

UNIVERSITÀ DEL SALENTO
DIPARTIMENTO DI MATEMATICA E FISICA
“ENNIO DE GIORGI”

Carlo Sempi

Secondo Corso di Probabilità



Quaderno 1/2017
Università del Salento - Coordinamento SIBA

QUADERNI DI MATEMATICA

Una pubblicazione a cura del
DIPARTIMENTO DI MATEMATICA E FISICA “ENNIO DE GIORGI”
UNIVERSITÀ DEL SALENTO

Comitato di Redazione

Angela Albanese
Francesco Catino
Domenico Perrone

I QUADERNI del Dipartimento di Matematica e Fisica “Ennio De Giorgi” della Università del Salento documentano gli aspetti di rilievo dell’attività di ricerca e didattica del Dipartimento. Nei Quaderni sono pubblicati articoli di carattere matematico che siano:

- (1) lavori di rassegna e monografie su argomenti di ricerca;
- (2) testi di seminari di interesse generale, tenuti da docenti o ricercatori del Dipartimento o esterni;
- (3) lavori di specifico interesse didattico.

La pubblicazione dei lavori è soggetta all’approvazione del Comitato di Redazione, che decide tenendo conto del parere di un *referee*, nominato di volta in volta sulla base delle competenze specifiche.

Quaderno 1/2017: e-ISBN 978–88–8305–129–6

Università del Salento - Coordinamento SIBA

Prefazione

Sono qui raccolte, con qualche ampliamento, le lezioni della parte avanzata del corso di Probabilità che ho tenuto agli studenti dell'Università di Lecce, prima, e del Salento, dopo, negli ultimi quarant'anni. La presentazione si è affinata nel corso degli anni e nuovi argomenti sono entrati a farne parte, mentre altri sono stati eliminati del tutto o ridotti in modo considerevole. Avverto gli studenti che gli esercizi costituiscono una parte integrale del corso; essi variano in difficoltà da quelli che sono una mera applicazione di quanto visto a lezione a altri più impegnativi. Mi sono divertito a introdurre alla fine di ogni capitolo cenni di storia della probabilità che spero stimolino a cercare approfondimenti nella letteratura.

Sono grato agli studenti che in questi anni mi hanno posto domande obbligandomi a uno sforzo di chiarezza del quale mi auguro sia rimasta traccia; un grazie particolare agli studenti che mi hanno suggerito migliorie che ho talvolta finito per includere.

Infine vorrei ringraziare di cuore Gianfausto Salvadori per i consigli e le discussioni che mi sono sempre stati preziosi.

Lecce, 25 maggio 2017

Indice

1	Elementi della Teoria della Misura	1
1.1	Tribú	1
1.2	Richiami di probabilità elementare	3
1.3	Probabilità e misura	5
1.4	Successioni di insiemi	7
1.5	Misure	8
1.6	Estensione di misure	12
1.7	Le misure di Borel–Stieltjes	20
1.8	Funzioni semplici e funzioni misurabili	22
1.9	La definizione dell’integrale	28
1.10	Proprietà dell’integrale	31
1.11	Misura immagine	36
1.12	Vocabolario	38
1.13	Gli spazi L^p	39
1.14	Misure definite da una densità	43
1.15	Misura prodotto	47
1.16	Completamento di misure	54
1.17	Appendice	55
1.18	Note al Capitolo 1	57
1.19	Esercizi sul Capitolo 1	61
2	La Convergenza Stocastica	69
2.1	I lemmi di Borel–Cantelli	69
2.2	Varî tipi di convergenza stocastica	73
2.3	La convergenza completa	82
2.4	Le convergenze vaga e stretta	87
2.5	Metriche	95
2.6	Altri tipi di convergenza per v.a.	99
2.7	Note al Capitolo 2	100
2.8	Esercizi sul Capitolo 2	102
3	Funzioni caratteristiche	117
3.1	Definizioni e proprietà elementari	117
3.2	La formula d’inversione	119
3.3	Funzioni caratteristiche notevoli	122
3.4	Funzioni caratteristiche e momenti	127
3.5	Funzioni caratteristiche e indipendenza	130

3.6	Il teorema di continuità	133
3.7	Individuazione delle f.c.	136
3.8	Funzione caratteristica di un vettore aleatorio	140
3.9	Note al Capitolo 3	142
3.10	Esercizi sul Capitolo 3	142
4	Teoremi Limite	153
4.1	TLC: condizioni sufficienti	153
4.2	TLC: condizioni necessarie	161
4.3	LGN: leggi deboli	167
4.4	LGN: leggi forti	170
4.5	Un'applicazione della LGN	180
4.6	Note al Capitolo 4	183
4.7	Esercizi sul Capitolo 4	184
5	Le Speranze Condizionate	195
5.1	La definizione	195
5.2	Leggi condizionate	196
5.3	Proprietà delle speranze condizionate	201
5.4	Distribuzioni condizionate regolari	208
5.5	Note al Capitolo 5	210
5.6	Esercizi sul Capitolo 5	211
6	Introduzione alle Martingale	215
6.1	Definizione ed esempi	215
6.2	Tempi d'arresto	221
6.3	Martingale e tempi d'arresto	224
6.4	Integrabilità uniforme	230
6.5	Convergenza delle martingale	235
6.6	Le martingale rovesciate	244
6.7	Applicazioni	245
6.7.1	Il teorema di Radon–Nikodym.	245
6.7.2	Legge 0–1 di Kolmogorov	247
6.7.3	Convergenza di serie di v.a..	247
6.7.4	Le LGN forti	250
6.7.5	Variabili scambiabili e il teorema di De Finetti	253
6.7.6	La rovina del giocatore	255
6.7.7	L'urna di Pólya	258
6.8	Il teorema di Burkholder	259
6.9	Note al Capitolo 6	264
6.10	Esercizi sul Capitolo 6	266

Capitolo 1

Elementi della Teoria della Misura

Per fissare il quadro nel quale opereremo richiamiamo il concetto di tribú.

1.1 Tribú

Definizione 1.1.1. Dato un insieme non vuoto Ω , si chiama *algebra di sottoinsiemi* di Ω , ogni famiglia $\mathcal{A} \subseteq \mathcal{P}(\Omega)$, non vuota, che sia stabile per la complementazione, per l'unione finita e tale che l'insieme vuoto appartenga ad \mathcal{A} ; cioè:

- (a) $\Omega \in \mathcal{A}$;
- (b) $A \in \mathcal{A} \implies A^c \in \mathcal{A}$;
- (c) $A, B \in \mathcal{A} \implies A \cup B \in \mathcal{A}$. ◇

Naturalmente, se \mathcal{A} è un'algebra (di sottoinsiemi di Ω) e se A e B sono in \mathcal{A} , allora vi appartengono anche $A \cap B$, $A \setminus B$, e $A \Delta B$; infatti

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

Definizione 1.1.2. Si chiama *tribú*, o σ -*algebra*, una famiglia \mathcal{F} di sottoinsiemi di Ω , $\mathcal{F} \subseteq \mathcal{P}(\Omega)$, che goda delle seguenti proprietà:

- (a) $\Omega \in \mathcal{F}$;
- (b) $A \in \mathcal{F} \implies A^c \in \mathcal{F}$;
- (c) $\forall n \in \mathbb{N} \quad A_n \in \mathcal{F} \implies \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F}$. ◇

Una tribú è dunque stabile rispetto all'operazione di unione numerabile. Usando le leggi di de Morgan è immediato dimostrare il seguente

Teorema 1.1.1. *Sia \mathcal{F} una tribú di sottoinsiemi di Ω . Allora*

(a) \mathcal{F} è stabile per le unioni finite:

$$A_i \in \mathcal{F} (i = 1, \dots, n) \implies \bigcup_{i=1}^n A_i \in \mathcal{F};$$

(b) \mathcal{F} è stabile rispetto alle intersezioni numerabili:

$$\forall n \in \mathbb{N} \quad A_n \in \mathcal{F} \implies \bigcap_{n \in \mathbb{N}} A_n \in \mathcal{F};$$

(c) \mathcal{F} è stabile rispetto alle intersezioni finite:

$$A_i \in \mathcal{F} \quad (i = 1, 2, \dots, n) \implies \bigcap_{i=1}^n A_i \in \mathcal{F}.$$

Si osservi che una tribú è anche un'algebra. Però non tutte le algebre sono anche tribú; si veda a tal fine l'Esercizio 1.2.

La classe delle tribú di sottoinsiemi di un insieme non vuoto Ω è ordinata, parzialmente, rispetto all'inclusione e contiene una piú piccola tribú, la *tribú banale*, $\mathcal{N} := \{\emptyset, \Omega\}$ ed una piú grande tribú, che è la famiglia delle parti $\mathcal{P}(\Omega)$, sicché, per ogni tribú \mathcal{F} , si ha $\mathcal{N} \subseteq \mathcal{F} \subseteq \mathcal{P}(\Omega)$.

Sia A un sottoinsieme proprio e non vuoto di Ω , cioè $A \neq \emptyset$ e $A \neq \Omega$; la famiglia $\mathcal{F}(A) := \{\emptyset, A, A^c, \Omega\}$ è un'algebra; è anzi, una tribú, poiché ogni algebra finita è anche una tribú, dato che ogni successione è necessariamente composta da un numero finito di insiemi distinti, sicché ogni unione numerabile è, di fatto, un'unione finita; essa è la piú piccola tribú che contenga A (e si dice *generata da A*). Infatti se \mathcal{G} è una tribú che contiene A , risulta, per definizione,

$$A \in \mathcal{G}, \quad A^c \in \mathcal{G}, \quad \emptyset \in \mathcal{G}, \quad \Omega \in \mathcal{G},$$

onde $\mathcal{F}(A) \subseteq \mathcal{G}$. Il teorema seguente è di dimostrazione banale.

Teorema 1.1.2. *Se $\{\mathcal{F}_\iota : \iota \in I\}$ è un'arbitraria famiglia di tribú di sottoinsiemi di Ω , è una tribú anche $\bigcap_{\iota \in I} \mathcal{F}_\iota$.*

Quest'ultimo risultato consente di risolvere il problema dell'esistenza della piú piccola tribú che contenga un'assegnata famiglia \mathcal{C} di sottoinsiemi di Ω . Tale tribú si indica con $\mathcal{F}(\mathcal{C})$ e si dice *generata da \mathcal{C}* ; essa è eguale all'intersezione di tutte le tribú in $\mathcal{P}(\Omega)$ che contengano \mathcal{C} . Si noti che la famiglia della quale si considera l'intersezione non è vuota perché vi appartiene almeno $\mathcal{P}(\Omega)$. Un esempio fondamentale di tale situazione è fornito dalla famiglia \mathcal{I} degli intervalli aperti della retta reale \mathbb{R} , $\mathcal{I} := \{]a, b[: a, b \in \mathbb{R}, a < b\}$; si osservi che la condizione $a \leq b$, in luogo di quella piú naturale $a < b$, fa sí che l'insieme vuoto \emptyset sia considerato come un particolare intervallo, ciò che è comodo. \mathcal{I} non è un'algebra (e pertanto neanche una tribú), poiché se, ad esempio, $a < b < c < d$, l'unione $]a, b[\cup]c, d[$ non è un intervallo. La tribú generata da \mathcal{I} si chiama *tribú di Borel* e la si denota con $\mathcal{B}(\mathbb{R})$ o, se non sorgono ambiguità, semplicemente con \mathcal{B} ; i suoi insiemi si chiamano *boreliani*. Vale il seguente utile

Teorema 1.1.3. *La tribú di Borel $\mathcal{B}(\mathbb{R})$ è generata da una qualsiasi delle seguenti famiglie:*

- (a) *le semirette del tipo $] - \infty, x]$ ($x \in \mathbb{R}$);*
- (b) *gli insiemi aperti di \mathbb{R} ;*
- (c) *gli insiemi chiusi di \mathbb{R} .*

Dimostrazione. Sia $\mathcal{B} = \mathcal{B}(\mathbb{R})$ e si indichi con \mathcal{B}_1 la tribú generata dalla famiglia indicata in (a). Si osservi che anche gli intervalli aperti a sinistra e chiusi a destra, cioè del tipo $]a, b]$, con a e b in \mathbb{R} , appartengono a \mathcal{B} . Infatti

$$]a, b] = \bigcap_{n \in \mathbb{N}} \left] a, b + \frac{1}{n} \right[.$$

Ora $]x, +\infty[= \bigcup_{n \in \mathbb{N}}]x, x + n]$ che appartiene a \mathcal{B} onde

$$]-\infty, x] =]x, +\infty[^c \in \mathcal{B},$$

e perciò $\mathcal{B}_1 \subseteq \mathcal{B}$. D'altra parte, $]x, +\infty[=]-\infty, x]^c \in \mathcal{B}_1$ onde, se $x < y$, $]x, y] =]-\infty, y] \cap]x, +\infty[\in \mathcal{B}_1$. Infine $]x, y[= \bigcup_{n \in \mathbb{N}}]x, y - 1/n] \in \mathcal{B}_1$. Dunque $\mathcal{B} \subseteq \mathcal{B}_1$ e quindi $\mathcal{B}_1 = \mathcal{B}$.

Si indichi con \mathcal{B}_2 la tribú generata dagli aperti. Poiché l'intervallo $]x, y[$ è esso stesso un aperto, si ha $\mathcal{B} \subseteq \mathcal{B}_2$. Se $A \subseteq \mathbb{R}$ è aperto, esiste, com'è noto, una successione $(]x_n, y_n[)$ di intervalli aperti tale che si possa rappresentare A nella forma $A = \bigcup_{n \in \mathbb{N}}]x_n, y_n[$ onde $A \in \mathcal{B}$ e quindi $\mathcal{B}_2 \subseteq \mathcal{B}$. Che \mathcal{B} sia generata anche dagli insiemi chiusi è ora immediato. \square

Se \mathcal{F} è la tribú generata dalla famiglia \mathcal{C} , non si può, in generale, dare una descrizione costruttiva degli elementi di \mathcal{F} partendo dagli elementi di \mathcal{C} .

Definizione 1.1.3. Si diranno *misurabili* (o, ove vi sia possibilità di confusione, *\mathcal{F} -misurabili*) gli insiemi appartenenti ad una prefissata tribú \mathcal{F} . Si dirà *spazio misurabile* la coppia (Ω, \mathcal{F}) costituita da un insieme non vuoto Ω e da una tribú \mathcal{F} di suoi sottoinsiemi.

1.2 Richiami di probabilità elementare

Richiamo qui concetti e definizioni già noti al lettore del corso introduttivo di probabilità. Alcuni dei concetti richiamati saranno in seguito estesi o generalizzati. La definizione di probabilità è formalizzata nella sezione successiva, Definizione 1.3.1. Dato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, una funzione misurabile $X : \Omega \rightarrow \mathbb{R}$ si dice *variabile aleatoria*. Tale concetto sarà poi esteso al caso di funzioni che assumano anche i valori $\pm\infty$. Si dice che una variabile aleatoria X è *discreta* quando assume un numero finito di valori.

In uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ due insiemi A e B in \mathcal{F} si dicono *indipendenti* se si ha

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

Due variabili aleatorie X e Y definite nello stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ si dicono *indipendenti* se, per ogni scelta di due boreliani A e B , si ha

$$\mathbb{P}(X^{-1}(A) \cap Y^{-1}(B)) = \mathbb{P}(X^{-1}(A)) \mathbb{P}(Y^{-1}(B)).$$

Questi concetti si estendono immediatamente a piú di due insiemi o di due variabili aleatorie.

Ricordiamo che l'indipendenza di insiemi è legata all'indipendenza di tribú.

Date n tribú $\mathcal{F}_1, \dots, \mathcal{F}_n$ contenute in \mathcal{F} , queste si dicono indipendenti se per ogni scelta di n insiemi $A_1 \in \mathcal{F}_1, \dots, A_n \in \mathcal{F}_n$ risulta

$$\mathbb{P}\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n \mathbb{P}(A_i).$$

Due insiemi A e B sono indipendenti se, e solo se, sono indipendenti le tribú $\mathcal{F}(A)$ e $\mathcal{F}(B)$ che essi generano. Per verificare che gli insiemi misurabili A_1, \dots, A_n siano indipendenti occorre allora controllare che risulti

$$\mathbb{P}\left(\bigcap_{i=1}^n \overline{A_i}\right) = \prod_{i=1}^n \mathbb{P}(\overline{A_i}),$$

ove, per ogni indice i , $\overline{A_i}$ può essere uno degli insiemi A_i, A_i^c, Ω o $A_i = \emptyset$.

Gli insiemi misurabili $(A_k)_{k \in \mathbb{N}}$ di una successione si dicono indipendenti se per ogni scelta di $n \geq 2$ sono indipendenti gli insiemi A_1, \dots, A_n .

Siano date n variabili aleatorie X_1, \dots, X_n nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ queste si dicono *indipendenti* se risulta

$$\mathbb{P}\left(\bigcap_{k=1}^n X^{-1}(B_k)\right) = \prod_{k=1}^n \mathbb{P}(X^{-1}(B_k))$$

per ogni scelta di n boreliani B_1, \dots, B_n .

Infine le variabili aleatorie $(X_k)_{k \in \mathbb{N}}$ di una successione si dicono indipendenti se sono indipendenti X_1, \dots, X_n per ogni scelta di $n \geq 2$.

Le variabili aleatorie discrete piú semplici sono quelle di Bernoulli: Y è una variabile aleatoria di Bernoulli se assume due soli valori, convenzionalmente chiamati *successo* e *fallimento* e indicati con s oppure 1 e f oppure 0; così $\mathbb{P}(Y = 1) = p$ e $\mathbb{P}(Y = 0) = q = 1 - p$. Il processo di Bernoulli, già incontrato nel primo corso di probabilità, è costituito da una successione (X_n) di variabili aleatorie indipendenti tutte con la stessa legge di Bernoulli, $\mathbb{P}(X_n = 1) = p$ e $\mathbb{P}(X_n = 0) = q = 1 - p$.

Dato un insieme $A \subset \Omega$ la sua *funzione indicatrice* $\mathbf{1}_A$ è così definita

$$\mathbf{1}_A(\omega) := \begin{cases} 1 & \omega \in A, \\ 0, & \omega \notin A. \end{cases}$$

Nel seguito si indicherà il limite a sinistra in $x_0 \in \mathbb{R}$ di una funzione $f : \mathbb{R} \rightarrow \mathbb{R}$, se esiste, mediante la notazione

$$\ell^- f(x_0) := \lim_{\substack{x \rightarrow x_0 \\ x < x_0}} f(x).$$

1.3 Probabilità e misura

Ricordiamo la definizione di probabilità già nota dal corso introduttivo.

Definizione 1.3.1. Dato uno spazio misurabile (Ω, \mathcal{F}) — vale a dire un insieme non vuoto Ω ed una tribú \mathcal{F} di suoi sottoinsiemi — si dice (*misura di*) *probabilità* su (Ω, \mathcal{F}) ogni funzione $\mathbb{P} : \mathcal{F} \rightarrow \mathbb{R}$ che soddisfaccia alle seguenti condizioni:

$$(P.1) \quad \mathbb{P}(A) \geq 0 \text{ per ogni insieme } A \in \mathcal{F};$$

$$(P.2) \quad \mathbb{P}(\Omega) = 1;$$

(P.3) per ogni successione $(A_n)_{n \in \mathbb{N}}$ di insiemi misurabili disgiunti ($A_n \in \mathcal{F}$, per ogni $n \in \mathbb{N}$, con $A_j \cap A_k = \emptyset$ ($j \neq k$)), vale la proprietà:

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n).$$

detta di *additività numerabile* o σ -*additività*. ◇

Anche nel caso discreto, a ben guardare, nasce la necessità di studiare situazioni piú complesse. Infatti, nello studio del processo di Bernoulli, viene naturale considerare spazi dei risultati Ω che non sono finiti o numerabili: cosí una generica “storia” può essere rappresentata mediante una successione

$$(x_1, \dots, x_n, \dots)$$

ove, per ogni $n \in \mathbb{N}$, $x_n = 0$ oppure $x_n = 1$; ma è noto che tali successioni possono essere poste in corrispondenza biunivoca con i numeri dell'intervallo $[0, 1]$, mediante

$$x = \sum_{n \in \mathbb{N}} \frac{x_n}{2^n},$$

se si adotta la convenzione che i numeri diadici siano rappresentati dalla successione che ha infiniti zeri. È inoltre noto che l'intervallo $[0, 1]$ non è numerabile.

Quando, nello studio della passeggiata aleatoria di Bernoulli, abbiamo considerato lo spazio di tutte le possibili storie, $\Omega := \{0, 1\}^{\mathbb{N}}$, abbiamo lasciato in sospeso la questione di quale sia la tribú da adottare per definire le probabilità atte a rispondere alle domande che ci ponemmo allora. Si consideri, per ogni $n \in \mathbb{N}$ il “segmento iniziale” $\Omega_n := \{0, 1\}^n$, ed, in questo si consideri la famiglia delle parti, $\mathcal{F}'_n := \mathcal{P}(\Omega_n)$. Si consideri ora la famiglia di sottoinsiemi di $\{0, 1\}^{\mathbb{N}}$ data da

$$\mathcal{F}_n := \mathcal{F}'_n \times \prod_{k=n}^{\infty} \{0, 1\};$$

evidentemente, la famiglia \mathcal{F}_n è adeguata per affrontare tutti i problemi che riguardano questioni sul processo di Bernoulli sino al tempo $t = n$. Si può quindi considerare la famiglia di sottoinsiemi di Ω definita da

$$\bigcup_{n \in \mathbb{N}} \mathcal{F}_n.$$

Questa, tuttavia non è una tribú; sarà, perciò naturale porci nella tribú generata da tale famiglia

$$\bigvee_{n \in \mathbb{N}} \mathcal{F}_n := \mathcal{F} \left(\bigcup_{n \in \mathbb{N}} \mathcal{F}_n \right).$$

In generale, si pone dunque il problema di definire le probabilità senza porre restrizioni sulla cardinalità di Ω . Storicamente, si sono presentate due strade che, entrambe, delinearemo, seppur succintamente.

La prima strada consiste nel partire, dato un insieme non vuoto Ω , da un'algebra \mathcal{A} di suoi sottoinsiemi; è naturale considerare un'algebra poiché, come si è visto, questa è stabile rispetto alle operazioni sugli insiemi, purché eseguite in numero finito. Una funzione $\mu : \mathcal{A} \rightarrow [0, +\infty]$ si dice *finitamente additiva* se

$$\mu \left(A \cup B \right) = \mu(A) + \mu(B), \quad (1.3.1)$$

per ogni coppia di insiemi disgiunti A e B di \mathcal{A} , $A \cap B = \emptyset$.

In particolare, una probabilità $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ è una funzione è finitamente additiva se obbedisce alla (1.3.1), che prende il nome di additività semplice e che si scrive

$$\mathbb{P} \left(A \cup B \right) = \mathbb{P}(A) + \mathbb{P}(B) \quad (A, B \in \mathcal{A}, A \cap B = \emptyset).$$

Vale il seguente

Teorema 1.3.1. (Tarski). *Ogni misura finitamente additiva e finita μ definita su un'algebra \mathcal{A} di sottoinsiemi di Ω può essere estesa a $\mathcal{P}(\Omega)$; vale a dire che esiste una misura finitamente additiva e finita $\nu : \mathcal{P}(\Omega) \rightarrow [0, 1]$ tale che risulti $\nu(A) = \mu(A)$ per ogni insieme A di \mathcal{A} .*

La dimostrazione si può dare agevolmente (ma non è la dimostrazione originale di Tarski) mediante il teorema di Hahn–Banach; esso presenta però l'inconveniente che l'estensione ν di μ non è, solitamente, unica. Di contro, nel caso delle probabilità, non si pongono restrizioni ai sottoinsiemi di Ω che possono essere riguardati come eventi, ai quali, cioè, si può attribuire una probabilità.

Vi è un secondo modo di procedere, che consiste nel richiedere che la funzione \mathbb{P} soddisfaccia, anziché all'assioma di additività semplice, a quello più forte di *additività numerabile* (o σ -*additività*), che abbiamo già incontrato.

Naturalmente, occorre che la famiglia di sottoinsiemi sulla quale è definita la probabilità sia tale che, se vi sono contenuti tutti gli insiemi di una successione di insiemi, vi appartenga anche la loro unione. Perciò è naturale definire la probabilità, in questa seconda accezione, in una tribú \mathcal{F} di sottoinsiemi di Ω .

Poiché adotteremo, come abbiamo già fatto, e come è tradizione, questo secondo punto di vista, la proprietà di additività numerabile rende opportuna una digressione attraverso le misure, dato che le probabilità sono particolari misure. Tuttavia, mettiamo in guardia il lettore dal pensare che la probabilità sia un mero esempio di misura con la restrizione che sia eguale a 1 la misura di tutto lo spazio, o come spesso si dice, sia una misura normalizzata. Bastano i soli concetti di indipendenza e di condizionamento per mostrare la ricchezza della Probabilità.

1.4 Successioni di insiemi

Le definizioni che seguono, e che non dipendono da una topologia, sono giustificate, *a posteriori*, dal legame con i concetti dello stesso nome per le funzioni a valori reali (si vedano gli esercizi). Per aiutare l'intuizione, si tenga presente che le definizioni ricalcano quelle per le successioni di numeri reali, se si fa corrispondere all'estremo superiore e all'estremo inferiore tra numeri reali l'unione e l'intersezione, rispettivamente, tra insiemi. Naturalmente, quest'analogia serve solo ad intuire i risultati, ma non a dimostrarli.

Definizione 1.4.1. Data una successione $(A_n)_{n \in \mathbb{N}}$ di sottoinsiemi di Ω l'insieme

$$\limsup_{n \rightarrow \infty} A_n := \bigcap_{n \in \mathbb{N}} \bigcup_{j \geq n} A_j$$

si chiama *limite superiore* o *massimo limite* di (A_n) mentre l'insieme

$$\liminf_{n \rightarrow \infty} A_n := \bigcup_{n \in \mathbb{N}} \bigcap_{j \geq n} A_j$$

si dice *limite inferiore* o *minimo limite* di (A_n) . \diamond

Gli insiemi $\limsup_{n \rightarrow \infty} A_n$ e $\liminf_{n \rightarrow \infty} A_n$ che si ottengono mediante operazioni numerabili sono misurabili se tale è ogni insieme della successione (A_n) .

L'insieme $\limsup_{n \rightarrow \infty} A_n$ è costituito da tutti, e soli, i punti di Ω che appartengono ad infiniti insiemi della successione (A_n) ; dire, infatti, che il punto ω appartiene a $\limsup_{n \rightarrow \infty} A_n = \bigcap_{n \in \mathbb{N}} \bigcup_{j \geq n} A_j$ equivale a dire che, per ogni numero naturale n , esiste un altro naturale $j \geq n$ tale che $\omega \in A_j$.

Analogamente, l'insieme $\liminf_{n \rightarrow \infty} A_n$ è costituito da tutti, e soli, i punti di Ω che appartengono definitivamente alla successione, tali cioè da appartenere a tutti gli insiemi della successione $\{A_n\}$ tranne, al più, ad un numero finito di essi. Infatti, dire che ω è nell'insieme $\liminf_{n \rightarrow \infty} A_n = \bigcup_{n \in \mathbb{N}} \bigcap_{j \geq n} A_j$ equivale a dire che esiste un naturale n tale che $\omega \in A_j$ per ogni $j \geq n$.

Valgono le seguenti inclusioni

$$\bigcap_{n \in \mathbb{N}} A_n \subseteq \liminf_{n \rightarrow +\infty} A_n \subseteq \limsup_{n \rightarrow +\infty} A_n \subseteq \bigcup_{n \in \mathbb{N}} A_n. \quad (1.4.1)$$

L'inclusione centrale segue da quanto precede. Per stabilire la (1.4.1), basta perciò mostrare una delle rimanenti inclusioni, per esempio la prima, l'altra dimostrandosi per passaggio al complementare.

Definizione 1.4.2. Si dirà che la successione (A_n) di sottoinsiemi di Ω *converge* se si ha

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

In tal caso, si scrive $\lim_{n \rightarrow +\infty} A_n = A$ (oppure $A_n \rightarrow A$). Una successione (A_n) si dirà *crescente* se è $A_n \subseteq A_{n+1}$ per ogni $n \in \mathbb{N}$, *decescente* se è $A_n \supseteq A_{n+1}$ per ogni $n \in \mathbb{N}$, *monotona* se è o crescente o decrescente. \diamond

Il teorema che segue è l'analogo del teorema sulle successioni reali monotone.

Teorema 1.4.1. *Ogni successione (A_n) monotona converge e risulta:*

(a) $\lim_{n \rightarrow +\infty} A_n = \bigcap_n A_n$ se (A_n) è decrescente;

(b) $\lim_{n \rightarrow +\infty} A_n = \bigcup_n A_n$ se (A_n) è crescente.

Dimostrazione. Basta dimostrare una sola delle due eguaglianze, l'altra ottenendosi per complementazione. Si supponga, dunque, che (A_n) sia crescente; allora

$$\begin{aligned} \limsup_{n \rightarrow +\infty} A_n &= \bigcap_{n \in \mathbb{N}} \bigcup_{j \geq n} A_j = \bigcap_{n \in \mathbb{N}} \bigcup_{j \in \mathbb{N}} A_j = \bigcup_{j \in \mathbb{N}} A_j, \\ \liminf_{n \rightarrow +\infty} A_n &= \bigcup_{n \in \mathbb{N}} \bigcap_{j \geq n} A_j = \bigcup_{n \in \mathbb{N}} A_n, \end{aligned}$$

che dimostra l'asserto. □

1.5 Misure

Diamo qui una piú ampia definizione di misura che consentirà per esempio lo studio delle estensioni di una misura.

Definizione 1.5.1. Sia \mathcal{C} una famiglia di sottoinsiemi di un insieme non vuoto Ω . Si dice *misura* su \mathcal{C} una funzione $\mu : \mathcal{C} \rightarrow \overline{\mathbb{R}}_+$ tale che:

- (a) non sia identicamente eguale a $+\infty$ (esiste cioè un insieme $A \in \mathcal{C}$ con $\mu(A) < +\infty$);
- (b) μ sia numerabilmente additiva, nel senso che, se (A_n) è una successione di insiemi disgiunti di \mathcal{C} , l'unione dei quali è ancora in \mathcal{C} ,

$$\forall n \in \mathbb{N} \quad A_n \in \mathcal{C} \quad A_i \cap A_j = \emptyset \quad (i \neq j) \quad \text{e} \quad \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{C},$$

si ha

$$\mu \left(\bigcup_{n \in \mathbb{N}} A_n \right) = \sum_{n \in \mathbb{N}} \mu(A_n). \quad (1.5.1)$$

Di solito, \mathcal{C} è un'algebra o una tribú; in quest'ultimo caso non occorre postulare che l'unione $\bigcup_{n \in \mathbb{N}} A_n$ appartenga a \mathcal{C} . ◇

Nel parlare di misure è piú comodo e naturale porsi, anziché in \mathbb{R} o in \mathbb{R}_+ , in $\overline{\mathbb{R}} := [-\infty, +\infty]$. Si noti che $\overline{\mathbb{R}}$ non è un gruppo giacché, per ogni reale x , è $x + \infty = +\infty$ e $x - \infty = -\infty$; non è definita l'operazione $\infty - \infty$.

Valgono, inoltre, le convenzioni $0 \cdot (\pm\infty) = 0$ e

$$\begin{aligned} \forall x \in \mathbb{R} \quad & -\infty < x < +\infty, \\ \forall x > 0 \quad & x \cdot (+\infty) = +\infty, \quad \text{e} \quad x \cdot (-\infty) = -\infty, \\ \forall x < 0 \quad & x \cdot (+\infty) = -\infty, \quad \text{e} \quad x \cdot (-\infty) = +\infty, \\ & (\pm\infty) \cdot (\pm\infty) = +\infty, \quad (\mp\infty) \cdot (\pm\infty) = -\infty. \end{aligned}$$

Se l'insieme vuoto \emptyset appartiene a \mathcal{C} , si ha $\mu(\emptyset) = 0$. Infatti sia $A \in \mathcal{C}$ tale che $\mu(A) < +\infty$ e si consideri la successione (A_n) di elementi di \mathcal{C} così definita: $A_1 := A$, $A_n = \emptyset$ per ogni $n \geq 2$; allora la (1.5.1) dà

$$\mu(A) = \mu(A) + \mu(\emptyset) + \cdots + \mu(\emptyset) + \cdots$$

cioè $\mu(\emptyset) = 0$. Similmente, si prova che μ è finitamente additiva. Se gli insiemi A , B e $A \cup B$ appartengono a \mathcal{C} , basta considerare la successione (A_n) definita da $A_1 := A$, $A_2 := B$, $A_n := \emptyset$ per $n \geq 3$. Dall'additività semplice scende anche che una misura è isotona: se $A \subseteq B$ e se gli insiemi A , B e $B \setminus A$ sono in \mathcal{C} , allora $\mu(A) \leq \mu(B)$.

Definizione 1.5.2. Dato un insieme non vuoto Ω , si dice *anello* una famiglia \mathcal{R} di sottoinsiemi di Ω che sia stabile rispetto all'intersezione ed alla differenza finita

$$A \cap B \in \mathcal{R} \quad \text{e} \quad A \Delta B \in \mathcal{R},$$

per ogni coppia A e B di insiemi di \mathcal{R} . ◇

Questa definizione assicura che il risultato di tutte le operazioni eseguite sopra un numero finito di insiemi dell'anello \mathcal{R} appartiene ancora ad \mathcal{R} (si vedano, a tal fine, gli esercizi). Si deve eccettuare la complementazione, a meno che Ω non appartenga a \mathcal{R} .

Prima di enunciare il prossimo risultato, che dà un criterio utile, spesso utilizzato, per stabilire per la σ -additività, conviene introdurre la definizione di funzione finitamente additiva.

La proprietà espressa dalla (1.3.1) vale, in particolare, per le probabilità. Dalla (1.3.1) scende, come immediata conseguenza, l'isotonia della funzione d'insieme μ , vale a dire che, se A e B sono due insiemi di \mathcal{R} con $A \subseteq B$, allora $\mu(A) \leq \mu(B)$.

Teorema 1.5.1. *Sia \mathcal{R} un anello di sottoinsiemi di Ω e sia $\mu : \mathcal{R} \rightarrow \overline{\mathbb{R}}_+$ una funzione d'insieme finitamente additiva; allora*

- (a) *se μ è σ -additiva, essa passa al limite lungo tutte le successioni crescenti di insiemi di \mathcal{R} ; in altre parole, se (A_n) è una successione di insiemi di \mathcal{R} con*

$$A_n \subseteq A_{n+1} \quad \text{e} \quad \bigcup_{n \in \mathbb{N}} A_n = A \in \mathcal{R},$$

allora $\lim_{n \rightarrow +\infty} \mu(A_n) = \mu(A)$;

- (b) *se μ è σ -additiva, essa passa al limite lungo tutte le successioni decrescenti di insiemi di \mathcal{R} per le quali esista un indice k tale che $\mu(A_k) < +\infty$; in altre parole, se (A_n) è una successione di insiemi di \mathcal{R} con*

$$A_n \supset A_{n+1}, \quad \bigcap_{n \in \mathbb{N}} A_n = A \in \mathcal{R}$$

ed esiste almeno un indice $k \in \mathbb{N}$ per il quale $\mu(A_k) < +\infty$, allora

$$\lim_{n \rightarrow +\infty} \mu(A_n) = \mu(A);$$

- (c) se μ passa al limite lungo tutte le successioni crescenti di insiemi di \mathcal{R} , cioè, se per ogni successione (A_n) di insiemi di \mathcal{R} con $A_n \subseteq A_{n+1}$ per ogni $n \in \mathbb{N}$, e $\cup_{n \in \mathbb{N}} A_n = A \in \mathcal{R}$, è $\lim_{n \rightarrow +\infty} \mu(A_n) = \mu(A)$, allora μ è σ -additiva;
- (d) se μ passa al limite lungo tutte le successioni decrescenti di insiemi di \mathcal{R} che tendono all'insieme vuoto \emptyset , cioè, se per ogni successione (A_n) di insiemi di \mathcal{R} con $A_n \supseteq A_{n+1}$ e $\cap_{n \in \mathbb{N}} A_n = \emptyset$, è $\lim_{n \rightarrow +\infty} \mu(A_n) = 0$, allora μ è σ -additiva.

Dimostrazione. (a) Si supponga che $A_n \uparrow A = \cup_{n \in \mathbb{N}} A_n$. Se esiste un indice k tale che $\mu(A_k) = +\infty$, si ha $\mu(A_n) \geq \mu(A_k) = +\infty$ per ogni $n \geq k$, sicché $\mu(A) = +\infty$ e $\mu(A_n) \rightarrow \mu(A)$. Possiamo quindi supporre che sia $\mu(A_n) < +\infty$ per ogni $n \in \mathbb{N}$. Si scriva A come unione disgiunta,

$$A = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \cdots = A_1 \cup \left(\bigcup_{n \in \mathbb{N}} (A_{n+1} \setminus A_n) \right),$$

onde, grazie alla σ -additività,

$$\begin{aligned} \mu(A) &= \mu(A_1) + \sum_{n \in \mathbb{N}} \mu(A_{n+1} \setminus A_n) \\ &= \mu(A_1) + \lim_{k \rightarrow +\infty} \sum_{n=1}^k \mu(A_{n+1} \setminus A_n) \\ &= \lim_{k \rightarrow +\infty} \left(\mu(A_1) + \sum_{n=1}^k \{ \mu(A_{n+1}) - \mu(A_n) \} \right) = \lim_{k \rightarrow +\infty} \mu(A_{k+1}). \end{aligned}$$

(b) Si supponga che esista un indice k tale che $\mu(A_k) < +\infty$ e si ponga, per ogni $n \geq k$, $C_n := A_k \setminus A_n$. La successione $(C_n)_{n \geq k}$ è tutta costituita da insiemi di \mathcal{R} , è crescente ed ammette come limite l'insieme $A_k \setminus A$; per quanto dimostrato in (a), si ha $\mu(C_n) \rightarrow \mu(A_k \setminus A) = \mu(A_k) - \mu(A)$. Poiché $\mu(C_n) = \mu(A_k) - \mu(A_n)$, segue che $\mu(A_n) \rightarrow \mu(A)$.

(c) Sia (A_n) una successione di insiemi disgiunti di \mathcal{R} e si supponga che anche l'unione $A = \cup_{n \in \mathbb{N}} A_n$ appartenga a \mathcal{R} . Posto $E_n := \cup_{j=1}^n A_j$, si ha che gli insiemi E_n appartengono a \mathcal{R} per ogni $n \in \mathbb{N}$ e che la successione (E_n) è crescente e tende a A , sicché

$$\sum_{j=1}^n \mu(A_j) = \mu(E_n) \xrightarrow{n \rightarrow +\infty} \mu(A) \quad \text{cioè} \quad \sum_{n \in \mathbb{N}} \mu(A_n) = \mu(A).$$

(d) Con la stessa notazione di (c), si ponga $F_n := A \setminus E_n = \cup_{j=n+1}^{\infty} A_j$; allora F_n appartiene a \mathcal{R} per ogni $n \in \mathbb{N}$, la successione (F_n) è decrescente e $\cap_{n \in \mathbb{N}} F_n = \emptyset$. Scende dall'ipotesi fatta che $\mu(F_n) \downarrow 0$, e, poiché $\mu(A) = \sum_{j=1}^n \mu(A_j) + \mu(F_n)$, si ha

$$\mu(A) = \sum_{j \in \mathbb{N}} \mu(A_j),$$

che conclude la dimostrazione. \square

È immediato il corollario

Corollario 1.5.1. *Per una funzione $\mu : \mathcal{R} \rightarrow \mathbb{R}_+$, finita e finitamente additiva, sono equivalenti le condizioni*

- (a) μ è σ -additiva;
- (b) $\forall n \in \mathbb{N} \quad A_n \in \mathcal{R}, A_n \uparrow A \in \mathcal{R} \implies \mu(A_n) \uparrow \mu(A)$;
- (c) $\forall n \in \mathbb{N} \quad A_n \in \mathcal{R}, A_n \downarrow A \in \mathcal{R} \implies \mu(A_n) \downarrow \mu(A)$.

Nel seguito chiameremo *spazio misurale* ogni terna $(\Omega, \mathcal{F}, \mu)$ formata da un insieme non vuoto Ω , da una tribú \mathcal{F} di sottoinsiemi di Ω e da una misura (σ -additiva) μ definita in \mathcal{F} . Se poi risulta $\mu(\Omega) = 1$, μ si dirà (*misura di*) *probabilità* e la terna $(\Omega, \mathcal{F}, \mu)$ si dirà *spazio di probabilità*; in tal caso, si usa solitamente \mathbb{P} in luogo di μ per indicare la probabilità.

È bene sottolineare che una probabilità \mathbb{P} è definita in una tribú \mathcal{F} che, in genere, differisce da $\mathcal{P}(\Omega)$ e che a \mathbb{P} si richiede di soddisfare all'assioma (P.3) di additività numerabile. Ci si può domandare se, come per le misure finitamente additive, sia possibile estendere una probabilità \mathbb{P} dalla tribú \mathcal{F} a tutta la famiglia delle parti $\mathcal{P}(\Omega)$; tale possibilità è esclusa, in generale, dal seguente teorema del quale non daremo la dimostrazione.

Teorema 1.5.2. (Ulam). *Se Ω è un insieme non numerabile $\text{card}(\Omega) > \aleph_0$, la sola misura μ definita su $\mathcal{P}(\Omega)$ e che si annulli su tutti i punti di Ω , vale a dire, tale che, per ogni $\omega \in \Omega$, $\mu(\{\omega\}) = 0$, è quella identicamente nulla, $\mu(A) = 0$ per ogni sottoinsieme A di Ω .*

A chi voglia trattare di probabilità si presenta dunque un'alternativa: considerare come eventi tutti i possibili sottoinsiemi di un insieme Ω ed usare probabilità finitamente additive, oppure, rinunciare a considerare come eventi tutti i sottoinsiemi di Ω , restringendo *a priori* la classe degli eventi ad una tribú \mathcal{F} , ma usare probabilità numericamente additive. La strada che tradizionalmente si segue è la seconda, soprattutto in virtù della maggiore ricchezza di risultati che deriva da un assioma piú forte (additività numerabile invece che semplice). Esistono però fenomeni nei quali solo l'uso di probabilità finitamente additive consente di trarre previsioni in accordo con le osservazioni; un esempio importante è costituito dalla cosiddetta legge di Benford, ossia il fatto che nelle serie di numeri la prima cifra significativa non sia distribuita uniformemente (si vedano a tal fine i lavori (Scozzafava, 1981) e (Fuchs & Letta, 1984). Tuttavia, come già detto, in queste lezioni adotteremo sistematicamente gli assiomi di Kolmogorov.

Il seguente esempio non contraddice al Teorema 1.5.2.

Esempio 1.5.1. Sia ω_0 un punto dell'insieme non vuoto Ω . Si dice *misura di Dirac concentrata in ω_0* la probabilità $\delta_{\omega_0} : \mathcal{P}(\Omega) \rightarrow [0, 1]$ definita, per ogni sottoinsieme A di Ω , da $\delta_{\omega_0}(A) := \mathbf{1}_A(\omega_0)$. Si osservi che la condizione del Teorema 1.5.2 è violata nell'unico punto ω_0 . ■

Osserviamo infine che le parti (b) e (c) del corollario 1.5.1 possono essere rafforzate: per una probabilità \mathbb{P} e per ogni successione $(A_n) \subseteq \mathcal{F}$, non necessariamente monotona, convergente a un insieme $A \in \mathcal{F}$, si ha $\mathbb{P}(A_n) \xrightarrow{n \rightarrow +\infty} \mathbb{P}(A)$. (Si vedano gli esercizi).

1.6 Estensione di misure

Ci proponiamo di affrontare in questa sezione il problema, che si pone spesso, di estendere una data misura da una famiglia \mathcal{C} ad una famiglia piú ampia che la contiene.

Si è già incontrata la definizione di anello e quella di algebra di una famiglia di sottoinsiemi di Ω . Introdurremo nel seguito altre famiglie di sottoinsiemi la cui considerazione è utile trattando di probabilità.

Definizione 1.6.1. Una famiglia \mathcal{C} di sottoinsiemi di Ω che sia stabile rispetto all'intersezione finita,

$$\forall A, B \in \mathcal{C} \quad A \cap B \in \mathcal{C}$$

si dice π -classe o π -sistema. ◇

Definizione 1.6.2. Si dice *semi-anello* una famiglia \mathcal{S} di sottoinsiemi di Ω che contenga l'insieme vuoto, che sia una π -classe e tale che per ogni coppia di insiemi A e B in \mathcal{S} esista una partizione finita di $A \setminus B$ in insiemi di \mathcal{S} :

- (a) $\emptyset \in \mathcal{S}$;
- (b) $A, B \in \mathcal{S} \implies A \cap B \in \mathcal{S}$;
- (c) per ogni coppia A e B di insiemi in \mathcal{S} , si può esprimere la loro differenza $A \setminus B$ come unione disgiunta di una famiglia finita di insiemi di \mathcal{S} . ◇

Un esempio notevole di semi-anello è dato, in \mathbb{R} , dalla famiglia \mathcal{I} degli intervalli $]a, b]$ aperti a sinistra e chiusi a destra; in \mathbb{R}^n è un semi-anello la famiglia \mathcal{I}^n dei rettangoli

$$R =]a_1, b_1] \times]a_2, b_2] \times \cdots \times]a_n, b_n] .$$

Definizione 1.6.3. Una famiglia non vuota di sottoinsiemi di Ω si dice *classe monotona* se è stabile rispetto ai limiti di successioni monotone. ◇

È evidente che una tribú è una classe monotona; viceversa, ogni algebra \mathcal{A} che sia anche una classe monotona è una tribú; infatti, se $A_n \in \mathcal{A}$ per ogni $n \in \mathbb{N}$, allora, posto $B_n := \cup_{j=1}^n A_j \in \mathcal{A}$, si ha $B_n \uparrow \cup_{n \in \mathbb{N}} A_n$. Un anello che sia anche una classe monotona non è necessariamente una tribú perché potrebbe non appartenervi l'insieme Ω .

Si dimostra facilmente che, data un'arbitraria famiglia non vuota \mathcal{C} di sottoinsiemi di Ω , esistono un piú piccolo anello ed una piú piccola classe monotona che contengono \mathcal{C} ; li si denotano rispettivamente con $\mathcal{R}(\mathcal{C})$ e con $m(\mathcal{C})$ e si chiamano rispettivamente *anello* e *classe monotona generati da \mathcal{C}* .

Teorema 1.6.1. (della classe monotona). *La classe monotona $m(\mathcal{A})$ e la tribú $\mathcal{F}(\mathcal{A})$ generate da un'algebra \mathcal{A} coincidono ($m(\mathcal{A}) = \mathcal{F}(\mathcal{A})$).*

Dimostrazione. L'inclusione $m(\mathcal{A}) \subseteq \mathcal{F}(\mathcal{A})$ è ovvia. Per dimostrare l'inclusione inversa, basta far vedere che $m(\mathcal{A})$ è una tribú e questo, a sua volta, scenderà dallo stabilire che $m(\mathcal{A})$ è un'algebra. A tale scopo, per ogni $B \subseteq \Omega$, si ponga

$$\mathcal{D}(B) := \left\{ A \subseteq \Omega : A \setminus B, B \setminus A, A \cup B \in m(\mathcal{A}) \right\}.$$

Se (A_n) è una successione decrescente di elementi di $\mathcal{D}(B)$ e $A = \bigcap_{n \in \mathbb{N}} A_n$, allora $(A_n \setminus B)$, $(B \setminus A_n)$, $(A_n \cup B)$ sono successioni monotone in $m(\mathcal{A})$ sicché i limiti $A \setminus B$, $B \setminus A$, $A \cup B$ appartengono a $m(\mathcal{A})$ onde A appartiene a $\mathcal{D}(B)$. Similmente si procede per una successione crescente. Perciò, $\mathcal{D}(B)$ è una classe monotona per ogni $B \subseteq \Omega$. Inoltre, se B è in \mathcal{A} , si ha $\mathcal{D}(B) \supset \mathcal{A}$. Poiché A appartiene a $\mathcal{D}(B)$ se, e solo se, B appartiene a $\mathcal{D}(A)$, necessariamente si ha $\mathcal{D}(B) \supset m(\mathcal{A})$ se B è in $m(\mathcal{A})$. Ciò significa che, se A e B sono in $m(\mathcal{A})$ vi sono anche $A \setminus B$, $B \setminus A$, $A \cup B$, sicché $m(\mathcal{A})$ è un anello; è, anzi, un'algebra perché Ω appartiene a \mathcal{A} e quindi a $m(\mathcal{A})$. \square

Il seguente risultato è talvolta utile perché caratterizza gli insiemi di un anello generato da un semi-anello.

Teorema 1.6.2. *L'anello $\mathcal{R}(\mathcal{S})$ generato da un semi-anello \mathcal{S} è costituito da tutti, e soli, gli insiemi che sono un'unione disgiunta e finita di insiemi di \mathcal{S} :*

$$\mathcal{R}(\mathcal{S}) = \bigcup_{n \in \mathbb{N}} \left\{ E = \bigcup_{j=1}^n A_j : A_j \in \mathcal{S}, A_i \cap A_j = \emptyset (i \neq j) \right\}.$$

Dimostrazione. è evidente che $\mathcal{R}(\mathcal{S})$ contiene gli insiemi della forma detta perché un anello è stabile rispetto alle unioni finite. Viceversa, sia

$$\mathcal{V} := \bigcup_{n \in \mathbb{N}} \left\{ E = \bigcup_{j=1}^n A_j : A_j \in \mathcal{S}, A_i \cap A_j = \emptyset (i \neq j) \right\}.$$

Poiché è ovvio che $\mathcal{V} \supset \mathcal{S}$, basta mostrare che \mathcal{V} è un anello. Siano E e F in \mathcal{V} con $E = \bigcup_{i=1}^n A_i$ e $F = \bigcup_{j=1}^m B_j$ e si ponga $C_{ij} := A_i \cap B_j$, ottenendo una famiglia disgiunta di insiemi di \mathcal{S} . Ora, $E \cap F = \bigcup_{i=1}^n \bigcup_{j=1}^m C_{ij}$, sicché \mathcal{V} è una π -classe. Per induzione, scende dalla definizione di semi-anello che

$$\begin{aligned} A_i &= \left(\bigcup_{j=1}^n C_{ij} \right) \cup \left(\bigcup_{k=1}^{r(i)} D_{ik} \right) & (i = 1, 2, \dots, n), \\ B_j &= \left(\bigcup_{i=1}^n C_{ij} \right) \cup \left(\bigcup_{k=1}^{s(j)} G_{jk} \right) & (j = 1, 2, \dots, m), \end{aligned}$$

ove $\{D_{ik} : 1 \leq i \leq n, 1 \leq k \leq r(i)\}$ e $\{G_{jk} : 1 \leq j \leq m, 1 \leq k \leq s(j)\}$ sono famiglie finite di insiemi disgiunti di \mathcal{S} . Perciò

$$E \Delta F = \left(\bigcup_{i=1}^n \bigcup_{k=1}^{r(i)} D_{ik} \right) \cup \left(\bigcup_{j=1}^m \bigcup_{k=1}^{s(j)} G_{jk} \right)$$

che è in \mathcal{V} ; \mathcal{V} è stabile anche rispetto alla differenza simmetrica ed è quindi un anello. \square

Siano date due classi \mathcal{C} e \mathcal{D} di sottoinsiemi di Ω ; se $\mu : \mathcal{C} \rightarrow \overline{\mathbb{R}}_+$ e $\mathcal{C} \subseteq \mathcal{D}$, si dice che la funzione $\nu : \mathcal{D} \rightarrow \overline{\mathbb{R}}_+$ *estende* μ se accade che $\nu(E) = \mu(E)$ per ogni insieme $E \in \mathcal{C}$.

Teorema 1.6.3. *Se \mathcal{S} è un semi-anello e $\mu : \mathcal{S} \rightarrow \mathbb{R}_+$ è finitamente additiva, esiste una sola estensione finitamente additiva ν di μ all'anello $\mathcal{R}(\mathcal{S})$ generato da \mathcal{S} . Se μ è una misura anche ν è una misura.*

Dimostrazione. Per il Teorema 1.6.2, ogni insieme E di $\mathcal{R} = \mathcal{R}(\mathcal{S})$ è della forma $E = \cup_{i=1}^n A_i$ ove $\{A_i\}$ è una partizione di E in \mathcal{S} . Si definisca

$$\nu(E) := \sum_{i=1}^n \mu(A_i). \quad (1.6.1)$$

La (1.6.1) è una buona definizione. Infatti, sia $E = \cup_{j=1}^m B_j$ un'altra decomposizione di E in insiemi disgiunti di \mathcal{S} . Posto $C_{ij} := A_i \cap B_j$, gli insiemi C_{ij} appartengono a \mathcal{S} e sono disgiunti; inoltre $A_i = \cup_{j=1}^m C_{ij}$ ($i = 1, 2, \dots, n$) e $B_j = \cup_{i=1}^n C_{ij}$ ($j = 1, 2, \dots, m$). Poiché μ è additiva in \mathcal{S} , si ha

$$\sum_{i=1}^n \mu(A_i) = \sum_{i=1}^n \sum_{j=1}^m \mu(C_{ij}) = \sum_{j=1}^m \mu(B_j),$$

sicché la (1.6.1) è una buona definizione.

Siano ora E_1 e E_2 due insiemi disgiunti di \mathcal{R} con decomposizioni $E_1 = \cup_{i=1}^n A_i$ e $E_2 = \cup_{j=1}^m B_j$. Una possibile decomposizione di $E_1 \cup E_2$ in insiemi disgiunti di \mathcal{S} è data da

$$E_1 \cup E_2 = \left(\bigcup_{i=1}^n A_i \right) \cup \left(\bigcup_{j=1}^m B_j \right),$$

sicché

$$\nu(E_1 \cup E_2) = \sum_{i=1}^n \mu(A_i) + \sum_{j=1}^m \nu(B_j) = \nu(E_1) + \nu(E_2),$$

vale a dire che ν è finitamente additiva.

Per dimostrare che ν è la sola estensione additiva di μ a \mathcal{R} , si supponga che ne esista un'altra $\tilde{\nu}$; se, come sopra, E è in \mathcal{R} , si ha

$$\begin{aligned} \tilde{\nu}(E) &= \tilde{\nu} \left(\bigcup_{i=1}^n A_i \right) = \sum_{i=1}^n \tilde{\nu}(A_i) \quad (\text{perché } \tilde{\nu} \text{ è additiva}) \\ &= \sum_{i=1}^n \mu(A_i) \quad (\text{perché } \tilde{\nu} \text{ estende } \mu) \\ &= \nu(E), \end{aligned}$$

sicché $\tilde{\nu} = \nu$.

Si supponga ora che μ sia una misura su \mathcal{S} e sia (E_n) una successione disgiunta di insiemi di \mathcal{R} tale che appartenga a \mathcal{R} la sua unione $E = \cup_{n \in \mathbb{N}} E_n$. Si introducano le unioni disgiunte di insiemi di \mathcal{R}

$$E = \bigcup_{r=1}^k A_r, \quad E_n = \bigcup_{i=1}^{k(n)} B_{ni} \quad (n \in \mathbb{N}).$$

Posto $C_{rni} := A_r \cap B_{ni}$ ($n \in \mathbb{N}; r = 1, 2, \dots, k; i = 1, 2, \dots, k(n)$), la famiglia $\{C_{rni}\}$ è composta di insiemi disgiunti di \mathcal{S} ; inoltre,

$$A_r = \bigcup_{n \in \mathbb{N}} \bigcup_{i=1}^{k(n)} C_{rni} \quad \text{e} \quad B_{ni} = \bigcup_{r=1}^k C_{rni}$$

sono decomposizioni in insiemi disgiunti di \mathcal{S} . Poiché μ è σ -additiva, si ha

$$\mu(A_r) = \sum_{n \in \mathbb{N}} \sum_{i=1}^{k(n)} \mu(C_{rni}) \quad \text{e} \quad \mu(B_{ni}) = \sum_{r=1}^k \mu(C_{rni}).$$

L'ordine di somma nelle serie a termini positivi può essere cambiato a piacimento, sicché

$$\begin{aligned} \nu(E) &= \sum_{r=1}^k \mu(A_r) = \sum_{r=1}^k \sum_{n \in \mathbb{N}} \sum_{i=1}^{k(n)} \mu(C_{rni}) \\ &= \sum_{n \in \mathbb{N}} \sum_{i=1}^{k(n)} \sum_{r=1}^k \mu(C_{rni}) = \sum_{n \in \mathbb{N}} \sum_{i=1}^{k(n)} \mu(B_{ni}) = \sum_{n \in \mathbb{N}} \nu(E_n). \end{aligned}$$

Perciò, anche l'estensione ν di μ è σ -additiva. □

Prima di poter enunciare il fondamentale teorema di estensione di Carathéodory occorrono alcune premesse.

Definizione 1.6.4. Dato un insieme non vuoto Ω , ogni funzione $\mu^* : \mathcal{P}(\Omega) \rightarrow \mathbb{R}_+$ tale che:

- (a) $\mu^*(\emptyset) = 0$;
- (b) μ^* sia isotona: se $E \subseteq F$ allora $\mu^*(E) \leq \mu^*(F)$;
- (c) μ^* sia numerabilmente subadditiva:

$$E \subseteq \bigcup_{n \in \mathbb{N}} E_n \implies \mu^*(E) \leq \sum_{n \in \mathbb{N}} \mu^*(E_n)$$

si dice *misura esterna* su Ω . ◇

Teorema 1.6.4. Sia μ^* una misura esterna su Ω e sia \mathcal{M} la famiglia dei sottoinsiemi di Ω definita da

$$\mathcal{M} := \left\{ E \subseteq \Omega : \forall A \subseteq \Omega \quad \mu^*(A) = \mu^*(A \cap E) + \mu^*(A \setminus E) \right\};$$

allora

- (a) \mathcal{M} è una tribú;
- (b) la restrizione μ di μ^* a \mathcal{M} è una misura.

Dimostrazione. (a) Dimostriamo dapprima che \mathcal{M} è stabile rispetto all'unione finita; a tal fine, basta far vedere che anche $E_1 \cup E_2$ appartiene a \mathcal{M} , se vi appartengono E_1 e E_2 . Poiché E_1 è in \mathcal{M} , si ha per ogni $A \subseteq \Omega$,

$$\mu^*(A) = \mu^*(A \cap E_1) + \mu^*(A \setminus E_1); \quad (1.6.2)$$

ma anche E_2 è in \mathcal{M} e dunque

$$\begin{aligned} \mu^*(A \setminus E_1) &= \mu^*[(A \setminus E_1) \cap E_2] + \mu^*[(A \setminus E_1) \setminus E_2] \\ &= \mu^*(A \cap E_1^c \cap E_2) + \mu^*[A \setminus (E_1 \cup E_2)]. \end{aligned} \quad (1.6.3)$$

Ora $(A \cap E_1^c \cap E_2) \cup (A \cap E_1) = A \cap (E_1 \cup E_2)$; perciò, sostituendo la (1.6.3) nella (1.6.2) si ottiene, ricorrendo alla subadditività di μ^* ,

$$\begin{aligned} \mu^*(A) &= \mu^*(A \cap E_1) + \mu^*(A \cap E_1^c \cap E_2) + \mu^*[A \setminus (E_1 \cup E_2)] \\ &\geq \mu^*[A \cap (E_1 \cup E_2)] + \mu^*[A \setminus (E_1 \cup E_2)]. \end{aligned}$$

Poiché la disuguaglianza nell'altro verso vale in virtù della subadditività di μ^* , si ha che $E_1 \cup E_2$ appartiene a \mathcal{M} .

Si vede subito che la stessa relazione che garantisce che E appartiene a \mathcal{M} assicura anche che vi appartenga il suo complementare E^c . È, poi, ovvio che appartiene a \mathcal{M} l'insieme vuoto \emptyset . Dunque \mathcal{M} è un'algebra. Per mostrare che essa è una tribù occorre far vedere che è in \mathcal{M} l'unione di ogni successione (E_n) di insiemi di \mathcal{M} . Basta, anzi, supporre che gli insiemi della successione siano disgiunti, perché avendo già mostrato che \mathcal{M} è un'algebra, si può sostituire ogni unione numerabile con un'unione numerabile disgiunta. Pertanto si supporrà che sia $E_j \cap E_k = \emptyset$ se $j \neq k$. Posto $F_n := \cup_{j=1}^n E_j$, si ha, per ogni $n \in \mathbb{N}$ e per ogni $A \subseteq \Omega$,

$$\mu^*(A \cap F_n) = \sum_{j=1}^n \mu^*(A \cap E_j). \quad (1.6.4)$$

Ciò è ovviamente vero per $n = 1$. Si supponga ora che la (1.6.4) valga per un naturale n . Poiché F_n appartiene a \mathcal{M} , prendendo $A \cap F_{n+1}$ in luogo di A nella definizione di \mathcal{M} , si ha, per $n + 1$,

$$\mu^*(A \cap F_{n+1}) = \mu^*(A \cap F_n) + \mu^*(A \cap E_{n+1}) = \sum_{j=1}^{n+1} \mu^*(A \cap E_j),$$

sicché la (1.6.4) vale per ogni $n \in \mathbb{N}$. Se $E = \cup_{n \in \mathbb{N}} E_n$, scende dall'isotonia di μ^* che

$$\mu^*(A \cap E) \geq \mu^*(A \cap F_n) = \sum_{j=1}^n \mu^*(A \cap E_j),$$

onde

$$\mu^*(A \cap E) \geq \sum_{n \in \mathbb{N}} \mu^*(A \cap E_n).$$

La subadditività di μ^* dà ora

$$\mu^*(A \cap E) = \sum_{n \in \mathbb{N}} \mu^*(A \cap E_n). \quad (1.6.5)$$

Così, per ogni sottoinsieme A di Ω e per ogni $n \in \mathbb{N}$, si ha

$$\mu^*(A) = \mu^*(A \cap F_n) + \mu^*(A \setminus F_n) \geq \sum_{j=1}^n \mu^*(A \cap E_j) + \mu^*(A \setminus E),$$

e, di qui,

$$\mu^*(A) \geq \mu^*(A \cap E) + \mu^*(A \setminus E),$$

sicchè

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \setminus E);$$

perciò E appartiene a \mathcal{M} .

Ponendo $A = \Omega$ nella (1.6.5) si vede che μ^* è σ -additiva in \mathcal{M} . \square

Subito dopo aver introdotto una nuova definizione, saremo in grado di enunciare il teorema d'estensione di Carathéodory.

Definizione 1.6.5. Una misura μ su (Ω, \mathcal{F}) si dice σ -finita se si può esprimere Ω come unione numerabile di insiemi misurabili ciascuno dei quali abbia misura finita, $\Omega = \cup_{n \in \mathbb{N}} E_n$, ove per ogni $n \in \mathbb{N}$, è $E_n \in \mathcal{F}$ e $\mu(E_n) < +\infty$. \diamond

Teorema 1.6.5. (Carathéodory). Sia \mathcal{R} un anello di sottoinsiemi di Ω . Si supponga che Ω si possa esprimere come unione di insiemi di \mathcal{R} , $\Omega = \cup_{n \in \mathbb{N}} E_n$, con $E_n \in \mathcal{R}$ per ogni $n \in \mathbb{N}$ e sia $\mu : \mathcal{R} \rightarrow \overline{\mathbb{R}}_+$ una misura su \mathcal{R} . Esiste allora una misura ν sulla tribú $\mathcal{F}(\mathcal{R})$ generata da \mathcal{R} che estende μ . Se μ è σ -finita, l'estensione ν è unica ed è anch'essa σ -finita.

Dimostrazione. Ogni sottoinsieme E di Ω può essere ricoperto dall'unione di una successione di insiemi di \mathcal{R} ; si può quindi definire

$$\mu^*(E) := \inf \left\{ \sum_{n \in \mathbb{N}} \mu(F_n) \right\},$$

eseguendo l'estremo inferiore su tutte le successioni (F_n) di insiemi di \mathcal{R} tali che $E \subseteq \cup_{n \in \mathbb{N}} F_n$. È evidente che la funzione μ^* cosídefinita su $\mathcal{P}(\Omega)$ è positiva, isotona, e tale che $\mu^*(\emptyset) = 0$. Sia ora $E \subseteq \cup_{n \in \mathbb{N}} E_n$. Se esiste almeno un indice n per il quale $\mu^*(E_n) = +\infty$, si ha banalmente,

$$\mu^*(E) \leq \sum_{n \in \mathbb{N}} \mu^*(E_n).$$

Se, invece, risulta $\mu^*(E_n) < +\infty$ per ogni $n \in \mathbb{N}$, si scelgano in \mathcal{R} , per ogni $\varepsilon > 0$, insiemi F_{nk} con $k \in \mathbb{N}$, in modo che sia, per ogni $n \in \mathbb{N}$,

$$E_n \subseteq \bigcup_{k \in \mathbb{N}} F_{nk} \quad \text{e} \quad \sum_{k \in \mathbb{N}} \mu^*(F_{nk}) < \mu^*(E_n) + \frac{\varepsilon}{2^n}.$$

Segue di qui che μ^* è una misura esterna. Si definisca ora \mathcal{M} come nel Teorema 1.6.4. Dimostriamo, in primo luogo, che \mathcal{M} contiene \mathcal{R} . Se $E \in \mathcal{R}$ e $\mu^*(A) < +\infty$, si scelga una successione (E_n) di insiemi di \mathcal{R} , tale che A sia contenuto nell'unione $\cup_{n \in \mathbb{N}} E_n$ e che

$$\mu^*(A) + \varepsilon \geq \sum_{n \in \mathbb{N}} \mu(E_n);$$

ora

$$\begin{aligned} \mu^*(A) + \varepsilon &\geq \sum_{n \in \mathbb{N}} \left[\mu(E_n \cap E) + \mu(E_n \setminus E) \right] \\ &\geq \mu^*(A \cap E) + \mu^*(A \setminus E), \end{aligned}$$

per la subadditività di μ^* . Se $\mu^*(A) = +\infty$, l'ultima disequaglianza è ovvia. Siccome ε è arbitrario, E appartiene a \mathcal{M} .

Per il Teorema 1.6.4, \mathcal{M} è una tribú, sicché essa include necessariamente la tribú $\mathcal{F}(\mathcal{R})$ generata da \mathcal{R} . Poiché la restrizione di μ^* a \mathcal{M} è una misura, l'ulteriore restrizione ν di μ^* a $\mathcal{F}(\mathcal{R})$ è ancora una misura. Per tutti gli insiemi E di \mathcal{R} si ha $\nu(E) = \mu^*(E) = \mu(E)$, sicché ν è effettivamente un'estensione di μ da \mathcal{R} a $\mathcal{F}(\mathcal{R})$.

Si supponga ora che μ (e quindi μ^*) sia σ -finita su \mathcal{R} ; allora si può scrivere Ω nella forma $\Omega = \cup_{n \in \mathbb{N}} E_n$ con (E_n) successione crescente di insiemi di \mathcal{R} e $\mu(E_n) < \infty$ per ogni $n \in \mathbb{N}$. Per dimostrare l'unicità dell'estensione, sia λ una qualsiasi estensione di μ a $\mathcal{F}(\mathcal{R})$ e si definisca, per $n \in \mathbb{N}$, $\mathcal{E}_n := \{E \in \mathcal{F}(\mathcal{R}) : E \subseteq E_n, \lambda(E) = \mu(E)\}$. È immediato verificare che \mathcal{E}_n è un'algebra; di piú, poiché le misure finite passano al limite lungo le successioni monotone, \mathcal{E}_n è una classe monotona, e quindi, in virtù del Teorema 1.6.1, $\mathcal{E}_n \supset \mathcal{F}(\mathcal{R}_n)$ ove $\mathcal{R}_n := \mathcal{R} \cap E_n = \{E \in \mathcal{R} : E \subseteq E_n\}$. Dunque, per ogni $n \in \mathbb{N}$, l'estensione di μ da \mathcal{R}_n a $\mathcal{F}(\mathcal{R}_n)$ è unica. Ma per ogni $E \in \mathcal{F}(\mathcal{R})$ si ha $(E \cap E_n) \uparrow E$, sicché applicando il Teorema 1.5.1 si ha l'unicità dell'estensione. \square

Nella dimostrazione dell'unicità nel teorema di Carathéodory avremmo potuto usare anche il successivo Teorema 1.6.7 che introduciamo perché vi faremo ricorso, sia pure tacitamente, nel seguito.

Definizione 1.6.6. Una famiglia \mathcal{L} di sottoinsiemi di Ω si dice essere un λ -sistema se sono verificate le condizioni:

- (a) Ω appartiene a \mathcal{L} ;
- (b) $A, B \in \mathcal{L}, A \subseteq B \implies B \setminus A \in \mathcal{L}$;
- (c) $A_n \in \mathcal{L} (n \in \mathbb{N}), A_n \uparrow A \implies A \in \mathcal{L}$. \diamond

Teorema 1.6.6. Se \mathcal{P} è un π -sistema di sottoinsiemi di Ω e se \mathcal{L} è un λ -sistema che contiene \mathcal{P} , allora \mathcal{L} contiene anche la tribú $\mathcal{F}(\mathcal{P})$ generata da \mathcal{P} .

Dimostrazione. Si indichi con $\mathcal{L}(\mathcal{P})$ il piú piccolo λ -sistema che contiene \mathcal{P} . Basta dimostrare che $\mathcal{L}(\mathcal{P})$ è una tribú.

Si osservi che, se un λ -sistema \mathcal{L} è stabile rispetto alle intersezioni finite, è una tribú. Infatti, \mathcal{L} è stabile rispetto alla complementazione perché, se A appartiene a \mathcal{L} , si ha $A^c = \Omega \setminus A \in \mathcal{L}$; inoltre \mathcal{L} è stabile rispetto alle unioni finite perché

$A \cup B = (A^c \cap B^c)^c$ ed infine è stabile rispetto alle unioni numerabili perché $\cup_{j=1}^n A_j \uparrow \cup_{n \in \mathbb{N}} A_n$. Basta, perciò, dimostrare che $\mathcal{L}(\mathcal{P})$ è stabile rispetto alle intersezioni finite.

Si prenda A in \mathcal{P} e si definisca

$$\mathcal{G}(A) := \left\{ B \subseteq \Omega : A \cap B \in \mathcal{L}(\mathcal{P}) \right\}.$$

Allora $\mathcal{G}(A)$ è un λ -sistema. Infatti Ω appartiene a $\mathcal{G}(A)$ perché $A \cap \emptyset = \emptyset$ è in \mathcal{P} e dunque anche in $\mathcal{L}(\mathcal{P})$. Se B e C appartengono a $\mathcal{G}(A)$ e $C \subseteq B$, si ha $A \cap (B \setminus C) = (A \cap B) \setminus (A \cap C)$ che appartiene a $\mathcal{L}(\mathcal{P})$; perciò $B \setminus C$ appartiene a $\mathcal{G}(A)$. Infine, se (B_n) è una successione di insiemi di $\mathcal{G}(A)$ che tende crescendo a B , si ha $B_n \cap A \uparrow B \cap A$ e dunque anche B appartiene a $\mathcal{G}(A)$. Poiché \mathcal{P} è contenuto in $\mathcal{G}(A)$, si ha $\mathcal{L}(\mathcal{P}) \subseteq \mathcal{G}(A)$. Perciò, se A appartiene a \mathcal{P} e B a $\mathcal{L}(\mathcal{P})$, si ha che $A \cap B$ appartiene a $\mathcal{L}(\mathcal{P})$ e, di qui, che $\mathcal{G}(B)$ contiene \mathcal{P} se B è in $\mathcal{L}(\mathcal{P})$; ma $\mathcal{G}(B)$ è un λ -sistema e dunque $\mathcal{G}(B)$ contiene $\mathcal{L}(\mathcal{P})$. Questo implica che, se B e C appartengono a $\mathcal{L}(\mathcal{P})$ vi appartiene anche $B \cap C$. \square

Siamo ora in grado di dimostrare il seguente

Teorema 1.6.7. *Dato uno spazio misurabile (Ω, \mathcal{F}) , se due misure μ_1 e μ_2 coincidono sugli insiemi di un π -sistema \mathcal{P} di sottoinsiemi di \mathcal{F} , $\mu_1(A) = \mu_2(A)$ per ogni $A \in \mathcal{P}$, allora coincidono sulla tribù $\mathcal{F}(\mathcal{P})$ generata da \mathcal{P} , $\mu_1(A) = \mu_2(A)$ per ogni $A \in \mathcal{F}(\mathcal{P})$.*

Dimostrazione. Sia A un insieme di \mathcal{P} tale che $\mu_1(A) = \mu_2(A) < +\infty$. Si ponga

$$\mathcal{L} := \left\{ B \in \mathcal{F} : \mu_1(A \cap B) = \mu_2(A \cap B) \right\}.$$

Banalmente, si ha che Ω appartiene a \mathcal{L} . Siano ora B e C due insiemi di \mathcal{P} con $B \supset C$; allora

$$\begin{aligned} \mu_1(A \cap (B \setminus C)) &= \mu_1(A \cap B) - \mu_1(A \cap C) \\ &= \mu_2(A \cap B) - \mu_2(A \cap C) = \mu_2(A \cap (B \setminus C)), \end{aligned}$$

sicché $B \setminus C$ appartiene a \mathcal{L} . Infine sia (B_n) una successione crescente di insiemi di \mathcal{L} ; se $B = \cup_{n \in \mathbb{N}} B_n$, si ha

$$\begin{aligned} \mu_1(A \cap B) &= \lim_{n \rightarrow +\infty} \mu_1(A \cap B_n) = \lim_{n \rightarrow +\infty} \mu_2(A \cap B_n) \\ &= \mu_2(A \cap B). \end{aligned}$$

Dunque \mathcal{L} è un λ -sistema. Allora il teorema precedente assicura che \mathcal{L} include la tribù $\mathcal{F}(\mathcal{P})$ generata da \mathcal{P} . Perciò se $A \in \mathcal{P}$ è tale che $\mu_1(A) = \mu_2(A) < +\infty$ e se B è in $\mathcal{F}(\mathcal{P})$, si ha $\mu_1(A \cap B) = \mu_2(A \cap B)$. Consideriamo una successione crescente (A_n) di insiemi di \mathcal{P} tali che $\cup_{n \in \mathbb{N}} A_n = \Omega$ e che $\mu_1(A_n) = \mu_2(A_n) < +\infty$ per ogni $n \in \mathbb{N}$. Allora, per ogni $B \in \mathcal{F}(\mathcal{P})$, si ha

$$\mu_1(A \cap B) = \lim_{n \rightarrow +\infty} \mu_1(A \cap B_n) = \lim_{n \rightarrow +\infty} \mu_2(A \cap B_n) = \mu_2(A \cap B),$$

che fornisce l'asserto. \square

Le π -classi ed i λ -sistemi si usano anche per lo studio dell'indipendenza.

Teorema 1.6.8. *Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, si considerino le π -classi $\mathcal{C}_1, \dots, \mathcal{C}_n$ di insiemi di \mathcal{F} ; se $\mathcal{C}_1, \dots, \mathcal{C}_n$ sono indipendenti, vale a dire se*

$$\mathbb{P} \left(\bigcap_{j=1}^n A_j \right) = \prod_{j=1}^n \mathbb{P}(A_j), \quad (1.6.6)$$

per ogni scelta di A_j in \mathcal{C}_j ($j = 1, \dots, n$), allora sono indipendenti anche le tribú $\mathcal{F}(\mathcal{C}_1), \dots, \mathcal{F}(\mathcal{C}_n)$ generate da $\mathcal{C}_1, \dots, \mathcal{C}_n$ rispettivamente.

Dimostrazione. Se \mathcal{C}_j non contiene già Ω si consideri la famiglia \mathcal{C}'_j ottenuta aggiungendo Ω a \mathcal{C}_j ; ogni nuova famiglia \mathcal{C}'_j è anch'essa una π -classe e la (1.6.6) vale per ogni scelta di A_j in \mathcal{C}'_j ($j = 1, 2, \dots, n$). Si fissino ora A_2, A_3, \dots, A_n in $\mathcal{C}'_2, \mathcal{C}'_3, \dots, \mathcal{C}'_n$ rispettivamente e sia \mathcal{L}_1 la famiglia degli insiemi di \mathcal{F} per i quali vale la (1.6.6). Si controlla immediatamente che \mathcal{L}_1 è un λ -sistema; inoltre, esso contiene \mathcal{C}'_1 , sicché, per il Teorema 1.6.6 contiene anche la tribú $\mathcal{F}(\mathcal{C}'_1) = \mathcal{F}(\mathcal{C}_1)$ generata da \mathcal{C}'_1 o da \mathcal{C}_1 . Pertanto, la (1.6.6) vale se A_1 è in $\mathcal{F}(\mathcal{C}_1)$ e A_j in \mathcal{C}_j per $j = 2, 3, \dots, n$. Si ripeta lo stesso ragionamento per gli insiemi di $\mathcal{F}(\mathcal{C}_1), \mathcal{C}_2, \dots, \mathcal{C}_n$ per concludere che la (1.6.6) vale quando gli insiemi A_1, A_2, \dots, A_n siano scelti in $\mathcal{F}(\mathcal{C}_1), \mathcal{F}(\mathcal{C}_2), \mathcal{C}_3, \dots, \mathcal{C}_n$ rispettivamente; si proceda poi nello stesso modo per i rimanenti indici. \square

1.7 Le misure di Borel–Stieltjes

Un esempio importante di spazio misurabile è dato da $(\mathbb{R}, \mathcal{B})$. Si dirà che una misura μ su \mathcal{B} è *localmente finita* se accade che, per ogni intervallo limitato I , sia $\mu(I) < +\infty$. Ovviamente, ogni misura finita su $(\mathbb{R}, \mathcal{B})$ è anche localmente finita.

Teorema 1.7.1. *Se $\mu : \mathcal{B} \rightarrow \overline{\mathbb{R}}_+$ è una misura localmente finita, allora la funzione $F : \mathbb{R} \rightarrow \mathbb{R}$ definita, a meno di una costante, da*

$$F(b) - F(a) := \mu([a, b]) \quad (a < b)$$

è crescente e continua a destra.

Dimostrazione. Si osservi che

$$F(x) = \begin{cases} F(0) + \mu([0, x]), & \text{se } x \geq 0, \\ F(0) - \mu([x, 0]), & \text{se } x < 0; \end{cases}$$

basta perciò fissare $F(0)$ arbitrariamente. Se $x' < x''$, si ha $F(x'') - F(x') = \mu([x', x'']) \geq 0$, sicché F è crescente. Sia ora (x_n) un'arbitraria successione reale che decresce a x ($x_n \downarrow x$). Allora $F(x_n) - F(x) = \mu([x, x_n]) \xrightarrow{n \rightarrow +\infty} 0$. \square

In effetti, in probabilità, si segue una via diversa per definire univocamente la funzione F che sarà detta *funzione di ripartizione*. Anziché fissare il valore $F(0)$ di F nell'origine, si pone eguale a zero il limite di F a $-\infty$.

Vale il seguente importante reciproco dell'ultimo teorema.

Teorema 1.7.2. *Se $F : \mathbb{R} \rightarrow \mathbb{R}$ è crescente e continua a destra, esiste un'unica misura $\mu : \mathcal{B} \rightarrow \mathbb{R}_+$ tale che $\mu(]a, b]) = F(b) - F(a)$ per ogni intervallo $]a, b]$ con $a < b$. Tale misura è localmente finita e si dice misura di Borel–Stieltjes associata a F ; la si indica spesso con μ_F .*

Dimostrazione. Sia \mathcal{R} l'anello delle unioni finite di intervalli disgiunti, aperti a sinistra e chiusi a destra, $\mathcal{R} := \cup_{n \in \mathbb{N}} \mathcal{R}_n$ ove

$$\mathcal{R}_n := \left\{ A : A = \bigcup_{i=1}^n]a_i, b_i], a_1 \leq b_1 \leq a_2 \leq b_2 \leq \cdots \leq a_n \leq b_n \right\}.$$

In \mathcal{R} si definisca $\nu : \mathcal{R} \rightarrow \mathbb{R}_+$ mediante

$$\nu(A) := \sum_{i=1}^n \{F(b_i) - F(a_i)\};$$

ν , così definita, è finitamente additiva; di piú, essa è una misura su \mathcal{R} come si dimostrerà di seguito.

Se A è in \mathcal{R} , esiste, per ogni $\varepsilon > 0$, un insieme B di \mathcal{R} tale che la sua chiusura \overline{B} sia contenuta in A , vale a dire $B \subseteq \overline{B} \subseteq A$, e che $\nu(A) - \nu(B) < \varepsilon$. Infatti, se $A =]a, b]$, B sarà del tipo $]a + h, b]$ con $h > 0$ sufficientemente piccolo (per la continuità a destra di F si ha

$$\lim_{h \downarrow 0} \nu(]a + h, b]) = \lim_{h \downarrow 0} [F(b) - F(a + h)] = F(b) - F(a) = \nu(A).$$

Si procede in maniera analoga se A è un'unione finita di intervalli. Sia ora (A_n) una successione decrescente di insiemi di \mathcal{R} con $\cap_{n \in \mathbb{N}} A_n = \emptyset$. Fissato $\varepsilon > 0$, si scelga, per ogni $n \in \mathbb{N}$, $B_n \in \mathcal{R}$ in modo che sia $B_n \subseteq \overline{B_n} \subseteq A_n$ e che $\nu(A_n) - \nu(\overline{B_n}) < \varepsilon 2^{-n}$. Allora, si ha

$$\bigcap_{n \in \mathbb{N}} \overline{B_n} = \emptyset.$$

Esiste $r \in \mathbb{N}$ tale che

$$\bigcap_{i=1}^r \overline{B_i} = \emptyset.$$

Per rendersene conto, si osservi che gli insiemi $\overline{\mathbb{R}} \setminus \overline{B_n}$ formano un ricoprimento aperto del compatto $\overline{\mathbb{R}} = [-\infty, +\infty]$; vi è, pertanto, un ricoprimento aperto finito di $\overline{\mathbb{R}}$:

$$\overline{\mathbb{R}} = \bigcup_{i=1}^r (\overline{\mathbb{R}} \setminus \overline{B_i}) = \bigcup_{i=1}^r (\overline{\mathbb{R}} \cap \overline{B_i}^c) = \overline{\mathbb{R}} \cap \left(\bigcup_{i=1}^r \overline{B_i}^c \right),$$

sicché $\cap_{i=1}^r \overline{B_i} = \emptyset$ e dunque, a maggior ragione, $\cap_{i=1}^r B_i = \emptyset$ e

$$A_r = A_r \cap \left(\bigcap_{i=1}^r B_i \right)^c = \bigcup_{i=1}^r (A_r \setminus B_i) \subseteq \bigcup_{i=1}^r (A_i \setminus B_i),$$

onde

$$\begin{aligned} \nu(A_r) &\leq \nu \left[\bigcup_{i=1}^r (A_i \setminus B_i) \right] \\ &\leq \sum_{i=1}^r \nu(A_i \setminus B_i) < \varepsilon \sum_{i=1}^r 2^{-i} = \varepsilon(1 - 2^{-r}) < \varepsilon. \end{aligned}$$

Perciò $\lim_{n \rightarrow +\infty} \nu(A_n) = 0$ e, in virtù del Teorema 1.5.1, ν è σ -additiva. L'anello \mathcal{R} soddisfa alle ipotesi del teorema di Carathéodory perché \mathbb{R} è l'unione numerabile degli intervalli $] -n, n]$ e perché ν è σ -finita, in quanto $\nu(] -n, n]) = F(n) - F(-n) < +\infty$. Esiste perciò un'unica misura μ su \mathcal{B} tale che $\mu(A) = \nu(A)$ per ogni insieme $A \in \mathcal{R}$. In particolare, $\mu(]a, b]) = \nu(]a, b]) = F(b) - F(a)$. \square

La funzione identità soddisfa alle ipotesi del Teorema 1.7.2: $F(x) = x$ per ogni $x \in \mathbb{R}$; la misura associata si dice di Borel–Lebesgue; essa fa corrispondere ad ogni intervallo $[a, b]$ la sua lunghezza $b - a$. Questa costruzione della misura di Lebesgue non è l'unica possibile.

1.8 Funzioni semplici e funzioni misurabili

Definizione 1.8.1. Dati due spazi misurabili (Ω, \mathcal{F}) e (Ω', \mathcal{F}') si dice che la funzione $f : \Omega \rightarrow \Omega'$ è *misurabile* (rispetto alle tribú \mathcal{F} e \mathcal{F}') se, per ogni insieme B di \mathcal{F}' , l'immagine inversa di B mediante f appartiene alla tribú \mathcal{F} , cioè

$$\forall B \in \mathcal{F}' \quad f^{-1}(B) \in \mathcal{F}.$$

Spesso si scrive $f : (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$ per indicare una funzione misurabile nel senso di questa definizione. Se poi f è una funzione da Ω in \mathbb{R} o in $\overline{\mathbb{R}}$ spesso si sottintende la tribú di Borel \mathcal{B} in \mathbb{R} o in $\overline{\mathbb{R}}$. In questo caso si scrive $f \in \mathcal{L}^0$, indicando lo spazio delle funzioni misurabili da Ω in \mathbb{R} o in $\overline{\mathbb{R}}$ con \mathcal{L}^0 o, se si vuole indicare esplicitamente la tribú considerata, con $\mathcal{L}^0(\Omega, \mathcal{F})$, oppure ancora con $\mathcal{L}^0(\mathcal{F})$. \diamond

Il seguente teorema è di dimostrazione immediata

Teorema 1.8.1. Siano (Ω, \mathcal{F}) , (Ω', \mathcal{F}') e $(\Omega'', \mathcal{F}'')$ tre spazi misurabili e siano $f : \Omega \rightarrow \Omega'$ e $g : \Omega' \rightarrow \Omega''$ due funzioni misurabili, rispettivamente rispetto alle coppie di tribú $\mathcal{F}, \mathcal{F}'$ e $\mathcal{F}', \mathcal{F}''$. Allora è misurabile rispetto a \mathcal{F} e a \mathcal{F}'' la funzione composta $g \circ f : \Omega \rightarrow \Omega''$.

Per verificare se una funzione sia misurabile si ricorre spesso al seguente criterio.

Teorema 1.8.2. Siano $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$ spazi misurabili e sia $\mathcal{A} \subseteq \mathcal{F}_2$ una famiglia di sottoinsiemi di Ω_2 che genera la tribú \mathcal{F}_2 , cioè $\mathcal{F}(\mathcal{A}) = \mathcal{F}_2$. Sono allora equivalenti, per una funzione $f : \Omega_1 \rightarrow \Omega_2$, le proprietà:

- (a) f è misurabile;
- (b) $\forall A \in \mathcal{A} \quad f^{-1}(A) \in \mathcal{F}_1$.

Dimostrazione. L'implicazione (a) \implies (b) è banale.

(b) \implies (a) Sia \mathcal{G} la famiglia dei sottoinsiemi di Ω_2 la cui immagine inversa appartiene a \mathcal{F}_1 : $\mathcal{G} := \{A \subseteq \Omega_2 : f^{-1}(A) \in \mathcal{F}_1\}$. Per ipotesi $\mathcal{A} \subseteq \mathcal{G}$. \mathcal{G} è una tribú: $\Omega_2 \in \mathcal{G}$ perché $f^{-1}(\Omega_2) = \Omega_1 \in \mathcal{F}_1$; se $A \in \mathcal{G}$, è $f^{-1}(A^c) = [f^{-1}(A)]^c \in \mathcal{F}_1$, cioè $A^c \in \mathcal{G}$. Infine, se A_n appartiene a \mathcal{G} per ogni $n \in \mathbb{N}$ si ha $f^{-1}(\cup_{n \in \mathbb{N}} A_n) = \cup_{n \in \mathbb{N}} f^{-1}(A_n) \in \mathcal{F}_1$, cioè $\cup_{n \in \mathbb{N}} A_n \in \mathcal{G}$. Ma allora, dalla definizione di tribú generata, scende che $\mathcal{G} \supseteq \mathcal{F}_2$, sicché f risulta misurabile. \square

Corollario 1.8.1. *Sia (Ω, \mathcal{F}) uno spazio misurabile; affinché la funzione $f : \Omega \rightarrow \mathbb{R}$ sia misurabile, rispetto a \mathcal{F} e a $\mathcal{B}(\mathbb{R})$, ognuna delle seguenti condizioni equivalenti è necessaria e sufficiente:*

- (a) $\{f \leq c\} := f^{-1}([-\infty, c]) = \{\omega \in \Omega : f(\omega) \leq c\}$ appartiene a \mathcal{F} per ogni $c \in \mathbb{R}$;
- (b) $\{f > c\} := f^{-1}(]c, +\infty]) = \{\omega \in \Omega : f(\omega) > c\}$ appartiene a \mathcal{F} per ogni $c \in \mathbb{R}$;
- (c) $\{f \geq c\} := f^{-1}([c, +\infty]) = \{\omega \in \Omega : f(\omega) \geq c\}$ appartiene a \mathcal{F} per ogni $c \in \mathbb{R}$;
- (d) $\{f < c\} := f^{-1}([-\infty, c[) = \{\omega \in \Omega : f(\omega) < c\}$ appartiene a \mathcal{F} per ogni $c \in \mathbb{R}$.

Dimostrazione. Ognuna delle famiglie indicate genera $\mathcal{B}(\mathbb{R})$. □

Definizione 1.8.2. Una funzione $f : \mathbb{R} \rightarrow \mathbb{R}$ (oppure da $\overline{\mathbb{R}}$ in $\overline{\mathbb{R}}$) che sia misurabile rispetto alle tribú di Borel si dice *boreliana*. ◇

Si osservi la notazione usata per la prima volta nell'enunciato del Corollario 1.8.1, $\{f \leq c\} := f^{-1}([-\infty, c]) = \{\omega \in \Omega : f(\omega) \leq c\}$ e simili; nel seguito essa sarà usata sistematicamente. Così, ad esempio, si scriverà $\{a \leq f \leq b\}$ o $\{f \in [a, b]\}$ in luogo delle notazioni, equivalenti, ma piú pesanti, $f^{-1}([a, b])$ oppure

$$\{\omega \in \Omega : a \leq f(\omega) \leq b\}.$$

Un'ampia classe di funzioni $f : \mathbb{R} \rightarrow \mathbb{R}$ misurabili rispetto alla tribú di Borel \mathcal{B} è data dalle funzioni continue (si vedano gli esercizi). Una funzione indicatrice $\mathbf{1}_A$ è misurabile se, e solo se, l'insieme A appartiene a \mathcal{F} . La classe delle funzioni misurabili è stabile rispetto a numerose operazioni eseguite in numero finito o numerabile, come dimostrano i due teoremi che seguono.

Teorema 1.8.3. *Siano $f, g : \Omega \rightarrow \overline{\mathbb{R}}$ funzioni misurabili e sia α un numero reale. Sono allora misurabili, se sono definite, anche le funzioni $\alpha \cdot f$, $f + g$, f^2 , $f^+ := f \vee 0$, $f^- := -(f \wedge 0)$, $|f|$, $f \cdot g$, $f \vee g$, $f \wedge g$, f^n per ogni $n \in \mathbb{N}$. Inoltre, se f è misurabile e positiva sono misurabili anche le funzioni $f^{1/n}$, per ogni $n \in \mathbb{N}$, e f^r per ogni numero razionale positivo r , $r \in \mathbb{Q}_+$; se $f \leq 0$ è misurabile e n è dispari, anche $f^{1/n}$ è misurabile.*

Dimostrazione. La dimostrazione che $\alpha \cdot f$ è misurabile è una banale applicazione del Corollario 1.8.1.

Per $\omega \in \Omega$ è ovvio che si ha $f(\omega) + g(\omega) > c$ se, e solo se, esiste un razionale q tale che

$$f(\omega) > q > c - g(\omega).$$

Si numerino i razionali

$$\mathbb{Q} = \bigcup_{n \in \mathbb{N}} \{q_n\}.$$

Allora

$$\{f + g > c\} = \bigcup_{n \in \mathbb{N}} \left(\{f > q_n\} \cap \{g > c - q_n\} \right);$$

per il Corollario 1.8.1, ciascuno degli insiemi dei quali si fa l'unione appartiene a \mathcal{F} , sicché vi appartiene anche la loro unione numerabile. Dunque, $f + g$ è misurabile.

$$\{f^2 \leq c\} = \begin{cases} \emptyset, & \text{se } c < 0, \\ \{f = 0\}, & \text{se } c = 0, \\ \{-\sqrt{c} \leq f \leq \sqrt{c}\}, & \text{se } c > 0; \end{cases}$$

poiché ognuno di questi insiemi è misurabile, f^2 è misurabile.

$$\{f^+ \geq c\} = \begin{cases} \Omega, & \text{se } c \leq 0, \\ \{f \geq c\}, & \text{se } c > 0, \end{cases}$$

sicché $\{f^+ \geq c\}$ è in \mathcal{F} per ogni c reale; f^+ è dunque misurabile. In maniera analoga si stabilisce la misurabilità di f^- . Poiché $|f| = f^+ + f^-$, anche $|f|$ risulta misurabile. La misurabilità delle altre funzioni scende ora da quanto è già stato dimostrato e dalle identità

$$\begin{aligned} f \cdot g &= \frac{1}{2} \{(f + g)^2 - f^2 - g^2\} \\ f \vee g &= \frac{1}{2} \{f + g + |f - g|\}, \quad f \wedge g = f + g - (f \vee g). \end{aligned}$$

Si è visto sopra che f^2 è misurabile. Per induzione si supponga che sia misurabile f^n con $n \in \mathbb{N}$. Allora, dalla misurabilità del prodotto di due funzioni misurabili scende la misurabilità di $f^n \cdot f = f^{n+1}$, sicché f^n è misurabile per ogni $n \in \mathbb{N}$.

Vogliamo, infine, studiare la misurabilità di $f^{1/n}$ quando f è positiva; si ha

$$\{f^{1/n} \geq c\} = \begin{cases} \Omega, & \text{se } c \leq 0, \\ \{f \geq c^n\}, & \text{se } c > 0. \end{cases}$$

Se f è negativa e n è dispari, $n = 2k - 1$ con $k \in \mathbb{N}$, si ha

$$\{f^{1/(2k-1)} \geq c\} = \{f \geq c^{2k-1}\}.$$

In ogni caso appartiene a \mathcal{F} . Dunque, $f^{1/n}$ è misurabile.

Infine, se r è un numero razionale positivo, si può scrivere $r = p/q$, con p e q naturali. L'asserto è ora conseguenza di quanto già visto. \square

Teorema 1.8.4. *Sia (f_n) una successione di funzioni misurabili da Ω in $\overline{\mathbb{R}}$; sono allora misurabili anche le seguenti funzioni:*

- (a) $\bigvee_{n \in \mathbb{N}} f_n := \sup_{n \in \mathbb{N}} f_n$;
- (b) $\bigwedge_{n \in \mathbb{N}} f_n := \inf_{n \in \mathbb{N}} f_n$,
- (c) $\limsup_{n \rightarrow +\infty} f_n$;
- (d) $\liminf_{n \rightarrow +\infty} f_n$;
- (e) $\lim_{n \rightarrow +\infty} f_n$ se la successione (f_n) converge.

Dimostrazione. (a) $\{\bigvee_{n \in \mathbb{N}} f_n \leq c\} = \bigcap_{n \in \mathbb{N}} \{f_n \leq c\}$ e l'asserto segue ora da 1.8.1.

(b) segue da (a) perché $\bigwedge_{n \in \mathbb{N}} f_n = -\bigvee_{n \in \mathbb{N}} (-f_n)$.

(c) $\limsup_{n \rightarrow +\infty} f_n = \inf_{n \rightarrow +\infty} g_n$ con $g_n := \sup_{k \geq n} f_k$ e g_n è misurabile per ogni $n \in \mathbb{N}$ in virtù di (a); l'asserto segue ora da (b). In maniera analoga si dimostra (d).

Quanto a (e), basta osservare che, se la successione (f_n) converge, si ha

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

Il risultato è così dimostrato. \square

Il seguente corollario è un'immediata conseguenza dei due teoremi precedenti.

Corollario 1.8.2. *Sia f una funzione misurabile positiva, $f \geq 0$ e x un qualsiasi reale positivo; è allora misurabile la funzione f^x .*

Dimostrazione. Basta considerare un'arbitraria successione $(r_n)_{n \in \mathbb{N}}$ di razionali positivi convergenti a x ed osservare che $\lim_{n \rightarrow +\infty} f^{r_n} = f^x$. \square

Definizione 1.8.3. Si dice (\mathcal{F}) -semplice ogni funzione $f : \Omega \rightarrow \mathbb{R}$ che si possa scrivere nella forma

$$f = \sum_{j=1}^n c_j \mathbf{1}_{A_j},$$

ove $A_j \in \mathcal{F}$ per $j = 1, \dots, n$, $A_j \cap A_k = \emptyset$ se $j \neq k$, $\bigcup_{j=1}^n A_j = \Omega$ e $c_j \in \mathbb{R}$ ($j = 1, \dots, n$). \diamond

Si può fare a meno della condizione $\bigcup_{j=1}^n A_j = \Omega$; basterebbe, infatti, porre $A_{n+1} := \Omega \setminus \bigcup_{j=1}^n A_j$ se fosse $\bigcup_{j=1}^n A_j \neq \Omega$. Allora, si può scrivere $f = \sum_{j=1}^{n+1} c_j \mathbf{1}_{A_j}$, ponendo $c_{n+1} := 0$. Ogni funzione semplice assume un numero finito di valori. La rappresentazione di una funzione semplice non è unica, fatto che si sfrutta, per esempio nella dimostrazione del successivo Teorema 1.8.5. Se è ovvio dal contesto quale sia la tribù alla quale si fa riferimento si dirà che una funzione è semplice, anziché \mathcal{F} -semplice.

Teorema 1.8.5. *Siano f e g due funzioni semplici; tali sono allora anche le funzioni αf per ogni $\alpha \in \mathbb{R}$, $f \pm g$, $f \cdot g$, $f \vee g := \max\{f, g\}$, $f \wedge g := \min\{f, g\}$.*

Dimostrazione. Siano c_1, \dots, c_m e d_1, \dots, d_n numeri reali, A_1, \dots, A_m e B_1, \dots, B_n partizioni misurabili di Ω , vale a dire $\bigcup_{j=1}^m A_j = \bigcup_{k=1}^n B_k = \Omega$, con A_j e B_k in \mathcal{F} per tutti gli indici j e k ; infine $A_i \cap A_j = B_i \cap B_j = \emptyset$ per $i \neq j$. Si consideri la partizione più fine di entrambe $(A_j \cap B_k)$ ($j = 1, \dots, m; k = 1, \dots, n$). È ora evidente che si possono rappresentare f e g ricorrendo alla stessa partizione $(A_j \cap B_k)$

$$f = \sum_{k=1}^n \sum_{j=1}^m c_j \mathbf{1}_{A_j \cap B_k},$$

$$g = \sum_{k=1}^n \sum_{j=1}^m d_k \mathbf{1}_{A_j \cap B_k},$$

sicché

$$f + g = \sum_{k=1}^n \sum_{j=1}^m (c_j + d_k) \mathbf{1}_{A_j \cap B_k},$$

che rende evidente l'asserto per la somma. Si procede in maniera analoga negli altri casi. \square

Teorema 1.8.6. *Ogni funzione \mathcal{F} -semplice è \mathcal{F} -misurabile.*

Dimostrazione. Sia $f = \sum_{j=1}^n c_j \mathbf{1}_{A_j}$; allora $\{f \leq c\} = \cup \{A_j : c_j \leq c\}$, che, come unione finita di insiemi di \mathcal{F} , appartiene a \mathcal{F} . \square

L'importanza delle funzioni semplici è data dal seguente

Teorema 1.8.7. *Ogni funzione $f : \Omega \rightarrow \overline{\mathbb{R}}_+$ misurabile e positiva è il limite di almeno una successione crescente di funzioni semplici positive.*

Dimostrazione. Per ogni intero positivo s si ponga

$$A_{k,n} := \begin{cases} \left\{ \frac{(k-1)}{2^n} \leq f < \frac{k}{2^n} \right\}, & \text{per } k = 1, \dots, 2^{2^n}, \\ \{f \geq 2^n\}, & \text{per } k = 0. \end{cases} \quad (1.8.1)$$

Poiché f è misurabile, tutti gli insiemi $A_{k,n}$ appartengono a \mathcal{F} . Inoltre, se $k \neq k'$ si ha

$$A_{k,n} \cap A_{k',n} = \emptyset \quad \text{e} \quad \Omega = \bigcup_{k=0}^{2^{2^n}} A_{k,n}.$$

La funzione $s_n : \Omega \rightarrow \mathbb{R}_+$ definita da

$$s_n := \sum_{k=1}^{2^{2^n}} \frac{k-1}{2^n} \mathbf{1}_{A_{k,n}} + 2^n \mathbf{1}_{A_{0,n}} \quad (1.8.2)$$

è semplice e soddisfa alla disuguaglianza $0 \leq s_n \leq f$.

Se $k > 0$, si ha $A_{k,n} = A_{2k-1,n+1} \cup A_{2k,n+1}$ che è un'unione disgiunta. Se ω è in $A_{2k-1,n+1}$, si ha $s_n(\omega) = s_{n+1}(\omega) = (k-1)/2^n$, mentre, se ω è in $A_{2k,n+1}$, è

$$s_{n+1}(\omega) = \frac{2k-1}{2^{n+1}} = \frac{2k-2}{2^{n+1}} + \frac{1}{2^{n+1}} = s_n(\omega) + \frac{1}{2^{n+1}},$$

sicché, in ogni caso, $s_{n+1}(\omega) \geq s_n(\omega)$. Se, poi, è $\omega \in A_{0,n}$, si osservi che

$$A_{0,n} = A_{0,n+1} \cup \left(\bigcup_{r=2^{2^{n+1}+1}}^{2^{2^{n+2}}} A_{r,n+1} \right),$$

e perciò $s_n(\omega) = 2^n \leq s_{n+1}(\omega)$. Dunque, $s_n(\omega) \leq s_{n+1}(\omega)$ per ogni $\omega \in \Omega$. Sia ora $\omega \in \Omega$ tale che $f(\omega) < +\infty$; esiste allora un numero naturale n tale $2^n > s_n(\omega)$; un tale punto appartiene ad un insieme $A_{k,n}$ con $k > 0$. Perciò $0 \leq f(\omega) - s_n(\omega) \leq 2^{-n}$, sicché $\lim_{n \rightarrow +\infty} s_n(\omega) = f(\omega)$. Se invece $f(\omega) = +\infty$, si ha $s_n(\omega) = 2^n$ per ogni $n \in \mathbb{N}$, sicché $\lim_{s \rightarrow +\infty} s_n(\omega) = +\infty$. \square

Prima di chiudere questa sezione diamo il seguente teorema che in letteratura è pure noto come “Teorema della classe monotona” e che costituisce una versione più sofisticata del Teorema 1.6.1.

Teorema 1.8.8. *Sia \mathcal{H} una classe di funzioni limitate da Ω in \mathbb{R} e si supponga che \mathcal{H} soddisfaccia alle proprietà:*

- (a) \mathcal{H} è uno spazio vettoriale;
- (b) la funzione identicamente eguale a 1 appartiene a \mathcal{H} ;
- (c) appartiene a \mathcal{H} ogni funzione limitata su Ω che sia l'estremo superiore di una successione (f_n) crescente di funzioni positive di \mathcal{H} .

Se, inoltre, sono in \mathcal{H} le funzioni indicatrici di ogni insieme in un dato π -sistema \mathcal{I} , allora appartiene a \mathcal{H} ogni funzione limitata definita in Ω che sia misurabile rispetto a $\mathcal{F}(\mathcal{I})$.

Dimostrazione. Sia f una funzione positiva misurabile rispetto a $\mathcal{F}(\mathcal{I})$ e si supponga che sia $0 \leq f(\omega) \leq K$ per ogni $\omega \in \Omega$. Come nell'eq. (1.8.2) si definisca per $s \in \mathbb{N}$

$$f_s := \sum_{s=1}^{K \wedge 2^{2s}} \frac{k-1}{2^s} \mathbf{1}_{A_{k,s}},$$

ove gli insiemi $A_{k,s}$ sono stati introdotti in (1.8.1). Poiché f è misurabile rispetto a $\mathcal{F}(\mathcal{I})$, ogni insieme $A_{k,s}$ è in $\mathcal{F}(\mathcal{I})$, sicché $\mathbf{1}_{A_{k,s}}$ appartiene a \mathcal{H} ; e, poiché \mathcal{H} è uno spazio vettoriale, f_s appartiene a \mathcal{H} per ogni $s \in \mathbb{N}$. Ma (f_s) tende crescendo a f sicché f è in \mathcal{H} .

Il caso generale scende dall'usuale decomposizione $f = f^+ - f^-$. □

Esempio 1.8.1. Si considerino lo spazio di probabilità $([0, 1[, \mathcal{B}([0, 1]), \lambda)$ e la funzione $R_n(x) := 2x_n - 1$ ove x_n è la n -esima cifra nello sviluppo binario di $x \in [0, 1]$:

$$x = \sum_{n \in \mathbb{N}} \frac{x_n}{2^n} \quad x_n \in \{0, 1\},$$

ove per evitare ambiguità non si considerano gli sviluppi con solo un numero finito di zeri. Evidentemente

$$R_1(x) = \begin{cases} -1, & x \in [0, 1/2[\\ 1, & x \in [1/2, 1[\end{cases}, \quad R_2(x) = \begin{cases} -1, & x \in [0, 1/4[\cup [1/4, 3/4[\\ 1, & x \in [1/4, 1/2[\cup [3/4, 1[\end{cases},$$

$$R_3(x) = \begin{cases} -1, & x \in [0, 1/8[\cup [1/4, 3/8[\cup [1/2, 5/8[\cup [3/4, 7/8[\\ 1, & x \in [1/8, 1/4[\cup [3/8, 1/2[\cup [5/8, 3/4[\cup [7/8, 1[\end{cases},$$

.....

$$R_n(x) = \begin{cases} -1, & x \in \bigcup_{k=0}^{2^{n-1}-1} \left[\frac{2k}{2^n}, \frac{2k+1}{2^n} \right[\\ 1, & x \in \bigcup_{k=0}^{2^{n-1}-1} \left[\frac{2k+1}{2^n}, \frac{2k+2}{2^n} \right] \end{cases}.$$

Le funzioni R_n sono misurabili e si chiamano *funzioni di Rademacher*; esse si usano in alcuni esempî. ■

1.9 La definizione dell'integrale

La definizione dell'integrale sarà data a tappe; prima per le funzioni semplici positive, quindi per le funzioni misurabili positive, ed infine sarà estesa alle funzioni che si diranno *integrabili*.

L'integrale per le funzioni semplici positive. Sia $(\Omega, \mathcal{F}, \mu)$ uno spazio misurale, sia

$$f = \sum_{j=1}^n c_j \mathbf{1}_{A_j}$$

una funzione semplice positiva, tale, cioè, che, per ogni indice j , si abbia $c_j \geq 0$ e gli insiemi A_j appartengano tutti a \mathcal{F} , siano disgiunti e costituiscano una partizione di Ω . Si definisca allora come integrale di f rispetto alla misura μ la somma

$$\int f \, d\mu := \sum_{j=1}^n c_j \mu(A_j). \quad (1.9.1)$$

Occorre mostrare che la definizione appena data (1.9.1) non dipende dalla particolare rappresentazione della funzione f . Si considerino due diverse rappresentazioni di f :

$$f = \sum_{j=1}^n c_j \mathbf{1}_{A_j} = \sum_{k=1}^r d_k \mathbf{1}_{B_k},$$

con $A_i \cap A_j = B_i \cap B_j = \emptyset$ se $i \neq j$, $\cup_{j=1}^n A_j = \cup_{k=1}^r B_k = \Omega$, $c_j, d_k \geq 0$, $A_i, B_k \in \mathcal{F}$ per ogni $i \leq n$ e per ogni $k \leq r$. Perciò,

$$\mu(A_i) = \sum_{j=1}^r \mu(A_i \cap B_j) \quad \text{e} \quad \mu(B_j) = \sum_{i=1}^n \mu(A_i \cap B_j).$$

Se $A_i \cap B_j \neq \emptyset$, accade che $f(\omega) = c_i = d_j$, per ogni $\omega \in A_i \cap B_j$; perciò,

$$\begin{aligned} \sum_{i=1}^n c_i \mu(A_i) &= \sum_{i=1}^n \sum_{j=1}^r c_i \mu(A_i \cap B_j) \\ &= \sum_{j=1}^r \sum_{i=1}^n d_j \mu(A_i \cap B_j) = \sum_{j=1}^r d_j \mu(B_j), \end{aligned}$$

di modo che la (1.9.1) è una buona definizione.

Siano ora f e g due funzioni semplici positive

$$f = \sum_{i=1}^n c_i \mathbf{1}_{A_i} \quad \text{e} \quad g = \sum_{j=1}^r d_j \mathbf{1}_{B_j};$$

ricorrendo alle rappresentazioni

$$f = \sum_{i=1}^n \sum_{j=1}^r c_i \mathbf{1}_{A_i \cap B_j} \quad \text{e} \quad g = \sum_{i=1}^n \sum_{j=1}^r d_j \mathbf{1}_{A_i \cap B_j}$$

si può scrivere $f + g = \sum_{i=1}^n \sum_{j=1}^r (c_i + d_j) \mathbf{1}_{A_i \cap B_j}$, onde

$$\begin{aligned} \int (f + g) \, d\mu &= \sum_{i=1}^n \sum_{j=1}^r (c_i + d_j) \mu(A_i \cap B_j) \\ &= \sum_{i=1}^n c_i \sum_{j=1}^r \mu(A_i \cap B_j) + \sum_{j=1}^r d_j \sum_{i=1}^n \mu(A_i \cap B_j) \\ &= \sum_{i=1}^n c_i \mu(A_i) + \sum_{j=1}^r d_j \mu(B_j) = \int f \, d\mu + \int g \, d\mu. \end{aligned}$$

Nella stessa maniera si dimostra che, se α e β sono reali positivi, si ha

$$\int (\alpha f + \beta g) \, d\mu = \alpha \int f \, d\mu + \beta \int g \, d\mu,$$

e che, se $0 \leq f \leq g$, è

$$\int f \, d\mu \leq \int g \, d\mu.$$

Funzioni misurabili positive. Se $f : \Omega \rightarrow \overline{\mathbb{R}}_+$ è misurabile, esiste una successione di funzioni semplici positive (s_n) con $s_n \uparrow f$ (Teorema 1.8.7). Per ogni $n \in \mathbb{N}$, è definito l'integrale $\int s_n \, d\mu$; inoltre, si è appena visto che la successione degli integrali

$$\left(\int s_n \, d\mu \right)_{n \in \mathbb{N}}$$

è crescente; essa ammette perciò limite. Si definisca

$$\int f \, d\mu := \lim_{n \rightarrow +\infty} \int s_n \, d\mu. \quad (1.9.2)$$

Occorre dimostrare che la (1.9.2) è una buona definizione, vale a dire che il valore dell'integrale non dipende dalla particolare successione (s_n) di funzioni semplici scelta per approssimare f . La dimostrazione fa uso del seguente

Lemma 1.9.1. *Sia (s_n) una successione crescente di funzioni semplici positive. Si supponga che esista una funzione semplice positiva $g : \Omega \rightarrow \mathbb{R}_+$ tale che*

$$\lim_{n \rightarrow +\infty} s_n \geq g.$$

Allora

$$\lim_{n \rightarrow +\infty} \int s_n \, d\mu \geq \int g \, d\mu. \quad (1.9.3)$$

Dimostrazione. Sia $g = \sum_{i=1}^k c_i \mathbf{1}_{A_i}$ e si supponga dapprima che sia $\int g \, d\mu = +\infty$. Esiste perciò un indice i tale che $c_i > 0$ e $\mu(A_i) = +\infty$. Preso ε in $]0, c_i[$, si ponga

$$E_n := \{s_n + \varepsilon \geq g\} \quad (n \in \mathbb{N});$$

la successione $(A_i \cap E_n)_{n \in \mathbb{N}}$ è crescente e tende a A_i , onde

$$\lim_{n \rightarrow +\infty} \mu(A_i \cap E_n) = +\infty.$$

Ora,

$$\int s_n \, d\mu \geq \int s_n \mathbf{1}_{A_i \cap E_n} \, d\mu \geq (c_i - \varepsilon) \mu(A_i \cap E_n) \xrightarrow{n \rightarrow \infty} +\infty,$$

ciò che stabilisce la (1.9.3), in queste ipotesi. Se poi è $\int g \, d\mu < +\infty$, si ponga

$$E := \{g > 0\} = \bigcup \{A_i : c_i > 0\} \in \mathcal{F}.$$

Poiché g è semplice, si ha $c := \min\{c_i : c_i > 0\} > 0$ e $\mu(E) < +\infty$. Fissato $\varepsilon \in]0, c[$, si ha

$$\begin{aligned} \int s_n \, d\mu &\geq \int s_n \mathbf{1}_{E \cap E_n} \, d\mu \geq \int (g - \varepsilon) \mathbf{1}_{E \cap E_n} \, d\mu \\ &= \int g \mathbf{1}_{E_n \cap E} \, d\mu - \varepsilon \mu(E_n \cap E) \geq \sum_{c_i > 0} \int g \mathbf{1}_{E_n \cap A_i} \, d\mu - \varepsilon \mu(E) \\ &= \sum_{c_i > 0} c_i \mu(E_n \cap A_i) - \varepsilon \mu(E), \end{aligned}$$

e, poiché $\mu(E_n \cap A_i) \rightarrow \mu(A_i)$,

$$\lim_{n \rightarrow +\infty} \int s_n \, d\mu \geq \sum_{c_i > 0} c_i \mu(A_i) - \varepsilon \mu(E) = \int g \, d\mu - \varepsilon \mu(E),$$

e, di qui, la (1.9.3) in virtù dell'arbitrarietà di ε . □

Siano ora (s_n) e (t_r) due successioni crescenti di funzioni semplici positive entrambe tendenti a f . Fissato $r \in \mathbb{N}$, si ha $\lim_{n \rightarrow +\infty} s_n = f \geq t_r$ onde, per il Lemma 1.9.1,

$$\lim_{n \rightarrow +\infty} \int s_n \, d\mu \geq \int t_r \, d\mu,$$

e, di qui,

$$\lim_{n \rightarrow +\infty} \int s_n \, d\mu \geq \lim_{r \rightarrow +\infty} \int t_r \, d\mu.$$

Scambiando i ruoli delle successioni (s_n) e (t_r) si ottiene la disuguaglianza nell'altro verso, sicché si può concludere che

$$\lim_{n \rightarrow +\infty} \int s_n \, d\mu = \lim_{r \rightarrow +\infty} \int t_r \, d\mu,$$

ciò che dimostra che la (1.9.2) è una buona definizione.

