P}\{\eta = 1\} = \frac{1}{3}, \quad \mathbf{P}\{\xi = 1, \eta = 1\} = 0,$$

e quindi  $\mathbf{P}\{\xi = 1, \eta = 1\} = 0 \neq \frac{1}{9} = \mathbf{P}\{\xi = 1\} \mathbf{P}\{\eta = 1\}$ . ◇

**I.6.25 Esempio:** Sia  $\xi$  una v.a. di Bernoulli con  $\mathbf{P}\{\xi = 1\} = p$ : si vede subito che, essendo  $\mathbf{E} \xi = p$  (vedi Esempio I.6.10)

$$\mathbf{V} \xi = \mathbf{E}(\xi - \mathbf{E} \xi)^2 = (1-p)^2 p + (0-p)^2 (1-p) = p(1-p).$$

Se allora  $\xi_1, \dots, \xi_n$  sono  $n$  v.a. di Bernoulli i.i.d. e se  $S^{(n)} = \xi_1 + \dots + \xi_n$ , si ha, a causa dell'indipendenza delle  $\xi_i$ , che

$$\mathbf{V} S^{(n)} = n p (1-p).$$

In particolare, al variare di  $p$ ,  $\mathbf{V} S^{(n)}$  assume valore massimo per  $p = \frac{1}{2}$  e vale  $\mathbf{V} S^{(n)} = \frac{n}{4}$ ; viceversa  $\mathbf{V} S^{(n)}$  diviene piccola quando  $p$  è prossimo ad 1 o a 0. Ad esempio, supponendo che  $n = 10$ , si ha

$p$	$\mathbf{E} S^{(10)}$	$\mathbf{V} S^{(10)}$
0,5	5	2,5
0,9	9	0,9
0,99	9,9	0,099

Questo esempio mostra, tra l'altro, la connessione che vi è tra il concetto di Var e quello di dispersione di una v.a. attorno al suo VdA. Infatti se le nostre 10 v.a. di Bernoulli sono distribuite con  $p = 0,5$ , ci attendiamo che  $\mathbf{E} S^{(10)} = 5$ , ma in questo caso l'incertezza sul risultato della misura di ogni v.a. è massima (può essere 0 oppure 1 con la stessa probabilità) e corrispondentemente  $\mathbf{V} S^{(10)} = 2,5$ . Se invece  $p = 0,9$  avremo  $\mathbf{E} S^{(10)} = 9$  con  $\mathbf{V} S^{(10)} = 0,9$  corrispondente al fatto che ora l'incertezza sul risultato di ogni misura è molto minore visto che in ciascun tentativo otterremo 1 con il 90% di probabilità. ◇

## I.7 Modello di Bernoulli e Legge dei grandi numeri

Abbiamo visto come le v.a. binomiali relative a modelli di Bernoulli possano essere sempre costruite come somme di v.a. di Bernoulli i.i.d. senza specificare i dettagli dello spazio di probabilità su cui esse sono definite (Osservazione I.5.17). Dobbiamo però anche ricordare che, al livello attuale della nostra discussione, tale spazio di probabilità non può che essere uno spazio finito; inoltre, come visto in Osservazione I.4.20, esso può essere sempre costruito come spazio prodotto diretto degli spazi di probabilità nei quali descriviamo i singoli tentativi di verifica del nostro modello di Bernoulli. Sarà quindi utile per gli sviluppi futuri riprendere alcune caratteristiche di questi spazi prodotto diretto indicando comunque in modo esplicito la dipendenza del nostro spazio di probabilità dal numero  $n$  dei tentativi di verifica.

**I.7.1 Osservazione:** Riprendiamo la definizione di Modello di Bernoulli come spazio prodotto diretto (Osservazione I.4.20) e stabiliamo la seguente notazione: siano  $(\Omega_j, \mathcal{A}_j, \mathbf{P}_j)_{j=1, \dots, n}$   $n$  copie del medesimo spazio di probabilità, dove  $\mathcal{A}_j = \{A_j, \overline{A}_j, \Omega_j, \emptyset_j\}$  è l'algebra degli eventi considerati, con  $A_j$  sottoinsieme di  $\Omega_j$ , e  $\mathbf{P}_j(A_j) = p$ . Conveniamo inoltre di indicare con  $B_j$  il generico evento di  $\mathcal{A}_j$ . Gli spazi così definiti descrivono ciascuno un tentativo di verifica di un dato evento (gli  $A_j$  sono  $n$  copie identiche dello stesso evento) e l'indice  $j$  ne individua solo la collocazione fra gli  $n$  tentativi eseguiti. Supponendo ora di eseguire  $n$  ripetizioni indipendenti dell'esperimento, lo spazio di probabilità adeguato a descriverle sarà lo spazio prodotto diretto degli  $n$  spazi dati. Pertanto  $\Omega^{(n)} = \Omega_1 \times \dots \times \Omega_n$  sarà lo spazio dei campioni che descrive gli  $n$  tentativi (cioè: un elemento di  $\Omega^{(n)}$  è una  $n$ -pla di risultati dei singoli tentativi) e  $R^{(n)} = B_1 \times \dots \times B_n$  saranno le parti di  $\Omega^{(n)}$  ottenute come prodotti cartesiani (**rettangoli**) di eventi relativi ai singoli tentativi<sup>1</sup>. Tali rettangoli sono costituiti da prodotti cartesiani dei simboli  $A_j, \overline{A}_j, \Omega_j, \emptyset_j$  (ad esempio:  $A_1 \times \Omega_2 \times \overline{A}_3 \times \dots \times A_{n-1} \times \Omega_n$ ) ciascuno dei quali indica il risultato del corrispondente tentativo<sup>2</sup>. Siccome però l'insieme di tali rettangoli non è un'algebra (le unioni di rettangoli non sono, in generale, dei rettangoli) indicheremo con  $\mathcal{A}^{(n)}$  l'algebra ottenuta eseguendo in tutti i modi possibili le unioni dei suddetti rettangoli e la considereremo come algebra dei nostri eventi. È facile inoltre verificare che ogni evento  $A^{(n)} \in \mathcal{A}^{(n)}$  può sempre essere decomposto nella unione di rettangoli disgiunti, cioè  $A^{(n)} = \bigcup_{i=1}^m R_i^{(n)}$ , con  $R_i^{(n)} \cap R_l^{(n)} = \emptyset$  se  $i \neq l$ . Se inoltre indichiamo con  $B_j^{(n)} = \Omega_1 \times \dots \times B_j \times \dots \times \Omega_n$  l'evento consistente

---

<sup>1</sup> Come a questo punto sarà già chiaro, da ora in poi con l'indice  $(n)$  indicheremo tutte le quantità che si riferiscono alla  $n$ -pla di tentativi di verifica del dato evento.

<sup>2</sup> Se compare  $A_j$  vuol dire che l'evento si è verificato al  $j$ -mo tentativo; se compare  $\overline{A}_j$  l'evento non si è verificato al  $j$ -mo tentativo; se compare  $\Omega_j$  semplicemente non abbiamo registrato i risultati del  $j$ -mo tentativo o non abbiamo neanche eseguito la prova; se infine compare un  $\emptyset_j$ , il nostro evento coincide con l'evento assurdo  $\emptyset^{(n)}$  dell'esperimento complessivo.

nell'esecuzione della verifica solo al  $j$ -mo tentativo, si prova facilmente che  $R^{(n)} = B_1 \times \dots \times B_n = B_1^{(n)} \cap \dots \cap B_n^{(n)}$ . Questa osservazione, assieme all'ipotesi che gli  $n$  tentativi di verifica sono indipendenti, ci servirà ora per definire la probabilità  $\mathbf{P}^{(n)}$  sullo spazio prodotto. Infatti, supponendo che  $\mathbf{P}^{(n)}(B_j^{(n)}) = \mathbf{P}_j(B_j)$  (dato che si tratta della medesima probabilità ambientata in due spazi diversi<sup>3</sup>), definiremo la probabilità dei rettangoli come

$$\begin{aligned} \mathbf{P}^{(n)}(R^{(n)}) &= \mathbf{P}^{(n)}(B_1^{(n)} \cap \dots \cap B_n^{(n)}) \\ &= \mathbf{P}^{(n)}(B_1^{(n)}) \cdot \dots \cdot \mathbf{P}^{(n)}(B_n^{(n)}) = \mathbf{P}_1(B_1) \cdot \dots \cdot \mathbf{P}_n(B_n); \end{aligned}$$

e quella degli altri eventi ottenuti da unioni di rettangoli disgiunti come:

$$\mathbf{P}^{(n)}(A^{(n)}) = \mathbf{P}^{(n)}\left(\bigcup_{i=1}^m R_i^{(n)}\right) = \sum_{i=1}^m \mathbf{P}^{(n)}(R_i^{(n)}).$$

In questo modo lo spazio  $(\Omega^{(n)}, \mathcal{A}^{(n)}, \mathbf{P}^{(n)})$  rimane completamente definito a partire dagli  $n$  spazi  $(\Omega_j, \mathcal{A}_j, \mathbf{P}_j)$ .

Le v.a. i.i.d. di Bernoulli  $\xi_j^{(n)}$  ( $j = 1, \dots, n$ ) che serviranno a costruire la nostra v.a. binomiale saranno allora definite su  $(\Omega^{(n)}, \mathcal{A}^{(n)}, \mathbf{P}^{(n)})$  come indicatori degli eventi  $A_j^{(n)}$ , cioè:  $\xi_j^{(n)} = I_{A_j^{(n)}}$ ; tali v.a. assumono valore 1 se si registra un successo al  $j$ -mo tentativo e 0 in caso contrario. Detto allora  $\mathbf{E}^{(n)}(\cdot)$  il VdA nello spazio delle  $n$ -ple di tentativi, si verifica facilmente che

$$\mathbf{E}^{(n)}\xi_j^{(n)} = \mathbf{P}^{(n)}(A_j^{(n)}) = \mathbf{P}_j(A_j) = p, \quad j = 1, \dots, n.$$

e che la v.a.

$$S^{(n)} = \sum_{j=1}^n \xi_j^{(n)}$$

che può assumere solo i valori interi  $k = 0, 1, \dots, n$ , rappresenta il *numero dei successi registrati negli  $n$  tentativi eseguiti*. ○

**I.7.2 Osservazione:** Sarà interessante notare che, dalla Proposizione I.6.5, si prova facilmente che

$$\mathbf{E}^{(n)}S^{(n)} = \sum_{j=1}^n \mathbf{E}^{(n)}\xi_j^{(n)} = np$$

---

<sup>3</sup> Nota che, mentre  $B_j$  è un evento in  $\Omega_j$  (cioè nello spazio dei campioni di un solo tentativo),  $B_j^{(n)}$  è un evento in  $\Omega^{(n)}$  (cioè nello spazio dei campioni delle  $n$ -ple di tentativi): essi quindi rappresentano la medesima affermazione (ad esempio, quando  $B_j = A_j$ : *si registra un successo nella  $j$ -ma verifica*), ma in due spazi di probabilità diversi.

cioè che il VdA della v.a. *frequenza dei successi*  $S^{(n)}/n$  coincide con la *probabilità di successo*  $p$  in ogni singolo tentativo:

$$\mathbf{E}^{(n)} \left( \frac{S^{(n)}}{n} \right) = p.$$

È però evidente che, essendo la frequenza dei successi una v.a. (e non un numero come il suo VdA), non potremo attenderci che il valore di  $S^{(n)}/n$  coincida con la probabilità di successo  $p$  (cioè con il suo VdA) ogni volta che eseguiamo una  $n$ -pla di tentativi di verifica. Anzi in generale, per una data  $n$ -pla di tentativi, risulterà  $S^{(n)}/n \neq p$ . Sarà pertanto interessante stimare di quanto possiamo attenderci che la v.a. frequenza dei successi  $S^{(n)}/n$  si discosti dal suo VdA, cioè dalla probabilità di successo  $p$ . Il problema così introdotto è particolarmente importante in quanto la sua soluzione indicherà con quale affidabilità è possibile stimare la probabilità di successo  $p$  (che in generale può non essere nota) tramite una osservazione (cioè una sola  $n$ -pla di tentativi di verifica) della frequenza dei successi  $S^{(n)}/n$ : quel che è accessibile alle nostre misure, infatti, è sempre e soltanto il conteggio delle frequenze e non il valore  $p$  delle probabilità *a priori*. Si potrebbe in breve dire che è il problema originario della **statistica** stabilire le condizioni sotto le quali un'osservazione della frequenza dei successi  $S^{(n)}/n$  consente di dedurre una **stima** affidabile di  $p$ .

Per introdurre la discussione su questo argomento notiamo che, siccome la v.a.

$$\left| \frac{S^{(n)}}{n} - p \right|$$

rappresenta la deviazione (in una  $n$ -pla di tentativi di verifica) della frequenza dei successi dal suo VdA (cioè dalla probabilità di successo  $p$ ), sarà sempre possibile, supponendo che sia  $p \neq 0$  e  $p \neq 1$ , scegliere  $\epsilon > 0$  abbastanza piccolo da far in modo che l'evento

$$\left\{ \left| \frac{S^{(n)}}{n} - p \right| \geq \epsilon \right\}$$

si verifichi con una probabilità diversa da zero. Infatti, basterà considerare un  $\epsilon$  tale che  $0 < \epsilon \leq \min(p, 1 - p)$  per avere che:

$$\left\{ \left| \frac{S^{(n)}}{n} - p \right| \geq \epsilon \right\} \supseteq \left\{ \frac{S^{(n)}}{n} = 1 \right\} \cup \left\{ \frac{S^{(n)}}{n} = 0 \right\}$$

e quindi

$$\begin{aligned} \mathbf{P}^{(n)} \left\{ \left| \frac{S^{(n)}}{n} - p \right| \geq \epsilon \right\} &\geq \mathbf{P}^{(n)} \left\{ \frac{S^{(n)}}{n} = 1 \right\} + \mathbf{P}^{(n)} \left\{ \frac{S^{(n)}}{n} = 0 \right\} \\ &= \mathbf{P}^{(n)} \{ \xi_1^{(n)} = 1, \dots, \xi_n^{(n)} = 1 \} + \mathbf{P}^{(n)} \{ \xi_1^{(n)} = 0, \dots, \xi_n^{(n)} = 0 \} \\ &= p^n + (1 - p)^n > 0. \end{aligned}$$

Ciò mostra che, se  $p \neq 0$  e  $p \neq 1$ , la deviazione risulterà non nulla con probabilità diversa da zero. Dagli stessi calcoli, però, risulta che il valore di  $p^n + (1-p)^n$  può essere reso più piccolo di qualunque numero positivo prendendo  $n$  abbastanza grande (ricordiamo, infatti, che stiamo considerando il caso in cui  $p < 1$  e  $1-p < 1$ ). Da questa osservazione nasce allora l'idea di provare a rendere *trascurabile* la probabilità che lo scarto fra frequenza dei successi e probabilità a priori superi un preassegnato valore  $\epsilon > 0$  scegliendo  $n$  *abbastanza grande*. Nel seguito di questo capitolo tenteremo di dare una formulazione precisa di questa affermazione (nota come **Legge dei grandi Numeri**), nella misura in cui ciò è possibile nel quadro di un modello finito di probabilità. Per ottenere questo scopo ci baseremo sulla forma elementare di un risultato generale noto come Diseguaglianza di Chebyshev<sup>4</sup> che avremo modo di applicare varie volte nei capitoli successivi.  $\circ$

**I.7.3 Proposizione (Diseguaglianza di Chebyshev):** Data, su uno spazio di probabilità  $(\Omega, \mathcal{A}, \mathbf{P})$ , una v.a. non negativa  $\xi \geq 0$ ,  $\forall \omega \in \Omega$ , la diseguaglianza

$$\mathbf{P}\{\xi \geq \epsilon\} \leq \frac{\mathbf{E}\xi}{\epsilon}$$

risulta sempre verificata qualunque sia il numero reale  $\epsilon > 0$ .

**Dimostrazione:** Siccome si verifica subito che  $\xi = \xi I_{\{\xi \geq \epsilon\}} + \xi I_{\{\xi < \epsilon\}} \geq \xi I_{\{\xi \geq \epsilon\}} \geq \epsilon I_{\{\xi \geq \epsilon\}}$  segue anche, dalle proprietà del VdA, che  $\mathbf{E}\xi \geq \epsilon \mathbf{E}(I_{\{\xi \geq \epsilon\}}) = \epsilon \mathbf{P}\{\xi \geq \epsilon\}$  da cui, data l'arbitrarietà di  $\epsilon$ , si ricava il risultato richiesto.  $\square$

**I.7.4 Corollario:** Data una v.a.  $\xi$ , le diseguaglianze

$$\mathbf{P}\{|\xi| \geq \epsilon\} \leq \frac{\mathbf{E}|\xi|}{\epsilon}, \quad \mathbf{P}\{|\xi| \geq \epsilon\} = \mathbf{P}\{\xi^2 \geq \epsilon^2\} \leq \frac{\mathbf{E}\xi^2}{\epsilon^2},$$

$$\mathbf{P}\{|\xi - \mathbf{E}\xi| \geq \epsilon\} \leq \frac{\mathbf{V}\xi}{\epsilon^2}$$

risultano sempre verificate qualunque sia il numero reale  $\epsilon > 0$ .

**Dimostrazione:** Banale.  $\square$

<sup>4</sup> Pafnutii L. Chebyshev (1821 - 1894), figlio di un nobile impoverito dalla carestia del 1840, è considerato, dopo Lobachevsky, il più importante matematico russo del XIX secolo. Allievo di V. J. Bunyakowskii (1804 - 1889), si laureò all'Università di Mosca all'età di venti anni e divenne presto professore di matematica all'Università di S. Pietroburgo dove insegnò dal 1847 al 1882. Diede importanti contributi alla teoria dei numeri, alla teoria dell'interpolazione, al calcolo delle variazioni, alla teoria dei numeri primi ed al calcolo delle probabilità. Fra i suoi allievi si annoverano matematici come A. Markov (1856 - 1922) e A. Ljapunov (1857 - 1918).



La diseguaglianza di Chebyshev con i suoi corollari ci mette in condizione di dare una dimostrazione molto semplice della cosiddetta **Legge dei grandi Numeri di J. Bernoulli**: varrà la pena ricordare, infatti, che la dimostrazione originale consisteva in una stima analitica piuttosto complicata delle *code* della distribuzione binomiale che non sarà qui riprodotta. Per far questo ricorderemo che la v.a.  $S^{(n)}$  è una v.a. binomiale sicché

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

Posto allora

$$\mathcal{N}_{n,\epsilon} = \left\{ k \in \mathbf{N} : \left| \frac{k}{n} - p \right| \geq \epsilon \right\}$$

definiremo il simbolo

$$p_{n,\epsilon} = \sum_{k \in \mathcal{N}_{n,\epsilon}} p_n(k) = \mathbf{P}^{(n)} \left\{ \left| \frac{S^{(n)}}{n} - p \right| \geq \epsilon \right\}$$

e passeremo ad enunciare la Legge dei grandi numeri.

**I.7.5 Teorema (Legge dei grandi Numeri di J. Bernoulli):** Comunque fissato  $\epsilon > 0$  risulta sempre:  $p_{n,\epsilon} \xrightarrow{n} 0$ .

**Dimostrazione:** Dai Corollari I.7.4 infatti, ponendo  $\xi = S^{(n)}/n$ , e ricordando i risultati dell'Esempio I.6.24 sulla varianza di una v.a. binomiale, risulta che

$$\begin{aligned} p_{n,\epsilon} &= \mathbf{P}^{(n)} \left\{ \left| \frac{S^{(n)}}{n} - p \right| \geq \epsilon \right\} \leq \frac{1}{\epsilon^2} \mathbf{V}^{(n)} \left( \frac{S^{(n)}}{n} \right) \\ &= \frac{1}{n^2 \epsilon^2} \mathbf{V}^{(n)} S^{(n)} = \frac{np(1-p)}{n^2 \epsilon^2} = \frac{p(1-p)}{n \epsilon^2} \leq \frac{1}{4n \epsilon^2} \end{aligned}$$

da cui discende facilmente la relazione richiesta. □

**I.7.6 Osservazione:** La Legge dei grandi Numeri enunciata nel Teorema I.7.5 non può essere immediatamente presa come una affermazione sulla probabilità di eventi relativi a *successioni infinite* di tentativi di verifica, nonostante il fatto che essa sembri, a prima vista, suggerire proprio questa interpretazione. In breve possiamo dire che, al punto di sviluppo della teoria al quale ci troviamo (cioè: avendo rigorosamente definito solo spazi *finiti* di probabilità), non siamo in alcun modo autorizzati ad interpretare il Teorema I.7.5 come un'affermazione del tipo: *al limite per  $n \rightarrow \infty$  la probabilità che lo scarto fra la frequenza osservata e la probabilità a priori dei successi sia maggiore o eguale ad un  $\epsilon > 0$  arbitrario è infinitesima*. Infatti per poter dire questo dovremmo aver già definito il significato di una probabilità su uno spazio dei campioni non finito (quello costituito dalle successioni infinite di tentativi di verifica), mentre invece noi siamo per ora

limitati alla considerazione di una successione  $\mathbf{P}^{(n)}$  di probabilità diverse su spazi finiti: niente ci assicura che sia possibile definire un'unica probabilità sullo spazio non finito delle successioni di tentativi e niente ci autorizza quindi ad estendere il significato del precedente Teorema al di là dei limiti della sua formulazione: esso non è altro che una affermazione analitica sul comportamento della successione  $p_n(k)$  delle distribuzioni binomiali (affermazione che avremmo potuto dimostrare, come appunto fece J. Bernoulli, con pure considerazioni analitiche e non probabilistiche) e non può essere immediatamente considerato come un enunciato sulle probabilità associate a successioni infinite di tentativi. Questa osservazione sarà ripresa più oltre per motivare il passaggio ad una forma più elaborata del calcolo delle probabilità.  $\circ$

**I.7.7 Esempio:** Dalla dimostrazione della Legge dei grandi Numeri possiamo ora ricavare una prima risposta alla domanda (tipica della Statistica Matematica): dato  $\epsilon > 0$  quale è il minimo valore di  $n$  che ci garantisce che

$$\mathbf{P}^{(n)} \left\{ \left| \frac{S^{(n)}}{n} - p \right| \geq \epsilon \right\} \leq \alpha$$

dove  $\alpha$  è un dato numero reale compreso fra 0 ed 1? Infatti dalla dimostrazione del Teorema I.7.5 sappiamo che

$$\mathbf{P}^{(n)} \left\{ \left| \frac{S^{(n)}}{n} - p \right| \geq \epsilon \right\} \leq \frac{1}{4n\epsilon^2}$$

sicché la disuguaglianza richiesta è garantita se  $\alpha \geq 1/4n\epsilon^2$  cioè se  $n \geq 1/4\alpha\epsilon^2$ . Così ad esempio, se richiediamo che la deviazione della frequenza dalla probabilità sia superiore ad  $\epsilon = 0,02$  con una probabilità inferiore ad  $\alpha = 0,05$ , sarà sufficiente eseguire un numero di osservazioni

$$n \geq \frac{1}{4\alpha\epsilon^2} = 12.500.$$

È possibile mostrare, comunque, che questo valore è eccessivo, nel senso che un numero inferiore di prove è largamente sufficiente: questo dipende dal fatto che la disuguaglianza di Chebyshev costituisce un limite superiore piuttosto grossolano per le deviazioni studiate.  $\diamond$

## I.8 Modello di Bernoulli e Teoremi limite

Abbiamo già osservato che i tentativi di dimostrare la Legge dei grandi Numeri in maniera diretta ed analitica (come nel caso della dimostrazione di J. Bernoulli) sono complicati da problemi di calcolo e di stime della distribuzione binomiale per grandi valori di  $n$ . Un gran progresso (precedente all'enunciazione della disegualianza di Chebyshev) nella tecnica di queste dimostrazioni risultò quindi dalla scoperta di de Moivre<sup>1</sup> (nel caso  $p = \frac{1}{2}$ ), perfezionata poi da Laplace<sup>2</sup> (nel caso generale  $0 < p < 1$ ), di semplici formule asintotiche per  $p_n(k)$  esprimibili tramite le cosiddette **funzioni di Gauss**<sup>3</sup> (**o di Laplace, o Normali**). Come vedremo più oltre la validità di queste ed altre formule asintotiche, (che, con il nome generico di **Teoremi limite**, comprendono anche il Teorema di Poisson<sup>4</sup>) si estende

---

<sup>1</sup> Abraham de Moivre (1667 - 1754), costretto ad abbandonare la Francia nel 1688 a causa della revoca dell'Editto di Nantes, trascorse tutta la sua vita a Londra dove impartiva lezioni di matematica in famiglie benestanti partecipando attivamente al dibattito scientifico di quegli anni. Il 12 novembre del 1733 presentò privatamente ad alcuni amici una breve memoria di sette pagine intitolata *Approximatio ad summam terminorum binomii  $(a + b)^n$  in seriem expansi*: oggi se ne conservano solo due copie e la sua traduzione fu inclusa nella seconda edizione (1738; la prima edizione risale al 1718) della sua opera sul Calcolo delle Probabilità: *The doctrine of chances*. In questo lavoro per la prima volta viene enunciato il teorema sull'approssimazione della distribuzione binomiale mediante una *curva normale*; esso mostra anche che, prima di Stirling, de Moivre aveva ottenuto dei risultati circa il valore di  $n!$  per grandi  $n$ .

<sup>2</sup> Pierre-Simon, Marchese di Laplace (1749 - 1827) fu uno dei fondatori dell'École Polytechnique e dell'École Normale di Parigi. La sua *Théorie analytique des probabilités* (1812) viene usualmente annoverata fra le opere più importanti nel campo del calcolo delle probabilità. Nella edizione del 1820 quest'opera era preceduta da una interessante introduzione che fu anche pubblicata separatamente con il titolo di *Essai philosophique sur les probabilités*. Nella *Théorie analytique* Laplace riprese le dimostrazioni di de Moivre generalizzandole, ma omise di far riferimento al suo predecessore: per questo motivo i risultati in questione sono spesso attribuiti a Laplace stesso invece che al loro originario scopritore.

<sup>3</sup> Carl Friedrich Gauss (1777 - 1855) trascorse tutta la sua carriera accademica come professore di astronomia a Göttingen in Germania, il che gli permise di dedicare tutto il suo tempo alla ricerca scientifica. Già nel 1801 i suoi lavori di astronomia lo avevano condotto a sviluppare dei metodi di approssimazione basati sul criterio dei *minimi quadrati* che furono pubblicati nel 1809 nella *Theoria motus corporum coelestium*. Nel 1821 (*Theoria combinationis observationum erroribus minimis obnoxiae*), in prosecuzione di queste ricerche, egli dimostrò che la cosiddetta *funzione di Gauss* è il limite della legge di probabilità della somma di errori aleatori piccoli ed indipendenti quando il loro numero diviene indefinitamente grande. Divenne così chiaro che il risultato di de Moivre e Laplace non era che un caso particolare di un comportamento molto generale che provava la natura universale della cosiddetta *legge gaussiana o normale*.

<sup>4</sup> Siméon-Denis Poisson (1781 - 1840) si formò all'École Polytechnique, nella quale successivamente insegnò, durante il primo periodo napoleonico. Le sue idee repubblicane, però, ne fecero un avversario del regime imperiale. Ciò gli giovò dopo la Restaurazione durante la quale servì come giurato in molti processi politici. Non è casuale, quindi, che Poisson sia stato indotto a studiare il Calcolo delle Probabilità dal problema di determinare la probabilità di un certo risultato nei giudizi dati a maggioranza (come avviene nei processi). Si tratta di una questione

ben al di là del semplice Modello di Bernoulli e rimane ancora oggi uno degli argomenti centrali del Calcolo delle Probabilità. In questo capitolo, comunque, ci limiteremo a fornire la formulazione esatta di queste proposizioni solo nel caso di distribuzioni binomiali rinviandone le generalizzazioni ad una fase più avanzata della nostra esposizione.

**I.8.1 Osservazione:** Sarà bene richiamare qui che per *funzione di Gauss o Normale* si intende la funzione

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

con un dato valore di  $\mu$  e  $\sigma$ . Tali funzioni hanno una caratteristica forma *a campana* con il massimo collocato in  $x = \mu$  e i due flessi in  $x = \mu \pm \sigma$  (Fig. I.8.1). Si chiama invece funzione di Gauss (o normale) *standard* la funzione di Gauss con  $\mu = 0$  e  $\sigma = 1$ :

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Rivestono una notevole importanza, inoltre, le cosiddette *funzioni Errore*, cioè una particolare classe di primitive delle funzioni di Gauss:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-(t-\mu)^2/2\sigma^2} dt.$$

Tali funzioni hanno la forma mostrata in Fig. I.8.1 con un flesso in  $x = \mu$  e con

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

In particolare, con  $\mu = 0$  e  $\sigma = 1$  si ottiene la funzione errore *standard*:

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

Nota che le funzioni errore non sono esprimibili mediante combinazioni finite di funzioni elementari, ma il loro uso è facilitato dalle tavole numeriche che ne elencano i valori (vedi Tavola I.8.1) e, al giorno d'oggi, dalle possibilità offerte dai moderni strumenti di calcolo automatico. Per l'uso delle Tavole per valori negativi dell'argomento sarà infine utile ricordare che è soddisfatta la relazione  $\Phi(-x) = 1 - \Phi(x)$ .  $\circ$

---

complessa che era stata già trattata da J. Bernoulli e da Laplace, lavori che, però, Poisson non trovava soddisfacenti; egli espose i suoi risultati nel 1837 nelle sue *Recherches sur la probabilité des jugements en matière criminelle et en matière civile* derivandoli solo come limite di distribuzioni binomiali. Le loro generalizzazioni hanno successivamente trovato vastissime applicazioni che vanno dalla gestione dei problemi di traffico fino alla teoria dei decadimenti radioattivi.

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

$x$	0	1	2	3	4	5	6	7	8	9
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7703	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8634	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9278	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9430	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9648	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9700	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9762	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9874	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9990	0,9993	0,9995	0,9997	0,9998	0,9998	0,9999	0,9999	1,0000

**Tavola I.8.1:** Funzione errore standard.

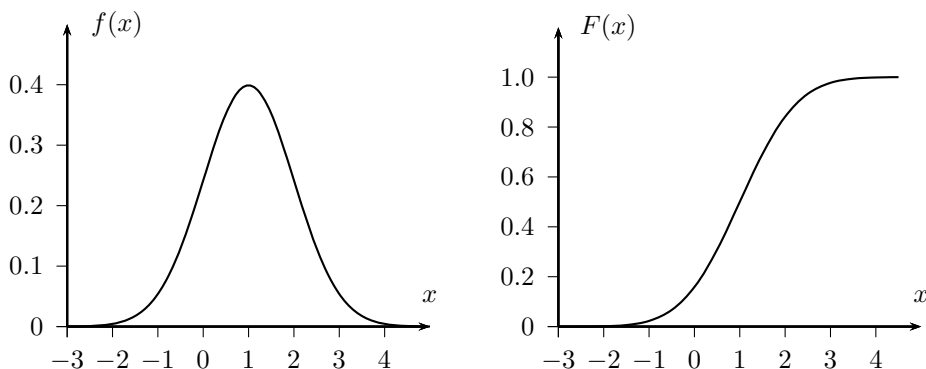


Fig. I.8.1 Funzione di Gauss e funzione errore ( $\mu = 1, \sigma = 1$ )

**I.8.2 Osservazione:** Prima di passare all'enunciazione dei vari teoremi sarà bene notare che essi prendono per lo più la forma di affermazioni riguardanti v.a. *standardizzate*, cioè v.a. con VdA nullo e Var pari ad 1. È facile vedere che data una qualunque v.a.  $\xi$  dotata di VdA  $\mathbf{E}\xi$  e di Var  $\mathbf{V}\xi$ , è possibile costruire una v.a. standardizzata mediante una semplice operazione. Infatti, definita la v.a.

$$\xi_0 = \frac{\xi - \mathbf{E}\xi}{\sqrt{\mathbf{V}\xi}},$$

si vede subito (da I.6.5 e I.6.19) che  $\mathbf{E}\xi_0 = 0$  e che  $\mathbf{V}\xi_0 = 1$ , cioè che essa è ora standardizzata. In particolare i Teoremi Limite di cui parleremo saranno delle proposizioni riguardanti  $S^{(n)}$  oppure la corrispondente v.a. standardizzata (viene qui ripresa la notazione del capitolo I.7)

$$\frac{S^{(n)} - \mathbf{E}^{(n)}S^{(n)}}{\sqrt{\mathbf{V}^{(n)}S^{(n)}}}.$$

Ricordando inoltre che  $S^{(n)}$  è una v.a. binomiale che assume gli  $n + 1$  valori  $k = 0, 1, \dots, n$ , con  $\mathbf{E}^{(n)}S^{(n)} = np$  e  $\mathbf{V}^{(n)}S^{(n)} = npq$  (dove  $q = 1 - p$ ), ci si accorge facilmente che la nuova v.a. standardizzata assume gli  $n + 1$  valori

$$x_k = \frac{k - np}{\sqrt{npq}}, \quad k = 0, 1, \dots, n.$$

In particolare si ha che la v.a. standardizzata ha la seguente DdP

$$\begin{aligned} \mathbf{P}^{(n)} \left\{ \frac{S^{(n)} - \mathbf{E}^{(n)}S^{(n)}}{\sqrt{\mathbf{V}^{(n)}S^{(n)}}} = x_k \right\} &= \mathbf{P}^{(n)} \{ S^{(n)} = k \} = p_n(k) \\ &= p_n(np + x_k\sqrt{npq}) = \binom{n}{k} p^k q^{n-k}, \\ \mathbf{P}^{(n)} \left\{ \left| \frac{S^{(n)} - \mathbf{E}^{(n)}S^{(n)}}{\sqrt{\mathbf{V}^{(n)}S^{(n)}}} \right| \leq x \right\} &= \sum_{k: |x_k| \leq x} p_n(k). \end{aligned}$$

I teoremi limite costituiscono delle formule asintotiche ( $n \rightarrow \infty$ ) particolarmente maneggevoli ed espresse in termini delle funzioni di Gauss o delle funzioni Errore per le distribuzioni delle v.a. considerate. Ad esempio il seguente enunciato del Teorema Limite Locale rappresenta la formulazione esatta dell'affermazione secondo la quale, per grandi valori di  $n$ , i valori della distribuzione binomiale  $p_n(k) = p_n(np + x_k\sqrt{npq})$  sono approssimati bene dalla funzione gaussiana

$$\frac{e^{-(k-np)^2/2npq}}{\sqrt{2\pi npq}} = \frac{1}{\sqrt{npq}} \frac{e^{-x_k^2/2}}{\sqrt{2\pi}}$$

per valori di  $k$  non troppo lontani da  $np$ , VdA di  $S^{(n)}$  (o per valori di  $x_k$  non troppo lontani da 0, VdA della v.a. standardizzata). Più precisamente, al crescere di  $n$ , l'ampiezza  $|k - np|$  (rispettivamente  $|x_k|$ ) dell'intervallo dei valori di  $k$  (di  $x_k$ ) attorno ad  $np$  (a 0) in cui l'approssimazione è correttamente applicabile deve crescere meno rapidamente di  $(npq)^{2/3}$  (di  $(npq)^{1/6}$ ).  $\circ$

**I.8.3 Teorema (Teorema Limite Locale - TLL):** Con le notazioni precedenti e per  $0 < p < 1$  risulta

$$\lim_n \left( \sup_{k: |k-np| \leq \phi(n)} \left| \frac{p_n(k)}{\frac{1}{\sqrt{2\pi npq}} e^{-(k-np)^2/2npq}} - 1 \right| \right) = 0$$

dove  $\phi(n) = o(npq)^{2/3}$ , oppure

$$\lim_n \left( \sup_{k: |x_k| \leq \psi(n)} \left| \frac{p_n(np + x_k\sqrt{npq})}{\frac{1}{\sqrt{2\pi npq}} e^{-x_k^2/2}} - 1 \right| \right) = 0$$

dove  $\psi(n) = o(npq)^{1/6}$ .

**Dimostrazione:** Omessa (vedi ad esempio: **A. N. Shirayev:** *Probability*, Springer, New York, 1984, pp. 56-59.)  $\square$

**I.8.4 Osservazione:** Ricordiamo qui che la notazione  $\phi(n) = o(g(n))$  ( $n \rightarrow \infty$ ) sta a significare che

$$\lim_n \frac{\phi(n)}{g(n)} = 0,$$

mentre  $\phi(n) = O(g(n))$  ( $n \in N$ ) indica che il rapporto  $\frac{\phi(n)}{g(n)}$  si mantiene limitato, cioè che esiste un numero reale e positivo  $M$  tale che

$$\left| \frac{\phi(n)}{g(n)} \right| \leq M, \quad \forall n \in N;$$

in particolare può risultare che  $\lim_n \frac{\phi(n)}{g(n)}$  esista e sia finito. Infine  $\phi(n) \sim g(n)$  ( $n \rightarrow \infty$ ) vuol dire che

$$\lim_n \frac{\phi(n)}{g(n)} = 1,$$

come già osservato in una nota all'Esempio I.3.4. ○

**I.8.5 Osservazione:** Sulla base dell'Osservazione precedente una seconda maniera di enunciare il TLL è la seguente: per  $n \rightarrow \infty$  si ha

$$\begin{aligned} \mathbf{P}^{(n)}\{S^{(n)} = k\} &\sim \frac{e^{-(k-np)^2/2npq}}{\sqrt{2\pi npq}}; & |k - np| \leq \phi(n) = o(npq)^{2/3}, \\ \mathbf{P}^{(n)}\left\{\frac{S^{(n)} - np}{\sqrt{npq}} = x_k\right\} &\sim \frac{1}{\sqrt{npq}} \frac{e^{-x_k^2/2}}{\sqrt{2\pi}}; & |x_k| \leq \psi(n) = o(npq)^{1/6}. \end{aligned}$$

In particolare val la pena di osservare qui che la convergenza dei due limiti del Teorema I.8.3 è anche uniforme in ogni intervallo di ampiezza finita (e costante) di valori di  $k$  attorno ad  $np$  (ovvero di valori di  $x_k$  attorno a 0). Se invece si vogliono eliminare le limitazioni imposte all'intervallo di convergenza dalle ipotesi del TLL si deve passare ad una seconda formulazione nota come Teorema Limite Integrale. Per far questo osserviamo innanzitutto che i numeri  $x_k$  (per  $p$  ed  $n$  dati) sono tutti equidistanti e poniamo

$$\Delta x_k = x_{k+1} - x_k = \frac{1}{\sqrt{npq}}.$$

In tal caso, per  $|x_k| \leq \psi(n) = o(npq)^{1/6}$  e indicando con  $\varphi(x)$  la funzione normale standard di I.8.1, potremo scrivere per  $n \rightarrow \infty$

$$\mathbf{P}^{(n)}\left\{\frac{S^{(n)} - np}{\sqrt{npq}} = x_k\right\} = p_n(np + x_k\sqrt{npq}) \sim \frac{\Delta x_k}{\sqrt{2\pi}} e^{-x_k^2/2} = \varphi(x_k)\Delta x_k.$$

Siccome per  $n \rightarrow \infty$  risulta anche  $\Delta x_k \rightarrow 0$ , l'insieme dei punti  $x_k$  tenderà a ricoprire tutta la retta reale e quindi potremo attenderci che in qualche senso opportunamente definito, comunque assegnati due numeri reali  $a$  e  $b$  (con  $a < b$ ), per  $n$  molto grande il valore di

$$\begin{aligned} P_n(a, b) &= \sum_{k:a < x_k \leq b} p_n(np + x_k\sqrt{npq}) = \mathbf{P}^{(n)}\left\{a < \frac{S^{(n)} - np}{\sqrt{npq}} \leq b\right\} \\ &\sim \sum_{k:a < x_k \leq b} \frac{\Delta x_k}{\sqrt{2\pi}} e^{-x_k^2/2} = \sum_{k:a < x_k \leq b} \varphi(x_k)\Delta x_k \end{aligned}$$

possa essere approssimato mediante

$$\int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \int_a^b \varphi(x) dx = \Phi(b) - \Phi(a),$$



dove  $\Phi(x)$  è la funzione errore standard definita in I.8.1. Prima di enunciare la forma esatta di questa affermazione osserviamo che nei Teoremi I.8.3 e I.8.6 (analogamente a quanto capitava per la Legge dei grandi numeri) si ha a che fare sempre con le  $p_n(k)$  (o con loro somme) e mai con probabilità su spazi non finiti dal momento che non abbiamo ancora gli strumenti per definirle.  $\circ$

**I.8.6 Teorema (Teorema Limite Integrale - TLI):** Se  $0 < p < 1$ , risulta che

$$\lim_n \left( \sup_{-\infty \leq a < b \leq +\infty} \left| P_n(a, b) - \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \right| \right) = 0,$$

dove è stato posto

$$P_n(a, b) = \sum_{k: a < x_k \leq b} p_n(np + x_k \sqrt{npq}),$$

con le notazioni definite in precedenza.

**Dimostrazione:** Omessa (vedi ad esempio: **A. N. Shirayev:** *Probability*; Springer, New York, 1984, pp. 59-62.)  $\square$

**I.8.7 Osservazione:** In particolare, dunque, comunque scelti  $a$  e  $b$  (con  $-\infty \leq a < b \leq +\infty$ ) otterremo sempre per  $n \rightarrow \infty$

$$\mathbf{P}^{(n)} \left\{ a < \frac{S^{(n)} - np}{\sqrt{npq}} \leq b \right\} \sim \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \Phi(b) - \Phi(a),$$

ovvero, ponendo  $A = np + a\sqrt{npq}$  e  $B = np + b\sqrt{npq}$ , anche

$$\mathbf{P}^{(n)} \{ A < S^{(n)} \leq B \} \sim \Phi \left( \frac{B - np}{\sqrt{npq}} \right) - \Phi \left( \frac{A - np}{\sqrt{npq}} \right).$$

La differenza fra i due teoremi limite qui enunciati sta nel fatto che mentre nel TLL si confrontano i singoli valori della distribuzione di probabilità (DdP) di una v.a. binomiale standardizzata con i valori di una funzione normale standard, nel TLI si confrontano somme della suddetta DdP su intervalli dati (anche illimitati) con integrali della normale standard sui medesimi intervalli. Come vedremo, i teoremi I.8.3 ed I.8.6 non sono gli unici teoremi limite: esistono enunciati molto più generali che riguardano anche v.a. non binomiali. Il più importante di questi, noto come **Teorema Limite Centrale (TLC)**, verrà ampiamente discusso in uno dei successivi capitoli.  $\circ$

**I.8.8 Esempio:** Supponiamo di lanciare un dado  $n = 12.000$  volte; qual è la probabilità che il numero di volte in cui il risultato è 6 sia compreso fra 1.800 e 2.100? Per un calcolo esatto dobbiamo ricordare che in ogni singolo lancio (i lanci

sono supposti indipendenti) la probabilità di ottenere 6 è  $\frac{1}{6}$ . Pertanto la v.a.  $S$  (ometteremo per brevità di indicare l'indice  $n = 12.000$  in tutti i simboli di questo esempio) che rappresenta *il numero di volte in cui il risultato è 6 su 12.000 lanci* sarà una v.a. binomiale con  $n = 12.000$  e  $p = \frac{1}{6}$  e quindi la probabilità richiesta sarà

$$P\{1.800 < S \leq 2.100\} = \sum_{1.800 < k \leq 2.100} \binom{12.000}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{12.000-k}.$$

Per evitare di calcolare fattoriali molto grandi potremo usare l'approssimazione fornita dal TLI nella forma espressa in Osservazione I.8.7 con  $A = 1.800$  e  $B = 2.100$ :

$$P\{1.800 < S \leq 2.100\} \sim \Phi(\sqrt{6}) - \Phi(-2\sqrt{6}) \sim 0,992.$$

Nel calcolo precedente abbiamo usato i valori della normale standard di Tavola I.8.1. ◇

**I.8.9 Osservazione:** Il TLI consente di dare una dimostrazione della Legge dei grandi Numeri di J. Bernoulli (vedi Capitolo I.7) diversa da quella ottenuta in I.7.5 mediante la Diseguaglianza di Chebyshev. In particolare vedremo anche, nel successivo esempio, che esso ci permette di dare una migliore stima della deviazione della frequenza (osservata) dei successi, dalla probabilità (a priori) di successo in ogni singolo tentativo. Con notazione ripresa da questo capitolo e dal precedente dovremo provare che  $p_{n,\epsilon} \xrightarrow{n} 0$  con  $\epsilon > 0$  arbitrario: per far questo osserviamo che comunque scelto  $\epsilon > 0$  esisterà sempre un  $\epsilon'$  tale che  $0 < \epsilon' < \epsilon$  e quindi

$$\begin{aligned} p_{n,\epsilon} &= \mathbf{P}^{(n)} \left\{ \left| \frac{S^{(n)}}{n} - p \right| \geq \epsilon \right\} \leq \mathbf{P}^{(n)} \left\{ \left| \frac{S^{(n)}}{n} - p \right| > \epsilon' \right\} \\ &= 1 - \mathbf{P}^{(n)} \left\{ \left| \frac{S^{(n)}}{n} - p \right| \leq \epsilon' \right\}. \end{aligned}$$

Ma, comunque scelto  $\epsilon' > 0$ , risulta anche

$$\begin{aligned} \mathbf{P}^{(n)} \left\{ \left| \frac{S^{(n)}}{n} - p \right| \leq \epsilon' \right\} &= \mathbf{P}^{(n)} \left\{ \left| \frac{S^{(n)} - np}{\sqrt{npq}} \right| \leq \epsilon' \sqrt{\frac{n}{pq}} \right\} \\ &= \mathbf{P}^{(n)} \left\{ -\epsilon' \sqrt{\frac{n}{pq}} \leq \frac{S^{(n)} - np}{\sqrt{npq}} \leq \epsilon' \sqrt{\frac{n}{pq}} \right\} \\ &\geq \mathbf{P}^{(n)} \left\{ -\epsilon' \sqrt{\frac{n}{pq}} < \frac{S^{(n)} - np}{\sqrt{npq}} \leq \epsilon' \sqrt{\frac{n}{pq}} \right\} \\ &= P_n(-\epsilon' \sqrt{n/pq}, \epsilon' \sqrt{n/pq}]. \end{aligned}$$

Siccome però ovviamente si ha

$$\begin{aligned} &\left| P_n(-\epsilon' \sqrt{n/pq}, \epsilon' \sqrt{n/pq}) - \frac{1}{\sqrt{2\pi}} \int_{-\epsilon' \sqrt{n/pq}}^{\epsilon' \sqrt{n/pq}} e^{-x^2/2} dx \right| \\ &\leq \sup_{-\infty \leq a < b \leq +\infty} \left| P_n(a, b] - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx \right|, \end{aligned}$$

e inoltre

$$\lim_n \frac{1}{\sqrt{2\pi}} \int_{-\epsilon' \sqrt{n/pq}}^{\epsilon' \sqrt{n/pq}} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2} dx = 1,$$

il TLI ci garantisce in sostanza che, con  $\epsilon' > 0$  arbitrario,

$$\lim_n P_n(-\epsilon' \sqrt{n/pq}, \epsilon' \sqrt{n/pq}) = 1$$

e quindi in definitiva anche che  $p_{n,\epsilon} \xrightarrow{n} 0$ . ○

**I.8.10 Esempio:** A questo punto possiamo dare una nuova risposta al problema posto nell'Esempio I.7.7. Infatti ora sappiamo che per grandi  $n$  risulta (con notazione ripresa dall'Osservazione I.8.1)

$$\begin{aligned} \mathbf{P}^{(n)} \left\{ \left| \frac{S^{(n)}}{n} - p \right| \geq \epsilon \right\} &\sim 1 - \frac{1}{\sqrt{2\pi}} \int_{-\epsilon \sqrt{n/pq}}^{\epsilon \sqrt{n/pq}} e^{-x^2/2} dx \\ &\leq 1 - \frac{1}{\sqrt{2\pi}} \int_{-2\epsilon\sqrt{n}}^{2\epsilon\sqrt{n}} e^{-x^2/2} dx = 1 - [\Phi(2\epsilon\sqrt{n}) - \Phi(-2\epsilon\sqrt{n})] \end{aligned}$$

sicch , se vogliamo determinare il pi  piccolo valore di  $n$  che rende

$$\mathbf{P}^{(n)} \left\{ \left| \frac{S^{(n)}}{n} - p \right| \geq \epsilon \right\} \leq \alpha,$$

baster  scegliere  $n$  in modo che  $\Phi(2\epsilon\sqrt{n}) - \Phi(-2\epsilon\sqrt{n}) \geq 1 - \alpha$ , cio  in modo che  $\Phi(2\epsilon\sqrt{n}) \geq 1 - \alpha/2$ . Dato che  $\Phi(x)$    crescente, se indichiamo con  $K(\alpha)$  il numero reale tale che  $\Phi(K(\alpha)) = 1 - \alpha/2$  (numero deducibile dalle tavole o da un calcolo numerico diretto), per soddisfare la nostra diseguaglianza baster  prendere il pi  piccolo intero  $n$  tale che  $2\epsilon\sqrt{n} \geq K(\alpha)$ , e cio 

$$n \geq \frac{K^2(\alpha)}{4\epsilon^2}.$$

Ricordiamo che invece la stima ottenuta in I.7.7 era

$$n \geq \frac{1}{4\epsilon^2\alpha}.$$

Con gli stessi dati numerici di I.7.7 abbiamo allora dalle tavole che  $K(0,05) = 1,96$ , da cui si ottiene che sar  sufficiente scegliere  $n = 2.400$  per essere sicuri che la probabilit  di registrare scarti superiori ad  $\epsilon = 0,02$  non superi  $\alpha = 0,05$ . Come si pu  vedere questa stima di  $n$    di molto inferiore a quella di 12.500 ottenuta in I.7.7 in base alla diseguaglianza di Chebyshev. ◇

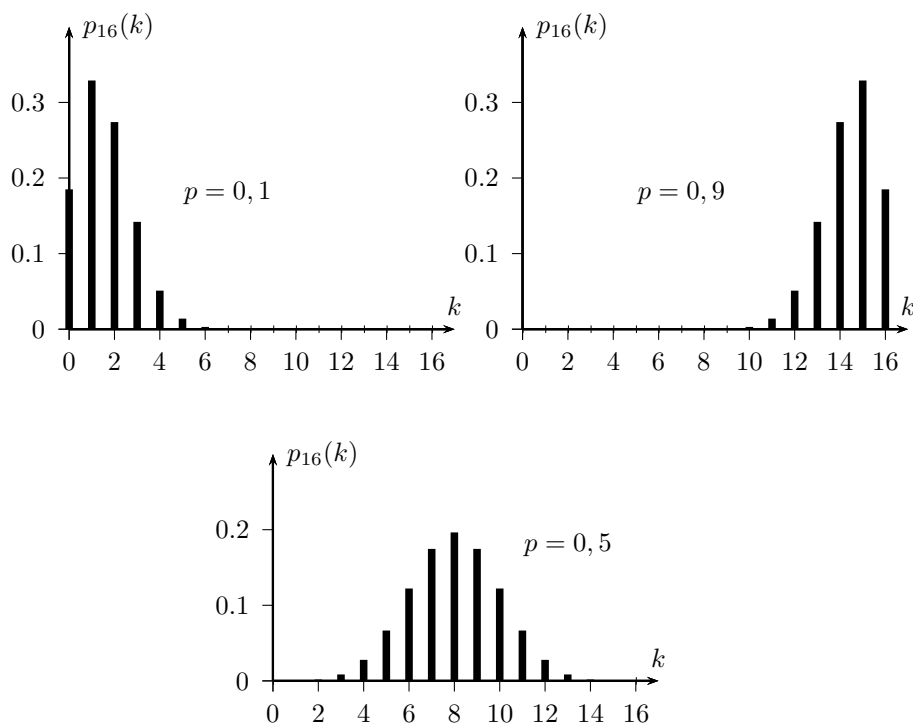


Fig. I.8.2 Esempi di distribuzioni binomiali con diversi valori di  $p$ .

**I.8.11 Osservazione:** Abbiamo già notato che il TLL pretende di approssimare i valori di una distribuzione binomiale mediante valori di funzioni normali e che la validità di tale approssimazione è sempre più dubbia man mano che i valori di  $k$  si allontanano dal valore d'attesa della v.a. binomiale data. Queste difficoltà del TLL divengono particolarmente evidenti quando il valore di  $p$  è molto prossimo a 0 o 1. Infatti una curva gaussiana è perfettamente simmetrica attorno al suo massimo, esattamente come capita per la distribuzione binomiale quando  $p = \frac{1}{2}$  (Fig. I.8.2). La simmetria della distribuzione binomiale, però, si perde quando  $p$  si discosta da  $\frac{1}{2}$  avvicinandosi a 0 oppure 1 (Fig. I.8.2). Per piccoli valori di  $p$  (o di  $1-p$ ) dunque non è ragionevole pretendere che una curva normale approssimi bene la distribuzione binomiale, se non nelle immediate vicinanze del suo massimo. In questi casi siamo dunque condotti a cercare altre approssimazioni asintotiche (per  $n \rightarrow \infty$ ) della distribuzione binomiale, anche perché, in un certo tipo di problemi pratici, saremo obbligati a costruire modelli probabilistici leggermente diversi da quello originario di Bernoulli. Più precisamente può capitare che la probabilità dell'evento  $A_j^{(n)}$  (si registra un successo al  $j$ -mo tentativo, vedi capitolo I.7) non sia sempre la stessa (cioè  $p$ ) al variare del numero di tentativi  $n$ , ma ne dipenda esplicitamente in modo che  $\mathbf{P}^{(n)}(A_j^{(n)}) = p(n)$ ; in particolare noi ci interesseremo al caso in cui  $p(n) \rightarrow 0$  per  $n \rightarrow \infty$  come esposto nel seguente esempio.  $\circ$

**I.8.12 Esempio:** Un centralino telefonico riceve, in istanti casuali, delle telefonate il cui numero medio è proporzionale alla lunghezza (in minuti) dell'intervallo considerato. In particolare è noto che, in un dato periodo della giornata, arrivano in media  $\alpha = 1,5$  telefonate al minuto (cioè 90 telefonate all'ora). Detta ora  $S$  la v.a. che indica il numero delle telefonate che arrivano in un intervallo di 3 minuti di quel periodo della giornata, ci chiediamo, sulla base delle precedenti informazioni, quale è la probabilità che in tale intervallo cadano  $k = 0, 1, 2, \dots$  telefonate: in pratica vogliamo calcolare la DdP di  $S$ . Ovviamente la nostra v.a. assumerà solo valori interi  $k = 0, 1, \dots$  ma, a differenza delle v.a. finora considerate (come le v.a. binomiali) i suoi valori non sono limitati: dato che nei 3 minuti considerati può arrivare un numero arbitrario di telefonate, l'insieme dei possibili valori di  $S$  coincide con tutto l'insieme dei numeri interi. Siccome però (Osservazione I.5.4) una v.a. definita su uno spazio finito di probabilità non può che assumere un numero finito di valori, è chiaro che ancora una volta ci troviamo in presenza di un problema il cui modello esatto non può essere dato a questo stadio della nostra trattazione. Cercheremo, pertanto, di trovare un opportuno metodo per approssimare i risultati esatti. Per far questo cominceremo con l'osservare che al nostro livello non ha senso tentare il calcolo partendo dalla probabilità di ricevere una telefonata in un preciso istante di tempo (che sarebbe sempre e comunque zero) e che pertanto ci conviene affrontare il problema con un metodo di approssimazioni successive. Siccome ci attendiamo che arrivino mediamente  $\alpha = 1,5$  telefonate al minuto, inizieremo col suddividere l'intervallo di tempo  $T = 3'$  in un numero  $n$  di sub-intervalli di egual durata e abbastanza piccoli da far in modo che in ciascuno di essi ci si attenda di ricevere non più di una telefonata. Ad esempio, se  $n = 9$ , i sub-intervalli dureranno  $20''$  ciascuno e siccome il numero medio di telefonate in tali intervalli è ora  $\alpha T/n = 1,5 \cdot 3'/9 = 0,5$ , in prima approssimazione potremo supporre che in ognuno dei sub-intervalli arriverà al più una telefonata. Il nostro modello può allora essere costruito con  $n = 9$  tentativi indipendenti di verificare che in ciascuno dei 9 sub-intervalli di  $20''$  cada o meno una telefonata, definendo le v.a. di Bernoulli  $\xi_j^{(9)}$  con  $j = 1, \dots, 9$  che assumono valore 1 se capita una telefonata nel  $j$ -mo sub-intervallo e valore 0 in caso contrario. Siccome inoltre il VdA del numero di telefonate in ogni subintervallo (che noi supponiamo essere 0 o 1) è 0,5, le nostre  $\xi_j^{(9)}$  avranno 0,5 come VdA e quindi (vedi Esempio I.6.10)  $p(9) = 0,5$  sarà anche il valore della probabilità con cui ogni  $\xi_j^{(9)}$  assume valore 1. In pratica otterremo un modello di Bernoulli con  $n = 9$  e  $p = p(9) = 0,5$ , sicché la v.a.  $S^{(9)} = \xi_1^{(9)} + \dots + \xi_9^{(9)}$  (che rappresenta il numero di telefonate in  $T = 3'$  se in ogni sub-intervallo capita al più una telefonata) assumerà i valori  $k = 0, 1, \dots, 9$  e sarà caratterizzata dalla seguente DdP binomiale:

$$\mathbf{P}^{(9)}\{S^{(9)} = k\} = \binom{9}{k} (0,5)^k (0,5)^{9-k} = \binom{9}{k} (0,5)^9.$$

Il difetto di questa approssimazione, come è chiaro, sta nel fatto che, per costruire un modello di Bernoulli, siamo stati costretti a supporre che in ogni sub-intervallo di  $20''$  arrivi al più una telefonata e di conseguenza abbiamo approssimato una

v.a.  $S$  i cui possibili valori sono infiniti, con una v.a.  $S^{(9)}$  i cui possibili valori sono solo 10. Questa osservazione, però, ci suggerisce anche la strada per migliorare la nostra approssimazione: se il numero  $n$  dei sub-intervalli in cui suddividiamo  $T = 3'$  aumenta, da un lato otteniamo delle v.a.  $S^{(n)}$  con un numero sempre crescente di possibili valori, e dall'altro, rendendo sempre più piccola la durata dei sub-intervalli, saremo sempre più ragionevolmente sicuri del fatto che in ogni sub-intervallo non arrivi più di una sola telefonata. Così, ad esempio, con  $n = 18$  sub-intervalli di  $10''$  ciascuno otteniamo  $p(18) = \alpha T/n = 1,5 \cdot 3'/18 = 0,25$  e quindi

$$\mathbf{P}^{(18)}\{S^{(18)} = k\} = \binom{18}{k} (0,25)^k (0,75)^{18-k}, \quad k = 0, 1, \dots, 18.$$

Potremo pertanto continuare a migliorare la nostra approssimazione aumentando il valore di  $n$ , nel qual caso risulterà  $p(n) = \alpha T/n = 1,5 \cdot 3'/n \xrightarrow{n \rightarrow \infty} 0$ . Notiamo che la successione  $p(n)$  è infinitesima, ma che  $np(n) = \alpha T = \lambda$  rimane costante lungo tutto il processo di limite per  $n \rightarrow \infty$ . La distribuzione limite ( $n \rightarrow \infty$ ) della successione di DdP

$$\mathbf{P}^{(n)}\{S^{(n)} = k\} = \binom{n}{k} p(n)^k (1 - p(n))^{n-k}$$

sarà infine ricavata nel seguente Teorema di Poisson. ◇

**I.8.13 Teorema (Teorema di Poisson):** Posto

$$p_n(k) = \begin{cases} \binom{n}{k} p^k q^{n-k}, & k = 0, 1, \dots, n, \\ 0 & , \quad k \geq n + 1, \end{cases}$$

con  $p = p(n) \rightarrow 0$  ( $n \rightarrow \infty$ ) e  $q = 1 - p(n)$ , se esiste  $\lambda > 0$  tale che  $\lim_n np(n) = \lambda$ , allora

$$\lim_n p_n(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, \dots;$$

l'insieme di numeri

$$\pi_\lambda(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, \dots;$$

prende anche il nome di **Distribuzione di Poisson**.

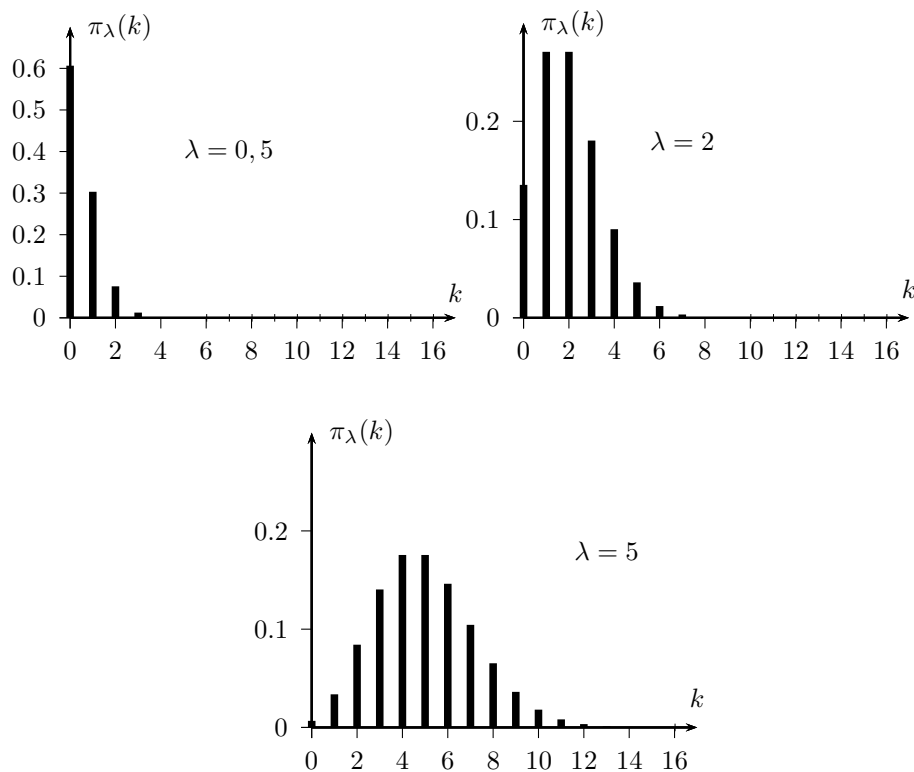
**Dimostrazione:** Le ipotesi fatte ci consentono di scrivere che  $p(n) = \lambda/n + o(n^{-1})$  sicché (con  $k = 0, 1, \dots, n$ )

$$\begin{aligned} p_n(k) &= \binom{n}{k} p^k q^{n-k} \\ &= \frac{n(n-1) \dots (n-k+1)}{k!} \left[ \frac{\lambda}{n} + o(n^{-1}) \right]^k \left[ 1 - \frac{\lambda}{n} + o(n^{-1}) \right]^{n-k}; \end{aligned}$$

ma si controlla anche facilmente, ricordando i limiti notevoli che definiscono il numero  $e$ , che

$$\begin{aligned} n(n-1)\dots(n-k+1) \left[ \frac{\lambda}{n} + o(n^{-1}) \right]^k &= \frac{n(n-1)\dots(n-k+1)}{n^k} [\lambda + o(1)]^k \\ &= \left(1 - \frac{1}{n}\right) \dots \left(1 - \frac{k-1}{n}\right) [\lambda + o(1)]^k \xrightarrow{n} \lambda^k, \\ \left[1 - \frac{\lambda}{n} + o(n^{-1})\right]^{n-k} &= \left[1 - \frac{\lambda}{n} + o(n^{-1})\right]^n \left[1 - \frac{\lambda}{n} + o(n^{-1})\right]^{-k} \xrightarrow{n} e^{-\lambda}, \end{aligned}$$

da cui segue immediatamente la tesi.  $\square$



**Fig. I.8.3** Esempî di distribuzioni di Poisson con diversi valori di  $\lambda$ .

**I.8.12 Esempio (continuazione):** Sulla base del precedente teorema, e tenendo conto del fatto che nel nostro esempio risulta  $np(n) = \alpha T = \lambda = 4,5$  (con  $n \in N$ ), potremmo dire che la probabilità di ricevere  $k$  telefonate in  $T$  minuti è

$$\pi_{4,5}(k) = \frac{(4,5)^k e^{-4,5}}{k!},$$

che è una particolare Distribuzione di Poisson (come esempi di grafici di Distribuzioni di Poisson vedi Fig. I.8.3). Nota però che (come nel caso della Osservazione I.7.6 sulla Legge dei Grandi Numeri) a rigore non possiamo affermare, a questo stadio della nostra discussione, che la v.a.  $S$  *numero di telefonate che arrivano nei 3 minuti considerati* sia distribuita secondo la DdP di Poisson; infatti tale v.a. dovrebbe essere definita su uno spazio di probabilità non finito, mentre finora noi abbiamo lavorato solo in spazi finiti. Pertanto il Teorema I.8.13 va per ora interpretato solo come una affermazione analitica sulla convergenza della successione delle  $p_n(k)$  verso  $\pi_\lambda(k)$  senza che a tale  $\pi_\lambda(k)$  possa, per ora, essere dato rigorosamente il significato di DdP di una v.a. Ciononostante, pur tenendo presenti queste osservazioni, forzeremo per il momento il significato dei termini (rinviando una definizione più precisa alla seconda parte di queste lezioni) e diremo che  $S$  è una **v.a. di Poisson di parametro  $\lambda$**  quando essa assume valori interi  $k = 0, 1, 2, \dots$  con probabilità date dalla Distribuzione di Poisson  $\pi_\lambda(k)$ . Estenderemo inoltre anche le definizioni di VdA e Var, date nel Capitolo I.6 per v.a. finite, al caso delle v.a. di Poisson ponendo

$$\mathbf{E}S = \sum_{k=0}^{\infty} k\pi_\lambda(k); \quad \mathbf{E}S^2 = \sum_{k=0}^{\infty} k^2\pi_\lambda(k); \quad \mathbf{V}S = \mathbf{E}S^2 - (\mathbf{E}S)^2.$$

Per calcolare tali quantità, a partire dalle seguenti forme equivalenti della serie esponenziale

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

si ha che

$$\begin{aligned} \mathbf{E}S &= \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda, \\ \mathbf{E}(S(S-1)) &= \sum_{k=0}^{\infty} k(k-1) e^{-\lambda} \frac{\lambda^k}{k!} = \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} = \lambda^2. \end{aligned}$$

Siccome  $\mathbf{E}(S(S-1)) = \mathbf{E}S^2 - \mathbf{E}S$ , ne segue anche che  $\mathbf{E}S^2 = \lambda^2 + \lambda$  e quindi che  $\mathbf{V}S = \mathbf{E}S^2 - (\mathbf{E}S)^2 = \lambda$ . Pertanto per una v.a. di Poisson di parametro  $\lambda$  risulta sempre  $\mathbf{E}S = \mathbf{V}S = \lambda$ .  $\diamond$

**I.8.14 Osservazione:** L'errore dell'approssimazione gaussiana alla distribuzione binomiale è tanto più piccolo quanto più  $npq$  è grande (Teorema I.8.3). D'altra parte se  $n$  è grande ma  $p$  è piccolo i valori della distribuzione binomiale possono essere approssimati con la distribuzione di Poisson di parametro  $\lambda = np$  (Teorema I.8.13): se il valore di  $\lambda$  resta piccolo solo tale approssimazione poissonniana può essere usata; se invece  $\lambda$  è grande potremo approssimare bene la distribuzione binomiale sia con funzioni di Gauss che con distribuzioni di Poisson. Questo



implica che per grandi valori di  $\lambda$  deve essere possibile anche approssimare la distribuzione di Poisson mediante funzioni di Gauss. In pratica, se  $S$  è una v.a. di Poisson di parametro  $\lambda$ , dai risultati dell'Esempio precedente avremo che la sua forma *standardizzata* sarà

$$\frac{S - \lambda}{\sqrt{\lambda}}$$

e la proprietà asintotica di cui stiamo discutendo potrebbe esprimersi (analogamente a quanto fatto per le v.a. binomiali in Osservazione I.8.2) dicendo che per grandi  $\lambda$  i valori della sua distribuzione sono ben approssimati dai valori della funzione di Gauss standard  $\varphi(x)$ . Come al solito, però, dovremo far attenzione al modo in cui tale proposizione viene enunciata dal momento che non abbiamo ancora correttamente definito v.a. su spazi non finiti di probabilità. Per questo motivo il teorema in questione sarà ancora una volta espresso come proprietà analitica dei numeri  $\pi_\lambda(k)$  in una forma per alcuni versi simile a quella del TLI I.8.6, mentre una dimostrazione più rigorosa sarà data nella Parte III di queste lezioni.  $\square$

**I.8.15 Teorema:** Se

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

si ha che

$$\lim_{\lambda \rightarrow +\infty} \sum_{\lambda + a\sqrt{\lambda} < k < \lambda + b\sqrt{\lambda}} \pi_\lambda(k) = \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx$$

quali che siano i valori di  $a$  e  $b$  con  $a < b$ .

**Dimostrazione:** Omessa (vedi ad esempio **W. Feller:** *An introduction to probability theory and its applications*; Wiley, New York, 1968; vol. I, p. 194).  $\square$



## I.9 Valore d'attesa condizionato

**I.9.1 Definizione:** Dati in uno spazio (finito) di probabilità  $(\Omega, \mathcal{A}, \mathbf{P})$  una decomposizione  $\mathcal{D} = \{D_1, \dots, D_k\}$  (vedi I.2.10) ed un evento  $A \in \mathcal{A}$ , chiameremo **probabilità condizionata di  $A$  rispetto alla decomposizione  $\mathcal{D}$**  la v.a.

$$\mathbf{P}(A | \mathcal{D}) = \sum_{i=1}^k \mathbf{P}(A | D_i) I_{D_i},$$

dove  $I_{D_i}$  sono gli indicatori degli atomi  $D_i$ . Conseguentemente tale v.a. assume il valore (costante)  $\mathbf{P}(A | D_i)$  su ogni atomo  $D_i$ .  $\triangle$

**I.9.2 Osservazione:** In un certo senso  $\mathbf{P}(A | \mathcal{D})$  è la v.a. che riassume tutte le possibili probabilità condizionate (nel senso usuale del termine)  $\mathbf{P}(A | D_i)$  rispetto agli atomi della data decomposizione e quindi gode di proprietà simili a quelle delle probabilità condizionate. Così, ad esempio, se  $A \cap B = \emptyset$  si ha

$$\mathbf{P}(A \cup B | \mathcal{D}) = \mathbf{P}(A | \mathcal{D}) + \mathbf{P}(B | \mathcal{D}).$$

Inoltre la formula della **probabilità totale** (vedi I.4.5) assume ora la forma

$$\mathbf{E}[\mathbf{P}(A | \mathcal{D})] = \mathbf{P}(A)$$

in quanto

$$\mathbf{E}[\mathbf{P}(A | \mathcal{D})] = \sum_{i=1}^k \mathbf{P}(A | D_i) \mathbf{P}(D_i) = \sum_{i=1}^k \mathbf{P}(A D_i) = \mathbf{P}(A),$$

come segue da I.9.1 e dalla definizione di VdA.  $\circ$

**I.9.3 Osservazione:** Indichiamo con  $\mathcal{D}_\eta$  la decomposizione indotta da una v.a.  $\eta$  (vedi l'inizio del capitolo I.6) nel senso che, detti  $y_j$  con  $j = 1, \dots, k$  i valori distinti di  $\eta$ , i suoi atomi sono  $D_j = \{\eta = y_j\}$ . Siccome è chiaro che in spazi finiti ogni v.a. induce una ed una sola decomposizione, per brevità useremo anche indicare la v.a.  $\mathbf{P}(A | \mathcal{D}_\eta)$  semplicemente con il simbolo  $\mathbf{P}(A | \eta)$ . Questa notazione è generalizzabile anche al caso in cui la nostra decomposizione  $\mathcal{D}_{\eta_1, \dots, \eta_m}$  è generata da  $m$  v.a.  $\eta_1, \dots, \eta_m$  nel senso che gli atomi sono gli eventi  $D_{y_1, \dots, y_m} = \{\eta_1 = y_1, \dots, \eta_m = y_m\}$ . In tal caso si usa indicare la v.a.  $\mathbf{P}(A | \mathcal{D}_{\eta_1, \dots, \eta_m})$  anche con il simbolo  $\mathbf{P}(A | \eta_1, \dots, \eta_m)$ .  $\circ$

**I.9.4 Esempio:** Siano  $\xi$  e  $\eta$  due v.a. i.i.d. (vedi I.5.15) che assumono solo i valori 0 ed 1 in modo tale che  $\mathbf{P}\{\xi = 1\} = \mathbf{P}\{\eta = 1\} = p$  e  $\mathbf{P}\{\xi = 0\} = \mathbf{P}\{\eta = 0\} = 1 - p$ . Vogliamo determinare esplicitamente la v.a.  $\mathbf{P}(\xi + \eta = k | \eta)$  dove ovviamente  $k = 0, 1, 2$ . Siccome  $\xi$  ed  $\eta$  sono indipendenti risulta sempre

$$\begin{aligned} \mathbf{P}(\xi + \eta = z | \eta = y) &= \frac{\mathbf{P}(\xi + \eta = z, \eta = y)}{\mathbf{P}(\eta = y)} = \frac{\mathbf{P}(\xi + y = z, \eta = y)}{\mathbf{P}(\eta = y)} \\ &= \frac{\mathbf{P}(\xi = z - y, \eta = y)}{\mathbf{P}(\eta = y)} = \frac{\mathbf{P}(\xi = z - y) \mathbf{P}(\eta = y)}{\mathbf{P}(\eta = y)} = \mathbf{P}(\xi + y = z), \end{aligned}$$

e pertanto da I.9.1 otteniamo

$$\begin{aligned}\mathbf{P}(\xi + \eta = k \mid \eta) &= \mathbf{P}(\xi + \eta = k \mid \eta = 0) I_{\{\eta=0\}} + \mathbf{P}(\xi + \eta = k \mid \eta = 1) I_{\{\eta=1\}} \\ &= \mathbf{P}(\xi = k) I_{\{\eta=0\}} + \mathbf{P}(\xi = k - 1) I_{\{\eta=1\}},\end{aligned}$$

sicch  avremo

$$\mathbf{P}(\xi + \eta = k \mid \eta) = \begin{cases} (1-p) I_{\{\eta=0\}}, & k = 0; \\ p I_{\{\eta=0\}} + (1-p) I_{\{\eta=1\}}, & k = 1; \\ p I_{\{\eta=1\}}, & k = 2, \end{cases}$$

ossia anche

$$\mathbf{P}(\xi + \eta = k \mid \eta) = \begin{cases} (1-p)(1-\eta), & k = 0; \\ p(1-\eta) + (1-p)\eta, & k = 1; \\ p\eta, & k = 2. \end{cases}$$

Nota che la v.a. ottenuta   esplicitamente data come una funzione della v.a. condizionante  $\eta$ .  $\diamond$

**I.9.5 Definizione:** Data una v.a.  $\xi$  che assuma valori  $x_j$  con  $j = 1, \dots, l$  in modo che

$$\xi = \sum_{j=1}^l x_j I_{A_j}, \quad A_j = \{\xi = x_j\},$$

e dato un evento  $B$  (con  $\mathbf{P}(B) > 0$ ), chiameremo **valore d'attesa condizionato rispetto all'evento  $B$**  il numero

$$\mathbf{E}(\xi \mid B) = \sum_{j=1}^l x_j \mathbf{P}(A_j \mid B).$$

Se inoltre  $\mathcal{D} = \{D_1, \dots, D_k\}$    una decomposizione data, la v.a.

$$\mathbf{E}(\xi \mid \mathcal{D}) = \sum_{j=1}^l x_j \mathbf{P}(A_j \mid \mathcal{D}) = \sum_{j=1}^l \sum_{i=1}^k x_j \mathbf{P}(A_j \mid D_i) I_{D_i} = \sum_{i=1}^k \mathbf{E}(\xi \mid D_i) I_{D_i}$$

prende il nome di **valore d'attesa condizionato rispetto alla decomposizione  $\mathcal{D}$** . Useremo inoltre le seguenti notazioni abbreviate

$$\mathbf{E}(\xi \mid \mathcal{D}_\eta) = \mathbf{E}(\xi \mid \eta), \quad \mathbf{E}(\xi \mid \mathcal{D}_{\eta_1, \dots, \eta_k}) = \mathbf{E}(\xi \mid \eta_1, \dots, \eta_k),$$

analoghe a quelle adottate per le probabilit  condizionate.  $\triangle$

**I.9.6 Osservazione:** In generale la v.a.  $\xi$  della precedente definizione non assume valori costanti sugli atomi di  $\mathcal{D}$ ; essa assume piuttosto valori costanti  $x_j$  sugli atomi

$A_j = \{\xi = x_j\}$  della decomposizione indotta. L'operazione di *valore d'attesa condizionato rispetto a  $\mathcal{D}$*  sostituisce al posto di  $\xi$  una nuova v.a. che invece assume valori costanti  $\mathbf{E}(\xi | D_i)$  proprio sugli atomi di  $\mathcal{D}$ . Notiamo inoltre che (vedi I.5.7 per le proprietà degli indicatori)

$$\begin{aligned} \mathbf{E}(\xi | D_i) &= \sum_{j=1}^l x_j \mathbf{P}(A_j | D_i) = \frac{1}{\mathbf{P}(D_i)} \sum_{j=1}^l x_j \mathbf{P}(A_j | D_i) \\ &= \frac{1}{\mathbf{P}(D_i)} \mathbf{E} \left[ \sum_{j=1}^l x_j I_{A_j} I_{D_i} \right] = \frac{\mathbf{E}(\xi I_{D_i})}{\mathbf{P}(D_i)} = \frac{\mathbf{E}(\xi I_{D_i})}{\mathbf{E} I_{D_i}}; \end{aligned}$$

una relazione che ci sarà utile in seguito. ○

**I.9.7 Proposizione:** Con le notazioni precedenti, le seguenti proprietà dei VdA condizionati

- 1)  $\mathbf{E}(a\xi + b\eta | \mathcal{D}) = a\mathbf{E}(\xi | \mathcal{D}) + b\mathbf{E}(\eta | \mathcal{D})$  ( $a$  e  $b$  costanti),
- 2)  $\mathbf{E}(\xi | \Omega) = \mathbf{E}\xi$ ,
- 3)  $\mathbf{E}(a | \mathcal{D}) = a$  ( $a$  costante),
- 4)  $\mathbf{E}(I_A | \mathcal{D}) = \mathbf{P}(A | \mathcal{D})$ ,
- 5)  $\mathbf{E}[\mathbf{E}(\xi | \mathcal{D})] = \mathbf{E}\xi$  (formula della probabilità totale),

sono sempre verificate.

**Dimostrazione:** Le verifiche sono elementari e sono lasciate per esercizio; come esempio dimostreremo solo la 5):

$$\begin{aligned} \mathbf{E}[\mathbf{E}(\xi | \mathcal{D})] &= \mathbf{E} \left[ \sum_{j=1}^l x_j \mathbf{P}(A_j | \mathcal{D}) \right] = \sum_{j=1}^l x_j \mathbf{E}[\mathbf{P}(A_j | \mathcal{D})] \\ &= \sum_{j=1}^l x_j \mathbf{P}(A_j) = \mathbf{E}\xi \end{aligned}$$

dove abbiamo fatto uso dei risultati esposti in I.9.2. □

**I.9.8 Definizione:** Una v.a.  $\eta$  si dice **misurabile** rispetto ad una data decomposizione  $\mathcal{D}$  ( $\mathcal{D}$ -mis) se  $\mathcal{D}_\eta \preceq \mathcal{D}$  cioè se  $\mathcal{D}$  è più fine di  $\mathcal{D}_\eta$  (vedi Definizione I.2.12 ed Osservazione I.5.4). △

**I.9.9 Osservazione:** Ovviamente ogni v.a.  $\eta$  è  $\mathcal{D}_\eta$ -mis, mentre se  $\mathcal{D} = \{\Omega\}$  le uniche v.a.  $\mathcal{D}$ -mis saranno quelle che assumono valore costante (arbitrario) su tutto  $\Omega$ . Notiamo tra l'altro che se  $\eta$  è  $\mathcal{D}$ -mis risulterà anche che gli atomi  $B_r = \{\eta = y_r\}$  (con  $r = 1, \dots, s$ ) di  $\mathcal{D}_\eta$  saranno decomponibili in atomi  $D_i$  di

$\mathcal{D}$ . Pertanto in questo caso  $\eta$  assume valori costanti anche sugli atomi  $D_i$  (anche se tali valori non sono tutti distinti come quelli assunti sulle  $B_r$ ) sicché potremo scrivere

$$\eta = \sum_{r=1}^s y_r I_{B_r} = \sum_{i=1}^k y'_i I_{D_i}$$

dove le  $y'_i$  saranno non tutte distinte. Riprendendo inoltre l'Osservazione I.9.6 potremo dire che in generale una v.a.  $\xi$  non sarà  $\mathcal{D}$ -mis rispetto ad una generica decomposizione  $\mathcal{D}$ , mentre  $\mathbf{E}(\xi | \mathcal{D})$  risulta sempre per definizione  $\mathcal{D}$ -mis.  $\circ$

**I.9.10 Proposizione:** Se  $\eta$  è  $\mathcal{D}$ -mis, qualunque sia  $\xi$  risulterà

$$\mathbf{E}(\xi\eta | \mathcal{D}) = \eta \mathbf{E}(\xi | \mathcal{D}).$$

In particolare:  $\mathbf{E}(\eta | \mathcal{D}) = \eta$  e  $\mathbf{E}(\eta | \mathcal{D}_\eta) = \eta$ .

**Dimostrazione:** Tenendo conto delle proprietà I.5.7 degli indicatori, per ipotesi potremo scrivere

$$\begin{aligned} \xi &= \sum_{j=1}^l x_j I_{A_j}, & \eta &= \sum_{i=1}^k y'_i I_{D_i}, \\ \xi\eta &= \sum_{j=1}^l \sum_{i=1}^k x_j y'_i I_{A_j D_i}, \end{aligned}$$

e pertanto, essendo disgiunti gli atomi delle decomposizioni,

$$\begin{aligned} \mathbf{E}(\xi\eta | \mathcal{D}) &= \sum_{j=1}^l \sum_{i=1}^k x_j y'_i \mathbf{P}(A_j D_i | \mathcal{D}) = \sum_{j=1}^l \sum_{i=1}^k x_j y'_i \sum_{m=1}^k \mathbf{P}(A_j D_i | D_m) I_{D_m} \\ &= \sum_{j=1}^l \sum_{i=1}^k x_j y'_i \mathbf{P}(A_j | D_i) I_{D_i}. \end{aligned}$$

Ma si ha anche (essendo  $I_{D_i} I_{D_m} = 0$  se  $i \neq m$ , e  $I_{D_i}^2 = I_{D_i}$ )

$$\begin{aligned} \eta \mathbf{E}(\xi | \mathcal{D}) &= \sum_{i=1}^k y'_i I_{D_i} \sum_{j=1}^l x_j \mathbf{P}(A_j | \mathcal{D}) = \sum_{i=1}^k y'_i I_{D_i} \sum_{m=1}^k \left[ \sum_{j=1}^l x_j \mathbf{P}(A_j | D_m) \right] I_{D_m} \\ &= \sum_{i=1}^k y'_i I_{D_i} \sum_{j=1}^l x_j \mathbf{P}(A_j | D_i) = \sum_{j=1}^l \sum_{i=1}^k x_j y'_i \mathbf{P}(A_j | D_i) I_{D_i} \end{aligned}$$

espressione coincidente con quella di  $\mathbf{E}(\xi\eta | \mathcal{D})$ .  $\square$

**I.9.11 Osservazione:** Date due decomposizioni  $\mathcal{D}_1$  e  $\mathcal{D}_2$ , con  $\mathcal{D}_1 \preceq \mathcal{D}_2$ , ed una generica v.a.  $\xi$ , la v.a.  $\mathbf{E}(\xi \mid \mathcal{D}_1)$  risulta sempre  $\mathcal{D}_1$ -mis e quindi anche  $\mathcal{D}_2$ -mis. Conseguentemente dalla Proposizione I.9.10 discende che

$$\mathbf{E}[\mathbf{E}(\xi \mid \mathcal{D}_1) \mid \mathcal{D}_2] = \mathbf{E}(\xi \mid \mathcal{D}_1).$$

Questa proprietà resta vera anche se si inverte l'ordine delle decomposizioni al primo membro, benché la dimostrazione, in questo secondo caso, sia meno immediata (vedi la successiva Proposizione I.9.12). In conclusione possiamo dire che, nell'esecuzione ripetuta VdA condizionati rispetto a date decomposizioni, prevale sempre l'azione della decomposizione meno fine  $\mathcal{D}_1$  qualunque sia l'ordine dei condizionamenti. Bisogna comunque avvertire che questa proprietà non è così generale come appare a prima vista perché, date due decomposizioni, non è sempre detto che una delle due sia meno fine dell'altra: quando nessuna delle due relazioni  $\mathcal{D}_1 \preceq \mathcal{D}_2$  oppure  $\mathcal{D}_2 \preceq \mathcal{D}_1$  è soddisfatta, la proprietà qui enunciata ovviamente non è verificata.  $\circ$

**I.9.12 Proposizione:** Date due decomposizioni  $\mathcal{D}_1$  e  $\mathcal{D}_2$ , con  $\mathcal{D}_1 \preceq \mathcal{D}_2$ , risulta sempre

$$\mathbf{E}[\mathbf{E}(\xi \mid \mathcal{D}_2) \mid \mathcal{D}_1] = \mathbf{E}(\xi \mid \mathcal{D}_1)$$

qualunque sia la v.a.  $\xi$ .

**Dimostrazione:** Siano  $\mathcal{D}_1 = \{D_1^1, \dots, D_m^1\}$  e  $\mathcal{D}_2 = \{D_1^2, \dots, D_n^2\}$  le due decomposizioni in questione. Siccome  $\mathcal{D}_1 \preceq \mathcal{D}_2$ , sarà  $m \leq n$  e inoltre ogni  $D_p^1$  (con  $p = 1, \dots, m$ ) si decomporrà in un numero finito di  $D_q^2$  (con  $q = 1, \dots, n$ ). Inoltre sia

$$\xi = \sum_{j=1}^l x_j I_{A_j}, \quad A_j = \{\xi = x_j\}$$

la v.a. considerata. Siccome

$$\mathbf{E}(\xi \mid \mathcal{D}_2) = \sum_{j=1}^l x_j \mathbf{P}(A_j \mid \mathcal{D}_2),$$

per dimostrare la proposizione basterà provare che

$$\mathbf{E}[\mathbf{P}(A_j \mid \mathcal{D}_2) \mid \mathcal{D}_1] = \mathbf{P}(A_j \mid \mathcal{D}_1), \quad j = 1, \dots, l.$$

Tenendo conto della Definizione I.9.1 e della 4) di I.9.7 si ottiene infatti

$$\begin{aligned} \mathbf{E}[\mathbf{P}(A_j \mid \mathcal{D}_2) \mid \mathcal{D}_1] &= \sum_{q=1}^n \mathbf{P}(A_j \mid D_q^2) \mathbf{E}(I_{D_q^2} \mid \mathcal{D}_1) \\ &= \sum_{q=1}^n \mathbf{P}(A_j \mid D_q^2) \mathbf{P}(D_q^2 \mid \mathcal{D}_1) = \sum_{q=1}^n \mathbf{P}(A_j \mid D_q^2) \sum_{p=1}^m \mathbf{P}(D_q^2 \mid D_p^1) I_{D_p^1} \end{aligned}$$

$$\begin{aligned}
 &= \sum_{p=1}^m I_{D_p^1} \sum_{q=1}^n \mathbf{P}(A_j | D_q^2) \mathbf{P}(D_q^2 | D_p^1) \\
 &= \sum_{p=1}^m I_{D_p^1} \sum_{q: D_q^2 \subseteq D_p^1} \mathbf{P}(A_j | D_q^2) \mathbf{P}(D_q^2 | D_p^1) \\
 &= \sum_{p=1}^m I_{D_p^1} \sum_{q: D_q^2 \subseteq D_p^1} \frac{\mathbf{P}(A_j D_q^2) \mathbf{P}(D_q^2)}{\mathbf{P}(D_q^2) \mathbf{P}(D_p^1)} = \sum_{p=1}^m I_{D_p^1} \mathbf{P}(A_j | D_p^1) = \mathbf{P}(A_j | \mathcal{D}_1),
 \end{aligned}$$

dove abbiamo fatto anche uso dell'osservazione secondo la quale (essendo  $\mathcal{D}_1 \preceq \mathcal{D}_2$ ) ogni  $D_q^2$  è interamente contenuto in un  $D_p^1$ .  $\square$

**I.9.13 Osservazione:** È facile verificare che se  $\xi$  è indipendente da  $\eta$  (nel senso di I.5.14 ed I.5.15) risulta sempre

$$\mathbf{E}(\xi | \eta) = \mathbf{E} \xi,$$

mentre da I.9.10 si ha  $\mathbf{E}(\xi\eta | \eta) = \eta \mathbf{E}(\xi | \eta)$  e  $\mathbf{E}(\eta | \eta) = \eta$ . Infine, dato che ovviamente risulta  $\mathcal{D}_{\eta_1} \preceq \mathcal{D}_{\eta_1 \eta_2}$ , si verifica anche facilmente da I.9.12 che la relazione

$$\mathbf{E}[\mathbf{E}(\xi | \eta_1 \eta_2) | \eta_1] = \mathbf{E}(\xi | \eta_1),$$

ed altre analoghe sono sempre verificate.  $\circ$

**I.9.14 Osservazione:** Date due v.a. con la solita notazione

$$\xi = \sum_{j=1}^l x_j I_{A_j}, \quad \eta = \sum_{i=1}^n y_i I_{B_i}$$

potremo scrivere che

$$\mathbf{E}(\xi | \eta) = \sum_{i=1}^n \mathbf{E}(\xi | B_i) I_{B_i};$$

cioè  $\mathbf{E}(\xi | \eta)$  assume il valore (costante)  $\mathbf{E}(\xi | B_i) = \mathbf{E}(\xi | \eta = y_i)$  su ogni  $B_i$ . Consideriamo ora la seguente funzione:

$$m(y) = \begin{cases} \mathbf{E}(\xi | B_i) = \mathbf{E}(\xi | \eta = y_i), & \text{se } y = y_i \text{ con } i = 1, \dots, n, \\ \text{valore arbitrario,} & \text{se } y \neq y_i \text{ con } i = 1, \dots, n; \end{cases}$$

ad esempio, in particolare, potremmo considerare la funzione

$$m(y) = \begin{cases} \mathbf{E}(\xi | B_i) = \mathbf{E}(\xi | \eta = y_i), & \text{se } y = y_i \text{ con } i = 1, \dots, n, \\ 0, & \text{se } y \neq y_i \text{ con } i = 1, \dots, n. \end{cases}$$

È ora possibile anche introdurre la v.a.  $m(\eta)$  (vedi capitolo I.5) che naturalmente risulterà definita senza ambiguità dato che i valori possibili di  $\eta$  sono proprio le



$y_i$  in cui  $m(y)$  è stata definita in modo univoco. Ma in conseguenza di questa definizione è facile ora vedere che  $m(\eta)$  assume (costantemente) proprio il valore  $\mathbf{E}(\xi | B_i) = \mathbf{E}(\xi | \eta = y_i)$  su ogni  $B_i$  e quindi che essa coincide con  $\mathbf{E}(\xi | \eta)$ . In definitiva possiamo dire che la v.a.  $\mathbf{E}(\xi | \eta)$  è una funzione  $m(\eta)$  della v.a.  $\eta$ ; più precisamente la funzione  $m(y)$  che vale  $\mathbf{E}(\xi | \eta = y)$  per tutti i valori di  $y$  per i quali risulta  $\mathbf{P}\{\eta = y\} \neq 0$ .  $\circ$

**I.9.15 Esempio:** Continuando l'Esempio I.9.4 potremo ora determinare il VdA condizionato  $\mathbf{E}(\xi + \eta | \eta)$  come funzione di  $\eta$ . Infatti da I.9.4 segue che

$$\mathbf{E}(\xi + \eta | \eta) = \sum_{k=0}^2 k\mathbf{P}(\xi + \eta = k | \eta) = p(1 - \eta) + (1 - p)\eta + 2p\eta = p + \eta.$$

Tale risultato poteva anche essere dedotto dalle proprietà dei VdA condizionati osservando che a causa della indipendenza di  $\xi$  da  $\eta$  si ha

$$\mathbf{E}(\xi + \eta | \eta) = \mathbf{E}(\xi | \eta) + \mathbf{E}(\eta | \eta) = \mathbf{E}\xi + \mathbf{E}(\eta | \eta) = p + \eta.$$

In questo caso la scelta più naturale per la funzione  $m(y)$  è ovviamente  $p + y$  per ogni  $y$ , ma l'essenziale è che  $m(y)$  assuma i valori  $p + y_i$  per  $y = y_i$  (se  $y_i$  sono i possibili valori di  $\eta$ ), mentre può assumere valori arbitrari per  $y \neq y_i$ .  $\diamond$

**I.9.16 Osservazione:** Come già osservato nel Capitolo I.2 c'è una corrispondenza biunivoca tra le decomposizioni  $\mathcal{D}$  di uno spazio finito  $\Omega$  e le algebre di parti di  $\Omega$ . Anzi, data  $\mathcal{D}$  è sempre possibile costruire la cosiddetta algebra generata da  $\mathcal{D}$ , cioè  $\mathcal{A} = \alpha(\mathcal{D})$ , e viceversa data un'algebra  $\mathcal{A}$  è sempre possibile determinare un'unica decomposizione  $\mathcal{D}$  che la genera. Per questo motivo, data una v.a.  $\xi$ , potremo sempre parlare (ed in modo univoco) sia di decomposizione indotta  $\mathcal{D}_\xi$  che di algebra indotta  $\mathcal{A}_\xi$ . Questo ci consente ora di dare alle notazioni adottate in questo capitolo una forma diversa che, a prima vista, può apparire solo come una inutile complicazione, ma che, come vedremo in dettaglio in seguito, si rivelerà essenziale per la definizione del concetto di condizionamento nel caso di spazio non finiti nel quale la corrispondenza fra algebre e decomposizioni non è più così semplice. La nuova notazione consiste semplicemente nel sostituire il simbolo  $\mathcal{A}$  al simbolo  $\mathcal{D}$  in tutti i condizionamenti definiti in precedenza. Così ad esempio scriveremo  $\mathbf{E}(\xi | \mathcal{A})$  per indicare  $\mathbf{E}(\xi | \mathcal{D})$  se  $\mathcal{A} = \alpha(\mathcal{D})$  e parleremo di **VdA condizionato rispetto all'algebra  $\mathcal{A}$** .  $\circ$



## I.10 Probabilità di Rovina

In questo capitolo esamineremo un altro esempio di quella classe di problemi che fanno uso di Modelli finiti di Bernoulli per formulare delle stime probabilistiche ed, eventualmente, per prendere delle decisioni. Naturalmente, siccome il nostro modello sarà fondato su spazi finiti di probabilità, dovremo esercitare tutta la necessaria cautela quando esprimeremo risultati validi *al limite per*  $n \rightarrow \infty$ , come abbiamo ripetutamente osservato in I.7.6 e I.8.

Riprenderemo dunque la notazione usata nel Capitolo I.7 semplificandola innanzitutto con la soppressione degli indici  $(n)$  in alto: così ad esempio  $(\Omega, \mathcal{A}, \mathbf{P})$  sarà lo spazio finito relativo agli  $n$  tentativi indipendenti di Bernoulli ed ogni  $\omega \in \Omega$  sarà una  $n$ -pla  $\{\omega_1, \dots, \omega_n\}$  di risultati di tali tentativi. Modificheremo, inoltre, la definizione delle v.a.  $\xi_i$  con  $i = 1, \dots, n$  in un modo che richiama la discussione dell'Esempio I.3.4 sulle *passegiate aleatorie*, cioè ora le  $\xi_i$ , invece di essere degli indicatori, saranno v.a. i.i.d. tali che

$$\xi_i = \begin{cases} +1, & \text{con probabilità } p, \\ -1, & \text{con probabilità } q = 1 - p, \end{cases}$$

In tal caso, posto

$$S_n(\omega) = \sum_{i=1}^n \xi_i(\omega)$$

e  $\nu(\omega) = \frac{1}{2}[n + S_n(\omega)]$ , il modello risulta definito dall'assegnazione

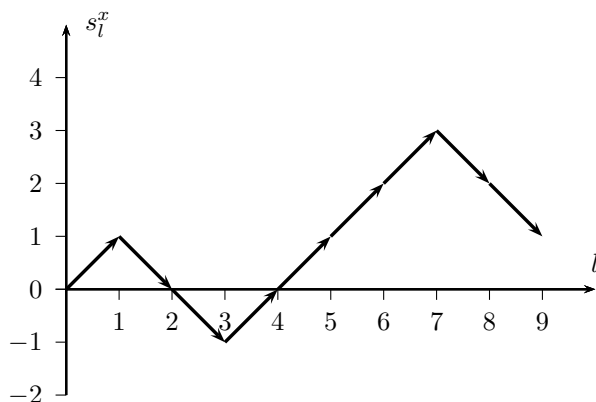
$$\mathbf{P}\{\omega\} = p^{\nu(\omega)}(1-p)^{n-\nu(\omega)},$$

(vedi l'analoga discussione in I.3.4). Indicheremo poi con  $x_i = \pm 1$  i valori numerici delle v.a.  $\xi_i$  e porremo infine

$$S_l = \sum_{i=1}^l \xi_i, \quad l = 1, \dots, n,$$
$$S_0 = 0,$$

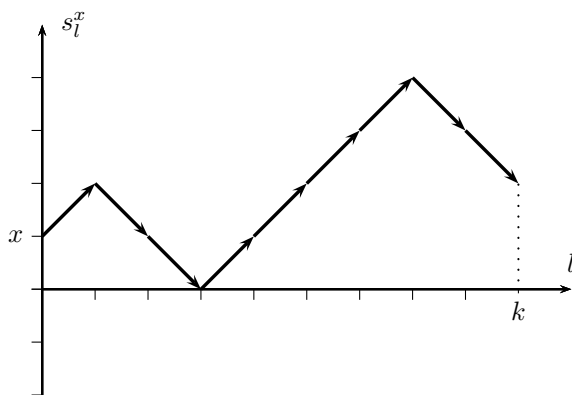
sicché avremo anche  $S_{l+1} = S_l + \xi_{l+1}$ . I generici valori numerici di  $S_l$  saranno indicati con  $s_l$ . Come già notato in I.3.4 questo modello consente di discutere problemi relativi alle traiettorie di percorsi aleatori unidimensionali, ma ora noi lo utilizzeremo per discutere l'andamento di un *Gioco d'Azzardo* fra due giocatori (che indicheremo con I e II): ad ogni *mano* di gioco ( $i = 1, \dots, n$  sarà l'indice che individua ogni mano in una partita di  $n$  mani) chi perde paga un'unità a chi vince. In tal caso la v.a.  $\xi_i$  rappresenterà la *vincita* (con segno, nel senso che, se negativa, essa rappresenta una *perdita*) di uno dei due giocatori (che convenzionalmente qui considereremo essere I) alla mano  $i$ -ma. Conseguentemente  $S_l$  rappresenterà la vincita totale (con segno) del giocatore I dopo  $l$  mani di gioco. Naturalmente ci

limiteremo ad esaminare le storie di gioco del solo giocatore I perchè, essendo il gioco qui descritto *a somma zero*, le vincite del giocatore II si possono dedurre da quelle di I per semplice differenza.



**Fig. I.10.1** Storia di gioco con vincita iniziale nulla.

Se ora scegliamo un ben determinato  $\omega$  ( $n$ -pla di mani di gioco) i numeri interi  $S_l(\omega) = s_l$  con  $l = 1, \dots, n$  rappresenteranno una ben determinata *storia di gioco* la cui traiettoria può essere rappresentata anche graficamente riportando su un diagramma cartesiano i valori interi  $s_l$  di  $S_l$  in funzione di  $l$  (vedi Fig. I.10.1).



**Fig. I.10.2** Storia di gioco con vincita iniziale  $x$ .

L'ipotesi  $S_0 = 0$  corrisponde alla supposizione che all'istante iniziale il giocatore I parta con vincita nulla. È quindi possibile generalizzare il nostro modello supponendo di osservare la storia del nostro gioco a partire dal momento in cui la vincita di I è già  $x$ , dove  $x$  è un numero intero: questa generalizzazione è significativa in quanto è possibile che un giocatore voglia prendere delle decisioni a partire da un determinato stadio avanzato del gioco e non dal suo effettivo inizio. In tal caso sarà utile definire le v.a.  $S_l^x = x + S_l$  le cui traiettorie  $s_l^x$  (vedi Fig. I.10.2)

rappresenteranno ora altrettante storie di gioco che iniziano con una vincita  $x$ . Consideriamo ora due numeri interi  $A \leq 0 \leq B$  (con  $A < B$  per non ridurci al caso banale  $A = B = 0$ , e con  $A \leq x \leq B$ ) e poniamoci il problema di stimare la probabilità che  $S_l^x$  raggiunga il valore  $A$  prima del valore  $B$  (o viceversa) e il numero di mani di gioco necessarie. Nel nostro modello di gioco d'azzardo questa domanda ha una traduzione in termini di **rovina di uno dei due giocatori**<sup>1</sup>. Infatti possiamo supporre che ciascuno dei due giocatori parta con una quantità di denaro iniziale: I parte con  $-A$  unità (ricorda che  $A \leq 0$ ) e II con  $B$  unità. Quando  $S_l^x = A$  il giocatore I è *rovinato* in quanto la sua *vincita* è di segno negativo (è in realtà una *perdita*) e di valore assoluto pari al suo capitale iniziale. In tal caso possiamo supporre che il gioco si arresti. Quando invece  $S_l^x = B$  il giocatore rovinato è II e il gioco si arresta egualmente. Graficamente è possibile esprimere questo problema riportando sul diagramma cartesiano delle storie di gioco due *barriere* in  $s = A$  ed  $s = B$ : il gioco si arresta quando la traiettoria tocca una delle due barriere (rovina di uno dei due giocatori: vedi Fig. I.10.3).

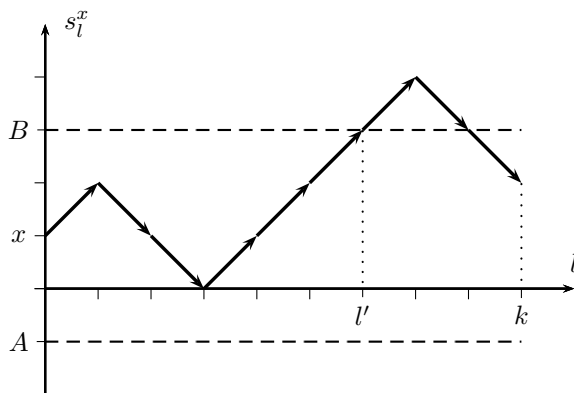


Fig. I.10.3 Tempo d'arresto  $\tau_k^x = l'$  del gioco.

Due osservazioni sono d'obbligo a questo punto: la prima è che, se il numero di mani  $n$  è finito e fissato dai due giocatori all'inizio della partita, è possibile (con probabilità diversa da zero) che il gioco termini senza che nessuno dei due si sia rovinato. Questo tipo di problemi con un numero finito (e fissato) di mani di gioco è ovviamente l'unico che è possibile affrontare in modo rigoroso al nostro livello di discussione, in quanto coinvolge solo spazi finiti di probabilità. Ciononostante,

<sup>1</sup> Siamo qui in presenza di una particolare versione del celebre problema della *suddivisione della posta in gioco* formulato già nel XIV e XV secolo da Pacioli, Tartaglia e Cardano e sottoposto dal Cavalier de Méré a Fermat e Pascal come già ricordato nel Capitolo I.1: *se due giocatori I e II giocano una successione di partite a testa e croce in modo che il primo dei due che vince un prefissato numero di partite guadagna la totalità della posta in gioco, e se per qualche motivo i due decidono di interrompere il gioco prima della fine, come si deve suddividere la posta per far in modo che la ripartizione sia equa?* Informazioni più precise su questo problema potranno essere reperite nella bibliografia indicata nel Capitolo I.1.

come abbiamo già fatto in precedenza per la Legge dei grandi Numeri e per i Teoremi limite, per raggiungere i risultati voluti non rispetteremo fino in fondo questa limitazione, rimandando ad una fase successiva la definizione di una base più precisa per le nostre deduzioni. La seconda osservazione riguarda invece il fatto che, se ci poniamo il problema di calcolare la **probabilità di rovina** di uno dei due giocatori (sia nel caso di un numero di mani di gioco limitato che nel caso di un numero di mani illimitato) dovremo tener conto del fatto che l'istante della rovina (ossia: la mano di gioco in cui uno dei giocatori si rovina) è anch'esso aleatorio e non può essere in alcun modo fissato a priori. Per questo motivo converrà introdurre una definizione che si rivelerà utile anche in seguito

**I.10.1 Definizione:** La v.a.

$$\tau_k^x = \min\{l : 0 \leq l \leq k ; S_l^x = A \text{ oppure } B\}, \quad 0 \leq k \leq n$$

prende il nome di **tempo d'arresto** (*stopping time*); se invece  $S_l^x$  non raggiunge mai né il valore  $A$  né il valore  $B$  per  $l = 0, 1, \dots, k$  si pone  $\tau_k^x = k$ .  $\triangle$

Naturalmente i possibili valori di  $\tau_k^x$  sono gli interi  $0, 1, \dots, k$  e rappresentano l'istante d'arresto del gioco per rovina di uno dei due giocatori, o per esaurimento del tempo a disposizione, entro  $k$  mani di gioco (con  $k \leq n$ ) e partendo con una vincita  $x$  di I. Questo concetto ci consente ora di dare un'opportuna definizione (con  $0 \leq l \leq k$ ) ad alcuni eventi utili per i calcoli successivi:

$$\begin{aligned} \{\tau_k^x = l\} &= \text{il gioco si arresta all}'l\text{-ma mano,} \\ \{\tau_k^x = l, S_l^x = A\} &= \text{il gioco si arresta all}'l\text{-ma mano per rovina di I,} \\ \{\tau_k^x = l, S_l^x = B\} &= \text{il gioco si arresta all}'l\text{-ma mano per rovina di II,} \\ A_k^x &= \bigcup_{l=0}^k \{\tau_k^x = l, S_l^x = A\} = \text{I si rovina entro la } k\text{-ma mano,} \\ B_k^x &= \bigcup_{l=0}^k \{\tau_k^x = l, S_l^x = B\} = \text{II si rovina entro la } k\text{-ma mano.} \end{aligned}$$

Pertanto i numeri

$$\alpha_k(x) = \mathbf{P}(A_k^x), \quad \beta_k(x) = \mathbf{P}(B_k^x)$$

rappresenteranno le probabilità di rovina, entro la  $k$ -ma mano, rispettivamente di I e II se si inizia con una vincita  $x$  di I. Ovviamente il nostro problema è quello di dare una stima per i numeri  $\alpha_k(x)$  e  $\beta_k(x)$ .

Per far questo osserviamo che il caso in cui  $x = A$  oppure  $x = B$  è banale: infatti in tal caso si partirebbe con uno dei due giocatori già rovinato ed il calcolo sarebbe immediato. In pratica basta osservare che  $A_k^A = B_k^B = \Omega$  per cui  $\alpha_k(A) = \beta_k(B) = 1$ . Peraltro, se supponiamo che  $A < x < B$ , è anche evidente che  $\alpha_0(x) = \beta_0(x) = 0$  visto che, se non si parte con uno dei due giocatori già rovinato, è anche impossibile che uno dei due si rovini in un numero nullo di mani di gioco (in

questo caso abbiamo infatti che  $A_0^x = B_0^x = \emptyset$ ). Consideriamo allora il caso in cui  $1 \leq k \leq n$  e  $x \neq A, B$ , e tentiamo di determinare le equazioni soddisfatte da  $\beta_k(x)$  (quelle per  $\alpha_k(x)$  si ricaveranno in modo analogo): siccome  $S_1^x = x + S_1 = x + \xi_1$  avremo anche, dalla formula della probabilità totale (vedi I.4.5), che

$$\begin{aligned}\beta_k(x) &= \mathbf{P}(B_k^x) = \mathbf{P}(B_k^x \mid \xi_1 = +1)\mathbf{P}(\xi_1 = 1) + \mathbf{P}(B_k^x \mid \xi_1 = -1)\mathbf{P}(\xi_1 = -1) \\ &= p\mathbf{P}(B_k^x \mid S_1^x = x + 1) + q\mathbf{P}(B_k^x \mid S_1^x = x - 1).\end{aligned}$$

Possiamo ora provare che

$$\begin{aligned}\mathbf{P}(B_k^x \mid S_1^x = x + 1) &= \mathbf{P}(B_{k-1}^{x+1}) = \beta_{k-1}(x + 1), \\ \mathbf{P}(B_k^x \mid S_1^x = x - 1) &= \mathbf{P}(B_{k-1}^{x-1}) = \beta_{k-1}(x - 1).\end{aligned}$$

Infatti, detto  $\mathcal{C}_k^x$  l'insieme delle traiettorie  $s_i^x$  di storie di gioco di  $k$  mani che stanno in  $B_k^x$  (cioè l'insieme delle  $(k+1)$ -ple  $(x, x+x_1, \dots, x+x_1+\dots+x_k)$  di numeri interi con  $x$  fissato,  $x_i = \pm 1$  e  $i = 1, \dots, k$ , tali che almeno una delle  $k+1$  componenti sia eguale a  $B$  senza che nessun'altra componente precedente sia eguale ad  $A$ ), potremo anche caratterizzare l'evento  $B_k^x$  nel modo seguente:

$$\begin{aligned}B_k^x &= \{\omega \in \Omega : (x, x + \xi_1, \dots, x + \xi_1 + \dots + \xi_k) \in \mathcal{C}_k^x\} \\ &= \{\omega \in \Omega : (x, x + S_1, \dots, x + S_k) \in \mathcal{C}_k^x\}.\end{aligned}$$

In questo modo, utilizzando l'ovvia relazione

$$\begin{aligned}\{\omega \in \Omega : (x, x + 1, x + 1 + \xi_2, \dots, x + 1 + \xi_2 + \dots + \xi_k) \in \mathcal{C}_k^x\} \\ = \{\omega \in \Omega : (x + 1, x + 1 + \xi_2, \dots, x + 1 + \xi_2 + \dots + \xi_k) \in \mathcal{C}_{k-1}^{x+1}\},\end{aligned}$$

e utilizzando i risultati dell'Esempio I.9.4, avremo che

$$\begin{aligned}\mathbf{P}(B_k^x \mid S_1^x = x + 1) &= \mathbf{P}(B_k^x \mid \xi_1 = 1) \\ &= \mathbf{P}((x, x + \xi_1, x + \xi_1 + \xi_2, \dots, x + \xi_1 + \xi_2 + \dots + \xi_k) \in \mathcal{C}_k^x \mid \xi_1 = 1) \\ &= \mathbf{P}((x, x + 1, x + 1 + \xi_2, \dots, x + 1 + \xi_2 + \dots + \xi_k) \in \mathcal{C}_k^x) \\ &= \mathbf{P}((x + 1, x + 1 + \xi_2, \dots, x + 1 + \xi_2 + \dots + \xi_k) \in \mathcal{C}_{k-1}^{x+1});\end{aligned}$$

se infine ricordiamo che le  $\xi_i$  sono tutte indipendenti ed identicamente distribuite, otteniamo il risultato richiesto:

$$\begin{aligned}\mathbf{P}(B_k^x \mid S_1^x = x + 1) &= \mathbf{P}((x + 1, x + 1 + \xi_1, \dots, x + 1 + \xi_1 + \dots + \xi_{k-1}) \in \mathcal{C}_{k-1}^{x+1}) \\ &= \mathbf{P}((x + 1, x + 1 + S_1, \dots, x + 1 + S_{k-1}) \in \mathcal{C}_{k-1}^{x+1}) \\ &= \mathbf{P}(B_{k-1}^{x+1}) = \beta_{k-1}(x + 1).\end{aligned}$$

La seconda relazione si prova in modo del tutto analogo. Pertanto si ottengono le seguenti *relazioni di ricorrenza*:

$$\beta_k(x) = p\beta_{k-1}(x + 1) + q\beta_{k-1}(x - 1), \quad 1 \leq k \leq n,$$

assieme alle condizioni

$$\begin{aligned}\beta_k(A) &= 0, & \beta_k(B) &= 1, & 0 \leq k \leq n \\ \beta_0(x) &= 0, & A < x < B,\end{aligned}$$

dovute alle ovvie osservazioni secondo cui se uno dei due giocatori parte rovinato, il suo avversario non si rovinerà mai durante il gioco e che se nessuno dei due parte rovinato il gioco non può arrestarsi senza aver giocato neanche una mano. Analogamente per  $\alpha_k(x)$  si ricava

$$\alpha_k(x) = p\alpha_{k-1}(x+1) + q\alpha_{k-1}(x-1), \quad 1 \leq k \leq n,$$

assieme alle condizioni

$$\begin{aligned}\alpha_k(A) &= 1, & \alpha_k(B) &= 0, & 0 \leq k \leq n \\ \alpha_0(x) &= 0, & A < x < B.\end{aligned}$$

Queste equazioni possono essere facilmente risolte per ricorrenza, a partire dalle condizioni date, e forniscono  $\alpha_1(x), \dots, \alpha_n(x), \beta_1(x), \dots, \beta_n(x)$ . Noi però lasceremo come esercizio per il lettore un esame più attento di queste soluzioni<sup>2</sup> e procederemo invece ad esaminare il nostro problema nell'ipotesi che  $n \rightarrow \infty$ , cioè nell'ipotesi che non vi sia limite sul numero delle mani di gioco che è possibile effettuare. Naturalmente anche qui si pone il problema della probabilità di rovina di uno dei due giocatori, ma esso sarà ora riformulato nel modo seguente: determinare la probabilità che, partendo con una vincita  $x$ , uno dei due giocatori si rovini in un numero di mani *qualsiasi purché finito*.

Rimandando come al solito ad un momento successivo la dimostrazione del fatto che uno spazio di probabilità che descriva un gioco con un numero di mani potenzialmente illimitato è in realtà possibile, ci limiteremo ad osservare che, essendo ovviamente  $B_{k-1}^x \subseteq B_k^x$ , risulta sempre che  $\beta_{k-1}(x) \leq \beta_k(x) \leq 1$ , sicché, con  $x$  arbitrario ma fissato, le  $\beta_k(x)$  costituiscono una successione monotona non decrescente e limitata. Un'analogha conclusione si può trarre sulla successione delle  $\alpha_k(x)$ . Ci attendiamo pertanto che  $\beta_k(x) \rightarrow \beta(x)$  e  $\alpha_k(x) \rightarrow \alpha(x)$ , per  $k \rightarrow \infty$ , dove  $\beta(x)$  e  $\alpha(x)$  rappresentano ora la probabilità di rovina (in un numero arbitrario ma finito di mani), rispettivamente di II e di I, quando il gioco comincia con una vincita  $x$  di I e si suppone di poter giocare un numero illimitato di mani<sup>3</sup>. In tal caso le nostre equazioni divengono

$$\beta(x) = p\beta(x+1) + q\beta(x-1), \quad A < x < B,$$

---

<sup>2</sup> In particolare sarebbe interessante mostrare che  $\alpha_n(x) + \beta_n(x) < 1$  per  $n$  finito e  $A < x < B$ , cioè che la probabilità dell'evento *nessuno dei due giocatori si rovina in  $n$  di mani* è diversa da zero. Mostriamo, invece, che quando il numero di mani è illimitato la probabilità di tale evento (che *non* è l'evento assurdo) è zero.

<sup>3</sup> Il ragionamento presentato nelle righe precedenti non può essere considerato conclusivo nella teoria degli spazi *finiti* di probabilità in quanto le maggiorazioni considerate presuppongono la possibilità di associare delle probabilità ad eventi di uno spazio *non finito* visto che si suppone di poter giocare un numero illimitato di mani.



con le condizioni  $\beta(A) = 0$  e  $\beta(B) = 1$ , e

$$\alpha(x) = p\alpha(x+1) + q\alpha(x-1), \quad A < x < B,$$

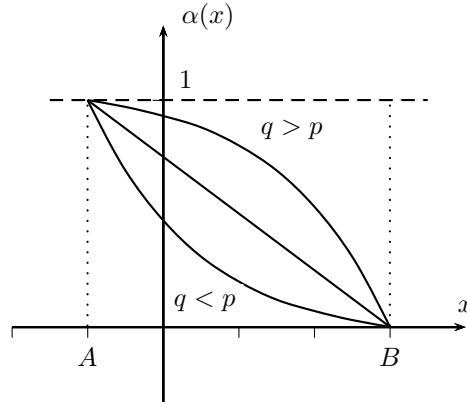
con le condizioni  $\alpha(A) = 1$  e  $\alpha(B) = 0$ . Per risolvere queste equazioni distingueremo due casi:  $p = q = \frac{1}{2}$  (gioco equo) e  $p \neq q$  (gioco non equo), trattandoli separatamente. Nel caso di *gioco non equo* ( $p \neq q$ ) si verifica facilmente per calcolo diretto che  $\beta(x) = a$  e  $\beta(x) = b(q/p)^x$  sono sempre soluzioni della prima equazione se  $a$  e  $b$  sono costanti reali arbitrarie. Inoltre si controlla che anche  $\beta(x) = a + b(q/p)^x$  è una soluzione e, imponendo le condizioni ai limiti, si ricava la soluzione del nostro problema:

$$\beta(x) = \frac{(q/p)^x - (q/p)^A}{(q/p)^B - (q/p)^A}, \quad A \leq x \leq B.$$

Analogamente la soluzione della seconda equazione è

$$\alpha(x) = \frac{(q/p)^B - (q/p)^x}{(q/p)^B - (q/p)^A}, \quad A \leq x \leq B.$$

Il grafico della funzione  $\alpha(x)$  per diversi valori del rapporto  $q/p$  è mostrato in Fig. I.10.4.



**Fig. I.10.4** Probabilità di rovina  $\alpha(x)$  del giocatore I.

Possiamo anche mostrare che queste sono le *uniche* soluzioni del nostro problema: infatti, ad esempio nel caso di  $\beta(x)$ , faremo vedere che ogni soluzione deve essere della forma  $a + b(q/p)^x$  e quindi, date le condizioni ai limiti, deve coincidere con quella da noi calcolata. Per verificarlo osserviamo che, se  $\tilde{\beta}(x)$  è una soluzione della nostra equazione con le condizioni  $\tilde{\beta}(A) = 0$  e  $\tilde{\beta}(B) = 1$ , essendo  $q/p \neq 0$ , potremo sempre determinare due numeri  $\tilde{a}$  e  $\tilde{b}$  tali che

$$\tilde{a} + \tilde{b}(q/p)^A = \tilde{\beta}(A) = 0, \quad \tilde{a} + \tilde{b}(q/p)^{A+1} = \tilde{\beta}(A+1);$$

inoltre, dalle equazioni soddisfatte da  $\tilde{\beta}(x)$ , si ha che

$$\begin{aligned}\tilde{\beta}(A+2) &= \frac{\tilde{\beta}(A+1)}{p} = \frac{\tilde{a} + \tilde{b}(q/p)^{A+1}}{p} \\ &= \tilde{a} + \tilde{a}q/p + \tilde{b}(q/p)^{A+1} + \tilde{b}(q/p)^{A+2} \\ &= \tilde{a} + \tilde{b}(q/p)^{A+2} + \frac{q}{p}[\tilde{a} + \tilde{b}(q/p)^A] \\ &= \tilde{a} + \tilde{b}(q/p)^{A+2} + \tilde{\beta}(A)q/p = \tilde{a} + \tilde{b}(q/p)^{A+2}.\end{aligned}$$

Iterando il procedimento si prova quindi che

$$\tilde{\beta}(x) = \tilde{a} + \tilde{b}(q/p)^x, \quad A \leq x \leq B$$

e imponendo le condizioni ai limiti si ritrova la soluzione data prima che risulta quindi unica.

Le soluzioni nel caso di *gioco equo* ( $p = q = \frac{1}{2}$ ) possono essere ricavate dalle soluzioni del caso precedente al limite per  $p \rightarrow \frac{1}{2}$  mediante la regola di l'Hôpital<sup>4</sup> (nota che le soluzioni del caso precedente non sono definite per  $p = q = \frac{1}{2}$ ), ma si può anche verificare per calcolo diretto che le uniche soluzioni (con  $A \leq x \leq B$ ) sono:

$$\beta(x) = \frac{x-A}{B-A}, \quad \alpha(x) = \frac{B-x}{B-A}.$$

Osserviamo in particolare che, in ambedue i casi, risulta sempre  $\alpha(x) + \beta(x) = 1$ , cioè (differentemente da quel che succede nel caso di un gioco con un numero limitato di mani), se il numero di mani possibili è illimitato la probabilità che nessuno dei due giocatori si rovini è zero. In altri termini: un gioco illimitato di questo tipo conduce alla rovina di uno dei due giocatori (in un tempo arbitrario ma finito) con probabilità eguale ad uno.

Le soluzioni trovate ci consentono ora di dare risposte ad alcuni semplici problemi di *strategia di gioco*; ad esempio chiediamoci se, supponendo che il gioco sia sfavorevole al giocatore I ( $q > p$ ) e partendo con  $x = 0$ , è possibile determinare una strategia di gioco che renda minima la probabilità di rovina  $\alpha(0)$  di I. Questa domanda è significativa perché il fatto che il gioco sia sfavorevole a I non implica necessariamente che I si rovinerà prima di II: pertanto una modifica nella *strategia delle puntate* che faccia diminuire la probabilità di rovina può essere estremamente preziosa per il giocatore I. Dalla discussione precedente sappiamo che nelle condizioni date la probabilità di rovina di I è

$$\alpha = \alpha(0) = \frac{(q/p)^B - 1}{(q/p)^B - (q/p)^A}.$$

<sup>4</sup> Guillaume François Antoine de l'Hospital, Marquis de St-Mesme (1661 - 1704), noto anche ai posteri come Marquis de l'Hôpital, discendeva da un'antica ed onorata famiglia di nobili francesi. Essendogli stata preclusa la carriera militare a causa della debolezza della sua vista, egli dedicò la seconda parte della vita ai suoi prediletti studi di matematica. Allievo di Jean Bernoulli giocò un ruolo importante nell'introduzione in Francia delle nuove idee sull'analisi con la sua opera *Analyse des infiniment petits* (1696).

Domandiamoci allora in che misura tale  $\alpha$  è modificata se si dimezza la posta giocata in ogni mano (se si adotta, cioè, una strategia di gioco più *prudente* con puntate più piccole). In questo caso i calcoli per ottenere il risultato richiesto sono esattamente gli stessi con la differenza che i possibili valori delle v.a.  $\xi_i$ , invece di essere  $\pm 1$ , saranno  $\pm \frac{1}{2}$ . Ci si convince però facilmente del fatto che la modifica proposta è formalmente equivalente a quella che si otterrebbe lasciando invariata la posta in gioco ad ogni mano e raddoppiando i capitali iniziali che ora varranno  $2A$  e  $2B$ . Se allora indichiamo con  $\alpha'$  la nuova probabilità di rovina di I, essendo  $q/p > 1$  ed  $A < 0$ , si avrà

$$\begin{aligned}\alpha' &= \frac{(q/p)^{2B} - 1}{(q/p)^{2B} - (q/p)^{2A}} = \frac{(q/p)^B - 1}{(q/p)^B - (q/p)^A} \cdot \frac{(q/p)^B + 1}{(q/p)^B + (q/p)^A} \\ &= \alpha \frac{(q/p)^B + 1}{(q/p)^B + (q/p)^A} > \alpha,\end{aligned}$$

cioè la probabilità di rovina sarà aumentata. Pertanto, nel caso di gioco sfavorevole, converrà sempre tenere alta la posta in gioco ad ogni mano per migliorare le proprie possibilità di vittoria finale.

L'ultimo problema che ci proporremo sarà quello di stimare la **durata media del gioco** intesa come numero medio di mani di gioco necessarie a far sì che la partita si arresti per rovina di uno dei due giocatori. Anche in questo caso eseguiremo il calcolo supponendo dapprima che il numero di mani sia limitato ad  $n$  ed eseguendo poi il limite per  $n \rightarrow \infty$ . Se  $n$  è un numero finito, è chiaro che ciò che noi vogliamo è calcolare

$$m_k(x) = \mathbf{E} \tau_k^x, \quad 0 \leq k \leq n, \quad A \leq 0 \leq B.$$

Siccome  $\tau_k^x$  assume solo i valori  $l = 0, 1, \dots, k$ , supponendo che  $A < x < B$  (in modo che  $\tau_k^x$  assuma valore 0 con probabilità nulla), avremo

$$\begin{aligned}m_k(x) &= \sum_{l=1}^k l \mathbf{P}(\tau_k^x = l) \\ &= \sum_{l=1}^k l [p \mathbf{P}(\tau_k^x = l \mid \xi_1 = 1) + q \mathbf{P}(\tau_k^x = l \mid \xi_1 = -1)] \\ &= \sum_{l=1}^k l [p \mathbf{P}(\tau_{k-1}^{x+1} = l-1) + q \mathbf{P}(\tau_{k-1}^{x-1} = l-1)] \\ &= \sum_{l=0}^{k-1} (l+1) [p \mathbf{P}(\tau_{k-1}^{x+1} = l) + q \mathbf{P}(\tau_{k-1}^{x-1} = l)] \\ &= pm_{k-1}(x+1) + qm_{k-1}(x-1) + p \sum_{l=0}^{k-1} \mathbf{P}(\tau_{k-1}^{x+1} = l) + q \sum_{l=0}^{k-1} \mathbf{P}(\tau_{k-1}^{x-1} = l) \\ &= pm_{k-1}(x+1) + qm_{k-1}(x-1) + 1.\end{aligned}$$

Pertanto le relazioni di ricorrenza saranno

$$m_k(x) = 1 + pm_{k-1}(x+1) + qm_{k-1}(x-1), \quad 0 \leq k \leq n,$$

con le condizioni

$$\begin{aligned} m_k(A) = m_k(B) &= 0, & 0 \leq k \leq n, \\ m_0(x) &= 0, & A < x < B. \end{aligned}$$

Siccome inoltre  $\tau_k^x \leq \tau_{k+1}^x$  (perché se il gioco si arresta prima di  $k$  esso si arresta anche prima di  $k+1$ ), avremo che  $m_k(x) \leq m_{k+1}(x)$  con  $x$  arbitrario ma fissato. Pertanto la successione  $m_k(x)$  risulta monotona non decrescente per cui sicuramente esiste il  $\lim_n m_n(x) = m(x)$  (il cui valore, però, può anche essere  $+\infty$  dato che la successione non è limitata). La funzione  $m(x)$  soddisferà allora l'equazione

$$m(x) = 1 + pm(x+1) + qm(x-1)$$

con la condizione  $m(A) = m(B) = 0$ . È possibile anche mostrare (ma noi non lo faremo) che  $m(x)$  assume sempre valori finiti se  $A < x < B$ . Per risolvere l'equazione distingueremo, come al solito, due casi: quello di gioco equo e quello di gioco non equo. Se il gioco non è equo ( $p \neq q$ ), si verifica per calcolo diretto che

$$m(x) = \frac{x}{q-p} + a + b \left(\frac{q}{p}\right)^x$$

è sempre soluzione della nostra equazione con  $a$  e  $b$  costanti arbitrarie. Le condizioni richieste conducono poi alla soluzione

$$m(x) = \frac{B\beta(x) + A\alpha(x) - x}{p-q}$$

dove  $\alpha(x)$  e  $\beta(x)$  sono le probabilità di rovina calcolate precedentemente. Nel caso di gioco equo ( $p = q = \frac{1}{2}$ ), invece, si verifica che la soluzione ha la forma generale

$$m(x) = a + bx - x^2$$

che, a causa delle condizioni assegnate, diviene

$$m(x) = (B-x)(x-A).$$

In particolare, se i due giocatori partono con il medesimo capitale iniziale ( $B = -A$ ), se il gioco è equo e se la vincita iniziale è supposta nulla ( $x = 0$ ), la durata media del gioco è  $m(0) = B^2$ : supponendo ad esempio  $B = 10$ , potremo attenderci in media che il gioco si arresti dopo 100 mani.

## I.11 Esercizi svolti

**I.11.1 Esercizio (A.A. Sveshnikov:** *Problems in probability theory, mathematical statistics and theory of random functions*; Dover, New York, 1978, p. 3): Dato  $(\Omega, \mathcal{A})$ , determinare l'evento  $X$  che soddisfa la relazione

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

dove  $A$  e  $B$  sono due eventi di  $\mathcal{A}$ .

**Soluzione:** Delle leggi di De Morgan (vedi I.2) si ricava che la relazione da imporre si può anche scrivere come

$$\overline{(X \cup A)(X \cup \bar{A})} = B;$$

Ne segue che

$$\bar{B} = (X \cup A)(X \cup \bar{A}) = XX \cup X\bar{A} \cup AX \cup A\bar{A} = X \cup XA \cup X\bar{A} = X$$

per cui  $X = \bar{B}$ . ◇

**I.11.2 Esercizio (A.A. Sveshnikov:** *Problems in probability theory, mathematical statistics and theory of random functions*; Dover, New York, 1978, p. 4): Un cubo con le facce colorate viene diviso in 1 000 cubetti di eguali dimensioni. I cubetti così ottenuti sono poi mescolati: determinare la probabilità che un cubetto estratto a caso abbia due facce colorate.

**Soluzione:** Su 1 000 cubetti ve ne sono 8 con tre facce colorate,  $8 \cdot 12 = 96$  con due facce colorate,  $64 \cdot 6 = 384$  con una faccia colorata e  $8^3 = 512$  non colorati. Pertanto

$$p = \frac{96}{1\,000} = 0,096$$

è la probabilità richiesta. ◇

**I.11.3 Esercizio (A.A. Sveshnikov:** *Problems in probability theory, mathematical statistics and theory of random functions*; Dover, New York, 1978, p. 4): Determinare la probabilità che le ultime due cifre del cubo di un numero intero  $n$  preso a caso (*numero casuale o aleatorio*) siano ambedue 1.

**Soluzione:** Se  $n$  ha  $k \geq 2$  cifre (con  $k = 1$  nessuna cifra soddisfa le richieste del problema) la sua rappresentazione decimale è

$$n = a_0 + a_1 \cdot 10 + a_2 \cdot 10^2 + \dots + a_k \cdot 10^k$$

dove  $a_j = 0, 1, \dots, 9$  (con  $j = 1, \dots, k$ ). Diremo che  $n$  è casuale o aleatorio quando le  $a_j$  sono v.a. i.i.d. che assumono i loro 10 valori interi con probabilità eguale a

$\frac{1}{10}$ . Siccome il cubo di  $n$  è della forma<sup>1</sup>

$$n^3 = a_0^3 + 3a_0^2a_1 \cdot 10 + (3a_0^2a_2 + 3a_0a_1^2) \cdot 10^2 + \dots$$

i valori delle ultime due cifre di  $n^3$  saranno determinati unicamente da  $a_0$  ed  $a_1$ . I casi possibili sono quindi dati dalle coppie di numeri  $(a_0, a_1)$  che sono in numero di  $10^2 = 100$ . I casi favorevoli, invece, sono quelli in cui  $a_0 = 1$  e l'ultima cifra di  $3a_1$  è 1: l'unica possibilità è allora che  $a_1 = 7$ . Pertanto c'è solo un caso favorevole su 100: la coppia  $(1, 7)$ , e la sua probabilità è  $p = 0,01$ . In conclusione potremo dire che tutti i numeri (con  $k \geq 2$ ) che finiscono con 71 hanno cubi che finiscono con 11 e che la probabilità di estrarne uno a caso è  $p = 0,01$ .  $\diamond$

**I.11.4 Esercizio (A.A. Sveshnikov:** *Problems in probability theory, mathematical statistics and theory of random functions*; Dover, New York, 1978, p. 12): Nella produzione di un'azienda si distinguono i pezzi difettosi da quelli buoni e, fra quelli buoni, i pezzi di prima e di seconda qualità. Supponendo che i pezzi difettosi siano il 4% e che fra quelli buoni solo il 75% sono di prima qualità, calcolare la probabilità che un pezzo scelto a caso su tutta la produzione sia di prima qualità.

**Soluzione:** Definiti gli eventi

$$A = \text{il pezzo è buono,} \quad B = \text{il pezzo è di prima qualità}$$

risulta evidentemente  $B \subseteq A$  ed inoltre

$$\mathbf{P}(A) = 0,96; \quad \mathbf{P}(B | A) = 0,75.$$

Pertanto

$$\mathbf{P}(B) = \mathbf{P}(AB) = \mathbf{P}(B | A)\mathbf{P}(A) = 0,72$$

è la probabilità richiesta.  $\diamond$

**I.11.5 Esercizio (A.A. Sveshnikov:** *Problems in probability theory, mathematical statistics and theory of random functions*; Dover, New York, 1978, p. 12): Si vuole ispezionare una partita di 100 oggetti esaminando solo un campione composto di 5 oggetti scelti a caso fra i 100: la partita viene respinta se fra i 5 oggetti se ne trova almeno uno difettoso. Supponendo che il 5% degli oggetti è difettoso, qual è la probabilità  $p$  che la partita venga respinta?

**Soluzione:** Definiti gli eventi

$$A = \text{la partita viene accettata} \\ A_k = \text{il } k\text{-mo oggetto ispezionato è buono,} \quad k = 1, \dots, 5$$

<sup>1</sup> Nota che quella scritta non è la rappresentazione decimale di  $n^3$  in quanto, ad esempio, sia  $a_0^3$  che  $3a_0^2a_1$  possono assumere valori maggiori di 9.

risulta evidentemente che  $A = A_1 A_2 A_3 A_4 A_5$  e che, detto  $q = \mathbf{P}(A)$ , la probabilità richiesta è  $p = 1 - q$ . Applicando inoltre la formula di moltiplicazione (vedi I.4.8) si ha

$$q = \mathbf{P}(A_1 A_2 A_3 A_4 A_5) = \mathbf{P}(A_5 | A_4 A_3 A_2 A_1) \cdot \dots \cdot \mathbf{P}(A_2 | A_1) \cdot \mathbf{P}(A_1)$$

dove chiaramente è

$$\begin{aligned} \mathbf{P}(A_1) &= \frac{95}{100}, & \mathbf{P}(A_2 | A_1) &= \frac{94}{99}, & \mathbf{P}(A_3 | A_2 A_1) &= \frac{93}{98}, \\ \mathbf{P}(A_4 | A_3 A_2 A_1) &= \frac{92}{97}, & \mathbf{P}(A_5 | A_4 A_3 A_2 A_1) &= \frac{91}{96}. \end{aligned}$$

Risulta facilmente allora che  $q = 0,77$  e  $p = 0,23$ . Nota che è possibile ottenere il medesimo risultato con qualche osservazione di calcolo combinatorio: quando estraiamo i 5 pezzi da esaminare in realtà eseguiamo un *campionamento senza rimessa* nel quale non teniamo conto dell'ordine degli oggetti estratti (vedi I.2.4). Siccome, essendo 100 i pezzi disponibili, i casi possibili (cioè i modi in cui possiamo estrarre i 5 pezzi) sono  $\binom{100}{5}$ , mentre i casi favorevoli ad  $A$  sono i  $\binom{95}{5}$  modi in cui possiamo estrarre 5 pezzi fra i 95 buoni, avremo

$$q = \mathbf{P}(A) = \frac{\binom{95}{5}}{\binom{100}{5}} = \frac{95!}{5!90!} \frac{5!95!}{100!} = \frac{95 \cdot 94 \cdot 93 \cdot 92 \cdot 91}{100 \cdot 99 \cdot 98 \cdot 97 \cdot 96} = 0,77$$

risultato che conferma il calcolo precedente.  $\diamond$

**I.11.6 Esercizio (A.A. Sveshnikov:** *Problems in probability theory, mathematical statistics and theory of random functions*; Dover, New York, 1978, p. 16): Si ripeta il calcolo dell'Esercizio I.11.5 esaminando però un campione di 50 pezzi scelti a caso fra quelli della partita ed accettando la partita se si trova non più di un pezzo difettoso.

**Soluzione:** Riprendendo le notazioni dell'Esercizio I.11.5, e definendo gli eventi

$$\begin{aligned} B_0 &= \text{non ci sono pezzi difettosi sui 50 esaminati} \\ B_1 &= \text{c'è un pezzo difettoso sui 50 esaminati,} \end{aligned}$$

questa volta avremo  $A = B_0 \cup B_1$ . Siccome inoltre  $B_0 \cap B_1 = \emptyset$ , dall'additività delle probabilità risulta anche che  $q = \mathbf{P}(A) = \mathbf{P}(B_0) + \mathbf{P}(B_1)$ . Questa volta, data l'ampiezza del campione, è consigliabile usare direttamente il calcolo combinatorio per valutare le probabilità richieste: essendo 100 i pezzi disponibili, i modi in cui possiamo estrarre i 50 pezzi (cioè il numero dei casi possibili) sono  $\binom{100}{50}$ . Analogamente i casi favorevoli a  $B_0$  sono i  $\binom{95}{50}$  modi in cui possiamo estrarre 50 pezzi fra i 95 buoni. Conseguentemente

$$\mathbf{P}(B_0) = \frac{\binom{95}{50}}{\binom{100}{50}} = 0,028.$$

Per  $B_1$ , invece, i casi favorevoli sono i  $\binom{95}{49}\binom{5}{1}$  modi in cui possiamo estrarre 49 pezzi fra i 95 buoni e 1 fra i 5 difettosi. Pertanto

$$\mathbf{P}(B_1) = \frac{\binom{95}{49}\binom{5}{1}}{\binom{100}{50}} = 0,153.$$

Ne segue che  $q = 0,181$ , cioè la probabilità di rifiutare la partita  $p = 1 - q = 0,819$  è aumentata rispetto al test precedente. Nota che, se avessimo solo aumentato l'ampiezza della campionatura da 5 a 50 senza spostare la soglia del test da *nessun pezzo campionato è difettoso a solo uno tra i pezzi campionati è difettoso*, avremmo avuto  $A = B_0$  e quindi  $q = \mathbf{P}(B_0) = 0,028$  e  $p = 0,972$ , cioè avremmo aumentato ancora di più la probabilità di rifiutare la partita. In pratica va osservato che un aumento dell'ampiezza del campione (da 5 a 50) rende comunque il test più severo, mentre un aumento del numero di pezzi difettosi tollerati lo rende ovviamente meno severo: il valore della probabilità richiesta in questi esercizi varia in ragione dell'equilibrio fra queste due tendenze in competizione fra loro. Sarà utile osservare infine che il calcolo di  $\mathbf{P}(B_1)$  è un esempio di **distribuzione ipergeometrica**, una distribuzione che incontreremo ancora nel seguito: dati  $N$  oggetti  $r$  dei quali di un tipo (difettosi) ed  $N - r$  di un altro (buoni), se ne estraggono  $n$  a caso e ci si domanda qual è la distribuzione della v.a.  $X$  che rappresenta il numero di pezzi del primo tipo estratti. Siccome l'evento  $X = k$  corrisponde al fatto che abbiamo estratto  $k$  oggetti del primo tipo (fra gli  $r$  disponibili) ed  $n - k$  del secondo (fra gli  $N - r$  disponibili), avremo la DdP

$$\mathbf{P}(X = k) = \frac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, 2, \dots$$

che prende il nome di distribuzione ipergeometrica. ◇

**I.11.7 Esercizio (A.A. Sveshnikov: *Problems in probability theory, mathematical statistics and theory of random functions*; Dover, New York, 1978, p. 18):** Una scatola contiene  $N$  palline:  $n$  bianche,  $m$  nere ed  $l$  rosse. Le palline vengono estratte una alla volta. Determinare la probabilità che una bianca sia estratta prima di una nera nei due casi:

- a) l'estrazione viene effettuata senza rimessa;
- b) l'estrazione viene effettuata con rimessa.

