

Un corso di Calcolo delle Probabilità.

Maurizio Pratelli

Anno Accademico 2015-16

Gli appunti che seguono corrispondono al materiale svolto nel corso “**Probabilità**” nell’anno accademico 2015-16. Si tratta di un corso semestrale, per studenti dal III anno di Matematica in sù, pensato per studenti che hanno già seguito l’insegnamento di “**Elementi di Probabilità e Statistica**”.

Questi appunti sono pertanto concepiti come una prosecuzione dell’insegnamento di **E.P.S.** sopra citato, e si suppone che il lettore sia a conoscenza delle definizioni e concetti introdotti in quel corso.

Il programma svolto ed i metodi di dimostrazione sono abbastanza vicini (a parte il breve capitolo sui *Processi stocastici*) al libro *J. Jacod, P. Protter “Probability Essentials”* (Springer collana Universitext), libro che è anche ricco di esercizi (alcuni dei quali sono stati svolti durante il corso).

Una ricca miniera di esercizi interessanti è il libro *G. Letta “Probabilità Elementare”* (Zanichelli).

Indice

1	Misura e integrazione	5
1.1	Costruzione di una probabilità.	5
1.2	Integrazione della variabili aleatorie reali.	10
1.3	Densità e teorema di Radon-Nikodym	14
1.4	Completamento di una Probabilità.	20
1.5	Appendice	21
1.5.1	Estensione dei risultati alle misure sigma-finite.	21
1.5.2	Nozioni essenziali sugli spazi di Hilbert.	22
1.5.3	Complementi su misurabilità (e legami con la teoria degli insiemi).	24
2	Indipendenza	29
2.1	Misurabilità	29
2.2	Indipendenza e prime proprietà.	31
2.3	Probabilità Prodotto	33
2.4	Borel–Cantelli	35
2.5	Appendice	37
3	Le funzioni caratteristiche.	41
3.1	Definizione e prime proprietà.	41
3.2	Inversione delle funzioni caratteristiche.	46
3.3	La funzione generatrice dei momenti.	50
3.4	Vettori aleatori gaussiani.	52
3.5	Appendice	54
4	Convergenza di variabili aleatorie.	55
4.1	Convergenza in probabilità e quasi certa.	55
4.2	Convergenza di misure sulla retta reale.	59
4.3	Convergenza in Legge di variabili aleatorie.	65
4.4	Appendice	70
4.4.1	Diseguaglianze negli spazi L^p	70

4.4.2	Convergenza di misure e “funzioni semicontinue”	71
5	Teoremi limite	75
5.1	Teoremi Limite Centrale.	75
5.2	Serie di variabili aleatorie indipendenti	77
5.3	Leggi dei Grandi Numeri.	81
5.4	Appendice	85
6	Speranza condizionale e alcuni complementi.	87
6.1	La speranza condizionale.	87
6.2	Esempi concreti di calcolo di speranza condizionale	90
6.3	Unione essenziale e un teorema di Halmos-Savage.	92
6.4	Sugli spazi di probabilità “non atomici”.	95
7	Breve introduzioni ai Processi Stocastici.	99
7.1	Prime definizioni.	99
7.2	Il Processo di Wiener (o Moto Browniano).	101
7.3	Il processo di Poisson.	106
7.4	Due parole sui Processi di Markov.	108

Capitolo 1

Nozioni di base di misura e integrazione.

1.1 Costruzione di una probabilità.

Diamo per scontata la definizione di *algebra* e σ -*algebra* di parti di un insieme Ω ; se $\mathcal{I} \subseteq \mathcal{P}(\Omega)$ è una famiglia di parti di Ω , indichiamo con $\sigma(\mathcal{I})$ la σ -algebra generata da \mathcal{I} .

Teorema 1.1.1 (Il teorema delle Classi Monotone). *Sia \mathcal{I} una famiglia di parti di Ω stabile per l'intersezione e contenente Ω e sia $\mathcal{M} \supseteq \mathcal{I}$ la più piccola famiglia di parti contenente \mathcal{I} e tale che*

- \mathcal{M} è stabile per unione numerabile crescente (è una “classe monotona”)
- \mathcal{M} è stabile per differenza (insiemistica)

Allora \mathcal{M} è una σ -algebra (in particolare $\mathcal{M} = \sigma(\mathcal{I})$).