Dalla stessa proprietà valida per le funzioni semplici positive si deduce che, se f e g sono funzioni misurabili e positive e se α e β sono numeri reali positivi, vale

$$\int (\alpha f + \beta g) \, d\mu = \alpha \int f \, d\mu + \beta \int g \, d\mu.$$

Funzioni misurabili integrabili. Sia $f : \Omega \rightarrow \overline{\mathbb{R}}$ misurabile; tali sono anche le funzioni positive f^+ e f^- . Se tanto f^+ quanto f^- hanno integrale finito, si ponga

$$\int f \, d\mu := \int f^+ \, d\mu - \int f^- \, d\mu.$$

Definizione 1.9.1. Si dice che una funzione $f : \Omega \rightarrow \overline{\mathbb{R}}$ è *integrabile* se tanto $\int f^+ d\mu$ quanto $\int f^- d\mu$ sono finiti; si dirà che f è *semi-integrabile* se solo uno dei due integrali $\int f^+ d\mu$ o $\int f^- d\mu$ è finito, precisamente f si dice *semi-integrabile inferiormente* se $\int f^+ d\mu = +\infty$ mentre $\int f^- d\mu < +\infty$ nel qual caso si porrà $\int f d\mu := +\infty$; si dice invece che f è *semi-integrabile superiormente* se $\int f^+ d\mu < +\infty$ e $\int f^- d\mu = +\infty$, nel qual caso si pone $\int f d\mu := -\infty$.

Se accade che sia $\int f^+ d\mu = \int f^- d\mu = +\infty$, non si definisce l'integrale $\int f d\mu$.

Se $A \in \mathcal{F}$ si definisce l'integrale esteso all'insieme A mediante

$$\int_A f d\mu := \int f \mathbf{1}_A d\mu,$$

se esiste l'integrale a secondo membro. \diamond

Dato lo spazio misurale $(\Omega, \mathcal{F}, \mu)$, l'insieme delle funzioni $f : \Omega \rightarrow \overline{\mathbb{R}}$ integrabili rispetto alla misura μ sarà denotato con $\mathcal{L}^1 = \mathcal{L}^1(\Omega, \mathcal{F}, \mu)$. Nel seguito si incontreranno le proprietà di tale spazio.

1.10 Proprietà dell'integrale

Dato uno spazio misurale $(\Omega, \mathcal{F}, \mu)$ si dice che una proprietà vale *quasi ovunque* (abbreviato in q.o. in italiano, a.e.=*almost everywhere* in inglese, p.p.=*presque partout* in francese) se l'insieme dei punti nei quali la proprietà non vale è contenuto in un insieme trascurabile, cioè misurabile e di misura nulla.

Teorema 1.10.1. *Siano $f, g : \Omega \rightarrow \overline{\mathbb{R}}$ funzioni integrabili; allora*

(a) *esiste finito l'integrale $\int_A f d\mu$;*

(b) *se $A \cap B = \emptyset$ con $A, B \in \mathcal{F}$, è*

$$\int_{A \cup B} f d\mu = \int_A f d\mu + \int_B f d\mu;$$

(c) *f è finita q.o. ($\mu(\{|f| = +\infty\}) = 0$);*

(d) *$f + g$ è integrabile e*

$$\int (f + g) d\mu = \int f d\mu + \int g d\mu;$$

(e) *$f \geq 0 \implies \int f d\mu \geq 0$ e $f \geq g \implies \int f d\mu \geq \int g d\mu$ (isotonía);*

(f) *$|f|$ è integrabile e*

$$\left| \int f d\mu \right| \leq \int |f| d\mu;$$

(g) *per ogni c reale, cf è integrabile e*

$$\int cf d\mu = c \int f d\mu;$$

- (h) $f \geq 0$ e $\int f \, d\mu = 0 \implies f = 0$ q.o.;
- (i) $f = g$ q.o. $\implies \int f \, d\mu = \int g \, d\mu$;
- (j) se $h : \Omega \rightarrow \overline{\mathbb{R}}$ è misurabile e $|h| \leq |f|$, h è integrabile.

Dimostrazione. (a) L'asserto scende dalle ovvie disuguaglianze

$$0 \leq f^+ \mathbf{1}_A \leq f^+ \quad \text{e} \quad 0 \leq f^- \mathbf{1}_A \leq f^-;$$

la (b) scende dalle relazioni

$$\begin{aligned} \mathbf{1}_{A \cup B} &= \mathbf{1}_A + \mathbf{1}_B, \\ f^+ \mathbf{1}_{A \cup B} &= f^+ \mathbf{1}_A + f^+ \mathbf{1}_B, \quad f^- \mathbf{1}_{A \cup B} = f^- \mathbf{1}_A + f^- \mathbf{1}_B. \end{aligned}$$

(c) Se f non fosse finita q.o., almeno uno dei due insiemi

$$E_1 := \{f = +\infty\} \quad \text{e} \quad E_2 := \{f = -\infty\}$$

avrebbe misura strettamente positiva; si supponga, per esempio, che sia $\mu(E_1) > 0$. Poiché la definizione di integrale implica che, se $f \geq 0$, sia anche $\int f \, d\mu \geq 0$, dalla disuguaglianza $0 \leq f^+ \mathbf{1}_{E_1} \leq f^+$ scende

$$\int f^+ \, d\mu \geq \int_{E_1} f^+ \, d\mu = +\infty,$$

sicché f non sarebbe integrabile.

(d) Per la (c), si può supporre che f e g siano finite in tutto Ω . Allora

$$f + g = (f + g)^+ - (f + g)^- \quad \text{e} \quad f + g = f^+ - f^- + g^+ - g^-,$$

onde, dal confronto, $(f + g)^+ + f^- + g^- = (f + g)^- + f^+ + g^+$. Poiché

$$(f + g)^+ \leq f^+ + g^+ \quad \text{e} \quad (f + g)^- \leq f^- + g^-,$$

si ha, intanto, che $(f + g)^+$ e $(f + g)^-$ sono integrabili. L'asserto è noto per le funzioni misurabili positive e perciò

$$\int (f + g)^+ \, d\mu + \int f^- \, d\mu + \int g^- \, d\mu = \int (f + g)^- \, d\mu + \int f^+ \, d\mu + \int g^+ \, d\mu,$$

relazione dalla quale scende facilmente quanto si voleva dimostrare.

(e) La prima asserzione è ovvia; quanto alla seconda si scriva $f = g + (f - g)$, ove $f - g \geq 0$. Perciò

$$\int f \, d\mu = \int g \, d\mu + \int (f - g) \, d\mu \geq \int g \, d\mu.$$

(f) Si applichi la (d) a $|f| = f^+ + f^-$ che risulta quindi integrabile; inoltre

$$\left| \int f \, d\mu \right| = \left| \int f^+ \, d\mu - \int f^- \, d\mu \right| \leq \int f^+ \, d\mu + \int f^- \, d\mu = \int |f| \, d\mu.$$

(g) Se $c = 0$, è $cf = 0$ e $c \int f \, d\mu = 0 = \int (cf) \, d\mu$. Se $c > 0$, è $(cf)^+ = cf^+$ e $(cf)^- = cf^-$, sicché

$$\int cf \, d\mu = \int cf^+ \, d\mu - \int cf^- \, d\mu = c \int f^+ \, d\mu - c \int f^- \, d\mu = c \int f \, d\mu.$$

Se, invece, $c < 0$, si ha $(cf)^+ = -cf^-$ e $(cf)^- = -cf^+$, onde

$$\begin{aligned} \int cf \, d\mu &= \int (cf)^+ \, d\mu - \int (cf)^- \, d\mu \\ &= \int (-c)f^- \, d\mu + c \int f^+ \, d\mu = c \int f \, d\mu. \end{aligned}$$

(h) Si osservi che, definendo, per ogni $n \in \mathbb{N}$, $A_n := \{f \geq 1/n\}$, si ha $A_n \in \mathcal{F}$ per ogni $n \in \mathbb{N}$ e si può scrivere $\{f > 0\} = \cup_{n \in \mathbb{N}} A_n$; quindi $\mu(f > 0) \leq \sum_{n \in \mathbb{N}} \mu(A_n)$. D'altro canto, per ogni $n \in \mathbb{N}$, si ha

$$\frac{\mathbf{1}_{A_n}}{n} \leq f \mathbf{1}_{A_n} \leq f,$$

onde

$$\int f \, d\mu \geq \frac{1}{n} \int \mathbf{1}_{A_n} \, d\mu = \frac{\mu(A_n)}{n} \geq 0;$$

perciò $\mu(A_n) = 0$ per ogni $n \in \mathbb{N}$, e, di conseguenza $\mu(\{f > 0\}) = 0$.

(i) $f = g$ q.o. implica $f^+ = g^+$ q.o. e $f^- = g^-$ q.o.. Gli insiemi $A_{k,s}$, che nel Teorema 1.8.7 servivano per costruire l'approssimazione di una funzione misurabile positiva mediante funzioni semplici, hanno, per f^+ e per g^+ , la stessa misura; perciò, se (s_n) e (s'_n) sono successioni crescenti di funzioni semplici tendono a f^+ e a g^+ rispettivamente, si ha $\int s_n \, d\mu = \int s'_n \, d\mu$ per ogni $n \in \mathbb{N}$ e, di qui, $\int f^+ \, d\mu = \int g^+ \, d\mu$. Similmente si mostra che $\int f^- \, d\mu = \int g^- \, d\mu$.

(j) Si ha $0 \leq h^+ \leq |f|$ e $0 \leq h^- \leq |f|$; perciò gli integrali delle funzioni semplici che approssimano h^+ e h^- sono crescenti e entrambe limitate superiormente. \square

Segue, in particolare, dalle proprietà (d) e (g) del Teorema 1.10.1 che l'insieme $\mathcal{L}^1(\Omega, \mathcal{F}, \mu)$ delle funzioni $f : \Omega \rightarrow \overline{\mathbb{R}}$ integrabili rispetto alla misura μ è uno spazio lineare.

I teoremi che chiudono questa sezione sono tra quelli di uso più frequente.

Teorema 1.10.2. (Teorema di Beppo Levi o di convergenza monotona). *Dato lo spazio misurale $(\Omega, \mathcal{F}, \mu)$, se (f_n) è una successione crescente di funzioni misurabili positive che converge puntualmente alla funzione f (necessariamente misurabile e positiva), si può passare al limite sotto il segno d'integrale*

$$\lim_{n \rightarrow +\infty} \int f_n \, d\mu = \int f \, d\mu,$$

nel senso che, se f è integrabile, si ha $\int f_n \, d\mu \rightarrow \int f \, d\mu$, mentre, se $\int f \, d\mu = +\infty$, allora $\int f_n \, d\mu \rightarrow +\infty$.

Dimostrazione. Per ogni $n \in \mathbb{N}$ sia $(s_{nk})_{k \in \mathbb{N}}$ una successione crescente di funzioni semplici positive tendente a $f_n := \lim_{k \rightarrow +\infty} s_{nk} = f_n$. Posto

$$g_k := \max\{s_{nk} : n \leq k\},$$

la successione (g_k) è crescente; inoltre, la funzione g_k è semplice e positiva. Perciò $g := \lim_{k \rightarrow +\infty} g_k$ è una funzione misurabile positiva. Ora, si ha, per ogni $n \in \mathbb{N}$ e per $k \geq n$, $s_{nk} \leq g_k \leq f_k \leq f$, sicché, facendo tendere k a $+\infty$, si ha, per ogni $n \in \mathbb{N}$, $f_n \leq g \leq f$, disuguaglianza dalla quale scende, facendo tendere n a $+\infty$, che $g = f$. Siccome

$$\int s_{nk} \, d\mu \leq \int g_k \, d\mu \leq \int f_k \, d\mu \quad (n \leq k),$$

e, per la definizione di integrale di una funzione misurabile positiva,

$$\sup_{n \in \mathbb{N}} \int g_k \, d\mu = \int g \, d\mu,$$

segue facendo tendere k a $+\infty$, che

$$\int f_n \, d\mu \leq \int g \, d\mu \leq \lim_{k \rightarrow +\infty} \int f_k \, d\mu;$$

si faccia ora tendere n a $+\infty$ per ottenere

$$\lim_{n \rightarrow +\infty} \int f_n \, d\mu \leq \int g \, d\mu \leq \lim_{k \rightarrow +\infty} \int f_k \, d\mu,$$

cioè

$$\lim_{n \rightarrow +\infty} \int f_n \, d\mu = \int g \, d\mu = \int f \, d\mu,$$

vale a dire l'asserto. □

Naturalmente lo stesso risultato vale se, anziché avere la convergenza puntuale di (f_n) a f , la successione (f_n) converge a f quasi ovunque.

Teorema 1.10.3. (Fatou). *Se (f_n) è una successione di funzioni misurabili limitata inferiormente da una funzione integrabile g , $f_n \geq g \in \mathcal{L}^1$ per ogni $n \in \mathbb{N}$, allora*

$$\int \liminf_{n \rightarrow +\infty} f_n \, d\mu \leq \liminf_{n \rightarrow +\infty} \int f_n \, d\mu.$$

Dimostrazione. Si supponga, dapprima, $g = 0$. Posto $g_n := \inf_{k \geq n} f_k$ ($n \in \mathbb{N}$), (g_n) risulta essere una successione crescente di funzioni misurabili positive con

$$\lim_{n \rightarrow +\infty} g_n = \liminf_{n \rightarrow +\infty} f_n.$$

Poiché $f_n \geq g_n$ per ogni $n \in \mathbb{N}$, il teorema di Beppo Levi dà

$$\liminf_{n \rightarrow +\infty} \int f_n \, d\mu \geq \lim_{n \rightarrow +\infty} \int g_n \, d\mu = \int \lim_{n \rightarrow +\infty} g_n \, d\mu = \int \liminf_{n \rightarrow +\infty} f_n \, d\mu.$$

Se, invece, g non è identicamente nulla, basta applicare quanto appena dimostrato alla successione (h_n) con $h_n := f_n - g \geq 0$, tenendo presente che

$$\int h_n \, d\mu = \int f_n \, d\mu - \int g \, d\mu \quad \text{e} \quad \liminf_{n \rightarrow +\infty} h_n = \liminf_{n \rightarrow +\infty} f_n - g,$$

che conclude la dimostrazione. □

Si applichi il lemma di Fatou alla successione $(-f_n)$.

Teorema 1.10.4. *Se (f_n) è una successione di funzioni misurabili limitata superiormente da una funzione integrabile g , $f_n \leq g \in \mathcal{L}^1$ per ogni $n \in \mathbb{N}$, allora*

$$\int \limsup_{n \rightarrow +\infty} f_n \, d\mu \geq \limsup_{n \rightarrow +\infty} \int f_n \, d\mu.$$

Teorema 1.10.5. (Teorema di Lebesgue o di convergenza dominata). *Dato lo spazio misurale $(\Omega, \mathcal{F}, \mu)$, sia $g : \Omega \rightarrow \overline{\mathbb{R}}_+$ una funzione integrabile e sia (f_n) una successione di funzioni misurabili $f_n : (\Omega, \mathcal{F}) \rightarrow (\overline{\mathbb{R}}, \mathcal{B})$ ($n \in \mathbb{N}$) che converge puntualmente (o quasi ovunque) alla funzione f ; se $|f_n| \leq g$ per ogni $n \in \mathbb{N}$, allora f è integrabile e si può passare al limite sotto il segno d'integrale:*

$$\lim_{n \rightarrow +\infty} \int f_n \, d\mu = \int f \, d\mu.$$

Dimostrazione. Se $f_n \geq 0$ e $f_n \rightarrow 0$, i Teoremi 1.10.3 e 1.10.4 danno

$$\begin{aligned} \int \limsup_{n \rightarrow +\infty} f_n \, d\mu &= \int \liminf_{n \rightarrow +\infty} f_n \, d\mu \leq \liminf_{n \rightarrow +\infty} \int f_n \, d\mu \\ &\leq \limsup_{n \rightarrow +\infty} \int f_n \, d\mu \leq \int \limsup_{n \rightarrow +\infty} f_n \, d\mu; \end{aligned}$$

le disuguaglianze scritte sono, in effetti, tutte eguaglianze e perciò

$$\lim_{n \rightarrow +\infty} \int f_n \, d\mu = 0.$$

In generale, si consideri la successione (h_n) definita, per ogni $n \in \mathbb{N}$, da $h_n := |f_n - f|$. Per questa successione sono verificate le condizioni della prima parte, $h_n \geq 0$ e $h_n \rightarrow 0$. Inoltre, vale $h_n \leq 2g$ per ogni $n \in \mathbb{N}$. Ora, per il Teorema 1.10.1 (f),

$$\left| \int f_n \, d\mu - \int f \, d\mu \right| \leq \int h_n \, d\mu \xrightarrow{n \rightarrow +\infty} 0,$$

che dà l'asserto. □

Esempio 1.10.1. Sia Ω un insieme costituito da un'infinità numerabile o da un numero finito di punti ω_n . Assegnare sulla famiglia delle parti di Ω una misura μ equivale a dare numeri positivi $\mu_n = \mu(\{\omega_n\})$. Per ogni sottoinsieme A di Ω , si ha

$$\mu(A) = \sum_{\omega_n \in A} \mu_n.$$

Naturalmente, si avrà $\sum_n \mu_n = \mu(\Omega)$. Ogni funzione $f : \Omega \rightarrow \mathbb{R}$ è misurabile rispetto alla tribù $\mathcal{P}(\Omega)$; ha senso, perciò, considerare l'integrale

$$\int |f| \, d\mu = \sum_n \int_{\{\omega_n\}} |f| \, d\mu = \sum_n |f(\omega_n)| \mu_n.$$

La funzione f è dunque integrabile se, e solo se, $\sum_n |f(\omega_n)| \mu_n < +\infty$; in tal caso è

$$\int f \, d\mu = \sum_n f(\omega_n) \mu_n. \quad (1.10.1)$$

In questo modo, ogni serie assolutamente convergente, oppure ogni somma finita di numeri reali, $\sum_n \alpha_n$ può essere riguardata come l'integrale di un'opportuna funzione rispetto alla *misura del contare*, vale a dire la misura ν definita sulla famiglia delle parti di \mathbb{N} mediante $\nu(\{n\}) = 1$ per ogni $n \in \mathbb{N}$. Basta, infatti, definire una funzione f mediante $f(n) := \alpha_n$ per avere, secondo la (1.10.1),

$$\sum_n \alpha_n = \sum_n f(n) = \sum_n f(n) \nu(\{n\}) = \int f \, d\nu.$$

La misura del contare non è finita ed attribuisce ad ogni sottoinsieme di \mathbb{N} misura eguale a $+\infty$ se l'insieme è infinito, ed eguale alla sua cardinalità se è finito. ■

1.11 Misura immagine

Siano $(\Omega, \mathcal{F}, \mu)$ uno spazio misurale, $(\Omega_1, \mathcal{F}_1)$ uno spazio misurabile e f una funzione misurabile tra (Ω, \mathcal{F}) e $(\Omega_1, \mathcal{F}_1)$. Si può allora definire una funzione ν da \mathcal{F}_1 in $\overline{\mathbb{R}}_+$ mediante

$$\nu(E) := \mu [f^{-1}(E)] \quad (E \in \mathcal{F}_1).$$

è immediato controllare che ν così definita è una misura; essa si dice *misura immagine di μ mediante f* . Poiché $\nu(\Omega_1) = \mu(\Omega)$, ν è finita se, e solo se, tale è anche μ . Si è soliti indicare la misura immagine mediante μf^{-1} .

Vale il seguente importante teorema sul “cambio di variabile”.

Teorema 1.11.1. *Siano (Ω, \mathcal{F}) e $(\Omega_1, \mathcal{F}_1)$ due spazi misurabili e μ una misura su (Ω, \mathcal{F}) ; se la funzione $g : \Omega_1 \rightarrow \overline{\mathbb{R}}$ è \mathcal{F}_1 -misurabile, è*

$$\int_{\Omega_1} g \, d(\mu f^{-1}) = \int_{\Omega} g \circ f \, d\mu, \quad (1.11.1)$$

nel senso che, se esiste uno dei due integrali, esiste anche l'altro e i due sono eguali.

Dimostrazione. Alla luce della definizione di integrale, ci si può limitare a considerare funzioni $g : \Omega_1 \rightarrow \overline{\mathbb{R}}$ positive. Sia, dapprima, $g = \mathbf{1}_E$ con $E \in \mathcal{F}_1$; allora

$$(g \circ f)(\omega) = \begin{cases} 1, & \text{se } \omega \in f^{-1}(E), \\ 0, & \text{se } \omega \in f^{-1}(E)^c; \end{cases}$$

perciò, $(\mathbf{1}_E \circ f) = \mathbf{1}_{f^{-1}(E)}$ con $f^{-1}(E) \in \mathcal{F}$ e

$$\int_{\Omega} g \circ f \, d\mu = \int_{\Omega} \mathbf{1}_{f^{-1}(E)} \, d\mu = (\mu f^{-1})(E) = \nu(E) = \int_{\Omega_1} \mathbf{1}_E \, d(\mu f^{-1}).$$

La (1.11.1) vale dunque per le funzioni indicatrici di insiemi di \mathcal{F}_1 e, come conseguenza della linearità degli integrali, per le funzioni \mathcal{F}_1 -semplici positive. Sia g

una funzione \mathcal{F}_1 -misurabile e positiva e sia (s_n) una successione crescente di funzioni semplici convergente a g ; allora $(s_n \circ f)$ tende a $g \circ f$ crescendo e, di piú, ogni funzione $s_n \circ f$ è semplice. Il ricorso alla definizione di integrale di una funzione positiva misurabile completa la dimostrazione. \square

La dimostrazione appena data è il prototipo di molte altre: si stabilisce il risultato per le funzioni semplici, quindi lo si estende alle funzioni semplici positive, quindi, ancora, alle funzioni misurabili positive ed infine alle funzioni integrabili, sicché, di fatto, la dimostrazione si riduce a controllare che il risultato sia vero per le sole funzioni indicatrici.

Sia μ una misura finita e $f : \Omega \rightarrow \overline{\mathbb{R}}$ una funzione misurabile; si definisca una funzione in \mathbb{R} una funzione F mediante $F(x) := \mu(\{f \leq x\})$. Si dimostra facilmente che F è crescente e continua a destra e che

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{e} \quad \lim_{x \rightarrow +\infty} F(x) = \mu(\Omega).$$

Per la dimostrazione basta osservare che, per la misura immagine $\nu = \mu \circ f^{-1}$ su $(\mathbb{R}, \mathcal{B})$, si ha $F(x) = \nu(] - \infty, x])$ e tenere presente il Teorema 1.7.1. Se μ_F è la misura di Stieltjes associata a F (Teorema 1.7.2), vale il

Teorema 1.11.2. *Sia $(\Omega, \mathcal{F}, \mu)$ è uno spazio misurabile e $f : \Omega \rightarrow \overline{\mathbb{R}}$ una funzione misurabile; se la funzione $F : \mathbb{R} \rightarrow [0, \mu(\Omega)]$ è definita, per ogni $x \in \mathbb{R}$ da $F(x) = \mu(f \leq x)$ risulta $\mu f^{-1} = \mu_F$; inoltre se $g : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ è misurabile vale l'eguaglianza*

$$\int_{\Omega} g \circ f \, d\mu = \int_{\mathbb{R}} g \, d\mu_F = \int_{\mathbb{R}} g \, d(\mu f^{-1}), \quad (1.11.2)$$

nel senso che, se esiste uno dei due integrali, esiste anche l'altro e i due sono eguali.

Dimostrazione. Alla luce del Teorema 1.6.7, per dimostrare l'eguaglianza

$$\mu_F(B) = (\mu f^{-1})(B)$$

per ogni boreliano B basta mostrare che essa è vera per ogni intervallo $]a, b]$ con $a, b \in \mathbb{R}$ e $a < b$; in tal caso,

$$(\mu f^{-1})(]a, b]) = \mu(a < f \leq b) = F(b) - F(a) = \mu_F(]a, b]),$$

sicché risulta effettivamente $\mu f^{-1} = \mu_F$.

Per dimostrare la (1.11.2), si ricorra al teorema del cambio di variabile 1.11.1; poiché $\mu f^{-1} = \mu_F$ la (1.11.1) dà appunto la (1.11.2). \square

Spesso l'integrale a secondo membro della (1.11.2) si scrive nella forma

$$\int_{\mathbb{R}} g \, d\mu_F = \int_{\mathbb{R}} g(x) \, dF(x) = \int_{\mathbb{R}} g \, dF,$$

che si dice *integrale di Lebesgue-Stieltjes della funzione f rispetto alla f.r. F* .

Se nella (1.11.2) si prende come integrando $g : \mathbb{R} \rightarrow \mathbb{R}$ la funzione $g(x) := x^k$ con $k \in \mathbb{Z}_+$ si ottiene la formula che è spesso utile per il calcolo effettivo dei momenti

$$\int_{\Omega} f^k \, d\mu = \int_{\mathbb{R}} x^k \, d\mu_F(x) = \int_{\mathbb{R}} x^k \, dF(x).$$

1.12 Vocabolario

Diversi dei concetti che si conoscono dai corsi elementari hanno in probabilità un nome diverso da quello che hanno in un altro contesto; si rende perciò necessario un “vocabolario” che consenta di tradurre i nomi.

Definizione 1.12.1. Si dirà *spazio di probabilità* o *spazio probabilizzato* uno spazio misurale $(\Omega, \mathcal{F}, \mathbb{P})$ nel quale la misura \mathbb{P} , detta, *probabilità* o *misura di probabilità* sia normalizzata, $\mathbb{P}(\Omega) = 1$. \diamond

Definizione 1.12.2. Dato uno spazio misurabile (Ω, \mathcal{F}) si dice *variabile aleatoria* (o *casuale*) (spesso abbreviato in v.a.) una funzione $X : \Omega \rightarrow \overline{\mathbb{R}}$ che sia misurabile rispetto alle tribù \mathcal{F} e \mathcal{B} . Qui, \mathcal{B} è la tribù di Borel in $\overline{\mathbb{R}}$. \diamond

Benché nella definizione di v.a. non vi sia menzione di una probabilità definita su (Ω, \mathcal{F}) , si suole parlare di v.a. sullo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ perché i problemi tipici di calcolo delle probabilità sono solitamente riconducibili alla forma: *Qual è la probabilità che una v.a. X assuma valori in un boreliano A ?*

Le v.a. sono, solitamente, indicate mediante le ultime lettere maiuscole dell’alfabeto, con suffissi, ove occorra.

Alla nozione di proprietà valida quasi ovunque, si sostituisce, con lo stesso significato, quella di proprietà valida *quasi certamente* (abbreviato in q.c.). Altri cambiamenti di linguaggio si incontreranno tra breve.

Nel seguito si considereranno quasi esclusivamente v.a. X q.c. finite, tali, cioè, che $\mathbb{P}(|X| = +\infty) = 0$. Perciò, di fatto, le v.a. saranno funzioni definite in Ω a valori in \mathbb{R} anziché in $\overline{\mathbb{R}}$. Per indicare che X è una v.a. scriveremo anche $X \in L^0$; e ricorreremo all’abuso di linguaggio che consiste nel considerare piuttosto le classi di equivalenza che non le singole funzioni. Un tale atteggiamento non può però essere adottato nel trattare di processi stocastici, che saranno però toccati solo marginalmente in queste lezioni.

Alla luce del Corollario 1.8.1, per verificare che una funzione $X : \Omega \rightarrow \mathbb{R}$ sia una v.a. basta, per esempio, mostrare che, per ogni $x \in \mathbb{R}$, appartiene alla tribù \mathcal{F} l’insieme $\{X < x\} = X^{-1}([-\infty, x[)$.

Definizione 1.12.3. Data una v.a. X su $(\Omega, \mathcal{F}, \mathbb{P})$ si dice *legge* o *distribuzione* di X la misura immagine su $(\mathbb{R}, \mathcal{B})$ di \mathbb{P} mediante X , cioè la misura \mathbb{P} su $(\mathbb{R}, \mathcal{B})$ definita da $\mathbb{P}_X(B) := \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B)$ per ogni insieme $B \in \mathcal{B}$. \diamond

Ogni v.a. X induce quindi una probabilità \mathbb{P} su \mathcal{B} . Si dice *funzione di ripartizione* (o *di distribuzione*), di solito abbreviato in f.r., della v.a. X la funzione F_X da \mathbb{R} in $[0, 1]$ definita da

$$F_X(x) := \mathbb{P}_X([-\infty, x]) = \mathbb{P}(X \leq x) \quad (x \in \mathbb{R}).$$

Si scriverà F , in vece di F_X , per la f.r. di una v.a. X tutte le volte che ciò non generi confusione.

Definizione 1.12.4. Due v.a. X e X' , non necessariamente definite sopra il medesimo spazio di probabilità, si dicono *isonome* o *identicamente distribuite*, o, ancora *somiglianti*, se hanno la stessa legge, cioè se $\mathbb{P}_X = \mathbb{P}_{X'}$. Due v.a. sono *isonome* se, e solo se, esse hanno la stessa f.r.. \diamond

Analogamente, se una v.a. X definita in $(\Omega, \mathcal{F}, \mathbb{P})$ ammette integrale finito, tale integrale si indicherà mediante la notazione

$$\mathbb{E}(X) := \int X \, d\mathbb{P},$$

che si dirà *speranza (matematica)* di X . Si dice *varianza* di X , se esiste, finita, la quantità

$$V(X) := \mathbb{E}[(X - \mathbb{E}(X))^2].$$

Teorema 1.12.1. (Diseguaglianza di Markov). *Sia X una v.a. sullo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$; se X è positiva ($X \geq 0$) si ha, quale che sia il numero reale $b > 0$,*

$$\mathbb{P}(X \geq b) \leq \frac{1}{b} \mathbb{E}(X).$$

Dimostrazione. Poiché l'integrale che definisce la speranza è positivo, si ha $0 \leq \mathbb{E}(X) \leq +\infty$. Se $\mathbb{E}(X) = +\infty$, non vi è nulla da dimostrare. Si supponga, perciò, che sia $\mathbb{E}(X) < +\infty$, vale a dire, si supponga che X sia integrabile, $X \in L^1$, e si ponga $A(b) := \{X > b\}$. Poiché è evidente che $X \geq X \cdot \mathbf{1}_{A(b)}$, risulta, dalla definizione di speranza, che

$$\begin{aligned} \mathbb{E}(X) &\geq \mathbb{E}(X \cdot \mathbf{1}_{A(b)}) \geq \mathbb{E}(b \cdot \mathbf{1}_{A(b)}) \\ &= b \mathbb{E}(\mathbf{1}_{A(b)}) = b \mathbb{P}(A(b)) = b \mathbb{P}(X \geq b), \end{aligned}$$

col che la diseguaglianza di Markov è provata. □

Il seguente è un corollario immediato della diseguaglianza di Markov.

Corollario 1.12.1. *Se c è un numero reale, $\varepsilon > 0$ e $n \in \mathbb{N}$, allora*

$$\mathbb{P}(|X - c| \geq \varepsilon) \leq \frac{\mathbb{E}(|X - c|^n)}{\varepsilon^n}. \quad (1.12.1)$$

Se, inoltre, la v.a. X ammette speranza e varianza finite si ha, per ogni numero reale $k > 0$,

$$\mathbb{P}\left(|X - \mathbb{E}(X)| \geq k\sqrt{V(X)}\right) \leq \frac{1}{k^2}. \quad (1.12.2)$$

Per la dimostrazione della (1.12.2) basta porre $c = \mathbb{E}(X)$ e $\varepsilon = k\sqrt{V(X)}$ nella (1.12.1). La (1.12.2) è nota con il nome di *diseguaglianza di Čebyšev*. Si noti che la diseguaglianza (1.12.1) ha senso anche quando non esiste finito il momento di ordine n : infatti se fosse $\mathbb{E}(|X - c|^n) = +\infty$ la (1.12.1) diverrebbe $\mathbb{P}(|X - c| \geq \varepsilon) < +\infty$, che è banalmente vera.

1.13 Gli spazî L^p

Gli spazî L^p introdotti da F. Riesz sono importanti in molti rami dell'Analisi, ma sono essenziali anche in Probabilità. Qui di seguito diamo la definizione e le proprietà che sono indispensabili in un primo approccio alla probabilità.

Definizione 1.13.1. Dato uno spazio misurale $(\Omega, \mathcal{F}, \mu)$, si indica con

$$\mathcal{L}^p = \mathcal{L}^p(\Omega, \mathcal{F}, \mu) \quad (p \in [1, +\infty])$$

l'insieme delle funzioni misurabili f per le quali è integrabile $|f|^p$. \diamond

Dunque

$$\mathcal{L}^p := \left\{ f \in \mathcal{L}^0 : \int |f|^p d\mu < +\infty \right\} .$$

Dato lo spazio misurale $(\Omega, \mathcal{F}, \mu)$, due funzioni misurabili, $f, g : \Omega \rightarrow \overline{\mathbb{R}}$ si dicono *eguali* (μ)-q.o. se $\mu(f \neq g) = 0$; si scrive $f \simeq_\mu g$. È immediato controllare che l'eguaglianza q.o. è una relazione d'equivalenza. È pertanto naturale considerarne le classi d'equivalenza. Si ha l'abitudine di scrivere f per indicare tanto la funzione f quanto la sua classe d'equivalenza rispetto all'eguaglianza q.o., \simeq_μ .

Si denota con $L^p = L^p(\Omega, \mathcal{F}, \mu)$ lo spazio quoziente di \mathcal{L}^p rispetto alla relazione d'eguaglianza q.o.

$$L^p(\Omega, \mathcal{F}, \mu) := \mathcal{L}^p(\Omega, \mathcal{F}, \mu) / \simeq_\mu .$$

In appendice richiamiamo per comodità la definizione ed alcune proprietà delle funzioni convesse che saranno utili in queste lezioni.

In questa sezione supporremo che la misura μ sullo spazio misurabile (Ω, \mathcal{F}) sia una probabilità. I risultati che seguono valgono, con opportune modifiche, anche se μ è una misura finita; se viene meno questa ipotesi non tutti continuano a valere.

Introduce una certa semplificazione nelle formule adottare la notazione, usuale in probabilità,

$$\mathbb{E}(f) := \int f d\mu$$

quando si considera l'integrale di una funzione misurabile f (che sia integrabile).

Teorema 1.13.1. (Jensen). *Sia (a, b) un intervallo di \mathbb{R} (con $-\infty \leq a < b \leq +\infty$), sia $\varphi : (a, b) \rightarrow \mathbb{R}$ convessa e sia f una funzione misurabile che assume valori in (a, b) . Allora, se f è integrabile, vale la diseguaglianza*

$$\varphi(\mathbb{E}(f)) \leq \mathbb{E}(\varphi \circ f) . \quad (1.13.1)$$

Dimostrazione. Si osservi, innanzi tutto, che $\mathbb{E}(f)$ è in (a, b) . Supponiamo, infatti che l'estremo a appartenga all'intervallo (a, b) ; in tal caso, non vi è nulla da provare perché, per ipotesi $f \geq a$ e, dunque, per l'isotonia dell'integrale, anche $\mathbb{E}(f) \geq a$. Supponiamo allora che l'intervallo (a, b) sia aperto in a , cioè che $f > a$. In tal caso, l'isotonia dà direttamente solo $\mathbb{E}(f) \geq a$, come prima, ma non necessariamente $\mathbb{E}(f) > a$, come asserito; se però fosse $\mathbb{E}(f) = a$, si avrebbe che è nullo l'integrale della funzione misurabile positiva $f - a$, e ciò accade solo se la funzione in questione è nulla q.o., vale a dire $f = a$ q.o., contrariamente all'ipotesi. Dunque, $\mathbb{E}(f) > a$. Analogamente si ragiona per l'altro estremo b .

Inoltre, φ è continua, sicché $\varphi \circ f$ è misurabile. Dalla (1.17.4) dell'Appendice segue che

$$(\varphi \circ f)(\omega) \geq \varphi(\mathbb{E}(f)) + \alpha \{f(\omega) - \mathbb{E}(f)\} ,$$

onde, integrando, si ottiene

$$\mathbb{E}(\varphi \circ f) \geq \varphi(\mathbb{E}(f)) + \alpha \{\mathbb{E}(f) - \mathbb{E}(f)\} ,$$

che conclude la dimostrazione. \square

Corollario 1.13.1. *Se $r \geq 1$ e $s > 0$*

$$\mathbb{E}^r(|f|) \leq \mathbb{E}(|f|^r), \quad (1.13.2)$$

$$\mathbb{E}^r(|f|^s) \leq \mathbb{E}(|f|^{rs}). \quad (1.13.3)$$

Dimostrazione. Si applichi la (1.13.1) alla funzione

$$x \mapsto \varphi(x) := x^r, \quad (x \geq 0, r \geq 1)$$

ottenendo così la (1.13.2). Applicando la disuguaglianza (1.13.2) così trovata alla funzione misurabile $g := |f|^s$, si ha la (1.13.3). \square

Il risultato che segue è conseguenza immediata del precedente. Lo isoliamo per metterne in risalto l'importanza per ciò che seguirà.

Corollario 1.13.2. *Se $0 < p < q$ risulta*

$$\mathbb{E}^{1/p}(|f|^p) \leq \mathbb{E}^{1/q}(|f|^q). \quad (1.13.4)$$

Dimostrazione. Nella (1.13.3) si ponga $r = q/p$ e $s = p$. \square

La (1.13.4) asserisce che $\mathcal{L}^q(\Omega, \mathcal{F}, \mu)$ è incluso in $\mathcal{L}^p(\Omega, \mathcal{F}, \mu)$ se $0 < p < q$, vale a dire $\mathcal{L}^q \subseteq \mathcal{L}^p$. Analogamente, si ha $L^q \subseteq L^p$. Ponendo $\|\varphi\|_p := \mathbb{E}^{1/p}(|\varphi|^p)$ per una funzione misurabile φ di L^p con $p \geq 1$, la (1.13.4) si scrive $\|\varphi\|_p \leq \|\varphi\|_q$ ($p < q$).

Lemma 1.13.1. *Siano x, y, α, β numeri strettamente positivi con $\alpha + \beta = 1$. Siano, inoltre, $s, t > 0$, $p, q > 1$ con $\frac{1}{p} + \frac{1}{q} = 1$. Allora,*

$$x^\alpha y^\beta \leq \alpha x + \beta y, \quad (1.13.5)$$

$$s t \leq \frac{s^p}{p} + \frac{t^q}{q}. \quad (1.13.6)$$

Dimostrazione. La (1.13.5) equivale alla disuguaglianza $\alpha \ln x + \beta \ln y \leq \ln(\alpha x + \beta y)$ che è ovvia perché la funzione logaritmo $\ln :]0, +\infty[\rightarrow \mathbb{R}$ è concava (ciò che vuol dire che $-\ln$ è convessa). Per ottenere la (1.13.6) basta poi porre

$$\alpha = \frac{1}{p}, \quad \beta = \frac{1}{q}, \quad x = s^p, \quad y = t^q,$$

concludendo così la dimostrazione. \square

Si osservi che la (1.13.5) contiene, come caso particolare, la ben nota disuguaglianza tra la media geometrica e la media aritmetica di due numeri positivi x e y ,

$$\sqrt{xy} \leq \frac{x+y}{2}.$$

Teorema 1.13.2. (Hölder). *Siano $p, q \in]1, +\infty[$ con*

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Se f appartiene a L^p , $f \in L^p$, e g appartiene a L^q , $g \in L^q$, allora il prodotto fg appartiene a L^1 e

$$\|fg\|_1 \leq \|f\|_p \|g\|_q, \quad (1.13.7)$$

cioè

$$\mathbb{E}(|f g|) \leq \mathbb{E}^{|f|^p} \mathbb{E}^{|g|^q}(|f|^q).$$

Dimostrazione. Non vi è nulla da dimostrare se $\mathbb{E}^{|f|^p} \mathbb{E}^{|g|^q}(|f|^q) = 0$. Si supponrà perciò che sia $\mathbb{E}^{|f|^p} \mathbb{E}^{|g|^q}(|f|^q) > 0$. Nella (1.13.6) si ponga

$$s = \frac{|f|}{\|f\|_p} \quad \text{e} \quad t = \frac{|g|}{\|g\|_q},$$

per ottenere

$$\frac{|fg|}{\|f\|_p \|g\|_q} \leq \frac{1}{p} \frac{|f|^p}{\|f\|_p^p} + \frac{1}{q} \frac{|g|^q}{\|g\|_q^q},$$

dalla quale si ricava la (1.13.7) integrando. \square

Per $p = q = 2$, si ha la disuguaglianza di Schwarz, valida per le funzioni di L^2 ,

$$\|f g\|_1 \leq \|f\|_2 \|g\|_2 \quad \left(\mathbb{E}(|f g|) \leq \mathbb{E}^{|f|} \mathbb{E}^{|g|}(|g|) \right).$$

Dalla disuguaglianza di Hölder segue quella di Minkowski.

Teorema 1.13.3. (Minkowski). *Se f e g appartengono a L^p con $p \geq 1$, allora*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p. \quad (1.13.8)$$

Dimostrazione. Non vi è nulla da dimostrare se $p = 1$; si supponga, perciò, che sia $p > 1$. È, in primo luogo, evidente che $|f + g|^p$ è integrabile, o, equivalentemente, che $f + g$ appartiene a L^p , perché

$$|f + g|^p \leq 2^p (|f|^p \vee |g|^p).$$

Dalla disuguaglianza di Hölder scende se, al solito, q è tale che $1/p + 1/q = 1$,

$$\begin{aligned} \mathbb{E}(|f + g|^p) &\leq \mathbb{E}(|f| |f + g|^{p-1}) + \mathbb{E}(|g| |f + g|^{p-1}) \\ &\leq \|f\|_p \mathbb{E}^{|f + g|^{p-1}} + \|g\|_p \mathbb{E}^{|f + g|^{p-1}} \\ &= (\|f\|_p + \|g\|_p) \mathbb{E}^{|f + g|^{p-1}}, \end{aligned}$$

che equivale alla (1.13.8). \square

La disuguaglianza di Minkowski assicura che L^p , con $p \geq 1$ è uno spazio vettoriale e che $\|\cdot\|_p : L^p \rightarrow \mathbb{R}_+$ è una norma su L^p (ma non su \mathcal{L}^p). Lo spazio L^p è uno spazio normato che risulta essere completo (Teorema di Fischer–Riesz). Ricordiamo che si chiama *spazio di Banach* uno spazio normato che sia completo rispetto alla topologia della norma. Dunque L^p per $p \geq 1$ è uno spazio di Banach, mentre L^2 è uno spazio di Hilbert con il prodotto interno definito da

$$\langle f, g \rangle := \mathbb{E}(f \bar{g}) = \int f \bar{g} dP.$$

1.14 Misure definite da una densità

Siano μ una misura su (Ω, \mathcal{F}) e $f : \Omega \rightarrow \overline{\mathbb{R}}_+$ una funzione misurabile positiva. Si definisca una funzione $\nu : \mathcal{F} \rightarrow \overline{\mathbb{R}}_+$ mediante

$$\nu(A) := \int_A f \, d\mu \quad (A \in \mathcal{F}), \quad (1.14.1)$$

allora ν è una misura su (Ω, \mathcal{F}) . Infatti, se $A = \cup_{n \in \mathbb{N}} A_n$ con $A_n \in \mathcal{F}$ per ogni $n \in \mathbb{N}$ e $A_i \cap A_j = \emptyset$ se $i \neq j$, si ha, e si vedano gli esercizi per giustificare taluni dei passaggi,

$$\nu(A) = \int \mathbf{1}_A f \, d\mu = \sum_{n \in \mathbb{N}} \int \mathbf{1}_{A_n} f \, d\mu = \sum_{n \in \mathbb{N}} \int f \, d\mu = \sum_{n \in \mathbb{N}} \nu(A_n).$$

La misura ν è detta *di densità f e di base μ* ; solitamente la si indica con $\nu = f \cdot \mu$.

Se la funzione f è integrabile, $f \in L^1(\mu)$, la misura ν è finita. Se poi, f è a valori in $\overline{\mathbb{R}}$ anziché in $\overline{\mathbb{R}}_+$, la (1.14.1) definisce in \mathcal{F} una funzione ν che è ancora numerabilmente additiva, anche se non con valori positivi. Una funzione $\nu : \mathcal{F} \rightarrow \overline{\mathbb{R}}$ che sia σ -additiva si dice *misura reale* o *con segno*. In queste lezioni non ci occuperemo delle misure reali.

Teorema 1.14.1. *Se $g : \Omega \rightarrow \overline{\mathbb{R}}$ è integrabile rispetto alla misura $\nu = f \cdot \mu$, è*

$$\int g \, d\nu = \int g f \, d\mu. \quad (1.14.2)$$

Dimostrazione. La (1.14.2) è ovvia quando g è una funzione indicatrice, $g = \mathbf{1}_A$ con $A \in \mathcal{F}$; in tal caso, infatti, la (1.14.2) coincide con la (1.14.1). La (1.14.2) vale allora per le funzioni semplici a causa della linearità dell'integrale e, in virtù della definizione di integrale, per le funzioni positive misurabili. Ma è allora anche vera per una generica funzione integrabile. \square

Lemma 1.14.1. *Per una funzione integrabile $f : \Omega \rightarrow \overline{\mathbb{R}}$ sono equivalenti le affermazioni:*

- (a) $f = 0$ μ -q.o.;
- (b) $\int_A f \, d\mu = 0$ per ogni $A \in \mathcal{F}$.

Dimostrazione. Poiché l'implicazione (a) \implies (b) è ovvia, basterà dimostrare che vale l'implicazione inversa (b) \implies (a). Preso $A := \{f \geq 0\}$, si ha

$$\int f^+ \, d\mu = \int \mathbf{1}_A f \, d\mu = 0,$$

sicché $f^+ = 0$ in virtù del Teorema 1.10.1(h). Similmente si procede per f^- . \square

Scende dal Lemma 1.14.1 che se ν_1 e ν_2 hanno rispettivamente densità f_1 e f_2 rispetto alla stessa base μ , $\nu_1 = f_1 \cdot \mu$ e $\nu_2 = f_2 \cdot \mu$, allora ν_1 e ν_2 sono eguali se, e solo se, le loro densità sono eguali μ -q.o..

Se μ e ν sono misure sullo stesso spazio misurabile (Ω, \mathcal{F}) , si dice che ν è *assolutamente continua* rispetto a μ , e si scrive $\nu \ll \mu$, se si ha $\nu(A) = 0$ per ogni insieme $A \in \mathcal{F}$ tale che $\mu(A) = 0$. È evidente che la (1.14.1) definisce una misura assolutamente continua rispetto a μ ; è notevole il fatto che, con questa costruzione, si ottengano *tutte* le misure assolutamente continue rispetto a μ : se ν è assolutamente continua rispetto a μ esiste una densità f tale che $\nu = f \cdot \mu$. È questo il contenuto del teorema di Radon–Nikodym del quale daremo di seguito la dimostrazione. Si tratta di un risultato fondamentale per la Probabilità e la Statistica.

Ricordiamo che si chiama *spazio di Hilbert* uno spazio vettoriale H dotato di *prodotto interno* $\langle \cdot, \cdot \rangle : H \times H \rightarrow \mathbb{C}$ e completo rispetto alla topologia della norma indotta dal prodotto interno, $\|x\| := \langle x, x \rangle^{1/2}$ ($x \in H$). Esempi di spazi di Hilbert sono dati dagli spazi euclidei \mathbb{R}^n e dallo spazio $L^2(\Omega, \mathcal{F}, \mu)$.

Dato uno spazio lineare V , si chiama *duale (topologico)* V' di V l'insieme, che è anch'esso uno spazio lineare, dei funzionali lineari e continui definiti su V , cioè

$$V' := \{F : V \rightarrow \mathbb{R} \text{ oppure } \mathbb{C}, F \text{ lineare e continuo}\}.$$

Il seguente, fondamentale, teorema sostanzialmente identifica il duale H' di uno spazio di Hilbert H con H stesso.

Teorema 1.14.2. (Riesz–Fréchet). *Sia H uno spazio di Hilbert; allora, ogni funzionale lineare e continuo F su H può essere rappresentato in un'unica maniera come prodotto interno con un opportuno vettore $z \in H$:*

$$\forall F \in H' \quad \exists z \in H \quad \forall x \in H \quad F(x) = \langle x, z \rangle; \quad (1.14.3)$$

inoltre $\|F\| = \|z\|$.

Dimostrazione. Se F è identicamente nullo si ponga $z = 0$. Se invece F non è identicamente nullo, il nucleo M di F ,

$$M = \ker F := \{x \in H : F(x) = 0\} = F^{-1}(\{0\}),$$

non coincide con tutto H , e, poiché F è lineare e continuo, è un sottospazio proprio chiuso di H . Pertanto il complemento ortogonale M^\perp non si riduce al solo vettore nullo; qui,

$$M^\perp := \{y \in H : \forall x \in M \langle x, y \rangle = 0\}.$$

Sia ora $z_0 \in M^\perp$ con z_0 diverso dal vettore nullo; allora,

$$\forall x \in H \quad F\left(x - \frac{F(x)}{F(z_0)} z_0\right) = F(x) - F(x) = 0,$$

cioè

$$x - \frac{F(x)}{F(z_0)} z_0$$

appartiene a M per ogni $x \in H$. Perciò,

$$\forall x \in H \quad 0 = \langle x - \frac{F(x)}{F(z_0)} z_0, z_0 \rangle = \langle x, z_0 \rangle - \frac{F(x)}{F(z_0)} \langle z_0, z_0 \rangle,$$

che equivale a

$$F(x) = \frac{F(z_0)}{\|z_0\|^2} \langle x, z_0 \rangle,$$

per ogni $x \in H$. Si ha perciò la (1.14.3) ponendo

$$z := \frac{\overline{F(z_0)}}{\|z_0\|^2} z_0.$$

Si è così dimostrata l'esistenza di un vettore con la proprietà dell'asserto. Per provarne l'unicità, si supponga che sia, per ogni $x \in H$,

$$\langle x, z \rangle = \langle x, z' \rangle, \quad \text{cioè} \quad \langle x, z - z' \rangle = 0;$$

in particolare, per $x = z - z'$, si ha $\|z - z'\| = 0$, sicché $z = z'$.

Infine, segue dalla disuguaglianza di Schwarz che

$$|F(x)| = |\langle x, z \rangle| \leq \|x\| \|z\|,$$

onde

$$\frac{|F(x)|}{\|x\|} \leq \|z\|,$$

sicché $\|F\| := \sup\{|F(x)| : x \in H, \|x\| \leq 1\} \leq \|z\|$.

D'altro canto, l'ultimo asserto $\|F\| = \|z\|$ scende ora dall'eguaglianza $|F(z)| = |\langle z, z \rangle| = \|z\|^2 = \|z\| \|z\|$. \square

Siamo ora in grado di dimostrare il teorema di Radon–Nikodym.

Teorema 1.14.3. (Radon–Nikodym). *Siano λ e μ due misure sullo spazio misurabile (Ω, \mathcal{F}) e siano λ finita e μ σ -finita. Sono allora equivalenti le seguenti condizioni:*

- (a) λ è assolutamente continua rispetto a μ , $\lambda \ll \mu$;
- (b) esiste un'unica funzione $h \in L^1(\mu)$ con $h \geq 0$ μ -q.o. tale che $\lambda = h \cdot \mu$, cioè tale che si abbia, per ogni $A \in \mathcal{F}$,

$$\lambda(A) = \int_A h \, d\mu.$$

Dimostrazione. L'implicazione (b) \implies (a) è ovvia. Basta, perciò, dimostrare l'implicazione (a) \implies (b). Si supponga dapprima che μ sia una misura finita. Ponendo $\nu := \lambda + \mu$, si definisce una nuova misura finita su (Ω, \mathcal{F}) . Sia ora $f \in L^2(\nu)$; la disuguaglianza di Schwarz dà

$$\begin{aligned} \left| \int f \, d\lambda \right| &\leq \int |f| \, d\lambda \leq \int |f| \, d\nu \\ &\leq \left\{ \int |f|^2 \, d\nu \right\}^{1/2} \left\{ \int d\nu \right\}^{1/2} = \|f\|_{L^2(\nu)} \sqrt{\nu(\Omega)}. \end{aligned}$$

Perciò l'applicazione $f \mapsto \int f \, d\lambda$ definisce un funzionale lineare e continuo di $L^2(\nu)$. Per il Teorema di Riesz–Fréchet, esiste allora $g \in L^2(\nu)$ tale che, per ogni f di $L^2(\nu)$, risulti

$$\int f \, d\lambda = \int f g \, d\nu. \quad (1.14.4)$$

Se $A \in \mathcal{F}$ e $\nu(A) > 0$ si ha, ricordando che $\nu = \lambda + \mu$ e che, come conseguenza, $\nu \ll \mu$,

$$0 \leq \lambda(A) = \int \mathbf{1}_A \, d\lambda = \int_A g \, d\nu \quad \text{e} \quad \lambda(A) \leq \nu(A),$$

donde, per ogni insieme $A \in \mathcal{F}$ con $\nu(A) > 0$,

$$0 \leq \frac{1}{\nu(A)} \int_A g \, d\nu \leq 1. \quad (1.14.5)$$

Quest'ultima relazione implica $0 \leq g \leq 1$ ν -q.o.. Infatti, posto $A := \{g > 1\}$, si ha

$$A = \bigcup_{n \in \mathbb{N}} A_n = \bigcup_{n \in \mathbb{N}} \left\{ g > 1 + \frac{1}{n} \right\}.$$

Se esistesse un indice $n \in \mathbb{N}$ tale che $\nu(A_n) > 0$, la (1.14.5) implicherebbe

$$\frac{1}{\nu(A_n)} \int_{A_n} g \, d\nu \geq 1 + \frac{1}{n},$$

una contraddizione, sicché si ha $\nu(A_n) = 0$ per ogni $n \in \mathbb{N}$, e, di qui, $\nu(g > 1) = 0$. Analogamente si dimostra che $\mu(g < 0) = 0$. Si può perciò assumere, senza perdita di generalità, che sia $0 \leq g(\omega) \leq 1$, per ogni $\omega \in \Omega$. Segue ora dalla (1.14.4) che, per ogni $f \in L^2(\nu)$, si ha

$$\int f \, d\lambda = \int f g \, d\nu = \int f g \, d\lambda + \int f g \, d\mu,$$

cioè

$$\int f(1-g) \, d\lambda = \int f g \, d\mu. \quad (1.14.6)$$

Se $S := \{\omega \in \Omega : g(\omega) = 1\}$, la (1.14.6) dà, per $f = \mathbf{1}_S$,

$$\mu(S) = \int \mathbf{1}_S g \, d\mu = \int_S g \, d\mu = \int_S (1-g) \, d\lambda = 0.$$

Per l'assoluta continuità di λ rispetto a μ si ha $\lambda(S) = 0$ e $\nu(S) = 0$. Si può, quindi supporre che sia $g(\omega) < 1$ per ogni $\omega \in \Omega$ e porre, nella (1.14.6), $f = \mathbf{1}_A \sum_{k=0}^n g^k$ ($A \in \mathcal{F}$) per ottenere

$$\int_A (1 - g^{n+1}) \, d\lambda = \int_A g \left(\sum_{k=0}^n g^k \right) \, d\mu.$$

Si applichi, ad entrambi i membri, il Teorema di Beppo Levi 1.10.2,

$$\lambda(A) = \int_A h \, d\mu,$$

ove

$$h := \lim_{n \rightarrow \infty} g \left(\sum_{k=0}^n g^k \right) = g \left(\sum_{k=0}^{\infty} g^k \right) = \frac{g}{1-g}.$$

Si osservi che $h \geq 0$ e che $\int |h| d\mu = \lambda(\Omega) < +\infty$, sicché h è integrabile, $h \in L^1(\mu)$. Per dimostrare l'unicità di h , a meno di equivalenze, si supponga che esista h' in $L^1(\mu)$ tale che per ogni $A \in \mathcal{F}$ sia

$$\lambda(A) = \int_A h' d\mu.$$

Ma allora

$$\forall A \in \mathcal{F} \quad \int_A h' d\mu = \int_A h d\mu,$$

sicché $h = h'$ μ -q.o..

Si supponga ora che μ sia σ -finita: in questo caso $\Omega = \cup_{n \in \mathbb{N}} E_n$ con $\mu(E_n) < +\infty$ per ogni $n \in \mathbb{N}$. Non è restrittivo supporre che gli insiemi E_n siano disgiunti; se così non è, si può sostituire la successione (E_n) con l'altra (F_n) ove $F_1 = E_1$ e, per $n \geq 2$,

$$F_n = E_n \setminus \left(\bigcup_{k=1}^{n-1} E_k \right).$$

Si considerino ora gli spazi misurabili (F_n, \mathcal{F}_n) ove \mathcal{F}_n è la traccia di \mathcal{F} su F_n , vale a dire $\mathcal{F}_n = \mathcal{F} \cap F_n := \{B \cap F_n : B \in \mathcal{F}\}$ e siano λ_n e μ_n le restrizioni di λ e μ a (F_n, \mathcal{F}_n) . Poiché è, evidentemente, $\lambda_n \ll \mu_n$, a tali misure si può applicare la prima parte della dimostrazione; si ottiene così una funzione $\varphi_n \in L^1(\mu_n)$ tale che per ogni $E \in \mathcal{F}$ sia

$$\lambda(E \cap F_n) = \lambda_n(E \cap F_n) = \int_{E \cap F_n} \varphi_n d\mu_n = \int_{E \cap F_n} \varphi_n d\mu.$$

Si definisca ora $\varphi := \sum_{n \in \mathbb{N}} \varphi_n \mathbf{1}_{F_n}$. Poiché in ogni punto $\omega \in \Omega$ in solo termine è diverso da zero, la funzione φ è integrabile, $\varphi \in L^1(\mu)$ e, per ogni $E \in \mathcal{F}$, risulta

$$\lambda(E) = \int_E \varphi d\mu,$$

che conclude la dimostrazione. □

1.15 Misura prodotto

Siano $(\Omega_1, \mathcal{F}_1, \mu_1)$ e $(\Omega_2, \mathcal{F}_2, \mu_2)$ due spazi di misura. La famiglia dei *rettangoli misurabili*

$$\mathcal{F}_1 \times \mathcal{F}_2 := \{A \times B : A \in \mathcal{F}_1, B \in \mathcal{F}_2\}$$

non è una tribú (si veda l'Esercizio 1.28). Si consideri allora la tribú $\mathcal{F}(\mathcal{F}_1 \times \mathcal{F}_2)$ generata da $\mathcal{F}_1 \times \mathcal{F}_2$; questa si dice *tribú prodotto* e la si denota con $\mathcal{F}_1 \otimes \mathcal{F}_2$. Ci si

pone il problema di stabilire in quali ipotesi si possa definire in $\mathcal{F}_1 \otimes \mathcal{F}_2$ una misura μ con la proprietà

$$\forall A \in \mathcal{F}_1 \quad \forall B \in \mathcal{F}_2 \quad \mu(A \times B) = \mu_1(A) \mu_2(B). \quad (1.15.1)$$

Ha interesse anche stabilire quando la misura μ , se esiste, sia unica.

Definizione 1.15.1. Per ogni sottoinsieme E di $\Omega_1 \times \Omega_2$ e per ogni punto $x \in \Omega_1$, l'insieme $E_x := \{y \in \Omega_2 : (x, y) \in E\} \subseteq \Omega_2$ si dice *sezione di E in x* . In maniera analoga si definisce la sezione $E_y \subseteq \Omega_1$ di E in $y \in \Omega_2$. \diamond

Teorema 1.15.1. *Le sezioni di ogni insieme E della tribú prodotto $\mathcal{F}_1 \otimes \mathcal{F}_2$ sono misurabili, vale a dire, E_x appartiene a \mathcal{F}_2 per ogni $x \in \Omega_1$ e E_y appartiene a \mathcal{F}_1 per ogni $y \in \Omega_2$.*

Dimostrazione. Posto

$$\mathcal{G} := \{E \subseteq \Omega_1 \times \Omega_2 : \forall x \in \Omega_1 \ E_x \in \mathcal{F}_2, \forall y \in \Omega_2 \ E_y \in \mathcal{F}_1\},$$

risulta $\mathcal{F}_1 \times \mathcal{F}_2 \subseteq \mathcal{G}$, cioè \mathcal{G} contiene i rettangoli misurabili; infatti

$$(A \times B)_x = \begin{cases} B, & \text{se } x \in A, \\ \emptyset, & \text{se } x \notin A, \end{cases}$$

sicché $(A \times B)_x \in \mathcal{F}_2$ per ogni $x \in \Omega_1$. Analogamente si procede per le sezioni rispetto ai punti $y \in \Omega_2$. Per dimostrare il teorema basterà stabilire l'inclusione $\mathcal{G} \supseteq \mathcal{F}_1 \otimes \mathcal{F}_2$; a tal fine si mostrerà che \mathcal{G} è una tribú. Data l'evidente simmetria tra le sezioni in $x \in \Omega_1$ e quelle in $y \in \Omega_2$, tratteremo solo le prime.

Evidentemente, tanto \emptyset quanto $\Omega_1 \times \Omega_2$ appartengono a \mathcal{G} . Se E è in \mathcal{G} , allora $(E^c)_x = (E_x)^c \in \mathcal{F}_2$ sicché E^c appartiene a \mathcal{G} e \mathcal{G} è stabile rispetto alla complementazione. Sia, infine, (E_n) una successione di insiemi di \mathcal{G} :

$$\left(\bigcup_{n \in \mathbb{N}} E_n \right)_x = \bigcup_{n \in \mathbb{N}} (E_n)_x \in \mathcal{F}_2,$$

sicché $\bigcup_{n \in \mathbb{N}} E_n$ appartiene a \mathcal{G} . \square

Teorema 1.15.2. *Se le misure μ_1 e μ_2 sono σ -finite, allora, per ogni insieme A di $\mathcal{F}_1 \otimes \mathcal{F}_2$, la funzione $x \mapsto \mu_2(A_x)$ è \mathcal{F}_1 -misurabile mentre la funzione $y \mapsto \mu_2(A_y)$ è \mathcal{F}_2 -misurabile.*

Dimostrazione. Sia \mathcal{C} la famiglia di insiemi A della tribú prodotto $\mathcal{F}_1 \otimes \mathcal{F}_2$ per i quali è vero che $x \mapsto \mu_2(A_x)$ è \mathcal{F}_1 -misurabile. \mathcal{C} include la famiglia $\mathcal{F}_1 \times \mathcal{F}_2$ dei rettangoli misurabili, $\mathcal{F}_1 \times \mathcal{F}_2 \subseteq \mathcal{C}$; infatti se $A = E \times F$ con $E \in \mathcal{F}_1$ e $F \in \mathcal{F}_2$, si ha $\mu_2(A_x) = \mu_2(F) \mathbf{1}_E(x)$ che è, ovviamente, \mathcal{F}_1 -misurabile. Se A^1, A^2, \dots, A^n sono rettangoli misurabili disgiunti, $A^i = E^i \times F^i$ con $E^i \in \mathcal{F}_1$ e $F^i \in \mathcal{F}_2$ ($i = 1, 2, \dots, n$), allora,

$$\mu_2 \left[\left(\bigcup_{i=1}^n A^i \right)_x \right] = \mu_2 \left(\bigcup_{i=1}^n A_x^i \right) = \sum_{i=1}^n \mu_2(A_x^i) = \sum_{i=1}^n \mu_2(F^i) \mathbf{1}_{E^i}(x),$$

che è misurabile rispetto a \mathcal{F}_1 , in quanto somma di un numero finito di funzioni misurabili. Perciò \mathcal{C} contiene l'anello \mathcal{R} delle unioni finite e disgiunte di rettangoli misurabili. Ma \mathcal{C} è una classe monotona perché se A_n è in \mathcal{C} per ogni $n \in \mathbb{N}$ e se $A_n \uparrow A$, allora $(A_n)_x \uparrow A_x$ onde $\mu_2[(A_n)_x] \rightarrow \mu_2(A_x)$: perciò $\mu_2(A_x)$, come limite di funzioni misurabili, è misurabile. Si ricorre allo stesso ragionamento se (A_n) è decrescente, anziché crescente. Si può quindi concludere che $\mathcal{C} = \mathcal{F}_1 \otimes \mathcal{F}_2$. \square

Teorema 1.15.3. *Se le misure μ_1 e μ_2 , su $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$ rispettivamente, sono σ -finite, esiste un'unica misura σ -finita μ definita in $\mathcal{F}_1 \otimes \mathcal{F}_2$ che verifichi la (1.15.1); essa si dice misura prodotto e si denota con $\mu = \mu_1 \otimes \mu_2$.*

Dimostrazione. Si definisca $\mu : \mathcal{F}_1 \otimes \mathcal{F}_2 \rightarrow \overline{\mathbb{R}}_+$ mediante

$$\mu(A) := \int_{\Omega_1} \mu_2(A_x) d\mu_1(x); \quad (1.15.2)$$

μ sarà la misura cercata. La definizione (1.15.2) è lecita in virtù del teorema precedente.

Sia (A^n) una successione di insiemi disgiunti di $\mathcal{F}_1 \otimes \mathcal{F}_2$. Poiché

$$\left[\bigcup_{n \in \mathbb{N}} A^n \right]_x = \bigcup_{n \in \mathbb{N}} A_x^n,$$

e poiché anche le sezioni sono disgiunte, si ha

$$\begin{aligned} \mu \left(\bigcup_{n \in \mathbb{N}} A^n \right) &= \int_{\Omega_1} \mu_2 \left[\left(\bigcup_{n \in \mathbb{N}} A^n \right)_x \right] d\mu_1(x) = \int_{\Omega_1} \sum_{n \in \mathbb{N}} \mu_2(A_x^n) d\mu_1(x) \\ &= \sum_{n \in \mathbb{N}} \int_{\Omega_1} \mu_2(A_x^n) d\mu_1(x) = \sum_{n \in \mathbb{N}} \mu(A^n), \end{aligned}$$

ciò che prova che μ è una misura. Per verificare che μ soddisfaccia alla (1.15.1) si ricordi che $\mu_2[(A \times B)_x] = \mu_2(B) \mathbf{1}_A(x)$ se A appartiene a \mathcal{F}_1 e B a \mathcal{F}_2 ; perciò

$$\mu(A \times B) = \int \mu_2(B) \mathbf{1}_A(x) d\mu_1(x) = \mu_2(B) \int \mathbf{1}_A d\mu = \mu_1(A) \mu_2(B).$$

Rimane dunque da controllare che μ sia σ -finita. Siccome tanto μ_1 quanto μ_2 sono σ -finite, esistono successioni (E_n) di insiemi di \mathcal{F}_1 e (F_r) di insiemi di \mathcal{F}_2 tali che

$$\bigcup_{n \in \mathbb{N}} E_n = \Omega_1, \quad \bigcup_{r \in \mathbb{N}} F_r = \Omega_2, \quad \text{e} \quad \mu_1(E_n) < +\infty, \quad \mu_2(F_r) < +\infty$$

per ogni n e per ogni r di \mathbb{N} . Perciò

$$\Omega_1 \times \Omega_2 = \bigcup_{n, r \in \mathbb{N}} E_n \times F_r$$

e

$$\forall n, r \in \mathbb{N} \quad \mu(E_n \times F_r) = \mu_1(E_n) \mu_2(F_r) < +\infty.$$

Si noti che, se \mathcal{R} è l'anello delle unioni finite e disgiunte di rettangoli misurabili, la restrizione ν di μ a \mathcal{R} è una misura tale che $\nu(E \times F) = \mu_1(E) \mu_2(F)$ per ogni $E \in \mathcal{F}_1$ e per ogni $F \in \mathcal{F}_2$; per il teorema di Carathéodory è unica l'estensione di ν a $\mathcal{F}_1 \otimes \mathcal{F}_2$ che è la tribù generata da \mathcal{R} . Ciò stabilisce l'unicità di μ . \square

Teorema 1.15.4. (Fubini–Tonelli). *Siano dati due spazi misurabili σ -finiti*

$$(\Omega_1, \mathcal{F}_1, \mu_1) \quad e \quad (\Omega_2, \mathcal{F}_2, \mu_2).$$

- (a) *Sia $f : \Omega_1 \times \Omega_2 \rightarrow \overline{\mathbb{R}}_+$ una funzione positiva misurabile rispetto alla tribù prodotto $\mathcal{F} := \mathcal{F}_1 \otimes \mathcal{F}_2$; allora, per ogni $x \in \Omega_1$, la funzione $y \mapsto f(x, y)$ è \mathcal{F}_2 -misurabile, la funzione $x \mapsto \int f(x, y) d\mu_2(y)$ è \mathcal{F}_1 -misurabile, ed inoltre si ha, se μ è la misura prodotto $\mu = \mu_1 \otimes \mu_2$,*

$$\int_{\Omega_1 \times \Omega_2} f d\mu = \int_{\Omega_1} d\mu_1(x) \int_{\Omega_2} f(x, y) d\mu_2(y). \quad (1.15.3)$$

- (b) *Se $f : \Omega_1 \times \Omega_2 \rightarrow \overline{\mathbb{R}}$ è \mathcal{F} -misurabile ed integrabile rispetto alla misura prodotto μ , allora la funzione $y \mapsto f(x, y)$ è integrabile rispetto a μ_2 per quasi ogni (rispetto alla misura μ_1) $x \in \Omega_1$, la funzione definita μ_1 -quasi ovunque da $x \mapsto \int f(x, y) d\mu_2(y)$ è integrabile rispetto a μ_1 e vale ancora la (1.15.3).*

Dimostrazione. Sia $f = \mathbf{1}_A$ con $A \in \mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$, allora ogni sezione A_x è in \mathcal{F}_2 e la funzione

$$y \mapsto \mathbf{1}_A(x, y) = \mathbf{1}_{A_x}(y)$$

è \mathcal{F}_2 -misurabile per ogni $x \in \Omega_1$. Inoltre, ricordando la (1.15.2), si ha

$$\int_{\Omega_1} d\mu_1(x) \int_{\Omega_2} \mathbf{1}_A(x, y) d\mu_2(y) = \int_{\Omega_1} \mu_2(A_x) d\mu_1(x) = \mu(A) = \int_{\Omega_1 \times \Omega_2} \mathbf{1}_A d\mu;$$

la (1.15.3) vale perciò per le funzioni indicatrici di insiemi di $\mathcal{F}_1 \otimes \mathcal{F}_2$. Per la linearità degli integrali, l'asserto vale anche per le funzioni \mathcal{F} -semplici. Sia ora (s_n) una successione di funzioni positive \mathcal{F} -semplici che tende crescendo alla funzione misurabile positiva f . L'applicazione $y \mapsto f(x, y)$ è \mathcal{F}_2 -misurabile perché $f(x, y) = \sup_{n \in \mathbb{N}} s_n(x, y)$. Analogamente la funzione

$$x \mapsto \int_{\Omega_2} f(x, y) d\mu_2(y) = \sup_{n \in \mathbb{N}} \int_{\Omega_2} s_n(x, y) d\mu_2(y)$$

risulta \mathcal{F}_1 -misurabile; il teorema di Beppo Levi dà ora

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} f d\mu &= \sup_{n \in \mathbb{N}} \int_{\Omega_1 \times \Omega_2} s_n d\mu = \sup_{n \in \mathbb{N}} \int_{\Omega_1} d\mu_1(x) \int_{\Omega_2} s_n(x, y) d\mu_2(y) \\ &= \int_{\Omega_1} d\mu_1(x) \left\{ \sup_{n \in \mathbb{N}} \int_{\Omega_2} s_n(x, y) d\mu_2(y) \right\} \\ &= \int_{\Omega_1} d\mu_1(x) \int_{\Omega_2} \sup_{n \in \mathbb{N}} s_n(x, y) d\mu_2(y) = \int_{\Omega_1} d\mu_1(x) \int_{\Omega_2} f(x, y) d\mu_2(y). \end{aligned}$$

- (b) Poiché f è integrabile, tali sono anche f^+ e f^- . Perciò la funzione

$$x \mapsto \int f^+(x, y) d\mu_2(y)$$

è finita per quasi ogni $x \in \Omega_1$ (rispetto a μ_1) per il Teorema 1.10.1(c); similmente si argomenta per f^- . Perciò la differenza

$$\int_{\Omega_2} f(x, y) d\mu_2(y) = \int_{\Omega_2} f^+(x, y) d\mu_2(y) - \int_{\Omega_2} f^-(x, y) d\mu_2(y)$$

è definita per quasi ogni $x \in \Omega_1$. Basta ora applicare la parte (a) separatamente alle funzioni positive f^+ e f^- . \square

Naturalmente, nelle stesse ipotesi, vale anche la relazione simmetrica

$$\int_{\Omega_1 \times \Omega_2} f d\mu = \int_{\Omega_2} d\mu_2(y) \int_{\Omega_1} f(x, y) d\mu_1(x).$$

Si osservi che, per poter scrivere la (1.15.3), occorre verificare che la funzione f sia integrabile rispetto alla misura prodotto. Si vedano, a tal fine, gli esercizi.

Il procedimento sin qui esposto si adatta, senza particolari difficoltà, al prodotto di un numero finito di misure, cioè alla misura prodotto sullo spazio misurabile

$$(\Omega_1 \times \Omega_2 \times \cdots \times \Omega_n, \mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \cdots \otimes \mathcal{F}_n).$$

Si presenta, invece, qualche difficoltà nel definire la misura prodotto nel caso di un prodotto numerabile (o di cardinalità maggiore) di spazi misurabili a meno che le misure non siano tutte di probabilità, caso che tratteremo di seguito.

Definizione 1.15.2. Sia $((\Omega_n, \mathcal{F}_n))_{n \in \mathbb{N}}$ una successione di spazi misurabili e sia $\Omega = \prod_{n \in \mathbb{N}} \Omega_n$ l'insieme di tutte le successioni

$$(\omega_1, \omega_2, \dots, \omega_n, \dots)$$

tali che ω_n appartenga a Ω_n per ogni $n \in \mathbb{N}$. Se E_j è in \mathcal{F}_j per $j = 1, 2, \dots, n$, l'insieme $E \subseteq \Omega$ definito da

$$E = E_1 \times E_2 \times \cdots \times E_n \times \prod_{j=n+1}^{\infty} \Omega_j \quad (1.15.4)$$

si dice *cilindro di base n -dimensionale* $E_1 \times E_2 \times \cdots \times E_n$,

Un cilindro con base n -dimensionale come quello della (1.15.4) può essere pensato come avente base di dimensione maggiore; per esempio, il cilindro E definito dalla (1.15.4) può essere pensato come un cilindro di base $(n+1)$ -dimensionale $E_1 \times E_2 \times \cdots \times E_n \times \Omega_{n+1}$.

Lemma 1.15.1. *La famiglia \mathcal{C} dei cilindri è un semi-anello di sottoinsiemi del prodotto cartesiano $\Omega = \prod_{n \in \mathbb{N}} \Omega_n$.*

Dimostrazione. Si considerino i due cilindri

$$E = E_1 \times E_2 \times \cdots \times E_n \times \prod_{j=n+1}^{\infty} \Omega_j,$$

$$F = F_1 \times F_2 \times \cdots \times F_k \times \prod_{j=k+1}^{\infty} \Omega_j,$$

ove $E_j \in \mathcal{F}_j$ per $j = 1, 2, \dots, n$ e $F_i \in \mathcal{F}_i$ per $i = 1, 2, \dots, k$. Non è restrittivo supporre, per fissare le idee, che sia $k < n$; si può quindi supporre che E e F abbiano entrambi la base di dimensione n e che sia $F_j = \Omega_j$ se $j = k + 1, \dots, n$.

Ora

$$E \cap F = \prod_{j=1}^n (E_j \cap F_j) \times \prod_{j=n+1}^{\infty} \Omega_j,$$

che è ancora un cilindro. D'altro canto,

$$E \setminus F = E \cap F^c = \prod_{j=1}^n (E_j \cap F_j^c) \times \prod_{j=n+1}^{\infty} \Omega_j,$$

che è anch'esso un cilindro. □

Teorema 1.15.5. *Se $((\Omega_n, \mathcal{F}_n, \mathbb{P}_n))_{n \in \mathbb{N}}$ è una successione di spazi di probabilità, esiste un'unica misura di probabilità sullo spazio misurabile (Ω, \mathcal{F}) , ove Ω è il prodotto cartesiano $\prod_{n \in \mathbb{N}} \Omega_n$ e \mathcal{F} è la tribù, $\mathcal{F} = \mathcal{F}(\mathcal{C})$, generata dai cilindri di $\prod_{n \in \mathbb{N}} \Omega_n$, tale che sul cilindro*

$$E = E_1 \times E_2 \times \cdots \times E_n \times \prod_{j=n+1}^{\infty} \Omega_j$$

assuma il valore

$$\mathbb{P} \left(\prod_{j=1}^n E_j \times \prod_{j=n+1}^{\infty} \Omega_j \right) = \mathbb{P}_1(E_1) \mathbb{P}_2(E_2) \cdots \mathbb{P}_n(E_n).$$

Dimostrazione. Si definisca \mathbb{P} sul semi-anello \mathcal{C} mediante

$$\mathbb{P}(E) := \mathbb{P}_1(E_1) \mathbb{P}_2(E_2) \cdots \mathbb{P}_n(E_n),$$

se E è il cilindro (1.15.4).

Per verificare che la funzione d'insieme \mathbb{P} così definita su \mathcal{C} è finitamente additiva, si supponga che F sia un cilindro che si può scrivere come unione disgiunta di altri due cilindri, $F = F_1 \cup F_2$. Esiste allora un numero naturale N tale che F , F_1 e F_2 possano essere scritti nella forma

$$E_1 \times E_2 \times \cdots \times E_N \times \prod_{j=N+1}^{\infty} \Omega_j.$$

Si può ora applicare il Teorema 1.15.3 per ottenere $\mathbb{P}(F) = \mathbb{P}(F_1) + \mathbb{P}(F_2)$ e quindi estendere \mathbb{P} all'anello \mathcal{R} generato da \mathcal{C} . Per poter applicare il teorema di estensione di Carathéodory occorre mostrare che \mathbb{P} è una misura su \mathcal{R} . A tal fine, basta provare che, per ogni successione decrescente (A_n) di insiemi di \mathcal{R} tale che $\mathbb{P}(A_n) \rightarrow \varepsilon > 0$, si ha $\bigcap_n A_n \neq \emptyset$. Sia (A_n) una successione siffatta. Si ponga $H_n := \prod_{j=n+1}^{\infty} \Omega_j$. Operando come abbiamo appena fatto, si può definire una funzione d'insieme prodotto $\nu^{(n)}$ definita sulla classe $\mathcal{R}^{(n)}$ delle unioni finite di cilindri

di $\mathcal{C}(\mathcal{F}_{n+1}, \mathcal{F}_{n+2}, \dots)$. Così, per ogni $n \in \mathbb{N}$ si può costruire \mathbb{P} sul semi-anello \mathcal{C} dei cilindri come prodotto delle $n + 1$ misure

$$\mathbb{P}_1, \mathbb{P}_2, \dots, \mathbb{P}_n \text{ e } \nu^{(n)}.$$

Sia $A_n(\omega_1)$ la sezione di A_n in $\omega_1 \in \Omega_1$,

$$A_n(\omega_1) := \{\omega \in H_1 : (\omega_1, \omega) \in A_n\};$$

per ogni $\omega_1 \in \Omega_1$ si ha $A_n(\omega_1) \in \mathcal{R}^{(1)}$. Posto

$$B_n := \left\{ \omega_1 : \nu^{(1)}(A_n(\omega_1)) > \frac{1}{2} \varepsilon \right\},$$

B_n è un sottoinsieme di Ω_1 e, di più, è l'unione finita di insiemi di \mathcal{F}_1 , sicché è esso stesso in \mathcal{F}_1 . Dalla decomposizione

$$A_n = (A_n \cap (B_n \times H_1)) \cup (A_n \cap (B_n^c \times H_1))$$

segue

$$\varepsilon \leq \mathbb{P}(A_n) \leq \mathbb{P}_1(B_n) + \frac{1}{2} \varepsilon [1 - \mathbb{P}_1(B_n)]$$

e di qui

$$\mathbb{P}_1(B_n) > \frac{1}{2} \varepsilon$$

per ogni $n \in \mathbb{N}$. Poiché (A_n) è decrescente, tale è anche la successione $\{B_n\}$. Perciò

$$\mathbb{P}_1(\cap_{n \in \mathbb{N}} B_n) \geq \frac{1}{2} \varepsilon.$$

Ora \mathbb{P}_1 è una misura; di conseguenza, esiste $\omega_1 \in \Omega_1$ tale che, per ogni $n \in \mathbb{N}$,

$$\nu^{(1)}(A_n(\omega_1)) > \frac{1}{2}.$$

Si fissi ω_1 e si ripeta il ragionamento per la successione $(A_n(\omega_1))$ di sottoinsiemi di H_1 . Si ha così un punto $\omega_2 \in \Omega_2$ tale che, per ogni $n \in \mathbb{N}$,

$$\nu^{(2)}(A_n(\omega_1, \omega_2)) > \frac{1}{4} \varepsilon.$$

Per induzione si ottiene un punto $(\omega_1, \omega_2, \dots)$ di $\prod_{n \in \mathbb{N}} \Omega_n$, tale che, per ogni coppia di numeri naturali k e n , sia

$$A_n(\omega_1, \omega_2, \dots, \omega_k) \neq \emptyset.$$

Ricordando però la definizione di A_n , che è in \mathcal{R} , si vede che $(\omega_1, \omega_2, \dots)$ appartiene ad A_n per ogni n , sicché

$$\cap_{n \in \mathbb{N}} A_n \neq \emptyset,$$

e, dunque, \mathbb{P} è una misura. □

1.16 Completamento di misure

La necessità di tenere conto dei sottoinsiemi di insiemi di misura nulla induce a considerare il cosiddetto completamento di una misura.

Teorema 1.16.1. *Sia μ una misura su (Ω, \mathcal{F}) e si definisca con $\mathcal{E}(\mu)$ la classe dei sottoinsiemi degli insiemi di μ -misura nulla*

$$\mathcal{E}(\mu) := \{E \subseteq \Omega : \exists N \in \mathcal{F}, E \subseteq N, \mu(N) = 0\},$$

e si ponga

$$\mathcal{F}(\mu) := \{A \cup E : A \in \mathcal{F}, E \in \mathcal{E}(\mu)\}.$$

Allora:

- (a) $\mathcal{F}(\mu)$ è una tribù;
- (b) la funzione $\bar{\mu} : \mathcal{F}(\mu) \rightarrow \bar{\mathbb{R}}_+$ definita da $\bar{\mu}(A \cup E) := \mu(A)$ è una misura su $(\Omega, \mathcal{F}(\mu))$;
- (c) $\bar{\mu}$ è completa, nel senso che, se $F \in \mathcal{F}(\mu)$ e $\bar{\mu}(F) = 0$, allora appartiene a $\mathcal{F}(\mu)$ ogni insieme $B \subseteq F$ (e necessariamente è $\mu(B) = 0$).

Dimostrazione. (a) $\mathcal{F}(\mu)$ è stabile rispetto alla complementazione; infatti, sia $F \in \mathcal{F}(\mu)$, e $F = A \cup E$, con $A \in \mathcal{F}$, $E \in \mathcal{E}(\mu)$, $E \subseteq N$, $N \in \mathcal{F}$, $\mu(N) = 0$; allora:

$$\begin{aligned} F^c &= A^c \cap E^c = [(A^c \cap N) \cup (A^c \cap N^c)] \cap E^c \\ &= (A^c \cap N \cap E^c) \cup (A^c \cap N^c \cap E^c) \\ &= (A^c \cap N \cap E^c) \cup (A^c \cap N^c). \end{aligned}$$

Ora $A^c \cap N^c \in \mathcal{F}$ e $A^c \cap N \cap E^c \subseteq N$, sicché $F^c \in \mathcal{F}(\mu)$. Se, poi, per ogni $n \in \mathbb{N}$, è $F_n \in \mathcal{F}(\mu)$, con $F_n = A_n \cup E_n$, $A_n \in \mathcal{F}$, $E_n \in \mathcal{E}(\mu)$, $E_n \subseteq N_n$, $N_n \in \mathcal{F}$, $\mu(N_n) = 0$, si ha

$$\bigcup_{n \in \mathbb{N}} F_n = (\bigcup_{n \in \mathbb{N}} A_n) \cup (\bigcup_{n \in \mathbb{N}} E_n),$$

ove

$$\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F} \quad \text{e} \quad \bigcup_{n \in \mathbb{N}} E_n \subseteq \bigcup_{n \in \mathbb{N}} N_n \in \mathcal{F}$$

e

$$\mu(\bigcup_{n \in \mathbb{N}} N_n) \leq \sum_{n \in \mathbb{N}} \mu(N_n) = 0;$$

dunque $\bigcup_{n \in \mathbb{N}} F_n$ è in \mathcal{F} e $\mathcal{F}(\mu)$ è stabile rispetto all'unione numerabile.

(b) Basta controllare che quella data sia una buona definizione. Siano, allora, $A_1 \cup E_1 = A_2 \cup E_2$ due diverse rappresentazioni dell'insieme $F \in \mathcal{F}(\mu)$:

$$F = A_1 \cup E_1 = A_2 \cup E_2,$$

con $A_i \in \mathcal{F}$, $E_i \in \mathcal{E}(\mu)$, $E_i \subseteq N_i \in \mathcal{F}$, $\mu(N_i) = 0$ ($i = 1, 2$). Si ha $A_1 \setminus A_2 \subseteq E_2$ e quindi

$$\mu(A_1) = \mu(A_1 \cap A_2) + \mu(A_1 \setminus A_2) = \mu(A_1 \cap A_2) \leq \mu(A_2);$$

simmetricamente si ha $\mu(A_2) \leq \mu(A_1)$ e dunque $\mu(A_1) = \mu(A_2)$.

(c) Sia $B \subseteq F \in \mathcal{F}(\mu)$ e $\bar{\mu}(F) = 0$. Scritto F nella forma $F = A \cup E$ con $A \in \mathcal{F}$, $E \in \mathcal{E}(\mu)$, $E \subseteq N \in \mathcal{F}$ e $\mu(N) = 0$, si ha $\mu(A) = 0$, sicché $B \subseteq A \cup N \in \mathcal{F}$ con $\mu(A \cup N) = 0$ onde $B \in \mathcal{E}(\mu)$ e dunque anche B appartiene a $\mathcal{F}(\mu)$. \square

Si dice che lo spazio di misura $(\Omega, \mathcal{F}(\mu), \bar{\mu})$ è il *completamento* di $(\Omega, \mathcal{F}, \mu)$ e che la tribú $\mathcal{F}(\mu)$ è il *completamento della tribú \mathcal{F} rispetto alla misura μ* .

Teorema 1.16.2. *Se $(\Omega, \mathcal{F}, \mu)$ è completo e $f = g$ q.o., allora f è misurabile se, e solo se, g è misurabile.*

Dimostrazione. Per ogni $c \in \mathbb{R}$ si ha $\{f < c\} \Delta \{g < c\} \subseteq \{f \neq g\}$; giacché, per ipotesi, $\mu(f \neq g) = 0$, gli insiemi $\{f < c\}$ e $\{g < c\}$ differiscono per un sottoinsieme di misura nulla che è in \mathcal{F} , perché \mathcal{F} è completa. Perciò

$$\{f < c\} \in \mathcal{F} \implies \{g < c\} \in \mathcal{F},$$

che dà l'asserto. \square

Il completamento della tribú di Borel è la tribú $\mathcal{L}(\mathbb{R})$ degli insiemi *misurabili secondo Lebesgue*, mentre il completamento della misura di Borel su $\mathcal{B}(\mathbb{R})$ si dice *misura di Lebesgue*. Le misure di Borel e di Lebesgue dunque coincidono su tutti gli insiemi boreliani, in particolare, sugli intervalli.

1.17 Appendice

In questa appendice rivedremo brevemente le proprietà delle funzioni convesse che servono per lo studio della probabilità.

Definizione 1.17.1. Sia $I =]a, b[$ con $-\infty \leq a < b \leq +\infty$ un intervallo aperto di \mathbb{R} . Una funzione $\varphi : I \rightarrow \mathbb{R}$ che, per ogni coppia di punti x e y di I , e per ogni $\alpha \in [0, 1]$ soddisfa alla disegualianza

$$\varphi(\alpha x + (1 - \alpha)y) \leq \alpha \varphi(x) + (1 - \alpha) \varphi(y). \quad (1.17.1)$$

si dice *convessa*. Si dice, invece, che φ è *concava* se è convessa la funzione $-\varphi$. \diamond

Si controlla facilmente per induzione che vale la seguente generalizzazione della (1.17.1); per ogni scelta di n punti x_j ($j = 1, 2, \dots, n$) in I e per ogni scelta di n numeri positivi α_j ($j = 1, 2, \dots, n$), $\alpha_j \geq 0$, con $\sum_{j=1}^n \alpha_j = 1$, si ha

$$\varphi\left(\sum_{j=1}^n \alpha_j x_j\right) \leq \sum_{j=1}^n \alpha_j \varphi(x_j).$$

Si considerino ora $x \in I$ e h_1 e h_2 sufficientemente piccoli affinché $x + h_2$ e $x - h_2$ appartengano a I e tali che $0 < h_1 < h_2$; scelto $\alpha = (h_2 - h_1)/h_2$ nella (1.17.1), si ha

$$\frac{h_2 - h_1}{h_2} \varphi(x) + \frac{h_1}{h_2} \varphi(x \pm h_2) \geq \varphi\left(\frac{h_2 - h_1}{h_2} x + \frac{h_1}{h_2} (x \pm h_2)\right) = \varphi(x \pm h_1).$$

Di qui si ricavano facilmente le disegualianze

$$\begin{aligned} \frac{\varphi(x + h_1) - \varphi(x)}{h_1} &\leq \frac{\varphi(x + h_2) - \varphi(x)}{h_2}, \\ \frac{\varphi(x - h_2) - \varphi(x)}{-h_2} &\leq \frac{\varphi(x - h_1) - \varphi(x)}{-h_1}. \end{aligned}$$

Queste ultime due relazioni mostrano che il rapporto incrementale destro e quello sinistro di φ sono rispettivamente crescente e decrescente, sicché esistono le derivate sinistra $D^-\varphi(x)$ e destra $D^+\varphi(x)$ di φ in $x \in I$. Inoltre per $h > 0$ e $h' > 0$ scende dalla (1.17.1)

$$\frac{\varphi(x-h) - \varphi(x)}{-h} \leq \frac{\varphi(x+h') - \varphi(x)}{h'}$$

che mostra che le derivate $D^-\varphi(x)$ e $D^+\varphi(x)$ sono finite. Di conseguenza ogni funzione convessa è continua in I . Segue da quanto detto sopra che si può scrivere

$$D^+\varphi(x) = \inf_{y>x} \frac{\varphi(y) - \varphi(x)}{y-x}, \quad (1.17.2)$$

$$D^-\varphi(x) = \sup_{y<x} \frac{\varphi(x) - \varphi(y)}{x-y}. \quad (1.17.3)$$

Perciò, se $x_1 < x_2$ è

$$\begin{aligned} D^-\varphi(x_1) &\leq D^+\varphi(x_1) \leq \frac{\varphi(x_2) - \varphi(x_1)}{x_2 - x_1} \\ &= \frac{\varphi(x_1) - \varphi(x_2)}{x_1 - x_2} \leq D^-\varphi(x_2) \leq D^+\varphi(x_2); \end{aligned}$$

ne consegue che le funzioni

$$t \mapsto D^+\varphi(t) \quad \text{e} \quad t \mapsto D^-\varphi(t)$$

sono entrambe isotone, e, pertanto, ciascuna di esse ammette al più un'infinità numerabile di punti di discontinuità.

Scende dalla (1.17.2) e dalla (1.17.3) che valgono le disequaglianze

$$\begin{aligned} \varphi(y) &\geq \varphi(x) + D^+\varphi(x)(y-x), & \text{se } y > x, \\ \varphi(y) &\geq \varphi(x) + D^-\varphi(x)(y-x), & \text{se } y < x, \end{aligned}$$

onde, per ogni $y \in I$ e per ogni $\alpha_x \in [D^-\varphi(x), D^+\varphi(x)]$,

$$\varphi(y) \geq \varphi(x) + \alpha_x (y-x). \quad (1.17.4)$$

In un punto x nel quale la funzione φ sia derivabile, $D^+\varphi(x) = D^-\varphi(x) = D\varphi(x)$, si ottiene la relazione elementare ben nota

$$\varphi(y) \geq \varphi(x) + D\varphi(x)(y-x),$$

che esprime il fatto geometrico che il grafico di φ giace sopra la retta tangente al grafico in x .

Posto, per x e s in I , $g_s(x) := \varphi(s) + \alpha_s(x-s)$, si ha, evidentemente, $\varphi(x) \geq g_s(x)$ e

$$\varphi(x) = \sup \{g_s(x) : s \in I\}.$$

Si noti che, dal punto di vista geometrico, g_s è una retta che si dice *linea di supporto di φ* . In probabilità è utile il seguente teorema, detto *della linea di supporto*.

Teorema 1.17.1. *Data una funzione convessa φ nell'intervallo aperto I esistono due successioni (a_n) e (b_n) di numeri reali tali che, per ogni $x \in I$, valga la rappresentazione*

$$\varphi(x) = \sup_{n \in \mathbb{N}} \{a_n x + b_n\},$$

sicché φ è l'estremo superiore di una famiglia numerabile di linee di supporto.

Dimostrazione. Si consideri l'insieme $\mathbb{Q} \cap I$ dei numeri razionali di I ; per ogni numero $q \in \mathbb{Q} \cap I$ è

$$\varphi(x) \geq g_q(x) = \varphi(q) + \alpha_q(x - q),$$

ove

$$D^- \varphi(q) \leq \alpha_q \leq D^+ \varphi(q).$$

Poiché $D^+ \varphi$ e $D^- \varphi$ sono limitate in ogni intervallo chiuso contenuto in I , la successione (α_q) è limitata. Pertanto esiste una successione (α_{q_n}) estratta da (α_q) che converge a $\alpha \in [D^- \varphi, D^+ \varphi]$ e

$$\lim_{n \rightarrow \mathbb{N}} g_{q_n}(x) = \varphi(x).$$

Poiché $\varphi(x) \geq g_{q_n}(x)$ per ogni $n \in \mathbb{N}$ si ha

$$\varphi(x) = \sup_{n \in \mathbb{N}} \{\varphi(q_n) + \alpha_{q_n}(x - q_n)\},$$

che prova l'asserto quando si sia posto

$$a_n := \alpha_{q_n} \quad \text{e} \quad b_n := \varphi(q_n) - \alpha_{q_n} q_n$$

per ogni $n \in \mathbb{N}$. □

1.18 Note al Capitolo 1

La teoria della misura come è qui succintamente presentata è scritta nello spirito del libro (Halmos, 1950). L'approccio di questo libro non è l'unico possibile; un altro, a prima vista completamente differente, si trova in (Bourbaki, 1965) e nei libri dei seguaci di questo "autore". Ritengo che il primo sia preferibile per gli usi del calcolo delle probabilità in questa scelta confortato dal notare che è anche quella di un maestro delle probabilità come Doob; si veda (Doob, 1994).

Quello che presento è il bagaglio minimo che deve essere a disposizione di uno studioso di probabilità; il solo argomento essenziale che ometto qui è l'equiintegrabilità, la cui introduzione rimando al Capitolo sulle martingale. Naturalmente esulano dallo scopo di queste lezioni lo studio delle misure reali, vale a dire delle funzioni $\mu : \mathcal{F} \rightarrow \mathbb{R}$ che siano σ -additive, e quello degli spazî di misura infinita, $\mu(\Omega) = +\infty$.

Nello scrivere queste lezioni mi sono basato soprattutto sui libri (Kingman & Taylor, 1966), (Ash, 1972) e (Rao, 1987).

Per la storia della teoria della misura di Lebesgue si veda (Hawkins, 1975). In generale, per la storia della misura e dell'integrazione, si vedano (Pier, 1994.b) e (Dieudonné, 1978.b). Un riferimento piú recente, ed eccellente, tanto per la storia quanto per tutti gli aspetti qui trattati, sono i due volumi (Pap, 2002). Un vero

trattato, che dedica grande attenzione agli aspetti della teoria che interessano la probabilità sono i due volumi (Bogachev, 2007).

Le idee fondamentali della teoria della misura sono dovute, in gran parte, a Lebesgue che, nella sua tesi di dottorato (Lebesgue, 1902), le introdusse in \mathbb{R}^n . Il concetto di misura svincolato dal sostegno costituito da \mathbb{R}^n , e chiamato talvolta “misura astratta”, fu introdotto e sviluppato da Fréchet (1915). Questi notò che non è essenziale che gli insiemi E per i quali è definita la misura $\mu(E)$ siano misurabili nel senso di Lebesgue: basta che essi costituiscano una tribú \mathcal{F} e che μ sia numerabilmente additiva su \mathcal{F} . Questo svincolava la costruzione della misura dalla topologia dello spazio \mathbb{R}^n e consentiva così di definire la misura su insiemi “astratti” qualsiasi Ω ; è il punto di vista moderno.

La bibliografia sulle probabilità vere e proprie è molto vasta; di seguito dò una selezione dei libri sulla Probabilità che ho tenuto presenti nello scrivere queste lezioni. Si tratta di un elenco parziale limitato alla letteratura in italiano, inglese e francese; su molti punti ho fatto riferimento ad altre fonti, che citerò nelle note ai vari capitoli. Qui ricordo:

(Ash, 1972), (Bauer, 1981 e 1986), (Breiman, 1968), (Chow & Teicher, 1978), (Chung, 1968), (Cramér, 1946), (Dudley, 1989), (Durrett, 1991), (Feller, 1971), (Fristedt & Gray, 1997), (Jacod & Protter, 2000), (Klenke, 2008), (Kolmogorov, 1933), (Koralov & Sinai, 2007), (Laha & Rohatgi, 1979), (Loève, 1963), (Métivier, 1968), (Neveu, 1964), (Rao, 1984), (Rényi, 1966), (Stromberg, 1994), (Tucker, 1967), (Williams, 1991).

Non si può tacere che, tra i libri citati sopra, tre hanno avuto una grande importanza dal punto di vista storico: (Kolmogorov, 1933), monografia densissima nella quale la probabilità ha trovato il suo assetto moderno e che riporta i risultati fondamentali, (Cramér, 1946) e (Feller, 1950) che sono stati, per lungo tempo, i soli testi di riferimento per gli studiosi.

Un approccio differente da quello tradizionale, basato sull’assiomatizzazione delle speranze anziché delle probabilità si può trovare in (Whittle, 1992).

Sarà bene tenere presenti le voci dell’enciclopedia (Kotz & Johnson, 1982). Si consultino anche i libri di esempi e controesempi, che sono sempre utili per mettere a cimento le proprie conoscenze, (Romano & Siegel, 1986), (Stoyanov, 1987) e (Székely, 1986).

Sezione 1.3 Per il Teorema 1.3.1 si vedano (Tarski, 1938) e (Horn & Tarski, 1948).

Per tutta la problematica riguardante le misure finitamente additive si consulti (Bhaskara Rao & Bhaskara Rao, 1983). Di seguito dò la dimostrazione del Teorema di Tarski basata sul Teorema di Hahn–Banach.

Dimostrazione. Sia B lo spazio di Banach di tutte le funzioni a valori reali limitate definite in Ω , munito della norma della convergenza uniforme

$$\|f\| := \sup_{\omega \in \Omega} |f(\omega)|.$$

Sia S il sottospazio di B costituito dalle funzioni semplici e si definisca il funzionale $I : S \rightarrow \mathbb{R}$ mediante

$$I(f) = I\left(\sum_{i=1}^n c_i \mathbf{1}_{A_i}\right) := \sum_{i=1}^n c_i \mu(A_i).$$

Tale funzionale è omogeneo: se α è un numero reale, si ha $I(\alpha f) = \alpha I(f)$. Se f e g sono due funzioni semplici

$$f = \sum_{i=1}^n c_i \mathbf{1}_{A_i} \quad \text{e} \quad g = \sum_{j=1}^k d_j \mathbf{1}_{B_j},$$

ove supporremo che sia

$$A_i \cap A_j = B_i \cap B_j = \emptyset \quad (i \neq k) \quad \text{e} \quad \bigcup_{i=1}^n A_i = \bigcup_{j=1}^k B_j = \Omega,$$

si può scrivere

$$f + g = \sum_{i=1}^n \sum_{j=1}^k (c_i + d_j) \mathbf{1}_{A_i \cap B_j},$$

onde

$$\begin{aligned} I(f + g) &= \sum_{i=1}^n \sum_{j=1}^k (c_i + d_j) \mu(A_i \cap B_j) \\ &= \sum_{i=1}^n \sum_{j=1}^k c_i \mu(A_i \cap B_j) + \sum_{i=1}^n \sum_{j=1}^k d_j \mu(A_i \cap B_j) \\ &= \sum_{i=1}^n c_i \mu \left[A_i \cap \left(\bigcup_{j=1}^k B_j \right) \right] + \sum_{j=1}^k d_j \mu \left[\left(\bigcup_{i=1}^n A_i \right) \cap B_j \right] \\ &= \sum_{i=1}^n c_i \mu(A_i) + \sum_{j=1}^k d_j \mu(B_j) = I(f) + I(g). \end{aligned}$$

Il funzionale I è dunque lineare; di piú, esso è limitato, perché

$$|I(f)| \leq \left(\max_{i=1,2,\dots,n} c_i \right) \sum_{i=1}^n \mu(A_i) = \|f\| \mu(\Omega);$$

di qui

$$\|I\| \leq \mu(\Omega). \quad (1.18.1)$$

Per il teorema di Hahn–Banach, si può prolungare il funzionale I come funzionale L a tutto B in modo da conservarne la norma, $\|L\| = \|I\|$. Per ogni sottoinsieme A di Ω si ponga

$$\nu(A) := L(\mathbf{1}_A).$$

La linearità di L dà l'additività semplice di ν ; inoltre, se A appartiene all'algebra \mathcal{A} si ha

$$\nu(A) = L(\mathbf{1}_A) = I(\mathbf{1}_A) = \mu(A),$$

sicché ν estende μ . Infine dimostriamo che ν è positiva. A tal fine, si supponga, se possibile, che esista un sottoinsieme B di Ω tale che $\nu(B) < 0$. In tal caso si avrebbe per il complementare B^c di B

$$\|L\| = \|L\| \|\mathbf{1}_{B^c}\| \geq |L(\mathbf{1}_{B^c})| \geq \nu(B^c) = \nu(\Omega) - \nu(B) > \mu(\Omega),$$

ciò che implica

$$\|I\| = \|L\| > \mu(\Omega),$$

che contraddice alla (1.18.1). \square

Sezione 1.5 Per il Teorema 1.5.2 si vedano (Banach & Kuratowski, 1929) e (Ulam, 1930). Avverto però che le varianti di questo teorema sono numerose.

Alle funzioni d'insieme finitamente additive, note anche come *cariche* è dedicato il libro (Bhaskara Rao & Bhaskara Rao, 1983).

Uno dei problemi che mal si prestano ad un modello numerabilmente additivo è quello della prima cifra decimale; per questo, si veda (Scozzafava, 1981).

Sezione 1.7 L'integrale di Stieltjes fu introdotto in (Stieltjes, 1894); l'integrale introdotto da Stieltjes si potrebbe chiamare di Riemann–Stieltjes. Questo autore utilizzava la nozione, comune in fisica, di distribuzione di massa: $\varphi(x)$ rappresenta la massa concentrata nell'intervallo $]0, x]$ e i salti di φ rappresentano la massa concentrata nei punti di discontinuità. Si deve a Radon (1913), l'aver riconosciuto che si potevano adattare i metodi di Borel e di Lebesgue, sostituendo alla lunghezza $b - a$ dell'intervallo $]a, b]$ la quantità $\varphi(b) - \varphi(a)$ per ottenere la misura di un aperto di \mathbb{R} .

Sezione 1.10 I risultati di questa sezione furono tutti introdotti dagli autori dei quali portano il nome, per l'integrale di Lebesgue, e dunque in un contesto differente da quello nel quale li presentiamo. Si vedano (Levi, 1906), (Fatou, 1906), (Lebesgue, 1902).

Sezione 1.14 Gli spazi di Hilbert furono introdotti e studiati dal punto di vista geometrico da (Schmidt, 1908). Gli spazi L^p furono introdotti e studiati da Riesz (1910). Fu introdotta da Jensen (1906) la diseuguaglianza che porta il suo nome. Hölder (1889) dimostrò la diseuguaglianza (1.13.7) nel caso delle somme finite; egli indicò chiaramente il suo debito verso Rogers (1888), tanto che qualche autore chiama la (1.13.7) diseuguaglianza di Rogers–Hölder. Anche la diseuguaglianza (1.13.8) fu dimostrata nel caso delle somme finite da Minkowski (1896). Entrambe le diseuguaglianze furono estese agli integrali da Riesz (1910).

In generale nel rintracciare gli autori delle diseuguaglianze occorre tenere presente il monito contenuto nell'introduzione di (Hardy et al., 1934): *... we speak of the inequalities of Schwarz, Hölder and Jensen, though all these inequalities can be traced further back; and we do not enumerate explicitly all the minor additions which are necessary for absolute completeness.*

Di fatto ci siamo occupati, sia pur brevemente, dei soli spazi L^p con $p \geq 1$, e solo di questi darò le applicazioni; ciò accade perché in uno spazio L^p con $0 < p < 1$ viene meno la diseuguaglianza triangolare e il solo funzionale lineare è quello nullo. Si veda, a questo proposito, (Day, 1940).

Mitrinović & Lacković (1985) attribuiscono l'origine del termine *convesso* a Hermite che lo introdusse nel 1881.

La dimostrazione del Teorema di Fischer–Riesz che gli spazi L^p , con $p \in [1, +\infty[$, sono completi può essere consultata, per esempio, in (Dunford & Schwartz, 1958) (III.6.6).

Il teorema di rappresentazione di Riesz–Fréchet è dovuto indipendentemente a Riesz (1907) e a Fréchet (1907); esso compare in due note pubblicate nello stesso fascicolo dei *Comptes Rendus*.

Le dimostrazioni del teorema di Radon–Nikodym si possono ricondurre a tre: quella originale degli autori dei quali porta il nome, (Radon, 1913) e (Nikodym, 1930), quella data in questo capitolo e dovuta a von Neumann (1940) ed infine quella basata sul teorema di convergenza per le martingale. Tutte e tre le dimostrazioni si possono trovare in (Rao, 1987).

Il teorema di Radon–Nikodym è essenziale per molte applicazioni anche alla Statistica matematica; a questo proposito, si veda l'articolo (Halmos & Savage, 1949).

Sezione 1.15 Lebesgue (1904) dimostrò che per una funzione limitata e misurabile f definita in un rettangolo $(a, b) \times (a, b)$ i due integrali iterati assumono lo stesso valore

$$\int_a^b dx \int_a^b f(x, y) dy = \int_a^b dy \int_a^b f(x, y) dx.$$

Lebesgue trovò quindi un caso particolare del teorema di Fubini; inoltre, prendendo come f una funzione indicatrice, si prova l'esistenza della misura prodotto, almeno sui rettangoli del tipo $(a, b) \times (a, b)$.

Per il teorema di Fubini si vedano (Fubini, 1907), (Tonelli, 1909) e l'ultimo lavoro di Fubini, pubblicato postumo, (Fubini, 1949), nel quale si spiega la genesi di tale teorema. Per le cautele necessarie nell'utilizzo del teorema di Fubini si veda l'articolo (Chatterji, 1986).

1.19 Esercizi sul Capitolo 1

1.1. (a) Sia (\mathcal{A}_n) una successione crescente di algebre di sottoinsiemi di Ω . Anche $\mathcal{A} := \cup_{n \in \mathbb{N}} \mathcal{A}_n$ è un'algebra.

(b) Se (\mathcal{F}_n) una successione crescente di tribú di sottoinsiemi di Ω , l'unione $\cup_{n \in \mathbb{N}} \mathcal{F}_n$ non è necessariamente una tribú.

1.2. (a) La famiglia dei sottoinsiemi di \mathbb{N} che o sono finiti o hanno il complementare finito, cioè

$$\mathcal{A} := \{A \subseteq \mathbb{N} : \text{card}(A) < \aleph_0 \text{ o } \text{card}(A^c) < \aleph_0\}$$

è un'algebra ma non una tribú.

(b) Se Ω non è finito o numerabile ($\text{card}(\Omega) > \aleph_0$), la famiglia dei sottoinsiemi di Ω che o sono finiti o numerabili o che hanno il complementare finito o numerabile, cioè

$$\mathcal{F} := \{A \subseteq \Omega : \text{card}(A) \leq \aleph_0 \text{ o } \text{card}(A^c) \leq \aleph_0\}$$

è una tribú.

1.3. Sia \mathcal{A} una famiglia di sottoinsiemi di Ω che goda delle seguenti proprietà:

(a) $\Omega \in \mathcal{A}$;

- (b) $A \in \mathcal{A} \implies A^c \in \mathcal{A}$;
- (c) \mathcal{A} è stabile rispetto all'unione finita ($A_1, A_2 \in \mathcal{A} \implies A_1 \cup A_2 \in \mathcal{A}$);
- (d) se $A_n \in \mathcal{A}$ per ogni $n \in \mathbb{N}$ e se gli insiemi della successione (A_n) sono a due a due disgiunti, allora $\cup_{n \in \mathbb{N}} A_n \in \mathcal{A}$.

Allora \mathcal{A} è una tribù.

1.4. Siano \mathcal{F} una tribù e A e B insiemi di \mathcal{F} tali che $A \subseteq B$ e $A \neq B$. Esiste allora una probabilità \mathbb{P} su \mathcal{F} tale che $\mathbb{P}(A) < \mathbb{P}(B)$.

1.5. Se per ogni n di \mathbb{N} si ha $A_n \in \mathcal{F}$ e $\mathbb{P}(A_n) = 1$, allora

$$\mathbb{P}\left(\bigcap_{n \in \mathbb{N}} A_n\right) = 1.$$

1.6. Per un'arbitraria successione $(A_n) \subseteq \mathcal{F}$ riesce

$$\mathbb{P}\left(\liminf_{n \rightarrow +\infty} A_n\right) \leq \liminf_{n \rightarrow +\infty} \mathbb{P}(A_n) \leq \limsup_{n \rightarrow +\infty} \mathbb{P}(A_n) \leq \mathbb{P}\left(\limsup_{n \rightarrow +\infty} A_n\right).$$

sicché $A_n \rightarrow A$ implica $\mathbb{P}(A_n) \rightarrow \mathbb{P}(A)$ anche se la successione non è monotona, in altre parole, le probabilità passano al limite lungo tutte le successioni convergenti di eventi. Si dia l'esempio di uno spazio di probabilità e di una successione di eventi per i quali le diseuguaglianze precedenti siano tutte strette.

1.7. $\mathbb{P}(A_n \Delta A) \rightarrow 0$ implica sia $\mathbb{P}(A_n) \rightarrow \mathbb{P}(A)$ sia $\mathbb{P}(A_n \cap A) \rightarrow \mathbb{P}(A)$.

1.8. Sia \mathcal{A} l'algebra dell'esercizio (1.2)(a). Se $\mu : \mathcal{A} \rightarrow \{0, 1\}$ è definita mediante

$$\mu(A) = \begin{cases} 0, & \text{se } \text{card}(A) < \aleph_0, \\ 1, & \text{se } \text{card}(A^c) < \aleph_0, \end{cases}$$

allora μ è finitamente additiva, ma non σ -additiva.

1.9. Nell'insieme \mathbb{N} dei numeri naturali si consideri la famiglia \mathcal{D} dei sottoinsiemi A di \mathbb{N} per i quali esiste il limite

$$\mu(A) := \lim_{n \rightarrow +\infty} \frac{\text{card}(A \cap \{1, \dots, n\})}{n}.$$

Tale limite, quando esiste, si dice *densità asintotica* dell'insieme A . Si mostri che

- (a) μ è finitamente, ma non σ -additiva su \mathcal{D} ;
- (b) \mathcal{D} contiene \emptyset e \mathbb{N} ed è stabile rispetto alla complementazione e alle unioni finite e disgiunte, mentre non è stabile rispetto alle unioni numerabili o alle unioni finite non disgiunte.

- (c) Si considerino infine gli insiemi $P := \{2n : n \in \mathbb{N}\}$ di tutti i numeri pari e $P^c := \{2n - 1 : n \in \mathbb{N}\}$ di tutti i numeri dispari. Se

$$F_n := \left(P^c \cap [2^{2n}, 2^{2n+1}] \right) \cup \left(P \cap [2^{2n+1}, 2^{2n+2}] \right)$$

e

$$F := \bigcup_{n \in \mathbb{Z}_+} F_n,$$

allora si ha tanto $P^c \in \mathcal{D}$ quanto $F \in \mathcal{D}$, ma $P^c \cap F \notin \mathcal{D}$, sicché \mathcal{D} non è un'algebra.

1.10. Nell'esercizio precedente si mostri che, per ogni $\alpha \in [0, 1]$, esiste un insieme $E \in \mathcal{D}$ tale che $\mu(E) = \alpha$.

1.11. Si dia l'esempio di una misura (necessariamente infinita) e di una successione (A_n) di insiemi misurabili per i quali $A_n \downarrow \emptyset$ ma tale che $(\mu(A_n))$ non converga a zero.

1.12. In uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, si scriva

$$A \sim B \quad \text{se} \quad P(A \Delta B) = 0$$

con A e B in \mathcal{F} ; " \sim " è una relazione d'equivalenza su \mathcal{F} compatibile con le operazioni di unione, intersezione e complementazione. Si scriva $[A]$ per indicare la classe di equivalenza di A . Inoltre, nell'insieme quoziente $\mathcal{F}^\sim := \mathcal{F}/\sim$, la funzione $d([A], [B]) := P(A \Delta B)$ ove $A \in [A]$ e $B \in [B]$, è una metrica.

1.13. Si mostri che una misura μ su (Ω, \mathcal{F}) è σ -subadditiva: se (A_n) è una successione di insiemi misurabili (non necessariamente disgiunti a due a due), allora

$$\mu \left(\bigcup_{n \in \mathbb{N}} A_n \right) \leq \sum_{n \in \mathbb{N}} \mu(A_n).$$

1.14. Per ogni anello \mathcal{R} si ha:

- (a) $\emptyset \in \mathcal{R}$;
 (b) $A, B \in \mathcal{R} \Rightarrow A \cup B \in \mathcal{R} \quad \text{e} \quad A \setminus B \in \mathcal{R}$.

1.15. Un'algebra è un anello che contiene Ω . Un anello è un'algebra se, e solo se, è stabile rispetto all'operazione di complementazione.

1.16. Ogni misura localmente finita su $(\mathbb{R}, \mathcal{B})$ è σ -finita.

1.17. In uno spazio topologico Ω si indica con $\mathcal{B}(\Omega)$ la tribú di Borel, cioè quella generata dagli insiemi aperti. Se Ω e Ω' sono spazi topologici e $f : \Omega \rightarrow \Omega'$ è una funzione continua, essa è misurabile rispetto a $\mathcal{B}(\Omega)$ e a $\mathcal{B}(\Omega')$.

1.18. Sia μ una misura infinita, ma σ -finita, definita sullo spazio misurabile (Ω, \mathcal{F}) . Allora per ogni $k > 0$ esiste un insieme A di \mathcal{F} tale che $\mu(A) \in]k, \infty[$.