**Soluzione:** Se definiamo gli eventi

- $A$  = una pallina bianca è estratta prima di una nera
- $B_k$  = una pallina bianca è estratta alla  $k$ -ma estrazione
- $R_k$  = una pallina rossa è estratta alla  $k$ -ma estrazione

si ha, nei due casi del problema,

$$A = \begin{cases} B_1 \cup (R_1 B_2) \cup (R_1 R_2 B_3) \cup \dots \cup (R_1 \dots R_l B_{l+1}), & \text{senza rimessa,} \\ B_1 \cup (R_1 B_2) \cup (R_1 R_2 B_3) \cup \dots \cup (R_1 \dots R_k B_{k+1}) \cup \dots, & \text{con rimessa,} \end{cases}$$



in quanto nel caso con rimessa il numero di volte in cui si può estrarre una pallina rossa è illimitato. Nota che in ambedue i casi si tratta di unioni di eventi disgiunti. Nel caso senza rimessa le estrazioni non sono indipendenti dato che il risultato delle estrazioni precedenti influenza le estrazioni successive (vedi la discussione degli esempi I.4.1 ed I.4.7). Pertanto, da I.4.8 e da conteggi elementari potremo ricavare che

$$\begin{aligned} \mathbf{P}(A) &= \mathbf{P}(B_1) + \mathbf{P}(B_2 | R_1)\mathbf{P}(R_1) + \dots \\ &\quad + \mathbf{P}(B_{l+1} | R_1 \dots R_l) \dots \mathbf{P}(R_l | R_1)\mathbf{P}(R_1) \\ &= \frac{n}{N} + \frac{l}{N} \frac{n}{N-1} + \frac{l}{N} \frac{l-1}{N-1} \frac{n}{N-2} + \dots + \frac{l}{N} \dots \frac{1}{N-l+1} \frac{n}{N-l} \\ &= \frac{n}{N} \left[ 1 + \sum_{k=1}^l \frac{\binom{l}{k}}{\binom{N-1}{k}} \right] = \frac{n}{N} \sum_{k=0}^l \frac{\binom{l}{k}}{\binom{N-1}{k}}. \end{aligned}$$

Nel caso con rimessa, invece, gli eventi  $B_k$  ed  $R_k$  sono tutti indipendenti e quindi

$$\begin{aligned} \mathbf{P}(A) &= \mathbf{P}(B_1) + \mathbf{P}(R_1)\mathbf{P}(B_2) + \dots + \mathbf{P}(R_1) \dots \mathbf{P}(R_k)\mathbf{P}(B_{k+1}) + \dots \\ &= \frac{n}{N} + \frac{l}{N} \frac{n}{N} + \frac{l^2}{N^2} \frac{n}{N} + \dots + \frac{l^k}{N^k} \frac{n}{N} + \dots \\ &= \frac{n}{N} \sum_{K=0}^{\infty} \left( \frac{l}{N} \right)^K = \frac{n}{N} \frac{1}{1 - \frac{l}{N}} = \frac{n}{N-l} = \frac{n}{n+m} \end{aligned}$$

è la probabilità richiesta . ◇

**I.11.8 Esercizio (A.A. Sveshnikov:** *Problems in probability theory, mathematical statistics and theory of random functions*; Dover, New York, 1978, p. 23): Data una scatola con  $n$  palline  $m$  ( $\leq n$ ) delle quali *vincono un premio*,  $n$  giocatori estraggono successivamente e senza rimessa una pallina a testa: stabilire se il gioco è equo.

**Soluzione:** Definiti gli eventi

$$A_k = \text{vinco un premio in una delle ultime } n - k \text{ estrazioni,} \quad k = 0, 1, \dots, n$$

dobbiamo stabilire se  $\mathbf{P}(A_k)$  dipende da  $k$  (gioco non equo) o meno (gioco equo). Per farlo definiamo gli eventi

$$B_{kl} = \text{nelle prime } k \text{ estrazioni sono stati vinti } l \text{ premi}$$

in cui  $l = 0, 1, \dots, k$  e comunque  $l \leq m$ . Dalla distribuzione ipergeometrica (Esercizio I.11.6) si ricava allora che

$$\mathbf{P}(B_{kl}) = \begin{cases} \frac{\binom{m}{l} \binom{n-m}{k-l}}{\binom{n}{k}}, & l \leq m; \\ 0, & l > m; \end{cases}$$

e inoltre ovviamente

$$\mathbf{P}(A_k | B_{kl}) = \begin{cases} (m-l)/(n-k), & \text{se } l \leq m; \\ 0, & \text{se } l > m. \end{cases}$$

Pertanto, siccome con  $k$  fissato ed al variare di  $l$  le  $B_{kl}$  realizzano una decomposizione dello spazio di probabilità, dalla formula della probabilità totale (Teorema I.4.5) e posto per brevità  $k \wedge m = \min\{k, m\}$ , si ricava che

$$\mathbf{P}(A_k) = \sum_{l=0}^{k \wedge m} \mathbf{P}(A_k | B_{kl}) \mathbf{P}(B_{kl}) = \sum_{l=0}^{k \wedge m} \frac{m-l}{n-k} \frac{\binom{m}{l} \binom{n-m}{k-l}}{\binom{n}{k}},$$

e siccome si verifica subito che  $(a-b) \binom{a}{b} = a \binom{a-1}{b}$ , si ha

$$\mathbf{P}(A_k) = \frac{m}{n} \sum_{l=0}^{k \wedge m} \frac{\binom{m-1}{l} \binom{n-m}{k-l}}{\binom{n-1}{k}}.$$

Infine, siccome (ad esempio se  $k < m$  per semplificare la notazione)

$$\begin{aligned} \sum_{l=0}^k \binom{m-1}{l} \binom{n-m}{k-l} x^{n-k-1} &= \frac{1}{k!} \sum_{l=0}^k \binom{k}{l} \frac{d^l}{dx^l} (x^{m-1}) \frac{d^{k-l}}{dx^{k-l}} (x^{n-m}) \\ &= \frac{1}{k!} \frac{d^k}{dx^k} (x^{m-1} x^{n-m}) = \frac{1}{k!} \frac{d^k}{dx^k} (x^{n-1}) \\ &= \binom{n-1}{k} x^{n-k-1}, \end{aligned}$$

e quindi ponendo  $x = 1$  si ha

$$\sum_{l=0}^{k \wedge m} \binom{m-1}{l} \binom{n-m}{k-l} = \binom{n-1}{k},$$

in definitiva si ottiene  $\mathbf{P}(A_k) = \frac{m}{n}$ , sicché il gioco risulta equo. ◇

**I.11.9 Esercizio (A.A. Sveshnikov: *Problems in probability theory, mathematical statistics and theory of random functions*; Dover, New York, 1978, p. 24):** Un recipiente contenente una particella è diviso in due scompartimenti (1 e 2), e la probabilità che la particella si trovi nello scompartimento 1 è  $p$ . Uno strumento di misura ha la probabilità  $q$  di rivelare la particella se usato nello scompartimento in cui essa si trova ( $q$  rappresenta l'efficienza dello strumento). Su  $n$  tentativi di rivelazione,  $m$  vengono eseguiti in 1 ed  $n - m$  in 2: determinare  $m$  in modo che la probabilità di rivelare la particella almeno una volta sia la più grande possibile.

**Soluzione:** Definiti gli eventi

$B_i$  = la particella si trova nello scompartimento  $i$

$A_m$  = la particella viene rivelata almeno una volta, misurando  $m$  volte in 1

con  $\mathbf{P}(B_1) = p$  e  $\mathbf{P}(B_2) = 1 - p$ , si ha

$$\mathbf{P}(A_m | B_1) = 1 - (1 - q)^m, \quad \mathbf{P}(A_m | B_2) = 1 - (1 - q)^{n-m},$$

dato che  $(1 - q)^k$  è la probabilità di non rivelare mai la particella eseguendo  $k$  misure nello scompartimento in cui essa si trova. Pertanto

$$\begin{aligned} \mathbf{P}(A_m) &= \mathbf{P}(A_m | B_1)\mathbf{P}(B_1) + \mathbf{P}(A_m | B_2)\mathbf{P}(B_2) \\ &= p[1 - (1 - q)^m] + (1 - p)[1 - (1 - q)^{n-m}]. \end{aligned}$$

Per rendere massima  $\mathbf{P}(A_m)$  consideriamo  $m$  come una variabile continua ed imponiamo che

$$\frac{d\mathbf{P}(A_m)}{dm} = -p(1 - q)^m \ln(1 - q) + (1 - p)(1 - q)^{n-m} \ln(1 - q) = 0.$$

Ne segue che il massimo è raggiunto in

$$m = \frac{n}{2} + \frac{\ln \frac{1-p}{p}}{2 \ln(1 - q)},$$

e quindi il numero di tentativi che rende massima la probabilità di rivelazione della particella è il numero intero più prossimo a questo valore. Nota che se  $p = 1/2$  risulta  $m = n/2$ , mentre se  $p < 1/2$  si ha  $m < n/2$  dato che  $\ln(1 - p)/p > 0$  e  $\ln(1 - q) < 0$ : ciò corrisponde al fatto intuitivo che conviene far meno misure nello scompartimento dove la particella soggiorna meno.  $\diamond$

**I.11.10 Esercizio (A.A. Sveshnikov: *Problems in probability theory, mathematical statistics and theory of random functions*; Dover, New York, 1978, p. 27):** Un telegrafo trasmette *punti* e *linee* ed è noto che la frequenza dei *punti* è  $\frac{5}{8}$  mentre quella delle *linee* è  $\frac{3}{8}$ . Disturbi di trasmissione modificano  $\frac{2}{5}$  dei *punti* in *linee* ed  $\frac{1}{3}$  delle *linee* in *punti*. Supponendo di ricevere un messaggio, calcolare la probabilità che il segno ricevuto sia proprio quello trasmesso nel caso in cui si riceve *punto* ed in quello in cui si riceve *linea*.

**Soluzione:** Per risolvere questo problema faremo uso del Teorema di Bayes (vedi I.4.9) definendo gli eventi

$$\begin{aligned} A &= \text{ricevo } \textit{punto}, & B &= \text{ricevo } \textit{linea}, \\ H_1 &= \text{è stato trasmesso } \textit{punto}, & H_2 &= \text{è stata trasmessa } \textit{linea}, \end{aligned}$$

con le rispettive probabilità

$$\begin{aligned} \mathbf{P}(H_1) &= \frac{5}{8}, & \mathbf{P}(H_2) &= \frac{3}{8}, \\ \mathbf{P}(A | H_1) &= \frac{3}{5}, & \mathbf{P}(A | H_2) &= \frac{1}{3}, \\ \mathbf{P}(B | H_1) &= \frac{2}{5}, & \mathbf{P}(B | H_2) &= \frac{2}{3}. \end{aligned}$$

Siccome per la formula della probabilità totale (vedi I.4.5) si ha

$$\begin{aligned}\mathbf{P}(A) &= \mathbf{P}(A | H_1)\mathbf{P}(H_1) + \mathbf{P}(A | H_2)\mathbf{P}(H_2) = \frac{1}{2} \\ \mathbf{P}(B) &= \mathbf{P}(B | H_1)\mathbf{P}(H_1) + \mathbf{P}(B | H_2)\mathbf{P}(H_2) = \frac{1}{2},\end{aligned}$$

la formula di Bayes ci consente di affermare che

$$\mathbf{P}(H_1 | A) = \frac{\mathbf{P}(A | H_1)\mathbf{P}(H_1)}{\mathbf{P}(A)} = \frac{3}{4}, \quad \mathbf{P}(H_2 | B) = \frac{\mathbf{P}(B | H_2)\mathbf{P}(H_2)}{\mathbf{P}(B)} = \frac{1}{2},$$

e pertanto la ricezione del segno *punto* è più affidabile della ricezione del segno *linea*.  $\diamond$

**I.11.11 Esercizio (F. Mosteller: *Fifty challenging problems in probability*; Dover, New York, 1987, p. 1):** Un cassetto contiene calzini rossi e bianchi ed è noto che, se si estraggono a caso due calzini, la probabilità che siano ambedue rossi è  $\frac{1}{2}$ . Qual è il più piccolo numero di calzini che rende possibile questa situazione? Qual è la risposta alla domanda precedente se è noto che il numero dei calzini bianchi è pari?

**Soluzione:** Se  $r$  è il numero di calzini rossi e  $b$  il numero di quelli bianchi, le condizioni del problema impongono che

$$\frac{r(r-1)}{(r+b)(r+b-1)} = \frac{1}{2},$$

cioè che

$$r^2 - (2b+1)r - b(b-1) = 0.$$

Ne segue che tra il numero di calzini rossi ed il numero di calzini bianchi intercorre la relazione

$$r = \frac{(2b+1) \pm \sqrt{8b^2+1}}{2}.$$

Possiamo ora affrontare il problema per tentativi ricordando che, naturalmente,  $r$  e  $b$  devono essere numeri interi e che il cassetto deve contenere almeno due calzini:

$b$	$r$	
0	1	i calzini sono meno di 2;
1	0 e 3	risposta alla prima domanda!
2	$\frac{5 \pm \sqrt{33}}{2}$	numeri non interi;
3	$\frac{7 \pm \sqrt{73}}{2}$	numeri non interi;
4	$\frac{9 \pm \sqrt{129}}{2}$	numeri non interi;
5	$\frac{11 \pm \sqrt{201}}{2}$	numeri non interi;
6	-2 e 15	risposta alla seconda domanda!

Pertanto la risposta alla prima domanda è  $b = 1$  ed  $r = 3$ , mentre la risposta alla seconda domanda è  $b = 6$  ed  $r = 15$ . Val la pena di osservare che la ricerca dei valori di  $r$  e  $b$  secondo lo schema così riportato costituiscono un tipico problema di teoria dei numeri.  $\diamond$

**I.11.12 Esercizio (F. Mosteller: *Fifty challenging problems in probability*; Dover, New York, 1987, p. 1):** Per incoraggiare la carriera tennistica di Mario suo padre gli promette un premio se egli riesce a vincere almeno due partite di seguito in una serie di tre partite giocate alternativamente con suo padre (P) e con il campione della loro associazione (C). Mario può scegliere di iniziare la serie con suo padre (PCP) o con il campione (CPC). Sapendo che il campione gioca meglio del padre, quale successione di partite gli converrà scegliere?

**Soluzione:** A prima vista la risposta sembra essere PCP perché in questo modo Mario giocherebbe due volte con l'avversario meno abile. Il fatto, però, che in questo modo l'avversario più abile venga a trovarsi in posizione centrale (separando così due eventuali vittorie contro il padre) rende il problema meno banale. Supponendo di indicare con  $p$  la probabilità che Mario batta suo padre e con  $c$  quella che egli batta il campione (quindi:  $p > c$ ) ed osservando che le successioni di vittorie (V) e sconfitte (S) favorevoli a Mario sono le seguenti

<i>avversari</i>	VVV	VVS	SVV
PCP	$pcp$	$pc(1-p)$	$(1-p)cp$
CPC	$cpc$	$cp(1-c)$	$(1-c)pc$

si ricava che le probabilità di vincere il premio nei due casi sono

$$\begin{array}{cc}
 \text{PCP} & \text{CPC} \\
 pc(2-p) & pc(2-p) + 2pc(p-c)
 \end{array}$$

e pertanto, siccome  $p > c$ , la successione CPC è più favorevole di PCP, contrariamente a quanto si poteva supporre da un esame superficiale.  $\diamond$

**I.11.13 Esercizio (F. Mosteller: *Fifty challenging problems in probability*; Dover, New York, 1987, p. 1):** Due dei membri di una giuria di tre persone emettono un verdetto giusto con probabilità  $p$ , mentre il terzo è indeciso ed emette i suoi verdetti lanciando una moneta (equa). La giuria prende le sue decisioni a maggioranza. Una seconda giuria è invece composta di una sola persona che emette un verdetto corretto con probabilità  $p$ . Confrontare le probabilità di ottenere una sentenza giusta dalle due giurie descritte.

**Soluzione:** Le due giurie hanno la medesima probabilità  $p$  di emettere un verdetto giusto. Infatti, nel caso della prima giuria, le combinazioni di verdetti individuali che danno (a maggioranza) una sentenza giusta e le loro probabilità sono (suppo-

nendo che i singoli verdetti siano indipendenti):

1° giurato	2° giurato	giurato indeciso	probabilità
giusto	giusto	giusto	$p^2 \cdot \frac{1}{2}$
giusto	giusto	sbagliato	$p^2 \cdot \frac{1}{2}$
giusto	sbagliato	giusto	$p(1-p) \cdot \frac{1}{2}$
sbagliato	giusto	giusto	$(1-p)p \cdot \frac{1}{2}$

e siccome la probabilità di ottenere una sentenza giusta è la somma di questa probabilità, il risultato è  $p$  come per la giuria composta di un solo giurato.  $\diamond$

**I.11.14 Esercizio** (A.A. Sveshnikov: *Problems in probability theory, mathematical statistics and theory of random functions*; Dover, New York, 1978, p. 44): Da una partita di 100 oggetti, 10 dei quali sono difettosi, viene estratto per un controllo un campione di 5 oggetti. Detta  $\xi$  la v.a.

$\xi$  = numero di oggetti difettosi presenti nel campione

determinarne la DdP.

**Soluzione:** La nostra v.a. assume solo i valori  $k = 0, 1, 2, 3, 4, 5$  con probabilità

$$p_k = \mathbf{P}\{\xi = k\} = \frac{\binom{10}{k} \binom{90}{5-k}}{\binom{100}{5}}$$

e quindi

$$\begin{array}{cccccc} k = & 0 & 1 & 2 & 3 & 4 & 5 \\ p_k = & 0,583 & 0,340 & 0,070 & 0,007 & 0,000 & 0,000 \end{array}$$

è la sua DdP.  $\diamond$

**I.11.15 Esercizio** (A.A. Sveshnikov: *Problems in probability theory, mathematical statistics and theory of random functions*; Dover, New York, 1978, p. 44): Vengono effettuati dei tests su un prodotto e si sa che la probabilità che il campione scelto superi il test è  $p = \frac{4}{5}$ . La successione dei tests si arresta quando viene individuato il primo campione che non supera il test. Definita la v.a.

$\nu$  = numero di tests effettuati prima dell'arresto

determinarne la DdP ed il VdA.

**Soluzione:** La v.a.  $\nu$  assume valori interi ma non limitati. Siccome l'evento  $\{\nu = k\}$  rappresenta l'affermazione *nei primi  $k-1$  tentativi il campione supera il test, ma non lo supera al  $k$ -mo*, avremo che

$$\mathbf{P}\{\nu = k\} = p^{k-1}(1-p) = \frac{1}{5} \left(\frac{4}{5}\right)^{k-1}.$$

Notiamo anche che  $\nu$  può assumere valori arbitrariamente grandi con probabilità non nulla, ma che

$$\lim_k \mathbf{P}\{\nu = k\} = 0.$$

Infine, se estendiamo la definizione data di VdA anche al caso in cui le v.a. assumono un numero infinito di valori, otteniamo

$$\mathbf{E}\nu = \sum_{k=1}^{\infty} k(1-p)p^{k-1} = (1-p) \frac{d}{dp} \sum_{k=0}^{\infty} p^k = (1-p) \frac{d}{dp} \left( \frac{1}{1-p} \right) = \frac{1}{1-p},$$

e pertanto nel nostro caso  $\mathbf{E}\nu = 5$ . ◇

**I.11.16 Esercizio** (A.A. Sveshnikov: *Problems in probability theory, mathematical statistics and theory of random functions*; Dover, New York, 1978, p. 44): Un automobilista percorre una strada con 4 semafori. Ogni semaforo lo arresta con una probabilità  $p = 0,5$ . Definita la v.a.

$\nu =$  numero dei semafori verdi prima del primo arresto

determinarne la DdP.

**Soluzione:** La v.a.  $\nu$  assume i valori  $k = 0, 1, 2, 3, 4$  con

$$p_k = \mathbf{P}\{\nu = k\} = \begin{cases} p(1-p)^k & \text{se } k = 0, 1, 2, 3; \\ (1-p)^4 & \text{se } k = 4. \end{cases}$$

Pertanto nel nostro caso

$$\begin{array}{cccccc} k & = & 0 & 1 & 2 & 3 & 4 \\ p_k & = & 0,5 & 0,25 & 0,125 & 0,062 & 0,062 \end{array}$$

è la DdP cercata. ◇





II

BASI MATEMATICHE  
del CALCOLO delle PROBABILITÀ



## II.1 Modelli di probabilità. Assiomi

Abbiamo visto ripetutamente nella parte I che i modelli *finiti* di probabilità si rivelano ben presto insufficienti (vedi Osservazione I.7.6 e la discussione del Capitolo I.10). Gli aspetti principali di questo problema possono essere messi in luce ancora una volta esaminando il seguente esempio.

**II.1.1 Esempio:** Prendiamo in considerazione un Modello di Bernoulli (vedi Osservazione I.4.20 e Capitolo I.7) per  $n$  tentativi successivi ed indipendenti di verifica di un dato evento in cui la probabilità di successo in ciascun tentativo sia  $p$ . È noto che il numero di elementi dello spazio  $\Omega$  (cioè il numero dei possibili risultati) è  $2^n$ . Vogliamo ora esaminare la possibilità di estendere il modello al caso di un numero infinito di tentativi, conservando sempre lo stesso valore di  $p$ . È ovvio che ora i possibili risultati del nostro esperimento (cioè gli elementi di  $\Omega$ ) saranno le successioni di simboli 0 ed 1 del tipo  $\omega = \{a_n\}_{n \in \mathbf{N}}$  dove le  $a_n$  assumono solo i valori 0 ed 1. Si può ora mostrare immediatamente che  $\Omega$  è non numerabile. Infatti gli elementi di  $\Omega$  risultano essere in corrispondenza biunivoca con gli elementi (non numerabili) dell'intervallo  $[0, 1)$  in quanto ogni numero  $0 \leq a < 1$  è suscettibile di un'unica *rappresentazione binaria* (contenente un numero infinito di termini nulli<sup>1</sup>) del tipo

$$a = \frac{a_1}{2} + \frac{a_2}{2^2} + \dots + \frac{a_n}{2^n} + \dots \quad (a_n = 0, 1),$$

e quindi è in corrispondenza biunivoca con le successioni infinite di 0 ed 1. Ciò significa che gli spazi  $\Omega$  necessari per descrivere un numero infinito di tentativi possono essere di natura piuttosto complicata e che la definizione elementare di un modello di probabilità su tali  $\Omega$  incontrerebbe subito delle gravi difficoltà. Se, riproducendo la costruzione dello spazio di probabilità descritta in I.3, volessimo attribuire delle probabilità  $p(\omega)$  ai risultati  $\omega \in \Omega$  del nostro esperimento composto di infiniti tentativi ci troveremmo subito in un grave imbarazzo: se ad esempio, per semplicità, avessimo  $p = \frac{1}{2}$ , tutti i risultati  $\omega$  sarebbero equiprobabili; siccome però deve risultare anche che  $\sum_{\omega \in \Omega} p(\omega)$  deve essere un numero finito (più precisamente deve essere eguale ad 1), e siccome  $\Omega$  non è numerabile, i valori di  $p(\omega)$  non potranno che essere tutti nulli (e la somma non potrà che essere zero), sicché

---

<sup>1</sup> Circa l'unicità della rappresentazione binaria dei numeri reali  $0 \leq a < 1$  sarà bene osservare che essa è garantita solo se si usa la convenzione in base alla quale un numero infinito di coefficienti  $a_n$  è nullo. Per convincersene basterà osservare che se  $a_n = 1, \forall n \in \mathbf{N}$  si ottiene  $a = 1$  e che se per un dato numero  $a$  solo un numero finito di  $a_n$  fosse nullo,  $a$  sarebbe eguale ad 1 meno un numero razionale e quindi sarebbe esso stesso un numero razionale; ma un numero razionale può sempre essere scritto come somma di un numero finito di potenze di  $\frac{1}{2}$ , cioè con un numero finito di  $a_n$  eguali ad 1 (e un numero infinito di  $a_n$  eguali a 0). Pertanto vi è una ambiguità nella rappresentazione binaria di tali numeri per rimuovere la quale si può usare la convenzione suddetta, o la convenzione opposta. Naturalmente con tale convenzione resta escluso il caso  $a = 1$  dato che la sua rappresentazione richiederebbe  $a_n = 1$  per ogni  $n$ .

nessuna definizione coerente di probabilità potrebbe essere costruita su questa base. Da questa discussione si ricava immediatamente che, nel caso generale, la definizione di  $\mathbf{P}(A)$  non può più essere costruita a partire da un'assegnazione di  $p(\omega)$  come nel caso elementare. Per questo motivo, in sostituzione del metodo elementare, svilupperemo nel seguito l'idea intuitiva secondo la quale, sempre restando nel caso equo  $p = \frac{1}{2}$ , la probabilità di ottenere risultati corrispondenti, ad esempio, all'insieme di punti  $[0, \frac{1}{2})$  è  $\frac{1}{2}$ : seguiremo cioè la via dell'associazione del concetto di probabilità con il concetto di *misura di insiemi*.  $\diamond$

**II.1.2 Definizione:** Diremo che una famiglia non vuota  $\mathcal{A} \subseteq \wp(\Omega)$  di parti di un insieme  $\Omega$  costituisce un **anello** quando

- a.  $A \cup B \in \mathcal{A}, \quad \forall A, B \in \mathcal{A};$
- b.  $A \setminus B \in \mathcal{A}, \quad \forall A, B \in \mathcal{A};$

inoltre  $\mathcal{A}$  si dice **algebra** se è un anello e se  $\Omega \in \mathcal{A}$ .  $\triangle$

**II.1.3 Proposizione:** Se  $\mathcal{A}$  è un anello valgono le seguenti affermazioni:

- I)  $\emptyset \in \mathcal{A};$
  - II)  $A \cap B \in \mathcal{A}, \quad \forall A, B \in \mathcal{A};$
- inoltre, se  $\mathcal{A}$  è anche un'algebra, risulta
- III)  $\overline{A} \in \mathcal{A}, \quad \forall A \in \mathcal{A}.$

**Dimostrazione:** Siccome  $\mathcal{A}$  contiene almeno un elemento, considerato  $A \in \mathcal{A}$ , dalla proprietà b. degli anelli risulta che  $A \setminus A = \emptyset \in \mathcal{A}$ . Inoltre dalle proprietà a. e b. deriva anche che

$$A \cap B = (A \cup B) \setminus [(A \setminus B) \cup (B \setminus A)] = (A \cup B) \setminus (A \Delta B)$$

è sicuramente un elemento di  $\mathcal{A}$ . Infine, se  $\mathcal{A}$  è anche un'algebra, da b. si ricava che  $\overline{A} = \Omega \setminus A$  è ancora un elemento di  $\mathcal{A}$ .  $\square$

**II.1.4 Esempio:** La famiglia  $\wp(\Omega)$  di tutte le parti di un generico insieme  $\Omega$  e tutti gli altri esempi presentati in I.2.9 sono ovviamente algebre di parti di  $\Omega$ . Viceversa la famiglia  $\mathcal{I}$  degli intervalli di  $\mathbf{R}$  non è un anello (e quindi neanche un'algebra) dato che non verifica a. e b. Per questo motivo sarà utile nel seguito estendere le nostre definizioni introducendo dei nuovi concetti.  $\diamond$

**II.1.5 Definizione:** Diremo che una famiglia  $\mathcal{A} \subseteq \wp(\Omega)$  di parti di un insieme  $\Omega$  costituisce un **semianello** quando contiene l'insieme  $\emptyset$  e

- a.  $A \cap B \in \mathcal{A}, \quad \forall A, B \in \mathcal{A};$
- b.  $A \setminus B = \bigcup_{k=1}^n A_k, \text{ con } A_k \in \mathcal{A} \text{ disgiunti,} \quad \forall A, B \in \mathcal{A};$

inoltre  $\mathcal{A}$  si dice **semialgebra** se è un semianello e se  $\Omega \in \mathcal{A}$ .  $\triangle$

**II.1.6 Proposizione:** Ogni anello è un semianello; viceversa la famiglia di tutte le unioni finite di elementi di un semianello  $\mathcal{A}$  è un anello.

**Dimostrazione:** Che ogni anello sia un semianello si verifica facilmente dal momento che, come si vede da II.1.3, esso contiene  $\emptyset$  e soddisfa a. di II.1.5; quanto a b. di II.1.5 essa è banalmente soddisfatta nel senso che  $A \setminus B \in \mathcal{A}$ . Per il viceversa dovremo provare che la seguente famiglia di parti di  $\Omega$

$$\mathcal{B} = \left\{ B \in \wp(\Omega) : B = \bigcup_{k=1}^n A_k, A_k \in \mathcal{A} \right\}$$

è un anello. Infatti sicuramente  $\mathcal{B}$  è non vuoto dato che  $\mathcal{A}$  contiene almeno  $\emptyset$ ; inoltre:

1. se  $B_1$  e  $B_2$  sono elementi di  $\mathcal{B}$  risulterà

$$B_1 = \bigcup_{k=1}^{n_1} A_k^1, B_2 = \bigcup_{l=1}^{n_2} A_l^2, \quad \text{con } A_k^1, A_l^2 \in \mathcal{A},$$

per cui  $B_1 \cup B_2$  sarà ancora unione finita di elementi di  $\mathcal{A}$  e quindi sarà un elemento di  $\mathcal{B}$ , cioè la proprietà a. degli anelli è soddisfatta;

2. con le stesse notazioni del punto 1. si verifica subito che

$$B_1 \setminus B_2 = \left( \bigcup_{k=1}^{n_1} A_k^1 \right) \setminus \left( \bigcup_{l=1}^{n_2} A_l^2 \right) = \bigcup_{k=1}^{n_1} \left[ \bigcap_{l=1}^{n_2} (A_k^1 \setminus A_l^2) \right];$$

siccome però  $A_k^1, A_l^2 \in \mathcal{A}$ , dalla proprietà b. dei semianelli risulterà sempre  $A_k^1 \setminus A_l^2 = \bigcup_{j=1}^m A_j^{kl}$  con  $A_j^{kl} \in \mathcal{A}$  e quindi

$$B_1 \setminus B_2 = \bigcup_{k=1}^{n_1} \bigcap_{l=1}^{n_2} \bigcup_{j=1}^m A_j^{kl};$$

basterà allora ricordare che<sup>2</sup> intersezioni finite di unioni finite sono unioni finite di intersezioni finite, per cui  $B_1 \setminus B_2$  risulta essere unione finita di intersezioni degli  $A_j^{kl}$  e quindi in definitiva (essendo  $\mathcal{A}$  un semianello) unione finita di elementi di  $\mathcal{A}$ , sicché  $B_1 \setminus B_2 \in \mathcal{B}$  e anche la proprietà b. degli anelli è soddisfatta.  $\square$

**II.1.7 Esempio:** È facile verificare che l'insieme  $\mathcal{I}$  degli intervalli di  $\mathbf{R}$  e l'insieme  $\mathcal{R}^n$  dei rettangoli di  $\mathbf{R}^n$  sono semianelli e quindi che, in base alla proposizione precedente, le famiglie delle unioni finite rispettivamente di intervalli o rettangoli sono degli anelli.  $\diamond$

---

<sup>2</sup> Per tutte le relazioni fra insiemi ricordate qui e nel seguito basterà consultare, ad esempio, i primi capitoli di **J. Dieudonné:** *Foundations of modern analysis*; Academic Press, New York, 1960, o di **A.N. Kolmogorov e S.V. Fomin:** *Elementi di Teoria delle Funzioni e di Analisi Funzionale*; Mir, Mosca, 1980.

**II.1.8 Definizione:** Data un'algebra  $\mathcal{A}$  di parti di un insieme  $\Omega$ , chiameremo **misura additiva** su  $\mathcal{A}$  ogni applicazione  $\mu : \mathcal{A} \rightarrow [0, +\infty]$  tale che, dati  $A, B \in \mathcal{A}$ ,

$$\mu(A \cup B) = \mu(A) + \mu(B), \quad \text{se } A \cap B = \emptyset;$$

se inoltre  $\mu(\Omega) < +\infty$  diremo che  $\mu$  è una **misura additiva finita** e se  $\mu(\Omega) = 1$  diremo che  $\mu$  è una **misura additiva di probabilità**.  $\triangle$

**II.1.9 Osservazione:** Nota che, comunque scelto  $A \in \mathcal{A}$ , risulterà sempre  $\mu(A) \leq \mu(\Omega)$  in quanto  $\mu$  è additiva e positiva, e  $\Omega = A \cup \bar{A}$ ; ne segue che se  $\mu$  è finita avremo anche  $\mu(A) < +\infty$  comunque scelto  $A \in \mathcal{A}$ . A questo punto potremmo pensare di definire un modello di probabilità come una terna  $(\Omega, \mathcal{A}, \mathbf{P})$  dove  $\Omega$  è un generico insieme,  $\mathcal{A}$  un'algebra di sue parti e  $\mathbf{P}$  una misura additiva di probabilità su  $\mathcal{A}$ . Ci accorgeremo subito, però, che così facendo otterremmo dei modelli di probabilità troppo generali per essere praticamente utili in quanto non potremmo affermare nulla di preciso circa le successioni *infinite* di sottinsiemi  $(A_n)_{n \in \mathbf{N}}$ . Infatti è facilmente verificabile che la proprietà a. delle algebre e l'additività delle misure possono essere immediatamente estese al caso di un numero finito (anche maggiore di 2) di sottinsiemi, ma che nulla si può dire a priori circa il caso di un numero infinito di sottinsiemi.  $\circ$

**II.1.10 Definizione:** Diremo che  $\mu$  è una **misura  $\sigma$ -additiva** (o anche semplicemente una **misura**) se essa è una misura additiva definita su un'algebra  $\mathcal{A}$  di parti di  $\Omega$  e se, comunque scelta una successione  $(A_n)_{n \in \mathbf{N}}$  di elementi disgiunti di  $\mathcal{A}$  tali che  $\bigcup_n A_n \in \mathcal{A}$ , risulta

$$\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n);$$

Diremo poi che  $\mu$  è una **misura finita** se  $\mu(\Omega) < +\infty$  e che è una **misura  $\sigma$ -finita** se  $\Omega$  può essere decomposto nell'unione  $\Omega = \bigcup_n A_n$ ,  $A_n \in \mathcal{A}$  di insiemi disgiunti tali che  $\mu(A_n) < +\infty$ ,  $\forall n \in \mathbf{N}$ . Una misura finita  $\mathbf{P}$  tale che  $\mathbf{P}(\Omega) = 1$  dicesi infine **misura di probabilità**.  $\triangle$

**II.1.11 Osservazione:** Notiamo che nella definizione di misura  $\sigma$ -additiva va esplicitamente richiesto che la successione  $(A_n)_{n \in \mathbf{N}}$  di elementi di  $\mathcal{A}$  sia tale che  $\bigcup_n A_n \in \mathcal{A}$ , in quanto ciò non è affatto garantito dalla proprietà a. delle algebre che riguarda soltanto unioni *finite*. Questa osservazione sarà ripresa più innanzi (vedi Osservazione II.1.14) per motivare l'introduzione del concetto di  $\sigma$ -algebra. Si ponga attenzione infine al fatto che, mentre una misura finita risulta sempre anche  $\sigma$ -finita, il viceversa non è vero: ad esempio l'usuale misura di Lebesgue sulla retta reale (quella, per intenderci, che ad ogni intervallo  $[a, b]$  della retta reale associa la misura  $|b - a|$ ) è ovviamente  $\sigma$ -finita, ma non finita come si mostra facilmente con semplici esempi.  $\circ$

**II.1.12 Proposizione:** Data una misura di probabilità  $\mathbf{P} : \mathcal{A} \rightarrow [0, 1]$  le seguenti proprietà sono sempre verificate:

- I)  $\mathbf{P}(\emptyset) = 0$ ;  
 II)  $\mathbf{P}(A \setminus B) = \mathbf{P}(A) - \mathbf{P}(AB)$ ,  $\forall A, B \in \mathcal{A}$ ;  
 III)  $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(AB)$ ,  $\forall A, B \in \mathcal{A}$ ;  
 IV)  $\mathbf{P}(A \Delta B) = \mathbf{P}(A) + \mathbf{P}(B) - 2\mathbf{P}(AB)$ ,  $\forall A, B \in \mathcal{A}$ ;  
 V)  $\mathbf{P}(B) \leq \mathbf{P}(A)$  se  $B \subseteq A$ , con  $A, B \in \mathcal{A}$ ;  
 VI)  $\mathbf{P}(\bigcup_n A_n) \leq \sum_n \mathbf{P}(A_n)$ ,  $\forall (A_n)_{n \in \mathbf{N}} \in \mathcal{A}^{\mathbf{N}}$  tale che  $\bigcup_n A_n \in \mathcal{A}$  (detta anche *subadditività*).

**Dimostrazione:**

I) Dall'additività di  $\mathbf{P}$  si ha, comunque scelto  $A \in \mathcal{A}$ , che

$$\mathbf{P}(A) = \mathbf{P}(A \cup \emptyset) = \mathbf{P}(A) + \mathbf{P}(\emptyset),$$

da cui discende immediatamente che  $\mathbf{P}(\emptyset) = 0$ .

II) Sempre per l'additività di  $\mathbf{P}$  si ha che

$$\mathbf{P}(A) = \mathbf{P}[(A \setminus B) \cup (AB)] = \mathbf{P}(A \setminus B) + \mathbf{P}(AB)$$

da cui si ricava subito che  $\mathbf{P}(A \setminus B) = \mathbf{P}(A) - \mathbf{P}(AB)$ .

III) La relazione

$$\begin{aligned} \mathbf{P}(A \cup B) &= \mathbf{P}[(A \setminus B) \cup (AB) \cup (B \setminus A)] = \mathbf{P}(A \setminus B) + \mathbf{P}(AB) + \mathbf{P}(B \setminus A) \\ &= \mathbf{P}(A) - \mathbf{P}(AB) + \mathbf{P}(AB) + \mathbf{P}(B) - \mathbf{P}(AB) \\ &= \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(AB) \end{aligned}$$

deriva dall'additività di  $\mathbf{P}$  e dalla proprietà II.

IV) La relazione

$$\begin{aligned} \mathbf{P}(A \Delta B) &= \mathbf{P}[(A \setminus B) \cup (B \setminus A)] = \mathbf{P}(A \setminus B) + \mathbf{P}(B \setminus A) \\ &= \mathbf{P}(A) - \mathbf{P}(AB) + \mathbf{P}(B) - \mathbf{P}(AB) = \mathbf{P}(A) + \mathbf{P}(B) - 2\mathbf{P}(AB) \end{aligned}$$

discende ancora dall'additività di  $\mathbf{P}$  e dalla proprietà II.

V) La relazione

$$\mathbf{P}(A) = \mathbf{P}[B \cup (A \setminus B)] = \mathbf{P}(B) + \mathbf{P}(A \setminus B) \geq \mathbf{P}(B)$$

si ricava dall'additività e dalla positività di valori di  $\mathbf{P}$ .

VI) In generale gli elementi di una successione  $(A_n)_{n \in \mathbf{N}} \in \mathcal{A}^{\mathbf{N}}$  (tale che  $\bigcup_n A_n \in \mathcal{A}$ ) non sono tra loro disgiunti. Per usare la proprietà di  $\sigma$ -additività di  $\mathbf{P}$  ci converrà allora definire un'altra successione di elementi di  $\mathcal{A}$  che risultano evidentemente due a due disgiunti:

$$\begin{aligned} B_1 &= A_1, \\ B_2 &= \overline{A_1} A_2 = A_2 \setminus A_1, \\ B_3 &= \overline{A_1} \overline{A_2} A_3 = \overline{(A_1 \cup A_2)} A_3 = A_3 \setminus (A_2 \cup A_1), \dots \end{aligned}$$

cioè in generale

$$B_n = \bar{A}_1 \dots \bar{A}_{n-1} A_n = \overline{\left( \bigcup_{k=1}^{n-1} A_k \right)} A_n = A_n \setminus \left( \bigcup_{k=1}^{n-1} A_k \right);$$

per tale successione si ha anche che

$$\begin{aligned} B_1 &= A_1, \\ B_1 \cup B_2 &= A_1 \cup (\bar{A}_1 A_2) = A_1 \cup A_2, \\ B_1 \cup B_2 \cup B_3 &= (A_1 \cup A_2) \cup [\overline{(A_1 \cup A_2)} A_3] = A_1 \cup A_2 \cup A_3, \dots \end{aligned}$$

cioè in generale

$$\bigcup_{k=1}^n B_k = \bigcup_{k=1}^n A_k,$$

e quindi in definitiva  $\bigcup_n B_n = \bigcup_n A_n$ . Dalla  $\sigma$ -additività di  $\mathbf{P}$  si ha allora che

$$\mathbf{P}\left(\bigcup_n A_n\right) = \mathbf{P}\left(\bigcup_n B_n\right) = \sum_n \mathbf{P}(B_n).$$

Siccome inoltre dalla definizione delle  $B_n$  si ha che  $B_n \subseteq A_n, \forall n \in \mathbf{N}$ , dalla proprietà V si ha anche che  $\mathbf{P}(B_n) \leq \mathbf{P}(A_n), \forall n \in \mathbf{N}$  e quindi

$$\mathbf{P}\left(\bigcup_n A_n\right) = \sum_n \mathbf{P}(B_n) \leq \sum_n \mathbf{P}(A_n),$$

che prova la subadditività di  $\mathbf{P}$ . □

**II.1.13 Teorema (criteri di  $\sigma$ -additività):** Se  $\mathbf{P} : \mathcal{A} \rightarrow [0, 1]$  è una misura additiva di probabilità, le seguenti quattro affermazioni sono equivalenti:

- 1)  $\mathbf{P}$  è una misura di probabilità (cioè è  $\sigma$ -additiva);
- 2) data una successione  $(A_n)_{n \in \mathbf{N}}, A_n \in \mathcal{A}, \forall n \in \mathbf{N}$  crescente ( $A_n \subseteq A_{n+1}$ ) e tale che  $\bigcup_n A_n \in \mathcal{A}$ , risulta  $\mathbf{P}(A_n) \xrightarrow{n} \mathbf{P}\left(\bigcup_n A_n\right)$ , e si dice che  $\mathbf{P}$  è *continua da sotto*;
- 3) data una successione  $(B_n)_{n \in \mathbf{N}}, B_n \in \mathcal{A}, \forall n \in \mathbf{N}$  decrescente ( $B_n \supseteq B_{n+1}$ ) e tale che  $\bigcap_n B_n \in \mathcal{A}$ , risulta  $\mathbf{P}(B_n) \xrightarrow{n} \mathbf{P}\left(\bigcap_n B_n\right)$ , e si dice che  $\mathbf{P}$  è *continua da sopra*;
- 4) data una successione  $(C_n)_{n \in \mathbf{N}}, C_n \in \mathcal{A}, \forall n \in \mathbf{N}$  decrescente ( $C_n \supseteq C_{n+1}$ ) e tale che  $\bigcap_n C_n = \emptyset$ , risulta  $\mathbf{P}(C_n) \xrightarrow{n} 0$ , e si dice che  $\mathbf{P}$  è *continua nell'origine*.

**Dimostrazione:** Per dimostrare il teorema proveremo la seguente catena di implicazioni:  $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 4 \Rightarrow 1$ , che evidentemente è equivalente alla tesi enunciata.



1  $\Rightarrow$  2: Dalla successione crescente  $(A_n)_{n \in \mathbf{N}}$  è possibile costruire la seguente successione di sottinsiemi disgiunti:

$$\begin{aligned} D_1 &= A_1, \\ D_2 &= A_2 \setminus A_1 = \overline{A_1} A_2, \\ D_3 &= A_3 \setminus A_2 = \overline{A_2} A_3, \dots \end{aligned}$$

per la quale si verifica facilmente che  $\bigcup_n A_n = \bigcup_n D_n$ . La  $\sigma$ -additività di  $\mathbf{P}$  implica il risultato richiesto:

$$\begin{aligned} \mathbf{P}\left(\bigcup_n A_n\right) &= \mathbf{P}\left(\bigcup_n D_n\right) = \sum_{k=1}^{\infty} \mathbf{P}(D_k) \\ &= \mathbf{P}(A_1) + \mathbf{P}(A_2) - \mathbf{P}(A_1) + \mathbf{P}(A_3) - \mathbf{P}(A_2) + \dots \\ &= \lim_n \mathbf{P}(A_n). \end{aligned}$$

2  $\Rightarrow$  3: Dato che  $(B_n)_{n \in \mathbf{N}}$  è decrescente si verifica facilmente che comunque scelto  $n \in \mathbf{N}$  si ha  $B_n = B_1 \setminus (B_1 \setminus B_n)$ , per cui, essendo anche  $B_1 \setminus B_n \subseteq B_1, \forall n \in \mathbf{N}$ , risulta

$$\mathbf{P}(B_n) = \mathbf{P}(B_1) - \mathbf{P}(B_1 \setminus B_n), \quad \forall n \in \mathbf{N}.$$

Se ora poniamo  $A_n = B_1 \setminus B_n$ , la successione  $(A_n)_{n \in \mathbf{N}}$  risulta crescente e tale che

$$\bigcup_n A_n = \bigcup_n (B_1 \setminus B_n) = B_1 \setminus \left(\bigcap_n B_n\right)$$

è certamente ancora un elemento di  $\mathcal{A}$  (in quanto  $\bigcap_n B_n \in \mathcal{A}$  per ipotesi); pertanto  $(A_n)_{n \in \mathbf{N}}$  soddisfa le ipotesi dell'enunciato 2) per cui sicuramente risulta

$$\lim_n \mathbf{P}(A_n) = \mathbf{P}\left(\bigcup_n A_n\right).$$

Dalla 2) e da  $\bigcap_n B_n \subseteq B_1$  segue allora la tesi:

$$\begin{aligned} \lim_n \mathbf{P}(B_n) &= \mathbf{P}(B_1) - \lim_n \mathbf{P}(A_n) = \mathbf{P}(B_1) - \mathbf{P}\left(\bigcup_n A_n\right) \\ &= \mathbf{P}(B_1) - \mathbf{P}\left[B_1 \setminus \left(\bigcap_n B_n\right)\right] = \mathbf{P}(B_1) - \mathbf{P}(B_1) + \mathbf{P}\left(\bigcap_n B_n\right) \\ &= \mathbf{P}\left(\bigcap_n B_n\right). \end{aligned}$$

3  $\Rightarrow$  4: La dimostrazione è banale dato che 4) può sempre essere considerato come un caso particolare di 3).

4  $\Rightarrow$  1: Per provare la  $\sigma$ -additività di  $\mathbf{P}$  consideriamo una successione  $(D_n)_{n \in \mathbf{N}}$  di elementi di  $\mathcal{A}$  due a due disgiunti e tali che  $\bigcup_n D_n \in \mathcal{A}$ . Siccome gli insiemi

$$\bigcup_{k=1}^n D_k, \quad \bigcup_{k>n} D_k$$

sono ambedue elementi di  $\mathcal{A}$  (il primo perché unione finita di elementi di  $\mathcal{A}$  e il secondo in quanto differenza fra elementi di  $\mathcal{A}$ ) e siccome per ipotesi  $\mathbf{P}$  è additiva, avremo che

$$\mathbf{P}\left(\bigcup_k D_k\right) = \mathbf{P}\left(\bigcup_{k=1}^n D_k\right) + \mathbf{P}\left(\bigcup_{k>n} D_k\right).$$

Se ora definiamo la successione

$$C_n = \bigcup_{k>n} D_k, \quad n \in \mathbf{N}$$

in modo che

$$\bigcup_{k=1}^n D_k = \left(\bigcup_k D_k\right) \setminus C_n,$$

ci accorgiamo subito che essa soddisfa le ipotesi dell'enunciato 4) in quanto essa risulta composta di elementi di  $\mathcal{A}$  (basterà osservare che  $C_n = (\bigcup_k D_k) \setminus (\bigcup_{k \leq n} D_k)$  con  $\bigcup_k D_k \in \mathcal{A}$  per ipotesi), è decrescente (dal momento che  $C_{n+1} = C_n \setminus D_{n+1} \subseteq C_n, \forall n \in \mathbf{N}$ ) e infine è tale che  $\bigcap_n C_n = \emptyset$  perché se  $\bigcap_n C_n$  risultasse non vuoto i  $D_k$  non potrebbero essere tutti disgiunti<sup>3</sup>. Pertanto da 4) segue che  $\lim_n \mathbf{P}(C_n) = 0$  e quindi

$$\begin{aligned} \sum_n \mathbf{P}(D_n) &= \lim_n \sum_{k=1}^n \mathbf{P}(D_k) = \lim_n \mathbf{P}\left(\bigcup_{k=1}^n D_k\right) \\ &= \lim_n \left[ \mathbf{P}\left(\bigcup_k D_k\right) - \mathbf{P}(C_n) \right] = \mathbf{P}\left(\bigcup_k D_k\right) - \lim_n \mathbf{P}(C_n) \\ &= \mathbf{P}\left(\bigcup_n D_n\right), \end{aligned}$$

che prova anche l'ultima implicazione. □

---

<sup>3</sup> Infatti, se  $\bigcap_n C_n$  fosse non vuoto esso conterrebbe almeno un elemento  $\omega \in \bigcap_n C_n$  cioè si avrebbe che  $\omega \in C_n, \forall n \in \mathbf{N}$ ; ma data la definizione di  $C_n$  ciò implicherebbe che  $\forall n \in \mathbf{N} \exists k > n \ni \omega \in D_k$ . Per  $n = 1$  potremmo allora trovare un  $k_1 > 1$  tale che  $\omega \in D_{k_1}$ ; preso però  $n > k_1$  potremmo anche trovare  $k_2 > n > k_1$  tale che  $\omega \in D_{k_2}$ . Ma questo implica che  $D_{k_1}$  e  $D_{k_2}$  (con  $k_1 < k_2$ ) non sarebbero disgiunti, contrariamente all'ipotesi fatta.

**II.1.14 Osservazione:** Abbiamo già fatto notare in II.1.11 che, per definire coerentemente una misura  $\sigma$ -additiva bisogna sempre richiedere esplicitamente che la successione  $(A_n)_{n \in \mathbb{N}}$  di elementi disgiunti di  $\mathcal{A}$  sia tale che  $\bigcup_n A_n \in \mathcal{A}$ . Per evitare di fare volta per volta questa esplicita ipotesi, e quindi per risolvere definitivamente il problema posto in II.1.9, è ora importante introdurre una nuova definizione che avrà il compito di restringere il numero delle algebre di parti di  $\Omega$  a quello delle algebre più opportune per la costruzione dei modelli di probabilità.  $\circ$

**II.1.15 Definizione:** Una famiglia  $\mathcal{F}$  di parti di  $\Omega$  viene detta  $\sigma$ -anello se è un anello e se soddisfa le seguente ulteriore condizione:

$$c. \bigcup_n A_n \in \mathcal{F}, \quad \forall (A_n)_{n \in \mathbb{N}} \in \mathcal{F}^{\mathbb{N}}$$

(che risulta ovviamente più forte di a. della Definizione II.1.2); inoltre  $\mathcal{F}$  viene detta  $\sigma$ -algebra se è un  $\sigma$ -anello e se  $\Omega \in \mathcal{F}$ . D'ora in poi gli elementi di una  $\sigma$ -algebra  $\mathcal{F}$  prenderanno anche il nome di **eventi**.  $\triangle$

**II.1.16 Osservazione:** Si verifica subito, con dimostrazione analoga a quella della Proposizione II.1.3, che in un  $\sigma$ -anello, comunque scelta una successione  $(A_n)_{n \in \mathbb{N}} \in \mathcal{F}^{\mathbb{N}}$ , risulta anche verificata la relazione  $\bigcap_n A_n \in \mathcal{F}$  che, quindi, non deve essere richiesta a parte esplicitamente.  $\circ$

**II.1.17 Definizione:** Si chiama **spazio misurabile** una coppia  $(\Omega, \mathcal{F})$  in cui  $\mathcal{F}$  è un  $\sigma$ -anello di parti di  $\Omega$ ; si chiama invece **spazio probabilizzabile** una coppia  $(\Omega, \mathcal{F})$  in cui  $\mathcal{F}$  è una  $\sigma$ -algebra di eventi di  $\Omega$ .  $\triangle$

**II.1.18 Definizione (Assiomi di Kolmogorov<sup>4</sup>):** Chiameremo **modello di probabilità** (o **spazio di probabilità**) una terna ordinata  $(\Omega, \mathcal{F}, \mathbf{P})$  in cui  $\Omega$  è un insieme di elementi  $\omega$  (detto anche **spazio dei campioni** o **degli eventi elementari**),  $\mathcal{F}$  è una  $\sigma$ -algebra di eventi di  $\Omega$  e  $\mathbf{P}$  è una misura di probabilità definita su  $\mathcal{F}$ .  $\triangle$

**II.1.19 Osservazione:** Sarà utile a questo punto introdurre alcune notazioni.

---

<sup>4</sup> Andrej N. Kolmogorov (1903 - 1987) è stato un importante matematico sovietico che ha dato contributi fondamentali in quasi tutti i principali campi della matematica e dell'applicazione di quest'ultima alla fisica, alla meccanica ed alla biologia. Assieme ai suoi allievi I. M. Gel'fand ed S. M. Nikol'skij egli è stato il fondatore della scuola sovietica di analisi funzionale. Il suo interesse per il calcolo delle probabilità risale al 1925, e la pubblicazione del suo *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Concetti fondamentali della Teoria delle Probabilità, Berlino 1933) segna la definitiva formulazione dei profondi legami tra il calcolo delle probabilità e la teoria metrica degli insiemi e delle funzioni che erano stati messi in luce per la prima volta da Émile Borel (1871 - 1956). Kolmogorov è anche uno dei fondatori della teoria dei processi stocastici stazionari che ha innumerevoli applicazioni nella scienza della natura. Va infine ricordata, nel periodo del dopoguerra, i suoi lavori in collaborazione con V. I. Arnol'd che portarono, nel 1957, alla soluzione del tredicesimo problema di Hilbert, e i suoi profondi interessi per la teoria dell'informazione, i sistemi dinamici e la meccanica teorica.

Innanzitutto ricordiamo che

$$\bigcup_n A_n = \{\omega \in \Omega : \exists n \in \mathbf{N} \ni \omega \in A_n\};$$

$$\bigcap_n A_n = \{\omega \in \Omega : \omega \in A_n \forall n \in \mathbf{N}\}.$$

Data inoltre la successione di eventi  $(A_n)_{n \in \mathbf{N}}$  crescente (decescente) per inclusione, cioè tale che  $A_n \subseteq A_{n+1}$ ,  $\forall n \in \mathbf{N}$  ( $A_n \supseteq A_{n+1}$ ,  $\forall n \in \mathbf{N}$ ), si usa anche indicare la relazione  $A = \bigcup_n A_n$  ( $A = \bigcap_n A_n$ ) con la notazione

$$A_n \uparrow A \quad (A_n \downarrow A);$$

in questo modo le ipotesi degli enunciati 2), 3) e 4) del Teorema II.1.13 potrebbero essere riscritti rispettivamente come

$$A_n \uparrow A, \quad B_n \downarrow B, \quad C_n \downarrow \emptyset.$$

Infine, data una successione di eventi  $(A_n)_{n \in \mathbf{N}}$ , le seguenti notazioni

$$\begin{aligned} \overline{\lim}_n A_n &= \limsup_n A_n = \bigcap_n \bigcup_{k \geq n} A_k \\ &= \{\omega \in \Omega : \forall n \in \mathbf{N}, \exists k \geq n \ni \omega \in A_k\} \\ &= \{\omega \in \Omega : \text{si verifica un numero infinito di eventi } A_n\}, \\ \underline{\lim}_n A_n &= \liminf_n A_n = \bigcup_n \bigcap_{k \geq n} A_k \\ &= \{\omega \in \Omega : \exists n \in \mathbf{N} \ni \omega \in A_k, \forall k \geq n\} \\ &= \{\omega \in \Omega : \text{tutti gli eventi } A_n, \text{ salvo un numero finito, si verificano}\}, \end{aligned}$$

saranno frequentemente utilizzate. ○

**II.1.20 Osservazione:** Nel seguito ci capiterà varie volte di dover provare che una certa famiglia  $\mathcal{A}$  di sottinsiemi di  $\Omega$  è un'algebra verificando le proprietà elencate nella Definizione II.1.2. A volte, però, può riuscire utile usare un insieme di proprietà equivalenti ma diverse da quelle di II.1.2. Ad esempio è possibile mostrare, completando l'enunciato della Proposizione II.1.3, che il seguente insieme di proprietà è equivalente a quello della Definizione II.1.2:

- a.  $\Omega \in \mathcal{A}$ ;
- b.  $\overline{A} \in \mathcal{A}, \quad \forall A \in \mathcal{A}$ ;
- c.  $A \cap B \in \mathcal{A}, \quad \forall A, B \in \mathcal{A}$ .

La dimostrazione di questa equivalenza è lasciata al lettore per esercizio. ○

## II.2 Spazi probabilizzabili

Il presente capitolo sarà dedicato all'analisi delle proprietà degli spazi probabilizzabili  $(\Omega, \mathcal{F})$  e ad un esame dettagliato di alcuni importanti esempi.

**II.2.1 Esempio:** Dato un insieme  $\Omega$  i più semplici esempi di  $\sigma$ -algebre sono i seguenti (vedi anche l'Esempio I.2.9):

$$\begin{aligned}\mathcal{F}_* &= \{\emptyset, \Omega\}, \\ \mathcal{F}_A &= \{A, \bar{A}, \emptyset, \Omega\}, \quad (A \subseteq \Omega), \\ \mathcal{F}^* &= \wp(\Omega).\end{aligned}$$

La  $\sigma$ -algebra  $\mathcal{F}_A$  prende il nome di  **$\sigma$ -algebra generata da  $A$**  ed è l'esempio più semplice di una situazione più generale: data una decomposizione finita o numerabile di  $\Omega$  (vedi Definizione I.2.10 estendendola facilmente al caso numerabile)  $\mathcal{D} = \{D_1, D_2, \dots\}$ , si verifica che la famiglia  $\mathcal{A} = \alpha(\mathcal{D})$  di tutte le unioni finite di elementi di  $\mathcal{D}$  è un'algebra; anzi si potrebbe dimostrare che essa è la più piccola algebra contenente  $\mathcal{D}$ . Generalizzando ancora le osservazioni precedenti, possiamo porci il problema di costruire, a partire da una data famiglia  $\mathcal{E} \subseteq \wp(\Omega)$  di parti di  $\Omega$ , la più piccola algebra  $\alpha(\mathcal{E})$  e la più piccola  $\sigma$ -algebra  $\sigma(\mathcal{E})$  contenenti  $\mathcal{E}$ : queste, se esistono, prendono rispettivamente i nomi di **algebra generata** e  **$\sigma$ -algebra generata da  $\mathcal{E}$** . Nel seguito verificheremo che questa costruzione è sempre possibile.  $\diamond$

**II.2.2 Proposizione:** Data una qualunque famiglia  $(\mathcal{F}_t)_{t \in T}$  di  $\sigma$ -algebre di parti di  $\Omega$  (l'insieme  $T$  degli indici è arbitrario), la famiglia intersezione  $\mathcal{F} = \bigcap_t \mathcal{F}_t$  è ancora una  $\sigma$ -algebra. Se le  $\mathcal{F}_t$  fossero solo algebre, anche  $\mathcal{F}$  sarebbe un'algebra.

**Dimostrazione:** Sicuramente  $\mathcal{F}$  contiene  $\Omega$  e  $\emptyset$  che, per definizione, sono elementi comuni a tutte le  $\sigma$ -algebre non vuote. Consideriamo ora una successione  $(A_n)_{n \in \mathbf{N}}$  di elementi di  $\mathcal{F}$ : per definizione di  $\mathcal{F}$  avremo allora che  $A_n \in \mathcal{F}_t, \forall n \in \mathbf{N}, \forall t \in T$ . Ne segue allora, dalle proprietà delle  $\sigma$ -algebre, che  $\bigcup_n A_n \in \mathcal{F}_t, \forall t \in T$  e quindi  $\bigcup_n A_n \in \mathcal{F}$ ; inoltre, dati  $A, B \in \mathcal{F}$ , risulterà  $A, B \in \mathcal{F}_t, \forall t \in T$  e quindi  $A \setminus B \in \mathcal{F}_t, \forall t \in T$ , cioè in definitiva  $A \setminus B \in \mathcal{F}$ .  $\square$

**II.2.3 Proposizione:** Data una famiglia  $\mathcal{E} \subseteq \wp(\Omega)$  di parti di  $\Omega$  esistono sempre l'algebra  $\alpha(\mathcal{E})$  e la  $\sigma$ -algebra  $\sigma(\mathcal{E})$  generate da  $\mathcal{E}$ .

**Dimostrazione:** Innanzitutto, siccome  $\mathcal{F}^* = \wp(\Omega)$  è sia un'algebra che una  $\sigma$ -algebra, sicuramente esistono almeno un'algebra ed una  $\sigma$ -algebra contenenti  $\mathcal{E}$ . Basterà allora definire  $\alpha(\mathcal{E})$  e  $\sigma(\mathcal{E})$  come le intersezioni rispettivamente di tutte

le algebre e di tutte le  $\sigma$ -algebre contenenti  $\mathcal{E}$ : la Proposizione II.2.2 ci garantisce allora che tali  $\alpha(\mathcal{E})$  e  $\sigma(\mathcal{E})$  sono un'algebra ed una  $\sigma$ -algebra mentre la loro costruzione ci assicura che si tratta proprio dell'algebra e della  $\sigma$ -algebra più piccole che contengono  $\mathcal{E}$ . Ne consegue che  $\alpha(\mathcal{E})$  e  $\sigma(\mathcal{E})$  sono l'algebra e la  $\sigma$ -algebra generate da  $\mathcal{E}$ .  $\square$

**II.2.4 Definizione:** Sia  $\mathcal{E}$  una famiglia di parti di  $\Omega$  e  $B \subseteq \Omega$  un generico sottinsieme di  $\Omega$ ; la famiglia di parti di  $B$  definita come

$$\mathcal{E} \cap B = \{A \in \wp(B) : A = E \cap B, E \in \mathcal{E}\}$$

prende il nome di **traccia** di  $\mathcal{E}$  su  $B$ .  $\triangle$

**II.2.5 Lemma:** Se  $\mathcal{F}$  è una  $\sigma$ -algebra di parti di  $\Omega$  e  $B \subseteq \Omega$  un sottinsieme di  $\Omega$ , la traccia  $\mathcal{F} \cap B$  è una  $\sigma$ -algebra di parti di  $B$ .

**Dimostrazione:** Basterà verificare che la famiglia  $\mathcal{F} \cap B$  di parti di  $B$  soddisfa le proprietà di una  $\sigma$ -algebra:

- 1) Innanzitutto  $B \in \mathcal{F} \cap B$ , dato che  $B = \Omega \cap B$  con  $\Omega \in \mathcal{F}$ .
- 2) Se  $A_1, A_2 \in \mathcal{F} \cap B$ , potremo sempre trovare  $C_1, C_2 \in \mathcal{F}$  tali che  $A_1 = C_1 \cap B$  e  $A_2 = C_2 \cap B$ ; ne segue che  $A_1 \cup A_2 = (C_1 \cup C_2) \cap B$  è un elemento di  $\mathcal{F} \cap B$  dato che  $C_1 \cup C_2 \in \mathcal{F}$ .
- 3) Con le stesse  $A_1$  ed  $A_2$  del punto 2) abbiamo inoltre che

$$\begin{aligned} A_1 \setminus A_2 &= A_1 \cap \overline{A_2} = (C_1 \cap B) \cap (\overline{C_2} \cup \overline{B}) = C_1 \cap B \cap \overline{C_2} \\ &= (C_1 \setminus C_2) \cap B, \end{aligned}$$

ed essendo  $C_1 \setminus C_2 \in \mathcal{F}$  questo prova anche che  $A_1 \setminus A_2 \in \mathcal{F} \cap B$ .

- 4) Se  $(A_n)_{n \in \mathbf{N}}$  è una successione di elementi di  $\mathcal{F} \cap B$ , potremo sempre determinare una successione  $(C_n)_{n \in \mathbf{N}}$  di elementi di  $\mathcal{F}$  tali che  $A_n = C_n \cap B, \forall n \in \mathbf{N}$ ; ne segue che  $\bigcup_n A_n = (\bigcup_n C_n) \cap B$  è elemento di  $\mathcal{F} \cap B$ , dato che  $\bigcup_n C_n \in \mathcal{F}$ .  $\square$

**II.2.6 Lemma:** Se  $\mathcal{F}$  è una  $\sigma$ -algebra di parti di  $\Omega$ ,  $B \subseteq \Omega$  un sottinsieme di  $\Omega$  e  $\mathcal{G}_B$  una  $\sigma$ -algebra di parti di  $B$ , la sottofamiglia di  $\mathcal{F}$

$$\mathcal{C}_B = \{A \in \mathcal{F} : A \cap B \in \mathcal{G}_B\}$$

è una  $\sigma$ -algebra di parti di  $\Omega$ .

**Dimostrazione:** Anche in questo caso dovremo verificare che  $\mathcal{C}_B$  soddisfa le proprietà di una  $\sigma$ -algebra:

- 1) Innanzitutto  $\Omega \in \mathcal{C}_B$ , dato che  $\Omega \in \mathcal{F}$  e  $\Omega \cap B = B \in \mathcal{G}_B$ .

- 2) Dati  $A_1, A_2 \in \mathcal{C}_B$ , certamente sarà  $A_1 \cup A_2 \in \mathcal{F}$ , dato che  $A_1, A_2 \in \mathcal{F}$ , e inoltre

$$(A_1 \cup A_2) \cap B = (A_1 \cap B) \cup (A_2 \cap B)$$

sarà un elemento di  $\mathcal{G}_B$  dato che  $A_1 \cap B, A_2 \cap B \in \mathcal{G}_B$ ; ne segue che  $A_1 \cup A_2 \in \mathcal{C}_B$ .

- 3) Dati  $A_1, A_2 \in \mathcal{C}_B$ , per le stesse ragioni del punto precedente risulterà  $A_1 \setminus A_2 \in \mathcal{F}$ , e inoltre

$$\begin{aligned} (A_1 \setminus A_2) \cap B &= A_1 \cap \overline{A_2} \cap B = (A_1 \cap \overline{A_2} \cap B) \cup (A_1 \cap \overline{B} \cap B) \\ &= (A_1 \cap B) \cap (\overline{A_2} \cup \overline{B}) = (A_1 \cap B) \setminus (A_2 \cap B) \end{aligned}$$

è un elemento di  $\mathcal{G}_B$  dato che per definizione  $A_1 \cap B$  e  $A_2 \cap B$  sono elementi di  $\mathcal{G}_B$ ; ne segue che  $A_1 \setminus A_2 \in \mathcal{C}_B$ .

- 4) Se  $(A_n)_{n \in \mathbf{N}}$  è una successione di elementi di  $\mathcal{C}_B$ , risulta innanzitutto  $\bigcup_n A_n \in \mathcal{F}$  dato che  $A_n \in \mathcal{F}, \forall n \in \mathbf{N}$ , e inoltre  $(\bigcup_n A_n) \cap B = \bigcup_n (A_n \cap B)$  risulta essere elemento di  $\mathcal{G}_B$  visto che  $A_n \cap B \in \mathcal{G}_B, \forall n \in \mathbf{N}$ ; ne segue che  $\bigcup_n A_n \in \mathcal{C}_B$ .  $\square$

**II.2.7 Proposizione:** Data una famiglia  $\mathcal{E}$  di parti di  $\Omega$  e un sottinsieme  $B \subseteq \Omega$ , l'eguaglianza

$$\sigma(\mathcal{E} \cap B) = \sigma(\mathcal{E}) \cap B$$

è sempre verificata, sicché la  $\sigma$ -algebra generata dalla traccia di  $\mathcal{E}$  su  $B$  coincide sempre con la traccia (su  $B$ ) della  $\sigma$ -algebra generata da  $\mathcal{E}$ .

**Dimostrazione:** La tesi viene dimostrata provando la mutua inclusione delle  $\sigma$ -algre  $\sigma(\mathcal{E} \cap B)$  e  $\sigma(\mathcal{E}) \cap B$ :

- A) Proviamo innanzitutto che  $\sigma(\mathcal{E} \cap B) \subseteq \sigma(\mathcal{E}) \cap B$ : dato che per definizione  $\mathcal{E} \subseteq \sigma(\mathcal{E})$ , risulta anche  $\mathcal{E} \cap B \subseteq \sigma(\mathcal{E}) \cap B$ ; ma dal Lemma II.2.5 sappiamo che  $\sigma(\mathcal{E}) \cap B$  è una  $\sigma$ -algebra e quindi, siccome  $\sigma(\mathcal{E} \cap B)$  è per definizione la più piccola  $\sigma$ -algebra contenente  $\mathcal{E} \cap B$ , non può che risultare  $\sigma(\mathcal{E} \cap B) \subseteq \sigma(\mathcal{E}) \cap B$ .
- B) Proviamo ora l'inclusione opposta  $\sigma(\mathcal{E}) \cap B \subseteq \sigma(\mathcal{E} \cap B)$ . Questa parte della dimostrazione è un primo esempio di uso del cosiddetto *principio degli insiemi appropriati*: per provare che le intersezioni  $A \cap B$  di eventi  $A \in \sigma(\mathcal{E})$  con  $B$  risultano tutte essere anche elementi di  $\sigma(\mathcal{E} \cap B)$  si definisce la famiglia degli elementi  $A$  di  $\sigma(\mathcal{E})$  che soddisfano la proprietà da noi richiesta

$$\mathcal{C}_B = \{A \in \sigma(\mathcal{E}) : A \cap B \in \sigma(\mathcal{E} \cap B)\}$$

e si dimostra che tale  $\mathcal{C}_B$  coincide proprio con  $\sigma(\mathcal{E})$ . Infatti, per definizione,  $\mathcal{C}_B \subseteq \sigma(\mathcal{E})$ . Inoltre il Lemma II.2.6 ci dice che  $\mathcal{C}_B$  è una  $\sigma$ -algebra di parti di  $\Omega$  e si vede subito che  $\mathcal{E} \subseteq \mathcal{C}_B$  in quanto, se  $A \in \mathcal{E} \subseteq \sigma(\mathcal{E})$ , dalla definizione di  $\mathcal{E} \cap B$  si ha che  $A \cap B \in \mathcal{E} \cap B \subseteq \sigma(\mathcal{E} \cap B)$ . Siccome però  $\sigma(\mathcal{E})$  è la più piccola  $\sigma$ -algebra contenente  $\mathcal{E}$  non potrà che risultare  $\sigma(\mathcal{E}) \subseteq \mathcal{C}_B$ . Conseguentemente  $\mathcal{C}_B = \sigma(\mathcal{E})$  e quindi la proposizione è completamente dimostrata.  $\square$

**II.2.8 Definizione:** Una famiglia  $\mathcal{M}$  di parti di  $\Omega$  si dice **classe monotona** se, comunque presa una successione (crescente o decrescente)  $(A_n)_{n \in \mathbf{N}}$  di suoi elementi con  $A_n \uparrow A$  o  $A_n \downarrow A$ , risulta sempre  $A \in \mathcal{M}$  (vedi Osservazione II.1.19). Data inoltre una famiglia  $\mathcal{E}$  di parti di  $\Omega$  la più piccola classe monotona  $\mu(\mathcal{E})$  contenente  $\mathcal{E}$ , che esiste, come affermato nella proposizione seguente, prende il nome di **classe monotona generata da  $\mathcal{E}$** .  $\triangle$

**II.2.9 Proposizione:** Data una famiglia  $\mathcal{E}$  di parti di  $\Omega$  esiste sempre la classe monotona  $\mu(\mathcal{E})$  generata da  $\mathcal{E}$ .

**Dimostrazione:** Analoga a quella della Proposizione II.2.3.  $\square$

**II.2.10 Lemma:** Un'algebra  $\mathcal{A}$  è anche una  $\sigma$ -algebra se e solo se essa è una classe monotona.

**Dimostrazione:**

- A) Se  $\mathcal{A}$  è una  $\sigma$ -algebra essa è sicuramente anche una classe monotona in quanto essa contiene tutte le unioni ed intersezioni numerabili dei suoi eventi.
- B) Se l'algebra  $\mathcal{A}$  è una classe monotona, presa una successione  $(A_n)_{n \in \mathbf{N}}$  di suoi elementi si definisce la successione crescente  $B_n = \bigcup_{k=1}^n A_k \in \mathcal{A}$  (ricorda che  $\mathcal{A}$  è un'algebra): ma allora  $B_n \uparrow \bigcup_n A_n$  e, per definizione di classe monotona,  $\bigcup_n A_n \in \mathcal{A}$ . Pertanto  $\mathcal{A}$  è  $\sigma$ -additiva e quindi è anche una  $\sigma$ -algebra.  $\square$

**II.2.11 Proposizione:** Se  $\mathcal{A}$  è un'algebra di parti di  $\Omega$  l'eguaglianza

$$\mu(\mathcal{A}) = \sigma(\mathcal{A})$$

è sempre verificata.

**Dimostrazione:** Dal punto A) del Lemma II.2.10 sappiamo che una  $\sigma$ -algebra è sempre anche una classe monotona e pertanto, per definizione di classe monotona generata da una famiglia di parti di  $\Omega$ , risulterà sicuramente  $\mu(\mathcal{A}) \subseteq \sigma(\mathcal{A})$ . Per dimostrare il teorema basterà allora provare che  $\mu(\mathcal{A})$  è anche una  $\sigma$ -algebra; ma siccome  $\mathcal{M} = \mu(\mathcal{A})$  è una classe monotona, dal Lemma II.2.10 si ricava che è sufficiente provare che  $\mathcal{M}$  è un'algebra. Per far questo utilizzeremo la definizione di algebra data mediante le proprietà elencate nell'Osservazione II.1.20:

- a. Sicuramente  $\Omega \in \mathcal{M}$  dato che, essendo  $\mathcal{A}$  un'algebra, risulta  $\Omega \in \mathcal{A} \subseteq \mu(\mathcal{A})$ .
- b. Per dimostrare che  $\overline{A} \in \mathcal{M}$  se  $A \in \mathcal{M}$  faremo uso ancora una volta del principio degli insiemi appropriati: definiamo a questo scopo la famiglia di eventi

$$\overline{\mathcal{M}} = \{B \in \mathcal{M} : \overline{B} \in \mathcal{M}\}$$



e proviamo che  $\overline{\mathcal{M}} = \mathcal{M}$ . Innanzitutto è evidente che  $\mathcal{A} \subseteq \overline{\mathcal{M}} \subseteq \mathcal{M}$ : infatti  $\overline{\mathcal{M}} \subseteq \mathcal{M}$  per definizione e inoltre, essendo  $\mathcal{A}$  un'algebra, comunque scelto  $A \in \mathcal{A}$  risulterà  $\overline{A} \in \mathcal{A}$ . Se allora riusciamo a dimostrare che  $\overline{\mathcal{M}}$  è anche una classe monotona avremo provato che  $\overline{\mathcal{M}} = \mathcal{M}$  dato che per definizione  $\mathcal{M}$  è la più piccola classe monotona contenente  $\mathcal{A}$ . Per far questo consideriamo una successione crescente  $(B_n)_{n \in \mathbf{N}}$  di elementi di  $\overline{\mathcal{M}}$  (la prova è analoga nel caso di successioni decrescenti) e proviamo che se  $B_n \uparrow B$  allora  $B \in \overline{\mathcal{M}}$ , cioè  $B \in \mathcal{M}$  e  $\overline{B} \in \mathcal{M}$ . Infatti, per definizione,  $B_n \in \mathcal{M}$  e  $\overline{B_n} \in \mathcal{M}$ ; pertanto  $(B_n)_{n \in \mathbf{N}}$  è una successione crescente di elementi di  $\mathcal{M}$ , mentre  $(\overline{B_n})_{n \in \mathbf{N}}$  è una successione decrescente di elementi di  $\mathcal{M}$ . Essendo  $\mathcal{M}$  una classe monotona si avrà allora che

$$B_n \uparrow B = \bigcup_n B_n \in \mathcal{M}; \quad \overline{B_n} \downarrow \bigcap_n \overline{B_n} = \overline{\left( \bigcup_n B_n \right)} = \overline{B} \in \mathcal{M}$$

e quindi  $B \in \overline{\mathcal{M}}$ .

- c. Per provare che  $A \cap B \in \mathcal{M}$  se  $A, B \in \mathcal{M}$ , preso un arbitrario  $A \in \mathcal{M}$  definiamo la famiglia

$$\mathcal{M}_A = \{B \in \mathcal{M} : A \cap B \in \mathcal{M}\},$$

e dimostriamo che  $\mathcal{M}_A = \mathcal{M}$ . Innanzitutto si verifica subito che  $\mathcal{M}_A$  è una classe monotona: presa infatti una successione crescente  $(B_n)_{n \in \mathbf{N}}$  di elementi di  $\mathcal{M}_A \subseteq \mathcal{M}$ , risulta  $A \cap B_n \in \mathcal{M}$ ,  $\forall n \in \mathbf{N}$  e quindi, essendo  $\mathcal{M}$  una classe monotona,  $A \cap B_n \uparrow \bigcup_n (A \cap B_n) \in \mathcal{M}$ ; ma  $\bigcup_n (A \cap B_n) = A \cap \left( \bigcup_n B_n \right)$ , per cui  $B_n \uparrow \bigcup_n B_n \in \mathcal{M}_A$ . Inoltre si controlla subito che, se  $A \in \mathcal{A}$ , risulta  $\mathcal{A} \subseteq \mathcal{M}_A \subseteq \mathcal{M}$  in quanto, essendo  $\mathcal{A}$  un'algebra, comunque scelto  $B \in \mathcal{A}$  si ha  $A \cap B \in \mathcal{A}$ . Siccome però  $\mathcal{M}$  contiene  $\mathcal{A}$ , si ottiene subito che  $B \in \mathcal{M}$  e  $A \cap B \in \mathcal{M}$ , cioè che  $B \in \mathcal{M}_A$ . In realtà si può dire di più: dato che  $\mathcal{M}$  è la più piccola classe monotona contenente  $\mathcal{A}$ , risulta anche  $\mathcal{M}_A = \mathcal{M}$  se  $A \in \mathcal{A}$ . Questa osservazione si rivela utile per estendere questa relazione anche al caso più generale in cui  $A \in \mathcal{M}$ . Infatti, comunque scelti  $A$  e  $B$  in  $\mathcal{M}$ ,

$$A \in \mathcal{M}_B \iff B \in \mathcal{M}_A$$

e in particolare, comunque scelti  $B \in \mathcal{A}$  e  $A \in \mathcal{M}$ ,

$$A \in \mathcal{M}_B = \mathcal{M} \iff B \in \mathcal{M}_A.$$

Ne segue che  $\mathcal{A} \subseteq \mathcal{M}_A$  ( $A \in \mathcal{M}$ ), in quanto se  $B \in \mathcal{A}$ , risulta  $\mathcal{M} = \mathcal{M}_B$  e quindi  $A \in \mathcal{M} = \mathcal{M}_B$  per cui anche  $B \in \mathcal{M}_A$ . Dunque  $\mathcal{A} \subseteq \mathcal{M}_A \subseteq \mathcal{M}$ ,  $\forall A \in \mathcal{M}$ . Siccome però  $\mathcal{M}$  è la più piccola classe monotona contenente  $\mathcal{A}$  si ha in definitiva che  $\mathcal{M}_A = \mathcal{M}$ ,  $\forall A \in \mathcal{M}$ .  $\square$

**II.2.12 Esempio:** Cominceremo con l'esaminare il caso in cui lo spazio dei campioni  $\Omega$  è l'insieme  $\mathbf{R}$  dei numeri reali  $(-\infty, +\infty)$ . Converremo inoltre di indicare con  $\mathcal{I}$  la famiglia degli **intervalli** (limitati o illimitati) chiusi a destra

$$I = (a, b], \quad -\infty \leq a < b \leq +\infty$$

con l'aggiunta<sup>1</sup> dell'insieme vuoto  $\emptyset$ ; inoltre, per convenzione, con il simbolo  $(a, +\infty]$  intenderemo indicare l'intervallo  $(a, +\infty)$ . In pratica  $\mathcal{I}$  contiene tutte le parti di  $\mathbf{R}$  del tipo

$$\emptyset, \quad (a, b], \quad (-\infty, b], \quad (a, +\infty] = (a, +\infty), \quad (-\infty, +\infty] = (-\infty, +\infty) = \mathbf{R}$$

con  $a, b \in \mathbf{R}$ . Siccome le intersezioni di intervalli chiusi a destra sono ancora intervalli chiusi a destra, e le differenze di intervalli chiusi a destra possono sempre essere ricostruite come unioni di un numero finito di intervalli chiusi a destra, si controlla facilmente che  $\mathcal{I}$  è una semialgebra (ricorda che  $\mathbf{R} \in \mathcal{I}$ ). Dato però che le unioni di intervalli chiusi a destra non sono, in generale, intervalli chiusi a destra,  $\mathcal{I}$  non è un'algebra. Se allora indichiamo con  $\mathcal{A}$  la famiglia di tutte le unioni finite di elementi di  $\mathcal{I}$ , la Proposizione II.1.6 ci autorizza ad affermare che  $\mathcal{A}$  è un'algebra. Un semplice esempio mostra però che  $\mathcal{A}$  non è una  $\sigma$ -algebra: infatti  $(0, 1 - \frac{1}{n}] \in \mathcal{A}, \forall n \in \mathbf{N}$ , ma  $\bigcup_n (0, 1 - \frac{1}{n}] = (0, 1)$  non è un elemento di  $\mathcal{A}$  perché non è chiuso a destra. È naturale allora prendere in considerazione la  $\sigma$ -algebra  $\sigma(\mathcal{A})$  generata da  $\mathcal{A}$ : in particolare essa conterrà tutte le parti di  $\mathbf{R}$  della forma

$$\{a\}, \quad [a, b], \quad [a, b), \quad (a, b], \quad (a, b),$$

dato che, ad esempio, gli insiemi di numeri reali

$$(a, b) = \bigcup_n \left( a, b - \frac{1}{n} \right], \quad [a, b] = \bigcap_n \left( a - \frac{1}{n}, b \right],$$

sono elementi di  $\sigma(\mathcal{A})$  in quanto unioni o intersezioni numerabili di elementi di  $\mathcal{A} \subseteq \sigma(\mathcal{A})$ . Si può inoltre dimostrare<sup>2</sup>, anche se noi ometteremo di farlo per brevità, che la medesima  $\sigma$ -algebra può essere generata anche partendo dalla semialgebra  $\mathcal{I}$  degli intervalli, cioè che  $\sigma(\mathcal{A}) = \sigma(\mathcal{I})$ ; anzi si può mostrare che si otterrebbe sempre la stessa  $\sigma$ -algebra anche partendo dalle famiglie degli intervalli aperti o chiusi o chiusi solo a sinistra. La  $\sigma$ -algebra costruita in una delle suddette equivalenti maniere viene poi in genere indicata con il simbolo  $\mathcal{B}(\mathbf{R})$  (o, a volte, anche semplicemente con  $\mathcal{B}$ ) e prende il nome di  **$\sigma$ -algebra di Borel**<sup>3</sup> di  $\mathbf{R}$ , e a

<sup>1</sup> Quest'aggiunta è necessaria perché intervalli chiusi solo a destra non sono mai vuoti dato che risulta sempre  $a < b$ .

<sup>2</sup> Vedi ad esempio **O. Lessi**: *Corso di Calcolo delle Probabilità*; Metria, Padova, 1990; p. 301.

<sup>3</sup> Émile Borel (1871 - 1956) è stato uno dei fondatori della teoria della misura di insiemi di punti assieme ad altri matematici francesi come Henri Lebesgue (1875 - 1941) e René Baire (1874 - 1932). Professore all'École Normale Supérieure dal 1896, si segnalò inizialmente per aver scoperto la prima prova elementare del teorema enunciato nel 1879 da Émile Picard (1856 - 1941) secondo il quale ogni funzione (di variabile complessa) intera non costante assume ogni valore complesso finito con al più una eccezione (vedi per esempio **B. Chabat**: *Introduction à l'Analyse Complexe*; MIR, Mosca, 1990, p. 234). Oltre ad aver dato contributi fondamentali alla moderna teoria delle funzioni di variabile reale ha scritto importanti lavori sulle serie divergenti e sulla

loro volta i suoi elementi vengono chiamati **boreliani di  $\mathbf{R}$** . Pertanto lo spazio probabilizzabile così ottenuto verrà, d'ora in poi indicato come

$$(\mathbf{R}, \mathcal{B}(\mathbf{R})).$$

A volte può riuscire utile usare uno spazio probabilizzabile costruito a partire dalla retta reale estesa  $\overline{\mathbf{R}} = [-\infty, +\infty]$ . Le regole delle operazioni algebriche su  $\overline{\mathbf{R}}$  vengono convenzionalmente completate con le seguenti prescrizioni:

$$\begin{aligned} x \cdot (\pm\infty) &= \begin{cases} \pm\infty, & \text{se } x > 0; \\ \mp\infty, & \text{se } x < 0; \end{cases} \\ 0 \cdot (\pm\infty) &= 0; \\ x \pm \infty &= \pm\infty, \quad \text{se } x \neq \mp\infty; \\ (\pm\infty) + (\pm\infty) &= \pm\infty; \end{aligned}$$

non viene invece assegnato alcun significato all'operazione  $+\infty - \infty$ . In questo caso l'insieme  $\mathcal{I}$  viene definito allo stesso modo come la semialgebra degli intervalli chiusi a destra, con la convenzione che  $(-\infty, b] = [-\infty, b]$ . La costruzione della  $\sigma$ -algebra generata da  $\mathcal{I}$  avviene nello stesso modo del caso precedente e lo spazio probabilizzabile così ottenuto si indica con il simbolo

$$(\overline{\mathbf{R}}, \mathcal{B}(\overline{\mathbf{R}})).$$

Notiamo infine che la nozione di  $\sigma$ -algebra di Borel è più generale di quella introdotta qui e può essere applicata al caso in cui  $\Omega$  è un generico spazio topologico: detta infatti  $\mathcal{A}$  la famiglia degli aperti di  $\Omega$ , chiameremo  **$\sigma$ -algebra di Borel di  $\Omega$**  la  $\sigma$ -algebra  $\mathcal{B}(\Omega) = \sigma(\mathcal{A})$  generata dalla famiglia degli aperti ed indicheremo con

$$(\Omega, \mathcal{B}(\Omega))$$

lo spazio probabilizzabile così ottenuto. Nel caso particolare da noi esaminato,  $\Omega = \mathbf{R}$ , abbiamo generato la  $\sigma$ -algebra  $\mathcal{B}(\mathbf{R})$  partendo dal sistema degli intervalli chiusi a destra invece che dal sistema degli intervalli aperti: le due costruzioni, come abbiamo già osservato sono però equivalenti; la motivazione della scelta verrà resa comunque più chiara successivamente nell'ambito della discussione sulle Funzioni di Distribuzione di Probabilità.  $\diamond$

**II.2.13 Esempio:** Consideriamo ora il caso in cui  $\Omega = \mathbf{R}^n = \mathbf{R}_1 \times \mathbf{R}_2 \times \dots \times \mathbf{R}_n$  è il prodotto cartesiano di  $n$  copie della retta reale; in tal caso gli elementi di  $\Omega$  sono le  $n$ -ple di numeri reali  $\omega = \{x_1, x_2, \dots, x_n\}$ . Come nell'esempio precedente,

---

Teoria dei giochi. La sua vita fu anche caratterizzata da un costante impegno politico: deputato (dal 1924 al 1936) e ministro della Marina (dal 1925 al 1940), fu arrestato (anche se brevemente) dal regime collaborazionista di Vichy nel 1940, prese successivamente parte alla Resistenza contro i nazisti e fu per questo decorato nel 1945.

anche in questo caso ci sono varie maniere equivalenti di definire la  $\sigma$ -algebra necessaria per costruire uno spazio probabilizzabile: si inizia definendo su ogni  $\mathbf{R}_k$  ( $k = 1, \dots, n$ ) la semialgebra  $\mathcal{I}_k$  degli intervalli chiusi a destra, l'algebra  $\mathcal{A}_k$  delle unioni finite di elementi di  $\mathcal{I}_k$  e la  $\sigma$ -algebra di Borel  $\mathcal{B}_k = \mathcal{B}(\mathbf{R}_k)$ . Si passa poi a definire le seguenti semialgre di parti di  $\mathbf{R}^n$  i cui elementi sono detti **rettangoli**:

$$\begin{aligned} \mathcal{I}^n &= \{I \subseteq \mathbf{R}^n : I = I_1 \times \dots \times I_n; I_k \in \mathcal{I}_k; k = 1, \dots, n\}, \\ \mathcal{R}^n &= \{B \subseteq \mathbf{R}^n : B = B_1 \times \dots \times B_n; B_k \in \mathcal{B}_k; k = 1, \dots, n\}. \end{aligned}$$

È facile accorgersi del fatto che gli elementi di  $\mathcal{I}^n$  sono rettangoli  $n$ -dimensionali i cui lati sono intervalli della retta reale, mentre gli elementi di  $\mathcal{R}^n$  sono rettangoli  $n$ -dimensionali i cui lati sono boreliani della retta reale. Indicheremo poi con  $\sigma(\mathcal{I}^n)$  e  $\sigma(\mathcal{R}^n)$  le  $\sigma$ -algre generate dalle suddette semialgre e con  $\mathcal{B}(\mathbf{R}^n)$  la  $\sigma$ -algebra di Borel generata a partire dalla famiglia degli aperti di  $\mathbf{R}^n$  munito della usuale topologia. La  $\sigma$ -algebra  $\mathcal{B}(\mathbf{R}^n)$  viene anche indicata con i simboli

$$\bigotimes_{k=1}^n \mathcal{B}_k = \mathcal{B}_1 \otimes \dots \otimes \mathcal{B}_n = \mathcal{B}(\mathbf{R}_1) \otimes \dots \otimes \mathcal{B}(\mathbf{R}_n).$$

La proposizione seguente mostra che tutte queste  $\sigma$ -algre sono coincidenti, sicché potremo indicare con il simbolo

$$(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$$

lo spazio probabilizzabile corrispondente al caso  $\Omega = \mathbf{R}^n$ . ◇

**II.2.14 Proposizione:**  $\sigma(\mathcal{I}^n) = \sigma(\mathcal{R}^n) = \mathcal{B}(\mathbf{R}^n)$ .

**Dimostrazione:** Ometteremo<sup>4</sup> la dimostrazione della eguaglianza con  $\mathcal{B}(\mathbf{R}^n)$  e ci limiteremo a dimostrare la prima eguaglianza nel caso  $n = 2$ .

- a) Innanzitutto si verifica banalmente che  $\sigma(\mathcal{I}^n) \subseteq \sigma(\mathcal{R}^n)$  in quanto  $\mathcal{I}_k \subseteq \mathcal{B}_k$  e quindi  $\mathcal{I}^n \subseteq \mathcal{R}^n$ .
- b) Per completare la dimostrazione sarà allora sufficiente mostrare che  $\sigma(\mathcal{R}^n) \subseteq \sigma(\mathcal{I}^n)$  e quindi, dato che  $\sigma(\mathcal{I}^n)$  è una  $\sigma$ -algebra, basterà mostrare che  $\mathcal{R}^n \subseteq \sigma(\mathcal{I}^n)$ . Limitandoci per brevità al caso  $n = 2$ , sarà utile introdurre le seguenti notazioni:

$$\begin{aligned} \mathcal{C}'_1 &= \{C' \subseteq \mathbf{R}^2 : C' = I_1 \times \mathbf{R}_2, I_1 \in \mathcal{I}_1\}, \\ \mathcal{C}'_2 &= \{C' \subseteq \mathbf{R}^2 : C' = \mathbf{R}_1 \times I_2, I_2 \in \mathcal{I}_2\}. \end{aligned}$$

---

<sup>4</sup> Per i dettagli vedi **M. Métivier**: *Notions Fondamentales de la Théorie des Probabilités*; Dunod, Paris, 1972, pp. 35 e 40.

Si tratta chiaramente delle famiglie delle strisce in  $\mathbf{R}^2$  che hanno intervalli per basi. Dato che  $\mathcal{I}_1$  e  $\mathcal{I}_2$  non sono delle  $\sigma$ -algebre,  $\mathcal{C}'_1$  e  $\mathcal{C}'_2$  non saranno  $\sigma$ -algebre. Notiamo anche che definendo

$$\mathcal{C}' = \{A \subseteq \mathbf{R}^2 : A = C'_1 \cap C'_2; C'_1 \in \mathcal{C}'_1, C'_2 \in \mathcal{C}'_2\},$$

risulta che  $\mathcal{C}' = \mathcal{I}^2$  dato che ogni  $A = C'_1 \cap C'_2 = (I_1 \times \mathbf{R}_2) \cap (\mathbf{R}_1 \times I_2) = I_1 \times I_2$  elemento di  $\mathcal{C}'$  è anche elemento di  $\mathcal{I}^2$ . Si definiscono poi anche le famiglie delle strisce con boreliani per basi:

$$\begin{aligned} \mathcal{C}_1 &= \{C \subseteq \mathbf{R}^2 : C = B_1 \times \mathbf{R}_2, B_1 \in \mathcal{B}_1\}, \\ \mathcal{C}_2 &= \{C \subseteq \mathbf{R}^2 : C = \mathbf{R}_1 \times B_2, B_2 \in \mathcal{B}_2\}, \end{aligned}$$

e si verifica facilmente che, essendo  $\mathcal{B}_1$  e  $\mathcal{B}_2$  delle  $\sigma$ -algebre, si tratta di  $\sigma$ -algebre di parti di  $\mathbf{R}^2$ . Data la evidente corrispondenza biunivoca tra gli elementi di  $\mathcal{C}'_1, \mathcal{C}'_2, \mathcal{C}_1, \mathcal{C}_2$  e quelli di  $\mathcal{I}_1, \mathcal{I}_2, \mathcal{B}_1, \mathcal{B}_2$ , si può anche far vedere che

$$\mathcal{C}_1 = \sigma(\mathcal{C}'_1); \quad \mathcal{C}_2 = \sigma(\mathcal{C}'_2).$$

Per mostrare ora che  $\mathcal{R}^2 \subseteq \sigma(\mathcal{I}^2)$  fissiamo un arbitrario  $B_1 \times B_2 \in \mathcal{R}^2$  e consideriamo

$$C_1 = B_1 \times \mathbf{R}_2 \in \mathcal{C}_1; \quad C_2 = \mathbf{R}_1 \times B_2 \in \mathcal{C}_2.$$

Tenendo conto della Proposizione II.2.7 avremo allora che

$$B_1 \times B_2 = C_1 \cap C_2 \in \mathcal{C}_1 \cap \mathcal{C}_2 = \sigma(\mathcal{C}'_1) \cap \mathcal{C}_2 = \sigma(\mathcal{C}'_1 \cap C_2).$$

D'altra parte, sempre in base alla stessa proposizione e al fatto ovvio per definizione che  $\mathcal{C}'_1 \cap \mathcal{C}'_2 \subseteq \mathcal{C}'$ , comunque scelto  $C'_1 \cap C_2 \in \mathcal{C}'_1 \cap \mathcal{C}_2$  si ha

$$C'_1 \cap C_2 \in \mathcal{C}'_1 \cap \mathcal{C}_2 = \mathcal{C}'_1 \cap \sigma(\mathcal{C}'_2) = \sigma(\mathcal{C}'_1 \cap \mathcal{C}'_2) \subseteq \sigma(\mathcal{C}') = \sigma(\mathcal{I}^2).$$

Pertanto  $\mathcal{C}'_1 \cap \mathcal{C}_2 \subseteq \sigma(\mathcal{I}^2)$  e quindi anche  $\sigma(\mathcal{C}'_1 \cap C_2) \subseteq \sigma(\mathcal{I}^2)$ , cioè  $B_1 \times B_2 \in \sigma(\mathcal{I}^2)$ .  $\square$

**II.2.15 Esempio:** Se con  $(\mathbf{R}_n)_{n \in \mathbf{N}}$  indichiamo una successione di copie dell'insieme  $\mathbf{R}$  dei numeri reali, con il simbolo  $\mathbf{R}^\infty = \mathbf{R}_1 \times \dots \times \mathbf{R}_n \times \dots$  indicheremo l'insieme delle successioni di numeri reali. Ci proponiamo ora di definire una opportuna  $\sigma$ -algebra di parti di  $\mathbf{R}^\infty$ . Per far questo cominceremo con l'indicare con

$$M = \{n_1, \dots, n_m\} = (n_l)_{l=1, \dots, m}; \quad n_l \in \mathbf{N}, \quad l = 1, \dots, m$$

gli insiemi finiti di indici estratti da  $\mathbf{N}$  e con  $\mathcal{M}$  la famiglia di tutti questi insiemi finiti di indici. Inoltre gli insiemi di  $m$ -ple di numeri reali

$$\mathbf{R}^{(M)} = \mathbf{R}_{n_1} \times \dots \times \mathbf{R}_{n_m}; \quad M \in \mathcal{M}$$

sono ovviamente casi particolari di spazi  $\mathbf{R}^m$  come quelli definiti nell'esempio II.2.13: in realtà lo spazio  $\mathbf{R}^{(M)}$  differisce dal generico  $\mathbf{R}^m$  solo per il fatto di recare traccia, in  $M$ , della collocazione (fra tutte quelle di un  $\mathbf{R}^\infty$ ) delle  $m$  rette reali usate per costruirlo. Naturalmente su ogni  $\mathbf{R}^{(M)}$  si potranno costruire le semialgebre  $\mathcal{I}^{(M)}$  ed  $\mathcal{R}^{(M)}$  e la  $\sigma$ -algebra  $\mathcal{B}(\mathbf{R}^{(M)})$ . Osserviamo inoltre che, fissato un generico  $M \in \mathcal{M}$ , ogni  $m$ -pla  $(x_n)_{n \in M} \in \mathbf{R}^{(M)}$  può sempre essere considerata come il risultato di una operazione di proiezione  $\pi_M$  eseguita sulle successioni  $x = (x_n)_{n \in \mathbf{N}} \in \mathbf{R}^\infty$ :

$$\pi_M(x) = (x_n)_{n \in M} = (x_{n_l})_{l=1, \dots, m}.$$

In pratica l'operazione  $\pi_M$  consiste nell'estrarre da  $x = (x_n)_{n \in \mathbf{N}} \in \mathbf{R}^\infty$  le  $m$  componenti relative alle collocazioni  $n_1, \dots, n_m$  di  $M$ . Chiameremo ora **cilindri con base** (di dimensione finita  $m$ )  $A^{(M)} \in \mathcal{B}(\mathbf{R}^{(M)})$ , con  $M \in \mathcal{M}$ , i seguenti sottinsiemi di  $\mathbf{R}^\infty$ :

$$C(A^{(M)}) = \{x \in \mathbf{R}^\infty : \pi_M(x) = (x_n)_{n \in M} \in A^{(M)}\} = \pi_M^{-1}(A^{(M)}).$$

Sostanzialmente il cilindro  $C(A^{(M)})$  consiste nell'insieme delle successioni  $x \in \mathbf{R}^\infty$  le cui componenti nelle collocazioni  $n_1, \dots, n_m$  costituiscono punti di  $A^{(M)} \in \mathcal{B}(\mathbf{R}^{(M)})$ . Naturalmente ogni cilindro con base in  $\mathcal{B}(\mathbf{R}^{(M)})$  è sempre suscettibile di infinite rappresentazioni equivalenti con basi in spazi più ampi:

$$C(A^{(M)}) = C(A^{(M)} \times \mathbf{R}^{(M')}); \quad M, M' \in \mathcal{M}; \quad M \cap M' = \emptyset.$$

Preso un arbitrario  $M \in \mathcal{M}$  potremo poi distinguere varie famiglie di cilindri secondo il loro tipo di base:

$$\begin{aligned} \mathcal{C}_{\mathcal{I}}^{(M)} &= \{C(A^{(M)})\}_{A^{(M)} \in \mathcal{I}^{(M)}}, \\ \mathcal{C}_{\mathcal{R}}^{(M)} &= \{C(A^{(M)})\}_{A^{(M)} \in \mathcal{R}^{(M)}}, \\ \mathcal{C}^{(M)} &= \{C(A^{(M)})\}_{A^{(M)} \in \mathcal{B}(\mathbf{R}^{(M)})}. \end{aligned}$$

Siccome evidentemente risulta  $\mathcal{I}^{(M)} \subseteq \mathcal{R}^{(M)} \subseteq \mathcal{B}(\mathbf{R}^{(M)})$ , si avrà anche che

$$\mathcal{C}_{\mathcal{I}}^{(M)} \subseteq \mathcal{C}_{\mathcal{R}}^{(M)} \subseteq \mathcal{C}^{(M)}.$$

Inoltre si verifica facilmente che  $\mathcal{C}_{\mathcal{I}}^{(M)}$  e  $\mathcal{C}_{\mathcal{R}}^{(M)}$  sono semialgebre di parti di  $\mathbf{R}^\infty$ , mentre  $\mathcal{C}^{(M)}$  è un'algebra di parti di  $\mathbf{R}^\infty$ . Infatti<sup>5</sup> se, ad esempio,  $A, B \in \mathcal{B}(\mathbf{R}^{(M)})$ , si ha

$$C(A) \cup C(B) = \pi_M^{-1}(A) \cup \pi_M^{-1}(B) = \pi_M^{-1}(A \cup B) = C(A \cup B),$$

---

<sup>5</sup> Per le proprietà generali delle applicazioni fra insiemi vedi, ad esempio, **M. Zamanski**: *Introduzione all'Algebra e all'Analisi moderna*; Feltrinelli, Milano, 1966, pp. 22-24.

e quindi anche  $C(A) \cup C(B) \in \mathcal{C}^{(M)}$  dato che  $A \cup B \in \mathcal{B}(\mathbf{R}^{(M)})$ . Tutte le altre proprietà si verificano allo stesso modo utilizzando le proprietà delle applicazioni inverse. È però anche evidente che la nostra ricerca di una adeguata  $\sigma$ -algebra di parti di  $\mathbf{R}^\infty$  non può partire con la definizione di  $\mathcal{C}^{(M)}$  dato che una famiglia di cilindri con basi di dimensione *finita e fissata* non è abbastanza ampia per i nostri scopi. Si definiscono quindi le seguenti famiglie di cilindri con *basi di dimensione finita, ma arbitraria*:

$$\mathcal{C}_{\mathcal{I}} = \bigcup_{M \in \mathcal{M}} \mathcal{C}_{\mathcal{I}}^{(M)}, \quad \mathcal{C}_{\mathcal{R}} = \bigcup_{M \in \mathcal{M}} \mathcal{C}_{\mathcal{R}}^{(M)}, \quad \mathcal{C} = \bigcup_{M \in \mathcal{M}} \mathcal{C}^{(M)},$$

per le quali, dato che comunque scelto  $M \in \mathcal{M}$  si ha  $\mathcal{C}_{\mathcal{I}}^{(M)} \subseteq \mathcal{C}_{\mathcal{R}}^{(M)} \subseteq \mathcal{C}^{(M)}$ , risulta che

$$\mathcal{C}_{\mathcal{I}} \subseteq \mathcal{C}_{\mathcal{R}} \subseteq \mathcal{C}.$$

Con metodi simili a quelli usati in precedenza si verifica poi che  $\mathcal{C}_{\mathcal{I}}$  e  $\mathcal{C}_{\mathcal{R}}$  sono semi algebre, mentre  $\mathcal{C}$  è un'algebra: ad esempio, tenendo conto delle solite proprietà delle applicazioni inverse  $\pi_{M \cup M'}^{-1}$ , presi arbitrariamente  $A^{(M)} \in \mathcal{B}(\mathbf{R}^{(M)})$  e  $B^{(M')} \in \mathcal{B}(\mathbf{R}^{(M')})$  si ha

$$\begin{aligned} C(A^{(M)}) \cup C(B^{(M')}) &= C(A^{(M)} \times \mathbf{R}^{(M' \setminus M)}) \cup C(\mathbf{R}^{(M \setminus M')} \times B^{(M')}) \\ &= C[(A^{(M)} \times \mathbf{R}^{(M' \setminus M)}) \cup (\mathbf{R}^{(M \setminus M')} \times B^{(M')})], \end{aligned}$$

e quindi  $C(A^{(M)}) \cup C(B^{(M')}) \in \mathcal{C}$  dato che  $(A^{(M)} \times \mathbf{R}^{(M' \setminus M)}) \cup (\mathbf{R}^{(M \setminus M')} \times B^{(M')})$  è sempre un elemento delle  $\sigma$ -algebra  $\mathcal{B}(\mathbf{R}^{(M \cup M')})$ . Non è invece possibile affermare che  $\mathcal{C}$  è una  $\sigma$ -algebra in quanto, data ad esempio una successione di basi  $A_n \in \mathcal{B}(\mathbf{R}^{(M_n)})$ ,  $\forall n \in \mathbf{N}$  (ciascuna con basi di dimensioni diverse e prese, per semplicità, in modo che le  $M_n$  siano tutte disgiunte), si ha che

$$\bigcup_n C(A_n) = C\left[\bigcup_n (\mathbf{R}^{(M_1)} \times \dots \times A_n \times \dots)\right]$$

dove  $\bigcup_n (\mathbf{R}^{(M_1)} \times \dots \times A_n \times \dots) \subseteq \mathbf{R}^{(\bigcup_n M_n)}$  e quindi è, in generale, una base di dimensione *non finita* ( $\bigcup_n M_n$  non è in generale un elemento di  $\mathcal{M}$ ); pertanto  $\bigcup_n C(A_n)$  non è in generale un elemento di  $\mathcal{C}$  anche se le  $C(A_n)$ , individualmente, lo sono. Siamo condotti quindi a considerare le  $\sigma$ -algebre generate da  $\mathcal{C}_{\mathcal{I}}$ ,  $\mathcal{C}_{\mathcal{R}}$  e  $\mathcal{C}$ : in una successiva proposizione si proverà che tali  $\sigma$ -algebre sono tutte eguali e coincidono anche con la  $\sigma$ -algebra generata dalla famiglia  $\mathcal{I}^\infty$  dei rettangoli di dimensione infinita i cui lati sono costituiti da intervalli:

$$\mathcal{I}^\infty = \{A \subseteq \mathbf{R}^\infty : A = I_1 \times \dots \times I_n \times \dots; I_n \in \mathcal{I}_n, n \in \mathbf{N}\}.$$

Infine, munito  $\mathbf{R}^\infty$  di una opportuna topologia (realizzata ad esempio mediante una metrica che qui, per brevità non discuteremo<sup>6</sup>), si può mostrare che tutte le

---

<sup>6</sup> Per i dettagli su questo argomento vedi, ad esempio, **A. N. Shiryaev**: *Probability*; Springer, New York, 1984, p. 145.

precedenti  $\sigma$ -algebre coincidono anche con la  $\sigma$ -algebra  $\mathcal{B}(\mathbf{R}^\infty)$  dei Boreliani di  $\mathbf{R}^\infty$  quella, cioè, generata dalla famiglia degli aperti di  $\mathbf{R}^\infty$ . Ciò posto con il simbolo

$$(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$$

indicheremo lo spazio probabilizzabile costruito, nel caso in cui  $\Omega = \mathbf{R}^\infty$ , mediante  $\sigma$ -algebre generate in uno qualunque dei modi descritti in precedenza. Sarà facile mostrare, a questo punto, che insiemi di successioni di  $\mathbf{R}^\infty$  del seguente tipo:

$$\begin{aligned} \{x \in \mathbf{R}^\infty : \sup_n x_n > a\}, & \quad \{x \in \mathbf{R}^\infty : \inf_n x_n < a\}, \\ \{x \in \mathbf{R}^\infty : \underline{\lim}_n x_n \leq a\}, & \quad \{x \in \mathbf{R}^\infty : \overline{\lim}_n x_n > a\}, \\ \{x \in \mathbf{R}^\infty : x \text{ converge}\}, & \quad \{x \in \mathbf{R}^\infty : \lim_n x_n > a\}, \end{aligned}$$

sono tutti elementi di  $\mathcal{B}(\mathbf{R}^\infty)$ . Ad esempio per il primo insieme si ha che

$$\{x \in \mathbf{R}^\infty : \sup_n x_n > a\} = \bigcup_n \{x \in \mathbf{R}^\infty : x_n > a\}$$

dove i cilindri  $\{x \in \mathbf{R}^\infty : x_n > a\}$  (con base ad una dimensione  $(a, +\infty)$  ed  $M = \{n\}$ ) sono tutti elementi di  $\mathcal{C}_{\mathcal{I}}$  e quindi sono elementi di  $\mathcal{B}(\mathbf{R}^\infty)$ ; siccome poi  $\mathcal{B}(\mathbf{R}^\infty)$  è una  $\sigma$ -algebra, l'unione numerabile di tali cilindri è ancora un elemento di  $\mathcal{B}(\mathbf{R}^\infty)$ .  $\diamond$

**II.2.16 Proposizione:**  $\sigma(\mathcal{C}_{\mathcal{I}}) = \sigma(\mathcal{C}_{\mathcal{R}}) = \sigma(\mathcal{C}) = \sigma(\mathcal{I}^\infty) = \mathcal{B}(\mathbf{R}^\infty)$ .

**Dimostrazione:** Per brevità dimostreremo solo le prime due eguaglianze: siccome, come già osservato,  $\mathcal{C}_{\mathcal{I}} \subseteq \mathcal{C}_{\mathcal{R}} \subseteq \mathcal{C}$ , potremo subito affermare che

$$\sigma(\mathcal{C}_{\mathcal{I}}) \subseteq \sigma(\mathcal{C}_{\mathcal{R}}) \subseteq \sigma(\mathcal{C}),$$

per cui per provare la tesi ci basterà dimostrare che  $\sigma(\mathcal{C}) \subseteq \sigma(\mathcal{C}_{\mathcal{I}})$ . Definita la seguente famiglia di basi (di dimensione finita) di cilindri:

$$\mathcal{G}^{(M)} = \{A^{(M)} \in \mathcal{B}(\mathbf{R}^{(M)}) : C(A^{(M)}) \in \sigma(\mathcal{C}_{\mathcal{I}})\}; \quad M \in \mathcal{M},$$

la dimostrazione procede in quattro punti:

- 1)  $\mathcal{G}^{(M)}$  è una  $\sigma$ -algebra di parti di  $\mathbf{R}^{(M)}$ : infatti, date  $A^{(M)}, B^{(M)} \in \mathcal{G}^{(M)}$ , a causa del fatto che  $C(A^{(M)}), C(B^{(M)}) \in \sigma(\mathcal{C}_{\mathcal{I}})$  e che  $\sigma(\mathcal{C}_{\mathcal{I}})$  è una  $\sigma$ -algebra, risulta che  $C(A^{(M)}) \cup C(B^{(M)}) \in \sigma(\mathcal{C}_{\mathcal{I}})$ ; d'altronde

$$\begin{aligned} C(A^{(M)}) \cup C(B^{(M)}) &= \pi_M^{-1}(A^{(M)}) \cup \pi_M^{-1}(B^{(M)}) = \pi_M^{-1}(A^{(M)} \cup B^{(M)}) \\ &= C(A^{(M)} \cup B^{(M)}), \end{aligned}$$



per cui  $C(A^{(M)} \cup B^{(M)})$  è ancora un elemento di  $\sigma(\mathcal{C}_{\mathcal{I}})$ , e ciò prova che  $A^{(M)} \cup B^{(M)} \in \mathcal{G}^{(M)}$ . Le altre proprietà delle  $\sigma$ -algebre si verificano allo stesso modo.

- 2)  $\mathcal{I}^{(M)} \subseteq \mathcal{G}^{(M)}$ ,  $\forall M \in \mathcal{M}$ : infatti, data  $A^{(M)} \in \mathcal{I}^{(M)}$ , risulta

$$C(A^{(M)}) \in \mathcal{C}_{\mathcal{I}}^{(M)} \subseteq \mathcal{C}_{\mathcal{I}} \subseteq \sigma(\mathcal{C}_{\mathcal{I}}),$$

e quindi  $A^{(M)} \in \mathcal{G}^{(M)}$ .

- 3) Si verifica subito, a questo punto, che

$$\bigcup_{M \in \mathcal{M}} \mathcal{B}(\mathbf{R}^{(M)}) \subseteq \bigcup_{M \in \mathcal{M}} \mathcal{G}^{(M)},$$

dato che, a causa dei punti 1) e 2) e della Proposizione II.2.14, comunque preso  $M \in \mathcal{M}$  risulta  $\mathcal{B}(\mathbf{R}^{(M)}) = \sigma(\mathcal{I}^{(M)}) \subseteq \mathcal{G}^{(M)}$ .

- 4) Si prova infine che  $\sigma(\mathcal{C}) \subseteq \sigma(\mathcal{C}_{\mathcal{I}})$ ; infatti, per definizione, sappiamo che  $\mathcal{C}$  è la famiglia dei cilindri con basi in  $\bigcup_{M \in \mathcal{M}} \mathcal{B}(\mathbf{R}^{(M)})$  (ossia: con basi, di dimensione finita, in qualche  $\mathcal{B}(\mathbf{R}^{(M)})$ ); detta allora  $\mathcal{C}'$  la famiglia dei cilindri con basi in  $\bigcup_{M \in \mathcal{M}} \mathcal{G}^{(M)}$ , dal punto 3) discende che  $\mathcal{C} \subseteq \mathcal{C}'$ . Ma si vede anche subito che  $\mathcal{C}' \subseteq \sigma(\mathcal{C}_{\mathcal{I}})$  in quanto, dato un cilindro  $C(A) \in \mathcal{C}'$ , la sua base  $A$  deve essere un elemento di qualche  $\mathcal{G}^{(M)}$  e quindi (per definizione di  $\mathcal{G}^{(M)}$ ) deve risultare che  $C(A) \in \sigma(\mathcal{C}_{\mathcal{I}})$ . Pertanto si ha che  $\mathcal{C} \subseteq \mathcal{C}' \subseteq \sigma(\mathcal{C}_{\mathcal{I}})$  e quindi  $\sigma(\mathcal{C}) \subseteq \sigma(\mathcal{C}_{\mathcal{I}})$ .  $\square$

**II.2.17 Esempio:** Detto  $T$  un arbitrario sottinsieme di  $\mathbf{R}$  (in generale non numerabile come, ad esempio, un intervallo  $[a, b]$ ), indicheremo con  $\mathbf{R}^T$  l'insieme delle funzioni definite da  $T$  in  $\mathbf{R}$ . Gli elementi di  $\mathbf{R}^T$  potranno essere indicati in diversi modi secondo l'opportunità della notazione:  $x$ ,  $x(\cdot)$ ,  $(x_t)_{t \in T}$ . Seguendo uno schema già utilizzato nel caso di  $\mathbf{R}^n$ , per definire una  $\sigma$ -algebra di parti di  $\mathbf{R}^T$  cominceremo con il prendere in considerazione le seguenti famiglie di sottinsiemi di  $T$ : indicheremo con  $\mathcal{S}$  la famiglia di tutti i sottinsiemi finiti  $S = \{t_1, \dots, t_s\}$  (con  $s \in \mathbf{N}$ ) di elementi di  $T$  e con  $\mathcal{S}'$  la famiglia di tutti i sottinsiemi finiti o numerabili  $S = \{t_1, \dots, t_k, \dots\}$  di elementi di  $T$ . Ovviamente risulterà  $\mathcal{S} \subseteq \mathcal{S}'$ . Se allora consideriamo un  $S \in \mathcal{S}'$ , detto  $\mathbf{R}^{(S)} = \mathbf{R}_{t_1} \times \dots \times \mathbf{R}_{t_k} \times \dots$  (che risulta essere un  $\mathbf{R}^s$ , se  $S$  è un insieme finito di  $s \in \mathbf{N}$  elementi, oppure un  $\mathbf{R}^\infty$  se  $S$  è numerabile), potremo sempre definire una  $\sigma$ -algebra  $\mathcal{B}(\mathbf{R}^{(S)})$  di parti di  $\mathbf{R}^{(S)}$  seguendo la costruzione esposta negli Esempi II.2.13 e II.2.15. Limitandoci per il momento al caso di  $S \in \mathcal{S}$  finiti e di cardinalità  $s$ , definiremo sul corrispondente  $\mathbf{R}^{(S)}$  le solite semialgebre dei rettangoli  $\mathcal{I}^{(S)}$  e  $\mathcal{R}^{(S)}$  e la  $\sigma$ -algebra  $\mathcal{B}(\mathbf{R}^{(S)})$  dei boreliani di  $\mathbf{R}^{(S)}$  generata, ad esempio, da una delle semialgebre di rettangoli. Siccome ogni elemento di  $\mathbf{R}^{(S)}$  è una  $s$ -pla di numeri reali  $(x_{t_k})_{k=1, \dots, s}$  che si può considerare come estratta da  $(x_t)_{t \in T} \in \mathbf{R}^T$  mediante la proiezione

$$\pi_S(x) = \pi_S[(x_t)_{t \in T}] = (x_{t_k})_{k=1, \dots, s} = (x_t)_{t \in S},$$

potremo definire in  $\mathbf{R}^T$  i **cilindri con base** (di dimensione finita  $s$ )  $A^{(S)} \in \mathcal{B}(\mathbf{R}^{(S)})$  ( $S \in \mathcal{S}$ ) come gli insiemi

$$C(A^{(S)}) = \pi_S^{-1}(A^{(S)}) = \{x \in \mathbf{R}^T : \pi_S(x) = (x_t)_{t \in S} \in A^{(S)}\},$$

ossia come gli insiemi delle funzioni  $x \in \mathbf{R}^T$  che, in un dato insieme finito di punti  $S \subseteq T$ , assumono valori tali che la  $s$ -pla di numeri  $(x_t)_{t \in S}$  risulta essere un punto di  $A^{(S)}$ . In pratica, se ad esempio avessimo scelto  $A^{(S)} = I_{t_1} \times \dots \times I_{t_s} \in \mathcal{I}^{(S)}$ ,  $C(A^{(S)})$  consisterebbe nell'insieme delle funzioni di  $\mathbf{R}^T$  che nei punti  $t_1, \dots, t_s$  assumono valori appartenenti agli intervalli  $I_{t_1}, \dots, I_{t_s}$ : un cilindro si rivela dunque come un insieme di funzioni che soddisfa delle particolari *restrizioni* in un dato numero di punti del suo insieme di definizione. Naturalmente, come nel caso dei cilindri in  $\mathbf{R}^\infty$ , anche i cilindri di  $\mathbf{R}^T$  sono suscettibili di un numero infinito di rappresentazioni equivalenti:

$$C(A^{(S)}) = C(A^{(S)} \times \mathbf{R}^{(S')}); \quad S, S' \in \mathcal{S}; \quad S \cap S' = \emptyset.$$

Distingueremo poi le varie famiglie di cilindri (con base di dimensione finita) secondo il tipo di base:

$$\begin{aligned} \mathcal{C}_{\mathcal{I}}^{(S)} &= \{C(A^{(S)})\}_{A^{(S)} \in \mathcal{I}^{(S)}}, \\ \mathcal{C}_{\mathcal{R}}^{(S)} &= \{C(A^{(S)})\}_{A^{(S)} \in \mathcal{R}^{(S)}}, \\ \mathcal{C}^{(S)} &= \{C(A^{(S)})\}_{A^{(S)} \in \mathcal{B}(\mathbf{R}^{(S)})}, \end{aligned}$$

ed esattamente come nel caso di  $\mathbf{R}^\infty$  si verifica che  $\mathcal{C}_{\mathcal{I}}^{(S)}$  e  $\mathcal{C}_{\mathcal{R}}^{(S)}$  sono semialgebre, che  $\mathcal{C}^{(S)}$  è un'algebra e che

$$\mathcal{C}_{\mathcal{I}}^{(S)} \subseteq \mathcal{C}_{\mathcal{R}}^{(S)} \subseteq \mathcal{C}^{(S)}; \quad S \in \mathcal{S}.$$

Definite quindi le famiglie di cilindri con basi di dimensione finita ma arbitraria

$$\mathcal{C}_{\mathcal{I}} = \bigcup_{S \in \mathcal{S}} \mathcal{C}_{\mathcal{I}}^{(S)}, \quad \mathcal{C}_{\mathcal{R}} = \bigcup_{S \in \mathcal{S}} \mathcal{C}_{\mathcal{R}}^{(S)}, \quad \mathcal{C} = \bigcup_{S \in \mathcal{S}} \mathcal{C}^{(S)},$$

si verifica che  $\mathcal{C}_{\mathcal{I}}$  e  $\mathcal{C}_{\mathcal{R}}$  sono semialgebre, che  $\mathcal{C}$  è un'algebra (ma non una  $\sigma$ -algebra) e che

$$\mathcal{C}_{\mathcal{I}} \subseteq \mathcal{C}_{\mathcal{R}} \subseteq \mathcal{C}.$$

Nella successiva proposizione proveremo che tali famiglie di cilindri generano tutte la medesima  $\sigma$ -algebra di parti di  $\mathbf{R}^T$  che indicheremo anche con il simbolo  $\mathcal{B}(\mathbf{R}^T)$ . Inoltre verrà dimostrato che tali  $\sigma$ -algebre coincidono anche con quella generata dalla seguente famiglia di cilindri con basi di *dimensione finita o numerabile* in un qualche  $\mathcal{B}(\mathbf{R}^{(S)})$  (si tratta dunque di una generalizzazione di  $\mathcal{C}$ ):

$$\begin{aligned} \mathcal{C}' &= \{C \in \wp(\mathbf{R}^T) : C = \pi_S^{-1}(B^{(S)}), B^{(S)} \in \mathcal{B}(\mathbf{R}^{(S)}), S \in \mathcal{S}'\} \\ &= \{C(B^{(S)})\}_{B^{(S)} \in \bigcup_{S \in \mathcal{S}'} \mathcal{B}(\mathbf{R}^{(S)})}, \end{aligned}$$

che risulta essere essa stessa una  $\sigma$ -algebra. Ciò posto, da ora in poi con il simbolo

$$(\mathbf{R}^T, \mathcal{B}(\mathbf{R}^T))$$

indicheremo lo spazio probabilizzabile costruito, nel caso in cui  $\Omega = \mathbf{R}^T$ , mediante  $\sigma$ -algre generate in uno qualunque dei modi descritti in precedenza. Il fatto che  $\mathcal{B}(\mathbf{R}^T)$  coincida con  $\mathcal{C}'$  sottolinea, peraltro, una caratteristica importante di tale  $\sigma$ -algebra:  $\mathcal{B}(\mathbf{R}^T)$  si rivela infatti come la famiglia di parti di  $\mathbf{R}^T$  che possono essere caratterizzate mediante restrizioni imposte ai valori di  $(x_t)_{t \in T}$  in un insieme di punti *finito o numerabile*. Ne segue, in particolare, che alcuni sottinsiemi di  $\mathbf{R}^T$ , caratterizzati dal comportamento della funzione  $(x_t)_{t \in T}$  in un insieme non numerabile di punti, non possono essere considerati come elementi di  $\mathcal{B}(\mathbf{R}^T)$ : ad esempio, con  $T = [0, 1]$ , si verifica che

$$\begin{aligned} A_1 &= \{x \in \mathbf{R}^{[0,1]} : \sup_{t \in [0,1]} x_t < a, a \in \mathbf{R}\} \\ A_2 &= \{x \in \mathbf{R}^{[0,1]} : \exists t \in [0, 1] \ni x_t = 0\} \\ A_3 &= \{x \in \mathbf{R}^{[0,1]} : x_t \text{ è continua in } t_0 \in [0, 1]\} \end{aligned}$$

non sono elementi di  $\mathcal{B}(\mathbf{R}^{[0,1]})$ . Siccome però tali insiemi rappresentano affermazioni provviste di significato e che comunque capita di fare quando si discute di funzioni aleatorie e processi stocastici, vale la pena di osservare che è possibile superare questa difficoltà limitandosi a classi di funzioni più ristrette di  $\mathbf{R}^T$ : ad esempio è possibile considerare l'insieme delle *funzioni continue* munito di un'opportuna metrica, e definire poi su questo la  $\sigma$ -algebra di Borel generata dagli aperti. In tal caso si può mostrare che sottinsiemi di funzioni come  $A_1, A_2$  ed  $A_3$  risulterebbero essere elementi di tale  $\sigma$ -algebra<sup>7</sup>.  $\diamond$

**II.2.18 Proposizione:**  $\mathcal{C}'$  è una  $\sigma$ -algebra e inoltre la relazione

$$\sigma(\mathcal{C}_{\mathcal{I}}) = \sigma(\mathcal{C}_{\mathcal{R}}) = \sigma(\mathcal{C}) = \mathcal{C}' = \mathcal{B}(\mathbf{R}^T)$$

è sempre verificata.

**Dimostrazione:** Proviamo innanzitutto che  $\mathcal{C}'$  è una  $\sigma$ -algebra: che si tratti di un'algebra si verifica immediatamente se si considerano due elementi  $C_1, C_2 \in \mathcal{C}'$  riconducendoli ad avere base in una stessa  $\sigma$ -algebra  $\mathcal{B}(\mathbf{R}^{(S_1 \cup S_2)})$ :

$$\begin{aligned} C_1 &= \pi_{S_1}^{-1}(B_1^{(S_1)}) = \pi_{S_1 \cup S_2}^{-1}(B_1^{(S_1)} \times \mathbf{R}^{(S_2 \setminus S_1)}); \\ C_2 &= \pi_{S_2}^{-1}(B_2^{(S_2)}) = \pi_{S_1 \cup S_2}^{-1}(\mathbf{R}^{(S_1 \setminus S_2)} \times B_2^{(S_2)}); \end{aligned}$$

<sup>7</sup> Per i dettagli di tale costruzione vedi ad esempio: **A. N. Shiriyayev:** *Probability*; Springer, New York, 1984, p. 148.

basterà allora utilizzare le solite proprietà delle applicazioni inverse e ricordarsi del fatto che  $\mathcal{B}(\mathbf{R}^{(S_1 \cup S_2)})$  è una  $\sigma$ -algebra per provare che unioni, intersezioni e differenze di  $C_1$  e  $C_2$  sono ancora elementi di  $\mathcal{C}'$ . Per provare che  $\mathcal{C}'$  è anche una  $\sigma$ -algebra dovremo poi considerare una successione  $(C_n)_{n \in \mathbf{N}}$  di elementi di  $\mathcal{C}'$  con  $C_n = \pi_{S_n}^{-1}(B^{(S_n)})$ ,  $\forall n \in \mathbf{N}$  dove  $S_n \in \mathcal{S}'$ . Posto allora  $S = \bigcup_n S_n$ , siccome  $S$ , in quanto unione numerabile di insiemi numerabili è ancora un insieme numerabile<sup>8</sup>, avremo che

$$C_n = \pi_S^{-1}(B^{(S_n)} \times \mathbf{R}^{(S \setminus S_n)}), \quad B^{(S_n)} \times \mathbf{R}^{(S \setminus S_n)} \in \mathcal{B}(\mathbf{R}^{(S)}),$$

con  $S \in \mathcal{S}'$  ed  $n \in \mathbf{N}$  arbitrario. Pertanto potremo sempre rappresentare i cilindri  $C_n$  in modo tale che abbiano tutti base in una stessa  $\sigma$ -algebra  $\mathcal{B}(\mathbf{R}^{(S)})$  con  $S \in \mathcal{S}'$  e quindi, dalle proprietà delle applicazioni inverse e dal fatto che  $\mathcal{B}(\mathbf{R}^{(S)})$  è una  $\sigma$ -algebra, potremo verificare facilmente che  $\bigcup_n C_n$  è ancora un elemento di  $\mathcal{C}'$ . Passeremo ora a provare che  $\sigma(\mathcal{C}_{\mathcal{I}}) = \sigma(\mathcal{C}_{\mathcal{R}}) = \sigma(\mathcal{C}) = \mathcal{C}'$ : dato che  $\mathcal{C}_{\mathcal{I}} \subseteq \mathcal{C}_{\mathcal{R}} \subseteq \mathcal{C}$ , si ha immediatamente che  $\sigma(\mathcal{C}_{\mathcal{I}}) \subseteq \sigma(\mathcal{C}_{\mathcal{R}}) \subseteq \sigma(\mathcal{C})$ . Inoltre si prova facilmente che  $\sigma(\mathcal{C}) \subseteq \mathcal{C}'$  in quanto  $\mathcal{C} \subseteq \mathcal{C}'$  (se  $C \in \mathcal{C}$  è un cilindro con base in qualche  $\mathcal{B}(\mathbf{R}^{(S)})$ , con  $S \in \mathcal{S} \subseteq \mathcal{S}'$ , dalla definizione di  $\mathcal{C}'$  risulta anche  $C \in \mathcal{C}'$ ) e  $\mathcal{C}'$  è una  $\sigma$ -algebra. Ciò premesso, per dimostrare la tesi della nostra proposizione sarà sufficiente provare che  $\mathcal{C}' \subseteq \sigma(\mathcal{C}_{\mathcal{I}})$ . Per far questo indichiamo con

$$\mathcal{G}^{(S)} = \{B^{(S)} \in \mathcal{B}(\mathbf{R}^{(S)}) : C(B^{(S)}) \in \sigma(\mathcal{C}_{\mathcal{I}})\}; \quad S \in \mathcal{S}',$$

la famiglia delle basi (di dimensione finita o numerabile) dei cilindri che risultano essere elementi di  $\sigma(\mathcal{C}_{\mathcal{I}})$ . La dimostrazione procede poi in quattro punti:

- 1) Seguendo uno schema analogo a quello del punto 1) della dimostrazione della Proposizione II.2.16 si prova innanzitutto che  $\mathcal{G}^{(S)}$  è una  $\sigma$ -algebra.
- 2) Si prova poi che

$$\mathcal{I}^{(S)} \subseteq \mathcal{G}^{(S)}, \quad S \in \mathcal{S}'.$$

La dimostrazione è molto semplice nel caso in cui  $S$  è un insieme finito (cioè se  $S \in \mathcal{S}$ ). Infatti in tal caso, se  $A^{(S)} \in \mathcal{I}^{(S)}$ , per definizione di  $\mathcal{C}_{\mathcal{I}}^{(S)}$  risulta  $C(A^{(S)}) \in \mathcal{C}_{\mathcal{I}}^{(S)} \subseteq \mathcal{C}_{\mathcal{I}} \subseteq \sigma(\mathcal{C}_{\mathcal{I}})$  e quindi  $A^{(S)} \in \mathcal{G}^{(S)}$ . Se invece  $S \in \mathcal{S}'$  è un insieme numerabile (per cui  $\mathcal{I}^{(S)}$  è un  $\mathcal{I}^\infty$ ), la dimostrazione è un po' più delicata. La difficoltà sorge dal fatto che, in questo caso,  $C(A^{(S)})$  (con  $A^{(S)} \in \mathcal{I}^{(S)}$ ) non è più un cilindro con base di dimensione finita e quindi non risulta essere un elemento di  $\mathcal{C}_{\mathcal{I}}$ . Si può però mostrare egualmente che  $C(A^{(S)}) \in \sigma(\mathcal{C}_{\mathcal{I}})$ , il che è comunque sufficiente per provare che  $A^{(S)} \in \mathcal{G}^{(S)}$  e quindi che  $\mathcal{I}^{(S)} \subseteq \mathcal{G}^{(S)}$  come visto nella prima parte di questo punto. Per provare che  $C(A^{(S)}) \in \sigma(\mathcal{C}_{\mathcal{I}})$ , anche se  $S \in \mathcal{S}'$  è numerabile, cominceremo con l'osservare che è sempre possibile determinare una successione  $(S_n)_{n \in \mathbf{N}}$  di elementi di  $\mathcal{S}$  (cioè finiti) tali che  $S_n \cap S_m = \emptyset$  se  $m \neq n$  e che  $\bigcup_n S_n = S$ . Infatti, se  $S$

---

<sup>8</sup> Vedi ad esempio **J. Dieudonné**: *Foundations of modern analysis*; Academic Press, New York, 1960, p. 13.

è numerabile esso sarà del tipo  $S = (t_n)_{n \in \mathbf{N}}$ : basterà quindi considerare ad esempio  $S_n = \{t_n\}$ ,  $\forall n \in \mathbf{N}$  per ottenere la successione cercata. Ovviamente, dato che  $S_n \in \mathcal{S}$ , sarà anche (in conseguenza della discussione svolta nella prima parte di questo punto)  $\mathcal{I}^{(S_n)} \subseteq \mathcal{G}^{(S_n)}$  con  $n \in \mathbf{N}$ . Se allora consideriamo  $A^{(S)} \in \mathcal{I}^{(S)}$ ,  $S \in \mathcal{S}'$ , ricordando che si tratta di rettangoli otterremo

$$A^{(S)} = A^{(S_1)} \times \dots \times A^{(S_n)} \times \dots = \bigcap_n [A^{(S_n)} \times \mathbf{R}^{(S \setminus S_n)}].$$

Dalle proprietà delle applicazioni inverse avremo allora che

$$\begin{aligned} C(A^{(S)}) &= \pi_S^{-1} \left( \bigcap_n [A^{(S_n)} \times \mathbf{R}^{(S \setminus S_n)}] \right) \\ &= \bigcap_n \pi_S^{-1} [A^{(S_n)} \times \mathbf{R}^{(S \setminus S_n)}] = \bigcap_n C(A^{(S_n)}). \end{aligned}$$

Siccome però  $A^{(S_n)} \in \mathcal{I}^{(S_n)} \subseteq \mathcal{G}^{(S_n)}$ ,  $\forall n \in \mathbf{N}$  si ha (per definizione di  $\mathcal{G}^{(S_n)}$ ) che  $C(A^{(S_n)}) \in \sigma(\mathcal{C}_{\mathcal{I}})$  e quindi  $C(A^{(S)})$ , in quanto intersezione numerabile di elementi della  $\sigma$ -algebra  $\sigma(\mathcal{C}_{\mathcal{I}})$ , è anche elemento di  $\sigma(\mathcal{C}_{\mathcal{I}})$ .

3) Posto ora

$$\mathcal{G} = \bigcup_{S \in \mathcal{S}'} \mathcal{G}^{(S)}$$

possiamo facilmente provare che

$$\bigcup_{S \in \mathcal{S}'} \mathcal{B}(\mathbf{R}^{(S)}) \subseteq \mathcal{G}.$$

Infatti dai punti 1) e 2) si ha subito che  $\sigma(\mathcal{I}^{(S)}) \subseteq \mathcal{G}^{(S)}$  per  $S \in \mathcal{S}'$  arbitrario; ma sia nel caso in cui  $S$  è finito (vedi Proposizione II.2.14:  $\sigma(\mathcal{I}^n) = \mathcal{B}(\mathbf{R}^n)$ ) sia in quello in cui è numerabile (vedi Proposizione II.2.16:  $\sigma(\mathcal{I}^\infty) = \mathcal{B}(\mathbf{R}^\infty)$ ) risulta  $\mathcal{B}(\mathbf{R}^{(S)}) = \sigma(\mathcal{I}^{(S)})$  e pertanto, comunque scelto  $S \in \mathcal{S}'$ , risulterà  $\mathcal{B}(\mathbf{R}^{(S)}) \subseteq \mathcal{G}^{(S)}$ , da cui segue il risultato richiesto.

4) Per provare finalmente che

$$\mathcal{C}' \subseteq \sigma(\mathcal{C}_{\mathcal{I}})$$

(il che, come sappiamo, prova anche la tesi della nostra Proposizione) indicheremo con

$$\mathcal{E} = \{C(A)\}_{A \in \mathcal{G}}$$

la famiglia dei cilindri le cui basi sono elementi di  $\mathcal{G}$  (cioè sono elementi di qualche  $\mathcal{G}^{(S)}$  con  $S$  finito o numerabile): l'inclusione richiesta sarà provata mediante la seguente catena di inclusioni:

$$\mathcal{C}' \subseteq \mathcal{E} \subseteq \sigma(\mathcal{C}_{\mathcal{I}}).$$

Infatti, siccome la definizione di  $\mathcal{C}'$  può essere data equivalentemente nella forma

$$\mathcal{C}' = \{C(B)\}_{B \in \bigcup_{S \in \mathcal{S}'} \mathcal{B}(\mathbf{R}^{(S)})},$$

dal punto 3) discende subito che  $\mathcal{C}' \subseteq \mathcal{E}$ . Inoltre, ogni cilindro  $C(A) \in \mathcal{E}$  ha per definizione la base  $A$  in  $\mathcal{G}$ , cioè in qualche  $\mathcal{G}^{(S)}$  con  $S \in \mathcal{S}'$ : ma questo, per definizione di  $\mathcal{G}^{(S)}$ , vuol dire che  $C(A) \in \sigma(\mathcal{C}_{\mathcal{I}})$  e quindi  $\mathcal{E} \subseteq \sigma(\mathcal{C}_{\mathcal{I}})$ .  $\square$

**II.2.19 Esempio:** Con il simbolo

$$\left( \prod_{t \in T} \Omega_t, \bigotimes_{t \in T} \mathcal{F}_t \right)$$

indicheremo lo spazio probabilizzabile ottenuto come **prodotto diretto** di altri spazi probabilizzabili. Più precisamente, se  $(\Omega_t, \mathcal{F}_t)$  con  $t \in T$  è una famiglia (in generale non numerabile) di spazi probabilizzabili in generale diversi fra loro, l'insieme  $\prod_{t \in T} \Omega_t$  sarà costituito dall'insieme delle funzioni  $(\omega_t)_{t \in T}$  con  $\omega_t \in \Omega_t$ . Seguendo poi la traccia dei casi discussi precedentemente definiremo prima la famiglia  $\mathcal{C}_{\mathcal{R}}$  dei cilindri le cui basi  $B = B^{(t_1)} \times \dots \times B^{(t_n)}$  sono rettangoli di dimensione finita i cui lati  $B^{(t_k)}$  si trovano nelle  $\sigma$ -algebre  $\mathcal{F}_{t_k}$ , e poi la  $\sigma$ -algebra prodotto diretto come

$$\bigotimes_{t \in T} \mathcal{F}_t = \sigma(\mathcal{C}_{\mathcal{R}}).$$

Naturalmente  $(\mathbf{R}^T, \mathcal{B}(\mathbf{R}^T))$ ,  $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$  e  $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$  sono tutti esempi di spazi prodotto diretto. In particolare  $(\mathbf{R}^T, \mathcal{B}(\mathbf{R}^T))$  corrisponde al caso in cui  $T$  è un qualunque sottinsieme (anche non numerabile) di  $\mathbf{R}$ ,  $\Omega_t = \mathbf{R}$  e  $\mathcal{F}_t = \mathcal{B}(\mathbf{R})$  qualunque sia  $t \in T$ ;  $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$  corrisponde al caso in cui  $T = \mathbf{N}$ ,  $\Omega_n = \mathbf{R}$  e  $\mathcal{F}_n = \mathcal{B}(\mathbf{R})$  qualunque sia  $n \in \mathbf{N}$ ; e infine  $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$  al caso in cui  $T$  è un insieme finito  $\{1, \dots, n\}$ ,  $\Omega_k = \mathbf{R}$  e  $\mathcal{F}_k = \mathcal{B}(\mathbf{R})$  qualunque sia  $k = 1, \dots, n$ .  $\diamond$

## II.3 Misure di probabilità (I Parte)

Questo capitolo ed il successivo saranno dedicati ad un'analisi dettagliata dei modi in cui è possibile definire delle misure di probabilità sui vari spazi probabilizzabili esaminati nel capitolo precedente iniziando, ovviamente, con il caso di  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ .

**II.3.1 Osservazione:** Supponiamo che  $\mathbf{P}(\cdot)$  sia una misura di probabilità definita su  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ , consideriamo i boreliani di  $\mathbf{R}$  della forma  $(-\infty, x] \in \mathcal{B}(\mathbf{R})$  e definiamo la funzione

$$F(x) = \mathbf{P}(-\infty, x], \quad \forall x \in \mathbf{R}.$$

Si verifica facilmente che  $F(x)$  soddisfa le seguenti tre proprietà:

- $D_1)$   $F(x)$  è non decrescente;
- $D_2)$   $\lim_{x \rightarrow +\infty} F(x) = 1$ ,  $\lim_{x \rightarrow -\infty} F(x) = 0$ ;
- $D_3)$   $F(x)$  è continua da destra e ammette limite da sinistra  $\forall x \in \mathbf{R}$  (*cadlag*).