Con “stabile per differenza” si intende che se A, B sono elementi di \mathcal{M} e $B \subseteq A$, allora anche $A \setminus B = A \cap B^c$ è un elemento di \mathcal{M} .

Dimostrazione. Poiché $\Omega \in \mathcal{M}$ e \mathcal{M} è stabile per passaggio al complementare (infatti $A^c = \Omega \setminus A$), è sufficiente provare che \mathcal{M} è stabile per intersezione: infatti di conseguenza è stabile per unione, e quindi per unione numerabile (che possiamo supporre crescente) poiché è una classe monotona.

Questa verifica viene fatta in due passi:

- se $B \in \mathcal{I}, C \in \mathcal{M} \implies B \cap C \in \mathcal{M}$
(infatti questa proprietà è vera se anche $C \in \mathcal{I}$, e la famiglia delle parti C che godono di questa proprietà è una *classe monotona* stabile per differenza).

- se $B \in \mathcal{M}, C \in \mathcal{M} \implies B \cap C \in \mathcal{M}$
(proprietà vera se $B \in \mathcal{I}$ e poi si prosegue come sopra)

La dimostrazione è così completa. □

Vediamo ora due conseguenze immediate: la facile dimostrazione è lasciata per esercizio.

Ricordiamo che si chiama **probabilità** una funzione \mathbf{P} σ -additiva definita su una σ -algebra di parti e tale che $\mathbf{P}(\Omega) = 1$.

Corollario 1.1.2. *Sia \mathcal{I} una famiglia di parti stabile per l'intersezione che genera la σ -algebra \mathcal{F} : due probabilità che coincidono su \mathcal{I} coincidono su \mathcal{F} (cioè sono eguali).*

Ricordiamo che dati due spazi misurabili (E, \mathcal{E}) e (F, \mathcal{F}) , si chiama σ -algebra prodotto (e si indica $\mathcal{E} \otimes \mathcal{F}$) la σ -algebra sul prodotto cartesiano $E \times F$ generata dagli insiemi della forma $A \times B$, $A \in \mathcal{E}$ e $B \in \mathcal{F}$ (detti “rettangoli misurabili”).

Corollario 1.1.3. *Se $C \in \mathcal{E} \otimes \mathcal{F}$, per ogni $x \in E$ la “sezione” $C_x = \{y \in F \mid (x, y) \in C\}$ è un elemento di \mathcal{F} (e analogamente per la sezione C_y).*

Infatti questa proprietà è vera se C è un rettangolo misurabile e poi si estende a tutti gli elementi di $\mathcal{E} \otimes \mathcal{F}$ col teorema delle classi monotone.

Se \mathcal{A} è un'algebra di parti di un insieme Ω , una funzione $\mathbf{m} : \mathcal{A} \rightarrow \mathbb{R}^+$ è detta

- *additiva* se, dati A e B disgiunti, si ha $\mathbf{m}(A \cup B) = \mathbf{m}(A) + \mathbf{m}(B)$ (o, equivalentemente, dati A e B qualsiasi si ha $\mathbf{m}(A \cup B) + \mathbf{m}(A \cap B) = \mathbf{m}(A) + \mathbf{m}(B)$)
- *σ -additiva* se, presa una successione $(A_n)_{n \geq 1}$ di elementi di \mathcal{A} a due a due disgiunti e supponendo che $\bigcup_{n \geq 1} A_n$ sia ancora un elemento di \mathcal{A} , si ha $\mathbf{m}(\bigcup_{n \geq 1} A_n) = \sum_{n \geq 1} \mathbf{m}(A_n)$.

Ricordiamo che se \mathbf{m} è *additiva*, la *σ -additività* equivale alla proprietà di *passaggio al limite sulle successioni crescenti di insiemi*, cioè se $A_n \uparrow A$ allora $\mathbf{m}(A_n) \uparrow \mathbf{m}(A)$.

Viceversa la *proprietà di passaggio al limite sulle successioni decrescenti* (cioè $A_n \downarrow A \implies \mathbf{m}(A_n) \downarrow \mathbf{m}(A)$) (unita alla additività semplice) è *più forte* della σ -additività, ed è equivalente ad essa se $\mathbf{m}(\Omega) < +\infty$.

Una funzione σ -additiva \mathbf{m} definita su una σ -algebra \mathcal{F} di parti di un insieme Ω a valori in \mathbb{R}^+ è chiamata *misura*, la misura è detta σ -finita se Ω è unione di una successione di elementi di \mathcal{F} aventi misura finita.