1.19. Si dimostri che, data una funzione semplice

$$s = \sum_{j=1}^k \alpha_j \mathbf{1}_{A_j},$$

ove gli insiemi A_j sono disgiunti, misurabili e formano una partizione di Ω , la funzione semplice s può essere rappresentata nella forma

$$s = \sum_{j=1}^k \beta_j \mathbf{1}_{C_j},$$

ove gli insiemi C_1, C_2, \dots, C_k sono misurabili e soddisfanno alle relazioni

$$C_1 \supset C_2 \supset \dots \supset C_k.$$

1.20. Siano \mathcal{C}_1 e \mathcal{C}_2 semi-anelli di sottoinsiemi degli insiemi Ω_1 e Ω_2 rispettivamente. Allora $\mathcal{C}_1 \times \mathcal{C}_2$ è un semi-anello di sottoinsiemi di $\Omega_1 \times \Omega_2$.

1.21. Se, per ogni $n \in \mathbb{N}$, le funzioni f_n sono integrabili, e $f_n \downarrow 0$, allora $\int f_n d\mu \downarrow 0$.

1.22. Si facciano discendere le disequaglianze dell'esercizio 3) dal lemma di Fatou.

1.23. Se, per ogni $n \in \mathbb{N}$, le funzioni $f_n : \Omega \rightarrow \overline{\mathbb{R}}_+$ sono misurabili (e positive), si ha

$$\int \left(\sum_n f_n \right) d\mu = \sum_n \int f_n d\mu.$$

1.24. Se $f, g : \Omega \rightarrow \overline{\mathbb{R}}$ sono integrabili si stabilisca la disequaglianza

$$\int (f \wedge g) d\mu \leq \int f d\mu \wedge \int g d\mu.$$

Si studii il caso dell'eguaglianza.

1.25. Sia $E = \cup_{n \in \mathbb{N}} E_n$ l'unione numerabile di insiemi disgiunti di \mathcal{F} . Se $f : \Omega \rightarrow \overline{\mathbb{R}}$ è integrabile, si mostri che

$$\int_E f d\mu = \sum_{n \in \mathbb{N}} \int_{E_n} f d\mu.$$

1.26. Sia λ la misura di Lebesgue su $(\mathbb{R}, \mathcal{B})$ e, per ogni $n \in \mathbb{N}$, si definisca $f_n : \mathbb{R} \rightarrow \mathbb{R}$ mediante $f_n := -n^2 \mathbf{1}_{]0, 1/n[}$. Si mostri che per tale successione non vale il lemma di Fatou.

1.27. Se $\nu = \mu f^{-1}$, la σ -finitezza di μ è condizione necessaria, ma non sufficiente per la σ -finitezza di ν .

1.28. Si dia l'esempio di due spazi misurabili $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$ tali che la famiglia $\mathcal{F}_1 \times \mathcal{F}_2$ dei rettangoli misurabili,

$$\mathcal{F}_1 \times \mathcal{F}_2 := \{A \times B : A \in \mathcal{F}_1, B \in \mathcal{F}_2\}$$

non sia una tribú.

1.29. Si considerino gli spazi misurabili $(\Omega_i, \mathcal{F}_i, \mu_i)$ con $i = 1, 2$, ove $\Omega_i = \mathbb{N}$, $\mathcal{F}_i = \mathcal{P}(\mathbb{N})$ e μ_i è la misura del contare (cioè $\mu_i(\{n\}) = 1$ per ogni $n \in \mathbb{N}$, $i = 1, 2$). Per la funzione $f : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{Z}$ definita da $f(n, n) = n$, $f(n, n+1) = -n$, $f(n, k) = 0$ se $k \neq n, n+1$, si mostri che

$$\int_{\mathbb{N}} d\mu_1 \int_{\mathbb{N}} f d\mu_2 \neq \int_{\mathbb{N}} d\mu_2 \int_{\mathbb{N}} f d\mu_1.$$

1.30. Si applichi il teorema di Fubini alla funzione

$$(x, y) \mapsto \varphi(x, y) := y \exp\{-(1+x^2)y^2\} \quad (x, y) \in \mathbb{R}_+^2,$$

per trovare che

$$\int_{\mathbb{R}_+} \exp(-t^2) dt = \frac{\sqrt{\pi}}{2}.$$

1.31. Gli integrali iterati nel teorema di Fubini possono essere entrambi finiti ma differenti. Sia $\Omega_1 = \Omega_2 = I := [0, 1]$ e $\mu_1 = \mu_2 = \lambda$, la (restrizione della) misura di Lebesgue ai boreliani di $[0, 1]$. Sia (δ_n) una successione strettamente crescente con $\delta_1 = 0$ e $\lim_n \delta_n = 1$ e sia $g_n : I \rightarrow \mathbb{R}$ una funzione continua con supporto contenuto in $[\delta_n, \delta_{n+1}]$ e tale che $\int g_n d\lambda = 1$ per ogni $n \in \mathbb{N}$. Si definisca $f : I \times I \rightarrow \mathbb{R}$ mediante

$$f(x, y) := \sum_{n=1}^{\infty} \{g_n(x) - g_{n+1}(x)\} g_n(y).$$

Allora

$$\int_I dx \int_I f(x, y) dy \neq \int_I dy \int_I f(x, y) dx.$$

1.32. Siano $\Omega_1 = \Omega_2 = I := [0, 1]$, μ_1 la (restrizione della) misura di Lebesgue ai boreliani di I e μ_2 la misura del contare su $\mathcal{B} = \mathcal{B}(I)$, cioè $\mu_2(A) = \text{card}(A)$ se l'insieme A è finito, mentre $\mu_2(A) = +\infty$ se A è infinito. Sia, infine, D la diagonale del quadrato $I \times I$, $D := \{(x, x) : x \in I\}$. Si mostri che D è in $\mathcal{B} \otimes \mathcal{B}$ e che se $f = \mathbf{1}_D$ risulta

$$\int_I d\mu_1 \int_I f d\mu_2 \neq \int_I d\mu_2 \int_I f d\mu_1.$$

1.33. Siano (Ω, \mathcal{F}) e $(\Omega_1, \mathcal{F}_1)$ due spazi misurabili e ν una misura su \mathcal{F}_1 . Per quasi ogni $y \in \Omega_1$ rispetto a ν , sia μ_y una misura su \mathcal{F} tale che, per ogni $E \in \mathcal{F}$, la funzione $y \mapsto \mu_y(E)$ sia \mathcal{F}_1 -misurabile. Se $\mu : \mathcal{F} \rightarrow \mathbb{R}$ è definita da

$$\mu(A) := \int_{\Omega_1} \mu_y(A) d\nu(y),$$

allora:

- (a) μ è una misura su \mathcal{F} ; ed è una probabilità, $\mu(\Omega) = 1$, se $\nu(\Omega_1) = \mu_y(\Omega) = 1$ per quasi ogni $y \in \Omega_1$;

(b) se $f : \Omega \rightarrow \bar{\mathbb{R}}_+$ è \mathcal{F} -misurabile, è \mathcal{F}_1 -misurabile la funzione $g : \Omega_1 \rightarrow \mathbb{R}$ definita da

$$g(y) := \int_{\Omega} f \, d\mu_y;$$

inoltre

$$\int_{\Omega} f \, d\mu = \int_{\Omega_1} g \, d\nu.$$

1.34. Si mostri che, per due misure μ e ν su (Ω, \mathcal{F}) , con ν finita, sono equivalenti le affermazioni:

- (a) $\mu \ll \nu$ (μ è assolutamente continua rispetto a ν);
 (b) ad ogni $\varepsilon > 0$ corrisponde $\delta = \delta(\varepsilon) > 0$ tale che $\mu(E) < \varepsilon$ per ogni insieme $E \in \mathcal{F}$ con $\nu(E) < \delta$.

È ineliminabile la richiesta che ν sia finita. Per esempio, si considerino, sullo spazio misurabile (Ω, \mathcal{F}) con $\Omega = \mathbb{N}$ e $\mathcal{F} = \mathcal{P}(\Omega)$, le misure definite da

$$\mu(A) := \sum_{n \in \mathbb{N}} 2^{-n} \quad \text{e} \quad \nu(A) := \sum_{n \in \mathbb{N}} 2^n.$$

1.35. Si mostri, mediante un esempio, che la densità di una misura di probabilità P rispetto ad un'altra misura di probabilità μ , con $P \ll \mu$, non è necessariamente limitata μ -q.c..

1.36. In uno spazio misurabile finito $(\Omega, \mathcal{F}, \mu)$, si supponga che una funzione φ di \mathcal{L}^1 sia tale che per ogni insieme misurabile $A \in \mathcal{F}$ con $\mu(A) > 0$ risulti

$$0 \leq \frac{1}{\mu(A)} \int_A \varphi \, d\mu \leq 1;$$

allora φ assume valori in $[0, 1]$, μ -q.o.

1.37. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ si considerino gli insiemi misurabili A_1, A_2, \dots, A_n . Vale allora la *diseguaglianza di Bonferroni*

$$\mathbb{P} \left(\bigcup_{j=1}^n A_j \right) \geq \sum_{j=1}^n \mathbb{P}(A_j) - \sum_{i \neq j} \mathbb{P}(A_i \cap A_j).$$

1.38. Quando vale l'eguaglianza nella disequaglianza di Hölder? e nella disequaglianza di Schwarz?

1.39. Se nello spazio $(\Omega, \mathcal{F}, \mu)$ la misura μ non è finita può non valere più l'inclusione $L^q \subseteq L^p$ se $q > p \geq 1$.

- (a) Se $\Omega = \mathbb{N}$, $\mathcal{F} = \mathcal{P}(\mathbb{N})$ e μ è la misura del contare, si studi la relazione tra gli spazi $l^p := L^p(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu)$ e $l^q := L^q(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu)$ con $q > p \geq 1$;
 (b) se $\Omega = \mathbb{R}_+$, $\mathcal{F} = \mathcal{B}(\mathbb{R}_+)$, e $\mu = \lambda$ è la misura di Lebesgue, si mostri che né $L^1 \subseteq L^2$ né $L^2 \subseteq L^1$.

1.40. Siano μ_j e ν_j due misure sullo spazio misurabile $(\Omega_j, \mathcal{F}_j)$ con $j = 1, 2$ e siano $\mu = \mu_1 \otimes \mu_2$ e $\nu = \nu_1 \otimes \nu_2$ le rispettive misure prodotto su (Ω, \mathcal{F}) con $\Omega = \Omega_1 \otimes \Omega_2$ e $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$. Se μ_j è assolutamente continua rispetto a ν_j , $\mu_j \ll \nu_j$ ($j = 1, 2$), con densità f_j ($\mu_j = f_j \cdot \nu_j$), si mostri che μ è assolutamente continua rispetto a ν e si calcoli la densità f di μ rispetto a ν .

1.41. Si dia l'esempio di uno spazio misurabile (Ω, \mathcal{F}) e di due misure definite su (Ω, \mathcal{F}) che coincidono sugli insiemi di una famiglia \mathcal{C} che genera \mathcal{F} ($\mathcal{F}(\mathcal{C}) = \mathcal{F}$).

1.42. Si chiama *distanza in variazione* di due misure di probabilità μ e ν su $(\mathbb{R}, \mathcal{B})$ la quantità

$$d_V(\mu, \nu) := \sup_{B \in \mathcal{B}} |\mu(B) - \nu(B)|.$$

In questo modo si è definita una distanza sullo spazio $\mathcal{M}_1(\mathbb{R}, \mathcal{B})$ delle misure di probabilità su $(\mathbb{R}, \mathcal{B})$; inoltre si ha

$$\forall \mu, \nu \in \mathcal{M}_1(\mathbb{R}, \mathcal{B}) \quad 0 \leq d_V(\mu, \nu) \leq 1.$$

Se le probabilità μ e ν sono assolutamente continue, con densità f e g rispettivamente, $\mu = f \cdot \lambda$ e $\nu = g \cdot \lambda$ (λ è la misura di Lebesgue su $(\mathbb{R}, \mathcal{B})$), allora è

$$d_V(\mu, \nu) \leq \int |f - g| d\lambda \leq 2 d_V(\mu, \nu).$$

Capitolo 2

La Convergenza Stocastica

2.1 I lemmi di Borel–Cantelli

I seguenti due risultati sono, nella loro semplicità, fondamentali.

Lemma 2.1.1. (Primo lemma di Borel–Cantelli). *Siano $(\Omega, \mathcal{F}, \mathbb{P})$ uno spazio di probabilità e $(A_n) \subseteq \mathcal{F}$ una successione di insiemi misurabili tale che*

$$\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) < +\infty.$$

Allora

$$\mathbb{P}\left(\limsup_{n \rightarrow +\infty} A_n\right) = 0.$$

Dimostrazione. La successione di insiemi $(\cup_{j \geq n} A_j)_{n \in \mathbb{N}}$ è decrescente, sicché

$$\begin{aligned} \mathbb{P}\left(\limsup_{n \rightarrow +\infty} A_n\right) &= \mathbb{P}\left(\bigcap_{n \in \mathbb{N}} \bigcup_{j \geq n} A_j\right) \\ &= \lim_{n \rightarrow +\infty} \mathbb{P}\left(\bigcup_{j \geq n} A_j\right) \leq \lim_{n \rightarrow +\infty} \sum_{j \geq n} \mathbb{P}(A_j) = 0, \end{aligned}$$

perché $\sum_{j \geq n} \mathbb{P}(A_j)$ è il resto $(n-1)$ -esimo della serie $\sum_{n \in \mathbb{N}} \mathbb{P}(A_n)$ che è, per ipotesi, convergente. \square

Lemma 2.1.2. (Secondo lemma di Borel–Cantelli). *Se gli eventi $A_n \in \mathcal{F}$ sono indipendenti e se $\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) = +\infty$, allora*

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

Dimostrazione. Come sopra

$$\mathbb{P}\left(\limsup_{n \rightarrow +\infty} A_n\right) = \lim_{n \rightarrow +\infty} \mathbb{P}\left(\bigcup_{j \geq n} A_j\right).$$

Poiché

$$\mathbb{P}\left(\bigcup_{j \geq n} A_j\right) = 1 - \mathbb{P}\left(\bigcap_{j \geq n} A_j^c\right) = 1 - \prod_{j \geq n} \mathbb{P}(A_j^c),$$

basterà mostrare che $\prod_{j \geq n} \mathbb{P}(A_j^c) \xrightarrow{n \rightarrow +\infty} 0$. A tal fine, si ricordi la diseuguaglianza $\ln x \leq x - 1$, nella quale il segno d'eguaglianza vale se, e solo se, $x = 1$. Posto, per $t \in [0, 1[$, $x = 1 - t$, si ha $\ln(1 - t) \leq -t$, onde, per $s \geq n$,

$$\begin{aligned} \ln \prod_{j=n}^s \mathbb{P}(A_j^c) &= \sum_{j=n}^s \ln \mathbb{P}(A_j^c) \\ &= \sum_{j=n}^s \ln(1 - \mathbb{P}(A_j)) \leq - \sum_{j=n}^s \mathbb{P}(A_j) \xrightarrow{s \rightarrow +\infty} -\infty, \end{aligned}$$

onde $\prod_{j \geq n} \mathbb{P}(A_j^c) \xrightarrow{n \rightarrow +\infty} 0$. □

Il seguente risultato è una conseguenza immediata dei lemmi di Borel–Cantelli. Esso è il prototipo delle cosiddette *leggi zero-uno*; sono, queste, teoremi che asseriscono che la probabilità di certi insiemi può assumere solo i valori zero o uno.

Teorema 2.1.1. (Legge zero-uno di Borel). *Se (A_n) è una successione di eventi indipendenti di \mathcal{F} , risulta*

$$\mathbb{P}\left(\limsup_{n \rightarrow +\infty} A_n\right) = 0 \quad \text{oppure} \quad \mathbb{P}\left(\limsup_{n \rightarrow +\infty} A_n\right) = 1,$$

secondo che risulti convergente o divergente la serie $\sum_{n \in \mathbb{N}} \mathbb{P}(A_n)$.

Ecco alcune applicazioni dei lemmi di Borel–Cantelli al gioco di testa o croce (o, se si preferisce, ad un processo di Bernoulli).

Esempio 2.1.1. Si consideri un'assegnata sequenza finita di risultati,

$$s := (x_1, \dots, x_k)$$

ove $x_i = 1$ oppure 0 ($i = 1, 2, \dots, k$) e si considerino gli eventi

$$A_n := \left\{ (x_n)_{n \in \mathbb{N}} \in \{0, 1\}^{\mathbb{N}} : (x_n, x_{n+1}, \dots, x_{n+k-1}) = s \right\}$$

e sia $p \in]0, 1[$ la probabilità di successo, $\mathbb{P}(x_n = 1) = p$; $\{0, 1\}^{\mathbb{N}}$ è l'insieme di tutte le successioni composte di 0 e di 1.

Per far vedere che si ha $\mathbb{P}(\limsup_{n \rightarrow +\infty} A_n) = 1$, non si può applicare il secondo lemma di Borel–Cantelli direttamente alla successione (A_n) perché gli eventi che la compongono non sono indipendenti. Si considerino, tuttavia, gli eventi

$$B_n := A_{(n-1)k+1} = \left\{ (x_n) \in \{0, 1\}^{\mathbb{N}} : (x_{(n-1)k+1}, x_{(n-1)k+2}, \dots, x_{nk}) = s \right\}.$$

Questi sono, evidentemente, indipendenti ed inoltre verificano l'inclusione

$$\limsup_{n \rightarrow +\infty} B_n \subseteq \limsup_{n \rightarrow +\infty} A_n.$$

Ora, $\mathbb{P}(B_n) = \mathbb{P}(B_1) = \mathbb{P}(s) > 0$ per ogni $n \in \mathbb{N}$, sicché $\sum_{n \in \mathbb{N}} \mathbb{P}(B_n) = +\infty$, ciò che implica

$$\mathbb{P}\left(\limsup_{n \rightarrow +\infty} B_n\right) = 1,$$

e, *a fortiori*, $\mathbb{P}(\limsup_{n \rightarrow +\infty} A_n) = 1$. Perciò, in una serie infinita di lanci di una moneta, è eguale a 1 la probabilità che un'assegnata sequenza finita di teste e di croci torni a verificarsi un numero infinito di volte.

Questo risultato complementa quello che si è già incontrato per il quale occorrerà aspettare, perché si realizzi per la prima volta la sequenza s , un tempo che è inversamente proporzionale alla probabilità di s . ■

Ancora, in un processo di Bernoulli, si ponga $Y_i(1) = 1$, $Y_i(0) = -1$ e $G_n = \sum_{i \leq n} Y_i$. Quando $G_n = 0$, si dirà che, al tempo n , il gioco è in parità, oppure, pensando alla passeggiata aleatoria, che al tempo n , il processo è nell'origine. Se $A_n := \{G_n = 0\}$, $\limsup_{n \rightarrow +\infty} A_n$ rappresenta l'evento “si ha parità infinite volte” oppure, pensando alla passeggiata aleatoria, l'evento “il processo torna infinite volte nell'origine”.

Teorema 2.1.2. *Se $p \neq 1/2$, allora si ha*

$$\mathbb{P}\left(\limsup_{n \rightarrow +\infty} A_n\right) = 0.$$

Dimostrazione. Ora $\mathbb{P}(A_{2n+1}) = 0$, mentre

$$\mathbb{P}(A_{2n}) = \mathbb{P}(G_{2n} = 0) = \binom{2n}{n} p^n q^n,$$

onde

$$\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) = \sum_{j \in \mathbb{N}} \mathbb{P}(A_{2j}) = \sum_{j \in \mathbb{N}} \binom{2j}{j} p^j q^j.$$

Quest'ultima serie è convergente, come subito ci si assicura, usando, ad esempio, il criterio del rapporto; l'asserto è ora un'immediata conseguenza del Lemma 2.1.1. □

Rimane da esaminare il caso $p = 1/2$. Stabiliremo prima il seguente risultato ausiliario

Lemma 2.1.3. *Per ogni $\alpha \in]0, 1[$ esiste una funzione $\varphi : \mathbb{N} \rightarrow \mathbb{N}$ tale che*

$$\mathbb{P}(|G_{\varphi(j)}| < j) \leq \alpha.$$

Dimostrazione. Per ogni $k \in \mathbb{Z}$, si ha

$$\lim_{n \rightarrow +\infty} \mathbb{P}(G_n = k) = 0;$$

perciò

$$\lim_{n \rightarrow +\infty} \sum_{|k| < j} \mathbb{P}(G_n = k) = 0.$$

Basta allora scegliere $\varphi(j)$ abbastanza grande da avere

$$\sum_{|k| < j} \mathbb{P}(G_{\varphi(j)} = k) \leq \alpha,$$

che conclude la dimostrazione. □

Teorema 2.1.3. *Se $p = 1/2$, allora si ha*

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

Dimostrazione. Come nell'esempio 2.1.1 non si può usare direttamente il secondo lemma di Borel–Cantelli perché gli eventi della successione (A_n) non sono indipendenti. Si consideri una successione strettamente crescente (n_j) estratta da quella dei numeri naturali e si supponga che sia $n_{j+1} - n_j \geq 2$. Per ogni $j \in \mathbb{N}$, sia s_j un naturale con $n_j < s_j < n_{j+1}$ e si definisca

$$B_j := \left\{ \sum_{k=n_j+1}^{s_j} Y_k \leq -n_j \right\} \cap \left\{ \sum_{k=s_j+1}^{n_{j+1}} Y_k \geq s_j \right\}.$$

Si noti che B_j è un insieme misurabile e che dipende dalle v.a.

$$Y_{n_j+1}, Y_{n_j+2}, \dots, Y_{n_{j+1}};$$

poiché le v.a. della successione (Y_n) sono indipendenti anche gli insiemi (B_j) lo sono.

Ora, poiché ogni v.a. Y_k assume solo i valori 1 e -1 , si ha

$$G_{n_j} = \sum_{k=1}^{n_j} Y_k \leq n_j \quad \text{e} \quad G_{s_j} \geq -s_j.$$

perciò, se $\Omega \in B_j$, vale tanto $G_{s_j}(\Omega) \leq 0$ quanto $G_{n_{j+1}}(\Omega) \geq 0$, sicché esiste un naturale n compreso tra $n_j + 1$ e n_{j+1} per il quale $G_n(\Omega) = 0$. Dunque

$$\limsup_{j \rightarrow +\infty} B_j \subseteq \limsup_{n \rightarrow +\infty} \{G_n = 0\}.$$

In virtù del secondo lemma di Borel–Cantelli, per dimostrare l'asserto, basta far vedere che è possibile scegliere i numeri n_j e s_j in modo che sia

$$\sum_{j \in \mathbb{N}} \mathbb{P}(B_j) = +\infty. \quad (2.1.1)$$

Si proceda ora a scegliere i numeri n_j e s_j come segue

$$n_1 = 1 \quad s_1 = 1 + \varphi(1),$$

e, per $j > 1$,

$$s_j = n_j + \varphi(n_j), \quad n_{j+1} = s_j + \varphi(s_j).$$

Ora

$$\mathbb{P}(B_j) = \mathbb{P} \left(\sum_{k=n_j+1}^{s_j} Y_k \leq -n_j \right) \mathbb{P} \left(\sum_{k=s_j+1}^{n_{j+1}} Y_k \geq s_j \right),$$

e di qui, per la simmetria della passeggiata aleatoria ($p = 1/2$),

$$\mathbb{P}(B_j) = \frac{1}{4} \mathbb{P} \left(\sum_{k=n_j+1}^{s_j} |Y_k| \geq n_j \right) \mathbb{P} \left(\sum_{k=s_j+1}^{n_{j+1}} |Y_k| \geq s_j \right);$$

ma i vettori aleatorî $(Y_{i+1}, Y_{i+2}, \dots, Y_{i+k})$ e (Y_1, Y_2, \dots, Y_k) hanno la stessa legge, sicché per il Lemma 2.1.3

$$\mathbb{P}(B_j) = \frac{1}{4} \mathbb{P} \left(\sum_{k=1}^{\varphi(n_j)} |Y_k| \geq n_j \right) \mathbb{P} \left(\sum_{k=1}^{\varphi(s_j)} |Y_k| \geq s_j \right) \geq \frac{1}{4} (1 - \alpha)^2 > 0,$$

ciò che prova la (2.1.1) e quindi l'asserto. \square

2.2 Varî tipi di convergenza stocastica

L'approccio piú ovvio alla convergenza di una successione di v.a. $(X_n)_{n \in \mathbb{N}}$ è di stabilire che la successione converga per ogni punto $\omega \in \Omega$. Per esempio, se X è una v.a. e $X_n := X + 1/n$, la successione $\{X_n\}$ cosí definita converge a X in ogni $\omega \in \Omega$. Tuttavia, questa definizione è eccessivamente restrittiva. Si consideri infatti un processo di Bernoulli nel quale sia $p \in]0, 1[$ la probabilità di un successo ad ogni prova; se $X_n = 1$ oppure 0 secondo che la n -esima prova abbia avuto come risultato un successo o un fallimento, e se $S_n := (1/n) \sum_{j \leq n} X_j$ indica la frequenza dei successi, la legge dei grandi numeri di Bernoulli asserisce che la probabilità che S_n si discosti da p può essere resa arbitrariamente piccola al tendere di n a $+\infty$. Però S_n non tende a p in ogni punto di $\Omega = \{0, 1\}^{\mathbb{N}}$; per esempio, per la successione $(0, 0, \dots, 0, \dots)$ si ha $S_n = 0$ per ogni $n \in \mathbb{N}$, mentre per la successione $(1, 1, \dots, 1, \dots)$ è $S_n = 1$ per ogni $n \in \mathbb{N}$. È facile dare esempi di successioni per le quali la frequenza dei successi è eguale ad un preassegnato numero (razionale). È cosí giustificata l'introduzione di altri tipi di convergenza. *Tutte le v.a. che compaiono nelle definizioni e nei teoremi di questa sezione si supporranno definite sopra il medesimo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$.*

Definizione 2.2.1. Si dice che la successione (X_n) converge *quasi certamente* alla v.a. X se esiste un insieme trascurabile $N \in \mathcal{F}$, $\mathbb{P}(N) = 0$, tale che per ogni $\omega \in N^c$ si abbia $X_n(\omega) \rightarrow X(\omega)$. Si scrive allora

$$X_n \xrightarrow[n \rightarrow +\infty]{q.c.} X.$$

Diremo questo tipo di convergenza *convergenza quasi certa*. \diamond

È opportuno porre la Definizione 2.2.1 in altri termini. Si definisca, per $\varepsilon > 0$ fissato arbitrariamente, l'insieme $A_{j,\varepsilon} := \{|X_j - X| \leq \varepsilon\}$. Poiché tanto X_j quanto X sono misurabili, si ha che $A_{j,\varepsilon}$ appartiene a \mathcal{F} . Si considerino, inoltre, gli insiemi, che sono tutti misurabili,

$$S_{n,\varepsilon} := \bigcap_{j \geq n} A_{j,\varepsilon}$$

e

$$S_\varepsilon := \bigcup_{n \in \mathbb{N}} S_{n,\varepsilon} = \bigcup_{n \in \mathbb{N}} \bigcap_{j \geq n} A_{j,\varepsilon} = \liminf_{n \rightarrow +\infty} A_{n,\varepsilon}.$$

Si osservi che, se $0 < \varepsilon < \varepsilon'$, allora, per ogni $j \in \mathbb{N}$, si ha $A_{j,\varepsilon} \subseteq A_{j,\varepsilon'}$, e, di conseguenza, $S_\varepsilon \subseteq S_{\varepsilon'}$. Sia, ora, (ε_n) una successione decrescente e infinitesima con $\varepsilon_n > 0$ per ogni $n \in \mathbb{N}$ e si ponga $S := \bigcap_{n \in \mathbb{N}} S_{\varepsilon_n} \in \mathcal{F}$. L'insieme S non dipende dalla

scelta della successione (ε_n) ; i punti di S sono tutti, e soli, i punti $\omega \in \Omega$ per i quali risulta $|X_n(\omega) - X(\omega)| \leq \varepsilon$, comunque si fissi $\varepsilon > 0$, per tutti gli indici n , tranne, al piú, un numero finito di essi; tale numero dipende in generale dal punto ω ; si può dunque scrivere $\mathbb{P}(S) = \mathbb{P}(\lim_{n \rightarrow +\infty} X_n = X)$. Pertanto, dire che la successione (X_n) converge q.c. alla v.a. X equivale a dire che $\mathbb{P}(S) = 1$.

Poiché la scelta della successione infinitesima (ε_n) è irrilevante, si può prendere, senza perdita di generalità, $\varepsilon_n = 1/n$. In questo modo, la condizione che esprime la convergenza q.c. può essere espressa nella forma

$$\mathbb{P} \left(\bigcap_{m \in \mathbb{N}} \bigcup_{n \in \mathbb{N}} \bigcap_{j \geq n} \left\{ |X_j - X| \leq \frac{1}{m} \right\} \right) = 1.$$

È spesso utile la seguente caratterizzazione

Teorema 2.2.1. *Per la successione (X_n) di v.a. sono equivalenti le proprietà:*

- (a) (X_n) converge q.c. alla v.a. X ;
- (b) per ogni $\delta > 0$, risulta $\lim_{n \rightarrow +\infty} \mathbb{P}(S_{n,\delta}) = 1$.

Dimostrazione. (a) \implies (b) Si è visto che $\mathbb{P}(S) = 1$ e poiché $S \subseteq S_{\varepsilon_n}$ per ogni $n \in \mathbb{N}$, si ha $\mathbb{P}(S_{\varepsilon_n}) = 1$ per ogni $n \in \mathbb{N}$. Fissato arbitrariamente $\delta > 0$, si scelga un elemento ε_k della successione (ε_n) in modo che sia $\varepsilon_k < \delta$. Segue dalla definizione che $S_{\varepsilon_k} \subseteq S_\delta$ così che $\mathbb{P}(S_\delta) = 1$. Poiché $S_\delta = \bigcup_n S_{n,\delta} = \lim_{n \rightarrow +\infty} S_{n,\delta}$, si ha l'asserto.

(b) \implies (a) Fissato arbitrariamente $\varepsilon > 0$, si ponga $\delta = \varepsilon 2^{-k}$; perciò, esiste un numero naturale $n_k = n(\varepsilon, k)$ tale che, per ogni $n \geq n_k$, sia

$$\mathbb{P}(S_{n,\varepsilon 2^{-k}}) > 1 - \varepsilon 2^{-k}.$$

Poiché

$$S = \bigcap_{k \in \mathbb{N}} S_{\varepsilon 2^{-k}} \supseteq \bigcap_{k \in \mathbb{N}} S_{n_k, \varepsilon 2^{-k}} = \left(\bigcup_{k \in \mathbb{N}} S_{n_k, \varepsilon 2^{-k}}^c \right)^c,$$

si ha

$$\begin{aligned} \mathbb{P}(S) &\geq 1 - \mathbb{P} \left(\bigcup_{k \in \mathbb{N}} S_{n_k, \varepsilon 2^{-k}}^c \right) \\ &\geq 1 - \sum_{k \in \mathbb{N}} \mathbb{P} \left(S_{n_k, \varepsilon 2^{-k}}^c \right) > 1 - \sum_{k \in \mathbb{N}} \varepsilon 2^{-k} = 1 - \varepsilon, \end{aligned}$$

che dà l'asserto. □

La condizione (b) dell'ultimo teorema si può scrivere in una delle forme equivalenti:

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\bigcap_{j=n}^{\infty} \{|X_j - X| \leq \delta\} \right) = 1; \quad (2.2.1)$$

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\sup_{j \geq n} |X_j - X| \leq \delta \right) = 1; \quad (2.2.2)$$

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\bigcup_{j=n}^{\infty} \{|X_j - X| > \delta\} \right) = 0; \quad (2.2.3)$$

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\sup_{j \geq n} |X_j - X| > \delta \right) = 0. \quad (2.2.4)$$

Il seguente semplice risultato illustra l'uso dei lemmi di Borel–Cantelli nello studio della convergenza quasi certa.

Teorema 2.2.2. *Per una successione (X_n) di v.a. definite sullo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sono equivalenti le seguenti condizioni:*

- (a) $X_n \rightarrow 0$ q.c.;
- (b) $\forall \varepsilon > 0 \quad \mathbb{P}(\limsup_{n \rightarrow +\infty} \{|X_n| \geq \varepsilon\}) = 0$;
- (c) $\forall \varepsilon > 0 \quad \mathbb{P}(\liminf_{n \rightarrow +\infty} \{|X_n| < \varepsilon\}) = 1$;

Dimostrazione. Poiché è ovvia l'equivalenza tra le condizioni (b) e (c), basta dimostrare che una di queste, per esempio la (c), è equivalente alla (a).

Ora, per il Teorema 2.2.1, $X_n \rightarrow 0$ q.c. se, e solo se, posto

$$S_{n,\varepsilon} := \bigcap_{j \geq n} \{|X_j| < \varepsilon\},$$

è $\lim_{n \rightarrow +\infty} \mathbb{P}(S_{n,\varepsilon}) = 1$, per ogni $\varepsilon > 0$, che è la (c) perché la successione $(S_{n,\varepsilon})$ è crescente. \square

Definizione 2.2.2. Si dice che una successione (X_n) di v.a. definite sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ è di *Cauchy rispetto alla convergenza* q.c. se esiste un insieme $N \in \mathcal{F}$ con $\mathbb{P}(N) = 0$ tale che, per ogni $\varepsilon > 0$ e per ogni $\omega \in N^c$, si possa determinare un naturale $n_0 = n_0(\varepsilon) \in \mathbb{N}$ con la proprietà che, per ogni scelta di n e m con $m, n \geq n_0$, si abbia $|X_n(\omega) - X_m(\omega)| < \varepsilon$. In termini più formali, se

$$\mathbb{P} \left(\bigcap_{m \in \mathbb{N}} \bigcup_{n \in \mathbb{N}} \bigcap_{j, k \geq n} \left\{ |X_j - X_k| < \frac{1}{m} \right\} \right) = 1,$$

allora (X_n) è una successione di Cauchy. \diamond

L'ultima equazione può essere scritta in forme differenti ma equivalenti; per esempio, equivale a richiedere che, per ogni $\varepsilon > 0$, valga

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\sup_{k \in \mathbb{N}} |X_{n+k} - X_n| \leq \varepsilon \right) = 1. \quad (2.2.5)$$

Teorema 2.2.3. *Per una successione (X_n) di v.a. definite sullo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sono equivalenti le affermazioni:*

- (a) (X_n) converge q.c.
- (b) (X_n) è di Cauchy rispetto alla convergenza quasi certa.

Dimostrazione. (a) \implies (b) Si supponga che esista una v.a. X tale $X_n \rightarrow X$ q.c.. Vogliamo dimostrare che per ogni $\varepsilon > 0$ vale la (2.2.5). Si introducano gli insiemi

$$B_{n,k}^\varepsilon := \{|X_{n+k} - X_n| \leq \varepsilon\}, \quad T_k^\varepsilon := \bigcap_{n \in \mathbb{N}} B_{n,k}^\varepsilon$$

$$S_n^\varepsilon := \bigcap_{k \geq n} T_k^\varepsilon, \quad S^\varepsilon := \bigcup_{n \in \mathbb{N}} S_n^\varepsilon.$$

Poiché $|X_{n+k} - X_n| \leq |X_{n+k} - X| + |X_n - X|$, si ha, se $A_{n,\varepsilon} := \{|X_n - X| \leq \varepsilon\}$,

$$A_{n+k,\varepsilon/2} \cap A_{n,\varepsilon/2} \subseteq B_{n,k}^\varepsilon,$$

sicché

$$S_n^{\varepsilon/2} \subseteq A_{n+k,\varepsilon/2} \cap A_{n,\varepsilon/2} \subseteq B_{n,k}^\varepsilon,$$

per ogni $k \in \mathbb{N}$. Di qui

$$S_n^{\varepsilon/2} \subseteq T_k^\varepsilon = \left\{ \sup_{k \in \mathbb{N}} |X_{n+k} - X_n| \leq \varepsilon \right\}.$$

Poiché, per il Teorema 2.2.1, si ha $\mathbb{P}(S_n^{\varepsilon/2}) \rightarrow 1$, segue anche la (2.2.5).

(b) \implies (a) Sia (X_n) una successione di Cauchy rispetto alla convergenza quasi certa, vale a dire sia vera la (2.2.5); si noti che, per ogni naturale $r \geq n$, si ha

$$T_k^{\varepsilon/2} \subseteq \{|X_{r+k} - X_n| \leq \varepsilon/2\} \cap \{|X_r - X_n| \leq \varepsilon/2\}$$

$$\subseteq \{|X_{r+k} - X_r| \leq \varepsilon\},$$

sicché

$$T_k^{\varepsilon/2} \subseteq B_{r,k}^\varepsilon$$

per ogni $r \geq n$ e per ogni $k \in \mathbb{N}$. Quindi

$$T_k^{\varepsilon/2} \subseteq T_k^\varepsilon,$$

sicché $T_n^{\varepsilon/2} \subseteq S_n^\varepsilon$ e perciò

$$\lim_{n \rightarrow +\infty} \mathbb{P}(T_n^{\varepsilon/2}) \leq \lim_{n \rightarrow +\infty} \mathbb{P}(S_n^\varepsilon) = \mathbb{P}(S^\varepsilon).$$

Quest'ultima relazione e la (2.2.5) implicano

$$\mathbb{P}(S^c) = 1.$$

Sia ora (ε_j) un'arbitraria successione infinitesima di numeri strettamente positivi e si consideri l'insieme

$$S := \bigcap_{j \in \mathbb{N}} S^{\varepsilon_j},$$

per il quale risulta $\mathbb{P}(S) = 1$. La successione $(X_n(\omega))$ soddisfa alla condizione di Cauchy per ogni $\omega \in S$. Esiste dunque una funzione Y tale che per ogni $\omega \in S$ sia $\lim_{n \rightarrow +\infty} X_n(\omega) = Y(\omega)$. Tale Y può essere estesa all'intero spazio Ω , ponendo

$$X(\omega) := \begin{cases} Y(\omega), & \omega \in S, \\ 0, & \omega \in S^c. \end{cases}$$

Poiché $\mathbb{P}(S^c) = 0$, si ha che X è una v.a. e che $X_n \rightarrow X$ q.c.. \square

La convergenza quasi certa non discende, in generale, da una metrica, anzi, come si dimostrerà oltre, non è neppure topologica. Vale il seguente teorema che non dimostriamo.

Teorema 2.2.4. *Sia L^0 lo spazio delle v.a. che sono quasi certamente finite su $(\Omega, \mathcal{F}, \mathbb{P})$. Affinché esista una metrica su L^0 che sia compatibile con la convergenza quasi certa occorre e basta che Ω sia l'unione numerabile punti $\{\omega_j : j \in \mathbb{N}\}$ con $\mathbb{P}(\{\omega_j\}) > 0$ per ogni $j \in \mathbb{N}$.*

Definizione 2.2.3. Si dice che la successione (X_n) di v.a. converge *in probabilità* alla v.a. X se, per ogni $\varepsilon > 0$, risulta

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0. \quad (2.2.6)$$

Si scriverà $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X$. Chiameremo questo tipo di convergenza *convergenza in probabilità*. \diamond

Vale la pena osservare che la definizione di convergenza in probabilità asserisce che, per ogni $\varepsilon > 0$ e per ogni $\delta > 0$, esiste un numero naturale $n_0 = n_0(\varepsilon, \delta)$, tale che per ogni $n \geq n_0$,

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq \delta.$$

Si osservi che si è già incontrato un risultato significativo nel quale compariva la convergenza in probabilità: si tratta della legge dei grandi numeri di Bernoulli.

Teorema 2.2.5. *La convergenza quasi certa implica la convergenza in probabilità.*

Dimostrazione. Se la successione (X_n) converge q.c. a X essa verifica la (2.2.4) e quindi, *a fortiori*, anche la (2.2.6). \square

Si vedrà piú avanti che non è necessariamente vero il viceversa; è tuttavia notevole il seguente risultato

Teorema 2.2.6. *Da ogni successione di v.a. (X_n) che converga in probabilità ad una v.a. X , si può estrarre una sottosuccessione che converge quasi certamente allo stesso limite X .*

Dimostrazione. Si supponga che (X_n) converga in probabilità a X e si scelga $n(k)$ in \mathbb{N} in modo che risulti $\mathbb{P}(|X_{n(k)} - X| > 2^{-k}) \leq 2^{-k}$; si sceglierà l'indice $n(k+1)$ in maniera analoga, con la sola condizione che sia $n(k+1) > n(k)$. Il primo lemma di Borel–Cantelli implica

$$\mathbb{P}\left(\limsup_{k \rightarrow +\infty} \{|X_{n(k)} - X| > 2^{-k}\}\right) = 0,$$

cioè

$$\mathbb{P}\left(\liminf_{k \rightarrow +\infty} \{|X_{n(k)} - X| \leq 2^{-k}\}\right) = 1.$$

Nella notazione di questa sezione si è appena dimostrato che, con riferimento alla successione $(X_{n(k)})$, per ogni $k \in \mathbb{N}$, si ha $\mathbb{P}(S_{2^{-k}}) = 1$, sicché

$$\mathbb{P}(S) = \mathbb{P}\left(\bigcap_{k \in \mathbb{N}} S_{2^{-k}}\right) = 1,$$

vale a dire che $X_{n(k)} \xrightarrow[k \rightarrow +\infty]{} X$ q.c.. □

Si vedrà oltre che la convergenza in probabilità discende da una metrica su L^0 . Nel seguito useremo liberamente questo fatto.

Definizione 2.2.4. Si dice che la successione $(X_n) \subseteq L^p$ converge alla v.a. X in L^p (oppure in media di ordine p) con $p \in]0, +\infty[$ se

$$\lim_{n \rightarrow +\infty} \mathbb{E}(|X_n - X|^p) = 0.$$

Questo modo di convergenza sarà detto *convergenza in L^p* . ◇

Nelle convergenze delle Definizioni 2.2.1, 2.2.3 e 2.2.4, (X_n) converge a X se, e solo se, $(X_n - X)$ converge alla v.a. q.c. nulla. Nelle dimostrazioni che seguono ci si limiterà dunque, senza perdita di generalità, ad esaminare il caso in cui la v.a. limite sia (q.c.) nulla.

Teorema 2.2.7. *Se $p < r$, allora la convergenza in L^r implica quella in L^p .*

Dimostrazione. È conseguenza immediata della (4.5.9). □

Teorema 2.2.8. *La convergenza in L^p con $p \in]0, +\infty[$ implica quella in probabilità.*

Dimostrazione. È conseguenza immediata della (1.9.4). □

Quest'ultimo risultato ammette un reciproco parziale. Si vedrà nel seguito che questo teorema è conseguenza di un risultato più generale (Teorema 6.4.2).

Teorema 2.2.9. *Se (X_n) converge a X in probabilità e se esiste una v.a. Y di L^p tale che $|X_n| \leq Y$, per ogni $n \in \mathbb{N}$, allora (X_n) converge a X anche in L^p .*

Dimostrazione. È, intanto, evidente che la condizione $|X_n| \leq Y$ per ogni $n \in \mathbb{N}$, con $Y \in L^p$, assicura che la successione (X_n) è in L^p . Basta far vedere che da ogni sottosuccessione di (X_n) se ne può estrarre un'altra convergente a X in L^p . Per il Teorema 2.2.6 esiste una sottosuccessione $(X_{n(k)})$ che converge a X q.c.. Poiché, evidentemente, si ha $|X| \leq Y$, risulta, per ogni $k \in \mathbb{N}$,

$$|X_{n(k)} - X|^p \leq 2^p Y^p,$$

sicché, per il teorema di convergenza dominata, si ha

$$\lim_{k \rightarrow +\infty} \mathbb{E}(|X_{n(k)} - X|^p) = 0,$$

che dà l'asserto. □

Gli esempi che seguono illustrano i rapporti tra i vari modi di convergenza sin qui introdotti.

Esempio 2.2.1. (La convergenza in L^p non implica la convergenza quasi certa). Sia $\Omega = [0, 1]$ dotato della misura di Lebesgue; si considerino le v.a.

$$X_{rs} := \mathbf{1}_{\left] \frac{s-1}{r}, \frac{s}{r} \right[} \quad (r \in \mathbb{N}; s = 1, \dots, r)$$

e le si pensino ordinate secondo l'ordine lessicografico sí da avere la successione

$$(X_{1,1}, X_{2,1}, X_{2,2}, X_{3,1}, X_{3,2}, X_{3,3}, \dots).$$

Questa converge in L^p per ogni $p \in]0, +\infty[$; infatti

$$\mathbb{E}(|X_{r,s}|^p) = \int |X_{r,s}|^p d\mathbb{P} = \int X_{r,s}^p d\mathbb{P} = \frac{1}{r}.$$

D'altra parte, se $\omega \neq 0$, $(X_{r,s}(\omega))$ è una successione reale composta da 0 e da 1, sicché essa converge se, e solo se, è definitivamente costante; però, essa assume, per ogni valore di r , esattamente una volta il valore 1 e $r - 1$ volte il valore 0, sicché non converge. Perciò $(X_{r,s})$ non converge in alcun punto di $]0, 1]$ e quindi non converge q.c.. Si osservi che, in virtù del Teorema 2.2.8, $(X_{r,s})$ tende in probabilità alla v.a. nulla, sicché essa fornisce anche l'esempio di una successione che converge in probabilità senza convergere quasi certamente. ■

Esempio 2.2.2. (La convergenza quasi certa non implica quella in L^p). Nello stesso spazio di probabilità dell'esempio precedente, si consideri la successione (X_n) ove

$$X_n := e^n \mathbf{1}_{\left] 0, \frac{1}{n} \right]}.$$

Si controlla subito che $\lim_{n \rightarrow +\infty} X_n(\omega) = 0$ se $\omega \neq 0$, onde $X_n \rightarrow 0$ q.c.. D'altra parte, per ogni $p > 0$ è

$$\mathbb{E}(|X_n|^p) = \int X_n^p d\mathbb{P} = \frac{e^{np}}{n} \rightarrow +\infty.$$

Si osservi che, in virtù del Teorema 2.2.7, (X_n) tende in probabilità alla v.a. nulla, sicché essa fornisce l'esempio di una successione di v.a. che converge in probabilità ma non in L^p . ■

Nei due esempi 2.2.1 e 2.2.2 si è visto che la convergenza quasi certa non implica né è implicata dalla convergenza in L^p . Qui di seguito, diamo l'esempio di una successione che converge sia in L^p sia quasi certamente. I rapporti tra i due tipi di convergenza restano così del tutto chiariti.

Esempio 2.2.3. (Una successione che converge sia quasi certamente sia in L^p). Sia X_n una v.a. che assume i valori $-1/n$ e $1/n$, entrambi con probabilità $1/2$. Pertanto, $\mathbb{E}(|X_n|^p) = 1/(n^p)$ e, quindi, $X_n \rightarrow 0$ in L^p . D'altro canto, sia $A_{j,\delta} := \{|X_j| \leq \delta\}$ e si noti che, per $j < k$, si ha $|X_j| > |X_k|$, onde $A_{j,\delta} \subseteq A_{k,\delta}$ e perciò

$$S_{n,\delta} = \bigcap_{j \geq n} A_{j,\delta} = A_{n,\delta}.$$

Fissato arbitrariamente $\delta > 0$, risulta

$$\mathbb{P}(S_{n,\delta}) = \mathbb{P}(A_{n,\delta}) = \mathbb{P}(|X_n| \leq \delta) = 1,$$

se $n \geq 1/\delta$, sicché $X_n \rightarrow 0$ q.c. in virtù del Teorema 2.2.1. ■

Esempio 2.2.4. (Se $s > r$, la convergenza in L^r non implica quella in L^s). Sia (X_n) una successione di v.a. tali che $X_n = n$ con probabilità $\mathbb{P}(X_n = n) = 1/n^s$ e $X_n = 0$ con probabilità $\mathbb{P}(X_n = 0) = 1 - 1/n^s$. Perciò, $\mathbb{E}(|X_n|^r) = n^r/n^s = 1/n^{s-r} \rightarrow 0$, sicché $X_n \rightarrow 0$ in L^r , ma $\mathbb{E}(|X_n|^s) = 1$ per ogni $n \in \mathbb{N}$, e quindi (X_n) non tende a zero in L^s . ■

La convergenza in probabilità discende da una metrica. Nel prossimo teorema si dà l'esempio di una tale metrica; poiché discende da una metrica si potranno usare i risultati noti dalla topologia, in primo luogo, si potrà parlare della *topologia* della convergenza in probabilità.

Teorema 2.2.10. *Sia dato lo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$. La funzione $d_{KF} : L^0 \times L^0 \rightarrow \mathbb{R}_+$ definita da*

$$d_{KF}(X, Y) := \inf \{ \varepsilon > 0 : \mathbb{P}(|X - Y| > \varepsilon) < \varepsilon \} \quad (2.2.7)$$

è una metrica su L^0 , detta metrica di Ky Fan; inoltre la successione di v.a. (X_n) converge in probabilità alla v.a. X se, e solo se, $d_{KF}(X_n, X) \xrightarrow{n \rightarrow +\infty} 0$.

Dimostrazione. Che d_{KF} sia positiva e simmetrica è una conseguenza immediata della definizione (2.2.7). Si supponga $d_{KF}(X, Y) = 0$; ciò vuol dire, in particolare che, per ogni $n \in \mathbb{N}$ risulta

$$\mathbb{P}(|X - Y| > 1/n^2) \leq \frac{1}{n^2},$$

onde, per il primo lemma di Borel–Cantelli,

$$\mathbb{P} \left(\limsup_{n \rightarrow +\infty} \left\{ |X - Y| > \frac{1}{n^2} \right\} \right) = 0,$$

o equivalentemente,

$$\mathbb{P} \left(\liminf_{n \rightarrow +\infty} \left\{ |X - Y| \leq \frac{1}{n^2} \right\} \right) = 1,$$

cioè $X = Y$ q.c., vale a dire $X = Y$ in L^0 .

Per stabilire la disuguaglianza triangolare, siano X , Y e Z tre v.a. e si ponga $\alpha := d_{KF}(X, Y)$ e $\beta := d_{KF}(Y, Z)$. Per ogni $\varepsilon > 0$, esistono allora due insiemi misurabili $E_\alpha(\varepsilon)$ e $F_\beta(\varepsilon)$ con $\mathbb{P}(E_\alpha(\varepsilon)) \leq \alpha + \varepsilon/2$ e $\mathbb{P}(F_\beta(\varepsilon)) \leq \beta + \varepsilon/2$, tali che $|X - Y| \leq \alpha + \varepsilon/2$ in $E_\alpha^c(\varepsilon)$ e $|Y - Z| \leq \beta + \varepsilon/2$ in $F_\beta^c(\varepsilon)$. Dunque nell'insieme

$$E_\alpha^c(\varepsilon) \cap F_\beta^c(\varepsilon) = \left(E_\alpha(\varepsilon) \cup F_\beta(\varepsilon) \right)^c$$

è

$$|X - Z| \leq |X - Y| + |Y - Z| \leq \alpha + \beta + \varepsilon.$$

Ora $\mathbb{P}(E_\alpha(\varepsilon) \cup F_\beta(\varepsilon)) \leq \mathbb{P}(E_\alpha(\varepsilon)) + \mathbb{P}(F_\beta(\varepsilon)) \leq \alpha + \beta + \varepsilon$. In conclusione,

$$d_{KF}(X, Z) \leq \alpha + \beta + \varepsilon = d_{KF}(X, Y) + d_{KF}(Y, Z) + \varepsilon,$$

onde l'asserto in virtù dell'arbitrarietà di ε .

Si supponga ora che $X_n \rightarrow X$ in probabilità; allora, per ogni $\varepsilon > 0$, si ha definitivamente

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq \varepsilon;$$

perciò, $d_{KF}(X_n, X) \rightarrow 0$.

Viceversa, si supponga che sia $\lim_{n \rightarrow +\infty} d_{KF}(X_n, X) = 0$. Si scelgano ora, arbitrariamente, $\varepsilon > 0$ e $\eta > 0$ e si scelga altresì $\delta < \varepsilon \wedge \eta = \min\{\varepsilon, \eta\}$. Per n sufficientemente grande, si ha $d_{KF}(X_n, X) < \delta$, che, a sua volta, implica $\mathbb{P}(|X_n - X| > \delta) < \delta$. Poiché $\eta > \delta$, si ha

$$\mathbb{P}(|X_n - X| > \eta) \leq \mathbb{P}(|X_n - X| > \delta) < \delta < \varepsilon,$$

cioè $X_n \rightarrow X$ in probabilità. □

Si osservi che da questo teorema scende come conseguenza l'unicità del limite in probabilità, e quindi, *a fortiori*, di quello quasi certo. La (2.2.7) non costituisce l'unica maniera di metrizzare la convergenza in probabilità su L^0 . Esiste un gran numero di tali metriche; si vedrà negli esercizi come costruirle.

Si è visto sopra che la convergenza in probabilità deriva da una metrica; in genere, essa è, però, incompatibile con l'esistenza di una norma.

Teorema 2.2.11. *Dato lo spazio di probabilità sono equivalenti le proprietà:*

- (a) *la convergenza in probabilità derivi da una norma su $L^0 = L^0(\Omega, \mathcal{F}, \mathbb{P})$;*
- (b) *Ω è l'unione di un numero finito di punti $\{\omega_i\}$ con $\mathbb{P}(\{\omega_i\}) > 0$.*

Teorema 2.2.12. *Lo spazio $L^0 = L^0(\Omega, \mathcal{F}, \mathbb{P})$ è completo rispetto alla convergenza in probabilità; in altre parole, (L^0, d_{KF}) è uno spazio metrico completo.*

Dimostrazione. Sia (X_n) una successione di Cauchy in probabilità, vale a dire che, per ogni $\varepsilon > 0$, esiste un indice $\nu \in \mathbb{N}$ tale che sia

$$d_{KF}(X_n, X_m) < \varepsilon$$

se $m, n \geq \nu$. Come nel Teorema 2.2.6 si può mostrare che è possibile estrarre da (X_n) una successione $(X_{n(k)})_{k \in \mathbb{N}}$ che sia di Cauchy rispetto alla convergenza q.c.. Esiste, dunque, una v.a. X tale che $\lim_{k \rightarrow +\infty} X_{n(k)} = X$ q.c. e quindi anche in probabilità. Per mostrare che l'intera successione (X_n) converge a X in probabilità si consideri che

$$d_{KF}(X_n, X) \leq d_{KF}(X_n, X_{n(k)}) + d_{KF}(X_{n(k)}, X) < 2\varepsilon,$$

se n e $n(k)$ sono abbastanza grandi. \square

2.3 La convergenza completa

Un tipo di convergenza che riveste particolare importanza in probabilità è la *convergenza completa*. È spesso importante, si ricordi il teorema integrale di de Moivre–Laplace, considerare, data una successione (X_n) di v.a. definite sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, il limite della probabilità $\mathbb{P}(X_n \in]a, b])$ che la v.a. X_n assuma valori nell'intervallo $]a, b]$. Poiché, ricorrendo alla f.r. F_n della v.a. X_n , tale probabilità si scrive

$$\mathbb{P}(X_n \in]a, b]) = F_n(b) - F_n(a),$$

è evidente l'interesse dello studio del limite della successione $(F_n(x))$, eventualmente con qualche restrizione sul punto x nel quale essa è calcolata.

Definizione 2.3.1. Si dice che la successione di f.r. (F_n) *converge completamente* alla f.r. F , se, per ogni punto di continuità x di F , si ha la convergenza puntuale di $(F_n(x))$ a $F(x)$. Se $C(F)$ indica l'insieme dei punti di continuità di F la convergenza completa si esprime nella forma

$$\forall x \in C(F) \quad \lim_{n \rightarrow +\infty} F_n(x) = F(x);$$

ciò che si indica con $F_n \xrightarrow{c} F$. \diamond

La dizione di convergenza completa non è adottata universalmente; piú usata è l'espressione *convergenza debole* che, però, in queste lezioni è riservata ad un altro tipo di convergenza.

Esistono modi equivalenti di porre la convergenza completa; alcuni si incontreranno in queste lezioni, altri sono dati negli esercizi. Lo spazio delle f.r. è indicato da \mathcal{D} .

Teorema 2.3.1. *Se (F_n) è una successione di f.r. di \mathcal{D} e F appartiene a \mathcal{D} , sono equivalenti le proprietà:*

(a) $F_n \xrightarrow{c} F$;

(b) *esiste un sottoinsieme denso D di \mathbb{R} tale che $\lim_{n \rightarrow +\infty} F_n(x) = F(x)$ per ogni $x \in D$.*

Dimostrazione. (a) \implies (b) Basta prendere $D = C(F)$.

(b) \implies (a) Sia F continua in x_0 , $x_0 \in C(F)$. Allora, comunque si fissi $\varepsilon > 0$, esiste $\delta = \delta(\varepsilon) > 0$ tale che

$$|F(x) - F(x_0)| < \frac{\varepsilon}{2}$$

se $|x - x_0| < \delta$. Poiché D è denso in \mathbb{R} , esistono due punti a e b di D che appartengono all'intervallo $]x_0 - \delta, x_0 + \delta[$ e tali che sia $a < x_0 < b$. Per ipotesi, esiste $n_0 \in \mathbb{N}$ tale che per ogni $n \geq n_0$, si abbia

$$|F_n(a) - F(a)| < \frac{\varepsilon}{2} \quad \text{e} \quad |F_n(b) - F(b)| < \frac{\varepsilon}{2}.$$

Allora, per ogni $n \geq n_0$, si ha

$$F_n(x_0) - F(x_0) \leq F_n(b) - F(x_0) < F(b) - F(x_0) + \frac{\varepsilon}{2} < \varepsilon$$

e

$$F_n(x_0) - F(x_0) \geq F_n(a) - F(x_0) > F(a) - F(x_0) - \frac{\varepsilon}{2} > -\varepsilon,$$

onde $|F_n(x_0) - F(x_0)| < \varepsilon$, se $n \geq n_0$. \square

Non si deve credere che il limite di una successione convergente di f.r. coincida (tranne, eventualmente, per un insieme numerabile) con una f.r. come mostrano gli esempi che seguono. Nel trattare della convergenza completa di f.r. è, perciò, indispensabile assicurarsi che il limite sia anch'esso una f.r..

Esempio 2.3.1. Sia $F_n = \varepsilon_n (= \mathbf{1}_{[n, \infty[})$ ($n \in \mathbb{N}$) la f.r. di una v.a. che assume q.c. il valore n ; allora $F_n(x) \rightarrow 0$ per ogni $x \in \mathbb{R}$ sicché la funzione identicamente nulla non è una f.r.. \blacksquare

Esempio 2.3.2. Se, per ogni $n \in \mathbb{N}$, F_n è la f.r. della legge uniforme in $(-n, n)$

$$F_n(x) := \frac{1}{2n} \int_{-\infty}^x \mathbf{1}_{(-n, n)}(t) dt = \begin{cases} 0, & x \leq -n, \\ \frac{(x+n)}{2n}, & x \in [-n, n], \\ 1, & x \geq n, \end{cases}$$

allora $F_n(x) \rightarrow 1/2$ per ogni $x \in \mathbb{R}$. \blacksquare

Esempio 2.3.3. Si consideri la successione di v.a. definita, per ogni $n \in \mathbb{N}$, da

$$X_n = \begin{cases} \frac{1}{n}, & n \text{ dispari,} \\ 0, & n \text{ pari.} \end{cases}$$

Si consideri la successione delle corrispondenti f.r. (F_n). Per ogni $x > 0$, si ha $F_n(x) = \mathbb{P}(X_n \leq x) \rightarrow 1$. Perciò $F_n \xrightarrow{c} \varepsilon_0$. Nell'origine si ha

$$F_n(0) = \mathbb{P}(X_n \leq 0) = \begin{cases} 0, & \text{se } n \text{ è dispari,} \\ 1, & \text{se } n \text{ è pari.} \end{cases}$$

In questo caso la successione ($F_n(0)$) non ammette neanche limite. \blacksquare

Il seguente è un concetto piú debole della convergenza completa quando sia applicato alle f.r..

Definizione 2.3.2. Si dice che una successione (f_n) di funzioni reali a valori reali crescenti (f_n da \mathbb{R} in \mathbb{R} , $x' < x'' \implies f_n(x') \leq f_n(x'')$) converge debolmente alla funzione (pure crescente) f se, per ogni $x \in C(f)$, è

$$\lim_{n \rightarrow +\infty} f_n(x) = f(x).$$

Tale modo di convergenza sarà detto *convergenza debole*. \diamond

Il seguente risultato è di frequente applicazione in Probabilità. Tuttavia, esso non è probabilistico in natura, sicché tutte le volte che lo si applicherà ad un insieme di f.r. occorrerà poi accertarsi che il limite di cui esso asserisce l'esistenza sia effettivamente una f.r..

Teorema 2.3.2. (Primo teorema di Helly). *Sia (f_n) una successione di funzioni crescenti ed uniformemente limitate, cioè $|f_n| \leq K$ per ogni $n \in \mathbb{N}$; esiste una successione estratta dalla data che converge debolmente ad una funzione g crescente, con $|g| \leq K$ e che può essere presa continua a destra.*

Dimostrazione. Supporremo, senza scapito di generalità, che ogni funzione f_n sia positiva e continua a destra. Sia \mathbb{Q} l'insieme dei numeri razionali che si supporrà numerato, $\mathbb{Q} = \{q_1, q_2, \dots\}$. La successione reale $(f_n(q_1))$ è limitata; esiste perciò una successione estratta $(f_{1n}(q_1))_{n \in \mathbb{N}}$ convergente ad un limite c_1 . Anche la successione di numeri reali $(f_{1n}(q_2))$ è limitata ed ha perciò una sottosuccessione che converge ad un limite c_2 ; la si indichi con $(f_{2n}(q_2))$. La successione di funzioni (f_{2n}) converge tanto in q_1 quanto in q_2 e si ha

$$\lim_{n \rightarrow +\infty} f_{2n}(q_1) = c_1 \quad \text{e} \quad \lim_{n \rightarrow +\infty} f_{2n}(q_2) = c_2.$$

Procedendo in questa maniera, si costruiscono successioni $(f_{kn})_{n \in \mathbb{N}}$ tutte estratte da quella data e con la proprietà che $\lim_{n \rightarrow +\infty} f_{kn}(q_j) = c_j$ per ogni $j \leq k$. Si consideri la successione "diagonale" (f_{nn}) e la funzione $\varphi : \mathbb{Q} \rightarrow \mathbb{R}$ definita da $\varphi(q_j) := c_j$. Si ha così $f_{nn}(q) \rightarrow \varphi(q)$ per ogni $q \in \mathbb{Q}$. Si definisca $g : \mathbb{R} \rightarrow \mathbb{R}$ mediante $g(x) := \inf\{\varphi(q) : q \in \mathbb{Q}, q > x\}$; la funzione g così definita è crescente e continua a destra. Infatti, sia x_0 un arbitrario numero reale; esiste allora, per ogni $\varepsilon > 0$, $q \in \mathbb{Q}$ con $q > x_0$ tale che $\varphi(q) - g(x_0) < \varepsilon$. Si consideri ora $x \in]x_0, q[$; dalla definizione di g segue che $g(x) \leq \varphi(q)$, sicché, per ogni $x \in]x_0, q[$, è $0 \leq g(x) - g(x_0) \leq \varphi(q) - g(x_0) < \varepsilon$, ciò che prova la continuità a destra di g . Sia $x \in \mathbb{R}$; allora per ogni $q \in \mathbb{Q}$ con $q > x$, si ha, per ogni $n \in \mathbb{N}$,

$$\limsup_{n \rightarrow +\infty} f_{nn}(x) \leq \limsup_{n \rightarrow +\infty} f_{nn}(q) = \varphi(q),$$

sicché $\limsup_{n \rightarrow +\infty} f_{nn}(x) \leq g(x)$. Analogamente, si ha, per ogni $q \in \mathbb{Q}$ con $q < x$,

$$\liminf_{n \rightarrow +\infty} f_{nn}(x) \geq \liminf_{n \rightarrow +\infty} f_{nn}(q) = \varphi(q),$$

e, dunque, se $y < x$ e $q \in \mathbb{Q}$ è tale che $q < x$,

$$\liminf_{n \rightarrow +\infty} f_{nn}(x) \geq \sup\{\varphi(q) : q \in \mathbb{Q}, q < x\} \geq g(y),$$

e $\liminf_{n \rightarrow +\infty} f_{nn}(x) \geq \ell^- g(x)$. In particolare, se g è continua in x , si ha

$$\lim_{n \rightarrow +\infty} f_{nn}(x) = g(x).$$

Poiché $0 \leq f \leq K$, è pure $0 \leq g \leq K$. \square

Quando il teorema di Helly si applica ad una successione di f.r., il limite non è necessariamente una f.r. perché potrebbero non essere soddisfatte le proprietà di limite a $-\infty$ e a $+\infty$.

Oltre a quelli incontrati in precedenza vi è un altro tipo di convergenza per una successione di v.a..

Definizione 2.3.3. Si dice che una successione di v.a. (X_n) converge in legge o in distribuzione alla v.a. X se, essendo F_n la f.r. di X_n e F la f.r. di X , la successione (F_n) converge completamente a F ; tale convergenza si indica con $X_n \xrightarrow{\mathcal{L}} X$. \diamond

Per la convergenza in legge di una successione di v.a. non occorre che tutte le v.a. della successione siano definite sopra il medesimo spazio di probabilità. La convergenza in legge non gode delle proprietà alle quali si è soliti pensare trattando di limiti; si vedrà che il limite in legge non è necessariamente unico, che il limite di una somma non sempre è eguale alla somma dei limiti e così via. Ovviamente se si considera la convergenza in legge in relazione ad altri tipi di convergenza, allora si richiederà che tutte le v.a. siano definite sullo stesso spazio di probabilità.

Teorema 2.3.3. (La convergenza in probabilità implica la convergenza in legge). Se (X_n) è una successione di v.a. sullo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ che converge in probabilità alla v.a. X , allora vi converge anche in legge,

$$X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X \implies X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X.$$

Dimostrazione. Sia $x' < x < x''$. Poiché

$$\begin{aligned} \{X \leq x'\} &= (\{X_n \leq x\} \cap \{X \leq x'\}) \cup (\{X_n > x\} \cap \{X \leq x'\}) \\ &\subseteq \{X_n \leq x\} \cup (\{X_n > x\} \cap \{X \leq x'\}), \end{aligned}$$

si ha

$$F(x') \leq F_n(x) + \mathbb{P}(X_n > x, X \leq x').$$

Poiché $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X$ e $x - x' > 0$, si ha

$$\mathbb{P}(X_n > x, X \leq x') \leq \mathbb{P}(|X_n - X| > x - x') \xrightarrow[n \rightarrow +\infty]{} 0.$$

Perciò

$$F(x') \leq \liminf_{n \rightarrow +\infty} F_n(x).$$

Scambiando i ruoli di X_n e X , si ha

$$F_n(x) \leq F(x'') + \mathbb{P}(X > x'', X_n \leq x)$$

e

$$\limsup_{n \rightarrow +\infty} F_n(x) \leq F(x'').$$

Perciò

$$F(x') \leq \liminf_{n \rightarrow +\infty} F_n(x) \leq \limsup_{n \rightarrow +\infty} F_n(x) \leq F(x''). \quad (2.3.1)$$

Si supponga ora che x sia un punto di continuità per F , $x \in C(F)$, e si facciano tendere x' e x'' a x , x' crescendo e x'' decrescendo, $x' \uparrow x$, $x'' \downarrow x$, ottenendo cosí dalla (2.3.1)

$$\lim_{n \rightarrow +\infty} F_n(x) = F(x),$$

cioè l'asserto. \square

Esempio 2.3.4. (La convergenza in legge non implica quella in probabilità). Sia X una v.a. che assume i valori 0 e 1, entrambi con probabilità eguale a $1/2$ e sia Y la v.a. definita da $Y := 1 - X$. Evidentemente, X e Y hanno la stessa f.r. $F = (\varepsilon_0 + \varepsilon_1)/2$. Si consideri la successione (X_n) con $X_n = Y$; essa converge in legge a X , ma poiché $|X_n - X| = |Y - X| = 1$, non vi converge in probabilità. \blacksquare

L'ultimo esempio stabilisce inoltre che il limite in legge non è unico perché la successione (X_n) lí considerata converge in legge sia a X sia a Y .

Teorema 2.3.4. *Per una successione di v.a. (X_n) sono equivalenti le condizioni:*

- (a) (X_n) converge in probabilità alla v.a. $X = a$ q.c.;
- (b) (X_n) converge in legge alla v.a. $X = a$ q.c..

Dimostrazione. L'implicazione (a) \implies (b) è contenuta nel Teorema 2.3.3.

(b) \implies (a) Se $X \rightarrow a$ in legge, si ha $\lim_{n \rightarrow +\infty} F_n(x) = \varepsilon_a(x)$ per ogni $x \neq a$ e quindi, per ogni $\varepsilon > 0$,

$$\begin{aligned} \mathbb{P}(X_n - a < -\varepsilon) &= \mathbb{P}(X_n < a - \varepsilon) \leq F_n(a - \varepsilon), \\ \mathbb{P}(X_n - a > \varepsilon) &= \mathbb{P}(X_n > a + \varepsilon) = 1 - F_n(a + \varepsilon), \end{aligned}$$

relazioni dalle quali scende immediatamente $\mathbb{P}(|X_n - a| > \varepsilon) \rightarrow 0$. \square

I risultati sulle relazioni tra i vari tipi di convergenza sono riassunti nella Figura 3.1, nella quale le frecce indicano le implicazioni.

È notevole e spesso utile il seguente

Teorema 2.3.5. (Skorohod). *Siano F_n ($n \in \mathbb{N}$) e F f.r. e si supponga che (F_n) converga completamente a F . Esistono allora uno spazio di probabilità $(\Omega, \mathcal{F}, \mu)$ e v.a. X_n ($n \in \mathbb{N}$) e X su $(\Omega, \mathcal{F}, \mu)$ tali che*

- (a) per ogni $n \in \mathbb{N}$, $F_{X_n} = F_n$ e $F_X = F$;
- (b) $\lim_{n \rightarrow +\infty} X_n = X$ μ -q.c..

Dimostrazione. Si ponga $\Omega =]0, 1[$, $\mathcal{F} = \mathcal{B}(]0, 1[)$ e sia μ la restrizione della misura di Lebesgue a (i boreliani di) $]0, 1[$. Per $\omega \in]0, 1[$, si ponga

$$X_n(\omega) := \inf\{x \in \mathbb{R} : \omega \leq F_n(x)\} \quad \text{e} \quad X(\omega) := \inf\{x \in \mathbb{R} : \omega \leq F(x)\}.$$

Risulta, allora, $\omega \leq F_n(x)$ se, e solo se, $X_n(\omega) \leq x$ e perciò

$$\mu(\{\omega : X_n(\omega) \leq x\}) = \mu(\{\omega : \omega \leq F_n(x)\}) = F_n(x),$$

sicché F_n è la f.r. della v.a. X_n ; similmente si dimostra che F è la f.r. di X . Fissato $\omega \in]0, 1[$ e dato $\varepsilon > 0$, si scelga un punto x di continuità per F , $x \in C(F)$ in modo che $X(\omega) - \varepsilon < x < X(\omega)$. Allora, è anche $F(x) < \omega$. Ora, $F_n(x) \rightarrow F(x)$ implica, per n sufficientemente grande, che sia $F_n(x) < \omega$, onde

$$X(\omega) - \varepsilon < x < X_n(\omega);$$

perciò $\liminf_{n \rightarrow +\infty} X_n(\omega) \geq X(\omega)$. Se $\omega' > \omega$, si scelga $x' \in C(F)$ tale che $X(\omega') < x' < X(\omega') + \varepsilon$; di qui $\omega < \omega' < F(x')$ e, perciò, per n sufficientemente grande, è $\omega < F_n(x')$, da cui scende

$$X_n(\omega) < x' < X(\omega') + \varepsilon.$$

Quindi, per $\omega' > \omega$, si ha $\limsup_{n \rightarrow +\infty} X_n(\omega) \leq X(\omega')$. Dunque, $X_n(\omega) \rightarrow X(\omega)$ se X è continua in ω . È però facile vedere che X è crescente in $]0, 1[$ sicché ha, al più, un'infinità numerabile di punti di discontinuità; inoltre, l'insieme dei punti di discontinuità di X ha misura nulla rispetto alla misura di Lebesgue. \square

La dimostrazione appena data del teorema di Skorohod usa in maniera essenziale l'ordinamento dei reali. Benché il teorema sia ancora valido in condizioni di maggiore generalità, la dimostrazione è, in quei casi, assai più complessa.

2.4 Le convergenze vaga e stretta

Definizione 2.4.1. Siano F_n e F ($n \in \mathbb{N}$) f.r. e siano \mathbb{P}_n e \mathbb{P} le misure di probabilità di Stieltjes su $(\mathbb{R}, \mathcal{B})$ generate da F_n e da F , $\mathbb{P}_n := \mu_{F_n}$ e $\mathbb{P} := \mu_F$. Se $F_n \xrightarrow[n \rightarrow +\infty]{c} F$, si dirà che \mathbb{P}_n converge completamente a \mathbb{P} ($\mathbb{P}_n \xrightarrow[n \rightarrow +\infty]{c} \mathbb{P}$). \diamond

Teorema 2.4.1. Sia $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ misurabile e si supponga che l'insieme $D(\varphi)$ dei punti di discontinuità di φ sia misurabile, $D(\varphi) \in \mathcal{B}$.

Se $\mathbb{P}(D(\varphi)) = 0$ e $\mathbb{P}_n \xrightarrow[n \rightarrow +\infty]{c} \mathbb{P}$, allora

$$\mathbb{P}_n \circ \varphi^{-1} \xrightarrow[n \rightarrow +\infty]{c} \mathbb{P} \circ \varphi^{-1}.$$

Dimostrazione. Siano $(\Omega, \mathcal{F}, \mu)$, X_n ($n \in \mathbb{N}$) e X come nel teorema di Skorohod. Si osservi che la legge della v.a. $\varphi \circ X$ è $\mathbb{P} \circ \varphi^{-1}$; infatti, per ogni boreliano A , è

$$\mu((\varphi \circ X) \in A) = \mu(X \in \varphi^{-1}(A)) = \mathbb{P}(\varphi^{-1}(A)).$$

Similmente $\varphi \circ X_n^{-1}$ ha legge $\mathbb{P}_n \circ \varphi^{-1}$.

Se $X(\omega)$ non appartiene a $D(\varphi)$, da $X_n(\omega) \rightarrow X(\omega)$ scende

$$\varphi[X_n(\omega)] \xrightarrow[n \rightarrow +\infty]{} \varphi[X(\omega)]$$

e, poiché

$$\mu(\{\omega : X(\omega) \in D(\varphi)\}) = \mathbb{P}(D(\varphi)) = 0,$$

si ha $\varphi \circ X_n \xrightarrow[n \rightarrow +\infty]{} \varphi \circ X$ μ -q.c.. Di conseguenza, si ha anche $\varphi \circ X_n \xrightarrow[n \rightarrow +\infty]{} \varphi \circ X$ in legge. Allora, la convergenza in legge di $\varphi \circ X_n$ a $\varphi \circ X$ equivale a dire che $\mathbb{P}_n \circ \varphi^{-1} \xrightarrow[n \rightarrow +\infty]{} \mathbb{P} \circ \varphi^{-1}$. \square

Il risultato che segue è essenziale per stabilire l'unicità del limite nelle convergenze che saranno introdotte nella successiva definizione 2.4.2.

Teorema 2.4.2. *Se due misure finite μ e ν su $(\mathbb{R}, \mathcal{B})$ sono tali che $\int f d\mu = \int f d\nu$ per ogni funzione continua e con il supporto compatto, $f \in C_c = C_c(\mathbb{R})$, allora esse sono eguali, $\mu = \nu$.*

Dimostrazione. Poiché la famiglia degli intervalli chiusi genera la tribù di Borel ed è stabile rispetto all'intersezione, basta far vedere che $\mu([a, b]) = \nu([a, b])$ per ogni coppia di numeri reali a, b con $a < b$. Esiste una successione $(f_n) \subseteq C_c$ tale che $f_n \rightarrow \mathbf{1}_{[a, b]}$ q.o.; una tale successione è, per esempio, data da

$$f_n(x) := \begin{cases} 0, & x \notin \left[a - \frac{1}{n}, b + \frac{1}{n} \right], \\ n \left(x - a + \frac{1}{n} \right), & x \in \left[a - \frac{1}{n}, a \right], \\ 1, & x \in [a, b], \\ n \left(b + \frac{1}{n} - x \right), & x \in \left[b, b + \frac{1}{n} \right]. \end{cases}$$

Poiché, per ogni $n \in \mathbb{N}$, $0 \leq f_n \leq f_1$ e poiché f_1 è integrabile sia rispetto a μ sia rispetto a ν , il teorema di convergenza dominata applicato due volte dà

$$\begin{aligned} \mu([a, b]) &= \int \mathbf{1}_{[a, b]} d\mu = \lim_{n \rightarrow +\infty} \int f_n d\mu \\ &= \lim_{n \rightarrow +\infty} \int f_n d\nu = \int \mathbf{1}_{[a, b]} d\nu = \nu([a, b]), \end{aligned}$$

che stabilisce l'eguaglianza delle misure μ e ν . □

Definizione 2.4.2. Siano μ_n ($n \in \mathbb{N}$) e μ misure su $(\mathbb{R}, \mathcal{B})$ finite; si dice che (μ_n) converge vagamente a μ , e si scrive $\mu_n \xrightarrow[n \rightarrow +\infty]{v} \mu$, se per ogni funzione f continua e con il supporto compatto, $f \in C_c(\mathbb{R})$, si ha

$$\lim_{n \rightarrow +\infty} \int f d\mu_n = \int f d\mu.$$

Si dice invece che (μ_n) converge strettamente a μ , e si scrive $\mu_n \xrightarrow[n \rightarrow +\infty]{s} \mu$, se, per ogni funzione f continua e limitata, $f \in C_b(\mathbb{R})$, si ha

$$\lim_{n \rightarrow +\infty} \int f d\mu_n = \int f d\mu.$$

Tali modi di convergenza si diranno *convergenza vaga* o *stretta*, rispettivamente. \diamond

Poiché $C_c \subseteq C_b$, la convergenza stretta implica la convergenza vaga. Il Teorema 2.4.2 assicura l'unicità del limite vago (e, di conseguenza, di quello stretto).

Teorema 2.4.3. *Per una successione (μ_n) di misure finite su $(\mathbb{R}, \mathcal{B})$ sono equivalenti le proprietà:*

(a) (μ_n) converge strettamente a μ ;

(b) (μ_n) converge vagamente a μ ed inoltre si ha $\lim_{n \rightarrow +\infty} \mu_n(\mathbb{R}) = \mu(\mathbb{R})$.

Dimostrazione. (a) \implies (b) Basta dimostrare il secondo asserto di (b). Poiché la funzione identicamente eguale a 1 è continua e limitata, $1 \in C_b$, è

$$\mu_n(\mathbb{R}) = \int 1 \, d\mu_n \rightarrow \int 1 \, d\mu = \mu(\mathbb{R}).$$

(b) \implies (a) Alla luce della definizione di integrale non è restrittivo dare la dimostrazione per le sole funzioni positive limitate superiormente da 1: $f \in C_b^+$, $0 \leq f \leq 1$. Si può trovare una successione $(f_r) \subseteq C_c$ tale che $f_r \uparrow f$. Per esempio, si consideri la funzione $g_r \in C_c$ definita da

$$g_r(x) := \begin{cases} 0, & x \notin [-(r+1), r+1], \\ x+r+1, & x \in [-(r+1), -r], \\ 1, & x \in [-r, r], \\ r+1-x, & x \in [r, r+1]. \end{cases}$$

La funzione $f_r := g_r \wedge f$ soddisfa al requisito di sopra. Allora, per ogni $r \in \mathbb{N}$,

$$\int f_r \, d\mu = \lim_{n \rightarrow +\infty} \int f_r \, d\mu_n = \liminf_{n \rightarrow +\infty} \int f_r \, d\mu_n \leq \liminf_{n \rightarrow +\infty} \int f \, d\mu_n,$$

onde

$$\int f \, d\mu = \sup_{r \in \mathbb{N}} \int f_r \, d\mu \leq \liminf_{n \rightarrow +\infty} \int f \, d\mu_n. \quad (2.4.1)$$

Applicando quest'ultima disuguaglianza alla funzione $1 - f$, che pure è continua ed assume valori in $[0, 1]$, si ottiene

$$\int (1 - f) \, d\mu \leq \liminf_{n \rightarrow +\infty} \int (1 - f) \, d\mu_n.$$

D'altro canto, ancora per la (2.4.1)

$$\begin{aligned} \int f \, d\mu &= \mu(\mathbb{R}) - \int (1 - f) \, d\mu = \limsup_{n \rightarrow +\infty} \mu_n(\mathbb{R}) - \int (1 - f) \, d\mu \\ &\geq \limsup_{n \rightarrow +\infty} \mu_n(\mathbb{R}) - \liminf_{n \rightarrow +\infty} \int (1 - f) \, d\mu_n \\ &\geq \limsup_{n \rightarrow +\infty} \left(\mu_n(\mathbb{R}) - \int (1 - f) \, d\mu_n \right) = \limsup_{n \rightarrow +\infty} \int f \, d\mu_n, \end{aligned}$$

che insieme alla (2.4.1) fornisce l'asserto. \square

Così, per le misure di probabilità, la convergenza vaga e quella stretta coincidono.

Se $A \in \mathcal{B}$ si dirà *frontiera* di A l'insieme $\partial A := \bar{A} - A^\circ$ (ove \bar{A} e A° sono rispettivamente la chiusura e l'interno di A). Se \mathbb{P} è una probabilità su $(\mathbb{R}, \mathcal{B})$ si dirà che $A \in \mathcal{B}$ è un *insieme di* (\mathbb{P}) -*continuità* se $\mathbb{P}(\partial A) = 0$. Il teorema che segue dà una formulazione di convergenza completa che non dipende dalle f.r..

Teorema 2.4.4. *Siano \mathbb{P}_n ($n \in \mathbb{N}$) e \mathbb{P} misure di probabilità su $(\mathbb{R}, \mathcal{B})$ e siano F_n e F le corrispondenti f.r.. Sono allora equivalenti le condizioni:*

- (a) $F_n \xrightarrow[n \rightarrow +\infty]{c} F$;
- (b) (\mathbb{P}_n) converge strettamente a \mathbb{P} ;
- (c) $\lim_{n \rightarrow +\infty} \mathbb{P}_n(A) = \mathbb{P}(A)$ per ogni insieme di \mathbb{P} -continuità.

Dimostrazione. (a) \implies (b) In virtù del teorema di Skorohod, esistono uno spazio di probabilità $(\Omega, \mathcal{F}, \mu)$ e v.a. X_n e X definite su di esso, tali che le leggi di X_n e di X siano, rispettivamente, \mathbb{P}_n e \mathbb{P} . Sia, inoltre, $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ una funzione misurabile, limitata e tale che l'insieme $D(\varphi)$ dei suoi punti di discontinuità sia misurabile e di probabilità nulla, $\mathbb{P}(D(\varphi)) = 0$. Poiché $\mu(X \in D(\varphi)) = (\mu \circ X^{-1})(D(\varphi)) = \mathbb{P}(D(\varphi)) = 0$, si ha $\varphi \circ X_n \xrightarrow[n \rightarrow +\infty]{} \varphi \circ X$ μ -q.c.. Il teorema di convergenza dominata dà

$$\int_{\mathbb{R}} \varphi d\mathbb{P}_n = \int_{\mathbb{R}} \varphi d(\mu \circ X_n^{-1}) = \int_{\Omega} \varphi \circ X_n d\mu \xrightarrow[n \rightarrow +\infty]{} \int_{\Omega} \varphi \circ X d\mu = \int_{\mathbb{R}} \varphi d\mathbb{P}.$$

Ciò mostra, in particolare, che (a) implica (b), poiché, se φ , oltre che essere limitata, è anche continua, è $D(\varphi) = \emptyset$ e dunque, *a fortiori*, $\mathbb{P}(D(\varphi)) = 0$.

(a) \implies (c) Preso $\varphi = \mathbf{1}_A$ con $A \in \mathcal{B}$ insieme di \mathbb{P} -continuità, risulta $D(\varphi) = \partial A$, sicché quanto visto sopra dà l'asserto.

(c) \implies (a) Considerato l'insieme $] -\infty, x]$, la sua frontiera è costituita dal singoletto $\{x\}$, sicché esso è di continuità per \mathbb{P} se, e solo se, $\mathbb{P}(\{x\}) = 0$, vale a dire se, e solo se, $x \in C(F)$. Perciò, se tale condizione è verificata, è

$$F_n(x) = \mathbb{P}_n(] -\infty, x]) \xrightarrow[n \rightarrow +\infty]{} \mathbb{P}(] -\infty, x]) = F(x).$$

Per completare la dimostrazione basta far vedere che (b) \implies (a). Si supponga allora che (\mathbb{P}_n) converga strettamente a \mathbb{P} . Non è difficile stabilire che l'insieme degli atomi di \mathbb{P} , cioè l'insieme $A \subseteq \mathbb{R}$ di punti con probabilità strettamente positiva, è, al più numerabile: se

$$A := \{x \in \mathbb{R} : \mathbb{P}(\{x\}) > 0\},$$

è $\text{card}(A) \leq \aleph_0$. Si prenda $D = A^c$; evidentemente, $\overline{D} = \mathbb{R}$. Siano a e b punti di D con $a < b$ e sia $f = \mathbf{1}_{]a, b]}$. Dato $\varepsilon > 0$, esiste $\delta = \delta(\varepsilon) > 0$ tale che $a + \delta < b - \delta$ e $\mathbb{P}(I) < \varepsilon$, ove

$$I :=]a - \delta, a + \delta[\cup]b - \delta, b + \delta[.$$

Si definiscano due funzioni $f_1, f_2 : \mathbb{R} \rightarrow [0, 1]$ di C_c mediante

$$f_1(x) := \begin{cases} 0, & x \in]-\infty, a] \cup [b, +\infty[, \\ \frac{x-a}{\delta}, & x \in]a, a+\delta] , \\ 1, & x \in]a+\delta, b-\delta] , \\ \frac{b-x}{\delta}, & x \in [b-\delta, b] . \end{cases}$$

e

$$f_2(x) := \begin{cases} 0, & x \in]-\infty, a - \delta] \cup [b + \delta, +\infty[, \\ \frac{x - a + \delta}{\delta}, & x \in]a - \delta, a] , \\ 1, & x \in]a, b] , \\ \frac{b + \delta - x}{\delta}, & x \in [b, b + \delta] . \end{cases}$$

Evidentemente, f_1 e f_2 sono in C_c e si ha $0 \leq f_1 \leq f \leq f_2 \leq f_1 + 1$. Allora

$$\int f_1 \, d\mathbb{P} \leq \int f \, d\mathbb{P} \leq \int f_2 \, d\mathbb{P} \quad \text{e} \quad \int f_1 \, d\mathbb{P}_n \leq \int f \, d\mathbb{P}_n \leq \int f_2 \, d\mathbb{P}_n$$

per ogni $n \in \mathbb{N}$. Poiché $\lim_{n \rightarrow +\infty} \int f_j \, d\mathbb{P}_n = \int f_j \, d\mathbb{P}$ ($j = 1, 2$) si ha, per n sufficientemente grande,

$$\left| \int f \, d\mathbb{P}_n - \int f \, d\mathbb{P} \right| \leq \int f_2 \, d\mathbb{P} - \int f_1 \, d\mathbb{P} + \varepsilon < P(I) + \varepsilon < 2\varepsilon ,$$

sicché

$$\lim_{n \rightarrow +\infty} (F_n(b) - F_n(a)) = \lim_{n \rightarrow +\infty} \int f \, d\mathbb{P}_n = \int f \, d\mathbb{P} = F(b) - F(a) . \quad (2.4.2)$$

Poiché $D = A^c = C(F)$, la (2.4.2) vale per ogni coppia a, b di punti di $C(F)$ con $a < b$. Per mostrare che $\lim_{n \rightarrow +\infty} F_n(x) = F(x)$ per ogni $x \in C(F)$ si scelgano a e b in $C(F)$ in modo da avere $a < x < b$, $F(a) < \varepsilon$ e $F(b) > 1 - \varepsilon$. Allora,

$$|F_n(x) - F(x)| \leq |F_n(x) - F_n(a) - F(x) + F(a)| + F_n(a) + F(a) ,$$

mentre dalla (2.4.2), con $b = x$, scende

$$\limsup_{n \rightarrow +\infty} |F_n(x) - F(x)| \leq F(a) + \limsup_{n \rightarrow +\infty} F_n(a) . \quad (2.4.3)$$

Ora,

$$\limsup_{n \rightarrow +\infty} F_n(a) \leq \limsup_{n \rightarrow +\infty} (1 - F_n(b) + F_n(a)) \leq 1 - F(b) + F(a) < 2\varepsilon ,$$

onde, sostituendo nella (2.4.3), si ha $\limsup_{n \rightarrow +\infty} |F_n(x) - F(x)| = 0$. \square

L'ultimo teorema consente di formulare in maniera differente, ma equivalente, la convergenza in legge di v.a.: una successione (X_n) di v.a. definite sopra il medesimo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ converge in legge alla v.a. X se, e solo se, per ogni funzione f di $C_b(\mathbb{R})$, o di $C_c(\mathbb{R})$, si ha

$$\mathbb{E}(f \circ X_n) \xrightarrow{n \rightarrow +\infty} \mathbb{E}(f \circ X) .$$

Infatti, per il teorema del cambio di variabile (3.8.2), si ha

$$\mathbb{E}(f \circ X_n) = \int_{\Omega} f \circ X_n \, d\mu = \int_{\mathbb{R}} f \, d(\mu X_n^{-1}) = \int_{\mathbb{R}} f \, d\mathbb{P}_n ,$$

dove \mathbb{P}_n è la legge di X_n . Analogamente si procede per $\mathbb{E}(f \circ X)$.

Risultati riguardanti la convergenza degli integrali si hanno anche nel caso della convergenza debole anziché completa; essi torneranno utili nel seguito.

Lemma 2.4.1. *Sia $(\varphi_n)_{n \in \mathbb{Z}_+}$ una successione di funzioni φ_n da \mathbb{R} in \mathbb{R} ($n \in \mathbb{Z}_+$) crescenti che converge debolmente a φ . Si ponga, per $a \in]0, +\infty[$,*

$$\begin{aligned}\delta\varphi &:= \varphi(+\infty) - \varphi(-\infty), \\ \delta\varphi_n &:= \varphi_n(+\infty) - \varphi_n(-\infty), \\ \delta\varphi_n(a) &:= \varphi_n(a) - \varphi_n(-a).\end{aligned}$$

Allora

$$\delta\varphi \leq \liminf_{n \rightarrow +\infty} \delta\varphi_n. \quad (2.4.4)$$

Inoltre, se $\delta\varphi_n(a)$ converge uniformemente a $\delta\varphi_n$ al tendere di a a $+\infty$, vale a dire se

$$\lim_{a \rightarrow +\infty} \sup_{n \in \mathbb{N}} (\delta\varphi_n - \delta\varphi_n(a)) = 0,$$

si ha

$$\lim_{n \rightarrow +\infty} \varphi_n(\pm\infty) = \varphi(\pm\infty).$$

Dimostrazione. Nelle disequaglianze $\varphi_n(-\infty) \leq \varphi_n(x) \leq \varphi_n(+\infty)$ si prenda x in $C(\varphi)$ e si passi al limite per $n \rightarrow \infty$ per ottenere

$$\limsup_{n \rightarrow +\infty} \varphi_n(-\infty) \leq \varphi(x) \leq \limsup_{n \rightarrow +\infty} \varphi_n(+\infty)$$

e

$$\liminf_{n \rightarrow +\infty} \varphi_n(-\infty) \leq \varphi(x) \leq \liminf_{n \rightarrow +\infty} \varphi_n(+\infty)$$

che insieme danno la (2.4.4).

Fissato, arbitrariamente, $\varepsilon > 0$, si scelga $\alpha = \alpha(\varepsilon) > 0$ in modo che

$$\delta\varphi_n - \delta\varphi_n(a) < \varepsilon$$

per ogni $n \in \mathbb{N}$ se $a \geq \alpha$. Allora, se tanto a quanto $-a$ sono punti di continuità per φ , è

$$\limsup_{n \rightarrow +\infty} \delta\varphi_n \leq \delta\varphi(a) + \varepsilon < +\infty.$$

In virtù della (2.4.4), ciò implica $\delta\varphi < +\infty$; poiché ε è arbitrario, risulta

$$\limsup_{n \rightarrow +\infty} \delta\varphi_n \leq \delta\varphi.$$

Questa disequaglianza e la (2.4.4) danno

$$\lim_{n \rightarrow +\infty} \delta\varphi_n = \delta\varphi.$$

Ora, le disequaglianze di sopra danno

$$\begin{aligned}\limsup_{n \rightarrow +\infty} \varphi_n(+\infty) &= \limsup_{n \rightarrow +\infty} (\delta\varphi_n + \varphi_n(-\infty)) \\ &\leq \delta\varphi + \limsup_{n \rightarrow +\infty} \varphi_n(-\infty) \leq \varphi(+\infty) \leq \liminf_{n \rightarrow +\infty} \varphi_n(+\infty)\end{aligned}$$

cioè $\lim_{n \rightarrow +\infty} \varphi_n(+\infty) = \varphi(+\infty)$. Di conseguenza si ha anche

$$\lim_{n \rightarrow +\infty} \varphi_n(-\infty) = \varphi(-\infty),$$

con il che la dimostrazione è conclusa. \square

Teorema 2.4.5. (Secondo teorema di Helly). *Sia (F_n) una successione uniformemente limitata di funzioni crescenti che converge debolmente a F . Se a e b sono punti di continuità per F ($a, b \in C(F)$), per ogni funzione continua $g \in C([a, b])$ è*

$$\lim_{n \rightarrow +\infty} \int_a^b g \, dF_n = \int_a^b g \, dF.$$

Dimostrazione. Dall'uniforme continuità di g in $[a, b]$ scende che, per ogni $\varepsilon > 0$, esistono punti

$$x_0, x_1, \dots, x_m$$

con $a_0 = x_0 < x_1 < \dots < x_m = b$ tali che

$$|g(x) - g(x_j)| < \varepsilon$$

per ogni $x \in [x_j, x_{j+1}]$ ($j = 0, 1, \dots, m-1$). Si definisca

$$g_\varepsilon := g(x_0) \mathbf{1}_{[x_0, x_1]} + \sum_{j=1}^{m-1} g(x_j) \mathbf{1}_{[x_j, x_{j+1}]}.$$

Non è, inoltre, restrittivo supporre che x_0, x_1, \dots, x_m siano punti di continuità per F . Per ogni $x \in [a, b]$, si ha $|g(x) - g_\varepsilon(x)| < \varepsilon$. Si consideri la norma della convergenza uniforme $\|g\| := \max\{|g(x)| : x \in [a, b]\}$; si può, allora, determinare $n_0 = n_0(\varepsilon) \in \mathbb{N}$ in modo che per ogni $n \geq n_0$ sia

$$|F_n(x_j) - F(x_j)| \leq \frac{\varepsilon}{\|g\|m} \quad (j = 0, 1, \dots, m).$$

Allora,

$$\begin{aligned} \left| \int_a^b g \, dF - \int_a^b g \, dF_n \right| &\leq \left| \int_a^b g \, dF - \int_a^b g_\varepsilon \, dF \right| \\ &\quad + \left| \int_a^b g_\varepsilon \, dF - \int_a^b g_\varepsilon \, dF_n \right| + \left| \int_a^b g_\varepsilon \, dF_n - \int_a^b g \, dF_n \right|. \end{aligned}$$

Si supponga che sia $|F| \leq K$ e $|F_n| \leq K$ per ogni $n \in \mathbb{N}$. Allora,

$$\left| \int_a^b g \, dF - \int_a^b g_\varepsilon \, dF \right| \leq \varepsilon \mu_F([a, b]) \leq 2\varepsilon K,$$

mentre

$$\left| \int_a^b g_\varepsilon \, dF_n - \int_a^b g \, dF_n \right| \leq \varepsilon \mu_{F_n}([a, b]) \leq 2\varepsilon K.$$

Inoltre, per $n \geq n_0$, si ha

$$\begin{aligned}
& \left| \int_a^b g_\varepsilon dF - \int_a^b g_\varepsilon dF_n \right| \\
&= \left| \sum_{j=0}^{m-1} g(x_j) (F(x_{j+1}) - F(x_j)) - \sum_{j=0}^{m-1} g(x_j) (F_n(x_{j+1}) - F_n(x_j)) \right| \\
&= \left| \sum_{j=0}^{m-1} g(x_j) (F(x_{j+1}) - F_n(x_{j+1})) - \sum_{j=0}^{m-1} g(x_j) (F(x_j) - F_n(x_j)) \right| \\
&\leq 2\|g\| m \frac{\varepsilon}{\|g\| m} = 2\varepsilon,
\end{aligned}$$

sicché, per $n \geq n_0$, si ha

$$\left| \int_a^b g dF - \int_a^b g dF_n \right| \leq (4K + 2)\varepsilon,$$

che dà l'asserto. \square

Teorema 2.4.6. *Se (F_n) è una successione uniformemente limitata di funzioni crescenti che converge debolmente a F e se $g \in C(\mathbb{R})$ è una funzione continua tale che $\lim_{|x| \rightarrow +\infty} g(x) = 0$, allora*

$$\lim_{n \rightarrow +\infty} \int_{\mathbb{R}} g dF_n = \int_{\mathbb{R}} g dF.$$

Dimostrazione. È ovvio che anche la funzione F è crescente e limitata. Si può supporre che sia $g \geq 0$ (altrimenti basta dimostrare il teorema separatamente per g^+ e per g^-). Allora se $a \in C(F)$ e $x \in \overline{\mathbb{R}}$, si definisca

$$\varphi_n(x) := \int_a^x g dF_n \quad \text{e} \quad \varphi(x) := \int_a^x g dF.$$

Per il Teorema 2.4.5, $\varphi_n(x)$ converge a $\varphi(x)$ per ogni $x \in C(F)$. Si scelga, ora, $a > 0$ sufficientemente grande perché sia $g(y) < \varepsilon$, per ogni y con $|y| > a$. Allora,

$$\begin{aligned}
\delta\varphi_n - \delta\varphi_n(y) &= \int_a^{+\infty} g dF_n - \int_a^{-\infty} g dF_n - \int_a^y g dF_n + \int_a^{-y} g dF_n \\
&= \int_y^{+\infty} g dF_n + \int_{-\infty}^{-y} g dF_n,
\end{aligned}$$

sicché, per ogni $n \in \mathbb{N}$, risulta

$$\delta\varphi_n - \delta\varphi_n(y) \leq 2\varepsilon K \quad (\text{se } |F_n| \leq K),$$

cioè

$$\limsup_{n \rightarrow +\infty} (\delta\varphi_n - \delta\varphi_n(y)) = 0.$$

L'asserto segue ora dal Lemma 2.4.1. \square

2.5 Metriche

Esaminiamo ora il problema dell'esistenza di una metrica per la convergenza completa. Una tale metrica è quella di Lévy.

Si indichi con \mathcal{D} l'insieme delle f.r., detto anche spazio delle f.r.; si noti che \mathcal{D} non è uno spazio vettoriale; per vederlo basta osservare, per esempio, che la somma di due f.r. non è una f.r.. A \mathcal{D} si può dare la struttura di spazio metrico introducendo la *metrica di Lévy*. Si denoterà con $(F, G; h)$ ove $h > 0$ la condizione

$$\forall x \in \mathbb{R} \quad F(x - h) - h \leq G(x) \leq F(x + h) + h.$$

Teorema 2.5.1. *Sia $d_L : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$ definita da*

$$d_L(F, G) := \inf\{h > 0 : \text{valgono sia } (F, G; h) \text{ sia } (G, F; h)\}.$$

Allora (\mathcal{D}, d_L) è uno spazio metrico e d_L una metrica detta metrica di Lévy.

Si osservi che per ogni coppia $F, G \in \mathcal{D}$ è $d_L(F, G) \leq 1$; infatti, per ogni $x \in \mathbb{R}$ si ha $F(x - 1) - 1 \leq 0 \leq G(x) \leq 1 \leq F(x + 1) + 1$. Alla dimostrazione è opportuno premettere il seguente

Lemma 2.5.1. *Se $d_L(F, G) = \alpha > 0$ valgono sia $(F, G; \alpha)$ sia $(G, F; \alpha)$.*

Dimostrazione. Sia $t > 0$ e si scelga y in modo che $x < y$; scende dalla definizione di d_L che

$$F(y - \alpha - t) - \alpha - t \leq G(y) \leq F(y + \alpha + t) + \alpha + t. \quad (2.5.1)$$

Nel limite $t \downarrow 0$, la (2.5.1) dà, per la continuità a destra di F ,

$$\ell^- F(y - \alpha) - \alpha \leq G(y) \leq F(y + \alpha) + \alpha,$$

per ogni $y \in \mathbb{R}$, onde, per $y \downarrow x$, si ha $F(x - \alpha) - \alpha \leq G(x) \leq F(x + \alpha) + \alpha$ cioè la $(F, G; \alpha)$. Analogamente si procede per la $(G, F; \alpha)$. \square

Dimostrazione del Teorema 2.5.1. Le condizioni $d_L(F, F) = 0$ per ogni $F \in \mathcal{D}$ e $d_L(F, G) = d_L(G, F)$ per ogni coppia $F, G \in \mathcal{D}$ sono di banale dimostrazione. Si supponga ora che $d_L(F, G) = 0$. Segue dalla definizione di d_L che $\ell^- F(x) \leq G(x)$ e $\ell^- G(x) \leq F(x)$ per ogni $x \in \mathbb{R}$, sicché F e G coincidono tranne, al più, nell'insieme al più numerabile dei loro punti di discontinuità. Per la continuità a destra delle f.r. si ha $F = G$.

Rimane da dimostrare la diseuguaglianza triangolare

$$d_L(F, H) \leq d_L(F, G) + d_L(G, H)$$

per ogni scelta delle f.r. F, G, H . Posto $\alpha = d_L(F, G)$ e $\beta = d_L(G, H)$, non vi è alcunché da dimostrare se $d_L(F, H) = 0$ oppure se $\alpha + \beta \geq 1$. Si può, perciò, supporre che $d_L(F, H) > 0$, che $\alpha + \beta < 1$, e che $\alpha > 0, \beta > 0$. In virtù del Lemma 2.5.1 si ha, per ogni $x \in \mathbb{R}$,

$$\begin{aligned} F(x - \alpha - \beta) - \alpha - \beta &\leq G(x - \beta) - \beta \\ &\leq H(x) \leq G(x + \beta) + \beta \leq F(x + \alpha + \beta) + \alpha + \beta \end{aligned}$$

sicché vale $(F, H; \alpha + \beta)$.

Similmente, si deduce che vale $(H, F; \alpha + \beta)$; perciò, è $d_L(F, H) \leq \alpha + \beta$. \square

Per il calcolo della metrica di Lévy è spesso utile la seguente osservazione. Se si completa, come in Figura 5.1, il grafico delle f.r. F e G tracciando segmenti verticali che congiungano i limiti a sinistra e a destra negli eventuali punti di discontinuità, si considerino i segmenti $P_c Q_c$ intercettati tra i due grafici, completati come appena detto, dalle rette parallele alla bisettrice del secondo e quarto quadrante $x + y = c$. Allora

$$d_L(F, G) = \sup \left\{ \frac{\overline{P_c Q_c}}{\sqrt{2}} : c \in \mathbb{R} \right\}.$$

La metrica di Lévy metrizza la convergenza completa.

Teorema 2.5.2. *Per una successione (F_n) di f.r. di \mathcal{D} sono equivalenti le affermazioni:*

- (a) (F_n) converge completamente a $F \in \mathcal{D}$;
- (b) $\lim_{n \rightarrow +\infty} d_L(F_n, F) = 0$.

Dimostrazione. (a) \implies (b) Fissato in maniera arbitraria $\varepsilon > 0$, si scelgano due punti a e b di continuità per F in modo tale che sia

$$F(a) < \frac{\varepsilon}{2} \quad \text{e} \quad F(b) < 1 - \frac{\varepsilon}{2}. \quad (2.5.2)$$

Per $i = 1, \dots, m$, si scelgano m punti di continuità per F nell'intervallo $]a, b[$, in modo che sia $a = a_0 < a_1 < \dots < a_m < a_{m+1} = b$ e $a_{j+1} - a_j < \varepsilon$ ($j = 0, 1, \dots, m$). Esiste allora un naturale $N = N(\varepsilon)$ tale che sia

$$|F_n(a_j) - F(a_j)| \leq \frac{\varepsilon}{2} \quad (2.5.3)$$

per ogni $n \geq N$ e per ogni $j = 0, 1, \dots, m + 1$. Per dimostrare che $d_L(F_n, F)$ tende a 0, basta far vedere che, per ogni $x \in \mathbb{R}$, valgono definitivamente le disequaglianze

$$F(x - \varepsilon) - \varepsilon \leq F_n(x) \leq F(x + \varepsilon) + \varepsilon. \quad (2.5.4)$$

Si distinguano tre casi:

(i) $x \in [a_{j-1}, a_j]$. Si prenda $n \geq N$ e si usi la (2.5.3) per ottenere

$$F_n(x) \leq F_n(a_j) \leq F(a_j) + \frac{\varepsilon}{2} \leq F(x + \varepsilon) + \frac{\varepsilon}{2} \leq F(x + \varepsilon) + \varepsilon,$$

e

$$F_n(x) \geq F_n(a_{j-1}) \geq F(a_{j-1}) - \frac{\varepsilon}{2} \geq F(x - \varepsilon) - \varepsilon,$$

sicché la (2.5.4) è, in questo caso, provata.

(ii) $x < a$. La prima delle (2.5.2) e la (2.5.3) danno, per $n \geq N$,

$$F_n(x) \leq F_n(a) \leq F(a) + \frac{\varepsilon}{2} \leq \varepsilon \leq F(x + \varepsilon) + \varepsilon,$$

e

$$F_n(x) \geq 0 \geq F(a) - \frac{\varepsilon}{2} \geq F(x - \varepsilon) - \varepsilon.$$

(iii) $x > b$. La seconda delle (2.5.2) e la (2.5.3) danno, per $n \geq N$,

$$F_n(x) \leq 1 \leq F(b) + \frac{\varepsilon}{2} \leq F(x + \varepsilon) + \varepsilon,$$

e

$$F_n(x) \geq F_n(b) \geq F(b) - \frac{\varepsilon}{2} \geq 1 - \varepsilon \geq F(x - \varepsilon) - \varepsilon.$$

(b) \implies (a) Sia x_0 un punto di continuità per F e si fissi $\varepsilon > 0$ in maniera arbitraria. Esiste allora $\delta = \delta(\varepsilon) > 0$ tale che $|x_0 - x| < \delta$ implichi

$$|F(x) - F(x_0)| < \varepsilon. \quad (2.5.5)$$

Posto $\eta < \min\{\varepsilon, \delta\}$, si scelga $N = N(\eta) = N(\varepsilon, \delta) \in \mathbb{N}$ in modo che si abbia $d_L(F_n, F) < \eta$ per ogni $n \geq N$. Scende allora dalla definizione di distanza di Lévy e dalla (2.5.5) che

$$F_n(x_0) \geq F(x_0 - \eta) - \eta \geq F(x_0) - 2\varepsilon,$$

e che

$$F_n(x_0) \leq F(x_0 + \eta) + \eta \leq F(x_0) + 2\varepsilon,$$

onde $|F_n(x_0) - F(x_0)| \leq 2\varepsilon$ se $n \geq N$. \square

Teorema 2.5.3. *Lo spazio metrico (\mathcal{D}, d_L) è completo.*

Dimostrazione. Sia $(F_n) \subseteq \mathcal{D}$ una successione di Cauchy; per il teorema di Helly si può estrarre dalla data una successione $(F_{n(j)})_{j \in \mathbb{N}}$ che converge debolmente alla funzione $G : \mathbb{R} \rightarrow [0, 1]$, crescente e continua a destra. Si supponga che, fissato ad arbitrio $\varepsilon > 0$, risulti $d_L(F_n, F_m) < \varepsilon$ se $m, n \geq N = N(\varepsilon)$. Si scelga ora x_m in modo che $F_m(x_m) < \varepsilon$ e si scelga $n(j)$ maggiore di N ; perciò $d_L(F_{n(j)}, F_m) < \varepsilon$. Se G è continua nel punto x e se $x < x_m - \varepsilon$, allora è

$$F_{n(j)}(x) \leq F_{n(j)}(x_m - \varepsilon) \leq F_m(x_m) + \varepsilon \leq 2\varepsilon,$$

onde $G(x) = \lim_{j \rightarrow +\infty} F_{n(j)}(x) \leq 2\varepsilon$; perciò $\lim_{x \rightarrow -\infty} G(x) = 0$. Si dimostra in maniera analoga che $\lim_{x \rightarrow +\infty} G(x) = 1$, sicché G è effettivamente una f.r. e la successione $(F_{n(j)})_{j \in \mathbb{N}}$ converge completamente a G , cioè $d_L(F_{n(j)}, G) \rightarrow 0$. Per dimostrare l'asserto, basta mostrare che tutta la successione (F_n) converge completamente a G ; a tal fine, si usi la disuguaglianza triangolare

$$d_L(F_n, G) \leq d_L(F_n, F_{n(j)}) + d_L(F_{n(j)}, G),$$

cioè l'asserto. \square

Vale la seguente importante disuguaglianza

Teorema 2.5.4. *Siano X e Y due v.a. definite sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e siano F e G , rispettivamente, le loro f.r.. Per le metriche di Lévy e di Ky Fan vale la seguente disuguaglianza*

$$d_L(F, G) \leq d_{KF}(X, Y). \quad (2.5.6)$$

Dimostrazione. Sia $\varepsilon > d_{KF}(X, Y)$; posto $A_\varepsilon := \{|X - Y| \leq \varepsilon\}$, si ha

$$\mathbb{P}(A_\varepsilon) > 1 - \varepsilon.$$

Pertanto

$$\begin{aligned} \{X \leq x\} \cap A_\varepsilon &\subseteq \{Y < x + \varepsilon\} \cap A_\varepsilon, \\ \{Y \leq y\} \cap A_\varepsilon &\subseteq \{X < y + \varepsilon\} \cap A_\varepsilon. \end{aligned}$$

Passando alle probabilità, dalla prima di queste inclusioni si ha

$$\mathbb{P}(\{X \leq x\} \cap A_\varepsilon) \leq \mathbb{P}(\{Y < x + \varepsilon\} \cap A_\varepsilon),$$

vale a dire

$$\begin{aligned} F(x) + \mathbb{P}(A_\varepsilon) - \mathbb{P}(\{X \leq x\} \cup A_\varepsilon) \\ \leq G(x + \varepsilon) + \mathbb{P}(A_\varepsilon) - \mathbb{P}(\{Y < x + \varepsilon\} \cup A_\varepsilon), \end{aligned}$$

onde

$$\begin{aligned} F(x) - G(x + \varepsilon) &\leq \mathbb{P}(\{X \leq x\} \cup A_\varepsilon) - \mathbb{P}(\{Y < x + \varepsilon\} \cup A_\varepsilon) \\ &\leq 1 - \mathbb{P}(A_\varepsilon) \leq 1 - (1 - \varepsilon) = \varepsilon, \end{aligned}$$

sicch 

$$F(x) \leq G(x + \varepsilon) + \varepsilon.$$

Similmente, dalla seconda inclusione, ponendo $y = x - \varepsilon$, si ottiene

$$\mathbb{P}(\{Y \leq x - \varepsilon\} \cap A_\varepsilon) \leq \mathbb{P}(\{X < x\} \cap A_\varepsilon),$$

vale a dire

$$G(x - \varepsilon) + \mathbb{P}(A_\varepsilon) - \mathbb{P}(\{Y \leq x - \varepsilon\} \cup A_\varepsilon) \leq F(x) + \mathbb{P}(A_\varepsilon) - \mathbb{P}(\{X < x\} \cup A_\varepsilon),$$

onde, come sopra,

$$\begin{aligned} G(x - \varepsilon) - F(x) &\leq \mathbb{P}(\{Y \leq x - \varepsilon\} \cup A_\varepsilon) - \mathbb{P}(\{X < x\} \cup A_\varepsilon) \\ &\leq 1 - \mathbb{P}(A_\varepsilon) \leq 1 - (1 - \varepsilon) = \varepsilon, \end{aligned}$$

e

$$G(x - \varepsilon) - \varepsilon \leq F(x);$$

le due disequaglianze ottenute bastano per concludere che $d_L(F, G) \leq \varepsilon$. Si faccia ora tendere ε a $d_{KF}(X, Y)$ decrescendo per avere l'asserto. \square

Basta applicare la (2.5.6) per ottenere per altra via il Teorema 2.3.3.

2.6 Altri tipi di convergenza per v.a.

I tipi di convergenza per v.a. considerati sin qui non esauriscono quelli introdotti, studiati ed usati nella letteratura scientifica; qui di seguito, accenneremo brevemente ad altri tipi di convergenza per v.a..

Definizione 2.6.1. Si dice che una successione (X_n) di v.a. definite nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ converge completamente alla v.a. X se

$$\lim_{n \rightarrow +\infty} \sum_{k=n}^{\infty} \mathbb{P}(|X_k - X| > \varepsilon) = 0,$$

per ogni $\varepsilon > 0$. ◇

La convergenza completa di v.a. è un concetto differente dalla convergenza completa di f.r. che si è studiata nella sezione 3.

La dimostrazione della seguente caratterizzazione della convergenza completa è rinviata agli esercizi.

Teorema 2.6.1. Sia (X_n) una successione di v.a. definite sullo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$. Sono allora equivalenti le seguenti affermazioni:

- (a) la successione (X_n) converge completamente a 0;
- (b) ogni successione (Y_n) di v.a., definite su un qualsiasi spazio di probabilità $(\Omega', \mathcal{A}, \mu)$ e tali che, per ogni $n \in \mathbb{N}$, Y_n abbia la stessa legge di X_n , converge q.c. rispetto alla misura di probabilità μ .

Definizione 2.6.2. Si dice che la successione (X_n) di v.a. definite sullo spazio $(\Omega, \mathcal{F}, \mathbb{P})$ converge quasi certamente uniformemente alla v.a. X se esiste un insieme $N \in \mathcal{F}$ con $\mathbb{P}(N) = 0$ tale che, per ogni $\omega \in N^c$, sia

$$\lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega) \quad \text{uniformemente in } \omega.$$

Si dice che (X_n) converge quasi uniformemente alla v.a. X se, per ogni $\varepsilon > 0$, esiste $N_\varepsilon \in \mathcal{F}$ con $\mathbb{P}(N_\varepsilon) < \varepsilon$ tale che (X_n) converga a X uniformemente in N_ε^c . ◇

Il seguente risultato è classico, dimostra che la convergenza quasi uniforme non è un nuovo concetto e fornisce una nuova caratterizzazione della convergenza quasi certa. La dimostrazione fa uso del teorema di Egorov non dimostrato in queste lezioni.

Teorema 2.6.2. In uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ la convergenza quasi certa e quella quasi uniforme sono equivalenti.

Diamo il teorema che chiarisce i rapporti tra i tipi di convergenza considerati; la dimostrazione è lasciata al lettore come esercizio.