Infatti se  $x' \leq x$  si ha  $(-\infty, x'] \subseteq (-\infty, x]$  e quindi, in base al punto V) della Proposizione II.1.12,

$$F(x') = \mathbf{P}(-\infty, x'] \leq \mathbf{P}(-\infty, x] = F(x),$$

che prova la proprietà  $D_1$ . Per provare  $D_2$  basterà considerare una successione numerica  $(x_n)_{n \in \mathbf{N}}$  con  $x_n \xrightarrow{n} -\infty$ ; in tal caso  $(-\infty, x_n] \downarrow \emptyset$  e quindi, in base al punto 4 del Teorema II.1.13,  $F(x_n) = \mathbf{P}(-\infty, x_n] \xrightarrow{n} 0$ . Analogamente se  $x_n \xrightarrow{n} +\infty$  risulterà  $(x_n, +\infty] \downarrow \emptyset$  e quindi (sempre per il Teorema II.1.13)  $F(x_n) = 1 - \mathbf{P}(x_n, +\infty] \xrightarrow{n} 1$ . Per provare  $D_3$ , infine, osserviamo innanzitutto che, essendo  $F(x)$  monotona e limitata fra 0 ed 1, essa ammette sempre limiti  $F(x^+)$  ed  $F(x^-)$  da destra e da sinistra in ogni  $x \in \mathbf{R}$ ; se allora  $(x_n)_{n \in \mathbf{N}}$  è una successione numerica con  $x_n \xrightarrow{n} x$  da destra, risulterà  $(-\infty, x_n] \downarrow (-\infty, x]$  e quindi dal punto 3 Teorema II.1.13 avremo

$$F(x^+) = \lim_n F(x_n) = \lim_n \mathbf{P}(-\infty, x_n] = \mathbf{P}(-\infty, x] = F(x),$$

cioè  $F(x)$  è continua da destra. Non si può invece dire, in generale, che  $F(x)$  sia anche continua da sinistra: se infatti  $x_n \xrightarrow{n} x$  da sinistra risulta che  $(-\infty, x_n] \uparrow (-\infty, x)$  e quindi

$$F(x^-) = \lim_n F(x_n) = \lim_n \mathbf{P}(-\infty, x_n] = \mathbf{P}(-\infty, x) \neq F(x),$$

dato che, essendo  $(-\infty, x] = (-\infty, x) \cup \{x\}$ , avremo in generale

$$F(x) = \mathbf{P}(-\infty, x] = \mathbf{P}(-\infty, x) + \mathbf{P}\{x\}.$$

Più precisamente potremo quindi affermare che

$$F(x^-) = F(x) - \mathbf{P}\{x\},$$

sicché  $F(x)$  risulterà continua anche da sinistra (e quindi, in definitiva, continua *tout court*) se e solo se  $\mathbf{P}\{x\} = 0$ . Potremo quindi concludere che se  $F(x)$  è discontinua in un punto  $x$  risulta

$$\mathbf{P}\{x\} = F(x) - F(x^-) = F(x^+) - F(x^-),$$

cioè che  $\mathbf{P}\{x\}$  rappresenta proprio il valore del salto di  $F(x)$  nel punto di discontinuità, e che  $F(x)$  è continua in  $x$  se e solo se è nulla la probabilità del boreliano  $\{x\} \in \mathcal{B}(\mathbf{R})$ . ○

**II.3.2 Definizione:** Si dice **Funzione di Distribuzione** (FdD), o anche **Funzione di Distribuzione Cumulativa**, su  $\mathbf{R}$  ogni funzione  $F(x)$  che soddisfa le proprietà  $D_1, D_2$  e  $D_3$ . △

**II.3.3 Osservazione:** La discussione svolta in II.3.1 mostra che ad ogni misura di probabilità  $\mathbf{P}(\cdot)$  su  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$  è associata una FdD  $F(x)$ . Nel seguito mostreremo che è vera anche l'affermazione opposta: ad ogni FdD definita su  $\mathbf{R}$  è possibile associare un'opportuna misura di probabilità  $\mathbf{P}(\cdot)$  su  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ . In questo modo viene stabilita una corrispondenza biunivoca fra le  $\mathbf{P}(\cdot)$  su  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$  e le FdD  $F(x)$  su  $\mathbf{R}$  sicché una probabilità su  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$  può considerarsi assegnata se e solo se viene data un'opportuna FdD  $F(x)$ . La prova di questa affermazione richiede l'uso di alcuni importanti risultati della Teoria della Misura e dell'Analisi che qui, però, ci limiteremo solo ad enunciare. ○

**II.3.4 Teorema (di Carathéodory):** Assegnata una misura ( $\sigma$ -additiva)  $\sigma$ -finita  $\mu_0$  su  $(\Omega, \mathcal{A})$ , dove  $\mathcal{A}$  è un'algebra di parti di  $\Omega$ , esiste sempre un'unica misura  $\mu$  su  $(\Omega, \sigma(\mathcal{A}))$  tale che

$$\mu(A) = \mu_0(A), \quad \forall A \in \mathcal{A},$$

tale, cioè, che sia un'estensione di  $\mu_0$  a tutto  $(\Omega, \sigma(\mathcal{A}))$ .

**Dimostrazione:** Omessa<sup>1</sup> □

<sup>1</sup> Vedi ad esempio M. Métivier: *Notions Fondamentales de la Théorie des Probabilités*; Dunod, Paris, 1972, p. 273. Constantin Carathéodory (1873 - 1950), figlio di un diplomatico di origine greca al servizio dell'Impero Ottomano, nacque a Berlino, ma trascorse la prima parte della sua vita a Bruxelles dove terminò gli studi all'École Militaire de Belgique nel 1895. Dopo aver collaborato alla progettazione di strade a Samo ed alla costruzione della diga di Asyut in Egitto, si dedicò agli studi matematici a Berlino (1902) e Göttingen (1904) laureandosi sotto la



**II.3.5 Teorema:** Data una FdD su  $\mathbf{R}$   $F(x)$ , esiste un'unica probabilità  $\mathbf{P}$  su  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$  tale che

$$\mathbf{P}(a, b] = F(b) - F(a)$$

comunque scelti  $a$  e  $b$  con  $-\infty \leq a < b \leq +\infty$ .

**Dimostrazione:** Abbiamo visto, nella discussione dell'Esempio II.2.12 che  $\mathcal{B}(\mathbf{R})$  può essere generata dall'algebra  $\mathcal{A}$  costituita dalle unioni finite di intervalli (disgiunti) chiusi a destra:  $\mathcal{B}(\mathbf{R}) = \sigma(\mathcal{A})$ . Inoltre su  $(\mathbf{R}, \mathcal{A})$  possiamo definire, mediante l'assegnata FdD  $F(x)$ , l'applicazione

$$\mathbf{P}_0 : \mathcal{A} \rightarrow [0, 1]$$

che ad ogni  $A \in \mathcal{A}$ , unione di intervalli disgiunti  $(a_1, b_1], \dots, (a_n, b_n]$ , associa il valore

$$\mathbf{P}_0(A) = \sum_{k=1}^n [F(b_k) - F(a_k)].$$

Si verifica facilmente che tale  $\mathbf{P}_0$  è additiva e  $\sigma$ -finita. Siamo dunque nelle condizioni di poter utilizzare il Teorema di Carathéodory II.3.4: se infatti riusciremo a provare che  $\mathbf{P}_0$  è anche  $\sigma$ -additiva il Teorema II.3.4 ci garantirà l'esistenza di un'unica probabilità  $\mathbf{P}$  su  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$  tale che, essendo questa estensione di  $\mathbf{P}_0$ , risulti anche

$$\mathbf{P}(a, b] = F(b) - F(a), \quad -\infty \leq a < b \leq +\infty.$$

La dimostrazione si ridurrà quindi alla prova del fatto che la  $\mathbf{P}_0$  da noi definita è  $\sigma$ -additiva: per far questo ricorremo al Teorema II.1.13 e ci limiteremo a provare che, data una successione  $(A_n)_{n \in \mathbf{N}}$  di elementi di  $\mathcal{A}$  decrescente e tale che  $A_n \downarrow \emptyset$ , risulta anche  $\mathbf{P}_0(A_n) \xrightarrow{n} 0$ . Assegnata dunque la successione  $(A_n)_{n \in \mathbf{N}}$  e dato un numero reale arbitrario  $\epsilon > 0$  la prova viene data dimostrando le seguenti sei successive affermazioni:

$$(1) \quad \exists M \in \mathbf{R}^+ \quad \exists' \quad \mathbf{P}_0(-M, M] > 1 - \frac{\epsilon}{2};$$

---

guida di H. Minkowski. Insegnò dapprima in varie università tedesche e, alla fine della Prima Guerra Mondiale, accettò una cattedra a Smirne (Izmir) offertagli dal governo Greco. Quando però nel 1922 i turchi Kemalisti riconquistarono la città, Carathéodory fu evacuato, in condizioni drammatiche, assieme a tutta la popolazione greca riuscendo a mettere in salvo la Biblioteca della sua Università. Dopo un ulteriore periodo di insegnamento ad Atene, si stabilì definitivamente a Monaco di Baviera. Cosmopolita e poliglotta per origini e per educazione, fu legato da stretta e personale amicizia con i più celebri matematici del suo tempo, tra i quali D. Hilbert e H. Schwarz. Diede importanti contributi al calcolo delle variazioni, alla teoria delle funzioni, all'assiomatizzazione del concetto di integrale ed all'impostazione assiomatica della termodinamica fenomenologica.

infatti, in base alle proprietà  $D_1$  e  $D_2$  delle FdD, sarà sempre possibile determinare un numero  $M > 0$  tale che risulti contemporaneamente  $1 - F(M) = |F(M) - 1| < \frac{\epsilon}{4}$  ed  $F(-M) < \frac{\epsilon}{4}$  e pertanto

$$\mathbf{P}_0(-M, M] = F(M) - F(-M) > 1 - \frac{\epsilon}{4} - \frac{\epsilon}{4} = 1 - \frac{\epsilon}{2}$$

come richiesto in (1).

$$(2) \quad \forall n \in \mathbf{N}, \quad \mathbf{P}_0(A_n) < \frac{\epsilon}{2} + \mathbf{P}_0(A'_n), \quad A'_n = A_n \cap (-M, M];$$

infatti, siccome  $A_n = (A_n \cap (-M, M]) \cup (A_n \cap \overline{(-M, M]})$  con  $(A_n \cap (-M, M]) \cap (A_n \cap \overline{(-M, M]}) = \emptyset$ , dall'additività di  $\mathbf{P}_0$  si ha che

$$\begin{aligned} \mathbf{P}_0(A_n) &= \mathbf{P}_0(A_n \cap (-M, M]) + \mathbf{P}_0(A_n \cap \overline{(-M, M]}) \\ &\leq \mathbf{P}_0(A'_n) + \mathbf{P}_0(\overline{(-M, M]}) = \mathbf{P}_0(A'_n) + 1 - \mathbf{P}_0(-M, M] < \frac{\epsilon}{2} + \mathbf{P}_0(A'_n), \end{aligned}$$

dove abbiamo tenuto anche conto del risultato del punto (1).

$$(3) \quad \forall n \in \mathbf{N} \exists B_n \in \mathcal{A} \ni [B_n] \subseteq A'_n \text{ e } \mathbf{P}_0(A'_n \setminus B_n) = \mathbf{P}_0(A'_n) - \mathbf{P}_0(B_n) < \frac{\epsilon}{2^{n+1}};$$

ricordiamo innanzitutto che con il simbolo  $[B_n]$  indichiamo la chiusura dell'insieme  $B_n$  di  $\mathcal{A}$ ; inoltre, siccome  $A'_n \in \mathcal{A}$ , in quanto intersezione di elementi di  $\mathcal{A}$ , esso risulterà essere unione finita di intervalli chiusi a destra del tipo  $(a, b]$ . Se allora  $A' = (a, b]$  ed  $x \in A'$  (cioè  $a < x \leq b$ ), posto  $B = (x, b]$  (e quindi  $[B] \subseteq A$ ) avremo, a causa della proprietà  $D_3$  delle FdD, che

$$\mathbf{P}_0(x, b] = F(b) - F(x) \longrightarrow F(b) - F(a) = \mathbf{P}_0(a, b] \quad (x \rightarrow a^+);$$

e pertanto (usando anche la seconda proprietà di II.1.12), comunque assegnato un numero  $\delta > 0$  potremo sempre trovare un  $x \in (a, b]$  (cioè un  $B \in \mathcal{A}$  con  $[B] \subseteq A'$ ) tale che

$$\mathbf{P}_0(A' \setminus B) = \mathbf{P}_0((a, b] \setminus (x, b]) = \mathbf{P}_0(a, b] - \mathbf{P}_0(x, b] < \delta.$$

Basterà poi scegliere  $\delta = \epsilon/2^{n+1}$  per ottenere (3). Ogni elemento della successione  $(B_n)_{n \in \mathbf{N}}$  con la proprietà (3) potrà quindi essere costruito usando ripetutamente (per un numero finito di volte) la costruzione qui descritta.

$$(4) \quad A'_n \downarrow \emptyset \quad \text{e} \quad \bigcap_n [B_n] = \emptyset;$$

infatti dalla definizione di  $A'_n$  data in (3) si vede che la successione delle  $A'_n$  è decrescente (dato che  $(A_n)_{n \in \mathbf{N}}$  è decrescente per ipotesi) e che  $A'_n \subseteq A_n$ ,  $\forall n \in \mathbf{N}$ . Pertanto  $\bigcap_n A'_n \subseteq \bigcap_n A_n = \emptyset$  e quindi  $A'_n \downarrow \emptyset$ . Della successione dei  $[B_n]$ , invece,

non si può dire che sia decrescente; siccome però  $[B_n] \subseteq A'_n, \forall n \in \mathbf{N}$ , risulta comunque che  $\bigcap_n [B_n] \subseteq \bigcap_n A'_n = \emptyset$  e quindi  $\bigcap_n [B_n] = \emptyset$ .

$$(5) \quad \exists \nu \in \mathbf{N} \ni' \bigcap_{n=1}^{\nu} B_n = \emptyset;$$

dalla catena di inclusioni

$$B_n \subseteq [B_n] \subseteq A'_n \subseteq (-M, M] \subseteq [-M, M]$$

si deduce che i chiusi  $[B_n]$  sono tutti contenuti nell'intervallo chiuso e limitato  $[-M, M]$ . Inoltre gli insiemi  $\overline{[B_n]}$ , in quanto complementari di chiusi, sono aperti di  $\mathbf{R}$ : definiti allora gli insiemi

$$D_n = [-M, M] \setminus [B_n] = [-M, M] \cap \overline{[B_n]} \subseteq \overline{[B_n]}$$

in modo che  $\bigcup_n D_n \subseteq \bigcup_n \overline{[B_n]}$ , in base al punto (4) si ha

$$\begin{aligned} \bigcup_n D_n &= \bigcup_n \left( [-M, M] \cap \overline{[B_n]} \right) = [-M, M] \cap \left( \bigcup_n \overline{[B_n]} \right) \\ &= [-M, M] \cap \left( \overline{\bigcap_n [B_n]} \right) = [-M, M] \cap \overline{\emptyset} = [-M, M], \end{aligned}$$

sicché in definitiva  $[-M, M] = \bigcup_n D_n \subseteq \bigcup_n \overline{[B_n]}$ , cioè gli aperti  $\overline{[B_n]}$  ricoprono  $[-M, M]$ . Ma  $[-M, M]$  è un intervallo chiuso di  $\mathbf{R}$  e quindi, per il Lemma di Heine-Borel<sup>2</sup> la famiglia degli aperti  $\overline{[B_n]}$  deve contenere un ricoprimento finito di  $[-M, M]$ , cioè che deve esistere un  $\nu \in \mathbf{N}$  tale che

$$[-M, M] \subseteq \bigcup_{n=1}^{\nu} \overline{[B_n]}.$$

Ne segue allora che

$$\bigcap_{n=1}^{\nu} [B_n] = \emptyset$$

---

<sup>2</sup> Vedi ad esempio **A. N. Kolmogorov e S. V. Fomin**: *Elementi di teoria delle funzioni e di analisi funzionale*; MIR, Mosca, 1980, p. 98. Il nome di Heinrich E. Heine (1821 - 1881) è storicamente legato a questo *Lemma del ricoprimento*, ma furono notevoli anche i suoi contributi all'analisi, ai metodi insiemistici cantoriani ed alla matematica applicata. Allievo di K. F. Gauss e K. G. Jacobi, insegnò per quasi tutta la sua vita nell'Università di Halle. Egli si occupò principalmente dei metodi analitici che permettono di padroneggiare l'equazione di Laplace nella teoria del potenziale e seppe incoraggiare molti allievi di talento fra cui G. Cantor. In un lavoro del 1872 dimostrò il teorema secondo il quale una funzione continua in un intervallo chiuso e limitato è anche uniformemente continua nello stesso intervallo. Per la dimostrazione egli si servì del principio che consente di ricoprire l'intervallo con un numero finito di aperti, principio che fu poi perfezionato da E. Borel e divenne uno dei cardini dell'analisi moderna.

in quanto se ciò non fosse vero risulterebbe

$$[-M, M] \cap \left( \overline{\bigcap_{n=1}^{\nu} [B_n]} \right) = [-M, M] \setminus \left( \bigcap_{n=1}^{\nu} [B_n] \right) \neq [-M, M]$$

a causa del fatto che

$$\bigcap_{n=1}^{\nu} [B_n] \subseteq [-M, M],$$

mentre invece noi sappiamo che, essendo

$$[-M, M] \subseteq \bigcup_{n=1}^{\nu} \overline{[B_n]},$$

si ha

$$[-M, M] \cap \left( \overline{\bigcap_{n=1}^{\nu} [B_n]} \right) = [-M, M] \cap \bigcup_{n=1}^{\nu} \overline{[B_n]} = [-M, M].$$

Siccome infine

$$\bigcap_{n=1}^{\nu} B_n \subseteq \bigcap_{n=1}^{\nu} [B_n] = \emptyset$$

l'affermazione (5) resta completamente dimostrata.

$$(6) \quad \forall n > \nu, \quad \mathbf{P}_0(A'_n) < \frac{\epsilon}{2};$$

infatti, essendo  $(A'_n)_{n \in \mathbf{N}}$  decrescente, risulterà  $A'_n \subseteq A'_\nu$ ,  $\forall n > \nu$  e quindi

$$\begin{aligned} \mathbf{P}_0(A'_n) &\leq \mathbf{P}_0(A'_\nu) = \mathbf{P}_0\left(A'_\nu \setminus \bigcap_{k=1}^{\nu} B_k\right) \leq \mathbf{P}_0\left(\bigcup_{k=1}^{\nu} (A'_\nu \setminus B_k)\right) \\ &\leq \sum_{k=1}^{\nu} \mathbf{P}_0(A'_\nu \setminus B_k) \leq \frac{\epsilon}{2} \sum_{k=1}^{\nu} \frac{1}{2^k} \leq \frac{\epsilon}{2} \sum_{k=1}^{\infty} \frac{1}{2^k} = \frac{\epsilon}{2}, \end{aligned}$$

dove abbiamo tenuto successivamente conto del fatto che è valido il risultato del punto (5); che essendo  $A'_\nu \subseteq A'_k$ ,  $\forall k \leq \nu$ , risulta

$$A'_\nu \setminus \bigcap_{k=1}^{\nu} B_k \subseteq \bigcup_{k=1}^{\nu} (A'_k \setminus B_k);$$

che  $\mathbf{P}_0$  è sub-additiva; che i risultati del punto (3) sono validi ed infine che

$$\sum_{k=1}^{\infty} 2^{-k} = 1.$$

La tesi complessiva del Teorema,  $\mathbf{P}_0(A_n) \xrightarrow{n} 0$ , segue allora dal fatto che dai punti (2) e (6) discende che  $\mathbf{P}_0(A_n) < \epsilon, \forall n > \nu$ , dove  $\nu \in \mathbf{N}$  viene fissato nel punto (5).  $\square$

**II.3.6 Definizione:** Diremo che  $\mu$  è una **misura di Lebesgue-Stieltjes**<sup>3</sup> (L-S) su  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$  se essa è ( $\sigma$ -additiva) e tale che la misura  $\mu(B)$  di ogni intervallo limitato  $B$  risulti finita. Chiameremo inoltre **Funzione di Distribuzione (Cumulativa) Generalizzata** su  $\mathbf{R}$  (FdDG) ogni funzione  $G(x)$  definita su  $\mathbf{R}$  che soddisfa le proprietà  $D_1$  e  $D_3$ .  $\triangle$

**II.3.7 Osservazione:** La conoscenza della FdD associata alla probabilità  $\mathbf{P}$  ci consente dunque di calcolare le probabilità di ogni evento  $B$  costituito dall'unione di alcuni intervalli disgiunti. Non abbiamo ancora mostrato, invece, come si calcola la probabilità di eventi generici  $B \in \mathcal{B}$ : si tratta di un argomento che sarà ripreso più oltre quando avremo introdotto i concetti relativi all'integrazione. Inoltre le misure di probabilità  $\mathbf{P}$  definite su  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$  tramite una FdD  $F(x)$ , in quanto misure finite, sono un caso particolare di misure di L-S che sono solo  $\sigma$ -finite<sup>4</sup>.

---

<sup>3</sup> Dopo i primi anni trascorsi a Beauvais, sua città natale, Henri Léon Lebesgue (1875 - 1941) iniziò i suoi studi di matematica all'École Normale Supérieure nel 1894. Fin da allora egli constatò che nei trattati di teoria delle funzioni "ci si limitava a poche considerazioni molto generali sulla nozione di funzione, per restringersi immediatamente a famiglie di funzioni molto particolari, per le quali le considerazioni generali potevano venire semplificate". La ragione era nel fatto che "l'analisi delle funzioni di variabile reale si riduceva allo studio di due operazioni funzionali, la derivazione e l'integrazione" mentre "le funzioni per le quali non si sapeva effettuare né l'una né l'altra operazione venivano di fatto escluse dall'analisi". Conseguentemente egli trovò necessario estendere la portata delle nozioni di derivata ed integrale: tra il 1899 ed il 1902, mentre insegnava al Liceo di Nancy, egli riuscì a realizzare in cinque lavori pubblicati sui *Comptes Rendus de l'Académie des Sciences* il programma delineato negli anni precedenti. Le sue ricerche vennero poi compendiate nel 1904 nelle *Leçons sur l'intégration et la recherche des fonctions primitives* pubblicate nella serie di monografie diretta da Émile Borel. Nel mondo matematico francese non mancarono ovviamente coloro che considerarono l'*integrale di Lebesgue* solo come uno strumento per trattare casi patologici e che videro come un sovvertimento dell'analisi l'introduzione della "mostruosità" delle funzioni senza derivate. La vittoria delle idee di Lebesgue non avvenne, pertanto, senza difficoltà: un disprezzo come quello manifestato a suo tempo dallo stesso C. Hermite (1822 - 1901) per "la piaga delle funzioni senza derivate" era duro a spegnersi. La reale portata delle nuove idee doveva, comunque, essere messa in luce dagli sviluppi della teoria delle funzioni e dell'analisi funzionale. Nel 1921 Lebesgue fu nominato professore al Collège de France e l'anno successivo fu eletto membro dell'Académie des Sciences di Parigi. Negli ultimi vent'anni di vita proseguì le sue ricerche prevalentemente da un punto di vista storico e pedagogico, impegnandosi attivamente nelle questioni riguardanti l'insegnamento secondario.

L'olandese Thomas Johannes Stieltjes (1856 - 1894) lavorò fino al 1883 presso l'osservatorio astronomico di Leiden. Nel 1886 conseguì un dottorato in matematica a Parigi ed ottenne una cattedra all'Università di Tolosa che conservò fino alla morte che lo colse appena trentottenne. Il concetto grazie al quale egli è principalmente ricordato, l'*integrale di Stieltjes*, fu introdotto nei suoi studi sulle frazioni continue. Gli altri suoi lavori includono scritti sulle armoniche sferiche, sui polinomi di Legendre nonché su molte questioni di astronomia teorica.

<sup>4</sup> Infatti  $\mathbf{R}$  può sempre essere decomposto in un insieme numerabile di intervalli finiti e disgiunti la cui misura di L-S è, per definizione, finita.

Ricordando ora che il Teorema di Carathéodory II.3.4 richiede che la misura  $\mu_0$  su  $(\Omega, \mathcal{A})$  sia solo  $\sigma$ -finita, è lecito attendersi che si possa stabilire una corrispondenza biunivoca fra misure di L-S ed un qualche tipo di funzioni che generalizzi le FdD  $F(x)$ . A questo scopo è stato introdotto il concetto di FdDG. Si vede subito che una FdDG  $G(x)$  gode delle stesse proprietà di una FdD fatta eccezione per  $D_2$  e che, pertanto, essa può assumere valori negativi o maggiori di 1 e i suoi limiti per  $x \rightarrow \pm\infty$  possono non essere finiti. È facile mostrare ora che, se  $\mu$  è una misura di L-S su  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ , la funzione  $G(x)$  definita da

$$G(x') - G(x) = \mu(x, x'], \quad (x < x')$$

è una FdDG<sup>5</sup>. La proposizione inversa, secondo la quale, data una FdDG  $G(x)$ , esiste sempre un'unica misura di L-S tale che

$$\mu(a, b] = G(b) - G(a), \quad (a < b)$$

richiede invece una discussione più accurata come nel caso della relazione fra misure di probabilità e FdD. ○

**II.3.8 Proposizione:** Data una FdDG su  $\mathbf{R}$   $G(x)$ , esiste un'unica misura di L-S  $\mu$  su  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$  tale che

$$\mu(a, b] = G(b) - G(a)$$

comunque scelti  $a$  e  $b$  con  $-\infty \leq a < b \leq +\infty$ .

**Dimostrazione:** Se ambedue i limiti di  $G(x)$  per  $x \rightarrow \pm\infty$  esistono finiti (anche se diversi da 0 ed 1) la dimostrazione è identica a quella del Teorema II.3.5. Se invece almeno uno di tali limiti non è finito la prova deve essere modificata. Per far questo si definisce sull'algebra  $\mathcal{A}$  delle unioni finite di intervalli di  $\mathbf{R}$  l'applicazione (ovviamente additiva e  $\sigma$ -finita) che ad ogni intervallo  $(a, b]$  associa la misura

$$\mu_0(a, b] = G(b) - G(a).$$

Nel caso di intervalli illimitati del tipo  $(-\infty, b]$ ,  $(a, +\infty]$ ,  $(-\infty, +\infty]$  il valore di  $\mu_0$  può anche essere  $+\infty$ ; così, ad esempio,

$$\mu_0(-\infty, b] = G(b) - \lim_{x \rightarrow -\infty} G(x)$$

può assumere valore  $+\infty$ . Per poter stabilire, tramite il Teorema II.3.4, che esiste un'estensione  $\mu$  di  $\mu_0$  a tutta  $\mathcal{B}(\mathbf{R})$ , dovremo ora provare che  $\mu_0$  è anche  $\sigma$ -additiva

---

<sup>5</sup> Nota che, diversamente dal caso delle FdD, la FdDG  $G(x)$  associata in questo modo a  $\mu$  è definita a meno di un fattore additivo arbitrario, in quanto, se  $G(x)$  è una FdDG definita dalla relazione precedente, anche  $G(x) + c$  lo è, con  $c$  numero reale arbitrario.

su  $\mathcal{A}$ . Consideriamo pertanto una successione  $(A_n)_{n \in \mathbf{N}}$  di elementi disgiunti di  $\mathcal{A}$  tale che  $A = \bigcup_n A_n \in \mathcal{A}$ . Dalla additività di  $\mu_0$  si ha allora che,  $\forall n \in \mathbf{N}$ ,

$$\mu_0(A) = \sum_{k=1}^n \mu_0(A_k) + \mu_0\left(\bigcup_{k>n} A_k\right) \geq \sum_{k=1}^n \mu_0(A_k),$$

e quindi anche

$$\mu_0(A) \geq \lim_n \sum_{k=1}^n \mu_0(A_k) = \sum_k \mu_0(A_k).$$

Se ora  $\sum_k \mu_0(A_k) = +\infty$ , la  $\sigma$ -additività di  $\mu_0$  è banalmente verificata visto che la disuguaglianza precedente impone anche che  $\mu_0(A) = +\infty$ . Se invece  $\sum_k \mu_0(A_k) < +\infty$  si definisce la seguente successione di FdDG:

$$G_n(x) = \begin{cases} G(-n), & \text{se } x < -n, \\ G(x), & \text{se } -n \leq x \leq n, \\ G(n), & \text{se } n < x, \end{cases}$$

per le quali si verifica subito che

$$\begin{aligned} \lim_n G_n(x) &= G(x), \quad \forall x \in \mathbf{R}; \\ \lim_n G_n(\pm n) &= \lim_n G(\pm n) = \lim_{x \rightarrow \pm\infty} G(x). \end{aligned}$$

Inoltre, comunque scelto  $n \in \mathbf{N}$ , si ha

$$\begin{aligned} \lim_{x \rightarrow -\infty} G_n(x) &= G_n(-n) = G(-n) > -\infty, \\ \lim_{x \rightarrow +\infty} G_n(x) &= G_n(n) = G(n) < +\infty, \end{aligned}$$

e pertanto, sulla base della prima parte della presente dimostrazione, ad ogni  $G_n(x)$  è associata una misura  $\sigma$ -additiva  $\mu_n$  su  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ . La successione di misure  $\sigma$ -additive  $\mu_n$  verifica inoltre le seguenti due proprietà:

$$(1) \quad \mu_n(B) \leq \mu_0(B), \quad \forall n \in \mathbf{N}, \forall B \in \mathcal{A};$$

infatti dall'additività di  $\mu_0$  si ha<sup>6</sup>

$$\begin{aligned} \mu_0(B) &= \mu_0(B \cap [-n, n]) + \mu_0(B \cap \overline{[-n, n]}) \\ &= \mu_n(B) + \mu_0(B \cap \overline{[-n, n]}) \geq \mu_n(B). \end{aligned}$$

---

<sup>6</sup> Qui abbiamo anche fatto uso dell'osservazione in base alla quale per insiemi  $A \in \mathcal{A}$  tali che  $A \subseteq [-n, n]$  risulta anche  $\mu_0(A) = \mu_n(A)$ , dato che in  $[-n, n]$  risulta sempre  $G_n(x) = G(x)$ , sicché  $\mu_0(B \cap [-n, n]) = \mu_n(B \cap [-n, n])$ . Inoltre, siccome  $\mu_n(D) = 0$  se  $D \cap [-n, n] = \emptyset$ , si ha anche  $\mu_n(B) = \mu_n(B \cap [-n, n]) + \mu_n(B \cap \overline{[-n, n]}) = \mu_n(B \cap [-n, n])$  dato che ovviamente  $B \cap \overline{[-n, n]}$  e  $[-n, n]$  sono disgiunti.

$$(2) \quad \lim_n \mu_n(B) = \mu_0(B), \quad \forall B \in \mathcal{A};$$

infatti se  $B$  è limitato, esisterà un certo indice  $\nu \in \mathbf{N}$  tale che per ogni  $n > \nu$  risulterà  $B \subseteq [-n, n]$  (cioè  $B \cap \overline{[-n, n]} = \emptyset$ ) e quindi  $\mu_n(B) = \mu_0(B)$ ,  $\forall n > \nu$ ; se invece  $B$  non è limitato risulta comunque che  $\mu_n(B) \xrightarrow{n} \mu_0(B)$ : ad esempio, se  $B = (-\infty, b]$ , si ha che

$$\mu_n(-\infty, b] = G_n(b) - \lim_{x \rightarrow -\infty} G_n(x) = G_n(b) - G_n(-n),$$

e quindi per le proprietà delle  $G_n$

$$\begin{aligned} \lim_n \mu_n(-\infty, b] &= \lim_n G_n(b) - \lim_n G_n(-n) \\ &= G(b) - \lim_{x \rightarrow -\infty} G(x) = \mu_0(-\infty, b]. \end{aligned}$$

Possiamo ora verificare la  $\sigma$ -additività di  $\mu_0$ : infatti, essendo  $A = \bigcup_k A_k \in \mathcal{A}$ , da (2) e dalla  $\sigma$ -additività delle  $\mu_n$  si ha

$$\mu_0(A) = \lim_n \mu_n(A) = \lim_n \mu_n\left(\bigcup_k A_k\right) = \lim_n \sum_k \mu_n(A_k);$$

ricordando poi che siamo nel caso in cui  $\sum_k \mu_0(A_k) < +\infty$ , che inoltre  $\mu_0(A) \geq \sum_k \mu_0(A_k)$  e che è valida la (1), si ha

$$\begin{aligned} 0 &\leq \mu_0(A) - \sum_k \mu_0(A_k) = \lim_n \sum_k \mu_n(A_k) - \sum_k \mu_0(A_k) \\ &= \lim_n \sum_k [\mu_n(A_k) - \mu_0(A_k)] \leq 0 \end{aligned}$$

e quindi  $\mu_0(A) = \sum_k \mu_0(A_k)$ , cioè  $\mu_0$  è anche  $\sigma$ -additiva. □

**II.3.9 Esempio:** Daremo ora alcuni esempi tipici di misure di L-S definiti a partire da opportune FdD (o FdDG). La misura  $\lambda$  definita su  $([0, 1], \mathcal{B}([0, 1]))$  tramite la FdD (vedi Fig. II.3.1)

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

si chiama **Misura di Lebesgue su  $[0, 1]$**  ed è una probabilità che ad ogni  $(a, b] \in \mathcal{B}([0, 1])$  associa la misura

$$\lambda(a, b] = b - a.$$

Diremo invece **Misura di Lebesgue su  $\mathbf{R}$**  la misura  $\lambda$  definita su  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$  dalla FdDG

$$G(x) = x$$



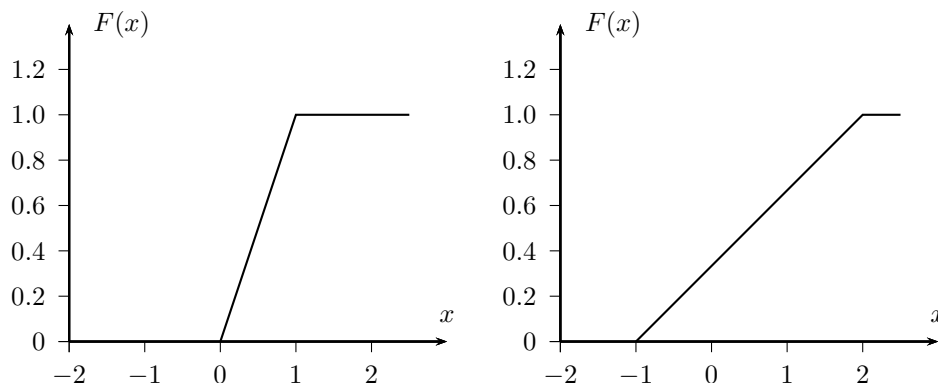
e che ad ogni  $(a, b] \in \mathcal{B}(\mathbf{R})$  associa la misura  $\lambda(a, b] = b - a$ . In questo secondo caso  $\lambda$  è  $\sigma$ -finita, ma non finita e quindi non è una misura di probabilità. Infine dato un intervallo  $[a, b] \subseteq \mathbf{R}$  si chiama **Distribuzione uniforme su  $[a, b]$**  la misura di probabilità  $\mu$  definita su  $([a, b], \mathcal{B}([a, b]))$  dalla FdD (vedi Fig. II.3.1 dove  $a = -1$  e  $b = 2$ )

$$F(x) = \begin{cases} 0, & \text{se } x < a, \\ (x - a)/(b - a), & \text{se } a \leq x \leq b, \\ 1, & \text{se } b < x, \end{cases}$$

che ad ogni intervallo  $(x_1, x_2] \in \mathcal{B}([a, b])$  associa la probabilità

$$\mu(x_1, x_2] = \frac{x_2 - x_1}{b - a}.$$

In tutti e tre questi esempi le FdD (oppure le FdDG) sono continue: ciò implica che le misure degli insiemi  $\{x\}$  ridotti a singoli punti sono nulle.  $\diamond$



**Fig. II.3.1** Misura di Lebesgue su  $[0, 1]$  e distribuzione uniforme su  $[-1, 2]$ .

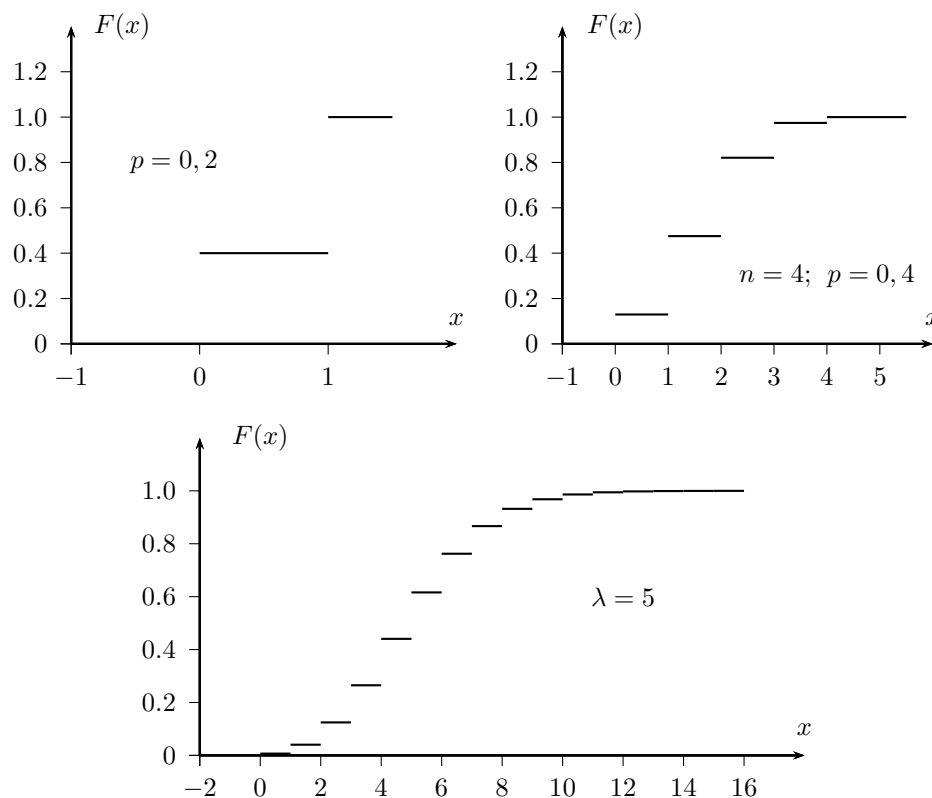
**II.3.10 Definizione:** Diremo che  $\mathbf{P}$  è una **misura di probabilità discreta** su  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$  se la sua FdD  $F(x)$  è costante a tratti e cambia valore in maniera discontinua nei punti  $x_1, x_2, \dots$  (che possono costituire un insieme numerico finito o numerabile) con  $\Delta F(x_i) = F(x_i) - F(x_i^-) > 0$ .  $\triangle$

**II.3.11 Osservazione:** La FdD di una misura di probabilità discreta, dunque, presenta un tipico andamento a scala (vedi Fig. II.3.2) dal quale si deduce subito che  $\mathbf{P}(a, b] = 0$  se  $(a, b]$  non contiene nessuno dei punti di discontinuità  $x_i$ ; se invece  $(a, b]$  contiene alcuni dei punti  $x_i$  si ha

$$\mathbf{P}(a, b] = \sum_{x_i \in (a, b]} \Delta F(x_i).$$

In ogni caso risulta ovviamente  $\mathbf{P}(a, b] = F(b) - F(a)$  e in particolare  $\mathbf{P}\{x\} = 0$  se  $F(x)$  è continua in  $x$ , mentre  $\mathbf{P}\{x_i\} = \Delta F(x_i)$  nei punti di discontinuità di  $F(x)$ .

La probabilità  $\mathbf{P}$  risulta quindi concentrata sui punti  $x_1, x_2, \dots$  e può quindi anche essere assegnata dando i numeri  $p_1, p_2, \dots$  con  $p_k = \mathbf{P}\{x_k\}$  che prendono anche il nome di **distribuzione di probabilità discreta**.  $\bigcirc$



**Fig. II.3.2** Distribuzioni di Bernoulli, binomiale e di Poisson.

**II.3.12 Esempio:** Abbiamo già incontrato alcuni casi di distribuzioni discrete nella discussione sulle v.a. del capitolo I.5: qui richiameremo solo i casi più notevoli. Una **distribuzione discreta uniforme** sull'insieme finito di punti  $\{x_1, \dots, x_n\}$  è definita da

$$p_k = \frac{1}{n}; \quad k = 1, \dots, n.$$

In particolare negli esempi che seguono le  $x_k$  sono numeri interi ( $x_k = k$ ). Una **distribuzione di Bernoulli** sui punti  $\{0, 1\}$  è assegnata dando un numero  $0 \leq p \leq 1$  e definendo (vedi Fig. II.3.2 dove  $p = 0,2$ )

$$p_1 = p; \quad p_0 = q = 1 - p.$$

Una **distribuzione binomiale** sui punti  $\{1, \dots, n\}$  è definita dalle probabilità (vedi Fig. II.3.2 dove  $p = 0,4$  ed  $n = 4$ )

$$p_k = \binom{n}{k} p^k q^{n-k}; \quad k = 1, \dots, n.$$

Infine una **distribuzione di Poisson** sull'insieme numerabile  $\mathbf{N}$  è assegnata tramite le probabilità (vedi Fig. II.3.2 dove  $\lambda = 5$ )

$$p_k = \frac{\lambda^k e^{-\lambda}}{k!}; \quad k \in \mathbf{N};$$

dove  $\lambda$  è un numero reale e positivo. ◇

**II.3.13 Definizione:** Date due misure  $\mu$  e  $\nu$  su un medesimo spazio probabilizzabile  $(\Omega, \mathcal{F})$ , si dice che  $\nu$  è **assolutamente continua (a.c.) rispetto a  $\mu$**  (e si scrive  $\nu \ll \mu$ ) quando assegnato un arbitrario  $A \in \mathcal{F}$  di misura  $\mu(A) = 0$ , risulta anche  $\nu(A) = 0$ . Nel caso particolare di misure su  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ , quando una misura di probabilità  $\mathbf{P}$  risulta a.c. rispetto alla misura di Lebesgue  $\lambda$  si usa anche dire, per brevità, che la FdD  $F(x)$  di  $\mathbf{P}$  è **assolutamente continua (a.c.)**. △

**II.3.14 Teorema (di Radon - Nikodym su  $\mathbf{R}$ ):** Una FdD  $F(x)$  su  $\mathbf{R}$  è a.c. se e solo se esiste una funzione  $f(x)$  definita su  $\mathbf{R}$  a valori non negativi, normalizzata ad 1:

$$\int_{-\infty}^{+\infty} f(x) dx = 1,$$

e per la quale risulti

$$F(x) = \int_{-\infty}^x f(t) dt.$$

Tale funzione  $f(x)$  prende il nome di **Funzione di Densità** (fdd) della FdD a.c.  $F(x)$ .

**Dimostrazione:** Omessa<sup>7</sup> □

**II.3.15 Osservazione:** È possibile dimostrare che se una FdD  $F(x)$  è a.c. essa è anche continua e risulta derivabile **quasi ovunque** (q.o.) (cioè è non derivabile

---

<sup>7</sup> Vedi ad esempio **M. Métivier:** *Notions Fondamentales de la Théorie des Probabilités*; Dunod, Paris, 1972, p. 288. J. Radon (1887 - 1956) mostrò nel 1913 come si poteva giungere ad una definizione molto generale di integrale che comprendeva tanto quella di Lebesgue quanto quella di Stieltjes e si basava su una nozione di misura che compendia la nozione stessa di funzione a variazione limitata. Per dettagli sulla teoria di Radon si può far utile riferimento al capitolo X di **M. Zamansky:** *Introduzione all'Algebra e all'Analisi moderna*; Feltrinelli, Milano, 1966. Nel 1930 O. Nikodym diede un'ulteriore generalizzazione del concetto di integrale introdotto da Radon.

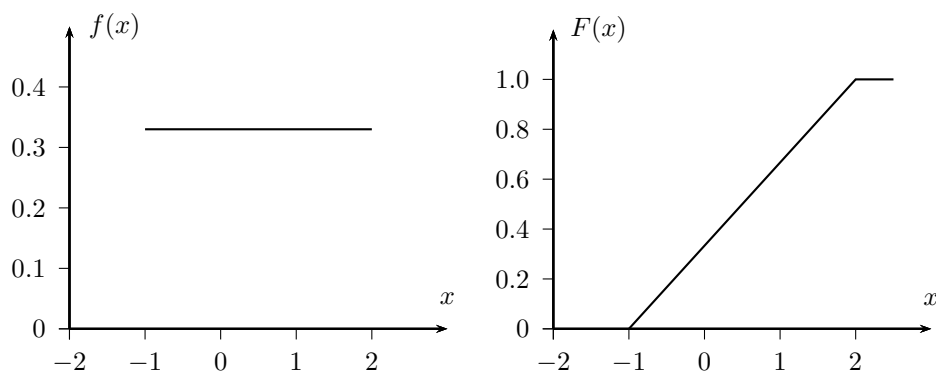
solo su un insieme di punti di misura di Lebesgue nulla). Il viceversa, invece, non è vero: esistono delle FdD  $F(x)$  che sono continue ma non sono a.c., come sarà mostrato in un successivo esempio. È facile inoltre verificare che, data una funzione  $f(x)$  definita su  $\mathbf{R}$ , a valori non negativi e tale che

$$\int_{-\infty}^{+\infty} f(x) dx = 1,$$

(l'integrale è qui inteso nel senso di Lebesgue), la funzione

$$F(x) = \int_{-\infty}^x f(t) dt$$

risulta essere una FdD continua e a.c. Il Teorema di Radon - Nikodym II.3.14 stabilisce il fatto notevole che risulta vero anche il viceversa: ogni FdD  $F(x)$  a.c. risulta essere la primitiva di un'opportuna fdd  $f(x)$ , sicché tutte e sole le FdD  $F(x)$  a.c. possono essere assegnate definendo una fdd  $f(x)$  su  $\mathbf{R}$ . Riassumendo potremo dire che fra tutte le FdD continue  $F(x)$  è possibile individuare l'importante classe delle funzioni a.c. che possono essere definite assegnando una fdd  $f(x)$  tale che  $F(x)$  sia la sua primitiva nel senso qui indicato.



**Fig. II.3.3** fdd e FdD di una distribuzione uniforme su  $[-1, 2]$ .

Nota anche che, quando  $F(x)$  è a.c. con fdd  $f(x)$ , le probabilità di intervalli  $(a, b]$  possono essere calcolati con la formula

$$\mathbf{P}(a, b] = F(b) - F(a) = \int_a^b f(t) dt$$

Naturalmente, se  $F(x)$  risulta essere anche derivabile (ovunque e non solo q.o.), la fdd non è altro che la derivata della FdD:

$$f(x) = F'(x).$$

Sarà utile ricordare qui, analogamente a quanto già detto in Osservazione II.3.15, che queste espressioni consentono di calcolare le probabilità di eventi costituiti da unioni di intervalli (disgiunti) dalla conoscenza della fdd  $f(x)$ : le formule più generali saranno fornite quando avremo discusso più in dettaglio i concetti relativi all'integrazione.  $\circ$

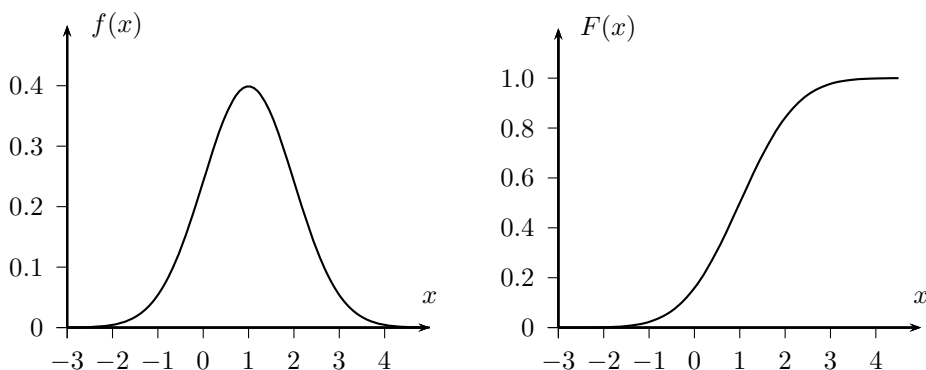
**II.3.16 Esempio:** Sono notevoli i seguenti casi di FdD a.c. dotate di fdd: la FdD **Uniforme** su  $[a, b]$  (vedi Fig. II.3.3 dove  $a = -1$  e  $b = 2$ ) già discussa nell'Esempio II.3.9 è ovviamente dotata di una densità che si ricava facilmente per derivazione e vale (vedi Fig. II.3.3)

$$f(x) = \begin{cases} 1/(b-a), & \text{se } a \leq x \leq b; \\ 0, & \text{altrove.} \end{cases}$$

Una **Distribuzione Normale (Gaussiana)**  $\mathcal{N}(m, \sigma^2)$  è caratterizzata dalla fdd (vedi Fig. II.3.4 dove  $m = -1$  e  $\sigma = 1$ )

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-m)^2/2\sigma^2}, \quad (\sigma > 0)$$

ed è già stata brevemente discussa nell'Esempio I.8.1 in relazione ai Teoremi limite.



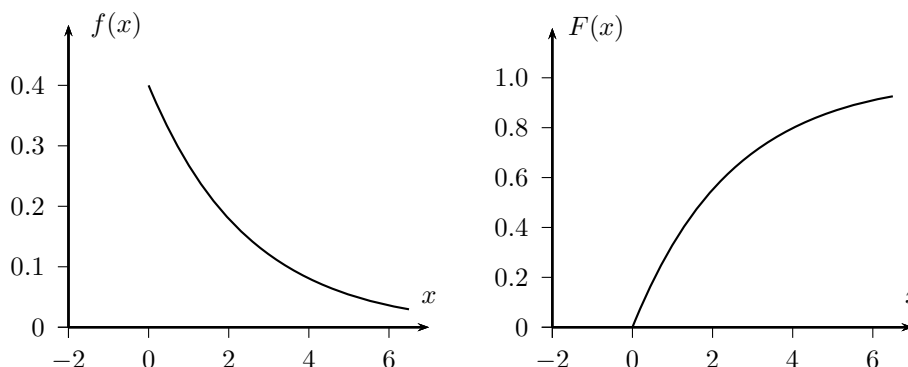
**Fig. II.3.4** fdd e FdD di una distribuzione normale ( $\mu = 1, \sigma = 1$ ).

In quel contesto è anche stato fatto notare che la corrispondente FdD (vedi Fig. II.3.4), ottenuta come primitiva di  $f(x)$ , non è esprimibile mediante combinazioni finite di funzioni elementari ma i suoi valori sono calcolabili numericamente e sono elencati in Tavole come la Tavola I.8.1. La particolare distribuzione normale  $\mathcal{N}(0, 1)$  prende il nome di distribuzione **Normale Standard** e la relativa FdD quello di **Funzione Errore**. Una **Distribuzione Esponenziale** è invece caratterizzata dalla fdd (vedi Fig. II.3.5 dove  $\lambda = 0, 4$ )

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{se } x \geq 0; \\ 0, & \text{se } x < 0, \end{cases}$$

e dalla corrispondente FdD (vedi Fig. II.3.5)

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{se } x \geq 0; \\ 0, & \text{se } x < 0. \end{cases}$$



**Fig. II.3.5** fdd e FdD di una distribuzione esponenziale ( $\lambda = 0,4$ ).

Infine la **Distribuzione di Cauchy**<sup>8</sup> ha come fdd (vedi Fig. II.3.6 dove  $a = 0,8$

---

<sup>8</sup> Augustin-Louis (Barone) Cauchy (1789-1857) trascorse l'infanzia, durante gli anni della rivoluzione, ad Arcueil vicino alle tenute del Marchese P.S. de Laplace che ebbe modo di conoscerlo (1800) e di apprezzarne il precoce genio matematico. Nel 1805 entrò all'École Polytechnique di Parigi e nel 1810 partì per un lungo soggiorno di lavoro a Cherbourg portando con sé, oltre alla *Mécanique céleste* di Laplace e al *Traité des fonctions analytiques* di Lagrange, anche le opere di Virgilio e l'*Imitazione di Cristo* di Tommaso da Kempis: indice dei suoi gusti letterari e delle sue tendenze religiose. Egli fu infatti un cattolico militante e a volte bigotto (narra Stendhal che durante una riunione dell'*Académie des Sciences* "dopo il discorso di un naturalista egli si alzò per impedire gli applausi dell'uditorio dicendo: 'anche se queste cose fossero vere, come penso che non siano, non bisognerebbe comunicarle al pubblico...Ciò può solamente pregiudicare la nostra santa religione' "). Anche in politica fu conservatore: fedele seguace dei Borboni di Francia, rifiutò di prestare giuramento alla *monarchia di luglio* nata dalla rivoluzione del 1830 e si esiliò volontariamente in Svizzera, a Torino e a Praga dove fu precettore del figlio di Carlo X. Tornato a Parigi nel 1838 dovette attendere fino al 1848 (anno della proclamazione della seconda Repubblica) prima di poter accettare una cattedra universitaria. Dispensato, assieme all'astronomo J.F. Arago, da Napoleone III dall'obbligo del giuramento proseguì nel suo insegnamento e nelle sue ricerche fino alla morte che lo colse per una bronchite nel 1857. Nell'immensa opera di Cauchy (789 fra articoli e memorie) si segnalano principalmente i suoi lavori di analisi contenuti prevalentemente nel suo celebre *Cours d'analyse de l'École Royale Polytechnique* (Parigi, 1821). Al di là dei risultati particolari colpisce soprattutto la rivoluzione metodologica in esso contenuta: nel secolo XVIII gli analisti, fiduciosi nei grandi sviluppi del calcolo infinitesimale, poco avevano badato alle perplessità che circondavano i suoi fondamenti. Ma più tardi tali perplessità, che invece di diminuire aumentavano, avevano generato un vero e proprio scetticismo. Cauchy pose a base del calcolo lo studio delle funzioni continue aprendo la strada ad un'analisi locale (il metodo del confronto noto anche come tecnica dell' $\epsilon$  e del  $\delta$ ) che fu poi definitivamente percorsa da K. Weierstraß impostando su basi completamente nuove e finalmente rigorose tutto l'edificio dell'analisi moderna. I risultati di Cauchy non si arrestano

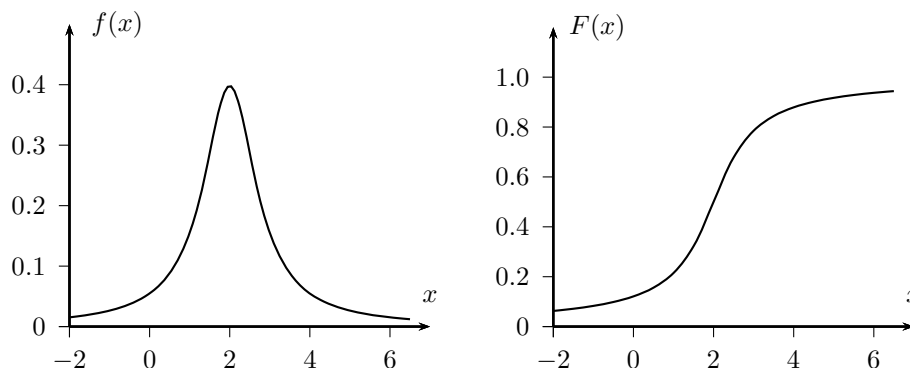
e  $b = 2$ )

$$f(x) = \frac{a}{\pi[a^2 + (x - b)^2]},$$

e come FdD (vedi Fig. II.3.6)

$$F(x) = \frac{1}{\pi} \left( \frac{\pi}{2} + \arctan \frac{x - b}{a} \right),$$

dove  $a$  è un numero reale positivo (diverso da 0). ○



**Fig. II.3.6** fdd e FdD di una distribuzione di Cauchy ( $a = 0,8$  e  $b = 2$ ).

**II.3.17 Definizione:** Diremo che  $\mathbf{P}$  è una **misura di probabilità singolare** su  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$  quando la sua FdD  $F(x)$  risulta essere continua ma non assolutamente continua. △

**II.3.18 Osservazione:** Abbiamo visto che le misure di probabilità a.c. rispetto alla misura di Lebesgue  $\lambda$  sono caratterizzate da una FdD  $F(x)$  a.c. e quindi da una fdd  $f(x)$ . Sappiamo inoltre che una  $F(x)$  a.c. è anche continua, mentre il viceversa non è, in generale, vero nel senso che esistono FdD  $F(x)$  continue (cioè non discrete) che non sono a.c. (cioè non ricavabili da una fdd  $f(x)$ ). Questa osservazione spiega la necessità di introdurre la Definizione II.3.17 e rende chiaro che una misura di probabilità singolare non può essere assegnata né tramite una distribuzione discreta di probabilità (cioè tramite un insieme di numeri  $p_k$  come

---

qui: egli ottenne importanti risultati sulla convergenza delle serie numeriche e di funzioni; diede per la prima volta (1823) una definizione di integrale come limite di somme integrali finite anche se la dimostrazione della convergenza non è del tutto rigorosa; ottenne importanti risultati nel campo delle equazioni differenziali, del calcolo delle probabilità e della fisica matematica. Egli va infine ricordato per aver dato inizio alla teoria delle funzioni (di variabile complessa) olomorfe con il suo celebre teorema integrale (1823) e la conseguente possibilità di rappresentare funzioni mediante la sua formula integrale.

quelli discussi nell'Esempio II.3.12), né tramite una fdd  $f(x)$  (come nell'Esempio II.3.16): l'unico modo di definirla è quello di assegnare una opportuna FdD continua  $F(x)$  la cui esistenza è comunque garantita.  $\circ$

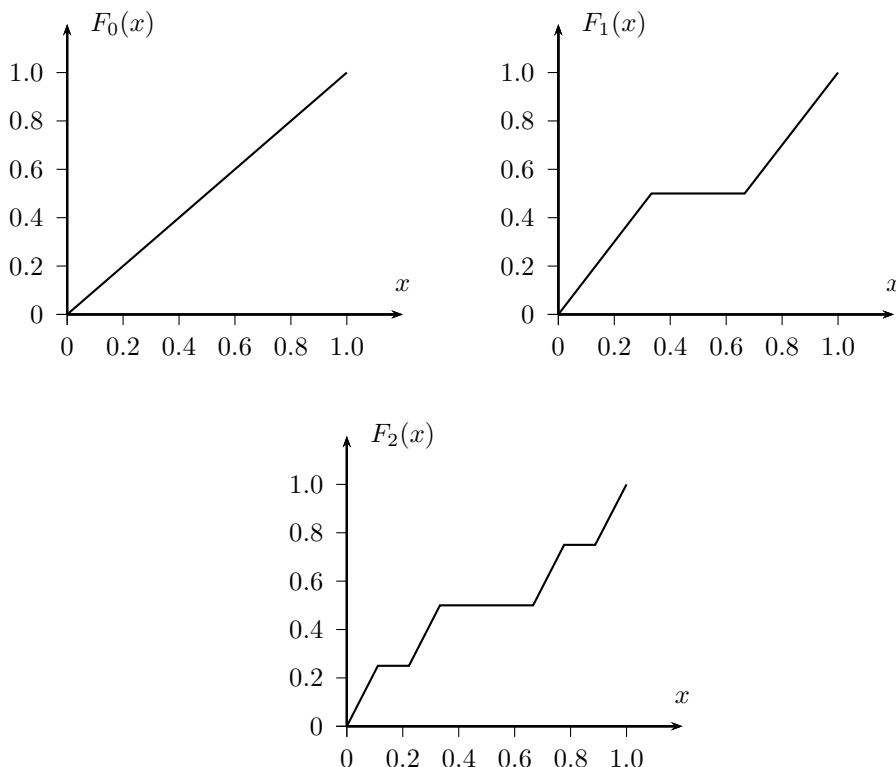


Fig. II.3.7 Funzioni di Cantor con  $n = 0, 1, 2$ .

**II.3.19 Esempio:** Costruiremo ora un esempio di misura di probabilità singolare sullo spazio  $([0, 1], \mathcal{B}([0, 1]))$  considerando una successione  $F_n(x)$  di FdD (definite su  $[0, 1]$ ) i cui primi due elementi sono

$$F_0(x) = x;$$

$$F_1(x) = \begin{cases} 3x/2, & \text{se } 0 \leq x \leq 1/3; \\ 1/2, & \text{se } 1/3 \leq x \leq 2/3; \\ (3x - 1)/2, & \text{se } 2/3 \leq x \leq 1; \end{cases}$$

mentre le successive  $F_n(x)$  si ottengono ricorrentemente suddividendo gli intervalli in cui la precedente FdD non è costante in tre parti eguali e sostituendovi dei nuovi valori che sono costanti nel terzo centrale, crescono linearmente nei due terzi laterali e si connettono tutti in maniera continua (vedi Fig. II.3.7). Questa



procedura (originariamente studiata da Georg Cantor<sup>9</sup>) conduce ad una FdD limite (la **funzione di Cantor**) che è continua ma non è a.c. Infatti, definite su  $[0, 1]$  la misura di probabilità  $\mu$  associata alla  $F(x)$  di Cantor e l'usuale misura di Lebesgue  $\lambda$ , e detto  $\mathcal{N}$  l'insieme dei *punti di crescita*<sup>10</sup> di  $F(x)$ , ci accorgiamo subito del fatto che  $\mu(\mathcal{N}) = 1$ , in quanto la misura di probabilità non può che essere concentrata sui punti di crescita di  $F(x)$  (la probabilità è nulla dove  $F(x)$  è costante), mentre risulta  $\lambda(\mathcal{N}) = 0$ . Infatti l'insieme  $\overline{\mathcal{N}}$  dei punti su cui  $F(x)$  è costante è costituito dall'unione degli intervalli  $(\frac{1}{3}, \frac{2}{3}), (\frac{1}{9}, \frac{2}{9}), (\frac{7}{9}, \frac{8}{9}), \dots$  e la sua misura è quindi

$$\lambda(\overline{\mathcal{N}}) = \frac{1}{3} + \frac{2}{9} + \frac{4}{27} + \dots = \frac{1}{3} \sum_{n=0}^{\infty} \left(\frac{2}{3}\right)^n = 1.$$

Pertanto risulta  $\mu(\mathcal{N}) = 1$  pur essendo  $\lambda(\mathcal{N}) = 0$ , cioè  $\mu$  non è a.c. rispetto a  $\lambda$  e quindi  $F(x)$  non è a.c. e non ammette una fdd.  $\diamond$

**II.3.20 Teorema (di Lebesgue - Nikodym):** Ogni misura di probabilità su  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$  può essere decomposta nella somma di una misura discreta, una a.c. rispetto alla misura di Lebesgue ed una singolare e la corrispondente FdD  $F(x)$  risulta essere una combinazione lineare

$$F(x) = p_1 F_1(x) + p_2 F_2(x) + p_3 F_3(x),$$

con  $F_1$  discreta,  $F_2$  a.c. e  $F_3$  singolare e con  $p_1, p_2, p_3$  numeri positivi tali che  $p_1 + p_2 + p_3 = 1$ .

**Dimostrazione:** Omessa<sup>11</sup> □

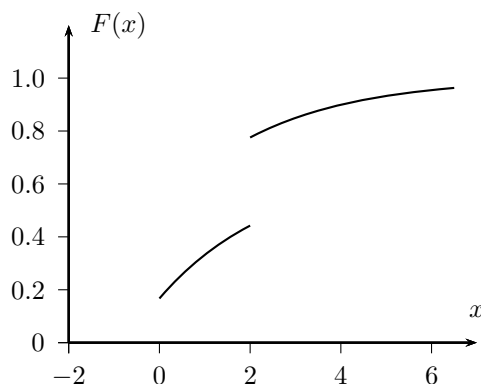
<sup>9</sup> Georg Cantor (1845 - 1918) nacque a San Pietroburgo da genitori danesi: il padre era un ricco mercante mentre la madre discendeva da una famiglia di musicisti. Stabilitosi in Germania dal 1856, studiò a Berlino con K. Weierstraß, E. Kummer e L. Kronecker e divenne professore a Halle (1867) dove rimase per il resto della sua carriera accademica. Cominciò in quel periodo a interessarsi della teoria dei numeri e già nel 1872 aveva introdotto una definizione dei numeri irrazionali mediante successioni convergenti di numeri razionali. Tali ricerche lo condussero allo studio delle caratteristiche degli insiemi infiniti sui quali, assieme a R. Dedekind, ottenne risultati fondamentali trovando però, da parte di Kronecker e dell'ambiente accademico, un'opposizione che lo obbligò a pubblicare i suoi lavori sulla rivista svedese *Acta Mathematica* diretta da G. Mittag-Leffler e gli impedì di essere nominato professore all'Università di Berlino. Dal 1895 egli elaborò la sua teoria dei numeri transfiniti e rimase attivo fino alla fine della sua vita nonostante i disturbi mentali che lo afflissero fin dal 1884.

<sup>10</sup> I punti di crescita di  $F(x)$  sono caratterizzati dal fatto che, comunque scelto  $\epsilon > 0$ , risulta sempre  $F(x + \epsilon) - F(x - \epsilon) > 0$ ; in sostanza si tratta dei punti  $x$  attorno ai quali la  $F(x)$  non si mantiene costante. Si potrebbe mostrare (vedi ad esempio **A.N. Kolmogorov** e **S.V. Fomin: Elementi di Teoria delle Funzioni e di Analisi Funzionale**; Mir, Mosca, 1980, p. 65) che l'insieme dei punti di crescita è un insieme chiuso (dunque un elemento di  $\mathcal{B}(\mathbf{R})$ ) composto di infiniti punti e non numerabile.

<sup>11</sup> Per maggiori dettagli vedi **M. Métivier: Notions Fondamentales de la Théorie des Probabilités**; Dunod, Paris, 1972, p. 290.

**II.3.21 Esempio:** Il Teorema II.3.20 rende chiaro, in definitiva, che i tre tipi di misure precedentemente discussi (discreto, continuo e singolare) esauriscono, assieme alle loro combinazioni, tutti i possibili casi di misure su  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ . Non ci resta che dare un semplice esempio di misura risultante da una combinazione del tipo indicato nel Teorema di Lebesgue-Nikodym: ci limiteremo a una combinazione di una misura discreta con FdD

$$F_1(x) = \begin{cases} 0, & \text{se } x < 0; \\ 1/3, & \text{se } 0 \leq x < 1; \\ 1, & \text{se } 1 \leq x; \end{cases}$$



**Fig. II.3.8** Distribuzione mista.

e di una misura continua esponenziale con FdD

$$F_2(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{se } 0 \leq x; \\ 0, & \text{se } x < 0; \end{cases}$$

con coefficienti  $p_1 = p_2 = \frac{1}{2}$ . La FdD  $F(x) = p_1 F_1(x) + p_2 F_2(x)$  risultante (vedi Fig. II.3.8) presenterà ovviamente delle discontinuità di prima specie in  $x = 0$  ed  $x = 1$  (a causa della presenza di una componente discreta), ma non risulterà costante nei tratti di continuità (a causa della presenza di una componente continua), come nel caso di una misura di probabilità puramente discreta del tipo discusso in II.3.10 e II.3.12.  $\diamond$

## II.4 Misure di probabilità (II Parte)

**II.4.1 Osservazione:** Se  $\mathbf{P}$  è una probabilità assegnata su  $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$  potremo sempre definire la seguente funzione di  $n$  variabili

$$F^{(n)}(x) = F^{(n)}(x_1, \dots, x_n) = \mathbf{P}(-\infty, x] = \mathbf{P}((-\infty, x_1] \times \dots \times (-\infty, x_n])$$

dove per brevità abbiamo posto  $x = (x_1, \dots, x_n)$  e  $(-\infty, x] = (-\infty, x_1] \times \dots \times (-\infty, x_n]$ . Se inoltre definiamo l'operatore *differenza*

$$\Delta_{a_i, b_i} F^{(n)}(x) = F^{(n)}(x_1, \dots, b_i, \dots, x_n) - F^{(n)}(x_1, \dots, a_i, \dots, x_n)$$

con  $a_i \leq b_i$ , è facile vedere che, posto  $a = (a_1, \dots, a_n)$ ,  $b = (b_1, \dots, b_n)$  (con  $a_i \leq b_i$ ) ed  $(a, b] = (a_1, b_1] \times \dots \times (a_n, b_n]$ , risulta

$$\mathbf{P}(a, b] = \Delta_{a_1, b_1} \dots \Delta_{a_n, b_n} F^{(n)}(x).$$

Infatti, se ad esempio ci limitiamo al caso  $n = 2$ , posto  $F(x_1, x_2) = F^{(2)}(x)$ , otteniamo

$$\begin{aligned} \Delta_{a_1, b_1} \Delta_{a_2, b_2} F^{(2)}(x) &= [F(b_1, b_2) - F(b_1, a_2)] - [F(a_1, b_2) - F(a_1, a_2)] \\ &= \mathbf{P}((-\infty, b_1] \times (-\infty, b_2]) - \mathbf{P}((-\infty, b_1] \times (-\infty, a_2]) - \\ &\quad \mathbf{P}((-\infty, a_1] \times (-\infty, b_2]) + \mathbf{P}((-\infty, a_1] \times (-\infty, a_2]) \\ &= \mathbf{P}((a_1, b_1] \times (a_2, b_2]). \end{aligned}$$

Pertanto, diversamente dal caso  $n = 1$  trattato nel capitolo precedente, ora  $\mathbf{P}(a, b]$  non coincide semplicemente con  $F^{(n)}(b) - F^{(n)}(a)$ , ma è data da una combinazione più complicata contenente i  $2^n$  termini prodotti dall'applicazione ripetuta dell'operatore differenza. Le funzioni  $F^{(n)}$  così definite godono, inoltre, di alcune proprietà che generalizzano quelle elencate in II.3.1 per il caso  $n = 1$ :

$D_1$ ) Comunque scelti  $a = (a_i)_{i=1 \dots n}$ ,  $b = (b_i)_{i=1 \dots n}$  (con  $a_i \leq b_i$ ;  $i = 1, \dots, n$ ), e data la positività di  $\mathbf{P}(a, b]$ , risulta sempre

$$\Delta_{a_1, b_1} \dots \Delta_{a_n, b_n} F^{(n)}(x) \geq 0.$$

In particolare  $F^{(n)}$  risulta non decrescente in ciascuna delle sue  $n$  variabili.

$D_2$ ) Risulta sempre

$$\lim_{x \rightarrow +\infty} F^{(n)}(x) = 1$$

dove il limite è inteso nel senso che *ciascuna* delle  $x_i$  tende verso  $+\infty$ ; analogamente, se  $y$  è un punto di  $\mathbf{R}^n$  con *almeno* una coordinata eguale a  $-\infty$ , risulta

$$\lim_{x \rightarrow y} F^{(n)}(x) = 0.$$

Questa proprietà discende dal fatto che  $(-\infty, +\infty] \times \dots \times (-\infty, +\infty] = \mathbf{R}^n$  ha probabilità 1 di realizzarsi, mentre ogni evento rettangolare nel quale almeno un fattore coincide con la parte vuota è la parte vuota di  $\mathbf{R}^n$  e quindi ha probabilità 0 di realizzarsi.

$D_3$ )  $F^{(n)}(x)$  risulta continua quando  $x^{(k)} \downarrow x$ , cioè

$$\lim_k F^{(n)}(x^{(k)}) = F^{(n)}(x)$$

quando i punti  $x^{(k)} = (x_1^{(k)}, \dots, x_n^{(k)})$  tendono ad  $x = (x_1, \dots, x_n)$  decrescendo, nel senso che  $x_i \leq x_i^{(k)} \leq x_i^{(k-1)}$ , con  $i = 1, \dots, n$ .

Sulla base di queste osservazioni possiamo ora dare delle definizioni ed enunciare delle proposizioni che generalizzano quelle date nel capitolo precedente per il caso  $n = 1$ . ○

**II.4.2 Definizione:** Chiameremo **Funzione di Distribuzione (FdD)** su  $\mathbf{R}^n$  ogni funzione  $F^{(n)}(x)$  che soddisfi  $D_1$ ,  $D_2$  e  $D_3$ . Chiameremo invece **Funzione di Distribuzione Generalizzata (FdDG)** su  $\mathbf{R}^n$  ogni funzione che soddisfi  $D_1$  e  $D_3$  senza essere vincolata da  $D_2$  a restare confinata tra 0 ed 1. △

**II.4.3 Teorema:** Data una FdD  $F^{(n)}(x)$  su  $\mathbf{R}^n$ , esiste un'unica probabilità  $\mathbf{P}$  su  $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$  tale che

$$\mathbf{P}(a, b] = \Delta_{a_1, b_1} \dots \Delta_{a_n, b_n} F^{(n)}(x)$$

comunque scelti  $a$  e  $b$  con  $-\infty \leq a_i < b_i \leq +\infty$  ( $i = 1, \dots, n$ ).

**Dimostrazione:** Analoga a quella del Teorema II.3.7. □

**II.4.4 Proposizione:** Data una FdDG  $G^{(n)}(x)$  su  $\mathbf{R}^n$ , esiste un'unica misura di Lebesgue-Stieltjes  $\mu$  su  $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$  (cioè una misura  $\sigma$ -additiva che assegna un valore finito ad ogni insieme limitato di  $\mathcal{B}(\mathbf{R}^n)$ ) tale che

$$\mu(a, b] = \Delta_{a_1, b_1} \dots \Delta_{a_n, b_n} G^{(n)}(x)$$

comunque scelti  $a$  e  $b$  con  $-\infty \leq a_i < b_i \leq +\infty$  ( $i = 1, \dots, n$ ).

**Dimostrazione:** Analoga a quella della Proposizione II.3.10. □

**II.4.5 Osservazione:** Data una FdD su  $\mathbf{R}^n$   $F^{(n)}(x)$ , si controlla facilmente che la funzione di  $n - 1$  variabili

$$F^{(n-1)}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = \lim_{x_i \rightarrow +\infty} F^{(n)}(x_1, \dots, x_i, \dots, x_n)$$

è ancora una FdD su  $\mathbf{R}^{n-1}$  nel senso che essa soddisfa ancora le proprietà  $D_1$ ,  $D_2$  e  $D_3$ . Naturalmente ciò rimane vero quale che sia la coordinata  $x_i$  scelta per eseguire il limite. Se poi questo limite viene eseguito su  $m(< n)$  variabili, si ottiene una  $F^{(n-m)}$  che è ancora una FdD su  $\mathbf{R}^{n-m}$ . Sulla base del Teorema II.4.3 potremo inoltre estendere queste osservazioni anche alle misure probabilità: se  $\mathbf{P}^{(n)}$  è una probabilità assegnata su  $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$ , potremo sempre assegnare delle probabilità sugli spazi  $(\mathbf{R}^k, \mathcal{B}(\mathbf{R}^k))$  (con  $k < n$ ) in maniera indotta definendo, ad esempio, per  $B^{(k)} \in \mathcal{B}(\mathbf{R}^k)$

$$\mathbf{P}^{(k)}(B^{(k)}) = \mathbf{P}^{(n)}(B^{(k)} \times \mathbf{R}_{k+1} \times \dots \times \mathbf{R}_n).$$

Le corrispondenti FdD saranno poi connesse dalla relazione

$$\lim_{x_{k+1} \rightarrow +\infty} \dots \lim_{x_n \rightarrow +\infty} F^{(n)}(x_1, \dots, x_n) = F^{(k)}(x_1, \dots, x_k).$$

Naturalmente la medesima costruzione sarebbe possibile se le  $k$  variabili di  $F^{(k)}$  fossero scelte fra quelle di  $F^{(n)}$  seguendo un ordine arbitrario e non coincidessero proprio con le prime  $k$ . ○

**II.4.6 Esempio:** Se  $F_1(x), \dots, F_n(x)$  sono  $n$  FdD su  $\mathbf{R}$  (sono, cioè, funzioni di una sola variabile  $x$ ), si verifica subito che la funzione di  $n$  variabili

$$F^{(n)}(x_1, \dots, x_n) = F_1(x_1) \cdot \dots \cdot F_n(x_n)$$

è una FdD su  $\mathbf{R}^n$ . Non è invece vero, in generale, che una generica FdD su  $\mathbf{R}^n$  sia il prodotto di  $n$  FdD su  $\mathbf{R}$ : ciò avviene solo in alcuni casi notevoli che saranno discussi nel seguito. Considerazioni analoghe sono ovviamente valide per le FdDG. In particolare, se (per  $i = 1, \dots, n$ ) le

$$F_i(x) = \begin{cases} 0, & \text{se } x < 0; \\ x, & \text{se } 0 \leq x \leq 1; \\ 1, & \text{se } 1 < x; \end{cases}$$

coincidono tutte con la FdD della misura di Lebesgue su  $[0, 1]$  e se le

$$G_i(x) = x$$

coincidono tutte con la FdDG della misura di Lebesgue su  $\mathbf{R}$ , le due funzioni

$$\begin{aligned} F^{(n)}(x) &= F_1(x_1) \cdot \dots \cdot F_n(x_n), \\ G^{(n)}(x) &= G_1(x_1) \cdot \dots \cdot G_n(x_n) = x_1 \cdot \dots \cdot x_n, \end{aligned}$$

sono la FdD e la FdDG delle misure di Lebesgue su  $[0, 1]^n$  ed  $\mathbf{R}^n$ . ◇

**II.4.7 Osservazione:** Quando la probabilità  $\mathbf{P}$  su  $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$  è a.c. rispetto alla misura di Lebesgue (vedi Definizione II.3.16), si dice che la FdD  $F^{(n)}$  è a.c. e

in tal caso una generalizzazione ad  $\mathbf{R}^n$  del Teorema II.3.14 ci garantisce che la FdD può essere calcolata mediante una **Funzione di Densità** (fdd)  $f^{(n)}(x_1, \dots, x_n)$  tramite la relazione

$$F^{(n)}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f^{(n)}(t_1, \dots, t_n) dt_1 \dots dt_n.$$

Come nel caso  $n = 1$  le fdd risultano essere funzioni non negative e tali che la relazione

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f^{(n)}(t_1, \dots, t_n) dt_1 \dots dt_n = 1$$

risulta sempre verificata. Inoltre si può verificare che

$$\mathbf{P}((a_1, b_1] \times \dots \times (a_n, b_n]) = \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} f^{(n)}(t_1, \dots, t_n) dt_1 \dots dt_n,$$

e che la fdd si può sempre ricavare dalla FdD con la relazione

$$f^{(n)}(x_1, \dots, x_n) = \frac{\partial^n}{\partial x_1 \dots \partial x_n} F^{(n)}(x_1, \dots, x_n)$$

quando  $F^{(n)}$  è derivabile. ○

**II.4.8 Esempio:** La distribuzione normale (Gaussiana) dell'esempio II.3.19 ammette una generalizzazione in  $\mathbf{R}^n$  che prende il nome di **Distribuzione Normale (Gaussiana) Multivariata**. Nel caso  $n = 1$  la fdd della  $\mathcal{N}(m, \sigma^2)$  era caratterizzata da due parametri numerici (il cui significato sarà discusso in dettaglio successivamente):  $m$  e  $\sigma > 0$ . Nel caso di  $\mathbf{R}^n$ , invece, bisognerà assegnare un vettore (colonna)  $m$  di numeri reali  $m_1, \dots, m_n$  detto **Vettore delle Medie** ed una matrice  $n \times n$   $\mathbf{R} = \|r_{ij}\|$ , simmetrica ( $r_{ij} = r_{ji}$ ) e non negativa<sup>1</sup> detta **Matrice delle Covarianze**: il motivo di tali denominazioni sarà chiarito successivamente. Nota, comunque, che per  $n = 1$  la matrice  $\mathbf{R}$  si riduce all'unico elemento  $\sigma^2$ . La distribuzione normale multivariata viene pertanto indicata con il simbolo  $\mathcal{N}(m, \mathbf{R})$  e, quando  $\mathbf{R}$  è positiva, posto  $\mathbf{R}^{-1} = \mathbf{A} = \|a_{ij}\|$ , essa è caratterizzata dalla fdd

$$f(x_1, \dots, x_n) = \sqrt{\frac{|\mathbf{A}|}{(2\pi)^n}} e^{-(x-m)^T \mathbf{A} (x-m)/2},$$

---

<sup>1</sup> La matrice  $\mathbf{R}$  dicesi non negativa se, comunque scelto un vettore  $x$  di numeri reali  $x_1, \dots, x_n$ , risulta sempre

$$\sum_{i,j=1}^n r_{ij} x_i x_j \geq 0;$$

la matrice si dice, invece, positiva se tale somma è sempre strettamente positiva (cioè è sempre diversa da zero). Se  $\mathbf{R}$  è positiva essa risulta anche non singolare nel senso che il suo determinante  $|\mathbf{R}| > 0$  è non nullo, e quindi è dotata di inversa  $\mathbf{R}^{-1}$ .

dove con  $|A|$  indichiamo il determinante di una matrice, con  $(x - m)^T$  il vettore riga trasposto di un vettore colonna, mentre fra vettori e matrici viene adottata l'usuale convenzione di prodotto righe per colonne sicché, ad esempio,

$$(x - m)^T A (x - m) = \sum_{i,j=1}^n (x_i - m_i) a_{ij} (x_j - m_j).$$

Sarà bene notare qui che i casi in cui  $\mathbf{R}$  non è invertibile (o, nel caso  $n = 1$ ,  $\sigma = 0$ ), non possono essere caratterizzati tramite una funzione di densità e saranno discussi in maniera dettagliata nel capitolo sulle Funzioni Caratteristiche. In particolare, quando  $n = 2$ , la densità normale bivariata può sempre essere messa nella forma

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \cdot \exp \left\{ -\frac{1}{2(1-r^2)} \left[ \frac{(x_1 - m_1)^2}{\sigma_1^2} - 2r \frac{(x_1 - m_1)(x_2 - m_2)}{\sigma_1\sigma_2} + \frac{(x_2 - m_2)^2}{\sigma_2^2} \right] \right\}$$

con un'opportuna definizione dei parametri  $\sigma_1$ ,  $\sigma_2$  e  $r$  per i quali valgono le limitazioni  $\sigma_i > 0$  e  $|r| < 1$ . ◇

**II.4.9 Osservazione:** L'estensione delle precedenti osservazioni al caso dello spazio  $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$  è meno immediata e richiede qualche cautela. Se  $\mathbf{P}$  fosse una probabilità data su  $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$ , sarebbe anche possibile definire un'intera famiglia di probabilità  $\mathbf{P}^{(M)}$  con  $M \in \mathcal{M}$  (useremo qui la notazione introdotta in II.2.15 e nelle pagine seguenti) sugli spazi  $(\mathbf{R}^{(M)}, \mathcal{B}(\mathbf{R}^{(M)}))$  di dimensione finita  $m$  pari alla cardinalità di  $M$ : basterà infatti porre

$$\mathbf{P}^{(M)}(B^{(M)}) = \mathbf{P}(C(B^{(M)})); \quad B^{(M)} \in \mathcal{B}(\mathbf{R}^{(M)}); \quad M \in \mathcal{M},$$

dove  $C(B^{(M)}) \in \mathcal{C}$  è il cilindro di  $\mathbf{R}^\infty$  con base di dimensione finita  $B^{(M)} \in \mathcal{B}(\mathbf{R}^{(M)})$ . In sostanza è sempre possibile definire la probabilità di  $B^{(M)}$  tramite la probabilità del cilindro (in  $\mathcal{B}(\mathbf{R}^\infty)$ ) che ha per base proprio  $B^{(M)}$ . Da tale costruzione discende anche il fatto che la famiglia (con  $M \in \mathcal{M}$ ) di spazi di probabilità  $(\mathbf{R}^{(M)}, \mathcal{B}(\mathbf{R}^{(M)}), \mathbf{P}^{(M)})$  esibisce la seguente importante proprietà detta **consistenza** (o anche **proiettività**) e già parzialmente discussa in II.4.5: dati due insiemi di indici finiti e disgiunti (cioè  $M, M' \in \mathcal{M}$  con  $M \cap M' = \emptyset$ ) avremo sempre che

$$\begin{aligned} \mathbf{P}^{(M \cup M')}(B^{(M)} \times \mathbf{R}^{(M')}) &= \mathbf{P}(C(B^{(M)} \times \mathbf{R}^{(M')})) = \mathbf{P}(C(B^{(M)})) \\ &= \mathbf{P}^{(M)}(B^{(M)}), \end{aligned}$$

data la possibilità di descrivere il medesimo cilindro in diverse rappresentazioni finito dimensionali. Bisogna ora, però, notare una importante differenza fra gli spazi  $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$  e quelli finito dimensionali  $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$ : mentre su un qualunque spazio di dimensione finita abbiamo appreso ad assegnare la probabilità

tramite opportune FdD (funzioni reali con un numero finito di variabili), è evidente che non è possibile estendere tale costruzione anche al caso di  $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$  in quanto non avrebbe alcun senso assegnare FdD con un numero infinito di variabili. Il problema degli strumenti analitici che ci consentono di assegnare una probabilità su  $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$  non potrà quindi essere affrontato in maniera diretta, ma, come vedremo, dovrà passare attraverso l'uso del teorema di estensione di Carathéodory. In pratica si sfrutterà la possibilità di definire in maniera diretta una probabilità  $\mathbf{P}_0$  sull'algebra  $\mathcal{C}$  dei cilindri con base di dimensione finita, ma arbitraria, per poi estenderla a  $\mathcal{B}(\mathbf{R}^\infty) = \sigma(\mathcal{C})$ . Si scopre però che l'assegnazione di  $\mathbf{P}_0$  non può essere fatta in maniera arbitraria: per risultare effettivamente estendibile  $\mathbf{P}_0$  deve essere costruita tramite l'assegnazione di una **famiglia consistente di spazi di probabilità**  $(\mathbf{R}^{(M)}, \mathcal{B}(\mathbf{R}^{(M)}), \mathbf{P}^{(M)})$  con  $M \in \mathcal{M}$ , dove la consistenza è intesa nel senso descritto prima. Riassumendo possiamo dire che, se una  $\mathbf{P}$  è assegnata su  $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$  è sempre possibile derivarne una famiglia consistente di spazi di probabilità finito dimensionali; viceversa (ma questa seconda affermazione è tutt'altro che banale e richiede la dimostrazione del successivo teorema) una volta assegnata una famiglia consistente di spazi di probabilità finito dimensionali essa definisce in maniera univoca una probabilità su  $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$ .  $\circ$

**II.4.10 Teorema (di Kolmogorov su  $\mathbf{R}^\infty$ ):** Data una famiglia consistente di spazi di probabilità  $(\mathbf{R}^{(M)}, \mathcal{B}(\mathbf{R}^{(M)}), \mathbf{P}^{(M)})$  con  $M \in \mathcal{M}$ , esiste sempre un'unica probabilità  $\mathbf{P}$  su  $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$  tale che

$$\mathbf{P}(C(B^{(M)})) = \mathbf{P}^{(M)}(B^{(M)})$$

comunque assegnato  $B^{(M)} \in \mathcal{B}(\mathbf{R}^{(M)})$  con  $M \in \mathcal{M}$ .

**Dimostrazione:** Se  $(\mathbf{R}^{(M)}, \mathcal{B}(\mathbf{R}^{(M)}), \mathbf{P}^{(M)})$  con  $M \in \mathcal{M}$  è una famiglia consistente, risulterà sempre

$$\mathbf{P}^{(M \cup M')}(B^{(M)} \times \mathbf{R}^{(M')}) = \mathbf{P}^{(M)}(B^{(M)}),$$

con  $M, M' \in \mathcal{M}$  e  $M \cap M' = \emptyset$ . Definiamo allora l'applicazione  $\mathbf{P}_0 : \mathcal{C} \rightarrow [0, 1]$  mediante la posizione

$$\mathbf{P}_0(C(B^{(M)})) = \mathbf{P}^{(M)}(B^{(M)})$$

dove  $C(B^{(M)})$  è il generico cilindro di  $\mathcal{C}$  con base  $B^{(M)} \in \mathcal{B}(\mathbf{R}^{(M)})$  di dimensione finita. Tale  $\mathbf{P}_0$  è innanzitutto ben definita, nel senso che (a causa della proprietà di consistenza ipotizzata) essa assegna ai cilindri di  $\mathcal{C}$  un valore che è indipendente dalla loro rappresentazione. Infatti:

$$\begin{aligned} \mathbf{P}_0(C(B^{(M)} \times \mathbf{R}^{(M')})) &= \mathbf{P}^{(M \cup M')}(B^{(M)} \times \mathbf{R}^{(M')}) = \mathbf{P}^{(M)}(B^{(M)}) \\ &= \mathbf{P}_0(C(B^{(M)})), \end{aligned}$$



dove  $C(B^{(M)} \times \mathbf{R}^{(M')})$  e  $C(B^{(M)})$  risultano essere rappresentazioni diverse del medesimo cilindro.  $\mathbf{P}_0$ , inoltre, risulta essere additiva su  $\mathcal{C}$ : detta infatti  $C(B^{(M_k)})$  con  $k = 1, \dots, n$  una famiglia finita di elementi disgiunti di  $\mathcal{C}$ , con basi  $B^{(M_k)}$  prese in spazi in genere diversi fra loro, potremo sempre modificarne la rappresentazione in modo da avere le basi tutte in un medesimo spazio. Posto infatti

$$M = \bigcup_{k=1}^n M_k \in \mathcal{M}; \quad M^k = M \setminus M_k \in \mathcal{M},$$

si vede subito che

$$C(B^{(M_k)}) = C(B^{(M_k)} \times \mathbf{R}^{(M^k)}) = C(D_k), \quad k = 1, \dots, n$$

dove

$$D_k = B^{(M_k)} \times \mathbf{R}^{(M^k)} \in \mathcal{B}(\mathbf{R}^{(M)}), \quad k = 1, \dots, n$$

sono ora tutti elementi disgiunti del medesimo spazio  $\mathcal{B}(\mathbf{R}^{(M)})$ . Dalle proprietà delle applicazioni inverse segue allora che

$$\bigcup_{k=1}^n C(D_k) = \bigcup_{k=1}^n \pi_M^{-1}(D_k) = \pi_M^{-1}\left(\bigcup_{k=1}^n D_k\right) = C\left(\bigcup_{k=1}^n D_k\right),$$

dove ovviamente  $\bigcup_k D_k \in \mathcal{B}(\mathbf{R}^{(M)})$ . Dalla addittività delle  $\mathbf{P}^{(M)}$  deriva allora che

$$\begin{aligned} \mathbf{P}_0\left(\bigcup_{k=1}^n C(D_k)\right) &= \mathbf{P}_0\left(C\left(\bigcup_{k=1}^n D_k\right)\right) = \mathbf{P}^{(M)}\left(\bigcup_{k=1}^n D_k\right) \\ &= \sum_{k=1}^n \mathbf{P}^{(M)}(D_k) = \sum_{k=1}^n \mathbf{P}_0(C(D_k)), \end{aligned}$$

cioè  $\mathbf{P}_0$  è additiva su  $\mathcal{C}$ . Tralasciando di dimostrare<sup>2</sup> che  $\mathbf{P}_0$  è anche  $\sigma$ -additiva su  $\mathcal{C}$ , ci limiteremo ad osservare in conclusione che la tesi del teorema discende allora dal Teorema II.3.4 che garantisce l'esistenza di un'unica estensione  $\mathbf{P}$  di  $\mathbf{P}_0$  a tutto  $\sigma(\mathcal{C}) = \mathcal{B}(\mathbf{R}^\infty)$ .  $\square$

**II.4.11 Osservazione:** Il teorema precedente può essere esteso anche al caso più generale di uno spazio  $(\prod_n \Omega_n, \otimes_n \mathcal{F}_n)$  prodotto diretto di una famiglia numerabile di spazi probabilizzabili  $(\Omega_n, \mathcal{F}_n)$  con  $n \in \mathbf{N}$ , nel senso che l'assegnazione di una famiglia numerabile e consistente di probabilità  $\mathbf{P}_n$  su ogni  $(\Omega_n, \mathcal{F}_n)$  è sufficiente per definire una probabilità  $\mathbf{P}$  sull'intero spazio prodotto. Naturalmente in questo caso la consistenza è definita da relazioni del tipo

$$\mathbf{P}^{(n+1)}(A^{(n)} \times \Omega_{n+1}) = \mathbf{P}^{(n)}(A^{(n)}); \quad A^{(n)} \in \bigotimes_{k=1}^n \mathcal{F}_k.$$

---

<sup>2</sup> Per i dettagli della dimostrazione vedi ad esempio **A.N. Shiryaev**: *Probability*; Springer, New York, 1984, pp. 161-2.

Comunque si può mostrare, mediante dei controesempi, che gli spazi  $(\Omega_n, \mathcal{F}_n)$  non possono essere del tutto arbitrari. Una condizione abbastanza generale, ma allo stesso tempo sufficiente a garantire la validità di un teorema di estensione di Kolmogorov sullo spazio prodotto diretto è che gli  $\Omega_n$  siano *spazi metrici, completi e separabili*<sup>3</sup>, cioè, secondo la terminologia accettata, degli **spazi polacchi**, e che le  $\mathcal{F}_n$  siano le  $\sigma$ -algebre dei Boreliani degli  $\Omega_n$ .  $\circ$

**II.4.12 Esempio:** Se  $(F_n(x))_{n \in \mathbf{N}}$  è una successione di FdD su  $\mathbf{R}$  possiamo definire un'altra successione di FdD (ciascuna su un  $\mathbf{R}^n$  diverso) ponendo

$$F^{(n)}(x_1, \dots, x_n) = F_1(x_1) \cdot \dots \cdot F_n(x_n); \quad n \in \mathbf{N}$$

(a questo proposito vedi anche l'Esempio II.4.6). Sulla base dei Teoremi II.3.5 e II.4.3 potremo allora costruire due successioni di spazi di probabilità: gli spazi unidimensionali  $(\mathbf{R}_n, \mathcal{B}(\mathbf{R}_n), \mathbf{P}_n)$  con le probabilità associate alle  $F_n$ ; e gli spazi  $n$ -dimensionali  $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n), \mathbf{P}^{(n)})$  con le probabilità associate alle  $F^{(n)}$ . Si può ora far vedere che la successione  $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n), \mathbf{P}^{(n)})$  costituisce una famiglia consistente. Infatti, se  $A_k = (a_k, b_k]$  sono intervalli di  $\mathbf{R}_k$  (con  $k = 1, \dots, n$ ), si vede subito che

$$\mathbf{P}^{(n)}(A_1 \times \dots \times A_n) = \mathbf{P}_1(A_1) \cdot \dots \cdot \mathbf{P}_n(A_n).$$

Infatti, nel caso  $n = 2$ , si ha

$$\begin{aligned} \mathbf{P}^{(2)}((a_1, b_1] \times (a_2, b_2]) &= \Delta_{a_1, b_1} \Delta_{a_2, b_2} F^{(2)}(x_1, x_2) \\ &= F^{(2)}(b_1, b_2) - F^{(2)}(b_1, a_2) - F^{(2)}(a_1, b_2) + F^{(2)}(a_1, a_2) \\ &= [F_1(b_1) - F_1(a_1)] \cdot [F_2(b_2) - F_2(a_2)] \\ &= \mathbf{P}_1(a_1, b_1] \cdot \mathbf{P}_2(a_2, b_2] \end{aligned}$$

dato che  $F^{(2)}(x_1, x_2) = F_1(x_1)F_2(x_2)$ . La precedente relazione si può poi anche estendere al caso in cui i lati del rettangolo siano generici boreliani (e non solo degli intervalli) e finalmente si può provare che

$$\mathbf{P}^{(n+1)}(B^{(n)} \times B_{n+1}) = \mathbf{P}^{(n)}(B^{(n)})\mathbf{P}_{n+1}(B_{n+1})$$

con  $B^{(n)} \in \mathcal{B}(\mathbf{R}^n)$  e  $B_{n+1} \in \mathcal{B}(\mathbf{R}_{n+1})$ . In particolare si ha allora che

$$\mathbf{P}^{(n+1)}(B^{(n)} \times \mathbf{R}_{n+1}) = \mathbf{P}^{(n)}(B^{(n)}); \quad B^{(n)} \in \mathcal{B}(\mathbf{R}^n),,$$

il che mostra che la famiglia di spazi di probabilità da noi costruita è consistente. Esiste pertanto, in base al Teorema II.4.10, un'unica misura di probabilità  $\mathbf{P}$  sullo spazio delle successioni numeriche  $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$  tale che

$$\mathbf{P}(x \in \mathbf{R}^\infty : (x_1, \dots, x_n) \in B^{(n)}) = \mathbf{P}^{(n)}(B^{(n)}).$$

---

<sup>3</sup> Per tutti questi concetti vedi, ad esempio, **A.N. Kolmogorov e S.V. Fomin**: *Elementi di Teoria delle Funzioni e di Analisi Funzionale*; Mir, Mosca, 1980, pp. 61 e 90.

In particolare si ha

$$\mathbf{P}(x \in \mathbf{R}^\infty : x_1 \leq a_1, \dots, x_n \leq a_n) = F_1(a_1) \cdot \dots \cdot F_n(a_n).$$

Come caso particolare possiamo considerare quello in cui tutte le  $F_n$  sono distribuzioni di Bernoulli identiche fra loro:

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

Questa assegnazione è sufficiente a definire una  $\mathbf{P}$  sullo spazio delle successioni  $(x_n)_{n \in \mathbf{N}}$  con  $x_n = 0, 1$  in modo tale che, posto  $p = 1 - q$ , risulti

$$\mathbf{P}(x \in \mathbf{R}^\infty : x_1 = a_1, \dots, x_n = a_n) = p^\nu q^{n-\nu}; \quad \nu = \sum_{k=1}^n a_k.$$

Tale  $\mathbf{P}$  è quindi l'estensione allo spazio delle successioni (infinite) di tentativi di verifica di un evento delle probabilità definite sugli spazi delle  $n$ -ple di tentativi. L'esistenza di tale estensione, ora garantita dal Teorema II.4.10, è proprio ciò che ci serviva nel capitolo I.7 per dare un significato probabilistico preciso alla Legge dei Grandi Numeri (vedi Osservazione I.7.6).  $\diamond$

**II.4.13 Osservazione:** Prendiamo infine in considerazione il caso dello spazio  $(\mathbf{R}^T, \mathcal{B}(\mathbf{R}^T))$  e supponiamo innanzitutto, come nel caso di  $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$ , che su di esso sia stata definita una probabilità  $\mathbf{P}$ . Anche in questo caso, come in II.4.9, la conoscenza di  $\mathbf{P}$  ci permetterà di definire una opportuna famiglia di probabilità  $\mathbf{P}^{(S)}$  sugli spazi di dimensione finita  $(\mathbf{R}^S, \mathcal{B}(\mathbf{R}^S))$  (dove, con notazione ripresa da II.2.17, sarà  $S \in \mathcal{S}$ ). Infatti, siccome  $\mathbf{P}$  è ovviamente definita anche sulla famiglia  $\mathcal{C}$  dei cilindri  $C(B^{(S)})$  con base  $B^{(S)} \in \mathcal{B}(\mathbf{R}^S)$  di dimensione finita ( $S \in \mathcal{S}$ ) ma arbitraria, potremo sempre definire

$$\mathbf{P}^{(S)}(B^{(S)}) = \mathbf{P}(C(B^{(S)})), \quad S \in \mathcal{S}.$$

Si verifica poi (a causa della indipendenza della probabilità che  $\mathbf{P}$  assegna ad un cilindro dalla rappresentazione di quest'ultimo) che la famiglia  $(\mathbf{R}^S, \mathcal{B}(\mathbf{R}^S), \mathbf{P}^{(S)})$  (con  $S \in \mathcal{S}$ ) così costruita è **consistente**, nel senso che, dati  $S_1, S_2 \in \mathcal{S}$  con  $S_1 \cap S_2 = \emptyset$  si ha

$$\mathbf{P}^{(S_1 \cup S_2)}(B^{(S_1)} \times \mathbf{R}^{(S_2)}) = \mathbf{P}(C(B^{(S_1)} \times \mathbf{R}^{(S_2)})) = \mathbf{P}(C(B^{(S_1)})) = \mathbf{P}^{(S_1)}(B^{(S_1)}).$$

Viceversa, partendo da una famiglia consistente di spazi di probabilità ci si chiede se è possibile definire su tutto  $(\mathbf{R}^T, \mathcal{B}(\mathbf{R}^T))$  una  $\mathbf{P}$  che risulti essere il prolungamento delle  $\mathbf{P}^{(S)}$ : ciò consentirebbe di garantire l'esistenza di una probabilità su uno spazio ad infinite dimensioni, come  $(\mathbf{R}^T, \mathcal{B}(\mathbf{R}^T))$ , mediante la definizione di

probabilità su spazi con un numero finito di dimensioni. La risposta al quesito posto è nel seguente teorema.  $\circ$

**II.4.14 Teorema (di Kolmogorov su  $\mathbf{R}^T$ ):** Data una famiglia consistente di spazi di probabilità  $(\mathbf{R}^S, \mathcal{B}(\mathbf{R}^{(S)}), \mathbf{P}^{(S)})$  (con  $S \in \mathcal{S}$ ), esiste sempre un'unica probabilità  $\mathbf{P}$  su  $(\mathbf{R}^T, \mathcal{B}(\mathbf{R}^T))$  tale che

$$\mathbf{P}(C(B^{(S)})) = \mathbf{P}^{(S)}(B^{(S)})$$

comunque assegnato  $B^{(S)} \in \mathcal{B}(\mathbf{R}^{(S)})$  con  $S \in \mathcal{S}$ .

**Dimostrazione:** La dimostrazione di questo teorema è sostanzialmente già contenuta in due proposizioni precedenti: la Proposizione II.2.18 ed il Teorema di Kolmogorov II.4.10. Innanzitutto osserviamo che, sebbene la famiglia di spazi di probabilità  $(\mathbf{R}^S, \mathcal{B}(\mathbf{R}^{(S)}), \mathbf{P}^{(S)})$  sia definita per ipotesi solo per  $S \in \mathcal{S}$  finiti, il Teorema di Kolmogorov II.4.10 ci consente immediatamente di estendere tale definizione anche al caso in cui  $S \in \mathcal{S}'$  è numerabile conservando anche la proprietà di consistenza. Potremo pertanto ritenere che ci sia stata assegnata una famiglia consistente con  $S \in \mathcal{S}'$ . In secondo luogo la Proposizione II.2.18 afferma che  $\mathcal{B}(\mathbf{R}^T) = \mathcal{C}'$ , cioè che ogni elemento  $A \in \mathcal{B}(\mathbf{R}^T)$  non è altro che un cilindro di  $\mathcal{C}'$  con base al più numerabile:

$$A = C(B^{(S)}); \quad B^{(S)} \in \mathcal{B}(\mathbf{R}^{(S)}); \quad S \in \mathcal{S}'.$$

Potremo allora, in modo naturale, definire  $\mathbf{P} : \mathcal{B}(\mathbf{R}^T) \rightarrow [0, 1]$  nel modo seguente:

$$\mathbf{P}(A) = \mathbf{P}(C(B^{(S)})) = \mathbf{P}^{(S)}(B^{(S)}).$$

La  $\mathbf{P}$  così definita è ovviamente un'estensione della famiglia di probabilità inizialmente definita su  $\mathcal{C}$  e pertanto non ci resta che verificare l'unicità di tale estensione. Per soddisfare le ipotesi del Teorema di Carathéodory dovremo allora solo verificare che il valore di  $\mathbf{P}(A)$  non dipende dalla particolare rappresentazione di  $A$  come cilindro, e che  $\mathbf{P}$  è additiva e  $\sigma$ -additiva.

a) Supponiamo che  $A \in \mathcal{B}(\mathbf{R}^T)$  ammetta, come cilindro di  $\mathcal{C}'$ , due distinte rappresentazioni

$$A = \begin{cases} C(B^{(S_1)} = C(B^{(S_1)} \times \mathbf{R}^{(S \setminus S_1)}) \\ C(B^{(S_2)} = C(B^{(S_2)} \times \mathbf{R}^{(S \setminus S_2)}) \end{cases}$$

con  $S_1, S_2 \in \mathcal{S}'$  e  $S = S_1 \cup S_2 \in \mathcal{S}'$ . Siccome due cilindri identici e con basi nello stesso spazio  $\mathcal{B}(\mathbf{R}^{(S)})$  non possono che avere basi coincidenti, risulterà

$$B^{(S_1)} \times \mathbf{R}^{(S \setminus S_1)} = B^{(S_2)} \times \mathbf{R}^{(S \setminus S_2)}.$$

Tenendo allora conto della consistenza della famiglia data si avrà

$$\begin{aligned} \mathbf{P}(C(B^{(S_1)})) &= \mathbf{P}^{(S_1)}(B^{(S_1)}) = \mathbf{P}^{(S)}(B^{(S_1)} \times \mathbf{R}^{(S \setminus S_1)}) \\ &= \mathbf{P}^{(S)}(B^{(S_2)} \times \mathbf{R}^{(S \setminus S_2)}) = \mathbf{P}^{(S_2)}(B^{(S_2)}) = \mathbf{P}(C(B^{(S_2)})) \end{aligned}$$

sicché il valore di  $\mathbf{P}$  non dipende dalla rappresentazione di  $A$ .

- b) Se  $M$  è un insieme finito o numerabile di indici, sia  $(A_k)_{k \in M}$  una famiglia finita o numerabile di elementi disgiunti di  $\mathcal{B}(\mathbf{R}^T)$ , cioè di cilindri con

$$A_k = C(B^{(S_k)}); \quad S_k \in \mathcal{S}'; \quad k \in M.$$

Posto poi  $S = \bigcup_k S_k \in \mathcal{S}'$  (ricorda che unioni numerabili di insiemi numerabili sono ancora insiemi numerabili) definiamo

$$D_k = B^{(S_k)} \times \mathbf{R}^{(S \setminus S_k)} \in \mathcal{B}(\mathbf{R}^{(S)}); \quad k \in M,$$

ed osserviamo che le  $A_k$  possono essere tutte rappresentate come cilindri aventi basi in un unico spazio  $\mathcal{B}(\mathbf{R}^{(S)})$ :

$$A_k = C(D_k); \quad D_k \in \mathcal{B}(\mathbf{R}^{(S)}); \quad k \in M.$$

Siccome le  $A_k$  sono disgiunte, tali dovranno risultare le loro basi e pertanto (ricordando l'additività e la  $\sigma$ -additività delle  $\mathbf{P}^{(S)}$ ) avremo

$$\begin{aligned} \mathbf{P}\left(\bigcup_{k \in M} A_k\right) &= \mathbf{P}\left(\bigcup_{k \in M} C(D_k)\right) = \mathbf{P}\left(C\left(\bigcup_{k \in M} D_k\right)\right) = \mathbf{P}^{(S)}\left(\bigcup_{k \in M} D_k\right) \\ &= \sum_{k \in M} \mathbf{P}^{(S)}(D_k) = \sum_{k \in M} \mathbf{P}(C(D_k)) = \sum_{k \in M} \mathbf{P}(A_k). \end{aligned}$$

Ne segue quindi che  $\mathbf{P}$  è additiva e  $\sigma$ -additiva. □

**II.4.15 Osservazione:** Il risultato precedente si estende anche ad arbitrarie famiglie  $(\Omega_t)_{t \in T}$  di *Spazi Polacchi* (vedi a questo proposito l'Osservazione II.4.11) munite delle  $\sigma$ -algebre dei loro boreliani  $(\mathcal{F}_t)_{t \in T}$ , nel senso che, data una famiglia di spazi probabilizzabili  $(\Omega_t, \mathcal{F}_t)_{t \in T}$  e costruita la famiglia degli spazi prodotto di dimensione finita  $(\Omega^{(S)}, \mathcal{F}^{(S)})_{S \in \mathcal{S}}$  con

$$\Omega^{(S)} = \prod_{t \in S} \Omega_t; \quad \mathcal{F}^{(S)} = \bigotimes_{t \in S} \mathcal{F}_t;$$

si assegna una famiglia  $(\mathbf{P}^{(S)})_{S \in \mathcal{S}}$  di probabilità consistenti, cioè tali che, se prendiamo  $S_1, S_2 \in \mathcal{S}$  con  $S_1 \cap S_2 = \emptyset$ ,

$$\mathbf{P}^{(S_1 \cup S_2)}(A^{(S_1)} \times \Omega^{(S_2)}) = \mathbf{P}^{(S_1)}(A^{(S_1)}),$$

e si dimostra che esiste sempre un'unica probabilità  $\mathbf{P}$  su  $(\Omega, \mathcal{F})$  con

$$\Omega = \prod_{t \in T} \Omega_t; \quad \mathcal{F} = \bigotimes_{t \in T} \mathcal{F}_t,$$

che risulta essere il prolungamento delle  $\mathbf{P}^{(S)}$ . Questo importante risultato è noto in letteratura sotto il nome di **Teorema di Kolmogorov e Bochner**<sup>4</sup>. Notiamo infine che in tutta la costruzione qui effettuata gli insiemi finiti o numerabili di indici  $S = \{t_1, \dots, t_k, \dots\}$  non sono ordinati nel senso che l'ordine dei punti  $t_k$  su cui vengono imposte le restrizioni alle funzioni  $(x_t)_{t \in T}$  è del tutto indifferente. Per questo motivo, senza perdita di generalità, potremo assumere d'ora in poi che le probabilità finito-dimensionali  $\mathbf{P}^{(S)}$  vengano definite su insiemi di indici ordinati in modo che  $t_1 < \dots < t_k < \dots$ .  $\circ$

**II.4.16 Esempio:** Consideriamo il caso in cui  $T = [0, +\infty)$ , sicché  $\mathbf{R}^T$  risulterà essere l'insieme delle funzioni  $(x_t)_{t \geq 0}$ , e mostiamo come è possibile definire un importante esempio di misura su  $(\mathbf{R}^{[0,+\infty)}, \mathcal{B}(\mathbf{R}^{[0,+\infty)}))$ . Consideriamo a questo scopo la seguente famiglia di fdd Gaussiane

$$\varphi_t(x|x_0) = \frac{1}{\sqrt{2\pi t}} e^{-(x-x_0)^2/2t}; \quad t > 0,$$

dove  $x_0$  è una costante reale. Detto ora  $S = \{t_1, \dots, t_s\} \in \mathcal{S}$  con  $0 < t_1 < \dots < t_s$  e  $A^{(S)} = A_1 \times \dots \times A_s \in \mathcal{I}^{(S)}$  un generico rettangolo con lati costituiti da intervalli definiamo la seguente probabilità su  $\mathcal{I}^{(S)}$ :

$$\mathbf{P}^{(S)}(A^{(S)}) = \int_{A_1} \dots \int_{A_s} \varphi_{t_1}(x_1|0) \varphi_{t_2-t_1}(x_2|x_1) \dots \varphi_{t_s-t_{s-1}}(x_s|x_{s-1}) dx_1 \dots dx_s,$$

che si rivela essere anche  $\sigma$ -additiva. Il teorema di Carathéodory ci consente allora di estendere  $\mathbf{P}^{(S)}$  a tutto  $\mathcal{B}(\mathbf{R}^{(S)})$ : si verifica che la famiglia  $(\mathbf{R}^S, \mathcal{B}(\mathbf{R}^{(S)}), \mathbf{P}^{(S)})$  (con  $S \in \mathcal{S}$ ) così costruita è consistente e quindi il Teorema II.4.14 ci garantisce l'esistenza di una probabilità  $\mathbf{P}$  definita su tutto  $(\mathbf{R}^{[0,+\infty)}, \mathcal{B}(\mathbf{R}^{[0,+\infty)}))$  in modo che sui cilindri risulti

$$\mathbf{P}(C(A^{(S)})) = \mathbf{P}^{(S)}(A^{(S)}).$$

La misura di probabilità così definita prende il nome di **misura di Wiener**<sup>5</sup> e gioca un ruolo molto importante nella teoria dei processi stocastici. Se, ad esempio, interpretiamo  $(x_t)_{t \geq 0}$  come la traiettoria (aleatoria) di un punto materiale, il

---

<sup>4</sup> L'enunciato e la dimostrazione di questo teorema possono essere trovati in **M. Métivier:** *Notions Fondamentales de la Théorie des Probabilités*; Dunod, Paris, 1972, p. 230. Salomon Bochner (1899 - 1982) cominciò ad interessarsi di matematica fin da ragazzo quando frequentava la scuola a Cracovia, sua città natale, che era allora sotto l'amministrazione austriaca. Studiò poi matematica a Berlino dove si laureò nel 1921 con E. Schmidt (1876 - 1959) ed I. Schur (1875 - 1941). Diede fondamentali contributi nello studio delle funzioni quasi-periodiche e dell'integrale di Fourier a proposito del quale scrisse il suo primo libro (*Fouriersche Integrale*, 1932) durante il suo soggiorno a Monaco che durò dal 1926 al 1933. In questo periodo ebbe modo di entrare in contatto con C. Carathéodory (1873 - 1950) che lo spinse ad occuparsi di funzioni di più variabili complesse e di geometria differenziale. Quando Hitler prese il potere nel 1933 Bochner emigrò negli Stati Uniti dove si stabilì definitivamente come professore prima a Princeton e poi alla Rice University.

<sup>5</sup> Il padre di Norbert Wiener (1894 - 1964), Leo, nato a Bielostock in Russia, era emigrato all'età di diciotto anni negli USA dove aveva fatto l'operaio, l'agricoltore, il venditore ambu-

cilindro  $C(A^{(S)})$  con base rettangolare rappresenterà il fascio di traiettorie che, agli istanti  $t_1 < \dots < t_s$ , passano attraverso le finestre  $A_1, \dots, A_s$ . La  $\mathbf{P}^{(S)}$  definita in precedenza assegna allora una probabilità a tali fasci di traiettorie in una maniera che può essere intuitivamente descritta in questo modo:  $\varphi_{t_k - t_{k-1}}(x_k | x_{k-1}) dx_k$  rappresenta la probabilità che la particella, partendo da  $x_{k-1}$  all'istante  $t_{k-1}$ , giunga in  $[x_k, x_k + dx_k]$  dopo un tempo  $t_k - t_{k-1}$ . In tal caso l'uso del prodotto delle fdd nella definizione di  $\mathbf{P}^{(S)}$  esprime l'ipotesi di indipendenza degli spostamenti della particella nei successivi intervalli di tempo  $[0, t_1], [t_1, t_2], \dots, [t_{s-1}, t_s]$ .

◇

---

lante, l'insegnante ed il docente di lingue straniere all'Università del Missouri. Dopo la nascita di Norbert, Leo divenne infine professore di lingue slave all'Università di Harvard. La personalità eccentrica del padre ebbe una fortissima influenza su Norbert che, assai precoce, leggeva i libri scientifici che trovava nella biblioteca del padre. A quindici anni cominciò i suoi studi di biologia a Harvard ma senza grande successo. Si avvicinò quindi alla filosofia e, attraverso questa, alla matematica. Dopo il dottorato conseguito con una tesi in logica matematica, Wiener lavorò a Cambridge sotto la direzione di B. Russell (1872 - 1970) che esercitò su di lui una grande influenza spingendolo a studiare matematica. Tornato negli USA ottenne nel 1919 un posto al Dipartimento di Matematica del MIT dove sviluppò i suoi primi importanti lavori sull'integrazione negli spazi funzionali: partendo dal modello di moto Browniano studiato da Einstein, Wiener sviluppò l'idea fondamentale di uno spazio di probabilità i cui punti sono funzioni. La moderna teoria dei processi stocastici ha origine da questi suoi studi. Wiener si dedicò successivamente a problemi di teoria del potenziale e di analisi armonica generalizzata in cui ottenne ancora risultati fondamentali. Se indicò nuove vie nella matematica pura, egli mantenne però anche uno stretto contatto con le scienze applicate e l'ingegneria. Durante la Seconda Guerra Mondiale lavorò alla progettazione di un meccanismo che potesse predire la posizione di un aereo, in modo da dirigere il tiro dell'artiglieria. Fu questo lo stimolo originale per i grandi successivi sviluppi della Teoria della Predizione e della Cibernetica. I problemi di controllo che si presentavano nello studio dell'operatore umano che muoveva un pezzo di artiglieria portarono Wiener a considerare questioni di neurofisiologia alla luce della teoria delle comunicazioni: fu l'applicazione di queste teorie ai processi biologici ed all'ingegneria che egli chiamò *cibernetica*, ossia teoria del *controllo e della comunicazione nell'animale e nella macchina*. Questi suoi lavori lo resero presto celebre, anche tra i non specialisti, soprattutto a causa delle sue doti di divulgatore e di unificatore di teorie già esistenti.





## II.5 Variabili aleatorie (I parte)

**II.5.1 Definizione:** Dati gli spazi probabilizzabili  $(\Omega, \mathcal{F})$  ed  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ , una funzione  $\xi : \Omega \rightarrow \mathbf{R}$  si dice  **$\mathcal{F}$ -misurabile** se

$$\xi^{-1}(B) = \{\xi \in B\} = \{\omega \in \Omega : \xi(\omega) \in B\} \in \mathcal{F}, \quad \forall B \in \mathcal{B}(\mathbf{R}).$$

Una funzione misurabile si dice anche **variabile aleatoria** (v.a.).  $\triangle$

**II.5.2 Osservazione:** Spesso, per indicare le  $\sigma$ -algebre rispetto alle quali la v.a. è misurabile, si usa scrivere

$$\xi : (\Omega, \mathcal{F}) \rightarrow (\mathbf{R}, \mathcal{B}(\mathbf{R})).$$

Infatti è evidente che la conoscenza delle  $\sigma$ -algebre  $\mathcal{F}$  e  $\mathcal{B}(\mathbf{R})$  è essenziale per poter decidere sulla misurabilità di una data funzione  $\xi : \Omega \rightarrow \mathbf{R}$ . Inoltre quando  $(\Omega, \mathcal{F}) = (\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$  con  $n \in \mathbf{N}$  le v.a. prendono anche il nome di **funzioni di Borel**. Ricordiamo anche che, ovviamente, il caso più semplice di v.a. è quello degli **indicatori**  $I_A(\omega)$  con  $A \in \mathcal{F}$ : è semplice, infatti, verificare che si tratta di funzioni misurabili rispetto ad  $\mathcal{F}$ .  $\circ$

**II.5.3 Osservazione:** Il concetto di misurabilità è fondamentale perché, data una probabilità  $\mathbf{P}$  su  $(\Omega, \mathcal{F})$ , consente di indurre delle nuove probabilità su  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$  definite tramite la v.a.  $\xi$ , e questo è reso possibile, come si vede dalla seguente Definizione, proprio dal fatto che  $\{\xi \in B\} \in \mathcal{F}$ ,  $\forall B \in \mathcal{B}(\mathbf{R})$ .  $\circ$

**II.5.4 Definizione:** Dato uno spazio di probabilità  $(\Omega, \mathcal{F}, \mathbf{P})$  ed una v.a.  $\xi : (\Omega, \mathcal{F}) \rightarrow (\mathbf{R}, \mathcal{B}(\mathbf{R}))$ , si dice **distribuzione di probabilità** (DdP) di  $\xi$  la misura di probabilità  $\mathbf{P}_\xi$  su  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$  indotta da  $\xi$  secondo la relazione

$$\mathbf{P}_\xi(B) = \mathbf{P}\{\xi \in B\}; \quad B \in \mathcal{B}(\mathbf{R}).$$

Si dice anche che  $\mathbf{P}_\xi$  è l'**immagine** di  $\mathbf{P}$  indotta da  $\xi$  su  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ . Inoltre la FdD  $F_\xi(x)$  di  $\mathbf{P}_\xi$ , cioè la funzione

$$F_\xi(x) = \mathbf{P}_\xi(-\infty, x] = \mathbf{P}\{\xi \leq x\} = \mathbf{P}\{\omega \in \Omega : \xi(\omega) \leq x\}; \quad x \in \mathbf{R},$$

prende il nome di **funzione di distribuzione** (FdD) di  $\xi$ .  $\triangle$

**II.5.5 Osservazione:** Le v.a. possono essere classificate secondo il tipo della loro distribuzione. Si dicono v.a. **discrete** quelle del tipo

$$\xi(\omega) = \sum_k x_k I_{A_k}(\omega),$$

dove le  $A_k$  costituiscono una decomposizione (finita o numerabile: vedi Definizione I.2.10) di  $\Omega$  in elementi di  $\mathcal{F}$ : per queste v.a. è facile infatti vedere che la DdP  $\mathbf{P}_\xi$  è

una misura di probabilità discreta (nel senso della Definizione II.3.10) concentrata su un insieme, al più numerabile, di punti  $x_k \in \mathbf{R}$ . In questo caso, posto

$$p_k = \mathbf{P}_\xi\{x_k\} = \mathbf{P}\{\xi = x_k\} = F_\xi(x_k) - F_\xi(x_k^-),$$

risulta

$$\mathbf{P}_\xi(B) = \sum_{k: x_k \in B} p_k; \quad B \in \mathcal{B}(\mathbf{R}).$$

Fra le v.a. discrete assumono poi un ruolo importante le cosiddette v.a. **semplici** che sono caratterizzate dal fatto di prendere solo un numero finito  $n$  di valori  $x_k$  e sono quindi della forma

$$\xi(\omega) = \sum_{k=1}^n x_k I_{A_k}(\omega),$$

dove le  $A_k$  costituiscono ora una decomposizione finita di  $\Omega$ . Una v.a. si dice invece **continua** se la FdD  $F_\xi(x)$  è una funzione continua di  $x$  (le v.a. discrete, dunque, non sono continue). Inoltre, come già osservato in precedenza (Definizione II.3.16 ed Osservazione II.3.18), un caso particolarmente rilevante di v.a. continue è quello delle v.a. **assolutamente continue** (a.c.) che sono caratterizzate dal fatto che  $F_\xi$  è a.c., cioè dall'esistenza di una **funzione di densità** (fdd), cioè di una funzione non negativa  $f_\xi(x)$  e tale che

$$F_\xi(x) = \int_{-\infty}^x f_\xi(y) dy, \quad x \in \mathbf{R},$$

dove l'integrale è inteso nel senso di Lebesgue. Naturalmente, dove  $F_\xi$  è derivabile, sussiste anche la relazione  $f_\xi(x) = F'_\xi(x)$ . Ricordiamo qui, comunque, che non tutte le v.a. continue sono anche a.c.: esistono infatti v.a. con FdD continue e singolari (vedi Definizione II.3.17). Nel seguito, però, non avremo quasi mai a che fare con il caso singolare, anzi va segnalato che spesso, nella letteratura di carattere applicativo, con il nome di v.a. continue si indicano le v.a. a.c., quelle caratterizzate, cioè, dall'esistenza di una fdd  $f_\xi(x)$ . Ovviamente tutte le osservazioni e gli esempi relativi alle FdD ed alle fdd citati in II.3 possono essere estesi in modo naturale alle FdD e fdd di v.a. ○

**II.5.6 Osservazione:** È facile verificare, dalla definizione di misurabilità, che se la funzione  $\xi : (\Omega, \mathcal{F}) \rightarrow (\mathbf{R}, \mathcal{B}(\mathbf{R}))$  è  $\mathcal{F}$ -misurabile essa è anche misurabile rispetto a qualunque altra  $\sigma$ -algebra  $\mathcal{F}''$  tale che  $\mathcal{F} \subseteq \mathcal{F}''$ . Viceversa niente garantisce che essa sia misurabile rispetto ad una  $\sigma$ -algebra  $\mathcal{F}' \subseteq \mathcal{F}$ . Come vedremo in seguito è anche possibile caratterizzare la più piccola (per inclusione)  $\sigma$ -algebra rispetto alla quale  $\xi$  risulta misurabile. Notiamo però a questo proposito che stabilire se  $\xi(\omega)$  è o meno  $\mathcal{F}$ -misurabile può non essere sempre un compito facile: per questo motivo il seguente lemma può risultare particolarmente utile. ○

**II.5.7 Lemma:** Dato uno spazio probabilizzabile  $(\Omega, \mathcal{F})$  e detta  $\mathcal{E}$  una famiglia di parti di  $\mathbf{R}$  tale che  $\sigma(\mathcal{E}) = \mathcal{B}(\mathbf{R})$ , una funzione  $\xi : \Omega \rightarrow \mathbf{R}$  risulta  $\mathcal{F}$ -misurabile se e solo se  $\xi^{-1}(E) \in \mathcal{F}$  comunque scelto  $E \in \mathcal{E}$ .

**Dimostrazione:** Se  $\xi$  è  $\mathcal{F}$ -misurabile si ha ovviamente che  $\xi^{-1}(E) \in \mathcal{F}$ ,  $\forall E \in \mathcal{E}$  in quanto  $\mathcal{E} \subseteq \sigma(\mathcal{E}) = \mathcal{B}(\mathbf{R})$ . Viceversa, se  $\xi^{-1}(E) \in \mathcal{F}$ ,  $\forall E \in \mathcal{E}$  per provare che  $\xi$  è  $\mathcal{F}$ -misurabile si userà in Principio degli insiemi appropriati: detta  $\mathcal{D}$  la famiglia di elementi di  $\mathcal{B}(\mathbf{R})$  che gode della proprietà richiesta

$$\mathcal{D} = \{D \in \mathcal{B}(\mathbf{R}) : \xi^{-1}(D) \in \mathcal{F}\} \subseteq \mathcal{B}(\mathbf{R}),$$

la dimostrazione si riduce a provare che  $\mathcal{D} = \mathcal{B}(\mathbf{R})$ . Per far questo osserviamo che

- a.  $\mathcal{D}$  è una  $\sigma$ -algebra: ciò discende direttamente dalle proprietà delle applicazioni inverse fra insiemi secondo le quali

$$\begin{aligned} \xi^{-1}\left(\bigcup_{\alpha} B_{\alpha}\right) &= \bigcup_{\alpha} \xi^{-1}(B_{\alpha}) \\ \xi^{-1}\left(\bigcap_{\alpha} B_{\alpha}\right) &= \bigcap_{\alpha} \xi^{-1}(B_{\alpha}) \\ \overline{\xi^{-1}(B)} &= \xi^{-1}(\overline{B}) \end{aligned}$$

con  $\alpha$  appartenente ad un arbitrario insieme di indici e con  $B_{\alpha}$  e  $B$  arbitrarie parti di  $\mathbf{R}$ .

- b.  $\mathcal{E} \subseteq \mathcal{D} \subseteq \mathcal{B}(\mathbf{R})$ : ciò deriva banalmente dal fatto che gli elementi di  $\mathcal{E}$  soddisfano per ipotesi la proprietà che definisce  $\mathcal{D}$  e che  $\mathcal{D} \subseteq \mathcal{B}(\mathbf{R})$  per definizione. Otteniamo quindi che  $\sigma(\mathcal{E}) \subseteq \sigma(\mathcal{D}) = \mathcal{D} \subseteq \mathcal{B}(\mathbf{R})$  e siccome per ipotesi  $\sigma(\mathcal{E}) = \mathcal{B}(\mathbf{R})$  risulta anche  $\mathcal{D} = \mathcal{B}(\mathbf{R})$ . □

**II.5.8 Corollario:** Nelle ipotesi del Lemma II.5.7, una v.a.  $\xi$  è  $\mathcal{F}$ -misurabile se e solo se  $\{\xi < x\} \in \mathcal{F}$ ,  $\forall x \in \mathbf{R}$  oppure  $\{\xi \leq x\} \in \mathcal{F}$ ,  $\forall x \in \mathbf{R}$ .

**Dimostrazione:** Basterà osservare che ambedue le famiglie di parti di  $\mathbf{R}$

$$\mathcal{E} = \{(-\infty, x)\}_{x \in \mathbf{R}}, \quad \mathcal{E} = \{(-\infty, x]\}_{x \in \mathbf{R}}$$

generano la  $\sigma$ -algebra  $\mathcal{B}(\mathbf{R})$ . □

**II.5.9 Lemma:** Date una funzione di Borel  $\phi(x)$  ( $x \in \mathbf{R}$ ) ed una v.a.  $\xi(\omega)$  ( $\omega \in \Omega$ ), la funzione composta

$$(\phi \circ \xi)(\omega) = \phi(\xi(\omega)) = \eta(\omega)$$

è ancora una v.a.

**Dimostrazione:** Dal fatto che

$$\{\eta \in B\} = \{\phi(\xi) \in B\} = \{\xi \in \phi^{-1}(B)\} = \xi^{-1}(\phi^{-1}(B))$$

si ottiene la tesi osservando che, essendo  $\phi$  e  $\xi$  misurabili, risulta  $\phi^{-1}(B) \in \mathcal{B}(\mathbf{R})$  e  $\xi^{-1}(\phi^{-1}(B)) \in \mathcal{F}$ , se  $B \in \mathcal{B}(\mathbf{R})$ .  $\square$

**II.5.10 Esempio:** Il lemma precedente consente di provare che, se  $\xi$  è una v.a., anche

$$\xi^n, \quad |\xi|, \quad \xi^+ = \max\{\xi, 0\}, \quad \xi^- = -\min\{\xi, 0\}$$

sono delle v.a. Nota che, naturalmente, ogni v.a.  $\xi$  è suscettibile di essere rappresentata come  $\xi = \xi^+ - \xi^-$ , dove  $\xi^+$  e  $\xi^-$  non sono mai entrambe diverse da zero sicché  $|\xi| = \xi^+ + \xi^-$ .  $\diamond$

**II.5.11 Osservazione:** Data una successione di v.a.  $(\xi_n)_{n \in \mathbf{N}}$  ci si può porre il problema di costruire altre v.a. del tipo

$$\lim_n \xi_n, \quad \overline{\lim}_n \xi_n, \quad \underline{\lim}_n \xi_n, \quad \sum_n \xi_n, \quad \sum_n |\xi_n|,$$

ma per farlo correttamente bisognerà affrontare le seguenti questioni:

- a) siccome le  $\xi_n(\omega)$  sono funzioni e non numeri, bisogna definire in che senso si intende la convergenza dei limiti eseguiti;
- b) se vogliamo studiare la possibilità che i risultati di queste operazioni siano nuove v.a., dovremo anche considerare la possibilità che i limiti eseguiti siano divergenti.

Dovremo dunque allargare la nostra definizione di v.a. per poter studiare, nei teoremi successivi, i seguenti due problemi:

1. se ogni v.a. possa essere considerata come limite di una opportuna successione di v.a.;
2. se ogni successione di v.a. che ammette limite converga verso una nuova v.a.

Prima di affrontare tali questioni vogliamo solo ricordare che il concetto di convergenza che qui introdurremo è solo il primo fra quelli che tratteremo in queste lezioni: infatti, siccome si tratta di limiti di funzioni, introdurremo diversi, e non equivalenti, modi di convergenza che saranno discussi in seguito.  $\circ$

**II.5.12 Definizione:** Una funzione misurabile  $\xi : (\Omega, \mathcal{F}) \rightarrow (\overline{\mathbf{R}}, \mathcal{B}(\overline{\mathbf{R}}))$ , dove con  $\overline{\mathbf{R}}$  si indica la retta reale estesa, dicesi **variabile aleatoria estesa**.  $\triangle$

**II.5.13 Definizione:** Diremo che una successione di v.a. estese  $(\xi_n)_{n \in \mathbf{N}}$  **converge puntualmente** se tutte le successioni numeriche (in  $\overline{\mathbf{R}}$ )  $(\xi_n(\omega))_{n \in \mathbf{N}}$  convergono (in  $\overline{\mathbf{R}}$ ) qualunque sia  $\omega \in \Omega$ .  $\triangle$

**II.5.14 Proposizione:** Se  $(\xi_n)_{n \in \mathbf{N}}$  è una successione di v.a. estese, allora anche

$$\sup_n \xi_n, \quad \inf_n \xi_n, \quad \overline{\lim}_n \xi_n, \quad \underline{\lim}_n \xi_n$$

sono v.a. estese.

**Dimostrazione:** Per dimostrare la proposizione basterà far appello al Corollario II.5.8 che è valido anche nel caso di v.a. estese e quindi basterà provare, ad esempio, che

$$\{\sup_n \xi_n \leq x\} \in \mathcal{F}, \quad \forall x \in \overline{\mathbf{R}}.$$

Nel caso in cui  $x \neq +\infty$  si ha che

$$\{\sup_n \xi_n \leq x\} = \overline{\{\sup_n \xi_n > x\}},$$

e il risultato discende dal fatto che

$$\{\sup_n \xi_n > x\} = \{\omega : \exists n \in \mathbf{N} \exists' \xi_n(\omega) > x\} = \bigcup_n \{\xi_n > x\} \in \mathcal{F};$$

infatti  $\mathcal{F}$  è una  $\sigma$ -algebra e  $\{\xi_n > x\}$  sono tutti suoi elementi. Se invece  $x = +\infty$  basterà osservare che, essendo  $\mathcal{F}$  una  $\sigma$ -algebra, risulta

$$\begin{aligned} \{\sup_n \xi_n = +\infty\} &= \{\omega : \forall m \in \mathbf{N} \exists n \in \mathbf{N} \exists' \xi_n(\omega) > m\} \\ &= \bigcap_m \bigcup_n \{\xi_n > m\} \in \mathcal{F}. \end{aligned}$$

Analogamente si ragiona per gli estremi inferiori notando che

$$\{\inf_n \xi_n \leq x\} = \bigcup_n \{\xi_n \leq x\} \in \mathcal{F}.$$

Infine dalle definizioni usuali

$$\overline{\lim}_n \xi_n = \inf_n \sup_{k \geq n} \xi_k \quad \underline{\lim}_n \xi_n = \sup_n \inf_{k \geq n} \xi_k$$

e dalla dimostrazione precedente si ricava immediatamente anche il risultato sul massimo e minimo limite.  $\square$

**II.5.15 Teorema:** Se la successione di v.a. estese  $(\xi_n)_{n \in \mathbf{N}}$  converge puntualmente verso  $\xi(\omega)$ , allora anche  $\xi$  è una v.a. estesa.

**Dimostrazione:** Se, comunque scelto  $\omega \in \Omega$ , la successione numerica in  $\overline{\mathbf{R}}$   $(\xi_n(\omega))_{n \in \mathbf{N}}$  converge verso il numero  $\xi(\omega) \in \overline{\mathbf{R}}$ , avremo

$$\xi(\omega) = \lim_n \xi_n(\omega) = \overline{\lim}_n \xi_n(\omega)$$

e quindi, in base alla Proposizione II.5.14,

$$\{\xi \leq x\} = \{\overline{\lim}_n \xi_n \leq x\} \in \mathcal{F} \quad \forall x \in \overline{\mathbf{R}}$$

per cui il Corollario II.5.8 ci garantisce che  $\xi$  è una v.a. estesa. □

**II.5.16 Teorema (di Lebesgue):** Se  $\xi$  è una v.a. estesa non negativa ( $\xi \geq 0$ ), è sempre possibile determinare una successione  $(\xi_n)_{n \in \mathbf{N}}$  di v.a. semplici tale che

$$0 \leq \xi_n \leq \xi_{n+1} \leq \xi, \quad \forall \omega \in \Omega, \quad \forall n \in \mathbf{N}$$

e tale che  $\xi_n \uparrow \xi$  nel senso della convergenza puntuale. Se invece  $\xi$  è una generica v.a. estesa è sempre possibile determinare una successione  $(\xi_n)_{n \in \mathbf{N}}$  di v.a. semplici tale che

$$|\xi_n| \leq |\xi|, \quad \forall \omega \in \Omega, \quad \forall n \in \mathbf{N}$$

e tale che  $\xi \xrightarrow{n} \xi$  nel senso della convergenza puntuale.

**Dimostrazione:** La dimostrazione di questo teorema, che, malgrado la sua semplicità, costituisce un passaggio essenziale per gli sviluppi successivi, si basa sulla costruzione diretta della successione di v.a. semplici che approssima la data v.a. estesa  $\xi$ . Notiamo innanzitutto che la seconda parte del teorema (quella in cui non si fa l'ipotesi che  $\xi$  sia non negativa) non è altro che un corollario della prima: infatti, con le definizioni dell'Esempio II.5.10, sappiamo che ogni  $\xi$  è suscettibile di essere rappresentata nella forma  $\xi = \xi^+ - \xi^-$ , dove  $\xi^+$  e  $\xi^-$  sono v.a. non negative per le quali vale la prima parte del teorema. Pertanto sarà sempre possibile determinare due successioni  $(\xi_n^\pm)_{n \in \mathbf{N}}$  di v.a. semplici, non negative, non decrescenti e tali che  $\xi_n^\pm \uparrow \xi^\pm$  nel senso della convergenza puntuale. Conseguentemente potremo sempre definire la successione  $\xi_n = \xi_n^+ - \xi_n^-$  (per la quale si ha ovviamente  $|\xi_n| = \xi_n^+ + \xi_n^- \leq \xi^+ + \xi^- = |\xi|$ ) che chiaramente convergerà puntualmente verso  $\xi = \xi^+ - \xi^-$ .