Teorema 2.6.3. Sia (X_n) una successione di v.a. definite sullo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$. Si considerino le affermazioni:

- (a) (X_n) converge q.c. uniformemente;

- (b) (X_n) converge completamente;
- (c) (X_n) converge quasi certamente, o, ciò che è lo stesso, quasi uniformemente.

Allora, valgono le implicazioni $(a) \implies (b) \implies (c)$, mentre non vale alcuna delle implicazioni inverse.

2.7 Note al Capitolo 2

Abbiamo avuto modo di usare tacitamente il seguente risultato, per il quale si veda, ad esempio, (Dixmier, 1981).

Teorema 2.7.1. *Sia (Ω, d) uno spazio metrico e sia (x_n) una successione di elementi di Ω . Se da ogni successione $(x_{n(k)})$ estratta da (x_n) se ne può estrarre un'altra $(x_{k(j)})$ convergente a x , allora tutta la successione (x_n) converge a x .*

Dimostrazione. Nelle condizioni enunciate si supponga, per assurdo, che (x_n) non converga a x . Ciò vuol dire che, per ogni $k \in \mathbb{N}$, esiste un elemento $x_{n(k)}$ della successione data tale che $d(x_{n(k)}, x) > 1/k$. Ma allora nessuna estratta da $(x_{n(k)})$ potrebbe convergere a x . \square

Sezione 2.1 Esistono numerose generalizzazioni dei lemmi di Borel–Cantelli. Si veda, per esempio, (Petrov, 2004) e la bibliografia lì citata.

Due esempi notevoli di leggi 0–1 di uso corrente nel calcolo delle probabilità sono quelle di Kolmogorov che si incontrerà più avanti (si veda la Sezione 6.7.2) e di Hewitt & Savage che si può leggere in molti libri di Probabilità, per esempio in (Bauer, 1996).

Sezione 2.2 Le relazioni tra i vari tipi di convergenza dati in questa sezione valgono in uno spazio di probabilità, oppure, più in generale in uno spazio di misura finita; le relazioni cambiano se si considera uno spazio di misura non necessariamente finita. Per questo si veda un buon libro di teoria della misura, per esempio (Kingman & Taylor, 1966).

Nel seguito quando si parlerà dell'integrabilità uniforme si vedrà che il Teorema 2.2.8 è un caso particolare di un risultato più generale.

La convergenza in misura fu introdotta da F. Riesz (1909).

Sezione 2.3 Il Teorema di Helly 2.3.2 fu introdotto da Helly (1921). Esso ha un'estensione al caso multidimensionale; se si dice *crescente* una funzione $F : \mathbb{R}^d \rightarrow [0, 1]$ che soddisfa alla disuguaglianza (d) (4.8.6), il teorema può essere generalizzato nella seguente maniera.

Teorema 2.7.2. *Da ogni successione (F_n) di f.r. d -dimensionali se ne può estrarre un'altra*

$$(F_{n(k)})_{k \in \mathbb{N}}$$

che converge debolmente ad una funzione $F : \mathbb{R}^d \rightarrow [0, 1]$ crescente. La funzione limite non è necessariamente una f.r..

Per la dimostrazione, oltreché per la condizione necessaria e sufficiente affinché la funzione limite F sia una f.r., si veda in (Feller, 1971) il Teorema 6.1 del Capitolo VIII.

La dimostrazione del teorema di Skorokhod si può trovare, nel caso \mathbb{R}^n , in (Billingsley, 1979), in generale in (Skorokhod, 1965), in (Rao, 1987) o in (Rogers & Williams, 1994); (Letta & Pratelli, 1997) ne hanno dato una dimostrazione generale riconducendosi al caso di \mathbb{R} .

Sezione 2.4 Le convergenze stretta e vaga furono introdotte da Alexandrov in tre lunghi articoli del tempo di guerra, (Alexandrov, 1940, 1941 e 1943). La forma definitiva è dovuta a Prokhorov (1956). In un ambito piú generale, vale a dire per la convergenza stretta negli spazî metrici, i due riferimenti tradizionali sono (Billingsley, 1968) e (Parthasarathy, 1967). Un approccio differente fu introdotto da Dudley (1966 e 1967); su questo è basato il libro (Pollard, 1984). Un approccio ancora differente, particolarmente utile nelle applicazioni statistiche dovuto a Hoffman–Jorgensen è adottato in (van der Vaart & Wellner, 1996).

Sezione 2.5 Ky Fan (1944) introdusse la metrica che ora porta il suo nome.

La convergenza quasi certa è, in generale, incompatibile con l'esistenza di una norma, si veda (Dugué, 1955); è compatibile se, e solo se, Ω è finito, si vedano (Marczewski, 1955), (Thomasian, 1956 e 1957). Si consultino questi ultimi riferimenti anche per il Teorema 2.2.11.

Per l'esistenza di numerose metriche su L^0 che metrizzano la convergenza in probabilità si consulti (Lukacs, 1975).

La metrica di Lévy fu introdotta da Paul Lévy nell'appendice al libro di Fréchet del 1936, ma si veda anche (Lévy, 1937).

Se si considerano le f.r. di Δ anziché quelle di \mathcal{D} , la metrica di Lévy non è piú adeguata a metrizzare la topologia della convergenza debole, che è quella rilevante in Δ ; si dice che una successione (F_n) di f.r. di δ converge debolmente alla f.r. F se per ogni $x \in \mathbb{R}$ che sia punto di continuità per F , $x \in C(F)$, si ha

$$\lim_{n \rightarrow +\infty} F_n(x) = F(x).$$

La metrica su Δ per la convergenza debole è una modifica di quella di Lévy e fu introdotta da Sibley (1971), si veda anche (Schweizer & Sklar, 1983). Altre metriche topologicamente equivalenti a quella di Sibley si possono trovare in (Sempi, 1982), (Taylor, 1985), (Sempi, 1986), (Pascali & Sempi, 1997).

Nei teoremi di questa sezione abbiamo parlato di metriche che metrizzano certi tipi di convergenza. Ricordiamo però che in uno spazio che soddisfaccia al primo assioma di numerabilità, vale a dire tale che ogni punto abbia una base di intorni numerabile, conoscere la convergenza delle successioni equivale a conoscerne la topologia (si veda, ad esempio, (Kelley, 1955)); perciò, avremmo potuto parlare egualmente bene della topologia della convergenza in probabilità in L^0 e della topologia della convergenza completa in \mathcal{D} o in Δ .

Sezione 2.6 Per questa sezione ho seguito il libro (Lukacs, 1975). La convergenza completa per le v.a. fu introdotta da Hsu & Robbins (1947). Per il teorema di Egorov si veda, per esempio, (Dunford & Schwartz, 1958).

2.8 Esercizi sul Capitolo 2

2.1. Sia X una v.a. distribuita uniformemente su $(0, 1)$. Se

$$A_n := \{X < 1/n\}$$

risulta $\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) = +\infty$ e $\mathbb{P}(\limsup_{n \rightarrow +\infty} A_n) = 0$ senza che ciò contraddica al secondo lemma di Borel–Cantelli.

2.2. La condizione che sia convergente la serie $\sum_{n \in \mathbb{N}} \mathbb{P}(A_n)$ è solo sufficiente, ma non necessaria affinché si abbia $\mathbb{P}(\limsup_{n \rightarrow +\infty} A_n) = 0$. Si costruiscano uno spazio di probabilità e, in questo, una successione di insiemi misurabili (A_n) tale che $\mathbb{P}(\limsup_{n \rightarrow +\infty} A_n) = 0$ e $\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) = +\infty$.

2.3. Si lancia una moneta equilibrata ($p = 1/2$) un numero infinito di volte. Qual è la probabilità di ottenere infinite volte due teste consecutivamente.

2.4. Sia (A_n) una successione di eventi indipendenti con

$$\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) = +\infty,$$

sia $\mathbf{1}_{A_n}$ la funzione indicatrice di A_n e si ponga $S_n := \sum_{j=1}^n \mathbf{1}_{A_j}$. Si mostri il seguente risultato, più forte del primo lemma di Bore–Cantelli:

$$\lim_{n \rightarrow +\infty} \frac{S_n}{\mathbb{E}(S_n)} = 1 \quad \text{q.c.}$$

(Suggerimento: si scelga una sottosuccessione $(n(j))$ tale che $\mathbb{E}(S_{n(j)}) \simeq j^2$ e si mostri prima il risultato per questa sottosuccessione, estendendolo poi all'intera successione).

2.5. Il secondo lemma di Borel–Cantelli vale se l'ipotesi che gli eventi della successione (A_n) siano indipendenti è sostituita da quella che siano *a due a due* indipendenti.

2.6. (a) Si mostri che, se in uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, la successione di eventi (A_n) soddisfa alle due condizioni

$$\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) = +\infty$$

$$\liminf_{n \rightarrow +\infty} \frac{\sum_{j,k=1}^n \mathbb{P}(A_j \cap A_k)}{\left\{ \sum_{j=1}^n \mathbb{P}(A_j) \right\}^2} = 1,$$

allora $\mathbb{P}(\limsup_{n \rightarrow +\infty} A_n) = 1$.

(b) Si deduca da (a) la conclusione dell'esercizio 2.6.

2.7. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia (X_n) una successione di v.a. indipendenti ed isonome, tutte di legge esponenziale di parametro 1, $X_n \sim \Gamma(1, 1)$. Si ponga, per $\alpha > 0$,

$$A_n := \{X_n > \alpha \ln n\} \quad \text{e} \quad U := \limsup_{n \rightarrow +\infty} \frac{X_n}{\ln n}.$$

(a) Si calcoli la probabilità

$$\mathbb{P} \left(\limsup_{n \rightarrow +\infty} A_n \right);$$

(b) si mostri che $\mathbb{P}(U = 1) = 1$.

2.8. (a) Per una successione (X_n) di v.a. nello spazio $(\Omega, \mathcal{F}, \mathbb{P})$ si definisca

$$X_n \xrightarrow[n \rightarrow +\infty]{} +\infty \quad \text{q.c.};$$

(b) si mostri che $X_n \rightarrow +\infty$ q.c. se, e solo se, per ogni $M > 0$, si ha

$$\mathbb{P} \left(\limsup_{n \rightarrow +\infty} \{X_n < M\} \right) = 0.$$

2.9. Si dia l'esempio di uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e di una successione (X_n) di v.a. definite in tale spazio, per le quali $\mathbb{E}(X_n) \rightarrow 0$, mentre non esiste alcuna successione $\{n(k) : k \in \mathbb{N}\}$ per la quale $X_{n(k)} \rightarrow 0$ in probabilità.

2.10. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ una v.a. $X : \Omega \rightarrow \overline{\mathbb{R}}$ si dice *limitata in probabilità* oppure *stocasticamente limitata* se, per ogni $\varepsilon > 0$ esiste un numero reale $M(\varepsilon) > 0$ tale che

$$\mathbb{P}(|X| \leq M(\varepsilon)) \geq 1 - \varepsilon.$$

Si dimostri che sono equivalenti le affermazioni:

(a) X è limitata in probabilità;

(b) X è q.c. finita.

2.11. Si dimostri il Teorema 2.2.5 in maniera diversa da quella proposta nel testo.

2.12. (a) Si studii la convergenza della successione di f.r. (F_n) , ove F_n è la f.r. della legge $N(0, n)$.

(b) Si studii la convergenza della successione (F_{σ_n}) , ove F_{σ_n} è la f.r. della legge $N(0, \sigma_n^2)$ e $\{\sigma_n\}$ è una successione infinitesima con $\sigma_n > 0$ per ogni $n \in \mathbb{R}$.

2.13. Si studii la convergenza completa della successione di f.r. (F_n) , ove F_n è la f.r. della legge di Poisson $\mathcal{P}(\lambda_n)$, nei due casi:

(a) $\lambda_n \xrightarrow[n \rightarrow +\infty]{} 0$;

(b) $\lambda_n \xrightarrow[n \rightarrow +\infty]{} +\infty$.

2.14. Siano F_n ($n \in \mathbb{N}$) e F f.r. di v.a. che assumono valori

$$x_1 < x_2 < \cdots < x_j < \dots$$

con probabilità date da $\mathbb{P}(X_n = x_j) = p_j^{(n)}$ e $\mathbb{P}(X = x_j) = p_j$. Allora $F_n \xrightarrow[n \rightarrow +\infty]{c} F$ se, e solo se, per ogni $j \in \mathbb{N}$, risulta $\lim_{n \rightarrow +\infty} p_j^{(n)} = p_j$.

2.15. Può una successione (F_n) ciascuna delle quali è assolutamente continua, convergere completamente a una f.r. che non è assolutamente continua?

2.16. (Teorema di Scheffé) Siano F_n e F f.r. assolutamente continue con densità, rispetto alla misura di Lebesgue, f_n e f , rispettivamente. Se $f_n \rightarrow f$ q.o. rispetto alla misura di Lebesgue riesce $F_n \xrightarrow[n \rightarrow +\infty]{c} F$. Vale lo stesso risultato se si suppone che $f_n \rightarrow f$ nel senso della misura di Lebesgue?

Si vedano a questo proposito (Scheffé, 1947) e (Sempi, 1989).

2.17. Una successione di f.r. assolutamente continue può convergere completamente anche se non converge q.o. la successione delle densità di probabilità. Non vale perciò il reciproco del teorema di Scheffé.

2.18. Sia (X_n) una successione di v.a. indipendenti definite sullo stesso spazio $(\Omega, \mathcal{F}, \mathbb{P})$, ognuna delle quali è distribuita uniformemente nell'intervallo $(0, 1)$. Si studi la convergenza della successione (V_n) ove

$$V_n := \bigvee_{j=1}^n X_j = \max\{X_1, X_2, \dots, X_n\}.$$

2.19. Sia (λ_n) una successione di numeri reali strettamente positivi tale che

$$\lim_{n \rightarrow +\infty} \lambda_n = +\infty.$$

Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia (X_n) una successione di v.a. indipendenti tale che, per ogni $n \in \mathbb{N}$, X_n abbia legge esponenziale di parametro λ_n , $X_n \sim \Gamma(1, \lambda_n)$. Si studi l'eventuale convergenza di (X_n) .

2.20. Sia (X_n) una successione di v.a. indipendenti definite sullo stesso spazio $(\Omega, \mathcal{F}, \mathbb{P})$, aventi la stessa legge esponenziale di parametro λ , $X_n \sim \Gamma(1, \lambda)$. Si studi l'eventuale convergenza in legge e in probabilità della successione (T_n) , ove

$$T_n := \frac{1}{\ln n} \bigvee_{j=1}^n X_j = \frac{1}{\ln n} \max\{X_1, X_2, \dots, X_n\}.$$

2.21. Per ogni $n \in \mathbb{N}$, si consideri la legge μ_n uniforme sui punti

$$0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}.$$

- (a) si mostri che (μ_n) converge completamente e si determini la misura di probabilità limite μ ;
- (b) si trovi un insieme misurabile A per il quale non vale la relazione

$$\lim_{n \rightarrow +\infty} \mu_n(A) = \mu(A).$$

2.22. Siano μ e μ_n , per ogni $n \in \mathbb{N}$, misure di probabilità su $(\mathbb{R}, \mathcal{B})$. Sono equivalenti le affermazioni:

- (a) $\mu_n \xrightarrow[n \rightarrow +\infty]{s} \mu$;
- (b) $\limsup_{n \rightarrow +\infty} \mu_n(C) \leq \mu(C)$ per ogni insieme chiuso C ;
- (c) $\liminf_{n \rightarrow +\infty} \mu_n(B) \geq \mu(B)$ per ogni insieme aperto B ;
- (d) per ogni funzione $f : \mathbb{R} \rightarrow \mathbb{R}$ lipschitziana, limitata e positiva, si ha

$$\liminf_{n \rightarrow +\infty} \int f \, d\mu_n \geq \int f \, d\mu;$$

- (e) per ogni funzione $f : \mathbb{R} \rightarrow \mathbb{R}$ *semicontinua inferiormente* e limitata inferiormente, si ha

$$\liminf_{n \rightarrow +\infty} \int f \, d\mu_n \geq \int f \, d\mu;$$

- (f) per ogni funzione $f : \mathbb{R} \rightarrow \mathbb{R}$ *semicontinua superiormente* e limitata superiormente, si ha

$$\limsup_{n \rightarrow +\infty} \int f \, d\mu_n \leq \int f \, d\mu.$$

Naturalmente, si possono corrispondentemente esprimere formulazione equivalenti della convergenza in legge di una successione (X_n) di v.a. alla v.a. X .

Ricordiamo che una funzione $f : \mathbb{R} \rightarrow \mathbb{R}$ si dice *semicontinua inferiormente* se, per ogni x_0 soddisfa alla condizione

$$\liminf_{x \rightarrow x_0} f(x) \geq f(x_0)$$

o, equivalentemente, se per ogni $c \in \mathbb{R}$ gli *insiemi di livello* $\{f \leq c\}$ sono chiusi, o, ancora se è chiuso l'*epigrafico* di f ,

$$\{(x, y) : f(x) \leq y\}.$$

Si dice che f è *semicontinua superiormente* se $-f$ è *semicontinua inferiormente*.

2.23. Ogni successione (X_n) di v.a. indipendenti definite sullo stesso spazio di probabilità ha probabilità eguale a zero o a uno di convergere quasi certamente. In particolare, se esse sono isonome (e non costanti) è nulla la probabilità che (X_n) converga.

2.24. Se (X_n) è una successione di v.a. sullo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ per la quale esiste $\delta > 0$ tale che sia convergente la serie

$$\sum_{n \in \mathbb{N}} \mathbb{E}(|X_n|^\delta),$$

allora $X_n \rightarrow 0$ q.c..

2.25. Sia X una v.a. su $(\Omega, \mathcal{F}, \mathbb{P})$ con legge uniforme su $(0, 1)$; per ogni $n \in \mathbb{N}$ sia X_n la v.a. definita sullo stesso spazio da

$$X_n := n \mathbf{1}_{\{X \leq 1/n\}}.$$

Si studi la convergenza della successione (X_n) .

2.26. Se $X_n \rightarrow X$ in L^p con $p \geq 1$ allora $\mathbb{E}(|X_n|^p) \rightarrow \mathbb{E}(|X|^p)$.

2.27. Siano (X_n) e (Y_n) due successioni di v.a. che convergono in legge a X e a Y , rispettivamente. Si mostri, con opportuni esempi, che, in generale, né $X_n + Y_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X + Y$ né $X_n Y_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} XY$.

2.28. Siano (X_n) , (Y_n) e (Z_n) successioni di v.a. definite sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$;

(a) (Teorema di Slutsky) se $X_n - Y_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0$ e $Y_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X$, allora

$$X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X;$$

(b) se $Y_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0$ e $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X$, allora $X_n Y_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0$;

(c) $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X$ e $Y_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} c$, allora $X_n + Y_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X + c$;

(d) se $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X$, $Y_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} a$ e $Z_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} c$, allora

$$X_n Y_n + Z_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} aX + c.$$

2.29. Se $X_n \rightarrow X$ in legge, allora $\mathbb{E}(|X|) \leq \liminf_{n \rightarrow +\infty} \mathbb{E}(|X_n|)$.

2.30. Se $X_n \rightarrow X$ in probabilità, allora $\varphi \circ X_n \rightarrow \varphi \circ X$ in probabilità per ogni funzione continua $\varphi : \mathbb{R} \rightarrow \mathbb{R}$. Il risultato può non essere vero se φ è solo misurabile.

2.31. Se $X_n \rightarrow X$ in legge e se $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ è continua, è vero che $\varphi \circ X_n$ converge in legge a $\varphi \circ X$?

2.32. Se (X_n) è una successione di v.a. definite sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, dominata da una v.a. integrabile Y , cioè $|X_n| \leq Y$, per ogni $n \in \mathbb{N}$, $Y \in L^0$, e se $X_n \rightarrow X$ in probabilità, allora

$$\int X_n d\mathbb{P} \rightarrow \int X d\mathbb{P}.$$

2.33. Siano Ω un insieme al più numerabile, cioè $\text{card}(\Omega) \leq \aleph_0$, sia $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ una tribù, e sia P una probabilità su \mathcal{F} . Per una successione (X_n) di v.a. sono equivalenti le proprietà:

(a) (X_n) converge q.c.;

(b) (X_n) converge in probabilità.

2.34. Dato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ si introducano le funzioni *essenzialmente limitate*, vale a dire, le funzioni $f : \Omega \rightarrow \mathbb{R}$ per le quali esiste una costante $k \geq 0$ tale che $\mathbb{P}(|f| > k) = 0$ e si definisca *l'estremo superiore essenziale* di f mediante

$$\text{ess sup } f := \inf\{k \geq 0 : \mathbb{P}(|f| > k) = 0\}.$$

Si pone

$$\|f\|_\infty := \text{ess sup } f.$$

Sia \mathcal{L}^∞ l'insieme delle funzioni essenzialmente limitate. Posto

$$L^\infty := \mathcal{L}^\infty / \simeq_{\mathbb{P}},$$

ove, al solito, $\simeq_{\mathbb{P}}$ è la relazione d'eguaglianza quasi certa, si mostri che

- (a) se f è in L^∞ con $\|f\|_\infty = c$, allora $|f| \leq c$ q.c.;
- (b) L^∞ è uno spazio vettoriale normato da $\|\cdot\|_\infty$ (anzi è anche completo, come si mostra nei corsi d'analisi matematica, sicché è uno spazio di Banach);
- (c) se f è in L^∞ allora essa appartiene anche a L^p per ogni $p \in [1, +\infty[$, sicché vale l'inclusione $L^\infty \subseteq L^p$;
- (d) se la successione di v.a. (X_n) di L^∞ converge alla v.a. X di L^∞ , allora converge allo stesso limite in L^p per ogni $p \in [1, +\infty[$;
- (e) se la successione di v.a. (X_n) di L^∞ converge alla v.a. X di L^∞ , allora converge allo stesso limite anche in probabilità;
- (f) il viceversa di (e) non è necessariamente vero; si dia l'esempio di uno spazio di probabilità e, su di questo, di una successione di v.a. $(X_n)_{n \in \mathbb{N}}$, ciascuna delle quali è limitata, tale che (X_n) converga in probabilità, ma non in L^∞ .

2.35. (Teorema di Pólya) Si mostri che se $F_n \xrightarrow[n \rightarrow +\infty]{c} F$ e se la f.r. F è continua, allora F_n converge uniformemente a F in \mathbb{R} .

2.36. Se $f \in C_b(\mathbb{R})$ e se $F_{n,\theta}$ è una f.r. avente media θ e varianza $\sigma_n^2(\theta)$ tale che $\lim_{n \rightarrow +\infty} \sigma_n^2(\theta) = 0$ per ogni $\theta \in \mathbb{R}$, allora, ponendo

$$\mathbb{E}_{n,\theta}(f) := \int_{\mathbb{R}} f(x) dF_{n,\theta}(x)$$

si ha $\lim_{n \rightarrow +\infty} \mathbb{E}_{n,\theta}(f) = f(\theta)$.

2.37. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ si consideri la famiglia

$$\mathcal{M}(P) := \left\{ f \in L^1 : f > 0 \text{ q.c.}, \mathbb{E}(f) = \int f d\mathbb{P} = 1 \right\}$$

delle densità delle misure di probabilità che sono equivalenti a P . (Due misure si dicono *equivalenti* quando ognuna di esse è assolutamente continua rispetto all'altra). Si dimostri che in $\mathcal{M}(P)$ la convergenza in probabilità e quella in L^1 coincidono.

2.38. Si calcoli il limite, al tendere di n a $+\infty$, della successione (F_n) , ove, per ogni $n \in \mathbb{N}$, F_n è la f.r. della legge di Student a n gradi di libertà.

2.39. Sia (X_n) una successione di v.a. indipendenti definite sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e tutte con distribuzione uniforme nell'intervallo $(0, 1)$. Per ogni $n \in \mathbb{N}$, si definisca $Y_n := \prod_{j=1}^n X_j$.

- (a) Si determini la legge di Y_n ;
- (b) si studi la convergenza di (Y_n) .

2.40. Sia (E_n) una successione di eventi indipendenti sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$; posto $p_n := \mathbb{P}(E_n)$ e $X_n := \mathbf{1}_{E_n}$, si diano condizioni necessarie e sufficienti sulla successione (p_n) affinché si abbia la convergenza $X_n \rightarrow 0$:

- (a) in probabilità;
- (b) quasi certamente.

2.41. Per ogni successione infinitesima (α_n) di reali strettamente positivi esiste una successione (X_n) di v.a. sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ che converge a zero in probabilità senza che $(\alpha_n X_n)$ converga a zero q.c..

2.42. Sia $(X_n)_{n \in \mathbb{N}}$ una successione di v.a. indipendenti ed isonome di L^1 . Per ogni $k \in \mathbb{N}$, sia \mathcal{F}_k la tribù generata dalle v.a. X_1, \dots, X_k , cioè

$$\mathcal{F}_k := \mathcal{F}(X_1, \dots, X_k).$$

Sia $N : \Omega \rightarrow \overline{\mathbb{Z}}_+$ una v.a. che assume valori interi positivi, tale che, per ogni $k \in \mathbb{N}$, risulti $\{N \leq k\} \in \mathcal{F}_k$ e tale che $\mathbb{E}(N) < +\infty$. Al solito, si ponga

$$S_n := \sum_{j=1}^n X_j \quad \text{e} \quad S_0 := 0.$$

Allora vale l'equazione di Wald che costituisce una generalizzazione della proprietà d'additività delle speranze

$$\mathbb{E}(S_N) = \mathbb{E}(X_1) \mathbb{E}(N)$$

ove S_N è la v.a. definita da

$$S_N(\omega) := \sum_{n \in \mathbb{N}} S_n(\omega) \mathbf{1}_{\{N=n\}}(\omega).$$

Si noti che S_N non è stata definita sull'evento trascurabile $\{N = +\infty\}$.

2.43. Siano (X_n) una successione di v.a. e (F_n) la corrispondente successione di f.r. e si supponga che (F_n) converga debolmente ad una funzione crescente F . Allora F è una f.r. se, e solo se, $\{X_n\}$ è stocasticamente limitata (si veda l'esercizio (2.10)).

2.44. Sia (X_n) una successione di v.a. indipendenti e si ponga, per semplicità, $F_n := F_{X_n}$. Delle due condizioni:

- (a) $\forall \varepsilon > 0 \quad \sum_{n \in \mathbb{N}} \{1 - F_n(\varepsilon) + F_n(-\varepsilon)\} < +\infty$,

(b) $X_n \rightarrow 0$ q.c.,

la prima implica la seconda; inoltre, se le v.a. della successione sono indipendenti, esse sono equivalenti.

2.45. Se \mathbb{P} e \mathbb{Q} sono due misure di probabilità sullo stesso spazio misurabile (Ω, \mathcal{F}) e \mathbb{Q} è assolutamente continua rispetto a \mathbb{P} , $\mathbb{Q} \ll \mathbb{P}$, la convergenza in probabilità rispetto a \mathbb{P} implica quella rispetto a \mathbb{Q} . Perciò se \mathbb{P} e \mathbb{Q} sono equivalenti, vale a dire se ciascuna di esse è assolutamente continua rispetto all'altra, $\mathbb{Q} \ll \mathbb{P}$ e $\mathbb{P} \ll \mathbb{Q}$, le due convergenze sono equivalenti.

Se invece $\mathbb{Q} \ll \mathbb{P}$, ma \mathbb{P} non è assolutamente continua rispetto a \mathbb{Q} , può accadere che $X_n \rightarrow 0$ in probabilità \mathbb{Q} , senza che X_n tenda a zero in probabilità \mathbb{P} .

2.46. Sia (F_n) una successione di f.r. e sia $x \in \mathbb{R}$. Le due proprietà seguenti si equivalgono:

(a) $F_n \xrightarrow[n \rightarrow +\infty]{c} \varepsilon_x$;

(b) se $\mathbb{P}_n := \mu_{F_n}$ è la misura di Borel–Stieltjes indotta da F_n , si ha

$$\mathbb{P}_n(A) \xrightarrow[n \rightarrow +\infty]{} 0$$

per ogni boreliano A tale che $x \notin \bar{A}$.

2.47. Sia (\mathbb{P}_n) una successione di misure di probabilità su $(\mathbb{R}, \mathcal{B})$ che converge completamente alla misura di probabilità \mathbb{P} ; allora per ogni $\varepsilon > 0$ esiste un compatto K di \mathbb{R} tale che $\mathbb{P}(K) > 1 - \varepsilon$ e $\mathbb{P}_n(K) > 1 - \varepsilon$ per ogni $n \in \mathbb{N}$.

2.48. (Ancora sul problema dei momenti) Siano X e Y v.a. a valori in $[0, 1]$ e si supponga che, per ogni $n \in \mathbb{Z}_+$, si abbia

$$\mathbb{E}(X^n) = \mathbb{E}(Y^n) .$$

Si mostri che:

(a) $\mathbb{E}(p \circ X) = \mathbb{E}(p \circ Y)$ per ogni polinomio p ;

(b) $\mathbb{E}(f \circ X) = \mathbb{E}(f \circ Y)$ per ogni funzione continua $f : [0, 1] \rightarrow \mathbb{R}$;

(c) X e Y hanno f.r. eguali.

2.49. Sia (F_n) una successione di f.r. con

$$F_n(x) = 0 \quad \text{per } x < 0 \quad \text{e} \quad F_n(1) = 1 ,$$

per ogni $n \in \mathbb{N}$. Si supponga che, per ogni $k \in \mathbb{Z}_+$, esista il limite

$$\alpha_k := \lim_{n \rightarrow +\infty} \int_{[0,1]} x^k dF_n(x) .$$

Allora (F_n) converge completamente ad una f.r. F tale che

$$\int_{[0,1]} x^k dF(x) = \alpha_k \quad (k \in \mathbb{Z}_+) .$$

2.50. Sullo spazio \mathcal{D} delle f.r. si considera talvolta la distanza della convergenza uniforme, detta, in questo ambito, *metrica di Kolmogorov*, definita da

$$d_K(F, G) := \sup\{|F(x) - G(x)| : x \in \mathbb{R}\} \quad (F, G \in \mathcal{D}).$$

Si calcoli la distanza di Kolmogorov $d_K(F, G)$ se $a < b$ e

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - \exp(-ax), & x \geq 0; \end{cases}$$

o

$$G(x) = \begin{cases} 0, & x < 0, \\ 1 - \exp(-bx), & x \geq 0; \end{cases}$$

2.51. Le due condizioni $(F, G; h)$ e $(G, F; h)$ che intervengono nella definizione della metrica di Lévy sono equivalenti. Inoltre, se ε_a è la f.r. della v.a. $X = a$ q.c., allora

$$d_L(\varepsilon_a, \varepsilon_b) = \min\{1, |b - a|\} \leq |b - a|.$$

2.52. Siano X e Y due v.a. discrete con

$$\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = 1/2 \quad \text{e} \quad \mathbb{P}(Y = 1) = 1/4, \mathbb{P}(Y = 2) = 3/4;$$

siano F e G sono le rispettive f.r. si calcolino $d_L(F, G)$ e $d_K(F, G)$.

2.53. Sia F la f.r. della distribuzione uniforme su $(-1, 1)$ e G la f.r. della distribuzione uniforme su $(0, 1)$. Si calcolino $d_L(F, G)$ e $d_K(F, G)$.

2.54. Si calcolino $d_L(F, G)$ e $d_K(F, G)$ se F e G sono date da

$$\begin{aligned} F(x) &= x \mathbf{1}_{[0,1]}(x) + \mathbf{1}_{[1,+\infty]}(x), \\ G(x) &= x^2 \mathbf{1}_{[0,1]}(x) + \mathbf{1}_{[1,+\infty]}(x). \end{aligned}$$

2.55. (a) Per ogni coppia di f.r. F e G si dimostri che

$$d_L(F, G) \leq d_K(F, G).$$

Si dia l'esempio di una successione $(F_n) \subseteq \mathcal{D}$ e di una f.r. F tali che

$$d_L(F_n, F) \rightarrow 0,$$

mentre la successione $(d_K(F_n, F))$ non tende a zero.

(b) Se la f.r. G è assolutamente continua con densità g , si ponga

$$A := \sup\{g(x) : x \in \mathbb{R}\},$$

in caso contrario si ponga $A := +\infty$; allora

$$d_K(F, G) \leq (A + 1) d_L(F, G).$$

2.56. (a) Le f.r. assolutamente continue sono dense in \mathcal{D} , rispetto alla topologia della metrica di Lévy. In questa topologia, anche le f.r. discrete sono dense in \mathcal{D} .

(b) Se F_1, F_2, G_1, G_2 sono f.r., vale la diseuguaglianza

$$d_L(F_1 * F_2, G_1 * G_2) \leq d_L(F_1, G_1) + d_L(F_2, G_2).$$

(c) Si dimostri che se (F_n) e (G_n) sono due successioni di f.r. che convergono completamente alle f.r. F e G , rispettivamente, allora $(F_n * G_n)$ converge completamente a $F * G$.

2.57. Si mostri che se $d_{KF}(X, Y) = \alpha > 0$, allora

$$\mathbb{P}(|X - Y| > \alpha) \leq \alpha.$$

2.58. La convergenza in probabilità per v.a. arbitrarie può essere ridotta a quella per v.a. uniformemente limitate mediante la trasformazione

$$X' := \arctan X.$$

Si mostri che $X_n \rightarrow X$ in probabilità se, e solo se $d_0(X_n, X) \rightarrow 0$, ove

$$d_0(X, Y) := \mathbb{E}(|\arctan X - \arctan Y|).$$

2.59. (a) Quella di Ky Fan non è l'unica metrica su L^0 che metrizza la convergenza in probabilità; di questa stessa proprietà gode anche la metrica definita da

$$d(X, Y) := \mathbb{E}\left(\frac{|X - Y|}{1 + |X - Y|}\right).$$

Si usi tale metrica per dimostrare che

(b) che se (X_n) tende a X in L^p ($p \geq 1$) vi converge anche in probabilità;

(c) che se (X_n) tende a X q.c. vi converge anche in probabilità.

2.60. Sia $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ una funzione continua, strettamente crescente, limitata, tale che $f(0) = 0$ e *subadditiva*, vale a dire tale che

$$\forall x, y \in \mathbb{R}_+ \quad f(x + y) \leq f(x) + f(y).$$

Allora

$$d_f(X, Y) := \mathbb{E}[f(|X - Y|)]$$

definisce una metrica su $L^0 = L^0(\Omega, \mathcal{F}, \mathbb{P})$ la cui topologia è quella della convergenza in probabilità.

Esempi di funzioni con le proprietà richieste sono

$$f_1(x) := 1 - e^{-x}, \quad (\text{a})$$

$$f_2(x) := \frac{kx}{1 + kx}, \quad (\text{b})$$

$$f_3(x) := \tanh x \quad (\text{c})$$

Si noti che la metrica dell'Esercizio 2.59 è un caso particolare di questo; basta considerare la funzione f_2 con $k = 1$.

2.61. Si dimostri direttamente l'implicazione (c) \implies (a) del Teorema 2.3.1.

2.62. Sia $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ una funzione uniformemente continua e limitata e si ponga

$$\|\varphi\| := \sup\{|\varphi(x)| : x \in \mathbb{R}\}.$$

Se $\varepsilon > 0$ e δ sono tali che

$$|\varphi(s) - \varphi(t)| < \varepsilon \quad \text{ogni qual volta} \quad |s - t| < \delta,$$

allora, quali che siano le v.a. X e Y in $(\Omega, \mathcal{F}, \mathbb{P})$ si ha

$$|\mathbb{E}(\varphi \circ X) - \mathbb{E}(\varphi \circ (X + Y))| \leq \varepsilon + 2\|\varphi\| \mathbb{P}(|Y| \geq \delta).$$

2.63. Una successione di leggi di probabilità (μ_n) sullo spazio misurabile $(\mathbb{R}, \mathcal{B})$ converge strettamente alla misura di probabilità μ se, e solo se, per ogni funzione $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ uniformemente continua e limitata si ha

$$\lim_{n \rightarrow +\infty} \int_{\mathbb{R}} \varphi d\mu_n = \int_{\mathbb{R}} \varphi d\mu.$$

2.64. Siano X_n ($n \in \mathbb{N}$), X, Y v.a. q.c. finite definite in $(\Omega, \mathcal{F}, \mathbb{P})$. Se

$$X_n + \alpha Y \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X + \alpha Y$$

per ogni $\alpha > 0$, allora

$$X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X.$$

2.65. Nello spazio $(\Omega, \mathcal{F}, \mathbb{P})$ sia (A_n) una successione di eventi. Allora

(a) vale la diseguaglianza

$$\mathbb{P}\left(\bigcup_{j=1}^n A_j\right) \geq \frac{E^2\left(\sum_{j=1}^n \mathbf{1}_{A_j}\right)}{\mathbb{E}\left[\left(\sum_{j=1}^n \mathbf{1}_{A_j}\right)^2\right]};$$

(b) se $\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) = +\infty$ e se esiste una costante $c > 0$ tale che si abbia, per ogni $j < k$,

$$\mathbb{P}\left(A_j \cap A_k\right) \leq c \mathbb{P}(A_j) \mathbb{P}(A_{k-j}),$$

allora $\mathbb{P}(\limsup_{n \rightarrow +\infty} A_n) > 0$.

2.66. Nello spazio $(\Omega, \mathcal{F}, \mathbb{P})$ sia (A_n) una successione di eventi tale che

$$\lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = 0 \quad \text{e} \quad \sum_{n \in \mathbb{N}} \mathbb{P}(A_n \setminus A_{n+1}) < +\infty;$$

allora $\mathbb{P}(\limsup_{n \rightarrow +\infty} A_n) = 0$.

2.67. Sia (μ_n) una successione di misure di probabilità su $(\mathbb{R}, \mathcal{B})$ che converge strettamente alla misura di probabilità μ . Se (f_n) è una successione di funzioni continue e limitate che converge uniformemente a f , si ha

$$\lim_{n \rightarrow +\infty} \int_{\mathbb{R}} f_n d\mu_n = \int_{\mathbb{R}} f d\mu.$$

2.68. Si mostri che se (X_n) converge in legge a X rispetto alla misura di probabilità P non è detto che vi converga rispetto ad ogni misura di probabilità Q equivalente a P .

2.69. Siano date una successione (X_n) di v.a. ed una v.a. X definite su $(\Omega, \mathcal{F}, \mathbb{P})$. Se per ogni boreliano A che sia di continuità per P_X si ha

$$\lim_{n \rightarrow +\infty} \mathbb{P}(\{X_n \in A\} \Delta \{X \in A\}) = 0,$$

allora (X_n) converge in legge a X .

2.70. Si consideri l'insieme $[0, 1]$ munito della misura di Lebesgue (ristretta ai boreliani di $[0, 1]$). Data la successione (X_n) di v.a. definite da $X_n(x) := x$ per $x \in [0, 1]$ e per $n \in \mathbb{N}$ e la v.a.

$$X(x) := \begin{cases} 1, & x = 0, \\ x + \frac{1}{2}, & x \in \left] 0, \frac{1}{2} \right[, \\ x - \frac{1}{2}, & x \in \left] \frac{1}{2}, 1 \right[. \end{cases}$$

Si dica se (X_n) converge a X in legge rispetto alla misura di probabilità Q che ha densità $f(x) := 2x$ rispetto alla misura di Lebesgue.

2.71. Sia (x_n) una successione di numeri reali convergente a x ,

$$\lim_{n \rightarrow +\infty} x_n = x \in \mathbb{R}$$

e si supponga che, per ogni $n \in \mathbb{N}$, sia $x_n \neq x$. Si consideri la successione di misure di probabilità (μ_n) con $\mu_n = \delta_{x_n}$, e la probabilità $\mu = \delta_x$, le misure di Dirac concentrate in x_n e in x rispettivamente. Si mostri che (μ_n) converge strettamente a μ e si dia l'esempio di un boreliano A tale che $\mu_n(A)$ non converga a $\mu(A)$.

2.72. Sia (X_n) una successione di v.a. positive di L^1 . Se è convergente la serie numerica

$$\sum_{n \in \mathbb{N}} \mathbb{E}(X_n),$$

allora converge q.c. la serie

$$\sum_{n \in \mathbb{N}} X_n.$$

2.73. Nello spazio $(\Omega, \mathcal{F}, \mathbb{P})$ siano (X_n) e (Y_n) due successioni di v.a. tali che

$$\sum_{n \in \mathbb{N}} \mathbb{P}(X_n \neq Y_n) < +\infty.$$

Allora, se (a_n) è una successione crescente di numeri reali con $a_n \rightarrow +\infty$, si ha

(a) la serie $\sum_{n \in \mathbb{N}} (X_n - Y_n)$ converge q.c.;

$$\frac{1}{a_n} \sum_{j=1}^n (X_j - Y_j) \xrightarrow{n \rightarrow +\infty} 0 \quad \text{q.c.}; \quad (\text{b})$$

(c) se $\frac{1}{a_n} \sum_{j=1}^n X_j$ converge q.c., allora anche $\frac{1}{a_n} \sum_{j=1}^n Y_j$ converge q.c. allo stesso limite.

2.74. Si dimostri il Teorema 2.6.1.

2.75. Si dimostri il Teorema 2.6.2.

2.76. Si dimostri il Teorema 2.6.3.

2.77. La convergenza completa implica la convergenza quasi certa; viceversa, se le v.a. della successione (X_n) sono indipendenti, la convergenza completa coincide con la convergenza quasi certa.

2.78. Per ogni $n \in \mathbb{N}$ sia

$$Z_n = X_n + Y_n,$$

dove, per ogni $n \in \mathbb{N}$, Y_n appartiene a L^2 ; si supponga, inoltre che sia

$$\mathbb{E}(Y_n) \xrightarrow{n \rightarrow +\infty} 0 \quad \text{e} \quad V(Y_n) \xrightarrow{n \rightarrow +\infty} 0.$$

Se (X_n) converge in legge a una v.a. X , allora anche (Z_n) converge in legge a X . (Questo risultato è dovuto a Cramér (1946)).

2.79. Per ogni $n \in \mathbb{N}$ sia

$$Z_n = X_n + Y_n$$

dove, per ogni $n \in \mathbb{N}$, Y_n appartiene a L^2 ; si supponga, inoltre che sia

$$\mathbb{E}(Y_n) \xrightarrow{n \rightarrow +\infty} 0 \quad \text{e} \quad V(Y_n) \xrightarrow{n \rightarrow +\infty} 0.$$

Per ogni $n \in \mathbb{N}$ sia U_n una v.a. indipendente da X_n . Se (X_n) e (U_n) convergono in legge a v.a. di f.r. rispettivamente eguali a F e a H , allora, per ogni coppia di numeri reali x e y , si ha

$$\lim_{n \rightarrow +\infty} \mathbb{P}(Z_n \leq x, U_n \leq y) = F(x)H(y).$$

(Per questo esercizio, si veda (Rényi, 1953)).

2.80. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia data una successione (X_n) di v.a. isonome e di L^1 . Se

$$X_n^* := \max \{|X_j| : j = 1, 2, \dots, n\}$$

allora è

$$\frac{X_n^*}{n} \xrightarrow{n \rightarrow +\infty} 0$$

tanto q.c. quanto in L^1 .

2.81. Date una misura di probabilità μ sullo spazio misurabile (Ω, \mathcal{F}) si definisce la *variazione* di μ mediante

$$\|\mu\| := \sup\{\mu(A) : A \in \mathcal{F}\}.$$

Se ν è un'altra misura di probabilità sullo stesso spazio misurabile, resta cosídefinita la *distanza in variazione* tra μ e ν :

$$\|\mu - \nu\| := \sup\{|\mu(A) - \nu(A)| : A \in \mathcal{F}\}.$$

Si mostri che, date μ e ν ,

(a) vale

$$\|\mu - \nu\| := \sup\{\mu(A) - \nu(A) : A \in \mathcal{F}\};$$

(b) esiste una misura λ su (Ω, \mathcal{F}) rispetto alla quale μ e ν sono assolutamente continue e, dunque, ammettono densità f e g , $\mu = f \cdot \lambda$ e $\nu = g \cdot \lambda$;

(c) esiste un insieme $B \in \mathcal{F}$ tale che $\|\mu - \nu\| = \mu(B) - \nu(B)$; inoltre, per le due densità di (b) vale

$$\|\mu - \nu\| = \int (f - g)^+ d\lambda = \frac{1}{2} \int |f - g| d\lambda;$$

(d) esistono due misure di probabilità σ e τ su (Ω, \mathcal{F}) tali che, per ogni $A \in \mathcal{F}$ sia

$$\mu(A) - \nu(A) = \|\mu - \nu\| (\sigma(A) - \tau(A));$$

(e) se $f : \Omega \rightarrow \mathbb{R}$ è limitata e la sua *oscillazione* è definita mediante

$$o(f) := \sup\{|f(\Omega) - f(\Omega')| : \Omega, \Omega' \in \Omega\},$$

allora vale la disuguaglianza

$$\left| \int f d\mu - \int f d\nu \right| \leq o(f) \|\mu - \nu\|.$$

2.82. Sia (μ_n) una successione di misure finite su $(\mathbb{R}, \mathcal{B})$ che converge strettamente (o, rispettivamente, vagamente) a μ . Se $\varphi : \mathbb{R} \rightarrow \mathbb{R}_+$ è continua e limitata, oppure, rispettivamente, continua e con il supporto compatto, allora, per le misure definite da $\nu_n := \varphi \cdot \mu_n$ ($n \in \mathbb{N}$) e da $\nu := \varphi \cdot \mu$ si ha che la successione (ν_n) converge strettamente (o, rispettivamente, vagamente) a ν .

2.83. Una famiglia \mathcal{H} di misure di probabilità sullo spazio misurabile $(\mathbb{R}, \mathcal{B})$ si dice *tesa* (*tight* in inglese) se, per ogni $\varepsilon > 0$ esiste un compatto K_ε con la proprietà che, per ogni misura $\mu \in \mathcal{H}$, si abbia

$$\mu(K_\varepsilon) > 1 - \varepsilon.$$

Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia (X_n) una successione di v.a. limitata in L^2 , vale a dire tale che

$$\sup_{n \in \mathbb{N}} \|X_n\|_2^2 = \sup_{n \in \mathbb{N}} \mathbb{E}(X_n^2) < +\infty.$$

Allora la successione (μ_n) delle leggi delle v.a. di (X_n) è tesa.

2.84. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia (X_n) una successione di v.a. che converge q.c. allav.a. X che è q.c. finita. Allora la v.a. $Y := \sup_{n \in \mathbb{N}} |X_n|$ è q.c. finita.

2.85. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia (X_n) una successione di v.a. normali e centrate, $X_n \sim N(0, \sigma_n^2)$ $n \in \mathbb{N}$; se (X_n) converge in probabilità, essa converge anche in L^2 .

Capitolo 3

Funzioni caratteristiche

3.1 Definizioni e proprietà elementari

Una funzione $f : \Omega \rightarrow \mathbb{C}$ si dice misurabile se sono misurabili sia la sua parte reale $\Re f$ sia la sua parte immaginaria $\Im f$, che sono funzioni a valori reali; analogamente si dirà che f è integrabile se tali sono $\Re f$ e $\Im f$. In tal caso, si definirà

$$\int f \, d\mu := \int \Re f \, d\mu + i \int \Im f \, d\mu.$$

Valgono per l'integrale delle funzioni a valori complessi le stesse proprietà che per l'integrale delle funzioni reali, con la sola eccezione di quelle che dipendono dall'essere \mathbb{R} un insieme totalmente ordinato. A titolo d'esempio, dimostriamo l'importante disuguaglianza

$$\left| \int f \, d\mu \right| \leq \int |f| \, d\mu.$$

Scritto l'integrale $\int f \, d\mu$ sotto forma trigonometrica

$$\int f \, d\mu = \rho e^{i\theta} \quad (\rho \text{ e } \theta \text{ reali e } \rho > 0),$$

si ha

$$\left| \int f \, d\mu \right| = \rho = \int e^{-i\theta} f \, d\mu,$$

e, poiché ρ è reale,

$$\left| \int f \, d\mu \right| = \int \Re(e^{-i\theta} f) \, d\mu \leq \int \left| \Re(e^{-i\theta} f) \right| \, d\mu \leq \int |f| \, d\mu.$$

Definizione 3.1.1. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia X una v.a. reale. La funzione $\varphi_X : \mathbb{R} \rightarrow \mathbb{C}$ definita da

$$\varphi_X(t) := \mathbb{E}[\exp(itX)] = \int_{\mathbb{R}} e^{itx} \, dF_X(x) \quad (t \in \mathbb{R}) \quad (3.1.1)$$

si dice *funzione caratteristica* (f.c.) di X . ◇

L'integrale che compare nella (3.1.1) esiste finito per ogni $t \in \mathbb{R}$ e per ogni v.a. X . La stessa (3.1.1) mostra, poi, che la f.c. φ_X dipende solo dalla f.r. F_X di X , o, ciò che è lo stesso, dalla legge \mathbb{P}_X . Ove non vi sia pericolo di confusione, si scriverà φ in luogo di φ_X , e, se opportuno, si parlerà della f.c. di una legge anziché di una v.a..

La funzione $\varphi : \mathbb{R} \rightarrow \mathbb{C}$ definita dalla (3.1.1) si dice anche *trasformata di Fourier–Stieltjes* della misura di Stieltjes μ_F generata dalla f.r. F ; si vedrà tra breve che, se la f.r. F è assolutamente continua con densità eguale a f , allora la trasformata di Fourier–Stieltjes di μ_F coincide con l'usuale trasformata di Fourier della densità f .

Teorema 3.1.1. *Per una f.c. φ valgono le seguenti proprietà:*

- (a) $\varphi(0) = 1$;
- (b) $|\varphi(t)| \leq 1$ per ogni $t \in \mathbb{R}$;
- (c) φ è uniformemente continua in \mathbb{R} .

Dimostrazione. (a) e (b) sono ovvie. Per la (c) si ha

$$\begin{aligned} |\varphi(t+h) - \varphi(t)| &= |\mathbb{E}[\exp\{i(t+h)X\} - \exp(itX)]| \\ &\leq \mathbb{E}[|\exp(itX)| |\exp(ihX) - 1|] = \mathbb{E}[|\exp(ihX) - 1|]. \end{aligned}$$

L'ultimo termine tende a zero al tendere di h a zero, in virtù del teorema di convergenza dominata; esso non dipende da t , ciò che assicura che la convergenza sia uniforme. \square

Teorema 3.1.2. *Se esiste $t_0 \in \mathbb{R}$ con $t_0 \neq 0$ tale che $\varphi_X(t_0) = 1$, allora la v.a. X assume q.c. i valori $2\pi k/t_0$ ($k \in \mathbb{Z}$).*

Dimostrazione. Da $\varphi_X(t_0) = 1$ scende $\mathbb{E}(\sin t_0 X) = 0$ e $\mathbb{E}(1 - \cos t_0 X) = 0$; poiché $1 - \cos t_0 X \geq 0$ segue che $\cos t_0 X = 1$ q.c. e perciò X è quasi certamente multipla di $2\pi/t_0$. \square

Proposizione 3.1.1. *Se la v.a. Y è una trasformazione affine della v.a. X , $Y = aX + b$ ($a, b \in \mathbb{R}$), si ha*

$$\varphi_Y(t) = \exp(itb) \varphi_X(at)$$

per ogni $t \in \mathbb{R}$.

Dimostrazione. Infatti

$$\varphi_Y(t) = \mathbb{E}(e^{itY}) = \mathbb{E}(e^{itb} e^{iatX}) = e^{itb} \varphi_X(at),$$

cioè l'asserto. \square

Le funzioni caratteristiche svolgono quattro ruoli importanti in probabilità, ruoli che saranno esaminati nel resto di questo capitolo:

- esiste una corrispondenza biunivoca tra funzioni caratteristiche e funzioni di ripartizione, sicché si potrà individuare una legge di probabilità indicandone la sua f.c. (Sezione 3.2);
- esiste un legame profondo tra i momenti di una legge di probabilità e le derivate della corrispondente f.c. (Sezione 3.4);
- le f.c. consentono di determinare in maniera semplice la legge della somma di variabili aleatorie indipendenti (Sezione 3.5);
- infine esse esemplificano lo studio della convergenza completa di successioni di f.r., o, ciò che è equivalente, della convergenza in legge di una successione di v.a. (Sezione 3.6).

3.2 La formula d'inversione

La definizione mostra che ad ogni f.r. corrisponde una f.c.; si vedrà qui di seguito che, viceversa, ad ogni f.c. corrisponde una f.r., sicché si viene a stabilire una biiezione tra la famiglia delle f.c. e quella delle f.r.. Sarà pertanto equivalente individuare una legge di probabilità mediante la sua f.r. oppure attraverso la sua f.c.. Faremo uso del seguente lemma, la cui dimostrazione è usuale nei corsi di analisi complessa.

Lemma 3.2.1. *La funzione $\psi : \mathbb{R}_+ \rightarrow \mathbb{R}$ definita da*

$$\psi(t) := \int_0^t \frac{\sin x}{x} dx \quad (3.2.1)$$

è limitata e risulta

$$\lim_{t \rightarrow +\infty} \psi(t) = \frac{\pi}{2}. \quad (3.2.2)$$

Si vedrà negli esercizi che la funzione $x \mapsto \frac{\sin x}{x}$ non è integrabile nel senso di Lebesgue.

La funzione ψ è stata definita su \mathbb{R}_+ ; tuttavia, tenendo conto delle simmetrie delle funzioni identità e seno, la sua definizione può essere estesa a tutto \mathbb{R} , ponendo

$$\psi(t) := -\psi(-t) \quad \text{se } t < 0.$$

Teorema 3.2.1. (d'inversione). *Sia φ la f.c. della f.r. F ; allora, per $a < b$, si ha*

$$\frac{1}{2\pi} \int_{-T}^T \frac{e^{-iat} - e^{-ibt}}{it} \varphi(t) dt \xrightarrow{T \rightarrow +\infty} \frac{F(b) + \ell^- F(b)}{2} - \frac{F(a) + \ell^- F(a)}{2}. \quad (3.2.3)$$

Dimostrazione. Si consideri l'integrale

$$\begin{aligned} I(T) &:= \frac{1}{2\pi} \int_{-T}^T \frac{e^{-iat} - e^{-ibt}}{it} \varphi(t) dt \\ &= \frac{1}{2\pi} \int_{-T}^T \frac{e^{-iat} - e^{-ibt}}{it} dt \int_{\mathbb{R}} e^{itx} dF(x) \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} dF(x) \int_{-T}^T \frac{e^{it(x-a)} - e^{it(x-b)}}{it} dt, \end{aligned}$$

ove si è fatto ricorso al teorema di Fubini, ciò che è lecito perché l'integrando ha modulo eguale a

$$\left| \frac{e^{-iat} - e^{-ibt}}{it} \right| = \left| \int_a^b e^{-its} ds \right| \leq b - a,$$

e

$$\int_{\mathbb{R}} dF \int_{-T}^T (b - a) dt = 2T(b - a) < +\infty.$$

Ora,

$$\begin{aligned} 2\pi I(T) &= \int_{\mathbb{R}} dF(x) \int_{-T}^T dt \int_{x-b}^{x-a} e^{its} ds = \int_{\mathbb{R}} dF(x) \int_{x-b}^{x-a} ds \int_{-T}^T e^{its} dt \\ &= 2 \int_{\mathbb{R}} dF(x) \int_{x-b}^{x-a} ds \int_0^T \cos ts dt = 2 \int_{\mathbb{R}} dF(x) \int_{T(x-b)}^{T(x-a)} \frac{\sin u}{u} du \\ &= 2 \int_{\mathbb{R}} (\psi[T(x-a)] - \psi[T(x-b)]) dF(x) \\ &= 2 \int_{\mathbb{R}} (\psi[T(x-a)] + \psi[T(b-x)]) dF(x), \end{aligned}$$

relazione che mostra che $I(T)$ è finito, in virtù della limitatezza di ψ . Al tendere di T a $+\infty$, l'integrando tende a: 0, se $x < a$; a $\pi/2$, se $x = a$; a π , se $x \in]a, b[$; a $\pi/2$, se $x = b$; a 0, se $x > b$. Il teorema di convergenza dominata assicura che si abbia

$$\begin{aligned} I(T) &\xrightarrow{T \rightarrow +\infty} \frac{1}{2} (F(a) - \ell^- F(a)) + (\ell^- F(b) - F(a)) \\ &\quad + \frac{1}{2} (F(b) - \ell^- F(b)) \\ &= \frac{1}{2} (F(b) + \ell^- F(b)) - \frac{1}{2} (F(a) + \ell^- F(a)), \end{aligned}$$

che conclude la dimostrazione. \square

Si osservi che nella (3.2.3) non è, in generale, possibile sostituire il limite per T che tende a $+\infty$ con l'integrale esteso a tutto \mathbb{R} . Infatti, in generale, per una funzione $g : \mathbb{R} \rightarrow \mathbb{R}$, si ha

$$\lim_{T \rightarrow +\infty} \int_{-T}^T g(t) dt \neq \int_{\mathbb{R}} g(t) dt.$$

Nella teoria delle funzioni, si definisce il *valore principale dell'integrale* di g come

$$\text{PV} \int_{\mathbb{R}} g(t) dt := \lim_{T \rightarrow +\infty} \int_{-T}^T g(t) dt;$$

quest'ultimo può esistere anche quando non esista l'integrale di g esteso a tutto \mathbb{R} .

Se la funzione di ripartizione F è continua in a e in b , allora la (3.2.3) si può scrivere nella forma

$$F(b) - F(a) = \lim_{T \rightarrow +\infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-iat} - e^{-ibt}}{it} \varphi(t) dt. \quad (3.2.4)$$

Il seguente corollario stabilisce l'asserita corrispondenza biunivoca tra f.r. e f.c..

Corollario 3.2.1. *Sono equivalenti le proprietà:*

- (a) *le f.r. F e G sono eguali, $F = G$;*
- (b) *le f.c. φ_F e φ_G sono eguali, $\varphi_F = \varphi_G$.*

Dimostrazione. Basta dimostrare l'implicazione (b) \implies (a), perché l'altra scende dalla definizione di f.c.. La (3.2.3) dà, per ogni coppia di numeri reali a e b con $a < b$:

$$(F(b) + \ell^- F(b)) - (F(a) + \ell^- F(a)) = (G(b) + \ell^- G(b)) - (G(a) + \ell^- G(a)).$$

Facendo tendere a a $-\infty$, si ottiene $F(b) + \ell^- F(b) = G(b) + \ell^- G(b)$ per ogni b reale; di qui segue che, se $b \in C(F) \cap C(G)$, allora $F(b) = G(b)$. Ma allora F e G , che, come f.r., sono entrambe continue a destra, coincidono. \square

Né la (3.2.3) né la (3.2.4) sono convenienti per i calcoli. Si hanno però formule piú semplici, e piú utili, se valgono ipotesi piú restrittive. In particolare, è utile il seguente risultato.

Teorema 3.2.2. *Se la f.c. φ è integrabile, cioè $\varphi \in \mathcal{L}^1$, allora la f.r. F che le corrisponde è assolutamente continua e ha densità data da*

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \varphi(t) dt \quad (x \in \mathbb{R}). \quad (3.2.5)$$

Dimostrazione. Nella (3.2.3) si prendano $b = x$ e $a = x - h$ con $h > 0$; poiché la condizione di integrabilità assicura che il valor principale coincida con l'integrale, cioè, formalmente, se $g \in L^1$, allora

$$\int_{\mathbb{R}} g(t) dt = \lim_{T \rightarrow +\infty} \int_{-T}^T g(t) dt,$$

si ha

$$F(x) + \ell^- F(x) - (F(x - h) + \ell^- F(x - h)) = \frac{1}{\pi} \int_{\mathbb{R}} \frac{e^{ith} - 1}{it} e^{-itx} \varphi(t) dt.$$

In quest'ultima relazione, si prenda il limite $h \downarrow 0$; il secondo membro tende a zero in virtù del teorema di convergenza dominata, mentre il primo ha come limite $F(x) - \ell^- F(x)$. Perciò, F è continua in \mathbb{R} e si può quindi usare la (3.2.4).

Dimostriamo ora che la f.r. F è assolutamente continua. A tal fine basta mostrare che essa è Lipschitziana. Per ogni x ed per ogni h in \mathbb{R} si ha

$$F(x+h) - F(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{1 - e^{-ith}}{it} e^{-itx} \varphi(t) dt.$$

Di qui

$$|F(x+h) - F(x)| \leq \frac{1}{2\pi} \int_{\mathbb{R}} \left| \frac{1 - e^{-ith}}{it} \right| |\varphi(t)| dt;$$

scrivendo, come sopra,

$$\left| \frac{1 - e^{-ith}}{it} \right| = \left| \int_0^h e^{-its} ds \right| \leq \int_0^{|h|} |e^{-its}| ds \leq |h|,$$

si ottiene

$$|F(x+h) - F(x)| \leq \frac{\|\varphi\|_1}{2\pi} |h|,$$

ciò che prova che F è assolutamente continua.

La (3.2.4), scritta con $b = x+h$ e $a = x$ ($h > 0$), dà

$$\frac{F(x+h) - F(x)}{h} = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{1 - e^{-ith}}{ith} e^{-itx} \varphi(t) dt;$$

esiste quindi il limite per $h \rightarrow 0$ che è eguale al secondo membro della (3.2.5). In maniera analoga si procede se $h < 0$. \square

3.3 Funzioni caratteristiche notevoli

Diamo di seguito la lista delle f.c. di alcune delle leggi di probabilità che si sono già incontrate.

Esempio 3.3.1. (v.a. costante q.c.). $X = a$; $\varphi(t) = \exp(iat)$. \blacksquare

Esempio 3.3.2. (v.a. di Bernoulli). $\varphi(t) = pe^{it} + q$. \blacksquare

Esempio 3.3.3. (Legge binomiale di parametri n e p). $X \sim bi(n, p)$:

$$\varphi(t) = \mathbb{E}(e^{itX}) = \sum_{j=0}^n \binom{n}{j} p^j q^{n-j} e^{itj} = (pe^{it} + q)^n;$$

questo risultato si sarebbe potuto ottenere facilmente come conseguenza di (3.5.6) e di (3.3.2). \blacksquare

Esempio 3.3.4. (Legge geometrica). Si ha

$$\varphi(t) = \sum_{n \in \mathbb{N}} pq^{n-1} e^{itn} = \frac{pe^{it}}{1 - qe^{it}}.$$

\blacksquare

Esempio 3.3.5. (Legge di Poisson).

$$\varphi(t) = e^{-\lambda} \sum_{n \in \mathbb{Z}_+} \frac{\lambda^n}{n!} e^{itn} = e^{-\lambda} \exp(\lambda e^{it}) = \exp\{\lambda (e^{it} - 1)\}.$$

■

Esempio 3.3.6. (Distribuzione uniforme in $(-a, a)$). Se $t \neq 0$ è

$$\begin{aligned} \varphi(t) &= \frac{1}{2a} \int_{-a}^a \cos tx \, dx + \frac{i}{2a} \int_{-a}^a \sin tx \, dx = \frac{1}{a} \int_0^a \cos tx \, dx \\ &= \frac{1}{at} [\sin tx]_{x=0}^{x=a} = \frac{\sin at}{at}. \end{aligned}$$

Se $t = 0$ si ha $\varphi(0) = 1$, come per ogni f.c.; si osservi che φ è continua nell'origine (in particolare). ■

Esempio 3.3.7. (Legge normale). Si consideri, innanzi tutto, il caso di una v.a. $X \sim N(0, 1)$:

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-x^2/2} e^{itx} \, dx = \frac{e^{-t^2/2}}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{1}{2}(x - it)^2\right) \, dx.$$

La funzione $z \mapsto f(z) := e^{-z^2/2}$ è analitica in tutto \mathbb{C} ; è perciò nullo il suo integrale esteso al contorno indicato in Fig. 3.1. Sia $z = \alpha - is$, con $s \in [0, |t|]$, un arbitrario punto del segmento chiuso DA ; allora

$$\left| \exp\left(-\frac{1}{2}(\alpha - is)^2\right) \right| = \exp\left(\frac{s^2 - \alpha^2}{2}\right) \leq \exp\left(\frac{t^2 - \alpha^2}{2}\right),$$

poiché $|e^z| = e^{\Re z}$. Perciò

$$\left| \int_D^A f(z) \, dz \right| \leq \int_0^{|t|} \left| \exp\left(-\frac{(a - is)^2}{2}\right) \right| \, ds \leq |t| \exp\left(-\frac{\alpha^2}{2}\right) e^{t^2/2},$$

donde, per ogni $t \in \mathbb{R}$,

$$\lim_{\alpha \rightarrow +\infty} \left| \int_D^A f(z) \, dz \right| = 0.$$

Similmente, per ogni $t \in \mathbb{R}$,

$$\lim_{\alpha \rightarrow +\infty} \left| \int_B^C f(z) \, dz \right| = 0.$$

Perciò

$$\lim_{\alpha \rightarrow +\infty} \int_B^A f(z) \, dz = \lim_{\alpha \rightarrow +\infty} \int_C^D f(z) \, dz;$$

ma

$$\lim_{\alpha \rightarrow +\infty} \int_B^A f(z) \, dz = \int_{\mathbb{R}} e^{-x^2/2} \, dx = \sqrt{2\pi},$$

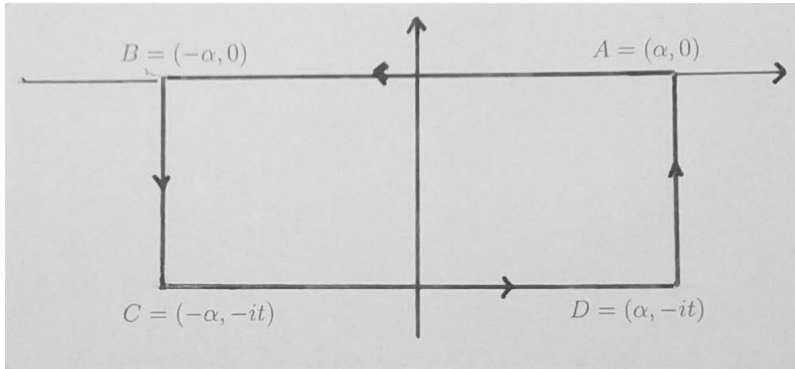


Figura 3.1: Il contorno d'integrazione per la f.c. della legge normale con $\alpha > 0$ e $t > 0$.

sicché

$$\lim_{\alpha \rightarrow +\infty} \int_C^D f(z) dz = \int_{\mathbb{R}} \exp\left(-\frac{(x-it)^2}{2}\right) dx = \sqrt{2\pi},$$

e dunque

$$\varphi_{N(0,1)}(t) = e^{-t^2/2}.$$

Poiché $X \sim N(m, \sigma^2)$ se $X = \sigma Y + m$ con $Y \sim N(0, 1)$, la Proposizione 3.1.1 dà

$$\varphi_{N(m, \sigma^2)}(t) = \exp\left(itm - \frac{1}{2}\sigma^2 t^2\right).$$

■

Esempio 3.3.8. (La legge gamma). $X \sim \Gamma(r, \lambda)$:

$$\varphi(t) = \frac{\lambda^r}{\Gamma(r)} \int_0^{+\infty} x^{r-1} \exp(-(\lambda - it)x) dx.$$

Per $r > 0$, la funzione $z \mapsto f(z) := z^{r-1}e^{-z}$ è analitica in ogni regione del piano complesso che non contenga l'origine. È così nullo il suo integrale calcolato lungo il contorno indicato in Fig. 3.2

Usando coordinate polari, con $\alpha = \lambda - it$, si ha

$$\begin{aligned} \left| \int_C^D z^{r-1} e^{-z} dz \right| &= \rho^r \left| \int_{\theta_0}^0 e^{ir\theta} \exp(-\rho e^{i\theta}) d\theta \right| \\ &\leq \rho^r \int_0^{|\theta_0|} \left| \exp(-\rho e^{i\theta}) \right| d\theta \\ &= \rho^r \int_0^{|\theta_0|} e^{-\rho \cos \theta} d\theta = \rho^r |\theta_0| e^{-\rho \cos \theta'}, \end{aligned}$$

per un opportuno $\theta' \in]0, |\theta_0|[$. Perciò

$$\lim_{\substack{\rho \rightarrow 0 \\ \rho > 0}} \left| \int_C^D z^{r-1} e^{-z} dz \right| = 0.$$

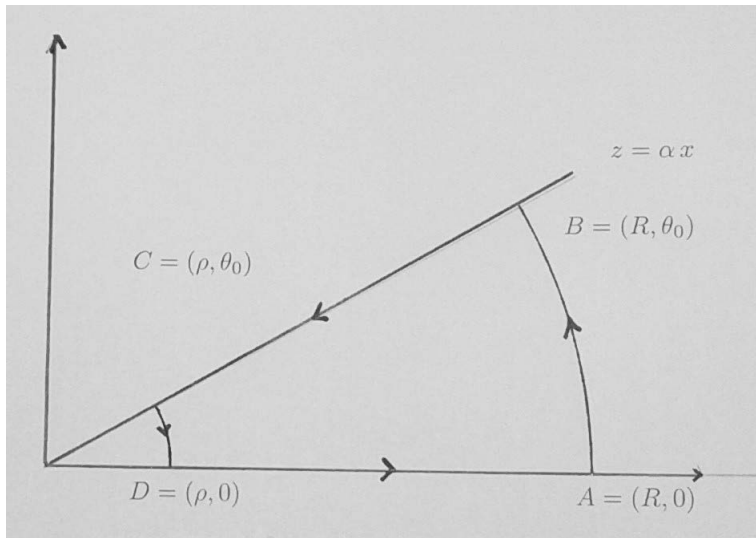


Figura 3.2: Il contorno d'integrazione per la legge gamma con $0 < \rho < R$ e $\theta_0 > 0$.

Analogamente,

$$\lim_{R \rightarrow +\infty} \left| \int_A^B z^{r-1} e^{-z} dz \right| = 0.$$

Scende pertanto dal teorema di Cauchy

$$\lim_{\substack{R \rightarrow +\infty \\ \rho \rightarrow 0, \rho > 0}} \int_C^B f(z) dz = \lim_{\substack{R \rightarrow +\infty \\ \rho \rightarrow 0, \rho > 0}} \int_D^A f(z) dz.$$

Ora

$$\lim_{\substack{R \rightarrow +\infty \\ \rho \rightarrow 0, \rho > 0}} \int_D^A f(z) dz = \int_0^{+\infty} x^{r-1} e^{-x} dx = \Gamma(r),$$

mentre

$$\lim_{\substack{R \rightarrow +\infty \\ \rho \rightarrow 0, \rho > 0}} \int_C^B f(z) dz = \alpha^r \int_0^{+\infty} x^{r-1} e^{-\alpha x} dx.$$

In definitiva, risulta

$$\begin{aligned} \varphi(t) &= \frac{\lambda^r}{\Gamma(r)} \int_0^{+\infty} x^{r-1} \exp(-(\lambda - it)x) dx \\ &= \frac{\lambda^r}{\Gamma(r)} \frac{\Gamma(r)}{\alpha^r} = \left(\frac{\lambda}{\alpha}\right)^r = \left(1 - i\frac{t}{\lambda}\right)^{-r}. \end{aligned}$$

■

Da quest'ultimo risultato si ha, in particolare, per la distribuzione esponenziale che ha legge $\Gamma(1, \lambda)$:

Esempio 3.3.9. (Legge esponenziale). $\varphi(t) = \left(1 - i\frac{t}{\lambda}\right)^{-1}$.

■

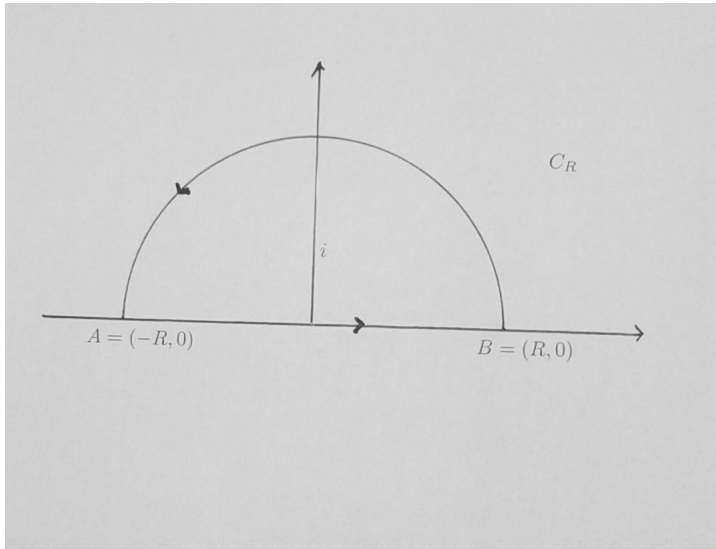


Figura 3.3: Il contorno d'integrazione per la legge di Cauchy e $t > 0$.

Esempio 3.3.10. (Legge di Cauchy). Si consideri dapprima il caso dei parametri $\alpha = 0$ e $\beta = 1$, $X \sim C(0, 1)$. Dovendo calcolare

$$\varphi(t) = \frac{1}{\pi} \int_{\mathbb{R}} \frac{\exp(itx)}{1+x^2} dx,$$

si consideri la funzione $f : \mathbb{C} \rightarrow \mathbb{C}$ definita da

$$f(z) := \frac{e^{itz}}{1+z^2},$$

che ha due poli semplici in $z = i$ e in $z = -i$. Supposto $t > 0$, si integri f lungo il contorno indicato in figura 3.3; sia $R > \sqrt{2}$ il raggio della semicirconferenza C_R .

Il teorema dei residui dà

$$\int_A^B + \int_{C_R} f(z) dz = 2\pi i r(i),$$

ove $r(i)$ è il residuo di f in $z = i$. Ora,

$$r(i) = \left. \frac{e^{itz}}{i+z} \right]_{z=i} = \frac{e^{-t}}{2i}$$

e

$$\begin{aligned} \left| \int_{C_R} f(z) dz \right| &= \left| \int_0^\pi \frac{\exp(itR e^{i\theta})}{1+R^2 \exp(2i\theta)} R i e^{i\theta} d\theta \right| \\ &\leq R \int_0^\pi \frac{|\exp(itR e^{i\theta})|}{\left[(1+R^2 \cos 2\theta)^2 + R^4 \sin^2 2\theta \right]^{1/2}} d\theta \\ &\leq \frac{R}{R^2-1} \int_0^\pi \exp(-Rt \sin \theta) d\theta = \pi \frac{R}{R^2-1} e^{-tRk} \end{aligned}$$

con $k \in]0, 1[$; perciò,

$$\lim_{R \rightarrow +\infty} \int_{C_R} f(z) \, dz = 0$$

e quindi

$$\int_{\mathbb{R}} \frac{e^{itx}}{1+x^2} \, dx = \pi e^{-t}.$$

Se, poi, $t < 0$, si integra lungo la semicirconfenza di raggio R contenuta nel semipiano delle ordinate negative e si ripetono le medesime considerazioni giungendo a

$$\int_{\mathbb{R}} \frac{e^{itx}}{1+x^2} \, dx = \pi e^t.$$

Perciò $\varphi(t) = e^{-|t|}$ per ogni $t \in \mathbb{R}$. Nel caso generale, $X \sim C(\alpha, \beta)$, dovendo calcolare

$$\varphi(t) = \frac{1}{\pi\beta} \int_{\mathbb{R}} \frac{e^{itx}}{1 + \left(\frac{x-\alpha}{\beta}\right)^2} \, dx,$$

si effettua il cambio di variabile $y = (x - \alpha)/\beta$ per ottenere

$$\varphi(t) = e^{-\beta|t|} e^{it\alpha}.$$

■

3.4 Funzioni caratteristiche e momenti

Occorre premettere la seguente generalizzazione del teorema sulla derivazione sotto il segno d'integrale.

Teorema 3.4.1. *Siano μ una misura su $(\mathbb{R}, \mathcal{B})$ e $(x, t) \mapsto f(x, t)$ una funzione definita in $\mathbb{R} \times]a, b[$ con $-\infty \leq a < b \leq +\infty$, a valori in \mathbb{C} . Se vale:*

- (a) *per ogni $t \in]a, b[$, la funzione $x \mapsto f(x, t)$ sia integrabile rispetto a μ ;*
- (b) *f sia derivabile rispetto a t ed esista una funzione $g : \mathbb{R} \rightarrow \mathbb{R}_+$ integrabile e tale che, per ogni $x \in \mathbb{R}$ e per ogni $t \in]a, b[$, sia $|\partial_2 f(x, t)| \leq g(x)$,*

Allora, la funzione

$$t \mapsto F(t) := \int_{\mathbb{R}} f(x, t) \, d\mu(x)$$

è derivabile e si ha

$$F'(t) = \int_{\mathbb{R}} \partial_2 f(x, t) \, d\mu(x). \quad (3.4.1)$$

In particolare, se μ è una misura di probabilità, la (3.4.1) si scrive

$$\frac{d}{dt} \mathbb{E}[f(\cdot, t)] = \mathbb{E}[\partial_2 f(\cdot, t)].$$

Dimostrazione. Sia (t_n) un'arbitraria successione tendente a t , con $t_n \neq t$ per ogni $n \in \mathbb{N}$, e si consideri il rapporto incrementale

$$\frac{F(t_n) - F(t)}{t_n - t} = \int_{\mathbb{R}} \frac{f(x, t_n) - f(x, t)}{t_n - t} d\mu(x).$$

Le funzioni $h_n : \mathbb{R} \rightarrow \mathbb{C}$ definite, per $n \in \mathbb{N}$, da

$$h_n(x) := \frac{f(x, t_n) - f(x, t)}{t_n - t},$$

sono misurabili e costituiscono una successione tendente a $\partial_2 f(x, t)$. D'altra parte, $h_n(x) = \partial_2 f[x, t + \theta_n(x)]$ e, pertanto, $|h_n| \leq g$, per ogni $n \in \mathbb{N}$, sicché l'asserto segue dal teorema di convergenza dominata. \square

Teorema 3.4.2. *Sia X una v.a. reale che ammetta momento di ordine $n \in \mathbb{N}$. Allora, la sua f.c. φ è derivabile n volte e risulta*

$$\mathbb{E}(X^k) = \frac{\varphi^{(k)}(0)}{i^k} = \frac{1}{i^k} \left[\frac{d^k \varphi}{dt^k} \right]_{t=0} \quad (k \leq n).$$

Dimostrazione. Sia $f(x, t) := e^{itx}$; applicando il Teorema 3.4.1 a questa funzione, si ha

$$\left| \partial_2^k f(x, t) \right| = \left| i^k x^k \exp(itx) \right| \leq |x|^k.$$

Per ipotesi, $\mathbb{E}(|X|^k) < +\infty$ se $k \leq n$. Di qui l'asserto. \square

Il reciproco del Teorema 3.4.2 non è valido in generale perché una f.c. può essere derivabile nell'origine senza che la corrispondente distribuzione abbia media finita.

Esempio 3.4.1. Si consideri la legge individuata dalla densità

$$f(x) := k^{-1} (|x|^2 \ln |x|)^{-1} \mathbf{1}_{\{|x| \geq 2\}}(x),$$

ove la costante k è tale che

$$k = \int_{\{|x| \geq 2\}} \frac{1}{|x|^2 \ln |x|} dx.$$

Tale legge non ha speranza finita; infatti

$$\begin{aligned} \int_{\mathbb{R}} |x| f(x) dx &= \frac{1}{k} \left(\int_{-\infty}^{-2} + \int_2^{+\infty} \frac{1}{|x| \ln |x|} dx \right) \\ &= \frac{2}{k} \int_2^{+\infty} \frac{1}{x \ln x} dx = \frac{2}{k} [\ln \ln x]_{x=2}^{x=+\infty} = +\infty. \end{aligned}$$

La sua f.c. è

$$\varphi(t) = \frac{2}{k} \int_2^{+\infty} \frac{\cos tx}{x^2 \ln x} dx.$$

Questa ammette derivata nell'origine $t = 0$:

$$\frac{\varphi(h) - \varphi(0)}{h} = \frac{2}{k} \int_2^{+\infty} \frac{\cos hx - 1}{h} \frac{1}{x^2 \ln x} dx \xrightarrow{h \rightarrow 0} 0.$$

La derivata non esiste se $t \neq 0$, perché

$$x \mapsto \frac{\sin tx}{x \ln x}$$

non è integrabile. Non esiste perciò neanche la derivata seconda di φ . ■

Vi è però un reciproco parziale del Teorema 3.4.2 che, alla luce dell'esempio precedente non può essere migliorato.

Teorema 3.4.3. *Se la f.c. φ di una v.a. X è derivabile un numero pari di volte, $2k$, allora X ha momento finito di ordine $2k$.*

Dimostrazione. Si supponga, dapprima, che sia $k = 1$. Si ponga

$$\alpha(t) := \Re\varphi(t) = \int_{\mathbb{R}} \cos xt \, dF(x),$$

e

$$\beta(t) := \Im\varphi(t) = \int_{\mathbb{R}} \sin xt \, dF(x);$$

allora $\alpha : \mathbb{R} \rightarrow \mathbb{R}$ è pari, mentre $\beta : \mathbb{R} \rightarrow \mathbb{R}$ è dispari. Le funzioni α e β sono derivabili e $\varphi' = \alpha' + i\beta'$; ora, α' è dispari, mentre β' è pari e perciò $\alpha'(0) = 0$. Si ha anche $\beta''(0) = 0$ perché β'' è nuovamente dispari. Quindi $\varphi''(0) = \alpha''(0)$. La formula di Taylor, applicata a α , dà $\alpha(t) = \alpha(0) + \frac{1}{2} \alpha''(0)t^2 + o(t^2)$, onde

$$\alpha''(0) = 2 \lim_{t \rightarrow 0} \frac{\alpha(t) - \alpha(0)}{t^2} = 2 \lim_{t \rightarrow 0} \frac{\alpha(t) - 1}{t^2}.$$

Sia (t_n) un'arbitraria successione infinitesima; dunque

$$\varphi''(0) = \alpha''(0) = 2 \lim_{n \rightarrow +\infty} \int_{\mathbb{R}} \frac{\cos t_n x - 1}{t_n^2} dF(x).$$

Definite le funzioni $h_n : \mathbb{R} \rightarrow \mathbb{R}_-$ mediante $h_n(x) := (\cos t_n x - 1)/t_n^2$, risulta $\lim_{n \rightarrow +\infty} h_n(x) = -x^2/2$. Poiché $h_n \leq 0$ per ogni $n \in \mathbb{N}$, applicando il lemma di Fatou, si ottiene

$$\begin{aligned} - \int x^2 dF(x) &= 2 \int \limsup_{n \rightarrow +\infty} h_n(x) dF(x) \\ &\geq \limsup_{n \rightarrow +\infty} 2 \int h_n(x) dF(x) = \varphi''(0), \end{aligned}$$

sicché la speranza $\mathbb{E}(|X|^2)$ è finita. Il teorema è così dimostrato per $k = 1$. Si proceda ora per induzione, supponendo che φ sia derivabile nell'origine $2k$ volte e

supponendo, altresí, di aver mostrato che X ammette momento di ordine $2k - 2$, con $k \geq 1$. Allora

$$\varphi^{(2k-2)}(t) = i^{(2k-2)} \int_{\mathbb{R}} x^{2k-2} e^{itx} dF(x) = (-1)^{k-1} \int_{\mathbb{R}} x^{2k-2} e^{itx} dF(x).$$

Posto

$$\alpha_{k-1}(t) := (-1)^{k-1} \int_{\mathbb{R}} x^{2k-2} \cos tx dF(x),$$

risulta, come sopra,

$$\varphi^{(2k)}(0) = 2 \lim_{t \rightarrow 0} \frac{\alpha_{k-1}(t) - \alpha_{k-1}(0)}{t^2};$$

da questo punto la dimostrazione ricalca quella del caso $k = 1$. \square

Corollario 3.4.1. *Se la f.c. φ della v.a. X è derivabile n volte, allora X ammette tutti i momenti di ordine $k \leq n$ se n è pari, mentre, se n è dispari, X ammette tutti i momenti di ordine $k \leq n - 1$.*

3.5 Funzioni caratteristiche e indipendenza

Prima di presentare i risultati riguardanti la f.c. della somma di variabili aleatorie indipendenti, è opportuno richiamare come lo stesso problema sia risolto senza l'ausilio delle f.c..

Siano X_1, \dots, X_n v.a. indipendenti definite nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, e sia F_j la f.r. di X_j ($j = 1, \dots, n$). Si consideri il vettore aleatorio $\mathbf{X} := (X_1, \dots, X_n)$ che prende valori in $(\mathbb{R}^n, \mathcal{B}^n)$; \mathbf{X} induce una misura probabilità $\mathbb{P}_{\mathbf{X}}$ su $(\mathbb{R}^n, \mathcal{B}^n)$ mediante

$$\mathbb{P}_{\mathbf{X}}(B) := \mathbb{P}(\mathbf{X} \in B) \quad (B \in \mathcal{B}^n).$$

Tale misura di probabilità si dice *legge di \mathbf{X}* .

In particolare, se t_1, \dots, t_n sono numeri reali, si consideri il boreliano

$$B =]-\infty, t_1] \times \dots \times]-\infty, t_n].$$

Poiché le v.a. X_1, \dots, X_n sono indipendenti si ha

$$\mathbb{P}_{\mathbf{X}}(]-\infty, t_1] \times \dots \times]-\infty, t_n]) = \mathbb{P} \left(\bigcap_{j=1}^n \{X_j \leq t_j\} \right) \quad (3.5.1)$$

$$= \prod_{j=1}^n \mathbb{P}(X_j \leq t_j) = \prod_{j=1}^n F_j(t_j). \quad (3.5.2)$$

Ma allora $\mathbb{P}_{\mathbf{X}}$ è la misura prodotto delle misure di Stieljes μ_1, \dots, μ_n indotte da F_1, \dots, F_n :

$$\mathbb{P}_{\mathbf{X}} = \mu_1 \otimes \dots \otimes \mu_n.$$

Pertanto, per ogni boreliano B di \mathbb{R}^n è

$$\mathbb{P}_{\mathbf{X}}(B) = \int_{\{\mathbf{x} \in B\}} dF_1(x_1) \dots dF_n(x_n).$$

Si può enunciare il risultato ottenuto nel seguente risultato.

Teorema 3.5.1. *Siano X_1, \dots, X_n v.a. indipendenti e q.c. finite definite nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e siano F_1, \dots, F_n le rispettive funzioni di ripartizione. Allora, per ogni $B \in \mathcal{B}(\mathbb{R}^n)$, si ha*

$$\mathbb{P}(\mathbf{X} \in B) = \int_{\{\mathbf{x} \in B\}} dF_1(x_1) \dots dF_n(x_n). \quad (3.5.3)$$

Inoltre se $g : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ è misurabile si ha

$$\mathbb{E}(g \circ \mathbf{X}) = \int_{\mathbb{R}^n} g(\mathbf{x}) dF_1(x_1) \dots dF_n(x_n), \quad (3.5.4)$$

nel senso che, se esiste l'integrale a primo o a secondo membro, allora esiste anche l'altro e i due sono eguali.

Come esempio significativo di applicazione del Teorema 3.5.1 si considerino due v.a. X e Y f.r. F e G , rispettivamente, e la loro somma $Z = X + Y$. La f.r. H di Z è, per ogni $t \in \mathbb{R}$,

$$\begin{aligned} H(t) &= \mathbb{P}(Z \leq t) = \mathbb{P}(X + Y \leq t) \\ &= \int_{\{(x,y):x+y \leq t\}} dF(x) dG(y) \\ &= \int_{\mathbb{R}} dG(y) \int_{x \in]-\infty, t-y]} dF(x) = \int_{\mathbb{R}} F(t-y) dG(y). \end{aligned}$$

Dunque

$$H(t) = \int_{\mathbb{R}} F(t-y) dG(y). \quad (3.5.5)$$

L'operazione definita si chiama *convoluzione (di Stieltjes)*; spesso si scrive $H = F * G$.

Se entrambe le v.a. X e Y sono assolutamente continue con densità f e g rispettivamente anche la loro somma $X + Y$ è assolutamente continua e si ha, com'è noto $f_{X+Y} = f * g$.

Possiamo ora ritornare alle f.c..

Teorema 3.5.2. *Se X_1 e X_2 sono v.a. indipendenti definite sopra il medesimo spazio di probabilità, allora vale, per ogni $t \in \mathbb{R}$,*

$$\varphi_{X_1+X_2}(t) = \varphi_{X_1}(t) \varphi_{X_2}(t). \quad (3.5.6)$$

Dimostrazione. Si ponga, per $(j = 1, 2)$,

$$Y_j := \cos tX_j = \Re \exp(itX_j), \quad Z_j := \sin tX_j = \Im \exp(itX_j).$$

I vettori aleatorî (Y_1, Z_1) e (Y_2, Z_2) sono indipendenti e reali; perciò se φ_1 e φ_2 sono le f.c. di X_1 e X_2 , rispettivamente,

$$\begin{aligned}\varphi_1(t) \varphi_2(t) &= \mathbb{E}(Y_1 + iZ_1) \mathbb{E}(Y_2 + iZ_2) \\ &= \mathbb{E}(Y_1) \mathbb{E}(Y_2) - \mathbb{E}(Z_1) \mathbb{E}(Z_2) \\ &\quad + i(\mathbb{E}(Y_1) \mathbb{E}(Z_2) + \mathbb{E}(Z_1) \mathbb{E}(Y_2)) \\ &= \mathbb{E}(Y_1 Y_2 - Z_1 Z_2 + i(Y_1 Z_2 + Z_1 Y_2)) \\ &= \mathbb{E}((Y_1 + iZ_1)(Y_2 + iZ_2)) \\ &= \mathbb{E}(\exp(itX_1) \exp(itX_2)) = \mathbb{E}(\exp[it(X_1 + X_2)]) \\ &= \varphi_{X_1+X_2}(t),\end{aligned}$$

che conclude la dimostrazione. \square

Il reciproco non è vero; si vedano gli esercizi. L'ultimo teorema si può enunciare in modo equivalente ricorrendo al concetto di convoluzione di leggi di probabilità.

Definizione 3.5.1. Siano μ_1 e μ_2 due leggi di probabilità su $(\mathbb{R}, \mathcal{B})$; si dice *convoluzione* $\mu_1 * \mu_2$ di μ_1 e μ_2 la funzione $\mu_1 * \mu_2 : \mathcal{B} \rightarrow \mathbb{R}_+$ definita da

$$(\mu_1 * \mu_2)(B) := \int_{\mathbb{R}} \mu_1(B - t) d\mu_2(t) \quad (B \in \mathcal{B}), \quad (3.5.7)$$

ove $B - t := \{y - t : y \in B\}$. \diamond

Teorema 3.5.3. Se μ_1 e μ_2 sono due misure di probabilità su $(\mathbb{R}, \mathcal{B})$, anche la loro convoluzione $\mu_1 * \mu_2$ è una misura di probabilità su $(\mathbb{R}, \mathcal{B})$; la f.r. di $\mu_1 * \mu_2$ è $F_1 * F_2$ ove F_1 e F_2 sono le f.r. di μ_1 e di μ_2 . Inoltre, se $g : \mathbb{R} \rightarrow \mathbb{R}$ è integrabile rispetto a $\mu_1 * \mu_2$ ($g \in L^1(\mu_1 * \mu_2)$), allora si ha

$$\int_{\mathbb{R}} g(t) d(\mu_1 * \mu_2)(t) = \int_{\mathbb{R}^2} g(x + y) d\mu_1(x) d\mu_2(y). \quad (3.5.8)$$

Dimostrazione. Segue immediatamente dalla (3.5.7) che $\mu_1 * \mu_2$ è una misura. Inoltre

$$(\mu_1 * \mu_2)(\mathbb{R}) = \int_{\mathbb{R}} \mu_1(\mathbb{R} - t) d\mu_2(t) = \int_{\mathbb{R}} \mu_1(\mathbb{R}) d\mu_2(t) = \int_{\mathbb{R}} d\mu_2(t) = 1.$$

Scende ancora dalla (3.5.7), ponendovi $B =]-\infty, x]$, che

$$\begin{aligned}F_{\mu_1 * \mu_2}(x) &= (\mu_1 * \mu_2)(] - \infty, x]) = \int_{\mathbb{R}} \mu_1(] - \infty, x] - t) d\mu_2(t) \\ &= \int_{\mathbb{R}} \mu_1(] - \infty, x - t]) d\mu_2(t) = \int_{\mathbb{R}} F_1(x - t) dF_2(t) \\ &= (F_1 * F_2)(x).\end{aligned}$$

Per dimostrare la (3.5.8), si supponrà, dapprima, che g sia la funzione indicatrice di un boreliano B , $g = \mathbf{1}_B$. Allora $\mathbf{1}_{B-y}(x) = \mathbf{1}_B(x+y)$, sicché

$$\begin{aligned} \int_{\mathbb{R}^2} \mathbf{1}_B(x+y) \, d\mu_1(x) \, d\mu_2(y) &= \int_{\mathbb{R}^2} \mathbf{1}_{B-y}(x) \, d\mu_1(x) \, d\mu_2(y) \\ &= \int_{\mathbb{R}} \mu_1(B-y) \, d\mu_2(y) = (\mu_1 * \mu_2)(B) \\ &= \int_{\mathbb{R}} \mathbf{1}_B(t) \, d(\mu_1 * \mu_2)(t); \end{aligned}$$

la (3.5.8) vale, dunque, per le funzioni indicatrici, e, quindi, per linearità, per le funzioni semplici; mediante il ragionamento oramai usuale, se ne stabilisce la validità anche per le funzioni misurabili positive e per le funzioni integrabili. \square

Teorema 3.5.4. *Siano μ, μ_1, μ_2 leggi di probabilità su $(\mathbb{R}, \mathcal{B})$; se $\mu = \mu_1 * \mu_2$, allora per le f.c. corrispondenti vale, per ogni $t \in \mathbb{R}$,*

$$\varphi(t) = \varphi_1(t) \varphi_2(t).$$

Mediante i teoremi della Sezione 3.2 e quelli della presente è risolto, in linea di principio, il problema di determinare la legge della somma di un numero finito di v.a. indipendenti di ciascuna delle quali sia nota la legge. Quest'ultima è, per la v.a. X_j , individuata dalla sua f.c. φ_j , sicché la f.c. della somma $S_n = \sum_{j \leq n} X_j$, che ne individua la legge, è data da $\varphi(t) = \prod_{j \leq n} \varphi_j(t)$ per ogni $t \in \mathbb{R}$.

3.6 Il teorema di continuità

I due teoremi che seguono, dovuti a Paul Lévy e a H. Cramér, sono noti sotto il nome complessivo di teorema di continuità e sono di larghissima applicazione nello studio di molti teoremi limite in teoria delle probabilità.

Teorema 3.6.1. *Sia (F_n) una successione di f.r. che converge completamente a $F \in \mathcal{D}$. Se φ_n e φ sono le f.c. di F_n e di F , rispettivamente, risulta, per ogni $t \in \mathbb{R}$,*

$$\lim_{n \rightarrow +\infty} \varphi_n(t) = \varphi(t).$$

Dimostrazione. L'asserto è un semplice corollario del Teorema (5.4.6) applicato alle funzioni di $C_b(\mathbb{R})$ definite da

$$\Re e^{itx} = \cos tx \quad \text{e} \quad \Im e^{itx} = \sin tx,$$

che stabilisce l'asserto. \square

Teorema 3.6.2. *Sia (F_n) una successione di f.r. e sia $\{\varphi_n\}$ la corrispondente successione di f.c.. Se, per ogni $t \in \mathbb{R}$, esiste il limite*

$$\varphi(t) := \lim_{n \rightarrow +\infty} \varphi_n(t)$$

e se la funzione $\varphi : \mathbb{R} \rightarrow \mathbb{C}$ così definita è continua in $t = 0$, allora φ è la f.c. di una f.r. F e inoltre si ha $F_n \xrightarrow[n \rightarrow +\infty]{c} F$.

Dimostrazione. Per il teorema di Helly, esistono una successione $(F_{n(j)})$ estratta da (F_n) e una funzione $F : \mathbb{R} \rightarrow [0, 1]$, crescente e continua a destra tale che $(F_{n(j)})$ converga debolmente a F . Per stabilire che F è una f.r., occorre mostrare che

$$\lim_{x \rightarrow +\infty} F(x) - \lim_{x \rightarrow -\infty} F(x) = 1,$$

oppure, ciò che è lo stesso, che, se μ_F è la misura di Stieltjes indotta da F su $(\mathbb{R}, \mathcal{B})$, allora $\mu_F(\mathbb{R}) = 1$. Poiché $\lim_{j \rightarrow +\infty} \varphi_{n(j)}(t) = \varphi(t)$ e $|\varphi_{n(j)}(t)| \leq 1$, il teorema di convergenza dominata dà, per ogni $\delta > 0$,

$$\frac{1}{\delta} \int_0^\delta \varphi(t) dt = \lim_{j \rightarrow +\infty} \frac{1}{\delta} \int_0^\delta \varphi_{n(j)}(t) dt. \quad (3.6.1)$$

In virtù del teorema di Fubini, l'integrale a secondo membro è

$$\begin{aligned} \frac{1}{\delta} \int_0^\delta \varphi_{n(j)}(t) dt &= \frac{1}{\delta} \int_0^\delta dt \int_{\mathbb{R}} e^{itx} dF_{n(j)}(x) \\ &= \int_{\mathbb{R}} dF_{n(j)}(x) \int_0^\delta \frac{e^{itx}}{\delta} dt = \int_{\mathbb{R}} \frac{e^{i\delta x} - 1}{i\delta x} dF_{n(j)}(x). \end{aligned}$$

Per il Teorema (5.4.13), è

$$\lim_{j \rightarrow +\infty} \int_{\mathbb{R}} \frac{e^{i\delta x} - 1}{i\delta x} dF_{n(j)}(x) = \int_{\mathbb{R}} \frac{e^{i\delta x} - 1}{i\delta x} dF(x),$$

sicché la (3.6.1) si scrive

$$\frac{1}{\delta} \int_0^\delta \varphi(t) dt = \int_{\mathbb{R}} \frac{e^{i\delta x} - 1}{i\delta x} dF(x).$$

Ora la continuità di φ in $t = 0$ dà

$$\lim_{\delta \rightarrow 0} \frac{1}{\delta} \int_0^\delta \varphi(t) dt = \varphi(0) = \lim_{n \rightarrow +\infty} \varphi_n(0) = 1. \quad (3.6.2)$$

D'altra parte, poiché

$$\lim_{\delta \rightarrow 0} \frac{e^{i\delta x} - 1}{i\delta x} = 1,$$

prendendo una successione infinitesima (δ_n) , il teorema di convergenza dominata dà, con ovvio significato dei simboli,

$$\lim_{n \rightarrow +\infty} \int_{\mathbb{R}} \frac{e^{i\delta_n x} - 1}{i\delta_n x} dF(x) = \int_{\mathbb{R}} dF(x) = F(+\infty) - F(-\infty);$$

quest'ultima differenza è, per la (3.6.2), eguale a 1. Perciò F è una f.r. e la successione $(F_{n(j)})_{j \in \mathbb{N}}$ converge completamente a F . Dunque φ è la f.c. di F .

Rimane da far vedere che tutta la successione (F_n) converge completamente a F . Se così non fosse, vi sarebbe $\varepsilon > 0$ tale che, per ogni $r \in \mathbb{N}$, esista $F_{n(r)}$ con $n(r) \geq r$,

con $d_L(F_{n(r)}, F) > \varepsilon$. Per il teorema di Helly, la sottosuccessione $(F_{n(r)})_{r \in \mathbb{N}} \subseteq (F_n)$ ne contiene un'altra, che si può supporre essere la stessa $(F_{n(r)})_{r \in \mathbb{N}}$, che converge debolmente a una funzione G crescente e continua a destra che, come sopra, si dimostra essere una f.r.. Ora $F \neq G$, ma $(\varphi_{n(r)})$ converge a φ sicché F e G sarebbero due f.r. differenti aventi la medesima f.c. ciò che è impossibile. \square

In virtù dei risultati del Capitolo 2 e di questo, è ovvio il seguente

Teorema 3.6.3. *La successione di v.a. (X_n) converge in legge alla v.a. X se, e solo se, si verifica una delle seguenti condizioni equivalenti (F_n e F sono le f.r. di X_n e X , rispettivamente):*

- (a) $d_L(F_n, F) \rightarrow 0$;
- (b) $\lim_{n \rightarrow +\infty} F_n(x) = F(x)$ per ogni $x \in C(F)$;
- (c) esiste un sottoinsieme denso di \mathbb{R} , $D \subseteq \mathbb{R}$, $\overline{D} = \mathbb{R}$, tale che

$$\forall x \in D \quad \lim_{n \rightarrow +\infty} F_n(x) = F(x);$$
- (d) $\lim_{n \rightarrow +\infty} \mathbb{E}(f \circ X_n) = \mathbb{E}(f \circ X)$ per ogni $f \in C_b(\mathbb{R})$;
- (e) $\lim_{n \rightarrow +\infty} \varphi_n(t) = \varphi(t)$ per ogni $t \in \mathbb{R}$, ove φ_n è la f.c. di X_n e φ è la f.c. di X .

Definizione 3.6.1. Una famiglia $\{f_\iota : \iota \in I\}$ di funzioni $f_\iota : \mathbb{R} \rightarrow \mathbb{C}$ ($\iota \in I$) si dice *equicontinua* se per ogni $\varepsilon > 0$, esiste $\delta = \delta(\varepsilon) > 0$ tale che $|h| < \delta$ implichi $|f_\iota(t+h) - f_\iota(t)| < \varepsilon$ per ogni $t \in \mathbb{R}$ e per ogni $\iota \in I$. \diamond

In particolare ogni funzione f_ι ($\iota \in I$) di una famiglia equicontinua è uniformemente continua.

Teorema 3.6.4. *Una successione (φ_n) di f.c. che converga ad una f.c. φ è equicontinua.*

Dimostrazione. Sia F_n la f.r. corrispondente a φ_n e F la f.r. corrispondente a φ . Per il Teorema 3.6.3, si ha $F_n \xrightarrow[n \rightarrow +\infty]{c} F$. Fissato $\varepsilon > 0$, esiste $a > 0$ tale che risulti $F_n(-a) < \varepsilon$ e $F_n(a) > 1 - \varepsilon$ per ogni $n \in \mathbb{N}$. Infatti esiste un punto $b > 0$ di continuità per F tale che $F(-b) < \varepsilon$ e $F(b) > 1 - \varepsilon$. Esiste pertanto $n_0 \in \mathbb{N}$ tale che per ogni $n \geq n_0$ sia $F_n(-b) < \varepsilon$ e $F_n(b) > 1 - \varepsilon$. Inoltre per ogni $j = 1, 2, \dots, n_0 - 1$ esiste $a_j > 0$ tale che $F_j(-a_j) < \varepsilon$ e $F_j(a_j) > 1 - \varepsilon$. Basta, allora scegliere $a := \max\{a_1, a_2, \dots, a_{n_0-1}, b\}$ per avere $F_n(-a) < \varepsilon$ e $F_n(a) > 1 - \varepsilon$ per ogni $n \in \mathbb{N}$. Inoltre esiste $\delta > 0$ tale che $|e^{i\delta a} - 1| < \varepsilon$. Allora, per $|h| < \delta$ e per ogni $n \in \mathbb{N}$, risulta

$$\begin{aligned} |\varphi_n(t+h) - \varphi_n(t)| &\leq \int_{\mathbb{R}} |e^{ihx} - 1| dF_n(x) \\ &= \int_{\{|x| < a\}} \dots + \int_{\{|x| \geq a\}} |e^{ihx} - 1| dF_n(x) \\ &\leq \int_{\{|x| < a\}} |e^{ihx} - 1| dF_n(x) + 2 \int_{\{|x| \geq a\}} dF_n(x) < 5\varepsilon, \end{aligned}$$

che dimostra l'equicontinuità di (φ_n) . \square

Teorema 3.6.5. *Nei Teoremi 3.6.1 e 3.6.2 la convergenza delle f.c. è uniforme in ogni intervallo chiuso e limitato $[a, b]$.*

Dimostrazione. Sia dato $\varepsilon > 0$; segue dall'equicontinuità di (φ_n) e dall'uniforme continuità di φ che esiste $\delta = \delta(\varepsilon) > 0$ tale che $|\varphi_n(t+h) - \varphi_n(t)| < \varepsilon$ e $|\varphi(t+h) - \varphi(t)| < \varepsilon$ se $|h| < \delta$, per ogni $n \in \mathbb{N}$ e per ogni $t \in \mathbb{R}$. Siano ora t_j ($j = 0, 1, \dots, r+1$) punti tali che $a = t_0 < t_1 < \dots < t_{r+1} = b$ e che $\max\{t_{j+1} - t_j : j = 0, 1, \dots, r\} < \delta$. Esiste allora $N \in \mathbb{N}$ tale che $|\varphi_n(t_j) - \varphi(t_j)| < \varepsilon$ per ogni j e per ogni $n \geq N$. Se $t \in [t_{j-1}, t_j]$, risulta, per $n \geq N$,

$$|\varphi_n(t) - \varphi(t)| \leq |\varphi_n(t) - \varphi_n(t_j)| + |\varphi_n(t_j) - \varphi(t_j)| + |\varphi(t_j) - \varphi(t)| < 3\varepsilon,$$

che conclude la dimostrazione. \square

Non è possibile eliminare, nell'enunciato del Teorema 3.6.2, la richiesta che la funzione limite φ sia continua in $t = 0$. Si considerino, infatti, i seguenti esempi.

Esempio 3.6.1. La f.c. della legge uniforme su $(-n, n)$ è

$$\varphi_n(t) = \frac{\sin nt}{nt}.$$

Ora

$$\lim_{n \rightarrow +\infty} \varphi_n(t) = \varphi(t) := \begin{cases} 0, & t \neq 0, \\ 1, & t = 0, \end{cases}$$

che non è una f.c.. \blacksquare

Esempio 3.6.2. Sia F_n la f.r. della legge esponenziale di parametro $1/n$

$$F_n(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-x/n}, & x \geq 0. \end{cases}$$

Allora $\lim_{n \rightarrow +\infty} F_n(x) = 0$ per ogni x di \mathbb{R} . Ora $\varphi_n(t) = (1 - int)^{-1}$ e dunque

$$\lim_{n \rightarrow +\infty} \varphi_n(t) = \varphi(t) := \begin{cases} 0, & t \neq 0, \\ 1, & t = 0, \end{cases}$$

che non è una f.c.. \blacksquare

3.7 Individuazione delle f.c.

Si pone la questione di riconoscere se un'assegnata funzione $\varphi : \mathbb{R} \rightarrow \mathbb{C}$ sia la f.c. di una v.a., o di una legge di probabilità.

Definizione 3.7.1. Una funzione $f : \mathbb{R} \rightarrow \mathbb{C}$ si dice *semidefinita positiva*, se si ha

$$\sum_{j=1}^n \sum_{k=1}^n f(t_j - t_k) z_j \bar{z}_k \geq 0,$$

per ogni scelta di n in \mathbb{N} , di n numeri reali t_1, t_2, \dots, t_n , e di n numeri complessi z_1, z_2, \dots, z_n . \diamond

Teorema 3.7.1. (Teorema di Bochner). *Per una funzione $\varphi : \mathbb{R} \rightarrow \mathbb{C}$ continua nell'origine e tale che $\varphi(0) = 1$ sono equivalenti le condizioni:*

- (a) φ è una f.c.;
 (b) φ è semidefinita positiva.

Dimostrazione. (a) \implies (b) Comunque si scelgano un numero naturale n , e, fissato n , n numeri reali t_1, t_2, \dots, t_n e n numeri complessi z_1, z_2, \dots, z_n , è

$$\begin{aligned} \sum_{j=1}^n \sum_{k=1}^n \varphi(t_j - t_k) z_j \bar{z}_k &= \sum_{j=1}^n \sum_{k=1}^n \int_{\mathbb{R}} \exp(i(t_j - t_k)) z_j \bar{z}_k dF(x) \\ &= \int_{\mathbb{R}} \sum_{j=1}^n \sum_{k=1}^n \{z_j \exp(it_j x)\} \overline{\{z_k \exp(it_k x)\}} dF(x) \\ &= \int_{\mathbb{R}} \left| \sum_{j=1}^n z_j \exp(it_j x) \right|^2 dF(x) \geq 0. \end{aligned}$$

(b) \implies (a) Se $f : \mathbb{R} \rightarrow \mathbb{C}$ è continua, allora, per ogni $T > 0$, si ha

$$\int_0^T \int_0^T \varphi(u - v) f(u) \overline{f(v)} du dv \geq 0; \quad (3.7.1)$$

infatti, l'integrale esiste sull'insieme compatto $[0, T] \times [0, T]$ ed è il limite delle somme di Riemann

$$\sum_j \sum_k \varphi(t_j - t_k) f(t_j) \overline{f(t_k)} \mathcal{D}t_j \mathcal{D}t_k,$$

ciascuna delle quali è positiva, perché φ è semidefinita positiva. Scegliendo ora $f(x) := \exp(-iux)$ nella (3.7.1), si ottiene

$$p_T(x) := \frac{1}{2\pi T} \int_0^T \int_0^T \varphi(u - v) \exp\{-i(u - v)x\} du dv \geq 0.$$

Si cambiino le variabili $s = u - v$, $t = v$. Si ottiene così

$$\begin{aligned} p_T(x) &= \frac{1}{2\pi T} \int_{-T}^0 \varphi(s) e^{-isx} ds \int_{-s}^T dt + \frac{1}{2\pi T} \int_0^T \varphi(s) e^{-isx} ds \int_0^{T-s} dt \\ &= \frac{1}{2\pi T} \int_{-T}^0 \varphi(s) e^{-isx} ds \int_{-s}^T dt + \frac{1}{2\pi T} \int_0^T \varphi(s) e^{-isx} ds \int_s^T dt \\ &= \frac{1}{2\pi T} \int_{-T}^T \varphi(s) e^{-isx} ds \int_{|s|}^T dt = \frac{1}{2\pi} \int_{-T}^T e^{-isx} \left(1 - \frac{|s|}{T}\right) \varphi(s) ds \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{-isx} \varphi_T(s) ds, \end{aligned}$$

ove si è posto

$$\varphi_T(t) := \begin{cases} \left[1 - \left(\frac{|t|}{T}\right)\right] \varphi(t), & \text{se } |t| \leq T, \\ 0, & \text{se } |t| > T. \end{cases} \quad (3.7.2)$$

Dimostreremo di seguito che:

- (a) p_T è una densità di probabilità su $(\mathbb{R}, \mathcal{B})$;
- (b) φ_T è, per ogni $T > 0$, una f.c..

Questi due punti bastano per concludere la dimostrazione del teorema; infatti scende dalla (3.7.2) che, per ogni $t \in \mathbb{R}$,

$$\lim_{T \rightarrow +\infty} \varphi_T(t) = \varphi(t);$$

per ipotesi, φ è continua nell'origine, sicché, per il Teorema 3.6.2, anche φ è una f.c..

(a) Poiché già si sa che p_T è positiva e che è continua, e, quindi, misurabile, basta mostrare che è eguale a 1 il suo integrale su \mathbb{R} . Si consideri la successione crescente (ψ_n) di funzioni $\psi_n : \mathbb{R} \rightarrow \mathbb{R}$ ($n \in \mathbb{N}$), definite da

$$\psi_n(x) := \begin{cases} 1 - \frac{|x|}{n}, & \text{se } |x| \leq n, \\ 0, & \text{se } |x| > n. \end{cases}$$

Allora, $\psi_n \uparrow 1$ al tendere di n a $+\infty$ e la funzione $\psi_n \cdot p_T : \mathbb{R} \rightarrow \mathbb{R}$ è continua. Dal teorema di Beppo Levi scende, ricorrendo anche al teorema di Fubini,

$$\begin{aligned} \int_{\mathbb{R}} p_T(x) \, dx &= \lim_{n \rightarrow +\infty} \int_{\mathbb{R}} \psi_n(x) p_T(x) \, dx \\ &= \lim_{n \rightarrow +\infty} \frac{1}{2\pi} \int_{\mathbb{R}} \psi_n(x) \, dx \int_{\mathbb{R}} e^{-itx} \varphi_T(t) \, dt \\ &= \lim_{n \rightarrow +\infty} \int_{\mathbb{R}} \varphi_T(t) \, dt \frac{1}{2\pi} \int_{-n}^n \left(1 - \frac{|x|}{n}\right) e^{-itx} \, dx \\ &= \lim_{n \rightarrow +\infty} \int_{\mathbb{R}} \varphi_T(t) \frac{1}{2\pi} \frac{[\sin(tn/2)]^2}{t^2 n/4} \, dt \\ &= \lim_{n \rightarrow +\infty} \frac{1}{\pi} \int_{\mathbb{R}} \varphi_T\left(\frac{2s}{n}\right) \frac{\sin^2 s}{s^2} \, ds = \frac{1}{\pi} \int_{\mathbb{R}} \varphi_T(0) \frac{\sin^2 s}{s^2} \, ds = 1. \end{aligned}$$

Abbiamo usato il fatto che $\varphi_T(0) = 1$ e il teorema di convergenza dominata, perché la funzione

$$s \mapsto \frac{\sin^2 s}{s^2}$$

è integrabile con l'integrale dato da

$$\int_{\mathbb{R}} \frac{\sin s^2}{s^2} \, ds = \pi,$$

per il quale si vedano gli esercizi.

(b) Per stabilire che φ_T è la f.c. di p_T , si usa l'integrabilità appena stabilita di p_T e il teorema di convergenza dominata

$$\begin{aligned} \int_{\mathbb{R}} e^{itx} p_T(x) \, dx &= \lim_{n \rightarrow +\infty} \int_{\mathbb{R}} e^{itx} \psi_n(x) p_T(x) \, dx \\ &= \lim_{n \rightarrow +\infty} \frac{1}{2\pi} \int_{\mathbb{R}} \psi_n(x) \, dx \int_{\mathbb{R}} e^{-(u-t)x} \varphi_T(u) \, du \\ &= \lim_{n \rightarrow +\infty} \frac{1}{2\pi} \int_{\mathbb{R}} \varphi_T(u) \left\{ \frac{2 \sin[(t-u)n/2]}{(t-u)n} \right\}^2 n \, du \\ &= \lim_{n \rightarrow +\infty} \frac{1}{\pi} \int_{\mathbb{R}} \varphi_T \left(t - \frac{2s}{n} \right) \left(\frac{\sin s}{s} \right)^2 \, ds = \varphi_T(t), \end{aligned}$$

che conclude la dimostrazione. \square

È conseguenza immediata del teorema di Bochner che se φ è una f.c., sono f.c. anche $\overline{\varphi}$, $|\varphi|^2$, $\Re\varphi$. Sempre per il teorema di Bochner, il prodotto di un numero finito di f.c. è ancora una f.c..

Una seconda caratterizzazione delle f.c. è data dal seguente teorema di Cramér, la dimostrazione del quale è, di fatto contenuta nel teorema precedente.

Teorema 3.7.2. *Per una funzione $\varphi : \mathbb{R} \rightarrow \mathbb{C}$ continua, limitata e tale che $\varphi(0) = 1$ sono equivalenti le seguenti proprietà:*

- (a) φ è una f.c.;
- (b) per ogni $T > 0$ e per ogni $x \in \mathbb{R}$, vale

$$\int_0^T ds \int_0^T \varphi(s-t) e^{i(s-t)x} \, dt \geq 0.$$

È talvolta utile il seguente criterio dovuto a Pólya (che non dimostreremo)

Teorema 3.7.3. *Se la funzione limitata $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ soddisfa alle seguenti condizioni:*

- (a) $\varphi(0) = 1$;
- (b) $\varphi(-t) = \varphi(t)$ per ogni $t \in \mathbb{R}$;
- (c) φ è convessa in $]0, +\infty[$;
- (d) $\lim_{t \rightarrow +\infty} \varphi(t) = 0$,

allora φ è la f.c. di una f.r. assolutamente continua.

3.8 Funzione caratteristica di un vettore aleatorio

Anche per i vettori aleatori è possibile definire la funzione caratteristica.

Definizione 3.8.1. Dato un vettore aleatorio $X : \Omega \rightarrow \mathbb{R}^n$ si dice funzione caratteristica di X la funzione $\varphi_X : \mathbb{R}^n \rightarrow \mathbb{C}$ definita, per $t \in \mathbb{R}^n$, da

$$\varphi_X(t) := \mathbb{E} [\exp(i\langle t, X \rangle)] , \quad (3.8.1)$$

dove $\langle a, b \rangle$ denota il prodotto interno dei vettori $a, b \in \mathbb{R}^n$. \diamond

Usando il teorema del cambio di variabili, si vede anche in questo caso, che la f.c. di un vettore aleatorio è, in effetti la f.c. della sua legge (multidimensionale). Infatti

$$\varphi_X(t) = \int_{\mathbb{R}^n} \exp(i\langle t, x \rangle) d\mu_F(x) \quad (t \in \mathbb{R}^n),$$

dove F è la f.r. del vettore X . Pertanto, si parlerà indifferentemente di f.c. di un vettore aleatorio o di una legge di probabilità multidimensionale.

Si supponga che sia dato un v.a. $X = (X_1, X_2, \dots, X_n)$; sia m il vettore delle speranze, cioè $m := (\mathbb{E}(X_1), \mathbb{E}(X_2), \dots, \mathbb{E}(X_n))$, e sia Γ la matrice di varianza-covarianza di X : $\gamma_{ii} = V(X_i)$ e $\gamma_{ij} = \text{Cov}(X_i, X_j)$ se $i \neq j$. Allora, per ogni $t \in \mathbb{R}^n$, la speranza della v.a. $\langle t, X \rangle$ è

$$\mathbb{E}(\langle t, X \rangle) = \sum_{j=1}^n t_j \mathbb{E}(X_j) = \langle t, m \rangle ,$$

mentre la sua varianza è

$$\begin{aligned} V(\langle t, X \rangle) &= \mathbb{E}(\langle t, X \rangle^2) - E^2(\langle t, X \rangle) = \mathbb{E}(\langle t, X \rangle^2) - \langle t, m \rangle^2 \\ &= \sum_{j,k=1}^n t_j t_k \mathbb{E}(X_j X_k) - \sum_{j,k=1}^n t_j t_k \mathbb{E}(X_j) \mathbb{E}(X_k) \\ &= \sum_{j=1}^n t_j^2 V(X_j) - \sum_{j \neq k} t_j t_k \text{Cov}(X_j, X_k) = \langle \Gamma t, t \rangle . \end{aligned}$$

Vale per le f.c. di vettori aleatori lo stesso risultato che per le f.c. di una v.a.: una f.c. individua in modo univoco una legge di probabilità multidimensionale. Del seguente teorema non daremo la dimostrazione.

Teorema 3.8.1. *Le f.c. di due vettori aleatori sono eguali se, e solo se, i due vettori hanno la stessa legge.*

È spesso utile il seguente criterio di indipendenza.

Teorema 3.8.2. *Sia $X = (X_1, \dots, X_n)$ un vettore aleatorio n -dimensionale. Sono allora equivalenti le affermazioni:*

- (a) *le componenti X_1, \dots, X_n di X sono indipendenti;*

(b) le f.c. φ di X e $\varphi_1, \dots, \varphi_n$ di X_1, \dots, X_n , rispettivamente, sono legate dalla relazione

$$\forall t = (t_1, \dots, t_n) \in \mathbb{R}^n \quad \varphi(t) = \prod_{j=1}^n \varphi_j(t_j).$$

Dimostrazione. (a) \implies (b) è

$$\begin{aligned} \varphi(t) &:= \mathbb{E}[\exp(i\langle t, X \rangle)] = \mathbb{E} \left[\exp \left(i \sum_{j=1}^n t_j X_j \right) \right] \\ &= \mathbb{E} \left[\prod_{j=1}^n e^{i t_j X_j} \right] = \prod_{j=1}^n \varphi_j(t_j). \end{aligned}$$

(b) \implies (a) Sia $Y = (Y_1, Y_2, \dots, Y_n)$ un vettore aleatorio n -dimensionale con componenti indipendenti e tale che, per ogni $j = 1, 2, \dots, n$, Y_j abbia la stessa legge, e quindi la stessa f.c., di X_j . Allora i vettori X e Y hanno, per la prima parte di questa dimostrazione, la stessa legge n -dimensionale; dunque le v.a. X_1, X_2, \dots, X_n sono indipendenti. \square

Definizione 3.8.2. Si dice che un vettore $X = (X_1, X_2, \dots, X_n)$ è *gaussiano* o *normale* se è gaussiana, per ogni $t \in \mathbb{R}^n$, la v.a. $\langle t, X \rangle$. \diamond

Teorema 3.8.3. (a) La legge di un vettore gaussiano X è determinata dal vettore delle speranze m e dalla matrice di covarianza Γ ; si scriverà

$$X \sim N_n(m, \Gamma).$$

(b) Se X ha legge $N_n(m, \Gamma)$, se A è una matrice $k \times n$ e se a è in \mathbb{R}^k , allora il vettore aleatorio $Y := AX + a$ ha legge $N_k(Am + a, A\Gamma A^t)$.

Dimostrazione. (a) La f.c. del vettore X è data da

$$\varphi_X(t) = \mathbb{E}[\exp(i\langle t, X \rangle)] = \exp \left(i\langle t, m \rangle - \frac{1}{2} \langle \Gamma t, t \rangle \right).$$

(b) Per quanto visto in (a) si ha, per $t \in \mathbb{R}^k$,

$$\begin{aligned} \varphi_Y(t) &= \mathbb{E}[\exp(i\langle t, Y \rangle)] \\ &= \mathbb{E}[\exp(i\langle t, AX \rangle + i\langle t, a \rangle)] = e^{i\langle t, a \rangle} \mathbb{E} \left(e^{i\langle t, AX \rangle} \right) \\ &= e^{i\langle t, a \rangle} \mathbb{E} \left(e^{i\langle A^T t, X \rangle} \right) \\ &= e^{i\langle t, a \rangle} \exp \left(i\langle A^T t, m \rangle - \frac{1}{2} \langle \Gamma A^T t, A^T t \rangle \right) \\ &= \exp \left(i\langle t, Am + a \rangle - \frac{1}{2} \langle A\Gamma A^T t, t \rangle \right), \end{aligned}$$

che è la f.c. di un vettore con legge $N_k(Am + a, A\Gamma A^t)$. \square

3.9 Note al Capitolo 3

Sezione 3.1 Le funzioni caratteristiche furono apparentemente usate per la prima volta da Lyapunov (1901) nella dimostrazione di un caso del Teorema del limite centrale (si veda il successivo Capitolo 4). Esse furono usate sistematicamente da Lévy in primo luogo per dare una dimostrazione piú semplice di quella originale del teorema di Lindeberg (di nuovo, si veda il successivo Capitolo 4); a tal proposito, si consultino (Lévy, 1922.a e 1922.b). Il riferimento d'obbligo per le f.c. è (Lukacs, 1970). Per le f.c. nell'ambito piú vasto delle trasformate di Fourier si veda, ad esempio, (Stromberg, 1994).

Sezione 3.2 La corrispondenza biunivoca tra funzioni caratteristiche e funzioni di ripartizione fu stabilita da Lévy (1922.c).

Sezione 3.6 Il teorema di continuità fu pubblicato da Lévy (1922.c); a questo teorema è associato anche il nome di Cramér perché la formulazione definitiva del teorema fu data in (Cramér, 1937). Per un'esposizione moderna si veda (Edwards, 1990).

Sezione 3.7 Il teorema di caratterizzazione delle f.c. è dovuto a Bochner (1952).

3.10 Esercizi sul Capitolo 3

3.1. Si ritrovino la media e la varianza delle seguenti leggi ricorrendo alle f.c.:

- (a) binomiale $Bi(n, p)$;
- (b) di Poisson;
- (c) geometrica;
- (d) $N(m, \sigma^2)$;
- (e) gamma $\Gamma(r, \lambda)$;
- (f) binomiale negativa;

3.2. Si calcoli la f.c. della legge uniforme sull'intervallo $(0, 1)$ e si controlli che essa si annulla per $t = 2\pi k$ con $k \in \mathbb{Z}$.

3.3. Si calcoli la f.c. della *legge triangolare* che è individuata dalla densità

$$f(x) := \frac{1}{\alpha} \left(1 - \frac{|x|}{\alpha} \right) \mathbf{1}_{(-\alpha, \alpha)}(x) \quad (\alpha > 0).$$

3.4. Si calcoli la f.c. della legge di Laplace di densità

$$f(x) := \frac{1}{2} \alpha \exp(-\alpha |x|) \quad (\alpha > 0, x \in \mathbb{R});$$

si mostri come si sarebbe potuto usare questo risultato per calcolare la f.c. della legge di Cauchy.

3.5. Si mostri che per ogni f.c. φ e per ogni $t \in \mathbb{R}$ vale

$$4\Re\{1 - \varphi(t)\} \geq \Re\{1 - \varphi(2t)\}.$$

Pertanto, se la successione (φ_n) di f.c. converge a 1 in un intervallo $[-T, T]$ essa converge a 1 su tutto \mathbb{R} .

3.6. Ricorrendo alle f.c. si dimostri il seguente teorema:

Teorema 3.10.1. *Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ siano X_1 e X_2 v.a. indipendenti. Allora*

- (a) *se $X_i \sim N(m_i, \sigma_i^2)$ ($i = 1, 2$), allora $X_1 + X_2 \sim N(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$;*
- (b) *se X_i ha legge binomiale di parametri n_i e p ($i = 1, 2$), allora $X_1 + X_2$ ha legge binomiale di parametri $n_1 + n_2$ e p ; cioè $X_i \sim Bi(n_i, p)$ ($i = 1, 2$) implica $X_1 + X_2 \sim Bi(n_1 + n_2, p)$;*
- (c) *se X_i ha legge di Poisson di parametro λ_i ($i = 1, 2$), cioè $X_i \sim \mathcal{P}(\lambda_i)$, allora $X_1 + X_2$ ha legge di Poisson di parametro $\lambda_1 + \lambda_2$, $X_1 + X_2 \sim \mathcal{P}(\lambda_1 + \lambda_2)$;*
- (d) *se $X_i \sim \Gamma(\alpha_i, \theta)$ ($i = 1, 2$), allora $X_1 + X_2$ ha legge $\Gamma(\alpha_1 + \alpha_2, \theta)$.*

3.7. Se φ_1, φ_2 e φ sono rispettivamente le f.c. delle v.a. X, Y e $X + Y$, si mostri, mediante un esempio, che può accadere che sia $\varphi(t) = \varphi_1(t)\varphi_2(t)$ per ogni $t \in \mathbb{R}$ senza che X e Y siano indipendenti (sicché la condizione (3.5.6) è una condizione necessaria, ma non sufficiente per l'indipendenza).
(Suggerimento: la legge di Cauchy.)

3.8. La f.c. φ di F è reale se, e solo se, F è simmetrica.

3.9. Facendo riferimento alla funzione ψ definita dalla (3.2.1), si mostri che non esiste nel senso di Lebesgue l'integrale

$$\int_{\mathbb{R}_+} \frac{\sin x}{x} dx.$$

3.10. Si calcoli $\int_0^T e^{-ux} \sin x dx$ e si usi il teorema di Fubini per calcolare

$$\int_0^T \frac{\sin x}{x} dx$$

ed ottenere così un'altra dimostrazione della (3.2.2).

3.11. Per il calcolo della f.c. della legge normale $N(0, 1)$ si può procedere, oltre che come nel testo, in altri modi che evitano il ricorso all'analisi complessa. Di seguito ne sono indicati due.

(a) Si osserva che

$$\varphi(t) = \sqrt{\frac{2}{\pi}} \int_0^{+\infty} e^{-x^2/2} \cos tx dx;$$

si ricava mediante derivazione (ed un'integrazione per parti) un'equazione differenziale per φ ; quest'ultima si può risolvere tenendo presente che si conosce la condizione $\varphi(0) = 1$;

(b) si ricorre allo sviluppo in serie di e^{itx} e si integra a termine a termine la serie che si ottiene.

3.12. Se (z_n) è una successione di numeri complessi che converge a $z \in \mathbb{C}$, si calcoli

$$\lim_{n \rightarrow +\infty} \left(1 + \frac{z_n}{n}\right)^n.$$

Questo risultato torna utile in molti casi in cui si vogliono dimostrare teoremi limite ricorrendo alla convergenza di f.c.. Lo si utilizzi per dimostrare direttamente il TLC (si veda il successivo Capitolo 4 per la terminologia) nel caso di una successione (X_n) di v.a. di L^2 indipendenti e isonome, risultato che stabiliremo come conseguenza del Teorema di Lindeberg nel Corollario 4.1.2.

3.13. L'insieme Δ^0 munito della convoluzione $*$ come operazione è un semigrupp commutativo con identità.

3.14. $E := \{\varepsilon_a : a \in \mathbb{R}\}$ è un sottosemigrupp commutativo di $(\Delta^0, *)$, detto *semigrupp di traslazione*.

3.15. Si individuino altri sottosemigruppi di $(\Delta^0, *)$ tenendo conto dell'Esercizio 3.6.

3.16. Se la v.a. X è assolutamente continua, allora la sua f.c. φ soddisfa alla proprietà

$$\lim_{|t| \rightarrow +\infty} \varphi(t) = 0$$

(teorema di Riemann–Lebesgue).

3.17. Le seguenti funzioni sono f.c.:

$$\varphi(t) = \frac{1}{1 + |t|}; \quad (\text{a})$$

$$\varphi(t) = \begin{cases} 1 - |t|, & 0 \leq |t| \leq 1, \\ 0, & |t| > 1; \end{cases} \quad (\text{b})$$

$$\varphi(t) = \begin{cases} 1 - |t|, & 0 \leq |t| \leq 1/2, \\ \frac{1}{4|t|}, & |t| > 1/2; \end{cases} \quad (\text{c})$$

$$\varphi(t) = \frac{1}{1 + |t|^\alpha} \quad (\alpha \in [0, 1]). \quad (\text{d})$$

3.18. Esistono due f.c. φ_1 e φ_2 con la proprietà che $\varphi_1(t) = \varphi_2(t)$ per ogni $t \in [-a, a]$, ove $a > 0$ senza che le corrispondenti f.r. F_1 e F_2 siano eguali (o, ciò che è lo stesso, senza che sia $\varphi_1(t) = \varphi_2(t)$ per ogni $t \in \mathbb{R}$).

3.19. La funzione $t \mapsto e^{-|t|^\alpha}$ con $\alpha \in]0, 1]$ è una f.c..

3.20. (a) Siano (F_n) e (G_n) due successioni di f.r. che convergono completamente alle f.r. F e G rispettivamente. Si mostri allora che $F_n * G_n \xrightarrow[n \rightarrow +\infty]{c} F * G$.

(b) Siano (X_n) e (Y_n) due successioni indipendenti di v.a. strettamente positive definite sullo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$. Se entrambe le successioni convergono in legge a v.a. strettamente positive X e Y , $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X$, $Y_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} Y$, si mostri che $X_n Y_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X Y$ e che $X_n / Y_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X / Y$.

3.21. Siano φ e ψ le f.c. corrispondenti alle f.r. F e G rispettivamente. Allora:

(a)

$$\bar{d}(F, G) := \sup_{t \in \mathbb{R}} \frac{|\varphi(t) - \psi(t)|}{1 + |t|}$$

è una metrica su Δ^0 ;

(b) la topologia di tale metrica è la stessa di quella di Lévy.

3.22. (a) Si mostri, partendo da una delle identità trigonometriche elementari che

$$\frac{\sin t}{t} = \prod_{n=1}^{\infty} \cos \frac{t}{2^n}.$$

(b) Siano (X_n) v.a. indipendenti tali che $\mathbb{P}(X_n = -\alpha) = \mathbb{P}(X_n = \alpha) = 1/2$. Si mostri che

$$Y_n := \sum_{j=1}^n \frac{X_j}{2^j}$$

converge in legge ad una v.a. distribuita uniformemente in $(-\alpha, \alpha)$.

3.23. Sia b un numero naturale con $b \geq 2$ e sia (X_n) una successione di v.a. indipendenti ed isonome che assumono valori nell'insieme

$$\{0, 1, \dots, b-1\}$$

e tali che per ogni $n \in \mathbb{N}$ sia $\mathbb{P}(X_n = j) = 1/b$ ($j = 0, 1, \dots, b-1$). Allora lo sviluppo in base b del numero di $[0, 1]$

$$X := \sum_{n \in \mathbb{N}} \frac{X_n}{b^n}$$

è uniformemente distribuito in $[0, 1]$.

3.24. Sia E un boreliano di \mathbb{R} e sia $(x, \lambda) \rightarrow F(x, \lambda)$ una funzione boreliana in \mathbb{R}^2 , tale che, per ogni $\lambda \in E$, $x \mapsto F_\lambda(x) := F(x, \lambda)$ sia una f.r.. Sia G una f.r. tale che $\mu_G(E) = 1$.

(a) $x \mapsto H(x) := \int_E F_\lambda(x) dG(\lambda)$ è una f.r., detta $(G-)$ mistura della famiglia $\{F_\lambda : \lambda \in E\}$. (La convoluzione è un caso particolare: $n = 1$ e $F_\lambda(x) = F(x - \lambda)$).

(b) Siano μ_H e μ_λ le misure di Borel–Stieltjes generate da H e da F_λ rispettivamente. Se $\varphi \in L^1(\mu_H)$, si mostri che

$$E_H(\varphi) = \int_E E_\lambda(\varphi) dG(\lambda).$$

(c) Se φ_λ è la f.c. di F_λ , allora per la f.c. φ_H di H vale

$$\varphi_H(t) = \int_E \varphi_\lambda(t) dG(\lambda).$$

- (d) In particolare se (φ_n) è una successione di f.c. e se $\sum_{n \in \mathbb{N}} c_n = 1$, con $c_n \geq 0$, allora

$$\sum_{n \in \mathbb{N}} c_n \varphi_n$$

è una f.c..

3.25. Si calcoli la f.c. della v.a. T_k , tempo del k -esimo successo in un processo di Bernoulli (X_n) sullo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e la usi per calcolare la varianza di T_k .

3.26. (a) La relazione $F * F_1 = F * F_2$ non implica $F_1 = F_2$;

(b) se, però, F è la f.r. della legge $N(0, 1)$, allora da $F * F_1 = F * F_2$ scende $F_1 = F_2$.

3.27. Se $E \subseteq \mathbb{R}$ è un semigruppò abeliano rispetto all'addizione, la famiglia

$$\{F_\lambda : \lambda \in E\}$$

si dice *stabile per addizione* se, per ogni $\lambda_1, \lambda_2 \in E$, si ha

$$F_{\lambda_1} * F_{\lambda_2} = F_{\lambda_1 + \lambda_2}.$$

(a) Sia, per $i = 1, 2$,

$$H_i(x) := \int_E F_\lambda(x) dG_i(\lambda)$$

la G_i -mistura della famiglia $\{F_\lambda : \lambda \in E\}$ stabile per addizione. Allora la convoluzione $H_1 * H_2$ è la $(G_1 * G_2)$ -mistura di $\{F_\lambda : \lambda \in E\}$.

(b) Se $E = \mathbb{Z}_+$ e la famiglia $\{F_n : n \in \mathbb{Z}_+\}$ è stabile per addizione, esiste una f.c. φ tale che $\varphi_n = \varphi^n$ per ogni $n \in \mathbb{Z}_+$.

3.28. (a) Siano $f_n : \mathbb{R} \rightarrow \mathbb{C}$ ($n \in \mathbb{N}$) e $f : \mathbb{R} \rightarrow \mathbb{C}$ funzioni equicontinue e si supponga che $f_n \rightarrow f$ uniformemente. Se $x_n \xrightarrow[n \rightarrow +\infty]{} x$, allora

$$\lim_{n \rightarrow +\infty} f_n(x_n) = f(x).$$

(b) Si supponga che $X_n \rightarrow X$ in legge. Se (a_n) e (b_n) sono successioni reali tali che $a_n \rightarrow a$ e $b_n \rightarrow b$, allora

$$a_n X_n + b_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} aX + b.$$

3.29. Siano X_j ($j = 1, 2, \dots, n$) v.a. indipendenti e tutte uniformemente distribuite nell'intervallo $(-1, 1)$. Si calcoli la densità della somma $S_n = \sum_{j=1}^n X_j$.

3.30. Si dimostri che se $\alpha > 0$, la funzione $h : \mathbb{R} \rightarrow \mathbb{R}$ definita da

$$h(x) := \frac{1}{\pi} \frac{1 - \cos \alpha x}{\alpha x^2}$$

è una densità di probabilità e se ne calcoli la f.c..

3.31. Sia X una v.a. nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$. Si mostri che, per ogni $x \in \mathbb{R}$, vale

$$\mathbb{P}(X = x) = \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T e^{-itx} \varphi(t) dt.$$

3.32. Siano X e Y due v.a. indipendenti ed isonome con f.c. comune φ . Siano $x_1, x_2, \dots \in \mathbb{R}$ i punti di probabilità strettamente positiva

$$P_X(\{x_j\}) = \mathbb{P}(X = x_j) > 0.$$

Allora si ha

$$\mathbb{P}(X - Y = 0) = \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T |\varphi(t)|^2 dt; \quad (\text{a})$$

e

$$\mathbb{P}(X - Y = 0) = \sum_{j \in \mathbb{N}} (P_X(\{x_j\}))^2, \quad (\text{b})$$

sicché

$$\lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T |\varphi(t)|^2 dt = \sum_{j \in \mathbb{N}} (P_X(\{x_j\}))^2;$$

(c) se la f.c. φ è in $L^2(\mathbb{R})$, allora P_X attribuisce probabilità nulla ad ogni punto di \mathbb{R} ($\forall x \in \mathbb{R} \quad P_X(\{x\}) = 0$).

3.33. Si determini la f.r. la cui f.c. è data da

$$\varphi(t) = C \sum_{j=2}^{\infty} \frac{\cos jt}{j^2 \ln j}$$

e si usi questa f.c. per mostrare che non è vero il reciproco del Teorema 3.4.2. (Occorre, naturalmente, specificare il valore della costante C).

3.34. Si mostri che per ogni t reale e per ogni $n \in \mathbb{Z}_+$ vale

$$\left| e^{it} - \sum_{j=0}^n \frac{(it)^j}{j!} \right| \leq \frac{|t|^{n+1}}{(n+1)!} \wedge \frac{2|t|^n}{n!}.$$

Se φ è la f.c. di una v.a. X , allora

$$\left| \varphi(t) - \sum_{j=0}^n \frac{(it)^j \mathbb{E}(X^j)}{j!} \right| \leq \mathbb{E} \left(\frac{|tX|^{n+1}}{(n+1)!} \wedge \frac{2|tX|^n}{n!} \right).$$

3.35. Le condizioni che seguono sono equivalenti per la successione (μ_n) dei momenti di una legge

(a) esiste $t > 0$ tale che

$$\sum_{n=1}^{\infty} \frac{\mu_{2n} t^{2n}}{(2n)!} < +\infty,$$

(b) esiste $t > 0$ tale che

$$\lim_{n \rightarrow +\infty} \frac{\mu_{2n} t^{2n}}{(2n)!} = 0,$$

(c) $\limsup_{n \rightarrow +\infty} \frac{1}{2n} \mu_{2n}^{\frac{1}{2n}} = A < +\infty$.

3.36. Se la successione dei momenti $(\mu_n)_{n \in \mathbb{Z}_+}$, con $\mu_0 = 1$, soddisfa a una delle condizioni equivalenti dell'esercizio precedente, essa individua una sola legge di probabilità. È noto che, se per $|t| \leq t_0$ esiste la funzione generatrice dei momenti della v.a. X ,

$$\psi(t) := \mathbb{E}(e^{tX}),$$

allora esistono finiti i momenti di ogni ordine $\mu_n := \mathbb{E}(X^n)$. Si dimostri che valgono le relazioni:

$$\psi_X(t) = \sum_{n=0}^{\infty} \mu_n \frac{t^n}{n!} \quad (|t| \leq t_0) \quad (3.10.1)$$

$$\mu_n = \psi_X^{(n)}(0) = \left[\frac{d^n \psi_X(t)}{dt^n} \right]_{t=0} \quad (3.10.2)$$

Inoltre la funzione generatrice dei momenti individua la legge di X , nel senso che, se X e Y sono due v.a. per le quali esiste $t_0 > 0$ tale che $\psi_X(t) = \psi_Y(t)$ per $|t| \leq t_0$, allora X e Y hanno la stessa legge.

3.37. Se per calcolare la f.c. della legge gamma si vuole evitare l'integrazione nel campo complesso, si sviluppi in serie la funzione $x \mapsto e^{it}$.

3.38. Sia X_p una v.a. con legge geometrica di parametro p . Per $\lambda > 0$ si consideri la successione $Y_n := \frac{1}{n} X_{\lambda/n}$. Si mostri che Y_n converge in legge a una v.a. esponenziale di parametro λ . Si studii anche la convergenza in legge delle v.a. definite, per ogni $n \in \mathbb{N}$, da

$$V_n := \frac{1}{n} X_p \quad \text{e da} \quad Z_n := \frac{1}{n} X_{p_n},$$

dando, per quest'ultima successione, una condizione sufficiente sulla successione (p_n) affinché Z_n converga in legge ad una v.a. non costante.

3.39. Siano (X_n) e (Y_n) due successioni di v.a. che convergono in legge a X e Y rispettivamente e tali che, per ogni $n \in \mathbb{N}$, X_n è indipendente da Y_n ; inoltre X è indipendente da Y . Si mostri che $X_n + Y_n$ tende in legge a $X + Y$. (è istruttivo paragonare questo risultato a quanto visto nell'esercizio 2.27).

3.40. Sia G una f.r. Sono equivalenti le affermazioni:

(a) $G = \varepsilon_0$;

(b) $F * G = F$ per ogni f.r. F .

3.41. Siano X e Y due v.a., definite sul medesimo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, indipendenti, isonome con f.r. comune F e in L^2 . Per ogni coppia (α, β) di numeri reali tali che $\alpha^2 + \beta^2 = 1$, anche la v.a. $\alpha X + \beta Y$ ha f.r. F .

- (a) Siano X_1, X_2, \dots, X_n v.a. indipendenti tutte con f.r. F . Si trovi una costante α_n tale che la v.a.

$$\alpha_n \sum_{j=1}^n X_j$$

abbia f.r. F .

- (b) Si determini F .

3.42. Sia (X_n) una successione di v.a. indipendenti e isonome, con legge individuata dalla f.c. di Linnik

$$\varphi_\alpha(t) := \frac{1}{1 + |t|^\alpha} \quad (\alpha \in]0, 1]).$$

- (a) Si studii la convergenza in legge delle v.a.

$$T_n := \frac{1}{n^{1/\alpha}} \sum_{j=1}^n X_j.$$

- (b) Sia N una v.a. di legge geometrica di parametro p , indipendente da quelle della successione (X_n) . La v.a.

$$Z_N := p^{1/\alpha} \sum_{j=1}^N X_j$$

ha legge di Linnik.

3.43. Sia (X_n) una successione di v.a. indipendenti tali che per ogni $j \in \mathbb{N}$ sia $X_j \sim \Gamma(1, c^j)$; in altre parole per ogni indice j la v.a. X_j sia esponenziale di parametro c^j . Si consideri la v.a.

$$Y_n := X_1 \wedge X_2 \wedge \dots \wedge X_n.$$

Allora

- (a) si determini la legge di Y_n ;
 (b) si studi la convergenza di (Y_n) .

3.44. Siano X e Y due v.a. indipendenti ed isonome centrate, entrambe con varianza eguale a 1, $\mathbb{E}(X) = \mathbb{E}(Y) = 0$, $\mathbb{E}(X^2) = \mathbb{E}(Y^2) = 1$. Se le v.a. $U := X + Y$ e $V := X - Y$ sono indipendenti, allora X , e quindi Y , ha legge $N(0, 1)$.

3.45. In uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia (X_n) una successione di v.a. indipendenti e isonome con legge esponenziale $\Gamma(1, \lambda)$ e sia, al solito,

$$S_n := \sum_{j=1}^n X_j, \quad S_0 := 0.$$

per $n \in \mathbb{Z}_+$ e per $t > 0$ si ponga

$$N_t := \sum_{j=1}^n \mathbf{1}_{\{S_j \leq t\}}$$

(a) Si determini la legge di N_t ;

(b) Se

$$T_t := (t' - t) \mathbf{1}_{\cup_{n \in \mathbb{Z}_+} \{S_n \leq t < S_{n+1} = t'\}},$$

la v.a. T_t ha la stessa legge di X_1 .

3.46. Siano X e Y due v.a. indipendenti con f.r. F_1 e F_2 rispettivamente. Si trovi la f.c. del prodotto XY . Si usi quindi tale risultato per provare che se (Y_n) è una successione di v.a. indipendenti da X che converge in legge a Y , allora (XY_n) converge in legge a XY .

3.47. In uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia data una v.a. X e si consideri il sottoinsieme J di \mathbb{R} definito da

$$J := \left\{ t \in \mathbb{R} : \int_{\Omega} e^{tX} d\mathbb{P} < +\infty \right\}.$$

J è non vuoto perché, per ogni v.a. X , $0 \in J$. Se J contiene un intorno dell'origine, allora si definisce la *funzione generatrice dei momenti* della v.a. X , $\psi : J \rightarrow \mathbb{R}_+$ mediante

$$\psi(t) := \mathbb{E}(e^{tX}) = \int e^{tX} d\mathbb{P} = \int_{\mathbb{R}} e^{tx} dF(x),$$

ove F è la f.r. di X . L'insieme J si dice *dominio proprio* della funzione generatrice dei momenti.

(a) Si mostri che se appartengono a J i punti $t_0 > 0$ e $-t_0$, allora J contiene l'intervallo $[-t_0, t_0]$;

(b) J è convesso;

(c) La funzione $g : J \rightarrow \mathbb{R}$ definita da $g(t) := \ln \psi(t)$ è convessa;

(d) la funzione generatrice dei momenti ψ è analitica in J , vale a dire che per ogni $t \in J$ vale

$$\psi(t) = \sum_{n \in \mathbb{Z}_+} \alpha_n \frac{t^n}{n!};$$

inoltre $\alpha_0 = 1$ e, per $n \in \mathbb{N}$

$$\alpha_n = \psi^{(n)}(0) = \int X^n d\mathbb{P},$$

sicché X ha momenti di tutti gli ordini;

(e) si dia l'esempio di una v.a. per la quale $J = \{0\}$;

(f) la condizione che J contenga un intervallo è essenziale; si trovi, ad esempio, l'insieme J per la v.a. X che ha densità data da

$$f(x) := \frac{1}{x^2} \mathbf{1}_{(1, +\infty)}(x);$$

(g) sono isonome due v.a. X e Y che hanno la stessa funzione generatrice dei momenti, $\psi_X = \psi_Y$.

3.48. Sia X una v.a. di L^1 con speranza eguale a m , sia ψ la sua funzione generatrice dei momenti e J il suo dominio proprio. Supposto che sia $J \neq \{0\}$, se $a \geq m$, vale la diseguaglianza dovuta a Cramér

$$\mathbb{P}(X \geq a) \leq \exp(-\inf\{at - \ln \psi(t) : t \geq 0, t \in J^0\}).$$

3.49. Si controlli che, per $t \neq 0$ non è integrabile la funzione

$$[2, +\infty[\ni x \mapsto \frac{\sin tx}{x \ln x}.$$

3.50. Si calcoli l'integrale

$$\int_{\mathbb{R}} \frac{\sin^2 s}{s^2} ds = \pi.$$

3.51. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, sia (X_n) una successione di v.a. indipendenti, tutte di legge $N(0, \sigma^2)$. Se θ è un numero reale, si definisca una nuova successione (Y_n) mediante

$$Y_1 := X_1, \quad Y_n := \theta Y_{n-1} + X_n \text{ se } n \geq 2.$$

(a) Si calcoli $\mathbb{E}(Y_n)$ ($n \in \mathbb{N}$);

(b) si calcoli $\text{Cov}(Y_n, Y_{n+1})$ ($n \in \mathbb{N}$);

(c) si determini la legge di Y_n ($n \in \mathbb{N}$);

(d) si studi la convergenza in legge della successione (Y_n) .

3.52. Dalla f.c. del vettore aleatorio $X = (X_1, X_2, \dots, X_n)$ si derivino le f.c. marginali delle componenti. In particolare si mostri che le componenti di un vettore aleatorio normale gaussiano sono ancora gaussiane.

3.53. Le v.a. X_n ($n \in \mathbb{N}$) e X abbiano f.c. φ_n e φ integrabili. Allora, tanto le v.a. X_n quanto X sono assolutamente continue; siano f_n e f le loro densità. Se è

$$\int_{\mathbb{R}} |\varphi_n(t) - \varphi(t)| dt \xrightarrow{n \rightarrow +\infty} 0,$$

allora

$$f_n(x) \xrightarrow{n \rightarrow +\infty} f(x) \quad \text{per ogni } x \in \mathbb{R}.$$

e (X_n) converge in legge a X .

3.54. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ per la successione

$$(X_{nk})_{n \in \mathbb{N}; k=1,2,\dots,n}$$

sono equivalenti le affermazioni:

$$\max_{k \leq n} P(|X_{nk}| \geq \varepsilon) \xrightarrow{n \rightarrow +\infty} 0, \quad (\text{a})$$

$$\max_{k \leq n} \int_{\Omega} \frac{X_{nk}^2}{1 + X_{nk}^2} d\mathbb{P} = \max_{k \leq n} \int_{\mathbb{R}} \frac{x^2}{1 + x^2} dF_{nk}(x) \xrightarrow{n \rightarrow +\infty} 0. \quad (\text{b})$$

Capitolo 4

Teoremi Limite

Una parte importante del Calcolo delle Probabilità è costituita da teoremi che studiano il limite di successioni di v.a.. Un gran numero di questi teoremi si può classificare in due grandi famiglie: Teoremi del Limite Centrale (TLC) e Leggi dei Grandi Numeri (LGN). Accanto a questi vi è poi lo studio della velocità di convergenza e delle oscillazioni. In queste lezioni ci limiteremo a dare alcuni esempi importanti di TLC e di LGN.

4.1 TLC: condizioni sufficienti

Il TLC studia la convergenza in legge di una successione di v.a. ad una v.a. (che può essere fittizia) che abbia legge $N(0, 1)$. Il TLC non è affatto un teorema, bensì piuttosto una classe di teoremi. Un esempio di TLC si è già incontrato nel corso introduttivo a proposito del teorema di de Moivre–Laplace. Si considererà qui il solo caso di una successione di v.a. indipendenti. Per una trattazione completa di questo problema nelle condizioni nelle quali ci siamo posti, si veda il libro di Loève citato in Bibliografia. Sia $(X_n)_{n \in \mathbb{N}}$ una successione di v.a. indipendenti e appartenenti a L^2 . Si ponga, per comodità di notazione,

$$m_j := \mathbb{E}(X_j), \quad \sigma_j^2 := V(X_j) \quad (j \leq n), \quad S_n := \sum_{j=1}^n X_j$$

onde

$$\mathbb{E}(S_n) = \sum_{j=1}^n m_j \quad \text{e} \quad c_n^2 := V(S_n) = \sum_{j=1}^n \sigma_j^2.$$

Si consideri ora la v.a.

$$T_n := \frac{S_n - \mathbb{E}(S_n)}{c_n} \quad (n \in \mathbb{N}),$$

che ha speranza e varianza rispettivamente eguali a 0 e a 1.

Definizione 4.1.1. Se X^* è una v.a. con legge $N(0, 1)$, si dice che la successione (X_n) obbedisce al TLC se la successione (T_n) converge a X^* in legge. \diamond

Il ruolo della v.a. X^* è del tutto marginale, tanto che spesso si scrive

$$T_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} N(0, 1).$$

Si cercheranno dapprima condizioni che assicurino la convergenza. Il risultato che segue è fondamentale.

Teorema 4.1.1. (Lindeberg). *Sia $(X_n)_{n \in \mathbb{N}}$ una successione di v.a. indipendenti di L^2 . Se, con le notazioni appena introdotte, è verificata la condizione di Lindeberg*

$$\lim_{n \rightarrow +\infty} \frac{1}{c_n^2} \sum_{j=1}^n \int_{\{|X_j - m_j| \geq \varepsilon c_n\}} (X_j - m_j)^2 d\mathbb{P} = 0, \quad (4.1.1)$$

allora la successione (T_n) converge in legge alla distribuzione $N(0, 1)$, vale a dire che la successione (X_n) obbedisce al TLC.

Si osservi che la condizione di Lindeberg (4.1.1) si può scrivere nella forma equivalente

$$\lim_{n \rightarrow +\infty} \frac{1}{c_n^2} \sum_{j=1}^n \int_{\{|x - m_j| \geq \varepsilon c_n\}} (x - m_j)^2 dF_j(x) = 0.$$

Prima di dare la dimostrazione di questo teorema, conviene metterne in luce alcune conseguenze particolarmente importanti nelle applicazioni; ciò sarà fatto in una serie di corollari.

Corollario 4.1.1. *Sia (X_n) una successione di v.a. indipendenti uniformemente limitate, cioè per ogni $n \in \mathbb{N}$ si ha $|X_n| \leq H$, e sia $\lim_{n \rightarrow +\infty} c_n^2 = +\infty$. Allora (X_n) obbedisce al TLC.*

Dimostrazione. Per ogni $j \leq n$ si ha, usando la disuguaglianza di Čebyšev,

$$\int_{\{|X_j - m_j| \geq \varepsilon c_n\}} (X_j - m_j)^2 d\mathbb{P} \leq 4H^2 \mathbb{P}(|X_j - m_j| \geq \varepsilon c_n) \leq \frac{4H^2 \sigma_j^2}{\varepsilon^2 c_n^2},$$

onde

$$0 \leq \frac{1}{c_n^2} \sum_{j=1}^n \int_{\{|X_j - m_j| \geq \varepsilon c_n\}} (X_j - m_j)^2 d\mathbb{P} \leq \frac{1}{c_n^2} \sum_{j=1}^n \frac{4H^2 \sigma_j^2}{\varepsilon^2 c_n^2} = \frac{4H^2}{\varepsilon^2 c_n^2},$$

sicché la condizione di Lindeberg (4.1.1) è verificata. \square

Corollario 4.1.2. *Sia (X_n) una successione di v.a. di L^2 indipendenti ed isonome. Allora (X_n) obbedisce al TLC.*

Dimostrazione. Poiché $\mathbb{E}(X_j) = m$ e $V(X_j) = \sigma^2$ per ogni $j \in \mathbb{N}$, si ha, indicando con X una qualsiasi v.a. della successione e con F la f.r. comune,

$$\begin{aligned} & \frac{1}{c_n^2} \sum_{j=1}^n \int_{\{|X_j - m_j| \geq \varepsilon c_n\}} (X_j - m_j)^2 d\mathbb{P} \\ &= \frac{1}{n\sigma^2} \sum_{j=1}^n \int_{\{|x - m| \geq \varepsilon \sigma \sqrt{n}\}} (x - m)^2 dF(x) \\ &= \frac{1}{\sigma^2} \int_{\{|x - m| \geq \varepsilon \sigma \sqrt{n}\}} (x - m)^2 dF(x). \end{aligned}$$

Posto $A_n := \{|X - m| \geq \varepsilon \sigma \sqrt{n}\}$, la successione (A_n) è decrescente e ha come limite l'intersezione $\bigcap_{n \in \mathbb{N}} A_n$; ora

$$\mathbb{P} \left(\bigcap_{n \in \mathbb{N}} A_n \right) = \mathbb{P}(\{\omega \in \Omega : |X(\omega)| = +\infty\}) = 0,$$

ma allora, ricordando che $(X - m)^2 \cdot \mathbb{P}$ è una misura finita, perché X appartiene a L^2 (e, dunque, anche a L^1), è

$$\lim_{n \rightarrow +\infty} \int_{A_n} (X - m)^2 \, d\mathbb{P} = 0,$$

che conclude la dimostrazione. \square

Il Corollario che abbiamo appena dimostrato riguarda il caso che in letteratura è solitamente indicato come il caso i.i.d., vale a dire di v.a. indipendenti ed identicamente distribuite.

Si osservi che il Corollario 4.1.2, ma anche 4.1.1, si applica, in particolare, al caso delle prove bernoulliane. Sia, infatti, S_n il numero di successi in n prove indipendenti. Posto $\sigma^2 := V(X_j) = pq$, si ha $\mathbb{E}(S_n) = np$, $c_n^2 = npq$, onde

$$T_n = \frac{S_n - np}{\sqrt{npq}};$$

si ritrova così il teorema integrale di de Moivre–Laplace.

Corollario 4.1.3. (Teorema di Lyapunov). *Se $\delta > 0$, sia (X_n) una successione di v.a. indipendenti di $L^{2+\delta}$. Se è verificata la condizione di Lyapunov*

$$\lim_{n \rightarrow +\infty} \frac{1}{c_n^{2+\delta}} \sum_{j=1}^n \mathbb{E}(|X_j - m_j|^{2+\delta}) = 0, \quad (4.1.2)$$

allora (X_n) obbedisce al TLC.

Dimostrazione. Si ha

$$\begin{aligned} \mathbb{E}(|X_j - m_j|^{2+\delta}) &= \int |X_j - m_j|^{2+\delta} \, d\mathbb{P} \\ &\geq \int_{\{|X_j - m_j| \geq \varepsilon c_n\}} |X_j - m_j|^{2+\delta} \, d\mathbb{P} \geq \varepsilon^\delta c_n^\delta \int_{\{|X_j - m_j| \geq \varepsilon c_n\}} (X_j - m_j)^2 \, d\mathbb{P}, \end{aligned}$$

sicché

$$\frac{1}{c_n^2} \sum_{j=1}^n \int_{\{|X_j - m_j| \geq \varepsilon c_n\}} (X_j - m_j)^2 \, d\mathbb{P} \leq \frac{1}{\varepsilon^\delta c_n^{2+\delta}} \sum_{j=1}^n \mathbb{E}(|X_j - m_j|^{2+\delta}).$$

L'asserto segue ora dalla (4.1.1) e dalla (4.1.2). \square

Dimostrazione del Teorema 4.1.1. Senza perdita di generalità, si può supporre che sia $m_j = 0$ per ogni $j \in \mathbb{N}$. Infatti, se il teorema è dimostrato con quest'ultima restrizione, si ponga $Y_j := X_j - m_j$ onde $\mathbb{E}(Y_j) = 0$ ($j \in \mathbb{N}$), $S'_n := \sum_{j=1}^n Y_j$ ($n \in \mathbb{N}$); ma allora si ha $T'_n = T_n$. Poiché

$$\int_{\{|X_j - m_j| \geq \varepsilon c_n\}} (X_j - m_j)^2 d\mathbb{P} = \int_{\{|Y_j| \geq \varepsilon c_n\}} Y_j^2 d\mathbb{P},$$

la condizione di Lindeberg vale per le successioni (Y_n) e (X_n) , onde sia T'_n sia T_n convergono in legge a $N(0, 1)$.

Sarà opportuno usare le seguenti relazioni, che sono facili conseguenze degli sviluppi in serie, nel campo complesso, delle funzioni in questione (si vedano gli esercizi)

$$e^{iy} = 1 + iy + \theta(y) \frac{y^2}{2} \quad (y \in \mathbb{R}, |\theta(y)| \leq 1), \quad (4.1.3)$$

$$e^{iy} = 1 + iy - \frac{y^2}{2} + \theta_1(y) \frac{|y|^3}{6} \quad (y \in \mathbb{R}, |\theta_1(y)| \leq 1), \quad (4.1.4)$$

$$\text{Log}(1+z) = z + \theta_2(z)|z|^2 \quad (z \in \mathbb{C}, |z| \leq 1/2, |\theta_2(z)| \leq 1), \quad (4.1.5)$$

ove, mediante Log, si è indicato il ramo principale del logaritmo. È noto, infatti, che, poiché la funzione $z \mapsto e^{iz}$ è periodica di periodo $i\pi$, un numero complesso $z = \rho e^{i\theta}$ ha come logaritmo, nel campo complesso, $\ln z = \ln \rho + i(\theta + n\pi)$, ove $n \in \mathbb{Z}$, sicché il logaritmo non è una funzione nel campo complesso. Per far sí che si possa definire il logaritmo di un numero complesso come funzione si considera *ramo principale* del logaritmo che corrisponde alla scelta del valore $n = 0$ nell'ultima formula; in questo modo, $\text{Log} 1 = 0$. Nello studio delle funzioni caratteristiche si sa che ogni funzione caratteristica φ è tale che $\varphi(0) = 1$; si impone, quindi, che il suo logaritmo si annulli nell'origine, $\ln \varphi(0) = 0$; ciò equivale a scegliere il ramo principale del logaritmo.

Se φ_j è la f.c. di X_j , la f.c. di T_n è

$$\begin{aligned} \varphi_{T_n}(t) &= \mathbb{E}(\exp(itT_n)) = \mathbb{E}(\exp(itS_n)/c_n) \\ &= \varphi_{S_n}\left(\frac{t}{c_n}\right) = \prod_{j=1}^n \varphi_j\left(\frac{t}{c_n}\right). \end{aligned}$$

Ora, per la (4.1.3) e per la (4.1.4), risulta, essendo F_j la f.r. di X_j ,

$$\begin{aligned} \varphi_j(t) &= \int_{\mathbb{R}} \exp(itx) dF_j(x) \\ &= \int_{\{|x| \geq \varepsilon c_n\}} \left(1 + itx + \theta \frac{t^2 x^2}{2}\right) dF_j(x) \\ &\quad + \int_{\{|x| < \varepsilon c_n\}} \left(1 + itx - \frac{t^2 x^2}{2} + \theta_1 \frac{|t|^3 |x|^3}{6}\right) dF_j(x). \end{aligned}$$

Si noti che, benché nell'ultimo integrale compaia la potenza $|x|^3$, e benché non si sia supposto che Y_j appartenga a L^3 , l'integrale esiste finito perché esso è esteso

all'insieme $\{|x| < \varepsilon c_n\}$, nel quale l'integrando è limitato. Poiché

$$\int_{\mathbb{R}} itx \, dF_j(x) = it \mathbb{E}(X_j) = 0,$$

si ha

$$\begin{aligned} \varphi_j\left(\frac{t}{c_n}\right) &= 1 + \left(\frac{t^2}{2c_n^2}\right) \int_{\{|x| \geq \varepsilon c_n\}} \theta x^2 \, dF_j(x) \\ &\quad - \left(\frac{t^2}{2c_n^2}\right) \int_{\{|x| < \varepsilon c_n\}} x^2 \, dF_j(x) + \left(\frac{|t|^3}{6c_n^3}\right) \int_{\{|x| < \varepsilon c_n\}} \theta_1 |x|^3 \, dF_j(x). \end{aligned} \quad (4.1.6)$$

Ora

$$\left| \frac{1}{2} \int_{\{|x| \geq \varepsilon c_n\}} \theta x^2 \, dF_j(x) \right| \leq \frac{1}{2} \int_{\{|x| \geq \varepsilon c_n\}} x^2 \, dF_j(x)$$

e perciò esiste $\theta_3 \in \mathbb{C}$ tale che $|\theta_3| \leq \frac{1}{2}$ e che

$$\frac{1}{2} \int_{\{|x| \geq \varepsilon c_n\}} \theta x^2 \, dF_j(x) = \theta_3 \int_{\{|x| \geq \varepsilon c_n\}} x^2 \, dF_j(x). \quad (4.1.7)$$

Analogamente,

$$\begin{aligned} \left| \frac{1}{6} \int_{\{|x| < \varepsilon c_n\}} \theta_1 |x|^3 \, dF_j(x) \right| &\leq \frac{1}{6} \int_{\{|x| < \varepsilon c_n\}} |x|^3 \, dF_j(x) \\ &\leq \frac{\varepsilon c_n}{6} \int_{\{|x| < \varepsilon c_n\}} |x|^2 \, dF_j(x); \end{aligned}$$

esiste perciò una costante $\theta_4 \in \mathbb{C}$ tale che $|\theta_4| \leq 1/6$ e che

$$\frac{1}{6} \int_{\{|x| < \varepsilon c_n\}} \theta_1 |x|^3 \, dF_j(x) = \theta_4 \varepsilon c_n \int_{\{|x| < \varepsilon c_n\}} x^2 \, dF_j(x). \quad (4.1.8)$$

Pertanto, posto per $n \in \mathbb{N}$ e per $j \leq n$,

$$\alpha_{nj} := \frac{1}{c_n^2} \int_{\{|x| \geq \varepsilon c_n\}} x^2 \, dF_j(x), \quad (4.1.9)$$

$$\beta_{nj} := \frac{1}{c_n^2} \int_{\{|x| < \varepsilon c_n\}} x^2 \, dF_j(x), \quad (4.1.10)$$

l'equazione (4.1.6) diviene $\varphi_j(t/c_n) = 1 + \gamma_{nj}$, ove, in virtù delle (4.1.7), (4.1.8), (4.1.9) e (4.1.10),

$$\gamma_{nj} := \theta_3 t^2 \alpha_{nj} - \frac{1}{2} t^2 \beta_{nj} + |t|^3 \varepsilon \theta_4 \beta_{nj}. \quad (4.1.11)$$

Alla luce della (4.1.9), la condizione di Lindeberg (4.1.1) si legge

$$\lim_{n \rightarrow +\infty} \sum_{j=1}^n \alpha_{nj} = 0.$$

Inoltre è

$$\sum_{j=1}^n (\alpha_{nj} + \beta_{nj}) = 1,$$

sicché l'ipotesi (4.1.1) dà

$$\lim_{n \rightarrow +\infty} \sum_{j=1}^n \beta_{nj} = 1.$$

Poiché $\beta_{nj} \leq \varepsilon^2$, segue dalla (4.1.11) che, per $\varepsilon < \sqrt{6}$, e per n abbastanza grande, diciamo per $n \geq n_0$,

$$\begin{aligned} \max\{|\gamma_{nj}| : j \leq n\} &= \max_{j \leq n} \left| \theta_3 t^2 \alpha_{nj} - \frac{\beta_{nj} t^2}{2} + |t|^3 \varepsilon \theta_4 \beta_{nj} \right| \\ &\leq \max_{j \leq n} \left(\frac{t^2 \alpha_{nj}}{2} + \frac{t^2 \beta_{nj}}{2} + \frac{|t|^3 \varepsilon^3}{6} \right) \leq \varepsilon^2 t^2 + \varepsilon |t|^3 \end{aligned}$$

e

$$\sum_{j=1}^n |\gamma_{nj}| \leq \sum_{j=1}^n \left(\frac{t^2 \alpha_{nj}}{2} + \frac{t^2 \beta_{nj}}{2} \right) + \sum_{j=1}^n \frac{|t|^3 \varepsilon \beta_{nj}}{6} \leq \frac{t^2}{2} + \varepsilon |t|^3.$$

Scende ora dalla (4.1.5) che

$$\begin{aligned} \frac{t^2}{2} + \text{Log } \varphi_{S_n} \left(\frac{t}{c_n} \right) &= \frac{t^2}{2} + \sum_{j=1}^n \text{Log } \varphi_j \left(\frac{t}{c_n} \right) \\ &= \frac{t^2}{2} + \sum_{j=1}^n \text{Log} (1 + \gamma_{nj}) = \frac{t^2}{2} + \sum_{j=1}^n (\gamma_{nj} + \theta_2 |\gamma_{nj}|^2), \end{aligned}$$

e dalla (4.1.11) che

$$\lim_{n \rightarrow +\infty} \sum_{j=1}^n \gamma_{nj} = -\frac{1}{2} t^2;$$

e, poiché

$$\sum_{j=1}^n |\gamma_{nj}|^2 \leq \left(\max_{j \leq n} |\gamma_{nj}| \right) \sum_{j=1}^n |\gamma_{nj}|,$$

segue che, scelto $\delta > 0$ arbitrariamente, si può determinare, subordinatamente a δ e a t , ε sufficientemente piccolo perché risulti, per $n \geq n_0$,

$$\left| \frac{t^2}{2} + \sum_{j=1}^n \text{Log} (1 + \gamma_{nj}) \right| < \delta.$$

Siccome la funzione esponenziale è continua, l'espressione

$$\begin{aligned} \exp\left(\frac{t^2}{2}\right) \prod_{j=1}^n \varphi_j\left(\frac{t}{c_n}\right) &= \exp\left(\frac{t^2}{2}\right) \prod_{j=1}^n (1 + \gamma_{nj}) \\ &= \exp\left(\frac{t^2}{2} + \sum_{j=1}^n \text{Log}(1 + \gamma_{nj})\right) \end{aligned}$$

tende a 1 al tendere di n a $+\infty$, cioè

$$\lim_{n \rightarrow +\infty} \varphi_{T_n}(t) = \exp\left(-\frac{t^2}{2}\right),$$

che è la f.c. di una v.a. con legge $N(0, 1)$. □

Il Corollario 4.1.2 è così importante per le applicazioni, in particolare alla Statistica, che vale la pena darne una dimostrazione diretta. Si supponga, senza perdita di generalità, che le v.a. della successione (X_n) siano centrate, $\mathbb{E}(X_n) = 0$ per ogni $n \in \mathbb{N}$. Allora

$$T_n = \frac{S_n}{\sigma \sqrt{n}}$$

e

$$\varphi_{T_n}(t) = \varphi_{S_n}\left(\frac{t}{\sigma \sqrt{n}}\right) = \prod_{j=1}^n \varphi\left(\frac{t}{\sigma \sqrt{n}}\right) = \varphi^n\left(\frac{t}{\sigma \sqrt{n}}\right),$$

ove φ è la f.c. comune delle v.a. della successione data. Pertanto,

$$\begin{aligned} \text{Log } \varphi_{T_n}(t) &= n \text{Log } \varphi\left(\frac{t}{\sigma \sqrt{n}}\right) \\ &= n \text{Log} \left\{ 1 + \left(\varphi\left(\frac{t}{\sigma \sqrt{n}}\right) - 1 \right) \right\}. \end{aligned} \tag{4.1.12}$$

Poiché X_1 è in L^2 e quindi ammette momento di ordine 2 finito, la f.c. φ è derivabile due volte in un intorno dell'origine, sicché

$$\varphi(t) = \varphi(0) + t \varphi'(0) + \frac{1}{2} t^2 \varphi''(0) + O(t^3).$$

Ma le v.a. della successione sono centrate sicché $\varphi'(0) = 0$; inoltre, $\varphi''(0) = -\sigma^2$, sicché

$$\begin{aligned} \varphi\left(\frac{t}{\sigma \sqrt{n}}\right) &= 1 + \frac{t^2}{2 \sigma^2 n} \varphi''(0) + O\left(\frac{1}{n^{3/2}}\right) \\ &= 1 - \frac{t^2}{2n} + O\left(\frac{1}{n^{3/2}}\right) \end{aligned} \tag{4.1.13}$$

Sostituendo la (4.1.13) nella (4.1.12) si ottiene

$$\text{Log } \varphi_{T_n}(t) = n \text{Log} \left\{ 1 - \frac{t^2}{2n} + O\left(\frac{1}{n^{3/2}}\right) \right\}$$

e di qui, sviluppando in serie il logaritmo,

$$\begin{aligned} \text{Log } \varphi_{T_n}(t) &= n \left\{ -\frac{t^2}{2n} + O\left(\frac{1}{n^{3/2}}\right) \right\} - \frac{n}{2} \left\{ -\frac{t^2}{2n} + O\left(\frac{1}{n^{3/2}}\right) \right\}^2 \\ &\quad + n O\left(\frac{1}{n^{3/2}}\right) \\ &= -\frac{t^2}{2} + O\left(\frac{1}{\sqrt{n}}\right) \xrightarrow{n \rightarrow +\infty} -\frac{t^2}{2}, \end{aligned}$$

cioè l'asserto.

Il TLC si può estendere a "schemi triangolari".

Teorema 4.1.2. *Nello spazio L^2 , per ogni $n \in \mathbb{N}$, siano X_{nj} n v.a. ($j = 1, 2, \dots, n$) indipendenti e centrate, $\mathbb{E}(X_{nj}) = 0$ ($n \in \mathbb{N}$; $j = 1, 2, \dots, n$). Si supponga che sia*

$$\sum_{j=1}^n \mathbb{E}(X_{nj}^2) = \sum_{j=1}^n V(X_{nj}) \xrightarrow{n \rightarrow +\infty} \sigma^2 > 0; \quad (4.1.14)$$

se, per ogni $\varepsilon > 0$, si ha

$$\lim_{n \rightarrow +\infty} \sum_{j=1}^n \int_{\{|X_{nj}| \geq \varepsilon\}} X_{nj}^2 \, d\mathbb{P} = 0, \quad (4.1.15)$$

allora la successione (S_n) con $S_n := \sum_{j=1}^n X_{nj}$ converge in legge a $N(0, \sigma^2)$.

Dimostrazione. Siano φ_{nj} e σ_{nj} rispettivamente la f.c. e la varianza di X_{nj} . È noto (si vedano gli esercizi del Capitolo 6) che, per ogni $\varepsilon > 0$,

$$\begin{aligned} \left| \varphi_{nj}(t) - \left(1 - \frac{1}{2} \sigma_{nj}^2 t^2 \right) \right| &\leq \mathbb{E} \left(\frac{|t|^2 |X_{nj}|^3}{j!} \wedge \frac{2|t|^2 |X_{nj}|^2}{2!} \right) \\ &\leq \frac{|t|^3}{6} \int_{\{|X_{nj}| \leq \varepsilon\}} |X_{nj}|^3 \, d\mathbb{P} + t^2 \int_{\{|X_{nj}| > \varepsilon\}} X_{nj}^2 \, d\mathbb{P}. \end{aligned}$$

Sommando su j e facendo tendere n a $+\infty$ si ottiene, ricorrendo alla (4.1.14) e alla (4.1.15),

$$\limsup_{n \rightarrow +\infty} \sum_{j=1}^n \left| \varphi_{nj}(t) - \left(1 - \frac{1}{2} \sigma_{nj}^2 t^2 \right) \right| \leq \frac{\varepsilon |t|^3 \sigma^2}{6},$$

che, per l'arbitrarietà di ε , dà

$$\sum_{j=1}^n \left| \varphi_{nj}(t) - \left(1 - \frac{1}{2} \sigma_{nj}^2 t^2 \right) \right| \xrightarrow{n \rightarrow +\infty} 0.$$

Si sa dagli esercizi che, se z_1, z_2, \dots, z_n e z'_1, z'_2, \dots, z'_n sono numeri complessi con $|z_j| \leq 1$ e $|z'_j| \leq 1$ ($j = 1, 2, \dots, n$), allora

$$\left| \prod_{j=1}^n z_j - \prod_{j=1}^n z'_j \right| \leq \sum_{j=1}^n |z_j - z'_j|.$$

Quest'ultima disequaglianza dà

$$\left| \prod_{j=1}^n \varphi_{nj}(t) - \prod_{j=1}^n \left(1 - \frac{1}{2} \sigma_{nj}^2 t^2 \right) \right| \xrightarrow{n \rightarrow +\infty} 0. \quad (4.1.16)$$

Si sa ancora (di nuovo, si vedano gli esercizi) che, se a_{nj} ($n \in \mathbb{N}$; $j = 1, 2, \dots, n$) sono numeri complessi tali che

$$\sum_{j=1}^n a_{nj} \xrightarrow{n \rightarrow +\infty} \alpha \quad \text{e} \quad \sum_{j=1}^n |a_{nj}|^2 \xrightarrow{n \rightarrow +\infty} 0, \quad (4.1.17)$$

allora

$$\prod_{j=1}^n (1 - a_{nj}) \xrightarrow{n \rightarrow +\infty} e^{-\alpha}. \quad (4.1.18)$$

Si prenda ora $a_{nj} := t^2 \sigma_{nj}^2 / 2$, sicché la (4.1.14) dà la prima delle (4.1.17) con $\alpha = t \sigma^2 / 2$. Si osservi che

$$\begin{aligned} \sigma_{nj}^2 &= \int_{\{|X_{nj}| \leq \varepsilon\}} \dots + \int_{\{|X_{nj}| > \varepsilon\}} X_{nj}^2 \, d\mathbb{P} \\ &\leq \varepsilon + \int_{\{|X_{nj}| > \varepsilon\}} X_{nj}^2 \, d\mathbb{P}. \end{aligned}$$

Perciò la prima delle (4.1.15) implica, per l'arbitrarietà di ε ,

$$\sup_{j \leq n} \sigma_{nj}^2 \xrightarrow{n \rightarrow +\infty} 0 \quad \text{e} \quad \sum_{j=1}^n \sigma_{nj}^2 \xrightarrow{n \rightarrow +\infty} 0.$$

Pertanto

$$\frac{t^2}{4} \sum_{j=1}^n \sigma_{nj}^2 \leq \frac{t^2}{4} \left(\sup_{j \leq n} \sigma_{nj}^2 \right) \sum_{j=1}^n \sigma_{nj}^2 \xrightarrow{n \rightarrow +\infty} 0,$$

ciò che assicura che sia verificata anche la seconda delle (4.1.17). La (4.1.18) dà dunque

$$\prod_{j=1}^n \left(1 - \frac{1}{2} t^2 \sigma_{nj}^2 \right) \xrightarrow{n \rightarrow +\infty} \exp \left(-\frac{1}{2} \sigma^2 t^2 \right);$$

l'asserto segue ora dalla (4.1.16). \square

4.2 TLC: condizioni necessarie

Prima di mostrare, con un esempio, che la condizione di Lindeberg (4.1.1) non è necessaria per la convergenza in legge a $N(0, 1)$, occorre provare il seguente

Lemma 4.2.1. *Sia $(X_n)_{n \in \mathbb{N}}$ una successione di v.a. indipendenti di L^2 che soddisfaccia alla condizione di Lindeberg (4.1.1). Risulta allora, per ogni $\varepsilon > 0$,*

$$\lim_{n \rightarrow +\infty} \max_{j \leq n} \mathbb{P} \left(\frac{|X_j - m_j|}{c_n} \geq \varepsilon \right) = 0. \quad (4.2.1)$$

Dimostrazione. Poiché una somma di termini positivi è maggiore di ciascuno dei suoi addendi e, dunque, in particolare del maggiore di essi, si ha

$$\begin{aligned} \frac{1}{c_n^2} \sum_{j=1}^n \int_{\{|X_j - m_j| \geq \varepsilon c_n\}} (X_j - m_j)^2 d\mathbb{P} \\ \geq \varepsilon^2 \sum_{j=1}^n \mathbb{P}(|X_j - m_j| \geq \varepsilon c_n) \geq \varepsilon^2 \max_{j \leq n} \mathbb{P}(|X_j - m_j| \geq \varepsilon c_n), \end{aligned}$$

che conclude la dimostrazione. \square

Si può ora costruire una successione di v.a. indipendenti tale che la (4.2.1) sia violata mentre la successione (T_n) converge in legge a $N(0, 1)$; alla luce del lemma appena dimostrato, si ha allora la convergenza di (T_n) a $N(0, 1)$ senza che sia soddisfatta la condizione di Lindeberg.

Esempio 4.2.1. Sia (X_n) una successione di v.a. indipendenti, ognuna delle quali abbia legge normale di speranza nulla e varianza $V(X_n) = \sigma_n^2 = 2^{n-2}$ se $n \geq 2$, mentre $V(X_1) = 1$. Ora, per $n \geq 2$, si ha

$$c_n^2 = \sum_{j=1}^n \sigma_j^2 = 1 + \sum_{j=2}^n 2^{j-2} = 1 + \sum_{j=0}^{n-2} 2^j = 1 + \frac{1 - 2^{n-1}}{1 - 2} = 2^{n-1};$$

dunque, X_n/c_n è una v.a. con legge normale di speranza nulla e varianza data da

$$V\left(\frac{X_n}{c_n}\right) = \frac{V(X_n)}{c_n^2} = \frac{2^{n-2}}{2^{n-1}} = \frac{1}{2}.$$

Pertanto

$$\max_{j \leq n} \mathbb{P}(|X_j| \geq \varepsilon c_n) \geq \mathbb{P}(|X_n| \geq \varepsilon c_n) = \frac{2}{\sqrt{\pi}} \int_{\varepsilon}^{+\infty} e^{-x^2} dx > 0,$$

che è una costante indipendente tanto da j quanto da n . Poiché la (4.2.1) non è verificata, non può esserlo neanche la (4.1.1). D'altra parte, però, se $T_n = S_n/c_n$, si ha

$$\begin{aligned} \varphi_{T_n}(t) &= \varphi_{S_n}\left(\frac{t}{c_n}\right) = \varphi_{S_n}\left(t 2^{-\frac{n-1}{2}}\right) = \prod_{j=1}^n \varphi_j\left(t 2^{-\frac{n-1}{2}}\right) \\ &= \exp\left(-\frac{t^2}{2^n}\right) \prod_{j=2}^n \exp\left(-\frac{2^{j-2} t^2}{2^n}\right) = \exp\left(-\frac{t^2}{2}\right); \end{aligned}$$

perciò T_n ha legge $N(0, 1)$ per ogni $n \in \mathbb{N}$, di modo che (T_n) converge in legge a $N(0, 1)$. \blacksquare

Se, di più, si impone che sia verificata la (4.2.1), la condizione di Lindeberg è necessaria oltreché sufficiente per la convergenza della successione (T_n) alla legge $N(0, 1)$.

Teorema 4.2.1. *Per una successione di v.a. indipendenti di L^2 sono equivalenti le affermazioni:*

- (a) *vale la condizione di Lindeberg (4.1.1);*
- (b) *la successione (T_n) converge in legge a $N(0, 1)$ e, inoltre, è verificata la (4.2.1).*

L'implicazione (a) \implies (b) è conseguenza del teorema 4.1.1 e del Lemma 4.2.1. La dimostrazione dell'implicazione inversa è dovuta a Feller. Occorrerà premetterle alcuni lemmi.

Lemma 4.2.2. (Diseguaglianza di troncamento). *Siano F una f.r. e φ la sua f.c.. Se $t > 0$, esiste $k \in]0, +\infty[$ tale che sia*

$$\int_{\{|x| \geq 1/t\}} dF(x) \leq \frac{k}{t} \int_0^t (1 - \Re\varphi(s)) ds. \quad (4.2.2)$$

Dimostrazione.

$$\begin{aligned} \frac{1}{t} \int_0^t (1 - \Re\varphi(s)) ds &= \frac{1}{t} \int_0^t ds \int_{\mathbb{R}} (1 - \cos sx) dF(x) \\ &= \int_{\mathbb{R}} dF(x) \int_0^t \frac{1 - \cos sx}{t} ds = \int_{\mathbb{R}} \frac{1}{t} \left[s - \frac{\sin sx}{x} \right]_{s=0}^{s=t} dF(x) \\ &= \int_{\mathbb{R}} \left(1 - \frac{\sin tx}{tx} \right) dF(x) \geq \left(\inf_{|t| \geq 1} \left\{ 1 - \frac{\sin t}{t} \right\} \right) \int_{\{|x| \geq 1/t\}} dF(x). \end{aligned}$$

Si controlla immediatamente che

$$\inf_{|t| \geq 1} \left\{ 1 - \frac{\sin t}{t} \right\} = 1 - \sin 1 \simeq 0.158529 > \frac{1}{7},$$

di modo che la (4.2.2) è valida con $k = 7$. □

La proprietà espressa dal prossimo lemma è conosciuta come *trascurabilità asintotica uniforme*.

Lemma 4.2.3. *Siano $(X_{nj})_{n \in \mathbb{N}, j \leq n}$ una successione di v.a. e (φ_{nj}) la corrispondente successione di f.c.. Sono allora equivalenti le condizioni:*

- (a) *per ogni $\varepsilon > 0$, la successione (X_n) soddisfa alla relazione*

$$\lim_{n \rightarrow +\infty} \max_{j \leq n} \mathbb{P}(|X_{nj}| \geq \varepsilon) = 0;$$

- (b) *$\lim_{n \rightarrow +\infty} \max_{j \leq n} |\varphi_{nj}(t) - 1| = 0$ uniformemente in ogni intervallo limitato.*

Dimostrazione. (a) \implies (b)

$$\begin{aligned} \max_{j \leq n} |\varphi_{nj}(t) - 1| &\leq \max_{j \leq n} \int_{\mathbb{R}} |e^{itx} - 1| \, dF_{nj}(x) \\ &\leq \max_{j \leq n} \int_{\{|x| < \varepsilon\}} |e^{itx} - 1| \, dF_{nj}(x) + \max_{j \leq n} \int_{\{|x| \geq \varepsilon\}} |e^{itx} - 1| \, dF_{nj}(x). \end{aligned}$$

Ora, poiché $|e^{ix} - 1| \leq |x|$, si ha

$$\begin{aligned} \max_{j \leq n} |\varphi_{nj}(t) - 1| &\leq \max_{j \leq n} \int_{\{|x| < \varepsilon\}} |tx| \, dF_{nj}(x) + 2 \max_{j \leq n} \int_{\{|x| \geq \varepsilon\}} dF_{nj}(x) \\ &\leq |t| \varepsilon + 2 \max_{j \leq n} \mathbb{P}(|X_{nj}| \geq \varepsilon). \end{aligned}$$

(b) \implies (a) In virtù della (4.2.2), risulta

$$\begin{aligned} \max_{j \leq n} \mathbb{P}(|X_{nj}| \geq \varepsilon) &= \max_{j \leq n} \int_{\{|x| \geq \varepsilon\}} dF_{nj}(x) \\ &\leq \max_{j \leq n} 7\varepsilon \int_0^{1/\varepsilon} (1 - \Re \varphi_{nj}(t)) \, ds \leq 7\varepsilon \int_0^{1/\varepsilon} \max_{j \leq n} |\varphi_{nj}(t) - 1| \, ds, \end{aligned}$$

poiché $|1 - \Re z| = |\Re(1 - z)| \leq |1 - z|$. Grazie all'ipotesi ed al teorema di convergenza dominata, vale la condizione (a). \square

Il seguente lemma serve a richiamare un risultato dell'analisi elementare.

Lemma 4.2.4. *Siano (a_n) e (b_n) due successioni di numeri reali tali che*

$$\lim_{n \rightarrow +\infty} (a_n + b_n) = 0;$$

allora si ha

$$\limsup_{n \rightarrow +\infty} |a_n| = \limsup_{n \rightarrow +\infty} |b_n|.$$

Dimostrazione. In virtù dell'ipotesi, si ha, definitivamente, $|a_n + b_n| < \varepsilon$, per ogni $\varepsilon > 0$; equivalentemente, $-\varepsilon < a_n + b_n < \varepsilon$, relazione dalla quale scendono le due disequaglianze, entrambe valide definitivamente,

$$-a_n < b_n + \varepsilon < |b_n| + \varepsilon \quad \text{e} \quad a_n < -b_n + \varepsilon < |b_n| + \varepsilon$$

onde $|a_n| < |b_n| + \varepsilon$. La disequaglianza inversa si ottiene scambiando i ruoli delle due successioni. \square

Fine della dimostrazione del Teorema 4.2.1. Rimane da dimostrare l'implicazione (b) \implies (a). Si mostrerà dapprima che

$$\lim_{n \rightarrow +\infty} \left(\frac{t^2}{2} + \sum_{j=1}^n \left(\varphi_j \left(\frac{t}{c_n} \right) - 1 \right) \right) = 0. \quad (4.2.3)$$

Poiché (T_n) converge in legge a $N(0, 1)$, si ha

$$\varphi_{T_n}(t) = \prod_{j=1}^n \varphi_j \left(\frac{t}{c_n} \right) = \prod_{j=1}^n \left(1 + \left[\varphi_j \left(\frac{t}{c_n} \right) - 1 \right] \right) \xrightarrow{n \rightarrow +\infty} \exp \left(-\frac{t^2}{2} \right).$$

Perciò

$$\lim_{n \rightarrow +\infty} \sum_{j=1}^n \text{Log} \left(1 + \left[\varphi_j \left(\frac{t}{c_n} \right) - 1 \right] \right) = -\frac{t^2}{2}$$

e, per la (4.1.5), con un opportuno ν_j tale che $|\nu_j| \leq 1$,

$$\lim_{n \rightarrow +\infty} \left(\frac{t^2}{2} + \sum_{j=1}^n \left(\varphi_j \left(\frac{t}{c_n} \right) - 1 \right) + \sum_{j=1}^n \nu_j \left| \varphi_j \left(\frac{t}{c_n} \right) - 1 \right|^2 \right) = 0,$$

onde, in virtù del Lemma 4.2.4,

$$\begin{aligned} & \limsup_{n \rightarrow +\infty} \left| \frac{t^2}{2} + \sum_{j=1}^n \left(\varphi_j \left(\frac{t}{c_n} \right) - 1 \right) \right| \\ &= \limsup_{n \rightarrow +\infty} \sum_{j=1}^n |\nu_j| \left| \varphi_j \left(\frac{t}{c_n} \right) - 1 \right|^2 \\ &\leq \limsup_{n \rightarrow +\infty} \left(\max_{j \leq n} \left| \varphi_j \left(\frac{t}{c_n} \right) - 1 \right| \right) \sum_{j=1}^n \left| \varphi_j \left(\frac{t}{c_n} \right) - 1 \right|. \end{aligned} \quad (4.2.4)$$

Ora, per l'ipotesi e per il Lemma 4.2.3, è

$$\lim_{n \rightarrow +\infty} \max_{j \leq n} \left| \varphi_j \left(\frac{t}{c_n} \right) - 1 \right| = 0,$$

e inoltre, tenendo conto della (4.1.3), si ha

$$\begin{aligned} \sum_{j=1}^n \left| \varphi_j \left(\frac{t}{c_n} \right) - 1 \right| &= \sum_{j=1}^n \left| \int_{\mathbb{R}} \left[\exp \left(\frac{itx}{c_n} \right) - 1 \right] dF_j(x) \right| \\ &= \sum_{j=1}^n \left| \int_{\mathbb{R}} \theta \frac{t^2 x^2}{2c_n^2} dF_j(x) \right| \leq \frac{t^2}{2c_n^2} \sum_{j=1}^n \sigma_j^2 = \frac{t^2}{2}. \end{aligned}$$

La (4.2.3) è dunque conseguenza della (4.2.4).

Si dimostrerà quindi che, per ogni $\varepsilon > 0$, si ha

$$\limsup_{n \rightarrow +\infty} \left(1 - \frac{1}{c_n^2} \sum_{j=1}^n \int_{\{|x| < \varepsilon c_n\}} x^2 dF_j(x) \right) \leq \frac{4}{\varepsilon^2 t^2}. \quad (4.2.5)$$

Per la (4.2.3), è

$$\lim_{n \rightarrow +\infty} \left(\frac{t^2}{2} - \sum_{j=1}^n \Re \left[1 - \varphi_j \left(\frac{t}{c_n} \right) \right] \right) = 0,$$

o, equivalentemente,

$$\lim_{n \rightarrow +\infty} \left\{ \frac{t^2}{2} - \sum_{j=1}^n \int_{\{|x| < \varepsilon c_n\}} \cdots - \sum_{j=1}^n \int_{\{|x| \geq \varepsilon c_n\}} \left(1 - \cos \frac{tx}{c_n} \right) dF_j(x) \right\} = 0,$$

di modo che, applicando il Lemma 4.2.4, si ottiene

$$\begin{aligned} \limsup_{n \rightarrow +\infty} \left| \frac{t^2}{2} - \sum_{j=1}^n \int_{\{|x| < \varepsilon c_n\}} \left(1 - \cos \frac{tx}{c_n} \right) dF_j(x) \right| \\ = \limsup_{n \rightarrow +\infty} \left| \sum_{j=1}^n \int_{\{|x| \geq \varepsilon c_n\}} \left(1 - \cos \frac{tx}{c_n} \right) dF_j(x) \right|. \end{aligned} \quad (4.2.6)$$

Ma

$$\begin{aligned} \sum_{j=1}^n \int_{\{|x| < \varepsilon c_n\}} \left(1 - \cos \frac{tx}{c_n} \right) dF_j(x) &\leq \sum_{j=1}^n \int_{\{|x| < \varepsilon c_n\}} \frac{t^2 x^2}{2c_n^2} dF_j(x) \\ &\leq \frac{t^2}{2c_n^2} \sum_{j=1}^n \sigma_j^2 = \frac{t^2}{2}, \end{aligned}$$

ciò che mostra come nella (4.2.6) si possano eliminare i valori assoluti. Dalla stessa (4.2.6) scende ora che

$$\begin{aligned} \limsup_{n \rightarrow +\infty} \frac{t^2}{2} \left(1 - \frac{1}{c_n^2} \sum_{j=1}^n \int_{\{|x| < \varepsilon c_n\}} x^2 dF_j(x) \right) \\ \leq \limsup_{n \rightarrow +\infty} \left(\frac{t^2}{2} - \sum_{j=1}^n \int_{\{|x| < \varepsilon c_n\}} \left(1 - \cos \frac{tx}{c_n} \right) dF_j(x) \right) \\ = \limsup_{n \rightarrow +\infty} \sum_{j=1}^n \int_{\{|x| \geq \varepsilon c_n\}} \left(1 - \cos \frac{tx}{c_n} \right) dF_j(x). \end{aligned}$$

D'altro canto, la disuguaglianza di Čebyšev dà

$$\begin{aligned} \sum_{j=1}^n \int_{\{|x| \geq \varepsilon c_n\}} \left(1 - \cos \frac{tx}{c_n} \right) dF_j(x) &\leq \sum_{j=1}^n \int_{\{|x| \geq \varepsilon c_n\}} dF_j(x) \\ &= 2 \sum_{j=1}^n \mathbb{P}(|X_j| \geq \varepsilon c_n) \leq \frac{2}{\varepsilon^2 c_n^2} \sum_{j=1}^n \sigma_j^2 = \frac{2}{\varepsilon^2}, \end{aligned}$$

sicché la (4.2.5) è dimostrata. Per dimostrare che vale la condizione di Lindeberg (4.1.1) basta ora far tendere t a $+\infty$ nella (4.2.5). \square

4.3 LGN: leggi deboli

In questa sezione e nella seguente supporremo, senza che ciò sia richiamato esplicitamente nel seguito, che sia assegnato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e che tutte le v.a. che saranno considerate siano definite in tale spazio.

Definizione 4.3.1. Sia $(X_n)_{n \in \mathbb{N}}$ una successione di v.a. di L^1 : si dice che la successione (X_n) obbedisce alla *Legge dei Grandi Numeri* (LGN) se, posto $S_n := \sum_{j=1}^n X_j$, la v.a.

$$Z_n := \frac{S_n - \mathbb{E}(S_n)}{n}$$

converge a zero. Si parlerà di LGN *debole* se la convergenza avviene in probabilità, di LGN *forte* se quasi certamente. \diamond

Teorema 4.3.1. (Markov). *Sia (X_n) una successione di v.a. indipendenti di L^2 con $V(X_n) = \sigma_n^2$ per ogni $n \in \mathbb{N}$. Se vale la condizione di Markov*

$$\lim_{n \rightarrow +\infty} \frac{1}{n^2} \sum_{j=1}^n \sigma_j^2 = 0, \quad (4.3.1)$$

(X_n) obbedisce alla legge debole dei grandi numeri.

Dimostrazione. Al solito non è restrittivo supporre che le v.a. della successione (X_n) siano tutte centrate, $\mathbb{E}(X_n) = 0$ ($n \in \mathbb{N}$). Se $Z_n := S_n/n$, si ha, per ogni $\varepsilon > 0$,

$$\mathbb{P}(|Z_n| \geq \varepsilon) \leq \frac{V(Z_n)}{\varepsilon^2} = \frac{V(S_n)}{\varepsilon^2 n^2} = \frac{1}{\varepsilon^2 n^2} \sum_{j=1}^n \sigma_j^2,$$

che dà l'asserto in virtù della (4.3.1). \square

L'ipotesi d'indipendenza è, in effetti, ridondante, perché basta supporre che le v.a. della successione (X_n) siano a due a due incorrelate. La condizione (4.3.1) è, in particolare, soddisfatta se le varianze sono uniformemente limitate: $\sigma_n^2 \leq A$ per ogni $n \in \mathbb{N}$. In quest'ultima ipotesi vale, però, la legge forte (si veda il successivo Teorema 4.4.1).

Teorema 4.3.2. *Se la successione (X_n) di v.a. di L^1 obbedisce alla LGN forte, allora la successione*

$$\left(\frac{X_n - \mathbb{E}(X_n)}{n} \right)$$

converge a 0 q.c..

Dimostrazione. Non è restrittivo supporre che le v.a. siano tutte centrate (vale a dire $\mathbb{E}(X_n) = 0$ per ogni $n \in \mathbb{N}$). Posto, al solito, $S_n := \sum_{j=1}^n X_j$, l'ipotesi è che si abbia

$$\frac{S_n}{n} \xrightarrow[n \rightarrow +\infty]{} 0 \quad \text{q.c..}$$

Ora

$$\frac{X_n}{n} = \frac{S_n}{n} - \frac{n-1}{n} \frac{S_{n-1}}{n-1},$$

onde l'asserto. \square

Mostriamo con un esempio che una successione (X_n) di v.a. può obbedire alla LGN debole senza obbedire alla LGN forte.

Esempio 4.3.1. Sia $(X_n)_{n \geq 2}$ una successione di v.a. indipendenti tale che

$$\mathbb{P}(X_n = -n) = \mathbb{P}(X_n = n) = \frac{1}{2n \ln n}, \quad \mathbb{P}(X_n = 0) = 1 - \frac{1}{n \ln n}.$$

Per ogni $n \geq 2$, X_n è in L^1 ; infatti,

$$\mathbb{E}(|X_n|) = \frac{1}{\ln n} < +\infty.$$

Inoltre

$$\mathbb{E}(X_n) = 0 \quad \text{e} \quad V(X_n) = \|X_n\|_2^2 = \frac{n}{\ln n}.$$

La funzione $x \mapsto \varphi(x) := x/\ln x$ è crescente per $x > e$. Perciò, se $n \geq 3$,

$$\begin{aligned} \frac{1}{n^2} \sum_{j=2}^n V(X_j) &= \frac{1}{n^2} \frac{2}{\ln 2} + \frac{1}{n^2} \sum_{j=3}^n V(X_j) \\ &\leq \frac{1}{n^2} \frac{2}{\ln 2} + \frac{1}{n^2} \frac{(n-2)n}{\ln n} \xrightarrow{n \rightarrow +\infty} 0. \end{aligned}$$

La successione (X_n) soddisfa, dunque, alla condizione di Markov (4.3.1) e verifica, quindi, la LGN debole. Essa non obbedisce, però, alla legge forte. Infatti, se così fosse, si avrebbe, come si è visto sopra, $X_n/n \rightarrow 0$ q.c..

Sia ora, per $\varepsilon > 0$, $A_n(\varepsilon) := \{|X_n| \geq n\varepsilon\}$. Gli eventi $A_n(\varepsilon)$ sono indipendenti.

È facile verificare che $X_n/n \xrightarrow{n \rightarrow +\infty} 0$ q.c. equivale a

$$\mathbb{P}\left(\liminf_{n \rightarrow +\infty} A_n^c(\varepsilon)\right) = 1.$$

Infatti, quest'ultima condizione equivale a dire che si ha definitivamente e q.c.

$\left|\frac{X_n}{n}\right| < \varepsilon$, cioè $\frac{X_n}{n} \rightarrow 0$ q.c.. D'altro canto, per ogni $\varepsilon \in]0, 1[$ risulta

$$\mathbb{P}(A_n(\varepsilon)) = \mathbb{P}(|X_n| \geq n\varepsilon) = \frac{1}{n \ln n},$$

onde

$$\sum_{n \geq 2} \mathbb{P}(A_n(\varepsilon)) = \sum_{n \geq 2} \frac{1}{n \ln n} = +\infty,$$

sicché il secondo lemma di Borel–Cantelli dà

$$\mathbb{P}\left(\limsup_{n \rightarrow +\infty} A_n(\varepsilon)\right) = 1,$$

ossia

$$\mathbb{P}\left(\liminf_{n \rightarrow +\infty} A_n^c(\varepsilon)\right) = 0.$$

■

Teorema 4.3.3. (Khinchin). *Se le v.a. della successione (X_n) sono in L^1 , indipendenti e isonome, la successione obbedisce alla LGN debole.*

Dimostrazione. Si può supporre, senza perdita di generalità, che le v.a. della successione siano centrate, cioè $\mathbb{E}(X_n) = 0$ per ogni $n \in \mathbb{N}$. In virtù del teorema (5.3.12), basta mostrare che $Z_n = \frac{1}{n} \sum_{j=1}^n X_j$ tende a zero in legge; ma ciò equivale, per il teorema (6.6.5), a mostrare che $\lim_{n \rightarrow +\infty} \varphi_{Z_n}(t) = 1$ per ogni $t \in \mathbb{R}$, oppure, equivalentemente, che $\lim_{n \rightarrow +\infty} \text{Log } \varphi_{Z_n}(t) = 0$ per ogni $t \in \mathbb{R}$. Sia φ la f.c. comune delle v.a. X_n . L'ipotesi d'indipendenza dà allora

$$\varphi_{Z_n}(t) = \left[\varphi \left(\frac{t}{n} \right) \right]^n,$$

onde

$$\text{Log } \varphi_{Z_n}(t) = n \text{Log } \varphi \left(\frac{t}{n} \right).$$

Dallo sviluppo (4.1.3) e da $\mathbb{E}(X_n) = 0$ segue $\varphi(t) = 1 + o(t)$, onde

$$\text{Log } \varphi_{Z_n}(t) = n \text{Log} \left[1 + o \left(\frac{t}{n} \right) \right] = n o \left(\frac{t}{n} \right).$$

Perciò $\lim_{n \rightarrow +\infty} \text{Log } \varphi_{Z_n}(t) = 0$. □

Si vedrà oltre che nelle stesse ipotesi dell'ultimo teorema vale la legge forte (Teorema 4.4.3).

Teorema 4.3.4. *Per una successione (X_n) di v.a. isonome e indipendenti di L^1 sono equivalenti le condizioni:*

(a) *esiste una successione (α_n) di numeri reali tale che*

$$\frac{S_n}{n} - \alpha_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0;$$

(b) $\lim_{n \rightarrow +\infty} n \mathbb{P}(|X_1| > n) = 0$.

Se vale una di queste condizioni si può prendere $\alpha_n = \mathbb{E}(X_1 \mathbf{1}_{\{|X_1| \leq n\}})$.

Dimostrazione. Dimostreremo la sola implicazione (b) \implies (a). Si considerino le v.a. troncate $X_{nk} := X_k \mathbf{1}_{\{|X_k| \leq n\}}$ e si ponga $S'_n := \sum_{k=1}^n X_{nk}$ e $\alpha_n := \mathbb{E}(X_1 \mathbf{1}_{\{|X_1| \leq n\}})$. Allora $\alpha_n = \mathbb{E}(S'_n/n)$. Poiché

$$\left\{ \left| \frac{S_n}{n} - \alpha_n \right| > \varepsilon \right\} \subseteq \left\{ \left| \frac{S'_n}{n} - \alpha_n \right| > \varepsilon \right\} \cup \{S_n \neq S'_n\},$$

si ha

$$\mathbb{P} \left(\left| \frac{S_n}{n} - \alpha_n \right| > \varepsilon \right) \leq \mathbb{P} \left(\left| \frac{S'_n}{n} - \alpha_n \right| > \varepsilon \right) + \mathbb{P}(S_n \neq S'_n). \quad (4.3.2)$$

Si osservi ora che

$$\{S_n \neq S'_n\} = \bigcup_{k=1}^n \{X_k \neq X_{nk}\},$$

sicché

$$\begin{aligned}\mathbb{P}(S_n \neq S'_n) &\leq \sum_{k=1}^n \mathbb{P}(X_k \neq X_{nk}) \\ &= \sum_{k=1}^n \mathbb{P}(|X_k| > n) = n \mathbb{P}(|X_1| > n) \xrightarrow{n \rightarrow +\infty} 0.\end{aligned}$$

Quanto all'altro termine al secondo membro della (4.3.2), si ha, dalla disuguaglianza di Čebyšev,

$$\mathbb{P}\left(\left|\frac{S'_n}{n} - \alpha_n\right| > \varepsilon\right) \leq \frac{V(S'_n)}{n^2 \varepsilon^2} = \frac{V(X'_{n1})}{n \varepsilon^2} \leq \frac{\mathbb{E}(X'^2_{n1})}{n \varepsilon^2}. \quad (4.3.3)$$

D'altro canto,

$$\mathbb{E}(X'^2_{n1}) = \int_0^{+\infty} 2t \mathbb{P}(X'_{n1} > t) dt = \int_0^n 2t \mathbb{P}(|X_1| > t) dt.$$

L'ipotesi fatta implica che sia

$$\begin{aligned}\lim_{n \rightarrow +\infty} \frac{\mathbb{E}(X'^2_{n1})}{n} &= \lim_{n \rightarrow +\infty} \frac{1}{n} \int_0^n 2t \mathbb{P}(|X_1| > t) dt \\ &= \lim_{n \rightarrow +\infty} 2n \mathbb{P}(|X_1| > n) = 0,\end{aligned}$$

e, quindi, l'asserto, in virtù della (4.3.3).

La dimostrazione dell'implicazione inversa si può trovare in (Feller, 1971) Teorema V.7.1.. \square

4.4 LGN: leggi forti

Diamo ora un primo esempio di LGN forte.

Teorema 4.4.1. (Rajchman, 1932). *Se le v.a. della successione (X_n) sono a due a due incorrelate e hanno varianze uniformemente limitate in L^2 (cioè $V(X_n) \leq H^2$ per ogni $n \in \mathbb{N}$), allora la successione (X_n) obbedisce alla LGN forte.*

Dimostrazione. Si può supporre, senza perdita di generalità, che la v.a. della successione data siano centrate, $\mathbb{E}(X_n) = 0$ per ogni $n \in \mathbb{N}$. Allora, l'ipotesi di limitatezza uniforme in L^2 dà $\sigma_n^2 \leq H^2$ per ogni $n \in \mathbb{N}$; così,

$$V\left(\frac{S_n}{n}\right) = \frac{V(S_n)}{n^2} = \frac{1}{n^2} \sum_{j=1}^n \sigma_j^2 \leq \frac{H^2}{n}.$$

Dalla disuguaglianza di Čebyšev si ottiene

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) \leq \frac{H^2}{n \varepsilon^2},$$

sicch , sommando su n , la serie a secondo membro diverge; limitandosi, per , alla sottosuccessione (S_{n^2}) , si ha

$$\sum_{n \in \mathbb{N}} \mathbb{P} \left(\left| \frac{S_{n^2}}{n^2} \right| > \varepsilon \right) \leq \frac{H^2}{\varepsilon^2} \sum_{n \in \mathbb{N}} \frac{1}{n^2},$$

che   convergente; adesso, il primo lemma di Borel–Cantelli assicura che sia

$$\mathbb{P} \left(\limsup_{n \rightarrow +\infty} \left\{ \left| \frac{S_{n^2}}{n^2} \right| > \varepsilon \right\} \right) = 0.$$

Perci   

$$\left| \frac{S_{n^2}}{n^2} \right| \leq \varepsilon$$

definitivamente fuorch  in un insieme B_ε , dipendente da ε , di probabilit  nulla, $\mathbb{P}(B_\varepsilon) = 0$. Pertanto

$$\frac{S_{n^2}}{n^2} \xrightarrow[n \rightarrow +\infty]{} 0 \quad \text{q.c..}$$

Se il numero naturale k non   un quadrato perfetto, esiste un altro numero naturale n tale che sia $n^2 < k < (n+1)^2$. Posto

$$D_n := \max \{ |S_k - S_{n^2}| : n^2 < k < (n+1)^2 \},$$

si ha, per $n^2 < k < (n+1)^2$,

$$\frac{|S_k|}{k} = \frac{|S_k - S_{n^2} + S_{n^2}|}{k} \leq \frac{|S_{n^2}| + D_n}{k} \leq \frac{|S_{n^2}| + D_n}{n^2};$$

baster  allora, per concludere la dimostrazione, far vedere che

$$\frac{D_n}{n^2} \xrightarrow[n \rightarrow +\infty]{} 0 \quad \text{q.c..}$$

Ora

$$D_n^2 \leq \sum_{k=n^2+1}^{(n+1)^2-1} (X_{n^2+1} + X_{n^2+2} + \cdots + X_k)^2,$$

onde, poich  le v.a. della successione sono centrate e a due a due incorrelate,

$$\begin{aligned} \mathbb{E}(D_n^2) &\leq \sum_{k=n^2+1}^{(n+1)^2-1} \{ \mathbb{E}(X_{n^2+1}^2) + \mathbb{E}(X_{n^2+2}^2) + \cdots + \mathbb{E}(X_k^2) \} \\ &\leq 2n \sum_{k=n^2+1}^{(n+1)^2-1} H^2 = 4n^2 H^2. \end{aligned}$$

Perci 

$$\mathbb{P} \left(\frac{D_n}{n^2} > \varepsilon \right) \leq \frac{\mathbb{E}(D_n^2)}{n^4 \varepsilon^2} \leq \frac{4H^2}{n^2 \varepsilon^2};$$

mediante quest'ultima diseguaglianza, ricorrendo nuovamente al primo lemma di Borel–Cantelli, si conclude che $D_n/n^2 \rightarrow 0$ q.c.. \square

Si osservi che il teorema di Rajchman si applica, in particolare, ad una successione di v.a. indipendenti ed isonome; vale, quindi il seguente

Corollario 4.4.1. *Una successione $(X_n) \subseteq L^2$ di v.a. indipendenti, isonome e con varianze uniformemente limitate obbedisce alla LGN.*

Applicato ad una successione di v.a. Bernoulliane indipendenti (X_n) nella quale sia p la probabilità di successo ad ogni prova, $\mathbb{P}(X_n = 1) = p$, questo corollario assicura che, nelle stesse ipotesi del Teorema (1.9.8) valga la LGN forte e non solo quella debole.

Osservazione 4.4.1. Si noti che nelle ipotesi del Teorema 4.4.1 si ha anche la convergenza in L^2 . Infatti, poiché le v.a. sono incorrelate, si ha

$$\begin{aligned} \left\| \frac{S_n - \mathbb{E}(S_n)}{n} \right\|_2 &= \mathbb{E} \left[\frac{(S_n - \mathbb{E}(S_n))^2}{n^2} \right] = \frac{V(S_n)}{n^2} \\ &= \frac{1}{n^2} \sum_{j=1}^n V(X_j) \leq \frac{H^2}{n} \xrightarrow{n \rightarrow +\infty} 0, \end{aligned}$$

che dà l'asserto. ■

Il risultato che segue rafforza la diseuguaglianza di Čebyšev nel caso della somma di v.a. indipendenti.

Teorema 4.4.2. (Diseuguaglianza di Kolmogorov).

Dimostrazione. Si ponga

$$A_1 := \{|S_1| \geq \varepsilon\},$$

e, per $j \geq 2$,

$$A_j := \{|S_j| \geq \varepsilon\} \cap \left(\bigcap_{k < j} \{|S_k| < \varepsilon\} \right).$$

Gli insiemi A_j così definiti sono misurabili e disgiunti. Inoltre

$$\bigcup_{j=1}^n A_j = \left\{ \max_{j \leq n} |S_j| \geq \varepsilon \right\}.$$

Si ha ora, procedendo come nella dimostrazione della diseuguaglianza di Čebyšev,

$$\begin{aligned} \mathbb{E}(S_n^2) &\geq \sum_{j=1}^n \int_{A_j} S_n^2 \, d\mathbb{P} = \sum_{j=1}^n \int_{A_j} \{S_j + (S_n - S_j)\}^2 \, d\mathbb{P} \\ &= \sum_{j=1}^n \int_{A_j} [S_j^2 + 2S_j(S_n - S_j) + (S_n - S_j)^2] \, d\mathbb{P} \\ &\geq \sum_{j=1}^n \int_{A_j} [S_j^2 + 2S_j(S_n - S_j)] \, d\mathbb{P}. \end{aligned}$$

Poiché A_j e S_j sono funzioni delle v.a. X_1, X_2, \dots, X_j , mentre la differenza $S_n - S_j$ è funzione delle v.a. $X_{j+1}, X_{j+2}, \dots, X_n$, che sono indipendenti dalle prime, le v.a. $S_j \mathbf{1}_{A_j}$ e $S_n - S_j$, sono pure indipendenti; perciò

$$\begin{aligned} \int_{A_j} S_j (S_n - S_j) \, d\mathbb{P} &= \mathbb{E} [(S_j \mathbf{1}_{A_j}) (S_n - S_j)] \\ &= \mathbb{E} [(S_j \mathbf{1}_{A_j})] \mathbb{E} [(S_n - S_j)] = 0. \end{aligned}$$

Quindi

$$\mathbb{E}(S_n^2) \geq \sum_{j=1}^n \int_{A_j} S_j^2 \, d\mathbb{P} \geq \sum_{j=1}^n \varepsilon^2 \mathbb{P}(A_j) = \varepsilon^2 \mathbb{P} \left(\max_{j \leq n} |S_j| \geq \varepsilon \right),$$

onde l'asserto. □

Dalla diseuguaglianza di Kolmogorov scende l'omonima LGN forte.

Teorema 4.4.3. *Sia (X_n) una successione di v.a. indipendenti e centrate di L^2 ; se è convergente la serie*

$$\sum_{n \in \mathbb{N}} \frac{V(X_n)}{n^2} < +\infty, \quad (4.4.1)$$

allora la successione (X_n) obbedisce alla LGN forte.

Dimostrazione. Si ponga, per $\varepsilon > 0$,

$$B_n := \bigcup_{2^n \leq j < 2^{n+1}} \{|S_j| > j \varepsilon\}.$$

Se convergesse la serie $\sum_{n \in \mathbb{N}} \mathbb{P}(B_n)$, il primo lemma di Borel–Cantelli darebbe

$$\mathbb{P} \left(\limsup_{n \rightarrow +\infty} B_n \right) = 0,$$

e cioè l'asserto. Basta perciò stabilire la convergenza di tale serie. Per un punto $\omega \in B_n$ esiste almeno un indice j con $2^n \leq j < 2^{n+1}$ tale che $|S_j(\omega)|/j > \varepsilon$, sicché

$$\mathbb{P}(B_n) \leq \mathbb{P} \left(\max_{2^n \leq j < 2^{n+1}} \frac{|S_j|}{j} > \varepsilon \right) \leq \mathbb{P} \left(\max_{2^n \leq j < 2^{n+1}} |S_j| > 2^n \varepsilon \right),$$

e perciò la diseuguaglianza di Kolmogorov dà

$$\mathbb{P}(B_n) \leq \sum_{j=2^n}^{2^{n+1}-1} \frac{V(X_j)}{2^{2n} \varepsilon^2};$$

di qui, poiché

$$\frac{1}{2^{2n}} \leq \frac{4}{j^2},$$

per $j \leq 2^{n+1}$ scende

$$\begin{aligned} \sum_{n \in \mathbb{N}} \mathbb{P}(B_n) &\leq \frac{1}{\varepsilon^2} \sum_{n \in \mathbb{N}} \sum_{j=2^n}^{2^{n+1}-1} \frac{\sigma_j^2}{2^{2n}} \leq \frac{4}{\varepsilon^2} \sum_{n \in \mathbb{N}} \sum_{j=2^n}^{2^{n+1}-1} \frac{\sigma_j^2}{j^2} \\ &\leq \frac{4}{\varepsilon^2} \sum_{j=2}^{\infty} \frac{\sigma_j^2}{j^2} < +\infty, \end{aligned}$$

che dà l'asserto. □

Se le altre condizioni dell'ultimo teorema sono verificate, la successione (X_n) può non obbedire alla LGN forte, se diverge la serie $\sum_{n \in \mathbb{N}} \sigma_n^2/n^2$.

Esempio 4.4.1. Sia (X_n) una successione di v.a. indipendenti tali che

$$\mathbb{P}(X_n = n) = \mathbb{P}(X_n = -n) = \frac{1}{2}.$$

Allora, per ogni $n \in \mathbb{N}$, si ha $\mathbb{E}(X_n) = 0$ e $\sigma_n^2 = n^2$ sicché

$$\sum_{n \in \mathbb{N}} \frac{\sigma_n^2}{n^2} = +\infty.$$

Se la successione obbedisse alla LGN forte, si avrebbe $X_n/n \rightarrow 0$ q.c., per il Teorema 4.3.2, onde $\mathbb{P}(\limsup_{n \rightarrow +\infty} \{|X_n| \geq \varepsilon n\}) = 0$ per ogni $\varepsilon > 0$. Il secondo lemma di Borel–Cantelli darebbe dunque

$$\sum_{n \in \mathbb{N}} \mathbb{P}(|X_n| \geq \varepsilon n) < +\infty,$$

mentre, per ogni $\varepsilon \in]0, 1[$, si ha, in realtà, $\mathbb{P}(|X_n| \geq \varepsilon n) = 1$ e, dunque

$$\sum_{n \in \mathbb{N}} \mathbb{P}(|X_n| \geq \varepsilon n) = +\infty.$$

■

Si osservi che, se le v.a. della successione (X_n) sono indipendenti, il Teorema di Rajchman 4.4.1 è una conseguenza immediata della LGN forte di Kolmogorov 4.4.3.

La versione della LGN forte piú utile per le applicazioni, in ispecie per quelle alla statistica matematica, è data nel seguente teorema.

Teorema 4.4.4. (Khinchin–Kolmogorov) *Una successione (X_n) di v.a. di L^1 indipendenti e isonome obbedisce alla LGN forte.*

Dimostrazione. Al solito, non è restrittivo supporre che le variabili della successione siano centrate, $\mathbb{E}(X_n) = 0$ per ogni $n \in \mathbb{N}$. Poiché le v.a. della successione hanno

tutte la stessa legge, si ha

$$\begin{aligned}
\sum_{n \in \mathbb{N}} \mathbb{P}(|X_n| \geq n) &= \sum_{n \in \mathbb{N}} \mathbb{P}(|X_1| \geq n) = \sum_{n \in \mathbb{N}} \sum_{j \geq n} \mathbb{P}(j \leq |X_1| < j+1) \\
&= \sum_{j \in \mathbb{N}} \sum_{n \leq j} \mathbb{P}(j \leq |X_1| < j+1) \\
&= \sum_{j \in \mathbb{N}} j \mathbb{P}(j \leq |X_1| < j+1) = \sum_{j \in \mathbb{N}} \int_{\{j \leq |X_1| < j+1\}} j \, d\mathbb{P} \\
&\leq \sum_{j \in \mathbb{N}} \int_{\{j \leq |X_1| < j+1\}} |X_1| \, d\mathbb{P} \leq \mathbb{E}(|X_1|) < +\infty.
\end{aligned}$$

Il primo lemma di Borel–Cantelli assicura che sia

$$\mathbb{P}\left(\limsup_{n \rightarrow +\infty} \{|X_n| \geq n\}\right) = 0.$$

Perciò, se si definiscono le v.a. “troncate” $Y_n := X_n \mathbf{1}_{\{|X_n| < n\}}$, si ha definitivamente $Y_n = X_n$ tranne che in un insieme di probabilità nulla.

Poiché le v.a. Y_n sono limitate, esse appartengono a L^2 . Si può quindi applicare il precedente Teorema 4.4.3 alla successione $(Y_n - \mathbb{E}(Y_n))$; per dimostrare che questa obbedisce alla LGN forte, basta verificare la convergenza della serie

$$\sum_{n \in \mathbb{N}} \frac{V(Y_n)}{n^2}.$$

Ora si ha

$$\begin{aligned}
\sum_{n \in \mathbb{N}} \frac{V(Y_n)}{n^2} &\leq \sum_{n \in \mathbb{N}} \frac{\mathbb{E}(Y_n^2)}{n^2} = \sum_{n \in \mathbb{N}} \frac{\mathbb{E}(X_1^2 \mathbf{1}_{\{|X_1| < n\}})}{n^2} \\
&= \sum_{n \in \mathbb{N}} \frac{1}{n^2} \sum_{j=1}^n \mathbb{E}(X_1^2 \mathbf{1}_{\{j-1 \leq |X_1| < j\}}) \\
&= \sum_{j \in \mathbb{N}} \mathbb{E}(X_1^2 \mathbf{1}_{\{j-1 \leq |X_1| < j\}}) \sum_{n=j}^{\infty} \frac{1}{n^2}.
\end{aligned}$$

Dalla maggiorazione elementare

$$\begin{aligned}
\sum_{n=k}^{+\infty} \frac{1}{n^2} &= \frac{1}{k^2} + \sum_{n \geq k+1} \frac{1}{n^2} \\
&\leq \frac{1}{k^2} + \sum_{n \geq k+1} \int_{n-1}^n \frac{dx}{x^2} \leq \frac{1}{k} + \int_k^{+\infty} \frac{dx}{x^2} = \frac{2}{k}, \tag{4.4.2}
\end{aligned}$$

segue che

$$\begin{aligned}
\sum_{n \in \mathbb{N}} \frac{V(Y_n)}{n^2} &\leq 2 \sum_{j \in \mathbb{N}} \frac{\mathbb{E}(X_1^2 \mathbf{1}_{\{j-1 \leq |X_1| < j\}})}{j} \\
&\leq 2 \sum_{j \in \mathbb{N}} \mathbb{E}(|X_1| \mathbf{1}_{\{j-1 \leq |X_1| < j\}}) = 2 \mathbb{E}(|X_1|) < +\infty.
\end{aligned}$$

D'altro canto, si ha

$$\mathbb{E}(Y_n) = \mathbb{E}(X_n \mathbf{1}_{\{|X_n| < n\}}) = \mathbb{E}(X_1 \mathbf{1}_{\{|X_1| < n\}}) \xrightarrow{n \rightarrow +\infty} \mathbb{E}(X_1) = 0,$$

e quindi, dato che la convergenza di una successione implica la sua convergenza allo stesso limite nel senso di CESÀRO,

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=1}^n \mathbb{E}(Y_j) = 0.$$

Si ha perciò che

$$\frac{1}{n} S'_n := \frac{1}{n} \sum_{j=1}^n Y_j \xrightarrow[n \rightarrow +\infty]{q.c.} 0.$$

Ora, fissato arbitrariamente $\varepsilon > 0$, si ha per $n \in \mathbb{N}$,

$$\{|S_n| < n\varepsilon\} \supset \{|S'_n| < n\varepsilon\} \cap \{S_n = S'_n\},$$

e dunque

$$\liminf_{n \rightarrow +\infty} \{|S_n| < n\varepsilon\} \supset \liminf_{n \rightarrow +\infty} (\{|S'_n| < n\varepsilon\}) \cap \liminf_{n \rightarrow +\infty} (\{S_n = S'_n\}).$$

Si è appena visto che $S'_n/n \rightarrow 0$ q.c., sicché

$$\mathbb{P} \left(\liminf_{n \rightarrow +\infty} \{|S'_n| < n\varepsilon\} \right) = 1;$$

per la prima parte della dimostrazione, poiché $Y_n = X_n$ q.c. e definitivamente, si ha

$$\mathbb{P} \left(\liminf_{n \rightarrow +\infty} \{S_n = S'_n\} \right) = 1.$$

Perciò

$$\mathbb{P} \left(\liminf_{n \rightarrow +\infty} \{|S_n| < n\varepsilon\} \right) = 1,$$

vale a dire,

$$\frac{S_n}{n} \xrightarrow[n \rightarrow +\infty]{} 0 \quad \text{q.c.},$$

onde l'asserto. □

Lo stesso risultato vale se, anziché supporre che le variabili della successione (X_n) siano indipendenti, esse sono *a due a due indipendenti*.

Teorema 4.4.5. (Etemadi) *Sia (X_n) una successione di v.a. di L^1 , isonome e a due a due indipendenti. Allora (X_n) obbedisce alla LGN forte*

$$\frac{S_n}{n} \xrightarrow[n \rightarrow +\infty]{q.c.} \mathbb{E}(X_1).$$

Dimostrazione. Posto, al solito, $S_n := \sum_{j=1}^n X_j$, si tratta di dimostrare che

$$\frac{S_n}{n} \xrightarrow{n \rightarrow +\infty} \mathbb{E}(X_1) \quad \text{q.c..}$$

Poiché anche (X_n^+) e (X_n^-) soddisfanno alle stesse ipotesi di (X_n) , e, per ogni $n \in \mathbb{N}$, $X_n = X_n^+ - X_n^-$, si può supporre che sia $X_n \geq 0$ ($n \in \mathbb{N}$). Si ha

$$\begin{aligned} \sum_{n \in \mathbb{N}} \mathbb{P}(X_n \geq n) &= \sum_{n \in \mathbb{N}} \mathbb{P}(X_1 \geq n) = \sum_{n \in \mathbb{N}} \sum_{j \geq n} \mathbb{P}(j \leq X_1 < j+1) \\ &= \sum_{j \in \mathbb{N}} \sum_{n \leq j} \mathbb{P}(j \leq X_1 < j+1) \\ &= \sum_{j \in \mathbb{N}} j \mathbb{P}(j \leq X_1 < j+1) = \sum_{j \in \mathbb{N}} \int_{\{j \leq X_1 < j+1\}} j \, d\mathbb{P} \\ &\leq \sum_{j \in \mathbb{N}} \int_{\{j \leq X_1 < j+1\}} X_1 \, d\mathbb{P} \leq \mathbb{E}(X_1) < +\infty. \end{aligned}$$

Il primo lemma di Borel–Cantelli assicura che sia

$$\mathbb{P} \left(\limsup_{n \rightarrow +\infty} \{X_n \geq n\} \right) = 0.$$

Perciò, se si definiscono le v.a. “troncate” $Y_n := X_n \mathbf{1}_{\{|X_n| \leq n\}}$, si ha definitivamente $Y_n = X_n$ tranne che su un insieme di probabilità nulla. Basta così far vedere che, se $S'_n := \sum_{j=1}^n Y_j$, si ha

$$\frac{S'_n}{n} \xrightarrow{n \rightarrow +\infty} \mathbb{E}(X_1) \quad \text{q.c..}$$

Scelti arbitrariamente $\varepsilon > 0$ e $\alpha > 1$, si ponga $k(n) := [\alpha^n]$ ($[x]$ è la parte intera di x). La diseguaglianza di Čebyšev e l’essere le v.a. a due a due indipendenti, e quindi, *a fortiori*, a due a due incorrelate, dà

$$\begin{aligned} \sum_{n \in \mathbb{N}} \left(\left| S'_{k(n)} - \mathbb{E} \left(S'_{k(n)} \right) \right| > \varepsilon k(n) \right) &\leq \frac{1}{\varepsilon^2} \sum_{n \in \mathbb{N}} \frac{V(S'_{k(n)})}{k^2(n)} \\ &= \frac{1}{\varepsilon^2} \sum_{n \in \mathbb{N}} \frac{1}{k^2(n)} \sum_{j=1}^{k(n)} V(Y_j) = \frac{1}{\varepsilon^2} \sum_{j=1}^{k(n)} V(Y_j) \sum_{n: k(n) \geq j} \frac{1}{k^2(n)}, \end{aligned}$$

ove, nell’ultimo passaggio, è stato usato il teorema di Fubini. Ovviamente, si ha $[\alpha^n] \geq \alpha^n/2$, per ogni $n \in \mathbb{N}$. Perciò

$$\begin{aligned} \sum_{n: k(n) \geq j} \frac{1}{k^2(n)} &= \sum_{n: k(n) \geq j} \frac{1}{[\alpha^n]^2} \leq 4 \sum_{n: k(n) \geq j} \frac{1}{\alpha^{2n}} \\ &\leq \frac{4}{j^2} \sum_{k \in \mathbb{N}} \frac{1}{\alpha^{2k}} = \frac{1}{j^2} \frac{1}{\alpha^2 - 1}. \end{aligned}$$

Dalla maggiorazione (4.4.2) segue, applicando il teorema di Fubini, che

$$\begin{aligned} \sum_{n \in \mathbb{N}} \frac{V(Y_n)}{n^2} &\leq \sum_{n \in \mathbb{N}} \frac{1}{n^2} \mathbb{E}(X_n^2 \mathbf{1}_{\{X_n < n\}}) = \sum_{n \in \mathbb{N}} \frac{1}{n^2} \sum_{j=1}^n \mathbb{E}(X_1^2 \mathbf{1}_{\{j-1 \leq X_1 < j\}}) \\ &= \sum_{j \in \mathbb{N}} \mathbb{E}(X_1^2 \mathbf{1}_{\{j-1 \leq X_1 < j\}}) \sum_{n=j}^{\infty} \frac{1}{n^2} \leq 2 \sum_{j \in \mathbb{N}} \frac{\mathbb{E}(X_1^2 \mathbf{1}_{\{j-1 \leq X_1 < j\}})}{j} \\ &\leq 2 \sum_{j \in \mathbb{N}} \mathbb{E}(X_1 \mathbf{1}_{\{j-1 \leq X_1 < j\}}) = 2 \mathbb{E}(X_1) < +\infty. \end{aligned}$$

Perciò

$$\begin{aligned} \sum_{n \in \mathbb{N}} \mathbb{P}\left(\left|S'_{k(n)} - \mathbb{E}\left(S'_{k(n)}\right)\right| > \varepsilon k(n)\right) &\leq \frac{4}{\varepsilon^2(\alpha^2 - 1)} \sum_{j \in \mathbb{N}} \frac{V(Y_j)}{j^2} \\ &\leq \frac{2 \mathbb{E}(X_1)}{\varepsilon^2(\alpha^2 - 1)} < +\infty. \end{aligned}$$

Il primo lemma di Borel–Cantelli assicura ora che sia

$$\frac{S'_{k(n)} - \mathbb{E}\left(S'_{k(n)}\right)}{k(n)} \xrightarrow[n \rightarrow +\infty]{} 0 \quad \text{q.c..}$$

Il Teorema 4.4.3 assicura che sia $\mathbb{E}(Y_n) \xrightarrow[n \rightarrow +\infty]{} \mathbb{E}(X_1)$, e, quindi, dato che la convergenza di una successione implica la sua convergenza allo stesso limite nel senso di Cesàro,

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=1}^n \mathbb{E}(Y_j) = 0.$$

Pertanto

$$\frac{\mathbb{E}\left(S'_{k(n)}\right)}{k(n)} = \frac{1}{k(n)} \sum_{j=1}^{k(n)} \mathbb{E}(Y_j) \xrightarrow[n \rightarrow +\infty]{} \mathbb{E}(X_1),$$

sicché

$$\frac{S'_{k(n)}}{k(n)} \xrightarrow[n \rightarrow +\infty]{} \mathbb{E}(X_1) \quad \text{q.c..}$$

Se il numero naturale k è tale che $k(n) < k < k(n+1)$, si ha, visto che le v.a. X_n , e quindi le Y_n , sono positive,

$$S'_{k(n)} \leq S'_k \leq S'_{k(n+1)};$$

di qui

$$\frac{k(n)}{k} \frac{S'_{k(n)}}{k(n)} \leq \frac{S'_k}{k} \leq \frac{S'_{k(n+1)}}{k(n+1)} \frac{k(n+1)}{k}. \quad (4.4.3)$$

Dalla definizione di $k(n)$ si ottiene

$$k(n) \leq \alpha^n < k(n) + 1 \leq k < k(n+1) \leq \alpha^{n+1},$$

onde

$$\frac{k(n+1)}{k} \leq \frac{\alpha^{n+1}}{k} < \frac{\alpha^{n+1}}{\alpha^n} = \alpha \quad \text{e} \quad \frac{k(n)}{k} > \frac{\alpha^n - 1}{k} > \frac{\alpha^n - 1}{\alpha^{n+1}}.$$

Passando al limite per $n \rightarrow +\infty$ nella (4.4.3), si ottiene

$$\frac{1}{\alpha} \mathbb{E}(X_1) \leq \liminf_{n \rightarrow +\infty} \frac{S'_k}{k} \leq \limsup_{n \rightarrow +\infty} \frac{S'_k}{k} \leq \alpha \mathbb{E}(X_1),$$

che dà l'asserto in virtù dell'arbitrarietà di $\alpha > 1$. \square

Il Teorema 4.4.4 è ora un corollario banale del Teorema 4.4.5.