Non resta quindi che dimostrare la prima parte del teorema: siccome la v.a. estesa  $\xi$  prende ora valori in  $[0, +\infty]$  converrà, per ogni  $n \in \mathbf{N}$  fissato, decomporre l'insieme  $[0, +\infty]$  nei seguenti  $n2^n + 1$  intervalli disgiunti

$$\left[0, \frac{1}{2^n}\right), \left[\frac{1}{2^n}, \frac{2}{2^n}\right), \dots, \left[\frac{n2^n - 1}{2^n}, n\right); [n, +\infty]$$

i primi  $n2^n$  dei quali di ampiezza eguale  $1/2^n$  e l'ultimo illimitato. Siccome  $\xi$  è misurabile e il suo valore deve cadere in uno dei suddetti intervalli, potremo decomporre  $\Omega$  nei seguenti  $n2^n + 1$  eventi disgiunti di  $\mathcal{F}$

$$B_k^{(n)} = \left\{ \xi \in \left[ \frac{k-1}{2^n}, \frac{k}{2^n} \right) \right\}; \quad k = 1, 2, \dots, n2^n;$$

$$B^{(n)} = \{ \xi \in [n, +\infty) \},$$

i cui indicatori saranno  $I_k^{(n)}$  e  $I^{(n)}$ . Possiamo quindi definire la seguente successione di v.a. (evidentemente semplici e non negative)

$$\xi_n = \sum_{k:1}^{n2^n} \frac{k-1}{2^n} I_k^{(n)} + nI^{(n)}$$

che, come ora proveremo, risulta non decrescente e convergente (in senso puntuale) verso  $\xi$ . Innanzitutto è facile verificare che, per ogni  $n \in \mathbf{N}$ , le  $\xi_n$  assumono i seguenti valori:

$$\xi_n(\omega) = \begin{cases} 0, & \text{se } \omega \in B_1^{(n)}, \text{ cioè se } \xi \in [0, 1/2^n); \\ 1/2^n, & \text{se } \omega \in B_2^{(n)}, \text{ cioè se } \xi \in [1/2^n, 2/2^n); \\ \dots & \dots \\ (n2^n - 1)/2^n, & \text{se } \omega \in B_{n2^n}^{(n)}, \text{ cioè se } \xi \in [(n2^n - 1)/2^n, n); \\ n, & \text{se } \omega \in B^{(n)}, \text{ cioè se } \xi \in [n, +\infty); \end{cases}$$

cioè, su ogni  $B_k^{(n)}$  e  $B^{(n)}$ , esse assumono un valore minore o eguale a tutti i possibili valori di  $\xi$  (l'estremo sinistro di ogni intervallo). Ne segue quindi che  $\xi_n \leq \xi$  comunque scelti  $n \in \mathbf{N}$  ed  $\omega \in \Omega$ . Inoltre risulta sempre  $\xi_{n+1} \geq \xi_n$  (comunque scelti  $n$  ed  $\omega$ ) dato che il passaggio da  $n$  ad  $n+1$  si realizza sempre mediante una suddivisione di  $[0, +\infty)$  consistente solo nell'aggiunta di nuovi punti ai precedenti, senza modificare la collocazione di questi ultimi. Così ad esempio si ha per gli intervalli finiti

$$\left[ \frac{k-1}{2^n}, \frac{k}{2^n} \right) = \left[ \frac{2k-2}{2^{n+1}}, \frac{2k-1}{2^{n+1}} \right) \cup \left[ \frac{2k-1}{2^{n+1}}, \frac{2k}{2^{n+1}} \right)$$

e per quelli illimitati

$$\begin{aligned} [n, +\infty) &= [n, n+1) \cup [n+1, +\infty) \\ &= \bigcup_{l:1}^{2^{n+1}} \left[ \frac{n2^{n+1} + l - 1}{2^{n+1}}, \frac{n2^{n+1} + l}{2^{n+1}} \right) \cup [n+1, +\infty). \end{aligned}$$

Basterà allora ricordare che, su ogni evento di una decomposizione di  $\Omega$ ,  $\xi_n$  e  $\xi_{n+1}$  assumono per definizione i valori dell'estremo sinistro del corrispondente intervallo di  $\mathbf{R}$  per ottenere che  $\xi_{n+1} \geq \xi_n$  (comunque scelti  $n$  ed  $\omega$ ). Resta quindi

da dimostrare che la nostra successione di v.a. semplici, non negative e non decrescenti converge puntualmente verso  $\xi$ . Preso allora un arbitrario  $\omega \in \Omega$  dobbiamo distinguere i due casi  $\xi(\omega) = +\infty$  e  $\xi(\omega) < +\infty$ : nella prima eventualità dovremo solo mostrare che  $(\xi_n(\omega))_{n \in \mathbf{N}}$  diverge; ma questo è evidente perché, se  $\xi(\omega) = +\infty$ , risulterà anche  $\xi(\omega) \in [n, +\infty]$ ,  $\forall n \in \mathbf{N}$  e quindi, dalle definizioni,  $\xi_n(\omega) = n$ ,  $\forall n \in \mathbf{N}$ , da cui risulta facilmente che  $\lim_n \xi_n(\omega) = +\infty$ . Se invece  $\xi(\omega) < +\infty$  potremo sempre trovare un  $\nu \in \mathbf{N}$  tale che  $\xi(\omega) < \nu$  e un  $k \in \{1, 2, \dots, \nu 2^\nu\}$  tale che

$$\xi(\omega) \in \left[ \frac{k-1}{2^\nu}, \frac{k}{2^\nu} \right).$$

Notiamo inoltre che con le scelte fatte si ha  $\xi_\nu(\omega) = (k-1)/2^\nu$ . Dato ora un arbitrario  $\epsilon > 0$  scegliamo  $\nu \in \mathbf{N}$  in modo da verificare le due relazioni  $\xi(\omega) < \nu$  (con  $\omega$  arbitrario ma fissato) e  $1/2^\nu < \epsilon$ . Se  $k$  viene poi scelto come descritto in precedenza si ottiene che  $\xi(\omega)$  cade in un intervallo di ampiezza  $1/2^\nu < \epsilon$  e di estremo sinistro  $(k-1)/2^\nu = \xi_\nu(\omega) \leq \xi(\omega)$ , sicché risulterà anche

$$\xi(\omega) - \xi_\nu(\omega) = |\xi(\omega) - \xi_\nu(\omega)| < \epsilon,$$

e siccome  $(\xi_n(\omega))_{n \in \mathbf{N}}$  è una successione numerica non decrescente con  $\xi_n(\omega) \leq \xi(\omega)$ , si avrà infine che

$$|\xi(\omega) - \xi_n(\omega)| < \epsilon, \quad \forall n > \nu,$$

cioè che  $\xi_n(\omega) \xrightarrow{n} \xi(\omega)$ . □

**II.5.17 Osservazione:** Sulla base dei teoremi precedenti si può anche provare che, se  $\xi$  e  $\eta$  sono v.a. (eventualmente anche estese), anche  $\xi \pm \eta$ ,  $\xi\eta$ ,  $\xi/\eta$  ... sono variabili aleatorie (estese), sempre che non assumano valori indeterminati del tipo  $\infty - \infty$ ,  $\infty/\infty$ ,  $0/0$  ... Infatti, in base al Teorema II.5.16, potremo sempre determinare delle successioni  $(\xi_n)_{n \in \mathbf{N}}$  e  $(\eta_n)_{n \in \mathbf{N}}$  di v.a. semplici convergenti (in senso puntuale) verso  $\xi$  e  $\eta$  e formare poi le successioni di v.a. semplici

$$\xi_n \pm \eta_n; \quad \xi_n \eta_n; \quad \frac{\xi_n}{\eta_n + \frac{1}{n} I_{\{\eta_n=0\}}}; \quad \dots$$

che convergono verso  $\xi \pm \eta$ ,  $\xi\eta$ ,  $\xi/\eta$  ... Il Teorema II.5.15 ci garantisce allora che queste ultime sono misurabili, cioè sono v.a. ○

**II.5.18 Osservazione:** Se  $\xi : (\Omega, \mathcal{F}) \rightarrow (\mathbf{R}, \mathcal{B}(\mathbf{R}))$  è una v.a. si può provare che la seguente famiglia di parti di  $\Omega$

$$\mathcal{F}_\xi = (\xi^{-1}(B))_{B \in \mathcal{B}(\mathbf{R})} \quad (\mathcal{F}_\xi \subseteq \mathcal{F})$$

è ancora una  $\sigma$ -algebra (verificarlo per esercizio) che prende il nome di  **$\sigma$ -algebra generata da  $\xi$** . È ovvio che questa risulta essere anche la più piccola  $\sigma$ -algebra



di parti di  $\Omega$  rispetto alla quale  $\xi$  è misurabile. È allora anche possibile precisare l'enunciato del Lemma II.5.9 nel modo seguente: se  $\xi$  è una v.a. e  $\phi(x)$  una funzione di Borel, anche  $\eta = \phi(\xi)$  è una v.a. che risulta anche essere  $\mathcal{F}_\xi$ -misurabile in quanto, siccome

$$(\phi^{-1}(B))_{B \in \mathcal{B}(\mathbf{R})} \subseteq \mathcal{B}(\mathbf{R}),$$

si ha che

$$(\eta^{-1}(B))_{B \in \mathcal{B}(\mathbf{R})} = (\xi^{-1}[\phi^{-1}(B)])_{B \in \mathcal{B}(\mathbf{R})} \subseteq (\xi^{-1}(A))_{A \in \mathcal{B}(\mathbf{R})} = \mathcal{F}_\xi,$$

ossia in pratica  $F_\eta \subseteq F_\xi$ . Il successivo importante Teorema asserisce che vale anche la proprietà inversa di quella appena enunciata: se, cioè,  $\eta$  è  $\mathcal{F}_\xi$ -misurabile, allora essa è certamente funzione di  $\xi$  tramite una ben determinata funzione di Borel  $\phi(x)$ . ○

**II.5.19 Teorema:** Se  $\eta$  è  $\mathcal{F}_\xi$ -misurabile, allora esiste una funzione di Borel  $\phi$  tale che  $\eta = \phi \circ \xi$ , cioè tale che  $\eta(\omega) = \phi[\xi(\omega)]$ ,  $\forall \omega \in \Omega$ .

**Dimostrazione:** Consideriamo innanzitutto il caso in cui la v.a.  $\eta$  sia un indicatore  $I_A(\omega)$ : per essere  $\mathcal{F}_\xi$ -misurabile deve risultare  $A \in \mathcal{F}_\xi$  e quindi anche  $A = \xi^{-1}(B)$  per qualche  $B \in \mathcal{B}(\mathbf{R})$  opportuno. Se allora consideriamo la funzione di Borel  $\chi_B(x)$  data dall'indicatore di  $B$ , si verifica facilmente che  $\eta(\omega) = I_A(\omega) = \chi_B(\xi(\omega))$ ,  $\forall \omega \in \Omega$  in quanto risulta

$$\chi_B(\xi(\omega)) = \begin{cases} 1, & \text{se } \xi(\omega) \in B, \text{ cioè se } \omega \in \xi^{-1}(B) = A; \\ 0, & \text{se } \xi(\omega) \notin B, \text{ cioè se } \omega \notin \xi^{-1}(B) = A. \end{cases}$$

Dunque abbiamo determinato una opportuna funzione di Borel  $\chi_B$  tale che  $\eta = I_A = \chi_B \circ \xi$ .

Il precedente ragionamento si estende facilmente al caso di v.a. semplici ed  $\mathcal{F}_\xi$ -misurabili del tipo

$$\sum_{k:1}^n a_k I_{A_k}(\omega), \quad A_k \in \mathcal{F}_\xi, \quad k = 1, \dots, n,$$

dato che queste non sono altro che combinazioni lineari finite di indicatori. Per completare la dimostrazione supporremo allora che  $\eta$  sia un'arbitraria v.a.  $\mathcal{F}_\xi$ -misurabile: il Teorema II.5.16 ci dice allora che esisterà una successione di v.a. semplici ed  $\mathcal{F}_\xi$ -misurabili  $(\eta_n)_{n \in \mathbf{N}}$  tale che  $\eta_n(\omega) \xrightarrow{n} \eta(\omega)$  comunque scelto  $\omega \in \Omega$ . Sulla base della discussione precedente, inoltre, sarà anche possibile determinare una successione di funzioni di Borel  $(\phi_n(x))_{n \in \mathbf{N}}$  tale che

$$\eta_n(\omega) = \phi_n[\xi(\omega)], \quad \forall n \in \mathbf{N}, \quad \forall \omega \in \Omega.$$

Definiamo ora i seguenti sottinsiemi di numeri reali

$$A = \{x \in \mathbf{R} : \exists \omega \in \Omega \ni x = \xi(\omega)\}$$

$$B = \{x \in \mathbf{R} : (\phi_n(x))_{n \in \mathbf{N}} \text{ converge verso un limite finito}\}$$

ed osserviamo che  $A \subseteq B$ : infatti se  $x \in A$ , essendo  $A$  null'altro che il codominio di  $\xi$ , per definizione esisterà un  $\omega \in \Omega$  tale che  $x = \xi(\omega)$  sicché

$$\phi_n(x) = \phi_n[\xi(\omega)] = \eta_n(\omega) \xrightarrow{n} \eta(\omega) \in \mathbf{R},$$

e quindi, ancora per definizione, si ha anche che  $x \in B$ . Se allora definiamo

$$\phi(x) = \lim_n [\phi_n(x)\chi_B(x)] = \begin{cases} \lim_n \phi_n(x), & \text{se } x \in B; \\ 0, & \text{se } x \notin B; \end{cases}$$

(dove  $\chi_B$  è la solita funzione di Borel indicatore di  $B$ ) ci si accorge subito del fatto che  $\eta(\omega) = \phi[\xi(\omega)]$  per ogni  $\omega \in \Omega$ , in quanto comunque scelto  $\omega \in \Omega$ , risulta  $\xi(\omega) \in A \subseteq B$  e quindi dalle definizioni date

$$\phi(\xi(\omega)) = \lim_n \phi_n[\xi(\omega)] = \lim_n \eta_n(\omega) = \eta(\omega).$$

Avendo quindi costruito la funzione cercata non ci resta che provare che si tratta di una funzione di Borel: siccome  $\phi$  è stata definita come limite delle funzioni  $\phi_n\chi_B$ , in base al Teorema II.5.15 basterà provare che queste ultime sono tutte di Borel, e siccome le  $\phi_n$  sono sicuramente di Borel per come sono state definite, basterà provare che  $\chi_B$  è di Borel, cioè che  $B \in \mathcal{B}(\mathbf{R})$ . Per verificare quest'ultima affermazione basterà osservare che, applicando il criterio di convergenza di Cauchy, dalla definizione di  $B$  otteniamo

$$B = \{x \in \mathbf{R} : \forall k \in \mathbf{N} \exists \nu \in \mathbf{N} \ni |\phi_n(x) - \phi_m(x)| < 1/k, \forall n, m > \nu\}$$

$$= \bigcap_k \bigcup_{\nu} \bigcap_{n, m > \nu} \{x \in \mathbf{R} : |\phi_n(x) - \phi_m(x)| < 1/k\}$$

e siccome risulta anche

$$\{x \in \mathbf{R} : |\phi_n(x) - \phi_m(x)| < 1/k\} \in \mathcal{B}(\mathbf{R}), \quad \forall k, n, m \in \mathbf{N}$$

ne segue immediatamente che  $B \in \mathcal{B}(\mathbf{R})$  dato che  $\mathcal{B}(\mathbf{R})$  è una  $\sigma$ -algebra. □

**II.5.20 Definizione:** Diremo che una proprietà è vera **P-quasi ovunque** (**P-q.o.**) se essa si verifica per ogni  $\omega \in \Omega$  fatta eccezione per un insieme di **P**-misura nulla, cioè se la probabilità che essa non si verifichi è nulla. In particolare diremo che due v.a.  $\xi$  ed  $\eta$  coincidono **P-q.o.** (e scriveremo  $\xi = \eta$ , **P-q.o.**) se  $\mathbf{P}\{\xi \neq \eta\} = 0$ . △

**II.5.21 Definizione:** Dato  $(\Omega, \mathcal{F}, \mathbf{P})$  diremo che  $D \in \mathcal{F}$  è un **atomo** rispetto a **P** quando, comunque scelto  $A \in \mathcal{F}$  con  $A \subseteq D$  risulta  $\mathbf{P}(A) = 0$  oppure  $\mathbf{P}(D \setminus A) = 0$ ;

cioè quando ogni evento contenuto in  $D$  o è trascurabile oppure coincide  $\mathbf{P}$ -q.o. con  $D$ .  $\triangle$

**II.5.22 Proposizione:** Dato uno spazio  $(\Omega, \mathcal{F}, \mathbf{P})$  con  $\mathcal{F} = \sigma(\mathcal{D})$  generata da una decomposizione  $\mathcal{D} = \{D_k\}_{k \in M}$  (finita o numerabile) in atomi  $D_k$  con  $\mathbf{P}(D_k) > 0$ , ogni v.a.  $\xi$   $\mathcal{F}$ -misurabile è della forma

$$\xi = \sum_{k \in M} x_k I_{D_k}, \quad \mathbf{P}\text{-q.o.},$$

cioè risulta costante  $\mathbf{P}$ -q.o. sugli atomi di  $\mathcal{D}$ .

**Dimostrazione:** Proveremo che se  $D$  è un atomo, allora  $\xi$  risulta  $\mathbf{P}$ -q.o. costante su di esso. Osserviamo innanzitutto che, al crescere di  $x \in \overline{\mathbf{R}}$ , le famiglie  $\{\xi < x\}$  e  $\{\xi \geq x\}$  di eventi di  $\mathcal{F}$  risultano rispettivamente non decrescente e non crescente (per inclusione) sicché, se  $D$  è un atomo, anche  $\mathbf{P}(D \cap \{\xi < x\})$  e  $\mathbf{P}(D \cap \{\xi \geq x\})$  risultano funzioni di  $x$  rispettivamente non decrescente e non crescente e hanno valori compresi tra 0 ed 1. Se allora poniamo

$$(1) \quad k = \sup\{x \in \overline{\mathbf{R}} : \mathbf{P}(D \cap \{\xi < x\}) = 0\}$$

potremo provare che  $\xi = k$   $\mathbf{P}$ -q.o. su  $D$ , cioè che  $\mathbf{P}(D \cap \{\xi \neq k\}) = 0$ . Data la monotonia di  $\mathbf{P}(D \cap \{\xi < x\})$  e per la definizione di  $k$  si ha ovviamente che

$$(2) \quad \mathbf{P}(D \cap \{\xi < x\}) = 0, \quad \forall x < k,$$

ma si può anche provare che

$$(3) \quad \mathbf{P}(D \cap \{\xi \geq x\}) = 0, \quad \forall x > k.$$

Infatti, siccome  $A \cap B = A \cap (\overline{A \cap B}) = A \cap (\overline{A \cap \overline{B}}) = A \setminus (A \cap \overline{B})$ , avremo che

$$(4) \quad D \cap \{\xi \geq x\} = D \setminus (D \cap \{\xi < x\})$$

e inoltre  $D \cap \{\xi < x\} \subseteq D$  con  $\mathbf{P}(D \cap \{\xi < x\}) > 0$  a causa di (1) e del fatto che  $x > k$ ; siccome però  $D$  è un atomo questo implica che  $\mathbf{P}(D \setminus (D \cap \{\xi < x\})) = 0$  e quindi da (4) si ottiene (3). Sulla base di queste osservazioni è possibile ora provare che

$$\mathbf{P}(D \cap \{\xi < k\}) = \mathbf{P}(D \cap \{\xi > k\}) = 0,$$

e siccome ovviamente

$$\mathbf{P}(D \cap \{\xi \neq k\}) = \mathbf{P}(D \cap \{\xi > k\}) + \mathbf{P}(D \cap \{\xi < k\}),$$

ne segue che  $\mathbf{P}(D \cap \{\xi \neq k\}) = 0$  sicché su ogni atomo  $D$  la  $\xi$  risulta  $\mathbf{P}$ -q.o. costante: siccome poi  $\Omega = \bigcup_k D_k$ , la tesi segue immediatamente. Per far questo definiamo allora gli eventi  $A_n = \{\xi < k - \frac{1}{n}\}$  e  $B_n = \{\xi \geq k + \frac{1}{n}\}$  e consideriamo le due successioni  $(A_n)_{n \in \mathbf{N}}$  e  $(B_n)_{n \in \mathbf{N}}$ . Tali successioni risultano ambedue non decrescenti (per inclusione) in quanto

$$\dots < k - \frac{1}{n} < k - \frac{1}{n+1} < \dots < k < \dots < k + \frac{1}{n+1} < k + \frac{1}{n} < \dots$$

e quindi  $A_n \subseteq A_{n+1}$  e  $B_n \subseteq B_{n+1}$  comunque scelto  $n \in \mathbf{N}$ . Inoltre, siccome

$$A_n \uparrow \{\xi < k\} = \bigcup_n A_n, \quad B_n \uparrow \{\xi > k\} = \bigcup_n B_n,$$

risulta che

$$D \cap A_n \uparrow D \cap \{\xi < k\} = \bigcup_n (D \cap A_n), \quad D \cap B_n \uparrow D \cap \{\xi > k\} = \bigcup_n (D \cap B_n),$$

e quindi, in base al Teorema II.1.13, si ha che

$$\mathbf{P}(D \cap \{\xi < k\}) = \lim_n \mathbf{P}(D \cap A_n), \quad \mathbf{P}(D \cap \{\xi > k\}) = \lim_n \mathbf{P}(D \cap B_n);$$

ma, siccome

$$k - \frac{1}{n} < k < k + \frac{1}{n}, \quad \forall n \in \mathbf{N},$$

da (2) e (3) si ricava che  $\mathbf{P}(A_n) = \mathbf{P}(B_n) = 0$ , sicché  $\mathbf{P}(D \cap A_n) = \mathbf{P}(D \cap B_n) = 0$ ,  $\forall n \in \mathbf{N}$ . Ne segue quindi facilmente, come annunciato, che  $\mathbf{P}(D \cap \{\xi < k\}) = \mathbf{P}(D \cap \{\xi > k\}) = 0$ , il che definitivamente prova la tesi.  $\square$

**II.5.23 Osservazione:** Il concetto di v.a. come funzione misurabile  $\xi : (\Omega, \mathcal{F}) \rightarrow (\mathbf{R}, \mathcal{B}(\mathbf{R}))$  può essere generalizzato supponendo che  $\xi$  prenda valori in spazi di natura diversa da  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ . In fondo, l'unica proprietà di  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$  indispensabile per la Definizione II.5.1 è che esso sia uno spazio probabilizzabile: le altre proprietà dei numeri reali sono usate solo nella discussione di esempi particolari. Pertanto esamineremo nel seguito alcune possibili generalizzazioni del concetto di v.a. supponendo che  $\xi$  prenda valori in insiemi di vario tipo.  $\circ$

**II.5.24 Definizione:** Dati due spazi probabilizzabili  $(\Omega, \mathcal{F})$  ed  $(E, \mathcal{E})$  diremo che una funzione  $X : (\Omega, \mathcal{F}) \rightarrow (E, \mathcal{E})$  è un **elemento aleatorio** se essa risulta  $\mathcal{F}/\mathcal{E}$ -misurabile, cioè se

$$X^{-1}(B) \in \mathcal{F}, \quad \forall B \in \mathcal{E};$$

a volte  $X$  viene chiamata anche variabile aleatoria con valori in  $E$ .  $\triangle$

**II.5.25 Esempio:** Supponiamo innanzitutto che  $(E, \mathcal{E}) = (\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$  con  $\mathbf{R}^n = \mathbf{R}_1 \times \dots \times \mathbf{R}_n$ . In questo caso i valori assunti da  $X$  in corrispondenza di un  $\omega \in \Omega$

sono le  $x = (x_1, \dots, x_n) = (x_k)_{k=1, \dots, n} \in \mathbf{R}^n$  e si usa dire che  $X$  è un **punto aleatorio** in  $\mathbf{R}^n$ . Osserviamo ora che, se  $\pi_k : \mathbf{R}^n \rightarrow \mathbf{R}_k$ , con  $k = 1, \dots, n$ , è la proiezione che estrae da  $x = (x_k)_{k=1, \dots, n}$  la  $k$ -ma coordinata  $x_k$  (nel senso che  $\pi_k(x) = x_k$ ), posto  $\xi_k = \pi_k \circ X$ , si verifica subito che le  $\xi_k$  sono delle v.a. in  $(\mathbf{R}_k, \mathcal{B}(\mathbf{R}_k))$  che prendono il nome di **componenti** del punto aleatorio  $X$ . Infatti, dato  $B \in \mathcal{B}(\mathbf{R}_k)$ , si ha

$$\begin{aligned} \{\xi_k \in B\} &= \{\xi_1 \in \mathbf{R}_1, \dots, \xi_k \in B, \dots, \xi_n \in \mathbf{R}_n\} \\ &= \{X \in \mathbf{R}_1 \times \dots \times B \times \dots \times \mathbf{R}_n\} \in \mathcal{F} \end{aligned}$$

dato che  $\mathbf{R}_1 \times \dots \times B \times \dots \times \mathbf{R}_n$  è certamente un elemento di  $\mathcal{B}(\mathbf{R}^n)$ . Se poi conveniamo di chiamare **vettore aleatorio** (vett.a.) ad  $n$  componenti una famiglia ordinata di  $n$  v.a.  $(\eta_1, \dots, \eta_n) = (\eta_k)_{k=1, \dots, n}$ , ci accorgiamo subito di aver mostrato che le componenti  $\xi_k$  di un punto aleatorio  $X$  costituiscono un vett.a.  $(\xi_k)_{k=1, \dots, n}$ . Viceversa possiamo anche mostrare che un vett.a. ad  $n$  componenti definisce sempre un punto aleatorio  $X$  in  $\mathbf{R}^n$  in quanto, generalizzando opportunamente il Lemma II.5.7 al caso di  $\mathbf{R}^n$ , siccome  $\mathcal{B}(\mathbf{R}^n)$  è generata dalla semialgebra dei rettangoli del tipo  $B_1 \times \dots \times B_n$ , basterà osservare che

$$\{X \in B_1 \times \dots \times B_n\} = \bigcap_{k=1}^n \{\xi_k \in B_k\} \in \mathcal{F}$$

per stabilire che in effetti  $X$  è un punto aleatorio. ◇

**II.5.26 Esempio:** Se  $\mathbf{C}$  è l'insieme dei numeri complessi  $z = x + iy$  possiamo considerare il caso in cui  $(E, \mathcal{E})$  è dato da  $(\mathbf{C}, \mathcal{B}(\mathbf{C}))$ . Data la nota isomorfia fra  $\mathbf{C}$  e  $\mathbf{R}^2$  si tratterà di un caso particolare dell'Esempio precedente per  $n = 2$ . Pertanto una **v.a. complessa**  $Z(\omega)$  sarà data tramite le sue componenti (parte reale e parte immaginaria) nella forma  $Z(\omega) = X(\omega) + iY(\omega)$ , dove  $X$  ed  $Y$  sono ordinarie v.a. reali. ◇

**II.5.27 Esempio:** Consideriamo ora il caso in cui  $(E, \mathcal{E})$  è lo spazio  $(\mathbf{R}^T, \mathcal{B}(\mathbf{R}^T))$  con  $T$  sottinsieme (in generale non finito, numerabile o non numerabile, in modo che  $\mathbf{R}^\infty$  ne sia un caso particolare) di  $\mathbf{R}$ . In tal caso un elemento aleatorio  $X$  associerà ad ogni  $\omega \in \Omega$  una intera funzione  $(x_t(\omega))_{t \in T}$  detta **traiettoria**, mentre l'elemento aleatorio stesso prende il nome di **funzione aleatoria**. Si vede però subito che, se  $\pi_t$  è la proiezione da  $\mathbf{R}^T$  in  $\mathbf{R}_t$  che da  $X$  estrae la v.a.  $\xi_t = \pi_t \circ X$  relativa al valore  $t \in T$ , la funzione aleatoria data può anche essere considerata come una famiglia  $(\xi_t)_{t \in T}$  di v.a. Se poi conveniamo di chiamare **processo stocastico** (p.s.) definito su  $T$  ogni famiglia di v.a.  $(\eta_t)_{t \in T}$ , la discussione precedente mostra che ogni funzione aleatoria  $X$  è rappresentabile come un p.s.  $(\xi_t)_{t \in T}$  con  $\xi_t = \pi_t \circ X$ . Viceversa si potrebbe mostrare che ogni p.s. definisce una funzione aleatoria. In un certo senso la differenza fra funzione aleatoria e p.s. consiste nel fatto che la funzione aleatoria  $X$  associa ad ogni  $\omega \in \Omega$  una funzione globale (traiettoria)  $(x_t)_{t \in T} \in \mathbf{R}^T$  ed è  $\mathcal{F}/\mathcal{B}(\mathbf{R}^T)$ -misurabile, mentre il p.s.  $(\xi_t)_{t \in T}$

associa ad ogni  $t \in T$  una v.a.  $\xi_t : (\Omega, \mathcal{F}) \rightarrow (\mathbf{R}, \mathcal{B}(\mathbf{R}))$  cioè una funzione  $\mathcal{F}/\mathcal{B}(\mathbf{R})$ -misurabile. In fondo  $X$  risulta essere funzione di due variabili  $t$  ed  $\omega$  (di natura diversa) e le due definizioni consistono nel mettere in evidenza alternativamente la dipendenza da ciascuna di queste variabili. Infine osserviamo che, quando  $T = \{0, 1, 2, \dots\} = \mathbf{N}$  è l'insieme dei numeri naturali si ricade nel caso di  $\mathbf{R}^\infty$ : il p.s. prende la forma  $(\eta_n)_{n \in \mathbf{N}}$  e viene chiamato **successione di v.a. aleatorie (successione aleatoria)** o anche p.s. a tempi discreti.  $\diamond$

**II.5.28 Definizione:** Dato un vett.a.  $X = (\xi_k)_{k=1, \dots, n}$  diremo **distribuzione di probabilità congiunta** (delle componenti) di  $X$  (DdP congiunta), o **immagine** di  $\mathbf{P}$  su  $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$  tramite  $X$ , la  $\mathbf{P}_X$  definita su  $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$  da

$$\mathbf{P}_X(B) = \mathbf{P}(X^{-1}(B)) = \mathbf{P}\{\omega : X(\omega) \in B\}$$

dove  $B$  è un generico elemento di  $\mathcal{B}(\mathbf{R}^n)$ ; chiameremo invece **distribuzioni di probabilità marginali** (delle componenti) di  $X$  (DdP marginali) le DdP  $\mathbf{P}_{\xi_k}$  definite sugli spazi  $(\mathbf{R}_k, \mathcal{B}(\mathbf{R}_k))$  da

$$\mathbf{P}_{\xi_k}(A) = \mathbf{P}(\xi_k^{-1}(A)) = \mathbf{P}\{\omega : \xi_k(\omega) \in A\}$$

dove  $A$  è un generico elemento di  $\mathcal{B}(\mathbf{R}_k)$ . Chiameremo inoltre **funzione di distribuzione congiunta** (FdD congiunta) di  $X$  la FdD di  $\mathbf{P}_X$ , cioè la funzione

$$\begin{aligned} F_X(x) &= F_X(x_1, \dots, x_n) = \mathbf{P}_X\{(-\infty, x_1] \times \dots \times (-\infty, x_n]\} \\ &= \mathbf{P}\{\omega : \xi_1 \leq x_1, \dots, \xi_n \leq x_n\} \end{aligned}$$

con  $x = (x_1, \dots, x_n) \in \mathbf{R}^n$ , e **funzioni di distribuzione marginali** (FdD marginali) delle componenti di  $X$  le FdD delle  $\mathbf{P}_{\xi_k}$ , cioè le funzioni

$$F_{\xi_k}(x_k) = \mathbf{P}_{\xi_k}\{(-\infty, x_k]\} = \mathbf{P}_X\{\mathbf{R}_1 \times \dots \times (-\infty, x_k] \times \dots \times \mathbf{R}_n\} = \mathbf{P}\{\omega : \xi_k \leq x_k\}$$

con  $x_k \in \mathbf{R}_k$  e  $k = 1, \dots, n$ .  $\triangle$

**II.5.29 Definizione:** Data una funzione aleatoria (o un p.s.)  $X = (\xi_t)_{t \in T}$  chiameremo **distribuzione di probabilità** (DdP) di  $X$ , o **immagine** di  $\mathbf{P}$  su  $(\mathbf{R}^T, \mathcal{B}(\mathbf{R}^T))$  tramite  $X$ , la probabilità  $\mathbf{P}_X$  definita su  $(\mathbf{R}^T, \mathcal{B}(\mathbf{R}^T))$  dalla relazione

$$\mathbf{P}_X(B) = \mathbf{P}\{\omega : X \in B\}$$

con  $B \in \mathcal{B}(\mathbf{R}^T)$ . Dato inoltre un sottinsieme  $S = \{t_1, \dots, t_s\}$  finito di  $T$  (con  $t_1 < \dots < t_s$ ) chiameremo **distribuzione di probabilità finito-dimensionale** di  $X$  la probabilità  $\mathbf{P}_X^{(S)}$  definita su  $(\mathbf{R}^{(S)}, \mathcal{B}(\mathbf{R}^{(S)}))$  da

$$\mathbf{P}_X^{(S)}(B) = \mathbf{P}\{\omega : (\xi_t)_{t \in S} \in B\} = \mathbf{P}(\pi_S^{-1}(B)) = \mathbf{P}(C(B))$$

con  $B \in \mathcal{B}(\mathbf{R}^{(S)})$ , e **funzione di distribuzione finito-dimensionale** di  $X$  la funzione

$$\begin{aligned} F_X^{(S)}(x) &= F_X^{(S)}(x_1, \dots, x_n) = \mathbf{P}_X^{(S)}\{(-\infty, x_1] \times \dots \times (-\infty, x_s]\} \\ &= \mathbf{P}\{\omega : \xi_{t_1} \leq x_1, \dots, \xi_{t_s} \leq x_s\} \end{aligned}$$

con  $x = (x_1, \dots, x_s) \in \mathbf{R}^{(S)}$ . △

**II.5.30 Osservazione:** Mostriamo qui di seguito che, data la DdP congiunta  $\mathbf{P}_X$  di un vett.a.  $X = (\xi_k)_{k=1, \dots, n}$  è sempre possibile ricavare in maniera unica le DdP marginali  $\mathbf{P}_{\xi_k}$  delle sue componenti. Sarà invece provato più oltre, con dei controesempi, che il viceversa non è, in generale, vero nel senso che, date le DdP marginali  $\mathbf{P}_{\xi_k}$ , la DdP congiunta  $\mathbf{P}_X$  non può, in generale, essere ricostruita in maniera univoca. Se la DdP congiunta  $\mathbf{P}_X$  è data, preso  $A \in \mathcal{B}(\mathbf{R}_k)$  e il rettangolo di  $\mathcal{B}(\mathbf{R}^n)$

$$\pi_k^{-1}(A) = \mathbf{R}_1 \times \dots \times A \times \dots \times \mathbf{R}_n$$

dove  $\pi_k$  è la proiezione tale che  $\xi_k = \pi_k(X)$ , abbiamo ovviamente

$$\begin{aligned} \xi_k^{-1}(A) &= \{\omega : \xi_k \in A\} = \{\omega : \xi_1 \in \mathbf{R}_1, \dots, \xi_k \in A, \dots, \xi_n \in \mathbf{R}_n\} \\ &= \{\omega : X \in \mathbf{R}_1 \times \dots \times A \times \dots \times \mathbf{R}_n\} = \{\omega : X \in \pi_k^{-1}(A)\} \\ &= X^{-1}(\pi_k^{-1}(A)) \end{aligned}$$

per cui dalla Definizione II.5.28 si ha

$$\mathbf{P}_{\xi_k}(A) = \mathbf{P}(\xi_k^{-1}(A)) = \mathbf{P}(X^{-1}(\pi_k^{-1}(A))) = \mathbf{P}_X(\pi_k^{-1}(A))$$

e quindi da  $\mathbf{P}_X$  è sempre possibile calcolare  $\mathbf{P}_{\xi_k}$ . Ovviamente, con lo stesso metodo, è anche possibile dedurre da  $\mathbf{P}_X$  le DdP congiunte di ogni vett.a. le cui componenti siano scelte (in numero minore di  $n$ ) fra le componenti di  $X$ . Le osservazioni precedenti possono inoltre essere estese anche alle relazioni che intercorrono tra la FdD congiunta e le FdD marginali di un vett.a.: da una FdD congiunta  $F_X(x_1, \dots, x_n)$  è sempre possibile ricavare le FdD marginali  $F_{\xi_k}(x_k)$ , ma in generale il viceversa non è possibile in modo univoco. Infatti è facile verificare, dai calcoli precedenti, che

$$F_{\xi_k}(x_k) = \lim_{x_j \rightarrow +\infty; j \neq k} F_X(x_1, \dots, x_k, \dots, x_n) = F_X(+\infty, \dots, x_k, \dots, +\infty).$$

Allo stesso modo si possono ricavare le FdD congiunte di ogni vett.a. le cui componenti siano scelte (in numero minore di  $n$ ) fra le componenti di  $X$ . Viceversa la conoscenza delle FdD marginali  $F_{\xi_k}(x_k)$  non permette di ricostruire in maniera univoca la FdD congiunta  $F_X$ ; in particolare è facilmente verificabile che, date le FdD marginali, la funzione

$$F_{\xi_1}(x_1) \cdot \dots \cdot F_{\xi_n}(x_n)$$

è una FdD congiunta che ha come marginali proprio le  $F_{\xi_k}(x_k)$  (per provarlo basterà adottare la regola di deduzione esposta prima tenendo conto della proprietà  $D_2$  dell'Osservazione II.3.1; vedi anche Esempio II.4.6), ma che essa, come vedremo nei successivi esempi, non è l'unica FdD congiunta che ammette quelle  $F_{\xi_k}(x_k)$  come marginali.  $\circ$

**II.5.31 Osservazione:** Se il vett.a.  $X$  ha FdD congiunta  $F_X$  a.c. (vedi a questo proposito la Definizione II.3.16 e l'Osservazione II.4.7) esisterà una **funzione di densità congiunta** (fdd congiunta)  $f_X(x_1, \dots, x_n)$  per la quale avremo

$$F_X(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_X(t_1, \dots, t_n) dt_1 \dots dt_n$$

con

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f_X(t_1, \dots, t_n) dt_1 \dots dt_n = 1.$$

Naturalmente sussiste anche la relazione inversa

$$f_X(x_1, \dots, x_n) = \frac{\partial^n}{\partial x_1 \dots \partial x_n} F_X(x_1, \dots, x_n).$$

In tal caso si mostra che anche le singole componenti  $\xi_k$  hanno FdD marginali  $F_{\xi_k}(x_k)$  a.c. con fdd  $f_{\xi_k}(x_k)$  che si ricavano dalla fdd congiunta  $f_X(x_1, \dots, x_n)$  nel modo seguente:

$$f_{\xi_k}(x_k) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f_X(t_1, \dots, x_k, \dots, t_n) dt_1 \dots dt_{k-1} dt_{k+1} \dots dt_n,$$

(dove le  $n-1$  integrazioni sono effettuate su tutte le variabili con l'eccezione di  $x_k$ ) e per le quali risulta anche  $f_{\xi_k}(x) = F'_{\xi_k}(x)$ . Infatti, sulla base dell'Osservazione II.5.30, si ha che

$$\begin{aligned} F_{\xi_k}(x_k) &= \mathbf{P}_{\xi_k}\{(-\infty, x_k]\} = \mathbf{P}_X\{\mathbf{R}_1 \times \dots \times (-\infty, x_k] \times \dots \times \mathbf{R}_n\} \\ &= F_X(+\infty, \dots, x_k, \dots, +\infty) \\ &= \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{x_k} \dots \int_{-\infty}^{+\infty} f_X(t_1, \dots, t_n) dt_1 \dots dt_n = \int_{-\infty}^{x_k} f_{\xi_k}(t_k) dt_k \end{aligned}$$

dove  $f_{\xi_k}(x_k)$  è l'integrale  $(n-1)$ -plo definito in precedenza. Queste  $f_{\xi_k}(x_k)$  prendono poi il nome di **funzioni di densità marginali** (fdd marginali) ed anche qui vale l'osservazione che, mentre dalla fdd congiunta  $f_X(x_1, \dots, x_n)$  è sempre possibile ricavare le fdd marginali  $f_{\xi_k}(x_k)$  (per integrazione sulle variabili che si vogliono eliminare, come mostrato prima), la ricostruzione della fdd congiunta a partire dalle marginali non può essere effettuata in modo univoco. In particolare si verifica anche in questo caso che, date le fdd marginali, la funzione

$$f_{\xi_1}(x_1) \cdot \dots \cdot f_{\xi_n}(x_n)$$



è una fdd congiunta che ammette le  $f_{\xi_k}(x_k)$  come marginali, ma che essa non è l'unica fdd con questa proprietà. Per le relazioni fra FdD e fdd nel caso unidimensionale si rinvia anche all'Osservazione II.3.18.  $\circ$

**II.5.32 Esempio:** Un caso importante di vett.a. dotato di fdd è costituito dai **vett.a. Gaussiani (o normali)** che sono caratterizzati da fdd Gaussiane multivariate del tipo

$$f(x_1, \dots, x_n) = \sqrt{\frac{|A|}{(2\pi)^n}} e^{-(x-m)^T A (x-m)/2},$$

già discusse nel corso dell'Esempio II.4.8. Val solo la pena di aggiungere nel presente contesto che ogni vett.a. gaussiano ha come fdd marginali delle fdd gaussiane del tipo

$$f_{\xi_k}(x_k) = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-(x_k - m_k)^2 / 2\sigma_k^2}, \quad (\sigma_k > 0)$$

come può essere visto con calcolo diretto applicando, ad esempio, la formula per la determinazione delle marginali esposta nell'Osservazione II.5.31 alla fdd gaussiana bivariata data alla fine dell'Esempio II.4.8. Non bisogna però pensare che un vett.a. con marginali gaussiane sia, solo per questo, anche un vett.a. gaussiano: la fdd congiunta, infatti, non è univocamente definita a partire dalle sue marginali e pertanto il nostro vett.a. potrebbe non avere una fdd gaussiana multivariata come fdd congiunta.  $\diamond$

**II.5.33 Esempio:** Mostriamo ora con un esempio che, come già più volte osservato, da FdD o fdd marginali non si può ricostruire in maniera univoca la FdD o la fdd congiunta. Supponiamo di considerare due vett.a. a due componenti  $X = (\xi_1, \xi_2)$  e  $Y = (\eta_1, \eta_2)$  ambedue dotati di FdD a.c. e quindi dotati di fdd. Supponiamo inoltre di sapere che le fdd marginali delle varie componenti coincidono per i due vett.a. e per ulteriore semplicità supponiamo che sia  $f_{\xi_1}(x) = f_{\xi_2}(x) = f_{\eta_1}(x) = f_{\eta_2}(x) = f(x)$  dove

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

Mostriamo ora che le fdd congiunte dei due vett.a.  $f_X(x, y)$  e  $f_Y(x, y)$  possono essere differenti pur ammettendo le fdd marginali indicate prima, cioè marginali coincidenti. Se infatti ad esempio le fdd congiunte fossero

$$f_X(x, y) = f_{\xi_1}(x)f_{\xi_2}(y) = \begin{cases} 1, & \text{se } (x, y) \in Q; \\ 0, & \text{se } (x, y) \notin Q; \end{cases}$$

$$f_Y(x, y) = \begin{cases} 2, & \text{se } (x, y) \in Q_1 \cup Q_2; \\ 0, & \text{se } (x, y) \notin Q_1 \cup Q_2; \end{cases}$$

dove  $Q = [0, 1] \times [0, 1]$ ,  $Q_1 = [0, \frac{1}{2}] \times [0, \frac{1}{2}]$  e  $Q_2 = [\frac{1}{2}, 1] \times [\frac{1}{2}, 1]$ , si verificherebbe immediatamente che le rispettive marginali coincidono tutte con  $f(x)$  dato che, ad esempio, usando le prescrizioni dell'Osservazione II.5.31 si ha

$$f_{\xi_1}(x) = \int_{-\infty}^{+\infty} f_X(x, y) dy = \begin{cases} 1, & \text{se } x \in [0, 1] \\ 0, & \text{se } x \notin [0, 1] \end{cases} = f(x),$$

e analogamente per le altre componenti di  $X$  ed  $Y$ . Ciò mostra esplicitamente che fdd congiunte differenti possono avere le medesime fdd marginali e quindi che dalla sola conoscenza delle marginali non si può risalire in maniera univoca alla fdd congiunta.  $\diamond$

**II.5.34 Osservazione:** Negli ultimi capitoli abbiamo appreso che, data una v.a.  $\xi$  è sempre possibile determinarne la DdP  $\mathbf{P}_\xi$  e conseguentemente la FdD  $F_\xi(x) = \mathbf{P}_\xi(-\infty, x]$  (vedi Osservazione II.3.1 e Definizione II.5.4). Inoltre se  $F_\xi$  è a.c. esiste anche una fdd  $f_\xi(x) = F'_\xi(x)$  per la quale sussistono le relazioni

$$F_\xi(x) = \int_{-\infty}^x f_\xi(t) dt, \quad F_\xi(b) - F_\xi(a) = \int_a^b f_\xi(t) dt,$$

(vedi Teorema II.3.17, Osservazione II.3.18 ed Osservazione II.5.5). Viceversa sappiamo anche che data una FdD  $F_\xi$  (o, se possibile, una fdd  $f_\xi$ ) è sempre possibile ricostruire in modo unico la DdP  $\mathbf{P}_\xi$  corrispondente (vedi Teorema II.3.5) per la quale risulta anche  $\mathbf{P}_\xi(a, b] = F_\xi(b) - F_\xi(a)$ . Ai fini pratici, però, sarebbe anche utile sapere in che modo, data  $F_\xi(x)$  (o, se possibile,  $f_\xi(x)$ ), si calcola  $\mathbf{P}_\xi(B)$  dove  $B$  è un generico elemento di  $\mathcal{B}(\mathbf{R})$  e non un intervallo del tipo  $(a, b]$ . La risposta a questa domanda sarà data e giustificata nel seguente capitolo sull'integrazione, ma, per non lasciare incompleta l'esposizione a questo punto, noi la anticiperemo osservando che essa può essere dedotta per analogia dalle seguenti relazioni vere nel caso  $B = (-\infty, x]$ :

$$\begin{aligned} \mathbf{P}_\xi(-\infty, x] &= F_\xi(x) = \int_{-\infty}^x dF_\xi(t) = \int_{(-\infty, x]} dF_\xi(t) \\ &= \int_{-\infty}^x f_\xi(t) dt = \int_{(-\infty, x]} f_\xi(t) dt, \end{aligned}$$

sulla base delle quali possiamo anticipare che nel caso di  $B \in \mathcal{B}(\mathbf{R})$  generico avremo

$$\mathbf{P}_\xi(B) = \int_B dF_\xi(x) = \int_{\mathbf{R}} I_B(x) dF_\xi(x) = \int_B f_\xi(x) dx = \int_{\mathbf{R}} I_B(x) f_\xi(x) dx$$

dove  $I_B(x)$  è l'indicatore dell'insieme  $B$ . Quanto qui esposto nel caso di v.a. a valori in  $\mathbf{R}$  può poi essere anche facilmente esteso al caso di vett.a. ad  $n$  componenti, sicchè, ad esempio la relazione

$$\begin{aligned} \mathbf{P}_X(B) &= \int_B f_X(x_1, \dots, x_n) dx_1 \dots dx_n \\ &= \int_{\mathbf{R}} I_B(x_1, \dots, x_n) f_X(x_1, \dots, x_n) dx_1 \dots dx_n \end{aligned}$$

permette di calcolare  $\mathbf{P}_X(B)$  per un generico insieme  $B \in \mathcal{B}(\mathbf{R}^n)$  nel caso in cui fosse nota la fdd  $f_X$ . ○

**II.5.35 Definizione:** Dato uno spazio di probabilità  $(\Omega, \mathcal{F}, \mathbf{P})$  ed una famiglia  $X$  di elementi aleatori  $X_\alpha : (\Omega, \mathcal{F}) \rightarrow (E_\alpha, \mathcal{E}_\alpha)$ , con  $\alpha \in \mathcal{I}$  (dove  $(E_\alpha, \mathcal{E}_\alpha)$  sono arbitrari spazi probabilizzabili e  $\mathcal{I}$  un arbitrario insieme di indici anche non finito o non numerabile) diremo che le componenti  $X_\alpha$  di  $X$  sono **elementi aleatori indipendenti** se, comunque scelto un insieme finito di indici  $\alpha_1, \dots, \alpha_m$ , risulta

$$\mathbf{P}(X_{\alpha_1} \in B_1, \dots, X_{\alpha_m} \in B_m) = \mathbf{P}(X_{\alpha_1} \in B_1) \cdot \dots \cdot \mathbf{P}(X_{\alpha_m} \in B_m)$$

dove i sottinsiemi  $B_k \in \mathcal{E}_{\alpha_k}$  sono arbitrari. Se in particolare  $\mathcal{I} = \{1, \dots, n\}$  e se  $(E_k, \mathcal{E}_k) = (\mathbf{R}_k, \mathcal{B}(\mathbf{R}_k))$  con  $k = 1, \dots, n$ , ci troveremmo nel caso di un vett.a.  $X = (\xi_k)_{k=1, \dots, n}$ : in tal caso diremo che le  $n$  componenti  $\xi_k$  sono **v.a. indipendenti** se

$$\begin{aligned} \mathbf{P}_X(B_1 \times \dots \times B_n) &= \mathbf{P}(\xi_1 \in B_1, \dots, \xi_n \in B_n) = \mathbf{P}(\xi_1 \in B_1) \cdot \dots \cdot \mathbf{P}(\xi_n \in B_n) \\ &= \mathbf{P}_{\xi_1}(B_1) \cdot \dots \cdot \mathbf{P}_{\xi_n}(B_n) \end{aligned}$$

comunque scelto un rettangolo  $B_1 \times \dots \times B_n \in \mathcal{B}(\mathbf{R}^n)$  con  $B_k \in \mathcal{B}(\mathbf{R}_k)$ . Nota che ciò equivale a richiedere che tutti gli eventi  $\{\xi_k \in B_k\}$  siano indipendenti al variare di  $B_k \in \mathcal{B}(\mathbf{R}_k)$  secondo la Definizione I.4.14. △

**II.5.36 Definizione:** Dati  $n$  spazi di probabilità  $(\Omega_k, \mathcal{F}_k, \mathbf{P}_k)$  con  $k = 1, \dots, n$  chiameremo **misura di probabilità prodotto** quella probabilità

$$\mathbf{P} = \mathbf{P}_1 \otimes \dots \otimes \mathbf{P}_n = \bigotimes_{k=1}^n \mathbf{P}_k$$

definita sullo spazio prodotto diretto

$$(\Omega, \mathcal{F}) = \left( \prod_{k=1}^n \Omega_k, \bigotimes_{k=1}^n \mathcal{F}_k \right)$$

per la quale risulti

$$\mathbf{P}(B) = \mathbf{P}(B_1 \times \dots \times B_n) = \mathbf{P}_1(B_1) \cdot \dots \cdot \mathbf{P}_n(B_n)$$

per ogni rettangolo  $B = B_1 \times \dots \times B_n$  con  $B_k \in \mathcal{F}_k$ ,  $k = 1, \dots, n$ . △

**II.5.37 Proposizione:** Dati  $n$  spazi di probabilità  $(\Omega_k, \mathcal{F}_k, \mathbf{P}_k)$  con  $k = 1, \dots, n$  esiste sempre un'unica misura di probabilità prodotto  $\mathbf{P} = \mathbf{P}_1 \otimes \dots \otimes \mathbf{P}_n$  definita sullo spazio prodotto diretto  $(\Omega, \mathcal{F})$ .

**Dimostrazione:** Daremo soltanto lo schema della dimostrazione per il caso in cui gli spazi  $(\Omega_k, \mathcal{F}_k)$  coincidono tutti con  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$  lasciandone i dettagli alla cura del lettore (per una dimostrazione più completa vedi ad esempio **R.B. Ash: *Real analysis and Probability***; Academic Press, 1972; p. 97). A partire dalle  $\mathbf{P}_k$  si costruiscano le FdD  $F_k(x)$  e la funzione  $F(x_1, \dots, x_n) = F_1(x_1) \cdot \dots \cdot F_n(x_n)$  che (come già osservato nell'Esempio II.4.6) è sicuramente una FdD sullo spazio prodotto  $(\Omega, \mathcal{F}) = (\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$ . Il Teorema II.4.3 ci permette allora di dire che esiste un'unica probabilità  $\mathbf{P}$  su  $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$  associata a tale  $F$ , sicché per dimostrare completamente la nostra proposizione non resta che da verificare con un calcolo diretto che la  $\mathbf{P}$  così costruita soddisfa la proprietà di fattorizzabilità delle misure prodotto e che è l'unica ad avere tale proprietà (vedi anche Esempio II.4.12).  $\square$

**II.5.38 Osservazione:** Si ricava subito dalle definizioni e dalla proposizione precedenti che, dato un vett.a.  $X = (\xi_k)_{k=1, \dots, n}$ , le sue componenti  $\xi_k$  risultano indipendenti quando la loro DdP congiunta  $\mathbf{P}_X$  risulta essere il prodotto  $\mathbf{P}_{\xi_1} \otimes \dots \otimes \mathbf{P}_{\xi_n}$  delle rispettive DdP marginali. Riprendendo inoltre la discussione delle Osservazioni II.5.30 e II.5.31 noteremo che l'unico caso in cui è possibile ricostruire la DdP congiunta in modo univoco partendo dalle DdP marginali è quello in cui le componenti di  $X$  risultano indipendenti. Infatti in questo caso basterà costruire la misura prodotto  $\mathbf{P}_{\xi_1} \otimes \dots \otimes \mathbf{P}_{\xi_n}$  per ottenere la richiesta DdP congiunta: ogni altra DdP congiunta che ammetta le medesime marginali non può che coincidere, sulla base della Proposizione II.5.37, con la suddetta probabilità prodotto.  $\circ$

**II.5.39 Definizione:** Dato uno spazio di probabilità  $(\Omega, \mathcal{F}, \mathbf{P})$  e una famiglia  $(\mathcal{F}_k)_{k=1, \dots, n}$  di sotto  $\sigma$ -algebre di  $\mathcal{F}$ , diremo che le  $\mathcal{F}_k$  sono  **$\sigma$ -algebre indipendenti** se

$$\mathbf{P}(A_1 \cap \dots \cap A_n) = \mathbf{P}_1(A_1) \cdot \dots \cdot \mathbf{P}_n(A_n)$$

comunque scelti gli eventi  $A_k \in \mathcal{F}_k$  con  $k = 1, \dots, n$ .  $\triangle$

**II.5.40 Proposizione:** Le componenti di un vett.a.  $X = (\xi_k)_{k=1, \dots, n}$  definito su  $(\Omega, \mathcal{F}, \mathbf{P})$  sono indipendenti se e solo se risultano indipendenti le  $\sigma$ -algebre  $\mathcal{F}_{\xi_k}$  da esse generate.

**Dimostrazione:** Supponiamo innanzitutto che le  $\sigma$ -algebre  $(\mathcal{F}_{\xi_k})_{k=1, \dots, n}$  siano indipendenti e proviamo che  $\mathbf{P}_X = \mathbf{P}_{\xi_1} \otimes \dots \otimes \mathbf{P}_{\xi_n}$ . Se infatti consideriamo  $B = B_1 \times \dots \times B_n \in \mathcal{B}(\mathbf{R}^n)$  con  $B_k \in \mathcal{B}(\mathbf{R}_k)$  e  $k = 1, \dots, n$ , definite le proiezioni  $\pi_k : \mathbf{R}^n \rightarrow \mathbf{R}_k$  mediante la relazione  $\pi_k(x) = x_k$  con  $x \in \mathbf{R}^n$ , si vede facilmente che  $\pi_k^{-1}(B_k) = \mathbf{R}_1 \times \dots \times B_k \times \dots \times \mathbf{R}_n$  e quindi anche che

$$B = B_1 \times \dots \times B_n = \pi_1^{-1}(B_1) \cap \dots \cap \pi_n^{-1}(B_n) = \bigcap_{k=1}^n \pi_k^{-1}(B_k).$$

Ne segue allora che, essendo  $\xi_k = \pi_k \circ X$  e quindi  $\xi_k^{-1} = X^{-1} \circ \pi_k^{-1}$ , risulta

$$\begin{aligned} \mathbf{P}_X(B) &= \mathbf{P}_X\left(\bigcap_{k=1}^n \pi_k^{-1}(B_k)\right) = \mathbf{P}\left[X^{-1}\left(\bigcap_{k=1}^n \pi_k^{-1}(B_k)\right)\right] \\ &= \mathbf{P}\left[\bigcap_{k=1}^n X^{-1}(\pi_k^{-1}(B_k))\right] = \mathbf{P}\left[\bigcap_{k=1}^n \xi_k^{-1}(B_k)\right] = \prod_{k=1}^n \mathbf{P}(\xi_k^{-1}(B_k)) \\ &= \prod_{k=1}^n \mathbf{P}_{\xi_k}(B_k) \end{aligned}$$

dove abbiamo usato il fatto che  $\xi_k^{-1}(B_k) \in \mathcal{F}_k$  e che le  $\mathcal{F}_{\xi_k}$  sono indipendenti. Ciò prova che  $\mathbf{P}_X = \mathbf{P}_{\xi_1} \otimes \dots \otimes \mathbf{P}_{\xi_n}$  e quindi, sulla base dell'Osservazione II.5.38, che le componenti di  $X$  sono indipendenti.

Supponiamo ora, viceversa, che  $\mathbf{P}_X = \mathbf{P}_{\xi_1} \otimes \dots \otimes \mathbf{P}_{\xi_n}$  (cioè che le componenti di  $X$  siano indipendenti) e proviamo che le  $\sigma$ -algebre generate sono anche indipendenti. Dati infatti gli eventi  $A_k \in \mathcal{F}_{\xi_k}$ , siano  $B_k \in \mathcal{B}(\mathbf{R}_k)$  gli insiemi tali che  $A_k = \xi_k^{-1}(B_k)$  (che esistono dato che le  $\mathcal{F}_{\xi_k}$  sono generate dalle  $\xi_k$ ). Si ottiene allora che

$$\begin{aligned} \mathbf{P}(A_1 \cap \dots \cap A_n) &= \mathbf{P}\left(\bigcap_{k=1}^n \xi_k^{-1}(B_k)\right) = \mathbf{P}\left[\bigcap_{k=1}^n X^{-1}(\pi_k^{-1}(B_k))\right] \\ &= \mathbf{P}\left[X^{-1}\left(\bigcap_{k=1}^n \pi_k^{-1}(B_k)\right)\right] = \mathbf{P}_X(B_1 \times \dots \times B_n) \\ &= \prod_{k=1}^n \mathbf{P}_{\xi_k}(B_k) = \prod_{k=1}^n \mathbf{P}(\xi_k^{-1}(B_k)) = \prod_{k=1}^n \mathbf{P}(A_k), \end{aligned}$$

il che prova l'indipendenza delle  $\mathcal{F}_k$ . □

**II.5.41 Teorema:** Dato un vett.a.  $X = (\xi_k)_{k=1, \dots, n}$  su  $(\Omega, \mathcal{F}, \mathbf{P})$  le seguenti due affermazioni sono equivalenti:

- (a) le componenti  $\xi_k$  sono indipendenti;
- (b)  $F_X(x_1, \dots, x_n) = F_{\xi_1}(x_1) \cdot \dots \cdot F_{\xi_n}(x_n)$ ;

esse, inoltre, sono anche equivalenti all'affermazione:

- (c)  $f_X(x_1, \dots, x_n) = f_{\xi_1}(x_1) \cdot \dots \cdot f_{\xi_n}(x_n)$ ,
- se  $F_X$  è a.c. ed è quindi dotata di densità.

**Dimostrazione:** Verifichiamo innanzitutto che (a)  $\implies$  (b): se, infatti, le componenti di  $X$  sono indipendenti la (b) segue immediatamente dal fatto che

$$\begin{aligned} F_X(x_1, \dots, x_n) &= \mathbf{P}_X((-\infty, x_1] \times \dots \times (-\infty, x_n]) = \prod_{k=1}^n \mathbf{P}_{\xi_k}(-\infty, x_k] \\ &= \prod_{k=1}^n F_{\xi_k}(x_k). \end{aligned}$$

Se poi  $F_X$  è a.c. da (b) si ha, essendo a.c. anche le FdD marginali (vedi Osservazione II.5.31),

$$\begin{aligned} F_X(x_1, \dots, x_n) &= F_{\xi_1}(x_1) \cdot \dots \cdot F_{\xi_n}(x_n) = \int_{-\infty}^{x_1} f_{\xi_1}(t_1) dt_1 \dots \int_{-\infty}^{x_n} f_{\xi_n}(t_n) dt_n \\ &= \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_{\xi_1}(t_1) \dots f_{\xi_n}(t_n) dt_1 \dots dt_n, \end{aligned}$$

da cui si ricava che la fdd di  $F_X$  è proprio  $f_{\xi_1}(x_1) \cdot \dots \cdot f_{\xi_n}(x_n)$ . Viceversa se (c) è vera risulta anche vera (b):

$$\begin{aligned} F_X(x_1, \dots, x_n) &= \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_{\xi_1}(t_1) \dots f_{\xi_n}(t_n) dt_1 \dots dt_n \\ &= \int_{-\infty}^{x_1} f_{\xi_1}(t_1) dt_1 \dots \int_{-\infty}^{x_n} f_{\xi_n}(t_n) dt_n = F_{\xi_1}(x_1) \cdot \dots \cdot F_{\xi_n}(x_n). \end{aligned}$$

Non ci resta quindi che da provare che (b)  $\implies$  (a), cioè che, se  $F_X$  si fattorizza nel prodotto delle sue marginali, allora risulta

$$\mathbf{P}(\xi_1 \in B_1, \dots, \xi_n \in B_n) = \mathbf{P}(\xi_1 \in B_1) \cdot \dots \cdot \mathbf{P}(\xi_n \in B_n),$$

comunque scelti i  $B_k \in \mathcal{B}(\mathbf{R}_k)$ . Questa affermazione è innanzitutto vera quando i  $B_k = I_k = (a_k, b_k]$  sono intervalli: infatti, posto per brevità  $(a, b] = (a_1, b_1] \times \dots \times (a_n, b_n]$ , si ha (tenendo conto anche del Teorema II.4.3 e dell'Esempio II.4.12)

$$\begin{aligned} \mathbf{P}(\xi_1 \in I_1, \dots, \xi_n \in I_n) &= \mathbf{P}(X \in I_1 \times \dots \times I_n) = \mathbf{P}(X \in (a, b]) = \mathbf{P}_X(a, b] \\ &= \Delta_{a_1, b_1} \dots \Delta_{a_n, b_n} F_X(x_1, \dots, x_n) \\ &= \prod_{k=1}^n \Delta_{a_k, b_k} F_{\xi_k}(x_k) = \prod_{k=1}^n [F_{\xi_k}(b_k) - F_{\xi_k}(a_k)] \\ &= \prod_{k=1}^n \mathbf{P}_{\xi_k}(a_k, b_k] = \prod_{k=1}^n \mathbf{P}(\xi_k \in I_k). \end{aligned}$$

Si dimostra poi che, se  $B_1$  è un generico elemento di  $\mathcal{B}(\mathbf{R}_1)$  e  $I_2, \dots, I_n$  sono intervalli, risulta ancora

$$\mathbf{P}(\xi_1 \in B_1, \xi_2 \in I_2, \dots, \xi_n \in I_n) = \mathbf{P}(\xi_1 \in B_1) \cdot \prod_{k=2}^n \mathbf{P}(\xi_k \in I_k).$$

Per provarlo si usa il principio degli insiemi appropriati (vedi Proposizione II.2.7): detta  $\mathcal{M}$  la famiglia di elementi  $B_1$  di  $\mathcal{B}(\mathbf{R}_1)$  per i quali la relazione precedente è verificata, ci basterà provare che  $\mathcal{M} = \mathcal{B}(\mathbf{R}_1)$ . Innanzitutto osserviamo che  $\mathcal{M}$  è non vuoto dato che, come abbiamo mostrato prima, tutti gli intervalli di  $\mathcal{B}(\mathbf{R}_1)$

sono elementi di  $\mathcal{M}$ . Data l'additività di  $\mathbf{P}$  ne segue, allora, che anche tutte le unioni finite di intervalli sono elementi di  $\mathcal{M}$  e quindi che

$$\mathcal{A} \subseteq \mathcal{M} \subseteq \mathcal{B}(\mathbf{R}_1)$$

se, come al solito (vedi Esempio II.2.12), con  $\mathcal{A}$  indichiamo l'algebra delle unioni finite di intervalli di  $\mathbf{R}_1$ . Se ora riuscissimo a provare che  $\mathcal{M}$  è una classe monotona (vedi Definizione II.2.8), tenendo conto della Proposizione II.2.11 e delle definizioni di  $\sigma$ -algebra e classe monotona generate da un'algebra, avremmo che

$$\sigma(\mathcal{A}) = \mu(\mathcal{A}) \subseteq \mathcal{M} \subseteq \mathcal{B}(\mathbf{R}_1),$$

e siccome (vedi Esempio II.2.12)  $\sigma(\mathcal{A}) = \mathcal{B}(\mathbf{R}_1)$ , resterebbe anche provato che  $\mathcal{M} = \mathcal{B}(\mathbf{R}_1)$ . Ci resta quindi solo da provare che  $\mathcal{M}$  è una classe monotona: data una successione crescente  $(B_1^k)_{k \in \mathbf{N}}$  di elementi di  $\mathcal{M}$ , con  $B_1^k \uparrow B_1 = \bigcup_k B_1^k$ , dobbiamo provare che  $B_1 \in \mathcal{M}$ . Infatti, introducendo le solite proiezioni  $\pi_i$  con  $i = 1, \dots, n$  come nella Proposizione II.5.40, si ha

$$\begin{aligned} \mathbf{P}(\xi_1 \in B_1, \xi_2 \in I_2, \dots, \xi_n \in I_n) &= \mathbf{P}\left(X \in \bigcup_k B_1^k \times I_2 \times \dots \times I_n\right) \\ &= \mathbf{P}\left(X \in \pi_1^{-1}\left(\bigcup_k B_1^k\right) \cap \pi_2^{-1}(I_2) \cap \dots \cap \pi_n^{-1}(I_n)\right) \\ &= \mathbf{P}\left(X \in \bigcup_k [\pi_1^{-1}(B_1^k) \cap \pi_2^{-1}(I_2) \cap \dots \cap \pi_n^{-1}(I_n)]\right) \\ &= \mathbf{P}\left(X \in \bigcup_k (B_1^k \times I_2 \times \dots \times I_n)\right) \\ &= \mathbf{P}_X\left(\bigcup_k (B_1^k \times I_2 \times \dots \times I_n)\right). \end{aligned}$$

Siccome inoltre la successione  $(B_1^k \times I_2 \times \dots \times I_n)_{k \in \mathbf{N}}$  è anch'essa crescente ed ogni  $B_1^k$  è elemento di  $\mathcal{M}$ , usando la continuità da sotto di  $\mathbf{P}_X$  e di  $\mathbf{P}_{\xi_1}$  (vedi Teorema II.1.13) si ha

$$\begin{aligned} \mathbf{P}(\xi_1 \in B_1, \xi_2 \in I_2, \dots, \xi_n \in I_n) &= \mathbf{P}_X\left(\bigcup_k (B_1^k \times I_2 \times \dots \times I_n)\right) \\ &= \lim_k \mathbf{P}_X(B_1^k \times I_2 \times \dots \times I_n) = \lim_k \mathbf{P}(\xi_1 \in B_1^k, \xi_2 \in I_2, \dots, \xi_n \in I_n) \\ &= \lim_k \mathbf{P}_{\xi_1}(B_1^k) \cdot \prod_{j=2}^n \mathbf{P}(\xi_j \in I_j) = \mathbf{P}_{\xi_1}(B_1) \cdot \prod_{j=2}^n \mathbf{P}(\xi_j \in I_j) \\ &= \mathbf{P}(\xi_1 \in B_1) \cdot \prod_{j=2}^n \mathbf{P}(\xi_j \in I_j), \end{aligned}$$

il che prova che  $B_1 \in \mathcal{M}$  e quindi che  $\mathcal{M}$  è una classe monotona. Ripetendo il ragionamento si può allora provare che, comunque scelti  $B_1$  e  $B_2$  in  $\mathcal{B}(\mathbf{R}_1)$  e  $\mathcal{B}(\mathbf{R}_2)$  si ha

$$\begin{aligned} & \mathbf{P}(\xi_1 \in B_1, \xi_2 \in B_2, \xi_3 \in I_3, \dots, \xi_n \in I_n) \\ &= \mathbf{P}(\xi_1 \in B_1) \cdot \mathbf{P}(\xi_2 \in B_2) \cdot \prod_{j=3}^n \mathbf{P}(\xi_j \in I_j), \end{aligned}$$

e infine, iterando l'operazione, la tesi del teorema può essere completamente dimostrata.  $\square$

**Esempio II.5.42:** Riprendendo l'Esempio II.5.33 osserveremo soltanto che, in base al Teorema II.5.41, possiamo ora affermare che le componenti del vett.a.  $X$  risultano indipendenti mentre quelle del vett.a.  $Y$  non lo sono. Faremo inoltre un secondo esempio di uso del Teorema II.5.41 nella discussione dell'indipendenza di v.a. Supporremo ancora di considerare due vett.a. a due componenti  $X = (\xi_1, \xi_2)$  e  $Y = (\eta_1, \eta_2)$  ambedue dotati di FdD a.c. e quindi dotati di fdd. Più precisamente supporremo che le loro fdd congiunte siano costanti (distribuzione uniforme) in due regioni diverse del piano  $\mathbf{R}^2$ :

$$f_X(x, y) = \begin{cases} 2/\pi, & \text{se } (x, y) \in D; \\ 0, & \text{se } (x, y) \notin D; \end{cases}$$

dove la regione  $D$  di  $\mathbf{R}^2$  è costituita dal semicerchio unitario caratterizzato da  $-1 \leq x \leq +1$  e  $0 \leq y \leq \sqrt{1-x^2}$ ; inoltre

$$f_Y(x, y) = \begin{cases} 1/2, & \text{se } (x, y) \in D'; \\ 0, & \text{se } (x, y) \notin D'; \end{cases}$$

dove ora  $D'$  è il rettangolo di  $\mathbf{R}^2$  caratterizzato da  $-1 \leq x \leq +1$  e  $0 \leq y \leq 1$ . Possiamo ora mostrare che le componenti di  $X$  non sono indipendenti mentre quelle di  $Y$  lo sono. Si ha infatti per le fdd marginali di  $X$  che (vedi Osservazione II.5.31)

$$f_{\xi_1}(x) = \begin{cases} \frac{2}{\pi} \sqrt{1-x^2}, & \text{se } x \in [-1, 1]; \\ 0, & \text{se } x \notin [-1, 1]; \end{cases} \quad f_{\xi_2}(y) = \begin{cases} \frac{4}{\pi} \sqrt{1-y^2}, & \text{se } y \in [0, 1]; \\ 0, & \text{se } y \notin [0, 1]; \end{cases}$$

mentre per le marginali di  $Y$  risulta

$$f_{\eta_1}(x) = \begin{cases} 1/2, & \text{se } x \in [-1, 1]; \\ 0, & \text{se } x \notin [-1, 1]; \end{cases} \quad f_{\eta_2}(y) = \begin{cases} 1, & \text{se } y \in [0, 1]; \\ 0, & \text{se } y \notin [0, 1]. \end{cases}$$

Si vede pertanto che  $f_X(x, y) \neq f_{\xi_1}(x)f_{\xi_2}(y)$ , mentre  $f_Y(x, y) = f_{\eta_1}(x)f_{\eta_2}(y)$ , e quindi, in base al Teorema II.5.41, possiamo concludere che le componenti di  $Y$  sono v.a. indipendenti mentre quelle di  $X$  non lo sono.  $\diamond$



## II.6 Valore d'attesa e integrazione

**II.6.1 Osservazione:** Nel caso di v.a. semplici  $\xi$  definite su uno spazio  $(\Omega, \mathcal{F}, \mathbf{P})$ , cioè v.a. della forma

$$\xi(\omega) = \sum_{k=1}^n x_k I_{A_k}(\omega),$$

con  $A_k \in \mathcal{F}$  disgiunti, il valore d'attesa è stato già definito, con le sue proprietà elementari, in I.6 come

$$\mathbf{E} \xi = \sum_{k=1}^n x_k \mathbf{P}(A_k).$$

In questo capitolo estenderemo questo concetto al caso di v.a. arbitrarie mediante gli strumenti della teoria dell'integrazione secondo Lebesgue<sup>1</sup>. Va subito osservato a questo proposito che la precedente definizione coincide con quella di **integrale secondo Lebesgue** di una funzione semplice,  $\mathcal{F}$ -misurabile e a valori reali, rispetto alla misura  $\mathbf{P}$ :

$$\mathbf{E} \xi = \int_{\Omega} \xi(\omega) \mathbf{P}(d\omega) = \int_{\Omega} \xi d\mathbf{P}.$$

Nell'estensione di questi concetti al caso in cui  $\xi$  non è una funzione semplice giocherà un ruolo essenziale il Teorema II.5.16 che consente di approssimare generiche v.a. mediante successioni di v.a. semplici. Infatti, nel caso in cui  $\xi$  è una generica v.a. non negativa è noto, dal citato Teorema, che esiste sempre una successione (non decrescente) di v.a. semplici e non negative  $(\xi_n)_{n \in \mathbf{N}}$  tale che  $\xi_n(\omega) \uparrow \xi(\omega)$  comunque scelto  $\omega \in \Omega$  (cioè in senso puntuale). Se allora consideriamo la successione numerica a termini non negativi  $(\mathbf{E} \xi_n)_{n \in \mathbf{N}}$ , siccome  $\xi_n \leq \xi_{n+1}$ , dalla Proposizione I.6.6 si ricava che essa è non decrescente e quindi ammette sicuramente un limite (eventualmente anche  $+\infty$ ). Questa osservazione ci consente di dare allora una definizione di valore d'attesa della quale, però, bisognerà anche provare la consistenza nel senso che dovremo provare che il valore d'attesa definito tramite questo metodo è indipendente dalla scelta delle successioni di v.a. semplici usate per approssimare  $\xi$ .  $\circ$

**II.6.2 Definizione:** Diremo **valore d'attesa (VdA) di una v.a. non negativa**  $\xi$  il numero (eventualmente anche  $+\infty$ )

$$\mathbf{E} \xi = \int_{\Omega} \xi d\mathbf{P} = \lim_n \mathbf{E} \xi_n$$

dove  $(\xi_n)_{n \in \mathbf{N}}$  è una successione di v.a. semplici tale che  $\xi_n \uparrow \xi$  comunque sia scelto  $\omega \in \Omega$  (in senso puntuale).  $\triangle$

---

<sup>1</sup> Per maggiori dettagli sulla teoria dell'integrazione vedi ad esempio **M. Métivier:** *Notions Fondamentales de la Théorie des Probabilités*; Dunod, Paris, 1972.

**II.6.3 Lemma:** Se  $\eta$  e  $\xi_n$  (con  $n \in \mathbf{N}$ ) sono v.a. semplici e non negative, e se  $\xi_n \uparrow \xi \geq \eta$  in senso puntuale, allora risulta anche  $\lim_n \mathbf{E} \xi_n \geq \mathbf{E} \eta$ .

**Dimostrazione:** Dato un arbitrario  $\epsilon > 0$ , consideriamo la successione di eventi  $A_n = \{\xi_n \geq \eta - \epsilon\}$ . Siccome  $\xi_n \uparrow \xi \geq \eta > \eta - \epsilon$  (in senso puntuale), è chiaro che risulta anche  $A_n \uparrow \Omega$  e  $\bar{A}_n \downarrow \emptyset$ . Inoltre, dal fatto che  $I_{A_n} + I_{\bar{A}_n} = 1$  (con  $n \in \mathbf{N}$ ) e che le  $\xi_n$  sono non negative, si ha che

$$\xi_n = \xi_n I_{A_n} + \xi_n I_{\bar{A}_n} \geq \xi_n I_{A_n} \geq (\eta - \epsilon) I_{A_n}, \quad \forall \omega \in \Omega$$

dato che, per definizione di  $A_n$ , risulta sempre  $\xi_n \geq \eta - \epsilon$  su tutto  $A_n$ . Dalle proprietà elementari dei VdA di v.a. semplici, posto per brevità

$$m = \max_{\omega \in \Omega} \eta(\omega)$$

(che esiste finito dato che  $\eta$  è una v.a. semplice), si ha ora che

$$\begin{aligned} \mathbf{E} \xi_n &\geq \mathbf{E} [(\eta - \epsilon) I_{A_n}] = \mathbf{E} (\eta I_{A_n}) - \epsilon \mathbf{P}(A_n) \\ &= \mathbf{E} \eta - \mathbf{E} (\eta I_{\bar{A}_n}) - \epsilon \mathbf{P}(A_n) \geq \mathbf{E} \eta - m \mathbf{P}(\bar{A}_n) - \epsilon \end{aligned}$$

dove abbiamo usato il fatto che  $\mathbf{E} (\eta I_{\bar{A}_n}) \leq m \mathbf{E} (I_{\bar{A}_n}) = m \mathbf{P}(\bar{A}_n)$  e che  $\mathbf{P}(A_n) \leq 1$ . Siccome  $\mathbf{P}(\bar{A}_n) \downarrow 0$  (dato che  $\bar{A}_n \downarrow \emptyset$  e  $\mathbf{P}$  è  $\sigma$ -additiva, vedi Teorema II.1.13), risulta  $\lim_n \mathbf{E} \xi_n \geq \mathbf{E} \eta - \epsilon$ , ed essendo  $\epsilon$  arbitrario si ha  $\lim_n \mathbf{E} \xi_n \geq \mathbf{E} \eta$ .  $\square$

**II.6.4 Osservazione:** Il lemma precedente ci mette in condizione di verificare la coerenza della Definizione II.6.2. Infatti, supponendo di considerare due diverse successioni di v.a. semplici e non negative tali che  $\xi_n \uparrow \xi$  e  $\eta_m \uparrow \xi$  (in senso puntuale), presa un'arbitraria  $\eta_m$  (con  $m$  arbitrario ma fissato) avremo che  $\xi_n \uparrow \xi \geq \eta_m$  e quindi

$$\lim_n \mathbf{E} \xi_n \geq \mathbf{E} \eta_m.$$

Data l'arbitrarietà di  $m$  risulta allora che

$$\lim_n \mathbf{E} \xi_n \geq \lim_m \mathbf{E} \eta_m.$$

Per ragioni di simmetria, però, questo ragionamento può essere ripetuto invertendo i ruoli sicché risulta anche che

$$\lim_m \mathbf{E} \eta_m \geq \lim_n \mathbf{E} \xi_n$$

e quindi in definitiva che

$$\lim_n \mathbf{E} \xi_n = \lim_m \mathbf{E} \eta_m.$$

Una volta definito il VdA per v.a. non negative, possiamo poi passare a discutere il caso di v.a. generiche ricordando che esse sono sempre suscettibili di essere

rappresentate come differenze di v.a. non negative (vedi Esempio II.5.10 e Teorema II.5.16) nella forma  $\xi = \xi^+ - \xi^-$ . ○

**II.6.5 Definizione:** Diremo che il **valore d'attesa (VdA)** di una v.a.  $\xi$  esiste (o che è definito) quando  $\min\{\mathbf{E}\xi^+, \mathbf{E}\xi^-\} < +\infty$ , cioè quando almeno uno dei due numeri (non negativi)  $\mathbf{E}\xi^+$ ,  $\mathbf{E}\xi^-$  è finito; in tal caso si definisce

$$\mathbf{E}\xi = \mathbf{E}\xi^+ - \mathbf{E}\xi^- = \int_{\Omega} \xi d\mathbf{P}.$$

Se ambedue i numeri  $\mathbf{E}\xi^+$ ,  $\mathbf{E}\xi^-$  sono finiti, anche  $\mathbf{E}\xi$  risulta finito e si dice che  $\xi$  è **integrabile**, cioè ha VdA finito. Si definiscono inoltre **momenti di ordine**  $r$  (con  $r > 0$  e intero) di una v.a.  $\xi$  i numeri (se esistono)

$$\mathbf{E}\xi^r = \int_{\Omega} \xi^r d\mathbf{P},$$

mentre i numeri

$$\mathbf{E}|\xi|^r = \int_{\Omega} |\xi|^r d\mathbf{P},$$

prendono il nome di **momenti assoluti di ordine**  $r$ . Siccome inoltre risulta  $|\xi| = \xi^+ + \xi^-$ , quando  $\xi$  è integrabile essa è anche **assolutamente integrabile** nel senso che  $\mathbf{E}|\xi|$  esiste finito ed è facile vedere che vale anche il viceversa, sicché una v.a. è integrabile se e solo se essa è assolutamente integrabile. Infine si dice **varianza (var.)** della v.a.  $\xi$  il numero  $\mathbf{V}\xi = \mathbf{E}(\xi - \mathbf{E}\xi)^2$ . △

**II.6.6 Osservazione:** Il concetto di VdA di una v.a. definito come integrale di Lebesgue di una funzione  $\mathcal{F}$ -misurabile può ora essere generalizzato: se  $\xi$  è una v.a. estesa (Definizione II.5.12) e  $\mu$  una generica misura su  $(\Omega, \mathcal{F})$  (Definizione II.1.10), è possibile riprodurre per l'integrale di Lebesgue la procedura di definizione del VdA praticamente senza variazioni. Si definisce prima l'integrale di Lebesgue nel caso di funzioni non negative e poi si estende la definizione al caso generale mediante la decomposizione in parti positiva e negativa. In questo modo è possibile parlare di integrale

$$\int_{\Omega} \xi(\omega) \mu(d\omega) = \int_{\Omega} \xi d\mu$$

in situazioni molto generali, ma val la pena di osservare che, nei casi in cui  $\mu$  non è una probabilità, tale integrale non può essere interpretato come un VdA. Un interessante caso particolare si ha quando  $(\Omega, \mathcal{F}) = (\mathbf{R}, \mathcal{B}(\mathbf{R}))$ , cioè quando siamo nel caso di funzioni di Borel  $\xi : (\mathbf{R}, \mathcal{B}(\mathbf{R})) \rightarrow (\mathbf{R}, \mathcal{B}(\mathbf{R}))$ : se la misura adottata è una probabilità  $\mathbf{P}$  essa sarà data tramite la FdD  $F(x)$  e il VdA assumerà la forma

$$\int_{\mathbf{R}} \xi(x) F(dx) = \int_{\mathbf{R}} \xi dF.$$

Se invece consideriamo una generica misura di Lebesgue-Stieltjes (vedi Definizione II.3.6) caratterizzata da una FdDG  $G(x)$ , l'integrale

$$\int_{\mathbf{R}} \xi(x) G(dx) = \int_{\mathbf{R}} \xi dG$$

prende il nome di **integrale di Lebesgue-Stieltjes** (L-S); se poi la misura scelta è proprio la misura di Lebesgue  $\lambda$  su  $\mathbf{R}$  (Esempio II.3.9) allora l'integrale

$$\int_{\mathbf{R}} \xi(x) dx$$

prende il nome di **integrale di Lebesgue** (L). Queste notazioni si estendono poi facilmente anche al caso in cui  $(\Omega, \mathcal{F}) = (\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$ : se  $\mathbf{P}$  è una probabilità su  $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$  con FdD  $F(x)$  ( $x \in \mathbf{R}^n$ ), si ha

$$\mathbf{E} \xi = \int_{\mathbf{R}^n} \xi d\mathbf{P} = \int_{\mathbf{R}^n} \xi(x_1, \dots, x_n) F(dx_1, \dots, dx_n) = \int_{\mathbf{R}^n} \xi dF;$$

se poi  $\mu$  è una misura di (L-S) su  $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$  con FdDG  $G(x)$  ( $x \in \mathbf{R}^n$ ), l'espressione

$$\int_{\mathbf{R}^n} \xi d\mu = \int_{\mathbf{R}^n} \xi(x_1, \dots, x_n) G(dx_1, \dots, dx_n) = \int_{\mathbf{R}^n} \xi dG$$

prende il nome di integrale di (L-S); se infine  $\lambda$  è la misura di (L) su  $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$ , allora

$$\int_{\mathbf{R}^n} \xi d\lambda = \int_{\mathbf{R}^n} \xi(x_1, \dots, x_n) dx_1, \dots, dx_n = \int_{\mathbf{R}^n} \xi dx$$

prende il nome di integrale di (L). Notiamo infine che, per il momento, i VdA di v.a.  $\xi, \eta, \dots$  da  $(\Omega, \mathcal{F})$  in  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$  sono definiti rispetto alla misura  $\mathbf{P}$  assegnata sullo spazio  $(\Omega, \mathcal{F})$ : vedremo in seguito come è possibile calcolare le medesime quantità utilizzando invece le DdP  $\mathbf{P}_\xi, \mathbf{P}_\eta, \dots$  (vedi Definizione II.5.4) Proiettate su  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$  dalle v.a. in questione.  $\circ$

**II.6.7 Osservazione:** Fino a questo punto abbiamo parlato sempre di integrali estesi a tutto  $\Omega$  (o in particolare a  $\mathbf{R}$  o  $\mathbf{R}^n$ ), ma è anche possibile definire integrali estesi a sottinsiemi di  $\Omega$ : se  $A \in \mathcal{F}$  è un sottinsieme di  $\Omega$ , si suole chiamare **integrale di Lebesgue esteso all'insieme  $A$**  l'integrale (nei due casi di una probabilità  $\mathbf{P}$  o di una generica misura  $\mu$ )

$$\int_A \xi d\mathbf{P} = \int_{\Omega} \xi I_A d\mathbf{P} = \mathbf{E}(\xi I_A), \quad \int_A \xi d\mu = \int_{\Omega} \xi I_A d\mu.$$

In particolare se  $\mu$  è una misura di (L-S) su  $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$  con FdDG  $G(x)$  ( $x \in \mathbf{R}^n$ ), e se  $A = (a_1, b_1] \times \dots \times (a_n, b_n]$ , scriveremo

$$\int_A \xi d\mu = \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} \xi(x_1, \dots, x_n) G(dx_1, \dots, dx_n),$$

mentre se  $\lambda$  è la misura di (L) su  $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$  scriveremo

$$\int_A \xi d\lambda = \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} \xi(x_1, \dots, x_n) dx_1, \dots, dx_n.$$

Nel caso di misure di probabilità  $\mathbf{P}$  su  $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$  con FdD  $F(x)$  avremo infine

$$\int_A \xi d\mathbf{P} = \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} \xi(x_1, \dots, x_n) F(dx_1, \dots, dx_n) = \mathbf{E}(\xi I_A).$$

Queste considerazioni ci consentono anche di cogliere le relazioni che intercorrono tra VdA di v.a. e probabilità di eventi e di dare una prima risposta alle questioni sollevate nell'Osservazione II.5.34: data la definizione di VdA per v.a. semplici (Osservazione II.6.1), nel caso di un indicatore  $I_A$  con  $A \in \mathcal{F}$  si ha  $\mathbf{E}(I_A) = \mathbf{P}(A)$ ; siccome però

$$\mathbf{E}(I_A) = \int_{\Omega} I_A d\mathbf{P} = \int_A d\mathbf{P}$$

perveniamo facilmente alla conclusione che

$$\mathbf{P}(A) = \int_A d\mathbf{P},$$

formula che consente il calcolo della probabilità di un evento mediante un integrale. In particolare, se  $(\Omega, \mathcal{F}) = (\mathbf{R}, \mathcal{B}(\mathbf{R}))$  e se  $F(x)$  è la FdD di  $\mathbf{P}$  avremo che la probabilità di un generico evento  $A$  si potrà calcolare mediante l'integrale

$$\mathbf{P}(A) = \int_A F(dx).$$

Se poi  $A = (a, b]$  è un intervallo l'espressione assume la forma

$$\mathbf{P}(A) = \int_a^b F(dx).$$

In modo analogo si mostra poi che

$$\mathbf{P}(A) = \int_A F(dx_1, \dots, dx_n) = \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} F(dx_1, \dots, dx_n)$$

è la formula appropriata nel caso in cui  $(\Omega, \mathcal{F}) = (\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$ . ○

**II.6.8 Proposizione:** Se  $\xi$  e  $\eta$  sono v.a. su  $(\Omega, \mathcal{F}, \mathbf{P})$ :

- (A) se  $\mathbf{E}\xi$  esiste e  $c \in \mathbf{R}$ , allora  $\mathbf{E}(c\xi) = c\mathbf{E}\xi$ ;
- (B) se  $\xi \leq \eta$ ,  $\forall \omega \in \Omega$ , allora  $\mathbf{E}\xi \leq \mathbf{E}\eta$ , con la precisazione che, se  $-\infty < \mathbf{E}\xi$  ( $\mathbf{E}\eta < +\infty$ ), l'esistenza di  $\mathbf{E}\eta$  ( $\mathbf{E}\xi$ ) è automaticamente garantita e la disuguaglianza è verificata;
- (C) se  $\mathbf{E}\xi$  esiste, allora  $|\mathbf{E}\xi| \leq \mathbf{E}|\xi|$ ;
- (D) se  $\mathbf{E}\xi$  esiste (finito), allora anche  $\mathbf{E}(\xi I_A)$  esiste (finito) con  $A \in \mathcal{F}$  arbitrario;
- (E) se  $\xi \geq 0$  e  $\eta \geq 0$  oppure se  $\mathbf{E}|\xi| < +\infty$  e  $\mathbf{E}|\eta| < +\infty$ , allora  $\mathbf{E}(\xi + \eta) = \mathbf{E}\xi + \mathbf{E}\eta$ .

**Dimostrazione:**

- (A) La prova è immediata per v.a. semplici. Se poi  $\xi \geq 0$ , sia  $(\xi_n)_{n \in \mathbf{N}}$  la successione di v.a. semplici che la approssima in base al Teorema II.5.16: se allora  $c \geq 0$ , si avrà anche che  $c\xi_n \uparrow c\xi$  e quindi che  $\mathbf{E}(c\xi) = \lim_n \mathbf{E}(c\xi_n) = c \lim_n \mathbf{E}\xi_n = c\mathbf{E}\xi$ . Nel caso generale si usa la rappresentazione  $\xi = \xi^+ - \xi^-$  e si nota che  $(c\xi)^+ = c\xi^+$ ,  $(c\xi)^- = c\xi^-$  quando  $c \geq 0$ , mentre  $(c\xi)^+ = -c\xi^-$ ,  $(c\xi)^- = -c\xi^+$  quando  $c < 0$ : in ogni caso avremo a che fare con v.a. non negative moltiplicate per costanti non negative alle quali si applica il ragionamento precedente.
- (B) Se  $0 \leq \xi \leq \eta$  la tesi deriva immediatamente dal Lemma II.6.3 in quanto in questo caso i VdA  $\mathbf{E}\xi$  e  $\mathbf{E}\eta$  sicuramente esistono (eventualmente infiniti); se allora  $(\xi_n)_{n \in \mathbf{N}}$  e  $(\eta_n)_{n \in \mathbf{N}}$  sono successioni tali che  $\xi_n \uparrow \xi$  e  $\eta_n \uparrow \eta$ , comunque fissato  $m \in \mathbf{N}$  si avrà  $\mathbf{E}\eta = \lim_n \mathbf{E}\eta_n \geq \mathbf{E}\xi_m$  e quindi anche  $\mathbf{E}\eta \geq \mathbf{E}\xi$ . Nel caso generale, invece, si osserva che, se  $\xi \leq \eta$ , risulta<sup>2</sup>  $\xi^+ \leq \eta^+$  e  $\xi^- \geq \eta^-$ , e pertanto anche  $\mathbf{E}\xi^+ \leq \mathbf{E}\eta^+$  e  $\mathbf{E}\xi^- \geq \mathbf{E}\eta^-$ . Supponendo ora che  $\mathbf{E}\xi$  esista, dobbiamo esaminare varie possibilità: se  $\mathbf{E}\xi = -\infty$ , allora  $\mathbf{E}\xi^- = +\infty$  (e  $\mathbf{E}\xi^+ < +\infty$ ), per cui nulla garantisce che  $\mathbf{E}\eta$  esista, ma se esiste il risultato è banale; se  $-\infty < \mathbf{E}\xi < +\infty$ , allora  $\mathbf{E}\xi^+ < +\infty$  (e  $\mathbf{E}\xi^- < +\infty$ ) e quindi anche  $\mathbf{E}\eta^- < +\infty$ , per cui  $\mathbf{E}\eta$  sicuramente esiste e si ha

$$\mathbf{E}\xi = \mathbf{E}\xi^+ - \mathbf{E}\xi^- \leq \mathbf{E}\eta^+ - \mathbf{E}\eta^- = \mathbf{E}\eta;$$

se infine  $\mathbf{E}\xi = +\infty$  allora  $\mathbf{E}\xi^+ = +\infty$  (e  $\mathbf{E}\xi^- < +\infty$ ), per cui anche  $\mathbf{E}\eta^+ = +\infty$  e  $\mathbf{E}\eta^- < +\infty$ , sicché  $\mathbf{E}\eta = +\infty$  e il risultato è banale. La discussione fatta partendo dall'ipotesi che esista  $\mathbf{E}\eta$  è analoga e quindi, con l'eccezione del caso  $\mathbf{E}\xi = -\infty$  ( $\mathbf{E}\eta = +\infty$ ) in cui l'esistenza di  $\mathbf{E}\xi$  ( $\mathbf{E}\eta$ ) non garantisce l'esistenza di  $\mathbf{E}\eta$  ( $\mathbf{E}\xi$ ) che va supposta espressamente, la tesi è sempre verificata nel senso che l'esistenza di uno dei due VdA implica l'esistenza dell'altro con la maggiorazione richiesta.

- (C) La dimostrazione deriva direttamente da (A) e (B) tenendo conto del fatto che  $-|\xi| \leq \xi \leq |\xi|$ , cioè che  $-\mathbf{E}|\xi| \leq \mathbf{E}\xi \leq \mathbf{E}|\xi|$ .
- (D) La dimostrazione deriva direttamente da (B) osservando che  $(\xi I_A)^\pm = \xi^\pm I_A \leq \xi^\pm$ .
- (E) Se  $\xi \geq 0$  e  $\eta \geq 0$ , dette  $(\xi_n)_{n \in \mathbf{N}}$  e  $(\eta_n)_{n \in \mathbf{N}}$  le successioni tali che  $\xi_n \uparrow \xi$  e  $\eta_n \uparrow \eta$ , risulta  $\mathbf{E}(\xi_n + \eta_n) = \mathbf{E}\xi_n + \mathbf{E}\eta_n$ , comunque scelto  $n \in \mathbf{N}$ , sicché dalla definizione di VdA di una v.a. non negativa e dal fatto che  $\xi_n + \eta_n \uparrow \xi + \eta$ , si ha che

$$\mathbf{E}(\xi_n + \eta_n) \uparrow \mathbf{E}(\xi + \eta), \quad \mathbf{E}\xi_n \uparrow \mathbf{E}\xi, \quad \mathbf{E}\eta_n \uparrow \mathbf{E}\eta,$$

e quindi anche  $\mathbf{E}(\xi + \eta) = \mathbf{E}\xi + \mathbf{E}\eta$ . Se invece si ha che  $\mathbf{E}|\xi| < +\infty$  e  $\mathbf{E}|\eta| < +\infty$ , le funzioni  $\xi^\pm$  e  $\eta^\pm$  hanno VdA finiti e la discussione si riconduce al caso precedente.  $\square$

<sup>2</sup> Ricorda, a questo proposito, che nella decomposizione in parti positiva e negativa almeno una delle due funzioni è nulla in ogni  $\omega \in \Omega$ .

**II.6.9 Proposizione:** Se  $\xi$  e  $\eta$  sono v.a. su  $(\Omega, \mathcal{F}, \mathbf{P})$ :

- (F) se  $\xi = 0$  ( $\mathbf{P}$ -q.o.), allora  $\mathbf{E} \xi = 0$ ; in particolare, se  $D \in \mathcal{F}$  con  $\mathbf{P}(D) = 0$  e  $\xi$  è una v.a., allora  $\mathbf{E}(\xi I_D) = \int_D \xi d\mathbf{P} = 0$ ;
- (G) se  $\xi = \eta$  ( $\mathbf{P}$ -q.o.) e  $\mathbf{E}|\xi| < +\infty$ , allora anche  $\mathbf{E}|\eta| < +\infty$  e  $\mathbf{E} \xi = \mathbf{E} \eta$ ;
- (H) se  $\xi \geq 0$  e  $\mathbf{E} \xi = 0$ , allora  $\xi = 0$  ( $\mathbf{P}$ -q.o.);
- (I) se  $\mathbf{E}|\xi| < +\infty$ ,  $\mathbf{E}|\eta| < +\infty$  e  $\mathbf{E}(\xi I_A) \leq \mathbf{E}(\eta I_A)$  ( $\forall A \in \mathcal{F}$ ), allora  $\xi \leq \eta$  ( $\mathbf{P}$ -q.o.); in particolare se  $\mathbf{E}(\xi I_A) = \mathbf{E}(\eta I_A)$  ( $\forall A \in \mathcal{F}$ ), allora  $\xi = \eta$  ( $\mathbf{P}$ -q.o.);
- (J) se  $\xi$  è una v.a. estesa e  $\mathbf{E}|\xi| < +\infty$ , allora  $|\xi| < +\infty$  ( $\mathbf{P}$ -q.o.).

**Dimostrazione:** Nota che le proprietà dei VdA enunciate in questa Proposizione si distinguono da quelle della precedente per il fatto che alcune delle loro affermazioni sono vere solo  $\mathbf{P}$ -q.o.

(F) Se  $\xi$  è una v.a. semplice

$$\xi = \sum_{k=1}^n x_k I_{A_k}, \quad (x_k \neq 0)$$

l'ipotesi  $\xi = 0$  ( $\mathbf{P}$ -q.o.) richiede che  $\mathbf{P}(A_k) = 0$  ( $k = 1, \dots, n$ ); ne segue per definizione che  $\mathbf{E} \xi = 0$ . Se invece  $\xi$  è una v.a. non negativa e  $(\xi_n)_{n \in \mathbf{N}}$  è la successione di v.a. tale che  $\xi_n \uparrow \xi$ , l'ipotesi che  $\xi = 0$  ( $\mathbf{P}$ -q.o.) implica che anche  $\xi_n = 0$  ( $\mathbf{P}$ -q.o.), e quindi che  $\mathbf{E} \xi_n = 0$  comunque scelto  $n \in \mathbf{N}$ . Ne segue che  $\mathbf{E} \xi = \lim_n \mathbf{E} \xi_n = 0$ . Infine nel caso generale la prova deriva dal fatto che  $\xi = \xi^+ - \xi^-$  e che  $\xi^+ \leq |\xi|$ ,  $\xi^- \leq |\xi|$  con  $|\xi| = 0$  ( $\mathbf{P}$ -q.o.).

- (G) Se  $D = \{\xi \neq \eta\}$ , l'ipotesi fatta richiede che  $\mathbf{P}(D) = 0$ ; pertanto da (E) ed (F) si ha  $\mathbf{E} \xi = \mathbf{E}(\xi I_D + \xi I_{\bar{D}}) = \mathbf{E}(\xi I_D) + \mathbf{E}(\xi I_{\bar{D}}) = \mathbf{E}(\xi I_{\bar{D}}) = \mathbf{E}(\eta I_{\bar{D}})$ . Siccome però si ha anche  $\mathbf{E}(\eta I_D) = 0$  si ha in definitiva che  $\mathbf{E} \xi = \mathbf{E}(\eta I_D) + \mathbf{E}(\eta I_{\bar{D}}) = \mathbf{E}(\eta I_D + \eta I_{\bar{D}}) = \mathbf{E} \eta$ .
- (H) Se poniamo  $A = \{\xi > 0\}$  e  $A_n = \{\xi \geq 1/n\}$ , è chiaro che  $A_n \uparrow A$  e che  $0 \leq \xi I_{A_n} \leq \xi I_A$ . Tenendo conto del fatto che  $\bar{A} = \{\xi = 0\}$  (e quindi che  $\xi I_{\bar{A}} = 0$ ), da (B) e dalle ipotesi risulta allora che  $0 \leq \mathbf{E}(\xi I_{A_n}) \leq \mathbf{E}(\xi I_A) = \mathbf{E}(\xi I_A) + \mathbf{E}(\xi I_{\bar{A}}) = \mathbf{E}(\xi I_A + \xi I_{\bar{A}}) = \mathbf{E} \xi = 0$ . Ne segue che  $\mathbf{E}(\xi I_{A_n}) = 0$  ( $\forall n \in \mathbf{N}$ ) e siccome  $\mathbf{E}(\xi I_{A_n}) \geq \frac{1}{n} \mathbf{P}(A_n)$  (dato che su  $A_n$  risulta  $\xi \geq 1/n$ ), anche  $\mathbf{P}(A_n) = 0$  ( $\forall n \in \mathbf{N}$ ). Ma dal Teorema II.1.13 sappiamo che  $\mathbf{P}(A) = \lim_n \mathbf{P}(A_n)$  e quindi  $\mathbf{P}(A) = 0$ , il che definitivamente prova la tesi.
- (I) Posto  $B = \{\xi > \eta\}$  si ha ovviamente  $\mathbf{E}(\eta I_B) \leq \mathbf{E}(\xi I_B)$ , e siccome per ipotesi deve risultare  $\mathbf{E}(\eta I_B) \geq \mathbf{E}(\xi I_B)$  si ricava che  $\mathbf{E}(\eta I_B) = \mathbf{E}(\xi I_B)$ . Da (E) segue allora che  $\mathbf{E}((\xi - \eta) I_B) = 0$  e quindi da (H) che  $(\xi - \eta) I_B = 0$  ( $\mathbf{P}$ -q.o.). Pertanto, dove  $\xi > \eta$  (cioè  $\xi - \eta \neq 0$ ) risulta  $I_B = 0$  ( $\mathbf{P}$ -q.o.) mentre per definizione  $I_B = 0$  dovunque  $\xi \leq \eta$ , per cui in definitiva  $I_B = 0$  ( $\mathbf{P}$ -q.o.) e quindi  $\mathbf{P}(B) = \mathbf{E}(I_B) = 0$ . Se invece per ipotesi risultasse  $\mathbf{E}(\xi I_A) = \mathbf{E}(\eta I_A)$  ( $\forall A \in \mathcal{F}$ ), si pone  $B = \{\xi \neq \eta\} = \{|\xi - \eta| > 0\}$  e siccome  $\mathbf{E}((\xi - \eta) I_A) = \mathbf{E}((\eta - \xi) I_A) = \mathbf{E}(|\xi - \eta| I_A) = 0$  ( $\forall A \in \mathcal{F}$ ), risulta anche  $\mathbf{E}(|\xi - \eta| I_B) = 0$ .

Da (H) si ha allora che  $|\xi - \eta|I_B = 0$  e quindi, come nel caso precedente,  $\mathbf{P}(B) = \mathbf{E}(I_B) = 0$ .

(J) Posto  $A = \{|\xi| = +\infty\}$ , se fosse  $\mathbf{P}(A) \neq 0$  sarebbe anche  $\mathbf{E}|\xi| \geq \mathbf{E}(|\xi|I_A) = +\infty \cdot \mathbf{P}(A) = +\infty$  contrariamente all'ipotesi  $\mathbf{E}|\xi| < +\infty$ .  $\square$

**II.6.10 Teorema (della convergenza monotona):** Date una successione di

v.a.  $(\xi_n)_{n \in \mathbf{N}}$  e due v.a.  $\xi$  ed  $\eta$ :

- (a) se  $\xi_n \uparrow \xi$ , se  $\mathbf{E}\eta > -\infty$  e se  $\xi_n \geq \eta, \forall \omega \in \Omega, \forall n \in \mathbf{N}$ , allora  $\mathbf{E}\xi_n \uparrow \mathbf{E}\xi$ ;  
 (b) se  $\xi_n \downarrow \xi$ , se  $\mathbf{E}\eta < +\infty$  e se  $\xi_n \leq \eta, \forall \omega \in \Omega, \forall n \in \mathbf{N}$ , allora  $\mathbf{E}\xi_n \downarrow \mathbf{E}\xi$ .

**Dimostrazione:** Omessa (A.N. Shiryayev: *Probability*; Springer, New York, 1984, p. 184). Questo Teorema è anche noto nella letteratura come Proprietà di Beppo Levi (1875 - 1928), dal nome del matematico italiano che la enunciò nel 1906.  $\square$

**II.6.11 Corollario:** Data una successione di v.a. non negative  $(\eta_n)_{n \in \mathbf{N}}$  risulta sempre verificata la relazione  $\mathbf{E}(\sum_n \eta_n) = \sum_n (\mathbf{E}\eta_n)$ .

**Dimostrazione:** La tesi segue da (E) di Proposizione II.6.8 e dall'applicazione del Teorema II.6.10 alla successione delle somme parziali per la quale si ha che

$$\xi_k = \sum_{n=1}^k \eta_n \uparrow \sum_n \eta_n = \xi, \quad (k \rightarrow \infty).$$

Nota che il Teorema II.6.10 è applicabile in quanto è sufficiente assumere che  $\eta = 0$  identicamente per avere che  $\mathbf{E}\eta = 0 > -\infty$  e  $\xi_n \geq 0, \forall \omega \in \Omega$  e  $\forall n \in \mathbf{N}$ .  $\square$

**II.6.12 Lemma (di Fatou):** Sia  $(\xi_n)_{n \in \mathbf{N}}$  una successione di v.a. ed  $\eta$  una v.a.:

- (a) se  $\mathbf{E}\eta > -\infty$  e se  $\xi_n \geq \eta, \forall \omega \in \Omega, \forall n \in \mathbf{N}$ , allora  $\mathbf{E}(\underline{\lim} \xi_n) \leq \underline{\lim} \mathbf{E}\xi_n$ ;  
 (b) se  $\mathbf{E}\eta < +\infty$  e se  $\xi_n \leq \eta, \forall \omega \in \Omega, \forall n \in \mathbf{N}$ , allora  $\mathbf{E}(\overline{\lim} \xi_n) \geq \overline{\lim} \mathbf{E}\xi_n$ ;  
 (c) infine, se  $\mathbf{E}\eta < +\infty$  e se  $|\xi_n| \leq \eta, \forall \omega \in \Omega, \forall n \in \mathbf{N}$ , allora è vera la catena di disuguaglianze  $\mathbf{E}(\underline{\lim} \xi_n) \leq \underline{\lim} \mathbf{E}\xi_n \leq \overline{\lim} \mathbf{E}\xi_n \leq \mathbf{E}(\overline{\lim} \xi_n)$ .

**Dimostrazione:** Omessa<sup>3</sup> (A.N. Shiryayev: *Probability*; Springer, New York, 1984, p. 185.)  $\square$

<sup>3</sup> P. Fatou (1878 - 1929) conseguì all'inizio del '900 rilevanti risultati nel settore dell'analisi e dell'integrazione esercitando una notevole influenza sui matematici contemporanei come H. Lebesgue e M. Riesz (1886 - 1969).



**II.6.13 Definizione:** Diremo che la successione  $(\xi_n)_{n \in \mathbf{N}}$  di v.a. **converge quasi ovunque rispetto alla probabilità  $\mathbf{P}$  ( $\mathbf{P}$ -q.o.)** alla v.a.  $\xi$ , e scriveremo

$$\xi_n \xrightarrow{n} \xi \quad (\mathbf{P}\text{-q.o.}) \quad \text{oppure} \quad \xi_n \xrightarrow{q.o.} \xi,$$

se l'insieme dei punti  $\omega \in \Omega$  tali che  $\xi_n(\omega)$  non converge verso  $\xi(\omega)$  ha probabilità zero, cioè se  $\mathbf{P}\{\omega : \xi_n \not\rightarrow \xi\} = 0$ .  $\triangle$

**II.6.14 Teorema (di Lebesgue sulla convergenza maggiorata):** Date una successione di v.a.  $(\xi_n)_{n \in \mathbf{N}}$  e due v.a.  $\xi$  ed  $\eta \geq 0$ , se  $\mathbf{E} \eta < +\infty$ , se  $|\xi_n| \leq \eta$  ( $\forall \omega \in \Omega$  e  $\forall n \in \mathbf{N}$ ) e se  $\xi_n \xrightarrow{n} \xi$  ( $\mathbf{P}$ -q.o.), allora risulta anche che  $\mathbf{E} |\xi| < +\infty$ ,  $\mathbf{E} \xi_n \xrightarrow{n} \mathbf{E} \xi$  e  $\mathbf{E} |\xi_n - \xi| \xrightarrow{n} 0$ . Inoltre, date una successione di v.a.  $(\xi_n)_{n \in \mathbf{N}}$  e due v.a.  $\xi$  ed  $\eta \geq 0$ , se  $|\xi_n| \leq \eta$  ( $\forall \omega \in \Omega$  e  $\forall n \in \mathbf{N}$ ), se esiste  $p > 0$  tale che  $\mathbf{E} \eta^p < +\infty$  e se  $\xi_n \xrightarrow{n} \xi$  ( $\mathbf{P}$ -q.o.), allora risulta anche che  $\mathbf{E} |\xi|^p < +\infty$  e  $\mathbf{E} |\xi_n - \xi|^p \xrightarrow{n} 0$ .

**Dimostrazione:** Omessa (**A.N. Shiryaev:** *Probability*; Springer, New York, 1984, p. 185). Nota che, se in particolare  $p \geq 1$ , siccome risulta  $|\mathbf{E}(\xi_n - \xi)|^p \leq \mathbf{E} |\xi_n - \xi|^p$  (vedi la successiva Diseguaglianza di Jensen II.6.19 con  $g(x) = |x|^p$ ), la seconda parte del Teorema implica anche che  $\mathbf{E} \xi_n \xrightarrow{n} \mathbf{E} \xi$ .  $\square$

**II.6.15 Teorema:** Date due v.a.  $\xi$  ed  $\eta$  indipendenti, se  $\mathbf{E} |\xi| < +\infty$  e  $\mathbf{E} |\eta| < +\infty$ , allora risulta anche che  $\mathbf{E} |\xi \eta| < +\infty$  e  $\mathbf{E} (\xi \eta) = \mathbf{E} \xi \cdot \mathbf{E} \eta$ .

**Dimostrazione:** Consideriamo dapprima il caso in cui  $\xi \geq 0$  e  $\eta \geq 0$ , poniamo

$$A_k^{(n)} = \left\{ \frac{k}{n} \leq \xi < \frac{k+1}{n} \right\}, \quad B_k^{(n)} = \left\{ \frac{k}{n} \leq \eta < \frac{k+1}{n} \right\}; \quad (k \geq 0, n \geq 1),$$

e definiamo le seguenti successioni di v.a. non negative:

$$\xi_n = \sum_{k=0}^{\infty} \frac{k}{n} I_{A_k^{(n)}}, \quad \eta_n = \sum_{k=0}^{\infty} \frac{k}{n} I_{B_k^{(n)}}.$$

Siccome (per ogni  $n$  fissato) gli eventi  $A_k^{(n)}$  sono disgiunti (e analogamente per i  $B_k^{(n)}$ ), in ogni  $\omega \in \Omega$  le serie che definiscono le nostre successioni constano di un solo termine non nullo sicché non ci sono problemi di convergenza. Ciononostante bisogna osservare che, a differenza delle successioni adottate nella dimostrazione del Teorema II.5.16 o nella definizione dei VdA, le  $(\xi_n)_{n \in \mathbf{N}}$  e  $(\eta_n)_{n \in \mathbf{N}}$  non sono composte di v.a. semplici e che l'uso di tali v.a. ci è consentito per il fatto che ora potremo far uso di risultati più avanzati (in particolare il Teorema II.6.14) per

discutere la convergenza della successione dei VdA. Infatti dalle definizioni date si vede subito che

$$|\xi_n| = \xi_n \leq \xi, \quad |\eta_n| = \eta_n \leq \eta, \quad \forall n \in \mathbf{N}, \forall \omega \in \Omega;$$

inoltre le limitazioni  $\mathbf{E} \xi < +\infty$  e  $\mathbf{E} \eta < +\infty$  sono garantite per ipotesi. Infine si vede facilmente che

$$\xi_n \xrightarrow{n} \xi, \quad \eta_n \xrightarrow{n} \eta; \quad (\mathbf{P}\text{-q.o.});$$

infatti, siccome  $\mathbf{E} |\xi| < +\infty$ , la (J) della Proposizione II.6.9 ci assicura che  $\xi$  è finita ( $\mathbf{P}\text{-q.o.}$ ); ma dovunque  $\xi$  è finita risulta anche che ( $\mathbf{P}\text{-q.o.}$ )

$$|\xi - \xi_n| = \xi - \xi_n = \xi - \sum_{k=0}^{\infty} \frac{k}{n} I_{A_k^{(n)}} \leq \sum_{k=0}^{\infty} \frac{k+1}{n} I_{A_k^{(n)}} - \sum_{k=0}^{\infty} \frac{k}{n} I_{A_k^{(n)}} = \frac{1}{n} \sum_{k=0}^{\infty} I_{A_k^{(n)}} = \frac{1}{n},$$

e quindi  $\xi_n \xrightarrow{n} \xi$  ( $\mathbf{P}\text{-q.o.}$ ). Analogamente si prova che  $\eta_n \xrightarrow{n} \eta$  ( $\mathbf{P}\text{-q.o.}$ ). Essendo dunque verificate le ipotesi del Teorema II.6.14 potremo affermare che

$$\mathbf{E} \xi_n \xrightarrow{n} \mathbf{E} \xi, \quad \mathbf{E} \eta_n \xrightarrow{n} \mathbf{E} \eta.$$

Siccome per ipotesi  $\xi$  e  $\eta$  sono indipendenti, tutti gli eventi  $A_k^{(n)}$  sono indipendenti dagli eventi  $B_l^{(n)}$  e pertanto (facendo anche uso del Corollario II.6.11) comunque scelto  $n \in \mathbf{N}$  abbiamo che

$$\mathbf{E} (\xi_n \eta_n) = \sum_{k,l=0}^{\infty} \frac{kl}{n^2} \mathbf{E} (I_{A_k^{(n)}} I_{B_l^{(n)}}) = \sum_{k,l=0}^{\infty} \frac{kl}{n^2} \mathbf{E} (I_{A_k^{(n)}}) \mathbf{E} (I_{B_l^{(n)}}) = \mathbf{E} \xi_n \cdot \mathbf{E} \eta_n.$$

Inoltre risulta anche

$$\mathbf{E} (\xi_n \eta_n) \xrightarrow{n} \mathbf{E} (\xi \eta)$$

in quanto si ha che (essendo  $|\eta - \eta_n| \leq 1/n$ ,  $|\xi - \xi_n| \leq 1/n$  e  $\eta_n \leq \eta$ )

$$\begin{aligned} |\mathbf{E} (\xi_n \eta_n) - \mathbf{E} (\xi \eta)| &\leq \mathbf{E} |\xi_n \eta_n - \xi \eta| \leq \mathbf{E} (|\xi(\eta - \eta_n)| + |\eta_n(\xi - \xi_n)|) \\ &= \mathbf{E} (|\xi| \cdot |\eta - \eta_n|) + \mathbf{E} (|\eta_n| \cdot |\xi - \xi_n|) \leq \frac{1}{n} \mathbf{E} |\xi| + \frac{1}{n} \mathbf{E} |\eta_n| \\ &= \frac{1}{n} \mathbf{E} \xi + \frac{1}{n} \mathbf{E} \eta_n \leq \frac{1}{n} \mathbf{E} \xi + \frac{1}{n} \mathbf{E} \eta = \frac{1}{n} (\mathbf{E} \xi + \mathbf{E} \eta) \end{aligned}$$

e quindi  $|\mathbf{E} (\xi_n \eta_n) - \mathbf{E} (\xi \eta)| \xrightarrow{n} 0$  dato che per ipotesi  $\mathbf{E} \xi$  e  $\mathbf{E} \eta$  sono finiti. In conclusione potremo dire che

$$\mathbf{E} (\xi \eta) = \lim_n \mathbf{E} (\xi_n \eta_n) = \lim_n (\mathbf{E} \xi_n \cdot \mathbf{E} \eta_n) = \left( \lim_n \mathbf{E} \xi_n \right) \left( \lim_m \mathbf{E} \eta_m \right) = \mathbf{E} \xi \cdot \mathbf{E} \eta.$$

Il caso generale si riconduce a questo usando le rappresentazioni  $\xi = \xi^+ - \xi^-$  e  $\eta = \eta^+ - \eta^-$  ed osservando che  $(\xi \eta)^+ = \xi^+ \eta^+ + \xi^- \eta^-$  e  $(\xi \eta)^- = \xi^+ \eta^- + \xi^- \eta^+$ .  $\square$

**II.6.16 Proposizione (Diseguaglianza di Chebyshev):** Se  $\xi$  è una v.a. non negativa la diseguaglianza

$$\mathbf{P}(\xi \geq \epsilon) \leq \frac{\mathbf{E} \xi}{\epsilon}$$

risulta sempre verificata qualunque sia il numero  $\epsilon > 0$ .

**Dimostrazione:** La proposizione discende dalle diseguaglianze

$$\mathbf{E} \xi \geq \mathbf{E} (\xi I_{\{\xi \geq \epsilon\}}) \geq \epsilon \mathbf{E} (I_{\{\xi \geq \epsilon\}}) = \epsilon \mathbf{P}(\xi \geq \epsilon)$$

dove  $\epsilon > 0$  è arbitrario. □

**II.6.17 Corollario:** Se  $\xi$  è una v.a. le diseguaglianze

$$\begin{aligned} \mathbf{P}(|\xi| \geq \epsilon) &\leq \frac{\mathbf{E} \xi^2}{\epsilon^2} \\ \mathbf{P}(|\xi - \mathbf{E} \xi| \geq \epsilon) &\leq \frac{\mathbf{E} (\xi - \mathbf{E} \xi)^2}{\epsilon^2} = \frac{\mathbf{V} \xi}{\epsilon^2} \end{aligned}$$

risultano sempre verificate qualunque sia il numero  $\epsilon > 0$ . Inoltre la diseguaglianza

$$\mathbf{P}(\xi \geq \epsilon) \leq \frac{\mathbf{E} \xi^p}{\epsilon^p}, \quad (p > 0)$$

è sempre verificata se  $\xi \geq 0$ .

**Dimostrazione:** Omessa (da svolgere per esercizio). □

**II.6.18 Proposizione (Diseguaglianza di Schwarz):** Se  $\xi$  e  $\eta$  sono v.a. tali che  $\mathbf{E} \xi^2 < +\infty$  e  $\mathbf{E} \eta^2 < +\infty$ , allora  $\mathbf{E} |\xi \eta| \leq +\infty$  e la diseguaglianza

$$(\mathbf{E} |\xi \eta|)^2 \leq \mathbf{E} \xi^2 \cdot \mathbf{E} \eta^2$$

è sempre verificata.

**Dimostrazione:** Il caso in cui  $\mathbf{E} \xi^2 \neq 0$  e  $\mathbf{E} \eta^2 \neq 0$  si tratta esattamente come nell'analoga proposizione elementare (vedi Proposizione I.6.12). Se invece, ad esempio,  $\mathbf{E} \xi^2 = 0$ , da (I) di Proposizione II.6.9 risulta  $\xi = 0$  ( $\mathbf{P}$ -q.o.) e quindi, da (F) della medesima Proposizione, anche  $\mathbf{E} |\xi \eta| = 0$  e pertanto la tesi è banalmente verificata anche in questo caso. □

**II.6.19 Proposizione (Diseguaglianza di Jensen):** Se  $g(x)$  è una funzione di Borel convessa (verso il basso) e se  $\xi$  è una v.a. con  $\mathbf{E}|\xi| < +\infty$ , allora la diseguaglianza

$$g(\mathbf{E}\xi) \leq \mathbf{E}(g(\xi))$$

è sempre verificata.

**Dimostrazione:** La diseguaglianza di Jensen<sup>4</sup> è una proprietà molto generale (si applica ad ogni funzione convessa) utile nella deduzione di diverse altre proposizioni. Se  $g(x)$  è convessa (verso il basso), comunque scelto  $x_0 \in \mathbf{R}$  esisterà un numero  $\lambda(x_0)$  tale che

$$g(x) \geq g(x_0) + (x - x_0)\lambda(x_0), \quad \forall x \in \mathbf{R}.$$

Sostituendo  $\mathbf{E}\xi$  ad  $x_0$  e  $\xi$  ad  $x$  si ha allora

$$g(\xi) \geq g(\mathbf{E}\xi) + (\xi - \mathbf{E}\xi)\lambda(\mathbf{E}\xi)$$

e quindi, prendendo i VdA di ambedue i membri,  $g(\mathbf{E}\xi) \leq \mathbf{E}(g(\xi))$ . □

**II.6.20 Corollario (Diseguaglianza di Lyapunov):** Se  $\xi$  è una v.a. e  $0 < s \leq t$ , allora risulta

$$(\mathbf{E}|\xi|^s)^{1/s} \leq (\mathbf{E}|\xi|^t)^{1/t};$$

ne segue in particolare che le diseguaglianze

$$\mathbf{E}|\xi| \leq (\mathbf{E}|\xi|^2)^{1/2} \leq \dots \leq (\mathbf{E}|\xi|^n)^{1/n} \leq \dots$$

sono sempre verificate.

**Dimostrazione:** Posto  $r = t/s \geq 1$  e  $\eta = |\xi|^s$ , si applica la diseguaglianza di Jensen alla funzione  $g(x) = |x|^r$  e si ottiene  $|\mathbf{E}\eta|^r \leq \mathbf{E}|\eta|^r$ , cioè  $(\mathbf{E}|\xi|^s)^{t/s} \leq \mathbf{E}|\xi|^t$  da cui segue immediatamente la tesi. La successiva catena di diseguaglianze non è che un caso particolare della diseguaglianza di Lyapunov<sup>5</sup>. □

<sup>4</sup> Il matematico danese Johan L. Jensen (1859-1925) fu con il suo collega A. K. Erlang (1878-1929) fra i pionieri che introdussero il telefono in Danimarca. Egli fu anche direttore della Compagnia dei Telefoni di Copenhagen dal 1890 al 1925.

<sup>5</sup> A. M. Lyapunov (1857 - 1918) è stato con A. A. Markov (1856 - 1922) uno degli allievi più fecondi di Chebyshev. I suoi contributi si rivelarono fondamentali, in particolare, nella derivazione del Teorema limite centrale, per il quale fornì, servendosi del metodo delle funzioni caratteristiche da lui stesso elaborato, una dimostrazione più funzionale di quella proposta un anno prima da Markov sulla base del metodo dei momenti. Lyapunov si distinse soprattutto nel settore della fisica matematica per le sue ricerche nella teoria delle equazioni differenziali, ponendo le basi della teoria della stabilità del moto. Molto importanti sono anche i suoi studi sulla teoria delle figure d'equilibrio di un fluido in rotazione.

**II.6.21 Proposizione (Diseguaglianza di Hölder):** Se  $1 < p < +\infty$ ,  $1 < q < +\infty$  con  $\frac{1}{p} + \frac{1}{q} = 1$  e se inoltre date due v.a.  $\xi$  ed  $\eta$  si ha  $\mathbf{E} |\xi|^p < +\infty$ ,  $\mathbf{E} |\eta|^q < +\infty$ , allora  $\mathbf{E} |\xi\eta| < +\infty$  e la disequaglianza

$$\mathbf{E} |\xi\eta| \leq (\mathbf{E} |\xi|^p)^{1/p} (\mathbf{E} |\eta|^q)^{1/q}$$

è sempre verificata.

**Dimostrazione:** Osserviamo innanzitutto che la disequaglianza di Schwarz è un caso particolare della disequaglianza di Hölder<sup>6</sup> per  $p = q = 2$ . Inoltre, quando almeno uno dei due numeri  $\mathbf{E} |\xi|^p$ ,  $\mathbf{E} |\eta|^q$  si annulla la dimostrazione è identica a quella dell'analogo caso nella Proposizione II.6.18. Potremo pertanto limitarci al caso in cui  $\mathbf{E} |\xi|^p \neq 0$ ,  $\mathbf{E} |\eta|^q \neq 0$  e porremo

$$\tilde{\xi} = \frac{|\xi|}{(\mathbf{E} |\xi|^p)^{1/p}}, \quad \tilde{\eta} = \frac{|\eta|}{(\mathbf{E} |\eta|^q)^{1/q}},$$

in modo che risulti  $\mathbf{E} |\tilde{\xi}|^p = \mathbf{E} |\tilde{\eta}|^q = 1$ . Osserviamo ora che, se  $a, b, x, y$  sono numeri positivi e  $a + b = 1$ , risulta

$$x^a y^b \leq ax + by.$$

Infatti, siccome  $g(x) = \ln x$  è concava (convessa verso l'alto; vedi a questo proposito l'analogica discussione nella proposizione II.6.19) si ha per ogni  $x_0 > 0$  e  $x > 0$  che

$$\ln x \leq \ln x_0 + \frac{x - x_0}{x_0}.$$

Ne segue, con  $a > 0$ ,  $b > 0$ ,  $y > 0$  e  $y_0 > 0$ , che

$$\begin{cases} a \ln x \leq a \ln x_0 + a x/x_0 - a \\ b \ln y \leq b \ln y_0 + b y/y_0 - b \end{cases}$$

e quindi sommando membro a membro

$$a \ln x + b \ln y \leq a \ln x_0 + b \ln y_0 + a \frac{x}{x_0} + b \frac{y}{y_0} - (a + b).$$

Se ora poniamo  $x_0 = y_0 = ax + by$  si ha

$$a \ln x + b \ln y \leq (a + b) \ln(ax + by) + 1 - (a + b),$$

<sup>6</sup> Il matematico tedesco O. L. Hölder (1859 - 1937) fu particolarmente attivo nel campo della Teoria dei gruppi nell'ambito della quale sviluppò la teoria astratta degli automorfismi. La disequaglianza qui riportata è stata da lui dimostrata nel caso delle serie; l'estensione agli integrali è invece dovuta a F. Riesz (1880 - 1956).

e se  $a + b = 1$  si ottiene

$$\ln(x^a y^b) = a \ln x + b \ln y \leq \ln(ax + by)$$

da cui deriva immediatamente la diseuguaglianza richiesta. Se allora poniamo  $x = \tilde{\xi}^p$ ,  $y = \tilde{\eta}^q$ ,  $a = 1/p$  e  $b = 1/q$  si ha

$$\tilde{\xi}\tilde{\eta} \leq \frac{1}{p}\tilde{\xi}^p + \frac{1}{q}\tilde{\eta}^q$$

e quindi (prendendo i VdA di ambedue i membri ed essendo  $\mathbf{E}\tilde{\xi}^p = \mathbf{E}\tilde{\eta}^q = 1$ )

$$\mathbf{E}(\tilde{\xi}\tilde{\eta}) \leq \frac{1}{p}\mathbf{E}\tilde{\xi}^p + \frac{1}{q}\mathbf{E}\tilde{\eta}^q = 1$$

da cui deriva immediatamente la tesi se si tiene conto della definizione di  $\tilde{\xi}$  e  $\tilde{\eta}$ .  $\square$

**II.6.22 Proposizione (Diseguaglianza di Minkowski):** Se  $\xi$  e  $\eta$  sono due v.a. tali che  $\mathbf{E}|\xi|^p < +\infty$  e  $\mathbf{E}|\eta|^p < +\infty$ , con  $1 \leq p < +\infty$ , allora anche  $\mathbf{E}|\xi + \eta|^p < +\infty$  e la diseuguaglianza

$$(\mathbf{E}|\xi + \eta|^p)^{1/p} \leq (\mathbf{E}|\xi|^p)^{1/p} + (\mathbf{E}|\eta|^p)^{1/p}$$

è sempre verificata.

**Dimostrazione:** Questa diseuguaglianza di Minkowski<sup>7</sup> giocherà un ruolo importante nella discussione degli spazi di v.a. con momento di ordine  $p$  finito, come sarà visto in un capitolo successivo. Nel caso  $p = 1$  la diseuguaglianza è banale dato che, essendo  $|\xi + \eta| \leq |\xi| + |\eta|$ , si ha

$$\mathbf{E}|\xi + \eta| \leq \mathbf{E}|\xi| + \mathbf{E}|\eta|.$$

Nel caso  $p > 1$ , preso  $q > 1$  tale che  $1/p + 1/q = 1$  (cioè tale che  $(p-1)q = p$ ), si ha innanzitutto che

$$(1) \quad |\xi + \eta|^p = |\xi + \eta| \cdot |\xi + \eta|^{p-1} \leq |\xi| \cdot |\xi + \eta|^{p-1} + |\eta| \cdot |\xi + \eta|^{p-1}.$$

---

<sup>7</sup> H. Minkowski (1864 - 1909), matematico tedesco di origine Lituana (era nato a Kaunas, che a quell'epoca faceva ancora parte dell'Impero Russo), è principalmente noto per aver suggerito ad Einstein (che era stato suo allievo al Politecnico di Zurigo) dei miglioramenti formali nella Teoria della Relatività che condussero alla sua formulazione in un unico spazio quadridimensionale, lo spazio-tempo, oggi noto come universo di Minkowski. Oltre che a Zurigo, Minkowski svolse la sua attività come professore a Göttingen dove ebbe modo di collaborare con matematici e fisici del livello di D. Hilbert e M. Born. Nel campo più strettamente matematico i suoi principali contributi riguardano la teoria geometrica dei numeri (*Geometrie der Zahlen*, 1896). La presente diseuguaglianza fu enunciata da Minkowski nel caso di somme finite e fu poi generalizzata da un altro dei suoi celebri allievi, l'ungherese F. Riesz (1880 - 1956).

Possiamo ora applicare la disuguaglianza di Hölder separatamente ai due prodotti  $|\xi| \cdot |\xi + \eta|^{p-1}$  e  $|\eta| \cdot |\xi + \eta|^{p-1}$  (con i dati valori di  $p$  e  $q$ ) ottenendo

$$\begin{aligned}
 \mathbf{E} (|\xi| \cdot |\xi + \eta|^{p-1}) &\leq (\mathbf{E} |\xi|^p)^{1/p} (\mathbf{E} |\xi + \eta|^{(p-1)q})^{1/q} \\
 (2) \qquad \qquad \qquad &= (\mathbf{E} |\xi|^p)^{1/p} (\mathbf{E} |\xi + \eta|^p)^{1/q} \\
 \mathbf{E} (|\eta| \cdot |\xi + \eta|^{p-1}) &\leq (\mathbf{E} |\eta|^p)^{1/p} (\mathbf{E} |\xi + \eta|^p)^{1/q}.
 \end{aligned}$$

Che l'applicazione della disuguaglianza di Hölder sia consentita dipende dal fatto che le ipotesi della Proposizione II.6.21 sono soddisfatte. Infatti  $\mathbf{E} |\xi|^p < +\infty$  e  $\mathbf{E} |\eta|^p < +\infty$  per ipotesi, mentre si prova che  $\mathbf{E} |\xi + \eta|^{p-1} < +\infty$ . Per dimostrarlo si utilizza la disuguaglianza

$$(a + b)^p \leq 2^{p-1}(a^p + b^p); \quad a, b > 0; \quad p \geq 1,$$

che si ottiene osservando che la funzione (con  $a, x > 0$ )

$$f(x) = (a + x)^p - 2^{p-1}(a^p + x^p)$$

assume il suo massimo  $f(a) = 0$  in  $x = a$ , sicché risulta sempre  $f(b) \leq 0$  (cioè la disuguaglianza richiesta) per ogni  $b > 0$ . Si ha allora che

$$|\xi + \eta|^p \leq (|\xi| + |\eta|)^p \leq 2^{p-1}(|\xi|^p + |\eta|^p),$$

da cui si ricava in base alle ipotesi del teorema che  $\mathbf{E} |\xi + \eta|^p < +\infty$  e dalla disuguaglianza di Lyapunov che  $\mathbf{E} |\xi + \eta|^{p-1} < +\infty$ . Pertanto tutte le v.a. di (2) sono integrabili e la disuguaglianza di Hölder è applicabile. Da (1) e (2) si ricava allora che

$$\begin{aligned}
 \mathbf{E} |\xi + \eta|^p &\leq \mathbf{E} (|\xi| \cdot |\xi + \eta|^{p-1}) + \mathbf{E} (|\eta| \cdot |\xi + \eta|^{p-1}) \\
 &\leq (\mathbf{E} |\xi + \eta|^p)^{1/q} \left[ (\mathbf{E} |\xi|^p)^{1/p} + (\mathbf{E} |\eta|^p)^{1/p} \right].
 \end{aligned}$$

Ora, se  $\mathbf{E} |\xi + \eta|^p = 0$  la disuguaglianza di Minkowski è banalmente vera; se invece  $\mathbf{E} |\xi + \eta|^p > 0$  la relazione precedente si può scrivere come

$$(\mathbf{E} |\xi + \eta|^p)^{1-1/q} \leq (\mathbf{E} |\xi|^p)^{1/p} + (\mathbf{E} |\eta|^p)^{1/p},$$

che coincide con la disuguaglianza di Minkowski dato che, per definizione,  $1 - 1/q = 1/p$ . □

**II.6.23 Osservazione:** In un successivo capitolo discuteremo la definizione rigorosa di VdA condizionati e probabilità condizionate, e in quell'occasione torneranno estremamente utili le osservazioni seguenti: se  $\xi$  è una v.a. con  $\mathbf{E} \xi$  definito, in base a (D) di Proposizione II.6.8 risulta ben definita anche l'applicazione  $\mathbf{Q} : \mathcal{F} \rightarrow \overline{\mathbf{R}}$  assegnata come

$$\mathbf{Q}(A) = \int_A \xi \, d\mathbf{P} = \int_{\Omega} \xi I_A \, d\mathbf{P} = \mathbf{E} (\xi I_A), \quad A \in \mathcal{F}.$$

Ovviamente  $\mathbf{Q}$  non è una probabilità (i suoi valori non cadono in  $[0, 1]$ ) e neanche una misura nel senso delle Definizioni II.1.8 e II.1.10 (i suoi valori non sono necessariamente positivi), ma è ancora  $\sigma$ -additiva: infatti, se  $\xi \geq 0$ , e se  $(A_k)_{k \in \mathbf{N}}$  sono eventi disgiunti di  $\mathcal{F}$ , posto  $A = \bigcup_k A_k$ , dal Corollario II.6.11 si ha che

$$\mathbf{Q}(A) = \mathbf{E}(\xi I_A) = \mathbf{E}\left(\sum_k \xi I_{A_k}\right) = \sum_k \mathbf{E}(\xi I_{A_k}) = \sum_k \mathbf{Q}(A_k);$$

inoltre, siccome in questo caso  $\mathbf{Q}(A) \geq 0$ , la  $\mathbf{Q}$  risulta essere una misura (positiva e  $\sigma$ -additiva) su  $\mathcal{F}$ . Se invece  $\xi$  è arbitraria, posto  $\xi = \xi^+ - \xi^-$  si definiscono su  $\mathcal{F}$  le due misure (positive)  $\mathbf{Q}^\pm(A) = \mathbf{E}(\xi^\pm I_A)$  (almeno una delle quali avrà valore finito dato che stiamo supponendo che  $\mathbf{E}\xi$  sia definito, vedi Definizione II.6.5) sicché avremo

$$\mathbf{Q}(A) = \mathbf{Q}^+(A) - \mathbf{Q}^-(A), \quad A \in \mathcal{F},$$

e  $\mathbf{Q}$  risulterà ben definita con valori reali e  $\sigma$ -additiva. Un'applicazione  $\sigma$ -additiva  $\mathbf{Q} : \mathcal{F} \rightarrow \overline{\mathbf{R}}$  prende il nome di **misura segnata** (cioè *con segno*, in quanto non definita positiva) e risulta sempre rappresentabile come differenza  $\mathbf{Q} = \mathbf{Q}_1 - \mathbf{Q}_2$  di altre due misure positive. Pertanto l'applicazione  $\mathbf{Q}(A) = \mathbf{E}(\xi I_A)$  da noi definita non è altro che un caso particolare di misura segnata. Per tale  $\mathbf{Q}$  si può ora mostrare facilmente che  $\mathbf{Q} \ll \mathbf{P}$ , cioè che  $\mathbf{Q}$  è a.c. rispetto a  $\mathbf{P}$  (vedi Definizione II.3.16). Per provarlo basterà ricordare che se  $\mathbf{P}(A) = 0$  dal punto (F) della Proposizione II.6.9 si ha anche  $\mathbf{Q}(A) = \mathbf{E}(\xi I_A) = 0$ . Pertanto  $\mathbf{Q} \ll \mathbf{P}$ ; ma ciò che è più interessante è che sussiste anche la proposizione inversa, secondo la quale, se  $\mathbf{Q} \ll \mathbf{P}$ , esiste sicuramente una v.a.  $\xi$  tale che

$$\mathbf{Q}(A) = \mathbf{E}(\xi I_A) = \int_A \xi d\mathbf{P}, \quad \forall A \in \mathcal{F}.$$

Tale proposizione, che sarà fondamentale nella definizione rigorosa del concetto di condizionamento, è enunciata nel seguente Teorema che generalizza il Teorema II.3.14. ○

**II.6.24 Teorema (di Radon - Nikodym):** Siano  $(\Omega, \mathcal{F})$  uno spazio misurabile,  $\lambda$  una misura  $\sigma$ -finita e  $\mu$  una misura segnata. Se  $\mu \ll \lambda$ , allora certamente esiste una funzione misurabile  $f : (\Omega, \mathcal{F}) \rightarrow (\overline{\mathbf{R}}, \mathcal{B}(\overline{\mathbf{R}}))$  tale che

$$\mu(A) = \int_A f(\omega) \lambda(d\omega), \quad A \in \mathcal{F};$$

tale  $f(\omega)$  che prende il nome di **densità**, o anche di **derivata di Radon-Nikodym**  $f = d\mu/d\lambda$ , è unica  $\lambda$ -q.o. nel senso che, se  $g(\omega)$  è un'altra densità per la medesima misura  $\mu$ , allora risulta  $\lambda\{\omega : f(\omega) \neq g(\omega)\} = 0$ .

**Dimostrazione:** Omessa (vedi ad esempio M. Métivier: *Notions Fondamentales de la Théorie des Probabilités*; Dunod, Paris, 1972, p. 288). □



**II.6.25 Teorema (cambiamento di variabili di integrazione):** Dato uno spazio di probabilità  $(\Omega, \mathcal{F}, \mathbf{P})$  e un elemento aleatorio  $X : (\Omega, \mathcal{F}) \rightarrow (\mathbf{E}, \mathcal{E})$ , detta  $\mathbf{P}_X$  la probabilità indotta su  $(\mathbf{E}, \mathcal{E})$  da  $X$  secondo la relazione  $\mathbf{P}_X(A) = \mathbf{P}\{\omega : X(\omega) \in A\}$ , con  $A \in \mathcal{E}$ , e data la funzione di Borel  $g : (\mathbf{E}, \mathcal{E}) \rightarrow (\mathbf{R}, \mathcal{B}(\mathbf{R}))$ , risulta per  $Y = g(X)$

$$\int_{X^{-1}(A)} Y d\mathbf{P} = \int_{X^{-1}(A)} g(X(\omega)) \mathbf{P}(d\omega) = \int_A g(x) \mathbf{P}_X(dx); \quad A \in \mathcal{E},$$

nel senso che se uno dei due integrali esiste, anche l'altro esiste e i due valori coincidono.