A completamento della disuguaglianza di Kolmogorov diamo la seguente disuguaglianza dovuta a Paul Lévy.

Lemma 4.4.1. (Disuguaglianza di Lévy). *Se per ogni $k \in \{1, \dots, n\}$ si ha*

$$\mathbb{P} \left(\left| \sum_{j=k}^n X_j \right| \geq \frac{\varepsilon}{2} \right) = \mathbb{P} \left(|S_n - S_{k-1}| \geq \frac{\varepsilon}{2} \right) \leq \delta, \quad (4.4.4)$$

allora

$$\mathbb{P} \left(\sup_{j \leq n} |S_j| \geq \varepsilon \right) \leq \frac{\delta}{1 - \delta}.$$

Dimostrazione. Si definiscano gli insiemi A_j come nella dimostrazione del Teorema 4.4.2. Se il punto ω appartiene a $A_j \cap \{|S_n| \leq \varepsilon/2\}$ si ha

$$-\frac{\varepsilon}{2} \leq S_n(\omega) = S_n(\omega) - S_j(\omega) + S_j(\omega) \leq S_n(\omega) - S_j(\omega) - \varepsilon;$$

e, poiché si ha contemporaneamente $S_n \leq \varepsilon/2$ e $S_j \geq \varepsilon$, anche

$$\frac{\varepsilon}{2} \geq S_n(\omega) - S_j(\omega) + \varepsilon,$$

sicché

$$\begin{aligned} \mathbb{P} \left(\left\{ \sup_{j \leq n} |S_j| \geq \varepsilon \right\} \cap \left\{ |S_n| \leq \frac{\varepsilon}{2} \right\} \right) &= \sum_{j=1}^n \mathbb{P} \left(A_j \cap \left\{ |S_n| \leq \frac{\varepsilon}{2} \right\} \right) \\ &\leq \sum_{j=1}^n \mathbb{P} \left(A_j \cap \left\{ |S_n - S_j| \geq \frac{\varepsilon}{2} \right\} \right) \\ &= \sum_{j=1}^n \mathbb{P}(A_j) \mathbb{P} \left(|S_n - S_j| \geq \frac{\varepsilon}{2} \right) \\ &\leq \delta \sum_{j=1}^n \mathbb{P}(A_j) = \delta \mathbb{P} \left(\sup_{j \leq n} |S_j| \geq \varepsilon \right). \end{aligned}$$

D'altro canto, è

$$\mathbb{P} \left(\left\{ \sup_{j \leq n} |S_j| \geq \varepsilon \right\} \cap \left\{ |S_n| > \frac{\varepsilon}{2} \right\} \right) \leq \mathbb{P} \left(|S_n| > \frac{\varepsilon}{2} \right) \leq \delta.$$

Sommando le due ultime disuguaglianze, si ottiene

$$\mathbb{P} \left(\sup_{j \leq n} |S_j| \geq \varepsilon \right) \leq \delta + \delta \mathbb{P} \left(\sup_{j \leq n} |S_j| \geq \varepsilon \right),$$

onde l'asserto. \square

Teorema 4.4.6. *Sia data una successione (X_n) di v.a. indipendenti definite nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$. Posto $S_n := \sum_{j=1}^n X_j$ ($n \in \mathbb{N}$), sono equivalenti le seguenti affermazioni:*

- (a) (S_n) converge in legge;
- (b) (S_n) converge in probabilità;
- (c) (S_n) converge q.c..

Dimostrazione. Basta, naturalmente, stabilire le implicazioni (a) \implies (b) \implies (c).

(a) \implies (b) Sia φ_n la f.c. di X_n . Allora

$$\varphi(t) := \prod_{n \in \mathbb{N}} \varphi_n(t)$$

è un prodotto infinito convergente, vale a dire φ è una f.c.. Siccome φ è continua nell'origine, ove è $\varphi(0) = 1$, essa è diversa da zero in un intervallo $[-T, T]$. Perciò, per ogni $t \in [-T, T]$, si ha

$$\lim_{\substack{n \rightarrow +\infty \\ m \rightarrow +\infty}} \prod_{j=m+1}^n \varphi_j(t) = 1. \quad (4.4.5)$$

Infatti $Z_n(t) := \sum_{j=1}^n \text{Log } \varphi_j(t)$ è una serie convergente, sicché soddisfa alla condizione di Cauchy; pertanto

$$\text{Log } \varphi(t) = \lim_{\substack{n \rightarrow +\infty \\ m \rightarrow +\infty}} \sum_{j=m+1}^n \text{Log } \varphi_j(t) = 0.$$

Dall'esercizio 3.5 segue che la relazione (4.4.5) vale per ogni $t \in \mathbb{R}$; ciò implica che la successione

$$(S_n - S_m)_{m, n \in \mathbb{N}, m < n}$$

converge in legge alla v.a. 0. Pertanto vi converge anche in probabilità, sicché

$$\lim_{\substack{n \rightarrow +\infty \\ m \rightarrow +\infty}} \mathbb{P}(|S_n - S_m| \geq \delta) = 0,$$

ciò che assicura che (S_n) sia una successione di Cauchy in probabilità e che, dunque, converga in probabilità.

(b) \implies (c) è una conseguenza immediata della disuguaglianza di Lévy. \square

4.5 Un'applicazione della LGN

Diamo qui un'applicazione della LGN che riveste grande importanza per la Teoria dell'Informazione (in forme meno semplici di quella qui presentata).

Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ si consideri una successione X_n di v.a. indipendenti e isonome, ognuna delle quali assume valori nell'insieme finito $S = \{s_1, s_2, \dots, s_r\}$. Si ponga $p_j := \mathbb{P}(X_n = s_j)$; poiché le v.a. hanno tutte la stessa legge, p_j non dipende da n . Si ponga inoltre

$$\Omega_n := S^n = \underbrace{S \times \dots \times S}_{n \text{ volte}}$$

e sulla famiglia delle parti di Ω_n si consideri la probabilità definita sui punti

$$\omega_n = (s^{(1)}, \dots, s^{(n)}) \in \Omega_n,$$

ove $s^{(j)} \in S$, mediante

$$\mathbb{P}_n(\{\omega_n\}) := \prod_{j=1}^n \mathbb{P}(s^{(j)}) = \prod_{j=1}^r p_j^{N_j^{(n)}(\omega)},$$

ove $N_j^{(n)}$ è la v.a. che conta quante volte nelle prime n prove le v.a. della successione X_n abbiano assunto il valore s_j ,

$$N_j^{(n)} := \sum_{k=1}^n \mathbf{1}_{\{X_k=s_j\}}.$$

Naturalmente, si può definire un'unica probabilità Q sul prodotto numerabile

$$\Omega_\infty = S^\infty = \prod_{n=1}^{\infty} S,$$

in modo che \mathbb{P}_n sia la restrizione di Q a $\mathcal{P}(\Omega_n)$ (si ricordi il Teorema 1.15.5).

Si chiama *entropia della successione* (X_n) , o, piú comunemente, *entropia di Shannon* della legge di probabilità (p_1, p_2, \dots, p_r) la quantità

$$H_r(p_1, \dots, p_r) := - \sum_{j=1}^r p_j \ln p_j. \quad (4.5.1)$$

In effetti, nella (4.5.1), si è soliti prendere i logaritmi in base 2 anziché in base e ; la differenza è, rispetto alla (4.5.1), data da una costante moltiplicativa.

Si osservi che l'entropia H_r dipende dalle probabilità p_1, \dots, p_r ma non dai valori assunti dalle v.a. della successione (X_n) .

Non è difficile provare, e per questo si vedano gli esercizi, che il massimo di H al variare di (p_1, \dots, p_r) tra le leggi di probabilità sull'insieme finito S , vale a dire quando $p_j \geq 0$ per ogni $j = 1, \dots, r$ e $\sum_{j=1}^r p_j = 1$, è dato da $\ln r$ e che il massimo è raggiunto dalla distribuzione uniforme su S ,

$$H_r(p_1, \dots, p_r) \leq H_r\left(\frac{1}{r}, \dots, \frac{1}{r}\right) = \ln r.$$

Con queste posizioni vale il seguente

Teorema 4.5.1. (MacMillan) *Per ogni $\alpha > 0$ e per ogni $\beta > 0$, esiste un naturale $n_0 := n_0(\alpha, \beta)$ tale che, per ogni $n \geq n_0$, si possa trovare un sottoinsieme C_n di Ω_n con le seguenti proprietà:*

(a) $Q(C_n) \geq 1 - \alpha;$

(b) per ogni $\omega \in C_n$ è

$$e^{-n(H_r+\beta)} \leq Q(\{\omega\}) = \mathbb{P}_n(\{\omega_n\}) \leq e^{-n(H_r-\beta)};$$

$$(c) \quad e^{n(H_r - \beta)} \leq \text{card}(C_n) \leq e^{n(H_r + \beta)},$$

ove $H_r := H_r(p_1, \dots, p_r)$.

Dimostrazione. (a) Si scelga $\delta = \delta(\beta) > 0$ in modo che sia

$$\delta \sum_{j=1}^r |\ln p_j| \leq \frac{\beta}{2},$$

e si ponga

$$C_n := \bigcap_{j=1}^r \left\{ \left| \frac{N_j^{(n)}}{n} - p_j \right| \leq \delta \right\}.$$

Segue dalla LGN debole (Teorema 4.3.3) che $Q(C_n) \xrightarrow{n \rightarrow +\infty} 1$, cioè $Q(C_n) \geq 1 - \alpha$ per n maggiore o eguale a un opportuno $n_1 = n_1(\alpha, \beta)$.

(b) Sia $\omega_n = (s^{(1)}, s^{(2)}, \dots, s^{(n)})$ un punto di C_n Allora

$$\begin{aligned} Q(\{\omega_n\} \times S \times S \times \dots) &= \mathbb{P}_n(\{\omega_n\}) = \prod_{j=1}^r p_j^{N_j^{(n)}(\omega_n)} \\ &= \exp \left(\sum_{j=1}^r N_j^{(n)}(\omega_n) \ln p_j \right) = \exp \left\{ -n \left(- \sum_{j=1}^r \frac{N_j^{(n)}}{n} \ln p_j \right) \right\} \\ &= \exp \left\{ -n \left(- \sum_{j=1}^r p_j \ln p_j - \sum_{j=1}^r \left(\frac{N_j^{(n)}}{n} - p_j \right) \ln p_j \right) \right\} \\ &= \exp \left\{ -n H_r + n \sum_{j=1}^r \left(\frac{N_j^{(n)}}{n} - p_j \right) \ln p_j \right\}. \end{aligned}$$

Poiché ω appartiene a C_n segue che

$$\left| \sum_{j=1}^r \left(\frac{N_j^{(n)}}{n} - p_j \right) \ln p_j \right| \leq \delta \sum_{j=1}^r |\ln p_j| \leq \frac{\beta}{2},$$

sicché

$$\begin{aligned} e^{-n(H_r + \beta)} &\leq \exp \left\{ -n \left(H_r + \frac{\beta}{2} \right) \right\} \leq Q(\{\omega\} \times S \times S \times \dots) \\ &= \mathbb{P}_n(\{\omega_n\}) \leq \exp \left\{ -n \left(H_r - \frac{\beta}{2} \right) \right\} \leq e^{-n(H_r - \beta)}. \end{aligned}$$

(c) Scende da (b) che

$$\begin{aligned} \exp \left\{ -n \left(H_r + \frac{\beta}{2} \right) \right\} \text{card}(C_n) &\leq Q(C_n) = \sum_{\omega \in C_n} Q(\{\omega\} \times S \times S \times \dots) \\ &\leq \exp \left\{ -n \left(H_r - \frac{\beta}{2} \right) \right\} \text{card}(C_n); \end{aligned}$$

e poiché

$$1 - \alpha \leq Q(C_n) = \sum_{\omega \in C_n} Q(\{\omega\} \times S \times S \times \dots) \leq 1,$$

si ha, per n abbastanza grande, diciamo per $n \geq n_2 = n_2(\alpha, \beta)$,

$$\begin{aligned} e^{n(H_r - \beta)} &\leq (1 - \alpha) e^{n(H_r - \beta/2)} \leq Q(C_n) e^{n(H_r - \beta/2)} \\ &\leq \text{card}(C_n) \leq Q(C_n) e^{n(H_r + \beta/2)} \leq e^{n(H_r + \beta)}, \end{aligned}$$

onde l'asserto ponendo $n_0 := n_1 \vee n_2$. \square

4.6 Note al Capitolo 4

Sezione 4.1 Il nome “Teorema del limite centrale” fu introdotto da Pólya nel 1920.

Come già detto, il primo caso di tale teorema fu dimostrato per l'approssimazione della legge binomiale per $p = 1/2$ da de Moivre nel 1733. La prima dimostrazione completa in un caso generale fu data da Lyapunov (1901). Il teorema che diamo nelle lezioni è dovuto al finlandese Lindeberg (1922). Il libro di Adams (2009) dà la storia dello sviluppo del TLC. Molto interessante anche l'articolo (Le Cam, 1986). Infine una notazione di linguaggio; in italiano, e in tutte le lingue neolatine, si oscilla tra le dizioni “Teorema del limite centrale” (qui adottata) e “Teorema centrale del limite”. Nella prima dizione si pone l'accento sul fatto che si considerano v.a. centrate, mentre nella seconda si sottolinea l'importanza del teorema in Probabilità.

Per il TLC il riferimento classico è (Gnedenko & Kolmogorov, 1954). Per una trattazione piú recente con un ampliamento dedicato alle v.a. con valori in uno spazio di Banach, si veda (Araujo & Giné, 1980).

Buona parte dei libri dedicati alla probabilità sorvola senza la dovuta attenzione sulla definizione del ramo principale del logaritmo. La questione è trattata in maniera rigorosa da Chung; si veda il Teorema 7.6.1 in (Chung, 1974), qui riprodotto.

Teorema 4.6.1. *Sia data la funzione $f : \mathbb{C} \rightarrow \mathbb{R}$ e si supponga che sia $f(0) = 1$, che esista $T > 0$ tale che f sia continua nell'intervallo $[-T, T]$ e che f non si annulli in tale intervallo. Esiste allora un'unica funzione $\text{Log} : [-T, T] \rightarrow \mathbb{C}$ tale che*

- (a) $\text{Log}(0) = 0$;
- (b) per ogni $t \in [-T, T]$ risulti $f(t) = \exp(\text{Log}(t))$.

Il risultato continua a valere se $[-T, T]$ è sostituito da \mathbb{R} .

Sezione 4.2 Le condizioni necessarie perché valga il TLC comparvero in due articoli di Feller (1935, 1937) scritti nel periodo che egli trascorse a Stoccolma dopo che fu obbligato a lasciare il suo posto all'Università di Kiel in seguito all'avvento del Nazismo e prima di stabilirsi definitivamente negli Stati Uniti.

Sezione 4.4 Sulle leggi dei grandi numeri si possono leggere il recente articolo di Seneta (2013) e i due articoli di Regazzini (2005, 2006), utilissimi per la storia dei teoremi che sono noti sotto questo nome e per il legame con la nozione stessa di probabilità.

Rajchman è uno dei matematici polacchi vittime del nazismo nella Polonia occupata durante la Seconda Guerra Mondiale. Si veda il primo numero della rivista *Fundamenta Mathematicae* pubblicato dopo la sospensione causata dalla guerra, *Fund. Math.* **33** (1945).

La diseguaglianza (??) fu introdotta da Kolmogorov (1928); essa è utilissima in tutte le questioni riguardanti somme di v.a. indipendenti. Nello stesso articolo Kolmogorov dà nel Teorema 11 una condizione necessaria e sufficiente perché valga la LGN debole; la dimostrazione dipende però dalla diseguaglianza di Kolmogorov. La condizione necessaria e sufficiente è costituita dai tre limiti

$$\begin{aligned} \lim_{n \rightarrow +\infty} \sum_{j=1}^n \mathbb{P}(|X_j| > n) &= 0 \\ \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=1}^n \mathbb{E} \left(X_j \mathbf{1}_{\{|X_j| \leq n\}} \right) &= 0 \\ \lim_{n \rightarrow +\infty} \frac{1}{n^2} \sum_{j=1}^n \mathbb{E} \left(X_j^2 \mathbf{1}_{\{|X_j| \leq n\}} \right) &= 0. \end{aligned}$$

Si osservi che le *Osservazioni* all'articolo contengono la correzione di un errore. Anche i sommi talvolta sbagliano!

Per il Teorema 4.4.5 si veda (Etemadi, 1981). Il Teorema 4.4.4 è di solito dimostrato direttamente e non come corollario del Teorema 4.4.5 di Etemadi. Per la dimostrazione usuale si veda, per esempio, (Rao, 1984); una dimostrazione ancora differente si trova in (Grimmett & Stirzaker, 2001).

Sezione 4.5 Questa sezione è ricalcata sulla trattazione di Sinai (1992). Per un'introduzione alla Teoria dell'Informazione si possono consultare (Ash, 1965), (Kullback, 1968), (Csiszár & Körner, 1986).

4.7 Esercizi sul Capitolo 4

4.1. Se $f : \mathbb{R} \rightarrow \mathbb{C}$ è derivabile n volte e le sue derivate sono continue nell'intervallo $I = [-a, a]$, allora si ha

$$f(t) = \sum_{j=0}^{n-1} \frac{f^{(j)}(0)t^j}{j!} + t^n \int_0^1 \frac{(1-s)^{n-1}}{(n-1)!} f^{(n)}(st) ds.$$

Se $|f^{(n)}(t)| \leq k$ per ogni $t \in I$, allora

$$f(t) = \sum_{j=0}^{n-1} \frac{f^{(j)}(0)t^j}{j!} + \frac{\theta(t)|t|^n}{n!} \quad \text{con } |\theta| \leq k.$$

4.2. Si mostri la validità delle (4.1.3) e delle (4.1.4).

4.3. Si mostri la validità della (4.1.5).

4.4. Si studi la convergenza della successione di v.a.

$$Z_n := \frac{\sqrt{n}X_0}{\sqrt{\sum_{j=1}^n X_j^2}},$$

ove le v.a. della successione $(X_n)_{n \in \mathbb{Z}_+}$ sono indipendenti e tutte di legge $N(0, 1)$. Si tenga presente l'esercizio (4.78) e si paragonino i risultati con quelli dell'esercizio (5.35).

4.5. (Un inverso della LGN di Khinchin) Sia (X_n) una successione di v.a. indipendenti ed isonome. Se

$$\frac{1}{n} \sum_{j=1}^n X_j$$

converge q.c. ad un limite finito, allora X_1 è in L^1 (e, quindi, lo sono tutte le v.a. della successione) ed il limite è eguale a $\mathbb{E}(X_1)$.

4.6. Sia (X_n) una successione di v.a. incorrelate e centrate di L^2 con $V(X_n) = 1$ per ogni $n \in \mathbb{N}$. Allora, se $\alpha > 1$, riesce

$$\frac{S_n}{n^\alpha} \xrightarrow{n \rightarrow +\infty} 0 \quad \text{q.c.}$$

4.7. Sia (X_n) una successione di v.a. incorrelate e centrate di L^2 con le varianze uniformemente limitate, tali cioè che

$$\sup_{n \in \mathbb{N}} V(X_n) < +\infty.$$

Se $\alpha > 1/2$, allora

$$\frac{S_n}{n^\alpha} \xrightarrow{n \rightarrow +\infty} 0 \quad \text{in probabilità.}$$

4.8. Se la successione di v.a. (X_n) di L^2 obbedisce al TLC e se è verificata la condizione

$$\lim_{n \rightarrow +\infty} \max_{j \leq n} \frac{\sigma_j^2}{c_n^2} = 0, \quad (\text{a})$$

allora la condizione di Lindeberg è verificata.

La condizione (a) equivale alle altre due, considerate congiuntamente,

$$\lim_{n \rightarrow +\infty} \frac{\sigma_n^2}{c_n^2} = 0 \quad \text{e} \quad \sum_{n \in \mathbb{N}} \sigma_n^2 = +\infty. \quad (\text{b})$$

4.9. Sia (X_n) una successione di v.a. indipendenti, tutte di legge esponenziale di parametro 1, $X_n \sim \Gamma(1, 1)$ per ogni $n \in \mathbb{N}$.

(a) Si introducano le v.a. $S_n := \sum_{j=1}^n X_j$ e se ne scriva la legge;

(b) dopo aver scritto la f.c. della v.a. ridotta

$$Z_n := \frac{S_n - \mathbb{E}(S_n)}{\sigma(S_n)},$$

si mostri direttamente che (X_n) obbedisce al TLC;

(c) sfruttando (b) e il teorema d'inversione per le f.c. integrabili si ottenga la formula di Stirling nella versione

$$\lim_{n \rightarrow +\infty} \frac{n^{n-\frac{1}{2}} e^{-n} \sqrt{2\pi}}{\Gamma(n)} = 1.$$

4.10. Sia (X_n) una successione di v.a. di L^2 che soddisfa alla condizione di Lindeberg. Si mostri che

$$\lim_{n \rightarrow +\infty} \sup_{k \leq n} \frac{V(X_n)}{V(S_n)} = 0,$$

ove, al solito, $S_n := \sum_{j=1}^n X_j$.

4.11. Sia (X_n) un processo di Bernoulli con

$$\mathbb{P}(X_n = 0) = \mathbb{P}(X_n = 1) = \frac{1}{2}.$$

Posto, al solito, $S_n = \sum_{j=1}^n X_j$, la LGN di Bernoulli assicura che, per ogni $\varepsilon > 0$, valga

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\left| \frac{S_n}{n} - \frac{1}{2} \right| \geq \varepsilon \right) = 0.$$

Mostriamo che la convergenza a zero avviene esponenzialmente. Si definiscano due funzioni $\varphi : [0, 1] \rightarrow \mathbb{R}$ e $\mathcal{I}_2 : \mathbb{R} \rightarrow \mathbb{R}$ mediante

$$\varphi(x) := \begin{cases} 0, & x = 0, \\ x \ln x, & x \in]0, 1] \end{cases},$$

e

$$\mathcal{I}_2(x) := \begin{cases} \varphi(x) + \varphi(1-x) + \ln 2, & x \in [0, 1], \\ +\infty, & x \notin [0, 1]. \end{cases}$$

Allora, se $A := \{x \in \mathbb{R} : |x - \frac{1}{2}| \geq \varepsilon\}$, si ha

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \ln \mathbb{P} \left(\left| \frac{S_n}{n} - \frac{1}{2} \right| \geq \varepsilon \right) = - \min_{x \in A} \mathcal{I}_2(x).$$

Per questo risultato, che è il primo teorema sulle grandi deviazioni, si veda (Cramér, 1938). Per un'introduzione all'argomento delle grandi deviazioni si può consultare (Dembo & Zeituni, 1998).

4.12. Sia (X_n) una successione di v.a. indipendenti, tutte di legge $N(0, 1)$ e tutte definite sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, e si consideri la successione (S_n) con $S_n := \sum_{j=1}^n X_j^2$. Si studi la convergenza in legge della successione

$$\left(\frac{S_n - n}{\sqrt{2n}} \right)_{n \in \mathbb{N}}.$$

4.13. Sia (X_n) una successione di v.a. indipendenti e centrate, tutte definite sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$; se (X_n) obbedisce alla LGN forte, allora, per ogni $\varepsilon > 0$,

$$\sum_{n \in \mathbb{N}} \mathbb{P}(|X_n| \geq n\varepsilon) < +\infty.$$

4.14. Sia data una successione (X_n) di v.a. indipendenti ed isonome sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$. Si consideri, per ogni $n \in \mathbb{N}$, la funzione $\bar{F}_n : \mathbb{R} \times \Omega \rightarrow [0, 1]$ definita da

$$\bar{F}_n(x, \omega) := \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{X_j \leq x\}}(\omega).$$

- (a) Per ogni $n \in \mathbb{N}$ e per ogni $x \in \mathbb{R}$, $\omega \mapsto \bar{F}_n(x, \omega)$ è una v.a.;
- (b) Per ogni $n \in \mathbb{N}$, esiste un insieme $N_n \in \mathcal{F}$ con $\mathbb{P}(N_n) = 0$, tale che, per ogni $\omega \in N_n^c$, la funzione $x \mapsto \bar{F}_n(x, \omega)$ sia una f.r.; di piú, tranne che per i punti di un insieme di probabilità nulla, $\{\bar{F}_n(\cdot, \omega) : n \in \mathbb{N}\}$ è una successione di f.r.;
- (c) $\bar{F}_n(x, \omega) \xrightarrow[n \rightarrow +\infty]{} F(x)$ q.c., ove F è la f.r. comune delle v.a. X_n ;
- (d) $\sqrt{n} \{\bar{F}_n(x, \omega) - F(x)\} \xrightarrow[n \rightarrow +\infty]{} N(0, F(x)\{1 - F(x)\})$ in legge.

La funzione \bar{F}_n è detta, nella Statistica Matematica, *funzione di ripartizione empirica*.

4.15. Sia (X_n) una successione di v.a. indipendenti e centrate di L^2 che soddisfa alla condizione di Lindeberg (4.1.1). Se $V(X_n) = \sigma_n^2$ per ogni $n \in \mathbb{N}$ e se Y è una v.a. di L^2 con $\mathbb{E}(Y) = 0$ e $V(Y) = 1$, la successione (Y_n) ove $Y_n := \sigma_n Y$ soddisfa alla condizione di Lindeberg.

4.16. Come preliminare alla dimostrazione di Trotter del TLC, dimostrazione che non fa ricorso alle f.r., si consideri la seguente situazione.

Sia X una v.a. sullo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e sia $U_b = U_b(\mathbb{R})$ lo spazio delle funzioni $f : \mathbb{R} \rightarrow \mathbb{R}$ limitate e uniformemente continue, munito della norma $\|f\| := \sup_{x \in \mathbb{R}} |f(x)|$. Si definisca in U_b l'operatore lineare T_X mediante

$$(T_X f)(t) := \mathbb{E}[f \circ (X + t)] = \int_{\mathbb{R}} f(x + t) dF_X(x) \quad (t \in \mathbb{R}).$$

Si mostri che

- (a) Per ogni $t \in \mathbb{R}$, $\|(T_X f)(t)\| \leq \|f\|$, sicché T_X è una contrazione;
- (b) per ogni coppia t_1, t_2 di numeri reali, è

$$|(T_X f)(t_1) - (T_X f)(t_2)| \leq \sup_{y \in \mathbb{R}} |f(y + t_1) - f(y + t_2)|,$$

sicché $T_X f \in U_b$ e $T_X : U_b \rightarrow U_b$;

- (c) se X_1 e X_2 sono indipendenti, $T_{X_1+X_2} = T_{X_1} T_{X_2} = T_{X_2} T_{X_1}$;

(d) se U_b^2 è il sottoinsieme di U_b formato dalle funzioni che sono derivabili due volte e tali che entrambe le derivate appartengano ancora a U_b , la condizione

$$\forall f \in U_b^2 \quad \lim_{n \rightarrow +\infty} \|T_{X_n} f - T_X f\| = 0$$

assicura che (X_n) converga in legge a X ;

(e) se X e Y sono indipendenti, allora

$$\forall f \in U_b \quad \|T_X^n f - T_Y^n f\| \leq n \|T_X f - T_Y f\|;$$

(f) se $X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_n$ sono indipendenti, allora, per ogni $f \in U_b$, si ha

$$\|T_{X_1} T_{X_2} \dots T_{X_n} f - T_{Y_1} T_{Y_2} \dots T_{Y_n} f\| \leq \sum_{j=1}^n \|T_{X_j} f - T_{Y_j} f\|.$$

Si veda (Trotter, 1959).

4.17. Ricorrendo ai risultati dell'esercizio precedente si dimostri il teorema di Lindeberg, prima nel caso di v.a. isonome e poi nel caso generale.

4.18. Sia U una v.a. con legge uniforme su $(0, 2\pi)$ e si definisca $X_n := \sin nU$ ($n \in \mathbb{N}$). Allora la successione (X_n) tende a zero q.c. nel senso di Cesàro.

4.19. Sia (X_n) una successione di v.a. indipendenti ed isonome di L^2 . Si mostri per $\{X_n\}$ la LGN debole è conseguenza del TLC.

4.20. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, si consideri il quadro

$$\begin{array}{cccc} X_{11}, & X_{12}, & \dots, & X_{1k_1} \\ X_{21}, & X_{22}, & \dots, & X_{2k_2} \\ \dots & \dots & \dots & \dots \\ X_{n1}, & X_{n2}, & \dots, & X_{nk_n} \end{array}$$

formato da v.a. $(\{X_{nj}\})$ con $n \in \mathbb{N}$, $j = 1, 2, \dots, k_n$, ove $\lim_{n \rightarrow +\infty} k_n = +\infty$. Per $\varepsilon > 0$ si considerino le condizioni

$$\forall j \quad \lim_{n \rightarrow +\infty} \mathbb{P}(|X_{nj}| > \varepsilon) = 0; \quad (\text{a})$$

$$\lim_{n \rightarrow +\infty} \max_{j=1,2,\dots,k_n} \mathbb{P}(|X_{nj}| > \varepsilon) = 0; \quad (\text{b})$$

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\max_{j=1,2,\dots,k_n} |X_{nj}| > \varepsilon \right) = 0; \quad (\text{c})$$

$$\lim_{n \rightarrow +\infty} \sum_{j=1}^{k_n} \mathbb{P}(|X_{nj}| > \varepsilon) = 0. \quad (\text{d})$$

Si mostri che

- (d) \implies (c) \implies (b) \implies (a);
- se le v.a. $\{X_{nj} : j = 1, 2, \dots, k_n\}$ di ogni riga sono indipendenti, allora le condizioni (c) e (d) si equivalgono;

– le implicazioni (d) \implies (c) \implies (b) sono strette.

4.21. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, per ogni $n \in \mathbb{N}$, la v.a. X_n abbia legge uniforme su $(-n, n)$. Si mostri che la successione (X_n) verifica la condizione di Lindeberg.

4.22. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, si considerino le v.a. a due a due incorrelate, definite, per $n \in \mathbb{N}$ e per $\alpha > 1$, da

$$\begin{aligned}\mathbb{P}(X_n = n^\alpha) &= \mathbb{P}(X_n = -n^\alpha) = \frac{1}{6n^{2(\alpha-1)}}, \\ \mathbb{P}(X_n = 0) &= 1 - \frac{1}{3n^{2(\alpha-1)}}.\end{aligned}$$

Si mostri che la successione (X_n) verifica la condizione di Lindeberg se, e solo se, $\alpha < 3/2$.

4.23. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia data la successione di v.a. (X_n) ; per ogni $j \geq 2$, la v.a. X_j dipende solo dalle v.a. X_{j-1} e X_{j+1} , mentre è indipendente dalle rimanenti. Se le varianze sono uniformemente limitate,

$$\forall n \in \mathbb{N} \quad V(X_n) \leq H,$$

allora vale la LGN debole.

4.24. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia data la successione di v.a. (X_n) tali che, per ogni $n \in \mathbb{N}$, sia

$$V(X_n) \leq H \quad \text{e} \quad \text{Cov}(X_j, X_k) < 0 \quad (j, k = 1, 2, \dots, n; j \neq k).$$

Allora (X_N) obbedisce alla LGN debole.

4.25. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia data la successione di v.a. (X_n) . Se le varianze delle v.a. della successione sono uniformemente limitate, $V(X_n) \leq H$ per ogni $n \in \mathbb{N}$, e se

$$\lim_{|j-k| \rightarrow +\infty} \text{Cov}(X_j, X_k) = 0,$$

allora (X_n) obbedisce alla LGN debole.

4.26. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia (X_n) una successione di v.a. incorrelate, isonome e centrate di L^2 . Si adatti la dimostrazione del Teorema di Rajchman e si mostri dapprima che

$$\text{se } \alpha > \frac{3}{4}, \text{ allora } \frac{S_n}{n^\alpha} \xrightarrow{n \rightarrow +\infty} 0 \text{ q.c.} \quad (\text{a})$$

e, quindi, adattando ancora quest'ultima dimostrazione, si mostri che

$$\frac{S_n}{n^\alpha} \xrightarrow{n \rightarrow +\infty} 0 \text{ q.c.} \quad \text{per ogni } \alpha > \frac{1}{2}.$$

4.27. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia (X_n) una successione di v.a. indipendenti, isonome e centrate di L^2 . Si dimostri che la successione (Z_n) , ove

$$Z_n := \frac{(S_n)^2}{n} = \frac{1}{n} \left(\sum_{j=1}^n X_j \right)^2$$

converge in legge e se ne trovi il limite.

4.28. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia (X_n) una successione di v.a. indipendenti ed isonome a valori in $[1, +\infty[$.

- (a) A quali condizioni deve soddisfare il parametro reale λ affinché, per $t \geq 1$ e per ogni $n \in \mathbb{N}$, sia possibile l'eguaglianza

$$\mathbb{P}(X_n \geq t) = t^{-\lambda} ?$$

- (b) Si calcolino la media e la varianza di X_n per quei valori di λ per i quali queste esistono.
 (c) Si mostri che converge q.c. la successione

$$\left\{ \left(\prod_{j=1}^n X_j \right)^{1/n} : n \in \mathbb{N} \right\}$$

e se ne determini il limite.

4.29. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia (X_n) una successione di vettori aleatori a valori in \mathbb{R}^k indipendenti ed isonomi con vettore delle medie eguale a \mathbf{m} e matrice di varianza-covarianza Γ . Vale, allora, la seguente versione vettoriale del TLC:

$$\frac{1}{\sqrt{n}} \left(\sum_{j=1}^n X_j - n \mathbf{m} \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} N_k(\mathbf{0}, \Gamma),$$

ove $\mathbf{0}$ è il vettore nullo in \mathbb{R}^k .

4.30. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ la successione (X_n) di v.a. indipendenti è tale che

$$\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = \frac{1}{2} \left(1 - \frac{1}{2^n} \right), \quad \mathbb{P}(X_n = 0) = \frac{1}{2^n}.$$

Allora (X_n) obbedisce al TLC.

4.31. Si mostri come far discendere il TLC nel caso di una successione (X_n) di v.a. indipendenti ed isonome, vale a dire il Corollario 4.1.2, dal Teorema 4.1.2.

4.32. Siano z_1, z_2, \dots, z_n e z'_1, z'_2, \dots, z'_n numeri complessi tali che, per ogni $j = 1, 2, \dots, n$, risulti $|z_j| \leq 1$ e $|z'_j| \leq 1$; allora,

$$\left| \prod_{j=1}^n z_j - \prod_{j=1}^n z'_j \right| \leq \sum_{j=1}^n |z_j - z'_j|.$$

4.33. (Convergenza alla legge di Poisson) Per ogni $n \in \mathbb{N}$, siano $X_{n1}, X_{n2}, \dots, X_{nn}$ v.a. bernoulliane indipendenti su $(\Omega, \mathcal{F}, \mathbb{P})$ con

$$\mathbb{P}(X_{nj} = 1) = p_{nj} \quad \mathbb{P}(X_{nj} = 0) = q_{nj}$$

e si supponga che sia

$$\sum_{j=1}^n p_{nj} \xrightarrow{n \rightarrow +\infty} \lambda > 0 \quad \text{e} \quad \max_{j \leq n} p_{nj} \xrightarrow{n \rightarrow +\infty} 0.$$

Allora, posto $S_n := \sum_{j=1}^n X_{nj}$, la successione (S_n) converge in legge alla distribuzione di Poisson di parametro λ , $\mathcal{P}(\lambda)$.

4.34. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia (X_n) una successione di v.a. tali che

$$\mathbb{P}(X_n = n) = \mathbb{P}(X_n = -n) = \frac{1}{2n^2}, \quad \mathbb{P}(X_n = 0) = 1 - \frac{1}{n^2}.$$

Obbedisce al TLC questa successione?

4.35. Sia (X_n) una successione di v.a. definite sullo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$. Se, al solito, $S_n := \sum_{j=1}^n X_j$, la successione (X_n) obbedisce alla LGN debole, se, e solo se, vale la condizione

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left(\frac{S_n^2}{n^2 + S_n^2} \right) = 0.$$

4.36. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia (X_n) una successione di v.a. indipendenti, isonome, tutte di legge di Cauchy di parametri 0 e 1, $X_n \sim C(0, 1)$ ($n \in \mathbb{N}$). Se $S_n := \sum_{j=1}^n X_j$, si mostri che *non* vale

$$\frac{S_n}{n} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0;$$

pertanto, la successione (X_n) non obbedisce alla LGN debole.

4.37. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ siano $(X_n)_{n \in \mathbb{N}}$ e $(Y_m)_{m \in \mathbb{N}}$ due successioni indipendenti, ciascuna formata da v.a. indipendenti di leggi $X_n \sim \mathcal{P}(n)$ ($n \in \mathbb{N}$) e $Y_m \sim \mathcal{P}(m)$ ($m \in \mathbb{N}$). Si mostri che la v.a.

$$Z_{n,m} := \frac{(X_n - n) - (Y_m - m)}{\sqrt{X_n + Y_m}}$$

è asintoticamente ben definita e che, per $n, m \rightarrow +\infty$, tende in legge alla distribuzione $N(0, 1)$.

4.38. Certi eventi accadono ai tempi T_1, T_2, \dots , ove, per ogni $n \in \mathbb{N}$,

$$T_n := \sum_{j=1}^n X_j;$$

Le v.a. X_n sono indipendenti, isonome, positive e in L^1 (hanno, cioè, speranza finita m). Per $t > 0$, sia $N(t) := \max\{n : T_n \leq t\}$ il numero di eventi realizzatisi entro il tempo t . Si mostri che valgono le affermazioni

$$N(t) \xrightarrow[t \rightarrow +\infty]{} 0 \quad \text{q.c.}, \quad (\text{a})$$

$$\frac{N(t)}{t} \xrightarrow[t \rightarrow +\infty]{} \frac{1}{m} \quad \text{q.c.} \quad (\text{b})$$

4.39. In uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ ogni successione (X_n) di v.a. indipendenti, isonome di L^2 obbedisce alla LGN debole.

4.40. Sia (X_n) una successione di v.a. indipendenti di legge

$$\mathbb{P}(X_n = n^\lambda) = \mathbb{P}(X_n = -n^\lambda) = \frac{1}{2},$$

ove $\lambda > 0$. Allora (X_n) obbedisce al TLC per ogni $\lambda > 0$, ma non obbedisce alla LGN debole se $\lambda \geq \frac{1}{2}$.

4.41. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, sia (X_n) una successione di v.a. limitata in L^2 , $\mathbb{E}(X_n^2) \leq c < +\infty$ per ogni $n \in \mathbb{N}$, e tale che $\mathbb{E}(X_j X_k) = 0$ se $j \neq k$. Se, al solito, $S_n := \sum_{j=1}^n X_j$, allora

$$\frac{S_n}{n} \xrightarrow[n \rightarrow +\infty]{} 0 \quad \text{in } L^2 \text{ e in probabilità.}$$

4.42. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, sia (X_n) una successione di v.a. indipendenti ed isonome di L^1 . Posto, per ogni $n \in \mathbb{N}$, $Y_n := e^{X_n}$, si mostri che la successione

$$\left(\left(\prod_{j=1}^n Y_j \right)^{1/n} \right)$$

converge q.c. ad una costante.

4.43. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, siano (X_n) e (Y_n) due successioni, ciascuna delle quali è formata da v.a. indipendenti ed isonome di L^1 ; si supponga, inoltre, che sia, per ogni $n \in \mathbb{N}$, $\mathbb{E}(X_n) = \alpha$ e $\mathbb{E}(Y_n) = \beta$. Si mostri che

$$\frac{\sum_{j=1}^n X_j}{\sum_{j=1}^n Y_j} \xrightarrow[n \rightarrow +\infty]{} \frac{\alpha}{\beta} \quad \text{q.c..}$$

4.44. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, sia (X_n) una successione di v.a. indipendenti ed isonome di L^p ; allora

$$\frac{1}{n} \sum_{j=1}^n X_j^p \xrightarrow[n \rightarrow +\infty]{} \mathbb{E}(X_1^p) \quad \text{q.c..}$$

4.45. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, sia (X_n) una successione di v.a. indipendenti ed isonome, tutte di legge $N(1, \sigma^2)$; allora

$$\frac{\sum_{j=1}^n X_j}{\sum_{j=1}^n X_j^2} \xrightarrow[n \rightarrow +\infty]{} \frac{1}{\sigma^2 + 1} \quad \text{q.c..}$$

4.46. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, sia (X_n) una successione di v.a. indipendenti, tutte di legge $C(0, 1)$. Si mostri che non esiste alcuna costante $\alpha \in \mathbb{R}$ tale che

$$\frac{S_n}{n} \xrightarrow[n \rightarrow +\infty]{} \alpha \quad \text{in probabilità o q.c.};$$

qui, al solito, $S_n := \sum_{j=1}^n X_j$. Si mostri, invece che S_n/n converge in legge e se ne trovi il limite.

4.47. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, sia (X_n) una successione di v.a. indipendenti ed isonome di L^1 a valori interi, $X : \Omega \rightarrow \mathbb{Z}$, con $\mathbb{E}(X_1) > 0$. Posto $S_n := \sum_{j=1}^n X_j$, si mostri che

$$S_n \xrightarrow[n \rightarrow +\infty]{} +\infty \quad \text{q.c..}$$

4.48. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, sia (X_n) di v.a. indipendenti ed isonome tali che $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = 0) = 1/2$. Se $S_n := \sum_{j=1}^n X_j$ si ponga $G_n := 2S_n - n$ (si tratta della v.a. che dà la posizione in una passeggiata aleatoria simmetrica); allora, per ogni $n \in \mathbb{N}$,

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\frac{G_n}{n} < x \right) = \varphi(x),$$

ove φ è la f.r. della legge $N(0, 1)$.

4.49. Sia (X_n) una successione di v.a. indipendenti tutte di legge di Laplace, con densità

$$f(x) = \frac{1}{2} e^{-|x|}.$$

Si mostri che

$$\sqrt{n} \left(\frac{\sum_{j=1}^n X_j}{\sum_{j=1}^n X_j^2} \right)$$

converge in legge a $N(0, 1/2)$.

4.50. Nello spazio di probabilità $([0, 1], \mathcal{B}([0, 1]), \lambda)$, λ è la (restrizione della) misura di Lebesgue, si consideri la successione (X_n) , ove, per ogni $n \in \mathbb{N}$,

$$X_n := n \mathbf{1}_{[0, 1/n]} + \mathbf{1}_{[1/n, 1]},$$

e sia Y una v.a. di legge $N(0, 1)$. Per $n \in \mathbb{N}$ si ponga $Z_n := X_n Y$. Si mostri, senza ricorrere al teorema di Lindeberg, che

- (a) $\mathbb{E}(X_n^2) \geq n$ ($n \in \mathbb{N}$);
- (b) $\mathbb{E}(Z_n) = 0$ ($n \in \mathbb{N}$);
- (c) $\lim_{n \rightarrow +\infty} V(Z_n) = +\infty$;
- (d) $Z_n \xrightarrow[n \rightarrow +\infty]{} N(0, 1)$.

4.51. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, sia (X_n) una successione di v.a. indipendenti ed isonome di L^2 con $\mathbb{E}(X_1) = 1$ e $V(X_1) = \sigma^2$. Si mostri che, se $S_n := \sum_{j=1}^n X_j$, allora

$$\frac{2}{\sigma} \left(\sqrt{S_n} - \sqrt{n} \right) \xrightarrow[n \rightarrow +\infty]{} N(0, 1) \quad \text{in legge.}$$

4.52. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, sia (X_n) una successione di v.a. indipendenti ed isonome, tutte di legge uniforme in $(-1, 1)$, $X_n \sim U(-1, 1)$. Introdotte, per ogni $n \in \mathbb{N}$, la v.a.

$$Y_n := \frac{\sum_{j=1}^n X_j}{\sum_{j=1}^n X_j^2 + \sum_{j=1}^n X_j^3},$$

si mostri che la successione $(\sqrt{n} Y_n)$ converge in legge.

4.53. (a) Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, sia (X_n) una successione di v.a. indipendenti e centrate di L^p con $p \in [1, 2]$. Se converge la serie

$$\sum_{n \in \mathbb{N}} \frac{\mathbb{E}(|X_n|^p)}{n^p},$$

allora converge q.c. la serie aleatoria

$$\sum_{n \in \mathbb{N}} \frac{X_n}{n}.$$

(b) Viceversa, se è divergente la serie

$$\sum_{n \in \mathbb{N}} \frac{\beta_n^{(p)}}{n^p}, \quad p \in [1, 2]$$

allora si possono trovare uno spazio di probabilità e, su questo, una successione (X_n) di v.a. indipendenti e centrate di L^1 , tali che, per ogni $n \in \mathbb{N}$, si abbia

$$\mathbb{E}(|X_n|^p) = \beta_n^{(p)},$$

e che la serie

$$\sum_{n \in \mathbb{N}} \frac{X_n}{n}$$

non converga q.c..

(Si veda (Marcinkiewicz & Zygmund, 1937)).

Capitolo 5

Le Speranze Condizionate

5.1 La definizione

La definizione di Speranza Condizionata (=SC) si dà usando il teorema di Radon–Nikodym. Ricordo ancora che dati uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e una sottotribú \mathcal{G} di \mathcal{F} si indica con $\mathbb{P}_{\mathcal{G}}$ la *restrizione* di \mathbb{P} a \mathcal{G} , vale a dire la misura di probabilità definita, per ogni insieme B di \mathcal{G} , da $\mathbb{P}_{\mathcal{G}}(B) = \mathbb{P}(B)$.

Teorema 5.1.1. *Sia \mathcal{G} una tribú contenuta in \mathcal{F} e sia X una v.a. di $L^1(\mathcal{F})$. Esiste allora una v.a. $Y \in L^1(\mathcal{G})$, unica a meno di equivalenze, tale che per ogni $B \in \mathcal{G}$ risulti*

$$\int_B X \, d\mathbb{P} = \int_B Y \, d\mathbb{P}_{\mathcal{G}}.$$

Dimostrazione. Si scriva la v.a. X come differenza delle sue parti positiva e negativa $X = X^+ - X^-$. Le applicazioni

$$B \mapsto \mathbb{E}(\mathbf{1}_B X^+) \quad \text{and} \quad B \mapsto \mathbb{E}(\mathbf{1}_B X^-)$$

definiscono su \mathcal{G} due misure reali assolutamente continue rispetto a $\mathbb{P}_{\mathcal{G}}$. Esistono allora due funzioni φ^+ e φ^- , uniche a meno di equivalenze, tali che sia, per ogni insieme B di \mathcal{G}

$$\mathbb{E}(\mathbf{1}_B X^{\pm}) = \int_B \varphi^{\pm} \, d\mathbb{P}_{\mathcal{G}}.$$

Basta ora prendere $Y := \varphi^+ - \varphi^-$ per avere l'asserto. □

Definizione 5.1.1. Si dice *Speranza Condizionata* (=SC) della v.a. $X \in L^1(\mathcal{F})$ data la tribú $\mathcal{G} \subseteq \mathcal{F}$ l'unica v.a. di $L^1(\mathcal{G})$, che sarà denotata da $\mathbb{E}(X \mid \mathcal{G})$ o da $\mathbb{E}_{\mathcal{G}}(X)$ tale che, per ogni insieme B di \mathcal{G} ,

$$\int_B X \, d\mathbb{P} = \int_B \mathbb{E}_{\mathcal{G}}(X) \, d\mathbb{P}_{\mathcal{G}}. \tag{5.1.1}$$

L'unicità si intende nello spazio quoziente $L^1(\mathcal{G})$ e non in $\mathcal{L}^1(\mathcal{G})$, in altre parole, a meno di equivalenze. Se \mathcal{G} è la tribú generata da una v.a. Y su $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathcal{G} = \mathcal{F}(Y)$, si scrive $\mathbb{E}(X \mid Y)$ in luogo di $\mathbb{E}_{\mathcal{F}(Y)}(X)$. ◇

In genere vi saranno molte v.a. \mathcal{G} -misurabili che soddisfanno alla (5.1.1); ognuna di esse si dice *versione* della SC. Due qualsiasi versioni della SC sono eguali q.c. (rispetto a \mathbb{P}).

Usualmente, e quando ciò non generi confusione, si confonderanno la misura di probabilità \mathbb{P} e la sua restrizione $\mathbb{P}_{\mathcal{G}}$ alla tribù \mathcal{G} e si scriverà

$$\forall B \in \mathcal{G} \quad \int_B X \, d\mathbb{P} = \int_B \mathbb{E}_{\mathcal{G}}(X) \, d\mathbb{P}.$$

In qualche caso è possibile estendere la definizione di SC a v.a. che non sono in $L^1(\mathcal{F})$ e che, quindi, non ammettono speranza finita. Per esempio, se X è una v.a. positiva, ma non di L^1 , vale a dire che si ha $X \geq 0$ e $\mathbb{E}(X) = +\infty$, non si ha alcuna difficoltà a definire la SC di X .

Il risultato che segue consente di definire le probabilità condizionate data una tribù $\mathcal{G} \subseteq \mathcal{F}$ e di stabilire il legame con le SC.

Teorema 5.1.2. *Esiste un'unica funzione $\mathbb{P}(\cdot | \mathcal{G}) \in L^1(\mathcal{G})$ detta probabilità condizionata data la tribù \mathcal{G} , tale che per ogni $B \in \mathcal{G}$ e per ogni $A \in \mathcal{F}$ valga*

$$\mathbb{P}(A \cap B) = \int_B \mathbb{P}(A | \mathcal{G}) \, d\mathbb{P}_{\mathcal{G}}; \quad (5.1.2)$$

vale inoltre

$$\mathbb{P}(A | \mathcal{G}) = \mathbb{E}(\mathbf{1}_A | \mathcal{G}). \quad (5.1.3)$$

Dimostrazione. Ponendo, per $A \in \mathcal{F}$ fissato, $\lambda(B) := \mathbb{P}(A \cap B)$ si definisce su \mathcal{G} una misura assolutamente continua rispetto a $\mathbb{P}_{\mathcal{G}}$. Il teorema di Radon–Nikodym assicura l'esistenza e l'unicità, a meno di equivalenze, di $\mathbb{P}(\cdot | \mathcal{G})$. Ponendo $X = \mathbf{1}_A$ nella (5.1.1) si ottiene la (5.1.3) in virtù dell'unicità q.c. sia di $\mathbb{E}(\cdot | \mathcal{G})$ sia di $\mathbb{P}(\cdot | \mathcal{G})$. \square

Anche in questo caso si scriverà, in luogo della (5.1.2),

$$\mathbb{P}(A \cap B) = \int_B \mathbb{P}(A | \mathcal{G}) \, d\mathbb{P}.$$

5.2 Leggi condizionate

Siamo in grado ora di costruire un modello per la situazione che segue. In uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, sia X una v.a. e sia Y una seconda v.a. la cui legge dipende dal valore x assunto dalla v.a. X . Per esempio, se $x \in [0, 1]$, si può pensare che la v.a. Y rappresenti il numero di teste in n prove bernoulliane indipendenti con probabilità x di successo. Perciò, vorremmo calcolare ciò che intuitivamente potremmo denotare mediante

$$\mathbb{P}(x, B) = \mathbb{P}(Y \in B | X = x);$$

si noti che la definizione di probabilità condizionata data nei corsi introduttivi di probabilità può non aver senso se l'evento $\{X = x\}$ ha probabilità nulla; nell'esempio appena fatto ciò accade addirittura per ogni $x \in [0, 1]$.

Teorema 5.2.1. *Siano (Ω, \mathcal{F}) e (Ω', \mathcal{F}') due spazi misurabili, \mathbb{P} una probabilità su (Ω, \mathcal{F}) e $X : \Omega \rightarrow \Omega'$ una funzione misurabile (rispetto alle tribú \mathcal{F} e \mathcal{F}'). Dato un insieme $B \in \mathcal{F}$ esiste allora una funzione misurabile $\varphi_B : \Omega' \rightarrow \mathbb{R}$, tale che per ogni $A \in \mathcal{F}'$ sia*

$$\mathbb{P}(\{X \in A\} \cap B) = \int_A \varphi_B \, d\mathbb{P}_X \quad (5.2.1)$$

ove $\mathbb{P}_X = \mathbb{P} \circ X^{-1}$ è la legge immagine di \mathbb{P} mediante X . Se ψ è un'altra funzione che soddisfa alla (5.2.1) allora $\psi = \varphi_B$ q.c. rispetto a \mathbb{P}_X . Si pone $\mathbb{P}(B \mid X = x) := \varphi_B(x)$.

Dimostrazione. Ponendo $\mu_B(A) := \mathbb{P}(\{X \in A\} \cap B)$ per ogni $A \in \mathcal{F}'$, si definisce una misura finita su \mathcal{F}' che è assolutamente continua rispetto a \mathbb{P}_X . Basta ora applicare il teorema di Radon–Nikodym. \square

Si noti che, scritta nella forma,

$$\mathbb{P}(\{X \in A\} \cap B) = \int_A \mathbb{P}(B \mid X = x) \, d\mathbb{P}_X(x),$$

ove $A \in \mathcal{F}'$ e $B \in \mathcal{F}$, la (5.2.1) è una generalizzazione del teorema delle probabilità totali. In particolare, se $\Omega' = \mathbb{R}$ e se $A = \mathbb{R}$, si ha $\{X \in \mathbb{R}\} = \Omega$ e, dunque,

$$\mathbb{P}(B) = \int_{\mathbb{R}} \mathbb{P}(B \mid X = x) \, d\mathbb{P}_X(x) = \int_{\mathbb{R}} \mathbb{P}(B \mid X = x) \, dF(x),$$

ove F è la f.r. della v.a. X .

Esempio 5.2.1. Sia X una v.a. discreta e siano $x_1, x_2, \dots, x_n, \dots$ i valori che essa assume, con probabilità date da $p_j = \mathbb{P}(X = x_j) > 0$ ($j \in \mathbb{N}$). Mostriamo che

$$\varphi_B(x_j) = \mathbb{P}(B \mid X = x_j) = \frac{\mathbb{P}(B \cap \{X = x_j\})}{\mathbb{P}(X = x_j)} \quad (j \in \mathbb{N}). \quad (5.2.2)$$

Poiché \mathbb{P}_X è concentrata sui singoletti $\{x_j\}$ non occorre specificare φ_B per $x \neq x_j$. Si ponga $\Omega' = \{x_1, x_2, \dots, x_n, \dots\}$, $\mathcal{F}' = \mathcal{P}(\Omega')$. Se A appartiene a \mathcal{F}' e se φ_B è definita dalla (5.2.2), si ha

$$\begin{aligned} \int_A \varphi_B \, d\mathbb{P}_X &= \int_{\Omega'} \varphi_B \mathbf{1}_A \, d\mathbb{P}_X = \sum_{n \in \mathbb{N}} \varphi_B(x_n) \mathbf{1}_A(x_n) \mathbb{P}_X(\{x_n\}) \\ &= \sum_{x_n \in A} \varphi_B(x_n) \mathbb{P}(X = x_n) = \sum_{x_n \in A} \mathbb{P}(B \cap \{X = x_n\}) \\ &= \mathbb{P}(B \cap \{X \in A\}). \end{aligned}$$

Nel caso discreto la definizione del teorema 5.2.1 coincide dunque con quella elementare. \blacksquare

Esempio 5.2.2. Siano X e Y due v.a. con legge congiunta assolutamente continua di densità f . In questo caso, per ogni $x \in \mathbb{R}$, l'evento $\{X = x\}$ è trascurabile, sicché non si può definire direttamente la probabilità condizionata

$$\mathbb{P}(Y \in D \mid X = x).$$

Tuttavia, si noti che

$$\begin{aligned} \mathbb{P}(Y \in D \mid x - h < X < x + h) &= \frac{\mathbb{P}(Y \in D, x - h < X < x + h)}{\mathbb{P}(x - h < X < x + h)} \\ &= \left(\int_{x-h}^{x+h} f_1(s) \, ds \right)^{-1} \left(\int_{x-h}^{x+h} ds \int_D f(s, t) \, dt \right), \end{aligned}$$

dove con

$$f_1(s) := \int_{\mathbb{R}} f(s, t) \, dt$$

si è indicata la densità marginale di X .

Ragionando intuitivamente, si ha che, se f è continua, l'ultima espressione scritta è approssimativamente eguale a

$$\frac{2h}{2hf_1(x)} \int_D f(x, t) \, dt = \int_D \frac{f(x, t)}{f_1(x)} \, dt.$$

Ciò porta a definire come

$$h(y \mid x) := \frac{f(x, y)}{f_1(x)}$$

la *densità condizionata di Y dato $\{X = x\}$* , o, in breve la densità di Y data X . A rigore $h(y \mid x)$ è definita solo se $f_1(x) \neq 0$; tuttavia, posto

$$S := \{(x, y) \in \mathbb{R}^2 : f_1(x) = 0\},$$

si ha che $\mathbb{P}[(X, Y) \in S] = 0$. Infatti,

$$\begin{aligned} \mathbb{P}[(X, Y) \in S] &= \int_S f(x, y) \, dx \, dy = \int_{\{x: f_1(x)=0\}} dx \int_{\mathbb{R}} f(x, y) \, dy \\ &= \int_{\{x: f_1(x)=0\}} f_1(x) \, dx = 0. \end{aligned}$$

Si osservi che, se $B \in \mathcal{F}$ e se $X = x$, allora $(x, y) \in B$ se, e solo se, $Y \in B_x$. Questo porta a pensare che sia

$$\varphi_B(x) = \int_{B_x} h(y \mid x) \, dy.$$

Poiché si può scrivere

$$\varphi_B(x) = \int_{\mathbb{R}} h(y \mid x) \mathbf{1}_B(x, y) \, dy,$$

si ha intanto che φ_B così definita è misurabile. Inoltre, se $A \in \mathcal{B}$,

$$\begin{aligned} \mathbb{P}(\{X \in A\} \cap B) &= \int_{\{(x,y) \in B, x \in A\}} f(x, y) \, dx \, dy \\ &= \int_{\mathbb{R}} \mathbf{1}_A(x) f_1(x) \, dx \int_{\mathbb{R}} \mathbf{1}_B(x, y) h(y | x) \, dy \\ &= \int_A f_1(x) \, dx \int_{B_x} h(y | x) \, dy \\ &= \int_A \varphi_B(x) f_1(x) \, dx = \int_A \varphi_B(x) \, d\mathbb{P}_X(x). \end{aligned}$$

Perciò $\varphi_B(x) = \mathbb{P}(B | X = x)$.

L'espressione $f(x, y) = f_1(x) h(y | x)$ si può leggere nei due versi: da un lato, come si è visto, essa serve a definire la densità condizionata di Y data X , se è nota la densità congiunta f del vettore aleatorio (X, Y) , e quindi anche la densità marginale f_1 di X . D'altro canto, se è noto che la v.a. X ha legge di densità f_1 , e se si sa che, subordinatamente all'evento $\{X = x\}$, Y ha densità $h(\cdot | x)$, allora il vettore aleatorio (X, Y) ha densità $(x, y) \mapsto f(x, y) = f_1(x) h(y | x)$. Infatti, ponendo, per $B \in \mathcal{B}$,

$$\mathbb{P}(x, B) := \int_B h(y | x) \, dy,$$

si vede che esiste un'unica misura di probabilità su $\mathcal{B}(\mathbb{R}^2)$ che soddisfa, per $A, B \in \mathcal{B}$, a

$$\mathbb{P}(X \in A, Y \in B) = \int_A \mathbb{P}(x, B) f_1(x) \, dx = \int_A f_1(x) \, dx \int_B h(y | x) \, dy.$$

■

Passando alla considerazione delle speranze, si pone il problema di dare una ragionevole definizione della speranza della v.a. Y sapendo che la v.a. X ha preso il valore x ; in simboli, si può dare un significato a $\mathbb{E}(Y | X = x)$.

Teorema 5.2.2. *Siano (Ω, \mathcal{F}) e (Ω', \mathcal{F}') due spazi misurabili, \mathbb{P} una probabilità su (Ω, \mathcal{F}) , $X : \Omega \rightarrow \Omega'$ una v.a. a valori in Ω' e $Y : \Omega \rightarrow \mathbb{R}$ una v.a. di $L^1(\mathcal{F})$. Esiste allora una funzione misurabile $\varphi : \Omega' \rightarrow \mathbb{R}$ tale che, per ogni $A \in \mathcal{F}'$,*

$$\int_{X^{-1}(A)} Y \, d\mathbb{P} = \int_A \varphi(x) \, d\mathbb{P}_X(x). \quad (5.2.3)$$

ove $\mathbb{P}_X = \mathbb{P} \circ X^{-1}$ è la legge X . Se ψ è un'altra funzione che soddisfa alla (5.2.3) allora $\psi = \varphi$ q.c. rispetto a \mathbb{P}_X . Si pone $\mathbb{E}(Y | X = x) := \varphi(x)$.

Dimostrazione. Se Y è positiva, si definisce una misura μ su \mathcal{F}' , ponendo, per ogni $A \in \mathcal{F}'$,

$$\mu(A) := \int_{X^{-1}(A)} Y \, d\mathbb{P}.$$

L'asserto segue applicando il teorema di Radon–Nikodym alla parte positiva e alla parte negativa di Y . \square

La funzione $\mathbb{E}(Y \mid X = x)$ appena introdotta si chiama *speranza di Y condizionata dall'evento $\{X = x\}$* oppure, piú comunemente, in specie nella Statistica Matematica, *funzione di regressione di Y su X* .

Le speranze condizionate includono le probabilità condizionate come caso particolare.

Corollario 5.2.1. *Sia X una v.a. su $(\Omega, \mathcal{F}, \mathbb{P})$ e sia $B \in \mathcal{F}$. Allora vale q.c. rispetto a $\mathbb{P}_X = \mathbb{P} \circ X$,*

$$\mathbb{E}(\mathbf{1}_B \mid X = x) = \mathbb{P}(B \mid X = x).$$

Dimostrazione. Basta porre $Y = \mathbf{1}_B$ nella (5.2.3). \square \square

Esempio 5.2.3. Come nell'esempio 5.2.1, siano X una v.a. discreta, $\{x_n : n \in \mathbb{N}\}$ i valori che questa assume, $\mathbb{P}(X = x_n) > 0$ le probabilità con le quali li assume. Si può supporre che $\Omega' = \{x_1, x_2, \dots, x_n, \dots\}$ e $\mathcal{F}' = \mathcal{P}(\Omega')$. Si ponga ora

$$\varphi(x_n) := \frac{1}{\mathbb{P}(X = x_n)} \int_{\{X=x_n\}} Y \, d\mathbb{P}.$$

Allora, per ogni $A \in \mathcal{F}'$, si ha

$$\begin{aligned} \int_{X^{-1}(A)} Y \, d\mathbb{P} &= \sum_{x_n \in A} \mathbb{P}(X = x_n) \frac{1}{\mathbb{P}(X = x_n)} \int_{\{X=x_n\}} Y \, d\mathbb{P} \\ &= \sum_{x_n \in A} \mathbb{P}(X = x_n) \varphi(x_n) = \int_A \varphi(x) \, d\mathbb{P}_X(x), \end{aligned}$$

onde, per ogni $n \in \mathbb{N}$,

$$\mathbb{E}(Y \mid X = x_n) = \frac{1}{\mathbb{P}(X = x_n)} \int_{\{X=x_n\}} Y \, d\mathbb{P}.$$

■

Esempio 5.2.4. Siano X e Y due v.a. con legge congiunta assolutamente continua di densità f e sia $h = h(y|x)$ la densità condizionata di Y data X . Per ogni $A \in \mathcal{B}$, si ha

$$\begin{aligned} \int_{X^{-1}(A)} Y \, d\mathbb{P} &= \int_{\{(x,y):x \in A\}} y f(x, y) \, dx \, dy \\ &= \int_A f_1(x) \, dx \int_{\mathbb{R}} y h(y|x) \, dy = \int_A d\mathbb{P}_X(x) \int_{\mathbb{R}} y h(y|x) \, dy. \end{aligned}$$

Perciò

$$\mathbb{E}(Y \mid X = x) = \int_{\mathbb{R}} y h(y|x) \, dy.$$

È spesso utile la seguente osservazione.

Se è noto che la speranza condizionata è $\varphi(x) := \mathbb{E}(Y | X = x)$ e se $\psi := \varphi \circ X$ allora $\mathbb{E}(Y | X) = \psi$; infatti dalla definizione di SC, dalla (5.2.3) e dal teorema del cambio di variabile (3.8.2) scende, per ogni $B = X^{-1}(A) \in \mathcal{F}(X)$, con $A \in \mathcal{B}$,

$$\begin{aligned} \int_{X^{-1}(A)} \mathbb{E}(Y | X) \, d\mathbb{P} &= \int_B Y \, d\mathbb{P} = \int_A \mathbb{E}(Y | X = x) \, d\mathbb{P}_X(x) \\ &= \int_{X^{-1}(A)} \varphi \circ X \, d\mathbb{P}. \end{aligned}$$

■

5.3 Proprietà delle speranze condizionate

In questa sezione X denoterà sempre una v.a. di $L^1(\mathcal{F})$ e \mathcal{G} una tribù contenuta in \mathcal{F} . Daremo solo le proprietà delle SC data una tribù \mathcal{G} , perché le proprietà della speranza della v.a. Y data un'altra v.a. X , $\mathbb{E}(Y | X = x)$, o quelle della probabilità condizionata $\mathbb{P}(B | X = x)$ sono del tutto analoghe.

Non è inutile ribadire che *tutte le relazioni che saranno date per le SC valgono quasi certamente rispetto alla misura di probabilità \mathbb{P} .*

Teorema 5.3.1. *Valgono le seguenti proprietà:*

- (a) $\mathbb{E}(\mathbb{E}(X | \mathcal{G})) = \mathbb{E}(X)$;
- (b) se X è \mathcal{G} -misurabile ($X \in L^1(\mathcal{G})$), allora $\mathbb{E}(X | \mathcal{G}) = X$;
- (c) se $X = a$ q.c., allora $\mathbb{E}(X | \mathcal{G}) = a$ (in particolare $\mathbb{E}(1 | \mathcal{G}) = 1$);
- (d) se $X \geq 0$, allora $\mathbb{E}(X | \mathcal{G}) \geq 0$;
- (e) se $c_1, c_2 \in \mathbb{R}$ e $X_1, X_2 \in L^1(\mathcal{F})$, allora

$$\mathbb{E}_{\mathcal{G}}(c_1 X_1 + c_2 X_2) = c_1 \mathbb{E}_{\mathcal{G}}(X_1) + c_2 \mathbb{E}_{\mathcal{G}}(X_2),$$

vale a dire che $\mathbb{E}_{\mathcal{G}}$ è un operatore lineare su $L^1(\mathcal{F})$;

- (f) se $X_1 \geq X_2$, allora $\mathbb{E}_{\mathcal{G}}(X_1) \geq \mathbb{E}_{\mathcal{G}}(X_2)$;
- (g) $|\mathbb{E}_{\mathcal{G}}(X)| \leq \mathbb{E}_{\mathcal{G}}(|X|)$;
- (h) se \mathcal{N} è la tribù banale, $\mathcal{N} := \{\Omega, \emptyset\}$, allora $\mathbb{E}_{\mathcal{N}}(X) = \mathbb{E}(X)$.

Dimostrazione. (a) Basta prendere $B = \Omega$ nella (5.1.1). La (b) scende dall'unicità q.c. delle derivate di Radon–Nikodym.

(c) Poiché ogni v.a. che sia q.c. costante è misurabile rispetto alla tribù \mathcal{N} , che è inclusa in ogni tribù di sottoinsiemi di Ω , e quindi anche in \mathcal{G} , la v.a. $X = a$ q.c. è misurabile rispetto a \mathcal{G} ; il risultato è dunque contenuto in (a).

(d) Per ogni $B \in \mathcal{G}$, si ha $\int_B X \, d\mathbb{P} \geq 0$.

(e) Per ogni $B \in \mathcal{G}$, si ha

$$\begin{aligned} \int_B \mathbb{E}_{\mathcal{G}}(c_1 X_1 + c_2 X_2) \, d\mathbb{P} &= \int_B (c_1 X_1 + c_2 X_2) \, d\mathbb{P} \\ &= c_1 \int_B \mathbb{E}_{\mathcal{G}}(X_1) \, d\mathbb{P} + c_2 \int_B \mathbb{E}_{\mathcal{G}}(X_2) \, d\mathbb{P} \\ &= \int_B \{c_1 \mathbb{E}_{\mathcal{G}}(X_1) + c_2 \mathbb{E}_{\mathcal{G}}(X_2)\} \, d\mathbb{P}. \end{aligned}$$

(f) Basta applicare la (d) alla v.a. $X_1 - X_2$ e tenere conto della linearità appena dimostrata.

(g) Basta applicare la (f) e la (e) alla diseuguaglianza $-|X| \leq X \leq |X|$.

(h) $\int_B X \, d\mathbb{P} = \int_B \mathbb{E}(X) \, d\mathbb{P}$ se $B = \emptyset$ oppure se $B = \Omega$; l'asserto segue ora dall'unicità q.c. di $\mathbb{E}_{\mathcal{G}}(X)$. \square

Osservazione 5.3.1. Le proprietà (g) e (a) dell'ultimo teorema mostrano che l'operatore $X \mapsto \mathbb{E}_{\mathcal{G}}(X)$ è a valori in $L^1(\mathcal{G})$, cioè $\mathbb{E}_{\mathcal{G}} : L^1(\mathcal{F}) \rightarrow L^1(\mathcal{G})$. Di più, esso è lineare, per la proprietà (e), suriettivo, per la proprietà (b) ed è una contrazione, come si vede integrando la (g) ed usando la (a):

$$\|\mathbb{E}_{\mathcal{G}}(X)\|_{L^1(\mathcal{G})} \leq \|X\|_{L^1(\mathcal{F})}.$$

Ciò rende interessante e naturale lo studio di $\mathbb{E}_{\mathcal{G}}$ come operatore su $L^1(\mathcal{F})$. Tale studio sarà intrapreso più avanti.

Val la pena di scrivere l'ultima diseuguaglianza in una notazione meno precisa, ma, certo, meno pesante,

$$\|\mathbb{E}_{\mathcal{G}}(X)\|_1 \leq \|X\|_1.$$

Valgono per le SC gli analoghi dei teoremi di Beppo Levi e di convergenza dominata.

Teorema 5.3.2. *Valgono le seguenti affermazioni:*

(a) *Sia (X_n) una successione crescente di v.a. di $L^1(\mathcal{F})$, positive che converge alla v.a. $X \in L^1(\mathcal{F})$, (se, per ogni $n \in \mathbb{N}$, $X_n \geq 0$, $X_{n+1} \geq X_n$, $X_n \xrightarrow[n \rightarrow +\infty]{} X$); allora*

$$\lim_{n \rightarrow +\infty} \mathbb{E}_{\mathcal{G}}(X_n) = \mathbb{E}_{\mathcal{G}}(X);$$

(b) *se per ogni $n \in \mathbb{N}$, $X_n \geq 0$, allora*

$$\mathbb{E}_{\mathcal{G}}\left(\sum_{n \in \mathbb{N}} X_n\right) = \sum_{n \in \mathbb{N}} \mathbb{E}_{\mathcal{G}}(X_n);$$

(c) *se (B_n) è una successione di insiemi misurabili e disgiunti (per ogni $n \in \mathbb{N}$, $B_n \in \mathcal{F}$, $B_j \cap B_k = \emptyset$, $j \neq k$), allora*

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} B_n \mid \mathcal{G}\right) = \sum_{n \in \mathbb{N}} \mathbb{P}(B_n \mid \mathcal{G}). \quad (5.3.1)$$

Dimostrazione. (a) Per ogni $B \in \mathcal{G}$ e per ogni $n \in \mathbb{N}$, è

$$\int_B X_n \, d\mathbb{P} = \int_B \mathbb{E}_{\mathcal{G}}(X_n) \, d\mathbb{P}.$$

Grazie al Teorema 5.3.1(f), la successione $(\mathbb{E}_{\mathcal{G}}(X_n))$ è crescente; essa tende perciò ad una funzione φ necessariamente \mathcal{G} -misurabile. Per il teorema di Beppo Levi si ha, perciò, per ogni $B \in \mathcal{G}$,

$$\int_B X \, d\mathbb{P} = \int_B \varphi \, d\mathbb{P},$$

onde $\varphi = \mathbb{E}_{\mathcal{G}}(X)$.

Le dimostrazioni dei punti (b) e (c) sono semplici applicazioni di (a). \square

Osservazione 5.3.2. Si osservi che l'eq. (5.3.1) *non* asserisce che $\mathbb{P}(\cdot | \mathcal{G})$ sia una probabilità (si veda la successiva Sezione 5.4). \blacksquare

Teorema 5.3.3. *Se (X_n) è una successione di v.a. di $L^1(\mathcal{F})$ che converge q.c. alla v.a. X e se esiste una v.a. $Y \in L^1(\mathcal{F})$ tale che, per ogni $n \in \mathbb{N}$, sia $|X_n| \leq Y$, allora*

$$\lim_{n \rightarrow +\infty} \mathbb{E}_{\mathcal{G}}(X_n) = \mathbb{E}_{\mathcal{G}}(X).$$

Dimostrazione. Posto $Y_n := \sup\{|X_j - X| : j \geq n\}$, la successione $\{Y_n\}$ è decrescente e converge a 0. Si osservi che, per ogni $n \in \mathbb{N}$, $\mathbb{E}_{\mathcal{G}}(X_n)$ e $\mathbb{E}_{\mathcal{G}}(X)$ sono in $L^1(\mathcal{G})$. Inoltre

$$|\mathbb{E}_{\mathcal{G}}(X_n) - \mathbb{E}_{\mathcal{G}}(X)| \leq \mathbb{E}_{\mathcal{G}}(|X_n - X|) \leq \mathbb{E}_{\mathcal{G}}(Y_n),$$

sicché basta mostrare che $\mathbb{E}_{\mathcal{G}}(Y_n) \rightarrow 0$. Da quanto precede segue che esiste una funzione φ , positiva e \mathcal{G} -misurabile, tale che $\mathbb{E}_{\mathcal{G}}(Y_n) \xrightarrow{n \rightarrow +\infty} \varphi$. Poiché $0 \leq Y_n \leq 2Y$, il teorema di convergenza dominata dà

$$0 \leq \int \varphi \, d\mathbb{P} \leq \int \mathbb{E}_{\mathcal{G}}(Y_n) \, d\mathbb{P} = \int Y_n \, d\mathbb{P} \xrightarrow{n \rightarrow +\infty} 0,$$

onde $\varphi = 0$ q.c.. \square

Valgono per le SC anche gli analoghi dei lemmi di Fatou.

Teorema 5.3.4. *Sia (X_n) una successione di v.a. di $L^1(\mathcal{F})$, tale che, per ogni $n \in \mathbb{N}$, sia $X_n \geq Y$ con $\mathbb{E}(Y) > -\infty$;*

(a) *se (X_n) è crescente e tende q.c. alla v.a. X , allora*

$$\mathbb{E}_{\mathcal{G}}(X_n) \xrightarrow{n \rightarrow +\infty} \mathbb{E}_{\mathcal{G}}(X);$$

(b) $\liminf_{n \rightarrow +\infty} \mathbb{E}_{\mathcal{G}}(X_n) \geq \mathbb{E}_{\mathcal{G}}(\liminf_{n \rightarrow +\infty} X_n)$.

Se, invece, (X_n) è una successione di v.a. di $L^1(\mathcal{F})$, tale che per ogni $n \in \mathbb{N}$, sia $X_n \leq Y$ con $\mathbb{E}(Y) < +\infty$; allora

(c) se (X_n) è decrescente e tende q.c. alla v.a. X ,

$$\mathbb{E}_{\mathcal{G}}(X_n) \xrightarrow{n \rightarrow +\infty} \mathbb{E}_{\mathcal{G}}(X);$$

(d) $\limsup_{n \rightarrow +\infty} \mathbb{E}_{\mathcal{G}}(X_n) \leq \mathbb{E}_{\mathcal{G}}(\limsup_{n \rightarrow +\infty} X_n)$.

Dimostrazione. Basta dimostrare (a) e (b), perché (c) e (d) scendono da quelle cambiando il segno a tutte le v.a. che intervengono.

(a) Posto $Z_n := X_n - X_1$, la successione (Z_n) è in $L^1(\mathcal{F})$, è positiva e crescente, e ammette come limite la v.a. $X - X_1$. Il teorema di Beppo Levi e la linearità delle SC danno

$$\mathbb{E}_{\mathcal{G}}(X_n) = \mathbb{E}_{\mathcal{G}}(Z_n) + \mathbb{E}_{\mathcal{G}}(X_1) \xrightarrow{n \rightarrow +\infty} \mathbb{E}_{\mathcal{G}}(X) - \mathbb{E}_{\mathcal{G}}(X_1) + \mathbb{E}_{\mathcal{G}}(X_1) = \mathbb{E}_{\mathcal{G}}(X).$$

(b) Se, per ogni $n \in \mathbb{N}$, si pone $V_n := \inf\{X_j : j \geq n\}$, la successione (V_n) tende crescendo a $V := \liminf_{n \rightarrow +\infty} X_n$, onde,

$$\begin{aligned} \mathbb{E}_{\mathcal{G}}(V) &= \mathbb{E}_{\mathcal{G}}(\liminf_{n \rightarrow +\infty} X_n) = \lim_{n \rightarrow +\infty} \mathbb{E}_{\mathcal{G}}(V_n) \\ &= \liminf_{n \rightarrow +\infty} \mathbb{E}_{\mathcal{G}}(V_n) \leq \liminf_{n \rightarrow +\infty} \mathbb{E}_{\mathcal{G}}(X_n), \end{aligned}$$

onde l'asserto. □

Le proprietà delle SC che seguono si riferiscono tutte a condizioni di regolarizzazione. Ad esse occorre premettere il seguente lemma di carattere tecnico. Ricordiamo che si dice *atomo* di uno spazio di probabilità (Ω, \mathcal{G}, P) un insieme $B \in \mathcal{G}$ tale che $P(B) > 0$ e che se $A \in \mathcal{G}$ e $A \subseteq B$ allora $P(A) = 0$ oppure $P(B \setminus A) = 0$.

Lemma 5.3.1. *Sia B un atomo di (Ω, \mathcal{G}, P) e sia X una v.a. a valori reali su tale spazio. Allora X è costante q.c. in B .*

Dimostrazione. Sia B un atomo; Si osservi, innanzi tutto, che si può supporre, che esistano numeri reali x per i quali $P(B \cap \{X < x\}) = 0$. Infatti, se così non fosse, si avrebbe, poiché B è un atomo, $P(B \cap \{X < x\}) = P(B)$ per ogni $x \in \mathbb{R}$; in altre parole, sarebbe $X = -\infty$ q.c. in B . Sia, perciò, $x \in \mathbb{R}$ tale che $P(B \cap \{X < x\}) = 0$; di conseguenza risulta $P(B \cap \{X < y\}) = 0$ per ogni $y < x$. Posto,

$$k := \sup\{x \in \mathbb{R} : P(B \cap \{X < x\}) = 0\},$$

si ha

$$P(B \cap \{X < k\}) = P\left(\bigcup_{\substack{q \in \mathbb{Q} \\ r < k}} (B \cap \{X < q\})\right) = 0.$$

Se $x > k$, risulta $P(B \cap \{X < x\}) = P(B)$, onde $P(B \cap \{X \geq x\}) = 0$, poiché B è un atomo. Perciò,

$$P(B \cap \{X > k\}) = P\left(\bigcup_{\substack{q \in \mathbb{Q} \\ q > k}} (B \cap \{X \geq q\})\right) = 0.$$

Pertanto, $X = k$ q.c. su B , cioè $P(B \cap \{X = k\}) = P(B)$. □

Siamo ora in grado di stabilire l'espressione della SC rispetto ad una tribú \mathcal{G} che sia generata da una partizione misurabile e al più numerabile di atomi.

Teorema 5.3.5. *Sia B un atomo di \mathcal{G} e sia X una v.a. di $L^1(\mathcal{F})$. Allora*

$$\mathbb{E}_{\mathcal{G}}(X) = \frac{1}{\mathbb{P}(B)} \int_B X \, d\mathbb{P} \quad \text{q.c. in } B.$$

Dimostrazione. Per il Lemma 5.3.1, $\mathbb{E}_{\mathcal{G}}(X)$ è costante q.c. in B . Perciò, la (5.1.1) dà

$$\mathbb{E}_{\mathcal{G}}(X) \mathbb{P}(B) = \int_B X \, d\mathbb{P},$$

cioè l'asserto. □

Corollario 5.3.1. *Se $\pi = \{B_n\}$ è una partizione, finita o misurabile, di Ω in insiemi \mathcal{F} -misurabili con $\mathbb{P}(B_n) > 0$, per ogni indice, e se \mathcal{G} è la tribú generata da π , $\mathcal{G} = \mathcal{F}(\pi)$, allora*

$$\mathbb{E}(X | \mathcal{G}) := \sum_n \frac{\mathbf{1}_{B_n}}{\mathbb{P}(B_n)} \int_{B_n} X \, d\mathbb{P}. \quad (5.3.2)$$

Teorema 5.3.6. *Siano X e Y due v.a. su $(\Omega, \mathcal{F}, \mathbb{P})$ tali che siano integrabili tanto X quanto il prodotto XY , $X \in L^1(\mathcal{F})$, $XY \in L^1(\mathcal{F})$; inoltre, Y sia \mathcal{G} -misurabile. Allora*

$$\mathbb{E}_{\mathcal{G}}(XY) = Y \mathbb{E}_{\mathcal{G}}(X). \quad (5.3.3)$$

Dimostrazione. Si supponga dapprima che Y sia una funzione indicatrice, $Y = \mathbf{1}_B$ con $B \in \mathcal{G}$; in tal caso, per ogni $A \in \mathcal{G}$ si ha

$$\begin{aligned} \int_A \mathbb{E}_{\mathcal{G}}(XY) \, d\mathbb{P} &= \int_A XY \, d\mathbb{P} = \int_{A \cap B} X \, d\mathbb{P} = \int_{A \cap B} \mathbb{E}_{\mathcal{G}}(X) \, d\mathbb{P} \\ &= \int_A \mathbf{1}_B \mathbb{E}_{\mathcal{G}}(X) \, d\mathbb{P} = \int_A Y \mathbb{E}_{\mathcal{G}}(X) \, d\mathbb{P}, \end{aligned}$$

sicché la (5.3.3) vale per le funzioni indicatrici di insiemi di \mathcal{G} e, quindi, per linearità, per le funzioni \mathcal{G} -semplici. Ricorrendo al Teorema 5.3.4(a), si dimostra la (5.3.3) per le funzioni \mathcal{G} -misurabili positive e, infine, mediante la decomposizione $Y = Y^+ - Y^-$, per tutte le funzioni \mathcal{G} -misurabili. □

In particolare se appartengono a $L^1(\mathcal{F})$ tanto X_1 e X_2 quanto il loro prodotto $X_1 X_2$, vale

$$\mathbb{E}_{\mathcal{G}}(X_1 \mathbb{E}_{\mathcal{G}}(X_2)) = \mathbb{E}_{\mathcal{G}}(X_1) \mathbb{E}_{\mathcal{G}}(X_2).$$

Se \mathcal{G}_1 e \mathcal{G}_2 sono due sottotribú distinte di \mathcal{F} , in generale gli operatori $\mathbb{E}_{\mathcal{G}_1}$ e $\mathbb{E}_{\mathcal{G}_2}$ non commutano come mostra il seguente

Esempio 5.3.1. Sia $\{B_1, B_2, B_3\}$ una partizione misurabile di Ω e si considerino le tribú

$$\mathcal{G}_1 = \mathcal{F}(\{B_1 \cup B_2, B_3\}), \quad \mathcal{G}_2 = \mathcal{F}(\{B_1, B_2 \cup B_3\})$$

e la v.a. $X = \mathbf{1}_{B_3}$. Allora $\mathbb{E}_{\mathcal{G}_2} \mathbb{E}_{\mathcal{G}_1} X = \mathbb{E}_{\mathcal{G}_2} X$ che non è \mathcal{G}_1 -misurabile e non può, perciò, coincidere con $\mathbb{E}_{\mathcal{G}_1} \mathbb{E}_{\mathcal{G}_2} X$. ■

Il teorema che segue dà l'esempio di una situazione importante nella quale $\mathbb{E}_{\mathcal{G}_1}$ e $\mathbb{E}_{\mathcal{G}_2}$ commutano.

Teorema 5.3.7. *Siano \mathcal{G}_1 e \mathcal{G}_2 due sottotribú di \mathcal{F} con $\mathcal{G}_1 \subseteq \mathcal{G}_2$ e sia $X \in L^1(\mathcal{F})$. Allora gli operatori $\mathbb{E}_{\mathcal{G}_1}$ e $\mathbb{E}_{\mathcal{G}_2}$ commutano e vale*

$$\mathbb{E}_{\mathcal{G}_2} \mathbb{E}_{\mathcal{G}_1} X = \mathbb{E}_{\mathcal{G}_1} \mathbb{E}_{\mathcal{G}_2} X = \mathbb{E}_{\mathcal{G}_1} X. \quad (5.3.4)$$

Dimostrazione. Poiché $\mathbb{E}_{\mathcal{G}_1} X$ è misurabile rispetto a \mathcal{G}_1 e quindi rispetto a \mathcal{G}_2 , si ha, per la 5.3.1(b)

$$\mathbb{E}_{\mathcal{G}_2} \mathbb{E}_{\mathcal{G}_1} X = \mathbb{E}_{\mathcal{G}_1} X.$$

Inoltre, per ogni A in \mathcal{G}_1 , e dunque anche in \mathcal{G}_2 , si ha

$$\int_A \mathbb{E}_{\mathcal{G}_1} X \, d\mathbb{P} = \int_A X \, d\mathbb{P} = \int_A \mathbb{E}_{\mathcal{G}_2} X \, d\mathbb{P} = \int_A \mathbb{E}_{\mathcal{G}_1} \mathbb{E}_{\mathcal{G}_2} X \, d\mathbb{P},$$

onde l'asserto. □

Si osservi che se $\mathcal{G}_1 = \mathcal{N}$ e $\mathcal{G}_2 = \mathcal{G}$, la (5.3.3) dà, poiché $\mathbb{E}_{\mathcal{N}} = E$, (Teorema 5.3.1(h)), la 5.3.1(a).

Se $\mathcal{G}_1 = \mathcal{G}_2 = \mathcal{G}$, la (5.3.4) dà $\mathbb{E}_{\mathcal{G}}^2 = \mathbb{E}_{\mathcal{G}} \mathbb{E}_{\mathcal{G}} = \mathbb{E}_{\mathcal{G}}$, sicché $\mathbb{E}_{\mathcal{G}}$ è una proiezione da $L^1(\mathcal{F})$ su $L^1(\mathcal{G})$.

Definizione 5.3.1. Una tribú $\mathcal{G} \subseteq \mathcal{F}$ ed una v.a. X su $(\Omega, \mathcal{F}, \mathbb{P})$ si dicono *indipendenti rispetto alla misura di probabilità* \mathbb{P} se sono indipendenti (rispetto a \mathbb{P}) \mathcal{G} e la tribú $\mathcal{F}(X) := \{X^{-1}(B) : B \in \mathcal{B}\}$ generata da X . ◇

Teorema 5.3.8. *Se la v.a. X di $L^1(\mathcal{F})$ è indipendente dalla tribú \mathcal{G} , si ha*

$$\mathbb{E}_{\mathcal{G}}(X) = \mathbb{E}(X).$$

Dimostrazione. In virtù dell'indipendenza di X e di \mathcal{G} , si ha, quale che sia $B \in \mathcal{G}$,

$$\int_B \mathbb{E}_{\mathcal{G}}(X) \, d\mathbb{P} = \mathbb{E}(X \mathbf{1}_B) = \mathbb{E}(X) \mathbb{E}(\mathbf{1}_B) = \mathbb{E}(X) \mathbb{P}(B) = \int_B \mathbb{E}(X) \, d\mathbb{P},$$

onde l'asserto. □

L'ultimo risultato si usa spesso nella seguente forma

Corollario 5.3.2. *Siano X e Y due v.a. indipendenti di $(\Omega, \mathcal{F}, \mathbb{P})$; allora*

$$\mathbb{E}(X | Y) = \mathbb{E}(X).$$

Dimostrazione. Infatti, è, per definizione, $\mathbb{E}(X | Y) := \mathbb{E}(X | \mathcal{F}(Y))$ e, inoltre, dire che X e Y sono indipendenti equivale ad affermare che X e la tribú $\mathcal{F}(Y)$ sono indipendenti. \square

Il seguente teorema, dovuto a Doob, è spesso utile.

Teorema 5.3.9. *Siano (Ω, \mathcal{F}) e (Ω', \mathcal{F}') due spazi misurabili e sia $Y : \Omega \rightarrow \Omega'$ una v.a.. Per una funzione $\varphi : \Omega \rightarrow \mathbb{R}$ sono equivalenti le proprietà:*

- (a) φ è misurabile rispetto a $\mathcal{F}(Y)$ e a \mathcal{B} ;
- (b) esiste una funzione misurabile $f : \Omega' \rightarrow \mathbb{R}$ tale che $\varphi = f \circ Y$.

Dimostrazione. Basta dimostrare l'implicazione (a) \implies (b). Sia dapprima φ una funzione indicatrice, $\varphi = \mathbf{1}_B$ con $B \in \mathcal{F}(Y)$; allora $B = Y^{-1}(A)$ con $A \in \mathcal{F}'$, sicché, se $f := \mathbf{1}_A$, si ha $f \circ Y = \mathbf{1}_{Y^{-1}(A)} = \mathbf{1}_B = \varphi$. L'asserto è dunque vero se φ è una funzione indicatrice. Se poi φ è una funzione semplice $\varphi = \sum_{j=1}^n \lambda_j \mathbf{1}_{B_j}$ con $\lambda_j \in \mathbb{R}$ e $B_j \in \mathcal{F}(Y)$ ($j = 1, 2, \dots, n$), allora $\mathbf{1}_{B_j} = f_j \circ Y$ come sopra, con $f_j := \mathbf{1}_{A_j}$ se $B_j = Y^{-1}(A_j)$ e $A_j \in \mathcal{F}'$; quindi, $f = \sum_{j=1}^n \lambda_j f_j$.

Sia ora φ una funzione positiva misurabile rispetto a $\mathcal{F}(Y)$ e a \mathcal{B} e sia (s_n) una successione crescente di funzioni semplici che tende a φ . Per quanto si è appena visto, per ogni $n \in \mathbb{N}$ esiste $f_n : \Omega' \rightarrow \mathbb{R}$ misurabile tale che $s_n = f_n \circ Y$. Si definisca $f := \limsup_{n \rightarrow +\infty} f_n$. Perciò

$$\varphi = \lim_{n \rightarrow +\infty} s_n = \limsup_{n \rightarrow +\infty} s_n = \limsup_{n \rightarrow +\infty} f_n \circ Y = f \circ Y.$$

Il caso generale segue ponendo $\varphi = \varphi^+ - \varphi^-$. \square

Teorema 5.3.10. *Sia Y una v.a. su $(\Omega, \mathcal{F}, \mathbb{P})$; e se $\mathcal{G} = \mathcal{F}(Y)$, esiste una funzione boreliana $f_X : \mathbb{R} \rightarrow \mathbb{R}$, che dipende da X , tale che sia*

$$\mathbb{E}(X | Y) = \mathbb{E}_{\mathcal{F}(Y)}(X) = f_X \circ Y.$$

Dimostrazione. È conseguenza immediata del Teorema 5.3.9. \square

Vale anche per le SC la diseguaglianza di Jensen.

Teorema 5.3.11. *Sia $\varphi : I \rightarrow \mathbb{R}$, con I intervallo (limitato o no, aperto o no) di \mathbb{R} , convessa e sia X una v.a. di $L^1(\mathcal{F})$ a valori in I . Se \mathcal{G} è una tribú contenuta in \mathcal{F} , vale la diseguaglianza di Jensen*

$$\mathbb{E}_{\mathcal{G}}(\varphi \circ X) \geq \varphi \circ \mathbb{E}_{\mathcal{G}}(X). \quad (5.3.5)$$

Dimostrazione. Sia $I = (a, b)$ con $-\infty \leq a < b \leq +\infty$. In primo luogo risulta $\mathbb{E}_{\mathcal{G}}(X) \in I$ q.c.. Infatti se, per esempio, è $X > a$ q.c., si ha

$$0 \geq \int_{\{\mathbb{E}_{\mathcal{G}} \leq a\}} (\mathbb{E}_{\mathcal{G}}(X) - a) \, d\mathbb{P} = \int_{\{\mathbb{E}_{\mathcal{G}} \leq a\}} (X - a) \, d\mathbb{P} \geq 0,$$

sicch  $\mathbb{P}(\mathbb{E}_{\mathcal{G}}(X) \leq a) = 0$, cio  $\mathbb{E}_{\mathcal{G}}(X) > a$ q.c.. Com'  noto, il *teorema della linea di supporto*, Teorema 1.17.1, asserisce che esistono due successioni di numeri reali (a_n) e (b_n) tali che, per ogni $x \in I$, si abbia

$$\varphi(x) = \sup\{a_n x + b_n : n \in \mathbb{N}\}. \quad (5.3.6)$$

Scende dalla (5.3.6) che, per ogni $n \in \mathbb{N}$, $f \circ X \geq a_n X + b_n$ onde

$$\mathbb{E}_{\mathcal{G}}(\varphi \circ X) \geq a_n \mathbb{E}_{\mathcal{G}}(X) + b_n.$$

Di qui, prendendo l'estremo superiore, si ottiene l'asserto. \square

Corollario 5.3.3. *Se X   in $L^p(\mathcal{F})$ (con $p \geq 1$) e se \mathcal{G}   una trib  contenuta in \mathcal{F} , allora $\mathbb{E}_{\mathcal{G}}(X)$   in $L^p(\mathcal{G})$ e vale la disuguaglianza*

$$\|\mathbb{E}_{\mathcal{G}}(X)\|_p \leq \|X\|_p \quad (5.3.7)$$

sicch  la restrizione di $\mathbb{E}_{\mathcal{G}}$ a $L^p(\mathcal{F})$   una proiezione e una contrazione su $L^p(\mathcal{G})$.

Dimostrazione. Basta applicare la (5.3.5) alla funzione convessa $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ definita da $\varphi(x) := |x|^p$ con $p \geq 1$: $|\mathbb{E}_{\mathcal{G}}(X)|^p \leq \mathbb{E}_{\mathcal{G}}(|X|^p)$. Integrando, e ricordando la 5.3.1(a), si ottiene la (5.3.7). \square

5.4 Distribuzioni condizionate regolari

Si   visto che se (B_n)   una successione di insiemi disgiunti di \mathcal{F} e se \mathcal{G}   una sottotrib  di \mathcal{F} , allora vale

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} B_n \mid \mathcal{G}\right) = \sum_{n \in \mathbb{N}} \mathbb{P}(B_n \mid \mathcal{G}) \quad \text{q.c.}$$

Ci  non implica che possiamo scegliere $\mathbb{P}(\cdot \mid \mathcal{G})$ in modo che sia una probabilit  per ogni (o quasi ogni) $\omega \in \Omega$. Si supponga, infatti che (B_n) sia una successione di insiemi disgiunti di \mathcal{F} ; allora

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} B_n \mid \mathcal{G}\right)(\omega) = \sum_{n \in \mathbb{N}} \mathbb{P}(B_n \mid \mathcal{G})(\omega)$$

tranne che per i punti ω di un insieme $N(B_1, B_2, \dots)$ di probabilit  nulla. Perci  l'insieme dei punti di Ω nei quali non vale l'additivit  numerabile  

$$M = \bigcup \{N(B_1, B_2, \dots) : B_1, B_2, \dots \text{insiemi disgiunti di } \mathcal{F}\}.$$

In genere tale insieme M   un'unione non numerabile di insiemi di probabilit  nulla e non   detto n  che appartenga a \mathcal{F} , n , se vi appartiene, che abbia probabilit  nulla, sicch  $\mathbb{P}(\cdot \mid \mathcal{G})$ potrebbe non essere numerabilmente additiva.

Nel seguito della sezione si daranno condizioni che assicurino l'additivit  numerabile.

Definizione 5.4.1. Siano Y una v.a. su $(\Omega, \mathcal{F}, \mathbb{P})$ e \mathcal{G} una sotto-tribú di \mathcal{F} . Si dice che una funzione

$$F : (\Omega, \mathbb{R}) \mapsto [0, 1]$$

è *funzione di ripartizione condizionata regolare per Y data \mathcal{G}* se sono soddisfatte le due condizioni

- (a) $F(\omega, \cdot)$ è una funzione di ripartizione per ogni $\omega \in \Omega$;
- (b) per ogni $t \in \mathbb{R}$ è $F(\omega, t) = \mathbb{P}(Y \leq t \mid \mathcal{G})(\omega)$ per quasi ogni $\omega \in \Omega$. \diamond

Teorema 5.4.1. Per ogni v.a. Y su $(\Omega, \mathcal{F}, \mathbb{P})$ e per ogni sotto-tribú \mathcal{G} di \mathcal{F} esiste una funzione di ripartizione condizionata regolare per Y data \mathcal{G} .

Dimostrazione. Per ogni razionale $q \in \mathbb{Q}$ si scelga una versione della probabilità condizionata $F_q(\omega) := \mathbb{P}(Y \leq q \mid \mathcal{G})(\omega)$. Si scriva l'insieme dei numeri razionali come unione numerabile

$$\mathbb{Q} = \bigcup_{n \in \mathbb{N}} \{q_n\}$$

e si ponga

$$A_{ij} := \{\omega \in \Omega : F_{q_j}(\omega) < F_{q_i}(\omega)\} \quad \text{e} \quad A := \bigcup \{A_{ij} : q_i < q_j\}.$$

Poiché la condizione $q_i < q_j$ implica $\mathbb{P}(Y \leq q_i \mid \mathcal{G}) \leq \mathbb{P}(Y \leq q_j \mid \mathcal{G})$ q.c., si ha $\mathbb{P}(A_{ij}) = 0$ e, quindi, $\mathbb{P}(A) = 0$.

Si definisca ora

$$B_i := \left\{ \omega \in \Omega : \lim_{n \rightarrow +\infty} F_{q_i + (1/n)}(\omega) \neq F_{q_i}(\omega) \right\} \quad \text{e} \quad B := \bigcup_{i \in \mathbb{N}} B_i.$$

Ora la successione di v.a.

$$(\mathbf{1}_{\{Y \leq q_i + (1/n)\}})_{n \in \mathbb{N}}$$

è decrescente e ha come limite $\mathbf{1}_{\{Y \leq q_i\}}$, sicché

$$F_{q_i + (1/n)} \xrightarrow[n \rightarrow +\infty]{} F_{q_i} \quad \text{q.c.}$$

Pertanto $\mathbb{P}(B_i) = 0$ per ogni $i \in \mathbb{N}$ e $\mathbb{P}(B) = 0$.

Inoltre, posto $C := \{\omega \in \Omega : \lim_{n \rightarrow +\infty} F_n(\omega) \neq 1\}$, si ha $\mathbb{P}(C) = 0$ poiché la successione di insiemi $(\{Y \leq n\})$ tende a Ω , sicché $\mathbb{P}(Y \leq n \mid \mathcal{G}) \rightarrow 1$ q.c. Similmente, ha probabilità nulla l'insieme $D := \{\omega \in \Omega : \lim_{n \rightarrow +\infty} F_n(\omega) \neq 0\}$.

Si definisca $E := A \cup B \cup C \cup D$ e

$$F(\omega, t) := \begin{cases} \lim_{q \rightarrow t, q > t} F_q(\omega), & \text{se } \omega \notin E, \\ G(t) \text{ per una qualsiasi f.r. } G, & \text{se } \omega \in E. \end{cases}$$

Si noti che ω non è in A , allora $q \mapsto F_q(\omega)$ è crescente sicché esiste il limite $\lim_{q \rightarrow t, q > t} F_q(\omega)$. Per $\omega \notin A \cup B$ e per $t \in \mathbb{Q}$, si ha $\lim_{q \rightarrow t, q > t} F_q(\omega) = F_t(\omega)$. Analogamente, per $\omega \notin A \cup C \cup D$ si ha $\lim_{t \rightarrow +\infty} F_t(\omega) = 1$ e $\lim_{t \rightarrow -\infty} F_t(\omega) = 0$.

F è una f.r. condizionata regolare per Y data \mathcal{G} . Si prenda $\omega \notin E$; allora $t \mapsto F(\omega, t)$ è crescente. Se $t < t' < q$ si ha $F(\omega, t) \leq F(\omega, t') \leq F(\omega, q) = F_q(\omega)$; ora $F_q(\omega)$ tende a $F(\omega, t)$ al tendere di q a t , sicché $F(\omega, \cdot)$ è continua a destra.

Se $q \leq t$ si ha

$$F(\omega, t) \geq F(\omega, q) = F_q(\omega) \xrightarrow{q \rightarrow +\infty} 1$$

sicché $F(\omega, t) \xrightarrow{t \rightarrow +\infty} 1$. In maniera analoga si mostra che $F(\omega, t) \xrightarrow{t \rightarrow -\infty} 0$. Pertanto $F(\omega, \cdot)$ è una f.r. e la condizione (a) della Definizione 5.4.1 è soddisfatta.

Per definizione di F si ha, per $q \in \mathbb{Q}$, $\mathbb{P}(Y \leq q | \mathcal{G})(\omega) = F(\omega, q) = F_q(\omega)$. Al tendere di q a t decrescendo si ha $F(\omega, t) \rightarrow F(\omega, t)$ per ogni $\omega \in \Omega$, a causa della continuità a destra; e, per la convergenza dominata delle SC si ha

$$\mathbb{P}(Y \leq q | \mathcal{G}) \xrightarrow{q \rightarrow t, q > t} \mathbb{P}(Y \leq t | \mathcal{G}) \quad \text{q.c.}$$

Perciò, fissato $t \in \mathbb{R}$ è $\mathbb{P}(Y \leq t | \mathcal{G}) = F(\omega, t)$ per quasi ogni $\omega \in \Omega$, ciò che prova la condizione (b) della Definizione 5.4.1 e conclude la dimostrazione. \square

Definizione 5.4.2. Siano Y una v.a. sullo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e \mathcal{G} una sotto-tribú di \mathcal{F} . La funzione $Q : \Omega \times \mathcal{B} \rightarrow [0, 1]$ si dice essere una *probabilità condizionata regolare per Y data \mathcal{G}* se sono verificate le seguenti condizioni:

(a) $Q(\omega, \cdot)$ è una misura di probabilità per ogni fissato $\omega \in \Omega$;

(b) $Q(\omega, B) = \mathbb{P}(Y \in B | \mathcal{G})(\omega)$ q.c. per ogni fissato $B \in \mathcal{B}$. \diamond

Teorema 5.4.2. Per ogni v.a. Y su $(\Omega, \mathcal{F}, \mathbb{P})$ e per ogni sotto-tribú \mathcal{G} di \mathcal{F} esiste una probabilità condizionata regolare per Y data \mathcal{G} .

Dimostrazione. Il Teorema 5.4.1 assicura che esista una f.r. condizionata regolare F per Y data \mathcal{G} . Si ponga, per $B \in \mathcal{F}$,

$$Q(\omega, B) = \int_{\{t \in B\}} dF(\omega, t).$$

Così, per ogni $\omega \in \Omega$, $Q(\omega, \cdot)$ è la misura di Stieltjes generata da $F(\omega, \cdot)$, e, come tale, $Q(\omega, \cdot)$ è una misura di probabilità.

Si ponga ora $\mathcal{C} := \{B \in \mathcal{B} : Q(\omega, B) = \mathbb{P}(Y \in B | \mathcal{G})(\omega) \text{ q.c.}\}$. Poiché $F(\omega, t) = \mathbb{P}(Y \leq t | \mathcal{G})(\omega)$, l'insieme \mathcal{C} contiene tutte le semirette $]-\infty, t]$ con $t \in \mathbb{R}$. Si vede subito che, se $a < b$, allora anche l'intervallo $]a, b]$ appartiene a \mathcal{C} ; allora vi appartengono anche tutte le unioni finite disgiunte di intervalli aperti a sinistra e chiusi a destra. Per il teorema della classe monotona si ha $\mathcal{C} = \mathcal{B}$. Dunque, Q è una probabilità condizionata regolare per Y data \mathcal{G} . \square

5.5 Note al Capitolo 5

Anche le SC furono introdotte da Kolmogorov nella sua monografia del 1933 già piú volte citata. Un approccio differente è presentato da Brown (1976). Lo studio delle SC è stato esteso a ambiti piú generali di quello nel quale si pongono queste lezioni:

o in spazi misurabili generali, o per v.a. che assumono valori in uno spazio di Banach (Diestel & Uhl, 1977)). Molto studiato è stato il problema della caratterizzazione delle SC; il lavoro pionieristico, in questo campo, è dovuto a Moy (1954). Tra le altre caratterizzazioni (Bahadur, 1955), (Šidák, 1957), (Rota, 1960), (Rao, 1965), (Douglas, 1965), (Olson, 1965), (Ando, 1966), (Pfanzagl, 1967). Nel mio quaderno (Sempi, 1986) ho dato una presentazione unificata, almeno dal punto di vista della notazione, di tali caratterizzazioni.

Si vedano (Pintacuda, 1989) e (Letta & Pratelli, 1997) per ampliamenti sull'importante Teorema 5.3.9.

5.6 Esercizi sul Capitolo 5

5.1. Sullo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, ove $\Omega =]0, 1[$, \mathcal{F} è la tribù di Borel e \mathbb{P} è la misura di Lebesgue, si consideri la partizione

$$\left] 0, \frac{1}{4} \right[, \quad \left[\frac{1}{4}, \frac{1}{2} \right[, \quad \left[\frac{1}{2}, \frac{3}{4} \right[, \quad \left[\frac{3}{4}, 1 \right[.$$

Si calcoli la SC rispetto alla tribù generata da tale partizione della v.a. $X(t) := \sin 2\pi t$.

5.2. (Generalizzazione del Teorema di Bayes) Sia \mathcal{G} una tribù contenuta in \mathcal{F} ; se G è in \mathcal{G} e A è in \mathcal{F} , allora

$$\mathbb{P}(G | A) = \frac{\int \mathbb{P}(A | \mathcal{G}) d\mathbb{P}}{\int \mathbb{P}(A | \mathcal{G}) d\mathbb{P}} .$$

Se la tribù \mathcal{G} è generata da una partizione, finita o numerabile, di Ω in insiemi misurabili,

$$\{B_1, B_2, \dots, B_n, \dots\},$$

allora si riottiene la formulazione classica del Teorema di Bayes,

$$\mathbb{P}(B_j | A) = \frac{\mathbb{P}(A | B_j) \mathbb{P}(B_j)}{\sum_n \mathbb{P}(A | B_n) \mathbb{P}(B_n)} .$$

5.3. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ siano X_1 e X_2 due v.a. indipendenti, entrambe di legge di Poisson di parametro $\lambda > 0$, $X_j \sim \mathcal{P}(\lambda)$ ($j = 1, 2$). Se $Y = X_1 + X_2$, si calcoli $\mathbb{P}(X_1 = k | Y)$.

5.4. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ siano X_1 e X_2 due v.a. indipendenti, entrambe di legge esponenziale di parametro $\lambda > 0$, $X_j \sim \Gamma(1, \lambda)$ ($j = 1, 2$). Se $Y = X_1 + X_2$, si calcoli $\mathbb{E}(X_1 | Y)$.

5.5. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ siano X_1 e X_2 due v.a. indipendenti entrambe di legge di Bernoulli di parametro p . Posto $Y := X_1 + X_2$ e $\mathcal{G} := \mathcal{F}(Y)$,

- (a) si trovino $\mathbb{E}_{\mathcal{G}}(X_1)$ e $\mathbb{E}_{\mathcal{G}}(X_2)$;
- (b) sono indipendenti $\mathbb{E}_{\mathcal{G}}(X_1)$ e $\mathbb{E}_{\mathcal{G}}(X_2)$?

5.6. Si dimostrino le diseguaglianze di Markov e di Čebyšev per le SC; se $\alpha > 0$ e se \mathcal{G} è una tribú contenuta in \mathcal{F} , allora

$$\begin{aligned}\mathbb{P}(|X| \geq \alpha \mid \mathcal{G}) &\leq \frac{1}{\alpha} \mathbb{E}_{\mathcal{G}}(|X|), \quad \text{se } X \in L^1 \\ \mathbb{P}(|X| \geq \alpha \mid \mathcal{G}) &\leq \frac{1}{\alpha^2} \mathbb{E}_{\mathcal{G}}(X^2), \quad \text{se } X \in L^2.\end{aligned}$$

5.7. Per le SC condizionate vale la diseguaglianza di Schwarz; se \mathcal{G} è una tribú contenuta in \mathcal{F} , e se le v.a. X e Y appartengono a L^2 , allora

$$\mathbb{E}_{\mathcal{G}}(XY) \leq \mathbb{E}_{\mathcal{G}}^{1/2}(X^2) \mathbb{E}_{\mathcal{G}}^{1/2}(Y^2).$$

5.8. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ siano X una v.a. di L^2 e \mathcal{G} una tribú contenuta in \mathcal{F} . Posto

$$V_{\mathcal{G}}(X) := \mathbb{E}_{\mathcal{G}} \left[(X - \mathbb{E}_{\mathcal{G}}(X))^2 \right],$$

si ha

$$V(X) = \mathbb{E} [V_{\mathcal{G}}(X)] + V(\mathbb{E}_{\mathcal{G}}(X)).$$

5.9. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia \mathcal{G} una tribú contenuta in \mathcal{F} . Se A è un insieme di \mathcal{F} , si definisca un insieme di \mathcal{G} mediante $B := \{\mathbb{E}_{\mathcal{G}}(\mathbf{1}_A) = 0\}$ e si mostri che, a meno di insiemi trascurabili si ha $B \subseteq A^c$, vale a dire $\mathbb{P}(A \cap B) = 0$.

5.10. Sia X una v.a. con legge $\Gamma(p, \lambda)$ ($\lambda > 0$) e S_x una v.a. con legge di Poisson di parametro x , ove x è il valore assunto da X . Si calcoli $\mathbb{P}(S = n)$; è questa una legge nota?

5.11. Sia (X, Y) un vettore aleatorio con legge normale, la cui densità è data da

$$f(x, y) := \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left(\frac{x^2}{\sigma_1^2} - \frac{2\rho xy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2} \right) \right\}.$$

Si calcolino la densità $h(y \mid x)$ di Y condizionata da $X = x$ e $\mathbb{E}(Y \mid X = x)$.

5.12. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia X una v.a. di L^2 . Se \mathcal{G} è una tribú contenuta in \mathcal{F} , si mostri che

(a) $V(\mathbb{E}_{\mathcal{G}}(X)) \leq V(X)$;

(b) se $Y := X \wedge \alpha$, allora

$$\mathbb{E}_{\mathcal{G}} \left[\{X - \mathbb{E}_{\mathcal{G}}(X)\}^2 \right] \geq \mathbb{E}_{\mathcal{G}} \left[\{Y - \mathbb{E}_{\mathcal{G}}(Y)\}^2 \right].$$

5.13. Si calcolino esplicitamente le SC dell'esempio 5.3.1.

5.14. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ siano X_1, X_2, \dots, X_n v.a. indipendenti ed isonome di L^1 . Posto, al solito, $S_n := \sum_{j=1}^n X_j$, sia \mathcal{G}_n la tribú generata da S_n . Si calcoli la SC $\mathbb{E}_{\mathcal{G}_n}(X_j)$ ($j = 1, 2, \dots, n$).

5.15. Sia T il triangolo di vertici $(0, 0)$, $(1, 0)$ e $(1, 1)$ e sia (X, Y) il vettore aleatorio avente legge uniforme su T ; in altre parole la densità di (X, Y) è data da

$$f(x, y) = 2 \cdot \mathbf{1}_T(x, y).$$

Si calcoli la SC $\mathbb{E}(Y | X)$.

5.16. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ siano X e Y v.a. indipendenti di L^1 ; se Y è centrata, allora

$$\mathbb{E}(|X|) \leq \mathbb{E}(|X + Y|).$$

5.17. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ siano X_1, X_2 e Y v.a. reali tali che i vettori aleatori (X_1, Y) e (X_2, Y) abbiano la stessa legge.

(a) Se $f : \mathbb{R} \rightarrow \mathbb{R}$ è misurabile e positiva (o limitata) si ha, q.c.,

$$\mathbb{E}(f \circ X_1 | Y) = \mathbb{E}(f \circ X_2 | Y)$$

(b) se $g : \mathbb{R} \rightarrow \mathbb{R}$ è misurabile e positiva (o limitata), e se

$$\varphi_j(X_j) := \mathbb{E}(g \circ Y | X_j) \quad (j = 1, 2),$$

allora $\varphi_1 = \varphi_2$ q.c. rispetto alla legge comune di X_1 e X_2 .

5.18. X_1 una v.a. con legge uniforme in $(0, 1)$. Se $X_1 = x_1$, la v.a. X_2 ha legge uniforme su $(0, x_1)$. In generale, se $X_1 = x_1, X_2 = x_2, \dots, X_k = x_k$, allora X_{k+1} ha legge uniforme su $(0, x_k)$. Si trovi $\mathbb{E}(X_n)$.

5.19. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, siano X_1, X_2 e X_3 tre v.a.; sono equivalenti le proprietà:

(a) per ogni $t \in \mathbb{R}$,

$$\mathbb{P}(X_3 \leq t | X_1, X_2) = \mathbb{P}(X_3 \leq t | X_2) \quad P\text{-q.c.};$$

(b) per ogni coppia s e t di numeri reali, è

$$\mathbb{P}(X_1 \leq s, X_3 \leq t | X_2) = \mathbb{P}(X_1 \leq s | X_2) \mathbb{P}(X_3 \leq t | X_2) \quad \mathbb{P}\text{-q.c..}$$

5.20. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ siano \mathcal{G} una sottotribù di \mathcal{F} e X e Y due v.a. tali che X sia indipendente da \mathcal{G} mentre Y è misurabile rispetto a \mathcal{G} . Allora si mostri che

(a) per ogni boreliano $A \in \mathcal{B}^2 = \mathcal{B}(\mathbb{R}^2)$ si ha

$$\mathbb{P}((X, Y) \in A | \mathcal{G}) = \mathbb{P}((X, Y) \in A | Y) \quad \mathbb{P}\text{-q.c.};$$

(b) per ogni boreliano $D \in \mathcal{B}^2$ si ha

$$\mathbb{P}(X + Y \in D | \mathcal{G}) = \mathbb{P}(X + Y \in D | Y) \quad \mathbb{P}\text{-q.c..}$$

5.21. Siano X e Y v.a. di L^1 . Se $\mathbb{E}(Y | X) = X$ e $\mathbb{E}(X | Y) = Y$, allora $X = Y$.

5.22. Siano X e Y due v.a. indipendenti entrambe con legge bernoulliana di parametro p , $X, Y \sim Bi(1, p)$. Introdotta la v.a. Z definita mediante

$$Z := \mathbf{1}_{\{X+Y=0\}},$$

si calcolino le SC $\mathbb{E}(X | Z)$ e $\mathbb{E}(Y | Z)$. Sono indipendenti?

5.23. Si dimostri che $Y \in L^1(\mathcal{G})$ è la SC $\mathbb{E}_{\mathcal{G}}(X)$ se la relazione

$$\int_A X \, d\mathbb{P} = \int_A Y \, d\mathbb{P}$$

vale per ogni insieme A in un π -sistema di sottoinsiemi \mathcal{C} di \mathcal{F} che contiene Ω e che genera la tribú \mathcal{G} , $\mathcal{G} = \mathcal{F}(\mathcal{C})$.

5.24. Siano dati uno spazio di probabilità $(\Omega_1, \mathcal{F}_1, P_1)$ ed uno spazio misurabile $(\Omega_2, \mathcal{F}_2)$. Si dice *probabilità di transizione da $(\Omega_1, \mathcal{F}_1)$ a $(\Omega_2, \mathcal{F}_2)$* una funzione $N : \Omega_1 \times \mathcal{F}_2 \rightarrow \mathbb{R}_+$ tale che

- per ogni $B \in \mathcal{F}_2$ la funzione $\omega_1 \mapsto N(\omega_1, B)$ sia \mathcal{F}_1 -misurabile;
- per ogni $\omega_1 \in \Omega_1$ la funzione $B \mapsto N(\omega_1, B)$ è una misura di probabilità.

Un altro nome per N è quello di *nucleo stocastico*. Allora:

- (a) la funzione $\omega_1 \mapsto \int_{\Omega_2} f(\omega_1, \omega_2) N(\omega_1, d\omega_2)$ è misurabile rispetto alla tribú \mathcal{F}_1 ;
- (b) esiste un'unica misura di probabilità Q su $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ tale che per ogni funzione $f : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}$ misurabile rispetto alla tribú prodotto $\mathcal{F}_1 \otimes \mathcal{F}_2$ e limitata, valga

$$\int_{\Omega_1 \times \Omega_2} f \, dQ = \int_{\Omega_1} d\mathbb{P}_1(\omega_1) \int_{\Omega_2} f(\omega_1, \omega_2) N(\omega_1, d\omega_2); \quad (*)$$

- (c) la (*) vale anche quando la funzione misurabile f sia positiva;
- (d) una funzione $f : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}$ è integrabile rispetto a Q se, e solo se, è

$$\int_{\Omega_1} d\mathbb{P}_1(\omega_1) \int_{\Omega_2} |f(\omega_1, \omega_2)| N(\omega_1, d\omega_2) < +\infty;$$

se f appartiene a $L^1(Q)$ allora vale la (*).

Si noti che da questo esercizio si riottiene il teorema di Fubini quando N non dipende da ω_1 , vale a dire quando $N(\omega_1, B) = \mathbb{P}_2(B)$ per ogni $\omega_1 \in \Omega_1$ e per ogni $B \in \mathcal{F}_2$, dove \mathbb{P}_2 è una misura di probabilità su $(\Omega_2, \mathcal{F}_2)$.

5.25. Nelle condizioni dell'esercizio precedente si considerino le tribú di sottoinsiemi di $\Omega_1 \times \Omega_2$ definite da

$$\widetilde{\mathcal{F}}_1 := \{A \times \Omega_2 : A \in \mathcal{F}_1\} \quad \text{e} \quad \widetilde{\mathcal{F}}_2 := \{\Omega_1 \times B : B \in \mathcal{F}_2\}.$$

Si dimostri che per una funzione $f : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}$ sono equivalenti le affermazioni

- (a) f è misurabile rispetto alla tribú $\widetilde{\mathcal{F}}_1$ (rispettivamente $\widetilde{\mathcal{F}}_2$);
- (b) f dipende dalla sola variabile ω_1 (rispettivamente ω_2) e come tale è misurabile rispetto alla tribú \mathcal{F}_1 (rispettivamente \mathcal{F}_2).

Capitolo 6

Introduzione alle Martingale

6.1 Definizione ed esempi

Benché queste lezioni trattino in prevalenza le martingale a tempo discreto, le definizioni di filtrazione, di martingala e di sottomartingala, sono date nel caso generale.

Ricordiamo che si chiama *diretto* ogni insieme preordinato D nel quale, per ogni coppia s e t di elementi di D , esista un elemento $u \in D$, che sia maggiore tanto di s quanto di t ; se si indica il preordine con “ \leq ”, sarà $s \leq u$ e $t \leq u$. Supporremo, inoltre, che esista in D un unico elemento t_0 tale che, per ogni $t \in D$, sia $t \leq t_0$. Tale elemento sarà indicato con $+\infty$. In un insieme diretto che non sia un sottoinsieme di \mathbb{R} (o di \mathbb{R}) l’espressione “ t tende all’infinito” o, in simboli, “ $t \rightarrow +\infty$ ”, significa che, per ogni $\tilde{t} \in D$ esiste $t \in D$ tale che $t \geq \tilde{t}$.

Definizione 6.1.1. Dati uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ ed un insieme diretto D , si dice *filtrazione* o *scala (stocastica)* una famiglia crescente $\{\mathcal{F}_t : t \in D\}$ di tribù contenute in \mathcal{F} : se s e t sono elementi di D con $s \leq t$, allora $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$.

L’indice t si interpreta come “tempo”; gli eventi di \mathcal{F}_t si dicono *anteriori al tempo t* .

Definizione 6.1.2. Una famiglia $\{X_t : t \in D\}$ di v.a. definite in $(\Omega, \mathcal{F}, \mathbb{P})$ tale che, per ogni t in D , sia misurabile rispetto a \mathcal{F}_t ,

$$\forall t \in D \quad X_t \in \mathcal{L}^0(\mathcal{F}_t).$$

si dice *adattata* alla filtrazione $\{\mathcal{F}_t\}$. ◇

Spesso, nelle applicazioni, è $D = \mathbb{N}$ oppure $D = \mathbb{Z}_+$, o, ancora $D = \mathbb{R}_+$. In termini un poco imprecisi, il concetto di filtrazione riflette il fenomeno dell’accumularsi delle conoscenze al passare del tempo: \mathcal{F}_t rappresenta le conoscenze acquisite sino al tempo t , incluso, sicché considerare una famiglia adattata significa far dipendere la v.a. X_t da ciò che è accaduto sino al tempo t , ma non da ciò che accadrà in futuro.

Si osservi che, data una successione (X_n) di v.a. sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, esiste sempre una filtrazione rispetto alla quale (X_n) è adattata: si tratta della *filtrazione naturale* che è data da

$$\mathcal{F}_1 := \mathcal{F}(X_1), \quad \mathcal{F}_2 := \mathcal{F}(X_1, X_2), \quad \dots, \quad \mathcal{F}_n := \mathcal{F}(X_1, X_2, \dots, X_n), \quad \dots$$

ove $\mathcal{F}(X_1, \dots, X_n)$ denota la tribú generata dalle prime n v.a. della successione (X_n) .

Definizione 6.1.3. Data una filtrazione $\{\mathcal{F}_t : t \in D\}$ sullo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, si dice *martingala*, una famiglia $\{X_t : t \in D\}$ di v.a. di $\mathcal{L}^1(\mathcal{F})$ che sia adattata alla filtrazione data e tale che per ogni coppia s, t di elementi di D con $s \leq t$, valga

$$\mathbb{E}(X_t | \mathcal{F}_s) = X_s. \quad (6.1.1)$$

Si parlerà allora della martingala $\{(X_t, \mathcal{F}_t)\}$. Se, invece della (6.1.1), vale la diseuguaglianza

$$\mathbb{E}(X_t | \mathcal{F}_s) \geq X_s, \quad (6.1.2)$$

$\{(X_t, \mathcal{F}_t)\}$ si dirà *sottomartingala*; se vale la diseuguaglianza

$$\mathbb{E}(X_t | \mathcal{F}_s) \leq X_s, \quad (6.1.3)$$

$\{(X_t, \mathcal{F}_t)\}$ si dirà *supermartingala*. \diamond

Si osservi che ponendo $\mathbb{E}_s(\cdot) := \mathbb{E}(\cdot | \mathcal{F}_s)$, le (6.1.1), (6.1.2) e (6.1.3) si possono scrivere nelle forme abbreviate, valide se $s \leq t$,

$$\mathbb{E}_s(X_t) = X_s, \quad \mathbb{E}_s(X_t) \geq X_s, \quad \mathbb{E}_s(X_t) \leq X_s.$$

Nel seguito adotteremo sistematicamente questa notazione.

Si noti che se $\{(X_t, \mathcal{F}_t)\}$ è una sottomartingala, allora $\{(-X_t, \mathcal{F}_t)\}$ è una supermartingala, e viceversa. Per questa ragione basta studiare le proprietà delle sottomartingale per ottenere facilmente quelle delle supermartingale.

Vale la pena di notare che, nel caso discreto $D = \mathbb{Z}_+$, la definizione di martingala si sarebbe potuta dare in modo leggermente differente, ma equivalente, richiedendo che per ogni $n \in \mathbb{N}$ valga

$$\mathbb{E}_{n-1}(X_n) = X_{n-1}. \quad (6.1.4)$$

È, infatti, evidente che la (6.1.1) implica la (6.1.4). Viceversa, si supponga vera quest'ultima relazione; allora per il Teorema 5.3.8 vale, per $k < n$,

$$\mathbb{E}_k(X_n) = \mathbb{E}_k \mathbb{E}_{n-1}(X_n) = \mathbb{E}_k(X_{n-1}) = \dots = \mathbb{E}_k(X_{k+1}) = X_k.$$

Considerazione analoghe valgono per le definizioni di sotto- e super-martingala.

La definizione di martingala, sotto- e super-martingala può essere introdotta in maniera differente, evitando il ricorso alle speranze condizionate. Al costo di una maggiore pesantezza di notazione, tale approccio ha il vantaggio di evitare il ricorso al teorema di Radon-Nikodym, implicito nell'uso delle speranze condizionate. Poiché è possibile dimostrare il teorema di Radon-Nikodym ricorrendo alle martingale non si sarà quindi compiuto un percorso circolare.

$\{(X_t, \mathcal{F}_t)\}$ è una martingala, una sotto- o una super-martingala se, e solo se, per ogni coppia s e t di elementi di D con $s \leq t$ e per ogni insieme $A \in \mathcal{F}_s$, si ha

rispettivamente

$$\int_A X_t \, d\mathbb{P} = \int_A X_s \, d\mathbb{P}, \quad (6.1.5)$$

$$\int_A X_t \, d\mathbb{P} \geq \int_A X_s \, d\mathbb{P}, \quad (6.1.6)$$

$$\int_A X_t \, d\mathbb{P} \leq \int_A X_s \, d\mathbb{P}. \quad (6.1.7)$$

Si noti che le (6.1.5), (6.1.6), (6.1.7) possono essere scritte nelle forme, a esse rispettivamente equivalenti,

$$\mathbb{E}((X_t - X_s) \mathbf{1}_A) = 0, \quad \mathbb{E}((X_t - X_s) \mathbf{1}_A) \geq 0, \quad \mathbb{E}((X_t - X_s) \mathbf{1}_A) \leq 0.$$

Teorema 6.1.1. *La famiglia delle speranze, $\{\mathbb{E}(X_t) : t \in D\}$, in particolare la successione delle speranze, se $D = \mathbb{N}$ o $D = \mathbb{Z}_+$, è*

- (a) *costante se $\{(X_t, \mathcal{F}_t)\}$ è una martingala,*
- (b) *crescente se $\{(X_t, \mathcal{F}_t)\}$ è una sottomartingala,*
- (c) *decrescente se $\{(X_t, \mathcal{F}_t)\}$ è una supermartingala.*

Dimostrazione. Basta porre $A = \Omega$, che appartiene ad ogni tribú, nella (6.1.5), (6.1.6) e (6.1.7), rispettivamente. \square

In particolare, nel caso di una successione, $(X_n, \mathcal{F}_n)_{n \in \mathbb{Z}_+}$, indicati con

$$D_n := X_n - X_{n-1}$$

gli incrementi di una martingala (X_n, \mathcal{F}_n) , si ha, per ogni $n \in \mathbb{N}$,

$$\mathbb{E}_{n-1}(D_n) = 0.$$

Se non è specificata la filtrazione, dicendo che la successione (X_t) è una martingala, si intende senz'altro che la filtrazione sia quella naturale. Analoghe considerazioni valgono per le sotto- e le super-martingale.

D'ora in poi, senza che ciò sia espressamente richiamato, supporremo sempre che le v.a. considerate siano definite sopra uno spazio di probabilità comune $(\Omega, \mathcal{F}, \mathbb{P})$.

Esempio 6.1.1. Una successione costante c di v.a. (cioè $X_t = c$ per ogni $n \in D$) è una martingala. \blacksquare

Esempio 6.1.2. Siano (\mathcal{F}_t) una filtrazione di $(\Omega, \mathcal{F}, \mathbb{P})$ e X una v.a. di $L^1(\mathcal{F})$; si ponga

$$X_t := \mathbb{E}_t(X).$$

Allora (X_t, \mathcal{F}_t) è una martingala; infatti per il Teorema 5.3.7, se $s < t$,

$$\mathbb{E}_s(X_t) = \mathbb{E}_s \mathbb{E}_t(X) = \mathbb{E}_s(X) = X_s.$$

Una tale martingala si dice *ereditaria*. \blacksquare

Esempio 6.1.3. Sia (X_n) una successione di v.a. indipendenti e centrate di $L^1(\mathcal{F})$ e si consideri la filtrazione naturale (\mathcal{F}_n) . Posto, come al solito, $S_n := \sum_{j=1}^n X_j$, si ha che (S_n, \mathcal{F}_n) è una martingala. Per controllarlo si usano l'additività delle speranze condizionate e il Teorema 5.3.8:

$$\begin{aligned}\mathbb{E}_n(S_{n+1}) &= \mathbb{E}_n(S_n + X_{n+1}) = \mathbb{E}_n(S_n) + \mathbb{E}_n(X_{n+1}) \\ &= S_n + \mathbb{E}_n(X_{n+1}) = S_n + \mathbb{E}(X_{n+1}) = S_n.\end{aligned}$$

Si osservi che poiché $X_n = S_n - S_{n-1}$ la filtrazione (\mathcal{F}_n) è anche la filtrazione naturale della successione (S_n) . ■

Esempio 6.1.4. Sia $(Y_n)_{n \in \mathbb{N}}$ una successione di v.a. indipendenti di $\mathcal{L}^1(\mathcal{F})$ con

$$\mathbb{E}(Y_n) = \alpha_n \neq 0 \quad \text{per ogni } n \in \mathbb{N}.$$

Si ponga $\mathcal{F}_n := \mathcal{F}(Y_1, Y_2, \dots, Y_n)$ e

$$X_n := \prod_{j=1}^n \frac{Y_j}{\alpha_j};$$

allora (X_n, \mathcal{F}_n) è una martingala:

$$\begin{aligned}\mathbb{E}_n(X_{n+1}) &= \mathbb{E}_n\left(\frac{X_n Y_{n+1}}{\alpha_{n+1}}\right) = \frac{X_n}{\alpha_{n+1}} \mathbb{E}_n(Y_{n+1}) \\ &= \frac{X_n}{\alpha_{n+1}} \mathbb{E}(Y_{n+1}) = X_n.\end{aligned}$$

Un martingala siffatta si dice *moltiplicativa*. ■

Teorema 6.1.2. *Le martingale rispetto alla stessa filtrazione (\mathcal{F}_t) formano uno spazio vettoriale (complesso) denotato con $\mathcal{M}(\mathcal{F}_t)$.*

Dimostrazione. Se (X_t, \mathcal{F}_t) è una martingala, per ogni $\alpha \in \mathbb{C}$, $(\alpha X_t, \mathcal{F}_t)$ è, ovviamente, una martingala. Se (X'_t) e (X''_t) sono due martingale rispetto alla stessa filtrazione (\mathcal{F}_t) , anche $(X'_t + X''_t)$ è una martingala rispetto alla stessa filtrazione:

$$\mathbb{E}_s(X'_t + X''_t) = \mathbb{E}_s(X'_t) + \mathbb{E}_s(X''_t) = X'_s + X''_s,$$

one l'asserto. □

È importante il caso particolare delle martingale *quadratiche*.

Definizione 6.1.4. Si dice che (X_n, \mathcal{F}_n) è una *martingala quadratica* se, oltre a soddisfare alla condizione (6.1.1), X_n appartiene a $\mathcal{L}^2(\mathcal{F}_n)$, per ogni $n \in \mathbb{Z}_+$. ◇

Teorema 6.1.3. *Sia (X_n, \mathcal{F}_n) una martingala quadratica; allora, per gli incrementi di martingala $D_n := X_n - X_{n-1}$ valgono le proprietà*

- (a) per ogni $n \in \mathbb{Z}_+$, è $\mathbb{E}_n(D_{n+1}^2) = \mathbb{E}_n[(X_{n+1} - X_n)^2] = \mathbb{E}_n(X_{n+1}^2) - X_n^2$;
- (b) se $j \neq k$, allora $\mathbb{E}(D_j D_k) = 0$; gli incrementi D_j sono dunque ortogonali se considerati come vettori di \mathcal{L}^2 , incorrelati se riguardati come v.a..

Dimostrazione. (a) Si ha

$$\begin{aligned}\mathbb{E}_n [(X_{n+1} - X_n)^2] &= \mathbb{E}_n (X_{n+1}^2) + \mathbb{E}_n (X_n^2) - 2\mathbb{E}_n (X_n X_{n+1}) \\ &= \mathbb{E}_n (X_{n+1}^2) + X_n^2 - 2X_n \mathbb{E}_n (X_{n+1}) \\ &= \mathbb{E}_n (X_{n+1}^2) + X_n^2 - 2X_n^2 = \mathbb{E}_n (X_{n+1}^2) - X_n^2.\end{aligned}$$

(b) Si supponga che sia $j < k$; allora

$$\begin{aligned}\mathbb{E}(D_j D_k) &= \mathbb{E}_{\mathcal{N}}(D_j D_k) = \mathbb{E}_{\mathcal{N}} \mathbb{E}_{k-1}(D_j D_k) \\ &= \mathbb{E}_{\mathcal{N}}(D_j \mathbb{E}_{k-1}(D_k)) = \mathbb{E}_{\mathcal{N}}(D_j (X_{k-1} - X_{k-1})) = 0,\end{aligned}$$

cioè l'asserto. □

Sia (X_n) una martingala quadratica; essa può essere espressa nella forma

$$X_n = X_0 + \sum_{j=1}^n (X_j - X_{j-1}) = X_0 + \sum_{j=1}^n D_j,$$

nella quale X_n risulta scritta come somma di elementi ortogonali di L^2 . Ricorrendo al Teorema 6.1.3 si ha

$$\mathbb{E}(X_n^2) = \mathbb{E}(X_0^2) + \sum_{j=1}^n \mathbb{E}[(X_j - X_{j-1})^2].$$

Perciò una martingala quadratica (X_n) è limitata in L^2 , vale a dire che si ha

$$\sup_{n \in \mathbb{N}} \|X_n\|_2 < +\infty,$$

se, e solo se,

$$\sum_{j=1}^n \mathbb{E}[(X_j - X_{j-1})^2] < +\infty.$$

Teorema 6.1.4. *Siano date $(X_t, \mathcal{F}_t)_{n \in D}$ una sottomartingala e una funzione convessa $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ tale che, per ogni $n \in D$, sia integrabile la v.a. $\varphi \circ X_t$. Se vale una delle seguenti condizioni:*

- (a) φ è crescente,
- (b) (X_t, \mathcal{F}_t) è una martingala,

allora $(\varphi \circ X_t, \mathcal{F}_t)$ è una sottomartingala.

Dimostrazione. In entrambe le ipotesi, la disuguaglianza di Jensen dà

$$\mathbb{E}_s(\varphi \circ X_t) \geq \varphi(\mathbb{E}_s(X_t)).$$

Se φ è crescente, $\varphi(\mathbb{E}_s(X_t)) \geq \varphi \circ X_s$ dalla definizione di sottomartingala; se, invece, (X_t, \mathcal{F}_t) è una martingala, $\varphi(\mathbb{E}_s(X_t)) = \varphi \circ X_s$. □

In particolare, se (X_t, \mathcal{F}_t) è una sottomartingala, anche (X_t^+, \mathcal{F}_t) è tale; se (X_t, \mathcal{F}_t) è una martingala, $(|X_t|^p, \mathcal{F}_t)$ è una sottomartingala per ogni p in $[1, +\infty[$.

Nel caso discreto, $D = \mathbb{Z}_+$, il prossimo risultato consentirà di ricondurre lo studio della convergenza di sotto- e super-martingale a quello delle martingale.

Teorema 6.1.5. (Decomposizione di Doob). *Se $(X_n, \mathcal{F}_n)_{n \in \mathbb{Z}_+}$ è una sottomartingala, per ogni $n \in \mathbb{Z}_+$, esistono v.a. Y_n e A_n tali che*

- (a) $X_n = Y_n + A_n$;
- (b) $(Y_n, \mathcal{F}_n)_{n \in \mathbb{Z}_+}$ è una martingala;
- (c) $A_0 = 0, A_n \leq A_{n+1}$;
- (d) per ogni $n \in \mathbb{N}$, A_n è misurabile rispetto a \mathcal{F}_{n-1} , $A_n \in L^0(\mathcal{F}_{n-1})$.

Le v.a. Y_n e A_n sono univocamente determinate. Un processo con le proprietà (c) e (d) si dice crescente.

Analogamente una supermartingala (X_n, \mathcal{F}_n) si può decomporre nella differenza $X_n = Y_n - A_n$ ove (Y_n, \mathcal{F}_n) è una martingala e $\{A_n\}$ è un processo crescente.

Dimostrazione. Si definisca per ricorrenza

$$A_0 := 0, \quad A_{n+1} := A_n + \mathbb{E}_n(X_{n+1}) - X_n \quad (n \in \mathbb{N}).$$

Dunque,

$$A_n := \sum_{j=1}^n (\mathbb{E}_{j-1}(X_j) - X_{j-1}),$$

che è misurabile rispetto a \mathcal{F}_{n-1} e che è una somma a termini positivi, poiché $\{(X_n, \mathcal{F}_n)\}$ è una sottomartingala; perciò $A_{n+1} \geq A_n$. Si ponga

$$Y_n := X_n - A_n = X_n - \sum_{j=1}^n (\mathbb{E}_{j-1}(X_j) - X_{j-1}).$$

Con queste definizioni si sono provate (a), (c) e (d).

Per dimostrare (b), si calcoli

$$\begin{aligned} \mathbb{E}_n(Y_{n+1}) &= \mathbb{E}_n(X_{n+1}) - \mathbb{E}_n \left\{ \sum_{j=1}^{n+1} (\mathbb{E}_{j-1}(X_j) - X_{j-1}) \right\} \\ &= \mathbb{E}_n(X_{n+1}) - \mathbb{E}_n(\mathbb{E}_n(X_{n+1}) - X_n) - \sum_{j=1}^n \mathbb{E}_n(\mathbb{E}_{j-1}(X_j) - X_{j-1}) \\ &= X_n - \sum_{j=1}^n (\mathbb{E}_{j-1}(X_j) - X_{j-1}) = Y_n, \end{aligned}$$

sicché $\{(Y_n, \mathcal{F}_n)\}$ è una martingala.

Resta da dimostrare l'unicità della decomposizione. Siano (Y'_n) e (A'_n) due successioni di v.a. con le stesse proprietà di (Y_n) e di (A_n) rispettivamente. Si ha intanto che $A'_0 = 0$ e $Y'_0 = Y_0$. Ragionando per induzione, si supponga di aver dimostrato

che, per $j = 1, 2, \dots, n$, vale $Y'_j = Y_j$ e $A'_j = A_j$. Allora, poiché tanto A'_{n+1} quanto A_{n+1} sono \mathcal{F}_n -misurabili, si ha

$$\begin{aligned} A'_{n+1} &= \mathbb{E}_n(A'_{n+1}) = \mathbb{E}_n(X_{n+1} - Y'_{n+1}) = \mathbb{E}_n(X_{n+1}) - Y'_n \\ &= \mathbb{E}_n(X_{n+1}) - Y_n = \mathbb{E}_n(X_{n+1} - Y_{n+1}) = \mathbb{E}_n(A_{n+1}) = A_{n+1}. \end{aligned}$$

Ne segue che anche $Y'_{n+1} = Y_{n+1}$. \square

Sia (X_n) una martingala quadratica tale che $X_0 = 0$; allora il Teorema 6.1.4 assicura che (X_n^2) sia una sottomartingala. Se ne consideri la decomposizione di Doob

$$X_n^2 = Y_n + A_n \quad (n \in \mathbb{N}),$$

ove (Y_n) è una martingala; poiché il processo (A_n) è crescente, esiste il suo limite quasi certo

$$A_\infty := \sup_{n \in \mathbb{N}} A_n.$$

Ora

$$\mathbb{E}(X_n^2) = \mathbb{E}(Y_n) + \mathbb{E}(A_n) = \mathbb{E}(Y_0) + \mathbb{E}(A_n) = \mathbb{E}(X_0) + \mathbb{E}(A_n) = \mathbb{E}(A_n),$$

sicché la sotto-martingala (X_n^2) è limitata in L^2 se, e solo se, $\mathbb{E}(A_\infty) < +\infty$. Si ha, inoltre, poiché A_n è misurabile rispetto a \mathcal{F}_{n-1} ,

$$\begin{aligned} A_n - A_{n-1} &= \mathbb{E}_{n-1}(A_n - A_{n-1}) = \mathbb{E}_{n-1}(X_n^2 - X_{n-1}^2) - \mathbb{E}_{n-1}(Y_n - Y_{n-1}) \\ &= \mathbb{E}_{n-1}(X_n^2 - X_{n-1}^2) = \mathbb{E}_{n-1}\left[(X_n - X_{n-1})^2\right], \end{aligned}$$

ove nell'ultimo passaggio si è usato il Teorema 6.1.3 (a).

6.2 Tempi d'arresto

In questa sezione ci occuperemo solo di filtrazioni numerabili e di successioni di v.a..

Definizione 6.2.1. Data una filtrazione $(\mathcal{F}_n)_{n \in \mathbb{Z}_+}$ in uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, si dice *tempo d'arresto rispetto alla filtrazione* (\mathcal{F}_n) ogni v.a.

$$T : \Omega \rightarrow \overline{\mathbb{Z}_+} := \mathbb{Z}_+ \cup \{+\infty\}$$

tale che, per ogni $n \in \mathbb{Z}_+$, $\{T \leq n\}$ appartenga a \mathcal{F}_n . \diamond

La condizione $\{T \leq n\} \in \mathcal{F}_n$ equivale all'altra $\{T = n\} \in \mathcal{F}_n$ per ogni $n \in \mathbb{N}$. Infatti, $\{T \leq n\} = \cup_{k \leq n} \{T = k\}$ e $\{T = k\}$ appartiene a \mathcal{F}_n per ogni $k \leq n$; viceversa,

$$\{T = n\} = \{T \leq n\} \setminus \{T \leq n-1\},$$

che appartiene alla tribú \mathcal{F}_n .

Esempio 6.2.1. Se (X_n) è una successione adattata alla filtrazione $\{\mathcal{F}_n\}$ e B è un boreliano di \mathbb{R} , allora il *tempo d'ingresso* in B ,

$$T_B := \inf\{n \in \overline{\mathbb{N}} : X_n \in B\}$$

è un tempo d'arresto. Infatti, per ogni $n \in \mathbb{N}$, è

$$\{T_B > n\} = \bigcap_{k=0}^n \{X_k \notin B\} \in \mathcal{F}_n,$$

onde $\{T_B \leq n\} = \{T_B > n\}^c \in \mathcal{F}_n$. ■

D'ora in poi si supponrà tacitamente che sia assegnata una filtrazione (\mathcal{F}_n) ; quando si parlerà di tempi d'arresto, si intenderà che questi siano considerati rispetto a tale filtrazione.

Un tempo d'arresto T si dice (q.c.) *finito* se $\mathbb{P}(T = +\infty) = 0$.

Dato un tempo d'arresto T si definisca una famiglia di sottoinsiemi di Ω mediante

$$\mathcal{F}_T := \{A \in \mathcal{F} : \forall n \in \mathbb{N} \quad A \cap \{T \leq n\} \in \mathcal{F}_n\}.$$

Teorema 6.2.1. *Per ogni tempo d'arresto T la famiglia \mathcal{F}_T è una tribù.*

Dimostrazione. L'appartenenza di Ω a \mathcal{F}_T scende direttamente dall'aver supposto che T sia un tempo d'arresto.

Si supponga ora che A sia un insieme di \mathcal{F}_T , cioè che sia $A \cap \{T \leq n\} \in \mathcal{F}_n$ per ogni $n \in \mathbb{N}$. Poiché $\{T > n\}$ è in \mathcal{F}_{n-1} e quindi anche in \mathcal{F}_n , appartiene a \mathcal{F}_n l'insieme

$$(A \cap \{T \leq n\}) \cup \{T > n\} = (A \cup \{T > n\}) \cap \Omega = A \cup \{T > n\},$$

e dunque anche il suo complementare $A^c \cap \{T \leq n\}$. Perciò anche A^c appartiene a \mathcal{F}_T .

Sia, infine, $(A_j)_{j \in \mathbb{N}}$ una successione di insiemi di \mathcal{F}_T ; pertanto, per ogni $n \in \mathbb{N}$, si ha

$$\left(\bigcup_{j \in \mathbb{N}} A_j \right) \cap \{T \leq n\} = \bigcup_{j \in \mathbb{N}} (A_j \cap \{T \leq n\}) \in \mathcal{F}_n.$$

Dunque, l'unione numerabile $\bigcup_{j \in \mathbb{N}} A_j$ appartiene a \mathcal{F}_T . □

Teorema 6.2.2. *Siano S e T tempi d'arresto. Allora*

- (a) T è \mathcal{F}_T -misurabile;
- (b) $S \wedge T$ e $S \vee T$ sono tempi d'arresto;
- (c) se $S \leq T$ allora $\mathcal{F}_S \subseteq \mathcal{F}_T$.

Dimostrazione. (a) Per ogni $c > 0$ si ha $\{T \leq c\} = \{T \leq [c]\}$ ove $[c]$ è la parte intera di c . Ora, per ogni $n \in \mathbb{N}$, è

$$\{T \leq [c]\} \cap \{T \leq n\} = \begin{cases} \{T \leq [c]\}, & \text{se } [c] \leq n, \\ \{T \leq n\}, & \text{se } [c] > n; \end{cases}$$

in entrambi i casi $\{T \leq [c]\} \cap \{T \leq n\}$ appartiene a \mathcal{F}_n , sicché T è misurabile rispetto a \mathcal{F}_T .

(b) Sia $n \in \mathbb{N}$; allora

$$\{S \wedge T \leq n\} = \{S \leq n\} \cup \{T \leq n\} \in \mathcal{F}_n,$$

e

$$\{S \vee T \leq n\} = \{S \leq n\} \cap \{T \leq n\} \in \mathcal{F}_n.$$

(c) Si supponga $S \leq T$ e $A \in \mathcal{F}_S$. Ora, $\{T \leq n\}$ è contenuto in $\{S \leq n\}$. Infatti la condizione $S \leq T$ implica che, per ogni $n \in \mathbb{N}$, valga l'inclusione

$$\{S > n\} \subseteq \{T > n\};$$

pertanto

$$\{T \leq n\} = \{T > n\}^c \subseteq \{S > n\}^c = \{S \leq n\}.$$

In definitiva, per ogni $n \in \mathbb{N}$, si ha

$$A \cap \{T \leq n\} = A \cap \{S \leq n\} \cap \{T \leq n\}$$

che appartiene a \mathcal{F}_n , poiché, per ipotesi, tanto $A \cap \{S \leq n\}$ quanto $\{T \leq n\}$ appartengono a \mathcal{F}_n . \square

Data una successione (X_n) adattata a (\mathcal{F}_n) e un tempo d'arresto T , si denoti con X_T la funzione $X_T : \Omega \rightarrow \mathbb{R}$ definita da

$$X_T(\omega) := X_{T(\omega)}(\omega) = \sum_{n \in \mathbb{N}} X_n(\omega) \mathbf{1}_{\{T=n\}}(\omega) + \infty \mathbf{1}_{\{T=+\infty\}}.$$

Si è così definita una v.a.; infatti se B è un boreliano della retta, si ha

$$X_T^{-1}(B) = \{X_T \in B\} = \bigcup_{n \in \mathbb{N}} (\{T = n\} \cap \{X_n \in B\}) \in \mathcal{F}.$$

Nelle stesse condizioni si dice *processo arrestato al tempo T* il processo stocastico¹ $X^T := (X_{T \wedge n})$, cioè $X_n^T := X_{T \wedge n}$ per ogni $n \in \mathbb{N}$.

Si noti che, mentre X_T è una v.a., X^T indica un processo stocastico.

Teorema 6.2.3. *Se T è un tempo d'arresto e $(X_n)_{n \in \mathbb{Z}_+}$ un processo adattato a $(\mathcal{F}_n)_{n \in \mathbb{Z}_+}$, vale la relazione*

$$X_n^T - X_{n-1}^T = (X_n - X_{n-1}) \mathbf{1}_{\{T \geq n\}} \quad (n \in \mathbb{N}), \quad (6.2.1)$$

detta identità d'arresto.

Dimostrazione. Se ω appartiene a $\{T < n\}$, è $X_{T \wedge n}(\omega) = X_{T \wedge (n-1)}(\omega)$; se, invece, ω è in $\{T \geq n\}$, allora $T(\omega) \wedge n = n$ e $T(\omega) \wedge (n-1) = n-1$. \square

Sarà utile il seguente lemma.

¹Un processo stocastico in tempo discreto è semplicemente una successione $(X_n)_{n \in \mathbb{N}}$ di variabili aleatorie definite sullo stesso spazio di probabilità.

Lemma 6.2.1. (Wald). *Siano $(Z_n)_{n \in \mathbb{Z}_+}$ una successione di v.a. tale che Z_0 sia integrabile, $\mathbb{E}(|Z_0|) < +\infty$, e tale che, per ogni $n \in \mathbb{N}$, sia*

$$|Z_n - Z_{n-1}| \leq \alpha,$$

ove $\alpha > 0$ è una costante, e T un tempo d'arresto a valori in $\bar{\mathbb{N}}$ pure integrabile, $\mathbb{E}(T) < +\infty$. Allora la successione Z^T arrestata al tempo T è dominata in L^1 , vale a dire che esiste $Y \in L^1$ tale che, per ogni $n \in \mathbb{Z}_+$, sia $|Z_{n \wedge T}| \leq Y$.

Dimostrazione. È immediato scrivere

$$Z_{n \wedge T} = Z_0 + \sum_{k \in \mathbb{N}} \mathbf{1}_{[1, n \wedge T]}(k) (Z_k - Z_{k-1}).$$

Di qui

$$|Z_{n \wedge T}| \leq |Z_0| + \alpha \sum_{k \in \mathbb{N}} \mathbf{1}_{[1, n \wedge T]}(k) = |Z_0| + \alpha (n \wedge T) \leq |Z_0| + \alpha T,$$

onde l'asserto. □

6.3 Martingale e tempi d'arresto

Questa sezione è dedicata interamente allo studio dei rapporti tra martingale e tempi d'arresto. Premettiamo una definizione e un risultato preliminare. Nel seguito, per comodità, supporremo che l'insieme dei tempi sia \mathbb{Z}_+ .

Definizione 6.3.1. Sia data una filtrazione $(\mathcal{F}_n)_{n \in \mathbb{Z}_+}$ sullo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$; un processo $(U_n)_{n \in \mathbb{Z}_+}$ si dice *prevedibile* se, per ogni $n \in \mathbb{N}$, U_n è misurabile rispetto a \mathcal{F}_{n-1} , $U_n \in L^0(\mathcal{F}_{n-1})$ per ogni $n \in \mathbb{N}$. ◇

Definizione 6.3.2. Si dice *compensatore* di un processo $(X_n)_{n \in \mathbb{Z}_+}$ un processo prevedibile $(V_n)_{n \in \mathbb{Z}_+}$ tale che $(X_n - V_n)$ sia un martingala. ◇

Naturalmente V_n sarà integrabile per ogni n . Se $(X_n - V_n)$ è una martingala si ha, per $n \in \mathbb{N}$,

$$\mathbb{E}_{n-1}(X_n - V_n) = X_{n-1} - V_{n-1},$$

onde

$$\mathbb{E}_{n-1}(X_n - X_{n-1}) = V_n - V_{n-1},$$

che basta per individuare il compensatore; infatti posto $V_0 = 0$ è

$$V_n = \sum_{j=1}^n (V_j - V_{j-1}) = \sum_{j=1}^n \mathbb{E}_{n-1}(X_j - X_{j-1}).$$

Esempio 6.3.1. Data una martingala (X_n) si sa dal Teorema 6.1.4 che (X_n^2) è una sotto-martingala. Allora gli incrementi

$$V_n - V_{n-1} = \mathbb{E}_{n-1}(X_n^2 - X_{n-1}^2) = \mathbb{E}_{n-1}[(X_n - X_{n-1})^2]$$

individuano il compensatore di (X_n^2) . ■

Diamo allora la definizione di trasformata di una martingala.

Definizione 6.3.3. Dati una martingala $(X_n, \mathcal{F}_n)_{n \in \mathbb{Z}_+}$ e un processo prevedibile $(U_n)_{n \in \mathbb{Z}_+}$, il processo $((U \cdot X)_n)_{n \in \mathbb{Z}_+}$ definito da

$$\begin{aligned} (U \cdot X)_0 &:= U_0 X_0 \\ (U \cdot X)_n &:= U_0 X_0 + \sum_{j=1}^n U_j (X_j - X_{j-1}) \quad (n \in \mathbb{N}) \end{aligned} \quad (6.3.1)$$

si dice *trasformata della martingala* (X_n) . \diamond

Teorema 6.3.1. *Con le stesse notazioni della Definizione 6.3.3 se nell'eq. (6.3.1) le variabili $(U \cdot X)_n$ sono integrabili, $(U \circ X)_n \in L^1$ per ogni $n \in \mathbb{N}$, allora*

- (a) *se (X_n, \mathcal{F}_n) è una martingala, tale è anche $((U \cdot X)_n, \mathcal{F}_n)$;*
- (b) *se (X_n, \mathcal{F}_n) è una sotto-martingala e se $U_n \geq 0$ per ogni $n \in \mathbb{N}$, è una sotto-martingala anche $((U \cdot X)_n, \mathcal{F}_n)$.*

Dimostrazione. Si osservi che, in virtù della (6.3.1) e del fatto che (U_n) è prevedibile, si ha, per $n \in \mathbb{N}$,

$$\begin{aligned} \mathbb{E}_{n-1} [(U \cdot X)_n] &= \mathbb{E}_{n-1} \left[U_0 X_0 + \sum_{j=1}^n U_j (X_j - X_{j-1}) \right] \\ &= U_0 X_0 + \sum_{j=1}^{n-1} U_j (X_j - X_{j-1}) + U_n \mathbb{E}_{n-1} (X_n - X_{n-1}). \end{aligned}$$

Se (X_n, \mathcal{F}_n) è una martingala, si ha immediatamente

$$\mathbb{E}_{n-1} [(U \cdot X)_n] = U_0 X_0 + \sum_{j=1}^{n-1} U_j (X_j - X_{j-1}) = (U \cdot X)_{n-1},$$

sicché $((U \cdot X)_n, \mathcal{F}_n)$ è una martingala, mentre, se (X_n, \mathcal{F}_n) è una sotto-martingala e se $U_n \geq 0$, allora

$$\mathbb{E}_{n-1} [(U \cdot X)_n] \geq U_0 X_0 + \sum_{j=1}^{n-1} U_j (X_j - X_{j-1}) = (U \cdot X)_{n-1},$$

sicché $((U \cdot X)_n, \mathcal{F}_n)$ è una sotto-martingala. \square

Si noti che basta che ogni U_n sia limitata affinché $(U \circ X)_n$ sia integrabile.

Teorema 6.3.2. *Se (X_n) è una sotto-martingala (o una martingala) e T è un tempo d'arresto, X^T è ancora una sotto-martingala (rispettivamente una martingala). In altre parole, arrestando una (sotto-)martingala si ottiene ancora una (sotto-)martingala.*

Dimostrazione. Basta infatti considerare il processo prevedibile definito da $U_0 = 1$ e, per $n \in \mathbb{N}$, da $U_n = \mathbf{1}_{\{T \geq n\}}$. Allora $(U \cdot X)_0 = X_0$, mentre per $n \in \mathbb{N}$, si ha, in virtù dell'identità d'arresto (6.2.1),

$$(U \cdot X)_n = X_0 + \sum_{j=1}^n (X_j - X_{j-1}) \mathbf{1}_{\{T \geq j\}} = X_0 + \sum_{j=1}^n (X_j^T - X_{j-1}^T) = X_n^T;$$

il risultato è ora una diretta conseguenza del Teorema 6.3.1. \square

Teorema 6.3.3. *Siano $(X_n, \mathcal{F}_n)_{n \in \mathbb{Z}_+}$ una sottomartingala e S e T due tempi d'arresto limitati con $S \leq T \leq N$ ($N \in \mathbb{N}$). Allora,*

$$\mathbb{E}(X_T | \mathcal{F}_S) \geq X_S.$$

Se (X_n, \mathcal{F}_n) è una martingala, nell'ultima relazione vale il segno d'eguaglianza.

Dimostrazione. Si ha, intanto, $|X_S| \leq \sum_{j=0}^N |X_j|$, sicché X_S è in $L^1(\mathcal{F})$. Se A è in \mathcal{F}_S e $j \leq N$, si ponga $A_j := A \cap \{S = j\} \in \mathcal{F}_j$ e, per $j \leq k \leq N$,

$$\begin{aligned} A(j, k) &:= A_j \cap \{T = k\}, \\ U(j, k) &:= \cup_{i \geq k} A(j, i) = A_j \cap \{T \geq k\}, \\ V(j, k) &:= U(j, k+1) = A_j \cap \{T > k\}. \end{aligned}$$

Tutti e tre questi insiemi appartengono a \mathcal{F}_k . Segue dalla definizione di sottomartingala che

$$\int_{V(j, k)} X_k \, d\mathbb{P} \leq \int_{V(j, k)} X_{k+1} \, d\mathbb{P};$$

e, poiché $U(j, k) = A(j, k) \cup V(j, k)$, che è un'unione disgiunta,

$$\begin{aligned} \int_{U(j, k)} X_k \, d\mathbb{P} &= \int_{A(j, k)} X_k \, d\mathbb{P} + \int_{V(j, k)} X_k \, d\mathbb{P} \\ &\leq \int_{A(j, k)} X_k \, d\mathbb{P} + \int_{U(j, k+1)} X_{k+1} \, d\mathbb{P}, \end{aligned}$$

onde,

$$\int_{U(j, k)} X_k \, d\mathbb{P} - \int_{U(j, k+1)} X_{k+1} \, d\mathbb{P} \leq \int_{A(j, k)} X_k \, d\mathbb{P} = \int_{A(j, k)} X_T \, d\mathbb{P},$$

cioè

$$\int_{U(j, k)} X_k \, d\mathbb{P} - \int_{U(j, k+1)} X_{k+1} \, d\mathbb{P} \leq \int_{A(j, k)} X_T \, d\mathbb{P}.$$

Si sommi ora, in quest'ultima disequaglianza, per k tra j e N e si osservi che è $A_j = U(j, j)$ e che $U(j, N+1) = \emptyset$:

$$\int_{A_j} X_S \, d\mathbb{P} = \int_{A_j} X_j \, d\mathbb{P} \leq \int_{A_j} X_T \, d\mathbb{P}.$$

Si sommi, infine, su j tra 0 e N per ottenere

$$\int_A X_S \, d\mathbb{P} \leq \int_A X_T \, d\mathbb{P},$$

onde l'asserto. \square

Sia T un tempo d'arresto limitato, $T \leq N$. Se (X_n) è una martingala, si può ricorrere al Teorema 6.3.3 per dimostrare che è, per ogni $n \in \mathbb{Z}_+$,

$$\mathbb{E}(X_T) = \mathbb{E}(X_N) = \mathbb{E}(X_0) = \mathbb{E}(X_n);$$

in altre parole, non si può aumentare la speranza della vincita arrestando un gioco equo ad un tempo aleatorio limitato T che dipende dall'informazione disponibile sino al momento dell'arresto.

Se (X_n) è una sottomartingala, vale

$$\mathbb{E}(X_0) \leq \mathbb{E}(X_T) \leq \mathbb{E}(X_N).$$

Si osservi che non è detto che X_T sia in L^1 anche se T è q.c. finito. Si veda, a tal proposito, il seguente

Esempio 6.3.2. Si ritorni all'esempio 6.1.4 e si consideri la martingala moltiplicativa

$$X_n := \prod_{j=1}^n (1 + 3R_j)$$

ove le R_j sono le funzioni di Rademacher. Si consideri il tempo d'arresto

$$T := \min\{n \in \mathbb{N} : R_n < 0\}.$$

Ora

$$\{T = n\} = \left[1 - \frac{1}{2^{n-1}}, 1 - \frac{1}{2^n}\right],$$

sicché $\mathbb{P}(T = n) = 2^{-n}$ e

$$\begin{aligned} \mathbb{E}(T) &= \sum_{n=1}^{\infty} \frac{n}{2^n} = \frac{1}{2} \sum_{n=1}^{\infty} n 2^{-(n-1)} = \frac{1}{2} \sum_{n=1}^{\infty} \left[\frac{d}{dt} t^n \right]_{t=1/2} \\ &= \frac{1}{2} \left[\frac{d}{dt} \sum_{n=0}^{\infty} t^n \right]_{t=1/2} = \frac{1}{2} \left[\frac{d}{dt} \frac{1}{1-t} \right]_{t=1/2} \\ &= \frac{1}{2} \left[\frac{1}{(1-t)^2} \right]_{t=1/2} = 2. \end{aligned}$$

Dunque T è q.c. finito. Non è difficile riconoscere che

$$X_T = -2 \cdot 4^{T-1} = -\frac{1}{2} 4^T,$$

onde

$$\mathbb{E}(|X_T|) = \frac{1}{2} \sum_{n \in \mathbb{N}} 4^n \mathbb{P}(T = n) = \frac{1}{2} \sum_{n \in \mathbb{N}} 2^n = +\infty.$$

■

Teorema 6.3.4. *Se (X_n) è una martingala limitata in L^1 e T è un tempo d'arresto q.c. finito, allora X_T appartiene a L^1 .*

Dimostrazione. Se è $\|X_n\|_1 \leq H$ per ogni $n \in \mathbb{N}$, allora,

$$\begin{aligned} \mathbb{E}(|X_{T \wedge n}|) &= \sum_{j=1}^n \mathbb{E}(|X_j| \mathbf{1}_{\{T=j\}}) + \mathbb{E}(|X_n| \mathbf{1}_{\{T>n\}}) \\ &= \sum_{j=1}^n \int_{\{T=j\}} |X_j| \, d\mathbb{P} + \mathbb{E}(|X_n| \mathbf{1}_{\{T>n\}}) \\ &\leq \sum_{j=1}^n \int_{\{T=j\}} \mathbb{E}_j(|X_n|) \, d\mathbb{P} + \int_{\{T>n\}} |X_n| \, d\mathbb{P} \\ &= \sum_{j=1}^n \int_{\{T=j\}} |X_n| \, d\mathbb{P} + \int_{\{T>n\}} |X_n| \, d\mathbb{P} \\ &= \mathbb{E}(|X_n|) = \|X_n\|_1 \leq H < +\infty. \end{aligned}$$

Però $X_{T \wedge n} \xrightarrow{n \rightarrow +\infty} X_T$, sicché il lemma di Fatou assicura che sia $\mathbb{E}(|X_T|) \leq H$, cioè che X_T sia in L^1 . \square

Tuttavia, anche quando $X_T \in L^1$, non è affatto detto che sia $\mathbb{E}(X_T) = \mathbb{E}(X_0)$.

Esempio 6.3.3. Si consideri un processo di Bernoulli. Si è visto nell'esempio 6.1.3 che (G_n) è una martingala rispetto alla filtrazione naturale. Fissato $j > 0$, si consideri il tempo di primo passaggio per la posizione $x = j$,

$$T_j := \min\{n \in \bar{\mathbb{N}} : G_n = j\};$$

il tempo d'arresto T_j è q.c. finito. Infatti

$$\{T_j = +\infty\} = \bigcap_{n=0}^{\infty} \{G_n < j\}$$

e si sa che

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|G_n| < j) = 0.$$

Ora $G_0 = 0$ e $G_{T_j} = j$ q.c., sicché G_{T_j} appartiene a L^1 e si ha banalmente $\mathbb{E}(G_{T_j}) = j \neq 0 = \mathbb{E}(G_0)$. \blacksquare

Il risultato che segue è fondamentale per i teoremi di convergenza.

Data una successione (X_n) di v.a. definite sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, si ponga

$$X_n^* := \max_{j \leq n} X_j.$$

Teorema 6.3.5. (Diseguaglianza massimale). *Se (X_n) è una sottomartingala, si ha, per ogni $\theta > 0$ e per ogni $n \in \mathbb{N}$,*

$$\theta \mathbb{P}(X_n^* \geq \theta) \leq \mathbb{E}(X_n \mathbf{1}_{\{X_n^* \geq \theta\}}) \leq \mathbb{E}(X_n^+). \quad (6.3.2)$$

Dimostrazione. Si ponga

$$S := \min\{j \leq n : X_j \geq \theta\} \mathbf{1}_{\{X_n^* \geq \theta\}} + n \mathbf{1}_{\{X_n^* < \theta\}}.$$

La v.a. S così definita è evidentemente un tempo d'arresto con $S \leq n$. Inoltre, $\{X_n^* \geq \theta\}$ appartiene a \mathcal{F}_S perché, per ogni $k \leq n$,

$$\{X_n^* \geq \theta\} \cap \{S \leq k\} = \{\max_{j \leq k} X_j \geq \theta\} = \{X_k^* \geq \theta\} \in \mathcal{F}_k.$$

Per il Teorema 6.3.3, con $T := n$, si ha $\mathbb{E}(X_n | \mathcal{F}_S) \geq X_S$. Poiché, nell'insieme $\{X_n^* \geq \theta\}$, è $X_S \geq \theta$, si ha

$$\begin{aligned} \theta \mathbb{P}(X_n^* \geq \theta) &= \mathbb{E}(\theta \mathbf{1}_{\{X_n^* \geq \theta\}}) \leq \int_{\{X_n^* \geq \theta\}} X_S \, d\mathbb{P} \\ &\leq \int_{\{X_n^* \geq \theta\}} X_n \, d\mathbb{P} \leq \mathbb{E}(X_n^+) , \end{aligned}$$

che conclude la dimostrazione. \square

Corollario 6.3.1. *Sia (X_n) una sotto-martingala positiva, $X_n \geq 0$ ($n \in \mathbb{Z}_+$); allora, per ogni $\theta > 0$ e per ogni $n \in \mathbb{Z}_+$, vale la disuguaglianza*

$$\mathbb{P}\left(\sup_{k \leq n} X_k \geq \theta\right) \leq \frac{1}{\theta} \mathbb{E}(X_n) = \frac{1}{\theta} \|X_n\|_1. \quad (6.3.3)$$

Il seguente risultato è notevole per l'importante corollario che permette di stabilire.

Lemma 6.3.1. *Siano X e Y dua v.a. positive tali che, per ogni $\theta > 0$, sia*

$$\theta \mathbb{P}(X \geq \theta) \leq \int_{\{X \geq \theta\}} Y \, d\mathbb{P}.$$

Allora, per $p > 1$ e per q tali che $(1/p) + (1/q) = 1$, si ha

$$\|X\|_p \leq q \|Y\|_p. \quad (6.3.4)$$

Dimostrazione. Com'è noto si ha

$$\mathbb{E}(X^p) = \int_0^{+\infty} p t^{p-1} \mathbb{P}(X \geq t) \, dt;$$

d'altro canto, si ha

$$\int_0^{+\infty} p t^{p-1} \mathbb{P}(X \geq t) \, dt \leq \int_0^{+\infty} p t^{p-2} \, dt \int_{\{X \geq t\}} Y \, d\mathbb{P}.$$

Ricorrendo al teorema di Fubini, si ha

$$\begin{aligned} \int_0^{+\infty} p t^{p-2} dt \int_{\{X \geq t\}} Y d\mathbb{P} &= \int_0^{+\infty} p t^{p-2} dt \int_{\Omega} \mathbf{1}_{\{X \geq t\}} Y d\mathbb{P} \\ &= \int_{\Omega} Y(\omega) d\mathbb{P}(\omega) \int_0^{X(\omega)} p t^{p-2} dt = \frac{p}{p-1} \int_{\Omega} Y X^{p-1} d\mathbb{P} \\ &= q \int_{\Omega} Y X^{p-1} d\mathbb{P} = q \mathbb{E}(X^{p-1}Y). \end{aligned}$$

Basta ora applicare la disuguaglianza di Hölder per avere

$$\mathbb{E}(X^p) \leq q \mathbb{E}(X^{p-1}Y) \leq q \|Y\|_p \|X^{p-1}\|_q. \quad (6.3.5)$$

Si supponga dapprima che X appartenga a L^p ; allora

$$\|X^{p-1}\|_q = \mathbb{E}^{1/q}(X^p),$$

sicché la (6.3.5) implica l'asserto. Se, invece, X non appartiene a L^p , vi appartiene senz'altro, per ogni $n \in \mathbb{N}$, $X \wedge n$, sicché si ha, per ogni $n \in \mathbb{N}$,

$$\|X \wedge n\|_p \leq q \|Y\|_p;$$

La (6.3.4) segue ora dal Teorema di convergenza monotona. \square

Corollario 6.3.2. *Siano $p > 1$ e q tali che $(1/p) + (1/q) = 1$. Sia (X_n) una sotto-martingala positiva e limitata uniformemente in L^p ,*

$$\sup_{n \in \mathbb{N}} \|X_n\|_p \leq H < +\infty.$$

Allora, se $X^* := \sup_{n \in \mathbb{N}} X_n$, X^* appartiene a L^p e vale la disuguaglianza

$$\|X^*\|_p \leq q \sup_{n \in \mathbb{N}} \|X_n\|_p.$$

Dimostrazione. Per ogni $n \in \mathbb{N}$ si definisca $X_n^* := \sup_{k \leq n} X_k$; segue allora dal Lemma 6.3.1 e dalla (6.3.3) che

$$\|X_n^*\|_p \leq q \|X_n\|_p \leq q \sup_{k \in \mathbb{N}} \|X_k\|_p.$$

Basta ora far tendere n a $+\infty$ ed applicare il teorema di convergenza monotona. \square

6.4 Integrabilità uniforme

Il concetto di *integrabilità uniforme*, o, come anche lo si chiama nella letteratura scientifica italiana, di *equi-integrabilità*, è fondamentale nella teoria delle martingale. Per questa ragione lo trattiamo qui benché sia un argomento appartenente piuttosto alla teoria dell'integrazione.

Definizione 6.4.1. Sia $\{X_\iota\}_{\iota \in I}$ un'arbitraria famiglia di v.a. reali definite sullo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$. La famiglia (X_ι) si dice *uniformemente integrabile* o *equi-integrabile* se

$$\lim_{c \rightarrow +\infty} \sup_{\iota \in I} \int_{\{|X_\iota| \geq c\}} |X_\iota| \, d\mathbb{P} = 0,$$

vale a dire se le speranze $\mathbb{E}(|X_\iota| \mathbf{1}_{\{|X_\iota| \geq c\}})$ tendono uniformemente a zero al tendere di c a $+\infty$. \diamond

È immediato controllare che, se $\{X_\iota\}$ è uniformemente integrabile, ogni v.a. X_ι è in L^1 . Se $|X_\iota| \leq Y$ per ogni $\iota \in I$ con $Y \in L^1$, allora $\{X_\iota\}$ è uniformemente integrabile; infatti

$$\int_{\{|X_\iota| \geq c\}} |X_\iota| \, d\mathbb{P} \leq \int_{\{|X_\iota| \geq c\}} Y \, d\mathbb{P}.$$

Ora

$$\mathbb{P}(|X_\iota| \geq c) \leq \frac{\mathbb{E}(|X_\iota|)}{c} \leq \frac{\mathbb{E}(Y)}{c},$$

che tende uniformemente a zero. In particolare, se le X_n sono uniformemente limitate, $\{X_n\}$ è uniformemente integrabile.

Lemma 6.4.1. *Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia X in L^1 e sia $(\mathcal{F}_\iota)_{\iota \in I}$ un'arbitraria famiglia di sottotribù di \mathcal{F} . Allora è uniformemente integrabile la famiglia $(X_\iota := \mathbb{E}_\iota(X) = \mathbb{E}(X | \mathcal{F}_\iota))_{\iota \in I}$.*

Dimostrazione. Per ogni $\iota \in I$ si ha

$$\mathbb{E}(|X_\iota|) = \mathbb{E}(|\mathbb{E}_\iota(X)|) = \mathbb{E}(|X|) < +\infty,$$

sicché (X_ι) è in L^1 . Inoltre

$$\int_{\{|X_\iota| \geq c\}} |X_\iota| \, d\mathbb{P} \leq \int_{\{|X_\iota| \geq c\}} \mathbb{E}_\iota(|X|) \, d\mathbb{P} = \int_{\{|X_\iota| \geq c\}} |X| \, d\mathbb{P}.$$

Poiché X_ι è integrabile, l'insieme $\{|X_\iota| \geq c\}$ tende ad un insieme di probabilità nulla al tendere di c a $+\infty$, sicché

$$\lim_{c \rightarrow +\infty} \sup_{\iota \in I} \int_{\{|X_\iota| \geq c\}} |X_\iota| \, d\mathbb{P} \leq \lim_{c \rightarrow +\infty} \int_{\{|X_\iota| \geq c\}} |X| \, d\mathbb{P} = 0,$$

da cui l'asserto. \square

Se la successione $(X_n)_{n \in \mathbb{Z}_+}$ è uniformemente integrabile, è anche uniformemente limitata in L^1 , cioè $\sup_{n \in \mathbb{Z}_+} \|X_n\|_1 < +\infty$; infatti, per ogni $\varepsilon > 0$ e per ogni n abbastanza grande,

$$\|X_n\|_1 = \mathbb{E}(|X_n|) = \int_{\{|X_n| < c\}} \dots + \int_{\{|X_n| \geq c\}} |X_n| \, d\mathbb{P} \leq c + \varepsilon.$$

Il viceversa non è necessariamente vero come dimostrato dal seguente

Esempio 6.4.1. Se $X_n := n \mathbf{1}_{[0, 1/n]}$, la successione (X_n) è uniformemente limitata in L^1 , ma non uniformemente integrabile rispetto alla restrizione λ della misura di Lebesgue ai boreliani di $[0, 1]$. Infatti, $\mathbb{E}(X_n) = 1$ per ogni n , ma, per $n > c$,

$$\int_{\{|X_n| \geq c\}} |X_n| \, d\lambda = n \lambda \left(\left[0, \frac{1}{n} \right] \right) = 1.$$

■

È utile il seguente criterio.

Teorema 6.4.1. Per una famiglia $(X_t)_{t \in D}$ di v.a. sono equivalenti le proprietà:

- (a) (X_t) è uniformemente integrabile;
 (b) (X_t) è uniformemente limitata in L^1 e, per ogni $\varepsilon > 0$, esiste $\delta = \delta(\varepsilon) > 0$ tale che, per ogni evento A con $\mathbb{P}(A) < \delta$, sia

$$\sup_{t \in D} \int_A |X_t| \, d\mathbb{P} < \varepsilon.$$

Dimostrazione. (a) \implies (b) Si è già visto che (X_t) è uniformemente limitata in L^1 .

$$\begin{aligned} \int_A |X_t| \, d\mathbb{P} &= \int_{A \cap \{|X_t| \geq c\}} \dots + \int_{A \cap \{|X_t| < c\}} |X_t| \, d\mathbb{P} \\ &\leq c \mathbb{P}(A) + \int_{A \cap \{|X_t| \geq c\}} |X_t| \, d\mathbb{P}. \end{aligned}$$

Per c sufficientemente grande l'ultimo integrale è minore di $\varepsilon/2$, sicché basta ora prendere $\delta = \varepsilon/(2c)$.

(b) \implies (a) Per la disegualianza di Markov si ha che, per ogni $c > 0$,

$$\mathbb{P}(|X_t| \geq c) \leq \frac{1}{c} \mathbb{E}(|X_t|) \leq \frac{H}{c}$$

ove $H := \sup_{t \in D} \mathbb{E}(|X_t|) = \sup_{t \in D} \|X_t\|_1$. Pertanto, se $c > H/\delta$, si ha

$$\sup_{t \in D} \int_{\{|X_t| \geq c\}} |X_t| \, d\mathbb{P} < \varepsilon$$

onde l'asserto. □

La proprietà espressa dalla (b) del Teorema 6.4.1 è detta di *continuità uniforme*.

Il concetto di integrabilità uniforme consente di chiarire del tutto i rapporti tra le convergenze in L^p e in probabilità per una successione di (X_n) di v.a. definite in uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$.

Teorema 6.4.2. Siano $(X_n)_{n \in \mathbb{N}}$ e X rispettivamente una successione e una v.a. di L^p con $p \in [1, +\infty[$. Sono equivalenti le proprietà:

(a) $X_n \xrightarrow[n \rightarrow +\infty]{} X$ in L^p ;

(b) $X_n \xrightarrow[n \rightarrow +\infty]{} X$ in probabilità e $(|X_n|^p)$ è uniformemente integrabile.

Dimostrazione. (a) \implies (b) Poiché si sa che la convergenza in L^p implica quella in probabilità, basta mostrare che $(|X_n|^p)$ è uniformemente integrabile. Poiché

$$\|X_n\|_p \leq \|X\|_p + \|X_n - X\|_p,$$

la successione (X_n) è uniformemente limitata in L^p o, equivalentemente, la successione $(|X_n|^p)$ è uniformemente limitata in L^1 . Fissato $\varepsilon > 0$, si prenda $\delta_1 = \delta_1(\varepsilon) > 0$ tale che $\mathbb{P}(A) < \delta_1$ implichi

$$\int_A |X|^p d\mathbb{P} < \frac{\varepsilon}{2^p};$$

tale δ_1 esiste perché $A \mapsto \mathbb{E}(|X|^p \mathbf{1}_A)$ è una misura finita su \mathcal{F} che è assolutamente continua rispetto a \mathbb{P} . Esiste inoltre $n_0 = n_0(\varepsilon) \in \mathbb{N}$ tale che sia

$$\|X_n - X\|_p^p < \frac{\varepsilon}{2^p},$$

per ogni $n \geq n_0$. Poiché vale la diseguaglianza

$$|X_n|^p \leq 2^{p-1} |X - X_n|^p + 2^{p-1} |X|^p, \quad (6.4.1)$$

si ha, se $n \geq n_0$,

$$\begin{aligned} \int_A |X_n|^p d\mathbb{P} &\leq 2^{p-1} \int_A |X_n - X|^p d\mathbb{P} + 2^{p-1} \int_A |X|^p d\mathbb{P} \\ &< 2^{p-1} \frac{\varepsilon}{2^p} + 2^{p-1} \frac{\varepsilon}{2^p} = \varepsilon. \end{aligned}$$

(b) \implies (a) Analogamente alla (6.4.1) si ha

$$|X - X_n|^p \leq 2^{p-1} |X_n|^p + 2^{p-1} |X|^p;$$

si può così concludere che, se $(|X_n|^p)$ è uniformemente integrabile se, e solo se, tale è anche la successione $(|X_n - X|^p)$. Fissato $\varepsilon \in]0, 1[$, si scelga $\delta = \delta(\varepsilon) > 0$ in modo che, se $\mathbb{P}(A) < \delta$, sia

$$\int_A |X_n - X|^p d\mathbb{P} < \varepsilon \quad (n \in \mathbb{N}).$$

Preso $n_0 \in \mathbb{N}$ in modo che $\mathbb{P}(|X_n - X| > \varepsilon) < \delta$, per ogni $n \geq n_0$, si ha

$$\begin{aligned} \mathbb{E}(|X_n - X|^p) &= \int_{\{|X_n - X| \leq \varepsilon\}} \dots + \int_{\{|X_n - X| > \varepsilon\}} |X_n - X|^p d\mathbb{P} \\ &\leq \varepsilon^p + \int_{\{|X_n - X| > \varepsilon\}} |X_n - X|^p d\mathbb{P} < \varepsilon^p + \varepsilon, \end{aligned}$$

che conclude la dimostrazione. \square

Si noti che il Teorema 2.2.9 che abbiamo dimostrato direttamente è ora un semplice corollario dell'ultimo risultato. Infatti, se $|X_n| \leq Y$ con $Y \in L^p$ per ogni $n \in \mathbb{N}$, la successione $(|X_n|^p)$ è uniformemente integrabile per il Lemma 6.4.1.

Definizione 6.4.2. Si dice che una martingala $(X_n, \mathcal{F}_n)_{n \in \mathbb{N}}$ ammette un *ultimo elemento* se esiste una v.a. $X_\infty \in L^1$ tale che, per ogni $n \in \mathbb{N}$, $\mathbb{E}_n(X_\infty) = X_n$. \diamond

Sia \mathcal{F}_∞ è la tribù generata dalla filtrazione $(\mathcal{F}_n)_{n \in \mathbb{N}}$, vale a dire

$$\mathcal{F}_\infty := \bigvee_{n \in \mathbb{N}} \mathcal{F}_n = \mathcal{F} \left(\bigcup_{n \in \mathbb{N}} \mathcal{F}_n \right);$$

si ponga $Y := \mathbb{E}(X_\infty | \mathcal{F}_\infty)$. Dunque, si ha anche $\mathbb{E}_n(Y) = X_n$ per ogni $n \in \mathbb{N}$, sicché non è restrittivo supporre, come faremo sempre nel seguito, che l'ultimo elemento X_∞ sia misurabile rispetto alla tribù \mathcal{F}_∞ . Si noti che, con tale convenzione, una martingala ha ultimo elemento se, e solo se, è ereditaria.

Per una martingala con un ultimo elemento si può usare la scrittura

$$(X_n, \mathcal{F}_n)_{n \in \bar{\mathbb{N}}}, \quad \text{oppure} \quad (X_n, \mathcal{F}_n)_{n \in \bar{\mathbb{Z}}_+},$$

ove $\bar{\mathbb{N}} := \mathbb{N} \cup \{+\infty\}$ e $\bar{\mathbb{Z}}_+ := \mathbb{Z}_+ \cup \{+\infty\}$ estendendo alla coppia di indici k e $+\infty$ la proprietà caratteristica delle martingale, $\mathbb{E}_k(X_n) = X_k$ se $k < n$.

La stessa definizione vale anche per le sotto- (o le super-)martingale; si dice che una sottomartingala ha un ultimo elemento X_∞ , e si scriverà $(X_n, \mathcal{F}_n)_{n \in \bar{\mathbb{N}}}$, se $\mathbb{E}_k(X_n) \geq X_k$, per ogni coppia di indici k e n , con $k < n \in \bar{\mathbb{N}}$. Anche in questo caso si può supporre che l'ultimo elemento X_∞ sia misurabile rispetto a \mathcal{F}_∞ .

Nella letteratura le martingale o le sottomartingale con un ultimo elemento si dicono anche *chiudibili (a destra)* o *chiuse (a destra)*.

Teorema 6.4.3. Siano $(X_n)_{n \in \mathbb{Z}_+}$ una martingala con ultimo elemento X_∞ e T un tempo d'arresto q.c. finito. Allora $\mathbb{E}(X_T) = \mathbb{E}(X_0)$.

Dimostrazione. Poiché T è q.c. finito, è

$$\mathbb{P} \left(\Omega \setminus \bigcup_{n \in \mathbb{Z}_+} \{T = n\} \right) = 0.$$

Allora

$$\begin{aligned} \mathbb{E}(X_T) &= \int X_T \, d\mathbb{P} = \sum_{n \in \mathbb{Z}_+} \int_{\{T=n\}} X_T \, d\mathbb{P} = \sum_{n \in \mathbb{Z}_+} \int_{\{T=n\}} X_n \, d\mathbb{P} \\ &= \sum_{n \in \mathbb{Z}_+} \int_{\{T=n\}} \mathbb{E}_n(X_\infty) \, d\mathbb{P} = \sum_{n \in \mathbb{Z}_+} \int_{\{T=n\}} X_\infty \, d\mathbb{P} \\ &= \int X_\infty \, d\mathbb{P} = \mathbb{E}(X_\infty). \end{aligned}$$

In particolare, per il tempo d'arresto $T = 0$ è $\mathbb{E}(X_0) = \mathbb{E}(X_\infty)$. \square

Teorema 6.4.4. *Per una martingala (X_n, \mathcal{F}_n) sono equivalenti le proprietà:*

- (a) (X_n, \mathcal{F}_n) è ereditaria;
- (b) (X_n) è uniformemente integrabile.

Dimostrazione. (a) \implies (b) Si supponga che (X_n, \mathcal{F}_n) sia ereditaria; esiste perciò una v.a. $X \in L^1(\mathcal{F})$ tale che, per ogni $n \in \mathbb{Z}_+$, sia $X_n = \mathbb{E}_t(X)$; il Lemma 6.4.1 dà ora l'asserto.

(b) \implies (a) Si supponga che (X_n, \mathcal{F}_n) sia uniformemente integrabile. Si introduca l'algebra $\mathcal{A} := \cup_{n \in \mathbb{Z}_+} \mathcal{F}_n$ e si definisca una funzione $\nu : \mathcal{A} \rightarrow \mathbb{R}$ mediante

$$\nu(A) := \int_A X_k \, d\mathbb{P} \quad \text{se } A \in \mathcal{F}_k.$$

Si osservi che, poiché (X_n, \mathcal{F}_n) è una martingala, per ogni $n \geq k$ si ha, se $A \in \mathcal{F}_k$,

$$\nu(A) = \int_A X_k \, d\mathbb{P} = \int_A X_n \, d\mathbb{P},$$

onde

$$\nu(A) = \lim_{n \rightarrow \infty} \int_A X_n \, d\mathbb{P};$$

ν è così ben definita ed è, per definizione, una misura finitamente additiva. Ma (X_n) è uniformemente integrabile, sicché ricorrendo ai Teoremi 6.4.1 e 1.5.1, si ha che ν è effettivamente una misura finita che è assolutamente continua rispetto a \mathbb{P} . Basta ora applicare il teorema di Radon–Nikodym e porre Y_∞ eguale alla densità $d\nu/d\mathbb{P}$ di ν rispetto a \mathbb{P} e $X_\infty := \mathbb{E}(Y_\infty | \mathcal{F}_\infty)$. Pertanto, per ogni $n \in \mathbb{Z}_+$ e per ogni $A \in \mathcal{F}_n$, è

$$\int_A X_n \, d\mathbb{P} = \nu(A) = \int_A X_\infty \, d\mathbb{P},$$

cioè $X_n = \mathbb{E}_n(X_\infty)$. □

6.5 Convergenza delle martingale

Considereremo dapprima la convergenza in L^p delle martingale. Questo studio, anche alla luce delle applicazioni che ce ne ripromettiamo, si può condurre per martingale $\{X_t : t \in D\}$ nelle quali l'insieme D degli indici è un insieme diretto.

Il seguente risultato di natura tecnica sarà utile nel seguito.

Lemma 6.5.1. *Per ogni $p \in [1, +\infty[$ il sottoinsieme V_p di L^p definito da*

$$V_p := \bigcup_{t \in D} L^p(\mathcal{F}_t)$$

è denso in $L^p(\mathcal{F}_\infty)$, ove

$$\mathcal{F}_\infty := \bigvee_{t \in D} \mathcal{F}_t$$

è la tribú generata dall'algebra $\mathcal{A} := \cup_{t \in D} \mathcal{F}_t$.

Dimostrazione. Basta mostrare che possono essere approssimate mediante elementi di V_p le funzioni indicatrici degli insiemi A di \mathcal{F}_∞ . Sia \mathcal{C}_p la classe di sottoinsiemi misurabili di \mathcal{F}_∞ le cui funzioni indicatrici possono essere approssimate mediante funzioni di V_p ,

$$\mathcal{C}_p := \{A \in \mathcal{F}_\infty : \forall \varepsilon > 0 \exists Y \in V_p \quad \|\mathbf{1}_A - Y\|_p < \varepsilon\}.$$

è immediato riconoscere che \mathcal{C}_p include l'algebra \mathcal{A} (gli insiemi di tale algebra hanno indicatrici che sono in V_p).

Inoltre, \mathcal{C}_p è una classe monotona. Si supponga che sia $A_n \in \mathcal{C}_p$ per ogni $n \in \mathbb{N}$ e $A_n \uparrow A$. Dato $\varepsilon > 0$, per ogni $n \in \mathbb{N}$, esiste $Y_n \in V_p$ tale che $\|\mathbf{1}_{A_n} - Y_n\|_1 < \varepsilon/2$. Si può usare il teorema di convergenza dominata per trovare un indice $n_0 = n_0(\varepsilon) \in \mathbb{N}$ tale che $\|\mathbf{1}_{A_n} - \mathbf{1}_A\|_1 < \varepsilon/2$ per ogni $n \geq n_0$. Dunque, se $n \geq n_0$, è

$$\|\mathbf{1}_A - Y_n\|_1 \leq \|\mathbf{1}_A - \mathbf{1}_{A_n}\|_1 + \|\mathbf{1}_{A_n} - Y_n\|_1 < \varepsilon.$$

Perciò A appartiene a \mathcal{C}_p . Analogamente si procede se $A_n \downarrow A$.

\mathcal{C}_p è così una classe monotona che contiene l'algebra \mathcal{A} ; per il teorema della classe monotona, si ha $\mathcal{C}_p = \mathcal{F}_\infty$. \square

Teorema 6.5.1. *Sia X una v.a. di L^p con $p \in [1, +\infty[$ e si consideri la martingala ereditaria $X_t := \mathbb{E}_t(X)$ ($t \in D$). Allora $\{X_t\}$ converge in L^p alla v.a.*

$$X_\infty := \mathbb{E}_\infty(X) = \mathbb{E}(X \mid \mathcal{F}_\infty),$$

ove \mathcal{F}_∞ è la tribù generata dall'algebra $\cup_{t \in D} \mathcal{F}_t$,

$$\mathcal{F}_\infty := \bigvee_{t \in D} \mathcal{F}_t.$$

Dimostrazione. Si supponga che esista un indice $s \in D$ per il quale Y appartenga a $L^p(\mathcal{F}_s)$; si ha allora, per ogni $t \geq s$, $\mathbb{E}_t(Y) = Y$. Perciò,

$$\forall Y \in \bigcup_{t \in D} L^p(\mathcal{F}_t) \quad \mathbb{E}_t(Y) \xrightarrow[t \rightarrow +\infty]{L^p} Y = \mathbb{E}_\infty(Y).$$

Sia ora X in $L^p(\mathcal{F}_\infty)$; in virtù del lemma precedente, per ogni $\varepsilon > 0$, si può prendere $Y_\varepsilon \in \cup_{t \in D} L^p(\mathcal{F}_t)$ con $\|X - Y_\varepsilon\|_p < \varepsilon$. Poiché \mathbb{E}_t è una contrazione su $L^p(\mathcal{F})$,

$$\begin{aligned} \|\mathbb{E}_t(X) - X\|_p &\leq \|\mathbb{E}_t(X) - \mathbb{E}_t(Y_\varepsilon)\|_p + \|\mathbb{E}_t(Y_\varepsilon) - Y_\varepsilon\|_p + \|Y_\varepsilon - X\|_p \\ &\leq 2\|X - Y_\varepsilon\|_p + \|\mathbb{E}_t(Y_\varepsilon) - Y_\varepsilon\|_p \\ &< 2\varepsilon + \|\mathbb{E}_t(Y_\varepsilon) - Y_\varepsilon\|_p. \end{aligned}$$

Basta, dunque, applicare la prima parte della dimostrazione per avere, per ogni X in $L^p(\mathcal{F}_\infty)$,

$$\mathbb{E}_t(X) \xrightarrow[t \rightarrow +\infty]{L^p} X = \mathbb{E}_\infty(X).$$

Infine, per ogni $X \in L^p(\mathcal{F})$, si ha

$$X_t = \mathbb{E}_t(X) = \mathbb{E}_t \mathbb{E}_\infty(X) \xrightarrow[t \rightarrow +\infty]{L^p} \mathbb{E}_\infty(X),$$

cioè l'asserto. \square

Nel trattare di convergenza in L^p di martingale occorre distinguere i due casi $p = 1$ e $p \in]1, +\infty[$. Tratteremo prima il caso $p > 1$.

Teorema 6.5.2. *Sia $(X_t, \mathcal{F}_t)_{t \in D}$ una martingala con $X_t \in L^p(\mathcal{F}_t)$ ($p > 1$) per ogni $t \in D$. Sono allora equivalenti le condizioni:*

- (a) *esiste una v.a. $X \in L^p$ tale che $X_t \rightarrow X$ in L^p ;*
- (b) *esiste una v.a. $X \in L^p(\mathcal{F}_\infty)$ tale che $X_t = \mathbb{E}_t(X)$ per ogni $t \in D$;*
- (c) $\sup_{t \in D} \|X_t\|_p < +\infty$.

Dimostrazione. (a) \implies (b) Poiché X è misurabile rispetto alla tribú \mathcal{F}_∞ , si ha $X \in L^p(\mathcal{F}_\infty)$. Se $A \in \mathcal{F}_s$, si ha, per ogni $n > k$,

$$\int_A X_s \, d\mathbb{P} = \int_A X_t \, d\mathbb{P}$$

e, di qui,

$$\int_A X_s \, d\mathbb{P} = \lim_{t \rightarrow +\infty} \int_A X_t \, d\mathbb{P} = \int_A X \, d\mathbb{P}$$

vale a dire

$$\forall n \in D \quad X_t = \mathbb{E}_t(X).$$

L'implicazione (b) \implies (a) è l'oggetto del teorema precedente.

(a) \implies (c) Il Teorema 6.4.2 implica che $(|X_t|^p)$ è uniformemente integrabile e il Teorema 6.4.1 che

$$\sup_{n \in D} \|X_t\|_p < +\infty.$$

(c) \implies (b) La martingala (X_t, \mathcal{F}_t) è uniformemente integrabile. Infatti, fissato $\varepsilon > 0$, si ponga

$$H := \sup_{t \in D} \|X_t\|_p, \quad \text{e} \quad \alpha := \frac{H^p}{\varepsilon}.$$

Poiché

$$\lim_{x \rightarrow +\infty} x^{p-1} = +\infty,$$

esiste $c > 0$ tale che $x^{p-1} \geq \alpha$ se $x \geq c$. Perciò,

$$\int_{\{|X_t| \geq c\}} |X_t| \, d\mathbb{P} \leq \frac{1}{\alpha} \int_{\{|X_t| \geq c\}} |X_t|^p \, d\mathbb{P} \leq \frac{H^p}{\alpha} = \varepsilon,$$

onde l'asserto. Ora il Teorema 6.4.4 assicura che esiste una v.a. X tale che $X_t = \mathbb{E}_t(X)$ per ogni $t \in D$. \square

La condizione (c) dell'ultimo teorema è necessaria e sufficiente per la convergenza in L^p con $p \in]1, +\infty[$. Tuttavia, l'analoga condizione per L^1 ,

$$\sup_{t \in D} \|X_t\|_1 < +\infty,$$

non basta ad assicurare la convergenza in L^1 , come mostra il seguente

Esempio 6.5.1. Sia (Y_n) una successione di v.a. indipendenti, isonome e positive di L^1 con

$$\mathbb{E}(Y_n) = 1 \quad \text{e} \quad \mathbb{P}(Y_n = 1) < 1$$

per ogni $n \in \mathbb{N}$. Posto $X_n := \prod_{j=1}^n Y_j$, (X_n) è una martingala moltiplicativa positiva (si ricordi l'Esempio 6.1.4) con $\mathbb{E}(X_n) = 1$. D'altro canto

$$\begin{aligned} \|X_{n+1} - X_n\|_1 &= \mathbb{E}(|X_{n+1} - X_n|) \\ &= \mathbb{E} \left[\left(\prod_{j=1}^n Y_j \right) |Y_{n+1} - 1| \right] \\ &= \mathbb{E}(|Y_{n+1} - 1|) = \mathbb{E}(|Y_1 - 1|) \neq 0, \end{aligned}$$

sicché (X_n) non converge in L^1 . Si vedrà nel seguito, Teorema 6.5.5, che X_n converge 0 q.c.. ■

Per la convergenza delle martingale in L^1 vale il seguente risultato, la cui dimostrazione è identica a quella del Teorema 6.5.2.

Teorema 6.5.3. *Per una martingala $(X_t, \mathcal{F}_t)_{t \in D}$ con $X_t \in L^1(\mathcal{F}_t)$ ($p > 1$) per ogni $t \in D$ sono equivalenti le condizioni:*

- (a) *esiste una v.a. $X \in L^1$ tale che $X_t \rightarrow X$ in L^1 ;*
- (b) *esiste una v.a. $X \in L^1(\mathcal{F}_\infty)$ tale che $X_t = \mathbb{E}_t(X)$ per ogni $t \in D$;*
- (c') *(X_t, \mathcal{F}_t) è uniformemente integrabile.*

Rimandiamo al seguito lo studio della convergenza in L^p con $p \in [1, +\infty[$ delle sottomartingale.

Affrontiamo ora lo studio della convergenza q.c. delle martingale. Tale studio si conduce con maggior semplicità nel caso delle martingale per le quali l'insieme degli indici è \mathbb{Z}_+ o \mathbb{N} .

Si è visto nel Capitolo 2 che, in generale, non esistono rapporti tra la convergenza quasi certa e quella in L^1 di una successione di v.a.. Se però la successione in questione forma una martingala la situazione è differente, come evidenziato dal prossimo risultato.

Teorema 6.5.4. *Ogni martingala (X_n, \mathcal{F}_n) che converge in L^1 converge anche quasi certamente.*

Dimostrazione. Si supponga che $X_n \xrightarrow{n \rightarrow +\infty} X$ in L^1 . È facile estrarre da \mathbb{N} una successione di naturali $(n(k))_{k \in \mathbb{N}}$, con $n(k) < n(k+1)$ per ogni $k \in \mathbb{N}$, in modo che sia convergente la serie

$$\sum_{k \in \mathbb{N}} k \|X_{n(k)} - X\|_1;$$

basta prendere, come $n(k)$, il più piccolo naturale tale che $\|X_{n(k)} - X\|_1 \leq 1/k^3$.

Per ogni $k \in \mathbb{N}$, $(X_n - X_{n(k)})_{n \geq n(k)}$ è una martingala alla quale si può applicare la diseguaglianza massimale (6.3.2), ottenendo, per ogni $s \in \mathbb{N}$ sufficientemente grande,

$$\mathbb{P} \left(\sup_{s \geq n \geq n(k)} |X_n - X_{n(k)}| > \frac{1}{k} \right) \leq k \|X_s - X_{n(k)}\|_1.$$

Facendo tendere s a $+\infty$, si ha

$$\mathbb{P} \left(\sup_{n \geq n(k)} |X_n - X_{n(k)}| > \frac{1}{k} \right) \leq k \|X - X_{n(k)}\|_1.$$

Allora, il primo lemma di Borel–Cantelli dà

$$\mathbb{P} \left(\limsup_{k \rightarrow +\infty} \left\{ \sup_{n \geq n(k)} |X_n - X_{n(k)}| > \frac{1}{k} \right\} \right) = 0,$$

vale a dire

$$\mathbb{P} \left(\liminf_{k \rightarrow +\infty} \left\{ \sup_{n \geq n(k)} |X_n - X_{n(k)}| \leq \frac{1}{k} \right\} \right) = 1,$$

sicché

$$\lim_{k \rightarrow +\infty} \sup_{n \geq n(k)} |X_n - X_{n(k)}| = 0 \quad \text{q.c.}$$

In altre parole, esiste un insieme $N \in \mathcal{F}$ con $\mathbb{P}(N) = 0$ tale che $(X_n(\omega))$ sia una successione di Cauchy per ogni $\omega \in N^c$. \square

Il seguente corollario è spesso noto come teorema di Lévy.

Corollario 6.5.1. *Ogni martingala uniformemente integrabile (X_n) converge q.c. e in L^1 a una v.a. $X_\infty \in L^1$ tale che $X_n = \mathbb{E}_n(X_\infty)$ per ogni $n \in \mathbb{N}$. X_∞ è l'ultimo elemento della martingala. In particolare, per una v.a. $X \in L^1$ è*

$$\lim_{n \rightarrow +\infty} \mathbb{E}_n(X) = \mathbb{E}_\infty(X).$$

Dimostrazione. Segue dal Teorema 6.4.4 che esiste una v.a. $X_\infty \in L^1$, che si può pensare misurabile rispetto a \mathcal{F}_∞ , tale che $X_n = \mathbb{E}_n(X_\infty)$ per ogni $n \in \mathbb{N}$. La convergenza in L^1 segue dal Teorema 6.5.3, mentre la convergenza q.c. è ora conseguenza immediata del teorema precedente. \square

Il risultato piú importante riguardante la convergenza q.c. delle martingale è dato dal seguente

Teorema 6.5.5. (Doob). *Se $(X_n, \mathcal{F}_n)_{n \in \mathbb{Z}_+}$ è una martingala limitata in L^1 , vale a dire tale che*

$$\sup_{n \in \mathbb{Z}_+} \|X_n\|_1 = \sup_{n \in \mathbb{Z}_+} \mathbb{E}(|X_n|) < +\infty,$$

allora (X_n) converge q.c..

Dimostrazione. Sia $H > 0$ e si ponga

$$T := \min\{n \in \overline{\mathbb{Z}}_+ : |X_n| \geq H\}.$$

Si è visto nell'esempio 6.2.1 che la v.a. T così definita è un tempo d'arresto. Per il Teorema 6.3.2, X^T è una martingala. Vogliamo provare che è uniformemente integrabile. Si consideri la v.a.

$$Y := |X_T| \mathbf{1}_{\{T < +\infty\}} + H \mathbf{1}_{\{T = +\infty\}}.$$

Ovviamente, $|X_n^T| = |X_{T \wedge n}| \leq Y$ per ogni $n \in \mathbb{Z}_+$, sicché basta dimostrare che Y appartiene a L^1 , vale a dire che $\mathbb{E}(Y) < +\infty$. Ora,

$$\int_{\{T=+\infty\}} Y \, d\mathbb{P} \leq H,$$

mentre, ricorrendo lemma di Fatou, si ha

$$\begin{aligned} \int_{\{T<+\infty\}} Y \, d\mathbb{P} &= \int_{\{T<+\infty\}} |X_T| \, d\mathbb{P} = \int_{\{T<+\infty\}} \lim_{n \rightarrow +\infty} |X_{T \wedge n}| \, d\mathbb{P} \\ &\leq \liminf_{n \rightarrow +\infty} \int_{\{T<+\infty\}} |X_{T \wedge n}| \, d\mathbb{P} \leq \sup_{n \in \mathbb{Z}_+} \|X_{T \wedge n}\|_1 \\ &\leq \sup_{n \in \mathbb{Z}_+} \|X_n\|_1 < +\infty. \end{aligned}$$

Dunque, Y appartiene a L^1 e X^T è uniformemente integrabile.

Nell'insieme $\{\sup_{n \in \mathbb{Z}_+} |X_n| < H\}$ si ha $X_n^T = X_{T \wedge n} = X_n$ e, quindi, (X_n) converge q.c. in virtù del Corollario 6.5.1.

La martingala (X_n) , per quanto appena visto, converge nell'insieme

$$B := \bigcup_{\substack{q \in \mathbb{Q} \\ q > 0}} \left\{ \sup_{n \in \mathbb{Z}_+} |X_n| < q \right\}.$$

Consideriamo il complementare di quest'ultimo insieme

$$B^c = \bigcap_{\substack{q \in \mathbb{Q} \\ q > 0}} \left\{ \sup_{n \in \mathbb{Z}_+} |X_n| \geq q \right\}.$$

Poiché

$$\left\{ \sup_{n \in \mathbb{Z}_+} |X_n| \geq q \right\} = \bigcup_{n \in \mathbb{Z}_+} \left\{ \max_{j \leq n} |X_j| \geq q \right\},$$

e poiché la successione formata dagli insiemi $\{\max_{j \leq n} |X_j| \geq q\}$ è crescente, dalla diseguaglianza massimale (6.3.2) applicata alla sottomartingala $(|X_n|)$ e dal teorema di Beppo Levi, si ha, per ogni razionale $q > 0$,

$$\begin{aligned} \mathbb{P}(B^c) &\leq \mathbb{P}\left(\sup_{n \in \mathbb{Z}_+} |X_n| \geq q\right) \leq \lim_{n \rightarrow +\infty} \mathbb{P}\left(\max_{j \leq n} |X_j| \geq q\right) \\ &\leq \lim_{n \rightarrow +\infty} \frac{1}{q} \mathbb{E}(|X_n|) = \frac{1}{q} \sup_{n \in \mathbb{Z}_+} \mathbb{E}(|X_n|), \end{aligned}$$

che tende a zero al tendere di q a $+\infty$; perciò, $\mathbb{P}(B^c) = 0$, e quindi $\mathbb{P}(B) = 1$, vale a dire che la martingala (X_n) converge q.c. \square

Si noti che, se (X_n, \mathcal{F}_n) è una martingala, $(|X_n|, \mathcal{F}_n)$ è una sottomartingala e, quindi, per il teorema (6.1.7), la successione $\{\mathbb{E}(|X_n|)\}$ è crescente, sicché la condizione del teorema precedente può essere posta nella forma equivalente

$$\lim_{n \rightarrow +\infty} \mathbb{E}(|X_n|) = \lim_{n \rightarrow +\infty} \|X_n\|_1 = \sup_{n \in \mathbb{Z}_+} \|X_n\|_1 < +\infty.$$

La condizione $\sup_{n \in \mathbb{Z}_+} \|X_n\|_1 < +\infty$ assicura che la martingala (X_n) converga q.c. ma non necessariamente in L^1 (si rammenti l'esempio 6.5.1); è inoltre una condizione sufficiente, ma non necessaria per la convergenza q.c. di una martingala, come si vedrà negli esercizi.

Teorema 6.5.6. *Una martingala $(X_n)_{n \in \mathbb{Z}_+}$ tale che i suoi incrementi siano maggiorati da una stessa variabile Z di L^1 converge quasi certamente nell'insieme $\{\sup_n X_n < +\infty\}$.*

Dimostrazione. Senza perdita di generalità si può supporre che sia $X_0 = 0$. Basta provare che, per ogni $a > 0$, la martingala (X_n) converge quasi certamente nell'insieme $\{\sup_n X_n \leq a\}$. Introdotto il tempo d'arresto

$$T := \inf\{n \in \overline{\mathbb{Z}}_+ : X_n > a\},$$

si ha

$$\left\{ \sup_n X_n \leq a \right\} = \{T = +\infty\}.$$

Poiché $T \geq 1$, è, per ogni $n \in \mathbb{N}$,

$$X_{n \wedge T} = (X_{n \wedge T} - X_{(n \wedge T) - 1}) - X_{(n \wedge T) - 1} \leq Z + a.$$

La martingala arrestata X^T è, dunque, limitata in L^1 e, per il Teorema 6.5.5, quasi certamente convergente. Poiché nell'insieme $\{T = +\infty\}$ la martingala X^T è eguale a (X_n) , anche quest'ultima converge quasi certamente. \square

Come conseguenza del lemma di Wald, Lemma 6.2.1, si ha

Corollario 6.5.2. *Siano $(X_n)_{n \in \mathbb{Z}_+}$ una martingala con gli incrementi uniformemente maggiorati in modulo, per ogni $n \in \mathbb{N}$, $|X_n - X_{n-1}| \leq H$, α una costante tale che $\mathbb{E}(X_0) < \alpha$ e T il tempo d'arresto definito da*

$$T(\omega) := \inf\{n \in \mathbb{N} : X_n(\omega) \geq \alpha\}.$$

Allora T non è in L^1 .

Dimostrazione. Si ragioni per assurdo e si supponga che T sia integrabile. In virtù del Lemma 6.2.1, la martingala arrestata X^T è allora dominata in L^1 , e pertanto, grazie al Teorema 6.5.5, convergente q.c. D'altronde da $\mathbb{E}(T) < +\infty$ scende $\mathbb{P}(T < +\infty) = 1$; inoltre vale l'inclusione

$$\{T < +\infty\} \subseteq \{X_T \geq \alpha\},$$

sicché nei punti di $\{T < +\infty\}$ la successione $(X_{n \wedge T})_{n \in \mathbb{Z}_+}$ coincide definitivamente con X_T . La martingala X^T è allora chiusa da X_T e si ha

$$\mathbb{E}(X_0) = \mathbb{E}(X_T) \geq \alpha,$$

una contraddizione. \square

Per estendere il teorema di convergenza alle sotto- e alle super-martingale ricorriamo alla decomposizione di Doob (Teorema 6.1.5).

Lemma 6.5.2. *Se (X_n, \mathcal{F}_n) è una sottomartingala limitata in L^1 , tale cioè che sia*

$$\sup_{n \in \mathbb{Z}_+} \|X_n\|_1 < +\infty,$$

nella decomposizione di Doob $X_n = Y_n + A_n$ ($n \in \mathbb{Z}_+$), il processo crescente (A_n) è uniformemente integrabile e (Y_n) e (A_n) sono limitati in L^1 . Se (X_n) è uniformemente integrabile, tali sono anche la martingala (Y_n) e il processo crescente (A_n) .

Dimostrazione. Basta dimostrare che (A_n) è uniformemente integrabile, le rimanenti affermazioni essendo di verifica banale.

Si ha $A_1 = 0 \leq A_2 \leq A_3 \leq \dots$ e

$$\mathbb{E}(A_n) = \mathbb{E}(X_n) - \mathbb{E}(Y_n) = \mathbb{E}(X_n) - \mathbb{E}(Y_0).$$

La successione (A_n) è crescente e quindi converge q.c. alla v.a.

$$Z = \sup_{n \in \mathbb{Z}_+} A_n \geq 0.$$

Il teorema di convergenza monotona dà

$$\mathbb{E}(Z) = \mathbb{E} \left(\sup_{n \in \mathbb{Z}_+} A_n \right) = \sup_{n \in \mathbb{Z}_+} \mathbb{E}(A_n) \leq \sup_{n \in \mathbb{Z}_+} \|X_n\|_1 + \|Y_0\|_1 < +\infty.$$

Si ha dunque $0 \leq A_n \leq Z \in L^1$ per ogni $n \in \mathbb{Z}_+$, sicché (A_n) è uniformemente integrabile. \square

Teorema 6.5.7. (a) *Una sotto- o super-martingala (X_n, \mathcal{F}_n) che sia limitata in L^1 ,*

$$\sup_{n \in \mathbb{Z}_+} \|X_n\|_1 < +\infty,$$

converge q.c. a una v.a. X_∞ .

(b) *Una sotto- o super-martingala (X_n, \mathcal{F}_n) che sia uniformemente integrabile converge tanto q.c. quanto in L^1 a una v.a. X_∞ .*

In entrambi i casi, se \mathcal{F}_∞ è la tribù generata dall'algebra $\mathcal{A} = \cup_{n \in \mathbb{Z}_+} \mathcal{F}_n$, allora $(X_n, \mathcal{F}_n)_{n \in \mathbb{Z}_+}$ è una sotto- (rispettivamente, super-) martingala chiusa a destra.

Dimostrazione. Considereremo solo il caso di una sotto-martingala.

(a) Il processo crescente (A_n) che compare nella decomposizione di Doob converge q.c. alla v.a. $Z \in L^1$ come nella dimostrazione del lemma precedente. Inoltre, (Y_n, \mathcal{F}_n) è una martingala limitata in L^1 che converge q.c., in virtù del Teorema 6.5.5; perciò, $X_n = Y_n + A_n$ converge q.c..

(b) Se (X_n) è uniformemente integrabile, tale è anche (Y_n) , per il lemma precedente; quindi (Y_n) , per il Corollario 6.5.1, converge q.c. e in L^1 . Quanto al processo crescente (Z_n) , esso converge a Z q.c. come nel lemma precedente e in L^1 per il teorema di convergenza monotona.

Per stabilire l'ultima affermazione, si prenda $n < k$; per ogni $A \in \mathcal{F}_n$, si ha

$$\int_A X_n \, d\mathbb{P} \leq \int_A X_k \, d\mathbb{P}.$$

Facendo tendere k a $+\infty$, la convergenza in L^1 dà

$$\int_A X_n \, d\mathbb{P} \leq \int_A X_\infty \, d\mathbb{P},$$

onde l'asserto. □

Diamo qui di seguito una condizione sufficiente affinché una sottomartingala sia uniformemente integrabile.

Teorema 6.5.8. *È uniformemente integrabile ogni sottomartingala positiva con ultimo elemento $(X_n, \mathcal{F}_n)_{n \in \mathbb{N}}$.*

Dimostrazione. Per ogni $n \in \mathbb{N}$, si ha

$$\int_{\{|X_n| \geq c\}} |X_n| \, d\mathbb{P} = \int_{\{X_n \geq c\}} X_n \, d\mathbb{P} \leq \int_{\{X_n \geq c\}} X_\infty \, d\mathbb{P};$$

la disuguaglianza di Markov dà ora

$$\mathbb{P}(X_n \geq c) \leq \frac{\mathbb{E}(X_n)}{c} \leq \frac{\mathbb{E}(X_\infty)}{c} \xrightarrow{c \rightarrow +\infty} 0,$$

uniformemente in n . □

Ritorniamo ora alla convergenza in L^p delle sottomartingale.

Teorema 6.5.9. *Sia (X_n, \mathcal{F}_n) una martingala, oppure una sottomartingala positiva, limitata in L^p con $p \in]1, +\infty[$,*

$$\sup_{n \in \mathbb{N}} \|X_n\|_p \leq H < +\infty;$$

allora (X_n) è uniformemente integrabile e pertanto converge ad una v.a. X_∞ di L^p sia q.c. sia in L^p .

Dimostrazione. Come nella dimostrazione dell'implicazione (c) \implies (b) del Teorema 6.5.2 si mostra che (X_n) è uniformemente integrabile. Perciò (X_n) converge q.c. a una v.a. X_∞ per il Teorema 6.5.5, se (X_n, \mathcal{F}_n) è una martingala, per il Teorema 6.5.7 se è una sottomartingala positiva.

In entrambi i casi, $(|X_n|^p, \mathcal{F}_n)_{n \in \mathbb{N}}$ è una sottomartingala positiva e, quindi, uniformemente integrabile. Basta ora applicare il Teorema 6.4.2 per avere la convergenza in L^p . □

6.6 Le martingale rovesciate

Definizione 6.6.1. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia $(\mathcal{F}_n)_{n \in \mathbb{Z}_+}$ una famiglia di sottotribú di \mathcal{F} tale che

$$\mathcal{F}_1 \supset \mathcal{F}_2 \supset \cdots \supset \mathcal{F}_n \supset \cdots$$

La successione (X_n) di variabili aleatorie adattate a (\mathcal{F}_n) , cioè tali che X_n sia misurabile rispetto a \mathcal{F}_n per ogni $n \in \mathbb{Z}_+$, si dice essere una *martingala rovesciata* se, per ogni $n \in \mathbb{N}$, si ha

$$\mathbb{E}(X_n | \mathcal{F}_{n+1}) = X_{n+1}.$$

Se invece si ha

$$\mathbb{E}(X_n | \mathcal{F}_{n+1}) \geq X_{n+1},$$

si parlerà di una *sotto-martingala rovesciata*. \diamond

Aniché parlare di una (sotto-)martingala rovesciata, si potrebbe parlare di una (sotto-)martingala rispetto alla filtrazione $(\mathcal{F}_{-n})_{n \in \mathbb{Z}_+}$. Nel seguito si porrà

$$\mathcal{F}_0 := \bigcap_{n \in \mathbb{N}} \mathcal{F}_n.$$

Il seguente è un esempio notevole di martingala rovesciata.

Esempio 6.6.1. Si consideri una successione $(X_n)_{n \in \mathbb{N}}$ di v.a. indipendenti, isonome e di L^1 sullo spazio $(\Omega, \mathcal{F}, \mathbb{P})$, si ponga $S_n := \sum_{j=1}^n X_j$ e $Z_{-n} := S_n/n$ e si introducano le tribú

$$\mathcal{F}_{-n} := \mathcal{F}(S_n, S_{n+1}, S_{n+2}, \dots) = \mathcal{F}(S_n, X_{n+1}, X_{n+2}, \dots).$$

La successione $(\mathcal{F}_{-n})_{n \in \mathbb{Z}_+}$ è decrescente e $(Z_{-n}, \mathcal{F}_{-n})$ è una martingala rovesciata. Si osservi, infatti, che, se $j, k \leq n+1$, allora

$$\mathbb{E}(X_j | \mathcal{F}_{-n-1}) = \mathbb{E}(X_k | \mathcal{F}_{-n-1}).$$

Perciò

$$\begin{aligned} \mathbb{E}(X_{n+1} | \mathcal{F}_{-n-1}) &= \frac{1}{n+1} \sum_{k=1}^{n+1} \mathbb{E}(X_k | \mathcal{F}_{-n-1}) \\ &= \frac{1}{n+1} \mathbb{E}(S_{n+1} | \mathcal{F}_{-n-1}) = \frac{S_{n+1}}{n+1}. \end{aligned}$$

Pertanto

$$\begin{aligned} \mathbb{E}(Z_{-n} | \mathcal{F}_{-n-1}) &= \mathbb{E}\left(\frac{S_{n+1}}{n} | \mathcal{F}_{-n-1}\right) - \mathbb{E}\left(\frac{X_{n+1}}{n} | \mathcal{F}_{-n-1}\right) \\ &= \frac{S_{n+1}}{n} - \frac{S_{n+1}}{n(n+1)} = \frac{S_{n+1}}{n+1} = Z_{-n-1}. \end{aligned}$$

Ciò mostra che $(Z_{-n}, \mathcal{F}_{-n})$ è una martingala rovesciata. \blacksquare

Lemma 6.6.1. *Ogni martingala rovesciata è uniformemente integrabile.*

Dimostrazione. Segue dalla definizione che, per ogni $n \in \mathbb{Z}_+$, si ha $X_n = \mathbb{E}(X_0 | \mathcal{F}_n)$; l'asserto segue ora dal Lemma 6.4.1. \square

Si può ora studiare la convergenza delle martingale rovesciate.

Teorema 6.6.1. *Per ogni martingala rovesciata (X_n, \mathcal{F}_n) vi è una variabile aleatoria X_∞ di L^1 tale che il limite*

$$\lim_{n \rightarrow +\infty} X_n = X_\infty$$

esista q.c. e in L^1 ; inoltre $\mathbb{E}(X_1 | X_\infty) = X_\infty$ e per ogni $n \in \mathbb{N}$, $\mathbb{E}(X_n) = \mathbb{E}(X_\infty)$.

Dimostrazione. In virtù del Lemma 6.6.1 e del Teorema 6.5.3 la martingala data converge in L^1 ad una variabile aleatoria X_∞ , e, poiché la convergenza in L^1 di una martingala implica la convergenza quasi certa allo stesso limite, Teorema 6.5.4, vi converge anche quasi certamente. Per il Corollario 6.5.1, X_∞ è l'ultimo elemento sicché vale $\mathbb{E}(X_1 | X_\infty) = X_\infty$. \square

Teorema 6.6.2. *Per ogni v.a. $X \in L^1$ si ha*

$$\lim_{n \rightarrow +\infty} \mathbb{E}_n(X) = \mathbb{E}_0(X). \quad (6.6.1)$$

Dimostrazione. Si ponga $Z_{-n} := \mathbb{E}_n(X)$; (Z_{-n} è una martingala rovesciata e per il Teorema 6.6.1 ($\mathbb{E}_n(X)$) converge a una v.a. Z di L^1 . Poiché il limite di $\mathbb{E}_n(X)$ per $n \geq k$ è misurabile rispetto a \mathcal{F}_k per ogni $k \in \mathbb{N}$, Z è misurabile rispetto a \mathcal{F}_0 . L'integrabilità uniforme di $\mathbb{E}_n(X)$ dà, per ogni insieme

$$\begin{aligned} \int_A Z \, d\mathbb{P} &= \lim_{n \rightarrow +\infty} \int_A \mathbb{E}_n(X) \, d\mathbb{P} = \lim_{n \rightarrow +\infty} \int_A \mathbb{E}_0[\mathbb{E}_n(X)] \, d\mathbb{P} \\ &= \lim_{n \rightarrow +\infty} \int_A \mathbb{E}_0(X) \, d\mathbb{P} = \int_A \mathbb{E}_0(X) \, d\mathbb{P}, \end{aligned}$$

sicché $Z = \mathbb{E}_0(X)$ q.o.. \square

6.7 Applicazioni

In questa sezione sono raccolte alcune delle numerose applicazioni delle martingale.

6.7.1 Il teorema di Radon–Nikodym.

Le martingale costituiscono uno strumento potente sia per la semplificazione che apportano alla dimostrazione di molti risultati sia perché consentono nuovi fruttuosi punti di vista in molti problemi di probabilità. In questa sezione il Teorema di Radon–Nikodym (Teorema 1.14.3) è dimostrato mediante il ricorso alle martingale.

Sia $(\Omega, \mathcal{F}, \mathbb{P})$ uno spazio di probabilità e sia ν una misura *finita* su (Ω, \mathcal{F}) assolutamente continua rispetto a \mathbb{P} , $\nu \ll \mathbb{P}$. Si consideri la famiglia Σ delle partizioni finite di Ω in insiemi misurabili. Se α e β sono in Σ , si dirà che $\alpha \leq \beta$ se β è un raffinamento di α ; Σ è così un insieme diretto. Se $\alpha \in \Sigma$, si indichi con \mathcal{F}_α la tribù

generata dalla partizione α . Si osservi che si ha $\mathcal{F} = \cup_{\alpha \in \Sigma} \mathcal{F}_\alpha$; infatti, un insieme $A \in \mathcal{F}$ appartiene alla partizione $\{A, A^c\}$ che genera la tribú $\mathcal{F}(A) = \{\emptyset, A, A^c, \Omega\}$.

Per ogni partizione $\alpha \in \Sigma$, si definisca

$$X_\alpha := \sum_{A \in \Sigma} \varphi(A) \mathbf{1}_A,$$

ove

$$\varphi(A) := \begin{cases} \frac{\nu(A)}{\mathbb{P}(A)}, & \text{se } \mathbb{P}(A) > 0, \\ 0, & \text{se } \mathbb{P}(A) = 0. \end{cases}$$

La famiglia $(X_\alpha, \mathcal{F}_\alpha)$ è una martingala; sia $\alpha \leq \beta$, cioè sia β una partizione di Ω piú fine di α ; allora

$$\begin{aligned} \mathbb{E}_\alpha(X_\beta) &= \sum_{A \in \beta} \frac{\nu(A)}{\mathbb{P}(A)} \mathbb{E}_\alpha(\mathbf{1}_A) = \sum_{A \in \beta} \frac{\nu(A)}{\mathbb{P}(A)} \sum_{B \in \alpha} \frac{\mathbf{1}_B}{\mathbb{P}(B)} \int_B \mathbf{1}_A \, d\mathbb{P} \\ &= \sum_{A \in \beta} \frac{\nu(A)}{\mathbb{P}(A)} \sum_{B \in \alpha} \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \mathbf{1}_B = \sum_{B \in \alpha} \frac{\mathbf{1}_B}{\mathbb{P}(B)} \sum_{A \in \beta} \frac{\nu(A)}{\mathbb{P}(A)} \mathbb{P}(A \cap B) \\ &= \sum_{B \in \alpha} \frac{\mathbf{1}_B}{\mathbb{P}(B)} \sum \left\{ \frac{\nu(A)}{\mathbb{P}(A)} \mathbb{P}(A) : A \in \beta, A \cap B \neq \emptyset \right\} \\ &= \sum_{B \in \alpha} \frac{\mathbf{1}_B}{\mathbb{P}(B)} \nu(B) = X_\alpha. \end{aligned}$$

Inoltre, $(X_\alpha, \mathcal{F}_\alpha)_{\alpha \in \Sigma}$ è uniformemente integrabile. Infatti, poiché è in \mathcal{F}_α l'insieme $\{|X_\alpha| \geq c\}$,

$$\begin{aligned} \int_{\{|X_\alpha| \geq c\}} |X_\alpha| \, d\mathbb{P} &= \int_{\{|X_\alpha| \geq c\}} X_\alpha \, d\mathbb{P} = \sum_{A \in \alpha} \frac{\nu(A)}{\mathbb{P}(A)} \int_{\{|X_\alpha| \geq c\}} \mathbf{1}_A \, d\mathbb{P} \\ &= \sum \{\nu(A) : A \in \alpha, A \cap \{X_\alpha \geq c\} \neq \emptyset\} = \nu(\{X_\alpha \geq c\}). \end{aligned}$$

D'altro canto,

$$\mathbb{P}(\{X_\alpha \geq c\}) \leq \frac{\mathbb{E}(X_\alpha)}{c} = \frac{\nu(\Omega)}{c};$$

di qui, e dall'essere ν assolutamente continua rispetto a \mathbb{P} , segue l'integrabilità uniforme di (X_α) . In virtù del Teorema 6.5.3, esiste $X \in L^1(\mathcal{F})$ tale che $X_\alpha = \mathbb{E}_\alpha(X)$. Sia ora $A \in \mathcal{F}$ e si consideri la partizione finita $\alpha = \{A, A^c\}$. Allora

$$\nu(A) = \int_A X_\alpha \, d\mathbb{P} = \int_A X \, d\mathbb{P}, \quad (6.7.1)$$

che esprime il teorema di Radon–Nikodym nelle ipotesi considerate. Per riportarsi alle ipotesi del Teorema (3.10.7), nelle quali si considera una misura finita μ in luogo della misura di probabilità \mathbb{P} , basta applicare quanto precede alla misura di probabilità definita, per ogni $A \in \mathcal{F}$ da

$$\mathbb{P}(A) := \frac{\mu(A)}{\mu(\Omega)}.$$

La 6.7.1 si scriverà ora

$$\nu(A) = \int_A \frac{X}{\mu(\Omega)} d\mu.$$

La densità X che compare nella 6.7.1 è quindi risultata divisa per il fattore $\mu(\Omega)$.

6.7.2 Legge 0–1 di Kolmogorov

Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia data una successione (X_n) di v.a. indipendenti di L^1 . Al solito sia (\mathcal{F}_n) la filtrazione naturale e $\mathcal{F}_\infty = \bigvee_n \mathcal{F}_n$ la tribù che essa genera. Si definiscano anche le tribù

$$\mathcal{T}_n := \mathcal{F}(X_{n+1}, X_{n+2}, \dots) \quad \text{e} \quad \mathcal{T} := \bigcap_n \mathcal{T}_n.$$

La tribù \mathcal{T} si dice *tribù terminale* di (X_n) e i suoi elementi *eventi terminali*. Esempi di eventi terminali sono dati da

$$\begin{aligned} A_1 &:= \left\{ \text{esiste } \lim_{n \rightarrow +\infty} X_n \right\}, \\ A_2 &:= \left\{ \sum_{n \in \mathbb{N}} X_n \text{ converge} \right\}, \\ A_3 &:= \left\{ \text{esiste } \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n X_k \right\}. \end{aligned}$$

Teorema 6.7.1. (Legge 0–1 di Kolmogorov) *Sia (X_n) una successione di v.a. indipendenti di L^1 e sia \mathcal{T} la tribù terminale di (X_n) . Per ogni evento terminale $A \in \mathcal{T}$ vale $\mathbb{P}(A) = 0$ oppure $\mathbb{P}(A) = 1$.*

Dimostrazione. Dato $A \in \mathcal{T}$ si ponga $X = \mathbf{1}_A$. Poiché A appartiene anche a \mathcal{F}_∞ , X risulta misurabile rispetto a \mathcal{F}_∞ , sicché segue dal teorema di convergenza di Lévy 6.5.1 che

$$X = \mathbb{E}(X \mid \mathcal{F}_\infty) = \lim_{n \rightarrow +\infty} \mathbb{E}(X \mid \mathcal{F}_n) \quad \text{q.c.}$$

D'altro canto X è misurabile rispetto a \mathcal{T}_n per ogni $n \in \mathbb{N}$ sicché essa è indipendente da \mathcal{F}_n ; perciò

$$\mathbb{E}(X \mid \mathcal{F}_n) = \mathbb{E}(X) = \mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A) \quad \text{q.c.}$$

Le ultime due relazioni implicano q.c. $\mathbb{P}(A) = X = \mathbf{1}_A$, che, essendo una funzione indicatrice, assume solo i valori 0 e 1. \square

6.7.3 Convergenza di serie di v.a..

Diamo ora alcune applicazioni delle martingale allo studio delle successioni di v.a. indipendenti.

Teorema 6.7.2. *Sia (X_n) una successione di v.a. indipendenti e centrate di L^2 . Si supponga che la successione sia uniformemente limitata in L^2 , $\mathbb{E}(\sup_n X_n^2) < \infty$. Sono allora equivalenti le proprietà:*

- (a) $\sum_{n \in \mathbb{N}} X_n$ converge q.c.;
- (b) $\sum_{n \in \mathbb{N}} V(X_n) = \sum_{n \in \mathbb{N}} \mathbb{E}(X_n^2) < \infty$.