**Dimostrazione:** Il teorema è facilmente provato per  $g(x) = I_B(x)$  ( $B \in \mathcal{E}$ ): infatti, essendo ovviamente  $I_B(X(\omega)) = I_{X^{-1}(B)}(\omega)$ , si ha

$$\begin{aligned} \int_A I_B(x) \mathbf{P}_X(dx) &= \int_{\mathbf{E}} I_A I_B d\mathbf{P}_X = \int_{\mathbf{E}} I_{A \cap B} d\mathbf{P}_X = \mathbf{P}_X(A \cap B) \\ &= \mathbf{P}[X^{-1}(A \cap B)] = \mathbf{P}[X^{-1}(A) \cap X^{-1}(B)] \\ &= \int_{\Omega} I_{X^{-1}(A) \cap X^{-1}(B)}(\omega) \mathbf{P}(d\omega) = \int_{X^{-1}(A)} I_{X^{-1}(B)}(\omega) \mathbf{P}(d\omega) \\ &= \int_{X^{-1}(A)} I_B(X(\omega)) \mathbf{P}(d\omega). \end{aligned}$$

Ne consegue che il teorema risulta vero anche per funzioni semplici a valori non negativi e, mediante i Teoremi II.5.16 e II.6.10, per generiche  $g(x)$  non negative. Nel caso generale si usa la rappresentazione  $g = g^+ - g^-$  e il teorema risulterà vero separatamente per  $g^+$  e per  $g^-$ ; inoltre, se  $\int_A g(x) \mathbf{P}_X(dx)$  è definito, almeno uno dei due integrali non negativi che lo definiscono è finito: supponiamo ad esempio che sia  $\int_A g^+(x) \mathbf{P}_X(dx) < +\infty$ . In tal caso la prima parte del teorema implica che sarà anche

$$\int_{X^{-1}(A)} g^+(X(\omega)) \mathbf{P}(d\omega) < +\infty,$$

sicché anche

$$\int_{X^{-1}(A)} g(X) d\mathbf{P}$$

esisterà e il suo valore coinciderà con  $\int_A g(x) \mathbf{P}_X(dx)$ . □

**II.6.26 Osservazione:** Il Teorema II.6.25 sul cambio di variabile di integrazione è particolarmente utile in quanto ci consente di fornire alcune semplici regole di calcolo mediante integrazioni che, nella maggior parte dei casi praticamente rilevanti, sono ordinarie integrazioni elementari. Cominceremo con l'esaminare il caso in cui  $(\mathbf{E}, \mathcal{E}) = (\mathbf{R}, \mathcal{B}(\mathbf{R}))$ , sicché  $X$  sarà una v.a.  $\xi$ ,  $g(x)$  una funzione di Borel secondo lo schema

$$(\Omega, \mathcal{F}) \xrightarrow{\xi} (\mathbf{R}, \mathcal{B}(\mathbf{R})) \xrightarrow{g} (\mathbf{R}, \mathcal{B}(\mathbf{R}))$$

e mediante queste potremo definire un'altra v.a.  $\eta = g(\xi)$ . Supporremo inoltre di indicare con  $F_\xi$  la FdD della DdP  $\mathbf{P}_\xi$  della v.a.  $\xi$ . In tal caso il Teorema II.6.25 ci dice che, comunque dato  $A \in \mathcal{B}(\mathbf{R})$ , risulta

$$\int_A g(x) F_\xi(dx) = \int_A g(x) \mathbf{P}_\xi(dx) = \int_{\xi^{-1}(A)} g(\xi(\omega)) \mathbf{P}(d\omega) = \int_{\xi^{-1}(A)} \eta d\mathbf{P}.$$

In particolare, quando  $A = \mathbf{R}$ , e quindi  $\xi^{-1}(A) = \xi^{-1}(\mathbf{R}) = \Omega$ , risulta

$$\mathbf{E} \eta = \mathbf{E} [g(\xi)] = \int_\Omega g(\xi) \mathbf{P}(d\omega) = \int_{\mathbf{R}} g(x) \mathbf{P}_\xi(dx) = \int_{\mathbf{R}} g(x) F_\xi(dx).$$

Nel caso in cui  $g(x) = x$  (in modo che  $\eta = \xi$ ) si ha quindi l'importante relazione

$$(1) \quad \mathbf{E} \xi = \int_{\mathbf{R}} x F_\xi(dx).$$

Val la pena notare inoltre che questa espressione consente di calcolare il VdA di  $\eta = g(\xi)$  in due modi distinti: mediante la FdD  $F_\eta$  di  $\eta$  oppure mediante la FdD  $F_\xi$  di  $\xi$ :

$$(2) \quad \mathbf{E} \eta = \mathbf{E} g(\xi) = \int_{\mathbf{R}} y F_\eta(dy) = \int_{\mathbf{R}} g(x) F_\xi(dx).$$

Un altro caso rilevante, poi, è quello in cui  $g(x) = I_B(x)$  ( $B \in \mathcal{B}(\mathbf{R})$ ), sicché  $\eta = g(\xi) = I_B(\xi) = I_{\xi^{-1}(B)}$  e quindi (vedi anche Osservazione II.6.7)

$$\begin{aligned} \mathbf{P}(\xi \in B) &= \mathbf{P}_\xi(B) = \mathbf{P}(\xi^{-1}(B)) = \mathbf{E} I_{\xi^{-1}(B)} = \int_\Omega I_{\xi^{-1}(B)} d\mathbf{P} = \int_\Omega I_B(\xi) d\mathbf{P} \\ &= \int_{\mathbf{R}} I_B(x) \mathbf{P}_\xi(dx) = \int_{\mathbf{R}} I_B(x) F_\xi(dx) = \int_B F_\xi(dx). \end{aligned}$$

Pertanto avremo in generale

$$(3) \quad \mathbf{P}(\xi \in B) = \int_B F_\xi(dx),$$

e in particolare, quando  $B$  è un intervallo,

$$(3a) \quad \mathbf{P}_\xi(a, b] = \mathbf{P}\{a < \xi \leq b\} = \int_a^b F_\xi(dx),$$

o anche

$$(3b) \quad F_\xi(x) = \int_{-\infty}^x F_\xi(dt).$$

Si può pertanto affermare che, per calcolare le quantità relative alla v.a.  $\xi$  o alle sue funzioni (cioè  $\mathbf{E} \xi$ ,  $\mathbf{E} g(\xi)$ ,  $\mathbf{P}(\xi \in B)$  ...) non è necessario conoscere la probabilità  $\mathbf{P}$  definita su  $(\Omega, \mathcal{F})$ , ma è sufficiente conoscere la DdP  $\mathbf{P}_\xi$  su  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$  o anche la sua FdD  $F_\xi(x)$  e calcolare gli opportuni integrali di Lebesgue-Stieltjes (vedi anche Osservazione II.6.6). Resta da notare che, se anche questi integrali di Lebesgue-Stieltjes risultano piuttosto astratti rispetto alle effettive applicazioni e ai più familiari metodi di calcolo, la situazione assume un aspetto definitivamente chiaro nel caso in cui la DdP  $\mathbf{P}_\xi$  su  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$  risulta a.c. rispetto alla misura di Lebesgue  $\lambda$  su  $\mathbf{R}$  (vedi II.3.13 e seguenti oltre che il Teorema II.6.24): mostreremo ora, infatti, che in tal caso, se  $f_\xi$  è la fdd di  $\mathbf{P}_\xi$ , comunque scelto  $A \in \mathcal{B}(\mathbf{R})$  risulta

$$(4) \quad \int_A g(x) F_\xi(dx) = \int_A g(x) f_\xi(x) dx,$$

dove l'ultima espressione ha la forma familiare di un integrale elementare. Per verificare (4) mostreremo innanzitutto che essa è vera per  $A = \mathbf{R}$ : infatti, nel caso in cui  $g(x) = I_B(x)$ , dal Teorema di Radon-Nikodym II.6.24 otteniamo

$$\begin{aligned} \int_{\mathbf{R}} I_B(x) F_\xi(dx) &= \mathbf{E} [I_B(\xi)] = \mathbf{E} (I_{\xi^{-1}(B)}) = \mathbf{P}(\xi^{-1}(B)) = \mathbf{P}_\xi(B) \\ &= \int_B f_\xi(x) dx = \int_{\mathbf{R}} I_B(x) f_\xi(x) dx; \end{aligned}$$

poi, seguendo la linea della dimostrazione del Teorema II.6.25, si osserva che la relazione è vera per funzioni  $g(x)$  semplici e, mediante il Teorema sulla convergenza monotona, per funzioni  $g(x)$  non negative ed infine arbitrarie. La (4) si ottiene poi osservando che, comunque preso  $A \in \mathcal{B}(\mathbf{R})$ , si avrebbe

$$\begin{aligned} \int_A g(x) F_\xi(dx) &= \int_{\xi^{-1}(A)} g(\xi) d\mathbf{P} = \int_{\Omega} I_{\xi^{-1}(A)} g(\xi) d\mathbf{P} \\ &= \int_{\Omega} g(\xi) I_A(\xi) d\mathbf{P} = \int_{\mathbf{R}} g(x) I_A(x) F_\xi(dx) \\ &= \int_{\mathbf{R}} g(x) I_A(x) f_\xi(x) dx = \int_A g(x) f_\xi(x) dx. \end{aligned}$$

Come conseguenza di (4) si ha che le relazioni (1), (2) e (3) assumono la seguente forma:

$$(5) \quad \mathbf{E} \xi = \int_{\mathbf{R}} x f_\xi(x) dx,$$

$$(6) \quad \mathbf{E} g(\xi) = \int_{\mathbf{R}} g(x) f_\xi(x) dx,$$

$$(7) \quad \mathbf{P}_\xi(B) = \int_B f_\xi(x) dx,$$

$$(7a) \quad \mathbf{P}_\xi(a, b] = \mathbf{P}\{a < \xi \leq b\} = \int_a^b f_\xi(x) dx,$$

$$(7b) \quad F_\xi(x) = \int_{-\infty}^x f_\xi(x) dx.$$

Tra l'altro questo risponde definitivamente alle domande poste nell'Osservazione II.5.34 mostrando che tutte le informazioni relative alla v.a. a.c.  $\xi$  ed alle sue funzioni potranno essere ricavate mediante integrazioni elementari che coinvolgono solamente la conoscenza delle opportune funzioni di densità.  $\circ$

**II.6.27 Osservazione:** Una importante generalizzazione delle formule presentate nell'osservazione precedente si ha nel caso in cui  $X = (\xi_k)_{k=1, \dots, n}$  sia un vett.a. ad  $n$  componenti e  $g(x_1, \dots, x_n)$  una funzione di  $n$  variabili  $g : \mathbf{R}^n \rightarrow \mathbf{R}$  secondo lo schema

$$(\Omega, \mathcal{F}) \xrightarrow{X} (\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n)) \xrightarrow{g} (\mathbf{R}, \mathcal{B}(\mathbf{R})).$$

Naturalmente in questo caso  $\eta = g(X)$  risulta una v.a.  $\eta : (\Omega, \mathcal{F}) \rightarrow (\mathbf{R}, \mathcal{B}(\mathbf{R}))$ ; ad esempio, se  $g(x_1, \dots, x_n) = x_1 + \dots + x_n$ , avremo  $\eta = \xi_1 + \dots + \xi_n$ . In questo caso il Teorema II.6.25 assume la forma

$$\begin{aligned} \int_A g(x_1, \dots, x_n) F_X(dx_1, \dots, dx_n) &= \int_A g(x_1, \dots, x_n) \mathbf{P}_X(dx_1, \dots, dx_n) \\ &= \int_{X^{-1}(A)} g(X(\omega)) \mathbf{P}(d\omega) = \int_{X^{-1}(A)} \eta d\mathbf{P}, \end{aligned}$$

dove  $A \in \mathcal{B}(\mathbf{R}^n)$  e  $F_X(x_1, \dots, x_n)$  è la FdD congiunta delle componenti di  $X$ . Quando  $A = \mathbf{R}^n$  si ottengono le relazioni

$$\mathbf{E} \eta = \mathbf{E} [g(X)] = \int_{\Omega} g(X) d\mathbf{P} = \int_{\mathbf{R}^n} g(x_1, \dots, x_n) F_X(dx_1, \dots, dx_n),$$

e siccome, in base alle formule dell'Osservazione II.6.26,  $\mathbf{E} \eta$  può essere anche calcolato mediante la conoscenza della FdD  $F_\eta(x)$ , si ha che

$$(1) \quad \mathbf{E} \eta = \mathbf{E} g(\xi) = \int_{\mathbf{R}} y F_\eta(dy) = \int_{\mathbf{R}^n} g(x_1, \dots, x_n) F_X(dx_1, \dots, dx_n).$$

Se in particolare  $g(x_1, \dots, x_n) = I_B(x_1, \dots, x_n)$  con  $B \in \mathcal{B}(\mathbf{R}^n)$ , essendo  $I_B(X) = I_{X^{-1}(B)}$  risulta anche

$$\begin{aligned} \mathbf{P}(X \in B) &= \mathbf{P}_X(B) = \mathbf{P}(X^{-1}(B)) = \mathbf{E} (I_{X^{-1}(B)}) = \int_{\Omega} I_B(X) d\mathbf{P} \\ &= \int_{\mathbf{R}^n} I_B(x_1, \dots, x_n) F_X(dx_1, \dots, dx_n) = \int_B F_X(dx_1, \dots, dx_n); \end{aligned}$$

pertanto si ha in generale che

$$(2) \quad \mathbf{P}(X \in B) = \int_B F_X(dx_1, \dots, dx_n),$$

e in particolare, se  $B = (a_1, b_1] \times \dots \times (a_n, b_n]$ ,

$$(2a) \quad \mathbf{P}_X((a_1, b_1] \times \dots \times (a_n, b_n]) = \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} F_X(dx_1, \dots, dx_n).$$

Inoltre, dato che

$$\begin{aligned} F_X(x_1, \dots, x_n) &= \mathbf{P}_X((-\infty, x_1] \times \dots \times (-\infty, x_n]) \\ &= \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} F_X(dt_1, \dots, dt_n), \end{aligned}$$

si ha anche che

$$(2b) \quad F_X(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} F_X(dt_1, \dots, dt_n).$$

Le DdP marginali (con  $B \in \mathcal{B}(\mathbf{R}_k)$ ) si ricavano poi come al solito da

$$\begin{aligned} \mathbf{P}_{\xi_k}(B) &= \mathbf{P}_X(\mathbf{R}_1 \times \dots \times B \times \dots \times \mathbf{R}_n) \\ &= \int_{-\infty}^{+\infty} \dots \int_B \dots \int_{-\infty}^{+\infty} F_X(dt_1, \dots, dt_k, \dots, dt_n), \end{aligned}$$

sicch  se si pone

$$F_{\xi_k}(x_k) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} F_X(dt_1, \dots, x_k, \dots, dt_n),$$

con  $n - 1$  integrazioni su tutte le variabili tranne che su  $x_k$ , si ha

$$\mathbf{P}_{\xi}(B) = \int_B F_{\xi_k}(dx_k)$$

e quindi

$$(2c) \quad F_{\xi_k}(x_k) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{x_k} \dots \int_{-\infty}^{+\infty} F_X(dt_1, \dots, dt_k, \dots, dt_n).$$

Ovviamente se  $\mathbf{P}_X$  risulta anche a.c. rispetto alla misura di Lebesgue su  $\mathbf{R}^n$  con fdd congiunta  $f_X(x_1, \dots, x_n)$ , come nel caso unidimensionale dell'Osservazione II.6.26, si prova innanzitutto che

$$(3) \quad \begin{aligned} \int_A g(x_1, \dots, x_n) F_X(dx_1, \dots, dx_n) \\ = \int_A g(x_1, \dots, x_n) f_X(x_1, \dots, x_n) dx_1, \dots, dx_n; \end{aligned}$$

come conseguenza (1) si riscrive come

$$(4) \quad \begin{aligned} \mathbf{E} \eta &= \mathbf{E} g(\xi) = \int_{\mathbf{R}} y f_Y(y) dy \\ &= \int_{\mathbf{R}^n} g(x_1, \dots, x_n) f_X(x_1, \dots, x_n) dx_1 \dots dx_n, \end{aligned}$$

mentre le relazioni (2) divengono

$$(5) \quad \mathbf{P}(X \in B) = \int_B f_X(x_1, \dots, x_n) dx_1 \dots dx_n,$$

$$(5a) \quad \mathbf{P}_X((a_1, b_1] \times \dots \times (a_n, b_n]) = \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} f_X(x_1, \dots, x_n) dx_1 \dots dx_n,$$

$$(5b) \quad F_X(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_X(t_1, \dots, t_n) dt_1 \dots dt_n.$$

Quanto alle fdd marginali basterà osservare che, posto

$$f_{\xi_k}(x_k) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f_X(t_1, \dots, x_k, \dots, t_n) dt_1 \dots dt_n,$$

dove le  $n - 1$  integrazioni sono eseguite su tutte le variabili fatta eccezione per la  $k$ -ma, si ha

$$(5c) \quad F_{\xi_k}(x_k) = \int_{-\infty}^{x_k} f_{\xi_k}(t_k) dt_k \\ = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{x_k} \dots \int_{-\infty}^{+\infty} f_X(t_1, \dots, t_k, \dots, t_n) dt_1 \dots dt_k \dots dt_n.$$

Pertanto anche nel caso di distribuzioni congiunte di vett.a. a.c. tutti i calcoli di interesse pratico potranno essere eseguiti a partire dalla conoscenza della opportuna funzione di densità. ○

**II.6.28 Osservazione:** Nel capitolo II.5 abbiamo discusso le relazioni fra il concetto di indipendenza di v.a. e quello di misure di probabilità prodotto: aggiungeremo ora alcune osservazioni importanti alla luce della teoria dell'integrazione. Innanzitutto è ovvio che la Definizione II.5.36 di probabilità prodotto si estende facilmente al caso di misure  $\sigma$ -finite generiche: dati  $n$  spazi misurabili  $(\Omega_k, \mathcal{F}_k)$  con misure  $\mu_k$   $\sigma$ -finite, chiameremo **misura prodotto** la misura  $\mu = \mu_1 \otimes \dots \otimes \mu_n$  su  $(\Omega, \mathcal{F})$  con

$$\Omega = \prod_{k=1}^n \Omega_k, \quad \mathcal{F} = \bigotimes_{k=1}^n \mathcal{F}_k$$

tale che per  $A = A_1 \times \dots \times A_n \in \mathcal{F}$  (con  $A_k \in \mathcal{F}_k$ ) sia

$$\mu(A) = \mu(A_1 \times \dots \times A_n) = \mu_1(A_1) \cdot \dots \cdot \mu_n(A_n).$$

Si può provare che tale  $\mu$  esiste ed è unica e che è  $\sigma$ -finita se lo sono le  $\mu_k$ . È interessante ora guardare un po' più da vicino come si costruiscono tali misure prodotto sia per mostrare, almeno per grandi linee, che esse esistono (e sono uniche), sia per far vedere che i valori della misura prodotto  $\mu(A) = \int_{\Omega} I_A d\mu = \int_A d\mu$  possono

sempre essere calcolati mediante integrazioni sulle misure  $\mu_k$ . Prenderemo in considerazione il caso  $n = 2$ , con  $\Omega = \Omega_1 \times \Omega_2$ ,  $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$  e con  $\mu_1$  e  $\mu_2$   $\sigma$ -finite. Dato  $B \in \mathcal{F}$  (in genere non un rettangolo) è utile definire le cosiddette  $\omega_1$ -sezione e  $\omega_2$ -sezione di  $B$ : fissato  $\omega_1 \in \Omega_1$ , la  $\omega_1$ -sezione di  $B$  è l'insieme di  $\Omega_2$

$$B_{\omega_1} = \{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in B\};$$

analogamente la  $\omega_2$ -sezione è l'insieme di  $\Omega_1$

$$B_{\omega_2} = \{\omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in B\}.$$

È evidente che  $B_{\omega_1}$  e  $B_{\omega_2}$  sono, rispettivamente, parti di  $\Omega_2$  e  $\Omega_1$  che, in generale, variano con  $\omega_1$  e  $\omega_2$ . In particolare se  $B = A_1 \times A_2$  è un rettangolo avremo che

$$B_{\omega_1} = \begin{cases} A_2, & \text{se } \omega_1 \in A_1; \\ \emptyset, & \text{se } \omega_1 \in \overline{A_1}; \end{cases} \quad B_{\omega_2} = \begin{cases} A_1, & \text{se } \omega_2 \in A_2; \\ \emptyset, & \text{se } \omega_2 \in \overline{A_2}; \end{cases}$$

e conseguentemente avremo

$$\mu_2(B_{\omega_1}) = \mu_2(A_2)I_{A_1}(\omega_1), \quad \mu_1(B_{\omega_2}) = \mu_1(A_1)I_{A_2}(\omega_2).$$

Nel caso più generale (quando  $B$  non è un rettangolo) invece si ha che  $\mu_2(B_{\omega_1})$  e  $\mu_1(B_{\omega_2})$  sono funzioni più complicate di  $\omega_1$  ed  $\omega_2$  rispettivamente, ma è comunque possibile provare che tali funzioni risultano  $\mathcal{F}_1$ -misurabili ed  $\mathcal{F}_2$ -misurabili. Osserveremo ora che le misure delle  $\omega$ -sezioni possono essere usate per esprimere i valori della misura prodotto  $\mu = \mu_1 \otimes \mu_2$ : infatti, nel caso dei rettangoli  $B = A_1 \times A_2$ , dalle osservazioni precedenti si ricava che

$$\mu(B) = \mu_1(A_1)\mu_2(A_2) = \begin{cases} \int_{\Omega_1} \mu_2(A_2)I_{A_1}(\omega_1)\mu_1(d\omega_1) = \int_{\Omega_1} \mu_2(B_{\omega_1})\mu_1(d\omega_1); \\ \int_{\Omega_2} \mu_1(A_1)I_{A_2}(\omega_2)\mu_2(d\omega_2) = \int_{\Omega_2} \mu_1(B_{\omega_2})\mu_2(d\omega_2). \end{cases}$$

È possibile poi dimostrare (anche se noi lo ometteremo) che in realtà questa espressione è valida per  $B \in \mathcal{F}$  generico. Queste osservazioni consentono anche di indicare il modo in cui la misura prodotto viene costruita: date  $\mu_1$  e  $\mu_2$   $\sigma$ -finite su  $(\Omega_1, \mathcal{F}_1)$  e  $(\Omega_2, \mathcal{F}_2)$ , comunque scelto  $B \in \mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$  si prova che  $\mu_2(B_{\omega_1})$  è  $\mathcal{F}_1$ -misurabile,  $\mu_1(B_{\omega_2})$  è  $\mathcal{F}_2$ -misurabile, e inoltre che

$$\int_{\Omega_1} \mu_2(B_{\omega_1})\mu_1(d\omega_1) = \int_{\Omega_2} \mu_1(B_{\omega_2})\mu_2(d\omega_2),$$

sicché è sempre possibile definire, in maniera coerente, una misura  $\mu : \mathcal{F} \rightarrow \overline{\mathbf{R}^+}$  ponendo per ogni  $B \in \mathcal{F}$

$$\mu(B) = \int_{\Omega_1} \mu_2(B_{\omega_1})\mu_1(d\omega_1) = \int_{\Omega_2} \mu_1(B_{\omega_2})\mu_2(d\omega_2).$$

È facile mostrare, sulla base delle osservazioni precedenti, che se  $B = A_1 \times A_2$  risulta  $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$  e si prova anche che la  $\mu$  da noi definita è l'unica con questa proprietà: essa prende quindi il nome di misura prodotto e si indica con  $\mu = \mu_1 \otimes \mu_2$ . Naturalmente tutto questo è vero anche quando abbiamo a che fare, in particolare, con delle probabilità. Questa discussione ci consente ora di enunciare un importante teorema, dovuto a G. Fubini<sup>8</sup>, che gioca un ruolo analogo a quello dei teoremi sulla riduzione di integrali doppi (o in generale multipli) di Riemann ad una successione di integrazioni semplici. Enunceremo successivamente un teorema che generalizza i risultati precedenti al caso in cui la misura  $\mu$  su  $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$  non è la misura prodotto  $\mu_1 \otimes \mu_2$  esaminata nella presente osservazione, ma è una generica misura. ○

**II.6.29 Teorema (di Fubini):** Facendo uso delle notazioni esposte in II.6.28, se  $\xi : (\Omega, \mathcal{F}) \rightarrow (\mathbf{R}, \mathcal{B}(\mathbf{R}))$  è una funzione misurabile risulta che

$$\int_{\Omega_2} \xi(\omega_1, \omega_2) \mu_2(d\omega_2)$$

è  $\mathcal{F}_1$ -misurabile e  $\mu_1$ -q.o. finita,

$$\int_{\Omega_1} \xi(\omega_1, \omega_2) \mu(d\omega_1)$$

è  $\mathcal{F}_2$ -misurabile e  $\mu_2$ -q.o. finita, e

$$\begin{aligned} \int_{\Omega} \xi d\mu &= \int_{\Omega_1} \left( \int_{\Omega_2} \xi(\omega_1, \omega_2) \mu_2(d\omega_2) \right) \mu_1(d\omega_1) \\ &= \int_{\Omega_2} \left( \int_{\Omega_1} \xi(\omega_1, \omega_2) \mu_1(d\omega_1) \right) \mu_2(d\omega_2) \end{aligned}$$

se  $\int_{\Omega} \xi d\mu$  esiste finito e se  $\mu$  è la misura prodotto  $\mu_1 \otimes \mu_2$ .

**Dimostrazione:** Omessa<sup>9</sup>. □

<sup>8</sup> Guido Fubini Ghiron (1879 - 1943), di famiglia piemontese residente a Venezia, fu professore di analisi in varie università italiane fra il 1903 ed il 1939 quando, a causa delle persecuzioni razziali, fu obbligato ad andare esule prima a Parigi e poi a Princeton negli USA dove morì nel 1943. Scienziato di attività vasta e multiforme (ha lasciato alcuni trattati e quasi 200 articoli) si è interessato soprattutto di geometria differenziale proiettiva, in particolare sotto l'influsso del suo maestro Luigi Bianchi. Nel campo dell'analisi fu uno dei primi ad apprezzare i nuovi concetti introdotti da Lebesgue e il teorema sugli integrali doppi qui riportato fu da lui enunciato nel 1907.

<sup>9</sup> Vedi ad esempio **A.N. Shiryaev**: *Probability*; Springer, New York, 1984, p. 196.



**II.6.30 Definizione:** Dati due spazi probabilizzabili  $(\Omega_1, \mathcal{F}_1)$  e  $(\Omega_2, \mathcal{F}_2)$  diremo che le misure  $\{\mu_2(\omega_1; \cdot)\}_{\omega_1 \in \Omega_1}$  definite su  $\mathcal{F}_2$  sono **uniformemente  $\sigma$ -finite** se è possibile determinare una successione  $(A_n)_{n \in \mathbf{N}}$ , con  $A_n \in \mathcal{F}_2$ , tale che  $\bigcup_n A_n = \Omega_2$  e che  $\forall n \in \mathbf{N} \exists M_n \in \mathbf{R}^+ \ni \mu_2(\omega_1; A_n) \leq M_n, \forall \omega_1 \in \Omega_1$ .  $\triangle$

**II.6.31 Teorema:** Dati due spazi probabilizzabili  $(\Omega_1, \mathcal{F}_1)$  e  $(\Omega_2, \mathcal{F}_2)$ , sia  $\mu_1$  una misura  $\sigma$ -finita su  $(\Omega_1, \mathcal{F}_1)$  e inoltre, per ogni  $\omega_1 \in \Omega_1$ , sia assegnata una misura  $\mu_2(\omega_1; \cdot)$  su  $(\Omega_2, \mathcal{F}_2)$  in modo tale che:

a. comunque fissato  $A \in \mathcal{F}_2$ , l'applicazione  $\mu_2(\cdot; A) : \Omega_1 \rightarrow \overline{\mathbf{R}}^+$  sia  $\mathcal{F}_1$ -misurabile;

b. le misure  $\{\mu_2(\omega_1; \cdot)\}_{\omega_1 \in \Omega_1}$  definite su  $\mathcal{F}_2$  siano uniformemente  $\sigma$ -finite; allora esiste sempre una ed una sola misura  $\sigma$ -finita  $\mu : \mathcal{F} \rightarrow \overline{\mathbf{R}}^+$ , con  $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$ , tale che

$$\mu(B) = \int_{\Omega_1} \mu_2(\omega_1; B_{\omega_1}) \mu_1(d\omega_1), \quad \forall B \in \mathcal{F};$$

in particolare quando  $B = A_1 \times A_2$

$$\mu(A_1 \times A_2) = \int_{\Omega_1} \mu_2(\omega_1; A_2) I_{A_1}(\omega_1) \mu_1(d\omega_1) = \int_{A_1} \mu_2(\omega_1; A_2) \mu_1(d\omega_1).$$

Inoltre si ha che

$$\int_{\Omega} \xi d\mu = \int_{\Omega_1 \times \Omega_2} \xi(\omega_1, \omega_2) \mu(d\omega_1, d\omega_2) = \int_{\Omega_1} \mu_1(d\omega_1) \int_{\Omega_2} \xi(\omega_1, \omega_2) \mu_2(\omega_1; d\omega_2)$$

se  $\xi : \Omega \rightarrow (\overline{\mathbf{R}}, \mathcal{B}(\overline{\mathbf{R}}))$  è una funzione misurabile con  $\Omega = \Omega_1 \times \Omega_2$  e  $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$ .

**Dimostrazione:** Omessa<sup>10</sup>.  $\square$

**II.6.32 Osservazione:** Notiamo che la generalizzazione che si ricava sulla base del Teorema II.6.31 consiste nell'idea che una delle due misure ( $\mu_2$  su  $\Omega_2$ ) possa dipendere dalla scelta di  $\omega_1 \in \Omega_1$ . Si vede infatti che, nel caso in cui  $\mu_2$  non dipende da  $\omega_1$ , il Teorema II.6.31 si riduce a quanto detto circa le misure prodotto nell'Osservazione II.6.28 e nel Teorema di Fubini. Inoltre, quando le misure  $\mu, \mu_1$  e  $\mu_2$  in questione sono delle probabilità  $\mathbf{P}, \mathbf{P}_1$  e  $\mathbf{P}_2$ , la differenza fra i due casi può essere discussa in termini di condizionamento ed indipendenza: se  $X = (\xi_1, \xi_2)$  è un vett.a. con due componenti,  $\mathbf{P}_X$  risulta essere una probabilità su  $(\mathbf{R}^2, \mathcal{B}(\mathbf{R}^2))$  prodotto diretto di  $(\mathbf{R}_1, \mathcal{B}(\mathbf{R}_1))$  e  $(\mathbf{R}_2, \mathcal{B}(\mathbf{R}_2))$ , mentre le sue marginali  $\mathbf{P}_{\xi_1}$  e  $\mathbf{P}_{\xi_2}$

<sup>10</sup> Vedi ad esempio **R.B. Ash:** *Real analysis and probability*; Academic Press, New York, 1972, p. 97.

sono probabilità rispettivamente su  $(\mathbf{R}_1, \mathcal{B}(\mathbf{R}_1))$  e  $(\mathbf{R}_2, \mathcal{B}(\mathbf{R}_2))$ . Se le componenti di  $X$  sono indipendenti  $\mathbf{P}_X$  è proprio la misura prodotto  $\mathbf{P}_{\xi_1} \otimes \mathbf{P}_{\xi_2}$  (vedi Osservazione II.5.38) e si può applicare il Teorema di Fubini. Se invece le componenti di  $X$  non sono indipendenti tutto ciò non accade, ma si mostrerà, nel successivo capitolo sul condizionamento, che ci troviamo comunque nella situazione del Teorema II.6.31 nel senso che potremo sempre fissare una delle due marginali come misura  $\mu_1$  (ad esempio  $\mathbf{P}_1 = \mathbf{P}_{\xi_1}$ ) e determinare successivamente una seconda probabilità  $\mathbf{P}_2(x_1; \cdot) : \mathcal{B}(\mathbf{R}_1) \rightarrow [0, 1]$  (che però non è la marginale  $\mathbf{P}_{\xi_2}$ , ma un'opportuna forma della probabilità condizionata) che soddisfi le condizioni del Teorema II.6.31. In tal caso risulterà (con la solita notazione in  $x_1$ -sezioni)

$$\mathbf{P}_X(B) = \int_{\mathbf{R}_1} \mathbf{P}_2(x_1; B_{x_1}) \mathbf{P}_1(dx_1), \quad \forall B \in \mathcal{B}(\mathbf{R}^2),$$

e in particolare, se  $B = A_1 \times A_2$ ,

$$\mathbf{P}_X(A_1 \times A_2) = \int_{\mathbf{R}_1} \mathbf{P}_2(x_1; A_2) I_{A_1}(x_1) \mathbf{P}_1(dx_1) = \int_{A_1} \mathbf{P}_2(x_1; A_2) \mathbf{P}_1(dx_1).$$

Come è evidente questa espressione non si fattorizza, in generale, nel prodotto  $\mathbf{P}_{\xi_1}(A_1)\mathbf{P}_{\xi_2}(A_2)$ ; solo nel caso in cui le  $\xi_1$  e  $\xi_2$  sono indipendenti risulta che  $\mathbf{P}_2(x_1; \cdot)$  non dipende da  $x_1$  e coincide con  $\mathbf{P}_{\xi_2}$  sicché in questo caso  $\mathbf{P}_X(A_1 \times A_2) = \mathbf{P}_{\xi_1}(A_1)\mathbf{P}_{\xi_2}(A_2)$ , come si ricava dall'equazione precedente. Nel caso generale, comunque, le relazioni contenute nelle proposizioni esposte consentono di calcolare  $\mathbf{P}_X(B)$  mediante la conoscenza di opportune probabilità (condizionate) sugli spazi fattori  $(\mathbf{R}_1, \mathcal{B}(\mathbf{R}_1))$  e  $(\mathbf{R}_2, \mathcal{B}(\mathbf{R}_2))$  e inoltre il VdA di v.a.  $\xi(x_1, x_2)$  (ossia integrali nella misura  $\mathbf{P}_X$  di funzioni misurabili  $\xi$  definite su  $\mathbf{R}^2$ ) mediante integrazioni successive eseguite sugli spazi ad una sola dimensione  $(\mathbf{R}_1, \mathcal{B}(\mathbf{R}_1))$  e  $(\mathbf{R}_2, \mathcal{B}(\mathbf{R}_2))$ , e questo anche nel caso in cui le componenti di  $X$  non siano indipendenti. In particolare è utile osservare a proposito del Teorema II.6.31 che, quando  $\xi(\omega_1, \omega_2) = I_B(\omega_1, \omega_2)$  con  $B \in \mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$ , la prima parte del teorema si ottiene facilmente dalla seconda in quanto, essendo  $I_B(\omega_1, \omega_2) = I_{B_{\omega_1}}(\omega_2)$ , risulta

$$\begin{aligned} \mu(B) &= \int_{\Omega} I_B d\mu = \int_{\Omega_1} \mu_1(d\omega_1) \int_{\Omega_2} I_B(\omega_1, \omega_2) \mu_2(\omega_1; d\omega_2) \\ &= \int_{\Omega_1} \mu_1(d\omega_1) \int_{\Omega_2} I_{B_{\omega_1}}(\omega_2) \mu_2(\omega_1; d\omega_2) = \int_{\Omega_1} \mu_1(d\omega_1) \mu_2(\omega_1; B_{\omega_1}). \end{aligned}$$

I risultati contenuti nel Teorema II.6.31 e nella presente osservazione possono, infine, essere anche generalizzati al caso di spazi prodotto di  $n$  spazi di probabilità nel modo che ora esporremo evitando, per semplicità, di usare la notazione in  $\omega_1$ -sezione,  $\omega_2$ -sezione,... ed adottando invece quella che fa uso degli indicatori. Dati dunque  $n$  spazi probabilizzabili  $(\Omega_1, \mathcal{F}_1), \dots, (\Omega_n, \mathcal{F}_n)$ , sia  $\mu_1$  una misura  $\sigma$ -finita su  $(\Omega_1, \mathcal{F}_1)$  e inoltre, comunque assegnato  $(\omega_1, \dots, \omega_k) \in \Omega_1 \times \dots \times \Omega_k$ , con  $k = 1, \dots, n-1$ , sia data la famiglia di misure  $\mu_{k+1}(\omega_1, \dots, \omega_k; \cdot)$  definite su  $\mathcal{F}_{k+1}$  con le seguenti proprietà:

- a) comunque scelto  $A \in \mathcal{F}_{k+1}$ , l'applicazione  $\mu_{k+1}(\dots; A) : \Omega_1 \times \dots \times \Omega_k \rightarrow \overline{\mathbf{R}}^+$  è misurabile;
- b) le misure  $\mu_{k+1}(\omega_1, \dots, \omega_k; \cdot)$  sono uniformemente  $\sigma$ -finite (vedi Definizione II.6.30) per  $k = 1, \dots, n-1$ .

Si ha allora che esiste sempre una misura  $\sigma$ -finita  $\mu : \mathcal{F} \rightarrow \mathbf{R}^+$  con  $\mathcal{F} = \mathcal{F}_1 \otimes \dots \otimes \mathcal{F}_n$  tale che per ogni rettangolo  $B = A_1 \times \dots \times A_n \in \mathcal{F}$  risulti

$$\begin{aligned} \mu(B) &= \mu(A_1 \times \dots \times A_n) \\ &= \int_{A_1} \mu_1(d\omega_1) \int_{A_2} \mu_2(\omega_1; d\omega_2) \dots \int_{A_n} \mu_n(\omega_1, \dots, \omega_{n-1}; d\omega_n); \end{aligned}$$

inoltre, comunque assegnata una v.a.  $\xi : (\Omega, \mathcal{F}) \rightarrow (\overline{\mathbf{R}}, \mathcal{B}(\overline{\mathbf{R}}))$ , con  $\Omega = \Omega_1 \times \dots \times \Omega_n$  e  $\mathcal{F} = \mathcal{F}_1 \otimes \dots \otimes \mathcal{F}_n$ , risulta (se l'integrale esiste) che

$$\begin{aligned} \int_{\Omega} \xi d\mu &= \int_{\Omega_1 \times \dots \times \Omega_n} \xi(\omega_1, \dots, \omega_n) \mu(d\omega_1, \dots, d\omega_n) \\ &= \int_{\Omega_1} \mu_1(d\omega_1) \int_{\Omega_2} \mu_2(\omega_1; d\omega_2) \dots \int_{\Omega_n} \xi(\omega_1, \dots, \omega_n) \mu_n(\omega_1, \dots, \omega_{n-1}; d\omega_n) \end{aligned}$$

relazione che in particolare consente di calcolare  $\mu(B)$  mediante un'integrazione anche quando  $B$  non è un rettangolo: posto infatti  $\xi = I_B$  si ha

$$\mu(B) = \int_{\Omega} I_B d\mu$$

che si calcola con la precedente formula di integrazioni successive. Infine, nel caso in cui le misure  $\mu_{k+1}(\cdot) = \mu_{k+1}(\omega_1, \dots, \omega_k; \cdot)$  non dipendano da  $\omega_1, \dots, \omega_k$ , la misura  $\mu$  è proprio la misura prodotto in quanto si verifica facilmente che  $\mu(A_1 \times \dots \times A_n) = \mu_1(A_1) \dots \mu_n(A_n)$ . in tal caso il Teorema di Fubini assume la forma

$$\int_{\Omega} \xi d\mu = \int_{\Omega_1} \mu_1(d\omega_1) \int_{\Omega_2} \mu_2(d\omega_2) \dots \int_{\Omega_n} \xi(\omega_1, \dots, \omega_n) \mu_n(d\omega_n),$$

e l'ordine delle integrazioni è irrilevante. ○

**II.6.33 Osservazione:** Sulla base delle relazioni sviluppate in questo capitolo daremo ora una (diversa e più semplice) prova della seconda parte del Teorema II.5.41 per il caso particolare in cui  $\mathbf{P}_X$  è a.c.: supponendo che  $f_X(x_1, \dots, x_n) = f_{\xi_1}(x_1) \dots f_{\xi_n}(x_n)$  proveremo che le componenti di  $X$  sono indipendenti. Preso infatti  $B = B_1 \times \dots \times B_n$ , da (5) dell'Osservazione II.6.27 risulta

$$\begin{aligned} \mathbf{P}_X(B) &= \int_{B_1} \dots \int_{B_n} f_X(x_1, \dots, x_n) dx_1 \dots dx_n = \prod_{k=1}^n \int_{B_k} f_{\xi_k}(x_k) dx_k \\ &= \prod_{k=1}^n \mathbf{P}_{\xi_k}(B_k), \end{aligned}$$

sicché  $\mathbf{P}_X$  è la misura prodotto delle marginali e le componenti di  $X$  risultano indipendenti in base all'Osservazione II.5.38. ○



## II.7 Condizionamento

**II.7.1 Osservazione:** Dato uno spazio di probabilità  $(\Omega, \mathcal{F}, \mathbf{P})$  ed assegnati due eventi  $B \in \mathcal{F}$  e  $D \in \mathcal{F}$  con  $\mathbf{P}(D) \neq 0$ , la definizione elementare di probabilità condizionata (vedi Definizione I.4.2)

$$\mathbf{P}(B|D) = \frac{\mathbf{P}(B \cap D)}{\mathbf{P}(D)}$$

può essere immediatamente estesa al caso generale. Allo stesso modo si può estendere il concetto (vedi Definizione I.9.1) di probabilità condizionata rispetto ad una decomposizione  $\mathcal{D} = \{D_k\}_{k=1, \dots, n}$  finita con  $\mathbf{P}(D_k) \neq 0$  o equivalentemente, rispetto all'algebra  $\mathcal{A} = \alpha(\mathcal{D}) \subseteq \mathcal{F}$  generata da tale decomposizione (vedi Osservazione I.9.16) mediante

$$\mathbf{P}(B|\mathcal{A}) = \mathbf{P}(B|\mathcal{D}) = \sum_k \mathbf{P}(B|D_k) I_{D_k}.$$

Ricordiamo a questo proposito che tali condizionamenti rispetto ad algebre definiscono delle nuove v.a. che, nel presente caso, risultano essere v.a. semplici essendo  $\mathcal{D}$  una decomposizione finita. Per discutere il caso generale (con  $\mathbf{P}(D)$  anche nullo) converrà però osservare che il concetto di probabilità condizionata può essere sempre formulato tramite quello di VdA condizionato. Per far questo mostriamo innanzitutto che il concetto di VdA condizionato può essere stabilito indipendentemente da quello di probabilità condizionata. Siano  $\xi$  una v.a. e  $D \in \mathcal{F}$  un evento con  $\mathbf{P}(D) \neq 0$ : se  $\xi$  è una v.a. semplice (che assume, quindi, solo un numero finito di valori) della forma

$$\xi = \sum_j x_j I_{B_j}$$

si può estendere immediatamente la Definizione I.9.5

$$\mathbf{E}(\xi|D) = \sum_j x_j \mathbf{P}(B_j|D).$$

Se invece  $\xi$  non è semplice la definizione precedente non può essere direttamente generalizzata, e pertanto conviene porla nella seguente forma:

$$\begin{aligned} \mathbf{E}(\xi|D) &= \sum_j x_j \mathbf{P}(B_j|D) = \sum_j x_j \frac{\mathbf{P}(B_j D)}{\mathbf{P}(D)} = \sum_j x_j \frac{\mathbf{E}(I_{B_j} I_D)}{\mathbf{P}(D)} \\ &= \frac{1}{\mathbf{P}(D)} \mathbf{E}\left(I_D \sum_j x_j I_{B_j}\right) = \frac{\mathbf{E}(\xi I_D)}{\mathbf{P}(D)}. \end{aligned}$$

Quest'ultima espressione, non facendo richiamo alla natura semplice della nostra v.a., è ovviamente estendibile al caso generale e può essere considerata come

definizione generale di VdA condizionato almeno nel caso in cui  $\mathbf{P}(D) \neq 0$ . In maniera analoga si estende anche il concetto di VdA condizionato rispetto ad un'algebra  $\mathcal{A} = \alpha(\mathcal{D}) \subseteq \mathcal{F}$  generata da una decomposizione  $\mathcal{D} = \{D_k\}_{k=1, \dots, n}$  finita con  $\mathbf{P}(D_k) \neq 0$ : si tratta della nuova v.a.

$$\mathbf{E}(\xi|\mathcal{A}) = \mathbf{E}(\xi|\mathcal{D}) = \sum_k \mathbf{E}(\xi|D_k)I_{D_k} = \sum_k \frac{\mathbf{E}(\xi I_{D_k})}{\mathbf{P}(D_k)} I_{D_k}$$

che risulta ovviamente misurabile rispetto all'algebra  $\mathcal{A} \subseteq \mathcal{F}$ . A questo punto è facile vedere che i concetti di probabilità condizionate rispetto ad eventi o algebre possono sempre essere ricavati da quelli dei corrispondenti VdA condizionati mediante le relazioni

$$\mathbf{P}(B|D) = \mathbf{E}(I_B|D), \quad \mathbf{P}(B|\mathcal{A}) = \mathbf{E}(I_B|\mathcal{A}),$$

sicché potremo limitarci a discutere inizialmente solo i concetti di VdA condizionati, riservandoci le opportune considerazioni sulle probabilità condizionate per un secondo momento. A questo punto dobbiamo quindi notare che i concetti di VdA condizionati  $\mathbf{E}(\xi|D)$  e  $\mathbf{E}(\xi|\mathcal{A})$  sono stati definiti per v.a. generiche, ma con due importanti limitazioni che riguardano gli eventi condizionanti: innanzitutto  $\mathcal{A}$  è un'algebra finita e non una  $\sigma$ -algebra come sarebbe opportuno che fosse nel caso generale; e in secondo luogo, e soprattutto, gli eventi condizionanti ( $D$  o gli elementi di  $\mathcal{A}$ ) devono essere eventi di probabilità diversa da zero se si vuole che le definizioni precedenti abbiano senso. L'opportunità di dare delle definizioni che siano applicabili anche nel caso in cui gli eventi condizionanti siano di probabilità zero è illustrata dal seguente esempio.  $\circ$

**II.7.2 Esempio:** Supponiamo che  $\pi$  sia una v.a. con valori in  $([0, 1], \mathcal{B}([0, 1]))$  distribuiti in maniera uniforme (vedi Esempio II.3.16) ed eseguiamo il seguente esperimento: se il valore di  $\pi$  è  $p \in [0, 1]$  si lancia  $n$  volte una moneta tale che la probabilità di ottenere *testa* sia proprio  $p$ . Sia inoltre  $\nu$  la v.a. che rappresenta il numero di volte in cui si ottiene *testa*: qual è la probabilità condizionata  $\mathbf{P}(\nu = k|\pi = p)$ , cioè la probabilità di ottenere  $k$  teste su  $n$  lanci indipendenti, condizionata dal fatto che, in ogni lancio, la probabilità di ottenere *testa* è  $p$ ? Dalle discussioni svolte in precedenza sulle distribuzioni binomiali è intuitivo riconoscere che la risposta dovrebbe essere  $\binom{n}{k} p^k (1-p)^{n-k}$ , ma bisogna anche dire che, fino a questo punto, la quantità  $\mathbf{P}(\nu = k|\pi = p)$  non è stata ancora correttamente definita in quanto  $\mathbf{P}(\pi = p) = 0$  dato che la distribuzione di probabilità scelta per  $\pi$  è continua (vedi la parte finale dell'Osservazione II.3.1).  $\diamond$

**II.7.3 Osservazione:** Per aggirare i problemi posti dal possibile annullamento della probabilità degli eventi condizionanti partiremo dal concetto di  $\mathbf{E}(\xi|\mathcal{A})$  definito nell'Osservazione II.7.1 e tenteremo di ridefinirlo in modo tale che gli eventi della decomposizione che genera  $\mathcal{A}$  (e soprattutto le loro probabilità) non compaiano più esplicitamente. Innanzitutto abbiamo notato che  $\mathbf{E}(\xi|\mathcal{A})$  risulta ovviamente essere una v.a.  $\mathcal{A}$ -misurabile, mentre  $\xi$  è soltanto  $\mathcal{F}$ -misurabile e quindi,

essendo  $\mathcal{A} \subseteq \mathcal{F}$ , in generale non  $\mathcal{A}$ -misurabile. Inoltre, per eliminare la presenza esplicita degli eventi  $D_k$ , osserviamo che, comunque preso  $A \in \mathcal{A}$ , questo risulterà costituito dall'unione finita di eventi  $D_k$  di  $\mathcal{D}$ . Detto allora  $N_A = \{j \in \mathbf{N} : D_j \subseteq A\}$ , si ha che

$$A = \bigcup_{k \in N_A} D_k, \quad \forall A \in \mathcal{A},$$

e quindi

$$\mathbf{E}(\xi|\mathcal{A})I_A = \sum_k \frac{\mathbf{E}(\xi I_{D_k})}{\mathbf{P}(D_k)} I_{D_k} I_A = \sum_{k \in N_A} \frac{\mathbf{E}(\xi I_{D_k})}{\mathbf{P}(D_k)} I_{D_k},$$

sicch  dalle propriet  elementari dei VdA risulta

$$\begin{aligned} \mathbf{E}[\mathbf{E}(\xi|\mathcal{A})I_A] &= \sum_{k \in N_A} \frac{\mathbf{E}(\xi I_{D_k})}{\mathbf{P}(D_k)} \mathbf{E}(I_{D_k}) = \sum_{k \in N_A} \mathbf{E}(\xi I_{D_k}) \\ &= \mathbf{E}\left(\xi \sum_{k \in N_A} I_{D_k}\right) = \mathbf{E}(\xi I_A). \end{aligned}$$

In conclusione possiamo asserire che, se  $\xi$    una generica v.a. e  $\mathcal{A}$  un'algebra finita di eventi, la v.a.  $\mathbf{E}(\xi|\mathcal{A})$  gode delle seguenti due propriet :

- (a)  $\mathbf{E}(\xi|\mathcal{A})$     $\mathcal{A}$ -misurabile (mentre  $\xi$  in generale non lo  );
- (b)  $\mathbf{E}[\mathbf{E}(\xi|\mathcal{A})I_A] = \mathbf{E}(\xi I_A)$ ,  $\forall A \in \mathcal{A}$ .

Si potrebbe inoltre mostrare che tali due propriet  sono caratteristiche del VdA condizionato da noi definito, nel senso che la v.a.  $\mathbf{E}(\xi|\mathcal{A})$  che le soddisfa esiste sempre ed   definita  $\mathbf{P}$ -q.o. in maniera unica dalle propriet  (a) e (b). Non verificheremo qui questa affermazione perch , siccome nelle propriet  (a) e (b) non compaiono pi  in maniera esplicita n  la richiesta che  $\mathcal{A}$  sia un'algebra generata da una decomposizione finita, n  l'ipotesi che gli atomi della decomposizione abbiano probabilit  non nulla, tenteremo ora di dare una definizione rigorosa del condizionamento proprio per mezzo di queste due propriet  e ci riserveremo quindi di verificare direttamente in quella sede che la definizione   ben posta.  $\circ$

**II.7.4 Definizione:** Sia  $\xi$  una v.a. definita su  $(\Omega, \mathcal{F}, \mathbf{P})$  e sia  $\mathcal{G} \subseteq \mathcal{F}$  una sotto  $\sigma$ -algebra di  $\mathcal{F}$ : se  $\xi \geq 0$  chiameremo **Valore d'Attesa di  $\xi$  condizionato dalla  $\sigma$ -algebra  $\mathcal{G}$**  la v.a. estesa e non negativa  $\mathbf{E}(\xi|\mathcal{G})$  tale che

- (a)  $\mathbf{E}(\xi|\mathcal{G})$     $\mathcal{G}$ -misurabile;
- (b)  $\mathbf{E}(\xi I_A) = \int_A \xi d\mathbf{P} = \int_A \mathbf{E}(\xi|\mathcal{G}) d\mathbf{P} = \mathbf{E}[\mathbf{E}(\xi|\mathcal{G})I_A]$ ,  $\forall A \in \mathcal{G}$ .

Se invece  $\xi$    generica,  $\mathbf{E}(\xi|\mathcal{G})$    una v.a. estesa che si considera definita se almeno una delle due v.a.  $\mathbf{E}(\xi^+|\mathcal{G})$  e  $\mathbf{E}(\xi^-|\mathcal{G})$  risulta finita ( $\mathbf{P}$ -q.o.), e in tal caso si pone

$$\mathbf{E}(\xi|\mathcal{G}) = \mathbf{E}(\xi^+|\mathcal{G}) - \mathbf{E}(\xi^-|\mathcal{G})$$

in tutti i punti  $\omega \in \Omega$  in cui questo valore   ben definito (cio  non   del tipo  $\infty - \infty$ ), mentre nel restante insieme di punti (che   di  $\mathbf{P}$ -misura nulla) si assegna a  $\mathbf{E}(\xi|\mathcal{G})$  un valore arbitrario, ad esempio 0. In particolare, se  $\mathcal{G} = \mathcal{F}_Y$    la  $\sigma$ -algebra

generata da un elemento aleatorio  $Y$ , chiameremo  $\mathbf{E}(\xi|\mathcal{F}_Y)$  **Valor d'Attesa di  $\xi$  condizionato da  $Y$**  e lo indicheremo, per brevità, con il simbolo  $\mathbf{E}(\xi|Y)$ .  $\triangle$

**II.7.5 Osservazione:** Per considerare questa definizione come ben posta dobbiamo ora mostrare che, per v.a. non negative, la v.a. estesa  $\mathbf{E}(\xi|\mathcal{G})$  in realtà esiste ed è univocamente determinata (**P**-q.o.) dalle nostre condizioni. Notiamo innanzitutto (vedi Osservazione II.6.23) che la relazione

$$\mathbf{Q}(A) = \mathbf{E}(\xi I_A) = \int_A \xi d\mathbf{P}, \quad A \in \mathcal{G}$$

definisce su  $(\Omega, \mathcal{G})$  una misura (positiva) che risulta a.c. rispetto alla restrizione della misura  $\mathbf{P}$  a  $(\Omega, \mathcal{G})$ . Il Teorema di Radon-Nikodym II.6.24 ci garantisce allora che esiste una v.a. estesa, unica **P**-q.o. e  $\mathcal{G}$ -misurabile, che indicheremo con  $\mathbf{E}(\xi|\mathcal{G})$  e tale che

$$\mathbf{E}(\xi I_A) = \mathbf{Q}(A) = \int_A \mathbf{E}(\xi|\mathcal{G}) d\mathbf{P} = \mathbf{E}[\mathbf{E}(\xi|\mathcal{G}) I_A].$$

A prima vista potrebbe sembrare artificioso introdurre una nuova v.a.  $\mathbf{E}(\xi|\mathcal{G})$  per giocare il ruolo di densità della misura  $\mathbf{Q}$  in quanto, dato che  $\mathbf{Q}(A) = \int_A \xi d\mathbf{P}$ , parrebbe naturale poter usare  $\xi$  stessa a questo scopo. Qui appare chiaramente, invece, l'importanza della proprietà (a) della Definizione II.7.4: infatti  $\xi$  non è, in generale, una v.a.  $\mathcal{G}$ -misurabile e quindi non può essere considerata (secondo il Teorema II.6.24) come densità di  $\mathbf{Q}$  rispetto a  $\mathbf{P}$  in quanto queste sono state intese come misure su  $(\Omega, \mathcal{G})$ . Resta inteso, ovviamente, che se  $\xi$  dovesse risultare anche  $\mathcal{G}$ -misurabile, allora si avrebbe banalmente  $\mathbf{E}(\xi|\mathcal{G}) = \xi$ , come sarà richiamato al punto (F) di II.7.9. Il VdA condizionato da  $\mathcal{G}$  rimane, pertanto, ben definito dalle condizioni date il II.7.4 ed in particolare potremo dire che

$$\mathbf{E}(\xi|\mathcal{G}) = \frac{d\mathbf{Q}}{d\mathbf{P}}$$

nel senso della derivata di Radon-Nikodym. Siccome però il Teorema II.6.24 definisce la densità solo **P**-q.o.,  $\mathbf{E}(\xi|\mathcal{G})$  non è unico, ma le differenti funzioni  $\mathcal{G}$ -misurabili  $\eta(\omega)$  tali che  $\mathbf{Q}(A) = \int_A \eta d\mathbf{P}$  (comunque preso  $A \in \mathcal{G}$ ) risultano diverse fra loro solo su insiemi di punti di **P**-misura nulla: ognuna di queste possibili v.a.  $\mathcal{G}$ -misurabili si dice **variante** del VdA di  $\xi$  condizionato da  $\mathcal{G}$ .

Anche il VdA  $\mathbf{E}\xi$  può essere considerato come un caso particolare di VdA condizionato se si esegue il condizionamento prendendo in considerazione la  $\sigma$ -algebra banale  $\mathcal{G}_* = \{\Omega, \emptyset\}$ . In tal caso, infatti, si ha  $\mathbf{E}(\xi|\mathcal{G}_*) = \mathbf{E}\xi$ , come si mostrerà al punto (E) della Proposizione II.7.9. Bisogna inoltre osservare che la v.a.  $\mathbf{E}(\xi|\mathcal{G})$  può risultare definita anche quando  $\mathbf{E}\xi$  non è definito. Supponiamo, ad esempio, che  $\xi$  sia una v.a. per la quale  $\mathbf{E}\xi^+ = \mathbf{E}\xi^- = +\infty$  (sicché  $\mathbf{E}\xi$  risulta non definito) e che  $\mathcal{G} = \mathcal{F}$ : in questo caso, siccome  $\xi^+$  e  $\xi^-$  sono  $\mathcal{G}$ -misurabili (in quanto  $\mathcal{G} = \mathcal{F}$ ), si ha ovviamente che

$$\mathbf{E}(\xi^+|\mathcal{G}) = \mathbf{E}(\xi^+|\mathcal{F}) = \xi^+, \quad \mathbf{E}(\xi^-|\mathcal{G}) = \mathbf{E}(\xi^-|\mathcal{F}) = \xi^-,$$



e quindi, dalla Definizione II.7.4,  $\mathbf{E}(\xi|\mathcal{G}) = \xi^+ - \xi^- = \xi$ .

Infine bisognerà mostrare che la definizione di  $\mathbf{E}(\xi|\mathcal{G})$  data in II.7.4 si riduce correttamente alle formule elementari data nell'Osservazione II.7.1 nel caso particolare in cui  $\mathcal{G} = \sigma(\mathcal{D})$  è la  $\sigma$ -algebra generata da una decomposizione finita di  $\Omega$  in eventi  $D_k$  con  $\mathbf{P}(D_k) > 0$  (vedi Definizione II.5.21 e Proposizione II.5.22). In pratica proveremo nella successiva Proposizione che in questo caso  $\mathbf{E}(\xi|\mathcal{G})$  vale  $\mathbf{E}(\xi|D_k)$   $\mathbf{P}$ -q.o. su ogni evento  $D_k$ . Per la precisione noteremo che date due v.a.  $\xi$  e  $\eta$  l'espressione  $\xi = \eta$ ,  $\mathbf{P}$ -q.o. su  $A$  vuol dire che  $\mathbf{P}(A \cap \{\xi \neq \eta\}) = 0$ .  $\circ$

**II.7.6 Proposizione:** Se  $\xi$  è una v.a., se  $\mathbf{E}\xi$  è definito e se  $\mathcal{G} = \sigma(\mathcal{D})$  dove  $\mathcal{D}$  è una decomposizione finita in eventi  $D_k$  con  $\mathbf{P}(D_k) > 0$ , si ha

$$\mathbf{E}(\xi|\mathcal{G}) = \mathbf{E}(\xi|D_k) = \frac{\mathbf{E}(\xi I_{D_k})}{\mathbf{P}(D_k)},$$

$\mathbf{P}$ -q.o. su  $D_k$ .

**Dimostrazione:** Siccome  $\mathbf{E}(\xi|\mathcal{G})$  è  $\mathcal{G}$ -misurabile la Proposizione II.5.22 ci assicura che  $\mathbf{E}(\xi|\mathcal{G})$  è costante ( $\mathbf{P}$ -q.o.) sugli eventi  $D_k$ . Dette allora  $c_k$  tali costanti, da (b) della Definizione II.7.4 si ha

$$\int_{D_k} \xi d\mathbf{P} = \int_{D_k} \mathbf{E}(\xi|\mathcal{G}) d\mathbf{P} = c_k \int_{D_k} d\mathbf{P} = c_k \mathbf{P}(D_k)$$

da cui, essendo  $\mathbf{P}(D_k) > 0$ , deriva

$$c_k = \frac{1}{\mathbf{P}(D_k)} \int_{D_k} \xi d\mathbf{P} = \frac{\mathbf{E}(\xi I_{D_k})}{\mathbf{P}(D_k)} = \mathbf{E}(\xi|D_k). \quad \square$$

**II.7.7. Definizione:** Dato un evento  $C \in \mathcal{F}$  chiameremo **Probabilità di  $C$  condizionata dalla  $\sigma$ -algebra  $\mathcal{G}$**  la v.a.

$$\mathbf{P}(C|\mathcal{G}) = \mathbf{E}(I_C|\mathcal{G}).$$

Inoltre, nel caso in cui  $\mathcal{G} = \mathcal{F}_Y$  risulta essere la  $\sigma$ -algebra generata da un elemento aleatorio  $Y$ , parleremo di **Probabilità di  $C$  condizionata da  $Y$**  ed useremo indicarla, per brevità, con il simbolo  $\mathbf{P}(C|Y) = \mathbf{P}(C|\mathcal{F}_Y)$ .  $\triangle$

**II.7.8 Osservazione:** Si ricava subito dalla Definizione precedente che, dato  $C \in \mathcal{F}$ ,  $\mathbf{P}(C|\mathcal{G})$  verifica le seguenti due proprietà:

(A)  $\mathbf{P}(C|\mathcal{G})$  è  $\mathcal{G}$ -misurabile;

(B) per  $A \in \mathcal{G}$  si ha  $\mathbf{P}(AC) = \mathbf{E}(I_A I_C) = \mathbf{E}[\mathbf{E}(I_C|\mathcal{G}) I_A] = \int_A \mathbf{P}(C|\mathcal{G}) d\mathbf{P}$ .

Bisognerà però guardarsi, a questo punto, dal pensare che, per ogni dato  $\omega \in \Omega$ , la  $\mathbf{P}(\cdot|\mathcal{G})(\omega)$  rappresenti una misura di probabilità su  $(\Omega, \mathcal{F})$ . Infatti mostreremo

più oltre che l'insieme dei punti  $\omega \in \Omega$  per i quali  $\mathbf{P}(\cdot | \mathcal{G})(\omega)$  non è una misura di probabilità non solo non è vuoto, ma non è neanche, in generale, di  $\mathbf{P}$ -misura nulla. Pertanto chiamare  $\mathbf{P}(C|\mathcal{G})$  con il nome di *probabilità* condizionata è, per il momento, strettamente parlando un abuso di linguaggio. Accontentandoci, per ora, di aver messo in evidenza il problema, rimandiamo una discussione approfondita di questo punto importante ad un momento successivo.  $\circ$

**II.7.9 Proposizione:** Dato  $(\Omega, \mathcal{F}, \mathbf{P})$ , se  $\xi$  e  $\eta$  sono v.a. con VdA definiti, se  $\mathcal{G}, \mathcal{G}_1, \mathcal{G}_2$  sono sotto  $\sigma$ -algebre di  $\mathcal{F}$  (ed in particolare  $\mathcal{G}_* = \{\Omega, \emptyset\}$ ) e se  $a, b, c$  sono costanti, si ha:

- (A)  $\xi = c, \mathbf{P}\text{-q.o.} \Rightarrow \mathbf{E}(\xi|\mathcal{G}) = c, \mathbf{P}\text{-q.o.};$
- (B)  $\xi \leq \eta, \mathbf{P}\text{-q.o.} \Rightarrow \mathbf{E}(\xi|\mathcal{G}) \leq \mathbf{E}(\eta|\mathcal{G}), \mathbf{P}\text{-q.o.};$
- (C)  $|\mathbf{E}(\xi|\mathcal{G})| \leq \mathbf{E}(|\xi| | \mathcal{G}), \mathbf{P}\text{-q.o.};$
- (D)  $\mathbf{E}(a\xi + b\eta|\mathcal{G}) = a\mathbf{E}(\xi|\mathcal{G}) + b\mathbf{E}(\eta|\mathcal{G}),$  se  $a\mathbf{E}\xi + b\mathbf{E}\eta$  è definito;
- (E)  $\mathbf{E}(\xi|\mathcal{G}_*) = \mathbf{E}\xi, \mathbf{P}\text{-q.o.};$
- (F)  $\mathbf{E}(\xi|\mathcal{F}) = \xi, \mathbf{P}\text{-q.o.};$
- (G)  $\mathcal{G}_1 \subseteq \mathcal{G}_2 \Rightarrow \mathbf{E}[\mathbf{E}(\xi|\mathcal{G}_2) | \mathcal{G}_1] = \mathbf{E}(\xi|\mathcal{G}_1), \mathbf{P}\text{-q.o.};$
- (H)  $\mathcal{G}_2 \subseteq \mathcal{G}_1 \Rightarrow \mathbf{E}[\mathbf{E}(\xi|\mathcal{G}_2) | \mathcal{G}_1] = \mathbf{E}(\xi|\mathcal{G}_2), \mathbf{P}\text{-q.o.};$
- (I)  $\mathbf{E}[\mathbf{E}(\xi|\mathcal{G})] = \mathbf{E}\xi;$
- (J) se  $\xi$  è indipendente da  $\mathcal{G}$ , allora  $\mathbf{E}(\xi|\mathcal{G}) = \mathbf{E}\xi, \mathbf{P}\text{-q.o.}$

**Dimostrazione:**

- (A) Basterà osservare che la v.a. identicamente eguale a  $c$  è ovviamente  $\mathcal{G}$ -misurabile e che inoltre, da (G) di II.6.9, risulta  $\mathbf{E}(\xi I_A) = \mathbf{E}(c I_A)$  comunque scelto  $A \in \mathcal{G}$ ; segue pertanto dalla Definizione II.7.4 che  $\mathbf{E}(\xi|\mathcal{G}) = c$ .
- (B) Se  $\xi \leq \eta$  (la limitazione  $\mathbf{P}\text{-q.o.}$  non è qui rilevante come risulta da (F) di II.6.9), da (B) di II.6.8 risulta  $\mathbf{E}(\xi I_A) \leq \mathbf{E}(\eta I_A)$  comunque scelto  $A \in \mathcal{G}$  e quindi, dalla definizione di VdA condizionato,  $\mathbf{E}[\mathbf{E}(\xi|\mathcal{G}) I_A] \leq \mathbf{E}[\mathbf{E}(\eta|\mathcal{G}) I_A]$ , per ogni  $A \in \mathcal{G}$ ; la tesi segue allora da (I) di II.6.9.
- (C) Segue dalla proprietà precedente dato che  $-|\xi| \leq \xi \leq |\xi|$ .
- (D) Per la linearità dei VdA (vedi II.6.8) si ha, comunque scelto  $A \in \mathcal{G}$ , che

$$\begin{aligned} \mathbf{E}[(a\xi + b\eta) I_A] &= \mathbf{E}(a\xi I_A) + \mathbf{E}(b\eta I_A) = \mathbf{E}[a\mathbf{E}(\xi|\mathcal{G}) I_A] + \mathbf{E}[b\mathbf{E}(\eta|\mathcal{G}) I_A] \\ &= \mathbf{E}[(a\mathbf{E}(\xi|\mathcal{G}) + b\mathbf{E}(\eta|\mathcal{G})) I_A], \end{aligned}$$

da cui segue la tesi in base alla Definizione II.7.4.

- (E) La v.a.  $\mathbf{E}(\xi|\mathcal{G}_*)$ , per essere  $\mathcal{G}_*$ -misurabile, deve essere  $\mathbf{P}\text{-q.o.}$  costante; detta allora  $c$  tale costante, la (b) della II.7.4 richiede che  $\mathbf{E}(c I_A) = \mathbf{E}(\xi I_A)$  con  $A \in \mathcal{G}_*$ : se in particolare si prende  $A = \Omega$  si ottiene subito che  $c = \mathbf{E}\xi$ .
- (F)  $\xi$  è  $\mathcal{F}$ -misurabile e inoltre, banalmente, si ha  $\mathbf{E}(\xi I_A) = \mathbf{E}(\xi I_A), \forall A \in \mathcal{F}$ , e quindi  $\mathbf{E}(\xi|\mathcal{F}) = \xi$ .
- (G) Se  $A \in \mathcal{G}_1 \subseteq \mathcal{G}_2$ , risultano contemporaneamente vere le due relazioni

$$\begin{aligned} \mathbf{E}[\mathbf{E}(\xi | \mathcal{G}_1) I_A] &= \mathbf{E}(\xi I_A) \\ \mathbf{E}[\mathbf{E}[\mathbf{E}(\xi | \mathcal{G}_2) | \mathcal{G}_1] I_A] &= \mathbf{E}[\mathbf{E}(\xi | \mathcal{G}_2) I_A] = \mathbf{E}(\xi I_A) \end{aligned}$$

e pertanto le due v.a.  $\mathcal{G}_1$ -misurabili  $\mathbf{E}(\xi | \mathcal{G}_1)$  e  $\mathbf{E}[\mathbf{E}(\xi | \mathcal{G}_2) | \mathcal{G}_1]$  sono tali che

$$\mathbf{E}[\mathbf{E}(\xi | \mathcal{G}_1)I_A] = \mathbf{E}(\mathbf{E}[\mathbf{E}(\xi | \mathcal{G}_2) | \mathcal{G}_1]I_A), \quad \forall A \in \mathcal{G}_1;$$

dalla (I) di II.6.9 segue allora  $\mathbf{E}(\xi | \mathcal{G}_1) = \mathbf{E}[\mathbf{E}(\xi | \mathcal{G}_2) | \mathcal{G}_1]$ , **P**-q.o.

- (H) Discende banalmente da (F) se si tiene conto del fatto che, essendo  $\mathcal{G}_2 \subseteq \mathcal{G}_1$ ,  $\mathbf{E}(\xi | \mathcal{G}_2)$  risulta anche, ovviamente,  $\mathcal{G}_1$ -misurabile.
- (I) Segue da (E) e (G) se si pone  $\mathcal{G}_1 = \mathcal{G}_* = \{\Omega, \emptyset\}$  e  $\mathcal{G}_2 = \mathcal{G}$ : infatti risulta che  $\mathbf{E}[\mathbf{E}(\xi | \mathcal{G})] = \mathbf{E}[\mathbf{E}(\xi | \mathcal{G}) | \mathcal{G}_*] = \mathbf{E}(\xi | \mathcal{G}_*) = \mathbf{E}\xi$ .
- (J) La v.a. costante eguale a  $\mathbf{E}\xi$  è  $\mathcal{G}$ -misurabile e inoltre, data l'indipendenza di  $\xi$  da  $\mathcal{G}$  (cioè da qualunque  $I_A$  con  $A \in \mathcal{G}$ ), si ha<sup>1</sup> dal Teorema II.6.15 che

$$\mathbf{E}(\xi I_A) = \mathbf{E}\xi \cdot \mathbf{E}I_A = \mathbf{E}(I_A \mathbf{E}\xi), \quad \forall A \in \mathcal{G},$$

e quindi  $\mathbf{E}(\xi | \mathcal{G}) = \mathbf{E}\xi$ , **P**-q.o. □

**II.7.10 Proposizione:** Se  $(\xi_n)_{n \in \mathbf{N}}$  è una successione di v.a. estese i teoremi di convergenza per i VdA assumono le forme seguenti per i VdA condizionati:

(a) se  $|\xi_n| \leq \eta$ ,  $\mathbf{E}\eta < +\infty$  e  $\xi_n \xrightarrow{n} \xi$ , **P**-q.o., allora  $\mathbf{E}(\xi_n | \mathcal{G}) \xrightarrow{n} \mathbf{E}(\xi | \mathcal{G})$  e inoltre  $\mathbf{E}(|\xi_n - \xi| | \mathcal{G}) \xrightarrow{n} 0$ , **P**-q.o.

(b) se  $\eta \leq \xi_n$  ( $\eta \geq \xi_n$ ),  $-\infty < \mathbf{E}\eta$  ( $\mathbf{E}\eta < +\infty$ ) e  $\xi_n \uparrow \xi$  ( $\xi_n \downarrow \xi$ ) **P**-q.o., allora  $\mathbf{E}(\xi_n | \mathcal{G}) \uparrow \mathbf{E}(\xi | \mathcal{G})$  ( $\mathbf{E}(\xi_n | \mathcal{G}) \downarrow \mathbf{E}(\xi | \mathcal{G})$ ), **P**-q.o.

(c) se  $\eta \leq \xi_n$  ( $\eta \geq \xi_n$ ),  $-\infty < \mathbf{E}\eta$  ( $\mathbf{E}\eta < +\infty$ ) **P**-q.o., allora  $\mathbf{E}(\underline{\lim}_n \xi_n | \mathcal{G}) \leq \underline{\lim}_n \mathbf{E}(\xi_n | \mathcal{G})$  ( $\mathbf{E}(\underline{\lim}_n \xi_n | \mathcal{G}) \geq \underline{\lim}_n \mathbf{E}(\xi_n | \mathcal{G})$ ), **P**-q.o.

(d) se  $\xi_n \geq 0$ , allora  $\mathbf{E}(\sum_n \xi_n | \mathcal{G}) = \sum_n \mathbf{E}(\xi_n | \mathcal{G})$ , **P**-q.o.

**Dimostrazione:** Omessa (vedi **A.N. Shiriyayev**: *Probability*; Springer, New York, 1984, p. 216.) □

**II.7.11 Proposizione:** Con le notazioni della Proposizione II.7.9 si ha:

(K) se  $\eta$  è  $\mathcal{G}$ -misurabile,  $\mathbf{E}|\eta| < +\infty$  e  $\mathbf{E}|\xi| < +\infty$  (e quindi anche  $\mathbf{E}|\xi\eta| < +\infty$ ), allora  $\mathbf{E}(\xi\eta | \mathcal{G}) = \eta\mathbf{E}(\xi | \mathcal{G})$ , **P**-q.o.

**Dimostrazione:** È facile mostrare che la tesi è vera per indicatori  $\eta = I_B$  che siano  $\mathcal{G}$ -misurabili (cioè tali che  $B \in \mathcal{G}$ ): infatti comunque scelto  $A \in \mathcal{G}$  si ha

$$\mathbf{E}(\xi\eta I_A) = \mathbf{E}(\xi I_{AB}) = \mathbf{E}(\mathbf{E}(\xi | \mathcal{G})I_{AB}) = \mathbf{E}(\eta\mathbf{E}(\xi | \mathcal{G})I_A)$$

<sup>1</sup> Nel caso in cui  $\mathbf{E}|\xi| < +\infty$  l'applicazione del Teorema II.6.15 è diretta; per il caso più generale dobbiamo invece osservare che il risultato del Teorema II.6.15 continua a valere anche se viene lasciata cadere l'ipotesi di finitezza dei VdA di  $\xi$  e  $\eta$  richiedendo invece che tali v.a. siano non negative.

da cui discende (K) per definizione di VdA condizionato. La proprietà si estende ovviamente anche alle v.a. semplici. Se ora  $\eta$  è una generica v.a.  $\mathcal{G}$ -misurabile con  $\mathbf{E}|\eta| < +\infty$ , sia  $(\eta_n)_{n \in \mathbf{N}}$  la successione di v.a. semplici e  $\mathcal{G}$ -misurabili che, secondo il Teorema II.5.16, verifica  $|\eta_n| \leq |\eta|$  e  $\eta_n \xrightarrow{n} \eta$ . Le osservazioni precedenti ci consentono allora di scrivere che

$$\mathbf{E}(\xi\eta_n | \mathcal{G}) = \eta_n \mathbf{E}(\xi | \mathcal{G}), \quad \mathbf{P}\text{-q.o.} \quad \forall n \in \mathbf{N}.$$

Ma siccome  $|\xi\eta_n| \leq |\xi\eta|$  e, per ipotesi,  $\mathbf{E}|\xi\eta| < +\infty$ , da (a) del Teorema II.7.10 si ha

$$\mathbf{E}(\xi\eta_n | \mathcal{G}) \xrightarrow{n} \mathbf{E}(\xi\eta | \mathcal{G}), \quad \mathbf{P}\text{-q.o.}$$

La proprietà (K) deriva allora dal fatto che è anche verificata la relazione

$$\eta_n \mathbf{E}(\xi | \mathcal{G}) \xrightarrow{n} \eta \mathbf{E}(\xi | \mathcal{G}), \quad \mathbf{P}\text{-q.o.}$$

e per dimostrare quest'ultima basterà provare<sup>2</sup> che  $|\mathbf{E}(\xi | \mathcal{G})| < +\infty$ ,  $\mathbf{P}$ -q.o. Osserviamo allora che da (I) della proposizione II.7.9 si ha  $\mathbf{E}[\mathbf{E}(|\xi| | \mathcal{G})] = \mathbf{E}|\xi| < +\infty$ ; pertanto la v.a.  $\mathbf{E}(|\xi| | \mathcal{G})$ , non negativa, risulta integrabile e quindi, da (J) di II.6.10, risulta anche  $\mathbf{E}(|\xi| | \mathcal{G}) < +\infty$ ,  $\mathbf{P}$ -q.o. Siccome infine da (C) di II.7.9 si ha  $|\mathbf{E}(\xi | \mathcal{G})| \leq \mathbf{E}(|\xi| | \mathcal{G})$ , si ottiene che  $|\mathbf{E}(\xi | \mathcal{G})| < +\infty$ ,  $\mathbf{P}$ -q.o.  $\square$

**II.7.12 Osservazione:** È importante ora esaminare in dettaglio il concetto di VdA di  $\xi$  condizionato da una v.a.  $\eta$ :  $\mathbf{E}(\xi|\eta) = \mathbf{E}(\xi|\mathcal{G}_\eta)$  (vedi Definizione II.7.4) è una nuova v.a.  $\mathcal{G}_\eta$ -misurabile, con  $\mathcal{G}_\eta \subseteq \mathcal{F}$  generata da  $\eta$ , sicché il Teorema II.5.19 garantisce l'esistenza di una funzione di Borel  $m(y) : \overline{\mathbf{R}} \rightarrow \overline{\mathbf{R}}$  tale che  $\mathbf{E}(\xi|\eta) = m(\eta)$ ,  $\forall \omega \in \Omega$ . Indicheremo tale funzione con il simbolo

$$m(y) = \mathbf{E}(\xi|\eta = y)$$

e la chiameremo **valore d'attesa di  $\xi$  condizionato dall'evento  $\{\eta = y\}$** . Notiamo inoltre che, comunque scelto  $A \in \mathcal{G}_\eta$ , la seguente catena di eguaglianze

$$\int_A \xi d\mathbf{P} = \mathbf{E}(\xi I_A) = \mathbf{E}[\mathbf{E}(\xi|\mathcal{G}_\eta)I_A] = \mathbf{E}[\mathbf{E}(\xi|\eta)I_A] = \mathbf{E}[m(\eta)I_A] = \int_A m(\eta) d\mathbf{P}$$

può essere riscritta diversamente tenendo conto del Teorema II.6.25 e del fatto che ogni  $A \in \mathcal{G}_\eta$  è del tipo  $A = \eta^{-1}(B)$  per un qualche  $B \in \mathcal{B}(\overline{\mathbf{R}})$ : infatti comunque scelto  $B \in \mathcal{B}(\overline{\mathbf{R}})$  si ha che

$$\int_A \xi d\mathbf{P} = \int_{\eta^{-1}(B)} \xi d\mathbf{P} = \int_{\eta^{-1}(B)} m(\eta) d\mathbf{P} = \int_B m(y) \mathbf{P}_\eta(dy).$$

---

<sup>2</sup> Questa precisazione è necessaria in quanto, ad esempio,  $\frac{1}{n} \cdot \infty = \infty$ ,  $\forall n \in \mathbf{N}$ , mentre  $0 \cdot \infty = 0$ , sicché dove l'attesa condizionata non è finita la convergenza non può essere immediatamente assicurata.

Questa osservazione ci permette allora di definire il VdA condizionato da  $\eta = y$  senza usare direttamente la v.a.  $\mathbf{E}(\xi|\eta)$  nel modo seguente: date le v.a. (anche estese)  $\xi$  e  $\eta$  con  $\mathbf{E}\xi$  definito chiameremo VdA di  $\xi$  condizionato da  $\eta = y$ , e lo indicheremo con  $\mathbf{E}(\xi|\eta = y)$ , ogni funzione di Borel  $m(y) : \overline{\mathbf{R}} \rightarrow \overline{\mathbf{R}}$  tale che

$$\int_{\eta^{-1}(B)} \xi d\mathbf{P} = \int_B m(y) \mathbf{P}_\eta(dy) = \int_B \mathbf{E}(\xi|\eta = y) \mathbf{P}_\eta(dy), \quad \forall B \in \mathcal{B}(\overline{\mathbf{R}}).$$

Questa seconda definizione è ben posta perché anche qui il Teorema II.6.24 di Radon-Nikodym ci garantisce l'esistenza di  $m(y)$  sulla base del fatto, facilmente verificato, che

$$\mathbf{Q}(B) = \int_{\eta^{-1}(B)} \xi d\mathbf{P}, \quad \forall B \in \mathcal{B}(\overline{\mathbf{R}})$$

è una misura segnata su  $(\overline{\mathbf{R}}, \mathcal{B}(\overline{\mathbf{R}}))$  che risulta essere a.c. rispetto alla probabilità  $\mathbf{P}_\eta$ . Inoltre è anche possibile verificare che, una volta definito il VdA condizionato  $m(y)$  in questo secondo modo, la v.a.  $m(\eta)$  coincide ( $\mathbf{P}$ -q.o.) con  $\mathbf{E}(\xi|\eta)$ . Infatti, ricordando che, comunque scelto  $A \in \mathcal{G}_\eta$ , esiste  $B \in \mathcal{B}(\overline{\mathbf{R}})$  tale che  $A = \eta^{-1}(B)$  e che in base al Teorema II.6.25 e alla seconda definizione

$$\int_{\eta^{-1}(B)} \xi d\mathbf{P} = \int_B m(y) \mathbf{P}_\eta(dy) = \int_{\eta^{-1}(B)} m(\eta) d\mathbf{P}, \quad \forall B \in \mathcal{B}(\overline{\mathbf{R}}),$$

otteniamo facilmente

$$\mathbf{E}(\xi I_A) = \int_A \xi d\mathbf{P} = \int_A m(\eta) d\mathbf{P} = \mathbf{E}[m(\eta) I_A], \quad \forall A \in \mathcal{G}_\eta,$$

da cui risulta  $m(\eta) = \mathbf{E}(\xi|\mathcal{G}_\eta) = \mathbf{E}(\xi|\eta)$  a causa della Definizione II.7.7. Ciò mostra in definitiva che le due definizioni qui date di VdA condizionato rispetto a  $\eta = y$  sono perfettamente equivalenti con la precisazione che, come sempre, le nostre funzioni di Borel sono solo  $\mathbf{P}_\eta$ -q.o. definite. Una volta stabilito il concetto di VdA condizionato da  $\eta = y$  è anche possibile dare la definizione di **probabilità di un evento  $A \in \mathcal{F}$  condizionata da  $\{\eta = y\}$**  come

$$\mathbf{P}(A|\eta = y) = \mathbf{E}(I_A|\eta = y).$$

Non si deve però pensare, come già osservato in II.7.8, che, comunque fissato  $y \in \overline{\mathbf{R}}$ , l'applicazione  $\mathbf{P}(\cdot|\eta = y) : \mathcal{F} \rightarrow [0, 1]$  sia una misura di probabilità: in effetti l'espressione *probabilità* condizionata è, a questo punto, impropria, ma per il momento, al solo scopo di poter discutere subito alcuni esempî concreti faremo l'ipotesi che  $\mathbf{P}(\cdot|\eta = y)$  sia una vera probabilità, rinviando alla fine di questo capitolo una discussione delle condizioni sotto cui ciò è effettivamente consentito. Osserviamo infine che anche  $\mathbf{P}(A|\eta = y)$ , come funzione di  $y$ , può essere definita in un secondo modo: infatti, siccome  $\mathbf{P}(A|\eta = y) = \mathbf{E}(I_A|\eta = y)$ , deve risultare

$$\int_{\eta^{-1}(B)} I_A d\mathbf{P} = \int_B \mathbf{P}(A|\eta = y) \mathbf{P}_\eta(dy), \quad \forall B \in \mathcal{B}(\overline{\mathbf{R}}),$$

e quindi, siccome

$$\int_{\eta^{-1}(B)} I_A d\mathbf{P} = \mathbf{E}[I_A I_{\eta^{-1}(B)}] = \mathbf{P}(A \cap \eta^{-1}(B)) = \mathbf{P}(A \cap \{\eta \in B\}),$$

potremo sempre dire che  $\mathbf{P}(A|\eta = y)$  è la funzione di Borel che soddisfa la relazione

$$(1) \quad \mathbf{P}(A \cap \{\eta \in B\}) = \int_B \mathbf{P}(A|\eta = y) \mathbf{P}_\eta(dy),$$

comunque scelto  $B \in \mathcal{B}(\overline{\mathbf{R}})$ . In particolare si ottiene

$$(2) \quad \mathbf{P}(A) = \int_{\mathbf{R}} \mathbf{P}(A|\eta = y) \mathbf{P}_\eta(dy),$$

e, se  $\eta$  è a.c.,

$$(3) \quad \mathbf{P}(A) = \int_{\mathbf{R}} \mathbf{P}(A|\eta = y) f_\eta(y) dy.$$

Naturalmente tutte queste notazioni sono anche estendibili al caso di condizionamenti rispetto a  $\{Y = y\}$  dove  $Y$  è un generico elemento aleatorio.  $\circ$

**II.7.13 Osservazione:** Le proprietà da (A) a (K) dei VdA condizionati rispetto a  $\sigma$ -algebre o v.a. si estendono facilmente anche ai VdA condizionati  $\mathbf{E}(\xi|\eta = y)$ . Osserviamo però che, nell'operare tale estensione, bisognerà aver cura di sostituire, ove necessario, l'espressione  $\mathbf{P}$ -q.o. che si riferisce alla misura  $\mathbf{P}$  su  $(\Omega, \mathcal{F})$  ed è appropriata al caso di v.a., con l'espressione  $\mathbf{P}_\eta$ -q.o. riferita alla misura  $\mathbf{P}_\eta$  su  $(\overline{\mathbf{R}}, \mathcal{B}(\overline{\mathbf{R}}))$  adatta per le funzioni di Borel. A questo proposito notiamo che, se  $f$  e  $g$  sono due funzioni di Borel, posto  $A = \{\omega \in \Omega : f(\eta) \neq g(\eta)\}$  e  $B = \{y \in \overline{\mathbf{R}} : f(y) \neq g(y)\}$  si ha facilmente che  $A = \eta^{-1}(B)$  e quindi che  $\mathbf{P}_\eta(B) = \mathbf{P}(\eta^{-1}(B)) = \mathbf{P}(A)$ ; pertanto  $\mathbf{P}(A) = 0$  equivale a  $\mathbf{P}_\eta(B) = 0$ , sicché l'affermazione  $f(\eta) = g(\eta)$ ,  $\mathbf{P}$ -q.o. sarà equivalente all'affermazione  $f(y) = g(y)$ ,  $\mathbf{P}_\eta$ -q.o. Esamineremo ora alcune delle forme particolari assunte dalle proprietà dei VdA condizionati da  $\eta = y$ .

(a) Da (K) di II.7.11 sappiamo che, comunque data una funzione di Borel  $f(y)$  risulta

$$\mathbf{E}[\xi f(\eta) | \eta] = f(\eta) \mathbf{E}(\xi|\eta), \quad \mathbf{P}\text{-q.o.},$$

dato che  $f(\eta)$  è  $\mathcal{G}_\eta$ -misurabile. Se ora ricordiamo che la v.a.  $\mathbf{E}(\xi|\eta)$  risulta dall'applicazione della funzione di Borel  $\mathbf{E}(\xi|\eta = y)$  alla v.a.  $\eta$ , dall'eguaglianza precedente potremo anche dedurre che

$$\mathbf{E}[\xi f(\eta) | \eta = y] = f(y) \mathbf{E}(\xi|\eta = y), \quad \mathbf{P}_\eta\text{-q.o.},$$

dato che  $\mathbf{E}[\xi f(\eta) | \eta]$  risulterà ora dall'applicazione della funzione di Borel  $f(y) \mathbf{E}(\xi|\eta = y)$  alla v.a.  $\eta$ .

- (b) Da (J) di II.7.9 si ha che, se  $\xi$  e  $\eta$  sono indipendenti, si ha  $\mathbf{E}(\xi|\eta) = \mathbf{E}\xi$  ( $\mathbf{P}$ -q.o.). Ne segue che anche  $\mathbf{E}(\xi|\eta = y) = \mathbf{E}\xi$  ( $\mathbf{P}_{\eta}$ -q.o.).
- (c) Se  $\xi$  e  $\eta$  sono due v.a. e  $\varphi(x, y)$  una funzione di Borel  $\varphi : (\mathbf{R}^2, \mathcal{B}(\mathbf{R}^2)) \rightarrow (\overline{\mathbf{R}}, \mathcal{B}(\overline{\mathbf{R}}))$ , si può provare che

$$\mathbf{E}[\varphi(\xi, \eta) \mid \eta = y] = \mathbf{E}[\varphi(\xi, y) \mid \eta = y], \quad \mathbf{P}_{\eta}\text{-q.o.}$$

Infatti la nostra affermazione è sicuramente vera nel caso in cui  $\varphi(x, y) = I_B(x, y)$ , dove  $B$  è il rettangolo  $B = B_1 \times B_2 \in \mathcal{B}(\mathbf{R}^2)$ , come si vede dal fatto che da (a) si ha che  $\mathbf{P}_{\eta}$ -q.o.

$$\begin{aligned} \mathbf{E}[I_B(\xi, \eta) \mid \eta = y] &= \mathbf{E}[I_{B_1}(\xi)I_{B_2}(\eta) \mid \eta = y] = I_{B_2}(y)\mathbf{E}[I_{B_1}(\xi) \mid \eta = y] \\ &= \mathbf{E}[I_{B_1}(\xi)I_{B_2}(y) \mid \eta = y] = \mathbf{E}[I_B(\xi, y) \mid \eta = y]. \end{aligned}$$

Il risultato generale si ricava poi applicando i teoremi sulla convergenza monotona, ma noi ometteremo la verifica completa di questa proprietà. In particolare, se  $\xi$  ed  $\eta$  sono indipendenti, da (b) segue anche che

$$\mathbf{E}[\varphi(\xi, \eta) \mid \eta = y] = \mathbf{E}[\varphi(\xi, y)], \quad \mathbf{P}_{\eta}\text{-q.o.}$$

in quanto anche  $\varphi(\xi, y)$  è una v.a. indipendente da  $\eta$  per ogni  $y \in \mathbf{R}$ . ○

**II.7.14 Osservazione:** Se  $\eta$  è una v.a. discreta con valori  $y_k$  e con  $\mathbf{P}(\eta = y_k) > 0$  e  $\sum_k \mathbf{P}(\eta = y_k) = 1$ , da (1) di Osservazione II.7.12 si ha che, dato un arbitrario  $A \in \mathcal{F}$ , risulta

$$\mathbf{P}(A \cap \{\eta = y_k\}) = \int_{\{y_k\}} \mathbf{P}(A|\eta = y) \mathbf{P}_{\eta}(dy) = \mathbf{P}(A|\eta = y_k) \cdot \mathbf{P}(\eta = y_k),$$

dato che su  $\{y_k\}$  la funzione  $\mathbf{P}(A|\eta = y)$  vale costantemente  $\mathbf{P}(A|\eta = y_k)$ . Ne segue pertanto che

$$\mathbf{P}(A|\eta = y_k) = \frac{\mathbf{P}(A \cap \{\eta = y_k\})}{\mathbf{P}(\eta = y_k)},$$

formula che coincide con quella che si assumerebbe come definizione nel caso elementare (vedi I.4 ed I.9). Se poi volessimo definire  $\mathbf{P}(A|\eta = y)$  per tutti i valori di  $y \in \mathbf{R}$  compresi quelli che non coincidono con nessun  $y_k$  (ricorda che in tal caso  $\mathbf{P}(\eta = y) = 0$ ), basterà osservare che a  $\mathbf{P}(A|\eta = y)$  può essere attribuito un valore arbitrario (ad esempio zero) in tutti i punti  $y$  diversi dai possibili valori  $y_k$  di  $\eta$ : infatti tali punti costituiscono un sottinsieme di  $\mathbf{R}$  di  $\mathbf{P}_{\eta}$ -misura nulla e pertanto su di essi il valore di  $\mathbf{P}(A|\eta = y)$  è irrilevante. Allo stesso modo si mostra che, se  $\xi$  è una v.a. con  $\mathbf{E}\xi$  definito,  $\mathbf{E}(\xi|\eta = y)$  è la funzione di Borel che assume i valori

$$\mathbf{E}(\xi|\eta = y_k) = \frac{1}{\mathbf{P}(\eta = y_k)} \int_{\{\eta = y_k\}} \xi d\mathbf{P} = \frac{\mathbf{E}(\xi I_{\{\eta = y_k\}})}{\mathbf{E} I_{\{\eta = y_k\}}}$$

nei punti  $y = y_k$ , e valori arbitrari (ad esempio nulli) nei punti  $y$  che non coincidono con nessuno degli  $y_k$ . ○

**II.7.15 Osservazione:** Se le v.a.  $\xi$  e  $\eta$  possiedono FdD congiunta a.c., cioè se sono dotate di fdd congiunta  $f_{\xi\eta}(x, y)$  tale che

$$\mathbf{P}[(\xi, \eta) \in D] = \int_D f_{\xi\eta}(x, y) dx dy, \quad \forall D \in \mathcal{B}(\mathbf{R}^2),$$

allora (facendo per ora l'ipotesi che le probabilità condizionate in questione esistono come vere misure di probabilità) è anche possibile definire una **densità di probabilità condizionata di  $\xi$  rispetto ad  $\eta$**  (o, viceversa, di  $\eta$  rispetto a  $\xi$ ) come fdd della DdP condizionata  $\mathbf{P}(\xi \in C | \eta = y) = \mathbf{P}_\xi(C | \eta = y)$  con  $C \in \mathcal{B}(\mathbf{R})$ . Infatti, se

$$f_\xi(x) = \int_{-\infty}^{+\infty} f_{\xi\eta}(x, y) dy, \quad f_\eta(y) = \int_{-\infty}^{+\infty} f_{\xi\eta}(x, y) dx$$

sono le fdd marginali di  $\xi$  e  $\eta$ , basterà definire (con una notazione che sarà ripresa e discussa alla fine di questo capitolo)

$$f_{\xi|\eta}(x | y) = \begin{cases} \frac{f_{\xi\eta}(x, y)}{f_\eta(y)} & \text{per } y \text{ tali che } f_\eta(y) \neq 0; \\ 0 & \text{per } y \text{ tali che } f_\eta(y) = 0; \end{cases}$$

e mostrare che comunque preso  $C \in \mathcal{B}(\mathbf{R})$

$$(1) \quad \mathbf{P}(\xi \in C | \eta = y) = \mathbf{P}_\xi(C | \eta = y) = \int_C f_{\xi|\eta}(x | y) dx$$

per aver provato che  $f_{\xi|\eta}(x | y)$  svolge proprio il ruolo di fdd della DdP  $\mathbf{P}_\xi(C | \eta = y)$  per ogni fissato  $y$ . Per dimostrare tutto ciò ci basterà allora verificare che la funzione di Borel  $\int_C f_{\xi|\eta}(x | y)$  soddisfa la definizione (1) di Osservazione II.7.12 di probabilità di  $A = \{\xi \in C\}$  condizionata da  $\eta = y$ . Infatti, comunque scelto  $B \in \mathcal{B}(\mathbf{R})$  si ha

$$\begin{aligned} \int_B \left( \int_C f_{\xi|\eta}(x | y) dx \right) \mathbf{P}_\eta(dy) &= \int_B \left( \int_C f_{\xi\eta}(x, y) dx \right) f_\eta(y) dy \\ &= \int_{C \times B} f_{\xi\eta}(x, y) f_\eta(y) dx dy \\ &= \int_{C \times B} f_{\xi\eta}(x, y) dx dy = \mathbf{P}(\xi \in C, \eta \in B); \end{aligned}$$

e quindi  $\int_C f_{\xi|\eta}(x | y)$  coincide proprio con  $\mathbf{P}(\xi \in C, \eta = y) = \mathbf{P}_\xi(C | \eta = y)$ ,  $\mathbf{P}_\eta$ -q.o. Allo stesso modo è possibile mostrare la validità di formule come

$$(2) \quad \mathbf{E}(\xi | \eta = y) = \int_{-\infty}^{+\infty} x f_{\xi|\eta}(x | y) dx$$



per i VdA condizionati. ○

**II.7.16 Esempio:** Supponiamo che il tempo durante il quale opera, senza rompersi, un componente di un dato apparato sia descritto da una v.a.  $\eta$  con fdd  $f_\eta(y)$ , con la convenzione che, se l'apparato comincia a funzionare all'istante  $y = 0$ , la  $f_\eta(y)$  sarà definita come identicamente nulla per  $y < 0$ . Poniamoci ora il problema di determinare la quantità  $\mathbf{E}(\eta - a | \eta \geq a)$  che rappresenta il VdA del tempo di vita del nostro componente (**vita media**) sotto l'ipotesi che esso sia ancora funzionante al tempo  $y = a > 0$ . Supponendo che  $\mathbf{P}(\eta \geq a) > 0$  (in modo che il nostro problema sia significativo), ricordando che per una data v.a.  $\xi$  e un dato evento  $D$  con  $\mathbf{P}(D) \neq 0$  vale la relazione (vedi Osservazione II.7.1)

$$\mathbf{E}(\xi | D) = \frac{\mathbf{E}(\xi I_D)}{\mathbf{P}(D)}$$

e tenendo conto delle formule (5) e (7) di II.6.26, si ha il risultato richiesto:

$$\mathbf{E}(\eta - a | \eta \geq a) = \frac{\mathbf{E}[(\eta - a)I_{\{\eta \geq a\}}]}{\mathbf{P}(\eta \geq a)} = \frac{\int_a^{+\infty} (y - a)f_\eta(y) dy}{\int_a^{+\infty} f_\eta(y) dy}.$$

In modo analogo (e facendo uso anche del fatto che, essendo  $F_\eta$  a.c., risulta sempre  $\mathbf{P}(\eta = a) = 0$ ) è anche possibile calcolare la FdD condizionata: per  $x > 0$  si ha

$$\begin{aligned} F_\eta(a + x | \eta \geq a) &= F_{\eta-a}(x | \eta - a \geq 0) = \mathbf{P}(\eta - a \leq x | \eta \geq a) \\ &= \mathbf{E}(I_{\{\eta - a \leq x\}} | \eta \geq a) = \frac{\mathbf{E}[I_{\{\eta - a \leq x\}} I_{\{\eta \geq a\}}]}{\mathbf{P}(\eta \geq a)} = \frac{\int_a^{a+x} f_\eta(y) dy}{\int_a^{+\infty} f_\eta(y) dy} \\ &= \frac{F_\eta(x + a) - F_\eta(a)}{1 - F_\eta(a)}, \end{aligned}$$

mentre per  $x \leq 0$  si ha  $F_\eta(a + x | \eta \geq a) = 0$ . È interessante esaminare questi risultati nel caso particolare in cui  $\eta$  è una v.a. esponenziale con fdd e FdD date nell'Esempio II.3.16. In tal caso si calcola infatti che

$$\mathbf{E}\eta = \frac{1}{\lambda}, \quad \mathbf{E}(\eta - a | \eta \geq a) = \frac{1}{\lambda},$$

cioè: la vita media del nostro componente condizionata dal fatto che esso ha funzionato fino all'istante  $y = a$  non dipende da  $a$  ed è sempre eguale a  $\mathbf{E}\eta$ . Analogamente si mostra che per  $x > 0$

$$F_\eta(a + x | \eta \geq a) = \frac{(1 - e^{-\lambda(a+x)}) - (1 - e^{-\lambda a})}{1 - (1 - e^{-\lambda a})} = 1 - e^{-\lambda x}$$

mentre  $F_\eta(a + x | \eta \geq a) = 0$  per  $x \leq 0$ , sicché, comunque scelto  $a \geq 0$  risulterà  $F_\eta(a + x | \eta \geq a) = F_\eta(x)$ , cioè la FdD di  $\eta - a$  condizionata da  $\eta \geq a$  non dipende

da  $a$  ma resta sempre eguale alla FdD non condizionata. Un risultato del tutto analogo vale per la fdd condizionata che si ricava per semplice derivazione:

$$f_{\eta-a}(x|\eta \geq a) = f_{\eta}(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Questo comportamento è caratteristico delle v.a. esponenziali nel senso che non vi sono altre distribuzioni a.c., diverse da quella esponenziale, dotate di queste proprietà.  $\diamond$

**II.7.17 Esempio (Ago di Buffon<sup>3</sup>):** Un ago di lunghezza unitaria viene lanciato a caso su un piano sul quale sono tracciate delle linee parallele a distanza unitaria fra loro: qual è la probabilità che l'ago intersechi una di queste linee? Innanzitutto, data la regolarità della ripetizione delle linee sul tavolo, sarà sufficiente esaminare il problema con due sole linee facendo l'ipotesi che il centro dell'ago cada fra queste ultime. La posizione dell'ago può essere allora caratterizzata mediante due v.a.: la distanza  $\xi$  del suo punto di mezzo dalla linea sinistra e l'angolo  $\theta$  che esso forma con una perpendicolare alle linee parallele. Che l'ago sia lanciato *a caso* vuol dire semplicemente che le due v.a.  $\xi$  e  $\theta$  sono distribuite in maniera uniforme (vedi Esempio II.3.16) rispettivamente negli intervalli  $[0, 1]$  e  $[-\frac{\pi}{2}, \frac{\pi}{2}]$ , cioè hanno fdd

$$f_{\xi}(x) = \begin{cases} 1, & \text{se } x \in [0, 1], \\ 0, & \text{altrove,} \end{cases} \quad f_{\theta}(t) = \begin{cases} 1/\pi, & \text{se } t \in [-\frac{\pi}{2}, \frac{\pi}{2}], \\ 0, & \text{altrove.} \end{cases}$$

Supporremo, inoltre, che  $\xi$  e  $\theta$  siano indipendenti. Posto ora

$$B = \left\{ (x, t) : x \leq \frac{1}{2} \cos t \text{ oppure } x \geq 1 - \frac{1}{2} \cos t, \text{ con } -\frac{\pi}{2} \leq t \leq \frac{\pi}{2} \right\}$$

si ha che l'evento del quale ci interessa stimare la probabilità è

$$A = \{ \text{l'ago interseca una linea} \} = \{ \omega \in \Omega : (\xi, \theta) \in B \}.$$

Il calcolo può essere eseguito in diverse maniere equivalenti: noi sceglieremo quella che fa uso, in successione, di (I) della Proposizione II.7.9, del Teorema II.6.25, di (c) dell'Osservazione II.7.13 e dell'uniformità della distribuzione di  $\theta$ :

$$\begin{aligned} \mathbf{P}(A) &= \mathbf{E} I_A = \mathbf{E} [I_B(\xi, \theta)] = \mathbf{E} \left( \mathbf{E} [I_B(\xi, \theta) \mid \theta] \right) \\ &= \int_{\Omega} \mathbf{E} [I_B(\xi, \theta) \mid \theta] \mathbf{P}(d\omega) = \int_{-\pi/2}^{\pi/2} \mathbf{E} [I_B(\xi, \theta) \mid \theta = t] \mathbf{P}_{\theta}(dt) \\ &= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \mathbf{E} [I_B(\xi, t)] dt. \end{aligned}$$

---

<sup>3</sup> Georges-Louis Leclerc de Buffon (1707 - 1788), naturalista francese e precursore della teoria dell'evoluzione con le sue idee sulle modificazioni indotte sulle specie dall'ambiente, è soprattutto noto per la sua monumentale *Histoire Naturelle* (36 volumi pubblicati dal 1749 al 1789), ma va ricordato qui anche per alcuni suoi lavori giovanili presentati all'Académie des Sciences sul calcolo delle probabilità (*Sur le jeu de franc-carreau*, 1732).

Basterà ora ricordare che, a causa dell'uniformità della distribuzione di  $\xi$ , si ha

$$\mathbf{E} [I_B(\xi, t)] = \mathbf{P}(\{\xi \leq \frac{1}{2} \cos t\} \cup \{\xi \geq 1 - \frac{1}{2} \cos t\}) = \frac{1}{2} \cos t + \frac{1}{2} \cos t = \cos t,$$

per ottenere

$$\mathbf{P}(A) = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \cos t \, dt = \frac{2}{\pi}.$$

Sarà interessante osservare che questo risultato può essere (ed è stato) usato per dare una stima *sperimentale* di  $\pi$ : se lanciamo l'ago  $N$  volte e definiamo  $N$  v.a. i.i.d.  $\eta_k$  (con  $k = 1, \dots, N$ ) in modo che siano eguali ad 1 se l'ago interseca una linea al  $k$ -mo lancio e a 0 nel caso contrario, e se  $p$  è la probabilità che l'ago intersechi una linea in ogni singolo lancio, detta  $\nu_N = \eta_1 + \dots + \eta_N$  la v.a. che indica il numero di volte in cui l'ago interseca una linea su  $N$  lanci, la Legge dei Grandi Numeri (vedi Teorema I.7.5) ci dice che, comunque scelto  $\epsilon > 0$ , risulterà

$$\lim_N \mathbf{P} \left\{ \left| \frac{\nu_N}{N} - p \right| > \epsilon \right\} = 0.$$

Potremo allora dire che, con  $N$  abbastanza grande, la v.a.  $\nu_N/N$  approssimerà bene il numero  $p$ . Sapendo allora, dalla discussione precedente, che  $p = 2/\pi$ , una buona approssimazione del valore di  $\pi$  sarà data da un valore misurato della v.a.  $2N/\nu_N$  con  $N$  abbastanza grande. Questo metodo per stimare  $\pi$  è stato usato varie volte nel corso della storia<sup>4</sup> e rappresenta il primo caso noto di applicazioni delle regolarità statistiche alla soluzione di problemi di analisi numerica: un metodo divenuto successivamente conosciuto come *metodo di Monte Carlo* e del quale avremo modo di riparlare nel corso di queste lezioni.  $\diamond$

**II.7.18 Osservazione:** Come già rilevato in II.7.8 e II.7.12, le probabilità condizionate  $\mathbf{P}(A|\mathcal{G})$ ,  $\mathbf{P}(A|\eta)$  e  $\mathbf{P}(A|\eta = y)$  non possono essere immediatamente interpretate come vere misure di probabilità per tutti gli  $\omega \in \Omega$  e  $y \in \mathbf{R}$ . Infatti, se  $(\xi_n)_{n \in \mathbf{N}}$  è una successione di v.a. non negative dalla (d) di II.7.10 si ha che

$$\mathbf{E} \left( \sum_n \xi_n \mid \mathcal{G} \right) = \sum_n \mathbf{E}(\xi_n | \mathcal{G}), \quad \mathbf{P}\text{-q.o.}$$

In particolare ponendo  $\xi_n = I_{A_n}$ , con  $A_n$  disgiunti, si ricava che l'eguaglianza

$$\mathbf{P} \left( \bigcup_n A_n \mid \mathcal{G} \right) = \bigcup_n \mathbf{P}(A_n | \mathcal{G})$$

risulta vera solo  $\mathbf{P}$ -q.o., sicché  $\mathbf{P}(\cdot | \mathcal{G})(\omega)$  non risulta  $\sigma$ -additiva per ogni fissato  $\omega \in \Omega$ . Si potrebbe però supporre che l'insieme di punti  $\omega$  su cui ciò avviene sia

---

<sup>4</sup> La stima fu tentata per la prima volta nel 1850 dall'astronomo svizzero R. Wolf (1816 - 1893) che, lanciando l'ago 5.000 volte ottenne come risultato 3,1596.

di  $\mathbf{P}$ -misura nulla, ma anche questo si rivela, in generale, falso. Infatti, se per brevità indichiamo con  $A$  la generica successione  $(A_n)_{n \in \mathbf{N}}$ , e con  $\mathcal{N}(A)$  l'insieme delle  $\omega \in \Omega$  per le quali non si verifica la  $\sigma$ -additività per la successione  $A$ , l'insieme di punti  $\omega$  su cui  $\mathbf{P}(\cdot | \mathcal{G})(\omega)$  non può essere considerata una probabilità è

$$\mathcal{N} = \bigcup_A \mathcal{N}(A);$$

ma siccome l'insieme di tutte le successioni non è numerabile, pur essendo ogni  $\mathcal{N}(A)$  di misura nulla non potremo affermare che anche  $\mathcal{N}$  è di misura nulla; pertanto  $\mathbf{P}(\cdot | \mathcal{G})(\omega)$  risulta non essere una probabilità su un insieme di punti di  $\mathbf{P}$ -misura in generale non nulla. D'altra parte è anche vero che sarebbe estremamente utile riuscire a definire  $\mathbf{P}(\cdot | \mathcal{G})(\omega)$  in modo che risulti una vera probabilità per ogni  $\omega \in \Omega$ . Infatti fino a questo punto abbiamo definito il VdA condizionato solo nella maniera rigorosa (ma indiretta) della Definizione II.7.4, mentre sarebbe concettualmente suggestivo poter dare una definizione diretta (analoga a quella del VdA in II.6.5) mediante un integrale su una opportuna misura di probabilità condizionata in modo tale che risulti ad esempio

$$\mathbf{E}(\xi | \mathcal{G})(\omega) = \int_{\Omega} \xi(\omega') \mathbf{P}(d\omega' | \mathcal{G})(\omega), \quad \mathbf{P}\text{-q.o.}$$

ed è evidente che, per far questo, dovremo riuscire a definire  $\mathbf{P}(\cdot | \mathcal{G})(\omega)$  in modo che sia una vera misura di probabilità comunque scelto  $\omega \in \Omega$ .  $\circ$

**II.7.19 Definizione:** Diremo che  $\mathbf{P}_{|\mathcal{G}}(\omega; A)$  è una **versione regolare della probabilità condizionata** rispetto alla  $\sigma$ -algebra  $\mathcal{G}$  se

- (a)  $\mathbf{P}_{|\mathcal{G}}(\omega; \cdot)$  è una probabilità su  $\mathcal{F}$ ,  $\forall \omega \in \Omega$ ;
- (b)  $\mathbf{P}_{|\mathcal{G}}(\cdot; A)$  è una variante della probabilità condizionata  $\mathbf{P}(A | \mathcal{G})$  (vedi II.7.7) comunque scelto  $A \in \mathcal{F}$ .

In particolare, se  $\mathcal{G} = \mathcal{G}_Y$  è generata da un elemento aleatorio  $Y$  useremo la notazione abbreviata  $\mathbf{P}_{|Y}(\omega; A)$ , e se  $\mathcal{G} = \mathcal{G}_\eta$  la notazione  $\mathbf{P}_{|\eta}(\omega; A)$ .  $\triangle$

**II.7.20 Osservazione:** Sulla base della Definizione precedente è ora possibile estendere la nomenclatura delle versioni regolari alle DdP e alle FdD. Se  $X : (\Omega, \mathcal{F}) \rightarrow (\mathbf{E}, \mathcal{E})$  è un elemento aleatorio, diremo che  $\mathbf{P}_{X|\mathcal{G}}(\omega; B)$  è una **versione regolare della DdP condizionata** di  $X$  rispetto a  $\mathcal{G}$  se

- (a)  $\mathbf{P}_{X|\mathcal{G}}(\omega; \cdot)$  è una probabilità su  $\mathcal{E}$ ,  $\forall \omega \in \Omega$ ;
- (b)  $\mathbf{P}_{X|\mathcal{G}}(\cdot; B)$  è una variante della DdP condizionata  $\mathbf{P}_X(B | \mathcal{G}) = \mathbf{P}(X \in B | \mathcal{G})$ , comunque scelto  $B \in \mathcal{E}$ .

In particolare, se  $\xi : (\Omega, \mathcal{F}) \rightarrow (\mathbf{R}, \mathcal{B}(\mathbf{R}))$  è una v.a., indicheremo con  $\mathbf{P}_{\xi|\mathcal{G}}(\omega; B)$  la suddetta versione regolare della DdP condizionata. Quando la  $\sigma$ -algebra  $\mathcal{G} = \mathcal{G}_\eta$  è generata da una v.a.  $\eta$  è anche possibile, tenendo conto della definizione di  $\mathbf{P}(A | \eta = y)$  data in II.7.12, trasferire questi concetti nel linguaggio del condizionamento rispetto all'evento  $\{\eta = y\}$ : diremo allora che  $\mathbf{P}_{|\eta}(y; A)$  è una **versione regolare della probabilità condizionata rispetto a  $\eta = y$**  quando

- (a)  $\mathbf{P}_{|\eta}(y; \cdot)$  è una probabilità su  $\mathcal{F}$ ,  $\forall y \in \mathbf{R}$ ;
- (b)  $\mathbf{P}_{|\eta}(\cdot; A)$  è una variante della probabilità condizionata  $\mathbf{P}(A|\eta = y)$ , comunque scelto  $A \in \mathcal{F}$ .

Data poi una seconda v.a.  $\xi$ , diremo che  $\mathbf{P}_{\xi|\eta}(y; B)$  è una **versione regolare della DdP di  $\xi$  condizionata rispetto a  $\eta = y$**  quando

- (a)  $\mathbf{P}_{\xi|\eta}(y; \cdot)$  è una DdP su  $\mathbf{R}$ ,  $\forall y \in \mathbf{R}$ ;
- (b)  $\mathbf{P}_{\xi|\eta}(\cdot; B)$  è una variante della DdP condizionata  $\mathbf{P}_{\xi}(B|\eta = y)$ , comunque scelto  $B \in \mathcal{B}(\mathbf{R})$ .

A partire da queste definizioni potremo ora anche introdurre delle **versioni regolari della FdD condizionate**: così  $F_{\xi|\mathcal{G}}(\omega; x)$  sarà la FdD della DdP condizionata regolare  $\mathbf{P}_{\xi|\mathcal{G}}(\omega; B)$ , mentre  $F_{\xi|\eta}(y; x)$  sarà la FdD della corrispondente  $\mathbf{P}_{\xi|\eta}(y; B)$ . Se  $F_{\xi|\eta}(y; x)$  risulta anche a.c., indicheremo con  $f_{\xi|\eta}(y; x)$  la **fdd condizionata regolare** di  $\xi$  rispetto a  $\eta = y$ , sicché avremo

$$\mathbf{P}(\xi \in B|\eta = y) = \int_B f_{\xi|\eta}(y; x) dx, \quad \mathbf{P}_{\eta\text{-q.o.}}$$

e quindi, tenendo conto della discussione nell'Esempio II.7.15,

$$f_{\xi|\eta}(y; x) = \frac{f_{\xi\eta}(x, y)}{f_{\eta}(y)}, \quad \mathbf{P}_{\eta\text{-q.o.}}$$

Aggiungeremo infine che talora (come si può vedere dalla notazione introdotta nell'Esempio II.7.15) al posto di  $F_{\xi|\eta}(y; x)$  e  $f_{\xi|\eta}(y; x)$  vengono usati i simboli  $F_{\xi|\eta}(x|y)$  e  $f_{\xi|\eta}(x|y)$ ; la scelta della notazione più adeguata dipenderà dalle particolari espressioni in questione. ○

**II.7.21 Teorema:** Dato un elemento aleatorio  $X : (\Omega, \mathcal{F}) \rightarrow (E, \mathcal{E})$ , se  $(E, \mathcal{E})$  è uno spazio polacco (vedi Osservazione II.4.11) e se  $\mathcal{G} \subseteq \mathcal{F}$  è una sotto  $\sigma$ -algebra di  $\mathcal{F}$ , esiste sempre una versione regolare  $\mathbf{P}_{X|\mathcal{G}}(\omega; B)$  della DdP condizionata di  $X$  rispetto a  $\mathcal{G}$ .

**Dimostrazione:** Omessa (vedi **A.N. Shiriyayev: *Probability***, Springer, New York, 1984, p. 228). Questo teorema assicura la legittimità di quanto abbiamo fatto finora, almeno per una classe di problemi abbastanza ampia da comprendere tutti i casi di interesse pratico: quella delle DdP condizionate (rispetto a generiche  $\sigma$ -algre) di elementi aleatori con valori in spazi polacchi. □

**II.7.22 Proposizione:** Se  $\mathbf{P}_{|\mathcal{G}}(\omega; A)$  è una versione regolare della probabilità condizionata rispetto a  $\mathcal{G}$ , risulta

$$\mathbf{E}(\xi|\mathcal{G})(\omega) = \int_{\Omega} \xi(\omega') \mathbf{P}_{|\mathcal{G}}(\omega; d\omega'), \quad \mathbf{P}\text{-q.o.}$$

per ogni v.a.  $\xi$  integrabile.

**Dimostrazione:** Il risultato è sicuramente vero per gli indicatori  $\xi = I_A$  con  $A \in \mathcal{F}$  in quanto, dalla Definizione II.7.19, risulta **P**-q.o.

$$\mathbf{E}(I_A|\mathcal{G})(\omega) = \mathbf{P}(A|\mathcal{G})(\omega) = \mathbf{P}_{|\mathcal{G}}(\omega; A) = \int_{\Omega} I_A \mathbf{P}_{|\mathcal{G}}(\omega; d\omega').$$

Ne segue allora che la proposizione è vera per v.a. semplici; se poi  $\xi \geq 0$ , detta  $(\xi_n)_{n \in \mathbf{N}}$  una successione di v.a. semplici tale che  $\xi_n \uparrow \xi$ , la (b) di Proposizione II.7.10 ci garantisce che  $\mathbf{E}(\xi_n|\mathcal{G}) \uparrow \mathbf{E}(\xi|\mathcal{G})$ , **P**-q.o., e siccome  $\mathbf{P}_{|\mathcal{G}}(\omega; \cdot)$  è una probabilità  $\forall \omega \in \Omega$ , il Teorema II.6.10 ci assicura che **P**-q.o. risulta anche

$$\mathbf{E}(\xi|\mathcal{G}) = \lim_n \mathbf{E}(\xi_n|\mathcal{G}) = \lim_n \int_{\Omega} \xi_n(\omega') \mathbf{P}_{|\mathcal{G}}(\omega; d\omega') = \int_{\Omega} \xi(\omega') \mathbf{P}_{|\mathcal{G}}(\omega; d\omega').$$

Il caso generale si riconduce a questo tramite la rappresentazione  $\xi = \xi^+ - \xi^-$ .  $\square$

**II.7.23 Osservazione:** Dobbiamo osservare che spesso si evita di distinguere fra le probabilità condizionate e le loro versioni regolari dato che l'esistenza di queste ultime, nei casi di interesse pratico, è garantita dal Teorema II.7.21. Nel seguito quindi utilizzeremo indifferentemente sia i simboli  $\mathbf{P}(A|\mathcal{G})$ ,  $\mathbf{P}(A|Y)$ ,  $\mathbf{P}(A|\eta)$ ,... che gli analoghi  $\mathbf{P}_{|\mathcal{G}}(\omega; A)$ ,  $\mathbf{P}_{|Y}(\omega; A)$ ,  $\mathbf{P}_{|\eta}(\omega; A)$ ... aggiungendo (qualora serva) la precisazione che si tratta di versioni regolari delle probabilità condizionate: la scelta della migliore grafia sarà dettata dal tipo di espressione in questione. In questo modo, ad esempio, la tesi della Proposizione II.7.22 si può anche riscrivere come

$$\mathbf{E}(\xi|\mathcal{G})(\omega) = \int_{\Omega} \xi(\omega') \mathbf{P}(d\omega'|\mathcal{G})(\omega), \quad \mathbf{P}\text{-q.o.}$$

coerentemente con quanto richiesto alla fine dell'Osservazione II.7.18. Va inoltre notato che, facendo uso del Teorema II.6.25 applicato alle versioni regolari delle probabilità condizionate, dalla Proposizione II.7.22 si ricava anche che

$$\mathbf{E}[g(\xi)|\mathcal{G})(\omega) = \int_{\Omega} g[\xi(\omega')] \mathbf{P}(d\omega'|\mathcal{G})(\omega) = \int_{\mathbf{R}} g(x) \mathbf{P}_{\xi}(dx|\mathcal{G})(\omega), \quad \mathbf{P}\text{-q.o.}$$

dove  $g(x)$  è una generica funzione di Borel. In particolare, date due v.a.  $\xi$  e  $\eta$  dotate di fdd congiunta  $f_{\xi\eta}(x, y)$ , se  $g(x)$  è una funzione di Borel tale che  $\mathbf{E}|g(\xi)| < +\infty$ , risulta

$$\mathbf{E}[g(\xi)|\eta = y] = \int_{-\infty}^{+\infty} g(x) f_{\xi|\eta}(x|y) dx$$

dove  $f_{\xi|\eta}(x|y)$  è la fdd condizionata discussa in Osservazione II.7.20. A questo punto siamo anche in grado di mostrare come la notazione introdotta in questo capitolo consente di esprimere l'enunciato del Teorema II.6.31 in termini di (versioni regolari di) probabilità condizionate, come già accennato nell'Osservazione II.6.32. Se  $\xi$  e  $\eta$  sono due v.a. con DdP probabilità congiunta  $\mathbf{P}_{\xi\eta}$  e se  $\varphi(x, y)$  è una funzione di Borel, sappiamo che

$$\mathbf{E}[\varphi(\xi, \eta)] = \int_{\mathbf{R}^2} \varphi(x, y) \mathbf{P}_{\xi\eta}(dx, dy);$$

dalla (I) di II.7.9, da II.6.25 e da (c) di II.7.13, d'altra parte, otteniamo

$$\begin{aligned} \mathbf{E} [\varphi(\xi, \eta)] &= \mathbf{E} [\mathbf{E} (\varphi(\xi, \eta) \mid \eta)] = \int_{\mathbf{R}} \mathbf{E} (\varphi(\xi, \eta) \mid \eta = y) \mathbf{P}_{\eta}(dy) \\ &= \int_{\mathbf{R}} \mathbf{E} (\varphi(\xi, y) \mid \eta = y) \mathbf{P}_{\eta}(dy) \\ &= \int_{\mathbf{R}} \left( \int_{\mathbf{R}} \varphi(x, y) \mathbf{P}_{\xi|\eta}(y; dx) \right) \mathbf{P}_{\eta}(dy), \end{aligned}$$

e quindi

$$\int_{\mathbf{R}^2} \varphi(x, y) \mathbf{P}_{\xi\eta}(dx, dy) = \int_{\mathbf{R}} \left( \int_{\mathbf{R}} \varphi(x, y) \mathbf{P}_{\xi|\eta}(y; dx) \right) \mathbf{P}_{\eta}(dy),$$

che è il risultato del Teorema II.6.31 nel caso di  $\Omega = \mathbf{R}^2 = \mathbf{R} \times \mathbf{R}$  con  $\mu_1 = \mathbf{P}_{\eta}$  e  $\mu_2 = \mathbf{P}_{\xi|\eta}$ . ○





## II.8 Variabili aleatorie (II Parte)

**II.8.1 Definizione:** Data una v.a.  $\xi$  su  $(\Omega, \mathcal{F}, \mathbf{P})$  con varianza  $\mathbf{V}\xi$ , chiameremo **Deviazione standard** la quantità  $\sigma = +\sqrt{\mathbf{V}\xi}$ . Inoltre se  $\xi$  è dotata di VdA condizionato  $\mathbf{E}(\xi|\mathcal{G})$ , è anche possibile definire la **Varianza condizionata rispetto alla  $\sigma$ -algebra  $\mathcal{G}$**  come

$$\mathbf{V}(\xi|\mathcal{G}) = \mathbf{E}[(\xi - \mathbf{E}(\xi|\mathcal{G}))^2 | \mathcal{G}].$$

Se  $\xi$  e  $\eta$  sono due v.a. su  $(\Omega, \mathcal{F}, \mathbf{P})$  chiameremo **Covarianza (Cov)** la quantità

$$\mathbf{cov}(\xi, \eta) = \mathbf{E}[(\xi - \mathbf{E}\xi)(\eta - \mathbf{E}\eta)] = \mathbf{E}(\xi\eta) - \mathbf{E}\xi \cdot \mathbf{E}\eta,$$

e, se  $\mathbf{V}\xi > 0$  e  $\mathbf{V}\eta > 0$ , chiameremo **Coefficiente di correlazione** la quantità

$$\rho(\xi, \eta) = \frac{\mathbf{cov}(\xi, \eta)}{\sqrt{\mathbf{V}\xi \cdot \mathbf{V}\eta}}.$$

Quando  $\mathbf{cov}(\xi, \eta) = 0$ , ovvero  $\rho(\xi, \eta) = 0$ , diremo che  $\xi$  e  $\eta$  sono due **v.a. non correlate**. Infine, se  $X = (\xi_1, \dots, \xi_n)$  è un vett.a. su  $(\Omega, \mathcal{F}, \mathbf{P})$  chiameremo **Matrice delle covarianze** di  $X$  la matrice  $n \times n$ ,  $\mathbf{R} = \|r_{ij}\|$  i cui elementi sono  $r_{ij} = \mathbf{cov}(\xi_i, \xi_j)$ . △

**II.8.2 Proposizione:** Se  $\xi$  e  $\eta$  sono v.a. non correlate si ha  $\mathbf{V}(\xi+\eta) = \mathbf{V}\xi + \mathbf{V}\eta$ .

**Dimostrazione:** La dimostrazione è identica a quella dell'analogia Proposizione I.6.22 enunciata nella prima parte di queste lezioni. □

**II.8.3 Osservazione:** Si vede facilmente che gli elementi diagonali della matrice delle covarianze  $\mathbf{R}$  coincidono con le varianze delle singole componenti di  $X$ :

$$r_{ii} = \mathbf{cov}(\xi_i, \xi_i) = \mathbf{E}(\xi_i - \mathbf{E}\xi_i)^2 = \mathbf{V}\xi_i.$$

Inoltre è chiaro che la matrice  $\mathbf{R}$  è simmetrica ( $r_{ij} = r_{ji}$ ) e si verifica subito che è anche definita non negativa:

$$\sum_{i,j=1}^n r_{ij} \lambda_i \lambda_j \geq 0; \quad \forall \lambda_i \in \mathbf{R}, \quad i = 1, \dots, n;$$

infatti si ha

$$\sum_{i,j=1}^n r_{ij} \lambda_i \lambda_j = \sum_{i,j=1}^n \lambda_i \lambda_j \mathbf{E}[(\xi_i - \mathbf{E}\xi_i)(\xi_j - \mathbf{E}\xi_j)] = \mathbf{E} \left[ \sum_{i=1}^n \lambda_i (\xi_i - \mathbf{E}\xi_i) \right]^2 \geq 0.$$

È interessante notare che queste proprietà sono caratteristiche delle matrici delle covarianze nel senso chiarito dalla proposizione seguente.  $\circ$

**II.8.4 Proposizione:** Condizione necessaria e sufficiente affinché una matrice  $n \times n$ ,  $R$  sia matrice delle covarianze di un vett.a.  $X = (\xi_1, \dots, \xi_n)$  è che la matrice sia simmetrica e definita non negativa, ovvero, equivalentemente, che esista una matrice  $n \times n$ ,  $C$  tale che  $R = CC^T$ , dove  $C^T$  è la trasposta della matrice  $C$ .

**Dimostrazione:** Abbiamo già visto nell'Osservazione II.8.3 che, se  $R$  è una matrice di covarianze, essa è anche simmetrica e definita non negativa. Qui mostriamo solo che, in tal caso, esiste sempre una matrice  $n \times n$ ,  $C$  tale che  $R = CC^T$ : infatti è noto dalla teoria delle forme quadratiche<sup>1</sup> che, se  $R$  è simmetrica e definita non negativa esiste sempre una matrice *ortogonale*  $O$  (tale cioè che  $OO^T = I$ , dove  $I$  è la matrice identità  $n \times n$ ) tale che

$$D = O^T R O = \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{pmatrix}; \quad d_i \geq 0, \quad i = 1, \dots, n,$$

sia una matrice diagonale con elementi diagonali non negativi. Se ora indichiamo con  $B$  la matrice diagonale i cui elementi diagonali sono  $b_i = +\sqrt{d_i}$ , con  $i = 1, \dots, n$ , in modo che  $D = B^2 = BB^T$ , e se poniamo  $C = OB$  otteniamo il risultato richiesto:

$$R = O D O^T = (OB)(B^T O^T) = CC^T.$$

Viceversa, se supponiamo che la matrice data sia della forma  $R = CC^T$  con  $C = \|c_{ij}\|$ , si vede immediatamente che essa è simmetrica:

$$R^T = (CC^T)^T = CC^T = R,$$

e definita non negativa in quanto, essendo

$$r_{ij} = (CC^T)_{ij} = \sum_{k=1}^n c_{ik}(C^T)_{kj} = \sum_{k=1}^n c_{ik}c_{jk},$$

risulta immediatamente

$$\sum_{i,j=1}^n r_{ij} \lambda_i \lambda_j = \sum_{i,j=1}^n \sum_{k=1}^n c_{ik}c_{jk} \lambda_i \lambda_j = \sum_{k=1}^n \left( \sum_{j=1}^n \lambda_j c_{jk} \right)^2 \geq 0.$$

---

<sup>1</sup> Vedi ad esempio **V.I. Smirnov**: *Corso di Matematica Superiore*, vol. III (parte Ia); Editori Riuniti, Roma, 1978, p. 118.

Resta da mostrare, quindi, che matrici  $R$  di questa forma possono sempre essere considerate come matrici delle covarianze di un qualche vett.a. Sia infatti  $Y = (\eta_1, \dots, \eta_n)$  un vett.a. con componenti indipendenti tali che  $\mathbf{E} \eta_j = 0$  e  $\mathbf{V} \eta_j = 1$  per  $j = 1, \dots, n$  (una simile scelta è sempre possibile, come sarà mostrato più oltre con alcuni esempi); in tal caso si vede subito che

$$\mathbf{cov}(\eta_i, \eta_j) = \mathbf{E}(\eta_i \eta_j) = \delta_{ij} = \begin{cases} 1, & \text{se } i = j; \\ 0, & \text{se } i \neq j. \end{cases}$$

È facile far vedere ora che il vett.a.  $X = CY$  con componenti  $\xi_j = \sum_k c_{jk} \eta_k$  ha come matrice delle covarianze proprio  $R = CC^T$ ; infatti, siccome

$$\mathbf{E} \xi_j = \sum_{k=1}^n c_{jk} \mathbf{E} \eta_k = 0,$$

si ha

$$\begin{aligned} \mathbf{cov}(\xi_i, \xi_j) &= \mathbf{E}(\xi_i \xi_j) = \sum_{k,l=1}^n c_{ik} c_{jl} \mathbf{E}(\eta_k \eta_l) = \sum_{k,l=1}^n c_{ik} c_{jl} \delta_{kl} \\ &= \sum_{k=1}^n c_{ik} c_{jk} = (CC^T)_{ij} = r_{ij}, \end{aligned}$$

che definitivamente prova la nostra affermazione. □

**II.8.5 Esempio:** Se  $\xi$  è una **v.a. Gaussiana (Normale)**  $\mathcal{N}(m, \sigma^2)$

$$f_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-m)^2/2\sigma^2}$$

con  $m \in \mathbf{R}$  e  $\sigma > 0$  (vedi anche Esempio II.3.16), si può ora mostrare con un calcolo diretto che i due parametri  $m$  e  $\sigma^2$  hanno una immediata interpretazione in termini di VdA e Var (vedi anche Esempio II.4.8):

$$m = \mathbf{E} \xi, \quad \sigma^2 = \mathbf{V} \xi.$$

Per verificarlo basterà osservare che, essendo

$$\int_{-\infty}^{\infty} e^{-t^2/2} dt = \sqrt{2\pi}, \quad \int_{-\infty}^{\infty} t e^{-t^2/2} dt = 0,$$

posto  $t = (x - m)/\sigma$  si ha

$$\begin{aligned} \mathbf{E} \xi &= \int_{-\infty}^{\infty} x f_\xi(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-(x-m)^2/2\sigma^2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma t + m) e^{-t^2/2} dt = \frac{m}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2/2} dt = m. \end{aligned}$$

Analogamente, utilizzando il fatto che integrando per parti risulta

$$\int_{-\infty}^{\infty} t^2 e^{-t^2/2} dt = \sqrt{2\pi},$$

si ha che

$$\begin{aligned} \mathbf{E} \xi^2 &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{(x-m)^2/2\sigma^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma t + m)^2 e^{-t^2/2} dt \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-t^2/2} dt + \frac{2m\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{-t^2/2} dt + \frac{m^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2/2} dt \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-t^2/2} dt + m^2 = \sigma^2 + m^2, \end{aligned}$$

e quindi

$$\mathbf{V} \xi = \mathbf{E} \xi^2 - (\mathbf{E} \xi)^2 = (\sigma^2 + m^2) - m^2 = \sigma^2.$$

Se poi  $X = (\xi, \eta)$  è un **vett.a. Gaussiano (Normale) bivariato**  $\mathcal{N}(m, \mathbf{R})$  (vedi anche Esempio II.4.8) sappiamo che la sua fdd congiunta ha la forma

$$\begin{aligned} f_X(x, y) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \cdot \\ &\exp \left\{ -\frac{1}{2(1-r^2)} \left[ \frac{(x-m_1)^2}{\sigma_1^2} - 2r \frac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2} \right] \right\} \end{aligned}$$

ed è caratterizzata da cinque parametri:  $m_1, m_2, \sigma_1, \sigma_2, r$ , con  $m_1, m_2 \in \mathbf{R}$ ,  $\sigma_1 > 0, \sigma_2 > 0$  e  $|r| \leq 1$ . Le fdd marginali, come si vede da un'applicazione delle relazioni di II.5.31, sono

$$\begin{aligned} f_\xi(x) &= \int_{-\infty}^{\infty} f_X(x, y) dy = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-(x-m_1)^2/2\sigma_1^2}, \\ f_\eta(y) &= \int_{-\infty}^{\infty} f_X(x, y) dx = \frac{1}{\sigma_2\sqrt{2\pi}} e^{-(y-m_2)^2/2\sigma_2^2}, \end{aligned}$$

da cui si deduce che le componenti di un vett.a. normale sono a loro volta v.a. normali. Un calcolo diretto mostra inoltre che l'interpretazione dei parametri che compaiono in  $f_X$  è la seguente:

$$\begin{aligned} m_1 &= \mathbf{E} \xi, & m_2 &= \mathbf{E} \eta, \\ \sigma_1^2 &= \mathbf{V} \xi, & \sigma_2^2 &= \mathbf{V} \eta, & r &= \rho(\xi, \eta). \end{aligned}$$

Si vede pertanto che il vettore delle medie  $m$  e la matrice delle covarianze  $\mathbf{R}$  sono, in termini di VdA, Var e Cov

$$m = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} = \begin{pmatrix} \mathbf{E} \xi \\ \mathbf{E} \eta \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2 r \\ \sigma_1\sigma_2 r & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \mathbf{V} \xi & \mathbf{cov}(\xi, \eta) \\ \mathbf{cov}(\xi, \eta) & \mathbf{V} \eta \end{pmatrix}.$$

Dal Teorema II.6.15 sappiamo che se due v.a. (anche non gaussiane)  $\xi$  e  $\eta$  sono indipendenti si ha che  $\mathbf{E}(\xi\eta) = \mathbf{E}\xi \cdot \mathbf{E}\eta$ ; dalla Definizione II.8.1 discende allora che  $\mathbf{cov}(\xi, \eta) = \rho(\xi, \eta) = 0$ , cioè le due v.a. sono anche non correlate. È stato però anche notato che il viceversa non è, in generale, vero (vedi Osservazione I.6.23 ed Esempio I.6.24): ci sono infatti v.a. non correlate che non risultano indipendenti. È allora interessante osservare che se le due v.a. sono distribuite in modo gaussiano bivariato (come nel presente Esempio), e se sono non correlate, esse risultano anche indipendenti. Naturalmente tale proprietà si estende anche ai vett.a. gaussiani multivariati con più di due componenti. In altre parole: per vett.a.  $X$  distribuiti in maniera gaussiana multivariata (e solo per questi) i concetti di indipendenza e di non correlazione delle componenti risultano equivalenti. Che ciò sia vero si vede, poi, dal fatto che se  $\xi$  e  $\eta$  sono non correlate si ha  $r = \rho(\xi, \eta) = 0$ ; in tal caso la fdd congiunta si riduce a

$$f_X(x, y) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-(x-m_1)^2/\sigma_1^2} e^{-(y-m_2)^2/\sigma_2^2},$$

e si vede subito che  $f_X(x, y) = f_\xi(x) \cdot f_\eta(y)$ , sicché, in base al Teorema II.5.41,  $\xi$  e  $\eta$  risultano anche indipendenti.  $\diamond$

**II.8.6 Osservazione:** Date due v.a.  $\xi$  e  $\eta$  può capitare che solo una di questa (ad esempio  $\xi$ ) sia accessibile all'osservazione: se fra le due v.a. c'è una relazione statistica è presumibile che delle misure eseguite sulla v.a.  $\xi$  diano delle informazioni anche sulla v.a. non osservabile  $\eta$ . Si pone allora il problema di dare una **stima** di  $\eta$  tramite  $\xi$ : in pratica ciò viene fatto considerando le v.a. della forma  $f(\xi)$ : ognuna di queste, al variare di  $f(\cdot)$ , prende il nome di **stimatore** di  $\eta$ . È necessario inoltre fissare un criterio di scelta fra tutti gli stimatori di  $\eta$ : usualmente si cerca di determinare il **miglior stimatore in media quadratica** (in m.q.) nel senso che, definito l'**errore quadratico medio** (e.q.m.) che si commette stimando  $\eta$  mediante  $f(\xi)$  come

$$\Delta[f] = \mathbf{E} [\eta - f(\xi)]^2,$$

si elegge a stimatore la v.a.  $f^*(\xi) = \eta^*$  che rende minimo tale e.q.m.:

$$\mathbf{E} [\eta - f^*(\xi)]^2 = \inf_f \mathbf{E} [\eta - f(\xi)]^2;$$

l'errore minimo commesso in questo caso è poi

$$\Delta^* = \mathbf{E} [\eta - f^*(\xi)]^2.$$

La determinazione di tale stimatore ottimale, cioè della funzione di Borel  $f^*$  che rende minimo l'e.q.m., è un problema variazionale che non è sempre di facile soluzione. Inizieremo quindi ad esaminare i casi in cui la forma delle funzioni  $f(\cdot)$  è esplicitamente limitata in qualche maniera: in questo modo in generale avremo non lo stimatore ottimale, ma il miglior stimatore possibile nel dato sottinsieme di funzioni di Borel. Come primo caso poniamoci, allora, il problema

di stimare  $\eta$  mediante una costante  $a$ . In questo caso estremamente semplice è implicito che, essendo lo stimatore  $f(\xi)$  una costante, non sarà neanche necessario misurare  $\xi$  per dare una stima di  $\eta$  e l'e.q.m. sarà

$$\Delta(a) = \mathbf{E}(\eta - a)^2 = \mathbf{E}\eta^2 - 2a\mathbf{E}\eta + a^2,$$

per cui, dalla condizione  $\Delta'(a) = 2(a - \mathbf{E}\eta) = 0$ , si ricava subito che il valore ottimale di  $a$  è proprio  $a^* = \mathbf{E}\eta$ : come ci si poteva attendere anche intuitivamente, la miglior stima in m.q. di  $\eta$  mediante un solo numero è proprio il suo VdA. Analogamente risulta familiare anche la forma dell'e.q.m. minimo:

$$\Delta^* = \Delta(a^*) = \mathbf{E}(\eta - a^*)^2 = \mathbf{E}(\eta - \mathbf{E}\eta)^2 = \mathbf{V}\eta.$$

In questo primo esempio, comunque, si nota subito che solo in un senso molto banale possiamo dire di aver eseguito una stima di  $\eta$  mediante  $\xi$ , dal momento che la v.a.  $\xi$  si riduce ad una costante la cui misura è, quindi, irrilevante. L'esempio successivo, pertanto, prenderà in considerazione il problema della **stima lineare in media quadratica**: in questo caso si determina il miglior stimatore esaminando le funzioni di Borel lineari della forma  $l(x) = a + bx$ . Otterremo così, avendo allargato il campo delle funzioni esaminate, una stima migliore della precedente pur non trattandosi ancora della miglior stima in m.q. in senso assoluto. In questo caso l'e.q.m. è

$$\Delta(a, b) = \mathbf{E}[\eta - (a + b\xi)]^2$$

e quindi, per determinare i valori che rendono minimo  $\Delta$ , dovremo risolvere il seguente sistema di equazioni lineari nelle incognite  $a$  e  $b$ :

$$\begin{cases} \frac{\partial \Delta}{\partial a} = -2\mathbf{E}[\eta - (a + b\xi)] = 0 \\ \frac{\partial \Delta}{\partial b} = -2\mathbf{E}[\eta\xi - (a + b\xi)\xi] = 0 \end{cases}$$

dal quale si ottiene facilmente

$$a^* = \mathbf{E}\eta - \frac{\mathbf{cov}(\xi, \eta)}{\mathbf{V}\xi} \mathbf{E}\xi, \quad b^* = \frac{\mathbf{cov}(\xi, \eta)}{\mathbf{V}\xi},$$

sicché il miglior stimatore lineare in m.q. sarà

$$l^*(\xi) = \mathbf{E}\eta + \frac{\mathbf{cov}(\xi, \eta)}{\mathbf{V}\xi} (\xi - \mathbf{E}\xi).$$

L'e.q.m. minimo commesso in questo caso è

$$\Delta^* = \Delta(a^*, b^*) = \mathbf{E}[\eta - l^*(\xi)]^2 = \mathbf{V}\eta - \frac{\mathbf{cov}^2(\xi, \eta)}{\mathbf{V}\xi} = [1 - \rho^2(\xi, \eta)] \mathbf{V}\eta.$$

È interessante osservare che, essendo  $\rho^2(\xi, \eta) \leq 1$ , il valore dell'e.q.m. minimo è tanto più piccolo quanto più il valore assoluto di  $\rho(\xi, \eta)$  si avvicina ad 1; cioè: la stima di  $\eta$  mediante  $\xi$  (anche nel semplice caso lineare) è tanto migliore quanto più  $\xi$  ed  $\eta$  risultano correlate. Viceversa, se  $\xi$  e  $\eta$  fossero non correlate (in particolare se fossero indipendenti), cioè se  $\mathbf{cov}(\xi, \eta) = \rho(\xi, \eta) = 0$ , la stima lineare ricadrebbe immediatamente nel caso della stima mediante una costante con  $a^* = \mathbf{E}\eta$ ,  $b^* = 0$  e  $\Delta^* = \mathbf{V}\eta$ . In altri termini: se  $\xi$  e  $\eta$  non sono correlate, un'osservazione di  $\xi$  non aggiunge nulla alla nostra conoscenza di  $\eta$ . Naturalmente l'esplorazione del problema della miglior stima in m.q. potrebbe proseguire considerando progressivamente delle funzioni  $f$  di tipo sempre più complicato (quadratiche, polinomiali,...) e caratterizzate da un numero sempre maggiore di parametri: sebbene questa sia una via che in pratica è spesso seguita (soprattutto da quando l'accresciuta potenza del calcolo automatico ha reso possibile seguirla in tempi ragionevoli), è importante sottolineare a questo punto che il nostro problema ha comunque una soluzione esatta anche nel caso più generale come espresso dall'enunciato del successivo Teorema, benché poi, in pratica, le difficoltà di calcolo siano solo spostate e la soluzione indicata abbia spesso solo un valore teorico e formale.  $\circ$

**II.8.7 Teorema:** Se  $\mathbf{E}\eta^2 < +\infty$ , il miglior stimatore in m.q. è  $\mathbf{E}(\eta \mid \xi)$  ed è quindi definito dalla funzione

$$f^*(x) = \mathbf{E}(\eta \mid \xi = x)$$

che prende il nome di **curva di regressione di  $\eta$  su  $\xi$** .

**Dimostrazione:** Se  $f(\xi)$  è il generico stimatore e  $f^*(\xi) = \mathbf{E}(\eta \mid \xi)$ , si ha:

$$\begin{aligned} \mathbf{E}[\eta - f(\xi)]^2 &= \mathbf{E}[(\eta - f^*(\xi)) + (f^*(\xi) - f(\xi))]^2 \\ &= \mathbf{E}[\eta - f^*(\xi)]^2 + \mathbf{E}[f^*(\xi) - f(\xi)]^2 + 2\mathbf{E}[(\eta - f^*(\xi))(f^*(\xi) - f(\xi))]. \end{aligned}$$

Da (I) di II.7.9 e (K) di II.7.11 si ha però che

$$\begin{aligned} \mathbf{E}[(\eta - f^*(\xi))(f^*(\xi) - f(\xi))] &= \mathbf{E}\left(\mathbf{E}[(\eta - f^*(\xi))(f^*(\xi) - f(\xi)) \mid \xi]\right) \\ &= \mathbf{E}\left((f^*(\xi) - f(\xi))\mathbf{E}[(\eta - f^*(\xi)) \mid \xi]\right) = 0, \end{aligned}$$

dato che  $\mathbf{E}[(\eta - f^*(\xi)) \mid \xi] = \mathbf{E}(\eta \mid \xi) - \mathbf{E}(f^*(\xi) \mid \xi) = \mathbf{E}(\eta \mid \xi) - f^*(\xi) = 0$ . Siccome inoltre, ovviamente,  $\mathbf{E}[f^*(\xi) - f(\xi)]^2 \geq 0$ , otteniamo in definitiva

$$\mathbf{E}[\eta - f(\xi)]^2 \geq \mathbf{E}[\eta - f^*(\xi)]^2$$

comunque scelta la funzione di Borel  $f$ .  $\square$

**II.8.8 Osservazione:** Il termine *regressione* ha la sua origine negli studi condotti da F. Galton<sup>2</sup> sulla relazione che intercorre fra le altezze dei genitori e dei figli in una data popolazione. I risultati delle sue ricerche possono essere espressi in termini di VdA condizionati: se le v.a.  $\xi$  e  $\eta$  rappresentano rispettivamente le altezze dei genitori e dei figli, e supponendo che esse abbiano lo stesso VdA e la stessa varianza (supponendo cioè che queste quantità non siano variate nel corso di una generazione),  $\mathbf{E}\xi = \mathbf{E}\eta = m$ ,  $\mathbf{V}\xi = \mathbf{V}\eta = \sigma^2$ , la Legge di Galton afferma che il VdA condizionato  $\mathbf{E}(\eta | \xi = x)$  risulta più piccolo (o rispettivamente più grande) di  $x$  se  $x \geq m$  (rispettivamente se  $x \leq m$ ). In pratica: se i genitori sono più alti della media i figli hanno la tendenza ad essere più bassi dei genitori; se, viceversa, i genitori sono più bassi della media i figli tendono ad essere più alti. In ambedue i casi potremo dire che l'altezza dei figli *regredisce* verso il VdA dell'altezza della popolazione. Il nome di *regressione* è da allora rimasto associato alla funzione  $f(x) = \mathbf{E}(\eta | \xi = x)$ .