Dimostrazione. (a) \implies (b) Si indichi, al solito, con (\mathcal{F}_n) la filtrazione naturale

$$\mathcal{F}_n = \mathcal{F}(X_1, \dots, X_n).$$

Si sa che, se $S_n := \sum_{j=1}^n X_j$, (S_n, \mathcal{F}_n) è una martingala. Poiché la serie $\sum_{n \in \mathbb{N}} X_n$ converge q.c., esiste $\alpha > 0$ tale che

$$\mathbb{P}\left(\sup_{n \in \mathbb{N}} |S_n| \leq \alpha\right) > 0.$$

Si introduca il tempo d'arresto T così definito: $T := \inf\{n \in \mathbb{N} : |S_n| > \alpha\}$ se un tale n esiste, altrimenti $T := +\infty$ se $|S_n| \leq \alpha$ per ogni $n \in \mathbb{N}$. In virtù del Teorema 6.3.2, S^T è ancora una martingala. Se $T > n$, è

$$S^T = S_n = S_{n-1} + (S_n - S_{n-1}) \leq \alpha + (S_n - S_{n-1}).$$

Se, invece, $T \leq n$, allora

$$S^T = S_{T-1} + (S_T - S_{T-1}) \leq \alpha + (S_T - S_{T-1}).$$

In ogni caso, $S^T \leq \alpha + Z$ ove

$$Z := \sup_{n \in \mathbb{N}} |S_n - S_{n-1}| = \sup_{n \in \mathbb{N}} |X_n|,$$

sicché, per ipotesi, Z è in L^2 , $\mathbb{E}(Z^2) < +\infty$. Ma allora S^T è una martingala quadratica che converge q.c. e in L^2 . Inoltre, per il Teorema 6.1.3,

$$\mathbb{E}(S_{T \wedge n}^2) = \sum_{j=1}^n \mathbb{E}\left[(S_{T \wedge n} - S_{T \wedge (n-1)})^2\right]$$

ove, per comodità di notazione, si è posto $S_{T \wedge 0} := 0$. Poiché $S^T = (S_{T \wedge n})$ converge in L^2 , la successione $(\mathbb{E}(S_{T \wedge n}^2))$ ammette limite e dunque l'ultima serie converge. Ma $S_{T \wedge n} - S_{T \wedge (n-1)} = X_n \mathbf{1}_{\{T \geq n\}}$, sicché è convergente la serie

$$\sum_{n \in \mathbb{N}} \mathbb{E}(X_n^2 \mathbf{1}_{\{T \geq n\}}),$$

e, di conseguenza, converge q.c. anche la serie

$$\sum_{n \in \mathbb{N}} \mathbb{E}(X_n^2 \mathbf{1}_{\{T \geq n\}} | X_1, \dots, X_{n-1}) = \sum_{n \in \mathbb{N}} \mathbb{E}_{n-1}(X_n^2 \mathbf{1}_{\{T \geq n\}}).$$

Infatti, se $Z_n \geq 0$ e $\sum_{n \in \mathbb{N}} \mathbb{E}(Z_n) < +\infty$, allora $\mathbb{E}(\sum_{n \in \mathbb{N}} Z_n) < +\infty$ e $\sum_{n \in \mathbb{N}} Z_n$ converge q.c.. Basta ora porre

$$Z_n := \mathbb{E}_{n-1}(X_n^2 \mathbf{1}_{\{T \geq n\}}).$$

Poiché $\mathbf{1}_{\{T \geq n\}}$ è misurabile rispetto alla tribù \mathcal{F}_{n-1} , converge q.c. la serie

$$\sum_{n \in \mathbb{N}} \mathbf{1}_{\{T \geq n\}} \mathbb{E}_{n-1}(X_n^2).$$

Si scelga, infine, ω in $\{\sup_{n \in \mathbb{N}} |S_n| \leq \alpha\}$ e tale che l'ultima serie converga. Allora $T(\omega) = +\infty \geq n$, sicché $\sum_{n \in \mathbb{N}} \mathbb{E}(X_n^2) < +\infty$.

(b) \implies (a) Si ha, in virtù della disuguaglianza di Schwartz

$$\mathbb{E}(|S_n|) = \int \left| \sum_{k=1}^n X_k \right| d\mathbb{P} \leq \left\{ \int \left(\sum_{k=1}^n X_k \right)^2 d\mathbb{P} \right\}^{1/2};$$

ora, poiché le X_k sono indipendenti e centrate,

$$\int \left(\sum_{k=1}^n X_k \right)^2 d\mathbb{P} = \sum_{k=1}^n \mathbb{E}(X_k^2) + \sum_{j \neq k} \mathbb{E}(X_j) \mathbb{E}(X_k) = \sum_{k=1}^n \mathbb{E}(X_k^2),$$

sicché

$$\mathbb{E}(|S_n|) \leq \left\{ \sum_{k=1}^n \mathbb{E}(X_k^2) \right\}^{1/2}.$$

Dunque, $\sup_{n \in \mathbb{N}} \mathbb{E}(|S_n|) < +\infty$ e la martingala (S_n) converge q.c. \square

Teorema 6.7.3. *Sia (X_n) una successione di v.a. indipendenti ed uniformemente limitate di L^2 . Sono allora equivalenti le proprietà:*

- (a) $\sum_{n \in \mathbb{N}} X_n$ converge q.c.;
- (b) sono convergenti entrambe le serie

$$\sum_{n \in \mathbb{N}} \mathbb{E}(X_n) \quad e \quad \sum_{n \in \mathbb{N}} V(X_n).$$

Dimostrazione. (a) \implies (b) Si costruisca la successione di v.a. indipendenti

$$(X_1, Y_1, X_2, Y_2, \dots, X_n, Y_n, \dots),$$

ove, per ogni $n \in \mathbb{N}$, Y_n ha la stessa legge di X_n . Allora $\mathbb{E}(X_n - Y_n) = 0$ e

$$\mathbb{E}[(X_n - Y_n)^2] = V(X_n - Y_n) = V(X_n) + V(Y_n) = 2V(X_n).$$

Poiché $\sum_{n \in \mathbb{N}} X_n$ converge q.c., altrettanto fa $\sum_{n \in \mathbb{N}} Y_n$. Perciò

$$\sum_{n \in \mathbb{N}} (X_n - Y_n)$$

converge q.c.; per il teorema precedente, $\sum_{n \in \mathbb{N}} V(X_n) < +\infty$ e, per lo stesso teorema, converge q.c. la serie $\sum_{n \in \mathbb{N}} (X_n - \mathbb{E}(X_n))$. Dunque, è q.c. convergente la serie $\sum_{n \in \mathbb{N}} \mathbb{E}(X_n)$.

(b) \implies (a) Per il teorema precedente, converge q.c. la serie

$$\sum_{n \in \mathbb{N}} (X_n - \mathbb{E}(X_n))$$

e, poiché è convergente $\sum_{n \in \mathbb{N}} \mathbb{E}(X_n)$, ne segue che converge q.c.

$$\sum_{n \in \mathbb{N}} X_n,$$

che stabilisce l'asserto. \square

Teorema 6.7.4. (delle tre serie) *Sia (X_n) una successione di v.a. indipendenti. Se $\alpha > 0$, si definisca, per ogni $n \in \mathbb{N}$,*

$$Y_n^\alpha := X_n \mathbf{1}_{\{|X_n| \leq \alpha\}}.$$

(a) *Se la serie $\sum_{n \in \mathbb{N}} X_n$ converge q.c., per ogni $\alpha > 0$, convergono le tre serie*

$$\sum_{n \in \mathbb{N}} \mathbb{P}(X_n \neq Y_n^\alpha), \quad \sum_{n \in \mathbb{N}} \mathbb{E}(Y_n^\alpha), \quad \sum_{n \in \mathbb{N}} V(Y_n^\alpha). \quad (6.7.2)$$

(b) *Se esiste $\alpha > 0$ tale che le tre serie (6.7.2) convergano, allora converge q.c. la serie*

$$\sum_{n \in \mathbb{N}} X_n.$$

Dimostrazione. (a) Per ipotesi, $X_n \rightarrow 0$ q.c., sicché si ha definitivamente $X_n = Y_n^\alpha$ q.c. per ogni $\alpha > 0$. Poiché $X_n \neq Y_n^\alpha$ solo per un numero finito di indici, la serie $\sum_{n \in \mathbb{N}} \mathbb{P}(X_n \neq Y_n^\alpha)$ converge. Le rimanenti due serie convergono in virtù del teorema precedente.

(b) La convergenza delle due ultime serie in (6.7.2) assicura, per il Teorema 6.7.3, la convergenza q.c. della serie $\sum_{n \in \mathbb{N}} Y_n^\alpha$. Ora, la convergenza di

$$\sum_{n \in \mathbb{N}} \mathbb{P}(X_n \neq Y_n^\alpha)$$

implica, in virtù del primo lemma di Borel–Cantelli,

$$\mathbb{P}\left(\liminf_{n \rightarrow +\infty} \{X_n \neq Y_n^\alpha\}\right) = 1;$$

in altre parole, si ha definitivamente $X_n = Y_n^\alpha$ q.c.; quindi, $\sum_{n \in \mathbb{N}} X_n$ converge q.c.. \square

6.7.4 Le LGN forti

Possiamo dimostrare la LGN forte di Kolmogorov 4.4.3 usando la convergenza delle martingale. Sarà utile premettere due lemmi.

Lemma 6.7.1. (Töplitz). *Sia (a_n) una successione di numeri reali positivi, $a_n \geq 0$ per ogni $n \in \mathbb{N}$. Posto*

$$b_n := \sum_{j=1}^n a_j,$$

si suppongano verificate le condizioni:

- (a) $b_n > 0$ per ogni $n \in \mathbb{N}$;
- (b) $\lim_{n \rightarrow +\infty} b_n = +\infty$.

Allora se (x_n) è una successione di numeri reali convergente a x ,

$$\lim_{n \rightarrow +\infty} x_n = x,$$

si ha

$$\lim_{n \rightarrow +\infty} \frac{1}{b_n} \sum_{j=1}^n a_j x_j = x.$$

Dimostrazione. Dalla convergenza della successione (x_n) segue che per ogni $\varepsilon > 0$ si può trovare $n_0 = n_0(\varepsilon) \in \mathbb{N}$ tale che, per ogni $n \geq n_0$ risulti $|x_n - x| < \varepsilon$; inoltre la successione $(|x_n - x|)$ essendo convergente è limitata, $|x_n - x| \leq H$ per ogni $n \in \mathbb{N}$. Perciò, per $n > n_0$ si ha

$$\begin{aligned} \left| \frac{1}{b_n} \sum_{k=1}^n a_k x_k - x \right| &\leq \frac{1}{b_n} \sum_{k=1}^{n_0-1} a_k |x_k - x| + \frac{1}{b_n} \sum_{k=n_0}^n a_k |x_k - x| \\ &\leq \frac{b_{n_0-1}}{b_n} H + \frac{\varepsilon}{b_n} \sum_{k=n_0}^n a_k \leq \varepsilon + \frac{b_{n_0-1}}{b_n} H \xrightarrow{n \rightarrow +\infty} \varepsilon, \end{aligned}$$

che dimostra l'asserto. □

Si noti che per $a_n = 1$ per ogni $n \in \mathbb{N}$ si ha la convergenza nel senso di Cesàro.

Lemma 6.7.2. (Kronecker). *Siano (x_n) una successione di numeri reali e (b_n) una successione crescente di numeri strettamente positivi, tale che*

$$\lim_{n \rightarrow +\infty} b_n = +\infty.$$

Allora la convergenza della serie

$$\sum_{n \in \mathbb{N}} \frac{x_n}{b_n} \tag{6.7.3}$$

implica

$$\lim_{n \rightarrow +\infty} \frac{1}{b_n} \sum_{k=1}^n x_k = 0.$$

Dimostrazione. Sia eguale a s la somma della serie (6.7.3); allora, posto

$$s_n := \sum_{k=1}^n \frac{x_k}{b_k},$$

si ha $\lim_{n \rightarrow +\infty} s_n = s$. Ora

$$\sum_{k=1}^n x_k = \sum_{k=1}^n b_k (s_k - s_{k-1}) = b_n s_n - b_1 s_0 - \sum_{k=1}^n s_{k-1} (b_k - b_{k-1}),$$

ove si è posto $b_0 = s_0 := 0$. Perciò, se $a_{k-1} := b_k - b_{k-1}$ si ottiene

$$\frac{1}{b_n} \sum_{k=1}^n x_k = s_n - \frac{1}{b_n} \sum_{k=1}^n a_{k-1} s_{k-1}.$$

Si osservi che

$$\sum_{k=1}^n a_{k-1} = \sum_{k=1}^n (b_k - b_{k-1}) = b_n,$$

sicché il lemma di Töplitz dà

$$\lim_{n \rightarrow +\infty} \frac{1}{b_n} \sum_{k=1}^n a_{k-1} s_{k-1} = s.$$

Infine

$$\lim_{n \rightarrow +\infty} \frac{1}{b_n} \sum_{k=1}^n x_k = s - s = 0,$$

che conclude la dimostrazione. \square

Teorema 6.7.5. *Sia (X_n) una successione di v.a. indipendenti e centrate di L^2 ; se è convergente la serie*

$$\sum_{n \in \mathbb{N}} \frac{V(X_n)}{n^2} < +\infty, \quad (6.7.4)$$

allora la successione (X_n) obbedisce alla LGN forte.

Dimostrazione. Basta dimostrare che converge quasi certamente la serie

$$\sum_{n \in \mathbb{N}} \frac{X_n}{n}, \quad (6.7.5)$$

perché allora il Lemma di Kronecker assicura che sia

$$0 = \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n X_k = \lim_{n \rightarrow +\infty} \frac{S_n}{n} :$$

ma la convergenza (6.7.5) segue ora dal Teorema 6.7.2. \square

Anche la LGN forte di Khinchin–Kolmogorov, Teorema 4.4.4, può essere dimostrata con l’ausilio delle martingale.

Teorema 6.7.6. *Una successione (X_n) di v.a. di L^1 indipendenti e isonome obbedisce alla LGN forte.*

Dimostrazione. Con la stessa notazione dell’Esempio 6.6.1, $(Z_{-n}, \mathcal{F}_{-n})$ è una martingala rovesciata. Per il Teorema 6.6.1 si ha

$$\lim_{n \rightarrow +\infty} \frac{S_n}{n} = \mathbb{E}(X_1 | \mathcal{F}_{-\infty}) \quad \text{q.c. e in } L^1.$$

Ora $\mathcal{F}_{-\infty}$ è la tribú terminale \mathcal{T} , che per la legge 0–1 di Kolmogorov (Teorema 6.7.1) è banale, sicché $\mathbb{E}(X_1 | \mathcal{F}_{-\infty})$ è costante; da $\mathbb{E}[\mathbb{E}(X_1 | \mathcal{F}_{-\infty})] = \mathbb{E}(X_1)$ segue l’asserto. \square

6.7.5 Variabili scambiabili e il teorema di De Finetti

Definizione 6.7.1. Una successione (X_n) di variabili aleatorie sullo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ si dice *scambiabile* se, per ogni permutazione π di \mathbb{N} che lascia invariati tutti i numeri con l'eccezione di al più un numero finito di essi, le successioni (X_n) e $(X_{\pi(n)})$ hanno la stessa legge, vale a dire, se per ogni $n \in \mathbb{N}$ e per ogni scelta di n numeri reali t_1, \dots, t_n si ha

$$\mathbb{P}(X_1 \leq t_1, \dots, X_n \leq t_n) = \mathbb{P}(X_{\pi(1)} \leq t_1, \dots, X_{\pi(n)} \leq t_n).$$

◇

Si noti in particolare, che tutte le variabili di una successione scambiabile hanno la stessa legge. Inoltre una successione di variabili aleatorie indipendenti e isonome è scambiabile.

Esempio 6.7.1. Siano $\Theta, X_1, \dots, X_n, \dots$ v.a. definite sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e si supponga che, subordinatamente a Θ , le v.a. X_j siano indipendenti e Bernoulliane con

$$\begin{cases} \mathbb{P}(X_j = 1) = \Theta, \\ \mathbb{P}(X_j = 0) = 1 - \Theta. \end{cases} \quad (j = 1, \dots, n)$$

Variabili aleatorie siffatte esistono; per rendersene conto, siano Θ, Z_1, Z_2, \dots v.a. indipendenti e Θ abbia un'arbitraria distribuzione in $[0, 1]$. Basta ora porre

$$X_j := \mathbf{1}_{\{Z_j \leq \Theta\}}.$$

Si ha allora

$$\mathbb{P}(X_1 = u_1, \dots, X_n = u_n \mid \Theta = \theta) = \theta^s (1 - \theta)^{n-s}, \quad (6.7.6)$$

ove per $j = 1, \dots, n$ è $u_j \in \{0, 1\}$ e $u_1 + \dots + u_n = s$. Integrando l'ultima espressione si ottiene

$$\mathbb{P}(X_1 = u_1, \dots, X_n = u_n) = \mathbb{E}(\Theta^s (1 - \Theta)^{n-s}), \quad (6.7.7)$$

sicché (X_n) è scambiabile. ■

Nell'esempio appena dato (X_n) è una mistura di processi di Bernoulli. Il seguente teorema di De Finetti afferma che *ogni* successione di v.a. che assumono i valori 0 e 1 è una mistura di processi di Bernoulli.

Teorema 6.7.7. *Se la successione (X_n) è scambiabile e se le v.a. X_n assumono valori in $\{0, 1\}$ esiste una v.a. Θ sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ tale che valgano le (6.7.6) e (6.7.7).*

Dimostrazione. Si ponga, al solito $S_n = \sum_{j=1}^n X_j$; per $k \leq N$ si ha

$$\mathbb{P}(S_n = k) = \sum' \mathbb{P}(X_1 = u_1, \dots, X_n = u_n),$$

ove la somma si esegue su tutte le sequenze (u_1, \dots, u_n) con $u_j \in \{0, 1\}$ e tali che $\sum_{j=1}^n u_j = k$. La scambiabilità assicura che siano eguali tutti i termini dell'ultima somma, e, poiché questa consiste di $\binom{n}{k}$ termini si ha

$$\mathbb{P}(X_1 = u_1, \dots, X_n = u_n \mid S_n = k) = \binom{n}{k}^{-1}.$$

Si supponga ora che sia $h \leq j \leq n$ e $\sum_{i=1}^j u_i = h \leq k \leq n$ e si sommi su u_{j+1}, \dots, u_n tali che $\sum_{i=j+1}^n u_i = k - h$ per ottenere

$$\begin{aligned} \mathbb{P}(X_1 = u_1, \dots, X_k = u_k \mid S_n = k) &= \binom{n-j}{k-h} \binom{n}{k}^{-1} \\ &= \frac{D_{k,h} d_{n-k,j-h}}{D_{n,j}} =: f_{j,s,n} \left(\frac{k}{n} \right). \end{aligned}$$

Quanto precede continua a valere se si aggiungono altri condizionamenti, $S_{n+1} = u_{n+1}, \dots, S_{n+r} = u_{n+r}$. Perciò

$$\mathbb{P}(X_1 = u_1, \dots, X_k = u_k \mid S_n, S_{n+1}, \dots, S_{n+r}) = f_{j,s,n} \left(\frac{S_n}{n} \right).$$

Sia $\mathcal{S}_n := \mathcal{F}(S_n, S_{n+1}, \dots)$ la tribú generata da S_n, S_{n+1}, \dots e sia $\mathcal{S} := \bigcap_{n \in \mathbb{N}} \mathcal{S}_n$ la tribú terminale. Si fissino j e u_1, \dots, u_j e si supponga che sia $u_1 + \dots + u_j = h$. Si faccia tendere r a $+\infty$; allora, per il Corollario 6.5.1 è

$$\mathbb{P}(X_1 = u_1, \dots, X_k = u_k \mid \mathcal{S}_n) = f_{j,s,n} \left(\frac{S_n}{n} \right).$$

Si faccia ora tendere n a $+\infty$; per il Teorema 6.6.2 si ha

$$\mathbb{P}(X_1 = u_1, \dots, X_k = u_k \mid \mathcal{S}) = \lim_{n \rightarrow +\infty} f_{j,s,n} \left(\frac{S_n}{n} \right) \quad \text{q.c.}$$

Tale limite vale fuori di un insieme trascurabile $N \in \mathcal{F}$, $\mathbb{P}(N) = 0$.

Sia ora $\omega \in N^c$ e si supponga, se possibile, che la successione $(S_n(\omega)/n)$ abbia due punti d'accumulazione. Ora

$$\left| \frac{S_n}{n} - \frac{S_{n+1}}{n+1} \right| = \begin{cases} \frac{S_n}{n(n+1)} \leq \frac{1}{n+1}, & \text{se } S_{n+1} = S_n, \\ \frac{S_n + n}{n(n+1)} \leq \frac{2}{n+1} & \text{se } S_{n+1} = S_n + 1. \end{cases}$$

In entrambi i casi la differenza considerata è minore di $2/n$; questo significa che i punti d'accumulazione sono densi in un intervallo I . Ora, $\lim_{\nu \rightarrow +\infty} x_{n_\nu} = x$ implica $\lim_{\nu \rightarrow +\infty} f_{j,s,n_\nu}(x_{n_\nu}) = x^s (1-x)^{j-s}$, sicché ne seguirebbe che $x^s (1-x)^{j-s}$ è costante in quell'intervallo I , una contraddizione. Dunque la successione converge a un limite $\Theta(\omega)$

$$\lim_{n \rightarrow +\infty} \frac{S_n(\omega)}{n} = \Theta(\omega).$$

Pertanto

$$\lim_{n \rightarrow +\infty} \mathbb{P}(X_1 = u_1, \dots, X_k = u_j \mid \mathcal{S}) = \Theta^h (1 - \Theta)^{j-h} \quad \text{q.c.}$$

Prendendo la speranza condizionata rispetto alla tribú $\mathcal{F}(\Theta)$ generata da Θ si ottiene la (6.7.6). \square

6.7.6 La rovina del giocatore

Il problema della *rovina del giocatore* è trattato nella maggior parte dei corsi introduttivi di Probabilità. Qui ne dò una trattazione piú raffinata basata sui tempi d'arresto e sulle martingale.

Si torni alla passeggiata aleatoria con l'interpretazione di due giocatori, siano G_1 e G_2 che ad ogni istante giocano una partita nella quale il giocatore G_1 vince € 1 con probabilità p , mentre il giocatore G_2 vince € 1 con probabilità q . Si ha quindi una successione $(X_n)_{n \in \mathbb{Z}_+}$ di variabili indipendenti con $\mathbb{P}(X_n = +1) = p$ e $\mathbb{P}(X_n = -1) = q$. Se il giocatore G_1 dispone inizialmente di € a e se $0 < a < b$, si vuole sapere quale sia la probabilità che il giocatore arrivi ad avere € b prima di perdere tutto il suo capitale iniziale di € a ; qui a e b sono numeri naturali. Posto, al solito, $S_n = \sum_{j=1}^n X_j$, si introducano i tempi d'arresto

$$T_b := \inf\{n : S_n = b\}, \quad T_{-a} := \inf\{n : S_n = -a\}.$$

Si ponga, inoltre,

$$T := T_b \wedge T_{-a}, \quad A := \{T_b < T_{-a}\}, \quad B := \{T_{-a} < T_b\}.$$

Al tempo n il capitale del giocatore G_1 è $a + S_n$, mentre il capitale del suo antagonista G_2 è $b - S_n$. T_{-a} rappresenta il tempo necessario perché il giocatore G_1 perda gli a € che possedeva inizialmente (la "rovina"); l'evento A significa che G_1 guadagna b € prima di perdere quanto aveva in partenza, mentre B rappresenta la rovina, vale a dire la perdita del capitale iniziale.

Studieremo dapprima il caso simmetrico, supponendo che X_n possa assumere anche il valore 0.

Nel teorema che segue si supporrà che sia $\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1)$ e inoltre che queste probabilità non siano necessariamente costanti:

$$p_n = \mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1).$$

Teorema 6.7.8. *Si supponga che sia*

$$\sum_{n \in \mathbb{N}} p_n = +\infty.$$

Allora

- (a) *quasi certamente la successione (S_n) non converge;*
- (b) *i tempi d'arresto T_b e T_{-a} sono entrambi quasi certamente finiti, ma nessuno di essi è integrabile;*

(c) è

$$\mathbb{P}(A) = \frac{a}{a+b} \quad e \quad \mathbb{P}(B) = \frac{b}{a+b};$$

(d) *T è integrabile se $\inf_{n \in \mathbb{N}} p_n = p > 0$;*

(e) *se si ha $p_n = p$ per ogni $n \in \mathbb{N}$, allora*

$$\mathbb{E}(T) = \frac{ab}{2p}.$$

Dimostrazione. (a) La successione (S_n) non converge quando in un numero infinito di termini si ha $|X_n| = 1$, sicché (S_n) non converge nell'insieme

$$\{\limsup_n \{|X_n| = 1\}\}.$$

Ma $\mathbb{P}(|X_n| = 1) = 2p_n$ e poiché la serie $\sum_n p_n$ per ipotesi diverge, il secondo lemma di Borel–Cantelli assicura che sia

$$\mathbb{P}\left(\limsup_{n \rightarrow +\infty} \{|X_n| = 1\}\right) = 1.$$

(b) È noto dall'Esempio 6.1.3 che (S_n) è una martingala; poiché i suoi incrementi sono limitati dalla costante 1, il Teorema 6.5.6 assicura che essa converge quasi certamente nell'insieme $\{\sup_n S_n < +\infty\}$ e, quindi, in particolare, in $\{T_b = +\infty\}$. Per quanto visto in (a) (S_n) converge in un insieme di probabilità nulla; dunque, $\mathbb{P}(T_b = +\infty) = 0$, sicché T_b è quasi certamente finito. Il Corollario 6.5.2 assicura che T_b non sia integrabile. Si procede in maniera simmetrica per T_{-a} .

(c) Poiché entrambi i tempi d'arresto T_{-a} e T_b sono quasi certamente finiti, tale è anche il tempo d'arresto $T = T_{-a} \wedge T_b$. La martingala arrestata S^T è limitata $-a \leq S^T \leq b$, sicché è uniformemente integrabile, e quindi, per il Teorema 6.4.4, ereditaria; pertanto converge per il Teorema 6.5.1. D'altro canto la successione $(S_{n \wedge T})_{n \in \mathbb{Z}_+}$ è definitivamente eguale a S_T nell'insieme $\{T < +\infty\}$ con $\mathbb{P}(T < +\infty) = 1$. Perciò

$$0 = \mathbb{E}(S_0) = \mathbb{E}(S_T) = \mathbb{E}(b \mathbf{1}_A - a \mathbf{1}_B) = p \mathbb{P}(A) - a \mathbb{P}(B).$$

Di qui e dall'ovvia condizione $\mathbb{P}(A) + \mathbb{P}(B) = 1$ scende l'asserto.

(d) Alla luce dell'Esempio 6.3.1 il compensatore della sotto-martingala (S_n^2) è

$$V_n = \sum_{j=1}^n \mathbb{E} \left[(S_n - S_{n-1})^2 \right] = \sum_{j=1}^n \mathbb{E}(X_n^2) = 2 \sum_{j=1}^n p_n \geq 2np.$$

Poiché $(S_n - V_n)_{n \in \mathbb{Z}_+}$ è una martingala nulla in 0, tale è anche $(S_n - V_n)^T$, sicché $\mathbb{E}(S_{n \wedge T}^2) = \mathbb{E}(V_{n \wedge T})$. Ora

$$2p \mathbb{E}(n \wedge T) \leq \mathbb{E}[V_{n \wedge T}] = \mathbb{E}[S_{n \wedge T}^2].$$

Si faccia tendere n a $+\infty$; utilizzando il teorema di convergenza monotona a primo membro e quello di convergenza dominata secondo membro, si ottiene

$$2p \mathbb{E}(T) \leq \mathbb{E}[S_T^2] = \mathbb{E}(b^2 \mathbf{1}_A + a^2 \mathbf{1}_B) = b^2 \frac{a}{a+b} + a^2 \frac{b}{a+b} = ab.$$

(e) Se si ha $p_n = p$ per ogni $n \in \mathbb{N}$, allora nelle relazioni precedenti si ha eguaglianza. \square

Esaminiamo ora il caso asimmetrico, nel quale supporremo che sia, per ogni $n \in \mathbb{N}$,

$$\mathbb{P}(X_n = 1) = p \in]0, 1[\quad \mathbb{P}(X_n = -1) = q = 1 - p.$$

Se $p > q$ sicché è favorito il giocatore G_1 .

Teorema 6.7.9. *Nelle ipotesi dette si ha*

$$\mathbb{P}(A) = \frac{\left(\frac{q}{p}\right)^{-a} - 1}{\left(\frac{q}{p}\right)^{-a} - \left(\frac{q}{p}\right)^b} \quad e \quad \mathbb{P}(B) = \frac{1 - \left(\frac{q}{p}\right)^b}{\left(\frac{q}{p}\right)^{-a} - \left(\frac{q}{p}\right)^b},$$

$$\mathbb{P}(T_{-a} < +\infty) = \left(\frac{q}{p}\right)^a,$$

$$\mathbb{E}(T_b) = \frac{b}{p-q}, \quad \mathbb{E}(T) = \frac{b\mathbb{P}(A) - a\mathbb{P}(B)}{p-q}.$$

Dimostrazione. In virtù della LGN forte di Khinchin–Kolmogorov (Teorema 4.4.4) la successione (S_n/n) converge quasi certamente a $p - q > 0$, poiché $\mathbb{E}(X_n) = p - q$; pertanto (S_n) converge q.c. a $+\infty$. Di conseguenza, il tempo d'arresto T_b è q.c. finito e, *a fortiori*, è q.c. finito T . Dall'Esercizio 6.6 segue che

$$W_n := \left(\frac{q}{p}\right)^{S_n}$$

è una martingala. La martingala arrestata W^T è positiva e limitata superiormente da $(q/p)^{-a}$. Pertanto essa è uniformemente integrabile; di conseguenza, per i Teoremi 6.5.2 e 6.5.4, converge q.c.. Ora, si è visto sopra che $\mathbb{P}(T < +\infty) = 1$; d'altro canto, in $\{T < +\infty\}$ si ha definitivamente $W_{n \wedge T} = W_T$, sicché $W_n = \mathbb{E}_n(X_T)$. Quindi

$$1 = \mathbb{E}(W_0) = \mathbb{E}(W_T) = \mathbb{E} \left[\left(\frac{q}{p}\right)^b \mathbf{1}_A + \left(\frac{q}{p}\right)^{-a} \mathbf{1}_B \right]$$

$$= \left(\frac{q}{p}\right)^b \mathbb{P}(A) + \left(\frac{q}{p}\right)^{-a} \mathbb{P}(B),$$

relazione che insieme a quella ovvia $\mathbb{P}(A) + \mathbb{P}(B) = 1$ porta alle annunciate espressioni per $\mathbb{P}(A)$ e $\mathbb{P}(B)$.

Si osservi che la successione $(T_b)_{b \in \mathbb{Z}_+}$ è crescente e, poiché $b \leq T_b$, si ha

$$\lim_{b \rightarrow +\infty} T_b = +\infty.$$

Perciò

$$\mathbb{P}(T_{-a} < +\infty) = \lim_{b \rightarrow +\infty} \mathbb{P}(T_{-a} < T_b) = \lim_{b \rightarrow +\infty} \frac{1 - \left(\frac{q}{p}\right)^b}{\left(\frac{q}{p}\right)^{-a} - \left(\frac{q}{p}\right)^b} = \left(\frac{q}{p}\right)^a.$$

Alla luce dell'ultimo risultato l'evento $\{T_{-a} = +\infty\}$ non è trascurabile, di modo che il tempo d'arresto T_{-a} non è integrabile.

Per l'Esercizio 6.5 la successione (M_n) con $M_n = S_n - n(p - q)$ è una martingala, nulla in 0 e con gli incrementi limitati. La martingala arrestata M^{T_b} è nulla in 0, sicché

$$0 = \mathbb{E}(M_{n \wedge T_b}) = \mathbb{E}[S_{n \wedge T_b} - (p - q)(n \wedge T_b)],$$

donde

$$(p - q) \mathbb{E}(n \wedge T_b) = \mathbb{E}(S_{n \wedge T_b}) \leq b.$$

Si faccia tendere n a $+\infty$ per ottenere

$$\mathbb{E}(T_b) \leq \frac{b}{p - q},$$

ciò che mostra che T_b , e, quindi, T è integrabile. Per il Lemma di Wald le due martingale arretrate M^{T_b} e M^T sono entrambe dominate in L^1 ; esse sono allora ereditarie con $M_{n \wedge T_b} = \mathbb{E}_n(M_{T_b})$ e $M_{n \wedge T} = \mathbb{E}_n(M_T)$. Da

$$\mathbb{E}(M_{T_b}) = \mathbb{E}(M_T) = 0,$$

scendono le relazioni

$$\begin{aligned} (p - q) \mathbb{E}(T_b) &= \mathbb{E}(S_{T_b}) = b, \\ (p - q) \mathbb{E}(T) &= \mathbb{E}(S_T) = b \mathbb{P}(A) - a \mathbb{P}(B), \end{aligned}$$

dalle quali segue l'asserto. □

6.7.7 L'urna di Pólya

Si supponga di avere un'urna che inizialmente contiene una pallina bianca e una pallina colorata. A ogni istante si estrare una pallina e la si rimette nell'urna insieme ad un'altra pallina dello stesso colore. Sia X_n la variabile aleatoria che assume i valori 1 e 0 secondo che, all' n -esima estrazione, si sia estratta una pallina bianca o una pallina colorata. In questo caso S_n indica il numero di *nuove* palline bianche presenti nell'urna al tempo n ; ovviamente, $S_0 = 0$. Il numero di palline bianche presenti nell'urna è $S_n + 1$.

Teorema 6.7.10. *La successione $(M_n)_{n \geq 0}$ definita da*

$$M_n := \frac{S_n + 1}{n + 2} \tag{6.7.8}$$

è una martingala.

Dimostrazione. Come in una passeggiata aleatoria, il numero di palline bianche presenti al tempo $n+1$ dipende da S_n , ma non da S_j per $j < n$. Pertanto considerata la filtrazione naturale (\mathcal{F}_n) si ha

$$\mathbb{E}_n(S_{n+1} + 1) = \mathbb{E}(S_{n+1} + 1 \mid S_n).$$

Ora

$$\mathbb{E}(S_{n+1} + 1 \mid S_n = k) = (k + 1) \frac{k + 1}{n + 2} + (n + 2 - k) \frac{k + 1}{n + 2} = (k + 1) \frac{n + 3}{n + 2},$$

sicché si può scrivere

$$\mathbb{E}_n(S_{n+1} + 1) = (S_n + 1) \frac{n + 3}{n + 2}.$$

Si è così dimostrato che $(M_n)_{n \geq 0}$ è una martingala. □

Teorema 6.7.11. Per ogni $n \geq 0$ la variabile aleatoria S_n ha legge uniforme in $\{0, \dots, n\}$.

Dimostrazione. Si proceda per induzione; l'affermazione è sicuramente vera per $n = 0$. Si supponga ora che sia vera per $n \in \mathbb{N}$, vale a dire che sia

$$\mathbb{P}(S_n = j) = \frac{1}{n+1}.$$

Per $n+1$ si ha

$$\begin{aligned} \mathbb{P}(S_{n+1} = j) &= \mathbb{P}(X_{n+1} = 1 \mid S_n = j-1) \mathbb{P}(S_n = j-1) \\ &\quad + \mathbb{P}(X_{n+1} = 0 \mid S_n = j) \mathbb{P}(S_n = j) \\ &= \frac{1}{n+1} \mathbb{P}(X_{n+1} = 1 \mid S_n = j-1) \\ &\quad + \frac{1}{n+1} \mathbb{P}(X_{n+1} = 0 \mid S_n = j) \\ &= \frac{1}{n+1} \frac{j}{n+2} + \frac{n-j+1}{n+2} \frac{1}{n+1} = \frac{1}{n+2}, \end{aligned}$$

che stabilisce l'asserto. \square

Corollario 6.7.1. La martingala del Teorema 6.7.10 converge quasi certamente a una variabile Z di legge uniforme in $[0, 1]$.

Dimostrazione. La martingala dell'eq. (6.7.8) è uniformemente limitata, e, a maggior ragione, uniformemente integrabile; in virtù dei Teoremi 6.5.3 e 6.5.4 essa converge quasi certamente a una variabile Z . Allora tende q.c. a Z anche (S_n/n) . Per ogni funzione $\varphi : [0, 1] \rightarrow \mathbb{R}$ continua il Teorema 6.7.11 assicura che si abbia

$$\mathbb{E} \left[\varphi \left(\frac{S_n}{n} \right) \right] = \frac{1}{n+1} \sum_{j=0}^n \varphi \left(\frac{j}{n} \right).$$

Si faccia tendere n a $+\infty$ per ottenere

$$\mathbb{E}[\varphi(Z)] = \lim_{n \rightarrow +\infty} \mathbb{E} \left[\varphi \left(\frac{S_n}{n} \right) \right] = \int_0^1 \varphi(x) dx;$$

l'arbitrarietà di φ dà ora l'asserto. \square

6.8 Il teorema di Burkholder

Il seguente risultato è importante per sé; noi ce ne serviremo nella dimostrazione del successivo teorema di Burkholder.

Teorema 6.8.1. (Krickeberg). Ogni martingala limitata in L^1 è la differenza di due martingale positive.

Dimostrazione. Sia (X_n) una martingala limitata in L^1 . Per ogni $n \in \mathbb{N}$, sia X_n^+ la parte positiva di X_n , $X_n^+ := X_n \vee 0$. Ovviamente si ha, per ogni $n \in \mathbb{N}$, $X_n \leq X_n^+$; di qui, per ogni $n \in \mathbb{N}$, scende

$$\mathbb{E}_n(X_{n+1}) \leq \mathbb{E}_n(X_{n+1}^+),$$

e, ricordando il Teorema 6.1.4,

$$\begin{aligned}\mathbb{E}_n (X_{n+k}^+) &= \mathbb{E}_n [\{\mathbb{E}_{n+k} (X_{n+k+1})\}^+] \\ &\leq \mathbb{E}_n \mathbb{E}_{n+k} (X_{n+k+1}^+) = \mathbb{E}_n (X_{n+k+1}^+).\end{aligned}$$

Fissato $n \in \mathbb{N}$, la successione $\{\mathbb{E}_n (X_{n+k+1}^+) : k \in \mathbb{N}\}$ è dunque crescente e positiva; pertanto, al tendere di k a $+\infty$ essa converge ad una v.a. positiva X'_n che risulta inoltre essere integrabile, perché

$$\mathbb{E} (\mathbb{E}_n (X_{n+k}^+)) = \mathbb{E} (X_{n+k}^+) \leq \mathbb{E} (|X_{n+k}|) < +\infty$$

in virtù dell'ipotesi.

Si ha perciò, per ogni $n \in \mathbb{N}$ e per ogni $k \in \mathbb{N}$,

$$X_n \leq X_n^+ = \mathbb{E}_n (X_n^+) \leq \mathbb{E}_n (X_{n+k}^+) \leq X'_n.$$

La successione (X'_n) è una martingala; infatti, applicando il teorema di convergenza monotona all'ovvia eguaglianza

$$\mathbb{E}_n (X_{n+k}^+) = \mathbb{E}_n \left(\mathbb{E}_{n+1} \left(X_{n+1+(k-1)}^+ \right) \right),$$

si ottiene facendo tendere k a $+\infty$,

$$X'_n = \mathbb{E}_n (X'_{n+1}),$$

sicché (X'_n) è una martingala positiva tale che $X'_n \geq X_n$ per ogni $n \in \mathbb{N}$.

Si ponga ora, per ogni $n \in \mathbb{N}$, $X''_n := X'_n - X_n$, che, per quanto appena detto, è positiva. Inoltre, come differenza di due martingale, (X''_n) è anch'essa una martingala. \square

Il seguente teorema è una generalizzazione del Teorema 6.5.5.

Teorema 6.8.2. (Burkholder). *Se la martingala $(X_n, \mathcal{F}_n)_{n \in \mathbb{Z}_+}$ è limitata in L^1 , cioè*

$$\sup_{n \in \mathbb{Z}_+} \|X_n\|_1 < +\infty$$

e se $(U_n, \mathcal{F}_n)_{n \in \mathbb{Z}_+}$ è un processo prevedibile, allora la martingala trasformata

$$((U \cdot X)_n)_{n \in \mathbb{Z}_+}$$

converge q.c. nell'insieme $\{\sup_{n \in \mathbb{Z}_+} |U_n| < +\infty\}$.

Dimostrazione. In virtù del teorema precedente non è restrittivo supporre che la martingala (X_n) sia positiva. Si ponga allora $X_n^* := \max_{j \leq n} X_j$ e, per $j \in \mathbb{N}$, $Z_j := X_j / X_j^*$; si definisca ora $Y_0 := 0$ e, per $n \in \mathbb{N}$,

$$Y_n := \sum_{j=1}^n \{Z_j - \mathbb{E}_{j-1} (Z_j)\}.$$

Controlliamo che (Y_n, \mathcal{F}_n) è una martingala; non vi è nulla da verificare per quanto riguarda la misurabilità di Y_n rispetto a \mathcal{F}_n e la sua integrabilità.

$$\begin{aligned} \mathbb{E}_n(Y_{n+1}) &= \mathbb{E}_n \left[\sum_{j=1}^{n+1} \{Z_j - \mathbb{E}_{j-1}(Z_j)\} \right] \\ &= \sum_{j=1}^n \{Z_j - \mathbb{E}_{j-1}(Z_j)\} + \mathbb{E}_n(Z_{n+1}) - \mathbb{E}_n(Z_{n+1}) = Y_n. \end{aligned}$$

Si noti che, per ogni indice $j \in \mathbb{N}$, è

$$\mathbb{E}_{j-1}(Z_j) \leq \mathbb{E}_{j-1} \left(\frac{X_j}{X_{j-1}^*} \right) = \frac{1}{X_{j-1}^*} \mathbb{E}_{j-1}(X_j) = Z_{j-1},$$

cioè

$$Z_{j-1} \geq \mathbb{E}_{j-1}(Z_j) \quad (6.8.1)$$

D'altro canto, è

$$\begin{aligned} E \left[\{Z_j - \mathbb{E}_{j-1}(Z_j)\}^2 \right] &= E(Z_j^2) - E \left[\{\mathbb{E}_{j-1}(Z_j)\}^2 \right] \\ &= E(Z_j^2) - E(Z_{j-1}^2) + E \left[Z_{j-1}^2 - \{\mathbb{E}_{j-1}(Z_j)\}^2 \right] \\ &= E(Z_j^2) - E(Z_{j-1}^2) + E \left[\{(Z_{j-1}) - \mathbb{E}_{j-1}(Z_j)\} \{(Z_{j-1}) + \mathbb{E}_{j-1}(Z_j)\} \right], \end{aligned}$$

e, per ogni $j \in \mathbb{N}$,

$$0 \leq Z_j \leq 1, \quad Z_{j-1} + \mathbb{E}_{j-1}(Z_j) \leq 2,$$

onde, per la (6.8.1),

$$E \left[\{Z_j - \mathbb{E}_{j-1}(Z_j)\}^2 \right] \leq E(Z_j^2) - E(Z_{j-1}^2) + 2 \{E(Z_{j-1}) - E(Z_j)\}. \quad (6.8.2)$$

Ora

$$\begin{aligned} E(Y_n^2) &= \sum_{j=1}^n E \left[(Z_j - \mathbb{E}_{j-1}(Z_j))^2 \right] \\ &\quad + \sum_{\substack{j,k=1 \\ j \neq k}}^n E \left[\{(Z_j - \mathbb{E}_{j-1}(Z_j))\} \{(Z_k - \mathbb{E}_{k-1}(Z_k))\} \right] \\ &= \sum_{j=1}^n E \left[(Z_j - \mathbb{E}_{j-1}(Z_j))^2 \right] \\ &\quad + \sum_{\substack{j,k=1 \\ j \neq k}}^n \{E(Z_j Z_k) - E(Z_j \mathbb{E}_{k-1}(Z_k)) - E(Z_k \mathbb{E}_{j-1}(Z_j)) \\ &\quad \quad + E[\mathbb{E}_{j-1}(Z_j) \mathbb{E}_{k-1}(Z_k)]\}; \end{aligned}$$

supposto che sia $j < k$, si ha

$$E(Z_j \mathbb{E}_{k-1}(Z_k)) = E(\mathbb{E}_{k-1}(Z_j Z_k)) = E(Z_j Z_k),$$

e

$$\begin{aligned}\mathbb{E}(Z_k \mathbb{E}_{j-1}(Z_j)) &= \mathbb{E}[\mathbb{E}_{k-1}(Z_k \mathbb{E}_{j-1}(Z_j))] \\ &= \mathbb{E}(\mathbb{E}_{k-1}(Z_k) \mathbb{E}_{j-1}(Z_j)) ,\end{aligned}$$

sicch 

$$\mathbb{E}(Y_n^2) = \sum_{j=1}^n \mathbb{E} \left[(Z_j - \mathbb{E}_{j-1}(Z_j))^2 \right] .$$

Pertanto, la (6.8.2) d 

$$\begin{aligned}\mathbb{E}(Y_n^2) &\leq \mathbb{E}(Z_n^2) - \mathbb{E}(Z_0^2) + 2\mathbb{E}(Z_0) - 2\mathbb{E}(Z_n) \\ &\leq \mathbb{E}(Z_n^2) + 2\mathbb{E}(Z_0) \leq 3 .\end{aligned}\tag{6.8.3}$$

Dunque, (Y_n, \mathcal{F}_n)   una martingala limitata in L^2 .

Si ponga ora, per $j \in \mathbb{N}$, $V_j := X_{j-1}^*$; per costruzione, il processo $(V_n)_{n \in \mathbb{Z}_+}$   prevedibile. Dimostriamo che

$$\begin{aligned}X_n &= V_{n+1} - \sum_{j=1}^n V_j \mathbb{E}_{j-1}(Z_{j-1} - Z_j) + (V \cdot Y)_n \\ &= X_n^* - \sum_{j=1}^n X_{j-1}^* \mathbb{E}_{j-1}(Z_{j-1} - Z_j) + (V \cdot Y)_n\end{aligned}\tag{6.8.4}$$

La (6.8.4)   di verifica banale se $n = 0$; in generale, si ha dalla definizione delle v.a. Z_j

$$\begin{aligned}\Delta X_n &:= X_n - X_{n-1} = X_n^* Z_n - X_{n-1}^* Z_{n-1} \\ &= Z_n (X_n^* - X_{n-1}^*) + X_{n-1}^* (Z_n - Z_{n-1}) ,\end{aligned}$$

e

$$Z_n (X_n^* - X_{n-1}^*) = X_n^* - X_{n-1}^* =: \Delta X_n^* .\tag{6.8.5}$$

Ci    sicuramente vero se $Z_n = 1$; se, invece, $Z_n < 1$, allora $X_n^* = X_{n-1}^*$ ed entrambi i membri della (6.8.5) sono nulli. Perci 

$$\begin{aligned}\Delta X_n &= \Delta X_n^* + X_{n-1}^* (Z_n - Z_{n-1}) \\ &= \Delta X_n^* + X_{n-1}^* \mathbb{E}_{n-1}(Z_n - Z_{n-1}) + X_{n-1}^* \{Z_n - \mathbb{E}_{n-1}(Z_n)\} \\ &= \Delta X_n^* - X_{n-1}^* \mathbb{E}_{n-1}(Z_{n-1} - Z_n) + X_{n-1}^* \{Z_n - \mathbb{E}_{n-1}(Z_n)\} ,\end{aligned}$$

ci  che dimostra la (6.8.4).

Ricorrendo ora alla (6.8.4) si pu  scrivere

$$\begin{aligned}(U \cdot X)_n &= U_0 X_0 + \sum_{j=1}^n U_j (X_j^* - X_{j-1}^*) \\ &\quad - \sum_{j=1}^n U_j X_{j-1}^* \mathbb{E}_{j-1}(Z_{j-1} - Z_j) + \sum_{j=1}^n U_j X_{j-1}^* \Delta Y_j\end{aligned}\tag{6.8.6}$$

La prima somma che compare nella (6.8.6) è convergente, al tendere di n a $+\infty$ nell'insieme

$$\{U^* \leq k\} \quad (k \in \mathbb{N})$$

perché è maggiorata da

$$k \sum_{j=1}^n (X_j^* - X_{j-1}^*) \leq k X_n^*$$

e $X^* = \sup_{n \in \mathbb{N}} X_n^* < +\infty$ q.c. dato che (X_n) è limitata in L^1 ; infatti, la disegualianza massimale (6.3.2) implica

$$\mathbb{P}(X^* = +\infty) \leq \mathbb{P}(X^* \geq \lambda) \leq \frac{1}{\lambda} \sup_{n \in \mathbb{N}} \mathbb{E}(X_n) \xrightarrow{\lambda \rightarrow +\infty} 0.$$

In definitiva, la somma in questione converge nell'insieme

$$\bigcup_{k \in \mathbb{N}} \{U^* \leq k\} = \{U^* < +\infty\}.$$

Per il termine successivo della (6.8.6) vale, per la (6.8.1),

$$\sum_{j=1}^n U_j X_{j-1}^* \mathbb{E}_{j-1}(Z_{j-1} - Z_j) \leq \sum_{j=1}^n |U_j| X_{j-1}^* \mathbb{E}_{j-1}(Z_{j-1} - Z_j),$$

che è la somma parziale n -esima di una serie a termine positivi. Poiché

$$\sum_{j=1}^n \mathbb{E}[\mathbb{E}_{j-1}(Z_{j-1} - Z_j)] = \sum_{j=1}^n \{\mathbb{E}(Z_{j-1}) - \mathbb{E}(Z_j)\} \leq 1$$

il teorema di Beppo Levi assicura che la serie in questione converge q.c. nell'insieme

$$\{U^* \leq k\} \cap \{X^* \leq h\};$$

essa, dunque, converge q.c. nell'insieme

$$\bigcup_{k, h \in \mathbb{N}} (\{U^* \leq k\} \cap \{X^* \leq h\}) = \{U^* < +\infty\}.$$

L'ultimo termine della (6.8.6) è

$$A_n := \sum_{j=1}^n U_j X_{j-1}^* \Delta Y_j;$$

e, poiché (Y_n) è una martingala, anche $(A_n, \mathcal{F}_n)_{n \in \mathbb{N}}$ è una martingala, com'è immediato controllare. Nell'insieme $\{U^* \leq k\} \cap \{X^* \leq h\}$ ($k, h \in \mathbb{N}$) essa coincide con

$$\mathbf{1}_{\{U^* \leq k\} \cap \{X^* \leq h\}} \sum_{j=1}^n U_j X_{j-1}^* \Delta Y_j \quad (6.8.7)$$

che è limitata in L^2 . Infatti, nell'insieme $\{U^* \leq k\} \cap \{X^* \leq h\}$, si ha

$$A_n^2 = \sum_{j=1}^n (U_j X_{j-1}^* \Delta Y_j)^2 + \sum_{\substack{j, k=1 \\ j \neq k}}^n U_j U_k X_{j-1}^* X_{k-1}^* \Delta Y_j \Delta Y_k.$$

Ora, per $j < k$, si ha, poiché (Y_n) è una martingala,

$$\begin{aligned} E [U_j U_k X_{j-1}^* X_{k-1}^* \Delta Y_j \Delta Y_k] &= \mathbb{E} [\mathbb{E}_{k-1} (U_j U_k X_{j-1}^* X_{k-1}^* \Delta Y_j \Delta Y_k)] \\ &= \mathbb{E} [U_j U_k X_{j-1}^* X_{k-1}^* \Delta Y_j \mathbb{E}_{k-1} (Y_k - Y_{k-1})] = 0. \end{aligned}$$

D'altro canto, abbiamo dimostrato che (Y_n) è una martingala limitata in L^2 , sicché, per il Teorema 6.1.3

$$\begin{aligned} \mathbb{E} (A_n^2) &= \sum_{j=1}^n \mathbb{E} [(U_j X_{j-1}^* \Delta Y_j)^2] \\ &\leq h^2 k^2 \sum_{j=1}^n \mathbb{E} [(\Delta Y_j)^2] = h^2 k^2 \sum_{j=1}^n \mathbb{E} [\mathbb{E}_{j-1} (\Delta Y_j)^2] \\ &= h^2 k^2 \sum_{j=1}^n \mathbb{E} [\mathbb{E}_{j-1} (Y_j^2) - Y_{j-1}^2] \\ &= h^2 k^2 \sum_{j=1}^n \{\mathbb{E} (Y_j^2) - \mathbb{E} (Y_{j-1}^2)\} \\ &= h^2 k^2 \mathbb{E} (Y_n^2) \leq 3 h^2 k^2. \end{aligned}$$

Così, la martingala (6.8.7) è limitata in L^2 e quindi converge in

$$\{U^* \leq k\} \cap \{X^* \leq h\}.$$

Perciò, come sopra, converge q.c. in $\{U^* < +\infty\}$. \square

Il teorema di Burkholder contiene come caso particolare il Teorema 6.5.5 sulla convergenza q.c. delle martingale limitate in L^1 . Basta, infatti prendere $U_n = 1$ per ogni $n \in \mathbb{N}$ per avere $(U \cdot X)_n = X_n$ per ogni $n \in \mathbb{N}$; di più, esso ne costituisce un'effettiva generalizzazione.

Si osservi che il teorema di Burkholder consente anche di dare una nuova dimostrazione del teorema di Doob; infatti, nella dimostrazione che abbiamo appena visto si è utilizzato solo il fatto che una martingala limitata in L^2 converge q.c., ma ciò è un'ovvia conseguenza dei Teoremi 6.5.2 e 6.5.4.

6.9 Note al Capitolo 6

Sezione 6.1 Il concetto di martingala fu introdotto esplicitamente da Ville (1939).

Il significato della parola “martingala” non è del tutto chiaro. La prima citazione di questo vocabolo sarebbe in Rabelais, *chausses à la martingale*. Il nome deriverebbe dal villaggio di Martigues nella Camargue. Tra gli altri significati ha quello di una parte dell'equipaggiamento di un cavallo da tiro; si veda in (Snell, 1982) la fotografia di Doob con una martingala, regalatagli dal suo primo studente Halmos. Come la parola sia venuta a significare una strategia di gioco, quella consistente nel puntare ogni volta il doppio della posta perso nella giocata precedente, non è noto. In quest'ultima accezione la parola “martingala” fu impiegata, secondo l'*Oxford Dictionary*, per la prima volta, almeno in inglese, da Thackeray nel 1854 nella frase *You have not*

played as yet? Do not do so; above all avoid a martingale if you do. Tuttavia la versione elettronica dello stesso dizionario dà una citazione precedente della parola, nello stesso significato, risalente al 1815. Benché l'uso di questa parola, in ispecie in francese, debba essere ancora anteriore, la piú antica citazione di questa parola che mi sia nota è di Giacomo Casanova (1989) nelle sue memorie scritte in francese alla fine del Settecento. Si veda [167].

La nozione di sotto-martingala fu introdotta da Snell (1952).

Le martingale divennero uno strumento fondamentale nel campo delle probabilità nei lavori di J.L. Doob, al quale è dedicata la monografia (Dellacherie & Meyer, 1980); nelle parole di questi autori, Doob *a démontré presque tous les résultats fondamentaux et . . . les a utilisés sur tous les champs de bataille du calcul des probabilités, de sorte qu'aucun probabiliste ne peut plus se permettre d'ignorer la théorie des martingales.*

La teoria della martingale si può studiare in molti libri; occorre però segnalare, tra tutti, quelli di Doob (1953) e di Dellacherie & Meyer (1980). Particolarmente lucida l'esposizione di Chatterji (1973). Molto buono per la chiarezza e la ricchezza di esempi, ai quali ho attinto liberamente, (Pintacuda, 1984), come pure, a livello piú alto, (Letta, 1984) e (Letta, 2016). Molto utili anche (Neveu, 1972), (Mazliak et al., 1999) e (Rogers & Williams, 1994).

Sono numerose le applicazioni delle martingale all'analisi; per un'introduzione a questo campo si può trovare un orientamento in (Letta & Pratelli, 1986) e (Durrett, 1991).

Le *amart* costituiscono una generalizzazione del concetto di martingala (il nome sta per *amart=asymptotic martingale*); si dice che un processo adattato (X_n) è un'amart, se considerato l'insieme diretto \mathcal{T} dei tempi d'arresto *finiti*, converge la famiglia $\{\mathbb{E}(X_T) : T \in \mathcal{T}\}$. Il concetto di amart fu introdotto da Austin et al. (1974). Si veda la monografia di Edgar & Sucheston (1992).

Un'altra generalizzazione rilevante è sicuramente quella costituita dalle martingale a valori vettoriali, il cui studio ha riflessi importanti per lo studio della geometria degli spazi di Banach; come introduzione si può consultare il Capitolo V in (Diestel & Uhl, 1977) e la bibliografia lí citata.

Sezione 6.2 L'importanza dei tempi d'arresto fu sottolineata da Doob. Ai tempi d'arresto e ai loro usi è dedicata la monografia (Egghe, 1984).

Sezione 6.3 L'importante teorema sulla trasformata di una martingala è dovuto a Burkholder (1966).

Sezione 6.4 Sull'integrabilità uniforme si veda l'ampio lavoro di Diestel (1991).

Sezione 6.5 La spiegazione della diversità della formulazione dell'ultima condizione nei Teoremi 6.5.2 e 6.5.3 trae origine dalla diversa natura delle condizioni che assicurano la compattezza relativa nella topologia debole degli spazi L^p nei casi $p > 1$ e $p = 1$; si vedano in proposito (Chatterji, 1973) e (Dunford & Schwartz, 1958).

Avvertiamo il lettore che la dimostrazione del fondamentale Teorema 6.5.5 è ancora inusuale nei libri di testo. La dimostrazione "canonica", che si può

trovare nella maggior parte dei libri di testo, a partire da (Doob, 1953), passa attraverso il teorema di Doob sul numero delle ascese (“upcrossings”) di una martingala. Per una dimostrazione di tale teorema si può consultare (Dubins, 1966). È interessante notare come la dimostrazione contenuta in (Doob, 1940) non ricorra ancora esplicitamente al concetto di ascensione. Qui abbiamo seguito essenzialmente (Lamb, 1973). Altri approcci sono stati introdotti da Isaac (1965), Báez–Duarte (1968), Chatterji (1973), Dinges (1973), Baxter (1974), Chen (1981), Al–Hussaini (1981). In alcuni libri di testo cominciano a comparire dimostrazioni differenti da quella “canonica”, per esempio in (Dudley, 1989) si segue la dimostrazione di Lamb, mentre Stromberg (1994) adatta la dimostrazione di Baxter. Ancora differente è la dimostrazione riportata da Petersen (1983), e da lui attribuita, senza citazione, a J. Horowitz.

Si è visto nella dimostrazione del Teorema 6.5.4 e del Corollario 6.5.1 che la disuguaglianza massimale (6.3.2) implica la convergenza quasi certa della successione $(X_n = \mathbb{E}_n(X))$; Chatterji (1980) mostra che la disuguaglianza massimale

- (a) non è necessariamente valida se l’insieme degli indici non è totalmente ordinato;
- (b) non è necessaria per la convergenza quasi certa di $X_n = \mathbb{E}_n(X)$.

Sottosezione 6.7.1 Del teorema di Radon–Nikodym esistono varie dimostrazioni; in queste lezioni se ne sono incontrate due, quella di von Neumann nella Sezione 1.14 e quella di questa sottosezione. Si veda l’ampia trattazione di (Rao, 1987).

Sezione 6.8 Per il Teorema 6.8.1 si veda Krickeberg (1956).

6.10 Esercizi sul Capitolo 6

6.1. Nello spazio misurabile $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ si considerino le v.a.

$$X_n := n \mathbf{1}_{B_n^c} - \mathbf{1}_{B_n},$$

ove $B_n := \{1, 2, \dots, n\}$. Si determini la filtrazione naturale.

6.2. Nello stesse condizioni e con le stesse definizioni dell’esercizio precedente si consideri in $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ la misura di probabilità definita per ogni $n \in \mathbb{N}$ da

$$\mathbb{P}(\{n\}) := \frac{1}{n(n+1)};$$

si mostri che (X_n) è una martingala.

6.3. Se (X_n) è una martingala e se

$$Y_n := n X_n - \sum_{j=1}^{n-1} X_j,$$

allora (Y_n) è anch’essa una martingala.

6.4. Sia (X_n) una successione di v.a. di $L^1(\mathcal{F})$. Se

$$Y_n := X_1 + \sum_{j=2}^n (X_j - \mathbb{E}_{j-1}(X_j)) ,$$

(Y_n) è una martingala.

6.5. Se (X_n, \mathcal{F}_n) è una martingala e se $j \leq k < n$, allora

$$\mathbb{E}[(X_n - X_k) X_j] = 0 .$$

6.6. Si consideri la passeggiata aleatoria $(X_n)_{n \in \mathbb{Z}_+}$, ove le v.a. X_n sono indipendenti e isonome con

$$\mathbb{P}(X_n = 1) = p \quad \mathbb{P}(X_n = -1) = q .$$

Si mostri che sono martingale le successioni (M_n) definita da

$$M_n := S_n - (p - q) n ,$$

e (V_n) definita da

$$W_n = \left(\frac{q}{p}\right)^{S_n} .$$

6.7. (a) Sia (X_n) una successione di v.a. indipendenti e centrate di L^1 . Si definisca, al solito, $S_n := \sum_{j=1}^n X_j$ e sia

$$T_n := \frac{S_n}{n} .$$

Allora (T_n) è una sottomartingala.

(b) Se (X_n) e (Y_n) sono due sottomartingale, è una sottomartingala anche (Z_n) ove, per ogni $n \in \mathbb{N}$, $Z_n := X_n \vee Y_n$.

6.8. In uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, sia (X_n) una successione di v.a. indipendenti ed isonome aventi legge $N(0, 1)$ e sia $\{\mathcal{F}_n\}$ la filtrazione naturale. Posto $S_n := \sum_{j=1}^n X_j$, si definisca, per ogni $n \in \mathbb{N}$,

$$Y_n := \exp\left(S_n - \frac{n}{2}\right) .$$

Allora la successione (Y_n) è una martingala.

6.9. Sia $(\Omega, \mathcal{F}, \mathbb{P})$ uno spazio di probabilità e sia μ una misura su (Ω, \mathcal{F}) . Per ogni $n \in \mathbb{N}$ sia $\{A_{nj} : j \in \mathbb{N}\}$ una partizione misurabile di Ω con $\mathbb{P}(A_{nj}) > 0$ per ogni $n, j \in \mathbb{N}$ e si consideri la tribù \mathcal{F}_n generata da $\{A_{nj} : j \in \mathbb{N}\}$. Si supponga, inoltre, che la partizione $\{A_{(n+1)j}\}$ sia un raffinamento della precedente, in altre parole, che per ogni $k \in \mathbb{N}$, l'insieme $A_{(n+1)j}$ sia contenuto in un insieme A_{nj} . Si definisca ora

$$X_n := \sum_{j \in \mathbb{N}} \frac{\mu(A_{nj})}{\mathbb{P}(A_{nj})} \mathbf{1}_{A_{nj}} .$$

Allora (X_n, \mathcal{F}_n) è una martingala.

6.10. Nello spazio di probabilità $([0, 1[, \mathcal{B}([0, 1]), \lambda)$, ove λ è la restrizione della misura di Lebesgue, si considerino le funzioni di Rademacher R_n . Se

$$X_n := \prod_{j=1}^n (1 + \alpha R_j),$$

si mostri che

- (a) (X_n) è una martingala;
- (b) se $|\alpha| < 1$, allora (X_n) è una martingala positiva.

6.11. Nello stesso spazio di probabilità dell'esercizio precedente si consideri, per ogni $n \in \mathbb{N}$, la tribú \mathcal{F}_n generata dagli intervalli

$$\mathcal{F}_n := \mathcal{F} \left(\left\{ \left[\frac{j-1}{2^n}, \frac{j}{2^n} \right] : j = 1, 2, \dots, 2^n \right\} \right).$$

Allora è una martingala la successione (X_n) , ove, per ogni $n \in \mathbb{N}$,

$$X_n := 2^n \mathbf{1}_{\left[1 - \frac{1}{2^n}, 1\right]}.$$

6.12. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia (X_n) una sottomartingala tale che sia costante la successione $(\mathbb{E}(X_n))$ delle speranze; allora (X_n) è una martingala.

6.13. Siano $(X_n)_{n \in \mathbb{Z}_+}$ e $(Y_n)_{n \in \mathbb{Z}_+}$ due martingale quadratiche rispetto alla stessa filtrazione $(\mathcal{F}_n)_{n \in \mathbb{Z}_+}$ nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$.

- (a) Per ogni $k \leq n$ si ha $\mathbb{E}_k(X_k Y_n) = X_k Y_k$;
- (b) $\mathbb{E}(X_n Y_n) - \mathbb{E}(X_0 Y_0) = \sum_{j=1}^n \mathbb{E}[(X_j - X_{j-1})(Y_j - Y_{j-1})]$.

6.14. Non tutte le successioni (X_n) che sono uniformemente integrabili sono tali che $|X_n| \leq Y$ per ogni $n \in \mathbb{N}$ ove Y è integrabile. Sia dia l'esempio di una successione (X_n) che sia uniformemente integrabile, ma per la quale non valga la condizione in esame.

6.15. Se $\alpha, \beta \geq 0$ e $p \geq 1$, allora

$$(\alpha + \beta)^p \leq 2^{p-1} (\alpha^p + \beta^p).$$

6.16. Ogni sottoinsieme limitato \mathcal{K} di L^p , con $p > 1$, costruito su uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ è uniformemente integrabile.

6.17. Si consideri lo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ con $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}([0, 1])$ e $P = \lambda$, la restrizione della misura di Lebesgue. Sia (x_n) una successione strettamente crescente di punti di $[0, 1]$ con $x_0 = 0 < x_1 < x_2 < \dots < x_n < \dots$ e $\lim_{n \rightarrow +\infty} x_n = 1$. Si prenda

$$f_n := \frac{1}{x_n - x_{n-1}} \mathbf{1}_{]x_{n-1}, x_n[} \quad (n \in \mathbb{N}).$$

La successione (f_n) è limitata in L^1 , ma non è uniformemente integrabile.

6.18. Sia X una v.a. di $L^1(\mathcal{F})$; allora, la famiglia

$$\{\mathbb{E}_{\mathcal{G}}(X) : \mathcal{G} \text{ è una sotto-tribú di } \mathcal{F}\}$$

è uniformemente integrabile.

6.19. Siano \mathcal{H}_1 e \mathcal{H}_2 due famiglie (contenute in L^1) uniformemente integrabili; è, allora, uniformemente integrabile anche la famiglia

$$\mathcal{H}_1 + \mathcal{H}_2 := \{f + g : f \in \mathcal{H}_1, g \in \mathcal{H}_2\}.$$

6.20. Sia $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ una funzione crescente e tale che

$$\lim_{x \rightarrow +\infty} \frac{\varphi(x)}{x} = +\infty.$$

Se \mathcal{K} è un insieme di L^1 per il quale esiste una costante $H > 0$ tale che, per ogni $X \in \mathcal{K}$,

$$\int \varphi \circ |X| \, d\mathbb{P} \leq H,$$

allora \mathcal{K} è uniformemente integrabile.

Questo esempio rappresenta una situazione tipica perché vale il seguente

Teorema 6.10.1. Per un sottoinsieme \mathcal{K} di L^1 sono equivalenti le proprietà:

- (a) \mathcal{K} è uniformemente integrabile;
- (b) esiste una funzione $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ pari e convessa tale che

$$\varphi(0) = 0 \quad e \quad \lim_{x \rightarrow +\infty} \frac{\varphi(x)}{x} = +\infty$$

per la quale

$$\sup_{X \in \mathcal{K}} \mathbb{E}(\varphi \circ |X|) < +\infty.$$

Per questo teorema, dovuto a de la Vallée Poussin (1915), si veda (Rao, 1981).

6.21. (a) Se (X_n, \mathcal{F}_n) è una martingala ereditaria (per ogni $n \in \mathbb{N}$ $X_n = \mathbb{E}_n(X)$ con $X \in L^1$), e se

$$X^* := \sup_{n \in \mathbb{N}} |X_n|,$$

allora vale la disuguaglianza, pure detta *massimale*: per ogni $\lambda > 0$, è

$$\mathbb{P}(X^* \geq \lambda) \leq \frac{1}{\lambda} \int_{\{X^* \geq \lambda\}} |X| \, d\mathbb{P} \leq \frac{1}{\lambda} \|X\|_1.$$

(b) Si supponga, inoltre, che $X \in L^p$ con $p > 1$; allora $X^* \in L^p$ e vale la disuguaglianza

$$\|X^*\|_p \leq q \|X\|_p$$

ove, al solito, $\frac{1}{p} + \frac{1}{q} = 1$.

6.22. Si faccia scendere la disuguaglianza di Kolmogorov (??) dalla disuguaglianza massimale (6.3.2).

6.23. Siano $(\Omega_j, \mathcal{F}_j)$ ($j = 1, 2, \dots, n$) spazi misurabili, e, per ogni j , μ_j e ν_j misure di probabilità su $(\Omega_j, \mathcal{F}_j)$ con $\mu_j \ll \nu_j$. Considerato lo spazio prodotto (Ω, \mathcal{F}) , ove

$$\Omega := \prod_{j=1}^n \Omega_j \quad \text{e} \quad \mathcal{F} := \mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \dots \otimes \mathcal{F}_n$$

sono lo spazio prodotto e la tribú prodotto, rispettivamente, e le misure prodotto

$$\mu := \mu_1 \otimes \mu_2 \otimes \dots \otimes \mu_n \quad \text{e} \quad \nu := \nu_1 \otimes \nu_2 \otimes \dots \otimes \nu_n;$$

per $j = 1, 2, \dots, n$ siano $\mu^{(j)}$ e $\nu^{(j)}$ le restrizioni di μ e ν alla tribú

$$\mathcal{F}^{(j)} := \{A \times \Omega_{j+1} \times \dots \times \Omega_n : A \in \mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \dots \otimes \mathcal{F}_j\}.$$

Allora,

- (a) μ è assolutamente continua rispetto a ν , $\mu \ll \nu$ e, per ogni $j = 1, 2, \dots, n$, $\mu^{(j)}$ è assolutamente continua rispetto a $\nu^{(j)}$, $\mu^{(j)} \ll \nu^{(j)}$, con densità

$$f_j := \frac{d\mu^{(j)}}{d\nu^{(j)}};$$

- (b) si calcoli $\mathbb{E}(f_{j+1} \mid \mathcal{F}^{(j)})$ nello spazio di probabilità $(\Omega, \mathcal{F}, \nu)$.

6.24. Se S e T sono tempi d'arresto, allora

$$\mathcal{F}_S \cap \mathcal{F}_T = \mathcal{F}_{S \wedge T}.$$

Appartengono a questa tribú gli insiemi $\{S \leq T\}$, $\{S < T\}$ e $\{S = T\}$.

6.25. Se S e T sono tempi d'arresto rispetto alla filtrazione $\{\mathcal{F}_n\}$, tale è anche $S+T$. Se k è un numero naturale, anche kT è un tempo d'arresto.

6.26. Se T è un tempo d'arresto, X_T è misurabile rispetto alla tribú \mathcal{F}_T .

6.27. Siano (X_n) una martingala quadratica, e S e T due tempi d'arresto limitati con $S \leq T \leq N$. Allora

- (a) tanto X_S quanto X_T appartengono a L^2 ;
 (b) $\mathbb{E}[(X_T - X_S)^2 \mid \mathcal{F}_S] = \mathbb{E}(X_T^2 \mid \mathcal{F}_S) - X_S^2$;
 (c) $\mathbb{E}[(X_T - X_S)^2] = \mathbb{E}(X_T^2) - \mathbb{E}(X_S^2)$.

6.28. Sia $(M_n)_{n \in \mathbb{Z}_+}$ una martingala quadratica. Allora (M_n^2) è una sottomartingala. Si consideri la decomposizione di Doob di (M_n^2) ,

$$M_n^2 = X_n + A_n,$$

dove (X_n) è una martingala e (A_n) un processo crescente. Si mostri che

$$\mathbb{E}(M_n^2) = \mathbb{E}(A_n).$$

6.29. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ siano (X_n) una successione di v.a. di L^1 e (\mathcal{F}_n) la filtrazione naturale. Se

$$\mathbb{E}_n(X_{n+1}) = \alpha X_n + \beta X_{n-1} \quad \text{per } n \in \mathbb{N}$$

con $\alpha > 0$, $\beta > 0$ e $\alpha + \beta = 1$ si cerchi $\gamma \in \mathbb{R}$ tale che

$$Y_n := \gamma X_n + X_{n-1} \quad \text{per } n \in \mathbb{N}, \quad Y_0 = X_0$$

sia una martingala rispetto alla filtrazione (\mathcal{F}_n) .

6.30. Per una successione $(X_n)_{n \in \mathbb{Z}_+}$ di L^1 adattata a una filtrazione (\mathcal{F}_n) sono equivalenti le seguenti proprietà:

- (a) (X_n) è una martingala;
- (b) per ogni tempo d'arresto limitato T si ha $\mathbb{E}(X_T) = \mathbb{E}(X_0)$.

6.31. Una sottomartingala (X_n) , positiva e limitata in L^2 , cioè

$$X_n \geq 0, \quad \mathbb{E}(X_n^2) \leq K \quad \text{per ogni } n \in \mathbb{N}, \text{ per un opportuno } K > 0,$$

converge in L^2 .

6.32. (a) Sia (X_n, \mathcal{F}_n) una sottomartingala limitata in L^1 ($\|X_n\|_1 \leq K$ per ogni $n \in \mathbb{N}$). Si mostri che la successione di integrali

$$\left(\int_A X_n \, d\mathbb{P} \right)$$

ammette limite finito per ogni insieme $A \in \mathcal{F}_n$ e che la funzione d'insieme $\nu_n : \mathcal{F}_n \rightarrow \mathbb{R}$, definita mediante

$$\nu_n(a) := \lim_{\substack{k \rightarrow n \\ k \geq n}} \int_A X_k \, d\mathbb{P} \quad A \in \mathcal{F}_n,$$

è una misura su \mathcal{F}_n finita e assolutamente continua rispetto alla restrizione P_n di P a \mathcal{F}_n .

(b) Esiste una martingala (U_n) rispetto alla stessa filtrazione (\mathcal{F}_n) tale che $X_n^+ \leq U_n$ per ogni $n \in \mathbb{N}$.

(c) Infine sia $(-X_n, \mathcal{F}_n)$ una martingala; sono allora equivalenti le affermazioni:

- (c1) $\sup_{n \in \mathbb{N}} \mathbb{E}(|X_n|) < +\infty$;
- (c2) vale la rappresentazione $X_n = U_n - V_n$ ($n \in \mathbb{N}$) ove (U_n) e (V_n) sono martingale *positive* rispetto alla filtrazione (\mathcal{F}_n) .

6.33. Si consideri una passeggiata aleatoria di Bernoulli. Con la notazione usuale, si mostri che per ogni $j > 0$ (ma anche per ogni $j < 0$ il tempo di primo passaggio per la posizione $x = j$,

$$T_j := \inf\{n \in \bar{\mathbb{N}} : G_n = j\}$$

è un tempo d'arresto q.c. finito.

6.34. Siano (X_n) una successione di v.a. indipendenti, isonome e centrate di $L^2(\mathcal{F})$, (\mathcal{F}_n) la filtrazione naturale e T un tempo d'arresto rispetto alla filtrazione (\mathcal{F}_n) integrabile. Posto, al solito, $S_n := \sum_{j=1}^n X_j$, si mostri che

- (a) che le v.a. $Y_n := X_n \mathbf{1}_{\{T \geq n\}}$ sono in L^2 e che sono ortogonali, vale a dire $\mathbb{E}(Y_k Y_n) = 0$ se $k \neq n$;
- (b) che la serie $\sum_{n \in \mathbb{N}} Y_n$ è convergente in L^2 ;
- (c) che $S_T = \sum_{n \in \mathbb{N}} Y_n$;
- (d) che vale la relazione, detta *teorema di Blackwell–Girsbick*,

$$\mathbb{E}(S_T^2) = \mathbb{E}(X_1^2) \mathbb{E}(T)$$

e che S_T è centrata.

6.35. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia $(X_n, \mathcal{F}_n)_{n \in \mathbb{Z}_+}$ una martingala limitata, $|X_n| \leq H$ per ogni $n \in \mathbb{Z}_+$. Si ponga

$$Y_n := \sum_{j=1}^n \frac{1}{j} (X_j - X_{j-1});$$

allora

- (a) $(Y_n, \mathcal{F}_n)_{n \in \mathbb{Z}_+}$ è una martingala;
- (b) $(Y_n, \mathcal{F}_n)_{n \in \mathbb{Z}_+}$ converge tanto q.c. quanto in L^2 .

6.36. Sia P la probabilità dell'esercizio (6.2) e (X_n) la martingala dello stesso esercizio, con $B_n := \{1, 2, \dots, n\}$. Si definisca una funzione $Y : \mathbb{N} \rightarrow \mathbb{Z}$ mediante

$$Y(k) := \begin{cases} k+1, & \text{se } k \text{ è pari,} \\ -k, & \text{se } k \text{ è dispari,} \end{cases}$$

e si ponga

$$Z_n(k) := Y(k) \mathbf{1}_{B_n}(k) + c_n \mathbf{1}_{B_n^c}(k),$$

ove $c_n = 1$ o $c_n = 0$ secondo che n sia dispari o pari, rispettivamente. Si mostri che, se $U_n = (-1)^n$ ($n \in \mathbb{N}$), allora

- (a) (Z_n) è la trasformata di (X_n) mediante (U_n) ,

$$Z_n = (U \cdot X)_n,$$

- (b) Z_n converge q.c. ma non è limitata in L^1 , sicché non si può usare il teorema di Doob.

6.37. Un'urna contiene inizialmente 2 palline, delle quali una è bianca e l'altra è nera. Ad ogni istante, si estrae una pallina e la si sostituisce con 2 dello stesso colore. Pertanto al tempo $t = n$ l'urna conterrà $n + 2$ palline, delle quali $B_n + 1$ sono bianche (B_n è il numero di palline bianche estratte sino al tempo $t = n$).

(a) Si mostri che

$$\mathbb{P}(B_n = k) = \frac{1}{n+1} \quad (k = 0, 1, \dots, n);$$

(b) posto

$$X_n := \varphi(n) (B_n + 1),$$

ove $\varphi(n)$ è un numero reale (che dipende da n), si calcoli $\varphi(n)$ in modo che $(X_n)_{n \in \mathbb{N}}$ sia una martingala rispetto alla filtrazione $\{\mathcal{F}_n : n \in \mathbb{N}\}$, ove \mathcal{F}_n è la tribù generata dalle variabili aleatorie B_1, B_2, \dots, B_n ;

(c) si mostri che con il valore di $\varphi(n)$ determinato sopra, la martingala (X_n) converge quasi certamente;

(d) si determini la legge della variabile aleatoria X , limite quasi certo di (X_n) .

6.38. Sia (X_n) una successione di variabili aleatorie integrabili a valori in \mathbb{Z} , indipendenti e con la stessa legge. Si supponga che, per ogni $n \in \mathbb{N}$, sia

$$\mathbb{E}(X_n) = m < 0, \quad \mathbb{P}(X_n = 1) > 0, \quad \mathbb{P}(X_n \geq 2) = 0.$$

Si ponga $S_0 := 0$ e, per $n \geq 1$, $S_n := \sum_{j=1}^n X_j$. Posto

$$W := \sup_{n \geq 0} X_n,$$

(a) si mostri che $W < +\infty$ q.c.;

(b) definita la trasformata di Laplace di X_1 ,

$$M(s) := \mathbb{E}(e^{sX_1}),$$

e posto $\psi(s) := \ln M(s)$, si mostri che è convessa la funzione ψ così definita;

(c) si mostri che $\psi(s) < +\infty$ per ogni $s \geq 0$, si calcolino il limite

$$\lim_{s \rightarrow +\infty} \psi(s)$$

e il valore della derivata $\psi'(0)$ e si mostri che esiste un unico punto $s_0 > 0$ tale che sia $\psi(s_0) = 0$;

(d) se, per ogni $n \geq 1$, $Z_n := \exp(s_0 S_n)$, ove s_0 è stato determinato in (c), si mostri che (Z_n) è una martingala rispetto alla filtrazione naturale e si calcoli il limite quasi certo di $\{Z_n\}$ per $n \rightarrow +\infty$;

(e) sia k un numero naturale, $k \in \mathbb{N}$ e sia τ_k il primo istante in cui (S_n) assume il valore k ,

$$\tau_k := \inf\{n \in \mathbb{N} \cup \{+\infty\} : S_n = k\}.$$

Si dimostri che, quasi certamente,

$$\lim_{n \rightarrow +\infty} Z_{n \wedge \tau_k} = e^{s_0 k} \mathbf{1}_{\{\tau_k < +\infty\}};$$

(f) si calcoli $\mathbb{P}(\tau_k < \infty)$ e si determini la legge di W .

6.39. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia (X_n) una successione di v.a. di legge $N(0, \sigma^2)$, con $\sigma > 0$. Sia $\{\mathcal{F}_n\}$ la filtrazione naturale e, per ogni $n \in \mathbb{N}$ di ponga, al solito, $S_n := \sum_{j=1}^n X_j$. Sia ψ la funzione generatrice dei momenti di X_1 ,

$$\psi(t) := \mathbb{E}[\exp(t X_1)] = \exp\left(\frac{1}{2} \sigma^2 t^2\right),$$

e si ponga, per ogni $n \in \mathbb{N}$,

$$Z_n^t := \exp\left(t S_n - \frac{1}{2} n \sigma^2 t^2\right).$$

- (a) Per ogni $t \in \mathbb{R}$, (Z_n^t) è una martingala rispetto alla filtrazione naturale;
- (b) si mostri che, per ogni $t \in \mathbb{R}$, (Z_n^t) converge q.c., ma non in L^1 .

6.40. Sia $(\mathcal{F}_n)_{n \in \mathbb{N}}$ una filtrazione nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$; per una successione adattata $(X_n, \mathcal{F}_n)_{n \in \mathbb{N}}$ di v.a. di L^1 sono equivalenti le seguenti due affermazioni:

- (a) $(X_n, \mathcal{F}_n)_{n \in \mathbb{N}}$ è una sottomartingala;
- (b) vale la seguente *decomposizione moltiplicativa*

$$\forall n \in \mathbb{N} \quad X_n = A_n M_n,$$

ove $(M_n, \mathcal{F}_n)_{n \in \mathbb{N}}$ è una martingala e (A_n) è un processo prevedibile crescente tale che $A_1 := 1$: per ogni $n \in \mathbb{N}$, $A_{n+1} \geq A_n$ e A_n è \mathcal{F}_n -misurabile.

Bibliografia

- [1] W.J. Adams (2009), *The life and times of the central limit theorem*, American Mathematical Society,
- [2] A.D. Alexandrov (1940), Additive set-functions in abstract spaces. I, *Mat. Sbornik N.S.* **8**, 307–348.
- [3] A.D. Alexandrov (1941), Additive set-functions in abstract spaces. II, *Mat. Sbornik N.S.* **9**, 563–628.
- [4] A.D. Alexandrov (1943), Additive set-functions in abstract spaces. III, *Mat. Sbornik N.S.* **13**, 196–238.
- [5] A.N. Al-Hussaini (1981), A projective limit view of L^1 -bounded martingales, *Rev. Roumaine Math. Pures Appl.* **26**, 51–54.
- [6] T. Ando (1966), Contractive projections in L^p spaces, *Pacific J. Math.* **17**, 391–405.
- [7] A. Araujo, E. Giné (1980), *The central limit theorem for real and Banach valued random variables*, Wiley, New York.
- [8] R.B. Ash, *Information theory*, Wiley-Interscience, New York, 1965; ristampa, Dover, New York, 1990.
- [9] R.B. Ash (1972) *Real analysis and probability*, Academic Press, New York.
- [10] D.G. Austin, G.A. Edgar, A. Ionescu Tulcea (1974) Pointwise convergence in terms of expectations, *Z. Wahrscheinlich. verw. Gebiete* **30**, 17–26.
- [11] L. Báez-Duarte (1968), Another look at the martingale theorem, *J. Math. Anal. Appl.* **85**, 551–559.
- [12] R.R. Bahadur (1955), Measurable subspaces and subalgebras, *Proc. Amer. Math. Soc.* **6**, 565–570.
- [13] S. Banach, C. Kuratowski (1929), Sur une généralisation du problème de la mesure, *Fund. Math.* **14**, 127–131.
- [14] H. Bauer (1981), *Probability theory and elements of measure theory*, Academic Press, New York–London.
- [15] H. Bauer (1996), *Probability theory*, de Gruyter, Berlin–New York.

- [16] J.R. Baxter (1974), Pointwise in terms of weak convergence, *Proc. Amer. Math. Soc.* **46**, 395–398.
- [17] K.P.S. Bhaskara Rao, M. Bhaskara Rao (1983), *Theory of charges. A study of finitely additive measures*, Academic Press, London–New York.
- [18] P. Billingsley (1979), *Probability and measure*, Wiley, New York; terza edizione, 1995.
- [19] S. Bochner (1932), *Vorlesungen über Fouriersche Integrale*, Akademische Verlagsgesellschaft, Leipzig; ristampa, Chelsea, New York, 1948.
- [20] V.I. Bogachev, *Measure theory*, Springer, Berlin–New York, 2007
- [21] N. Bourbaki (1965) *Éléments de mathématique. Livre VI: Intégration. Chapitres 1,2,3 et 4*, (2nde éd.) Hermann, Paris.
- [22] L. Breiman (1968) *Probability*, Addison-Wesley, Reading MA.
- [23] J.R. Brown (1976), *Ergodic theory and topological dynamics*, Academic Press, New York.
- [24] D.L. Burkholder (1966), Martingale transforms, *Ann. Math. Statist.* **37**, 1494–1504.
- [25] S.D. Chatterji (1973), Les martingales et leurs applications analytiques, in *École d'été de probabilités: processus stochastiques*, J.L. Bretagnolle, S.D. Chatterji, P.-A. Meyer, Eds., Springer (Lecture Notes in Mathematics 307), Berlin – Heidelberg – New York, pp. 27–164.
- [26] S.D. Chatterji (1980), Some comments on the maximal inequality in martingale theory. *Measure theory, Oberwolfach 1979 (Proc. Conf., Oberwolfach, 1979)*, *Lecture Notes in Math.* **794**, Springer, Berlin, pp. 361-364; Erratum, *Measure theory, Oberwolfach 1981 (Oberwolfach, 1981)*, *Lecture Notes in Math.* **945**, Springer, Berlin–New York, 1982, p. 431.
- [27] S.D. Chatterji (1986), Elementary counter-examples in the theory of double integrals, *Atti Sem. Mat. Fis. Univ. Modena* **34**, 17–38.
- [28] L.H.Y. Chen (1981), Martingale convergence via the square function, *Proc. Amer. Math. Soc.* **83**, 125–127.
- [29] Y. S. Chow, H. Teicher (1978) *Probability theory. Independence, interchangeability, martingales*, Springer, New York–Berlin.
- [30] K. L. Chung (1968), *A Course in probability theory*, Harcourt Brace, New York; seconda edizione, Academic Press, New York–London, 1974.
- [31] H. Cramér (1937), *Random variables and probability distributions*, Cambridge University Press.
- [32] H. Cramér (1938), Sur un nouveau théorème–limite de la théorie des probabilités, *Actual. Sci. Indust.* **736**, 5–53; anche in (Cramér, 1994), pp. 895–913.

- [33] H. Cramér (1946), *Mathematical methods of statistics*, Princeton University Press.
- [34] H. Cramér (1994), *Collected works*, A. Martin-Löf, Ed., Springer, Berlin–Heidelberg–New York.
- [35] I. Csiszár, J. Körner (1986), *Information theory. Coding theorems for discrete memoryless systems*, Akadémiai Kiadó, Budapest and Academic Press, Orlando FL.
- [36] C. Dellacherie, P.A. Meyer (1980), *Probabilités et potentiel. Chapitres V à VIII. Théorie des martingales*, Hermann, Paris.
- [37] A. Dembo, O. Zeituni (1998), *Large deviations techniques and applications*, Springer, New York.
- [38] J. Diestel (1991), Uniform integrability: an introduction, *Rend. Ist. Mat. Univ. Trieste* **23**, 41–80.
- [39] J. Diestel, J.J. Uhl jr. (1977) *Vector measures*, American Mathematical Society (Mathematical Surveys 15), Providence, RI.
- [40] J. Dieudonné (1978.a) *Abrégé d'histoire des mathématiques 1700–1900, Vol. II*, Hermann, Paris.
- [41] J. Dieudonné (1978.b) *Intégration et mesure*, in Dieudonné (1978.a), pp. 267–277.
- [42] H. Dinges (1970), Inequalities leading to a proof of the classical martingale convergence theorem, in *Martingales*, Springer (Lecture Notes in Mathematics 190), Berlin–Heidelberg–New York, pp. 9–12.
- [43] J. Dixmier (1981), *Topologie générale*, Presses Universitaires de France, Paris; traduzione inglese *General topology*, Springer, New York, 1984.
- [44] J.L. Doob (1940), Regularity properties of certain families of chance variables, *Trans. Amer. Math. Soc.* **47**, 455–486.
- [45] J.L. Doob (1953), *Stochastic processes*, Wiley, New York.
- [46] J.L. Doob (1994.a) *Measure theory*, Springer (GTM 143), New York–Berlin–Heidelberg.
- [47] R.G. Douglas (1965), Contractive projections on an L^1 space, *Pacific J. Math.* **15**, 443–462.
- [48] L.E. Dubins (1969), A note on upcrossings of martingales, *Ann. Math. Statist.* **37**, 728.
- [49] R.M. Dudley (1966), Weak convergence of measures on nonseparable metric spaces and empirical measures on Euclidean spaces, *Illinois J. Math.* **10**, 109–126.

- [50] R.M. Dudley (1967), Measures on nonseparable metric spaces, *Illinois J. Math.* **11**, 449–453.
- [51] R.M. Dudley (1989), *Real analysis and probability*, Wadsworth & Brooks/Cole, Pacific Grove CA.
- [52] D. Dugué (1955), L'existence d'une norme est incompatible avec la convergence en probabilité, *C.R. Acad. Sci. Paris* **240**, 1307–1308.
- [53] N. Dunford, J.T. Schwartz (1958) *Linear operators. I*, Wiley Interscience, New York.
- [54] R. Durrett (1991), *Probability: theory and examples*, Duxbury Press, Belmont CA.
- [55] G.A. Edgar, L. Sucheston (1992), *Stopping times and directed processes*, (Encyclopedia of mathematics and Its Applications 47), Cambridge University Press.
- [56] D.A. Edwards (1990), A proof of the Lévy–Cramér continuity theorem for probability measures, *Expo. Math.* **8**, 185–192.
- [57] L. Egghe (1984), *Stopping times techniques for analysts and probabilists*, (London Mathematical Society Lecture Notes Series 100), Cambridge University Press.
- [58] N. Etemadi (1981), An elementary proof of the strong law of large numbers, *Z. Wahrscheinlich. verw. Gebiete* **55**, 119–122.
- [59] P.J.L. Fatou (1906), Séries trigonométriques et séries de Taylor, *Acta Math.* **30**, 335–340.
- [60] W. Feller (1935) Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung, *Mat. Zeit.* **40**, 521–559.
- [61] W. Feller (1971), *An Introduction to probability theory and its applications. vol II*, Wiley, New York (2nd. ed.).
- [62] M. Fréchet (1907), Sur les ensembles de fonctions et les opérateurs linéaires, *C.R. Acad. Sci. Paris* **144**, 1414–1416.
- [63] M. Fréchet (1915), Sur l'intégrale d'une fonctionnelle étendue à un ensemble abstrait, *Bull. Soc. Math. France* **43**, 249–267.
- [64] B. Fristedt, L. Gray (1997) *A modern approach to probability theory*, Birkhäuser, Boston.
- [65] A. Fuchs, G. Letta (1984), *Boll. Unione Mat. Ital.*, (6–B) **3** 451–461.
- [66] B.V. Gnedenko, A.N. Kolmogorov (1954) *Limit distribution for sums of independent random variables*, revised ed., Addison–Wesley, Reading MA, 1968.

- [67] G.R. Grimmett, & D.R. Stirzaker, *Probability and random processes*, Oxford University Press, 2001 (3rd ed.).
- [68] P. Hall (1981), A comedy of errors: the canonical form for a stable characteristic function, *Bull. London Math. Soc.* **13**, 23–27.
- [69] P.R. Halmos & L.J. Savage (1949), Application of the Radon–Nikodym theorem to the theory of sufficient statistics, *Ann. Math. Statist.* **20**, 225–241; anche in (Halmos, 1983), pp. 95–111 e in (Savage, 1981), pp. 163–179.
- [70] P.R. Halmos (1950) *Measure theory*, Van Nostrand Reinhold, New York; ristampa, Springer, New York (GTM 15).
- [71] P.R. Halmos (1983) *Selecta: research contributions*, D.E Sarason, N.A Friedman, Eds., Springer, New York.
- [72] G. Hardy, J.E. Littlewood, G. Pólya (1934), *Inequalities*, Cambridge University Press; 2nd ed., 1952.
- [73] T. Hawkins (1975) *Lebesgue's theory of integration*, Chelsea, New York.
- [74] E. Helly (1921), Über lineare Funktionaloperationen, *Sitzungsberichte der Naturwiss. Klasse Akad. Wiss. Wien* **121**, 265–295.
- [75] O. Hölder (1889), Über einen Mittelwertsatz, *Nach. Akad. Wiss. Göttingen Math. Phys. Kl.* 38–47.
- [76] A. Horn, A. Tarski (1948), Measures on boolean algebras, *Trans. Amer. Math. Soc.* **64**, 467–497.
- [77] P.L. Hsu, H. Robbins (1947), Complete convergence and the law of large numbers, *Proc. Nat. Acad. Sci. U.S.A.* **33**, 25–31; anche in (Hsu, 1983), pp. 229–240.
- [78] R. Isaac (1965), A proof of the martingale convergence theorem, *Proc. Amer. Math. Soc.* **16**, 842–844.
- [79] J. Jacod, Ph. Protter (2000) *Probability essentials*, Springer, Berlin–Heidelberg–New York.
- [80] J.L.W.V. Jensen (1906), Sur les fonctions convexes et les inégalités entre les valeurs moyennes, *Acta Math.* **30**, 175–193.
- [81] J.L. Kelley (1955), *General topology*, Van Nostrand, New York; ristampa, Springer, New York–Heidelberg–Berlin.
- [82] A. Ya. Khinchin (1937), Una nuova derivazione di una formula del Sig. P. Lévy, *Bull. Moskov. Gos. Univ.* **1**, 1–5 (in russo).
- [83] A.Ya. Khinchin, P. Lévy (1936), Sur les lois stables, *C.R. Acad. Sci. Paris* **202**, 374–376; anche in (Lévy, 1973–1980) vol. III, pp. 345–346.
- [84] J.F.C. Kingman, S.J. Taylor, *Introduction to measure and probability*, Cambridge University Press, 1966.

- [85] A. Klenke (2008), *Probability theory. A comprehensive treatise*, Springer, London.
- [86] A.N. Kolmogorov, Über die Summen durch den Zufall bestimmter unabhängiger Größen, *Mat. Ann.* **99**, 309–319, **102**, 484–488 (1928); anche in (Kolmogorov, 1992) pp. 15–31.
- [87] A. N. Kolmogorov, Sulla forma generale di un processo stocastico omogeneo, *Atti. R. Accad. Lincei Rend. Cl. Sci. Fis. Mat. Nat. (6)* **15**, 805–808, 866–869 (1932); traduzione inglese in (Kolmogorov, 1992) pp. 121–127.
- [88] A.N. Kolmogorov (1992) *Selected works of A.N. Kolmogorov. Vol. II: Probability theory and mathematical statistics*, Kluwer, Dordrecht.
- [89] L. B. Korolov, Ya. G. Sinai (2007), *Theory of probability and random processes*, 2nd ed., Springer, Berlin–Heidelberg–New York.
- [90] S. Kotz, N.L Johnson (1982–1988), *Encyclopedia of statistical sciences*, Wiley, New York.
- [91] K. Krickeberg (1965), *Probability theory*, Addison–Wesley, Reading MA.
- [92] S. Kullback, *Information theory and statistics*, Dover, New York, 1968.
- [93] Ky Fan (1944), Entfernung zweier zufälliger Größen und die Konvergenz nach Wahrscheinlichkeit, *Math. Z.* **49**, 681–683.
- [94] R.G. Laha, V.K. Rohatgi (1979), *Probability theory*, Wiley, New York.
- [95] C.W. Lamb (1973), A short proof of the martingale convergence theorem, *Proc. Amer. Math. Soc.* **38**, 215–217.
- [96] H. Lebesgue (1902), Intégrale, longueur, aire, *Ann. Mat. Pura Appl. (3)* **7**, 231–359.
- [97] L. Le Cam (1986), The central limit theorem around 1935, *Statistical Science* **1**, 78–96.
- [98] G. Letta (1984), *Martingales et intégration stochastique*, Quaderni della Scuola Normale Superiore, Pisa.
- [99] G. Letta (2016), *Elementi di teoria dei processi stocastici*, Quaderni dell’Unione Matematica Italiana 56, Unione Matematica Italiana, Bologna.
- [100] G. Letta, M. Pratelli (1986), *Probability and analysis. Varenna (Como) 1985*, Berlin–Heidelberg: Springer (LNM1206).
- [101] G. Letta, L. Pratelli (1997), Le théorème de Skorohod pour des lois de Radon sur un espace métrisable, *Rend. Accad. Naz. Sci. XL* **21**, 157–162.
- [102] B. Levi (1906), Sopra l’integrazione delle serie, *Rend. Ist. Lombardo Sci. Lett. (2)* **39**, 775–780.

- [103] P. Lévy (1922.a), Sur le rôle de la lois de Gauss dans la théorie des erreurs, *C.R. Acad. Sci. Paris* **174**, 855–857; anche in (Lévy, 1973–1980) vol. III, pp. 9–11.
- [104] P. Lévy (1922.b), Sur la lois de Gauss, *C.R. Acad. Sci. Paris* **174**, 1682–1684; anche in (Lévy, 1973–1980) vol. III, pp. 12–13.
- [105] P. Lévy (1922.c), Sur la détermination des lois de probabilité par leurs fonctions caractéristiques, *C.R. Acad. Sci. Paris* **175**, 854–856; anche in (Lévy, 1973–1980) vol. III, pp. 333–335.
- [106] P. Lévy (1924), Théorie des erreurs. La loi de Gauss et les lois exceptionnelles, *Bull. Soc. Math. France* **52**, 49–85; anche in (Lévy, 1973–1980) vol. III, pp. 14–49.
- [107] P. Lévy (1934), Sur les intégrales dont les éléments sont des variables léatoires indépendentes, *Ann. R. Scuola Norm. Sup. Pisa* (2) **3**, 337–366; anche in (Lévy, 1973–1980) vol. IV, pp. 9–38.
- [108] P. Lévy (1935), Observation sur un précédent mémoire de l’auteur, *Ann. R. Scuola Norm. Sup. Pisa* (2) **4**, 217–218; anche in (Lévy, 1973–1980) vol. IV, pp. 39–40.
- [109] P. Lévy (1937), Distance de deux variables aléatoires et distance de deux lois de probabilités, *Calcul des Probabilités et ses Applications* **1** 331–337; anche in (Lévy, 1973–1980) vol. IV, pp. 39–40.
- [110] P. Lévy (1973–1980), *Œuvres de Paul Lévy*, Gauthier–Villars, Paris.
- [111] J.W. Lindeberg (1922), Über das Exponentialgesetze in der Wahrscheinlichkeitsrechnung, *Math. Z.* **15**, 211–225.
- [112] M. Loève (1963), *Probability theory*, Van Nostrand, New York; quarta edizione in due volumi *Probability Theory I, II*, Springer, New York–Heidelberg–Berlin, 1977–78.
- [113] E. Lukacs (1970), *Characteristic functions*, II ed., Griffin, London.
- [114] E. Lukacs (1975a), *Stochastic convergence*, Academic Press, New York–San Francisco–London.
- [115] A.M Lyapunov (1901), Nouvelle forme du théorème sur la limite des probabilités, *Mem. Acad. Sci. St. Petersburg* **12**, No. 5, 1–24.
- [116] E. Marczewski (1955), Remarks on the convergence of measurable sets and measurable functions, *Colloq. Math.* **3**, 118–124.
- [117] J. Marcinkiewicz, A. Zygmund (1937), Sur les fonctions indépendantes, *Fund. Math.* **29**, 60–90.
- [118] L. Mazliak, P. Priouret, P. Baldi (1999), *Martingales et chaînes de Markov*, Hermann, Paris.

- [119] M. Métivier (1968), *Notions fondamentales de théorie des probabilités*, Dunod, Paris.
- [120] H. Minkowski (1896), *Geometrie der Zahlen*, Teubner, Leipzig.
- [121] D.S. Mitrinović, I.B. Lacković (1985), Hermite and convexity, *Aequationes Math.* **28**, 229–232.
- [122] S.T.C. Moy (1954), Characterizations of conditional expectation as a transformation of function spaces, *Pacific J. Math.* **4**, 47–63.
- [123] J. von Neumann (1940), On rings of operators, *Ann. Math.* **41** 94–161; anche in (von Neumann, 1961), Vol. III, pp. 161–228.
- [124] J. von Neumann (1961), *Collected works*, Pergamn Press, Oxford–London.
- [125] J. Neveu (1964), *Bases mathématiques du calcul des probabilités*, Masson, Paris.
- [126] J. Neveu (1972), *Martingales à temps discret*, Masson, Paris.
- [127] O. Nikodym (1930), Sur une généralisation des mesures de M. J. Radon, *Fund. Math.* **15**, 131–179.
- [128] M.P. Olson (1965), Characterization of conditional probability, *Pacific J. Math.* **15**, 971–983.
- [129] E. Pap, ed. (2002), *Handbook of measure theory, Vol. I & II*, Elsevier North–Holland, Amsterdam.
- [130] K.R. Parthasarathy (1967), *Probability measures on metric spaces*, Academic Press, New York.
- [131] E. Pascali, C. Sempì (1997), Two Lévy–type metrics for distribution functions, *Ric. Mat.* **46**, 49–60.
- [132] K. Petersen (1983), *Ergodic theory*, Cambridge University Press.
- [133] V.V. Petrov (2004), A generalization of the Borel–Cantelli lemma, *Statist. Probab. Lett.* **67**, 233–239.
- [134] J. Pfanzagl (1967), Characterizations of conditional expectations, *Ann. Math. Statist.* **38**, 415–421.
- [135] J.–P. Pier (1994.a) *Development of mathematics 1900–1950*, Birkhäuser, Basel–Boston–Berlin.
- [136] J.–P. Pier (1994.b), *Intégration et mesure 1900–1950*, in (Pier, 1994.a), pp. 517–564.
- [137] N. Pintacuda (1984) *Secondo corso di probabilità. La teoria matematica*, Muzio, Padova.
- [138] N. Pintacuda (1989), On Doob’s measurability lemma, *Bull. Un. Mat. Ital. A* (7) **3**, 237–241.

- [139] D. Pollard (1984), *Convergence of stochastic processes*, Springer, New York.
- [140] G. Pólya, Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem, *Math. Z* **8**, 171–181 (1920).
- [141] U.V. Prokhorov (1956), Convergence of random processes and limit theorems in probability, *Teor. Veroyatnost. i Primenem.* **1**, 177–238.
- [142] J. Radon (1913), Theorie und Anwendungen der absolut additiven Mengenfunktionen, *Sitzungs Ber. der math. naturwiss. Klasse der Akad. der Wiss. (Wien)* **122**, 1295–1438.
- [143] M.M. Rao (1965), Conditional expectations and closed projections, *Nederl. Akad. Wetensch. Proc. Ser. A* **68**=*Indag. Math.* **27**, 100–112.
- [144] M.M. Rao (1981), *Foundations of stochastic analysis*, Academic Press, New York.
- [145] M.M. Rao (1984) *Probability theory with applications*, Academic Press, New York–London.
- [146] M.M. Rao (1987), *Measure theory and integration*, Wiley, New York.
- [147] E. Regazzini, Leggi dei grandi numeri e dintorni. Annotazioni preliminari, *Boll. Un. Mat. Ital.* **8–A**, 1–22 (2005).
- [148] E. Regazzini, Leggi dei grandi numeri e dintorni. Risultati classici, *Boll. Un. Mat. Ital.* **9–A**, 89–130 (2006).
- [149] A. Rényi (1953), On the theory of order statistics, *Acta Math. Acad. Sci. Hung.* **4**, 191–231; anche in (Turán, 1976), Vol. 1, pp. 328–364.
- [150] A. Rényi (1966), *Calcul des probabilités*, Dunod, Paris.
- [151] F. Riesz (1907), Sur une espèce de géométrie analytique des fonctions sommables, *C.R. Acad. Sci. Paris* **144**, 1409–1411.
- [152] F. Riesz (1909), Sur les suites de fonctions mesurables, *C.R. Acad. Sci. Paris* **148**, 1303–1305.
- [153] F. Riesz (1910), Untersuchungen über Systeme integrierbarer Funktionen, *Math. Ann.* **69**, 449–497.
- [154] L.J. Rogers (1888), An extension of a certain theorem in inequalities, *Messenger Math.* **17**, 145–150.
- [155] L.C.G. Rogers, D. Williams (1994), *Diffusions, Markov processes and martingales. Vol. 1: Foundations*, Wiley, Chichester (2nd ed.).
- [156] J.P. Romano, A.F. Siegel (1986) *Counterexamples in probability and statistics*, Wadsworth & Brooks/Cole, Monterey CA.
- [157] G.C. Rota (1960), On the representation of averaging operators, *Rend. Sem. Mat. Univ. Padova* **30**, 52–64.

- [158] L.J. Savage (1981), *The writings of Leonard Jimmie Savage — A memorial selection*, The American Statistical Association and the Institute of Mathematical Statistics.
- [159] H. Scheffé (1947), A useful convergence theorem for probability distributions, *Ann. Math. Statist.* **18**, 434–438.
- [160] E. Schmidt (1908), Über die Auflösung lineare Gleichungen mit unendlich vielen Unbekannten, *Rend. Circ. Mat. Palermo* **25**, 53–77.
- [161] B. Schweizer, A. Sklar (1983), *Probabilistic metric spaces*, Elsevier North–Holland, New York.
- [162] R. Scozzafava (1981), Un esempio concreto di probabilità non σ -additiva: la distribuzione della prima cifra significativa dei dati statistici, *Boll. Un. Mat. Ital.* (5) **18–A**, 403–410.
- [163] C. Sempi (1982), On the space of distribution functions, *Riv. Mat. Univ. Parma* (4) **8**, 243–250.
- [164] C. Sempi (1986a), Orlicz metrics for weak convergence of distribution functions, *Riv. Mat. Univ. Parma* (4) **12**, 289–292.
- [165] C. Sempi (1986b), *Le speranze condizionate e le loro caratterizzazioni*, Quaderni del Dipartimento dell’Università di Lecce.
- [166] C. Sempi (1989), Variations on a theme by Scheffé, *Rend. Ist. Mat. Univ. Trieste* **21**, 219–223.
- [167] C. Sempi (2016), Martingala: l’origine di un nome, *Ithaca: Viaggio nella Scienza* **VIII**, 95–99.
- [168] E. Seneta, A tricentenary history of the law of large numbers, *Bernoulli* **19**, 1088–1121 (2013).
- [169] D.A. Sibley (1971), A metric for weak convergence of distribution functions, *Rocky Mountain J. Math.* **1**, 427–430.
- [170] Ya.G. Sinai, *Probability theory. An introductory course*, Springer, Berlin, 1992.
- [171] A.V. Skorokhod (1965), *Studies in the theory of random processes*, Addison–Wesley, Reading MA; ristampa, Dover, New York, 1982.
- [172] J.L. Snell (1952), Applications of martingale system theorems, *Trans. Amer. Math. Soc.* **73**, 293–312.
- [173] J.L. Snell (1982), Gambling, probability and martingales, *Math. Intelligencer* **4**, 629–632.
- [174] T. Stieltjes (1894), Recherches sur les fractions continues, *Ann. Fac. Sci. Toulouse* **8**, 1–122; anche in (Stieltjes, 1993) vol. II, pp. 406–570; traduzione inglese pp. 609–745.

- [175] T.J. Stieltjes (1993), *Œuvres complètes. Collected papers*, G. van Dijk, ed., Springer, New York – Heidelberg – Berlin.
- [176] J.M. Stoyanov (1987), *Counterexamples in probability*, Wiley, Chichester–New York.
- [177] K. Stromberg (1994), *Probability for analysts*, Chapman & Hall, New York–London.
- [178] G.J. Székely (1986), *Paradoxes in probability theory and mathematical statistics*, Reidel, Dordrecht.
- [179] A. Tarski (1938), Algebraische Fassung des Massesproblems, *Fund. Math.* **31**, 47–66.
- [180] M.D. Taylor (1985), New metrics for weak convergence of distribution functions, *Stochastica* **9** 5–17.
- [181] A.J. Thomasian (1956), Distances et normes sur les espaces de variables aléatoires, *C.R. Acad. Sci. Paris* **242**, 447–448.
- [182] A.J. Thomasian (1957), Metrics and norms for spaces of random variables, *Ann. Math. Statist.* **28**, 512–514.
- [183] H.F. Trotter (1959) *An elementary proof of the central limit theorem*, *Arch. Math.* **10**, 226–234.
- [184] H.G. Tucker (1967) *A graduate course in probability*, Academic Press, New York.
- [185] P. Turán, ed. (1976) *Selected papers of Alfréd Rényi*, Akadémiai Kiadó, Budapest.
- [186] S. Ulam (1930), Zur Masstheorie in der allgemeinen mengenlehre, *Fund. Math.* **16**, 140–150; anche in (Ulam, 1974), pp. 9–19.
- [187] S. Ulam (1974), *Sets, numbers and universes*, W.A. Beyer, J. Mycielski, G.–C. Rota, eds., MIT Press, Cambridge MA–London.
- [188] A.W. van der Vaart, J.A. Wellner (1996), *Weak convergence and empirical processes. With applications to statistics*, Springer, New York.
- [189] Ch. J. de la Vallée Poussin (1915), Sur l'intégrale de Lebesgue, *Trans. Amer. Math. Soc.* **16** 435–501.
- [190] S.R.S. Varadhan (2000) *Probability theory*, Courant Institute of Mathematical Sciences, New York and Amer. Math. Soc., Providence RI (Courant Lecture Notes 7).
- [191] J. Ville (1939) *Étude critique de la notion de collectif*, Gauthier–Villars, Paris.
- [192] D. Williams (1991) *Probability with martingales*, Cambridge University Press.

Indice analitico

- Additività numerabile 5
- Algebra di insiemi 1
- Anello 9
- Atomo 204

- Cilindro 51
- Classe monotona 12
- Compensatore 224
- Completamento 55
- Convergenza
 - completa 82
 - debole 84
 - in L^p 78
 - in legge 85
 - in probabilità 77
 - quasi certa 73
 - q.c. uniforme 99
 - stretta 88
 - vaga 88
- Convoluzione 131

- Densità
 - asintotica 62
 - condizionata 198
- Diseguaglianza
 - di Čebyšev 39
 - di Hölder 42
 - di Jensen 40
 - di Kolmogorov 172
 - di Lévy 179
 - di Markov 39
 - di Minkowski 42
 - di Schwarz 42
 - di troncamento 163
 - massimale 228
- Distanza in variazione 115
- Distribuzione 38
- Distribuzione condizionata regolare 209

- Entropia di Shannon 181
- Equazione di Wald 108

- Equi-integrabilità 231
- Estremo superiore essenziale 107

- Famiglia
 - adattata 215
 - equicontinua 135
 - equi-integrabile 231
 - tesa 115
 - uniformemente integrabile(6.4.1)
- Filtrazione 215
- Formula d' inversione 119
- Frontiera 89
- Funzione/i
 - boreliana 23
 - caratteristica 117
 - convessa 55
 - di regressione 200
 - di ripartizione 38
 - di ripartizione empirica 187
 - essenzialmente limitata 107
 - di RADEMACHER 27
 - generatrice dei momenti 150
 - integrabile 31
 - misurabile 22
 - semidefinita positiva 136
 - semplice 25
 - subadditiva 111

- Identità d'arresto 223
- Insieme
 - di livello 105
 - di \mathbb{P} -continuità 89
 - diretto 215
 - misurabile 3
- Integrabilità uniforme 231
- Integrale
 - continuità uniforme 232
 - per le funzioni semplici positive 28
 - per le funzioni misurabili 29

- λ -sistema 18

- Legge dei Grandi Numeri (LGN) debole
 - di Khinchin 169
 - di Markov 167
- Legge dei Grandi Numeri (LGN) forte
 - di Etemadi 176
 - di Khinchin–Kolmogorov 174
 - di Kolmogorov 173
 - di Rajchman 170
- Legge (di probabilità) 38
- Legge zero–uno
 - di Borel 70
- Lemma
 - di Borel–Cantelli 69
 - di Kronecker 251
 - di Töplitz (6.7.1)
 - di Wald 224
- Martingala 216
 - ereditaria 217
 - moltiplicativa 218
 - quadratica 218
 - trasformata 225
- Massimo limite 7
- Metrica
 - di Kolmogorov 110
 - di Ky Fan 80
 - di Lévy 95
- Minimo limite 7
- Misura
 - assolutamente continua 44
 - definita da una densità 43
 - definizione 8
 - del contare 36
 - di Borel–Stieltjes 21
 - di Dirac 11
 - di LEBESGUE 55
 - esterna 15
 - immagine 36
 - finitamente additiva 6
 - localmente finita 20
 - prodotto 49
 - reale 43
 - σ -finita 17
- Nucleo stocastico 214
- π -classe 12
- Probabilità
 - condizionata 196
 - di transizione 214
- Processo
 - arrestato al tempo T 223
 - crescente 220
 - prevedibile 224
- Prodotto interno 44
- Quasi certamente 38
- Quasi ovunque 31
- Ramo principale del logaritmo 156
- Restrizione di una probabilità 195
- Rettangoli misurabili 47
- Semi–anello 12
- Sezione 48
- σ -algebra 1
- Sottomartingala 216
 - decomposizione di Doob 220
- Spazio
 - duale 44
 - L^p 40
 - di Hilbert 44
 - di probabilità 38
 - misurabile 3
- Speranza 39
- Speranza condizionata 195
- Successione
 - convergente 7
 - monotona 7
- Supermartingala 216
- Tempo d'ingresso 221
- Tempo d'arresto rispetto
 - definizione 221
 - finito 222
- Teorema
 - del cambio di variabile 36
 - della classe monotona 12, 27
 - della linea di supporto 57
 - di Beppo Levi 33
 - di Bochner 137
 - di Carathéodory 17
 - di convergenza di Doob 239
 - di convergenza dominata 35
 - di convergenza monotona 33
 - di Fatou 34
 - di Feller 163
 - di Fubini–Tonelli 50

- di Helly 84, 93
- di Lindeberg 154
- di Lyapunov 155
- di MacMillan 181
- di Pólya 107
- di Radon–Nikodym 45
- di Riemann–Lebesgue 144
- di Riesz–Fréchet 44
- di Scheffé 104
- di Skorohod 86
- di Slutsky 106
- di Tarski 6
- di Ulam 11
- di de Moivre–Laplace 155
- Trasformata di Fourier–Stieltjes 118
- Tribú
 - banale 2
 - definizione 1
 - di Borel 2
 - generata 2
 - prodotto 47
- Variabile/i aleatoria/e
 - definizione 38
 - indipendenti 4
 - isonome 38
- Varianza 39
- Versione 196
- Vettore gaussiano 141