Abbiamo già notato che, sebbene la soluzione del nostro problema data dal precedente Teorema sia abbastanza semplice, il suo calcolo effettivo può risultare complicato, soprattutto se si tiene conto del fatto che il problema della miglior stima in m.q. può essere posto anche nel caso in cui alla v.a.  $\xi$  si sostituisce un elemento aleatorio  $X$  con valori in  $(E, \mathcal{E})$ : in tal caso, infatti, dovremo determinare non una funzione  $f(x)$  di una sola variabile, ma una generica funzione misurabile  $f : (E, \mathcal{E}) \rightarrow (\mathbf{R}, \mathcal{B}(\mathbf{R}))$ . In particolare, se  $X = (\xi_1, \dots, \xi_n)$  è un vett.a., lo stimatore ottimale ha la forma  $\mathbf{E}(\eta | X) = \mathbf{E}(\eta | \xi_1, \dots, \xi_n)$  e la superficie di regressione è la funzione di  $n$  variabili  $f^*(x_1, \dots, x_n) = \mathbf{E}(\eta | \xi_1 = x_1, \dots, \xi_n = x_n)$  la cui determinazione può essere piuttosto complessa. Per questo motivo, come vedremo meglio più oltre, ci si accontenta spesso di determinare la miglior stima *lineare* in m.q., cioè ci si limita a ricercare il miglior stimatore di  $\eta$  fra le combinazioni lineari delle  $\xi_j$ . Sorge a questo punto in modo naturale la domanda seguente: fermo restando che in generale la stima lineare di  $\eta$  mediante  $\xi_1, \dots, \xi_n$  non è la miglior stima possibile in m.q., esistono casi particolari in cui la miglior stima in m.q. è data proprio dalla stima lineare? In altri termini: ci sono casi in cui la curva di regressione diviene una retta in modo che si possa parlare di **regressione lineare**? La risposta (affermativa) a tale domanda è contenuta nella successiva proposizione che, ancora una volta, mette in rilievo la particolarità e l'importanza di vett.a. che siano distribuiti in modo Gaussiano (normale) multivariato.  $\bigcirc$

---

<sup>2</sup> Sir Francis Galton (1822-1911) nacque a Birmingham in un ambiente intellettuale estremamente favorevole agli studi scientifici: suo nonno era membro della Royal Society e suo cugino era niente meno che Charles Darwin. Egli si dimostrò presto uno studioso geniale e polivalente: geografo, esploratore, meteorologo, genetista e psicologo, cominciò ad interessarsi di statistica verso il 1865 con lo scopo di descrivere, e possibilmente quantificare, il comportamento umano e la sua evoluzione, come mostrato dall'esempio descritto nel testo. Per questo motivo Galton è considerato come uno dei padri della biometria e dell'eugenetica. Fu molto legato, verso la fine della sua vita, ad un altro grande statistico, Karl Pearson, e con lui collaborò alla fondazione della rivista *Biometrika* (1901).



**II.8.9 Proposizione:** Se  $X = (\xi, \eta)$  è un vett.a. Gaussiano bivariato con  $\mathbf{V}\xi > 0$ , la miglior stima in m.q. di  $\eta$  mediante  $\xi$  coincide con la miglior stima lineare e precisamente risulta

$$f^*(\xi) = \mathbf{E}(\eta | \xi) = \mathbf{E}\eta + \frac{\mathbf{cov}(\xi, \eta)}{\mathbf{V}\xi} (\xi - \mathbf{E}\xi),$$

$$\Delta^* = \mathbf{E}[\eta - \mathbf{E}(\eta | \xi)]^2 = [1 - \rho^2(\xi, \eta)] \mathbf{V}\eta.$$

In tal caso si parla di **regressione lineare** di  $\eta$  su  $\xi$ .

**Dimostrazione:** Tenendo conto delle relazioni contenute in II.8.5 e II.7.15 si ottiene facilmente che

$$f_{\eta|\xi}(y|x) = \frac{1}{\sigma_2 \sqrt{2\pi(1-r^2)}} e^{-(y-m(x))^2/2\sigma_2^2(1-r^2)}$$

dove abbiamo posto, per semplicità,  $m(x) = m_2 + \sigma_2 r(x - m_1)/\sigma_1$ . Pertanto, ricordando le formule dell'Osservazione II.7.23 e il fatto che, per ogni  $x$  fissato,  $f_{\eta|\xi}(y|x)$  è una densità Gaussiana in  $y$  con VdA  $m(x)$  e Var  $\sigma_2^2(1-r^2)$ , otteniamo facilmente che

$$\mathbf{E}(\eta | \xi = x) = \int_{-\infty}^{\infty} y f_{\eta|\xi}(y|x) dy = m(x)$$

$$\mathbf{V}(\eta | \xi = x) = \mathbf{E}\left[(\eta - \mathbf{E}(\eta | \xi = x))^2 \mid \xi = x\right] = \int_{-\infty}^{\infty} (y - m(x))^2 f_{\eta|\xi}(y|x) dy$$

$$= \sigma_2^2(1-r^2).$$

Ne segue quindi che la curva di regressione è lineare e che la miglior stima in m.q. ha la forma

$$\mathbf{E}(\eta | \xi) = m(\xi) = \mathbf{E}\eta + \frac{\mathbf{cov}(\xi, \eta)}{\mathbf{V}\xi} (\xi - \mathbf{E}\xi).$$

Quanto all'errore  $\Delta^*$  che si commette con questa stima, usando successivamente (I) di II.7.9, (6) di Osservazione II.6.26 e (c) di II.7.13 si ottiene

$$\begin{aligned} \Delta^* &= \mathbf{E}(\eta - \mathbf{E}(\eta | \xi))^2 = \mathbf{E}\left(\mathbf{E}[(\eta - \mathbf{E}(\eta | \xi))^2 \mid \xi]\right) \\ &= \int_{-\infty}^{\infty} \mathbf{E}\left(\mathbf{E}[(\eta - \mathbf{E}(\eta | \xi))^2 \mid \xi = x]\right) f_{\xi}(x) dx \\ &= \int_{-\infty}^{\infty} \mathbf{E}\left(\mathbf{E}[(\eta - \mathbf{E}(\eta | \xi = x))^2 \mid \xi = x]\right) f_{\xi}(x) dx \\ &= \int_{-\infty}^{\infty} \mathbf{V}(\eta | \xi = x) f_{\xi}(x) dx = \sigma_2^2(1-r^2), \end{aligned}$$

dove si è fatto anche uso del precedente calcolo di  $\mathbf{V}(\eta | \xi = x)$ . □

**II.8.10 Osservazione:** Se  $\xi$  è una v.a. con FdD  $F_\xi(x)$  (ed eventualmente dotata di fdd  $f_\xi(x)$ ), data una funzione di Borel  $\varphi(x)$  ci porremo ora il problema di **determinare la FdD  $F_\eta(y)$  (ed eventualmente la fdd  $f_\eta(y)$ ) della v.a.  $\eta = \varphi(\xi)$  in termini delle funzioni  $F_\xi$  (eventualmente  $f_\xi$ ) e  $\varphi$ .** In sostanza ciò viene fatto notando che

$$F_\eta(y) = \mathbf{P}(\eta \leq y) = \mathbf{P}(\varphi(\xi) \leq y) = \mathbf{P}(\xi \in \varphi^{-1}(-\infty, y]) = \int_{\varphi^{-1}(-\infty, y]} F_\xi(dx),$$

che fornisce la relazione richiesta. Nella pratica si ottengono formule di facile uso quando le v.a. coinvolte sono tutte dotate di fdd  $f_\xi$  e  $f_\eta$ . Supporremo innanzitutto che  $\xi$  prenda valori che cadono nell'intervallo  $[a, b]$  (in modo che  $f_\xi(x)$  sia nulla al di fuori di tale intervallo) e che la funzione  $\varphi(x)$  sia definita, derivabile e strettamente crescente ( $\varphi'(x) > 0$ ) su tutto  $[a, b]$ . Detto allora  $[\alpha, \beta]$  l'insieme dei valori assunti da  $\varphi(x)$  (codominio) e indicata per semplicità con  $h(y) = \varphi^{-1}(y)$  la funzione inversa di  $\varphi$  (la cui esistenza è garantita dalle ipotesi fatte), per  $y \in [\alpha, \beta]$  l'ipotesi di monotonìa implica, con il cambiamento di variabile  $t = \varphi(x)$  (ossia  $x = h(t)$ ), che

$$\begin{aligned} F_\eta(y) &= \mathbf{P}(\eta \leq y) = \mathbf{P}(\varphi(\xi) \leq y) = \mathbf{P}(\xi \leq \varphi^{-1}(y)) = \mathbf{P}(\xi \leq h(y)) \\ &= \int_{-\infty}^{h(y)} f_\xi(x) dx = \int_a^{h(y)} f_\xi(x) dx = \int_\alpha^y f_\xi(h(t)) h'(t) dt \\ &= \int_{-\infty}^y f_\xi(h(t)) h'(t) dt, \end{aligned}$$

da cui si ottiene immediatamente che, per  $y \in [\alpha, \beta]$ ,

$$f_\eta(y) = f_\xi(h(y)) h'(y),$$

mentre per i valori di  $y$  che non vengono assunti da  $\eta$  si ha  $f_\eta(y) = 0$ . Se invece  $\varphi$  fosse strettamente decrescente un analogo ragionamento condurrebbe alla conclusione che per  $y \in [\alpha, \beta]$

$$f_\eta(y) = -f_\xi(h(y)) h'(y).$$

In pratica, se si fa l'ipotesi che  $\varphi$  sia strettamente monotona, per ambedue i casi potremo scrivere

$$f_\eta(y) = f_\xi(h(y)) |h'(y)| I_{[\alpha, \beta]}(y).$$

Osserviamo inoltre che, indicata con  $x_1(y)$  la soluzione (unica nelle nostre ipotesi) dell'equazione  $\varphi(x) = y$  con  $y \in [\alpha, \beta]$ , e ricordando che per un ben noto risultato dell'analisi (vedi ad esempio **V.I. Smirnov**: *Corso di matematica superiore*, vol. I; Editori Riuniti, Roma, 1977, p. 114)

$$h'(y) = \frac{1}{\varphi'(x_1(y))},$$

è anche possibile riscrivere la relazione precedente come

$$f_{\eta}(y) = \frac{f_{\xi}(x_1(y))}{|\varphi'(x_1(y))|} \cdot I_{[\alpha, \beta]}(y).$$

Se invece  $\varphi$  risulta non strettamente monotona su tutto l'insieme dei valori assunti da  $\xi$ , in molti casi di utilità pratica si potrà decomporre il suo insieme di definizione nell'unione di  $n$  intervalli  $[a_k, b_k]$  all'interno dei quali essa è derivabile e strettamente monotona, con derivata non nulla in ogni punto. Se allora  $h_k(y)$  indica la funzione inversa della restrizione di  $\varphi(x)$  all'intervallo  $[a_k, b_k]$ , e se con  $[\alpha_k, \beta_k]$  indichiamo il suo codominio, si verifica facilmente che

$$f_{\eta}(y) = \sum_{k=1}^n f_{\xi}(h_k(y)) |h'_k(y)| \cdot I_{[\alpha_k, \beta_k]}(y)$$

dato che in ogni intervallo sono vere le relazioni precedenti. In pratica nella somma indicata saranno presenti solo gli  $m \leq n$  addendi per i quali gli indicatori sono diversi da zero, cioè quelli per i quali  $y \in [\alpha_k, \beta_k]$ . Se allora, fissato  $y$ , indichiamo con  $x_1(y), \dots, x_m(y)$  le  $m$  soluzioni dell'equazione  $\varphi(x) = y$ , un'osservazione analoga a quella del caso precedente ci consentirà di scrivere la nostra relazione nella forma

$$f_{\eta}(y) = \sum_{j=1}^m \frac{f_{\xi}(x_j(y))}{|\varphi'(x_j(y))|} \cdot I_{[\alpha, \beta]}(y),$$

nella quale, in definitiva, compaiono solo i termini non nulli della somma precedente. ○

**II.8.11 Esempio:** Se  $\eta = a\xi + b$ , con  $a > 0$ , si ha

$$F_{\eta}(y) = \mathbf{P}\left(\xi \leq \frac{y-b}{a}\right) = F_{\xi}\left(\frac{y-b}{a}\right).$$

Se invece fosse  $a < 0$  si avrebbe

$$F_{\eta}(y) = \mathbf{P}\left(\xi \geq \frac{y-b}{a}\right) = 1 - \mathbf{P}\left(\xi < \frac{y-b}{a}\right) = 1 - F_{\xi}\left(\frac{y-b}{a}\right) + \mathbf{P}\left(\xi = \frac{y-b}{a}\right).$$

Se  $\xi$  è dotata di fdd  $f_{\xi}(x)$ , si ha  $h(y) = (y-b)/a$  (con  $a \neq 0$ ), e quindi

$$f_{\eta}(y) = \frac{1}{|a|} f_{\xi}\left(\frac{y-b}{a}\right),$$

formula valida sia per  $a > 0$  che per  $a < 0$ . In particolare, se  $\xi \sim \mathcal{N}(m, \sigma^2)$  è una v.a. normale (nel seguito useremo spesso il simbolo  $\sim$  per abbreviare l'espressione

è distribuita come), la relazione precedente mostra che, posto  $a = 1/\sigma$  e  $b = -m/\sigma$ , si ha

$$\eta = a\xi + b = \frac{\xi - m}{\sigma} \sim \mathcal{N}(0, 1),$$

cioè che  $\eta$  è una v.a. normale standard. ◇

**II.8.12 Esempio:** Se  $\eta = \xi^2$  si ha ovviamente che  $F_\eta(y) = 0$  per  $y < 0$ , mentre per  $y \geq 0$

$$F_\eta(y) = \mathbf{P}(\xi^2 \leq y) = \mathbf{P}(-\sqrt{y} \leq \xi \leq \sqrt{y}) = F_\xi(\sqrt{y}) - F_\xi(-\sqrt{y}) + \mathbf{P}(\xi = -\sqrt{y}).$$

Se  $\xi$  è dotata di fdd  $f_\xi(x)$  dovremo osservare che la funzione  $\varphi(x) = x^2$  non è monotona su  $(-\infty, \infty)$ . Converrà allora suddividere l'asse reale nei due intervalli  $(-\infty, 0)$  e  $(0, \infty)$  sui quali le restrizioni di  $\varphi(x) = x^2$  sono strettamente monotone. Si possono allora definire due funzioni inverse:  $h_1(y) = -\sqrt{y}$  (con  $y > 0$ ) per la restrizione su  $(-\infty, 0)$  e  $h_2(y) = \sqrt{y}$  (con  $y > 0$ ) per la restrizione su  $(0, \infty)$ . Ciò posto, dalle formule di Osservazione II.8.10 otteniamo

$$f_{\xi^2}(y) = f_\eta(y) = \begin{cases} \frac{1}{2\sqrt{y}} [f_\xi(\sqrt{y}) + f_\xi(-\sqrt{y})], & y > 0 \\ 0, & y \leq 0. \end{cases}$$

Questo risultato poteva anche essere ottenuto per derivazione dalla formula trovata in precedenza per  $F_\eta(y)$ , tenendo conto anche del fatto che, essendo  $F_\xi(x)$  a.c. e quindi continua, risulta anche  $\mathbf{P}(\xi = -\sqrt{y}) = 0$ . In particolare, se  $\xi \sim \mathcal{N}(0, 1)$  si ha

$$f_{\xi^2}(y) = \begin{cases} \frac{1}{\sqrt{2\pi y}} e^{-y/2}, & y > 0 \\ 0, & y \leq 0. \end{cases}$$

Le seguenti relazioni

$$f_{|\xi|}(y) = \begin{cases} f_\xi(y) + f_\xi(-y), & y > 0 \\ 0, & y \leq 0, \end{cases}$$

$$f_{+\sqrt{|\xi|}}(y) = \begin{cases} 2y[f_\xi(y^2) + f_\xi(-y^2)], & y > 0 \\ 0, & y \leq 0, \end{cases}$$

possono essere infine dimostrate nello stesso modo delle precedenti. ◇

**II.8.13 Esempio:** Se  $\eta = e^\xi$  e  $\xi \sim \mathcal{N}(m, \sigma^2)$ , si ha che (verificarlo per esercizio)

$$f_\eta(y) = \begin{cases} \frac{1}{\sigma y \sqrt{2\pi}} e^{-(\ln y - m)^2 / 2\sigma^2}, & y > 0 \\ 0, & y \leq 0. \end{cases}$$

Una v.a. distribuita in questo modo prende anche il nome di **log-normale**.  $\diamond$

**II.8.14 Osservazione:** Ci proponiamo ora di generalizzare alcuni risultati della Osservazione II.8.10 al caso di funzioni di più di una v.a. Siano  $\xi$  e  $\eta$  due v.a. con FdD congiunta  $F_{\xi\eta}(x, y)$  e sia  $\varphi(x, y)$  una funzione di Borel: ragionando come nell'Osservazione II.8.10, e indicando per brevità con  $\{\varphi(x, y) \leq z\}$  l'insieme  $\varphi^{-1}(-\infty, z]$  delle coppie  $(x, y) \in \mathbf{R}^2$  per le quali vale la relazione  $\varphi(x, y) \leq z$ , potremo concludere che la FdD della v.a.  $\zeta = \varphi(\xi, \eta)$  può essere calcolata a partire dalla conoscenza di  $\varphi$  e  $F_{\xi\eta}$  mediante la relazione

$$\begin{aligned} F_{\zeta}(z) &= \int_{\varphi^{-1}(-\infty, z]} F_{\xi\eta}(dx, dy) = \int_{\{\varphi(x, y) \leq z\}} F_{\xi\eta}(dx, dy) \\ &= \int_{\mathbf{R}^2} I_{\{\varphi(x, y) \leq z\}}(x, y) F_{\xi\eta}(dx, dy). \end{aligned}$$

In particolare, se  $\varphi(x, y) = x + y$  e se le v.a.  $\xi$  e  $\eta$  sono indipendenti, in modo che  $F_{\xi\eta}(x, y) = F_{\xi}(x)F_{\eta}(y)$ , dal Teorema di Fubini II.6.29 avremo che

$$\begin{aligned} F_{\zeta}(z) &= \int_{\mathbf{R}^2} I_{\{x+y \leq z\}}(x, y) F_{\xi}(dx)F_{\eta}(dy) \\ &= \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} I_{\{y \leq z-x\}}(y) F_{\eta}(dy) \right] F_{\xi}(dx) = \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{z-x} F_{\eta}(dy) \right] F_{\xi}(dx) \\ &= \int_{-\infty}^{\infty} F_{\eta}(z-x) F_{\xi}(dx), \end{aligned}$$

o analogamente, invertendo l'ordine delle integrazioni,

$$F_{\zeta}(z) = \int_{-\infty}^{\infty} F_{\xi}(z-y) F_{\eta}(dy).$$

Se allora, date due FdD  $F$  e  $G$ , chiamiamo **convoluzione** di  $F$  e  $G$  la funzione<sup>3</sup>

$$H(x) = F(x) * G(x) = G(x) * F(x) = \int_{-\infty}^{\infty} F(x-y) G(dy) = \int_{-\infty}^{\infty} G(x-y) F(dy),$$

potremo dire di aver mostrato che la FdD della somma  $\zeta$  di due v.a. indipendenti  $\xi$  e  $\eta$  è la convoluzione delle loro FdD marginali:

$$F_{\zeta}(x) = F_{\xi}(x) * F_{\eta}(x) = F_{\eta}(x) * F_{\xi}(x).$$

Se poi le nostre due v.a. sono anche dotate di fdd  $f_{\xi}$  e  $f_{\eta}$ , dalla relazione

$$\begin{aligned} F_{\zeta}(z) &= \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{z-x} f_{\eta}(y) dy \right] f_{\xi}(x) dx = \int_{-\infty}^{\infty} \left[ \int_{-\infty}^z f_{\eta}(t-x) dt \right] f_{\xi}(x) dx \\ &= \int_{-\infty}^z \left[ \int_{-\infty}^{\infty} f_{\eta}(t-x) f_{\xi}(x) dx \right] dt, \end{aligned}$$

<sup>3</sup> È facile verificare che la funzione che si ottiene dalla convoluzione di due FdD è ancora una FdD secondo la Definizione II.3.2.

e da quella analoga ottenuta invertendo l'ordine delle integrazioni, si ottiene che

$$f_{\zeta}(z) = \int_{-\infty}^{\infty} f_{\eta}(z-x)f_{\xi}(x) dx = \int_{-\infty}^{\infty} f_{\xi}(z-y)f_{\eta}(y) dy,$$

e se chiamiamo **convoluzione** di due fdd  $f$  e  $g$  la fdd

$$h(x) = f(x) * g(x) = g(x) * f(x) = \int_{-\infty}^{\infty} f(x-y)g(y) dy = \int_{-\infty}^{\infty} g(x-y)f(y) dy,$$

potremo affermare che la fdd di una v.a.  $\zeta$  somma di due v.a.  $\xi$  e  $\eta$  indipendenti è la convoluzione delle rispettive fdd marginali

$$f_{\zeta}(x) = f_{\xi}(x) * f_{\eta}(x) = f_{\eta}(x) * f_{\xi}(x).$$

Le considerazioni precedenti si generalizzano facilmente al caso di somme di più di due v.a.: se ad esempio  $\xi_1, \dots, \xi_n$  sono  $n$  v.a. indipendenti dotate di fdd, la fdd della loro somma  $\eta = \xi_1 + \dots + \xi_n$  si ottiene eseguendo la convoluzione delle rispettive marginali secondo la relazione

$$f_{\eta}(x) = f_{\xi_1}(x) * \dots * f_{\xi_n}(x).$$

Infine, sempre con gli stessi metodi, si può mostrare che se  $\xi$  e  $\eta$  sono v.a. indipendenti dotate di fdd con marginali  $f_{\xi}(x)$  e  $f_{\eta}(y)$ , le FdD delle v.a.  $\zeta = \xi \cdot \eta$  e  $\tau = \xi/\eta$  sono date da

$$F_{\zeta}(z) = \int_{\{xy \leq z\}} F_{\xi\eta}(dx, dy) = \int_{\mathbf{R}^2} I_{\{xy \leq z\}}(x, y) f_{\xi}(x) f_{\eta}(y) dx dy,$$

$$F_{\tau}(t) = \int_{\{x/y \leq t\}} F_{\xi\eta}(dx, dy) = \int_{\mathbf{R}^2} I_{\{x/y \leq t\}}(x, y) f_{\xi}(x) f_{\eta}(y) dx dy,$$

mentre le loro fdd sono

$$f_{\zeta}(z) = \int_{-\infty}^{\infty} f_{\xi}\left(\frac{z}{y}\right) f_{\eta}(y) \frac{dy}{|y|},$$

$$f_{\tau}(t) = \int_{-\infty}^{\infty} f_{\xi}(ty) f_{\eta}(y) |y| dy,$$

come può essere verificato con qualche cambiamento di variabili. ○

**II.8.15 Esempio:** Se  $\xi_1, \dots, \xi_n$  sono v.a. i.i.d. con fdd uniformi su  $[-1, 1]$ , cioè con

$$f_{\xi_k}(x) = f(x) = \begin{cases} 1/2 & , \text{ se } |x| \leq 1 \\ 0 & , \text{ se } |x| > 1 \end{cases}; \quad k = 1, \dots, n,$$

si mostra facilmente con un calcolo diretto che

$$f_{\xi_1+\xi_2}(x) = \begin{cases} \frac{2-|x|}{4}, & |x| \leq 2, \\ 0, & |x| > 2, \end{cases}$$

$$f_{\xi_1+\xi_2+\xi_3}(x) = \begin{cases} \frac{(3-|x|)^2}{16}, & 1 \leq |x| \leq 3, \\ \frac{3-x^2}{8}, & 0 \leq |x| \leq 1, \\ 0, & |x| > 3, \end{cases}$$

e quindi per induzione si mostra anche che

$$f_\eta(x) = \begin{cases} \frac{1}{2^n(n-1)!} \sum_{k=0}^{[(n+x)/2]} (-1)^k \binom{n}{k} (n+x-2k)^{n-1}, & |x| \leq n, \\ 0, & |x| > n, \end{cases}$$

se  $\eta = \xi_1 + \dots + \xi_n$  e se con  $[\alpha]$  indichiamo la parte intera del numero reale  $\alpha$ .  $\diamond$

**II.8.16 Esempio:** Se  $\xi \sim \mathcal{N}(m_1, \sigma_1^2)$  e  $\eta \sim \mathcal{N}(m_2, \sigma_2^2)$  sono due v.a. normali indipendenti, posto per brevità

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

in modo che

$$f_\xi(x) = \frac{1}{\sigma_1} f\left(\frac{x-m_1}{\sigma_1}\right), \quad f_\eta(y) = \frac{1}{\sigma_2} f\left(\frac{y-m_2}{\sigma_2}\right),$$

mediante calcolo diretto si ottiene che

$$f_{\xi+\eta}(x) = \frac{1}{\sqrt{\sigma_1^2 + \sigma_2^2}} f\left(\frac{x-(m_1+m_2)}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right),$$

cioè che  $\xi + \eta \sim \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$ . In altri termini: la somma di due v.a. normali indipendenti è ancora una v.a. normale con VdA pari alla somma dei VdA iniziali e Var pari alla somma delle Var iniziali. Questo importante risultato, che si estende anche ad un numero arbitrario di v.a., prende il nome di **proprietà riproduttiva delle v.a. normali**.  $\diamond$

**II.8.17 Esempio:** Se  $\xi_1, \dots, \xi_n$  sono v.a. i.i.d. tutte con fdd normale standard, cioè  $\xi_k \sim \mathcal{N}(0, 1)$ ,  $k = 1, \dots, n$ , tenendo conto dei risultati dell'Esempio II.8.12 e

applicando ripetutamente la regola sulle convoluzioni dell'Osservazione II.8.14 alle  $f_{\xi_k^2}(x)$  si ottiene la seguente fdd della v.a.  $\chi_n^2 = \xi_1^2 + \dots + \xi_n^2$  che prende il nome di **distribuzione del  $\chi^2$  con  $n$  gradi di libertà**:

$$f_{\chi_n^2}(x) = \begin{cases} \frac{x^{n/2-1}e^{-x/2}}{2^{n/2}\Gamma(n/2)}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

dove  $\Gamma(x)$  è la funzione Gamma definita da

$$\Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t} dt$$

e che soddisfa le proprietà

$$\Gamma(x) = (x-1)\Gamma(x-1), \quad \Gamma(1) = 1, \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi},$$

sicch  avremo

$$\Gamma\left(\frac{n}{2}\right) = \begin{cases} \frac{(n-2)!!}{2^{n/2}}, & \text{se } n \text{   pari,} \\ \frac{(n-2)!!}{2^{(n-1)/2}} \sqrt{\pi}, & \text{se } n \text{   dispari.} \end{cases}$$

Si chiama poi **distribuzione  $\chi$  con  $n$  gradi di libert ** la fdd

$$f_{\chi_n}(x) = \begin{cases} \frac{2x^{n-1}e^{-x^2/2}}{2^{n/2}\Gamma(n/2)}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

della v.a.  $\chi_n = +\sqrt{\chi_n^2}$  che si ottiene tenendo conto dei risultati dell'Esempio II.8.12. ◇

**II.8.18. Esempio:** Se  $\xi_0, \xi_1, \dots, \xi_n$  sono  $n+1$  v.a. normali i.i.d. con fdd  $\mathcal{N}(0, \sigma^2)$  (con  $\sigma > 0$ ), posto

$$\tau = \frac{\xi_0}{\sqrt{(\xi_1^2 + \dots + \xi_n^2)/n}} = \frac{\xi_0/\sigma}{\sqrt{(\xi_1^2 + \dots + \xi_n^2)/n\sigma^2}} = \frac{\xi_0/\sigma}{\sqrt{\chi_n^2/n}},$$

osserviamo che la v.a.  $\chi_n^2 = (\xi_1^2 + \dots + \xi_n^2)/\sigma^2$  segue una distribuzione del  $\chi^2$  con  $n$  gradi di libert  (vedi Esempio II.8.17) dato che ogni v.a.  $\xi_k/\sigma \sim \mathcal{N}(0, 1)$    normale standard come segue dall'Esempio II.8.11. Utilizzando allora i risultati dell'Esempio II.8.17 e dell'Osservazione II.8.14 si ricava allora che la fdd di  $\tau$   

$$f_\tau(t) = \frac{1}{\sqrt{\pi n}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}$$



che prende il nome di **distribuzione  $t$  di Student**<sup>4</sup>. È interessante osservare che essa è indipendente dal valore  $\sigma$  della Var delle v.a. iniziali.  $\diamond$

**II.8.19 Osservazione:** Le v.a. caratterizzate da distribuzioni  $\chi^2$  oppure  $t$  di Student non sono state introdotte qui in maniera del tutto arbitraria. La loro importanza è enorme in statistica in quanto esse permettono di stimare, sotto particolari ipotesi, la confidenza con cui certe affermazioni possono essere fatte o la probabilità di commettere errori accettando o rifiutando determinate ipotesi. Purtroppo per ragioni di spazio nel corso di queste lezioni non sarà possibile dare neanche un breve accenno ai problemi della **stima statistica** o dei **test di ipotesi**, sicché ci limiteremo a rinviare gli interessati a qualcuno dei numerosi testi disponibili come ad esempio **P. Baldi**, *Calcolo delle Probabilità e Statistica*, McGraw-Hill, Milano, 1992.  $\circ$

---

<sup>4</sup> Gli utilizzatori della cosiddetta distribuzione di Student ignorano spesso che *Student* era lo pseudonimo che lo statistico inglese William Sealy Gosset (1876-1937) usò per firmare le sue numerose pubblicazioni. Diversamente dalla maggior parte dei suoi colleghi Gosset, terminati gli studi universitari, decise di non intraprendere la carriera accademica e trovò un impiego presso la celebre birreria Guinness di Dublino (che più di una volta aveva assunto giovani scienziati usciti da Oxford o Cambridge) dove ebbe occasione di utilizzare la grande massa di dati sperimentali disponibili, riflettendo approfonditamente sulla teoria degli errori. Nel 1905 egli incontrò Karl Pearson (1857-1936), un altro importante statistico inglese, che lo invitò a trascorrere un anno di studi (1906-7) presso il laboratorio di Biometria dell'Università di Londra. Fu proprio nel corso di questo anno sabbatico che Gosset elaborò le basi dei lavori che lo resero celebre. Le sue due pubblicazioni fondamentali (*The probable error of a mean* e *Probable error of a correlation coefficient*) apparvero così nel 1908 sulla rivista *Biometrika* diretta da K. Pearson. Fino alla sua morte Gosset continuò a dare importanti contributi scientifici esercitando sui suoi colleghi, tramite la sua fitta corrispondenza, un notevole influsso, che il figlio di K. Pearson, Egon (1895-1980) e lo statistico polacco Jerzy Neyman (1894-1981) riconobbero esplicitamente a proposito, ad esempio, della teoria dei tests statistici e della comparazione di verosimiglianze di ipotesi. Precursore degli statistici industriali moderni Gosset fu sempre considerato dagli statistici come un consigliere scientifico della sua azienda e da questa come un dipendente che consacrava il suo tempo libero alla ricerca.



## II.9 Convergenza. Spazi vettoriali di variabili aleatorie

**II.9.1 Definizione:** Data una successione di v.a.  $(\xi_n)_{n \in \mathbf{N}}$  definite su  $(\Omega, \mathcal{F}, \mathbf{P})$ , diremo che essa **converge in probabilità** alla v.a.  $\xi$  ( $\xi_n \xrightarrow{\mathbf{P}} \xi$ ) se

$$\mathbf{P}\{|\xi_n - \xi| > \epsilon\} \xrightarrow{n} 0, \quad \forall \epsilon > 0;$$

diremo invece che essa vi **converge quasi ovunque (q.o.)**, o anche **quasi certamente** o **con probabilità uno**, ( $\xi_n \xrightarrow{q.o.} \xi$  o anche  $\xi_n \xrightarrow{n} \xi$ , **P-q.o.**), se, detto  $\{\xi_n \not\rightarrow \xi\}$  l'insieme di punti  $\omega \in \Omega$  per i quali  $\xi_n(\omega)$  non converge a  $\xi(\omega)$ , risulta

$$\mathbf{P}\{\xi_n \not\rightarrow \xi\} = 0;$$

diremo inoltre che essa vi **converge in  $L^p$** , o anche **in media di ordine  $p$** , con  $0 < p < +\infty$  ( $\xi_n \xrightarrow{L^p} \xi$ ), se

$$\mathbf{E}|\xi_n - \xi|^p \xrightarrow{n} 0;$$

in particolare se  $p = 2$  si dice che  $(\xi_n)_{n \in \mathbf{N}}$  **converge in media quadratica (m.q.)** e si usa la notazione  $\xi_n \xrightarrow{m.q.} \xi$  oppure  $\text{li.m.}\xi_n = \xi$ , dove *li.m.* sta per *limite in media*. diremo infine che essa vi **converge in distribuzione** ( $\xi_n \xrightarrow{d} \xi$ ) se

$$\mathbf{E}f(\xi_n) \xrightarrow{n} \mathbf{E}f(\xi), \quad \forall f \in \mathcal{C}(\mathbf{R})$$

dove  $\mathcal{C}(\mathbf{R})$  indica l'insieme delle funzioni limitate e continue su  $\mathbf{R}$ . △

**II.9.2 Definizione:** Data una successione di misure di probabilità  $(\mathbf{P}_n)_{n \in \mathbf{N}}$  definite su uno spazio  $(E, \mathcal{E})$  (con  $E$  spazio metrico), diremo che essa **converge debolmente** alla probabilità  $\mathbf{P}$  su  $(E, \mathcal{E})$  ( $\mathbf{P}_n \xrightarrow{w} \mathbf{P}$ ) se

$$\int_E f(x)\mathbf{P}_n(dx) \xrightarrow{n} \int_E f(x)\mathbf{P}(dx), \quad \forall f \in \mathcal{C}(E)$$

dove con  $\mathcal{C}(E)$  indicheremo l'insieme delle funzioni limitate e continue definite su  $E$  a valori in  $\mathbf{R}$ . △

**II.9.3 Definizione:** Data una successione di FdD  $(F_n(x))_{n \in \mathbf{N}}$  su  $\mathbf{R}$  diremo che essa **converge in generale** alla FdD  $F(x)$  ( $F_n \implies F$ ) se

$$F_n(x) \xrightarrow{n} F(x), \quad \forall x \in P_C(F)$$

dove  $P_C(F)$  è l'insieme dei punti  $x \in \mathbf{R}$  per i quali  $F(x)$  è continua. △

**II.9.4 Osservazione:** Notiamo innanzitutto che mentre le definizioni di II.9.1 sono relative a successioni di v.a., quelle di II.9.2 e II.9.3 si riferiscono a successioni

di misure di probabilità o a FdD. Inoltre, affinché le definizioni di convergenza in probabilità, quasi ovunque e in  $L^p$  abbiano senso, le v.a.  $\xi_n$  devono essere tutte definite sul medesimo spazio<sup>1</sup>  $(\Omega, \mathcal{F}, \mathbf{P})$ , mentre invece la definizione di convergenza in distribuzione può essere data anche quando le v.a.  $\xi_n$  sono definite su spazi di probabilità diversi. Infatti quest'ultimo tipo di convergenza, mediante il Teorema II.6.25, può anche essere definito richiedendo che

$$\int_E f(x) \mathbf{P}_{\xi_n}(dx) \xrightarrow{n} \int_E f(x) \mathbf{P}_{\xi}(dx), \quad \forall f \in \mathcal{C}(\mathbf{R})$$

e risulta quindi essere equivalente alla convergenza debole  $\mathbf{P}_{\xi_n} \xrightarrow{w} \mathbf{P}_{\xi}$  della relativa successione di DdP. Ovviamente, quando come nel nostro caso, le v.a.  $\xi_n$  e  $\xi$  prendono valori in  $\mathbf{R}$  (cioè quando  $E = \mathbf{R}$ ), la convergenza debole  $\mathbf{P}_{\xi_n} \xrightarrow{w} \mathbf{P}_{\xi}$  della successione di DdP può essere equivalentemente pensata come convergenza debole  $F_{\xi_n} \xrightarrow{w} F_{\xi}$  della relativa successione di FdD, indicando con il simbolo  $F_n \xrightarrow{w} F$  il fatto che

$$\int_E f(x) F_n(dx) \xrightarrow{n} \int_E f(x) F(dx), \quad \forall f \in \mathcal{C}(\mathbf{R}).$$

Non si deve però pensare che la convergenza debole di FdD  $F_{\xi_n} \xrightarrow{w} F_{\xi}$  equivale alla convergenza ordinaria  $F_{\xi_n}(x) \xrightarrow{n} F_{\xi}(x)$  in ogni punto  $x \in \mathbf{R}$ : mentre, infatti, quest'ultima implica la convergenza debole, il viceversa non è garantito come verrà mostrato qui di seguito con un controesempio tratto dalla discussione sulla Legge dei Grandi Numeri e sui Teoremi Limite svolta in I.7 e I.8. Una successiva Proposizione chiarirà infine che la convergenza debole di FdD equivale non alla convergenza ordinaria in ogni punto  $x \in \mathbf{R}$  ma solo alla convergenza in generale  $F_n \implies F$ . ○

**II.9.5 Esempio:** Supponiamo che  $(\xi_n)_{n \in \mathbf{N}}$  sia una successione di v.a., definite su  $(\Omega, \mathcal{F}, \mathbf{P})$ , che assumono solo i valori 1 e 0 rispettivamente con probabilità  $p$  e  $q = 1 - p$ ; supponiamo poi che tali  $\xi_n$ , oltre che essere identicamente distribuite, siano anche indipendenti e definiamo  $S^{(n)} = \xi_1 + \dots + \xi_n$ . È noto dalle precedenti discussioni (vedi ad esempio l'Osservazione I.5.17) che tale  $S^{(n)}$  è una v.a. binomiale con

$$\mathbf{P}(S^{(n)} = k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, \dots, n.$$

Sappiamo inoltre dalla Legge dei grandi Numeri (Teorema I.7.5) che

$$\frac{S^{(n)}}{n} \xrightarrow{\mathbf{P}} p, \quad (n \rightarrow \infty)$$

---

<sup>1</sup> Ricorda in proposito la discussione svolta in I.7.1 e I.7.2 circa la formulazione della Legge dei Grandi Numeri di J. Bernoulli I.7.5 che veniva espressa mediante affermazioni sulla convergenza in probabilità.

e, come vedremo fra breve, la convergenza in probabilità implica sempre anche la convergenza in distribuzione, per cui dette

$$F_n(x) = \mathbf{P} \left( \frac{S^{(n)}}{n} \leq x \right); \quad F(x) = \begin{cases} 1, & \text{se } x \geq p; \\ 0, & \text{se } x < p; \end{cases}$$

rispettivamente le FdD di  $S^{(n)}/n$  e della v.a. identicamente eguale a  $p$ , la Legge dei Grandi Numeri ci dice anche che  $F_n \xrightarrow{w} F$ . Questo però non implica che  $F_n(x)$  converga a  $F(x)$  in ogni punto  $x \in \mathbf{R}$ : mostreremo ora infatti che  $F_n(x) \xrightarrow{n} F(x)$  solo nei punti in cui la FdD limite  $F(x)$  è continua, cioè in tutte le  $x$  con l'eccezione di  $x = p$ . Infatti

$$F_n(x) = \mathbf{P} \left( \frac{S^{(n)}}{n} \leq x \right) = \mathbf{P} \left( \frac{S^{(n)} - np}{\sqrt{npq}} \leq \sqrt{n} \frac{x - p}{\sqrt{pq}} \right)$$

e siccome ovviamente

$$\left| F_n(x) - \int_{-\infty}^{\sqrt{n} \frac{x-p}{\sqrt{pq}}} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt \right| \leq \sup_{-\infty \leq b \leq +\infty} \left| F_n(b) - \int_{-\infty}^b \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt \right|$$

si ottiene facilmente dal Teorema Limite Integrale I.8.6 che, comunque fissato  $x$ , risulta

$$\lim_n \left| F_n(x) - \int_{-\infty}^{\sqrt{n} \frac{x-p}{\sqrt{pq}}} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt \right| = 0.$$

Pertanto, detta  $\Phi(x)$  la funzione errore standard (vedi I.8.1), otteniamo che

$$\lim_n F_n(x) = \lim_n \Phi \left( \sqrt{n} \frac{x - p}{\sqrt{pq}} \right), \quad \forall x \in \mathbf{R}.$$

Dall'andamento della funzione  $\Phi$  (vedi ad esempio Figura I.8.1 e Tavola I.8.1) otteniamo allora che

$$\lim_n F_n(x) = \begin{cases} 0, & \text{se } x < p, \\ 1/2, & \text{se } x = p, \\ 1, & \text{se } x > p, \end{cases}$$

e quindi  $\lim_n F_n(x) = F(x)$  nei punti  $x \neq p$  in cui  $F$  è continua, ma  $\lim_n F_n(x) = 1/2$  in  $x = p$  mentre  $F(p) = 1$ . È interessante notare che la situazione messa in evidenza dal presente esempio è del tutto generale nel senso che la convergenza debole di una successione di FdD verso un'altra FdD non implica la convergenza puntuale della successione di funzioni se non per i punti in cui la FdD limite risulta continua. La formulazione esatta di questa affermazione è contenuta nel seguente Teorema in base al quale (e in base all'Osservazione II.9.4) potremo dire che la convergenza in distribuzione di una successione di v.a.  $(\xi_n \xrightarrow{d} \xi)$  equivale sia alla convergenza debole della corrispondente successione di FdD  $(F_{\xi_n} \xrightarrow{w} F_{\xi})$  che alla convergenza in generale delle medesime FdD  $(F_{\xi_n} \implies F_{\xi})$ .  $\diamond$

**II.9.6 Teorema:** Data la successione di FdD  $(F_n)_{n \in \mathbf{N}}$  e la FdD  $F$ , si ha  $F_n \xrightarrow{w} F$  se e solo se  $F_n \implies F$ .

**Dimostrazione:** Omessa (vedi ad esempio **A.N. Shiriyayev:** *Probability*; Springer, New York, 1984, p. 311). □

**II.9.7 Teorema:** Se  $\xi_n$ , con  $n \in \mathbf{N}$ , e  $\xi$  sono v.a., le seguenti implicazioni sono sempre verificate:

- (a)  $\xi_n \xrightarrow{q.o.} \xi \implies \xi_n \xrightarrow{\mathbf{P}} \xi$ ;
- (b)  $\xi_n \xrightarrow{L^p} \xi \implies \xi_n \xrightarrow{\mathbf{P}} \xi$ ,  $p > 0$ ;
- (c)  $\xi_n \xrightarrow{\mathbf{P}} \xi \implies \xi_n \xrightarrow{d} \xi$ ;
- (d)  $\xi_n \xrightarrow{d} C \implies \xi_n \xrightarrow{\mathbf{P}} C$ , se  $C$  è una costante.

**Dimostrazione:**

- (a) Introdotta, per brevità, la notazione  $A_n^\epsilon = \{|\xi_n - \xi| > \epsilon\}$  con  $n \in \mathbf{N}$  ed  $\epsilon > 0$ , si verifica subito che, essendo

$$\bigcup_{k \geq n} A_k^\epsilon = \{\exists k \geq n \ni |\xi_k - \xi| > \epsilon\},$$

risulta anche

$$\bigcup_{k \geq n} A_k^\epsilon = \left\{ \sup_{k \geq n} |\xi_k - \xi| > \epsilon \right\}, \quad \forall n \in \mathbf{N}, \quad \forall \epsilon > 0.$$

Siccome inoltre

$$\{|\xi_n - \xi| > \epsilon\} \subseteq \left\{ \sup_{k \geq n} |\xi_k - \xi| > \epsilon \right\}, \quad \forall n \in \mathbf{N}, \quad \forall \epsilon > 0,$$

per provare la convergenza  $\xi_n \xrightarrow{\mathbf{P}} \xi$ , cioè  $\mathbf{P}(|\xi_n - \xi| > \epsilon) \xrightarrow{n} 0, \forall \epsilon > 0$ , sarà sufficiente provare che

$$\mathbf{P} \left( \sup_{k \geq n} |\xi_k - \xi| > \epsilon \right) \xrightarrow{n} 0, \quad \forall \epsilon > 0,$$

ossia equivalentemente che

$$\lim_n \mathbf{P} \left( \bigcup_{k \geq n} A_k^\epsilon \right) = 0, \quad \forall \epsilon > 0.$$

D'altra parte, sempre con la stessa notazione, dobbiamo osservare che

$$A_n^\epsilon \subseteq \{|\xi_n - \xi| \geq \epsilon\}, \quad \forall n \in \mathbf{N}, \quad \forall \epsilon > 0,$$

e quindi che risulterà anche

$$\begin{aligned} \{\xi_n \not\rightarrow \xi\} &= \overline{\{\xi_n \rightarrow \xi\}} = \overline{\{\forall \epsilon > 0 \exists n \in \mathbf{N} \exists' \forall k \geq n : |\xi_k - \xi| < \epsilon\}} \\ &= \bigcap_{\epsilon > 0} \bigcup_{n \in \mathbf{N}} \bigcap_{k \geq n} \{|\xi_n - \xi| < \epsilon\} = \bigcup_{\epsilon > 0} \bigcap_{n \in \mathbf{N}} \bigcup_{k \geq n} \{|\xi_n - \xi| \geq \epsilon\} \\ &\supseteq \bigcup_{\epsilon > 0} \bigcap_{n \in \mathbf{N}} \bigcup_{k \geq n} A_k^\epsilon = \bigcup_{\epsilon > 0} A^\epsilon \end{aligned}$$

dove per brevità abbiamo posto

$$A^\epsilon = \bigcap_{n \in \mathbf{N}} \bigcup_{k \geq n} A_k^\epsilon = \overline{\lim}_n A_n^\epsilon.$$

Pertanto l'ipotesi  $\xi_n \xrightarrow{q.o.} \xi$  (cioè  $\mathbf{P}(\xi_n \not\rightarrow \xi) = 0$ ) implica che

$$\mathbf{P}\left(\bigcup_{\epsilon > 0} A^\epsilon\right) = 0,$$

e quindi, anche in base alle osservazioni precedenti, la proposizione (a) sarà dimostrata se riusciremo a provare che, comunque scelto  $\epsilon > 0$ , risulta

$$\mathbf{P}\left(\bigcup_{\epsilon' > 0} A^{\epsilon'}\right) = 0 \implies \lim_n \mathbf{P}\left(\bigcup_{k \geq n} A_k^\epsilon\right) = 0.$$

Per raggiungere questo scopo notiamo che  $(A^{\epsilon'})_{\epsilon' > 0}$  è una famiglia non numerabile di eventi che risulta decrescente per inclusione<sup>2</sup>. È facile, però, estrarne la famiglia numerabile, e crescente per inclusione,  $(A^{1/m})_{m \in \mathbf{N}}$  per la quale risulta

$$A^{1/m} \uparrow \bigcup_{m \in \mathbf{N}} A^{1/m} = \bigcup_{\epsilon > 0} A^\epsilon$$

e che ci consentirà di lavorare con unioni numerabili di eventi. Dal punto (2) del Teorema II.1.13 e dalle nostre ipotesi, allora, si ha

$$\lim_m \mathbf{P}(A^{1/m}) = \mathbf{P}\left(\bigcup_{m \in \mathbf{N}} A^{1/m}\right) = \mathbf{P}\left(\bigcup_{\epsilon > 0} A^\epsilon\right) = 0,$$

e siccome  $\mathbf{P}(A^{1/m})$  è una successione numerica positiva e crescente, otteniamo anche che  $\mathbf{P}(A^{1/m}) = 0$  comunque scelto  $m \in \mathbf{N}$ . Ne segue<sup>3</sup> allora che  $\mathbf{P}(A^\epsilon) = 0$ , comunque scelto  $\epsilon > 0$ , e siccome per definizione di  $A^\epsilon$  si ha

$$\bigcup_{k \geq n} A_k^\epsilon \downarrow A^\epsilon = \bigcap_{n \in \mathbf{N}} \bigcup_{k \geq n} A_k^\epsilon,$$

<sup>2</sup> Infatti, se  $\epsilon' \leq \epsilon$ , si ha  $A_n^{\epsilon'} = \{|\xi_n - \xi| > \epsilon'\} \supseteq \{|\xi_n - \xi| > \epsilon\} = A_n^\epsilon$ , per ogni  $n \in \mathbf{N}$ .

<sup>3</sup> Basta osservare che, comunque scelto  $\epsilon > 0$ , esiste sempre un  $m \in \mathbf{N}$  tale che  $1/m < \epsilon$  e quindi tale che  $P(A^\epsilon) \leq P(A^{1/m})$ .

dal punto (3) del Teorema II.1.13 avremo anche che

$$\lim_n \mathbf{P} \left( \bigcup_{k \geq n} A_k^\epsilon \right) = \mathbf{P}(A^\epsilon) = 0, \quad \forall \epsilon > 0,$$

il che, come già osservato, è sufficiente a provare l'affermazione (a).

(b) Dal Corollario II.6.17 alla Diseguaglianza di Chebyshev si ha che

$$\mathbf{P}\{|\xi_n - \xi| > \epsilon\} \leq \mathbf{P}\{|\xi_n - \xi| \geq \epsilon\} \leq \frac{\mathbf{E}|\xi_n - \xi|^p}{\epsilon^p}, \quad p > 0,$$

per cui se  $\xi_n \xrightarrow{L^p} \xi$ , cioè se  $\mathbf{E}|\xi_n - \xi|^p \xrightarrow{n} 0$ , si ottiene anche, comunque scelto  $\epsilon > 0$ , che  $\mathbf{P}\{|\xi_n - \xi| > \epsilon\} \xrightarrow{n} 0$ .

(c) Sia  $f(x)$  un'arbitraria funzione limitata e continua di  $\mathcal{C}(\mathbf{R})$  con  $|f(x)| \leq c$  e  $c \geq 1$  (scelta, sempre possibile, effettuata per comodità), e sia  $\epsilon > 0$  un numero arbitrario: siccome dal Corollario II.6.17 sappiamo che  $\mathbf{P}(|\xi| > N) \leq \mathbf{E}|\xi|/N$ , potremo sempre trovare  $N > 0$  abbastanza grande per far in modo che  $\mathbf{P}(|\xi| > N) < \epsilon/4c$ ; inoltre, siccome  $f(x)$  è uniformemente continua in  $[-N, N]$ , per il Teorema di Cantor<sup>4</sup> sappiamo che sicuramente esiste un  $\delta > 0$  tale che, comunque presi  $x \in [-N, N]$  ed  $y$  tale che  $|x - y| \leq \delta$ , risulta  $|f(x) - f(y)| \leq \epsilon/2c$ . Ciò posto, tenendo conto del fatto che  $c \geq 1$  e che  $|f(x) - f(y)| \leq |f(x)| + |f(y)| \leq 2c$ , se per brevità usiamo la notazione  $\mathbf{E}(\xi; A) = \mathbf{E}(\xi I_A)$ , otteniamo che

$$\begin{aligned} \mathbf{E}|f(\xi_n) - f(\xi)| &= \mathbf{E}(|f(\xi_n) - f(\xi)|; |\xi_n - \xi| \leq \delta, |\xi| \leq N) \\ &\quad + \mathbf{E}(|f(\xi_n) - f(\xi)|; |\xi_n - \xi| \leq \delta, |\xi| > N) \\ &\quad + \mathbf{E}(|f(\xi_n) - f(\xi)|; |\xi_n - \xi| > \delta) \\ &\leq \frac{\epsilon}{2c} \mathbf{P}(|\xi_n - \xi| \leq \delta, |\xi| \leq N) + 2c \mathbf{P}(|\xi_n - \xi| \leq \delta, |\xi| > N) \\ &\quad + 2c \mathbf{P}(|\xi_n - \xi| > \delta) \\ &\leq \frac{\epsilon}{2} + 2c \mathbf{P}(|\xi| > N) + 2c \mathbf{P}(|\xi_n - \xi| > \delta) \\ &\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} + 2c \mathbf{P}(|\xi_n - \xi| > \delta) = \epsilon + 2c \mathbf{P}(|\xi_n - \xi| > \delta). \end{aligned}$$

---

<sup>4</sup> Vedi ad esempio **C. Miranda**: *Lezioni di Analisi Matematica - I*, Liguori, Napoli, 1967, p. 229. Notiamo, per la precisione, che la definizione di uniforme continuità richiederebbe che, fissato  $\epsilon > 0$ , si possa trovare un  $\delta' > 0$  tale che, comunque presi  $x, y \in [-N, N]$  con  $|x - y| < \delta'$ , risulti  $|f(x) - f(y)| < \epsilon$ . Noi invece ne usiamo una versione leggermente diversa nella quale solo ad  $x$  si richiede esplicitamente di cadere in  $[-N, N]$ . Si vede però nel nostro caso che, siccome  $f(x) \in \mathcal{C}(R)$ , fissato  $\epsilon$  e scelto l'opportuno  $\delta'$ ,  $f(x)$  risulta uniformemente continua anche in  $[-N - \delta', N + \delta']$ , e quindi potremo sempre trovare un altro  $\delta \leq \delta'$  tale che, se  $x, y \in [-N - \delta', N + \delta']$  con  $|x - y| < \delta$ , sia  $|f(x) - f(y)| < \epsilon$ . In tal caso se prendiamo  $x \in [-N, N]$  ed  $y$  tale che  $|x - y| < \delta \leq \delta'$ , abbiamo anche che  $x, y \in [-N - \delta', N + \delta']$  e quindi in ogni caso risulta  $|f(x) - f(y)| < \epsilon$ .



Siccome per ipotesi  $\mathbf{P}(|\xi_n - \xi| > \delta) \xrightarrow{n} 0$ , per  $n$  abbastanza grande si potrà sempre avere  $\mathbf{E}|f(\xi_n) - f(\xi)| \leq 2\epsilon$ : data l'arbitrarietà di  $\epsilon$  e il fatto ovvio che

$$|\mathbf{E}f(\xi_n) - \mathbf{E}f(\xi)| = |\mathbf{E}[f(\xi_n) - f(\xi)]| \leq \mathbf{E}|f(\xi_n) - f(\xi)|,$$

questo implica che  $\mathbf{E}f(\xi_n) \xrightarrow{n} \mathbf{E}f(\xi)$  e, data l'arbitrarietà di  $f(x)$ , ciò equivale a  $\xi_n \xrightarrow{d} \xi$ .

- (d) La conclusione del punto (c) non può, in generale, essere invertita: la convergenza in distribuzione non implica quella in probabilità. In un caso rilevante, però, questo è possibile: se  $(\xi_n)_{n \in \mathbf{N}}$  converge in distribuzione verso una costante  $C$ , allora essa vi converge anche in probabilità. Supponiamo infatti che  $\xi_n \xrightarrow{d} C$ : per definizione questo vuol dire che  $\mathbf{E}f(\xi_n) \xrightarrow{n} \mathbf{E}f(C)$ , comunque scelta la funzione  $f \in \mathcal{C}(\mathbf{R})$ . Dato  $\epsilon > 0$ , consideriamo allora in particolare la seguente funzione in  $\mathcal{C}(\mathbf{R})$ :

$$f(x) = \begin{cases} \frac{C-x}{\epsilon}, & x \in [C-\epsilon, C], \\ \frac{x-C}{\epsilon}, & x \in [C, C+\epsilon], \\ 1, & |x-C| > \epsilon. \end{cases}$$

Tale funzione risulta non negativa con  $f(C) = 0$ , per cui l'ipotesi di convergenza in distribuzione implica che

$$|\mathbf{E}f(\xi_n) - \mathbf{E}f(\xi)| = |\mathbf{E}f(\xi_n) - \mathbf{E}f(C)| = |\mathbf{E}f(\xi_n)| = \mathbf{E}f(\xi_n) \xrightarrow{n} 0.$$

Siccome inoltre, posto  $B = \overline{[C-\epsilon, C+\epsilon]}$ , si verifica immediatamente che  $I_B(x) \leq f(x)$ , risulta anche che

$$\mathbf{P}(|\xi_n - C| > \epsilon) = \mathbf{E}I_{\{|\xi_n - C| > \epsilon\}} = \mathbf{E}I_B(\xi_n) \leq \mathbf{E}f(\xi_n) \xrightarrow{n} 0$$

e quindi che  $\xi_n \xrightarrow{\mathbf{P}} \xi$ . □

**II.9.8 Esempio:** Mostriamo ora con qualche esempio che implicazioni diverse da quelle del Teorema II.9.7 non sono in generale garantite. Consideriamo innanzitutto uno spazio di probabilità in cui  $\Omega = [0, 1]$ ,  $\mathcal{F} = \mathcal{B}([0, 1])$  e  $\mathbf{P}$  sia la usuale misura di Lebesgue. Posto

$$A_n^k = \left[ \frac{k-1}{n}, \frac{k}{n} \right], \quad \xi_n^k = I_{A_n^k}; \quad k = 1, \dots, n; \quad n \in \mathbf{N},$$

si verifica subito che la successione

$$\xi_1^1; \xi_2^1, \xi_2^2; \dots; \xi_n^1, \dots, \xi_n^n; \dots$$

converge in probabilità e in media di ordine  $p$  verso  $\xi = 0$  in quanto

$$\mathbf{P}(|\xi_n^k - \xi| > \epsilon) = \mathbf{P}(\xi_n^k > \epsilon) = \frac{1}{n} \xrightarrow{n} 0, \quad \forall \epsilon > 0 (\epsilon \leq 1);$$

$$\mathbf{E}(|\xi_n^k - \xi|^p) = \mathbf{E}(\xi_n^k)^p = \frac{1}{n} \xrightarrow{n} 0;$$

ma è anche evidente che essa non converge a  $\xi = 0$  in nessuno dei punti  $\omega \in [0, 1]$  (sicché  $\mathbf{P}(\xi_n^k \neq 0) = 1$ ) dato che, comunque preso  $\omega$ , la successione numerica  $\xi_n^k(\omega)$  (che assume solo i valori 0 e 1) non si stabilizza mai definitivamente sul valore 0 in quanto vi saranno sempre elementi successivi che assumeranno valore 1. Sullo stesso spazio di probabilità si può anche considerare la seguente successione di v.a.:

$$\xi_n(\omega) = \begin{cases} e^n, & \omega \in [0, 1/n], \\ 0, & \omega \in (1/n, 1]. \end{cases}$$

Di questa si verifica facilmente la convergenza q.o. dato che, comunque fissato  $\omega > 0$ , da un certo  $n$  in poi la successione  $(\xi_n(\omega))_{n \in \mathbf{N}}$  è identicamente nulla; pertanto l'unico punto in cui non c'è convergenza è  $\omega = 0$  e quindi  $\mathbf{P}(\xi_n \not\rightarrow 0) = \mathbf{P}(\{0\}) = 0$  essendo  $\mathbf{P}$  la misura di Lebesgue. Segue allora dal Teorema II.9.7 che la successione  $(\xi_n)_{n \in \mathbf{N}}$  converge a zero anche in probabilità. Si vede però subito che essa non vi converge in media di ordine  $p$  in quanto  $\mathbf{E}|\xi_n|^p = e^{np}/n \xrightarrow{n} +\infty$  comunque scelto  $p > 0$ .

Sarà interessante a questo punto individuare delle ipotesi supplementari che garantiscano, se soddisfatte, anche implicazioni diverse da quelle espresse nel Teorema II.9.7: alcune fra le più note di tali condizioni supplementari sono raccolte nel successivo Teorema.  $\diamond$

**II.9.9 Teorema:**

- (1) Se  $\xi_n \xrightarrow{\mathbf{P}} \xi$ , allora esiste sempre una successione estratta  $(\xi_{n_k})_{k \in \mathbf{N}}$  tale che  $\xi_{n_k} \xrightarrow{\text{q.o.}} \xi$ ;
- (2) se  $\xi_n \xrightarrow{L^p} \xi$ , allora esiste sempre una successione estratta  $(\xi_{n_k})_{k \in \mathbf{N}}$  tale che  $\xi_{n_k} \xrightarrow{\text{q.o.}} \xi$ ;
- (3) se  $\xi_n \xrightarrow{\text{q.o.}} \xi$ , e se esiste una v.a.  $\eta \geq 0$  con  $\mathbf{E}|\eta| < +\infty$  tale che  $|\xi_n - \xi| < \eta$ , allora risulta anche  $\xi_n \xrightarrow{L^p} \xi$ .

**Dimostrazione:** Omessa (vedi ad esempio **A.N. Shiriyayev:** *Probability*; Springer, New York, 1984, p. 256, oppure **N. Bouleau:** *Probabilités de l'ingénieur*; Hermann, Paris, 1986; p. 164).  $\square$

**II.9.10 Definizione:** Una successione  $(\xi_n)_{n \in \mathbf{N}}$  di v.a. è una **successione di Cauchy P-q.o.**, o **in media di ordine  $p$** , o **in probabilità** se accade, rispettivamente, che  $|\xi_n - \xi_m| \xrightarrow{n,m} 0$  P-q.o., o  $\mathbf{E}|\xi_n - \xi_m|^p \xrightarrow{n,m} 0$ , o  $\mathbf{P}(|\xi_n - \xi_m| > \epsilon) \xrightarrow{n,m} 0$  comunque scelto  $\epsilon > 0$ .  $\triangle$

**II.9.11 Proposizione (criterio di Cauchy):** Condizione necessaria e sufficiente affinché una successione di v.a.  $(\xi_n)_{n \in \mathbf{N}}$  converga  $\mathbf{P}$ -q.o., o in probabilità, o (se  $\mathbf{E} |\xi_n|^p < +\infty$ ) in media di ordine  $p$  (verso una v.a.  $\xi$  con  $\mathbf{E} |\xi|^p < +\infty$ ), è che essa risulti essere, rispettivamente, una successione di Cauchy  $\mathbf{P}$ -q.o., o in probabilità, o in media di ordine  $p$ .

**Dimostrazione:** Omessa (vedi ad esempio **A.N. Shiriyayev:** *Probability*; Springer, New York, 1984, pp. 256-8). □

**II.9.12 Osservazione:** Denoteremo con il simbolo  $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbf{P})$  l'insieme delle v.a. definite su  $(\Omega, \mathcal{F}, \mathbf{P})$  con  $\mathbf{E} |\xi|^p < +\infty$ , con  $p > 0$ . Quando  $p \geq 1$  la disuguaglianza di Minkowski (Proposizione II.6.22) consente immediatamente di affermare che  $\mathcal{L}^p$  è uno spazio vettoriale. Inoltre, posto

$$\|\xi\|_p = (\mathbf{E} |\xi|^p)^{1/p},$$

si vede che, con  $c$  costante arbitraria,

$$\|\xi\|_p \geq 0, \quad \|c\xi\|_p = |c| \cdot \|\xi\|_p.$$

Pertanto l'applicazione  $\|\cdot\|_p$  definita su  $\mathcal{L}^p$  ha tutte le caratteristiche di una **norma**<sup>5</sup> eccetto una: per poter affermare che  $\|\cdot\|_p$  è una norma dovrebbe accadere che la condizione  $\|\xi\|_p = 0$  implichi che  $\xi = 0$ ,  $\forall \omega \in \Omega$ , mentre noi sappiamo (da (H) di Proposizione II.6.9) che  $\|\xi\|_p = 0$  implica soltanto  $\xi = 0$ ,  $\mathbf{P}$ -q.o. In queste condizioni si usa dire che  $\|\cdot\|_p$  definisce su  $\mathcal{L}^p$  una **seminorma**. Se vogliamo invece definire una vera norma sarà utile introdurre su  $\mathcal{L}^p$  una relazione d'equivalenza che considera equivalenti due vettori  $\xi$  e  $\eta$  di  $\mathcal{L}^p$  quando  $\xi = \eta$ ,  $\mathbf{P}$ -q.o. ed indicare con il simbolo  $L^p(\Omega, \mathcal{F}, \mathbf{P})$  lo spazio vettoriale delle classi d'equivalenza definite su  $\mathcal{L}^p$  dalla suddetta relazione<sup>6</sup>. Da ogni elemento di  $L^p$  potremo poi prendere una v.a. che funga da rappresentante della sua classe d'equivalenza (ad esempio la v.a. identicamente nulla per la classe delle v.a. nulle  $\mathbf{P}$ -q.o.) e, con una piccola forzatura, potremo considerare  $L^p$  ancora come uno spazio di funzioni piuttosto che come uno spazio di classi d'equivalenza. Conseguentemente, nel caso  $p \geq 1$ , da ora in poi considereremo senz'altro gli  $L^p(\Omega, \mathcal{F}, \mathbf{P})$  come spazi vettoriali normati. Sarà anche utile osservare che la citata disuguaglianza di Minkowski permette anche di ricavare immediatamente la cosiddetta disuguaglianza triangolare

$$\|\xi - \eta\|_p = \|(\xi - \zeta) + (\zeta - \eta)\|_p \leq \|\xi - \zeta\|_p + \|\zeta - \eta\|_p.$$

---

<sup>5</sup> Per i concetti di base sugli spazi vettoriali e normati sarà sufficiente consultare un buon testo di analisi funzionale come ad esempio **A.N. Kolmogorov e S.V. Fomin:** *Elementi di Teoria delle Funzioni e di Analisi Funzionale*; Mir, Mosca, 1980, cap. III, p. 119.

<sup>6</sup> Lasciamo come esercizio per il lettore la verifica del fatto che quella da noi definita è una relazione d'equivalenza e che  $L^p$  è uno spazio vettoriale.

Infine, tenendo conto dei risultati del Teorema II.9.11, potremo anche affermare che ogni  $L^p$  è uno spazio normato completo (nel senso che ogni successione di Cauchy, secondo la norma, di elementi di  $L^p$  converge, secondo la norma, verso un elemento di  $L^p$ ) e quindi, come suol dirsi, è uno **spazio di Banach**<sup>7</sup>.  $\circ$

**II.9.13 Osservazione:** Nel caso in cui  $0 < p < 1$  la disuguaglianza di Minkowski non è più garantita e quindi gli spazi  $L^p$  non possono più essere strutturati né come spazi normati né come spazi vettoriali. Ciononostante è possibile mostrare che essi possono essere dotati di una **metrica**<sup>8</sup> (non dedotta, questa volta, da una norma) definita come

$$d(\xi, \eta) = \mathbf{E} |\xi - \eta|^p.$$

Si verifica facilmente, infatti, che tale definizione soddisfa le proprietà richieste ad una metrica. In particolare, per verificare la disuguaglianza triangolare (comunque

---

<sup>7</sup> Stefan Banach (1892-1945) era figlio di un ferroviere di Cracovia, città polacca che a quell'epoca faceva parte dell'Impero Austro-Ungarico. Abbandonato dai genitori fu allevato da una lavandaia dalla quale prese il nome e fin dall'età di 16 anni dovette guadagnarsi da vivere dando lezioni private. Durante i suoi studi al Politecnico le sue attitudini per la matematica furono scoperte casualmente (1916) e valorizzate da H. Steinhaus con il quale egli iniziò una collaborazione ininterrotta. Nel 1920 si trasferì al Politecnico di Leopoli (città che all'epoca si trovava entro i confini della Polonia uscita dalla prima guerra mondiale e che attualmente, col nome di L'vov si trova in Ucraina) dove, divenuto professore nel 1922, rimase per il resto della sua vita. La sua attività scientifica fu fecondissima anche se le sue pubblicazioni rimasero in numero piuttosto limitato (56 memorie originali più alcune monografie tra cui ricordiamo *Théorie des opérations linéaires*, 1932) a causa della sua riluttanza a mettere per iscritto in forma organica i propri contributi. In collaborazione con Steinhaus fondò e diresse la rivista *Studia Mathematica* che ancora oggi si pubblica in Polonia. Uomo di sana e robusta vitalità, alieno dalle arti e dalle lettere e senza interessi politici, egli non rispecchiava l'immagine dello scienziato ascetico: la sede preferita per le sue elaborazioni era, infatti, il caffè o l'osteria dove lavorava per ore con colleghi e allievi. Al Café Scotland di Leopoli era depositato presso il capocameriere un quaderno, il cosiddetto *Taccuino scozzese*, che, a richiesta, veniva fornito ai clienti e che conteneva i problemi aperti e, quando si trovavano, le soluzioni. Esso risulta scritto principalmente da Banach, S. Ulam, S. Mazur e altri matematici di Leopoli, ma contiene anche contributi di matematici quali M.-R. Fréchet o J. von Neumann. Banach e la sua scuola sono considerati i fondatori della moderna Analisi Funzionale. Partite con i lavori di V. Volterra, I. Fredholm e D. Hilbert all'inizio del secolo, le idee della nuova teoria erano nell'aria dagli anni venti e furono sistematizzate quasi contemporaneamente dai lavori di N. Wiener, E. Helly, H. Hahn e Banach. Fra questi Banach e Wiener adottarono un approccio più astratto che permetteva di unificare le più diverse trattazioni e di derivare teoremi in modo sorprendentemente rapido. Gli sviluppi nel campo dell'analisi funzionale che si fondarono sull'opera di Banach sono stati imponenti sicché essa è divenuta un'entità teorica con estese ramificazioni inseparabile dall'analisi matematica convenzionale. Allo scoppio della seconda guerra mondiale Leopoli fu annessa all'URSS e subito dopo, nel 1941, fu invasa dai tedeschi: le autorità naziste eliminarono ogni vita culturale nella città e fucilarono immediatamente più di trenta scienziati polacchi. Banach si salvò, ma non svolse più alcun lavoro scientifico: trovò un lavoro in un istituto di vaccinazione e sopravvisse alla guerra per morire poi subito dopo per un cancro polmonare.

<sup>8</sup> Per le nozioni fondamentali sugli spazi metrici vedi **A.N. Kolmogorov e S.V. Fomin:** *Elementi di Teoria delle Funzioni e di Analisi Funzionale*; Mir, Mosca, 1980, cap. II, pag. 50.

scelti  $\xi, \eta, \zeta$ )

$$\mathbf{d}(\xi, \eta) \leq \mathbf{d}(\xi, \zeta) + \mathbf{d}(\zeta, \eta),$$

basterà osservare che la funzione

$$F(x) = (a+x)^p - (a^p + x^p), \quad a, x \geq 0, \quad 0 < p < 1,$$

risulta sempre decrescente con  $F(0) = 0$ , sicché  $F(x) \leq 0$  per  $x \geq 0$  e quindi

$$(a+b)^p \leq a^p + b^p, \quad a, b \geq 0, \quad 0 < p < 1,$$

da cui segue immediatamente che

$$\begin{aligned} \mathbf{d}(\xi, \eta) &= \mathbf{E} |\xi - \eta|^p \leq \mathbf{E} (|\xi - \zeta| + |\zeta - \eta|)^p \leq \mathbf{E} |\xi - \zeta|^p + \mathbf{E} |\zeta - \eta|^p \\ &= \mathbf{d}(\xi, \zeta) + \mathbf{d}(\zeta, \eta). \end{aligned}$$

Anche tali spazi, inoltre, risultano completi rispetto alla convergenza definita mediante la loro metrica. ○

**II.9.14 Osservazione:** Gli spazi  $L^p$  con  $p \geq 1$  sono dunque degli spazi di Banach, ma per essi si intende sempre che  $p < +\infty$ , cioè che  $p$  abbia un valore finito. Se però chiamiamo **estremo superiore essenziale** di una v.a.  $\xi$  il numero

$$\text{ess sup}|\xi| = \inf\{0 \leq c \leq +\infty : \mathbf{P}(|\xi| > c) = 0\},$$

cioè l'estremo inferiore (eventualmente  $+\infty$ ) dei numeri  $c$  tali che  $|\xi| \leq c$ , **P**-q.o., è interessante notare che l'applicazione

$$\|\xi\|_\infty = \text{ess sup}|\xi|$$

definisce una norma sullo spazio<sup>9</sup>  $L^\infty(\Omega, \mathcal{F}, \mathbf{P})$  delle v.a. con estremo superiore essenziale finito. Tale spazio risulta, inoltre, completo, cioè è anch'esso uno spazio di Banach. L'uso dei simboli  $\|\cdot\|_\infty$  e  $L^\infty$  è poi giustificato dalle seguenti osservazioni: dalla disuguaglianza di Lyapunov (Proposizione II.6.20) si deduce immediatamente che, per  $1 \leq p \leq q < +\infty$ , si ha  $\|\xi\|_1 \leq \|\xi\|_p \leq \|\xi\|_q$ . Ne segue che, dati  $1 \leq p \leq q$ , se  $\xi \in L^q$  anche  $\xi \in L^p$ , e inoltre se una successione converge in  $L^q$  essa converge anche in  $L^p$ , sicché

$$L^1 \supseteq L^p \supseteq L^q, \quad 1 \leq p \leq q < +\infty.$$

Si può però anche mostrare che

$$\|\xi\|_p \leq \|\xi\|_\infty, \quad 1 \leq p < +\infty.$$

---

<sup>9</sup> Per la precisione  $L^\infty$  è lo spazio delle classi d'equivalenza di v.a. rispetto alla relazione che considera equivalenti le v.a. quasi ovunque eguali: per brevità non abbiamo ripreso qui questa discussione per la quale rimandiamo all'Osservazione II.9.12.

Infatti se  $\|\xi\|_\infty = +\infty$  la disuguaglianza è ovvia; se invece  $\|\xi\|_\infty < +\infty$ , dalla definizione di estremo superiore essenziale si ricava che l'insieme numerico  $S = \{0 \leq c < +\infty : \mathbf{P}(|\xi| > c) = 0\}$  è non vuoto. In tal caso, comunque scelto  $c \in S$  si avrà  $\mathbf{P}(|\xi| > c) = 0$  e quindi  $\mathbf{E}(|\xi|^p I_{\{|\xi| > c\}}) = 0$ , sicché

$$\begin{aligned} \mathbf{E} |\xi|^p &= \mathbf{E} (|\xi|^p I_{\{|\xi| \leq c\}}) + \mathbf{E} (|\xi|^p I_{\{|\xi| > c\}}) = \mathbf{E} (|\xi|^p I_{\{|\xi| \leq c\}}) \\ &\leq c^p \mathbf{E} (I_{\{|\xi| \leq c\}}) \leq c^p. \end{aligned}$$

Ne consegue che

$$\|\xi\|_p = (\mathbf{E} |\xi|^p)^{1/p} \leq c, \quad \forall c \in S,$$

e quindi, essendo  $\|\xi\|_\infty = \inf S$ , anche che  $\|\xi\|_p \leq \|\xi\|_\infty$ . Pertanto la catena di disuguaglianze ed inclusioni ricavata dalla disuguaglianza di Lyapunov può essere estesa anche al caso  $p = +\infty$ : se  $1 \leq p \leq q \leq +\infty$  risulta

$$\|\xi\|_1 \leq \|\xi\|_p \leq \|\xi\|_q \leq \|\xi\|_\infty$$

e quindi

$$L^1 \supseteq L^p \supseteq L^q \supseteq L^\infty$$

giustificando così anche le notazioni adottate. In conclusione potremo affermare che gli spazi  $L^p(\Omega, \mathcal{F}, \mathbf{P})$  con  $0 < p \leq +\infty$  sono tutti spazi metrici completi; se poi  $1 \leq p \leq +\infty$  la metrica viene realizzata tramite la norma definendola come  $\mathbf{d}(\xi, \eta) = \|\xi - \eta\|_p$  e gli spazi  $L^p$  sono tutti spazi di Banach.  $\circ$

**II.9.15 Osservazione:** Fra gli spazi di Banach  $L^p$  con  $p \geq 1$  un ruolo particolarmente rilevante è giocato dal caso  $p = 2$ , cioè dallo spazio  $L^2(\Omega, \mathcal{F}, \mathbf{P})$ : è facile infatti controllare che la norma  $\|\cdot\|_2$  (che d'ora in poi indicheremo semplicemente con  $\|\cdot\|$ ) può essere realizzata tramite il **prodotto scalare**<sup>10</sup>

$$(\xi, \eta) = \mathbf{E} (\xi\eta)$$

che, come si verifica facilmente, soddisfa tutte le proprietà necessarie. La connessione tra la norma ed il prodotto scalare in  $L^2$  è poi data da

$$\|\xi\| = \sqrt{(\xi, \xi)} = \sqrt{\mathbf{E} \xi^2}.$$

Siccome  $L^2$  è anche uno spazio completo rispetto alla sua norma esso risulta anche essere uno **spazio di Hilbert**<sup>11</sup>. L'introduzione del concetto di prodotto scalare consente di usare in probabilità non solo ben noti metodi dell'analisi funzionale, ma

<sup>10</sup> Per i concetti relativi ai prodotti scalari e ai cosiddetti Spazi Euclidei, cioè gli spazi vettoriali dotati di prodotto scalare, vedi ad esempio **A.N. Kolmogorov e S.V. Fomin: *Elementi di Teoria delle Funzioni e di Analisi Funzionale***; Mir, Mosca, 1980, p. 142.

<sup>11</sup> Per i concetti relativi agli spazi di Hilbert vedi **A.N. Kolmogorov e S.V. Fomin: *Elementi di Teoria delle Funzioni e di Analisi Funzionale***; Mir, Mosca, 1980, p. 151. David

anche idee e concetti mutuati dalla geometria come sarà mostrato nelle osservazioni successive. Diremo allora che due v.a.  $\xi, \eta \in L^2$  sono **ortogonali** quando  $(\xi, \eta) = \mathbf{E}(\xi\eta) = 0$  e diremo che una famiglia di v.a. di  $L^2$  forma un **sistema ortogonale** quando comunque prese in essa due v.a. distinte esse risultano ortogonali. Se inoltre in un sistema ortogonale risulta anche  $\|\xi\| = 1$  per ogni elemento, si dice che la famiglia forma un **sistema ortonormale**. Notiamo infine che, se due v.a. sono non correlate si ha

$$(\xi, \eta) = \mathbf{E}(\xi\eta) = \mathbf{E}\xi \cdot \mathbf{E}\eta$$

per cui esse risulteranno ortogonali se almeno una delle due ha VdA nullo. ○

**II.9.16 Osservazione:** Riprendiamo, nel contesto di una teoria  $L^2$ , la discussione del problema della **stima lineare in m.q.** già affrontato nell'Osservazione II.8.6. Dati in  $L^2(\Omega, \mathcal{F}, \mathbf{P})$  un sistema ortonormale di v.a.  $\mathcal{M} = \{\eta_1, \dots, \eta_n\}$  e una generica v.a.  $\xi$ , poniamoci il problema di determinare il miglior stimatore lineare (in m.q., cioè secondo la metrica determinata dalla norma di  $L^2$ ) di  $\xi$  in termini delle  $\eta_k$ : si tratta cioè di determinare i numeri reali  $a_k$  di  $\sum_k a_k \eta_k$  in modo tale che questa combinazione lineare sia la più vicina possibile a  $\xi$  secondo la norma in  $L^2$  (cioè in m.q.). Siccome a causa della ortonormalità delle v.a. di  $\mathcal{M}$  (cioè del fatto che  $(\eta_k, \eta_l) = \delta_{kl} = 1$  oppure  $0$  secondo che  $k = l$  o  $k \neq l$ ) per ogni scelta

---

Hilbert (1862-1943) era nato a Königsberg (attualmente in Russia, ma all'epoca parte della Prussia Orientale) in una famiglia benestante che avrebbe voluto avviarlo agli studi giuridici. Egli però nel 1880 optò per gli studi matematici e si laureò nella sua città natale nel 1885. Intraprese subito dopo una serie di viaggi di studio all'estero che lo condussero a contatto con H. Minkowski, F. Klein, H. Poincaré, K. Weierstraß e L. Kronecker, per citarne solo alcuni. Iniziò allora la sua prodigiosa attività in numerosi campi della ricerca: teoria degli invarianti algebrici, fondamenti della geometria, logica, teoria dei numeri, calcolo delle variazioni, equazioni integrali e alle derivate parziali. Egli divenne professore a Königsberg nel 1894 e subito dopo, nel 1895, a Göttingen dove rimase per il resto della sua vita. All'inizio del nuovo secolo Hilbert aveva rinnovato profondamente diversi settori della matematica senza però perdere mai un punto di vista complessivo sui problemi trattati. Nella relazione *Matematische Probleme* che lesse al II Congresso Internazionale di Matematica (Parigi, 1900) egli elencò ventitré problemi non risolti, corrispondenti ad altrettanti settori di ricerca che costituirono alcuni dei principali obiettivi degli sforzi successivi stimolando la scoperta di nuovi metodi e nuove teorie. A Göttingen, intanto, egli e i suoi collaboratori avevano creato un attivissimo centro di ricerca matematica che proseguì la sua attività anche nel periodo della prima guerra mondiale anche se il progressivo isolamento della Germania e il crescente clima nazionalistico furono di ostacolo alla cooperazione internazionale. Ancora negli anni venti il suo impegno nella ricerca rimase notevole e i lavori suoi e dei suoi allievi (M. Born ed H. Weyl) influenzarono profondamente anche la nascita e la sistematizzazione della Meccanica Quantistica. Ma con la crisi della Repubblica di Weimar e l'ascesa del Nazionalsocialismo (1933) iniziò, con il grande esodo, il declino dei centri di cultura tedeschi, e innanzitutto di Göttingen. Stanco, malato e amareggiato Hilbert trascorse gli ultimi anni della sua vita in una solitudine sempre più tragica, nonostante i riconoscimenti in patria e all'estero. Egli si spense nel 1943 per i postumi di una caduta: poco dopo i russi entravano in Königsberg, l'antica città dei Cavalieri Teutonici quasi totalmente distrutta, mentre il Terzo Reich si avviava verso la sua fine.

delle  $a_k$  risulta

$$\begin{aligned}
 \left\| \xi - \sum_{k=1}^n a_k \eta_k \right\|^2 &= \mathbf{E} \left| \xi - \sum_{k=1}^n a_k \eta_k \right|^2 = \left( \xi - \sum_{k=1}^n a_k \eta_k, \xi - \sum_{l=1}^n a_l \eta_l \right) \\
 &= \|\xi\|^2 - 2 \sum_{k=1}^n a_k (\xi, \eta_k) + \sum_{k,l=1}^n a_k a_l (\eta_k, \eta_l) \\
 &= \|\xi\|^2 - 2 \sum_{k=1}^n a_k (\xi, \eta_k) + \sum_{k=1}^n a_k^2 \\
 &= \|\xi\|^2 - \sum_{k=1}^n |(\xi, \eta_k)|^2 + \sum_{k=1}^n |a_k - (\xi, \eta_k)|^2 \\
 &\geq \|\xi\|^2 - \sum_{k=1}^n |(\xi, \eta_k)|^2,
 \end{aligned}$$

è facile accorgersi che il valore minimo dell'e.q.m.  $\Delta = \|\xi - \sum_k a_k \eta_k\|^2$  (vedi II.8.6) viene raggiunto quando  $a_k = (\xi, \eta_k)$ , con  $k = 1, \dots, n$ . Pertanto il miglior stimatore lineare (in m.q.) di  $\xi$  mediante i vettori di  $\mathcal{M}$  è la v.a.

$$\tilde{\xi} = \sum_{k=1}^n (\xi, \eta_k) \eta_k$$

e l'errore quadratico medio che si commette con questa scelta è, come si deduce dal calcolo precedente,

$$\tilde{\Delta} = \mathbf{E} |\xi - \tilde{\xi}|^2 = \|\xi\|^2 - \sum_{k=1}^n |(\xi, \eta_k)|^2.$$

Questo risultato generalizza quello più semplice ottenuto in Osservazione II.8.6 sulla miglior approssimazione lineare di una v.a. tramite una sola v.a.  $\eta$ . Ricordiamo che la miglior stima *lineare* non coincide, in generale, con la miglior stima in media quadratica che, in generale, risulta essere una funzione *non lineare* delle v.a. di  $\mathcal{M}$ . Per tenere conto delle analogie e delle differenze dei due casi si usa spesso indicare il miglior stimatore lineare con il simbolo

$$\tilde{\xi} = \tilde{\mathbf{E}}(\xi | \eta_1, \dots, \eta_n)$$

che prende anche il nome di **valor d'attesa condizionato in senso lato** per contrapporlo al ben noto simbolo  $\mathbf{E}(\xi | \eta_1, \dots, \eta_n)$  del VdA condizionato (in senso stretto) che rappresenta, come enunciato dal Teorema II.8.7, la miglior stima (non lineare) in  $L^2$  di  $\xi$  tramite le v.a. di  $\mathcal{M}$ . In altri termini, mentre in  $\mathbf{E}(\xi | \eta_1, \dots, \eta_n) = f^*(\eta_1, \dots, \eta_n)$  la  $f^*$  è in generale una funzione non lineare che fornisce il miglior stimatore (in assoluto) di  $\xi$  tramite  $\eta_k$ , in  $\tilde{\mathbf{E}}(\xi | \eta_1, \dots, \eta_n) =$



$\sum_k (\xi, \eta_k) \eta_k = \tilde{f}(\eta_1, \dots, \eta_n)$  la  $\tilde{f}$  è la combinazione lineare che fornisce il miglior stimatore lineare di  $\xi$  tramite  $\eta_k$ . Osserviamo infine che a questa soluzione del problema della stima lineare può essere dato anche un ben preciso significato geometrico noto come **Principio di Ortogonalità**: se con  $M$  indichiamo il sottospazio di  $L^2$  sviluppato dal sistema ortonormale  $\mathcal{M}$  (cioè il sottospazio che contiene tutte le combinazioni lineari di elementi di  $\mathcal{M}$ ), la precedente discussione mostra che ogni  $\xi \in L^2$  ammette, in modo unico, la seguente decomposizione ortogonale:

$$\xi = \tilde{\xi} + (\xi - \tilde{\xi}) = \tilde{\xi} + \xi'$$

dove  $\tilde{\xi} \in M$ , mentre  $\xi' = \xi - \tilde{\xi}$  è ortogonale ad  $M$  (nel senso che è ortogonale a tutti gli elementi di  $M$  come si verifica facilmente con un calcolo diretto tenendo conto della definizione di  $\tilde{\xi}$ ). È pertanto naturale dire che  $\tilde{\xi}$  rappresenta la **proiezione ortogonale** di  $\xi$  su  $M$  e quindi che la miglior stima lineare in m.q. di  $\xi$  tramite  $\mathcal{M}$  è fornita proprio dalla sua proiezione ortogonale sul sottospazio  $M$ .  $\circ$

**II.9.17 Osservazione:** Nell'Osservazione II.9.16 l'ipotesi che  $\mathcal{M}$  fosse un sistema ortonormale giocava un ruolo importante per semplificare la determinazione del miglior stimatore lineare permettendo di individuarlo come la proiezione ortogonale  $\tilde{\xi}$  di  $\xi$  su  $M$ . Tuttavia il problema della stima lineare può essere affrontato e risolto anche senza quest'ipotesi di ortonormalità, benché a costo di una certa complicazione dei calcoli. È conveniente innanzitutto introdurre il seguente concetto: diremo che le v.a. della famiglia  $\mathcal{M} = \{\eta_1, \dots, \eta_n\}$  di  $L^2$  sono **linearmente indipendenti**<sup>12</sup> se l'equazione

$$\sum_{k=1}^n a_k \eta_k = 0, \quad \mathbf{P}\text{-q.o.}$$

è soddisfatta solo quando le  $a_k$  sono tutte nulle. È facile mostrare che un sistema ortonormale di v.a. è composto di v.a. linearmente indipendenti: il viceversa, invece, non è sempre vero. Ciò posto, vogliamo far vedere che il problema della miglior stima lineare (in m.q.) di  $\xi$  mediante un sistema  $\mathcal{M}$  di v.a. linearmente indipendenti (ma non, in generale, ortonormali) può sempre essere ricondotto ad un problema di stima lineare mediante un opportuno sistema ortonormale di v.a. Infatti, come notato alla fine dell'Osservazione II.9.16, la soluzione del problema della miglior stima lineare consiste nella ricerca della proiezione ortogonale di  $\xi$  sul sottospazio  $M$  generato da  $\mathcal{M}$ . Tale sottospazio  $M$  (e quindi ogni proiezione ortogonale su di esso), comunque, può essere generato da più di un sistema di vettori di  $L^2$ . In particolare mostreremo ora che, comunque dato in  $L^2$  un sistema  $\mathcal{M} = \{\eta_1, \dots, \eta_n\}$  di v.a. linearmente indipendenti sarà sempre possibile determinare un sistema ortonormale di v.a.  $\mathcal{L} = \{\zeta_1, \dots, \zeta_n\}$  tale che i sottospazi

---

<sup>12</sup> Val la pena di osservare qui che, nonostante il nome possa ingenerare confusione, il concetto di *indipendenza lineare* di v.a. nello spazio di Hilbert  $L^2$  non coincide affatto con quello di *indipendenza statistica* (o *indipendenza tout-court*) da noi introdotto in II.5.35.

generati dai due sistemi coincidano:  $M(\eta_1, \dots, \eta_n) = L(\zeta_1, \dots, \zeta_n)$ . Per mostrarlo consideriamo, per semplicità, il caso in cui  $\mathbf{E} \eta_k = 0$  per tutte le v.a. di  $\mathcal{M}$ , sicché la matrice delle covarianze (vedi Definizione II.8.1) avrà la forma

$$\mathbf{R} = \mathbf{E}(YY^T), \quad r_{kl} = \mathbf{E}(\eta_k \eta_l)$$

dove  $Y = (\eta_1, \dots, \eta_n)$  è il vett.a. di componenti  $\eta_k$ . Sappiamo (vedi la dimostrazione di Proposizione II.8.4) che  $\mathbf{R}$  può essere diagonalizzata da una matrice ortogonale  $\mathbf{O}$  mediante la relazione

$$\mathbf{O}^T \mathbf{R} \mathbf{O} = \mathbf{D} = \begin{pmatrix} d_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & d_n \end{pmatrix}$$

I numeri  $d_k$  coincidono con gli **autovalori** di  $\mathbf{R}$ , cioè con le radici dell'equazione  $|\mathbf{R} - \lambda \mathbf{I}| = 0$ , dove  $|\cdot|$  indica il determinante di una matrice. Si prova facilmente<sup>13</sup> che, se le  $\eta_k$  sono linearmente indipendenti, allora  $|\mathbf{R}| \neq 0$  e, di conseguenza,  $d_k > 0$  (ricordiamo che  $\mathbf{R}$  è anche definita non negativa: vedi Osservazione II.8.3). Potremo allora definire prima la matrice

$$\mathbf{B} = \begin{pmatrix} \sqrt{d_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sqrt{d_n} \end{pmatrix}$$

in modo tale che  $\mathbf{D} = \mathbf{B}\mathbf{B}$ , e poi il vett.a.

$$\mathbf{Z} = \mathbf{B}^{-1} \mathbf{O}^T \mathbf{Y}$$

e verificare che le sue componenti  $\zeta_k$  sono ortonormali: infatti

$$\begin{aligned} \mathbf{E}(\mathbf{Z}\mathbf{Z}^T) &= \mathbf{B}^{-1} \mathbf{O}^T \mathbf{E}(YY^T) \mathbf{O} \mathbf{B}^{-1} = \mathbf{B}^{-1} \mathbf{O}^T \mathbf{R} \mathbf{O} \mathbf{B}^{-1} \\ &= \mathbf{B}^{-1} \mathbf{D} \mathbf{B}^{-1} = \mathbf{B}^{-1} \mathbf{B} \mathbf{B} \mathbf{B}^{-1} = \mathbf{I}, \end{aligned}$$

cioè  $\mathbf{E}(\zeta_k, \zeta_l) = \delta_{kl}$ . Le componenti di  $\mathbf{Z}$ , che sono combinazioni lineari delle componenti di  $\mathbf{Y}$ , costituiscono allora il sistema ortonormale  $\mathcal{L}$  che, ovviamente, genera il medesimo sottospazio  $M$  generato da  $\mathcal{M}$ , e pertanto, se la stima lineare di  $\xi$  mediante i vettori di  $\mathcal{M}$  consiste nella ricerca della proiezione ortogonale  $\tilde{\xi}$  di  $\xi$  su  $M$ , dato che  $M$  è anche generato da  $\mathcal{L}$  tale problema equivale a quello della stima lineare di  $\xi$  mediante il sistema ortonormale  $\mathcal{L} = \{\zeta_1, \dots, \zeta_n\}$ .  $\circ$

---

<sup>13</sup> Infatti, dalla relazione che definisce l'indipendenza lineare delle  $\eta_k$  e dalla definizione della matrice  $r_{jk}$  si ricava facilmente che il sistema di equazioni lineari  $\sum_k r_{jk} a_k = 0$  deve avere solo la soluzione banale e questo richiede che il determinante dei coefficienti sia diverso da zero.

## II.10 Funzioni caratteristiche

**II.10.1 Osservazione:** Come già accennato nell'Esempio II.5.26 è possibile considerare v.a.  $\zeta$  che prendono valori complessi ed estendere conseguentemente molte delle definizioni e proprietà delle v.a. reali. Le v.a. complesse sono costituite da una parte reale ed una immaginaria, che sono a loro volta delle v.a. reali, secondo l'espressione

$$\zeta = \xi + i\eta.$$

Diremo inoltre che il VdA  $\mathbf{E}\zeta$  esiste se esistono  $\mathbf{E}\xi$  e  $\mathbf{E}\eta$  e in tal caso porremo

$$\mathbf{E}\zeta = \mathbf{E}\xi + i\mathbf{E}\eta.$$

Due v.a. complesse  $\zeta_1 = \xi_1 + i\eta_1$  e  $\zeta_2 = \xi_2 + i\eta_2$  sono indipendenti se lo sono i vett.a.  $(\xi_1, \eta_1)$  e  $(\xi_2, \eta_2)$ : la definizione si estende poi facilmente a qualunque numero di v.a. complesse. Lo spazio  $L^2(\Omega, \mathcal{F}, \mathbf{P})$  di v.a. complesse sarà poi definito come lo spazio di Hilbert delle  $\zeta$  tali che  $\mathbf{E}|\zeta|^2 < +\infty$ , dove  $|\zeta|^2 = \xi^2 + \eta^2$ ; in tale spazio il prodotto scalare sarà definito come  $(\zeta_1, \zeta_2) = \mathbf{E}(\bar{\zeta}_1, \zeta_2)$  se  $\bar{\zeta}_1 = \xi - i\eta$  è la v.a. complessa coniugata di  $\zeta = \xi + i\eta$ . Le altre definizioni e proprietà si estendono in maniera naturale.  $\circ$

**II.10.2 Definizione:** Chiameremo **Funzione caratteristica (f.c.)** del vett.a.  $X = (\xi_1, \dots, \xi_n)$  la funzione

$$\varphi_X(t) = \varphi_X(t_1, \dots, t_n) = \mathbf{E}(e^{i(t, X)}); \quad t \in \mathbf{R}^n$$

cioè il VdA della v.a. complessa  $\zeta = e^{i(t, X)}$ , dove  $(t, X) = \sum_k t_k \xi_k$ .  $\triangle$

**II.10.3 Osservazione:** Esplicitamente la f.c.  $\varphi_X(t)$  può essere scritta come

$$\varphi_X(t) = \int_{\mathbf{R}^n} e^{i(t, x)} F_X(d^n x) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} e^{i(t, x)} F_X(dx_1 \dots dx_n), \quad t \in \mathbf{R}^n$$

Ne consegue che la f.c. di  $X$  è determinata anche soltanto dalla conoscenza della FdD congiunta  $F_X(x_1 \dots x_n)$ ; se poi  $X$  è dotato di densità si ha

$$\varphi_X(t) = \int_{\mathbf{R}^n} e^{i(t, x)} f_X(x) d^n x = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} e^{i(t, x)} f_X(x_1 \dots x_k) dx_1 \dots dx_n.$$

Nel caso in cui il vett.a.  $X$  si riduce ad una sola componente  $\xi$  la f.c. diviene

$$\varphi_\xi(t) = \mathbf{E}(e^{it\xi}) = \int_{\mathbf{R}} e^{itx} F_\xi(dx), \quad t \in \mathbf{R},$$

e se  $\xi$  è dotata di densità

$$\varphi_\xi(t) = \int_{-\infty}^{+\infty} e^{itx} f_\xi(x) dx, \quad t \in \mathbf{R}.$$

Come si vede, dunque, la f.c. di una v.a. (o di un vett.a.) non è altro che la **Trasformata di Fourier**<sup>1</sup> della sua FdD o della sua fdd.

Esamineremo ora le prime semplici proprietà delle f.c. Se  $\xi$  è una v.a. con f.c.  $\varphi_\xi(t)$ , e se definiamo la v.a.  $\eta = a\xi + b$ , dove  $a$  e  $b$  sono numeri, si ottiene facilmente che

$$\varphi_\eta(t) = \mathbf{E}(e^{it\eta}) = e^{itb} \mathbf{E}(e^{i(at)\xi}) = e^{itb} \varphi_\xi(at).$$

Se invece  $\xi_1, \dots, \xi_n$  sono  $n$  v.a. indipendenti e se poniamo  $S^{(n)} = \xi_1 + \dots + \xi_n$ , si vede subito che

$$\begin{aligned} \varphi_{S^{(n)}}(t) &= \mathbf{E}(e^{itS^{(n)}}) = \mathbf{E}(e^{it\xi_1} \dots e^{it\xi_n}) \\ &= \mathbf{E}(e^{it\xi_1}) \dots \mathbf{E}(e^{it\xi_n}) = \varphi_{\xi_1}(t) \cdot \dots \cdot \varphi_{\xi_n}(t). \end{aligned}$$

Questa proprietà sarà particolarmente utile per la dimostrazione dei Teoremi limite: notiamo infatti che, mentre (come visto in Osservazione II.8.14) la densità  $f_{S^{(n)}}(x)$  della somma di  $n$  v.a. indipendenti è la convoluzione  $f_{\xi_1}(x) * \dots * f_{\xi_n}(x)$  delle densità delle singole v.a., la f.c.  $\varphi_{S^{(n)}}(t)$  è il prodotto nel senso ordinario  $\varphi_{\xi_1}(t) \cdot \dots \cdot \varphi_{\xi_n}(t)$  delle corrispondenti f.c. Questo mostra chiaramente i vantaggi che si hanno lavorando con le f.c.: siccome infatti le convoluzioni sono operazioni integrali, esse risultano più difficili da usare rispetto ai prodotti ordinari.  $\circ$

---

<sup>1</sup> I concetti relativi alla Trasformazione di Fourier possono essere trovati in vari testi di Analisi e di Analisi Funzionale; ad esempio si veda **A.N. Kolmogorov e S.V. Fomin: *Elementi di Teoria delle Funzioni e di Analisi Funzionale***; Mir, Mosca, 1980, cap. VIII, p. 400 e seg. Jean-Baptiste Joseph Fourier (1768-1830), di umili origini, ebbe un'eccellente formazione nella scuola reale militare di Auxerre, sua città natale, dove successivamente insegnò ed iniziò le sue prime ricerche. La rivoluzione del 1789 segnò profondamente la sua vita e la sua carriera: favorevole alle nuove idee egli era stato nominato membro del Comité révolutionnaire della sua città, ma la sua difesa delle vittime del Terrore gli attirarono odii e persecuzioni. Dopo la caduta di Robespierre (agosto 1794) egli riprese la sua attività scientifica a Parigi dove fu notato da P.S. de Laplace, G.L. Lagrange e G. Monge il quale gli fece ottenere (1795) un posto di professore all'École Polytechnique. Nel 1798 partecipò alla spedizione d'Egitto e fu incaricato di raccogliere il materiale necessario per un'opera collettiva su quel paese. Rientrato nel 1801 fu nominato da Napoleone prefetto dell'Isère; contemporaneamente coordinò la redazione della progettata *Description de l'Égypte*, opera monumentale per la quale egli stesso scrisse una *Préface historique*. Tale attività non gli impedì di continuare il suo lavoro scientifico: dedicatosi allo studio della propagazione del calore, egli presentò nel 1807 all'Académie des Sciences una memoria sulla soluzione dell'equazione di diffusione in cui per la prima volta rappresentò alcune funzioni arbitrarie sotto forma di sviluppi in serie trigonometriche, ma la pubblicazione di queste idee fu ritardata dalle obiezioni di Lagrange e Poisson. La caduta di Napoleone lo allontanò ben presto dalla vita politica attiva permettendogli di dedicarsi ai suoi studi: fu proprio nel periodo 1816-1830 infatti che Fourier rivelò progressivamente i suoi importanti risultati dando alle stampe la *Théorie analytique de la chaleur* (1822) e le sue precedenti memorie sulla diffusione del calore. Egli si dedicò anche a ricerche di analisi infinitesimale, di algebra e di calcolo delle probabilità esercitando inoltre, nella sua qualità di *secrétaire perpétuel* dell'Académie des Sciences, una profonda influenza su molti giovani matematici di valore come J.P. Dirichlet e J.C. Sturm.

**II.10.4 Esempio:** Considereremo ora alcuni esempi di calcolo di f.c. di v.a. discrete. Se  $\xi = m$  ( $\mathbf{P}$ -q.o.) è una v.a. q.o. costante si vede immediatamente che, essendo  $\mathbf{P}(\xi = m) = 1$  e  $\mathbf{P}(\xi \neq m) = 0$ , risulta

$$\varphi_{\xi}(t) = \mathbf{E}(e^{imt}) = e^{imt}.$$

Se invece  $\xi$  è una v.a. di Bernoulli con  $\mathbf{P}(\xi = 1) = p$  e  $\mathbf{P}(\xi = 0) = q$ , con  $p + q = 1$  e  $0 < p < 1$ , è presto visto che

$$\varphi_{\xi}(t) = p e^{it} + q.$$

Se ora  $\xi_1, \dots, \xi_n$  sono  $n$  v.a. di Bernoulli indipendenti e tutte identicamente distribuite, posto

$$S^{(n)} = \sum_{k=1}^n \xi_k, \quad T^{(n)} = \frac{S^{(n)} - np}{\sqrt{npq}} = \sum_{k=1}^n \eta_k, \quad \eta_k = \frac{\xi_k - p}{\sqrt{npq}} = \frac{\xi_k}{\sqrt{npq}} - \sqrt{\frac{p}{nq}},$$

si ottiene facilmente dai risultati di Osservazione II.10.3 che

$$\begin{aligned} \varphi_{\eta_k}(t) &= \mathbf{E}(e^{it\eta_k}) = e^{-it\sqrt{p/nq}} \varphi_{\xi_k}\left(\frac{t}{\sqrt{npq}}\right) = e^{-it\sqrt{p/nq}} \left( p e^{it/\sqrt{npq}} + q \right) \\ &= p e^{it\sqrt{q/np}} + q e^{-it\sqrt{p/nq}} \end{aligned}$$

e quindi

$$\varphi_{T^{(n)}}(t) = \mathbf{E}\left(e^{itT^{(n)}}\right) = \prod_{k=1}^n e^{it\eta_k} = \left( p e^{it\sqrt{q/np}} + q e^{-it\sqrt{p/nq}} \right)^n.$$

Sarà anche interessante osservare che in questo caso, usando gli sviluppi in serie degli esponenziali, si ha

$$\begin{aligned} \varphi_{T^{(n)}}(t) &= \left[ p \left( 1 + it\sqrt{\frac{q}{np}} - \frac{t^2}{2} \frac{q}{np} \right) + q \left( 1 - it\sqrt{\frac{p}{nq}} - \frac{t^2}{2} \frac{p}{nq} \right) + o\left(\frac{1}{n}\right) \right]^n \\ &= \left( 1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right) \right)^n \xrightarrow{n \rightarrow \infty} e^{-t^2/2}. \end{aligned}$$

Infine, se  $\xi$  è una v.a. di Poisson (vedi Esempio II.3.12) si ha che

$$\varphi_{\xi}(t) = \sum_{k=0}^{\infty} e^{itk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{it})^k}{k!} = e^{\lambda(e^{it} - 1)}. \quad \diamond$$

**II.10.5 Esempio:** Calcoleremo qui le f.c. di alcune v.a. continue dotate di densità. Se  $\xi$  è una v.a. uniformemente distribuita in  $[a, b]$  (vedi Esempio II.3.16), la sua f.c. è

$$\varphi_{\xi}(t) = \int_a^b \frac{e^{ixt}}{b-a} dx = e^{i(b+a)t/2} \frac{\sin \frac{b-a}{2} t}{\frac{b-a}{2} t};$$

in particolare se  $\xi$  è uniforme su  $[-1, 1]$  risulta

$$\varphi_{\xi}(t) = \frac{\sin t}{t}.$$

Se  $\xi$  è esponenziale (vedi Esempio II.3.16) la sua f.c. è

$$\varphi_{\xi}(t) = \int_0^{+\infty} \lambda e^{-\lambda x} e^{ixt} dx = \frac{\lambda}{\lambda - it} = \frac{\lambda^2 + i\lambda t}{\lambda^2 + t^2};$$

mentre se  $\xi$  è un'esponenziale *bilatera*, cioè con  $f_{\xi}(x) = \frac{\lambda}{2} e^{-\lambda|x|}$  come fdd, la sua f.c. è

$$\varphi_{\xi}(t) = \int_{-\infty}^{+\infty} \lambda e^{-\lambda|x|} e^{ixt} dx = \frac{\lambda^2}{\lambda^2 + t^2}.$$

Se infine  $\xi \sim \mathcal{N}(m, \sigma^2)$  (vedi Esempio II.3.16) con  $|m| < +\infty$  e  $\sigma^2 > 0$  conviene porre  $\eta = (\xi - m)/\sigma$  ed osservare (vedi Esempio II.8.11) che  $\eta \sim \mathcal{N}(0, 1)$ . Dalla discussione di Osservazione II.10.3, essendo  $\xi = \sigma\eta + m$ , risulta allora che

$$\varphi_{\xi}(t) = e^{itm} \varphi_{\eta}(\sigma t),$$

sicché sarà sufficiente calcolare la f.c. di una v.a. normale standard  $\mathcal{N}(0, 1)$ . Usando le proprietà di convergenza dello sviluppo in serie di potenze degli esponenziali avremo infatti che

$$\begin{aligned} \varphi_{\eta}(t) &= \mathbf{E}(e^{it\eta}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{itx} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2} \sum_{n=0}^{\infty} \frac{(itx)^n}{n!} dx \\ &= \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^n e^{-x^2/2} dx; \end{aligned}$$

se poi si ricorda che

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^n e^{-x^2/2} dx = \begin{cases} 0, & n = 2k + 1 \\ \frac{(2k)!}{2^k k!}, & n = 2k \end{cases}$$

si ottiene

$$\varphi_{\eta}(t) = \sum_{n=0}^{\infty} \frac{(it)^{2k}}{(2k)!} \frac{(2k)!}{2^k k!} = \sum_{n=0}^{\infty} \left(-\frac{t^2}{2}\right)^k \frac{1}{k!} = e^{-t^2/2}$$

da cui in definitiva

$$\varphi_{\xi}(t) = e^{itm - t^2 \sigma^2 / 2}.$$

In particolare se  $\xi \sim \mathcal{N}(0, \sigma^2)$  la fdd e la f.c. sono rispettivamente

$$f_{\xi}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}, \quad \varphi_{\xi}(t) = e^{-t^2 \sigma^2 / 2}$$

caso che mette bene in evidenza sia il fatto che la f.c. di una fdd Gaussiana è ancora una funzione Gaussiana, sia la relazione inversa che sussiste fra la larghezza (varianza)  $\sigma^2$  della fdd e la larghezza  $1/\sigma^2$  della f.c. Questa osservazione è di particolare importanza per le applicazioni ed assume tutto il suo significato se considerata in connessione con il cosiddetto *Principio di indeterminazione* della meccanica quantistica<sup>2</sup>. Osserviamo infine che i risultati dell'Esempio II.10.4 e quelli esposti in questo Esempio mostrano che  $\varphi_{T^{(n)}}(t) \xrightarrow{n \rightarrow \infty} \varphi_\eta(t)$  dove  $\eta \sim \mathcal{N}(0, 1)$ ; in altre parole: la successione delle f.c. delle  $T^{(n)}$  converge punto a punto verso la f.c. di una v.a. normale standard. Come vedremo in seguito questo è il primo esempio di dimostrazione del *Teorema Limite Centrale* con il metodo delle f.c.  $\diamond$

**II.10.6 Teorema:** Se  $\xi$  è una v.a. con FdD  $F(x)$  e f.c.  $\varphi(t)$ , risultano verificate le seguenti proprietà:

- (1)  $|\varphi(t)| \leq \varphi(0) = 1; \quad t \in \mathbf{R};$
- (2)  $\varphi(t)$  è uniformemente continua in  $\mathbf{R}$ ;
- (3)  $\varphi(t) = \overline{\varphi(-t)}$ ;
- (4)  $\varphi(t)$  è reale se e solo se  $F(x)$  è *simmetrica*, cioè se  $\int_B F(dx) = \int_{-B} F(dx)$  con  $B \in \mathcal{B}(\mathbf{R})$  e  $-B = \{x : -x \in B\}$ ;
- (5) se per un dato  $n \geq 1$  risulta  $\mathbf{E}|\xi|^n < +\infty$ , allora esistono tutte le derivate  $\varphi^{(r)}(t)$  con  $r \leq n$  e inoltre

$$\varphi^{(r)}(t) = \int_{\mathbf{R}} (ix)^r e^{itx} F(dx), \quad \varphi^{(r)}(0) = i^r \mathbf{E} \xi^r;$$

$$\varphi(t) = \sum_{r=0}^n \frac{(it)^r}{r!} \mathbf{E} \xi^r + \frac{(it)^n}{n!} \epsilon_n(t) = \sum_{r=0}^n \frac{t^r}{r!} \varphi^{(r)}(0) + \frac{(it)^n}{n!} \epsilon_n(t)$$

dove  $|\epsilon_n(t)| \leq 3\mathbf{E}|\xi|^n$  e  $\epsilon_n(t) \rightarrow 0$  per  $t \rightarrow 0$ ;

- (6) se  $\mathbf{E}|\xi|^n < +\infty, \forall n \in \mathbf{N}$  e se inoltre

$$\overline{\lim}_n \frac{(\mathbf{E}|\xi|^n)^{1/n}}{n} = \frac{1}{R} < +\infty$$

allora risulta

$$\varphi(t) = \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \mathbf{E} \xi^n = \sum_{n=0}^{\infty} \frac{t^n}{n!} \varphi^{(n)}(0)$$

in tutti i punti tali che  $|t| < R/3$ .

**Dimostrazione:** Analizzeremo in successione le sei affermazioni:

<sup>2</sup> Vedi ad esempio P. Caldirola, R. Cirelli e G.M. Prosperi: *Introduzione alla Fisica Teorica*; UTET, Torino, 1982, pp. 342-5.

(1) Per la prova basterà osservare che

$$|\varphi(t)| = \left| \int_{-\infty}^{+\infty} e^{itx} F(dx) \right| \leq \int_{-\infty}^{+\infty} F(dx) = \varphi(0) = 1.$$

(2) Posto che

$$|\varphi(t+h) - \varphi(t)| = |\mathbf{E}[e^{it\xi}(e^{ih\xi} - 1)]| \leq \mathbf{E}|e^{ih\xi} - 1|,$$

e tenendo conto del fatto che le ipotesi del Teorema II.6.14 sono verificate<sup>3</sup> si ha

$$|\varphi(t+h) - \varphi(t)| \rightarrow 0, \quad (h \rightarrow 0)$$

in maniera indipendente da  $t \in R$ . Ne segue che  $\varphi(t)$  è uniformemente continua in  $R$ .

(3) La verifica è banale e viene lasciata al lettore.

(4) La v.a.  $\xi$  ha una FdD simmetrica se e solo se la sua FdD coincide con quella di  $-\xi$  (cioè se  $\xi$  e  $-\xi$  sono identicamente distribuite), come si vede subito osservando che

$$\begin{aligned} \int_{(-\infty, x]} F_{\xi}(dx) &= F_{\xi}(x), \\ \int_{-(\infty, x]} F_{\xi}(dx) &= \int_{[-x, +\infty)} F_{\xi}(dx) \\ &= \mathbf{P}(\xi \in [-x, +\infty)) = \mathbf{P}(-\xi \in (-\infty, x]) = F_{-\xi}(x). \end{aligned}$$

Pertanto  $F$  è simmetrica se e solo se

$$\varphi(t) = \mathbf{E}(e^{it\xi}) = \mathbf{E}(e^{it(-\xi)}) = \mathbf{E}(e^{-it\xi}) = \overline{\varphi(t)}.$$

(5) Questa proposizione è strettamente connessa con l'uso della Formula di Taylor<sup>4</sup>. Per quel che riguarda la derivabilità della f.c., preso in considerazione il

<sup>3</sup> Il fatto che qui non si abbia a che fare con una successione non crea alcuna difficoltà: le ipotesi del Teorema possono essere infatti verificate nella forma  $|e^{i\xi/n}| = 1 < +\infty$  e  $e^{i\xi/n} \xrightarrow{n} 0$ , **P**-q.o. In tal caso il Teorema II.6.14 garantisce che  $\mathbf{E}|e^{ih\xi} - 1| \rightarrow 0$  per  $h \rightarrow 0$ .

<sup>4</sup> Vedi ad esempio **C. Miranda**: *Lezioni di Analisi Matematica - I*; Liguori, Napoli, 1967, p. 319. Il nome del matematico inglese Brook Taylor (1667-1752) è familiare a chiunque conosca i rudimenti del calcolo essendo il suo teorema uno dei primi strumenti di cui si fa uso. Educato a Cambridge egli divenne presto segretario della Royal Society. A soli 34 anni, però, si dimise per dedicarsi agli studi: i suoi risultati più famosi sono contenuti nel lavoro *Methodus Incrementorum Directa et Inversa* (Londra, 1715), ma egli si interessò anche di altri argomenti pubblicando varie memorie sulla prospettiva, i logaritmi e le serie. Al nome di Taylor, inoltre, è associato anche quello dello scozzese Colin MacLaurin (1698-1746): entrato all'Università di Glasgow a soli undici anni, egli aveva pubblicato i suoi primi importanti lavori già nel 1719. Nel 1725 divenne assistente e poi professore all'Università di Edinburgo. L'identità che porta il suo nome è contenuta nel suo *Treatise of Fluxions* (Edinburgo, 1742), ma egli scrisse anche di algebra, geometria e fisica.



rapporto incrementale

$$\frac{\varphi(t+h) - \varphi(t)}{h} = \mathbf{E} \left( e^{it\xi} \frac{e^{ih\xi} - 1}{h} \right)$$

osserviamo che, siccome

$$\left| \frac{\sin \alpha}{\alpha} \right| \leq 1, \quad \alpha \in \mathbf{R},$$

risulta anche

$$|e^{ihy} - 1| = \sqrt{2(1 - \cos hy)} = 2 \left| \sin \frac{hy}{2} \right| \leq 2 \left| \frac{hy}{2} \right| = |hy|,$$

e quindi

$$\left| e^{it\xi} \frac{e^{ih\xi} - 1}{h} \right| = \left| \frac{e^{ih\xi} - 1}{h} \right| \leq |\xi|.$$

Conseguentemente, essendo per ipotesi  $\mathbf{E}|\xi| < +\infty$ , il Teorema II.6.14 ci garantisce che possiamo passare al limite sotto segno di VdA sicché:

$$\begin{aligned} \varphi'(t) &= \lim_{h \rightarrow 0} \frac{\varphi(t+h) - \varphi(t)}{h} = \mathbf{E} \left( e^{it\xi} \lim_{h \rightarrow 0} \frac{e^{ih\xi} - 1}{h} \right) \\ &= \mathbf{E} (e^{it\xi} [i\xi e^{ih\xi}]_{h=0}) = \mathbf{E} (i\xi e^{it\xi}). \end{aligned}$$

Questo stabilisce la tesi per  $r = 1$ : notiamo che in pratica questa proposizione consente di eseguire la derivata direttamente sotto segno di integrale. Le derivate di ordine successivo si ricavano per induzione, e la relazione

$$\varphi^{(r)}(0) = i^r \mathbf{E} \xi^r$$

è un'immediata conseguenza di tale risultato. Quanto all'ultima parte della proposizione ricorderemo che la Formula di Taylor per una funzione  $f(x)$ , dotata di derivate in un intorno del punto  $x = a$  fino a quella di ordine  $n$  incluso, afferma che comunque scelto  $x$  in tale intorno è sempre possibile determinare un opportuno numero  $\theta \in (0, 1)$  (dipendente da  $x$ ) in modo tale che in tutto un intorno di  $x = a$  risulti

$$\begin{aligned} f(x) &= f(a) + (x-a)f'(a) + \dots \\ &\quad + \frac{(x-a)^{n-1}}{(n-1)!} f^{(n-1)}(a) + \frac{(x-a)^n}{n!} f^{(n)}(a + (x-a)\theta). \end{aligned}$$

Applicando questo risultato alle funzioni  $\cos y$  e  $\sin y$  (parti reale e immaginaria di  $e^{iy}$ ) si ha che potremo sempre determinare  $\theta_1$  e  $\theta_2$  (dipendenti da  $y$ ) in  $(0, 1)$  in modo tale che in un intorno di  $y = 0$  risulti

$$e^{iy} = \sum_{k=0}^{n-1} \frac{(iy)^k}{k!} + \frac{(iy)^n}{n!} [\cos(y\theta_1) + i \sin(y\theta_2)].$$

Pertanto, calcolando la nostra espressione in  $y = t\xi$ , si ha che

$$e^{it\xi} = \sum_{k=0}^{n-1} \frac{(it\xi)^k}{k!} + \frac{(it\xi)^n}{n!} [\cos(t\xi\theta_1) + i \sin(t\xi\theta_2)],$$

dove  $\theta_1$  e  $\theta_2$ , che dipendevano da  $y$ , sono ora funzioni di  $\xi$  e quindi sono v.a. Ponendo allora

$$\epsilon_n(t) = \mathbf{E} [\xi^n (\cos(t\xi\theta_1) + i \sin(t\xi\theta_2) - 1)],$$

potremo scrivere, come richiesto nella tesi,

$$\begin{aligned} \varphi(t) &= \mathbf{E} (e^{it\xi}) = \sum_{k=0}^{n-1} \frac{(it)^k}{k!} \mathbf{E} \xi^k + \frac{(it)^n}{n!} [\mathbf{E} \xi^n + \epsilon_n(t)] \\ &= \sum_{k=0}^n \frac{(it)^k}{k!} \mathbf{E} \xi^k + \frac{(it)^n}{n!} \epsilon_n(t). \end{aligned}$$

Inoltre si vede subito che

$$\begin{aligned} |\epsilon_n(t)| &\leq \mathbf{E} (|\xi|^n |\cos(t\xi\theta_1) + i \sin(t\xi\theta_2) - 1|) \\ &\leq \mathbf{E} (|\xi|^n (|\cos(t\xi\theta_1)| + |\sin(t\xi\theta_2)| + 1)) \leq 3\mathbf{E} |\xi|^n. \end{aligned}$$

Infine, siccome per  $\theta_1, \theta_2 \in (0, 1)$  si ha

$$\xi^n (\cos(t\xi\theta_1) + i \sin(t\xi\theta_2) - 1) \rightarrow 0, \quad (t \rightarrow 0)$$

per ogni  $\omega \in \Omega$ , e visto che

$$|\xi^n (\cos(t\xi\theta_1) + i \sin(t\xi\theta_2) - 1)| \leq 3|\xi|^n$$

con  $\mathbf{E} |\xi|^n < +\infty$ , il Teorema II.6.14 ci consente ancora una volta di passare al limite sotto il segno di VdA sicché, come richiesto,  $\epsilon_n(t) \rightarrow 0$  per  $t \rightarrow 0$ .

- (6) Consideriamo un numero  $t_0$  tale che  $0 < t_0 < R/3$ , in modo che per ipotesi avremo

$$\overline{\lim}_n \frac{(\mathbf{E} |\xi|^n)^{1/n}}{n} = \frac{1}{R} < \frac{1}{3t_0}$$

ossia anche

$$\overline{\lim}_n \frac{(\mathbf{E} |\xi|^n (3t_0)^n)^{1/n}}{n} < 1.$$

Inoltre dalla formula di Stirling (vedi Esempio I.3.4)

$$n! = \sqrt{2\pi n} \left(\frac{n}{e}\right)^n e^{\theta_n/12n}; \quad \theta_n \in (0, 1),$$

si ha che  $n!e^n > n^n$ , cioè

$$\frac{1}{e(n!)^{1/n}} < \frac{1}{n}$$

per cui risulterà anche

$$\overline{\lim}_n \left( \frac{\mathbf{E} |\xi|^n (3t_0)^n}{n!} \right)^{1/n} < e.$$

Secondo un noto teorema<sup>5</sup> se per una serie di potenze  $\sum_n a_n y^n$  risulta

$$\overline{\lim}_n \sqrt[n]{|a_n|} = l,$$

la serie ha raggio di convergenza pari a  $r = 1/l$ . Nel nostro caso potremo allora dire che la serie

$$\sum_n \frac{\mathbf{E} |\xi|^n (3t_0)^n}{n!} y^n$$

converge per  $y < 1/e$  e quindi in particolare per  $y = 1/3$ ; in altre parole potremo affermare che la serie numerica (a termini positivi)

$$\sum_n \frac{\mathbf{E} |\xi|^{nt_0^n}}{n!}$$

sicuramente converge. Ciò posto, siccome per  $|t| \leq t_0$  si ha

$$\left| \frac{(it)^n}{n!} \mathbf{E} \xi^n \right| \leq \frac{|t|^n}{n!} \mathbf{E} |\xi|^n \leq \frac{\mathbf{E} |\xi|^{nt_0^n}}{n!},$$

la serie di potenze

$$\sum_{n=0}^{\infty} \frac{(it)^n}{n!} \mathbf{E} \xi^n$$

risulta maggiorata da una serie numerica convergente e quindi è uniformemente ed assolutamente convergente in  $|t| \leq t_0$ . Per determinare la somma di tale serie ricordiamo che da (5) si ha

$$\varphi(t) = \sum_{k=0}^n \frac{(it)^k}{k!} \mathbf{E} \xi^k + R_n(t), \quad n \in \mathbf{N}$$

e che, per  $|t| \leq t_0$ , si ha

$$|R_n(t)| = \left| \frac{(it)^n}{n!} \epsilon_n(t) \right| \leq 3 \frac{|t|^n}{n!} \mathbf{E} |\xi|^n \leq 3 \frac{t_0^n}{n!} \mathbf{E} |\xi|^n.$$

<sup>5</sup> Vedi ad esempio **C. Miranda**: *Lezioni di Analisi Matematica - I*; Liguori, Napoli, 1967, p. 490.

Siccome la convergenza della corrispondente serie numerica richiede che la successione dei suoi termini sia infinitesima

$$\frac{t_0^n}{n!} \mathbf{E} |\xi|^n \xrightarrow{n \rightarrow \infty} 0,$$

ciò implica che

$$\varphi(t) = \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \mathbf{E} \xi^n, \quad |t| \leq t_0$$

in quanto per  $|t| \leq t_0$  il resto  $n$ -mo  $R_n(t)$  risulta infinitesimo se  $n \rightarrow \infty$ . Data l'arbitrarietà di  $t_0 \in (0, R/3)$  potremo quindi affermare che la tesi risulta verificata per  $|t| < R/3$ .  $\square$

**II.10.7 Osservazione:** Con una dimostrazione del tutto analoga a quella data il punto (5) del precedente Teorema può essere esteso al caso di sviluppi attorno al punto  $t = s$  invece che  $t = 0$ . In tal caso (vedi anche il successivo Teorema II.10.13) si avrà

$$\varphi(t) = \sum_{k=0}^n \frac{i^k (t-s)^k}{k!} \mathbf{E} (\xi^k e^{is\xi}) + \frac{i^n (t-s)^n}{n!} \epsilon_n(t-s)$$

dove  $|\epsilon_n(t-s)| \leq 3\mathbf{E} |\xi|^n$  e inoltre  $\epsilon_n(t-s) \rightarrow 0$  per  $t \rightarrow s$ .  $\circ$

**II.10.8 Teorema di Unicità:** Se  $F(x)$  e  $G(x)$  sono due FdD dotate della stessa f.c., cioè se

$$\int_{-\infty}^{+\infty} e^{itx} F(dx) = \int_{-\infty}^{+\infty} e^{itx} G(dx), \quad t \in \mathbf{R},$$

allora risulta  $F(x) = G(x), \forall x \in \mathbf{R}$ .

**Dimostrazione:** Omessa<sup>6</sup>. In pratica il Teorema afferma che se le f.c. di due v.a.  $\xi$  e  $\eta$  coincidono in ogni punto (cioè se  $\varphi_\xi(t) = \varphi_\eta(t)$  per ogni  $t \in \mathbf{R}$ ), allora le due v.a. sono identicamente distribuite.  $\square$

<sup>6</sup> Vedi ad esempio **A.N. Shiryaev:** *Probability*; Springer, New York, 1984, p. 280.

**II.10.9 Teorema (Formula di inversione):** Se  $F(x)$  è una FdD con f.c.  $\varphi(t)$ , allora:

(A) comunque dati  $a$  e  $b$  con  $a < b$  risulta

$$\frac{F(b) + F(b^-)}{2} - \frac{F(a) + F(a^-)}{2} = \frac{1}{2\pi} \lim_{T \rightarrow +\infty} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt;$$

(B) inoltre si ha anche

$$f(x) = \frac{1}{2\pi} \lim_{T \rightarrow +\infty} \int_{-T}^T e^{-itx} \varphi(t) dt,$$

se  $F(x)$  è dotata di fdd  $f(x)$ .

**Dimostrazione:** Omessa<sup>7</sup>. Questo Teorema mostra come si calcolano  $F$  e  $f$  a partire da una data f.c.  $\varphi$ ; il Teorema di Unicità ci garantisce poi che tali  $F$  e  $f$  sono uniche.  $\square$

**II.10.10 Osservazione:** L'espressione (A) del Teorema precedente consente di tenere conto dei casi in cui  $a$  e  $b$  (o almeno uno dei due) sono punti di discontinuità per  $F(x)$ . Se invece  $a$  e  $b$  sono punti di continuità per  $F(x)$ , la formula di inversione (A) si semplifica in

$$F(b) - F(a) = \frac{1}{2\pi} \lim_{T \rightarrow +\infty} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt.$$

Inoltre, se  $F(x)$  è dotata di densità, e supponendo senza prova di poter scambiare l'ordine delle integrazioni, si vede subito che la formula (B) implica (A):

$$\begin{aligned} F(b) - F(a) &= \int_a^b f(x) dx = \int_a^b \left[ \frac{1}{2\pi} \lim_{T \rightarrow +\infty} \int_{-T}^T e^{-itx} \varphi(t) dt \right] dx \\ &= \frac{1}{2\pi} \lim_{T \rightarrow +\infty} \int_{-T}^T \left[ \int_a^b e^{-itx} dx \right] \varphi(t) dt \\ &= \frac{1}{2\pi} \lim_{T \rightarrow +\infty} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt. \end{aligned}$$

Val la pena di osservare infine che il limite per  $T \rightarrow +\infty$  delle formule (A) e (B) non definisce un integrale improprio (per l'esistenza del quale sarebbe necessaria l'esistenza separata dei limiti nei due estremi di integrazione), ma il cosiddetto *valor principale* dell'integrale. Naturalmente se l'integrale improprio esistesse

<sup>7</sup> Vedi ad esempio **A.N. Shiryaev**: *Probability*; Springer, New York, 1984, p. 281.

esso coinciderebbe con il suo valor principale, ma quest'ultimo può esistere anche quando il primo non esiste.  $\circ$

**II.10.11 Teorema:** Condizione necessaria e sufficiente affinché le componenti di un vett.a.  $X = (\xi_1, \dots, \xi_n)$  siano indipendenti è che

$$\varphi_X(t_1, \dots, t_n) = \varphi_{\xi_1}(t_1) \cdot \dots \cdot \varphi_{\xi_n}(t_n),$$

cioè che la sua f.c.  $\varphi_X(t_1, \dots, t_n)$  sia il prodotto delle f.c.  $\varphi_{\xi_k}(t_k)$  delle sue componenti.

**Dimostrazione:** Che  $\varphi_X$  sia il prodotto delle  $\varphi_{\xi_k}$  se le sue componenti sono indipendenti è un fatto banale alla luce del Teorema II.6.15 e della Definizione II.10.2. Viceversa, supponendo che  $\varphi_X(t_1, \dots, t_n) = \varphi_{\xi_1}(t_1) \cdot \dots \cdot \varphi_{\xi_n}(t_n)$ , indicando con  $F(x_1, \dots, x_n)$  la FdD congiunta di  $X$  e posto  $G(x_1, \dots, x_n) = F_{\xi_1}(x_1) \cdot \dots \cdot F_{\xi_n}(x_n)$ , il Teorema di Unicità II.10.8 ci consente di dedurre che  $F(x_1, \dots, x_n) = G(x_1, \dots, x_n)$  dal fatto che (usando anche il Teorema di Fubini)

$$\begin{aligned} \int_{\mathbf{R}^n} e^{i(t,x)} G(dx_1, \dots, dx_n) &= \prod_{k=1}^n \int_{-\infty}^{+\infty} e^{it_k x_k} F_{\xi_k}(dx_k) = \prod_{k=1}^n \varphi_{\xi_k}(t_k) \\ &= \varphi_X(t_1, \dots, t_n) = \int_{\mathbf{R}^n} e^{i(t,x)} F(dx_1, \dots, dx_n). \end{aligned}$$

Ma  $F = G$  vuol dire  $F(x_1, \dots, x_n) = F_{\xi_1}(x_1) \cdot \dots \cdot F_{\xi_n}(x_n)$  e questo, in base al Teorema II.5.41 equivale al fatto che le componenti di  $X$  sono indipendenti.  $\square$

**II.10.12 Osservazione:** Supponiamo che tutti i momenti  $m_n = \mathbf{E} \xi^n$ ,  $n \geq 1$ , di una data v.a.  $\xi$  esistano e ci siano noti: il Teorema II.10.6 ci dice che vi è una stretta relazione fra tali momenti e la f.c.  $\varphi(t)$  della nostra v.a.; inoltre i Teoremi II.11.8 e II.11.9 ci dicono che la f.c.  $\varphi(t)$  determina in modo unico la FdD  $F(x)$  della  $\xi$ . Ha quindi senso porsi la seguente domanda che prende il nome di **Problema dei Momenti**: è possibile risalire in modo unico dalla conoscenza dei momenti  $(m_n)_{n \in \mathbf{N}}$  di una v.a.  $\xi$  alla sua FdD  $F(x)$ ? In particolare il problema dell'unicità può essere posto in questa forma: se fossero date due FdD  $F(x)$  e  $G(x)$  tali che

$$\int_{-\infty}^{+\infty} x^n F(dx) = \int_{-\infty}^{+\infty} x^n G(dx), \quad n \geq 1$$

è possibile affermare che  $F(x) = G(x)$  in ogni punto  $x$ ? In realtà si potrebbe mostrare con dei controesempi (vedi **A.N. Shiryaev**: *Probability*, Springer, New York, 1984, p. 292) che la risposta è in generale negativa: è possibile, cioè, costruire FdD distinte che danno luogo alla medesima successione di momenti. Sarà allora importante determinare sotto quali condizioni il problema dei momenti ammette una ed una sola soluzione: le ipotesi necessarie sono riassunte nel successivo Teorema.  $\circ$

**II.10.13 Teorema:** Data una v.a.  $\xi$  con FdD  $F(x)$  e posto

$$m_n = \mathbf{E} \xi^n, \quad \mu_n = \mathbf{E} |\xi|^n,$$

se i suoi momenti assoluti  $\mu_n$  risultano tutti finiti e se

$$\overline{\lim}_n \frac{\mu_n^{1/n}}{n} < +\infty$$

allora i momenti  $m_n$  determinano la FdD in maniera unica.

**Dimostrazione:** Da (6) di Teorema II.11.6 sappiamo che, nelle nostre ipotesi, esiste sicuramente un numero  $t_0 > 0$  tale che la f.c. di  $\xi$ ,  $\varphi(t) = \mathbf{E}(e^{it\xi})$  può essere rappresentata mediante lo sviluppo in serie

$$\varphi(t) = \sum_{k=0}^{\infty} \frac{(it)^k}{k!} m_k, \quad |t| \leq t_0.$$

Pertanto i momenti  $m_n$  determinano (in modo unico) la f.c. esclusivamente nell'intervallo  $[-t_0, t_0]$ . Per poter usare i Teoremi II.10.8 e II.10.9, però, a noi occorre conoscere  $\varphi(t)$  su tutto  $\mathbf{R}$ : sarà necessario, quindi, mostrare che i momenti  $m_n$  determinano univocamente  $\varphi(t)$  anche al di fuori dell'intervallo  $[-t_0, t_0]$ . Per far questo consideriamo due numeri  $t, s$  tali che  $|s| \leq t_0/2$  e  $|t-s| \leq t_0$ . Dal fatto ovvio che  $|t| = |(t-s) + s| \leq |t-s| + |s|$ , segue allora che  $|t| \leq 3t_0/2$ . Riprendendo ora la dimostrazione del punto (5) del Teorema II.10.6 (vedi anche Osservazione II.10.7), dalla formula di Taylor di  $e^{iy}$ , che è valida per qualunque valore di  $y$ , otteniamo

$$e^{i(t-s)\xi} = \sum_{k=0}^{n-1} \frac{[i(t-s)\xi]^k}{k!} + \frac{[i(t-s)\xi]^n}{n!} (\cos[(t-s)\xi\theta_1] + i \sin[(t-s)\xi\theta_2]),$$

sicché seguendo i medesimi passaggi di quella dimostrazione e tenendo conto dell'espressione per le derivate  $\varphi^{(r)}(t)$  date nel medesimo Teorema, si ottiene

$$\begin{aligned} \varphi(t) &= \sum_{k=0}^n \frac{(t-s)^k}{k!} \mathbf{E} [(i\xi)^k e^{is\xi}] + \frac{[i(t-s)]^n}{n!} \epsilon_n(t-s) \\ &= \sum_{k=0}^n \frac{(t-s)^k}{k!} \varphi^{(k)}(s) + R_n(t-s), \end{aligned}$$

dove  $\epsilon_n$  ed  $R_n$  hanno le medesime definizioni date ai punti (5) e (6) del Teorema II.10.6 e godono delle medesime proprietà anche se sono ora calcolate in  $t-s$  invece che in  $t$ . In particolare nelle nostre ipotesi, dato che  $|t-s| \leq t_0$ , avremo che  $R_n(t-s) \xrightarrow{n} 0$  e quindi

$$\varphi(t) = \sum_{n=0}^{\infty} \frac{(t-s)^n}{n!} \varphi^{(n)}(s), \quad |t| \leq \frac{3t_0}{2}.$$

Se ora osserviamo che i numeri  $\varphi^{(n)}(s)$  sono derivabili dalla conoscenza dei momenti<sup>8</sup>  $m_n$ , possiamo trarre la conclusione che i momenti determinano  $\varphi(t)$  nell'intervallo esteso  $|t| \leq 3t_0/2$ . Iterando questo procedimento<sup>9</sup> potremo affermare, allora, che i momenti  $m_n$  consentono di ricostruire  $\varphi(t)$  su tutto  $\mathbf{R}$  e quindi, in base ai Teoremi II.10.8 e II.10.9, di ricostruire univocamente la FdD  $F(x)$  da cui derivano.  $\square$

**II.10.14 Corollario:** I momenti di una v.a. determinano sempre la sua FdD  $F(x)$  se questa è concentrata su un intervallo finito.

**Dimostrazione:** Basterà mostrare che le ipotesi del Teorema II.10.13 sono verificate quando  $F(x)$  è costante al di fuori dell'intervallo finito  $[a, b]$ . Infatti, posto  $M = \max(|a|, |b|)$ , si ha

$$\mu_n = \int_a^b |x|^n F(dx) \leq M^n \int_a^b F(dx) = M^n$$

sicché risulterà

$$\overline{\lim}_n \frac{\mu_n^{1/n}}{n} \leq \overline{\lim}_n \frac{M}{n} = \lim_n \frac{M}{n} = 0. \quad \square$$

**II.10.15 Osservazione:** Come sarà mostrato fra breve la grande utilità delle f.c. risiede nella possibilità delle loro applicazioni alle dimostrazioni dei Teoremi limite. La stretta relazione che sussiste fra la FdD  $F(x)$  e la f.c.  $\varphi(t)$  di una v.a. induce infatti a pensare che la convergenza debole (vedi le Definizioni II.9.1-2-3) di una successione di FdD  $(F_n(x))_{n \in \mathbf{N}}$  possa essere analizzata mediante lo studio della convergenza (puntuale) della corrispondente successione di f.c.  $(\varphi_n(t))_{n \in \mathbf{N}}$ . Tale possibilità sarebbe particolarmente utile nel caso, più volte incontrato, di successioni di somme di v.a. indipendenti  $\xi_n$  del tipo  $S^{(n)} = \sum_k \xi_k$ . Infatti sappiamo che, mentre le FdD  $F_n(x)$  di tali somme si calcolano mediante prodotti di convoluzione (vedi Osservazione II.8.14) delle FdD marginali  $F_{\xi_k}(x)$ , le corrispondenti f.c.  $\varphi_n(t)$  sono i prodotti ordinari (molto più semplici da usare, quindi) delle f.c. individuali  $\varphi_{\xi_k}(t)$  (vedi Osservazione II.10.3). Gli importanti risultati riassunti nei Teoremi seguenti individuano le condizioni sotto le quali è possibile affermare che

<sup>8</sup> Infatti  $\varphi(t)$  è data in  $|t| \leq t_0$  da una serie (uniformemente convergente) costruita a partire dai momenti; pertanto in  $|t| \leq t_0$  le derivate  $\varphi^{(k)}(t)$  possono essere calcolate derivando la serie termine a termine, sicché i momenti  $m_n$  determinano anche tutte le derivate  $\varphi^{(k)}(t)$  in  $|t| \leq t_0$ . Siccome  $|s| \leq t_0/2 < t_0$ , è chiaro che i numeri  $\varphi^{(k)}(s)$  saranno tutti calcolabili a partire dai momenti  $m_n$ .

<sup>9</sup> Dalla costruzione qui esposta si ricava immediatamente che all' $n$ -ma iterazione l'estremo superiore dell'intervallo di convergenza sarà collocato in  $t_n = (3/2)t_{n-1} = (3/2)^2 t_{n-2} = \dots = (3/2)^n t_0$ , sicché sarà facile mostrare che  $t_n \xrightarrow{n} +\infty$  in modo che, al limite, l'intervallo di convergenza tenderà a ricoprire tutto  $\mathbf{R}$ .



la convergenza debole  $F_n \xrightarrow{w} F$  è equivalente alla convergenza (semplice punto a punto)  $\varphi_n(t) \rightarrow \varphi(t)$  delle corrispondenti f.c. ○

**II.10.16 Teorema di continuità (P. Lévy):** Data una successione di FdD  $(F_n(x))_{n \in \mathbf{N}}$  e data la corrispondente successione di f.c.  $(\varphi_n(t))_{n \in \mathbf{N}}$  si ha che:

(a) se  $F_n \xrightarrow{w} F$  e se  $F(x)$  risulta essere una FdD, allora anche  $\varphi_n(t) \xrightarrow{n} \varphi(t)$  in ogni  $t \in \mathbf{R}$ , e  $\varphi(t)$  risulta essere la f.c. di  $F(x)$ ;

(b) se la funzione  $\varphi(t) = \lim_n \varphi_n(t)$  esiste per ogni  $t \in \mathbf{R}$  e se  $\varphi(t)$  è continua in  $t = 0$ , allora  $\varphi(t)$  è la f.c. di una FdD  $F(x)$  e risulta  $F_n \xrightarrow{w} F$ .

**Dimostrazione:** Omessa<sup>10</sup> (vedi **A.N. Shirayev:** *Probability*; Springer, New York, 1984, p. 320). □

**II.10.17 Corollario:** Data una successione di FdD  $(F_n(x))_{n \in \mathbf{N}}$  e data la corrispondente successione di f.c.  $(\varphi_n(t))_{n \in \mathbf{N}}$ , se  $F(x)$  è una FdD e  $\varphi(t)$  la sua f.c., allora risulta  $F_n \xrightarrow{w} F$  se e solo se  $\varphi_n(t) \xrightarrow{n} \varphi(t)$  in ogni  $t \in \mathbf{R}$ .

**Dimostrazione:** Banale: lasciata al lettore come esercizio. □

---

<sup>10</sup> Paul Lévy (1886-1971), ingegnere minerario e matematico francese, si è segnalato nella sua lunga carriera per numerosi lavori di calcolo delle probabilità, analisi funzionale (l'introduzione di questo termine, tra l'altro, sembra essere dovuta proprio a lui) ed equazioni differenziali alle derivate parziali. È stato professore all'École Nationale Supérieure des Mines dal 1914 al 1951 e all'École Polytechnique dal 1920 al 1959. Tra i suoi numerosi trattati ricorderemo solo *Processus Stochastiques et Mouvement Brownien* (Paris, 1948).



## II.11 Sistemi gaussiani

**II.11.1 Osservazione:** Famiglie di v.a. distribuite congiuntamente in maniera gaussiana (normale) giocano un ruolo estremamente importante in probabilità e statistica. Questo è motivato innanzitutto, come vedremo nei capitoli successivi, dal cosiddetto *Teorema Limite Centrale* che generalizza i risultati di De Moivre e Laplace già esposti nel capitolo I.8 e afferma che somme di un gran numero di v.a. (distribuite in maniera arbitraria e soggette a condizioni non troppo restrittive) sono distribuite in maniera gaussiana. È questa la base teorica della cosiddetta *legge degli errori* secondo la quale gli errori casuali sulle misure fisiche, che derivano da somme di un gran numero di errori elementari indipendenti e incontrollabili, si distribuiscono in maniera gaussiana. Inoltre le v.a. gaussiane sono caratterizzate da fdd determinate mediante un numero piuttosto piccolo di *parametri* (il che costituisce un gran vantaggio nella costruzione di semplici modelli probabilistici) e dall'equivalenza, non garantita per altre v.a., fra le proprietà di *indipendenza* e *non correlazione* (vedi Esempio II.8.5). Infine le v.a. gaussiane hanno anche *momenti finiti* di ogni ordine e possono quindi essere analizzate con le tecniche di analisi funzionale esaminate in II.9.; in particolare le curve di regressione fra v.a. gaussiane sono delle rette (vedi Proposizione II.8.9). Per tutte queste ragioni sarà utile fornire qui ulteriori informazioni su queste v.a. ricordando che le precedenti sezioni in cui se ne è fatta menzione sono le seguenti: Capitolo I.8 (introduzione della funzione di Gauss; Teoremi limite di De Moivre e Laplace); Esempio II.3.16 (introduzione della fdd di Gauss e della sua FdD o funzione errore); Esempio II.4.8 (introduzione della fdd gaussiana bivariata e multivariata); Esempio II.5.32 (vett.a. gaussiani multivariati); Esempio II.8.5 (medie, varianze e covarianze di vett.a. gaussiani; equivalenza fra indipendenza e non correlazione); Proposizione II.8.9 (linearità della miglior stima in m.q. per v.a. gaussiane multivariata); Esempi II.8.16-17-18 (proprietà riproduttive delle v.a. gaussiane; distribuzioni del  $\chi^2$  e di Student); Esempio II.10.5 (f.c. delle v.a. gaussiane). ○

**II.11.2 Osservazione:** Come è noto la fdd di una v.a. gaussiana  $\xi \sim \mathcal{N}(m, \sigma^2)$  è, per  $\sigma > 0$ ,

$$f_{\xi}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-m)^2/2\sigma^2}.$$

Dato che  $\sigma^2$  rappresenta la Var della nostra v.a. (vedi Esempio II.8.5), quando  $\sigma \downarrow 0$  la distribuzione della v.a.  $\xi$  converge intuitivamente verso quella di una v.a. degenerare con un solo valore, cioè tale che  $\mathbf{P}(\xi = m) = 1$  e  $\mathbf{P}(\xi \neq m) = 0$ . D'altra parte è noto che una simile v.a. limite, tipicamente discreta, sarebbe caratterizzata da una distribuzione non continua e quindi non dotata di fdd (vedi ad esempio Osservazione II.5.5). Pertanto siamo obbligati a distinguere nettamente il caso in cui  $\sigma > 0$  (in cui  $\xi \sim \mathcal{N}(m, \sigma^2)$  è dotata di fdd gaussiana) dal caso  $\sigma = 0$  (in cui  $\xi$  degenera attorno al valore  $\xi = m$  e non può essere descritta da una fdd) e ad ammettere che le due descrizioni non sfumano con continuità una nell'altra quando  $\sigma \downarrow 0$ . Per porre rimedio a questa situazione che ci obbliga sempre a distinguere

i due casi e a trattarli separatamente osserveremo che nel capitolo precedente abbiamo imparato che una v.a., oltre che mediante le sue FdD o fdd, può essere descritta altrettanto bene tramite la sua f.c.  $\varphi_\xi(t)$  (vedi Teoremi II.10.8-9), e che le f.c. delle v.a. qui considerate sono (Esempi II.10.4-5):

$$\varphi_\xi(t) = \begin{cases} e^{itm} & , \quad \sigma = 0, \\ e^{itm-t^2\sigma^2/2} & , \quad \sigma > 0. \end{cases}$$

Si vede subito ora che la f.c. per  $\sigma = 0$  si ottiene per continuità da quella per  $\sigma > 0$  al limite per  $\sigma \downarrow 0$ : questa è una semplice illustrazione del cosiddetto fenomeno dell'*attrazione delle f.c.* che sarà utilizzato nel seguito per dare una definizione del concetto di vett.a. gaussiano che non sia forzata a distinguere tra il caso degenere e quello non degenere. Va notato a questo proposito che, per vett.a. con più di una componente, la difficoltà di dare una descrizione unitaria (mediante, cioè, una sola funzione) di tutti i casi possibili potrebbe essere accresciuta dal fatto che le componenti mostrino comportamenti differenti: non è esclusa, infatti, l'eventualità che solo alcune delle componenti siano degeneri e che quindi la distribuzione sia intrinsecamente mista. ○

**II.11.3 Definizione:** Diremo che  $X = (\xi_1, \dots, \xi_n)$  è un **vett.a. normale (o Gaussiano)** con vettore delle medie  $m = (m_1, \dots, m_n) \in \mathbf{R}^n$  e matrice delle covarianze  $R = \|r_{kl}\|$  (simmetrica e definita non negativa, vedi Osservazione II.8.3) se la sua f.c. ha la forma

$$\varphi_X(t) = \varphi_X(t_1, \dots, t_n) = e^{i(t,m)} e^{-(t,Rt)/2}, \quad t \in \mathbf{R}^n$$

dove con  $(a, b) = \sum_k a_k b_k$  indichiamo l'usuale prodotto scalare Euclideo di vettori  $a, b \in \mathbf{R}^n$ . In tal caso scriveremo anche  $X \sim \mathcal{N}(m, R)$ . △

**II.11.4 Osservazione:** La definizione precedente ovviamente si riconduce alla discussione svolta nella Osservazione II.11.2 quando il vett.a.  $X$  possiede solo una componente ( $n = 1$ ): in tal caso  $m$  è un numero che rappresenta il VdA di tale componente e  $R$  si riduce ad un solo elemento coincidente con la sua  $\text{Var } \sigma^2$ . Mostriamo ora in maniera esplicita che anche nel caso generale ( $n > 1$ ) la funzione assegnata in Definizione II.11.3 è sempre una f.c. che, nel caso non degenere, coincide con la f.c. di un vett.a. Gaussiano definito, come in Esempio II.4.8, mediante la sua fdd. I due casi (quello di  $X$  degenere e quello di  $X$  non degenere) sono distinti dal fatto che  $R$  risulti o meno singolare. Se infatti  $R$  è una matrice non singolare (cioè risulta  $|R| \neq 0$ , dove  $|R|$  indica il determinante della matrice  $R$ ), possiamo definire la sua inversa  $A = R^{-1}$  e mostrare che la  $\varphi_X(t)$  della Definizione II.11.3 coincide con la f.c. di un vett.a. Gaussiano dotato di densità normale multivariata (vedi Esempio II.4.8)

$$f(x_1, \dots, x_n) = \sqrt{\frac{|A|}{(2\pi)^n}} e^{-(x-m)^T A (x-m)/2}$$

dove con T indichiamo, come al solito, l'operazione di trasposizione. Dovremo cioè verificare che

$$\varphi_X(t) = \int_{\mathbf{R}^n} e^{i(t,x)} f(x) d^n x,$$

ovvero, in pratica, che

$$I = \sqrt{\frac{|A|}{(2\pi)^n}} \int_{\mathbf{R}^n} e^{i(t,x-m)} e^{-(x-m)^T A(x-m)/2} d^n x = e^{-(t,Rt)/2}.$$

Per eseguire il calcolo consideriamo la matrice ortogonale O che diagonalizza R (vedi dimostrazione della Proposizione II.8.4) e poniamo

$$D = O^T R O = \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{pmatrix}.$$

Siccome R è non singolare esisterà la sua inversa  $A = R^{-1}$  e inoltre avremo che  $d_i > 0$  per  $i = 1, \dots, n$ . Siccome O è ortogonale ( $O^{-1} = O^T$ ) e, in quanto tale, lascia invariante il determinante delle matrici, risulterà

$$O^T A O = (O^T R O)^{-1} = D^{-1} = \begin{pmatrix} 1/d_1 & 0 & \dots & 0 \\ 0 & 1/d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/d_n \end{pmatrix},$$

$$|A| = |R^{-1}| = |D^{-1}| = \frac{1}{d_1 \dots d_n}.$$

Eseguiamo ora, nell'integrale multiplo I, la trasformazione di coordinate (rotazione senza inversioni di assi) definita da

$$r = O^T(x - m) \quad s = O^T t,$$

il cui determinante Jacobiano (se  $O_{jk}$  sono gli elementi di matrice di O)

$$J = \frac{\partial(x_1, \dots, x_n)}{\partial(r_1, \dots, r_n)}, \quad \frac{\partial x_j}{\partial r_k} = \frac{\partial}{\partial r_k} \left( m_j + \sum_{i=1}^n O_{ji} r_i \right) = O_{jk}$$

vale chiaramente  $J = |O| = 1$  dato che per l'ortogonalità si ha  $O^T O = I$ , dove I è la matrice identità, e dato che non consideriamo inversioni di assi. Avremo allora

$$\begin{aligned} i(t, x - m) - \frac{1}{2}(x - m, A(x - m)) &= i(Os, Or) - \frac{1}{2}(Or, AOr) \\ &= i(Os)^T Or - \frac{1}{2}(Or)^T AOr = is^T r - \frac{1}{2} r^T O^T AOr = i \sum_{k=1}^n s_k r_k - \frac{1}{2} \sum_{k=1}^n \frac{r_k^2}{d_k}, \end{aligned}$$

e siccome (vedi Esempio II.10.5)

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{itx} e^{-x^2/2\sigma^2} dx = e^{-t^2\sigma^2/2},$$

avremo in definitiva, come volevamo verificare, che

$$\begin{aligned} I &= \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{d_1 \cdots d_n}} \int_{\mathbf{R}^n} \prod_{k=1}^n (e^{is_k r_k} e^{-r_k^2/2d_k}) d^n r \\ &= \prod_{k=1}^n \frac{1}{\sqrt{2\pi d_k}} \int_{-\infty}^{+\infty} e^{is_k r_k} e^{-r_k^2/2d_k} dr_k = \prod_{k=1}^n e^{-s_k^2 d_k/2} \\ &= e^{-s^T D s/2} = e^{-s^T O^T R O s/2} = e^{-t^T R t/2} = e^{-(t, R t)/2}. \end{aligned}$$

Invece, se  $R$  è una matrice singolare (se, cioè,  $|R| = 0$ ), la  $\varphi_X(t)$  data in Definizione II.11.3 non potrà più essere considerata come f.c. di un vett.a. dotato di fdd gaussiana (multivariata) a causa del fatto che ora  $R$  non è dotata di inversa. Ciononostante si può mostrare che  $\varphi_X(t)$  risulta ancora essere f.c. di un vett.a. Infatti, preso  $\epsilon > 0$ , definiamo la matrice  $R_\epsilon = R + \epsilon I$ : essa risulta evidentemente ancora simmetrica e definita non negativa, ma a differenza di  $R$  risulta anche non singolare per ogni  $\epsilon > 0$ . Ne segue allora che, per ogni  $\epsilon > 0$ , ci ritroviamo nel caso discusso precedentemente sicché

$$\varphi_\epsilon(t) = e^{i(t, m)} e^{-(t, R_\epsilon t)/2}$$

sarà la f.c. del vett.a.  $X_\epsilon \sim \mathcal{N}(m, R_\epsilon)$ . Siccome inoltre risulta ovviamente

$$\lim_{\epsilon \rightarrow 0} \varphi_\epsilon(t) = e^{i(t, m)} e^{-(t, R t)/2} = \varphi_X(t),$$

e la funzione limite è continua in  $t = (0, \dots, 0)$ , il Teorema di Continuità II.10.16 ci garantisce che  $\varphi_X(t)$  è f.c. di un vett.a. In pratica i vett.a. così caratterizzati avranno alcune componenti degeneri ( $\mathbf{P}(\xi_k = m_k) = 1$ ) e quindi non potranno essere descritti tramite una fdd, ma avendo per f.c. il limite (continuo) di una successione di f.c. di vett.a. Gaussiani vengono considerati anche essi a pieno titolo dei vett.a. Gaussiani. ○

**II.11.5 Osservazione:** Un calcolo diretto, analogo a quello indicato nell'Esempio II.8.5, mostra infine che il significato del vettore  $m = (m_1, \dots, m_n)$  e della matrice  $R = \|r_{kl}\|$  è il seguente:

$$m_k = \mathbf{E} \xi_k; \quad r_{kl} = \mathbf{cov}(\xi_k, \xi_l).$$

Le forme esplicite di  $m$  e  $R$  per il caso  $n = 2$  sono riportate nell'Esempio II.8.5, mentre, nel caso non degenero, la corrispondente matrice  $A$  è

$$A = R^{-1} = \frac{1}{(1-r^2)\sigma_1^2\sigma_2^2} \begin{pmatrix} \sigma_2^2 & -r\sigma_1\sigma_2 \\ -r\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}$$

la cui forma ci permette di verificare che la fdd data tramite A nell'Osservazione precedente coincide con quella assegnata ai vett.a. Gaussiani bivariati in Esempio II.8.5. È inoltre possibile verificare che se  $X \sim \mathcal{N}(m, \mathbf{R})$ , allora anche le sue componenti  $\xi_k \sim \mathcal{N}(m_k, \sigma_k^2)$  sono Gaussiane con  $\sigma_k^2 = r_{kk}$ . La prova è molto semplice se si osserva che in generale, dato un arbitrario vett.a.  $X = (\xi_1, \dots, \xi_n)$  con f.c. congiunta e marginali

$$\varphi_X(t_1, \dots, t_n) = \mathbf{E}(e^{i(t, X)}), \quad \varphi_{\xi_k}(t_k) = \mathbf{E}(e^{it_k \xi_k}),$$

si ha anche  $\varphi_{\xi_k}(t_k) = \varphi_X(0, \dots, t_k, \dots, 0)$ . Pertanto per un vett.a. normale  $X \sim \mathcal{N}(m, \mathbf{R})$  assegnato, secondo la Definizione II.11.3, tramite la sua f.c.  $\varphi_X$  si ha che le f.c. marginali sono del tipo

$$\varphi_{\xi_k}(t_k) = e^{it_k m_k} e^{-t_k^2 \sigma_k^2 / 2},$$

cioè, sempre in base alla Definizione II.11.3, sono del tipo  $\xi_k \sim \mathcal{N}(m_k, \sigma_k^2)$ . ○

**II.11.6 Proposizione:** Le componenti del vett.a. Gaussiano  $X = (\xi_1, \dots, \xi_n)$  sono indipendenti se e solo se esse sono non correlate.

**Dimostrazione:** Questo risultato è stato già discusso, in forma semplificata, nell'Esempio II.8.5: in questa sede estenderemo la dimostrazione al caso più generale comprendente (grazie alla nostra nuova Definizione II.11.3) anche i casi degeneri. Innanzitutto il Teorema II.6.15 ci garantisce che se le  $\xi_k$  sono indipendenti esse sono anche non correlate. Viceversa, se le componenti di un vett.a. Gaussiano sono non correlate la matrice delle covarianze  $\mathbf{R}$  risulta diagonale con  $r_{kl} = \delta_{kl} \sigma_k^2$  (con  $\delta_{kl} = 1$  oppure 0 secondo che  $k = l$  oppure  $k \neq l$ ) e pertanto la sua f.c. ha la forma

$$\varphi_X(t_1, \dots, t_n) = e^{i(t, m)} e^{-\sum_k t_k^2 \sigma_k^2 / 2} = \prod_{k=1}^n (e^{it_k m_k} e^{-t_k^2 \sigma_k^2 / 2}) = \prod_{K=1}^n \varphi_{\xi_k}(t_k)$$

dove le  $\varphi_{\xi_k}(t_k)$  sono, come visto nell'Osservazione II.11.5, le f.c. delle singole componenti. Ne segue, in base al Teorema II.10.11, che le componenti di  $X$  sono indipendenti. □

**II.11.7 Proposizione:** Un vett.a.  $X = (\xi_1, \dots, \xi_n)$  è Gaussiano  $\mathcal{N}(m, \mathbf{R})$  se e solo se ogni v.a.  $(\lambda, X)$ , con  $\lambda \in \mathbf{R}^n$ , è Gaussiana  $\mathcal{N}((\lambda, m), (\lambda, \mathbf{R}\lambda))$ .

**Dimostrazione:** Se  $X \sim \mathcal{N}(m, \mathbf{R})$ , posto

$$Y = (\lambda, X) = \sum_{k=1}^n \lambda_k \xi_k, \quad \lambda = (\lambda_1, \dots, \lambda_n) \in \mathbf{R}^n$$

si ha, con  $s \in \mathbf{R}$  e  $s\lambda = (s\lambda_1, \dots, s\lambda_n)$ , che

$$\varphi_Y(s) = \mathbf{E}(e^{isY}) = \mathbf{E}(e^{i(s\lambda, X)}) = e^{i(s\lambda, m)} e^{-(s\lambda, R s\lambda)/2} = e^{is(\lambda, m)} e^{-s^2(\lambda, R\lambda)/2},$$

e quindi dalla Definizione II.11.3 risulta  $Y \sim \mathcal{N}((\lambda, m), (\lambda, R\lambda))$ . Notiamo in particolare che se  $\lambda = (0, \dots, 1, \dots, 0)$  (cioè  $\lambda_j = \delta_{jk}$  per un qualche  $k$ ) si ha  $Y = \xi_k$  e la proposizione si riduce ad affermare che  $\xi_k \sim \mathcal{N}(m_k, r_{kk})$ , risultato già ottenuto per altra via in Osservazione II.11.5. Viceversa, se  $Y = (\lambda, X)$  è una v.a. Gaussiana del tipo  $\mathcal{N}(\mathbf{E}Y, \mathbf{V}Y)$  comunque dato  $\lambda \in \mathbf{R}^n$ , avremo dalla Definizione II.11.3 che

$$\varphi_Y(1) = \mathbf{E}(e^{iY}) = e^{i\mathbf{E}Y} e^{\mathbf{V}Y/2}.$$

In particolare, ponendo  $\lambda = t = (t_1, \dots, t_n) \in \mathbf{R}^n$  e usando come al solito le notazioni  $m_k = \mathbf{E}\xi_k$ ,  $r_{kl} = \mathbf{cov}(\xi_k, \xi_l)$ ,  $R = \|r_{kl}\|$ , avremo  $Y = (t, X)$  e

$$\begin{aligned} \mathbf{E}Y &= \mathbf{E}[(t, X)] = \sum_{k=1}^n t_k \mathbf{E}\xi_k = (t, m) \\ \mathbf{V}Y &= \mathbf{V}[(t, X)] = \mathbf{E}[(t, X)^2 - (\mathbf{E}(t, X))^2] = \sum_{k,l=1}^n t_k t_l [\mathbf{E}(\xi_k \xi_l) - \mathbf{E}\xi_k \mathbf{E}\xi_l] \\ &= \sum_{k,l=1}^n t_k t_l \mathbf{cov}(\xi_k, \xi_l) = \sum_{k,l=1}^n t_k t_l r_{kl} = (t, R t). \end{aligned}$$

Ne segue che

$$\varphi_X(t) = \mathbf{E}(e^{i(t, X)}) = \mathbf{E}(e^{iY}) = \varphi_Y(1) = e^{i\mathbf{E}Y} e^{\mathbf{V}Y/2} = e^{i(t, m)} e^{-(t, R t)},$$

cioè che  $X \sim \mathcal{N}(m, R)$ . □

**II.11.8 Proposizione:** Un vett.a.  $X = (\xi_1, \dots, \xi_n)$  con  $m_k = \mathbf{E}\xi_k = 0$ , per  $k = 1, \dots, n$ , e con componenti  $\xi_k$  *linearmente* indipendenti (vedi Osservazione II.9.17) è un vett.a. Gaussiano se e solo se esiste un altro vett.a. Gaussiano  $Y = (\eta_1, \dots, \eta_n) \sim \mathcal{N}(0, I)$  con componenti normali standard *statisticamente* indipendenti  $\eta_k \sim \mathcal{N}(0, 1)$  e una matrice  $n \times n$  non singolare  $C$  tali che risulti  $X = CY$ . In tal caso, se  $X \sim \mathcal{N}(0, R)$ , la matrice delle covarianze soddisfa la relazione  $R = CC^T$ .

**Dimostrazione:** Questa proposizione afferma in sostanza che le componenti di un vett.a. Gaussiano (con medie nulle) risultano sempre combinazioni lineari omogenee di v.a. normali standard indipendenti  $\eta_k \sim \mathcal{N}(0, 1)$ . Inoltre la richiesta di indipendenza lineare delle componenti di  $X$  non è realmente limitativa: infatti se essa non fosse rispettata ci sarebbero  $r \geq 1$  relazioni lineari fra le  $\xi_k$  e quindi  $r$  delle componenti di  $X$  potrebbero essere determinate a partire dalle restanti  $n - r$  (che sono invece linearmente indipendenti). In tal caso sarà sufficiente applicare



il risultato alle  $n - r$  componenti linearmente indipendenti: le rimanenti  $r$ , ricavate successivamente dalle relazioni che le legano alle prime  $n - r$ , risulteranno poi ancora combinazioni lineari omogenee delle componenti (normali standard e statisticamente indipendenti) di  $Y$ . Infine se anche l'ipotesi  $m_k = 0$  non fosse rispettata la Proposizione sarebbe ancora valida in una forma leggermente modificata. Infatti sarebbe sufficiente osservare che essendo le componenti (ovviamente con attese nulle) del vett.a.  $X - m$  combinazioni lineari omogenee delle componenti (normali standard e statisticamente indipendenti) di un opportuno vett.a.  $Y$ , le componenti di  $X$  risulterebbero ora combinazioni lineari *non omogenee* delle componenti di  $Y$  secondo la relazione  $X = CY + m$ . In definitiva la proposizione afferma che un vett.a.  $X$  è Gaussiano se e solo se le sue componenti risultano combinazioni lineari (eventualmente non omogenee) di v.a. normali standard statisticamente indipendenti.

Sulla base delle ipotesi fatte, dato che le  $\xi_k$  sono linearmente indipendenti, la matrice  $R$  delle covarianze di  $X$  risulterà non singolare (cioè  $|R| > 0$ , ricorda infatti che  $R$  è non negativa; vedi a questo proposito le Osservazioni II.8.3 e II.9.17). Infatti siccome per ipotesi  $\sum_k c_k \xi_k = 0$  se e solo se  $c_k = 0$  per ogni  $k$ , il sistema omogeneo di equazioni nelle incognite  $c_k$

$$\sum_{k=1}^n c_k (\xi_k, \xi_l) = \sum_{k=1}^n c_k \mathbf{E}(\xi_k \xi_l) = \sum_{k=1}^n c_k r_{kl} = 0$$

ammetterà solo la soluzione banale  $c_1 = \dots = c_n = 0$ , il che richiede che il determinante  $|r_{kl}| = |R|$  dei suoi coefficienti sia diverso da zero. Se allora  $O$  è la matrice ortogonale ( $OO^T = O^T O = I$ ) che diagonalizza  $R$ , la matrice diagonale  $D = O^T R O$  avrà elementi diagonali non nulli  $d_k > 0$ . Se poi  $B$  è la matrice diagonale tale che  $B^2 = D$  (gli elementi diagonali di  $B$  sono cioè  $+\sqrt{d_k}$ : vedi Teorema II.8.4), risulta

$$Y = B^{-1} O^T X \sim \mathcal{N}(0, I)$$

ossia il vett.a.  $Y = (\eta_1, \dots, \eta_n)$  è Gaussiano con componenti normali standard  $\eta_k \sim \mathcal{N}(0, 1)$  indipendenti (cioè non correlate: vedi Proposizione II.11.6). Infatti si ha per la f.c. di  $Y$  che

$$\begin{aligned} \varphi_Y(t) &= \mathbf{E}(e^{i(t, Y)}) = \mathbf{E}(e^{i(t, B^{-1} O^T X)}) = \mathbf{E}(e^{i(OB^{-1} t, X)}) = \varphi_X(OB^{-1} t) \\ &= e^{-(OB^{-1} t, ROB^{-1} t)/2} = e^{-t^T B^{-1} O^T R O B^{-1} t/2} = e^{-t^T B^{-1} D B^{-1} t/2} = e^{-(t, t)/2}, \end{aligned}$$

cioè  $Y \sim \mathcal{N}(0, I)$  e inoltre evidentemente  $X = CY$  con  $C = OB$ . Infine si ha

$$R = O D O^T = O B^2 O^T = (OB)(OB)^T = C C^T.$$

Viceversa se  $Y \sim \mathcal{N}(0, I)$  e  $X = CY$  con  $C$  non singolare, si ha

$$\begin{aligned} \varphi_X(t) &= \mathbf{E}(e^{i(t, X)}) = \mathbf{E}(e^{i(t, CY)}) = \mathbf{E}(e^{i(C^T t, Y)}) = \varphi_Y(C^T t) \\ &= e^{-(C^T t, C^T t)/2} = e^{-(t, C C^T t)/2} = e^{-(t, R t)/2}, \end{aligned}$$

cioè  $X \sim \mathcal{N}(0, R)$  avendo posto  $R = C C^T$ . □

## II.12 Esercizi svolti

**II.12.1 Esercizio (A.A. Sveshnikov: *Problems in probability theory, mathematical statistics and theory of random functions*; Dover, New York, 1978, p. 63):** La fdd di una v.a.  $X$  vale

$$f(x) = \frac{x}{a^2} e^{-x^2/2a^2}, \quad x \geq 0,$$

ed è nulla per  $x < 0$ . Determinarne il VdA la Var e i *momenti centrati* del terzo e del quarto ordine:

$$\mu_3 = \mathbf{E}(X - \mathbf{E}X)^3 \quad \mu_4 = \mathbf{E}(X - \mathbf{E}X)^4.$$

**Soluzione:** Ricordando che per gli integrali ( $n = 1, 2, \dots$ )

$$J_n = \int_0^\infty t^n e^{-t^2} dt$$

valgono le relazioni

$$J_{2k} = \frac{1}{2} \Gamma\left(k + \frac{1}{2}\right) = \frac{(2k-1)!!}{2^{k+1}} \sqrt{\pi}, \quad J_{2k+1} = \frac{1}{2} \Gamma(k+1) = \frac{k!}{2},$$

si ha facilmente

$$\mathbf{E}X = \int_0^\infty x f(x) dx = \frac{1}{a^2} \int_0^\infty x^2 e^{-x^2/2a^2} dx = 2\sqrt{2}a \int_0^\infty t^2 e^{-t^2} dt = a\sqrt{\frac{\pi}{2}};$$

$$\mathbf{V}X = \mathbf{E}X^2 - (\mathbf{E}X)^2 = 4a^2 J_3 - a^2 \frac{\pi}{2} = a^2 \left(2 - \frac{\pi}{2}\right);$$

$$\mu_3 = \mathbf{E}X^3 - 3\mathbf{E}X\mathbf{E}X^2 + 2(\mathbf{E}X)^3 = 4\sqrt{2}a^3 J_4 - 12a^3 \sqrt{\frac{\pi}{2}} J_3 + a^3 \pi \sqrt{\frac{\pi}{2}}$$

$$= a^3(\pi - 3)\sqrt{\frac{\pi}{2}};$$

$$\mu_4 = \mathbf{E}X^4 - 4\mathbf{E}X\mathbf{E}X^3 + 6(\mathbf{E}X)^2\mathbf{E}X^2 - 3(\mathbf{E}X)^4 = a^4 \left(8 - \frac{3}{4}\pi^2\right).$$

La fdd qui considerata prende anche il nome di *legge di Rayleigh* dal nome del fisico inglese John W.S. Rayleigh (1842-1919) che si distinse per i suoi importanti studi in acustica, elettrodinamica, meccanica dei fluidi, termodinamica e che ebbe il Premio Nobel per la chimica nel 1904 per la scoperta del gas inerte argon.  $\diamond$

**II.12.2 Esercizio (A.A. Sveshnikov: *Problems in probability theory, mathematical statistics and theory of random functions*; Dover, New York, 1978, p. 87):** Il vett.a.  $(X, Y)$  è dotato di fdd eguale a

$$f(x, y) = 0,5 \sin(x + y), \quad 0 \leq x \leq \frac{\pi}{2}, \quad 0 \leq y \leq \frac{\pi}{2},$$

e nulla altrove. Determinare i VdA di  $X$  ed  $Y$  e la matrice delle covarianze.

**Soluzione:** Il VdA di  $X$  è dato da

$$\begin{aligned} \mathbf{E}X &= 0,5 \int_0^{\pi/2} \int_0^{\pi/2} x \sin(x+y) dx dy \\ &= 0,5 \int_0^{\pi/2} x \left[ -\cos\left(x + \frac{\pi}{2}\right) + \cos x \right] dx = \frac{\pi}{4} = 0,785, \end{aligned}$$

e inoltre  $\mathbf{E}Y = \mathbf{E}X$  a causa della simmetria della fdd. Anche le due varianze  $\mathbf{V}X = \mathbf{V}Y$  coincidono e valgono

$$\begin{aligned} \mathbf{V}X &= 0,5 \int_0^{\pi/2} \int_0^{\pi/2} x^2 \sin(x+y) dx dy - \frac{\pi^2}{16} \\ &= 0,5 \int_0^{\pi/2} x^2 \left[ -\cos\left(x + \frac{\pi}{2}\right) + \cos x \right] dx - \frac{\pi^2}{16} = \frac{\pi^2}{16} + \frac{\pi}{2} - 2 = 0,188. \end{aligned}$$

La covarianza è invece data da

$$\begin{aligned} \mathbf{cov}(X, Y) &= 0,5 \int_0^{\pi/2} \int_0^{\pi/2} xy \sin(x+y) dx dy - \frac{\pi^2}{16} \\ &= 0,5 \int_0^{\pi/2} x \left[ \sin\left(x + \frac{\pi}{2}\right) - \sin x - \frac{\pi}{2} \cos\left(x + \frac{\pi}{2}\right) \right] dx - \frac{\pi^2}{16} \\ &= \frac{\pi}{2} - 1 - \frac{\pi^2}{16} = -0,046, \end{aligned}$$

sicché

$$\|r_{ij}\| = \begin{pmatrix} 0,188 & -0,046 \\ -0,046 & 0,188 \end{pmatrix}$$

sarà la nostra matrice delle covarianze. ◇

**II.12.3 Esercizio (E. Wentzel e L. Ovcharov: *Applied problems in probability theory*; Mir, Mosca, 1986, p. 122):** Una v.a.  $X$  ha fdd  $f_1(x)$  con probabilità  $p_1$  ed  $f_2(x)$  con probabilità  $p_2$  ( $p_1 + p_2 = 1$ ). Determinare la fdd e la FdD di  $X$  e calcolarne il VdA.

**Soluzione:** Dalla formula della probabilità totale (vedi Proposizione I.4.5) si ha

$$F(x) = \mathbf{P}(X \leq x) = p_1 F_1(x) + p_2 F_2(x)$$

dove

$$F_1(x) = \int_{-\infty}^x f_1(t) dt \quad F_2(x) = \int_{-\infty}^x f_2(t) dt,$$

sicché per derivazione si ha

$$f(x) = F'(x) = p_1 f_1(x) + p_2 f_2(x).$$

Conseguentemente per il VdA si ha

$$\mathbf{E} X = \int_{-\infty}^{+\infty} x f(x) dx = p_1 \int_{-\infty}^{+\infty} x f_1(x) dx + p_2 \int_{-\infty}^{+\infty} x f_2(x) dx. \quad \diamond$$

**II.12.4 Esercizio (E. Wentzel e L. Ovcharov:** *Applied problems in probability theory*; Mir, Mosca, 1986, p. 124): Un campo di punti di tipo poissonniano (omogeneo) è una collezione di punti diffusa aleatoriamente su un piano in modo tale che 1) gli eventi del tipo *in una data regione del piano cadono n punti* sono tutti indipendenti se tali regioni non si sovrappongono e 2) la probabilità che in una regione di area  $s$  cadano  $n \geq 0$  punti è

$$P_n = e^{-\lambda s} \frac{(\lambda s)^n}{n!}$$

dove la costante  $\lambda$  si dice *intensità* del campo e rappresenta il VdA del numero di punti che cadono in una regione di area unitaria. Dato un campo poissonniano determinare la fdd, il VdA e la Var della distanza (aleatoria)  $R$  di un punto del campo dal suo vicino più prossimo.

**Soluzione:** Per determinare la FdD  $F(r)$  di  $R$  tracciamo attorno ad un generico punto del piano un cerchio di raggio  $r$ : in tal caso l'evento  $\{R \leq r\}$  coinciderà con l'evento *nel cerchio cade almeno un punto* (oltre quello che si trova nel centro) sicché, essendo ora  $s = \pi r^2$ , si avrà per  $r \geq 0$

$$F(r) = \mathbf{P}(R \leq r) = \sum_{n \geq 1} P_n = 1 - P_0 = 1 - e^{-\lambda \pi r^2}.$$

Ne segue, per derivazione, che la fdd è

$$f(r) = 2\pi\lambda r e^{-\lambda \pi r^2}$$

cioè  $R$  è una v.a. di Rayleigh (vedi Esercizio II.12.1) con  $a = (2\pi\lambda)^{-1/2}$ . I risultati dell'Esercizio II.12.1 ci consentono allora di concludere che

$$\mathbf{E} R = \frac{1}{2\sqrt{\lambda}}, \quad \mathbf{V} R = \frac{4 - \pi}{4\pi\lambda}. \quad \diamond$$

**II.12.5 Esercizio (A.A. Svishnikov:** *Problems in probability theory, mathematical statistics and theory of random functions*; Dover, New York, 1978, p. 71): La misura della distanza da un dato oggetto è una v.a. a causa di un errore *sistematico* di 50 m (nel verso della diminuzione della distanza) e di un errore *casuale* distribuito in maniera normale con media nulla e  $\sigma = 100$  m. Determinare (1) la probabilità che l'errore non ecceda 150 m in valore assoluto, e (2) la probabilità che la distanza misurata non ecceda quella reale.

**Soluzione:** Se  $X$  è la v.a. che rappresenta l'errore totale, esso sarà distribuito in modo normale secondo  $X \sim \mathcal{N}(-50, 10.000)$  e quindi, tenendo conto delle osservazioni fatte in Esempio II.8.11, la v.a.  $Y = (X + 50)/100$  sarà normale standard  $Y \sim \mathcal{N}(0, 1)$ . Ne segue che (vedi Tavola I.8.1)

$$\begin{aligned} \mathbf{P}(|X| < 150) &= \mathbf{P}(-150 < X < 150) = \mathbf{P}(-1 < Y < 2) = \Phi(2) - \Phi(-1) \\ &= \Phi(2) + \Phi(1) - 1 = 0,8185. \end{aligned}$$

Inoltre la probabilità che la distanza misurata non ecceda quella reale coincide con la probabilità che  $X < 0$ , sicché

$$\mathbf{P}(X < 0) = \mathbf{P}(Y < 0,5) = \Phi(0,5) = 0,6915. \quad \diamond$$

**II.12.6 Esercizio (E. Wentzel e L. Ovcharov: *Applied problems in probability theory*; Mir, Mosca, 1986, p. 130):** Data una v.a.  $X \sim \mathcal{N}(0, \sigma^2)$  e preso un intervallo  $[a, b]$  che non contiene l'origine, determinare il valore di  $\sigma$  che rende massimo il valore di  $\mathbf{P}(X \in [a, b])$ .

**Soluzione:** Siccome  $Y = X/\sigma \sim \mathcal{N}(0, 1)$ , la dipendenza di  $\mathbf{P}(X \in [a, b])$  da  $\sigma$  è data dalla relazione

$$\mathbf{P}(X \in [a, b]) = \Phi(b/\sigma) - \Phi(a/\sigma).$$

Siccome risulta

$$\Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

avremo che

$$\frac{d}{d\sigma} \mathbf{P}(X \in [a, b]) = \frac{1}{\sqrt{2\pi}} \left[ -\frac{b}{\sigma^2} e^{-b^2/2\sigma^2} + \frac{a}{\sigma^2} e^{-a^2/2\sigma^2} \right],$$

e quindi imponendo

$$\frac{d}{d\sigma} \mathbf{P}(X \in [a, b]) = 0$$

avremo l'equazione

$$b e^{-b^2/2\sigma^2} = a e^{-a^2/2\sigma^2}$$

la cui soluzione

$$\sigma = \sqrt{\frac{b^2 - a^2}{2(\ln b - \ln a)}} = \sqrt{\frac{b+a}{2} \frac{b-a}{\ln b - \ln a}}$$

fornisce il risultato richiesto. ◇

**II.12.7 Esercizio (E. Wentzel e L. Ovcharov: *Applied problems in probability theory*; Mir, Mosca, 1986, p. 130):** È necessario approssimare la v.a.

$X \sim \mathcal{N}(m, \sigma^2)$  con una v.a.  $Y$  distribuita uniformemente in  $[a, b]$ : determinare i valori di  $a$  e  $b$  in modo tale che  $\mathbf{E}X = \mathbf{E}Y$  e  $\mathbf{V}X = \mathbf{V}Y$ .

**Soluzione:** La v.a.  $Y$  è dotata della seguente fdd

$$f(y) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & \text{altrove} \end{cases}$$

per cui si ha facilmente

$$\mathbf{E}Y = \frac{a+b}{2}, \quad \mathbf{V}Y = \frac{(b-a)^2}{12}.$$

Imponendo allora le condizioni richieste

$$\frac{a+b}{2} = m, \quad \frac{(b-a)^2}{12} = \sigma^2$$

si ottengono i seguenti valori per gli estremi dell'intervallo:

$$a = m - \sqrt{3}\sigma, \quad b = m + \sqrt{3}\sigma. \quad \diamond$$

**II.12.8 Esercizio (E. Wentzel e L. Ovcharov: *Applied problems in probability theory*; Mir, Mosca, 1986, p. 134):** Un voltaggio aleatorio  $V$  caratterizzato da una fdd  $f(v)$  passa attraverso un limitatore di voltaggio che taglia tutti i valori più piccoli di  $v_1$  (innalzandoli a tale valore) e più grandi di  $v_2$  (abbassandoli a tale valore) con  $v_1 < v_2$ . Determinare la distribuzione del voltaggio aleatorio  $U$  che si ottiene in uscita.

**Soluzione:** La v.a.  $U$  coincide con  $V$  se  $v_1 < V < v_2$  mentre vale  $v_1$  se  $V < v_1$  e  $v_2$  se  $v_2 < V$ . Conseguentemente  $U$  è distribuita come  $V$  su tutti i valori  $u \in (v_1, v_2)$  e inoltre assume con probabilità diversa da zero i due valori  $v_1$  e  $v_2$ . Ovviamente

$$\mathbf{P}(U = v_1) = p_1 = \mathbf{P}(V < v_1) = \int_{-\infty}^{v_1} f(v) dv = F_V(v_1),$$

$$\mathbf{P}(U = v_2) = p_2 = \mathbf{P}(V > v_2) = \int_{v_2}^{+\infty} f(v) dv = 1 - F_V(v_2).$$

Pertanto  $U$  è una v.a. *mista* nel senso che è dotata di densità fra  $v_1$  e  $v_2$  e inoltre assume, con probabilità non nulla, i due valori discreti  $v_1$  e  $v_2$ . La sua FdD  $F(u)$  è quindi nulla per  $u < v_1$ ; vale

$$F(u) = \int_{-\infty}^u f(t) dt = F_V(u)$$

per  $v_1 \leq u < v_2$ , e resta costantemente eguale ad 1 per  $v_2 \leq u$ . Ne segue che, ad esempio, il VdA di  $U$  sarà

$$\mathbf{E}U = v_1 p_1 + v_2 p_2 + \int_{v_1}^{v_2} f(v) dv. \quad \diamond$$

**II.12.9 Esercizio (E. Wentzel e L. Ovcharov: *Applied problems in probability theory*; Mir, Mosca, 1986, p. 141):** Il tempo di vita di un apparecchiatura è una v.a.  $T$  dotata di fdd  $f(t)$  ( $t > 0$ ). All'istante  $t_0$ , se l'apparecchiatura non si è ancora rotta, viene effettuata una manutenzione preventiva e successivamente essa opera con un tempo di vita (aleatorio)  $T_1$  con fdd  $f_1(t)$ . Se invece l'apparecchiatura si rompe ad un istante  $s < t_0$  essa viene immediatamente riparata e successivamente opera con un tempo di vita (aleatorio)  $T_2$  con fdd  $f_2(t)$ . Non è prevista nessun'altra riparazione. Determinare il VdA del tempo di vita complessivo  $\Theta$  della nostra apparecchiatura (escludendo il tempo necessario per riparazioni e manutenzioni).

**Soluzione:** Se  $T$  assume un valore  $t < t_0$  si ha

$$\mathbf{E}(\Theta | T = t) = t + t_2 \quad \text{dove} \quad t_2 = \int_0^{+\infty} t f_2(t) dt;$$

se invece  $T$  assume un valore  $t \geq t_0$  si ha

$$\mathbf{E}(\Theta | T = t) = t_0 + t_1 \quad \text{dove} \quad t_1 = \int_0^{+\infty} t f_1(t) dt.$$

Pertanto

$$\mathbf{E}(\Theta | T = t) = \begin{cases} t + t_2, & \text{se } t < t_0; \\ t_0 + t_1, & \text{se } t \geq t_0. \end{cases}$$

L'attesa non condizionata sarà allora

$$\begin{aligned} \mathbf{E}\Theta &= \mathbf{E}[\mathbf{E}(\Theta | T)] = \int_0^{+\infty} \mathbf{E}(\Theta | T = t) f(t) dt \\ &= \int_0^{t_0} (t + t_2) f(t) dt + \int_{t_0}^{+\infty} (t_0 + t_1) f(t) dt \\ &= \int_0^{t_0} t f(t) dt + t_2 F(t_0) + (t_0 + t_1)(1 - F(t_0)) \end{aligned}$$

dove

$$F(t) = \int_0^t f(s) ds \quad t > 0$$

è la FdD della v.a.  $T$ . ◇

**II.12.10 Esercizio (E. Wentzel e L. Ovcharov:** *Applied problems in probability theory*; Mir, Mosca, 1986, p. 153): Un vett.a. con due componenti  $(X, Y)$  ha fdd congiunta

$$f(x, y) = \frac{a}{1 + x^2 + x^2y^2 + y^2} = \frac{a}{(1 + x^2)(1 + y^2)};$$

(1) determinare il coefficiente  $a$  e le fdd marginali  $f_X(x)$  e  $f_Y(y)$ ; (2) discutere l'indipendenza di  $X$  e  $Y$ ; (3) determinare la probabilità che il punto aleatorio  $(X, Y)$  cada nel quadrato  $Q$  del piano  $x, y$  con centro nell'origine e lati paralleli agli assi coordinati e di lunghezza  $b = 2$ .

**Soluzione:** Il coefficiente  $a$  si calcola imponendo la condizione di normalizzazione

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$$

da cui si ricava che  $a = 1/\pi^2$ . Le fdd marginali sono

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy = \frac{1}{\pi(1 + x^2)} \quad f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx = \frac{1}{\pi(1 + y^2)},$$

da cui discende anche che  $X$  ed  $Y$  sono indipendenti dato che la fdd congiunta  $f(x, y)$  si fattorizza nel prodotto  $f_X(x)f_Y(y)$ . Infine

$$\mathbf{P}((X, Y) \in Q) = \int_{-1}^{+1} \int_{-1}^{+1} \frac{dx dy}{\pi^2(1 + x^2)(1 + y^2)} = \frac{1}{4}. \quad \diamond$$

**II.12.11 Esercizio (E. Wentzel e L. Ovcharov:** *Applied problems in probability theory*; Mir, Mosca, 1986, p. 161): Dato il vett.a.  $(X, Y)$ , sia noto che  $X$  è esponenziale con fdd  $f_X(x) = \lambda e^{-\lambda x}$  per  $x > 0$ . Per un dato valore  $X = x > 0$  sappiamo che anche  $Y$  è esponenziale in modo che  $f_{Y|X}(y|x) = x e^{-xy}$  per  $y > 0$ . determinare la fdd congiunta  $f(x, y)$ , la marginale  $f_Y(y)$  e la fdd condizionata  $f_{X|Y}(x|y)$ .

**Soluzione:** La fdd congiunta è

$$f(x, y) = f_{Y|X}(y|x)f_X(x) = \begin{cases} \lambda x e^{-(\lambda+y)x}, & \text{se } x > 0 \text{ e } y > 0; \\ 0, & \text{se } x \leq 0 \text{ oppure } y \leq 0; \end{cases}$$

la fdd marginale di  $Y$  è allora

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx = \begin{cases} \lambda/(\lambda + y)^2, & \text{se } y > 0, \\ 0, & \text{se } y \leq 0; \end{cases}$$

infine per  $y > 0$  si ha

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} = \begin{cases} x(\lambda + y)^2 e^{-(\lambda+y)x}, & \text{se } x > 0, \\ 0, & \text{se } x \leq 0; \end{cases} \quad \diamond$$



**II.12.12 Esercizio (A.A. Sveshnikov:** *Problems in probability theory, mathematical statistics and theory of random functions*; Dover, New York, 1978, p. 80): La probabilità dell'evento  $A$  condizionata da  $X = x$  è espressa da

$$\mathbf{P}(A|X = x) = \begin{cases} 1 - e^{-kx} & \text{per } x \geq 0, \\ 0 & \text{per } x < 0. \end{cases}$$

Determinare  $\mathbf{P}(A)$  sotto l'ipotesi che  $X \sim \mathcal{N}(m, \sigma^2)$ .

**Soluzione:** Basterà osservare che

$$\mathbf{P}(A) = \int_{-\infty}^{+\infty} \mathbf{P}(A|X = x)f(x) dx, \quad \text{dove} \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-m)^2/2\sigma^2}.$$

Si ha in tal caso che, introducendo la solita funzione errore standard  $\Phi(x)$ ,

$$\mathbf{P}(A) = \Phi\left(\frac{m}{\sigma}\right) - \frac{1}{\sigma\sqrt{2\pi}} \int_0^{+\infty} e^{-(x-m)^2/2\sigma^2} e^{-kx} dx$$

e siccome

$$-\frac{(x-m)^2}{2\sigma^2} - kx = -\frac{(x-m+k\sigma^2)^2}{2\sigma^2} - k\left(m - \frac{k\sigma^2}{2}\right)$$

dalla relazione

$$\frac{1}{\sigma\sqrt{2\pi}} \int_0^{+\infty} e^{-(x-m+k\sigma^2)^2/2\sigma^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-(m-k\sigma^2)/\sigma}^{+\infty} e^{-t^2/2} dt = \Phi\left(\frac{m-k\sigma^2}{\sigma}\right)$$

si ottiene il risultato richiesto

$$\mathbf{P}(A) = \Phi\left(\frac{m}{\sigma}\right) - \Phi\left(\frac{m-k\sigma^2}{\sigma}\right) e^{-km+k^2\sigma^2/2}. \quad \diamond$$

**II.12.13 Esercizio (A.A. Sveshnikov:** *Problems in probability theory, mathematical statistics and theory of random functions*; Dover, New York, 1978, p. 108): Determinare il VdA e la Var della v.a.  $Y = e^{aX}$  se  $X$  è una v.a. binomiale.

**Soluzione:** Dato che  $X$  assumerà solo i valori interi  $0, 1, \dots, n$  si ha facilmente con ovvio significato dei simboli che

$$\begin{aligned} \mathbf{E}Y &= \sum_{k=0}^n e^{ak} \binom{n}{k} p^k q^{n-k} = (q + pe^a)^n \\ \mathbf{V}Y &= \mathbf{E}Y^2 - (\mathbf{E}Y)^2 = \sum_{k=0}^n e^{2ak} \binom{n}{k} p^k q^{n-k} - (\mathbf{E}Y)^2 \\ &= (q + pe^{2a})^n - (q + pe^a)^{2n}. \end{aligned} \quad \diamond$$

**II.12.14 Esercizio (A.A. Sveshnikov:** *Problems in probability theory, mathematical statistics and theory of random functions*; Dover, New York, 1978, p. 109): Lo schermo di un radar è un cerchio di raggio  $a$ . A causa del rumore alcuni punti luminosi possono apparire in maniera casuale e la loro distribuzione sullo schermo è uniforme. Determinare il VdA e la Var della distanza (aleatoria)  $R$  dei punti luminosi dal centro dello schermo.

**Soluzione:** Il punto aleatorio  $(X, Y)$  sul piano dello schermo dotato di un sistema di assi con origine nel centro è caratterizzato da una fdd uniforme del tipo

$$f(x, y) = \begin{cases} 1/\pi a^2 & \text{se } x^2 + y^2 \leq a^2, \\ 0 & \text{se } x^2 + y^2 > a^2. \end{cases}$$

Ne segue che per la v.a.  $R = \sqrt{X^2 + Y^2}$ , detto  $D$  il cerchio  $x^2 + y^2 \leq a^2$ , si ha con un cambiamento di variabili in coordinate polari che

$$\begin{aligned} \mathbf{E} R &= \frac{1}{\pi a^2} \int \int_D \sqrt{x^2 + y^2} dx dy = \frac{1}{\pi a^2} \int_0^{2\pi} d\phi \int_0^a r^2 dr = \frac{2a}{3}, \\ \mathbf{V} R &= \frac{1}{\pi a^2} \int \int_D (x^2 + y^2) dx dy - (\mathbf{E} R)^2 \\ &= \frac{1}{\pi a^2} \int_0^{2\pi} d\phi \int_0^a r^3 dr - \frac{4a^2}{9} = \frac{a^2}{18}. \end{aligned} \quad \diamond$$

**II.12.15 Esercizio (A.A. Sveshnikov:** *Problems in probability theory, mathematical statistics and theory of random functions*; Dover, New York, 1978, p. 117): Determinare la fdd della v.a.  $Y = +\sqrt{|X|}$  se  $X \sim \mathcal{N}(0, 1)$ .

**Soluzione:** Siccome la funzione  $y = \varphi(x) = +\sqrt{|x|}$  non è monotona bisognerà considerare separatamente gli intervalli su cui la restrizione di  $\varphi(x)$  risulta invertibile. Posto allora

$$y = \begin{cases} \varphi_1(x) = \sqrt{x} & \text{se } x \geq 0 \\ \varphi_2(x) = \sqrt{-x} & \text{se } x < 0 \end{cases}$$

siccome  $\varphi_1$  e  $\varphi_2$  risultano ora monotone potremo definire le funzioni inverse (per  $y \geq 0$ )

$$h_1(y) = \varphi_1^{-1}(y) = y^2, \quad h_2(y) = \varphi_2^{-1}(y) = -y^2.$$

Dai risultati di Osservazione II.8.10 si ha allora che per  $y \geq 0$

$$f_Y(y) = f_X(y^2)|h_1'(y)| + f_X(-y^2)|h_2'(y)| = \frac{4y}{\sqrt{2\pi}} e^{-y^4/2},$$

mentre per  $y < 0$  si ha  $f_Y(y) = 0$ . ◇

**II.12.16 Esercizio (E. Wentzel e L. Ovcharov:** *Applied problems in probability theory*; Mir, Mosca, 1986, p. 172): Due v.a.  $X$  ed  $Y$  sono nella seguente relazione:

$Y = 2 - 3X$ . Se  $\mathbf{E}X = -1$  e  $\mathbf{V}X = \sigma_X^2 = 4$ , determinare VdA e Var di  $Y$  e covarianza e coefficiente di correlazione delle due v.a.

**Soluzione:** Si ha facilmente che

$$\mathbf{E}Y = 2 - 3\mathbf{E}X = 5, \quad \mathbf{V}Y = \sigma_Y^2 = (-3)^2\mathbf{V}X = 36;$$

inoltre

$$\mathbf{cov}(X, Y) = \mathbf{E}(XY) - \mathbf{E}X\mathbf{E}Y = \mathbf{E}[X(2 - 3X)] + 5 = 2\mathbf{E}X - 3\mathbf{E}X^2 + 5$$

e siccome  $\mathbf{E}X^2 = \mathbf{V}X + (\mathbf{E}X)^2 = 4 + 1 = 5$  si ha  $\mathbf{cov}(X, Y) = -12$ . Il coefficiente di correlazione è allora

$$\rho(X, Y) = -\frac{12}{\sigma_X\sigma_Y} = -1$$

risultato prevedibile sulla base della Proposizione I.6.21. ◇

**II.12.17 Esercizio (E. Wentzel e L. Ovcharov: *Applied problems in probability theory*; Mir, Mosca, 1986, p. 178):** Una v.a.  $X$  ha fdd  $f(x)$ : data la v.a.  $Y = \min\{X, a\}$  dove  $a$  è un numero dato, determinare VdA e Var di  $Y$ .

**Soluzione:** Siccome ovviamente

$$\min\{x, a\} = \begin{cases} x & \text{se } x < a \\ a & \text{se } x \geq a \end{cases}$$

si ha

$$\mathbf{E}Y = \int_{-\infty}^a xf(x) dx + a \int_a^{+\infty} f(x) dx = \int_{-\infty}^a xf(x) dx + a\mathbf{P}(X > a).$$

Inoltre da

$$\mathbf{E}Y^2 = \int_{-\infty}^a x^2 f(x) dx + a^2 \int_a^{+\infty} f(x) dx = \int_{-\infty}^a x^2 f(x) dx + a^2\mathbf{P}(X > a),$$

si ottiene anche il secondo risultato mediante  $\mathbf{V}Y = \mathbf{E}Y^2 - (\mathbf{E}Y)^2$ . ◇

**II.12.18 Esercizio (E. Wentzel e L. Ovcharov: *Applied problems in probability theory*; Mir, Mosca, 1986, p. 195):** Un foro può prodursi con egual probabilità in una qualsiasi posizione sulle sei pareti di un contenitore cubico di carburante provocando la perdita di tutto il liquido che si trova ad un'altezza superiore al foro. Il serbatoio è normalmente pieno per  $3/4$ : determinare il VdA della quantità di carburante presente dopo l'apparizione del foro.

**Soluzione:** Per semplicità assumeremo che la lunghezza del lato del serbatoio sia unitaria: se  $X$  è l'altezza alla quale si produce il foro e  $Y$  il carburante residuo, avremo

$$Y = \begin{cases} X & \text{se } 0 \leq X \leq 0,75 \\ 0,75 & \text{se } 0,75 \leq X \leq 1. \end{cases}$$

Si ha allora che, data l'uniformità della distribuzione della posizione del foro,

$$\mathbf{P}(Y = 0,75) = \mathbf{P}(X \geq 0,75) = \frac{1}{6} + 4 \frac{1}{6} 0,25 = \frac{1}{3}.$$

Inoltre si ha ovviamente che

$$\mathbf{P}(Y = 0) = \mathbf{P}(X = 0) = \frac{1}{6}.$$

Se infine il foro compare ad un'altezza  $0 < X \leq 0,75$ , allora  $Y = X$  e la v.a.  $Y$  sarà distribuita uniformemente fra 0 e 0,75. Per determinare il valore (costante) della relativa fdd basterà osservare che fra 0 e 0,75 si troverà distribuita uniformemente la probabilità residua  $1 - \frac{1}{3} - \frac{1}{6} = \frac{1}{2}$  non impegnata nei due casi esaminati in precedenza. Pertanto il valore costante della fdd sarà  $0,5/0,75 = \frac{2}{3}$ . Ne segue che

$$\mathbf{E} Y = 0,75 \frac{1}{3} + \frac{2}{3} \int_0^{0,75} x dx = 0,44. \quad \diamond$$

**II.12.19 Esercizio (E. Wentzel e L. Ovcharov: *Applied problems in probability theory*; Mir, Mosca, 1986, p. 196):** Fissato un punto  $a$  in  $[0, 1]$  si consideri un punto aleatorio  $X$  uniformemente distribuito in tale intervallo. Posto  $R = |a - X|$ , determinare il coefficiente di correlazione fra  $X$  ed  $R$  e il valore di  $a$  per il quale le due v.a. risultano non correlate.

**Soluzione:** Per calcolare  $\mathbf{cov}(X, R) = \mathbf{E}(XR) - \mathbf{E}X\mathbf{E}R$  osserviamo che

$$\begin{aligned} \mathbf{E}(XR) &= \mathbf{E}(X|a - X|) = \int_0^1 x|a - x| dx \\ &= \int_0^a x(a - x) dx - \int_a^1 x(a - x) dx = \frac{a^3}{3} - \frac{a}{2} + \frac{1}{3}, \\ \mathbf{E}X &= \frac{1}{2}, \\ \mathbf{E}R &= \int_0^1 |a - x| dx = \int_0^a (a - x) dx - \int_a^1 (a - x) dx = a^2 - a + \frac{1}{2}, \end{aligned}$$

sicché si ottiene

$$\mathbf{cov}(X, R) = \frac{a^3}{3} - \frac{a^2}{2} + \frac{1}{12}.$$

Siccome inoltre  $\mathbf{V}X = \sigma_X^2 = 1/12$  e da

$$\mathbf{E}R^2 = \int_0^1 (a - x)^2 dx = a^2 - a + \frac{1}{3}$$

risulta  $\mathbf{V}R = \sigma_R^2 = \mathbf{E}R^2 - (\mathbf{E}R)^2 = 2a^3 - a^4 - a^2 + \frac{1}{12}$ , abbiamo che

$$\rho(X, R) = \frac{\mathbf{cov}(X, R)}{\sigma_X \sigma_R} = 2\sqrt{3} \frac{\frac{a^3}{3} - \frac{a^2}{2} + \frac{1}{12}}{\sqrt{2a^3 - a^4 - a^2 + \frac{1}{12}}}.$$

Per determinare il valore di  $a$  che rende  $X$  ed  $R$  non correlate osserveremo infine che l'equazione

$$\frac{a^3}{3} - \frac{a^2}{2} + \frac{1}{12} = 0$$

ha in  $[0, 1]$  solo la radice  $a = 1/2$  che sarà quindi il valore richiesto.  $\diamond$

**II.12.20 Esercizio (E. Wentzel e L. Ovcharov: *Applied problems in probability theory*; Mir, Mosca, 1986, p. 200):** La v.a.

$$Z = \sum_{k=1}^Y X_k$$

è la somma di un numero aleatorio  $Y$  di v.a.  $X_k$  i.i.d. con VdA  $m$  e  $\text{Var } \sigma^2$ ; il numero degli addendi  $Y$  è una v.a. intera, indipendente dalle  $X_k$  con VdA  $n$  e  $\text{Var } \mu^2$ . Determinare VdA e  $\text{Var}$  di  $Z$

**Soluzione:** Dalla relazione

$$\mathbf{E}(Z | Y = j) = \sum_{k=1}^j \mathbf{E}X_k = jm$$

si ha facilmente che

$$\mathbf{E}Z = \sum_j \mathbf{E}(Z | Y = j) \mathbf{P}(Y = j) = m \sum_j j \mathbf{P}(Y = j) = m \mathbf{E}Y = mn.$$

Inoltre siccome

$$\begin{aligned} \mathbf{E}(Z^2 | Y = j) &= \mathbf{E} \left( \sum_{k=1}^j X_k \right)^2 = \mathbf{E} \left( \sum_{k=1}^j X_k^2 + 2 \sum_{k < l=1}^j X_k X_l \right) \\ &= \sum_{k=1}^j \mathbf{E}X_k^2 + 2 \sum_{k < l=1}^j \mathbf{E}X_k \mathbf{E}X_l = j(\sigma^2 + m^2) + j(j-1)m^2 \\ &= j\sigma^2 + j^2m^2, \end{aligned}$$

si ha

$$\begin{aligned} \mathbf{E}Z^2 &= \sum_j \mathbf{E}(Z^2 | Y = j) \mathbf{P}(Y = j) = \sigma^2 \sum_j j \mathbf{P}(Y = j) + m^2 \sum_j j^2 \mathbf{P}(Y = j) \\ &= \sigma^2 \mathbf{E}Y + m^2 \mathbf{E}Y^2 = \sigma^2 n + m^2(\mu^2 + n^2) \end{aligned}$$

sicch  in definitiva

$$\mathbf{V}Z = \mathbf{E}Z^2 - (\mathbf{E}Z)^2 = \sigma^2 n + m^2 \mu^2.$$

Se ad esempio  $Y$  fosse una v.a. di Poisson con parametro  $\lambda$  si avrebbe

$$\mathbf{E}Z = \lambda m, \quad \mathbf{V}Z = \lambda(\sigma^2 + m^2). \quad \diamond$$

**II.12.21 Esercizio (E. Wentzel e L. Ovcharov: *Applied problems in probability theory*; Mir, Mosca, 1986, p. 223):** Determinare la fdd della v.a.  $Y = 1/X$  se  $X$    una v.a. di Cauchy con fdd

$$f(x) = \frac{1}{\pi(1+x^2)}.$$

**Soluzione:** Poich  la funzione  $y = \varphi(x) = 1/x$    monotona, posto  $x = h(y) = 1/y$  si ha

$$g(y) = \frac{1}{\pi(1+1/y)^2} \frac{1}{y^2} = \frac{1}{\pi(1+y^2)}$$

sicch  l'inversa di una v.a. di Cauchy risulta essere ancora una v.a. di Cauchy.  $\diamond$

**II.12.22 Esercizio (E. Wentzel e L. Ovcharov: *Applied problems in probability theory*; Mir, Mosca, 1986, p. 224):** Il raggio  $R$  di un cerchio   una v.a. con distribuzione di Rayleigh

$$f(r) = \frac{r}{a^2} e^{-r^2/2a^2} \quad (r > 0).$$

Determinare la distribuzione dell'area  $S$  del cerchio.

**Soluzione:** Siccome  $S = \pi R^2$  e  $s = \varphi(r) = \pi r^2$    monotona su  $[0, +\infty)$ , la fdd dell'area  $S$  sar 

$$g(s) = \frac{1}{2\pi a^2} e^{-s/2\pi a^2},$$

cio   $S$  risulta esponenziale con parametro  $1/2\pi a^2$ .  $\diamond$

**II.12.23 Esercizio (A.A. Sveshnikov: *Problems in probability theory, mathematical statistics and theory of random functions*; Dover, New York, 1978, p. 77):** Data la v.a.  $X$  con fdd

$$f(x) = \frac{1}{2} e^{-|x|}$$

calcolarne i momenti di ogni ordine mediante l'uso della f.c.  $\varphi(t)$ .

**Soluzione:** La f.c. di  $X$    (vedi Esempio II.10.5)

$$\varphi(t) = \frac{1}{1+t^2},$$

e i momenti  $m_k = \mathbf{E} X^k$  possono essere calcolati (vedi Teorema II.10.6) come coefficienti dello sviluppo in serie di McLaurin di  $\varphi(t)$ . Siccome dallo sviluppo in serie geometrica si ha

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

è facile ottenere che

$$m_k = \begin{cases} k! & \text{per } k \text{ pari,} \\ 0 & \text{per } k \text{ dispari.} \end{cases} \quad \diamond$$





III  
SUCCESSIONI  
di VARIABILI ALEATORIE



## III.1 Teoremi Limite e Leggi dei Grandi Numeri

**III.1.1 Teorema (Teorema Limite Centrale per v.a. i.i.d.):** Se  $(\xi_n)_{n \in \mathbf{N}}$  è una successione di v.a. indipendenti, identicamente distribuite, di quadrato integrabile ( $\mathbf{E} \xi_n^2 < +\infty$ ) e non degeneri ( $\mathbf{V} \xi_n = \sigma^2 > 0$ ), risulta

$$\frac{S^{(n)} - \mathbf{E} S^{(n)}}{\sqrt{\mathbf{V} S^{(n)}}} \xrightarrow{d} \mathcal{N}(0, 1)$$

dove  $S^{(n)} = \xi_1 + \dots + \xi_n$ .

**Dimostrazione:** Ricordando che la convergenza in distribuzione di una successione di v.a. equivale alla convergenza in generale della corrispondente successione di FdD (vedi Teorema II.9.6) la tesi del nostro Teorema consiste nell'affermare che, posto

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

risulta

$$\mathbf{P} \left( \frac{S^{(n)} - \mathbf{E} S^{(n)}}{\sqrt{\mathbf{V} S^{(n)}}} \leq x \right) \xrightarrow{n} \Phi(x), \quad \forall x \in \mathbf{R},$$

in quanto la FdD limite (cioè la cosiddetta funzione errore) è continua in ogni punto  $x$ . Per provare il Teorema faremo poi uso del Corollario II.10.17: dato che le nostre v.a. sono tutte i.d. poniamo

$$m = \mathbf{E} \xi_n, \quad \sigma^2 = \mathbf{V} \xi_n, \quad \varphi(t) = \mathbf{E} (e^{it(\xi_n - m)}),$$

ed osserviamo che quindi (vedi anche Proposizione II.8.2)

$$\mathbf{E} S^{(n)} = nm, \quad \mathbf{V} S^{(n)} = n\sigma^2;$$

inoltre, detta  $\varphi_n(t)$  la f.c. della v.a.

$$\frac{S^{(n)} - \mathbf{E} S^{(n)}}{\sqrt{\mathbf{V} S^{(n)}}} = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^n (\xi_k - m),$$

dall'indipendenza delle  $\xi_n$  si ha

$$\begin{aligned} \varphi_n(t) &= \mathbf{E} \left( e^{it(S^{(n)} - \mathbf{E} S^{(n)})/\sqrt{\mathbf{V} S^{(n)}}} \right) = \mathbf{E} \left( \prod_{k=1}^n e^{it(\xi_k - m)/\sigma\sqrt{n}} \right) \\ &= \prod_{k=1}^n \mathbf{E} \left( e^{it(\xi_k - m)/\sigma\sqrt{n}} \right) = \left[ \varphi \left( \frac{t}{\sigma\sqrt{n}} \right) \right]^n. \end{aligned}$$

Siccome  $\mathbf{E} \xi_n^2 < +\infty$ , e siccome  $\varphi'(0) = \mathbf{E}(\xi_n - m) = 0$ , dal punto (5) del Teorema II.10.6 sappiamo che

$$\varphi(t) = 1 - \frac{\sigma^2 t^2}{2} + o(t^2), \quad (t \rightarrow 0)$$

sicché per  $t$  fissato e  $n \rightarrow \infty$  risulterà

$$\varphi_n(t) = \left[ 1 - \frac{\sigma^2 t^2}{2\sigma^2 n} + o\left(\frac{1}{n}\right) \right]^n \xrightarrow{n} e^{-t^2/2}.$$

Siccome nell'Esempio II.10.5 è stato mostrato che  $e^{-t^2/2}$  è la f.c. di una v.a. normale standard, il Teorema resta provato sulla base del Corollario II.10.17.  $\square$

**III.1.2 Osservazione:** Una forma semplificata (limitata a v.a. di Bernoulli) del precedente Teorema era stata già discussa negli Esempî II.10.4 e II.10.5: qui abbiamo generalizzato questo risultato al caso di v.a. i.i.d. È però possibile anche modificare alcune di queste ipotesi sostituendole con condizioni che consentono di dimostrare i Teoremi Limite che saranno brevemente analizzati nel seguito. Discuteremo innanzitutto il caso del Teorema di Poisson, già dimostrato nella sua forma elementare come limite di distribuzioni binomiali (vedi Teorema I.8.13), e successivamente daremo (anche se senza dimostrazione e non nella loro formulazione più generale) alcune fra le condizioni più usuali che garantiscono la convergenza in distribuzione di somme di v.a. indipendenti verso v.a. normali. Per generalizzare la dimostrazione del Teorema di Poisson cominceremo con il ricordare che i modelli probabilistici ai quali tipicamente si applica questa analisi sono quelli (vedi Esempio I.8.12) caratterizzati da una successione di esperimenti in ciascuno dei quali si considerano  $n$  v.a. indipendenti che possono assumere valori 0 oppure 1. Fissato, cioè,  $n \in \mathbf{N}$  si considerano  $n$  v.a.  $\xi_1^{(n)}, \dots, \xi_n^{(n)}$  tali che

$$\mathbf{P}(\xi_k^{(n)} = 1) = p_k^{(n)}, \quad \mathbf{P}(\xi_k^{(n)} = 0) = q_k^{(n)}, \quad p_k^{(n)} + q_k^{(n)} = 1, \quad k = 1, \dots, n.$$

La v.a.  $S^{(n)} = \xi_1^{(n)} + \dots + \xi_n^{(n)}$  rappresenta allora il *conteggio* del numero di *successi* (consistenti nel fatto che le v.a. considerate valgano 1) su  $n$  tentativi indipendenti. Ciò che caratterizza i modelli Poissonniani più generali è il fatto che non solo al variare di  $n$  (cioè all'aumentare del numero dei tentativi eseguiti) le  $p_k^{(n)}$  variano ma, per ogni  $n$  fissato, le  $\xi_k^{(n)}$  non sono neanche identicamente distribuite. In tali condizioni le v.a.  $S^{(n)}$  non sono binomiali come nel caso elementare studiato in I.8.13. In sostanza avremo a che fare con schemi triangolari del tipo

$$\begin{array}{ll} \xi_1^{(1)}, & p_1^{(1)}; \\ \xi_1^{(2)}, \xi_2^{(2)}, & p_1^{(2)}, p_2^{(2)}; \\ \vdots & \vdots \\ \xi_1^{(n)}, \dots, \xi_n^{(n)}, & p_1^{(n)}, \dots, p_n^{(n)}; \\ \dots & \dots \end{array}$$

in ciascuna riga del quale le  $\xi_k^{(n)}$  sono indipendenti. Il seguente Teorema stabilisce prima di tutto le condizioni sotto le quali la successione delle  $S^{(n)}$  converge in distribuzione ( $n \rightarrow \infty$ ) verso una v.a. di Poisson  $\eta \sim \pi(\lambda)$  (con  $\lambda > 0$ ) ossia verso una v.a. la cui distribuzione è

$$\mathbf{P}(\eta = m) = e^{-\lambda} \frac{\lambda^m}{m!}, \quad m \in \mathbf{N},$$

e il successivo mostra poi in che senso una distribuzione di Poisson  $\pi(\lambda)$  può essere approssimata da una distribuzione normale quando  $\lambda \rightarrow +\infty$ .  $\circ$

**III.1.3 Teorema (di Poisson):** Date le v.a.  $\xi_k^{(n)}$  con  $k = 1, \dots, n$  e  $n \in \mathbf{N}$  tali che, comunque scelto  $n \in \mathbf{N}$  le  $\xi_1^{(n)}, \dots, \xi_n^{(n)}$  siano indipendenti e

$$\mathbf{P}(\xi_k^{(n)} = 1) = p_k^{(n)}, \quad \mathbf{P}(\xi_k^{(n)} = 0) = q_k^{(n)}, \quad p_k^{(n)} + q_k^{(n)} = 1, \quad k = 1, \dots, n,$$

se risulta

$$\max_{1 \leq k \leq n} p_k^{(n)} \xrightarrow{n} 0, \quad \sum_{k=1}^n p_k^{(n)} \xrightarrow{n} \lambda > 0,$$

allora risulta

$$S^{(n)} \xrightarrow{d} \pi(\lambda)$$

dove  $S^{(n)} = \xi_1^{(n)} + \dots + \xi_n^{(n)}$ .

**Dimostrazione:** Dall'Esempio II.10.4 sappiamo che le f.c. delle  $\xi_k^{(n)}$  sono tutte del tipo  $\varphi_n^{(k)}(t) = p_k^{(n)} e^{it} + q_k^{(n)}$  e quindi che, data l'indipendenza delle  $\xi_k^{(n)}$  (con  $k = 1, \dots, n$  per un dato  $n \in \mathbf{N}$ ), la f.c. di  $S^{(n)}$  ha la forma

$$\varphi_{S^{(n)}}(t) = \mathbf{E}(e^{itS^{(n)}}) = \prod_{k=1}^n (p_k^{(n)} e^{it} + q_k^{(n)}) = \prod_{k=1}^n [1 + p_k^{(n)}(e^{it} - 1)].$$

Ne segue che, essendo per ipotesi  $p_k^{(n)} \xrightarrow{n} 0$  (per  $k = 1, \dots, n$ ), si ha

$$\begin{aligned} \ln \varphi_{S^{(n)}}(t) &= \sum_{k=1}^n \ln[1 + p_k^{(n)}(e^{it} - 1)] = \sum_{k=1}^n [p_k^{(n)}(e^{it} - 1) + o(p_k^{(n)})] \\ &\xrightarrow{n} \lambda(e^{it} - 1), \end{aligned}$$

per cui, data la continuità della funzione logaritmo,  $\varphi_{S^{(n)}}(t) \xrightarrow{n} e^{\lambda(e^{it} - 1)}$ , cioè  $\varphi_{S^{(n)}}(t)$  converge puntualmente verso la f.c. (continua) di una v.a. di Poisson  $\pi(\lambda)$  (vedi Esempio II.10.4) e pertanto in base al Corollario II.10.17  $S^{(n)} \xrightarrow{d} \pi(\lambda)$ .  $\square$

**III.1.4 Teorema:** Se  $S \sim \pi(\lambda)$  è una v.a. di Poisson si ha che

$$\frac{S - \lambda}{\sqrt{\lambda}} \xrightarrow{d} \mathcal{N}(0, 1)$$

per  $\lambda \rightarrow +\infty$ .

**Dimostrazione:** La dimostrazione fa ancora uso delle proprietà delle f.c. Detta

$$\varphi(t) = e^{\lambda(e^{it} - 1)}$$

la f.c. di  $S$ , per la f.c.  $\varphi_\lambda$  di  $(S - \lambda)/\sqrt{\lambda}$  si ha dallo sviluppo in serie di Taylor di un esponenziale che

$$\begin{aligned} \varphi_\lambda &= \mathbf{E} \left( e^{it(S-\lambda)/\sqrt{\lambda}} \right) = e^{-it\sqrt{\lambda}} \mathbf{E} \left( e^{itS/\sqrt{\lambda}} \right) = e^{-it\sqrt{\lambda}} \varphi \left( \frac{t}{\sqrt{\lambda}} \right) \\ &= \exp \left[ -it\sqrt{\lambda} + \lambda \left( e^{it/\sqrt{\lambda}} - 1 \right) \right] \\ &= \exp \left[ -it\sqrt{\lambda} - \lambda + \lambda \left( 1 + \frac{it}{\sqrt{\lambda}} - \frac{t^2}{2\lambda} + o\left(\frac{1}{\lambda}\right) \right) \right] \\ &= e^{-t^2/2} e^{\lambda o(1/\lambda)} \rightarrow e^{-t^2/2} \end{aligned}$$

per  $\lambda \rightarrow +\infty$ . Ricordando poi che  $e^{-t^2/2}$  è la f.c. di una v.a. normale standard  $\mathcal{N}(0, 1)$ , il Teorema resta provato sulla base del Corollario II.10.17.  $\square$

**III.1.5 Teorema (Condizioni di J. W. Lindeberg - 1922):** Data una successione  $(\xi_n)_{n \in \mathbf{N}}$  di v.a. indipendenti, di quadrato integrabile ( $\mathbf{E} \xi_n^2 < +\infty$ ) e non degeneri ( $\mathbf{V} \xi_n = \sigma_n^2 > 0$ ), e posto  $S^{(n)} = \xi_1 + \dots + \xi_n$ , per garantire la tesi del Teorema Limite Centrale

$$\frac{S^{(n)} - \mathbf{E} S^{(n)}}{\sqrt{\mathbf{V} S^{(n)}}} \xrightarrow{d} \mathcal{N}(0, 1)$$

è sufficiente che sia verificata la seguente condizione di Lindeberg

$$\frac{1}{V_n^2} \sum_{k=1}^n \int_{\{x: |x-m_k| \geq \epsilon V_n\}} |x - m_k|^2 F_{\xi_k}(dx) \xrightarrow{n} 0 \quad \forall \epsilon > 0$$

dove  $m_n = \mathbf{E} \xi_n$  e  $V_n^2 = \mathbf{V} S^{(n)} = \sigma_1^2 + \dots + \sigma_n^2$ .

**Dimostrazione:** Omessa (vedi ad esempio **A.N. Shiriyayev**: *Probability*; Springer, New York, 1984, p. 331).  $\square$

**III.1.6 Teorema (Condizioni di Lyapunov):** Data una successione  $(\xi_n)_{n \in \mathbf{N}}$  di v.a. indipendenti, di quadrato integrabile ( $\mathbf{E} \xi_n^2 < +\infty$ ) e non degeneri ( $\mathbf{V} \xi_n = \sigma_n^2 > 0$ ), e posto  $S^{(n)} = \xi_1 + \dots + \xi_n$ , per garantire la tesi del Teorema Limite Centrale

$$\frac{S^{(n)} - \mathbf{E} S^{(n)}}{\sqrt{\mathbf{V} S^{(n)}}} \xrightarrow{d} \mathcal{N}(0, 1)$$

è sufficiente che esista un  $\delta > 0$  tale che sia verificata la seguente condizione di Lyapunov

$$\frac{1}{V_n^{2+\delta}} \sum_{k=1}^n \mathbf{E} |\xi_k - m_k|^{2+\delta} \xrightarrow{n} 0$$

dove  $m_n = \mathbf{E} \xi_n$  e  $V_n^2 = \mathbf{V} S^{(n)} = \sigma_1^2 + \dots + \sigma_n^2$ .

**Dimostrazione:** Omessa (vedi ad esempio **A.N. Shirayev**: *Probability*; Springer, New York, 1984, p. 331). □

**III.1.7 Esempio:** Consideriamo la successione di v.a. indipendenti  $(\xi_n)_{n \in \mathbf{N}}$  con  $\mathbf{P}(\xi_n = n) = \mathbf{P}(\xi_n = -n) = 1/2$ . In tal caso si ha facilmente che  $\mathbf{E} \xi_n = 0$  mentre  $\mathbf{V} \xi_n = n^2$ . Le nostre v.a. non sono, quindi, identicamente distribuite; inoltre, posto  $S^{(n)} = \xi_1 + \dots + \xi_n$ , si ha<sup>1</sup>

$$V_n^2 = \mathbf{V} S^{(n)} = \sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6}.$$

Il  $\delta$  richiesto dalla condizione di Lyapunov è allora  $\delta = 1$ ; infatti

$$\mathbf{E} |\xi_n|^3 = n^3, \quad \sum_{k=1}^n \mathbf{E} |\xi_k|^3 = \sum_{k=1}^n k^3 = \frac{n^2(n+1)^2}{4},$$

per cui in definitiva per  $n \rightarrow \infty$

$$\begin{aligned} \frac{1}{V_n^{2+\delta}} \sum_{k=1}^n \mathbf{E} |\xi_k|^{2+\delta} &= \frac{1}{V_n^3} \sum_{k=1}^n \mathbf{E} |\xi_k|^3 \\ &= \frac{n^2(n+1)^2}{4} \frac{6^{3/2}}{[n(n+1)(2n+1)]^{3/2}} \sim \frac{3\sqrt{3}}{4} \frac{n^4}{n^{9/2}} \end{aligned}$$

e quindi

$$\frac{1}{V_n^3} \sum_{k=1}^n \mathbf{E} |\xi_k|^3 \xrightarrow{n} 0.$$

---

<sup>1</sup> Per il valore della somma delle potenze dei primi  $n$  numeri interi vedi **I.S. Gradshteyn e I.M. Ryzhik**: *Table of integrals, series and products*; Academic Press, San Diego, 1980, p. 1.

Ne segue che  $S^{(n)} \xrightarrow{d} \mathcal{N}(0, 1)$  per  $n \rightarrow \infty$ . ◇

**III.1.8 Teorema (Legge debole dei Grandi Numeri):** Data una successione  $(\xi_n)_{n \in \mathbf{N}}$  di v.a. i.i.d. e integrabili ( $\mathbf{E}|\xi_n| < +\infty$ ), definita  $S^{(n)} = \xi_1 + \dots + \xi_n$  risulta

$$\frac{S^{(n)}}{n} \xrightarrow{\mathbf{P}} m$$

dove abbiamo posto  $m = \mathbf{E} \xi_n$ .

**Dimostrazione:** Ricordiamo innanzitutto che la convergenza in probabilità verso una costante (nel nostro caso il numero  $m$ ) è equivalente alla convergenza in distribuzione verso la medesima costante (vedi punto (d) del Teorema II.9.7). Ne segue che potremo limitarci a dimostrare la convergenza in distribuzione di  $(\xi_n)_{n \in \mathbf{N}}$  facendo uso del solito Corollario II.10.17. Se allora indichiamo con  $\varphi(t)$  la f.c. comune a tutte le  $\xi_n$  e con  $\varphi_{S^{(n)}}(t)$  la f.c. di  $S^{(n)}$  avremo che

$$\varphi_{S^{(n)}}(t) = \mathbf{E} (e^{itS^{(n)}/n}) = \prod_{k=1}^n \mathbf{E} (e^{it\xi_k}) = \left[ \varphi \left( \frac{t}{n} \right) \right]^n .$$

D'altra parte, dato che le nostre v.a. sono integrabili, dal punto (5) del Teorema II.10.6 si ha che  $\varphi(t) = 1 + itm + o(t)$ , per  $(t \rightarrow 0)$ ; ossia, per un dato  $t$  e per  $n \rightarrow \infty$ ,

$$\varphi \left( \frac{t}{n} \right) = 1 + i \frac{t}{n} m + o \left( \frac{1}{n} \right) .$$

Ne segue che, comunque fissato  $t \in \mathbf{R}$ , avremo

$$\varphi_{S^{(n)}}(t) = \left[ 1 + i \frac{t}{n} m + o \left( \frac{1}{n} \right) \right]^n \xrightarrow{n \rightarrow \infty} e^{itm} .$$

Siccome  $e^{itm}$  è la f.c. della v.a.  $\mathbf{P}$ -q.o. costantemente eguale ad  $m$  (vedi Esempio II.10.4) il Corollario II.10.17 ci garantisce la validità della tesi. □

**III.1.9 Osservazione:** La tesi del Teorema III.1.8 può essere facilmente riformulata come

$$\frac{S^{(n)} - \mathbf{E} S^{(n)}}{n} \xrightarrow{\mathbf{P}} 0$$

a causa del fatto che le nostre  $\xi_n$  sono identicamente distribuite. Tale nuova espressione della Legge debole dei Grandi Numeri è particolarmente utile per generalizzare questo importante risultato alle situazioni in cui le  $\xi_n$  non sono più identicamente distribuite: in tal caso, infatti, le attese  $\mathbf{E} \xi_n$  non sono più tutte uguali e quindi, mentre non ha più senso chiedersi se  $S^{(n)}/n$  converge verso il comune VdA delle  $\xi_n$ , resta significativo il problema della convergenza verso 0 di  $(S^{(n)} - \mathbf{E} S^{(n)})/n$ . In particolare nel Teorema seguente mostreremo che l'ipotesi



di identica distribuzione può essere trascurata richiedendo un adeguato comportamento delle varianze  $\mathbf{V}\xi_n$ . Più precisamente verrà richiesto che esse siano **uniformemente limitate**, cioè che si possa determinare un numero  $C > 0$  tale che  $\mathbf{V}\xi_n < C$  comunque scelto  $n \in \mathbf{N}$ . Notiamo infine che, nel caso di v.a. i.i.d., siccome  $\mathbf{V}S^{(n)} = n\sigma^2$  se  $\sigma^2 = \mathbf{V}\xi_n$ , la Legge debole dei Grandi Numeri può essere anche espressa come

$$\frac{S^{(n)} - \mathbf{E}S^{(n)}}{\mathbf{V}S^{(n)}} \xrightarrow{\mathbf{P}} 0,$$

formulazione che mette bene in luce le analogie e le differenze con la tesi del Teorema Limite Centrale III.1.1. ○

**III.1.10 Teorema:** Se  $(\xi_n)_{n \in \mathbf{N}}$  è una successione di v.a. indipendenti, integrabili ( $\mathbf{E}|\xi_n| < +\infty$ ) e dotate di varianze uniformemente limitate, si ha

$$\frac{S^{(n)} - \mathbf{E}S^{(n)}}{n} \xrightarrow{\mathbf{P}} 0$$

dove abbiamo posto  $S^{(n)} = \xi_1 + \dots + \xi_n$ .

**Dimostrazione:** La dimostrazione di questo Teorema si basa sull'uso della Disuguaglianza di Chebyshev (vedi Corollario II.6.17): a causa dell'ipotesi di indipendenza abbiamo infatti, con  $\epsilon > 0$  arbitrario, che<sup>2</sup>

$$\begin{aligned} \mathbf{P}\left(\left|\frac{S^{(n)} - \mathbf{E}S^{(n)}}{n}\right| \geq \epsilon\right) &\leq \frac{1}{\epsilon^2} \mathbf{V}\left(\frac{S^{(n)} - \mathbf{E}S^{(n)}}{n}\right) = \frac{1}{n^2\epsilon^2} \mathbf{V}\left(\sum_{k=1}^n (\xi_k - \mathbf{E}\xi_k)\right) \\ &= \frac{1}{n^2\epsilon^2} \sum_{k=1}^n \mathbf{V}(\xi_k - \mathbf{E}\xi_k) = \frac{1}{n^2\epsilon^2} \sum_{k=1}^n \mathbf{V}\xi_k \leq \frac{nC}{n^2\epsilon^2} \\ &= \frac{C}{n\epsilon^2} \xrightarrow{n} 0, \end{aligned}$$

e il Teorema è dimostrato per definizione di convergenza in probabilità. □

**III.1.11 Osservazione:** La Legge dei Grandi Numeri gioca un ruolo fondamentale nella Teoria della Probabilità in quanto consente, ad esempio, di stimare i VdA di v.a. tramite medie aritmetiche dei risultati di un gran numero di osservazioni indipendenti<sup>3</sup> di tali v.a. Così il VdA  $\mathbf{E}\xi$  di una v.a. può essere stimato considerando la successione  $(\xi_n)_{n \in \mathbf{N}}$  delle infinite possibili osservazioni indipendenti

<sup>2</sup> Nella deduzione si usa la Proposizione II.8.2 ed il fatto, banalmente dimostrabile, che, se  $a$  è una costante,  $\mathbf{V}(\xi - a) = \mathbf{V}\xi$ .

<sup>3</sup> Vedremo nel capitolo sulle Catene di Markov che questo risultato può addirittura essere esteso, sotto particolari condizioni (*ergodicità*), al caso in cui le v.a.  $\xi_n$  della successione non sono indipendenti.

di  $\xi$  (le  $\xi_n$  sono dunque v.a. indipendenti e tutte i.d. come  $\xi$  stessa: ognuna di esse rappresenta il risultato della  $n$ -ma osservazione) e calcolando la media aritmetica

$$\frac{S^{(n)}}{n} = \frac{\xi_1 + \dots + \xi_n}{n}.$$

La Legge debole dei Grandi Numeri ci dice allora che, se  $n$  è abbastanza grande, *con grande confidenza* potremo affermare che il valore della media aritmetica (ottenuto con una sola serie di  $n$  osservazioni) si discosta da  $\mathbf{E}\xi$  di una quantità arbitrariamente piccola al crescere di  $n$ . L'espressione *con grande confidenza*, però, sta qui ad indicare un punto problematico: i Teoremi III.1.8 e III.1.10, infatti, assicurano la convergenza di  $S^{(n)}/n$  verso  $\mathbf{E}\xi$  solamente in probabilità e non garantiscono affatto la convergenza  $\mathbf{P}$ -q.o. (vedi Teorema II.9.7 ed Esempio II.9.8). Ciò sta ad indicare che, secondo la Legge debole dei Grandi Numeri, la probabilità che la successione  $S^{(n)}/n$  non converga verso  $\mathbf{E}\xi$  potrebbe, a rigore, essere diversa da zero. In altre parole: se non disponessimo di risultati più forti dei Teoremi III.1.8 e III.1.10, con probabilità non nulla ci potrebbe capitare di stimare  $\mathbf{E}\xi$  mediante le osservazioni di una successione le cui medie aritmetiche *non convergono* verso  $\mathbf{E}\xi$ , e questo creerebbe una situazione particolarmente delicata per tutte le applicazioni del Calcolo delle Probabilità. Per questo motivo grandi sforzi sono stati dedicati alla ricerca di Leggi dei Grandi Numeri in forma forte (vere cioè  $\mathbf{P}$ -q.o.) che assicurassero la correttezza delle procedure empiriche con probabilità 1. Nel seguito esporremo (anche se dimostreremo solo quella dovuta a F.P. Cantelli (1875-1966): egli ne diede una prima versione basandosi su un lemma, pubblicato nel 1917 in una memoria per l'Accademia Nazionale dei Lincei, che faceva uso di alcuni precedenti lavori di E. Borel) alcune delle possibili formulazioni di questa Legge forte dei Grandi numeri che si differenziano per le ipotesi su  $(\xi_n)_{n \in \mathbf{N}}$  necessarie a garantire il risultato. ○

**III.1.12 Lemma (di Borel - Cantelli):**

- (a) Se  $(A_n)_{n \in \mathbf{N}}$  è una successione di eventi e se  $\sum_n \mathbf{P}(A_n) < +\infty$ , allora  $\mathbf{P}(\overline{\lim} A_n) = 0$ ;  
 (b) se invece  $\sum_n \mathbf{P}(A_n) = +\infty$ , e se gli  $A_n$  sono anche indipendenti, allora  $\mathbf{P}(\overline{\lim} A_n) = 1$ .

**Dimostrazione:**

- (a) Ricordando che

$$\bigcup_{k \geq n} A_k$$

è una successione di eventi decrescente per inclusione, dal punto 3) del Teorema II.1.13 e dalla subadditività (punto VI del Teorema II.1.12) si ha che

$$\mathbf{P}(\overline{\lim} A_n) = \mathbf{P}\left(\bigcap_{n \in \mathbf{N}} \bigcup_{k \geq n} A_k\right) = \lim_n \mathbf{P}\left(\bigcup_{k \geq n} A_k\right) \leq \lim_n \sum_{k \geq n} \mathbf{P}(A_k).$$

Siccome per ipotesi la serie  $\sum_n \mathbf{P}(A_n)$  è convergente, la successione dei resti parziali deve essere infinitesima, sicché

$$\mathbf{P}(\overline{\lim} A_n) \leq \lim_n \sum_{k \geq n} \mathbf{P}(A_k) = 0$$

e quindi, come richiesto,  $\mathbf{P}(\overline{\lim} A_n) = 0$ .

- (b) Se le  $A_n$  sono indipendenti, tali risulteranno anche le  $\overline{A}_n$  sicché, comunque presi  $n, m \in \mathbf{N}$  (con  $m \geq n$ ) si avrà

$$\mathbf{P}\left(\bigcap_{k=n}^m \overline{A}_k\right) = \prod_{k=n}^m \mathbf{P}(\overline{A}_k).$$

Siccome al crescere di  $m \geq n$  la successione di eventi

$$\bigcap_{k=n}^m \overline{A}_k$$

è chiaramente decrescente per inclusione, dal Teorema II.1.13 si ha

$$\begin{aligned} \prod_{k \geq n} \mathbf{P}(\overline{A}_k) &= \lim_m \prod_{k=n}^m \mathbf{P}(\overline{A}_k) = \lim_m \mathbf{P}\left(\bigcap_{k=n}^m \overline{A}_k\right) \\ &= \mathbf{P}\left(\bigcap_{m \geq n} \bigcap_{k=n}^m \overline{A}_k\right) = \mathbf{P}\left(\bigcap_{k \geq n} \overline{A}_k\right). \end{aligned}$$

Osservando ora che per  $0 \leq x < 1$  risulta  $\ln(1-x) \leq -x$ , e che a causa delle nostre ipotesi (siccome i resti parziali di una serie a termini positivi divergente sono anch'essi tutti divergenti)

$$\sum_{k \geq n} \mathbf{P}(A_k) = +\infty,$$

avremo che

$$\begin{aligned} \ln \mathbf{P}\left(\bigcap_{k \geq n} \overline{A}_k\right) &= \ln \prod_{k \geq n} [1 - \mathbf{P}(A_k)] = \sum_{k \geq n} \ln[1 - \mathbf{P}(A_k)] \\ &\leq - \sum_{k \geq n} \mathbf{P}(A_k) = -\infty. \end{aligned}$$

Ne segue che, comunque scelto  $n \in \mathbf{N}$ , si ha

$$0 = \mathbf{P}\left(\bigcap_{k \geq n} \overline{A}_k\right) = \mathbf{P}\left(\overline{\bigcup_{k \geq n} A_k}\right) = 1 - \mathbf{P}\left(\bigcup_{k \geq n} A_k\right)$$

e quindi

$$\mathbf{P}\left(\bigcup_{k \geq n} A_k\right) = 1, \quad \forall n \in \mathbf{N}.$$

Siccome però

$$\bigcup_{k \geq n} A_k$$

è una successione decrescente per inclusione, si ha in definitiva

$$\mathbf{P}(\overline{\lim} A_n) = \mathbf{P}\left(\bigcap_{n \in \mathbf{N}} \bigcup_{k \geq n} A_k\right) = \lim_n \mathbf{P}\left(\bigcup_{k \geq n} A_k\right) = 1. \quad \square$$

**III.1.13 Lemma:** Condizione sufficiente a garantire la convergenza  $\mathbf{P}$ -q.o. di una successione di v.a.  $(\xi_n)_{n \in \mathbf{N}}$  verso la v.a.  $\xi$  è che risulti

$$\sum_n \mathbf{P}(|\xi_n - \xi| \geq \epsilon) < +\infty$$

comunque scelto  $\epsilon > 0$ .

**Dimostrazione:** Posto per brevità  $B_n^\epsilon = \{|\xi_n - \xi| \geq \epsilon\}$  e  $B^\epsilon = \overline{\lim}_n B_n^\epsilon$ , si ha che

$$\begin{aligned} \{\xi_n \not\rightarrow \xi\} &= \overline{\{\xi_n \rightarrow \xi\}} = \overline{\{\forall \epsilon > 0 \exists n \in \mathbf{N} \exists' \forall k \geq n : |\xi_k - \xi| < \epsilon\}} \\ &= \bigcap_{\epsilon > 0} \bigcup_{n \in \mathbf{N}} \bigcap_{k \geq n} \{\overline{|\xi_n - \xi| < \epsilon}\} = \bigcup_{\epsilon > 0} \bigcap_{n \in \mathbf{N}} \bigcup_{k \geq n} \{|\xi_n - \xi| \geq \epsilon\} \\ &= \bigcup_{\epsilon > 0} \bigcap_{n \in \mathbf{N}} \bigcup_{k \geq n} B_k^\epsilon = \bigcup_{\epsilon > 0} B^\epsilon. \end{aligned}$$

Inoltre dalla famiglia non numerabile (decescente<sup>4</sup> per inclusione)  $(B^\epsilon)_{\epsilon > 0}$  si può facilmente costruire la famiglia numerabile (crescente per inclusione)  $(B^{1/m})_{m \in \mathbf{N}}$  tale che

$$B^{1/m} \uparrow \bigcup_{m \in \mathbf{N}} B^{1/m} = \bigcup_{\epsilon > 0} B^\epsilon.$$

Dal Teorema II.1.13 abbiamo allora che

$$\mathbf{P}(\xi_n \not\rightarrow \xi) = \mathbf{P}\left(\bigcup_{m \in \mathbf{N}} B^{1/m}\right) = \lim_m \mathbf{P}(B^{1/m}),$$

<sup>4</sup> Infatti, se  $\epsilon' \leq \epsilon$ , si ha  $B_n^{\epsilon'} = \{|\xi_n - \xi| \geq \epsilon'\} \supseteq \{|\xi_n - \xi| \geq \epsilon\} = B_n^\epsilon$ , comunque scelto  $n \in \mathbf{N}$ .

ma siccome per ipotesi

$$\sum_n \mathbf{P}(B_n^\epsilon) < +\infty, \quad \forall \epsilon > 0$$

dal Lemma III.1.12 di Borel - Cantelli si ha che

$$\mathbf{P}(B^\epsilon) = \mathbf{P}(\overline{\lim} B_n^\epsilon) = 0, \quad \forall \epsilon > 0$$

cioè  $\mathbf{P}(B^{1/m}) = 0$  per ogni  $m$ , e quindi in definitiva

$$\mathbf{P}(\xi_n \not\rightarrow \xi) = \lim_m \mathbf{P}(B^{1/m}) = 0$$

cioè  $\xi_n \xrightarrow{\text{q.o.}} \xi$ . □

**III.1.14 Teorema (Legge forte dei Grandi Numeri di Cantelli):** Se  $(\xi_n)_{n \in \mathbf{N}}$  è una successione di v.a. indipendenti con quarto momento finito ( $\mathbf{E} \xi_n^4 < +\infty$ ) e per la quale esista un numero  $C > 0$  tale che

$$\mathbf{E} |\xi_n - \mathbf{E} \xi_n|^4 \leq C, \quad \forall n \in \mathbf{N}$$

allora risulta

$$\frac{S^{(n)} - \mathbf{E} S^{(n)}}{n} \xrightarrow{\text{q.o.}} 0$$

dove abbiamo posto  $S^{(n)} = \xi_1 + \dots + \xi_n$ .

**Dimostrazione:** Senza perdita di generalità<sup>5</sup> possiamo supporre che  $\mathbf{E} \xi_n = 0$  per ogni  $n \in \mathbf{N}$  in modo che la tesi da dimostrare si semplifica in

$$\frac{S^{(n)}}{n} \xrightarrow{\text{q.o.}} 0.$$

Il Lemma III.1.13 ci dice che questo risultato è garantito se si riesce a provare che

$$\sum_n \mathbf{P} \left( \left| \frac{S^{(n)}}{n} \right| \geq \epsilon \right) < +\infty, \quad \forall \epsilon > 0,$$

mentre a sua volta la Diseguaglianza di Chebyshev (Corollario II.6.17) afferma che, comunque scelto  $n \in \mathbf{N}$ , si ha

$$\mathbf{P} \left( \left| \frac{S^{(n)}}{n} \right| \geq \epsilon \right) \leq \frac{1}{\epsilon^4} \mathbf{E} \left( \frac{S^{(n)}}{n} \right)^4,$$

---

<sup>5</sup> Se  $\mathbf{E} \xi_n \neq 0$  basterà, come al solito, prendere in considerazione le v.a. con VdA nulli  $\xi_n - \mathbf{E} \xi_n$  e dimostrare il teorema per queste: il risultato per le  $\xi_n$  segue poi facilmente.

per cui il teorema sarà dimostrato se si prova che

$$\sum_n \mathbf{E} \left( \frac{S^{(n)}}{n} \right)^4 < +\infty.$$

Nello sviluppo di  $(S^{(n)})^4$  i termini sono tutti di uno dei seguenti tipi:

$$\xi_i^4, \quad \xi_i^3 \xi_j, \quad \xi_i^2 \xi_j^2, \quad \xi_i^2 \xi_j \xi_k, \quad \xi_i \xi_j \xi_k \xi_l,$$

ma, siccome le  $\xi_n$  sono tutte indipendenti con  $\mathbf{E} \xi_n = 0$ , nella espressione di  $\mathbf{E} (S^{(n)})^4$  tutti i termini si annulleranno fatta eccezione per quelli del tipo  $\mathbf{E} \xi_i^4$  e  $\mathbf{E} \xi_i^2 \cdot \mathbf{E} \xi_j^2$ . Ricordando ora che (vedi **M. Abramowitz e I.A. Stegun: Handbook of mathematical functions**; Dover, New York, 1972, p. 823)

$$(x_1 + \dots + x_m)^n = \sum_{n_1 + \dots + n_m = n} \binom{n}{n_1 \dots n_m} x_1^{n_1} \cdot \dots \cdot x_m^{n_m}$$

dove i coefficienti multinomiali sono dati da

$$\binom{n}{n_1 \dots n_m} = \frac{n!}{n_1! \dots n_m!},$$

osservando che nel nostro caso (con  $n = 4$ ) siamo interessati solo ai termini in cui (a) solo uno fra i numeri  $n_1, \dots, n_m$  vale 4 mentre gli altri sono nulli (tutti i termini di questo tipo sono moltiplicati per lo stesso coefficiente multinomiale  $4!/4! = 1$ ), oppure (b) solo due (diversi) fra i numeri  $n_1, \dots, n_m$  valgono 2 mentre gli altri sono nulli (tutti i termini di questo tipo sono moltiplicati per lo stesso coefficiente multinomiale  $4!/2! \cdot 2! = 6$ ), e tenendo infine conto delle ipotesi fatte e della Diseguaglianza di Lyapunov (Corollario II.6.20 applicato a  $\xi_i^2$  con  $s = 1$  e  $t = 2$ ), si ha che

$$\begin{aligned} \mathbf{E} (S^{(n)})^4 &= \sum_{i=1}^n \mathbf{E} \xi_i^4 + 6 \sum_{i < j} \mathbf{E} \xi_i^2 \cdot \mathbf{E} \xi_j^2 \leq nC + 6 \sum_{i < j} \sqrt{\mathbf{E} \xi_i^4 \cdot \mathbf{E} \xi_j^4} \\ &\leq nC + 6 \sum_{i < j} C = nC + 6 \frac{n(n-1)}{2} C = (3n^2 - 2n)C < 3n^2C, \end{aligned}$$

e pertanto

$$\sum_n \mathbf{E} \left( \frac{S^{(n)}}{n} \right)^4 \leq \sum_n \frac{3n^2C}{n^4} = 3C \sum_n \frac{1}{n^2} < +\infty,$$

il che definitivamente prova il teorema. Notiamo infine che solo l'uso della quarta potenza nelle ipotesi e nella Diseguaglianza di Chebyshev ha consentito di ottenere delle maggiorazioni mediante serie convergenti: se ad esempio avessimo usato la

seconda potenza la serie maggiorante (in questo caso  $\sum_n (1/n)$ ) sarebbe stata divergente e non avremmo potuto raggiungere la nostra conclusione.  $\square$

**III.1.15 Teorema (Legge forte dei Grandi Numeri di Kolmogorov):**

Se  $(\xi_n)_{n \in \mathbf{N}}$  è una successione di v.a. indipendenti con secondo momento finito ( $\mathbf{E} \xi_n^2 < +\infty$ ) e se  $(b_n)_{n \in \mathbf{N}}$  è una successione divergente di numeri positivi tale che

$$\sum_n \frac{\mathbf{V} \xi_n}{b_n} < +\infty,$$

allora, posto  $S^{(n)} = \xi_1 + \dots + \xi_n$ , si ha

$$\frac{S^{(n)} - \mathbf{E} S^{(n)}}{b_n} \xrightarrow{\text{q.o.}} 0;$$

in particolare se  $b_n = n$  la condizione

$$\sum_n \frac{\mathbf{V} \xi_n}{n} < +\infty,$$

assicura che

$$\frac{S^{(n)} - \mathbf{E} S^{(n)}}{n} \xrightarrow{\text{q.o.}} 0,$$

cioè garantisce la validità della Legge forte dei Grandi Numeri.

**Dimostrazione:** Omessa (vedi ad esempio **A.N. Shirayev:** *Probability*; Springer, New York, 1984, p. 364). Si tratta ancora, come nel caso del Teorema di Cantelli III.1.14, di un risultato valido per successioni di v.a. indipendenti ma non necessariamente i.d. In questo caso l'uso di metodi più potenti di dimostrazione consente di indebolire le ipotesi del Teorema di Cantelli.  $\square$

**III.1.16 Teorema (Legge forte dei Grandi Numeri di Kolmogorov):**

Se  $(\xi_n)_{n \in \mathbf{N}}$  è una successione di v.a. i.i.d. e assolutamente integrabili ( $\mathbf{E} |\xi_n| < +\infty$ ), allora risulta

$$\frac{S^{(n)}}{n} \xrightarrow{\text{q.o.}} m$$

dove abbiamo posto  $S^{(n)} = \xi_1 + \dots + \xi_n$  e  $\mathbf{E} \xi_n = m$ .

**Dimostrazione:** Omessa (vedi ad esempio **A.N. Shirayev:** *Probability*; Springer, New York, 1984, p. 366). In questa formulazione della Legge forte dei Grandi Numeri si fa a meno anche dell'ipotesi di esistenza (e finitezza) dei secondi momenti, ma il prezzo da pagare è l'introduzione dell'ipotesi di identica distribuzione delle v.a. della successione  $(\xi_n)_{n \in \mathbf{N}}$  che non era richiesta nel Teorema III.1.15. Le conclusioni di questo teorema restano valide anche se il VdA  $\mathbf{E} \xi_n$  non è finito:

supponiamo, ad esempio, che  $\mathbf{E} \xi_n^- < +\infty$  e  $\mathbf{E} \xi_n^+ = +\infty$  e consideriamo, per un dato  $c > 0$ , le v.a. *troncate* i.i.d.  $\xi_n^c = \xi_n I_{\{\xi_n \leq c\}}$  per le quali il nostro Teorema vale dato che ovviamente  $\mathbf{E} |\xi_n^c| < +\infty$ . Posto allora  $S_c^{(n)} = \xi_1^c + \dots + \xi_n^c$  si ha

$$\lim_n \frac{S^{(n)}}{n} \geq \lim_n \frac{S_c^{(n)}}{n} = \mathbf{E} \xi_1^c, \quad \mathbf{P}\text{-q.o.}$$

Siccome però

$$\lim_{c \rightarrow +\infty} \mathbf{E} \xi_1^c = \mathbf{E} \xi_1 = +\infty$$

risulta anche

$$\frac{S^{(n)}}{n} \xrightarrow{\text{q.o.}} +\infty.$$

Al contrario le conclusioni del nostro Teorema non sono più garantite se il VdA delle v.a.  $\xi_n$  non esiste (vedi Definizione II.6.3) come mostra la discussione del seguente esempio nel quale viene esplicitamente costruito un modello per il quale la Legge dei Grandi Numeri non vale.  $\square$

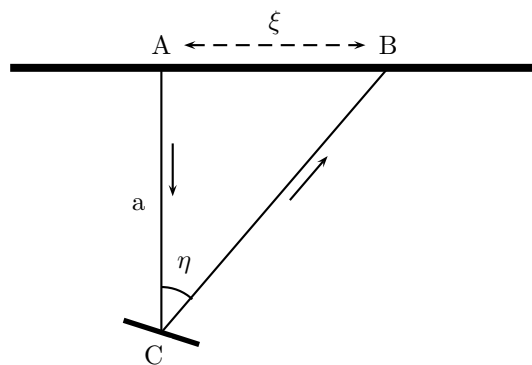


Fig. III.1.1 Costruzione di v.a. distribuite secondo Cauchy.

**III.1.17 Esempio:** Supponiamo che un raggio di luce emesso da una sorgente luminosa collocata in A su una parete (vedi Fig. III.1.1) colpisca uno specchio collocato in B ad una distanza  $a$  e libero di muoversi ruotando attorno ad un perno. La posizione dello specchio è aleatoria in modo tale che l'angolo  $\eta$  di riflessione del raggio sia una v.a. distribuita uniformemente fra  $-\frac{\pi}{2}$  e  $\frac{\pi}{2}$ . Sia infine  $\xi = a \tan \eta$  la distanza (con segno) da A del punto C in cui il raggio riflesso colpisce la parete. Mostriamo innanzitutto che  $\xi$  è una v.a. assolutamente continua dotata di fdd di Cauchy (vedi Esempio II.3.16): infatti per ipotesi

$$f_\eta(y) = \begin{cases} \frac{1}{\pi}, & \text{se } y \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \\ 0, & \text{altrove} \end{cases}$$



mentre  $\xi$  risulta dall'applicazione ad  $\eta$  della funzione monotona  $x = g(y) = a \tan y$  definita su  $(-\frac{\pi}{2}, \frac{\pi}{2})$ . Dai risultati esposti in Osservazione II.8.10, posto  $y = h(x) = g^{-1}(x) = \arctan(y/a)$  ed osservato che

$$h'(x) = \frac{a}{a^2 + x^2}$$

si ottiene per la fdd di  $\xi$

$$f_{\xi}(x) = f_{\eta}(h(x))h'(x) = \frac{1}{\pi} \frac{a}{a^2 + x^2}$$

dato che  $f_{\eta}(y)$  assume costantemente il valore  $\frac{1}{\pi}$  in  $[-\frac{\pi}{2}, \frac{\pi}{2}]$  e che  $h(x) \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ . Possiamo ora supporre di ripetere, in maniera indipendente, varie misure di  $\xi$  ottenendo una successione  $(\xi_n)_{n \in \mathbf{N}}$  di v.a. i.i.d. tutte dotate della medesima fdd di Cauchy e della medesima f.c. che, come si vede con un calcolo diretto<sup>6</sup>, è

$$\varphi(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{a}{a^2 + x^2} e^{itx} dx = e^{-a|t|}.$$

Se però definiamo  $S^{(n)} = \xi_1 + \dots + \xi_n$  vediamo subito che la f.c. di  $S^{(n)}/n$  è, comunque scelto  $n$ ,

$$\varphi_n(t) = [\varphi(t/n)]^n = (e^{-a|t|/n})^n = e^{-a|t|} = \varphi(t)$$

sicché si ha anche  $\varphi_n(t) \xrightarrow{n} \varphi(t)$  in ogni  $t \in \mathbf{R}$ . Dal Corollario II.10.17 segue allora non solo che

$$\frac{S^{(n)}}{n} \xrightarrow{d} \xi,$$

ma addirittura che tutte le  $S^{(n)}/n$  sono distribuite allo stesso modo di  $\xi$ , dove  $\xi$  è la nostra v.a. di Cauchy: questa ha, quindi, la proprietà di non obbedire ad una Legge dei Grandi Numeri nel senso che medie aritmetiche di sue misure ripetute non mostrano una tendenza (né in probabilità né  $\mathbf{P}$ -q.o. altrimenti sarebbe vero anche in distribuzione) a concentrare i loro valori attorno ad un numero al crescere di  $n$ . D'altra parte il fatto che le  $S^{(n)}/n$  siano tutte v.a. di Cauchy mostra anche che nemmeno il Teorema limite centrale è rispettato, dal momento che la distribuzione delle  $S^{(n)}/n$  non approssima una distribuzione normale per  $n \rightarrow \infty$ . È possibile mostrare che questo comportamento apparentemente anomalo che sembra violare i Teoremi precedenti è dovuto al fatto che  $\xi$  non ha un VdA definito e, quindi, non rispetta le ipotesi di tali Teoremi. Infatti, dato che  $\xi^+ = \xi I_{[0, +\infty)}(\xi)$

---

<sup>6</sup> Il calcolo è più convenientemente eseguito con una integrazione nel campo complesso, vedi ad esempio **B. Chabat**: *Introduction à l'Analyse Complexe - I*; MIR, Mosca, 1990, p. 141.

e  $\xi^- = -\xi I_{(-\infty, 0]}(\xi)$ , e data la simmetria della fdd di Cauchy ( $f_\xi(-x) = f_\xi(x)$ ), si ha

$$\mathbf{E} \xi^+ = \mathbf{E} \xi^- = \frac{a}{\pi} \int_0^{+\infty} \frac{x}{a^2 + x^2} dx = \left[ \frac{1}{2} \ln(a^2 + x^2) \right]_0^{+\infty} = +\infty.$$

D'altra parte sarebbe anche possibile mostrare che

$$\int_{-\infty}^{+\infty} \frac{x}{a^2 + x^2} dx$$

non è definito nemmeno nel senso dell'integrale improprio secondo Riemann in quanto

$$\lim_{M \rightarrow +\infty} \int_{-M}^0 \frac{x}{a^2 + x^2} dx = -\infty, \quad \lim_{N \rightarrow +\infty} \int_0^N \frac{x}{a^2 + x^2} dx = +\infty.$$

Siamo quindi nel caso in cui  $\mathbf{E} \xi_n$  non è definito sicché nessuna delle precedenti Leggi dei Grandi Numeri è applicabile.  $\diamond$

**III.1.18 Esempio:** Supponiamo di considerare una funzione continua  $g(x) : [0, 1] \rightarrow [0, 1]$  e di voler calcolare

$$\int_0^1 g(x) dx$$

in modo numerico, senza cioè calcolare la primitiva di  $g(x)$ . Mostriamo ora come è possibile far questo sfruttando le regolarità statistiche secondo un metodo genericamente noto come **Monte Carlo** (vedi anche Esempio II.7.17). Consideriamo il vett.a.  $Y = (\xi, \eta)$  a valori in  $(\mathbf{R}^2, \mathcal{B}(\mathbf{R}^2))$  con componenti indipendenti ed uniformemente distribuite in  $[0, 1]$  in modo che

$$f_\xi(x) = \begin{cases} 1, & x \in [0, 1] \\ 0, & \text{altrove} \end{cases} \quad f_\eta(y) = \begin{cases} 1, & y \in [0, 1] \\ 0, & \text{altrove} \end{cases}$$

$$f_{\xi\eta}(x, y) = \begin{cases} 1, & (x, y) \in [0, 1] \times [0, 1] \\ 0, & \text{altrove} \end{cases}$$

sicché  $Y$  risulta uniformemente distribuito sul quadrato  $[0, 1] \times [0, 1]$ . Posto allora

$$A = \{(x, y) \in [0, 1] \times [0, 1] : g(x) \geq y\}, \quad B = \{(\xi, \eta) \in A\} = \{g(\xi) \geq \eta\}$$

(nota che  $A$  è un sottinsieme di  $\mathbf{R}^2$  mentre  $B$  è un sottinsieme di  $\Omega$ ) si ha che  $\zeta = I_B = I_A(\xi, \eta) = I_{[0, 1]}(g(\xi) - \eta)$  è la v.a. che vale 1 se il punto aleatorio  $(\xi, \eta)$  cade in  $A$  e 0 se esso cade altrove. Si ha allora

$$\begin{aligned} \mathbf{E} \zeta = \mathbf{P}(B) &= \int_{\mathbf{R}^2} f_{\xi\eta}(x, y) I_{[0, 1]}(g(x) - y) dx dy = \int_0^1 \int_0^1 I_{[0, 1]}(g(x) - y) dx dy \\ &= \int_0^1 \left[ \int_0^1 I_{[0, 1]}(g(x) - y) dy \right] dx = \int_0^1 \left[ \int_0^{g(x)} dy \right] dx = \int_0^1 g(x) dx. \end{aligned}$$

Pertanto il nostro integrale può essere calcolato stimando  $\mathbf{E}\zeta$  tramite la Legge forte dei Grandi Numeri<sup>7</sup>: detta  $(\zeta_n)_{n \in \mathbf{N}}$  la successione di v.a. i.i.d. che si ottiene eseguendo una successione di scelte dei punti aleatori  $(\xi, \eta)$ , avremo  $\zeta_n = 1$  se all' $n$ -mo tentativo  $(\xi, \eta) \in A$  e  $\zeta = 0$  altrimenti. Pertanto  $S^{(n)} = \zeta_1 + \dots + \zeta_n$  rappresenterà il numero (aleatorio) di volte in cui  $(\xi, \eta) \in A$  su  $n$  tentativi: la Legge dei Grandi Numeri ci dice allora che per  $n$  abbastanza grande  $S^{(n)}/n$  è una buona stima di  $\mathbf{E}\zeta$ , ovvero del valore dell'integrale che volevamo calcolare. In pratica il valore approssimato dell'integrale si ottiene calcolando la frequenza con cui un punto scelto a caso nel quadrato  $[0, 1] \times [0, 1]$  cade nella regione  $A$ , frequenza che, anche intuitivamente, stima la frazione di area del quadrato rappresentata da  $A$ , cioè il valore dell'integrale. Un secondo metodo di integrazione numerica con il metodo di Monte Carlo sarà discusso in un Esercizio.  $\diamond$

---

<sup>7</sup> Le ipotesi del Teorema III.1.16 sono rispettate dato che  $\mathbf{E}\zeta$  coincide con il nostro integrale il cui valore è finito essendo  $g(x)$  una funzione continua in  $[0, 1]$ .



## III.2 Martingale

**III.2.1 Definizione:** Dato uno spazio di probabilità  $(\Omega, \mathcal{F}, \mathbf{P})$  diremo che una famiglia di  $\sigma$ -algebre  $(\mathcal{F}_n)_{n \geq 0}$  è una **Filtrazione** se essa risulta crescente per inclusione nel senso che  $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots \subseteq \mathcal{F}_n \subseteq \dots \subseteq \mathcal{F}$ . Data allora una successione di v.a.  $(X_n)_{n \geq 0}$  su  $(\Omega, \mathcal{F}, \mathbf{P})$  ed una filtrazione  $(\mathcal{F}_n)_{n \geq 0}$ , diremo che  $X = (X_n, \mathcal{F}_n)_{n \geq 0}$  è una **Successione stocastica** quando  $X_n$  è  $\mathcal{F}_n$ -misurabile comunque scelto  $n \geq 0$ . In questo caso si dice anche che la successione  $(X_n)_{n \geq 0}$  è **adattata** alla filtrazione  $(\mathcal{F}_n)_{n \geq 0}$ .  $\triangle$

**III.2.2 Definizione:** Una successione stocastica  $X = (X_n, \mathcal{F}_n)_{n \geq 0}$  si dice **prevedibile** quando  $X_n$  risulta  $\mathcal{F}_{n-1}$ -misurabile; si dice invece **crescente** quando  $\mathbf{P}$ -q.o. risulta  $X_0 = 0$  e  $X_n \leq X_{n+1}$ , comunque scelto  $n \geq 0$ .  $\triangle$

**III.2.3 Definizione:** Diremo che una successione stocastica  $X = (X_n, \mathcal{F}_n)_{n \geq 0}$  è una **Martingala** quando  $\mathbf{E}|X_n| < +\infty$  e inoltre

$$\mathbf{E}(X_{n+1}|\mathcal{F}_n) = X_n, \quad \mathbf{P}\text{-q.o.}$$

diremo invece che essa è rispettivamente una **Submartingala** o una **Supermartingala** quando, essendo  $\mathbf{E}|X_n| < +\infty$ , risulta

$$\begin{aligned} \mathbf{E}(X_{n+1}|\mathcal{F}_n) &\geq X_n, & \mathbf{P}\text{-q.o.} & \text{ (submartingala)} \\ \mathbf{E}(X_{n+1}|\mathcal{F}_n) &\leq X_n, & \mathbf{P}\text{-q.o.} & \text{ (supermartingala)} \end{aligned} \quad \triangle$$

**III.2.4 Osservazione:** In una successione stocastica  $X = (X_n, \mathcal{F}_n)_{n \geq 0}$  la v.a.  $X_n$  è non solo  $\mathcal{F}_n$ -misurabile, ma anche  $\mathcal{F}_{n+k}$ -misurabile con  $k \geq 0$  a causa del fatto che una filtrazione è crescente; viceversa, per la stessa ragione, non si può dire in generale che essa è anche  $\mathcal{F}_{n-k}$ -misurabile con  $k \geq 1$ . Da un altro punto di vista potremmo dire che in una successione stocastica e rispetto ad una data  $\mathcal{F}_n$  sono misurabili tutte le  $X_k$  con  $k = 0, 1, \dots, n$ , mentre in generale non lo sono quelle con  $k > n$ . Questo riflette il fatto che  $\mathcal{F}_n$  rappresenta l'informazione che si ottiene osservando la successione stocastica *fino all'istante  $n$* , nel senso che in  $\mathcal{F}_n$  acquistano senso tutte le proposizioni relative alle  $X_k$  con  $k \leq n$ , mentre in generale nulla si può dire sulle  $X_k$  con  $k > n$ . Queste osservazioni rendono anche comprensibile la definizione data di successione prevedibile: si tratta di quelle per le quali l'osservazione condotta fino all'istante  $n$  consente anche delle previsioni relative all'istante  $n + 1$ .  $\circ$

**III.2.5 Osservazione:** Le relazioni usate per definire le martingale (o le sub- e supermartingale) riguardano soltanto istanti immediatamente successivi ( $n$  ed  $n + 1$ ), ma si vede facilmente che esse sono estendibili ad istanti successivi arbitrari ( $n$  ed  $m > n$ ), nel senso che, ad esempio per una martingala, è sempre vero anche che

$$\mathbf{E}(X_m|\mathcal{F}_n) = X_n, \quad \mathbf{P}\text{-q.o.}; \quad m > n.$$

Infatti, tenendo conto di (H) della Proposizione II.7.9, del fatto che  $\mathcal{F}_n \subseteq \mathcal{F}_{m-1}$  e della proprietà delle martingale, si ha **P**-q.o.

$$\mathbf{E}(X_m|\mathcal{F}_n) = \mathbf{E}[\mathbf{E}(X_m|\mathcal{F}_{m-1}) | \mathcal{F}_n] = \mathbf{E}(X_{m-1}|\mathcal{F}_n);$$

sicché iterando questa derivazione si ottiene

$$\mathbf{E}(X_m|\mathcal{F}_n) = \mathbf{E}(X_{m-1}|\mathcal{F}_n) = \dots = \mathbf{E}(X_{n+1}|\mathcal{F}_n) = X_n, \quad \mathbf{P}\text{-q.o.}$$

Prendendo il VdA di queste relazioni si ha anche che per una martingala

$$\mathbf{E} X_0 = \mathbf{E} X_1 = \dots = \mathbf{E} X_n = \dots$$

Naturalmente se  $(X_n, \mathcal{F}_n)_{n \geq 0}$  è una sub- o supermartingala la catena di relazioni precedenti è vera se si sostituiscono gli opportuni segni di disuguaglianza. Notiamo inoltre che se una martingala  $X = (X_n, \mathcal{F}_n)_{n \geq 0}$  è anche prevedibile, dalla Proposizione II.7.11 si ha che  $\mathbf{E}(X_{n+1}|\mathcal{F}_n) = X_{n+1}$  e quindi che  $X_n = X_{n+1}$ , **P**-q.o. Pertanto una martingala prevedibile è una successione (**P**-q.o.) costante. Analogamente una submartingala prevedibile è (**P**-q.o.) crescente nel senso che  $X_{n+1} \geq X_n$ , **P**-q.o.

Sarà inoltre utile ricordare che la proprietà che definisce una martingala può essere espressa in modo diverso tenendo conto della Definizione II.7.4 di VdA condizionato e dell'Osservazione II.7.5. Infatti sappiamo che  $\mathbf{E}(X_{n+1}|\mathcal{F}_n)$  è la v.a. tale che, comunque scelto  $A \in \mathcal{F}_n$ , risulta

$$\int_A X_{n+1} d\mathbf{P} = \mathbf{E}(I_A X_{n+1}) = \mathbf{E}[I_A \mathbf{E}(X_{n+1}|\mathcal{F}_n)] = \int_A \mathbf{E}(X_{n+1}|\mathcal{F}_n) d\mathbf{P},$$

mentre la Definizione III.2.3 di martingala richiede che (**P**-q.o.)  $\mathbf{E}(X_{n+1}|\mathcal{F}_n) = X_n$ . Se ne deduce allora facilmente che per una martingala deve risultare<sup>1</sup>

$$\int_A X_{n+1} d\mathbf{P} = \int_A X_n d\mathbf{P}, \quad \forall A \in \mathcal{F}_n, \quad n \geq 0.$$

Si può poi provare il viceversa: se la relazione precedente è vera, allora la successione stocastica  $X = (X_n, \mathcal{F}_n)_{n \geq 0}$  è una martingala, sicché tale relazione può essere usata come definizione equivalente a quella data in III.2.3. In modo del tutto analogo si mostra che la proprietà

$$\int_A X_{n+1} d\mathbf{P} \geq \int_A X_n d\mathbf{P}, \quad \forall A \in \mathcal{F}_n, \quad n \geq 0,$$

---

<sup>1</sup> Ricordiamo che questa relazione non implica che  $X_{n+1} = X_n$  (**P**-q.o.), come potrebbe sembrare da un'applicazione affrettata di (I) della Proposizione II.6.9. Infatti qui  $X_{n+1}$  non è  $\mathcal{F}_n$ -misurabile e quindi la proposizione citata non è applicabile.

definisce una submartingala, e che la relazione con la diseguaglianza invertita caratterizza le supermartingale.  $\circ$

**III.2.6 Osservazione:** Data una successione  $(X_n)_{n \geq 0}$  di v.a. è sempre possibile definire la cosiddetta **filtrazione naturale** i cui elementi  $\mathcal{F}_n^X = \sigma\{X_0, \dots, X_n\}$  sono le  $\sigma$ -algebre generate dalle prime  $n + 1$  v.a. della successione. L'interesse di tale concetto è nell'osservazione che ogni  $X_n$  risulterà ora, per definizione,  $\mathcal{F}_n^X$ -misurabile in modo che  $(X_n, \mathcal{F}_n^X)_{n \geq 0}$  sarà automaticamente una successione stocastica secondo la definizione III.2.1. Quando nel seguito, per brevità, diremo che  $X = (X_n)_{n \geq 0}$  è una successione stocastica o una martingala (o una sub- o supermartingala) sottintenderemo sempre che la successione  $(X_n)_{n \geq 0}$  è tale rispetto alla filtrazione naturale.  $\circ$

**III.2.7 Esempio:** Sia  $(\xi_n)_{n \geq 0}$  una successione di v.a. indipendenti e sia  $\mathcal{F}_n = \sigma\{\xi_0, \dots, \xi_n\}$  la sua filtrazione naturale: se  $\mathbf{E} \xi_n = 0$  per ogni  $n \geq 0$ , si vede subito che posto

$$X_n = \sum_{k=0}^n \xi_k$$

la successione  $X = (X_n, \mathcal{F}_n)_{n \geq 0}$  risulta essere una martingala. Infatti  $(\mathcal{F}_n)_{n \geq 0}$  è anche la filtrazione naturale di  $(X_n)_{n \geq 0}$ , e inoltre, data l'indipendenza delle  $\xi_n$ , si ha

$$\mathbf{E}(X_{n+1} | \mathcal{F}_n) = \mathbf{E}(\xi_{n+1} + X_n | \mathcal{F}_n) = \mathbf{E}(\xi_{n+1} | \mathcal{F}_n) + X_n = \mathbf{E} \xi_{n+1} + X_n = X_n.$$

Inoltre, se  $(\eta_n)_{n \geq 0}$  è una successione di v.a. indipendenti e se non facciamo ipotesi sul valore di  $\mathbf{E} \eta_n$ , potremo sempre costruire una martingala nel modo seguente: detta  $\mathcal{F}_n = \sigma\{\eta_0, \dots, \eta_n\}$  la sua filtrazione naturale e posto

$$Y_n = \sum_{k=0}^n \eta_k$$

avremo che  $Y = (Y_n, \mathcal{F}_n)_{n \geq 0}$  risulterà essere una successione stocastica ma non, in generale, una martingala. Se però consideriamo ora la successione delle  $\xi_n = \eta_n - \mathbf{E} \eta_n$ , queste risulteranno indipendenti con  $\mathbf{E} \xi_n = 0$ , sicché posto

$$X_n = \sum_{k=0}^n \xi_k = Y_n - \mathbf{E} Y_n$$

la successione stocastica  $X = (X_n, \mathcal{F}_n)_{n \geq 0}$  risulterà ora essere una martingala per quanto detto all'inizio di questo Esempio. In sostanza, data una generica successione di v.a. indipendenti  $(\eta_n)_{n \geq 0}$  potremo sempre ottenere una martingala costruendo la successione delle somme  $Y_n$  e depurandola poi dei VdA. In altri termini: le fluttuazioni (attorno ai loro VdA) delle somme di v.a. indipendenti sono sempre delle martingale (rispetto alla loro filtrazione naturale).

Ci sono anche altri metodi per costruire in modo semplice delle martingale: se  $(\xi_n)_{n \geq 0}$  è una successione di v.a. indipendenti,  $(\mathcal{F}_n)_{n \geq 0}$  la sua filtrazione naturale e se  $\mathbf{E} \xi_n = 1$  per ogni  $n \geq 0$ , posto

$$X_n = \prod_{k=0}^n \xi_k$$

si vede che  $X = (X_n, \mathcal{F}_n)_{n \geq 0}$  è una martingala. Infatti si ha

$$\begin{aligned} \mathbf{E}(X_{n+1} | \mathcal{F}_n) &= \mathbf{E}(\xi_{n+1} \xi_n \dots \xi_0 | \mathcal{F}_n) = \mathbf{E}(\xi_{n+1} | \mathcal{F}_n) \xi_n \dots \xi_0 \\ &= \xi_0 \dots \xi_n \mathbf{E} \xi_{n+1} = \prod_{k=0}^n \xi_k = X_n. \end{aligned}$$

Se poi  $\xi$  è una v.a. integrabile ( $\mathbf{E} |\xi| < +\infty$ ) e  $(\mathcal{F}_n)_{n \geq 0}$  è un'arbitraria filtrazione, posto  $X_n = \mathbf{E}(\xi | \mathcal{F}_n)$  si verifica subito che anche  $(X_n, \mathcal{F}_n)_{n \geq 0}$  è una martingala; infatti  $\mathbf{E}(\xi | \mathcal{F}_n)$  è ovviamente  $\mathcal{F}_n$ -misurabile e inoltre (vedi Proposizione II.7.9), essendo  $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ , si ha (**P**-q.o.)

$$\mathbf{E}(X_{n+1} | \mathcal{F}_n) = \mathbf{E}[\mathbf{E}(\xi | \mathcal{F}_{n+1}) | \mathcal{F}_n] = \mathbf{E}(\xi | \mathcal{F}_n) = X_n.$$

In modo del tutto analogo possiamo costruire sub- o supermartingale da successioni di v.a. indipendenti: ad esempio se le  $\xi_n$  sono v.a. non negative ( $\mathbf{E} \xi_n \geq 0$ ), e se si pone

$$X_n = \sum_{k=0}^n \xi_k$$

si verifica con un calcolo simile a quello iniziale che  $X = (X_n, \mathcal{F}_n)_{n \geq 0}$  è ora una submartingala. Infine, se  $(X_n, \mathcal{F}_n)_{n \geq 0}$  è una generica martingala e  $g(x)$  una funzione convessa (con  $\mathbf{E} |g(X_n)| < +\infty$  per  $n \geq 0$ ), allora  $(g(X_n), \mathcal{F}_n)_{n \geq 0}$  è una supermartingala. Infatti dalla Diseguaglianza di Jensen (Proposizione II.6.19) si ha ora che

$$\mathbf{E}(g(X_{n+1}) | \mathcal{F}_n) \geq g(\mathbf{E}(X_{n+1} | \mathcal{F}_n)) = g(X_n). \quad \diamond$$

**III.2.8 Esempio:** Il concetto ed anche il nome di *martingala* derivano, come vedremo meglio in seguito, dal mondo del gioco d'azzardo: supponiamo che le v.a.  $\xi_n$  rappresentino la vincita di un giocatore alla mano  $n$ -ma di un certo gioco e supponiamo che esse, diversamente da quanto supposto negli esempî esaminati in precedenza, non siano indipendenti. È possibile infatti pensare che la somma di denaro puntata alla mano  $(n+1)$ -ma sia stabilita sulla base della conoscenza dell'andamento del gioco fino alla mano  $n$ -ma. Ad esempio la *martingala* classica è una strategia di puntate che consiste nel raddoppio della posta in gioco ogni volta che (nella mano precedente) la scommessa è stata persa, e nell'arresto del gioco immediatamente dopo la prima mano vincente. Discuteremo più in dettaglio nel seguito questo tipo di strategia e la sua connessione con il concetto di martingala



esposto in Definizione III.2.3; qui ci limiteremo invece ad esaminare le relazioni fra questa definizione ed il concetto di **gioco assolutamente equo**, intendendo con questo un gioco nel quale la conoscenza della storia passata non permette ad un giocatore di migliorare (in media) le proprie possibilità di vittoria scegliendo un'adeguata strategia. Detta in particolare  $(\mathcal{F}_n)_{n \geq 1}$  la filtrazione naturale della nostra successione  $(\xi_n)_{n \in \mathbf{N}}$  di v.a. che rappresentano le vincite al gioco, diremo che la successione stocastica  $(\xi_n, \mathcal{F}_n)_{n \geq 1}$  è assolutamente equa quando

$$\mathbf{E} \xi_1 = 0, \quad \mathbf{E} (\xi_{n+1} | \mathcal{F}_n) = 0, \quad n \geq 1, \quad \mathbf{P}\text{-q.o.}$$

(da cui segue anche  $0 = \mathbf{E} \xi_1 = \dots = \mathbf{E} \xi_n = \dots$ ). In queste condizioni, infatti, è illusorio supporre che l'osservazione del gioco fino alla mano  $n$ -ma possa permettere di escogitare una strategia di puntate che (in media) rechi vantaggio ad uno dei due giocatori.

La stretta relazione fra le martingale ed i giochi assolutamente equi è ora messa in evidenza dal fatto che, se la costante  $C$  rappresenta il capitale iniziale di uno dei due giocatori, e la successione delle v.a.

$$X_n = \sum_{k=1}^n \xi_k + C$$

il capitale di quel giocatore dopo l' $n$ -ma mano di gioco, e se il gioco è assolutamente equo, non solo si ha che  $\mathbf{E} X_n = C$  per ogni  $n \geq 1$ , ma anche che la successione stocastica  $(X_n, \mathcal{F}_n)_{n \geq 1}$  è una martingala dato che le  $X_n$  sono evidentemente  $\mathcal{F}_n$ -misurabili e che inoltre, data l'assoluta equità del gioco,

$$\mathbf{E} (X_{n+1} | \mathcal{F}_n) = \mathbf{E} (\xi_{n+1} | \mathcal{F}_n) + \mathbf{E} (X_n | \mathcal{F}_n) = X_n.$$

Val la pena di notare che il presente risultato non è un corollario degli esempi precedenti in quanto, a differenza di quanto accadeva nei casi discussi nell'Esempio III.2.7, le nostra  $\xi_n$  non sono più indipendenti. Inoltre è anche possibile affermare, viceversa, che data una generica martingala  $(X_n, \mathcal{F}_n)_{n \geq 0}$ , essa può sempre essere posta nella forma

$$X_n = \sum_{k=1}^n \xi_k + C$$

dove le  $\xi_n$  sono v.a.  $\mathcal{F}_n$ -misurabili tali che la successione stocastica  $(\xi_n, \mathcal{F}_n)_{n \geq 1}$  risulti assolutamente equa. Infatti basterà porre

$$C = \mathbf{E} X_1, \quad \xi_1 = X_1 - \mathbf{E} X_1, \quad \dots, \quad \xi_{n+1} = X_{n+1} - X_n, \quad \dots \quad \forall n \geq 1$$

ed osservare che  $\xi_n$  è  $\mathcal{F}_n$ -misurabile, che

$$X_n = (X_n - X_{n-1}) + \dots + (X_2 - X_1) + (X_1 - \mathbf{E} X_1) + \mathbf{E} X_1 = \sum_{k=1}^n \xi_k + C,$$

che  $\mathbf{E} \xi_1 = 0$ , ed infine che (essendo  $X_n$  una martingala)  $\mathbf{P}$ -q.o. si ha

$$\mathbf{E}(\xi_{n+1}|\mathcal{F}_n) = \mathbf{E}(X_{n+1}|\mathcal{F}_n) - \mathbf{E}(X_n|\mathcal{F}_n) = X_n - X_n = 0.$$

In conclusione potremo dire che il concetto di martingala esposto nella Definizione III.2.3 è strettamente connesso con i giochi assolutamente equi. In particolare nei giochi classici di tipo *testa o croce* con puntate fisse e moneta ben bilanciata la successione delle vincite  $(\xi_n)_{n \in \mathbf{N}}$  è composta di v.a. indipendenti a media nulla (come quelle dell'Esempio III.2.7): in questo caso la successione delle vincite dopo  $n$  mani  $X_n$  è una martingala. Se ora si complica il gioco permettendo al giocatore di stabilire l'ammontare della posta ad ogni mano secondo una strategia che tiene conto delle mani già giocate, la successione delle vincite è composta di v.a. che non sono più indipendenti, ma, come sarà dimostrato più oltre, se la moneta con cui si gioca è ben bilanciata, il gioco resta assolutamente equo e conseguentemente la successione delle vincite dopo  $n$  mani  $X_n$  continua ad essere una martingala. Aggiungeremo infine che, in modo analogo, i concetti di sub- e supermartingala sono direttamente connessi rispettivamente con giochi d'azzardo vantaggiosi o svantaggiosi, tali cioè che ( $\mathbf{P}$ -q.o.) risulti

$$\begin{aligned} \mathbf{E}(\xi_{n+1}|\mathcal{F}_n) &\geq 0, \quad n \geq 1 \quad (\text{vantaggioso - submartingala}) \\ \mathbf{E}(\xi_{n+1}|\mathcal{F}_n) &\leq 0, \quad n \geq 1 \quad (\text{svantaggioso - supermartingala}) \quad \diamond \end{aligned}$$

**III.2.9 Definizione:** Una v.a.  $\tau$  che assume valori in  $\{0, 1, \dots, +\infty\}$  si chiama **Tempo di Markov** rispetto ad una data filtrazione  $(\mathcal{F}_n)_{n \geq 0}$  quando

$$\{\tau = n\} \in \mathcal{F}_n, \quad n \geq 0;$$

se inoltre  $\mathbf{P}(\tau < +\infty) = 1$ ,  $\tau$  si chiama **Tempo d'arresto**. △

**III.2.10 Osservazione:** Un tempo di Markov è una v.a. (cioè è  $\mathcal{F}$ -misurabile), ma in generale non è  $\mathcal{F}_n$ -misurabile rispetto a nessuna delle  $\sigma$ -algebre della filtrazione  $(\mathcal{F}_n)_{n \geq 0}$ . Infatti la sua definizione richiede solo che l'evento  $\{\tau = n\}$  sia un elemento di  $\mathcal{F}_n$  (con  $n \geq 0$  arbitrario) e questo non basta a stabilire la  $\mathcal{F}_n$ -misurabilità. Un particolare tempo d'arresto è stato discusso nel Capitolo I.10 in connessione con il problema della rovina di un giocatore. Per costruirne un esempio un po' più generale, data una successione stocastica  $(X_n, \mathcal{F}_n)_{n \geq 0}$ , prendiamo un evento  $B \in \mathcal{B}(\mathbf{R})$  e definiamo il *tempo di primo impatto* in  $B$  come la v.a.

$$\tau_B = \inf\{n \geq 0 : X_n \in B\}$$

con l'accorgimento di porre  $\tau_B = +\infty$  se risulta  $\{n \geq 0 : X_n \in B\} = \emptyset$ . Si verifica ora facilmente che si tratta di un tempo di Markov (che generalizza quello del Capitolo I.10) dato che

$$\{\tau_B = n\} = \{X_0 \notin B, \dots, X_{n-1} \notin B, X_n \in B\}$$

è sicuramente un evento di  $\mathcal{F}_n$  visto che per ipotesi  $X_0, \dots, X_n$  sono tutte  $\mathcal{F}_n$ -misurabili. D'altra parte  $\{\tau_B = +\infty\}$  risulta essere un evento di  $\mathcal{F}$  dato che

$$\{\tau_B = +\infty\} = \bigcap_n \{X_n \notin B\} = \bigcap_n \overline{\{X_n \in B\}}$$

e che  $\{X_n \in B\}$  sono tutti eventi della  $\sigma$ -algebra  $\mathcal{F}$ . ○

**III.2.11 Osservazione:** Data una successione stocastica  $(X_n, \mathcal{F}_n)_{n \geq 0}$  ed un tempo di Markov (rispetto alla stessa filtrazione  $(\mathcal{F}_n)_{n \geq 0}$ )  $\tau$ , definiamo

$$X_\tau = \sum_{n \in \mathbf{N}} X_n I_{\{\tau = n\}}$$

intendendo, in particolare, con questo che  $X_\tau = 0$  sull'evento  $\{\tau = +\infty\}$ . Si verifica ora che  $X_\tau$  è una v.a.  $\mathcal{F}$ -misurabile. Infatti gli eventi  $\{\tau = n\}$  sono tutti disgiunti al variare di  $n$ , sicché preso  $\omega \in \Omega$  al più uno solo degli indicatori presenti nella definizione è diverso da zero<sup>2</sup>: se allora  $\tau(\omega) = n$ , si ha  $X_\tau(\omega) = X_n(\omega)$  dove  $n = \tau(\omega)$ ; se invece  $\tau(\omega) = +\infty$  si ha  $X_\tau(\omega) = 0$ . In pratica  $X_\tau$  è la nostra successione stocastica osservata in istanti aleatori, nel senso che per un dato  $\omega$  il valore di  $X_\tau$  è  $X_n(\omega)$  dove  $n = \tau(\omega)$ . Pertanto, preso l'evento<sup>3</sup>  $B \in \mathcal{B}(\mathbf{R})$  si ha

$$\{X_\tau \in B\} = \bigcup_{n \in \mathbf{N}} \{X_n \in B, \tau = n\} \in \mathcal{F}$$

dato che  $\{X_n \in B, \tau = n\} \in \mathcal{F}$  per ogni  $n \geq 0$ . Ciò posto sappiamo che se  $(X_n, \mathcal{F}_n)_{n \geq 0}$  è una martingala si ha  $\mathbf{E} X_n = \mathbf{E} X_0$  per ogni  $n \geq 0$  (e analogamente, con le appropriate disequaglianze, per le sub- e supermartingale): vogliamo ora esaminare se tale proprietà delle martingale (o delle sub- e supermartingale) è conservata quando ad  $n$  sostituiamo un tempo di Markov  $\tau$  nel senso descritto in questa osservazione. In altre parole vogliamo esaminare se è ancora possibile affermare la validità della relazione  $\mathbf{E} X_\tau = \mathbf{E} X_0$  o delle analoghe disequaglianze nei casi appropriati. L'Esempio seguente mostrerà che in generale la risposta è negativa mentre la Proposizione successiva chiarirà sotto quali condizioni tale affermazione può essere fatta. ○

**III.2.12 Esempio:** Supponiamo che  $(\eta_n)_{n \in \mathbf{N}}$  sia una successione di v.a. i.i.d. con  $\mathbf{P}(\eta_n = +1) = p$ ,  $\mathbf{P}(\eta_n = -1) = q$  e  $p + q = 1$  e, per fissare le idee, stabiliamo che, per un dato gioco d'azzardo,  $\{\eta_n = +1\}$  rappresenti l'evento *il giocatore vince l'n-ma mano* e che  $\{\eta_n = -1\}$  rappresenti l'evento *il giocatore perde l'n-ma mano*.

<sup>2</sup> Non si può escludere che siano anche tutti nulli nel senso che se  $\tau(\omega) = +\infty$  risulterà  $X_\tau = 0$ .

<sup>3</sup> Per semplicità supporremo che  $B$  non contenga il numero 0; se invece lo contenesse basterebbe riprodurre lo stesso ragionamento aggiungendo l'unione con l'evento  $\{\tau = +\infty\} \in \mathcal{F}$ .

Se all' $n$ -ma mano il giocatore ha scommesso una quantità di denaro<sup>4</sup> pari a  $V_n$ , la v.a.  $\xi_n = \eta_n V_n$  rappresenterà la vincita all' $n$ -ma mano e

$$X_n = \sum_{k=1}^n \xi_k = \xi_n + X_{n-1}, \quad X_0 = 0$$

la vincita complessiva dopo  $n$  mani. Considerata ora la filtrazione  $\mathcal{F}_0 = \{\emptyset, \Omega\}$ ,  $\mathcal{F}_n = \sigma\{\eta_1, \dots, \eta_n\}$ , dato che  $V_n$  dipende dall'andamento del gioco fino all'istante  $(n-1)$ -mo<sup>5</sup> è facile verificare che  $V_n$  è  $\mathcal{F}_{n-1}$ -misurabile sicché la successione stocastica  $(V_n, \mathcal{F}_n)_{n \geq 1}$ , che rappresenta la strategia del nostro giocatore, risulta prevedibile. Per la successione stocastica  $(\xi_n, \mathcal{F}_n)_{n \geq 1}$ , invece, potremo dire che, siccome

$$\mathbf{E}(\xi_{n+1} | \mathcal{F}_n) = \mathbf{E}(V_{n+1} \eta_{n+1} | \mathcal{F}_n) = V_{n+1} \mathbf{E}(\eta_{n+1} | \mathcal{F}_n) = V_{n+1} \mathbf{E} \eta_{n+1} = (p - q) V_{n+1}$$

essa risulta assolutamente equa se  $p = q = 1/2$  (nel qual caso, come già abbiamo notato, la successione  $X_n$  delle somme è una martingala), favorevole se  $p > q$  (con  $X_n$  submartingala) e sfavorevole se  $p < q$  (con  $X_n$  supermartingala).

Consideriamo ora la seguente strategia di puntate (nota con il nome di *martingala*<sup>6</sup>):

$$V_1 = 1, \quad n = 1$$

$$V_n = \begin{cases} 2^{n-1}, & \text{se } \eta_1 = \dots = \eta_{n-1} = -1 \\ 0, & \text{altrimenti} \end{cases}, \quad n \geq 2$$

<sup>4</sup> Qui supporremo che tale quantità di denaro sia una v.a., cioè una quantità positiva che non è prefissata ma dipende dall'andamento del gioco.

<sup>5</sup> Cioè è funzione di  $\eta_1, \dots, \eta_{n-1}$  e di  $V_1, \dots, V_{n-1}$ , ma siccome a loro volte anche le  $V_j$  sono funzioni delle  $\eta_1, \dots, \eta_{n-1}$ , in definitiva  $V_n$  è funzione di  $\eta_1, \dots, \eta_{n-1}$ .

<sup>6</sup> La parola francese *martingale* designa comunemente una cintura adoperata come ornamento di capi di vestiario e con questo significato si trova usata fin dal XV secolo. L'origine è poco chiara: potrebbe derivare dallo spagnolo (di origine araba) *almartaga* che significa appunto *briglia, correggia* (vedi *Nouveau dictionnaire étymologique et historique*, Larousse, Paris, 1985), oppure dal provenzale *martegalo* con un riferimento agli abitanti di Martigues, cittadina situata presso Marsiglia alle foci del Rodano (vedi *Dictionnaire historique de la langue française*, Le Robert, Paris, 1992). Nella seconda interpretazione si osserva inoltre che gli abitanti di Martigues godono di una reputazione regionale di ingenuità e di stravaganza, sicché l'espressione *à la martingale* esprime fin dalle prime testimonianze letterarie un'idea di assurdità e di illogicità. Nel campo del gioco d'azzardo la parola *martingala* si trova invece adoperata dalla fine del XVIII secolo, e questo uso appare inizialmente connesso con il significato peggiorativo illustrato prima: la strategia che consiste nel raddoppiare la posta se si perde nella mano precedente contiene, in un certo senso, degli elementi di assurdità. Questa evoluzione semantica è però giudicata poco sicura. Peraltro a partire dal XIX secolo *martingala* indica piuttosto una combinazione più o meno scientifica di puntate che garantisca dei vantaggi: significato che sembra allontanarsi da qualunque idea di assurdità. Bisogna precisare, infine, che nel campo del gioco d'azzardo il nome si riferisce alla strategia seguita indipendentemente dall'assoluta equità del gioco (e quindi indipendentemente dal fatto che la successione delle vincite sia una vera martingala nel senso della Definizione III.2.3): si potrà cioè giocare alla martingala anche in situazioni nelle quali i modelli matematici più adatti sarebbero quelli delle sub- e supermartingale.

in base alla quale il nostro giocatore scommette  $V_1 = 1$  alla prima mano, raddoppia la posta ogni volta che perde e abbandona il gioco non appena vince una mano. Se nelle prime  $n$  mani egli perde e vince per la prima volta alla  $(n + 1)$ -ma mano egli esce dal gioco con una vincita positiva in quanto

$$X_{n+1} = V_{n+1} - \sum_{k=1}^n V_k = 2^n - \sum_{k=1}^n 2^{k-1} = 2^n - (2^n - 1) = 1.$$

Si pone inoltre il problema di stimare *quando* il gioco si arresta: posto  $\tau = \inf\{n \geq 1 : X_n = 1\}$ , si ottiene facilmente che

$$\mathbf{P}(\tau = n) = \mathbf{P}(\eta_1 = -1, \dots, \eta_{n-1} = -1, \eta_n = 1) = q^{n-1}p$$

e quindi, essendo gli eventi  $\{\tau = n\}$  tutti disgiunti,

$$\mathbf{P}(\tau < +\infty) = \mathbf{P}\left(\bigcup_{n \geq 1} \{\tau = n\}\right) = \sum_{n \geq 1} \mathbf{P}(\tau = n) = \sum_{n \geq 1} q^{n-1}(1-q) = 1.$$

Ciò mostra che  $\tau$  è un tempo d'arresto cioè che il gioco si arresta in un tempo arbitrario ma finito (con probabilità 1). Il VdA di questo tempo d'arresto si calcola anche facilmente e vale

$$\mathbf{E} \tau = \sum_n n p q^{n-1} = p \frac{d}{dq} \sum_n q^n = p \frac{d}{dq} \left( \frac{1}{1-q} \right) = \frac{p}{(1-q)^2} = \frac{1}{p}.$$

Ma vi è di più: posto

$$X_\tau = \sum_{n \in \mathbf{N}} X_n I_{\{\tau = n\}}$$

siccome  $\{\tau = n\} = \{X_1 = -1, \dots, X_{n-1} = -1 - 2^{n-1}, X_n = 1\}$ , si ha essendo  $\{\tau = n\} \subseteq \{X_n = 1\}$  che

$$\{X_\tau = 1\} = \bigcup_{n \geq 1} \{X_n = 1, \tau = n\} = \bigcup_{n \geq 1} \{\tau = n\}$$

e quindi con un calcolo analogo al precedente

$$\mathbf{P}(X_\tau = 1) = \sum_{n \geq 1} \mathbf{P}(\tau = n) = 1.$$

Ne segue che  $X_\tau = 1$  (**P**-q.o.) e quindi che  $\mathbf{E} X_\tau = 1$ , pur essendo  $0 = \mathbf{E} X_0 = \dots = \mathbf{E} X_n = \dots$  nel caso equo (martingala) e  $0 = \mathbf{E} X_0 \geq \dots \geq \mathbf{E} X_n \geq \dots$  nel caso sfavorevole (supermartingala). Questo è il tipo di risultati che ha stabilito, per generazioni di giocatori, il fascino della strategia della martingala: anche nel caso di giochi sfavorevoli ( $p < q$ ) infatti questa strategia sembra consentire di terminare il gioco in vantaggio e in un tempo finito (benché arbitrario) con probabilità 1;

dato che la stazionarietà (o la monotonia) nel tempo dei VdA di  $X_n$  sembra essere posta in discussione se si sostituisce  $n$  con un tempo d'arresto  $\tau$  suggerendo così la possibilità di guadagni certi.

Qualche dubbio sulla reale applicabilità di questi risultati viene però insinuato dall'osservazione che pur essendo  $\tau$  finito ( $\mathbf{P}$ -q.o.) il suo VdA può essere piuttosto grande dato che  $\mathbf{E}\tau = 1/p$ . Se  $p = 1/2$ , come nel caso di un gioco assolutamente equo (*testa o croce* con una moneta ben bilanciata), il VdA  $\mathbf{E}\tau = 2$  è piuttosto piccolo e potrebbe convenire rischiare; se però  $p \ll 1$ , come ad esempio quando si punta su un numero alla roulette,  $\mathbf{E}\tau$  può essere piuttosto grande e si deve mettere in conto il rischio di perdere una cifra dell'ordine di  $2^{\mathbf{E}\tau}$  prima di poter uscire dal gioco con una vincita di 1. Ad esempio per una roulette con 36 numeri la probabilità di vincere scommettendo su un numero è  $1/36$  e conseguentemente il capitale necessario per poter tentare di vincere 1 unità con una martingala è  $2^{36} \simeq 6,8 \times 10^{10}$ . Da un punto di vista formale il paradosso nasce dal fatto che la regola  $\mathbf{E}X_n = \mathbf{E}X_0$  delle martingale (o  $\mathbf{E}X_n \leq \mathbf{E}X_0$  delle supermartingale) non si trasferisce automaticamente alla v.a.  $X_\tau$  con la conseguenza che la strategia della martingala sembra consentire al giocatore di chiudere il gioco con una vincita di 1 in un tempo finito (con probabilità 1). In realtà le relazioni  $\mathbf{E}X_\tau = \mathbf{E}X_0$  e  $\mathbf{E}X_\tau \leq \mathbf{E}X_0$  (che eliminerebbero il paradosso facendo sfumare ogni speranza di sicure vincite alla martingala) può essere garantita solo sotto alcune condizioni che saranno esaminate nella Proposizione seguente, mentre violazioni di tale relazione, come quelle espone nel presente Esempio, non possono che sorgere in situazioni che non garantiscano le suddette condizioni e che potremmo descrivere come *fisicamente non realizzabili*. Infatti, nella discussione precedente, abbiamo supposto che il nostro giocatore abbia a disposizione un tempo di gioco e un capitale illimitati. D'altra parte, se è vero che risulta  $\mathbf{P}(\tau < +\infty) = 1$ , dobbiamo ricordare che ciò non implica che esista un numero finito  $N$  tale che  $\mathbf{P}(\tau \leq N) = 1$  in quanto per ogni numero  $N$  si avrà:

$$\mathbf{P}(\tau \leq N) = \sum_{n=1}^N pq^{n-1} = 1 - \sum_{n>N} pq^{n-1} < 1.$$

Mostriamo tra breve che i risultati del nostro calcolo sarebbero stati diversi se fossimo costretti a sopporre l'esistenza di un  $N$  tale che  $\mathbf{P}(\tau \leq N) = 1$ . Inoltre si può vedere facilmente che, se il capitale  $C$  a disposizione del giocatore per le puntate è limitato ( $C < +\infty$ ), esiste sempre un numero  $N$  tale che  $\mathbf{P}(\tau \leq N) = 1$ . Infatti in questo caso dovremo modificare la definizione del nostro tempo d'arresto:

$$\tau = \inf\{n \geq 1 : X_n = 1, \text{ oppure } X_n < -C\}$$

per tenere conto dell'eventualità che il gioco finisca per rovina del giocatore. Conseguentemente il tempo  $\tau$  non può superare il valore  $n$  per il quale il valore della perdita supera  $C$  e che si ricava da

$$\sum_{k=0}^{n-1} 2^k = \frac{1-2^n}{1-2} = 2^n - 1 > C,$$

sicché deve risultare (indicando con  $[x]$  la parte intera di un numero  $x$ )

$$\tau \leq N = [\log_2(C - 1)] + 1, \quad \mathbf{P}\text{-q.o.}$$

Infatti o il giocatore vince, ed esce dal gioco, entro tale istante oppure si rovina ed il gioco si chiude comunque. In conclusione, se il capitale iniziale  $C$  a disposizione del giocatore non è illimitato, il tempo di gioco  $\tau$  non può superare ( $\mathbf{P}\text{-q.o.}$ ) un certo numero  $N$  e in tali condizioni la Proposizione che segue ci consente di affermare che le proprietà delle martingale (o delle sub- e supermartingale) si estendono automaticamente anche alla v.a.  $X_\tau$  cosicché in particolare  $\mathbf{E} X_\tau = 0$  per i giochi assolutamente equi e  $\mathbf{E} X_\tau \leq 0$  per quelli sfavorevoli. In tal caso il gioco si arresta con probabilità 1 in un tempo inferiore ad  $N = [\log_2(C - 1)] + 1$ , ma non obbligatoriamente con la vittoria del nostro giocatore: bisogna infatti tenere in conto l'eventualità che egli si rovini prima di vincere la prima mano. L'enunciato che stiamo per discutere è una versione semplificata, ma ancora adeguata al livello degli esempi qui presentati, dell'importante Teorema d'arresto di J.L. Doob (vedi ad esempio **A.N. Shiryaev**: *Probability*; Springer, New York, 1984, p. 457), il matematico americano che maggiormente contribuì, assieme a P. Lévy, allo sviluppo della teoria delle martingale lungo gli anni '40 e '50.  $\diamond$

**III.2.13 Proposizione:** Data una martingala  $(X_n, \mathcal{F}_n)_{n \geq 0}$  (rispettivamente sub- o supermartingala) ed un tempo d'arresto  $\tau$ , se esiste un numero intero  $N < +\infty$  tale che  $\tau \leq N$  ( $\mathbf{P}\text{-q.o.}$ ), allora si ha che

$$\mathbf{E}(X_\tau | \mathcal{F}_0) = X_0 \quad \mathbf{P}\text{-q.o.}$$

(rispettivamente  $\mathbf{E}(X_\tau | \mathcal{F}_0) \geq X_0$  e  $\mathbf{E}(X_\tau | \mathcal{F}_0) \leq X_0$ ,  $\mathbf{P}\text{-q.o.}$ ), e in particolare che  $\mathbf{E} X_\tau = \mathbf{E} X_0$  (rispettivamente  $\mathbf{E} X_\tau \geq \mathbf{E} X_0$  e  $\mathbf{E} X_\tau \leq \mathbf{E} X_0$ ).

**Dimostrazione:** Data la Definizione II.7.4 dei VdA condizionati, sarà sufficiente provare che, comunque scelto  $A \in \mathcal{F}_0$ , si ha nei tre casi rispettivamente

$$\mathbf{E}(X_\tau I_A) = \mathbf{E}(X_0 I_A), \quad \mathbf{E}(X_\tau I_A) \geq \mathbf{E}(X_0 I_A), \quad \mathbf{E}(X_\tau I_A) \leq \mathbf{E}(X_0 I_A).$$

Per brevità dimostreremo la proposizione solo nel caso delle supermartingale (terzo caso: quello del gioco sfavorevole nel quale il paradosso discusso in precedenza è più evidente) osservando che gli altri due casi si dimostrano in modo identico sostituendo ove necessario gli opportuni segni di eguaglianza o disequaglianza. Preso allora  $A \in \mathcal{F}_0$  (per cui sarà anche  $A \in \mathcal{F}_k$  per ogni  $k \geq 0$ ) e usando le proprietà delle supermartingale abbiamo che

$$\begin{aligned} \mathbf{E}(X_0 I_A) &= \mathbf{E}(X_0 I_A I_{\{\tau=0\}}) + \mathbf{E}(X_0 I_A I_{\{\tau \geq 1\}}) \\ &\geq \mathbf{E}(X_\tau I_A I_{\{\tau=0\}}) + \mathbf{E}(\mathbf{E}(X_1 | \mathcal{F}_0) I_A I_{\{\tau \geq 1\}}). \end{aligned}$$

D'altra parte, siccome in generale comunque preso  $k \geq 0$  si ha

$$\begin{aligned} \{\tau \geq k + 1\} &= \{\tau > k\} = \overline{\{\tau \leq k\}} = \overline{\{\tau = 0\}} \cup \dots \cup \overline{\{\tau = k\}} \\ &= \overline{\{\tau = 0\}} \cap \dots \cap \overline{\{\tau = k\}} \in \mathcal{F}_k, \end{aligned}$$

si avrà anche che  $B = A \cap \{\tau \geq 1\} \in \mathcal{F}_0$  per cui, sempre per la definizione di VdA condizionato, si potrà scrivere che  $\mathbf{E}(\mathbf{E}(X_1|\mathcal{F}_0)I_B) = \mathbf{E}(X_1I_B)$ ; ne segue allora che

$$\mathbf{E}(X_0I_A) \geq \mathbf{E}(X_\tau I_A I_{\{\tau=0\}}) + \mathbf{E}(X_1 I_A I_{\{\tau \geq 1\}}).$$

Ripetendo  $m$  volte la stessa derivazione otterremo allora, comunque scelto  $m \geq 0$ , che

$$\mathbf{E}(X_0I_A) \geq \mathbf{E}(X_\tau I_A I_{\{0 \leq \tau \leq m\}}) + \mathbf{E}(X_{m+1} I_A I_{\{\tau \geq m+1\}}).$$

Per ipotesi, però, sappiamo che se  $m \geq N$  allora si ha  $\mathbf{P}(\tau \geq m+1) = 0$  e  $\mathbf{P}(0 \leq \tau \leq m) = 1$ , sicché non appena  $m$  supera  $N$  otteniamo la relazione  $\mathbf{E}(X_\tau I_A) \leq \mathbf{E}(X_0 I_A)$  che, data l'arbitrarietà di  $A \in \mathcal{F}_0$ , dimostra la nostra tesi. La relazione  $\mathbf{E} X_\tau \leq \mathbf{E} X_0$  si ricava infine prendendo i VdA di ambo i membri della relazione precedente.  $\square$

**III.2.14 Osservazione:** Risultati particolarmente importanti nella teoria delle Martingale sono i cosiddetti teoremi di convergenza. Non abbiamo qui la possibilità di analizzarli né di soffermarci sul loro significato per cui ci limiteremo ad enunciare solo uno a titolo di esempio rinviando ad altri manuali per una trattazione completa (vedi ad esempio **A.N. Shiryaev**: *Probability*; Springer, New York, 1984, cap. VII, § 4, p. 476).  $\circ$

**III.2.15 Teorema (di convergenza delle martingale):** Se  $(X_n, \mathcal{F}_n)_{n \geq 0}$  è una martingala con  $\mathbf{E}(X_n^2) < C < +\infty$  per ogni  $n \in \mathbf{N}$ , allora esiste una v.a.  $X$  tale che  $X_n \xrightarrow{q.o.} X$  e che  $\mathbf{E} X_n = \mathbf{E} X$  per ogni  $n \in \mathbf{N}$ .

**Dimostrazione:** Omessa (vedi ad esempio **W. Feller**: *An introduction to probability theory and its applications - Vol. II*; J.Wiley & Sons, New York, 1971, p. 242.)  $\square$



## III.3 Catene di Markov

**III.3.1 Definizione:** Una successione stocastica  $X = (X_n, \mathcal{F}_n)_{n \geq 0}$  si dice **Catena di Markov**<sup>1</sup> se, comunque scelto  $B \in \mathcal{B}(\mathbf{R})$ , risulta

$$\mathbf{P}(X_n \in B | \mathcal{F}_m) = \mathbf{P}(X_n \in B | X_m), \quad \mathbf{P}\text{-q.o.}, \quad 0 \leq m < n.$$

Diremo inoltre, in forma abbreviata, che  $(X_n)_{n \geq 0}$  è una Catena di Markov se  $\mathcal{F}_n = \sigma\{X_0, \dots, X_n\}$  e  $(\mathcal{F}_n)_{n \geq 0}$  è la filtrazione naturale. In generale le  $X_n$  (finora intese come v.a. a valori reali) possono prendere valori anche in uno spazio misurabile generico  $(E, \mathcal{E})$  che prende il nome di **Spazio delle fasi**. Inoltre la Catena di Markov  $X$  si dice **discreta** quando l'insieme  $E$  dello spazio delle fasi è al più numerabile mentre si dice **finita** se  $E$  è finito. Gli elementi di  $E$  si dicono anche **stati** della catena.  $\triangle$

---

<sup>1</sup> Andrej Andreevič Markov (1856-1922), figlio di un funzionario zarista allontanato dall'impiego a causa del suo rifiuto di rendersi complice delle attività illegali dei suoi superiori, studiò matematica all'Università di Pietroburgo a partire dal 1874 sotto la guida di P.L. Chebyshev. Laureatosi nel 1878 e conseguito il dottorato nel 1884, si interessò inizialmente di frazioni continue e dei valori limite di determinati integrali. Professore all'Università di Pietroburgo dal 1886 e membro ordinario dell'Accademia delle Scienze della medesima città dal 1896, i suoi interessi scientifici si orientarono verso la teoria della probabilità dalla fine degli anni '90. In particolare, continuando i lavori di Chebyshev, sviluppò il cosiddetto metodo dei momenti per la dimostrazione del *Teorema limite centrale* che a quell'epoca non era stato ancora provato in maniera rigorosa e del quale non si conoscevano gli esatti limiti di applicabilità. Nel 1900 pubblicò il suo corso di *Calcolo delle probabilità* nel quale inserì la nuova dimostrazione. Sebbene il suo metodo dei momenti sia stato rapidamente soppiantato dal più flessibile metodo delle funzioni caratteristiche (introdotto già nel 1901 da A.M. Ljapunov) i suoi lavori rappresentarono una pietra miliare nella trasformazione della teoria delle probabilità in una disciplina matematica rigorosa. Durante il primo decennio del nuovo secolo Markov si distinse anche per alcune iniziative politiche con un carattere di aperta ostilità nei confronti del governo zarista. In particolare ricorderemo la sua proposta, lanciata nel 1913, di celebrare in quell'anno, anziché il terzo centenario della dinastia dei Romanov, il secondo centenario della *Legge dei grandi numeri* pubblicata nel 1713 nell'*Ars Coniectandi* di J. Bernoulli. Dal 1905 si volse allo studio delle grandezze aleatorie dipendenti (*Estensione della Legge dei grandi numeri a grandezze dipendenti l'una dall'altra*, 1906) mostrando che l'indipendenza non è una condizione necessaria per l'applicazione della Legge dei grandi numeri. Tra gli esempi riportati in quest'opera figurano le cosiddette *Catene di Markov* caratterizzate dal fatto che il passato e il futuro risultano indipendenti rispetto ad un presente noto. Nei successivi articoli, pubblicati in un periodo che arriva fino al 1912, discusse i concetti di ergodicità e le condizioni sotto le quali una catena di Markov verifica anche i Teoremi limite. Questi suoi lavori posero le basi di un ramo importantissimo del calcolo delle probabilità che generalizzava ai fenomeni casuali il principio classico secondo il quale lo stato attuale del sistema determina la sua evoluzione nel tempo. Non è chiaro se Markov stesso si fosse reso conto dell'enorme importanza pratica di queste nuove idee; l'unico esempio di applicazione riportato nei suoi lavori è, infatti, l'analisi dell'alternarsi di vocali e consonanti in un testo letterario. Nel 1915 partecipò all'elaborazione della riforma dell'insegnamento della matematica nelle scuole medie e nel 1917 prese parte ad una missione scientifica di un anno all'interno della Russia. Al suo ritorno a Pietrogrado, però, la sua salute prese a declinare finché una setticemia lo portò alla tomba nell'estate del 1922.

**III.3.2 Osservazione:** La proprietà di Markov espressa nella Definizione precedente può anche essere data in un certo numero di modi equivalenti. Così ad esempio si può mostrare che  $(X_n, \mathcal{F}_n)_{n \geq 0}$  è una Catena di Markov se e solo se, comunque scelta una funzione di Borel  $g(x)$  limitata risulta

$$\mathbf{E}(g(X_n) \mid \mathcal{F}_m) = \mathbf{E}(g(X_n) \mid X_m), \quad \mathbf{P}\text{-q.o.}, \quad 0 \leq m < n.$$

Se poi  $(\mathcal{F}_n)_{n \geq 0}$  è la filtrazione naturale di  $(X_n)_{n \geq 0}$  ed  $Y$  è una v.a.  $\mathcal{F}_n$ -misurabile, cioè se  $Y = g(X_0, \dots, X_n)$ , si prova anche che

$$\mathbf{E}(Y \mid \mathcal{F}_m) = \mathbf{E}(Y \mid X_m), \quad \mathbf{P}\text{-q.o.}, \quad 0 \leq m \leq n.$$

Tale affermazione si prova innanzitutto in maniera banale per le funzioni di Borel della forma

$$g(x_0, \dots, x_n) = I_{A_0 \times \dots \times A_n}(x_0, \dots, x_n), \quad A_0, \dots, A_n \in \mathcal{B}(\mathbf{R}),$$

e si estende poi<sup>2</sup>, prima alle funzioni del tipo  $g(x_0, \dots, x_n) = I_B(x_0, \dots, x_n)$  con  $B \in \mathcal{B}(\mathbf{R}^n)$  e quindi a tutte le altre. Un altro modo equivalente di esprimere la proprietà di Markov consiste nell'affermazione secondo la quale il *passato* ed il *futuro* della nostra catena sono (condizionalmente) indipendenti se è dato il *presente*. In altre parole: se  $m$  è l'istante presente,  $A \in \mathcal{F}_{m-1}$  è un arbitrario evento del passato e  $B \in \sigma\{X_k : k \geq m+1\}$  è un arbitrario evento del futuro, si può mostrare che la proprietà di Markov equivale a richiedere che

$$\mathbf{P}(AB \mid X_m) = \mathbf{P}(A \mid X_m)\mathbf{P}(B \mid X_m), \quad \mathbf{P}\text{-q.o.}$$

Non daremo, per brevità, una dimostrazione generale di questa affermazione e ci limiteremo piuttosto ad illustrarla con qualche esempio. ○

**III.3.3 Osservazione:** Mostreremo ora con quali strumenti analitici è possibile costruire una Catena di Markov finita. Supponiamo che  $E = \{a_1, \dots, a_N\}$  sia un insieme finito (che giocherà il ruolo di insieme degli stati per la nostra catena) e che  $(E, \mathcal{E})$  sia il corrispondente spazio delle fasi e consideriamo le seguenti funzioni definite rispettivamente su  $E$  e  $E^2$ :

$$p_0(x) \geq 0, \quad p_n(x, y) \geq 0; \quad n \geq 1, \quad x, y \in E$$

tali che risulti

$$\sum_{x \in E} p_0(x) = 1, \quad \sum_{y \in E} p_n(x, y) = 1; \quad n \geq 1, \quad x \in E.$$

---

<sup>2</sup> Si tratta di una dimostrazione che fa appello ai risultati sulle classi monotone contenuti nel Capitolo II.2. Un esempio di dimostrazioni eseguite con questa tecnica è stato fornito dalla prova del Teorema II.5.41.

Ovviamente, essendo  $E$  finito,  $p_0(x)$  costituisce un vettore con  $N$  componenti e  $p_n(x, y)$  una matrice  $N \times N$  tale che la somma degli elementi di ciascuna riga<sup>3</sup> sia sempre eguale ad 1: matrici di questo genere prendono anche il nome di **matrici stocastiche**. È ora sempre possibile determinare<sup>4</sup> uno spazio  $(\Omega, \mathcal{F}, \mathbf{P})$  ed una successione di v.a.  $(X_n)_{n \geq 0}$  su di esso a valori in  $(E, \mathcal{E})$  in modo tale che la probabilità  $\mathbf{P}$  risulti definita a partire da  $p_0$  e  $p_n$  secondo le seguenti prescrizioni (con  $n \geq 1$ ):

$$\begin{aligned} \mathbf{P}(X_0 = x_0) &= p_0(x_0), \\ \mathbf{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) &= p_0(x_0)p_1(x_0, x_1) \dots p_n(x_{n-1}, x_n). \end{aligned}$$

Si controlla facilmente, infatti, che tale prescrizione soddisfa le ipotesi di consistenza previste dal Teorema di Kolmogorov II.4.10 e quindi che essa garantisce l'esistenza di una estensione a tutto  $(\Omega, \mathcal{F})$ .

Esamineremo ora alcune conseguenze della nostra definizione. Innanzitutto determiniamo la DdP  $p_n(x) = \mathbf{P}(X_n = x)$  di ciascuna delle  $X_n$  in termini delle funzioni  $p_0(x)$  e  $p_n(x, y)$ : siccome l'evento  $\{X_n = x\}$  risulta unione (finita) di eventi disgiunti secondo la relazione

$$\{X_n = x\} = \bigcup_{x_0 \in E} \dots \bigcup_{x_{n-1} \in E} \{X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x\}$$

si ha facilmente che

$$\begin{aligned} p_n(x) &= \mathbf{P}(X_n = x) = \sum_{x_0 \in E} \dots \sum_{x_{n-1} \in E} \mathbf{P}(X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x) \\ &= \sum_{x_0 \in E} \dots \sum_{x_{n-1} \in E} p_0(x_0)p_1(x_0, x_1) \dots p_{n-1}(x_{n-2}, x_{n-1})p_n(x_{n-1}, x). \end{aligned}$$

Allo stesso modo si possono poi calcolare anche le DdP congiunte di ogni altra combinazione finita delle v.a.  $X_n$ . Le nostre  $X_n$  sono quindi, in generale, non identicamente distribuite e, come vedremo subito, non indipendenti. Infatti, se si escludono i punti  $x$  tali che  $\mathbf{P}(X_{n-1} = x) = 0$  (che costituiscono, ovviamente, un insieme di punti di misura nulla in modo che le nostre conclusioni saranno

---

<sup>3</sup> Nota che il ruolo delle righe e delle colonne in questo tipo di matrici non è, in generale, simmetrico come sarà reso chiaro in seguito dalla loro interpretazione in termini di probabilità di transizione. Per questo motivo non vi è, in generale, alcun vincolo sulla somma degli elementi delle colonne di  $p_n(x, y)$  analogo a quello qui enunciato sulla somma degli elementi delle righe.

<sup>4</sup> In particolare lo spazio probabilizzabile opportuno  $(\Omega, \mathcal{F})$  potrà sempre essere costruito come spazio prodotto con  $\Omega = E^\infty$  con la corrispondente  $\sigma$ -algebra. In questo caso le v.a.  $X_k$  associeranno ad ogni elemento di  $E^\infty$  (cioè ad ogni successione  $x = (x_n)_{n \in \mathbf{N}}$  di elementi di  $E$ ) il numero  $x_k$  corrispondente alla componente  $k$ -ma di  $x$ . La misura di probabilità su tale  $(\Omega, \mathcal{F})$  viene poi costruita secondo le indicazioni del testo.

comunque vere **P**-q.o.), abbiamo dalla definizione elementare di probabilità condizionata che

$$\begin{aligned} \mathbf{P}(X_n = y | X_{n-1} = x) &= \frac{\mathbf{P}(X_{n-1} = x, X_n = y)}{\mathbf{P}(X_{n-1} = x)} \\ &= \frac{\sum_{x_0 \in E} \cdots \sum_{x_{n-2} \in E} p_0(x_0) \cdots p_{n-1}(x_{n-2}, x) p_n(x, y)}{\sum_{x_0 \in E} \cdots \sum_{x_{n-2} \in E} p_0(x_0) \cdots p_{n-1}(x_{n-2}, x)} \\ &= p_n(x, y), \end{aligned}$$

relazione che mette in evidenza il ruolo di probabilità condizionate assegnato alle funzioni  $p_n(x, y)$  in questo modello. A questo punto si può mostrare che, se consideriamo la filtrazione naturale  $(\mathcal{F}_n)_{n \geq 0}$  determinata dalla successione  $(X_n)_{n \geq 0}$ , la successione stocastica  $X = (X_n, \mathcal{F}_n)_{n \geq 0}$  costituisce una Catena di Markov. Infatti (sempre **P**-q.o.) si ha

$$\begin{aligned} &\mathbf{P}(X_n = y | X_{n-1} = x, X_{n-2} = x_{n-2}, \dots, X_0 = x_0) \\ &= \frac{\mathbf{P}(X_n = y, X_{n-1} = x, X_{n-2} = x_{n-2}, \dots, X_0 = x_0)}{\mathbf{P}(X_{n-1} = x, X_{n-2} = x_{n-2}, \dots, X_0 = x_0)} \\ &= \frac{p_0(x_0) \cdots p_{n-1}(x_{n-2}, x) p_n(x, y)}{p_0(x_0) \cdots p_{n-1}(x_{n-2}, x)} \\ &= p_n(x, y) = \mathbf{P}(X_n = y | X_{n-1} = x), \end{aligned}$$

e pertanto

$$\mathbf{P}(X_n = y | \mathcal{F}_{n-1}) = \mathbf{P}(X_n = y | X_{n-1}, \dots, X_0) = \mathbf{P}(X_n = y | X_{n-1}), \quad \mathbf{P}\text{-q.o.}$$

Questa relazione può facilmente essere estesa dalla coppia di istanti  $n-1, n$  ad una generica coppia di istanti  $n, m$  (con  $m > n$ ) e mostra quindi che  $(X_n, \mathcal{F}_n)_{n \geq 0}$  è realmente una Catena di Markov. In particolare è possibile verificare su questo modello che, come affermato nell'Osservazione III.3.2, la proprietà di Markov equivale all'indipendenza del passato dal futuro, condizionatamente rispetto ad un presente noto. Per far questo mostriamo innanzitutto che, comunque scelto  $k \geq 1$  si ha **P**-q.o.

$$\begin{aligned} &\mathbf{P}(X_{n+k} = x_{n+k}, \dots, X_{n+1} = x_{n+1} | \mathcal{F}_n) \\ &= \mathbf{P}(X_{n+k} = x_{n+k}, \dots, X_{n+1} = x_{n+1} | X_n). \end{aligned}$$

Infatti, posto per brevità  $A_n = \{X_n = x_n\}$ , e usando più volte la proprietà di Markov si ha

$$\begin{aligned} \mathbf{P}(A_{n+k} \cdots A_{n+1} | A_n \cdots A_0) &= \frac{\mathbf{P}(A_{n+k} \cdots A_0)}{\mathbf{P}(A_n \cdots A_0)} \\ &= \frac{\mathbf{P}(A_{n+k} \cdots A_0)}{\mathbf{P}(A_{n+1} \cdots A_0)} \frac{\mathbf{P}(A_{n+1} \cdots A_0)}{\mathbf{P}(A_n \cdots A_0)} \end{aligned}$$

$$\begin{aligned}
 &= \mathbf{P}(A_{n+k} \dots A_{n+2} | a_{n+1} \dots A_0) \mathbf{P}(A_{n+1} | A_n \dots A_0) \\
 &= \mathbf{P}(A_{n+k} \dots A_{n+2} | A_{n+1} \dots A_0) \mathbf{P}(A_{n+1} | A_n) \\
 &\dots \\
 &= \mathbf{P}(A_{n+k} | A_{n+k-1}) \dots \mathbf{P}(A_{n+2} | A_{n+1}) \mathbf{P}(A_{n+1} | A_n) \\
 &= \mathbf{P}(A_{n+k} | A_{n+k-1} \dots A_n) \dots \mathbf{P}(A_{n+2} | A_{n+1} A_n) \mathbf{P}(A_{n+1} | A_n) \\
 &= \frac{\mathbf{P}(A_{n+k} \dots A_n)}{\mathbf{P}(A_{n+k-1} \dots A_n)} \dots \frac{\mathbf{P}(A_{n+2} A_{n+1} A_n)}{\mathbf{P}(A_{n+1} A_n)} \frac{\mathbf{P}(A_{n+1} A_n)}{\mathbf{P}(A_n)} \\
 &= \frac{\mathbf{P}(A_{n+k} \dots A_n)}{\mathbf{P}(A_n)} = \mathbf{P}(A_{n+k} \dots A_{n+1} | A_n),
 \end{aligned}$$

da cui discende la relazione richiesta. Se ora  $N = \{X_n = x_n\}$  è un evento del *presente*,  $P = \{X_0 = x_0, \dots, X_{n-1} = x_{n-1}\}$  uno del *passato* e  $F = \{X_{n+1} = x_{n+1}, \dots, X_{n+k} = x_{n+k}\}$  uno del futuro della nostra catena, la relazione appena dimostrata ci dice che  $\mathbf{P}(F | NP) = \mathbf{P}(F | N)$  e quindi si ha che

$$\begin{aligned}
 \mathbf{P}(FP | N) &= \frac{\mathbf{P}(FPN)}{\mathbf{P}(N)} = \frac{\mathbf{P}(FPN)}{\mathbf{P}(PN)} \frac{\mathbf{P}(PN)}{\mathbf{P}(N)} = \mathbf{P}(F | PN) \mathbf{P}(P | N) \\
 &= \mathbf{P}(F | N) \mathbf{P}(P | N),
 \end{aligned}$$

cioè che  $P$  e  $F$  sono indipendenti se  $N$  è dato.

Abbiamo quindi mostrato che è sempre possibile costruire una Catena di Markov basandosi sulle funzioni  $p_0(x)$  e  $p_n(x, y)$  (con  $n \geq 1$  e  $x, y \in E$ ) ed è d'altra parte possibile mostrare che per ogni data Catena di Markov finita le funzioni  $p_0(x) = \mathbf{P}(X_0 = x)$  e  $p_n(x, y) = \mathbf{P}(X_n = y | X_{n-1} = x)$  soddisfano tutte le relazioni esaminate nella presente discussione. In pratica potremo dire che tutte le Catene di Markov finite sono ben definite una volta che siano note la funzione  $p_0(x)$ , che prende il nome di **distribuzione iniziale** della catena, e la matrice stocastica  $p_n(x, y)$  che prende il nome di matrice della **probabilità di transizione** della catena. Tutte le altre quantità relative alla nostra catena, come ad esempio le funzioni  $p_n(x)$  che prendono il nome di **distribuzioni all'istante  $n$ -mo**, saranno poi calcolabili a partire dalle  $p_0(x)$  e  $p_n(x, y)$ . In sostanza potremo dire che una Catena di Markov è definita se e solo se sono assegnate la sua distribuzione iniziale e la famiglia delle sue probabilità di transizione ad ogni istante. È intuitivo, a questo punto, che a causa della grande semplificazione che ciò comporta assumono rilievo particolare le Catene di Markov finite per le quali le matrici stocastiche  $p_n(x, y)$  non dipendono da  $n$ , cioè le catene le cui probabilità di transizione sono indipendenti dall'istante  $n$  preso in considerazione. Tali catene si dicono **catene omogenee** ed hanno la caratteristica di essere ben definite una volta che siano assegnate la distribuzione iniziale  $p_0(x)$  ed un'unica matrice stocastica  $p(x, y)$  che giocherà il ruolo di matrice delle probabilità di transizione in ogni istante di tempo. In tutto il resto di questo capitolo, e salvo avviso contrario, avremo praticamente a che fare solo con Catene di Markov finite ed omogenee. Per semplificare la notazione, inoltre, converrà osservare che siccome le  $X_n$  prendono valori in un insieme finito  $E$ , sarà comodo d'ora in poi supporre che  $E$  sia proprio l'insieme

$\{1, 2, \dots, N\}$  o, secondo i casi, l'insieme  $\{-N, \dots, -1, 0, 1, \dots, N\}$ . In pratica sostituiamo all'insieme  $E$  l'insieme degli indici dei suoi elementi, sicché  $p_0(x)$  e  $p(x, y)$  saranno caratterizzati da variabili  $x, y$  che sono dei numeri interi  $i, j$  per cui converrà usare piuttosto le notazioni  $p_i^{(0)} = p_0(i)$  e  $p_{ij} = p(i, j)$ . Pertanto nel seguito<sup>5</sup> indicheremo con  $\pi^{(0)} = (p_i^{(0)})_{i \in E}$  il vettore DdP iniziale ( $p_i^{(0)}$  essendo la probabilità che, inizialmente, la catena si trovi nello stato  $i$ -mo), e con  $\Pi = (p_{ij})_{i, j \in E}$  la matrice delle probabilità di transizione ( $p_{ij}$  essendo la probabilità di transizione dallo stato  $i$ -mo allo stato  $j$ -mo in ogni istante di tempo). La nostra discussione precedente mostra allora che una Catena di Markov finita ed omogenea sarà completamente specificata una volta che sia definito lo spazio delle fasi e una volta che sia assegnata la coppia  $(\pi, \Pi)$ . Termineremo queste prime osservazioni ricordando che una rappresentazione particolarmente suggestiva delle Catene di Markov finite ed omogenee può essere data mediante dei **grafici** sui quali si rappresentano gli stati di  $E$  mediante dei punti sul piano e le transizioni mediante frecce dotate di un peso: la freccia dallo stato  $i$ -mo allo stato  $j$ -mo porterà il peso  $p_{ij}$  per indicare che la transizione considerata avviene con probabilità  $p_{ij}$  (vedi Figura III.3.1) Le frecce omesse, naturalmente, indicano che la transizione è vietata (ossia avviene con probabilità 0).  $\circ$

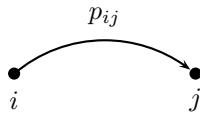


Fig. III.3.1 Rappresentazione grafica di transizioni markoviane.

**III.3.4 Esempio:** Un semplicissimo esempio di Catena di Markov può essere dato supponendo che  $E = \{0, 1, 2\}$  e che la matrice di transizione sia la matrice stocastica

$$\Pi = \|p_{ij}\| = \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 0 & 1/2 \\ 2/3 & 0 & 1/3 \end{pmatrix}$$

Il grafico corrispondente è riportato in Fig. III.3.2. Da questo (o dalla semplice osservazione degli elementi della prima riga della matrice di transizione) si può facilmente notare che lo stato 0 è caratterizzato dal fatto che una volta arrivata in esso la catena vi permane indefinitamente. Stati di questo genere vengono anche detti **assorbenti**. Semplici casi di evoluzione della catena possono essere elaborati dal lettore.

<sup>5</sup> Adotteremo inoltre la convenzione di considerare sempre i vettori come *righe* e di eseguire i prodotti *righe per colonne*.

Un altro esempio può essere costruito su uno spazio delle fasi finito del tipo  $E = \{-N, \dots, -1, 0, 1, \dots, N\}$  assegnando come distribuzione iniziale  $p_i^{(0)} = \delta_{0,i}$  (la catena, cioè, parte dallo stato 0 con probabilità 1) e come matrice delle transizioni  $(2N + 1) \times (2N + 1)$   $\|p_{ij}\|$  (con  $i, j$  che prendono valori tra  $-N$  ed  $N$ )

$$p_{ij} = \begin{cases} \delta_{N,j}, & \text{se } i = N; \\ p\delta_{i+1,j} + q\delta_{i-1,j}, & \text{se } |i| < N; \\ \delta_{-N,j}, & \text{se } i = -N. \end{cases}$$

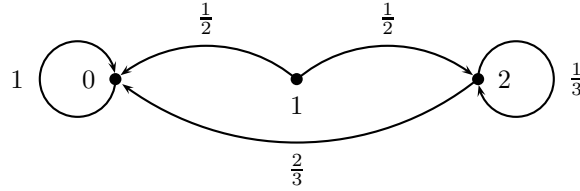


Fig. III.3.2 Esempio di catena di Markov a tre stati.

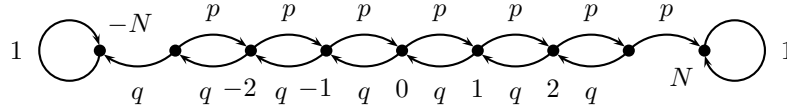


Fig. III.3.3 Esempio di catena di Markov a  $2N + 1$  stati.

Esplicitamente la matrice delle transizioni assume la forma

$$\Pi = \|p_{ij}\| = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ q & 0 & p & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & q & 0 & p & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & q & 0 & \dots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 & p & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & q & 0 & p & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & q & 0 & p \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 1 \end{pmatrix}$$

e il corrispondente grafico è riportato in Fig. III.3.3 per  $N = 3$ . Anche in questo caso si vede che gli stati  $\pm N$  sono assorbenti. Una semplice osservazione mostra che questa catena può essere adoperata per descrivere il gioco d'azzardo discusso nel Capitolo I.10 nel caso in cui ambedue i giocatori partono con lo stesso capitale iniziale  $N$  e giocano un'unità per volta: l'assorbitività degli stati estremi

corrisponde allora alla rovina di uno dei due giocatori. A questo scopo il modello sarà costruito nel modo seguente: detta  $(\xi_n)_{n \in \mathbf{N}}$  la successione di v.a. indipendenti di Bernoulli con

$$\mathbf{P}(\xi_n = +1) = p, \quad \mathbf{P}(\xi_n = -1) = q, \quad n \geq 1, \quad (p + q = 1)$$

che descrive la successione delle vincite di uno dei giocatori ad ogni mano, la successione delle vincite totali sarà

$$X_0 = 0; \\ X_n = \begin{cases} \sum_{k=1}^n \xi_k, & \text{se } |X_j| < N, \quad j = 1, \dots, n-1; \\ \pm N, & \text{altrimenti;} \end{cases}$$

si controlla facilmente allora che  $(X_n)_{n \geq 0}$  è una Catena di Markov dato che  $X_{n+1} = X_n + \xi_{n+1}$  sicché  $X_{n+1}$  dipende dal suo passato solo tramite la v.a. immediatamente precedente  $X_n$ :

$$\begin{aligned} \mathbf{P}(X_{n+1} = j | X_n = i, X_{n-1} = x_{n-1} \dots, X_0 = x_0) &= \mathbf{P}(X_n + \xi_{n+1} = j | X_n = i) \\ &= \mathbf{P}(X_{n+1} = j | X_n = i); \end{aligned}$$

inoltre si verifica facilmente che le probabilità di transizione

$$p_{ij} = \mathbf{P}(X_{n+1} = j | X_n = i) = \mathbf{P}(\xi_{n+1} = j - i)$$

sono indipendenti da  $n$  (la catena è omogenea dato che le  $\xi_n$  sono identicamente distribuite) ed hanno esattamente la forma precedentemente indicata.  $\diamond$

**III.3.5 Osservazione:** Abbiamo notato che la markovianità dell'esempio precedente era dovuta al fatto che la nostra successione era della forma  $X_{n+1} = X_n + \xi_{n+1}$  con  $(\xi_n)_{n \in \mathbf{N}}$  successione di v.a. indipendenti. Si verifica facilmente che questo non è che un caso particolare di un comportamento più generale: data una successione  $(\xi_n)_{n \in \mathbf{N}}$  di v.a. indipendenti ed una successione di funzioni di Borel  $f_n : \mathbf{R}^2 \rightarrow \mathbf{R}$  a due variabili, si può sempre costruire una Catena di Markov tramite le **relazioni di ricorrenza**

$$X_{n+1} = f_n(X_n, \xi_{n+1}), \quad n \geq 0,$$

dove  $X_0$  è un'arbitraria v.a. che descrive la situazione iniziale della catena. Il risultato discende direttamente dal fatto che, per sua definizione,  $X_{n+1}$  dipende dal suo passato solo tramite l'ultima v.a.  $X_n$ . In generale le catene ottenute in questo modo sono non omogenee: nell'Esempio III.3.4 la catena risultante era invece omogenea a causa della semplicità delle funzioni date ( $f_n(x, y) = x + y$  per  $n \geq 1$ ) e del fatto che le  $\xi_n$  erano identicamente distribuite oltre che indipendenti. Val la pena di osservare che questo modo di costruire Catene di Markov è una



generalizzazione aleatoria dei metodi comunemente usati per generare successioni deterministiche di stati tramite relazioni di ricorrenza<sup>6</sup>. Ad esempio per movimenti unidimensionali, data una successione di funzioni  $f_n : \mathbf{R} \rightarrow \mathbf{R}$  e uno stato iniziale arbitrario  $x_0 \in \mathbf{R}$ , la traiettoria (deterministica) può essere definita dalle relazioni di ricorrenza  $x_{n+1} = f_n(x_n)$  con  $n \geq 0$ . ○

**III.3.6 Esempio:** Useremo ora le indicazioni fornite nell'Osservazione III.3.5 per costruire un altro esempio di Catena di Markov: supponiamo che ad una stazione di taxi arrivi una vettura alla volta ad intervalli di tempo unitari e che essa riparta subito portando via un passeggero (se c'è) o vuota (se nessuno è in coda). Supponiamo inoltre che all'istante  $n$ -mo arrivino  $\xi_n$  passeggeri e che le  $\xi_n$  siano tutte v.a. indipendenti. Detta allora  $X_n$  la v.a. che rappresenta la lunghezza della coda dei passeggeri in attesa all'istante  $n$ -mo, e supponendo che  $X_0 = 0$ , avremo che

$$X_{n+1} = (X_n - 1)^+ + \xi_{n+1} = \begin{cases} \xi_{n+1}, & \text{se } X_n = 0; \\ X_n - 1 + \xi_{n+1}, & \text{se } X_n \geq 1; \end{cases}$$

(dove come al solito si pone  $x^+ = \max\{x, 0\}$ ). Ne segue che  $(X_n)_{n \geq 0}$  è una Catena di Markov in quanto definita tramite le relazioni di ricorrenza  $X_{n+1} = f_n(X_n, \xi_{n+1})$  con  $f_n(x, y) = (x - 1)^+ + y$  e con le  $\xi_n$  v.a. indipendenti. ◇

**III.3.8 Osservazione:** Le nozioni introdotte ci consentono ora di studiare le **equazioni di evoluzione** di una Catena di Markov. Supponiamo che la nostra catena omogenea sia caratterizzata dalle distribuzioni  $(\pi, \Pi)$  delle quali sappiamo che  $p_{ij}$  rappresenta la probabilità di trovarsi in  $j$  in un dato istante (arbitrario a causa dell'omogeneità) se ci si trovava in  $i$  all'istante immediatamente precedente. In un certo senso potremo dunque dire che le  $p_{ij}$  rappresentano le **probabilità di transizione ad un passo** sicché converrà introdurre anche la nozione di **probabilità di transizione a  $m$  passi** così definita

$$p_{ij}^{(m)} = \mathbf{P}(X_{n+m} = j \mid X_n = i), \quad n \geq 0$$

che rappresenta la probabilità di trovarsi in  $j$  in un dato istante se  $m$  istanti prima ci si trovava in  $i$ . L'omogeneità della nostra Catena di Markov implica che le  $p_{ij}^{(m)}$  sono costanti al variare di  $n$ , cioè non dipendono dall'istante  $n$  nel quale inizia la transizione. Indicheremo inoltre con  $\Pi^{(m)}$  la corrispondente **matrice delle transizioni a  $m$  passi**, e con  $\pi^{(n)}$  il vettore con elementi

$$p_i^{(n)} = \mathbf{P}(X_n = i)$$

---

<sup>6</sup> Nonostante la semplicità di definizione di tali *sistemi dinamici* il loro comportamento può essere estremamente complesso ed occupa oggi un importante settore della ricerca in fisica teorica e matematica. Un'utile introduzione a questa affascinante problematica può essere trovata in **R.M. May**: *Simple mathematical models with very complicated dynamics*; Nature 261 (1976) 459.

che rappresenta la DdP della catena all'istante  $n$ -mo. Ovviamente avremo anche che

$$\begin{aligned} p_{ij}^{(1)} &= p_{ij}, & \Pi^{(1)} &= \Pi, \\ p_i^{(0)} &= p_i, & \pi^{(0)} &= \pi, \end{aligned}$$

essendo  $(\pi, \Pi)$  le quantità che definiscono la nostra catena.

Mostreremo ora che le matrici  $\Pi^{(m)}$  soddisfano le seguenti relazioni note come **equazioni di Chapman-Kolmogorov**<sup>7</sup>:

$$\Pi^{(l+m)} = \Pi^{(l)}\Pi^{(m)}, \quad l, m \geq 1,$$

ossia, in termini espliciti,

$$p_{ij}^{(l+m)} = \sum_{k \in E} p_{ik}^{(l)} p_{kj}^{(m)}, \quad l, m \geq 1, \quad i, j \in E.$$

Infatti dalle proprietà delle probabilità condizionate e delle Catene di Markov si ha

$$\begin{aligned} p_{ij}^{(l+m)} &= \mathbf{P}(X_{l+m} = j \mid X_0 = i) = \sum_{k \in E} \mathbf{P}(X_{l+m} = j, X_l = k \mid X_0 = i) \\ &= \sum_{k \in E} \mathbf{P}(X_{l+m} = j \mid X_l = k, X_0 = i) \mathbf{P}(X_l = k \mid X_0 = i) \\ &= \sum_{k \in E} \mathbf{P}(X_l = k \mid X_0 = i) \mathbf{P}(X_{l+m} = j \mid X_l = k) = \sum_{k \in E} p_{ik}^{(l)} p_{kj}^{(m)}. \end{aligned}$$

Le equazioni di Chapman-Kolmogorov esprimono il fatto (del tutto intuitivo se si pensa ad  $X_n$  come ad una traiettoria nel senso usuale del termine) che, nell'andare dallo stato  $i$  allo stato  $j$  in  $l+m$  passi, in un dato istante intermedio  $l$  la catena si troverà, con una certa probabilità, in uno degli altri stati. Due casi particolarmente importanti sono le seguenti due relazioni

$$\Pi^{(m+1)} = \Pi^{(m)}\Pi, \quad \Pi^{(m+1)} = \Pi\Pi^{(m)},$$

dette rispettivamente **equazione in avanti** ed **equazione all'indietro**. In modo del tutto analogo si dimostra che poi per le DdP ad ogni istante  $\pi^{(m)}$  valgono le analoghe relazioni

$$\pi^{(l+m)} = \pi^{(l)}\Pi^{(m)}$$

---

<sup>7</sup> Il fisico e matematico inglese Sydney Chapman (1888-1970) è noto soprattutto per le sue ricerche sulla teoria dei gas sul geomagnetismo e sulla fisica dell'atmosfera. Nel periodo tra il 1912 ed il 1917 egli generalizzò la teoria dei gas di J.C. Maxwell (1867) studiando teoricamente e sperimentalmente la loro diffusione termica. Tali studi si dimostrarono poi preziosi per le sue successive ricerche di geomagnetismo e fisica dell'atmosfera.

ossia in termini di componenti

$$p_j^{(l+m)} = \sum_{k \in E} p_k^{(l)} p_{kj}^{(m)}, \quad l \geq 0, m \geq 1, \quad j \in E,$$

dette anche loro rispettivamente **equazione in avanti** ed **equazione all'indietro** per le DdP. In particolare avremo anche

$$\pi^{(m+1)} = \pi^{(m)} \Pi = \pi^{(1)} \Pi^{(m)} = \pi \Pi^{(m)}.$$

Da successive applicazioni di queste equazioni di evoluzione segue poi che  $\Pi^{(m)} = \Pi^m$  e quindi

$$\pi^{(m)} = \pi \Pi^{(m-1)} = \pi \Pi^m, \quad m \geq 1,$$

cioè:  $\Pi^{(m)}$  coincide con la potenza  $m$ -ma della matrice  $\Pi$  e la DdP all'istante  $m$ -mo si ottiene per applicazione ripetuta della matrice  $\Pi$ , mettendo ancora una volta in evidenza il fatto che la conoscenza della matrice  $\Pi$  determina completamente<sup>8</sup> l'evoluzione temporale della catena una volta che sia data la distribuzione iniziale  $\pi$ . Sulla base di tutte queste osservazioni diremo ora che  $\bar{\pi} = (\bar{p}_i)_{i \in E}$  è una **distribuzione stazionaria** se

$$\bar{\pi} = \bar{\pi} \Pi$$

ossia se

$$\sum_{k \in E} \bar{p}_k p_{kj} = \bar{p}_j, \quad j \in E.$$

In tal caso infatti l'applicazione, anche ripetuta di  $\Pi$  a  $\bar{\pi}$  non modifica la distribuzione e si avrà ovviamente

$$\bar{\pi}^{(m)} = \bar{\pi}, \quad m \geq 1.$$

Si verifica infine in tal caso che la Catena di Markov omogenea  $(X_n)_{n \geq 0}$  costruita con  $(\bar{\pi}, \Pi)$  risulta anche essere una **Catena di Markov stazionaria** nel senso che tutte le distribuzioni congiunte di famiglie finite di suoi elementi del tipo  $(X_n, X_{n+1}, \dots, X_{n+m})$  (con  $n, m \geq 0$ ) non dipendono dall'istante iniziale  $n$  e coincidono tutte con la distribuzione congiunta di  $(X_0, X_1, \dots, X_m)$ .  $\circ$

**III.3.9 Osservazione:** Prima di passare a discutere alcune importanti applicazioni delle Catene di Markov finite vogliamo mostrare che molti dei concetti fin qui esposti possono essere immediatamente estesi anche al caso di **Catene di Markov generali** con spazio delle fasi  $(E, \mathcal{E})$  non elementare né finito. Data una Catena di Markov  $(X_n, \mathcal{F}_n)_{n \geq 0}$  con spazio delle fasi  $(E, \mathcal{E})$ , siano  $\mathbf{P}_{X_{n+1}|X_n}(\omega; B)$  la versione regolare della DdP condizionata di  $X_{n+1}$  rispetto a  $X_n$  e  $\mathbf{P}_{X_{n+1}|X_n}(x; B)$

---

<sup>8</sup> Naturalmente questa affermazione va intesa nel senso che  $\Pi$  determina completamente (e deterministicamente) la DdP dello stato della catena all'istante  $m$ , non lo stato stesso: si tratta pur sempre di un'evoluzione di elementi aleatori.

la versione regolare della DdP condizionata di  $X_{n+1}$  rispetto a  $X_n = x$  (nel senso della Definizione II.7.19 e dell'Osservazione II.7.20)<sup>9</sup>. Seguendo poi le indicazioni dell'Osservazione II.7.23 useremo per brevità le notazioni  $\mathbf{P}_n(B | X_n)$  e  $\mathbf{P}_n(B | x)$  rispettivamente al posto di  $\mathbf{P}_{X_{n+1}|X_n}(\omega; B)$  e di  $\mathbf{P}_{X_{n+1}|X_n}(x; B)$  dovunque ciò non dia adito ad equivoci. Ricorderemo inoltre che le v.a.  $\mathbf{P}_n(B | X_n)$  possono essere ottenute calcolando le funzioni (di  $x$ )  $\mathbf{P}_n(B | x)$  nelle v.a.  $X_n$ . Le  $\mathbf{P}_n(B | x)$  prendono anche il nome di **funzioni di transizione** della Catena di Markov e diremo che la catena è **omogenea** quando tali funzioni di transizione non dipendono da  $n$ , cioè quando risultano tutte eguali al variare di  $n$ . In tal caso sarà conveniente usare semplicemente il simbolo  $\mathbf{P}(B | x)$  (o sinteticamente  $\Pi$ ) per indicare la funzione di transizione della catena. Se indichiamo poi con  $\pi(B)$  (o sinteticamente  $\pi$ ) la **distribuzione iniziale**  $\mathbf{P}(X_0 \in B)$  della catena, potremo dire, analogamente al caso finito, che una Catena di Markov è data quando sono assegnate la distribuzione iniziale  $\pi(B)$  e la funzione di transizione  $\mathbf{P}(B | x)$ , ossia la coppia  $(\pi, \Pi)$ : infatti tutte le proprietà probabilistiche di  $(X_n, \mathcal{F}_n)_{n \geq 0}$  potranno essere dedotte dalla conoscenza di tale  $(\pi, \Pi)$ . Ad esempio, se per semplicità scegliamo  $(\mathbf{E}, \mathcal{E}) = (\mathbf{R}, \mathcal{B}(\mathbf{R}))$ , si può mostrare che tutte le proprietà probabilistiche di un numero finito di componenti della catena possono essere espresse mediante  $(\pi, \Pi)$ , nel senso che, dato un arbitrario  $A \in \mathcal{B}(\mathbf{R}^{n+1})$ , risulta

$$\begin{aligned} \mathbf{P}((X_0, \dots, X_n) \in A) \\ = \int_{\mathbf{R}} \pi(dx_0) \int_{\mathbf{R}} \mathbf{P}(dx_1 | x_0) \dots \int_{\mathbf{R}} \mathbf{P}(dx_n | x_{n-1}) I_A(x_0, \dots, x_n), \end{aligned}$$

e inoltre, comunque data una funzione di Borel  $g(x_0, \dots, x_n)$ , si ha

$$\mathbf{E}(g(X_0, \dots, X_n)) = \int_{\mathbf{R}} \pi(dx_0) \int_{\mathbf{R}} \mathbf{P}(dx_1 | x_0) \dots \int_{\mathbf{R}} \mathbf{P}(dx_n | x_{n-1}) g(x_0, \dots, x_n).$$

Infatti posto  $B = \{(X_0, \dots, X_n) \in A\} \in \mathcal{F}$ , si ha  $\mathbf{P}((X_0, \dots, X_n) \in A) = \mathbf{E} I_B$  e con un uso ripetuto delle proprietà dei VdA condizionati e delle Catene di Markov risulta<sup>10</sup>

$$\begin{aligned} \mathbf{E} I_B &= \mathbf{E} [\mathbf{E}(I_B | \mathcal{F}_0)] = \mathbf{E} [\mathbf{E}(\mathbf{E}(I_B | \mathcal{F}_1) | \mathcal{F}_0)] = \dots \\ &= \mathbf{E} \left[ \mathbf{E} \left( \mathbf{E} \left( \dots \mathbf{E}(\mathbf{E}(I_B | \mathcal{F}_{n-1}) | \mathcal{F}_{n-2}) \dots | \mathcal{F}_1 \right) \middle| \mathcal{F}_0 \right) \right] \\ &= \mathbf{E} \left[ \mathbf{E} \left( \mathbf{E} \left( \dots \mathbf{E}(\mathbf{E}(I_B | X_{n-1}) | X_{n-2}) \dots | X_1 \right) \middle| X_0 \right) \right]. \end{aligned}$$

<sup>9</sup> Ricordiamo qui per non doverlo poi ripetere ogni volta che, per come sono definite le versioni regolari, tutte le affermazioni che seguono sono vere q.o. nella misura di probabilità appropriata all'espressione considerata.

<sup>10</sup> Nota che la procedura si arresta all' $n$ -ma iterazione perché  $\mathbf{E}(I_B | \mathcal{F}_n) = \mathbf{E}(I_B | \mathcal{F}_{n+1}) = \dots = I_B$  dato che  $I_B$  è per costruzione misurabile rispetto a tutte le  $\sigma$ -algebre della filtrazione dall'indice  $n$  in poi.

La relazione richiesta segue allora dall'applicazione dei risultati della Proposizione II.7.22 e dell'Osservazione II.7.23 e dal fatto che  $I_B = I_A(X_0, \dots, X_n)$ . La seconda relazione si ottiene poi dalla prima ricordando che ogni funzione di Borel  $g(x_0, \dots, x_n)$  può essere approssimata con una successione di funzioni semplici. È anche possibile mostrare (ma ne tralascieremo la verifica per brevità) che comunque assegnate  $(\pi, \Pi)$ , dove  $\pi$  è una probabilità su  $(E, \mathcal{E})$  e  $\Pi = (\mathbf{P}(B|x))_{x \in \mathbf{R}}$  una famiglia di probabilità su  $(E, \mathcal{E})$  tale che  $\mathbf{P}(B|\cdot)$  sia  $\mathcal{B}(\mathbf{R})$ -misurabile per ogni  $B \in \mathcal{E}$ , esiste sempre una Catena di Markov omogenea che abbia  $\pi$  come distribuzione iniziale e  $\Pi$  come famiglia di funzioni di transizione. Ciò mostra in definitiva che ogni Catena di Markov omogenea è caratterizzata da una coppia  $(\pi, \Pi)$  che, nel caso di catene finite, si riduce ad un vettore ad  $N$  componenti  $p_i^{(0)}$  e ad una matrice stocastica  $N \times N$   $p_{ij}$  secondo quanto mostrato in precedenza.

Restringendoci sempre al caso  $(E, \mathcal{E}) = (\mathbf{R}, \mathcal{B}(\mathbf{R}))$  mostreremo ora come si scrivono le **equazioni di Chapman-Kolmogorov** per Catene di Markov generali. Per far questo bisognerà introdurre, accanto alle probabilità e funzioni di transizione ad un passo

$$\begin{aligned} \mathbf{P}_n^{(1)}(B|X_n) &= \mathbf{P}_n(B|X_n) = \mathbf{P}(X_{n+1} \in B|X_n) \\ \mathbf{P}_n^{(1)}(B|x) &= \mathbf{P}_n(B|x) = \mathbf{P}(X_{n+1} \in B|X_n = x) \end{aligned}$$

**le probabilità e funzioni di transizione a  $m$  passi**

$$\begin{aligned} \mathbf{P}_n^{(m)}(B|X_n) &= \mathbf{P}(X_{n+m} \in B|X_n) \\ \mathbf{P}_n^{(m)}(B|x) &= \mathbf{P}(X_{n+m} \in B|X_n = x). \end{aligned}$$

Nel caso in cui la catena è omogenea tutte le probabilità e funzioni di transizione a  $m$  passi  $\mathbf{P}_n^{(m)}(B|X_n)$  e  $\mathbf{P}_n^{(m)}(B|x)$  coincidono al variare di  $n$  sicché sarà sufficiente assegnare una sola famiglia di funzioni di transizione a  $m$  passi  $\mathbf{P}^{(m)}(B|x)$ ; come al solito le corrispondenti probabilità di transizione  $\mathbf{P}^{(m)}(B|X_n)$  si otterranno calcolando le funzioni (di  $x$ )  $\mathbf{P}^{(m)}(B|x)$  nella v.a.  $X_n$ . Con queste notazioni possiamo ora mostrare che le equazioni di Chapman-Kolmogorov per Catene di Markov omogenee assumono la forma

$$\mathbf{P}^{(l+m)}(B|X_n) = \int_{\mathbf{R}} \mathbf{P}^{(m)}(B|y)\mathbf{P}^{(l)}(dy|X_n), \mathbf{P}\text{-q.o.}$$

Infatti posto  $C = \{X_{l+m+n} \in B\}$  abbiamo

$$\begin{aligned} \mathbf{P}^{(l+m)}(B|X_n) &= \mathbf{E}(I_C|X_n) = \mathbf{E}[\mathbf{E}(I_C|\mathcal{F}_{l+n})|X_n] \\ &= \mathbf{E}[\mathbf{E}(I_C|X_{l+n})|X_n] = \mathbf{E}[\mathbf{P}(X_{l+m+n} \in B|X_{l+n})|X_n] \\ &= \mathbf{E}[\mathbf{P}^{(m)}(B|X_{l+n})|X_n] = \int_{\mathbf{R}} \mathbf{P}^{(m)}(B|y)\mathbf{P}^{(l)}(dy|X_n). \end{aligned}$$

Le versioni regolari delle probabilità condizionate possono poi anche essere sempre scelte in modo che si possa scrivere

$$\mathbf{P}^{(l+m)}(B|x) = \int_{\mathbf{R}} \mathbf{P}^{(m)}(B|y)\mathbf{P}^{(l)}(dy|x).$$

Sarà utile ricordare infine che, nel caso in cui la distribuzione iniziale e le funzioni di transizione di una catena omogenea possano essere espresse mediante densità

$$\pi(B) = \int_B f(y) dy, \quad \mathbf{P}^{(m)}(B|x) = \int_B f^{(m)}(y|x) dy,$$

tutte le espressioni qui menzionate potranno essere riscritte in una forma analoga a quella corrispondente nel caso di catene finite. Così ad esempio

$$\begin{aligned} & \mathbf{P}((X_0, \dots, X_n) \in A) \\ &= \int_{\mathbf{R}} f(x_0) dx_0 \int_{\mathbf{R}} f(x_1|x_0) dx_1 \dots \int_{\mathbf{R}} f(x_n|x_{n-1}) dx_n I_A(x_0, \dots, x_n), \\ & \mathbf{E}(g(X_0, \dots, X_n)) \\ &= \int_{\mathbf{R}} f(x_0) dx_0 \int_{\mathbf{R}} f(x_1|x_0) dx_1 \dots \int_{\mathbf{R}} f(x_n|x_{n-1}) dx_n g(x_0, \dots, x_n); \end{aligned}$$

mentre per le equazioni di Chapman-Kolmogorov si ha

$$f^{(l+m)}(x'|x) = \int_{\mathbf{R}} f^{(m)}(x'|y) f^{(l)}(y|x) dy. \quad \circ$$

**III.3.10 Esempio:** Consideriamo una Catena di Markov omogenea a due stati (cioè con spazio delle fasi  $E = \{0, 1\}$ ) caratterizzata dalla seguente matrice di transizione

$$\Pi = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}, \quad \begin{cases} p_{00} + p_{01} = 1 \\ p_{10} + p_{11} = 1. \end{cases}$$

È facile mostrare con un calcolo diretto che, essendo  $\Pi$  una matrice stocastica, risulta

$$\begin{aligned} \Pi^{(2)} = \Pi^2 &= \begin{pmatrix} p_{00}^2 + p_{01}p_{10} & p_{01}(p_{00} + p_{11}) \\ p_{10}(p_{00} + p_{11}) & p_{11}^2 + p_{01}p_{10} \end{pmatrix} \\ &= \frac{1}{2 - p_{00} - p_{11}} \begin{pmatrix} 1 - p_{11} & 1 - p_{00} \\ 1 - p_{11} & 1 - p_{00} \end{pmatrix} \\ &\quad + \frac{(p_{00} + p_{11} - 1)^2}{2 - p_{00} - p_{11}} \begin{pmatrix} 1 - p_{00} & -(1 - p_{00}) \\ -(1 - p_{11}) & 1 - p_{11} \end{pmatrix} \end{aligned}$$

sicché per induzione si può mostrare che

$$\begin{aligned} \Pi^{(n)} = \Pi^n &= \frac{1}{2 - p_{00} - p_{11}} \begin{pmatrix} 1 - p_{11} & 1 - p_{00} \\ 1 - p_{11} & 1 - p_{00} \end{pmatrix} \\ &\quad + \frac{(p_{00} + p_{11} - 1)^n}{2 - p_{00} - p_{11}} \begin{pmatrix} 1 - p_{00} & -(1 - p_{00}) \\ -(1 - p_{11}) & 1 - p_{11} \end{pmatrix}. \end{aligned}$$

Escludendo per ora i casi degeneri<sup>11</sup> in cui  $p_{00}$  e  $p_{11}$  sono ambedue eguali a 0 oppure eguali ad 1, cioè supponendo che  $|p_{00} + p_{11} - 1| < 1$ , si vede subito che

$$\Pi^n \xrightarrow{n} \frac{1}{2 - p_{00} - p_{11}} \begin{pmatrix} 1 - p_{11} & 1 - p_{00} \\ 1 - p_{11} & 1 - p_{00} \end{pmatrix} = \begin{pmatrix} \bar{p}_0 & \bar{p}_1 \\ \bar{p}_0 & \bar{p}_1 \end{pmatrix},$$

dove abbiamo posto

$$\bar{p}_0 = \frac{1 - p_{11}}{2 - p_{00} - p_{11}}, \quad \bar{p}_1 = \frac{1 - p_{00}}{2 - p_{00} - p_{11}}, \quad \bar{p}_0 + \bar{p}_1 = 1.$$

In altre parole la matrice limite è composta di righe tutte eguali sicché

$$p_{i0}^{(n)} \xrightarrow{n} \bar{p}_0, \quad p_{i1}^{(n)} \xrightarrow{n} \bar{p}_1, \quad (i = 0, 1).$$

Praticamente le  $p_{ij}^{(n)}$  tendono a dei limiti  $\bar{p}_j$  che non dipendono più da  $i$  sicché potremo concludere che l'influenza dello stato iniziale  $i$  sulla probabilità di trovare la catena in uno dei due stati 0, 1 diventa trascurabile al crescere di  $n$ . Indicate allora rispettivamente con

$$\pi = (p_0, p_1), \quad \bar{\pi} = (\bar{p}_0, \bar{p}_1)$$

la DdP iniziale e la DdP limite (per  $n \rightarrow \infty$ ) ci accorgiamo che esse godono di alcune importanti proprietà:

- (a) se  $|p_{00} + p_{11} - 1| < 1$ , qualunque DdP iniziale  $\pi$  evolve verso la medesima DdP finale  $\bar{\pi}$  nel senso che

$$\pi \Pi^n = \pi^{(n)} \xrightarrow{n} (p_0, p_1) \begin{pmatrix} \bar{p}_0 & \bar{p}_1 \\ \bar{p}_0 & \bar{p}_1 \end{pmatrix} = (\bar{p}_0, \bar{p}_1) = \bar{\pi},$$

per cui  $\bar{\pi}$  rappresenta la DdP limite di qualunque DdP iniziale; si dice anche, in questo caso, che la catena *perde memoria* dello stato dal quale è partita;

- (b) la DdP limite  $\bar{\pi}$  risulta anche essere stazionaria in quanto con un calcolo diretto si vede che  $\bar{\pi} \Pi = \bar{\pi}$ ; pertanto, se la DdP iniziale coincide proprio con  $\bar{\pi}$  essa resta eguale a se stessa lungo tutta l'evoluzione della catena; se invece la DdP iniziale è generica la catena, dopo un numero abbastanza elevato di passi, approssima una DdP stazionaria che, per quanto detto in (a), è indipendente dallo stato iniziale;
- (c) si verifica facilmente che se  $p_{ij} > 0$  per tutti i valori di  $i, j$ , cioè se tutte le transizioni sono permesse, allora anche  $\bar{p}_j > 0$  per tutti i valori di  $j$ .

---

<sup>11</sup> Si tratta del caso in cui la catena oscilla deterministicamente fra i due stati ( $p_{00} = p_{11} = 0$ ) e del caso in cui la catena degenera in due cicli completamente separati permanendo indefinitamente in 0 oppure 1 ( $p_{00} = p_{11} = 1$ ). Maggiori dettagli su queste catene saranno forniti in uno degli esempi successivi.

Le proprietà qui esposte non sono tipiche solo della semplice Catena di Markov esaminata, ma risultano caratteristiche per tutta un'importante categoria di catene finite che saranno discusse nelle Definizioni e nel Teorema seguenti.  $\diamond$

**III.3.11 Definizione:** Una Catena di Markov finita (con  $E = \{1, \dots, N\}$ ) e omogenea con matrice delle transizioni  $\Pi = \|p_{ij}\|$  si dice **ergodica**<sup>12</sup> se, dette  $p_{ij}^{(n)}$  le sue probabilità di transizione ad  $n$  passi, i limiti

$$\bar{p}_j = \lim_n p_{ij}^{(n)}, \quad i = 1, \dots, N$$

esistono, sono indipendenti da  $i$  e costituiscono una DdP  $\bar{\pi}$  (cioè risulta  $\sum_j \bar{p}_j = 1$ ) tale che  $\bar{p}_j > 0$  per  $j = 1, \dots, N$ . Tale DdP  $\bar{\pi}$  si dice anche **distribuzione ergodica**.  $\triangle$

**III.3.12 Definizione:** Una Catena di Markov finita ed omogenea si dice **transitiva** se esiste un numero intero  $n_0$  tale che

$$\min_{i,j} p_{ij}^{(n_0)} > 0. \quad \triangle$$

**III.3.13 Teorema (Teorema ergodico di A.A. Markov):** Data una Catena di Markov finita (con  $E = \{1, \dots, N\}$ ) e omogenea con matrice delle transizioni  $\Pi = \|p_{ij}\|$  essa risulta ergodica se e solo se è transitiva; inoltre la distribuzione ergodica  $\bar{\pi}$  è unica e risulta essere una distribuzione stazionaria.

**Dimostrazione:** Dimostriamo innanzitutto che se la catena è transitiva essa è anche ergodica. Posto per brevità

$$m_j^{(n)} = \min_i p_{ij}^{(n)}, \quad M_j^{(n)} = \max_i p_{ij}^{(n)},$$

si dimostra facilmente che  $m_j^{(n)} \leq m_j^{(n+1)}$  e  $M_j^{(n)} \geq M_j^{(n+1)}$  in quanto dalle equazioni di Chapman-Kolmogorov si ha

$$\begin{aligned} m_j^{(n+1)} &= \min_i p_{ij}^{(n+1)} = \min_i \sum_{k \in E} p_{ik} p_{kj}^{(n)} \\ &\geq \min_i \left[ \left( \sum_{k \in E} p_{ik} \right) \min_l p_{lj}^{(n)} \right] = \min_l p_{lj}^{(n)} = m_j^{(n)}, \end{aligned}$$

---

<sup>12</sup> L'aggettivo *ergodico* deriva dal greco  $\epsilon\rho\gamma\acute{\omega}\delta\eta\varsigma$  che vuol dire *faticoso, laborioso, difficile* e fu usato inizialmente nel dibattito sulla meccanica statistica degli anni '80 del XIX secolo da J.C. Maxwell e L. Boltzmann per indicare i sistemi meccanici la cui orbita nello spazio delle fasi passa attraverso tutti i punti di una superficie di energia costante.



e analogamente per  $M_j^{(n)}$ . Pertanto, con  $j$  fissato, le successioni (in  $n$ )  $m_j^{(n)}$  e  $M_j^{(n)}$  risultano rispettivamente monotona non decrescente e non crescente, e siccome ovviamente

$$m_j^{(n)} \leq p_{ij}^{(n)} \leq M_j^{(n)}, \quad \forall i \in E,$$

la convergenza verso un limite indipendente da  $i$  sarà assicurata se riusciremo a dimostrare che, comunque scelto  $j \in E$ , risulta

$$\lim_n (M_j^{(n)} - m_j^{(n)}) = 0, \quad \forall j \in E.$$

Per far questo cominciamo con l'osservare che, fissato  $j$ , la successione  $M_j^{(n)} - m_j^{(n)}$  è monotona non crescente al variare di  $n$  e che quindi, per provarne la convergenza, basterà studiare la convergenza di una successione estratta. Posto allora

$$\alpha = \min_{i,j} p_{ij}^{(n_0)}$$

con  $0 < \alpha < 1$  (dato che la catena è, per ipotesi, transitiva), dalle equazioni di Chapman-Kolmogorov si ha ancora

$$\begin{aligned} p_{ij}^{(n_0+n)} &= \sum_{k \in E} p_{ik}^{(n_0)} p_{kj}^{(n)} = \sum_{k \in E} (p_{ik}^{(n_0)} - \alpha p_{jk}^{(n)}) p_{kj}^{(n)} + \sum_{k \in E} \alpha p_{jk}^{(n)} p_{kj}^{(n)} \\ &= \sum_{k \in E} (p_{ik}^{(n_0)} - \alpha p_{jk}^{(n)}) p_{kj}^{(n)} + \alpha p_{jj}^{(2n)}; \end{aligned}$$

ma siccome  $\alpha \leq p_{ik}^{(n_0)}$  (per ogni  $i, j \in E$ ) e  $p_{jk}^{(n)} \leq 1$ , si ha anche  $p_{ik}^{(n_0)} - \alpha p_{jk}^{(n)} \geq 0$ , per cui, essendo  $p_{kj}^{(n)} \geq m_j^{(n)}$  per definizione, risulta

$$\begin{aligned} p_{ij}^{(n_0+n)} &\geq m_j^{(n)} \sum_{k \in E} (p_{ik}^{(n_0)} - \alpha p_{jk}^{(n)}) + \alpha p_{jj}^{(2n)} \\ &= m_j^{(n)} \left( \sum_{k \in E} p_{ik}^{(n_0)} - \alpha \sum_{k \in E} p_{jk}^{(n)} \right) + \alpha p_{jj}^{(2n)} = m_j^{(n)} (1 - \alpha) + \alpha p_{jj}^{(2n)} \end{aligned}$$

e in definitiva

$$m_j^{(n_0+n)} \geq m_j^{(n)} (1 - \alpha) + \alpha p_{jj}^{(2n)}.$$

In modo analogo si prova che

$$M_j^{(n_0+n)} \leq M_j^{(n)} (1 - \alpha) + \alpha p_{jj}^{(2n)}$$

e sottraendo le due relazioni si ha

$$M_j^{(n_0+n)} - m_j^{(n_0+n)} \leq (M_j^{(n)} - m_j^{(n)}) (1 - \alpha).$$

Ne segue anche, allora, che preso un intero  $l \geq 1$  si ha

$$\begin{aligned} M_j^{(ln_0+n)} - m_j^{(ln_0+n)} &\leq (M_j^{[(l-1)n_0+n]} - m_j^{[(l-1)n_0+n]})(1-\alpha) \\ &\leq \dots \leq (M_j^{(n_0)} - m_j^{(n_0)})(1-\alpha)^l \end{aligned}$$

sicch , per  $l \rightarrow \infty$ , la successione estratta  $M_j^{(ln_0+n)} - m_j^{(ln_0+n)}$  risulta infinitesima dato che  $0 < 1-\alpha < 1$  per definizione. Pertanto la successione  $M_j^{(n)} - m_j^{(n)}$  risulta infinitesima e quindi i limiti

$$\bar{p}_j = \lim_n p_{ij}^{(n)}$$

esistono per ogni  $i, j \in E$  e sono indipendenti da  $i$ . Inoltre risulta  $\bar{p}_j > 0$  per ogni  $j \in E$  in quanto si ha anche

$$\bar{p}_j = \lim_n m_j^{(n)}$$

con  $m_j^{(n)} \geq m_j^{(n_0)} \geq \alpha > 0$  per ogni  $n \geq n_0$ . La catena   quindi ergodica secondo la Definizione III.3.11.

Supponendo ora che la catena sia ergodica proveremo che essa   anche transitiva. Se infatti la catena   ergodica comunque fissato  $j$  deve risultare

$$\lim_n p_{ij}^{(n)} = \bar{p}_j > 0, \quad \forall i \in E,$$

cio  deve succedere che

$$\forall j \in E, \forall \epsilon > 0, \exists \nu_{\epsilon,j} \exists' \bar{p}_j + \epsilon > p_{ij}^{(n)} > \bar{p}_j - \epsilon, \forall n \geq \nu_{\epsilon,j}, \forall i \in E.$$

Siccome per   $\bar{p}_j > 0$  esister  certamente un  $\epsilon_0 > 0$  tale che  $\bar{p}_j - \epsilon_0 > 0$  e in corrispondenza di tale  $\epsilon_0$  potremo affermare che

$$\forall j \in E, \exists \nu_{0,j} \exists' p_{ij}^{(n)} > \bar{p}_j - \epsilon_0 > 0, \forall n \geq \nu_{0,j}, \forall i \in E.$$

Posto allora

$$\nu_0 = \max_{j \in E} \nu_{0,j}$$

e scelto un arbitrario  $n_0 \geq \nu_0$  risulter  anche  $p_{ij}^{(n_0)} > \bar{p}_j - \epsilon_0 > 0$ , comunque scelti  $i, j \in E$ , e quindi in definitiva per il nostro  $n_0$  si ha

$$\min_{i,j \in E} p_{ij}^{(n_0)} > 0,$$

il che mostra che la catena   transitiva.

Che la distribuzione ergodica sia stazionaria si mostra passando al limite per  $n \rightarrow \infty$  nell'equazione di Chapman-Kolmogorov

$$p_{ij}^{(n+1)} = \sum_{k \in E} p_{ik}^{(n)} p_{kj}.$$

Infatti se la catena è ergodica al limite si ottiene

$$\bar{p}_j = \sum_{k \in E} \bar{p}_k p_{kj},$$

cioè  $\bar{\pi} \Pi = \bar{\pi}$ , il che mostra che  $\bar{\pi}$  è stazionaria. L'unicità della distribuzione ergodica e stazionaria  $\bar{\pi}$  si prova infine osservando che se ve ne fosse un'altra  $\bar{\pi}'$  l'applicazione ripetuta delle equazioni di Chapman-Kolmogorov implicherebbe che

$$\bar{p}'_j = \sum_{k \in E} \bar{p}'_k p_{kj} = \dots = \sum_{k \in E} \bar{p}'_k p_{kj}^{(n)} = \dots, \quad \forall n \geq 1, \quad \forall j \in E$$

e quindi passando al limite per  $n \rightarrow \infty$  si avrebbe

$$\bar{p}'_j = \sum_{k \in E} \bar{p}'_k \bar{p}_j = \bar{p}_j, \quad \forall j \in E$$

cioè  $\bar{\pi}' = \bar{\pi}$ . □

**III.3.14 Osservazione:** Il Teorema ergodico di Markov mostra che l'ergodicità e la transitività delle catene sono due proprietà perfettamente equivalenti sicché potremmo dire che la progressiva perdita di memoria dello stato iniziale (ossia l'ergodicità) in una catena è una caratteristica strettamente legata alla accessibilità di qualunque stato a partire da qualunque altro in un numero finito di passi (ossia la transitività). Una ovvia conseguenza dell'ergodicità è anche il fatto che la distribuzione  $\bar{\pi}$  risulta anche essere il limite cui tende la successione delle  $\pi^{(n)}$  della catena; infatti

$$p_k^{(n)} = \sum_{j \in E} p_j p_{jk}^{(n)} \xrightarrow{n} \sum_{j \in E} p_j \bar{p}_k = \bar{p}_k.$$

Il Teorema mostra anche che la distribuzione ergodica è stazionaria; bisogna però guardarsi dal pensare che l'esistenza di distribuzioni stazionarie sia una caratteristica solo delle Catene di Markov ergodiche. Mostriamo ora con un esempio, infatti, che distribuzioni stazionarie (ed uniche) possono esistere anche per catene non ergodiche. Consideriamo infatti il caso, da noi escluso nella discussione dell'Esempio III.3.10, in cui  $p_{00} = p_{11} = 0$  e conseguentemente  $p_{01} = p_{10} = 1$ ; esso è allora caratterizzato da una matrice delle transizioni del tipo

$$\Pi = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

sicché per ogni  $n \geq 0$  si ha

$$\Pi^{2n} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \Pi^{2n+1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Ne segue che la successione delle  $p_{ij}^{(n)}$  risulta oscillante per  $n \rightarrow \infty$ , il  $\lim_n p_{ij}^{(n)}$  non esiste e la catena non è, quindi, ergodica. Ciononostante la relazione  $\bar{\pi} \Pi = \bar{\pi}$  ossia

$$\bar{p}_j = \sum_{k \in E} \bar{p}_k p_{kj}$$

che definisce le distribuzioni stazionarie ha in questo caso la soluzione banale  $\bar{p}_1 = \bar{p}_2$  che (tenendo conto della relazione  $\bar{p}_1 + \bar{p}_2 = 1$ ) ammette come unica soluzione la distribuzione stazionaria  $\bar{\pi} = (1/2, 1/2)$ . Pertanto una distribuzione stazionaria (unica) esiste anche per la catena non ergodica data.  $\circ$

**III.3.15 Osservazione:** Le Catene di Markov ergodiche sono particolarmente importanti perché per esse si può dimostrare una Legge dei Grandi Numeri. La rilevanza di questo risultato sta nel fatto che una Catena di Markov costituisce una successione di v.a. non indipendenti sicché il Teorema che dimostreremo non può in nessun caso essere considerato come una conseguenza dei teoremi già enunciati sulla Legge dei Grandi Numeri (Teoremi III.1.7, III.1.9, III.1.13, III.1.14 e III.1.15) che valgono solo per successioni di v.a. indipendenti. Per introdurre la discussione su questo punto consideriamo una Catena di Markov finita  $(X_n)_{n \geq 0}$  su uno spazio delle fasi  $E = \{0, 1, \dots, n\}$ . Preso un sottinsieme  $A \subseteq E$ , la v.a.  $I_A(X_k)$  assumerà valore 1 solo se la catena si trova in  $A$  all'istante  $k$ -mo sicché

$$\nu_A(n) = \frac{1}{n+1} \sum_{k=0}^n I_A(X_k)$$

sarà la v.a. il cui valore rappresenta la **frequenza di ritrovamenti** della catena in  $A$  durante i primi  $n$  istanti di tempo. Ovviamente  $A$  può anche coincidere con un solo stato ( $A = \{j\}$ ), ed in questo caso parleremo di frequenza di ritrovamenti in uno stato. Si vede ora facilmente che

$$\mathbf{E}(I_{\{j\}}(X_k) | X_0 = i) = \mathbf{P}(X_k \in \{j\} | X_0 = i) = p_{ij}^{(k)},$$

e quindi si potrà introdurre la notazione seguente:

$$\mathbf{E}(I_A(X_k) | X_0 = i) = \mathbf{P}(X_k \in A | X_0 = i) = \sum_{j \in A} p_{ij}^{(k)} = p_{i,A}^{(k)}.$$

Potremo pertanto scrivere che

$$\mathbf{E}(\nu_{\{j\}}(n) | X_0 = i) = \frac{1}{n+1} \sum_{k=0}^n p_{ij}^{(k)}, \quad \mathbf{E}(\nu_A(n) | X_0 = i) = \frac{1}{n+1} \sum_{k=0}^n p_{i,A}^{(k)}.$$

Se ora supponiamo che la catena sia ergodica le successioni numeriche  $p_{ij}^{(k)}$  e  $p_{i,A}^{(k)}$  risulteranno convergenti, per  $k \rightarrow \infty$ , rispettivamente verso la distribuzione ergodica  $\bar{\pi} = (\bar{p}_j)_{j \in E}$  e verso la probabilità di essere in  $A$  secondo tale distribuzione, cioè

$$\bar{p}_A = \sum_{j \in A} \bar{p}_j.$$

Per un noto risultato dell'analisi<sup>13</sup> potremo allora dire che

$$\mathbf{E}(\nu_{\{j\}}(n) | X_0 = i) \xrightarrow{n} \bar{p}_j, \quad \mathbf{E}(\nu_A(n) | X_0 = i) \xrightarrow{n} \bar{p}_A.$$

Pertanto la probabilità di trovare la catena in  $A$  (o nello stato  $j$ ) dopo un numero  $n$  abbastanza elevato di passi è approssimata arbitrariamente bene dal VdA della frequenza di ritrovamenti in  $A$  nei primi  $n$  istanti di tempo. Inoltre il VdA di tale frequenza di ritrovamenti (pur essendo condizionata dall'evento  $\{X_0 = i\}$ ) perde progressivamente memoria dello stato iniziale  $i$ . In realtà per Catene di Markov ergodiche è possibile provare più di questo: non solo i VdA (condizionati) delle frequenze di ritrovamenti in  $A$  approssimano la probabilità di presenza  $\bar{p}_A$  (calcolata secondo la distribuzione ergodica  $\bar{\pi}$ ), ma le stesse v.a.  $\nu_A(n)$  approssimano la  $\bar{p}_A$  secondo la convergenza in probabilità. Questa affermazione è l'oggetto della seguente Legge dei Grandi Numeri per successioni di v.a. dipendenti che formano Catene di Markov ergodiche.  $\circ$

**III.3.16 Teorema (Legge dei Grandi Numeri di A.A. Markov):** Data una Catena di Markov finita ed ergodica risulta che (con le notazioni dell'Osservazione precedente )

$$\lim_n \mathbf{P}(|\nu_A(n) - \bar{p}_A| > \epsilon) = 0, \quad \forall \epsilon > 0,$$

cioè

$$\nu_A(n) \xrightarrow{\mathbf{P}} \bar{p}_A, \quad (n \rightarrow \infty)$$

qualunque sia la distribuzione iniziale e per ogni  $A \subseteq E$ .

**Dimostrazione:** Mostriamo innanzitutto (con notazioni riprese dalla dimostrazione del Teorema III.3.13) che è sempre possibile trovare due numeri  $C > 1$  e  $\rho < 1$  tali che, risulti

$$|p_{ij}^{(n)} - \bar{p}_j| \leq C\rho^n, \quad \forall n \geq 1, \quad \forall i, j \in E.$$

Infatti, siccome dal Teorema III.3.13 risulta  $m_j^{(n)} \uparrow \bar{p}_j$  e  $M_j^{(n)} \downarrow \bar{p}_j$  (per  $n \rightarrow \infty$ ), e per le definizioni stesse di  $m_j^{(n)}$  e  $M_j^{(n)}$ , si hanno le disequaglianze

$$m_j^{(n)} \leq \bar{p}_j \leq M_j^{(n)}, \quad m_j^{(n)} \leq p_{ij}^{(n)} \leq M_j^{(n)}, \quad \forall n \geq 1, \quad \forall i, j \in E,$$

da cui segue che

$$|p_{ij}^{(n)} - \bar{p}_j| \leq M_j^{(n)} - m_j^{(n)}, \quad \forall n \geq 1, \quad \forall i, j \in E.$$

---

<sup>13</sup> Vedi ad esempio **C. Miranda:** *Lezioni di Analisi Matematica - I*; Liguori, Napoli, 1967, p. 187.

Essendo la nostra catena ergodica (e quindi transitiva) sappiamo che esiste un  $n_0$  tale che

$$\min_{i,j \in E} p_{ij}^{(n_0)} = \alpha > 0.$$

Utilizzando ora alcuni risultati secondari ottenuti nella dimostrazione del Teorema III.3.13, siccome (essendo per definizione  $M_j^{(k)} \leq 1$ ,  $m_j^{(k)} \geq 0$ ,  $M_j^{(k)} - m_j^{(k)} \geq 0$  e  $0 < \alpha < 1$ ) si ha

$$0 \leq M_j^{(k)} - m_j^{(k)} \leq 1 < \frac{1}{1-\alpha},$$

avremo che per  $n \geq n_0$  varrà la seguente disequaglianza<sup>14</sup>

$$\begin{aligned} M_j^{(n)} - m_j^{(n)} &= M_j^{(ln_0+k)} - m_j^{(ln_0+k)} \leq (M_j^{(k)} - m_j^{(k)})(1-\alpha)^l \\ &= (M_j^{(k)} - m_j^{(k)})(1-\alpha)^{[n/n_0]} \leq (1-\alpha)^{[n/n_0]-1}. \end{aligned}$$

Siccome inoltre  $n/n_0 - 1 \leq [n/n_0]$  e quindi  $n/n_0 - 2 \leq [n/n_0] - 1$ , potremo affermare che per ogni  $i, j \in E$  e per ogni  $n \geq n_0$  risulterà

$$\begin{aligned} |p_{ij}^{(n)} - \bar{p}_j| &\leq (1-\alpha)^{[n/n_0]-1} \\ &\leq (1-\alpha)^{n/n_0-2} = (1-\alpha)^{-2} [(1-\alpha)^{1/n_0}]^n = C' \rho^n \end{aligned}$$

dove abbiamo posto

$$C' = (1-\alpha)^{-2} > 1, \quad \rho = (1-\alpha)^{1/n_0} < 1.$$

Essendo d'altra parte l'insieme numerico delle restanti  $|p_{ij}^{(n)} - \bar{p}_j|$  con  $n \leq n_0$  composto solo da un numero finito di elementi al variare di  $i, j \in E$  e di  $n \leq n_0$ , sarà sempre possibile trovare un altro  $C \geq C' > 1$  abbastanza grande per far in modo che la disequaglianza

$$|p_{ij}^{(n)} - \bar{p}_j| \leq C \rho^n$$

sia verificata per ogni  $i, j \in E$  e per ogni  $n \geq 1$ .

Ciò posto possiamo passare a dimostrare la tesi principale del Teorema: siccome (vedi Osservazione III.3.15) si ha

$$\nu_A(n) = \sum_{j \in A} \nu_{\{j\}}(n), \quad \bar{p}_A = \sum_{j \in A} \bar{p}_j$$

è facile vedere che sarà sufficiente provare

$$\lim_n \mathbf{P}(|\nu_{\{j\}}(n) - \bar{p}_j| > \epsilon) = 0, \quad \forall \epsilon > 0, \quad \forall j \in E.$$

---

<sup>14</sup> Per  $n \geq n_0$  poniamo  $n = ln_0 + k$  con  $l = [n/n_0]$  e indicando come al solito con  $[n/n_0]$  la parte intera del numero razionale  $n/n_0$ . Naturalmente risulterà anche  $k = 0, 1, \dots, n_0 - 1$ .

Dalla Formula della probabilità totale (Proposizione I.4.5) si ricava, inoltre, che basterà allora provare

$$\lim_n \mathbf{P}(|\nu_{\{j\}}(n) - \bar{p}_j| > \epsilon \mid X_0 = i) = 0, \quad \forall \epsilon > 0, \quad \forall i, j \in \mathbf{E},$$

in quanto il condizionamento può essere rimosso alla fine della dimostrazione prendendo il VdA su tutti i valori  $i$  di  $X_0$ . La disuguaglianza di Chebyshev (Corollario I.7.4) mostra, d'altra parte, che

$$\mathbf{P}(|\nu_{\{j\}}(n) - \bar{p}_j| > \epsilon \mid X_0 = i) < \frac{\mathbf{E}(|\nu_{\{j\}}(n) - \bar{p}_j|^2 \mid X_0 = i)}{\epsilon^2}$$

sicché il nostro Teorema sarà dimostrato se riusciremo a provare che

$$\lim_n \mathbf{E}(|\nu_{\{j\}}(n) - \bar{p}_j|^2 \mid X_0 = i) = 0, \quad \forall i, j \in \mathbf{E}.$$

Per far questo, tenendo conto del fatto che dalle definizioni si ha

$$\nu_{\{j\}}(n) - \bar{p}_j = \frac{1}{n+1} \sum_{k=0}^n I_{\{j\}}(X_k) - \bar{p}_j = \frac{1}{n+1} \sum_{k=0}^n [I_{\{j\}}(X_k) - \bar{p}_j],$$

osserveremo innanzitutto che

$$\begin{aligned} \mathbf{E}(|\nu_{\{j\}}(n) - \bar{p}_j|^2 \mid X_0 = i) &= \frac{1}{(n+1)^2} \mathbf{E} \left( \left[ \sum_{k=0}^n (I_{\{j\}}(X_k) - \bar{p}_j) \right]^2 \mid X_0 = i \right) \\ &= \frac{1}{(n+1)^2} \sum_{k,l=0}^n m_{ij}^{(k,l)} \leq \frac{1}{(n+1)^2} \sum_{k,l=0}^n |m_{ij}^{(k,l)}|, \end{aligned}$$

dove abbiamo posto per brevità

$$\begin{aligned} m_{ij}^{(k,l)} &= \mathbf{E} \left( [I_{\{j\}}(X_k) - \bar{p}_j] \cdot [I_{\{j\}}(X_l) - \bar{p}_j] \mid X_0 = i \right) \\ &= \mathbf{E} (I_{\{j\}}(X_k) I_{\{j\}}(X_l) \mid X_0 = i) \\ &\quad + \bar{p}_j^2 - \bar{p}_j \left[ \mathbf{E} (I_{\{j\}}(X_k) \mid X_0 = i) + \mathbf{E} (I_{\{j\}}(X_l) \mid X_0 = i) \right] \\ &= \mathbf{P}(X_k = j, X_l = j \mid X_0 = i) + \bar{p}_j^2 - \bar{p}_j (p_{ij}^{(k)} + p_{ij}^{(l)}). \end{aligned}$$

Risulta inoltre dalla proprietà di Markov che

$$\begin{aligned} \mathbf{P}(X_k = j, X_l = j \mid X_0 = i) &= \begin{cases} \mathbf{P}(X_k = j \mid X_l = j, X_0 = i) \mathbf{P}(X_l = j \mid X_0 = i), & \text{se } k \geq l, \\ \mathbf{P}(X_l = j \mid X_k = j, X_0 = i) \mathbf{P}(X_k = j \mid X_0 = i), & \text{se } k \leq l, \end{cases} \\ &= \begin{cases} \mathbf{P}(X_k = j \mid X_l = j) \mathbf{P}(X_l = j \mid X_0 = i) = p_{jj}^{(k-l)} p_{ij}^{(l)}, & \text{se } k \geq l, \\ \mathbf{P}(X_l = j \mid X_k = j) \mathbf{P}(X_k = j \mid X_0 = i) = p_{jj}^{(l-k)} p_{ij}^{(k)}, & \text{se } k \leq l, \end{cases} \\ &= p_{jj}^{(t)} p_{ij}^{(s)}, \end{aligned}$$

avendo posto

$$t = |k - l|, \quad s = \min\{k, l\},$$

e pertanto si ha che

$$m_{ij}^{(k,l)} = p_{jj}^{(t)} p_{ij}^{(s)} + \bar{p}_j^2 - \bar{p}_j (p_{ij}^{(k)} + p_{ij}^{(l)}).$$

Dalla diseguaglianza provata nella parte iniziale sappiamo però che comunque scelti  $n \geq 1$  e  $i, j \in E$  si ha  $p_{ij}^{(n)} = \bar{p}_j + \epsilon_{ij}^{(n)}$  con  $|\epsilon_{ij}^{(n)}| \leq C\rho^n$ , e pertanto risulta

$$\begin{aligned} m_{ij}^{(k,l)} &= (\bar{p}_j + \epsilon_{jj}^{(t)})(\bar{p}_j + \epsilon_{ij}^{(s)}) + \bar{p}_j^2 - \bar{p}_j (2\bar{p}_j + \epsilon_{ij}^{(k)} + \epsilon_{ij}^{(l)}) \\ &= \bar{p}_j (\epsilon_{jj}^{(t)} + \epsilon_{ij}^{(s)} - \epsilon_{ij}^{(k)} - \epsilon_{ij}^{(l)}) + \epsilon_{jj}^{(t)} \epsilon_{ij}^{(s)}. \end{aligned}$$

Ne segue che

$$\begin{aligned} |m_{ij}^{(k,l)}| &\leq C\bar{p}_j(\rho^t + \rho^s + \rho^k + \rho^l) + C^2\rho^{s+t} \\ &\leq C(\rho^t + \rho^s + \rho^k + \rho^l) + C^2\rho^{s+t}, \end{aligned}$$

e siccome

$$s + t = \begin{cases} k, & \text{se } k \geq l, \\ l, & \text{se } k \leq l, \end{cases}$$

ponendo  $C_1 = C + C^2 > 1$  otterremo

$$|m_{ij}^{(k,l)}| \leq C_1(\rho^t + \rho^s + \rho^k + \rho^l), \quad k, l \geq 1, \quad i, j \in E.$$

Conseguentemente per ogni  $n \geq 1$  avremo

$$\mathbf{E} (|\nu_{\{j\}}(n) - \bar{p}_j|^2 \mid X_0 = i) \leq \frac{C_1}{(n+1)^2} \sum_{k,l=0}^n (\rho^t + \rho^s + \rho^k + \rho^l).$$



Ma con un calcolo diretto, essendo  $\rho < 1$ , si ha

$$\begin{aligned}
 \sum_{k,l=0}^n \rho^k &= \sum_{k,l=0}^n \rho^l = (n+1) \sum_{k=0}^n \rho^k = (n+1) \frac{1-\rho^{n+1}}{1-\rho} \leq \frac{n+1}{1-\rho} < \frac{2(n+1)}{1-\rho}; \\
 \sum_{k,l=0}^n \rho^t &= \sum_{k=0}^n \sum_{l=0}^n \rho^{|k-l|} = \sum_{k=0}^n \sum_{l=0}^k \rho^{k-l} + \sum_{k=0}^n \sum_{l=k+1}^n \rho^{l-k} \\
 &= \sum_{r=0}^n \sum_{k=r}^n \rho^r + \sum_{r=1}^n \sum_{k=0}^{n-r} \rho^r \\
 &= \sum_{r=0}^n (n-r+1)\rho^r + \sum_{r=1}^n (n-r+1)\rho^r \\
 &= \sum_{r=0}^n 2(n-r+1)\rho^r - (n+1) \\
 &= 2(n+1) \sum_{r=0}^n \rho^r - 2 \sum_{r=0}^n r\rho^r - (n+1) \leq 2(n+1) \sum_{r=0}^n \rho^r \\
 &= 2(n+1) \frac{1-\rho^{n+1}}{1-\rho} < \frac{2(n+1)}{1-\rho}; \\
 \sum_{k,l=0}^n \rho^s &= \sum_{k=0}^n \sum_{l=0}^n \rho^{\min\{k,l\}} = \sum_{k=0}^n \sum_{l=0}^k \rho^l + \sum_{k=0}^n \sum_{l=k+1}^n \rho^k \\
 &= \sum_{l=0}^n \sum_{k=l}^n \rho^l + \sum_{k=0}^n \sum_{l=k+1}^n \rho^k = \sum_{k=0}^n \sum_{l=k}^n \rho^k + \sum_{k=0}^n \sum_{l=k+1}^n \rho^k \\
 &= \sum_{k=0}^n (n-k+1)\rho^k + \sum_{k=0}^n (n-k)\rho^k \\
 &= 2 \sum_{k=0}^n (n-k+1)\rho^k - \sum_{k=0}^n \rho^k \\
 &= 2(n+1) \sum_{k=0}^n \rho^k - 2 \sum_{k=0}^n k - \sum_{k=0}^n \rho^k \rho^k \leq 2(n+1) \sum_{k=0}^n \rho^k \\
 &= 2(n+1) \frac{1-\rho^{n+1}}{1-\rho} < \frac{2(n+1)}{1-\rho};
 \end{aligned}$$

e pertanto risulta

$$\mathbf{E} (|\nu_{\{j\}}(n) - \bar{p}_j|^2 \mid X_0 = i) \leq \frac{8C_1}{(n+1)(1-\rho)},$$

da cui deriva il risultato richiesto dato che  $\mathbf{E} (|\nu_{\{j\}}(n) - \bar{p}_j|^2 \mid X_0 = i) \xrightarrow{n} 0$  indipendentemente da  $i, j \in E$ .  $\square$



## III.4 Esercizi svolti

**III.4.1 Esercizio (E. Wentzel e L. Ovcharov: *Applied problems in probability theory*; Mir, Mosca, 1986, p. 243):** Data la successione di v.a.  $X_k$  ( $k = 1, 2, \dots$ ) indipendenti ed uniformemente distribuite negli intervalli  $[-k, k]$  e definita la successione delle medie aritmetiche

$$Y_n = \frac{1}{n} \sum_{k=1}^n X_k$$

stabilire se  $Y_n$  converge verso 0 per  $n \rightarrow \infty$  (Legge dei grandi numeri).

**Soluzione:** Siccome  $\mathbf{E} X_k = 0$  per ogni  $k = 1, 2, \dots$ , è facile verificare che anche  $\mathbf{E} Y_n = 0$  per ogni  $n = 1, 2, \dots$ . Tuttavia l'applicazione del Teorema III.1.9 o del Teorema III.1.14 (quella dei Teoremi III.1.7 e III.1.15 non può essere presa in considerazione dato che le v.a.  $X_k$  non sono identicamente distribuite) è impedita dal fatto che la successione delle varianze  $\mathbf{V} X_k = \mathbf{E} X_k^2 = k^2/3$  non soddisfa le ipotesi di questi teoremi.  $\diamond$

**III.4.2 Esercizio (E. Wentzel e L. Ovcharov: *Applied problems in probability theory*; Mir, Mosca, 1986, p. 244):** Un computer genera casualmente delle cifre binarie indipendenti con 0 e 1 equiprobabili. La successione dei simboli è suddivisa in gruppi consistenti ciascuno dello stesso simbolo (ad esempio 000 111 0 1 00 1111 00 11 ...) e di lunghezza aleatoria. Se  $n$  è il numero dei gruppi di simboli, determinare il comportamento per  $n \rightarrow \infty$  della v.a.  $Y_n$  che rappresenta il numero medio di simboli per gruppo.

**Soluzione:** Se  $X_i$  è il numero di simboli nell' $i$ -mo gruppo, la v.a. alla quale siamo interessati ha la forma

$$Y_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Se  $p = 0,5$  è la probabilità che sia generato il simbolo 1 (con  $q = 1 - p = 0,5$ ), le v.a.  $X_i$  risulteranno tutte i.i.d. con la seguente distribuzione (**distribuzione geometrica**):  $X_i$  può assumere qualunque valore intero (a partire da 1) e si verifica facilmente che

$$\mathbf{P}(X_i = k \mid \text{se il primo simbolo è } 1) = p^{k-1}q$$

$$\mathbf{P}(X_i = k \mid \text{se il primo simbolo è } 0) = q^{k-1}p$$

sicché dalla formula della probabilità totale  $\mathbf{P}(X_i = k) = p^k q + q^k p = 2^{-k}$ . Ne segue allora che

$$\mathbf{E} X_i = \sum_{k=1}^{\infty} k 2^{-k} = 2, \quad \mathbf{V} X_i = \sum_{k=1}^{\infty} k^2 2^{-k} - 2^2 = 2,$$

e quindi anche

$$\mathbf{E} Y_n = \frac{2n}{n} = 2, \quad \mathbf{V} Y_n = \frac{2n}{n^2} = \frac{2}{n}.$$

Pertanto, essendo le  $X_i$  i.i.d. e (assolutamente) integrabili, la Legge forte dei grandi numeri III.1.15 ci consente di dire che  $\mathbf{P}\text{-q.o.} Y_n \xrightarrow{n} \mathbf{E} X_i = 2$ .  $\diamond$

**III.4.3 Esercizio (E. Wentzel e L. Ovcharov: *Applied problems in probability theory*; Mir, Mosca, 1986, p. 244):** Una fabbrica produce delle partite di  $n = 10000$  pezzi e la probabilità che uno di tali pezzi sia difettoso è  $p = 0,05$ . Le cause dei difetti sono indipendenti per i diversi pezzi. I pezzi difettosi sono accumulati in un recipiente: determinare il numero di pezzi per i quali tale recipiente deve essere progettato in modo che esso non risulti insufficiente con una probabilità di 0,99.

**Soluzione:** Il numero  $X$  di pezzi difettosi ha una distribuzione binomiale, ma siccome  $n$  è grande potremo calcolare il nostro risultato supponendo (vedi il Teorema limite integrale I.8.6) che risulti  $X \sim \mathcal{N}(np, npq) = \mathcal{N}(500, 475)$ . Bisognerà allora determinare il valore di  $x$  in modo tale che  $\mathbf{P}(X < x) = 0,99$ . Posto allora  $Y = (X - 500)/\sqrt{475} \sim \mathcal{N}(0, 1)$ , siccome  $\mathbf{P}(X < x) = \mathbf{P}(Y < (x - 500)/\sqrt{475})$ , dalla Tavola I.8.1 si ottiene la relazione approssimata

$$\frac{x - 500}{\sqrt{475}} \simeq 2,33$$

ovvero  $x \simeq 551$ .  $\diamond$

**III.4.4 Esercizio (E. Wentzel e L. Ovcharov: *Applied problems in probability theory*; Mir, Mosca, 1986, p. 245):** Una coda di  $n = 60$  persone attende di ritirare del denaro ad uno sportello bancario: la quantità di denaro prelevata da ciascuno è una v.a. con VdA  $m = 50.000$  Lit ed una deviazione standard  $\sigma = 20.000$  Lit. I prelevamenti sono indipendenti. Determinare l'ammontare del denaro che deve essere in cassa per soddisfare le richieste di tutti con una probabilità di 0,95, e la quantità di denaro residuo (con probabilità 0,95) se in cassa ci sono inizialmente 3.500.000 Lit.

**Soluzione:** Se  $X_i$  è la somma prelevata dall' $i$ -mo cliente, il denaro necessario sarà

$$Y = \sum_{i=1}^{60} X_i$$

con

$$\mathbf{E} Y = \sum_{i=1}^{60} \mathbf{E} X_i = 60m, \quad \mathbf{V} Y = \sum_{i=1}^{60} \mathbf{V} X_i = 60\sigma^2.$$

Sulla base del Teorema limite centrale (Teorema III.1.1) potremo supporre che approssimativamente  $Y \sim \mathcal{N}(60m, 60\sigma^2)$ , ossia che  $Z = (Y - 60m)/\sigma\sqrt{60} \sim \mathcal{N}(0, 1)$ . Ne segue che la condizione  $\mathbf{P}(Y < y) = \mathbf{P}(Z < (y - 60m)/\sigma\sqrt{60}) = 0,95$

è garantita dalla scelta (vedi Tavola I.8.1)  $y \simeq 60m + 1,65 \sqrt{60} \sigma \simeq 3.256.000$  Lit. Pertanto, se in cassa ci fossero inizialmente 3.500.000 Lit il denaro residuo sarebbe 244.000 con probabilità 0,95.  $\diamond$

**III.4.5 Esercizio (S. Lipschutz:** *Calcolo delle probabilità*; collana *Schaum's*, McGraw-Hill, Milano, 1994, p. 119): Si supponga che 300 errori di stampa siano distribuiti a caso in un libro di 500 pagine: si determinino le probabilità che una pagina contenga (a) 2 errori, (b) più di 2 errori.

**Soluzione:** Il numero  $X$  degli errori che cadono su una data pagina è una v.a. binomiale con  $p = 1/500$  ed  $n = 300$ . Non sarebbe corretto, però, questa volta approssimare i calcoli con una distribuzione normale (Teorema limite locale I.8.3) perché il valore di  $p$  è piuttosto piccolo (vedi Osservazione I.8.11). In tal caso converrà piuttosto adottare l'approssimazione fornita dal Teorema di Poisson I.8.13: la distribuzione di  $X$  sarà approssimativamente poissonniana con parametro  $\lambda = np = 3/5 = 0,6$ . Ne segue allora che

$$\mathbf{P}(X = 2) \simeq \frac{(0,6)^2 e^{-0,6}}{2} \simeq 0,099.$$

Allo stesso modo si calcola che  $\mathbf{P}(X = 0) \simeq 0,549$  e che  $\mathbf{P}(X = 1) \simeq 0,329$  e pertanto  $\mathbf{P}(X \geq 2) = 1 - \mathbf{P}(X = 0) - \mathbf{P}(X = 1) \simeq 0,122$ .  $\diamond$

**III.4.6 Esercizio (A.A. Sveshnikov:** *Problems in probability theory, mathematical statistics and theory of random functions*; Dover, New York, 1978, p. 172): È possibile calcolare il valore numerico approssimato dell'integrale

$$I = \int_a^{+\infty} \frac{\sin x}{x} dx, \quad a > 0$$

mediante il metodo di Monte Carlo? Nota che l'integrale assegnato (nel senso di un integrale improprio di Riemann) esiste come una ben determinata funzione del parametro  $a$  (vedi ad esempio **E.T. Whittaker e G.N. Watson:** *A course of modern analysis*; Cambridge, Cambridge, 1963, p. 352).

**Soluzione:** Riformuliamo innanzitutto la discussione svolta in Esempio III.1.17 semplificando il metodo di Monte Carlo nel modo seguente: il calcolo dell'integrale

$$I = \int_0^1 g(x) dx$$

può essere eseguito osservando che, se  $\xi$  è una v.a. uniformemente distribuita su  $[0, 1]$ , si ha per  $\eta = g(\xi)$

$$\mathbf{E} \eta = \mathbf{E} [g(\xi)] = \int_0^1 g(x) dx.$$

Se la Legge dei grandi numeri è applicabile si potrà allora valutare  $I$  stimando il VdA di  $\eta = g(\xi)$  mediante una media aritmetica: detta  $(\xi_k)_{k \in \mathbf{N}}$  una successione di v.a. i.i.d. uniformi su  $[0, 1]$ ,  $\mathbf{P}$ -q.o. avremo che

$$\frac{1}{n} \sum_{k=1}^n g(\xi_k) \xrightarrow{n} \mathbf{E} g(\xi) = I.$$

Per discutere il nostro problema converrà inizialmente eseguire il cambiamento di variabile di integrazione  $y = a/x$  mediante il quale si ottiene

$$I = \int_0^1 \frac{1}{y} \sin \frac{a}{y} dy.$$

Possiamo dunque pensare di calcolare il nostro integrale approssimandolo con la successione

$$I_n = \frac{1}{n} \sum_{k=1}^n \frac{1}{\eta_k} \sin \frac{a}{\eta_k}$$

dove le  $\eta_k$  sono v.a. i.i.d. uniformi su  $[0, 1]$ . Per poter applicare una qualche forma della Legge dei grandi numeri (Teorema III.1.7 o III.1.15), comunque, dobbiamo preliminarmente verificare l'assoluta integrabilità delle v.a.  $\frac{1}{\eta_k} \sin \frac{a}{\eta_k}$  (ossia l'esistenza del loro VdA): essendo le  $\eta_k$  tutte uniformemente distribuite su  $[0, 1]$  ciò equivale a richiedere la convergenza assoluta (nel senso di Lebesgue) dell'integrale

$$I = \int_0^1 \frac{1}{y} \sin \frac{a}{y} dy,$$

che, trattandosi di un integrale improprio, non è garantita dalla esistenza dell'integrale di Riemann. Si ha allora che, posto  $s = [a/\pi] + 1$  dove come al solito  $[x]$  rappresenta la parte intera del numero reale  $x$ , risulta

$$\begin{aligned} \int_0^1 \left| \frac{1}{y} \sin \frac{a}{y} \right| dy &= \int_a^{+\infty} \frac{|\sin x|}{x} dx \geq \sum_{k=s}^{+\infty} \int_{k\pi}^{(k+1)\pi} \frac{|\sin x|}{x} dx = \sum_{k=s}^{+\infty} \int_0^\pi \frac{\sin y}{t + k\pi} dt \\ &> \sum_{k=s}^{+\infty} \frac{1}{(k+1)\pi} \int_0^\pi \sin t dt = \frac{2}{\pi} \sum_{k=s}^{+\infty} \frac{1}{k+1} = +\infty. \end{aligned}$$

Conseguentemente il nostro integrale non è assolutamente convergente sicché niente garantisce l'applicabilità della Legge dei grandi numeri e del metodo di Monte Carlo.  $\diamond$

**III.4.7 Esercizio (A.A. Sveshnikov:** *Problems in probability theory, mathematical statistics and theory of random functions*; Dover, New York, 1978, p. 177): Quanti campioni bisogna scegliere per calcolare l'integrale

$$I = \int_0^{\pi/2} \cos x dx$$

col metodo di Monte Carlo in modo che l'errore relativo non superi 0,05 con probabilità 0,9?

**Soluzione:** L'integrale

$$\frac{2}{\pi} I = \frac{2}{\pi} \int_0^{\pi/2} \cos x \, dx$$

può essere considerato come il VdA della v.a.  $\eta = \cos \xi$  con la v.a.  $\xi$  uniformemente distribuita su  $[0, \pi/2]$ . Siccome è facile verificare che le condizioni di applicabilità della Legge dei Grandi numeri sono soddisfatte, il nostro integrale sarà approssimabile (**P**-q.o.) con la successione di v.a.

$$X_n = \frac{\pi}{2n} \sum_{k=1}^n \cos \xi_k$$

con le  $\xi_k$  i.i.d. uniformi su  $[0, \pi/2]$ . Siccome inoltre si vede subito che  $\mathbf{E} X_n = I$ , il Teorema limite centrale (Teorema III.1.1) ci dice che

$$T_n = \frac{X_n - I}{\sqrt{\mathbf{V} X_n}} \sim \mathcal{N}(0, 1)$$

dato che le v.a.  $\cos \xi_k$  son i.i.d. con varianza finita non nulla. In particolare, dato che

$$\begin{aligned} \mathbf{V} \cos \xi_k &= \mathbf{E} \cos^2 \xi_k - (\mathbf{E} \cos \xi_k)^2 = \frac{2}{\pi} \int_0^{\pi/2} \cos^2 x \, dx - \left( \frac{2}{\pi} \int_0^{\pi/2} \cos x \, dx \right)^2 \\ &= \frac{\pi^2 - 4}{2\pi^2}, \end{aligned}$$

si ha

$$\mathbf{V} X_n = \frac{\pi^2}{4n} \mathbf{V} \cos \xi_1 = \frac{\pi^2 - 8}{8n},$$

e quindi

$$\begin{aligned} \mathbf{P}(|X_n - I| < 0,05) &= \mathbf{P}\left(|T_n| < 0,05 \sqrt{\frac{8n}{\pi^2 - 8}}\right) \\ &= \Phi\left(0,05 \sqrt{\frac{8n}{\pi^2 - 8}}\right) - \Phi\left(-0,05 \sqrt{\frac{8n}{\pi^2 - 8}}\right) \\ &= 2\Phi\left(0,05 \sqrt{\frac{8n}{\pi^2 - 8}}\right) - 1 = 0,9. \end{aligned}$$

Dalla Tavola I.8.1 si ha allora che deve risultare

$$0,05 \sqrt{\frac{8n}{\pi^2 - 8}} \simeq 1,65$$

e risolvendo questa equazione nell'incognita  $n$  si ottiene che  $n$  deve essere almeno eguale a 252 per soddisfare i requisiti del nostro problema.  $\diamond$

**III.4.8 Esercizio (S. Lipschutz:** *Calcolo delle probabilità*; collana *Schaum's*, McGraw-Hill, Milano, 1994, p. 141): Uno studioso che sta eseguendo un esperimento sul comportamento di alcune cavie osserva che in ogni prova l'80% degli animali che nella prova precedente hanno scelto di andare a destra andranno a destra anche nella successiva, mentre solo il 60% di quelli che avevano scelto di andare a sinistra andranno a destra successivamente. Se durante la prima prova le due direzioni sono state scelte in modo equiprobabile, cosa si prevede per la seconda e la terza prova? Quale sarà la distribuzione delle due direzioni dopo molte prove?

**Soluzione:** La catena di Markov delle prove ha solo due stati che indicheremo con  $D$  ed  $S$  (nell'ordine) e la matrice delle transizioni fra stati è

$$\Pi = \begin{pmatrix} 0,8 & 0,2 \\ 0,6 & 0,4 \end{pmatrix}$$

come si deduce facilmente dalle informazioni del problema. La distribuzione iniziale, invece, è data dal vettore  $\pi = (1/2, 1/2)$ . Le distribuzioni di probabilità per la seconda e la terza prova saranno allora

$$\begin{aligned} (0,5 \quad 0,5) \begin{pmatrix} 0,8 & 0,2 \\ 0,6 & 0,4 \end{pmatrix} &= (0,7 \quad 0,3), \\ (0,7 \quad 0,3) \begin{pmatrix} 0,8 & 0,2 \\ 0,6 & 0,4 \end{pmatrix} &= (0,74 \quad 0,26). \end{aligned}$$

Chiaramente la nostra catena è omogenea e transitiva e quindi il Teorema ergodico III.3.13 garantisce l'esistenza di una (unica) distribuzione limite che è anche stazionaria. Per determinare tale distribuzione  $\bar{\pi}$  basterà quindi risolvere il sistema di due equazioni  $\bar{\pi}\Pi = \bar{\pi}$ : un semplice calcolo mostra allora che  $\bar{\pi} = (3/4, 1/4)$  sicché, a lungo andare, è prevedibile che il 75% delle cavie andrà a destra ed il 25% a sinistra. Varrà la pena osservare che, come mostrato nel teorema ergodico, quest'ultimo risultato sulla distribuzione limite è indipendente dalla distribuzione iniziale.  $\diamond$

**III.4.9 Esercizio (S. Lipschutz:** *Calcolo delle probabilità*; collana *Schaum's*, McGraw-Hill, Milano, 1994, p. 143): Le urne  $A$  e  $B$  contengono rispettivamente due palline bianche e tre palline rosse: si estrae di volta in volta una pallina da ciascun'urna riponendovele dopo averle scambiate fra loro. Supponendo che lo stato della catena  $a_i$  sia individuato dal numero  $i$  di palline rosse nell'urna  $A$ , determinare la probabilità che lo stato sia  $a_2$  dopo tre transizioni e al limite per un gran numero di scambi.

**Soluzione:** Gli stati della nostra catena sono tre:  $a_0, a_1, a_2$  e la matrice delle transizioni può essere determinata come segue: se la catena è nello stato  $a_0$  essa



sicuramente passerà nello stato  $a_1$  per cui la prima riga è  $(0, 1, 0)$ ; se invece la catena è in  $a_1$  essa passerà in  $a_0$  con probabilità  $p_{10} = \frac{1}{2} \frac{1}{3} = \frac{1}{6}$  (bisogna estrarre una pallina rossa da  $A$  ed una bianca da  $B$ ) in  $a_2$  con probabilità  $p_{12} = \frac{1}{2} \frac{2}{3} = \frac{1}{3}$  (bisogna estrarre una pallina bianca da  $A$  ed una rossa da  $B$ ) e resterà quindi in  $a_1$  con probabilità  $p_{11} = 1 - \frac{1}{6} - \frac{1}{3} = \frac{1}{2}$ ; se infine la catena è in  $a_2$  con un ragionamento analogo si vede che essa non può andare in  $a_0$  mentre va in  $a_1$  con  $p_{21} = \frac{2}{3}$  e resta in  $a_2$  con  $p_{22} = \frac{1}{3}$ . La matrice delle transizioni è quindi

$$\Pi = \begin{pmatrix} 0 & 1 & 0 \\ 1/6 & 1/2 & 1/3 \\ 0 & 2/3 & 1/3 \end{pmatrix}$$

e siccome lo stato iniziale della catena è  $a_0$ , la distribuzione iniziale e quella dopo tre transizioni saranno

$$\pi^{(0)} = (1, 0, 0), \quad \pi^{(3)} = (1/12, 23/36, 5/18).$$

Per determinare la distribuzione ergodica bisognerà trovare la distribuzione stazionaria come soluzione del sistema di tre equazioni  $\bar{\pi}\Pi = \bar{\pi}$ . Si vede facilmente che tale distribuzione è

$$\bar{\pi} = (1/10, 6/10, 3/10)$$

sicché a lungo andare la probabilità che in  $A$  vi siano due palline rosse è  $0,3$ .  $\diamond$



# Indice analitico

additività .....	26
ago di Buffon .....	252
algebra .....	126
- finita .....	24
- generata .....	24, 135
algebre indipendenti .....	42
anello .....	126
assiomi di Kolmogorov .....	133
assoluta continuità .....	165
atomo .....	24, 196
Banach S. (1892-1945) .....	286
Bayes T. (1702-1761) .....	41
Bernoulli J. (1654-1705) .....	46
Bochner S. (1899-1982) .....	184
boreliani .....	141
Borel E. (1871-1956) .....	140
Buffon G.-L. (1707-1788) .....	252
Bunyakovskii V.J. (1804-1889) .....	66
<i>cadlag</i> .....	153
campionamento	
- con rimessa .....	19
- senza rimessa .....	21
Cantelli F.P. (1875-1966) .....	340
Cantor G. (1845-1918) .....	170
Carathéodory C. (1873-1950) .....	154
catena di Markov .....	363
- discreta .....	363
- ergodica .....	378
- finita .....	363
- omogenea .....	367, 374
- stazionaria .....	373
- transitiva .....	378
Cauchy A.-L. (1789-1857) .....	168
Chapman S. (1888-1970) .....	372
Chebyshev P.L. (1821-1894) .....	74
cilindro .....	144, 148
classe monotona .....	138
- generata .....	138
coefficiente di correlazione .....	69, 259
combinazioni .....	21
condizioni di Lindeberg .....	336
condizioni di Lyapunov .....	337

consistenza di spazi di probabilità .....	178, 181
continuità (da sotto, da sopra, nell'origine) .....	130
convergenza	
- debole .....	275
- in distribuzione .....	277
- in generale .....	277
- in $L^p$ .....	277
- in media quadratica [m.q.] .....	277
- in probabilità .....	277
- puntuale .....	190
- <b>P</b> -quasi ovunque [ <b>P-q.o.</b> ] .....	219, 277
convoluzione .....	271
covarianza .....	68, 259
criterio di convergenza di Cauchy .....	285
decomposizione .....	24
De Mére (1607-1684) .....	15
De Moivre A. (1667-1754) .....	77
De Morgan A. (1806-1871) .....	23
derivata (o densità) di Radon-Nikodym .....	226
deviazione standard .....	67, 259
diagrammi di Venn .....	22
diffusione .....	34
diseguaglianza	
- di Chebyshev .....	74, 221
- di Hölder .....	223
- di Jensen .....	222
- di Lyapunov .....	222
- di Minkowski .....	224
- di Schwarz .....	66, 221
disposizioni .....	20
distribuzione	
- binomiale .....	31, 164
- del $\chi^2$ .....	274
- di Bernoulli .....	164
- di Cauchy .....	168
- di Poisson .....	88, 165
- di probabilità [ <b>DdP</b> ] .....	50, 187
- - congiunta .....	54, 200
- - discreta .....	164
- - finito-dimensionale .....	200
- - marginale .....	55, 200
- ergodica .....	378
- esponenziale .....	167
- esponenziale bilatera .....	296

- geometrica .....	389
- ipergeometrica .....	114
- multinomiale .....	34
- stazionaria .....	373
- normale (Gaussiana) .....	167, 176
- $t$ di Student .....	275
- uniforme .....	163
- uniforme discreta .....	164
Doob J.L. ....	361
durata media di un gioco .....	109
elemento aleatorio .....	198
elementi aleatori indipendenti .....	205
equazioni di Chapman-Kolmogorov .....	372, 375
equiprobabilità .....	13, 27
errore quadratico medio ( <b>eqm</b> ) .....	263
estremo superiore essenziale .....	287
eventi	
- disgiunti (incompatibili) .....	23
- elementari .....	19
- condizionatamente indipendenti .....	44
- indipendenti .....	42, 43
evento .....	133
Fatou P. (1878-1929) .....	218
Fermat P. (1608-1665) .....	16
filtrazione .....	351
- naturale .....	353
formula	
- di Bayes .....	41
- di inversione delle funzioni caratteristiche .....	303
- di Stirling .....	34
- di Taylor .....	298
Fourier J.-B. (1768-1830) .....	294
Fubini Ghiron G. (1879-1943) .....	234
funzione	
- aleatoria .....	199
- caratteristica .....	293
- di Borel .....	187
- di Cantor .....	171
- di densità [ <b>fdd</b> ] .....	165, 176, 188
- - congiunta .....	202
- - condizionata da $\eta = y$ .....	250
- - marginale .....	202
- di distribuzione [ <b>FdD</b> ] .....	52, 154, 174, 187
- - generalizzata [ <b>FdDG</b> ] .....	159, 174

- - congiunta .....	54, 200
- - finito-dimensionale .....	200
- - marginale .....	200
- di Gauss .....	78, 167
- di transizione .....	374
- di variabile aleatoria .....	57, 189, 268
- errore .....	78, 167
Galton F. (1822-1911) .....	266
Gauss C.F. (1777-1855) .....	77
gioco assolutamente equo .....	355
Gosset W.S. (1876-1937) .....	275
Heine H.E. (1821-1881) .....	157
Hilbert D. (1862-1943) .....	288
Hölder O.L. (1859-1937) .....	223
immagine di $\mathbf{P}$ .....	187, 200
indicatore .....	49, 187
indipendenza lineare .....	291
integrale	
- di Lebesgue .....	62, 211, 214
- di Lebesgue-Stieltjes .....	214
Jensen J.L. (1859-1925) .....	222
Kolmogorov A.N. (1903-1987) .....	133
Laplace P.S. (1749-1827) .....	77
Lebesgue H.L. (1875-1941) .....	159
legge dei grandi numeri	
- di J.Bernoulli .....	75
- debole .....	338
- forte (Cantelli) .....	343
- forte (Kolmogorov) .....	345
- per catene di Markov .....	383
leggi di De Morgan .....	23
lemma	
- di Borel-Cantelli .....	340
- di Fatou .....	218
- di Heine-Borel (del ricoprimento) .....	157
L'Hôpital G.F.A. (1661-1704) .....	108
limitatezza uniforme .....	339
Lindeberg J.W. ....	336
Lévy P. (1886-1971) .....	307
Lyapunov A.M. (1857-1918) .....	222
Markov A.A. (1856-1922) .....	363
martingala .....	351, 358

matrice	
- delle covarianze	176, 259
- delle transizioni	368, 371
- stocastica	365
McLaurin C. (1698-1746)	298
metrica	286
Minkowski H. (1864-1909)	224
misura	128
- additiva	128
- - finita	128
- - di probabilità	128
- finita	128
- di Lebesgue	162, 175
- di Lebesgue-Stieltjes	159
- di probabilità	128
- - discreta	163
- - prodotto	205
- - singolare	169
- di Wiener	184
- prodotto	232
- segnata	226
- $\sigma$ -additiva	128
- $\sigma$ -finita	128
misurabilità	50, 95, 187
misure uniformemente $\sigma$ -finite	235
modello di Bernoulli	45
modello di probabilità	133
moltiplicazione, formula di	40
momento	213
momento assoluto	213
Monte Carlo	253, 348, 391
norma	285
$\omega$ -sezione	233
operatore differenza	173
ortogonalità	289
<b>P</b> -quasi ovunque [ <b>P</b> -q.o.]	196
passeggiata aleatoria	32
permutazioni	21
Poisson S.D. (1781-1840)	77
principio degli insiemi appropriati	137
principio di ortogonalità	291

probabilità	
- definizione classica	13, 27
- condizionata	37
- - da una decomposizione	93
- - da una $\sigma$ -algebra	243
- - da un elemento aleatorio	243
- - da $\eta = y$	247
- congiunta	37
- di rovina	104
- di transizione	367, 371
- totale, formula della	39, 93, 95
problema dei momenti	304
processo stocastico [p.s.]	32, 199
prodotto scalare	288
proiezione ortogonale	291
proprietà riproduttiva	273
punto aleatorio	199
punto di crescita	171
Radon J. (1887-1956)	165
Rayleigh J.W.S. (1842-1919)	317
regressione	265
- lineare	267
retta reale estesa	141
rettangolo	142
ripartizioni	20
Schwarz H.A. (1843-1921)	66
semialgebra	126
semianello	126
seminorma	285
$\sigma$ -algebra	133
- di Borel	140
- generata	135, 194
$\sigma$ -algebre indipendenti	206
$\sigma$ -anello	133
sistema ortonormale	289
spazio	
- finito di probabilità	26
- dei campioni (o degli eventi elementari)	19, 133
- delle fasi	363
- di Banach	286
- di Hilbert	288
- di probabilità	133
- misurabile	133
- polacco	180



- probabilizzabile .....	133
- prodotto .....	46, 57, 183
stato .....	363
- assorbente .....	368
Stieltjes T.J. (1856-1894) .....	159
stima	
- di parametri .....	73, 76, 85
- in media quadratica .....	263
- lineare in media quadratica .....	264, 289
Stirling J. (1692-1770) .....	34
Student .....	(vedi Gosset W.S.)
subadditività .....	26, 129
submartingala .....	351
successione	
- di variabili aleatorie .....	200
- di Cauchy .....	284
- stocastica .....	351
- - adattata .....	351
- - prevedibile .....	351
- - cerescente .....	351
supermartingala .....	351
Taylor B. (1667-1752) .....	298
tempo	
- d'arresto .....	104, 356
- di primo impatto .....	356
- di Markov .....	356
teorema	
- della convergenza monotona .....	218
- della convergenza maggiorata .....	219
- di Bayes .....	40
- di Carathéodory .....	154
- di continuità di P.Lévy .....	307
- di Fubini .....	234
- di Kolmogorov su $\mathbf{R}^\infty$ .....	178
- di Kolmogorov su $\mathbf{R}^T$ .....	182
- di Kolmogorov e Bochner .....	184
- di Lebesgue .....	192
- di Lebesgue-Nikodym .....	171
- di Poisson .....	88, 335
- di Radon-Nikodym .....	165, 226
- di unicità .....	302
- ergodico (Markov) .....	378
- limite centrale .....	333
- limite integrale .....	83

- limite locale .....	81
- sul cambio di variabili di integrazione .....	227
traccia su un insieme .....	136
traiettoria .....	199
trasformata di Fourier .....	294
valore d'attesa [ <b>VdA</b> ] .....	61, 211, 213
- condizionato .....	94
- - da una decomposizione .....	94
- - da un'algebra .....	99
- - da una $\sigma$ -algebra .....	241
- - da un elemento aleatorio .....	242
- - da $\eta = y$ .....	246
- - in senso lato .....	290
variabile aleatoria [ <b>v.a.</b> ] .....	49, 187
- assolutamente continua .....	188
- di Bernoulli .....	59
- binomiale .....	59
- complessa .....	199
- continua .....	188
- discreta .....	187
- degenerare .....	67
- estesa .....	190
- integrabile .....	213
- log-normale .....	270
- normale (Gaussiana) .....	203, 261, 310
- di Poisson .....	90
- semplice .....	61, 188
- standardizzata .....	80
variabili aleatorie	
- identicamente distribuite [ <b>i.d.</b> ] .....	55
- indipendenti .....	56, 205
- indipendenti e identicamente distribuite [ <b>i.i.d.</b> ] .....	57
- linearmente indipendenti .....	291
- non correlate .....	69, 259
variante di un VdA condizionato .....	242
varianza [ <b>Var</b> ] .....	67, 213
- condizionata .....	259
Venn J. (1834-1923) .....	22
versione regolare	
- della probabilità condizionata .....	254
- della distribuzione condizionata .....	254
- della funzione di distribuzione condizionata .....	255
- della funzione di densità condizionata .....	255
vettore aleatorio [ <b>vett.a.</b> ] .....	54, 199

- normale (Gaussiano) .....	203, 261, 310
vettore delle medie .....	176, 261
vita media .....	251
Wiener N. (1894-1964) .....	184
Wolf R. (1816-1893) .....	253