Il nostro scopo è ora dimostrare il seguente

Teorema 1.1.4 (Teorema di prolungamento). *Sia \mathbf{P} una funzione σ -additiva definita su un'algebra \mathcal{A} di parti di un insieme Ω tale che $\mathbf{P}(\Omega) = 1$: \mathbf{P} si prolunga (in un sol modo) alla σ -algebra \mathcal{F} generata da \mathcal{A} .*

Prima di dimostrare il teorema, introduciamo alcuni risultati preliminari.

Lemma 1.1.5. *Siano $(A_n)_{n \geq 1}$ e $(A'_n)_{n \geq 1}$ due famiglie crescenti di elementi di \mathcal{A} e supponiamo che si abbia $\bigcup_{n \geq 1} A_n \subseteq \bigcup_{n \geq 1} A'_n$: allora vale la disuguaglianza*

$$\lim_{n \rightarrow \infty} \mathbf{P}(A_n) \leq \lim_{n \rightarrow \infty} \mathbf{P}(A'_n)$$

Dimostrazione. Per ogni fissato n si ha $A_n = \bigcup_{m \geq 1} (A_n \cap A'_m)$ e di conseguenza $\mathbf{P}(A_n) = \lim_{m \rightarrow \infty} \mathbf{P}(A_n \cap A'_m) \leq \lim_{m \rightarrow \infty} \mathbf{P}(A'_m)$. È facile a questo punto completare la dimostrazione. \square

Indichiamo con \mathcal{B} la classe degli insiemi che siano *unione di una successione crescente* di elementi di \mathcal{A} e **definiamo**, se $B = \bigcup_{n \geq 1} A_n$ e $A_n \subseteq A_{n+1}$, $\mathbf{P}(B) = \lim_{n \rightarrow \infty} \mathbf{P}(A_n)$. Il lemma 1.1.5 mostra che questa definizione prolunga la funzione d'insieme \mathbf{P} da \mathcal{A} a \mathcal{B} in modo non ambiguo, cioè non dipende dalla particolare successione crescente $(A_n)_{n \geq 1}$ di insiemi scelta per rappresentare B .

Lemma 1.1.6. *La funzione \mathbf{P} definita su \mathcal{B} gode delle seguenti proprietà:*

- a) se $B_n \uparrow B$, $\mathbf{P}(B_n) \uparrow \mathbf{P}(B)$;
- b) se B_1, B_2 sono elementi di \mathcal{B} , anche $(B_1 \cup B_2)$ e $(B_1 \cap B_2)$ sono elementi di \mathcal{B} e vale l'eguaglianza

$$\mathbf{P}(B_1 \cup B_2) + \mathbf{P}(B_1 \cap B_2) = \mathbf{P}(B_1) + \mathbf{P}(B_2)$$

In particolare \mathbf{P} definita su \mathcal{B} è σ -additiva.

Dimostrazione. Cominciamo a provare l'affermazione a) : per ogni n , scriviamo $B_n = \bigcup_{m \geq 1} B_{n,m}$ e definiamo $D_n = B_{1,n} \cup \dots \cup B_{n,n}$.

Per ogni n , $D_n \subseteq B_n$ e quindi $\lim_{n \rightarrow \infty} \mathbf{P}(D_n) \leq \lim_{n \rightarrow \infty} \mathbf{P}(B_n)$. Viceversa $D_n \uparrow B$ (poichè $D_n \supseteq B_{k,n}$ qualunque sia $k \leq n$), e quindi $\lim_{n \rightarrow \infty} \mathbf{P}(B_n) \geq$

$\mathbf{P}(B)$. Poichè la disuguaglianza nell'altro senso è immediata, segue l'eguaglianza.

Proviamo ora b) : consideriamo due successioni crescenti di elementi di \mathcal{A} tali che $B_{1,n} \uparrow B_1$ e $B_{2,n} \uparrow B_2$. Per ogni n fissato vale l'eguaglianza

$$\mathbf{P}(B_{1,n} \cup B_{2,n}) + \mathbf{P}(B_{1,n} \cap B_{2,n}) = \mathbf{P}(B_{1,n}) + \mathbf{P}(B_{2,n})$$

Questa eguaglianza va al limite e si ottiene la proprietà b). \square

Definiamo ora, dato $C \subseteq \Omega$, $\mathbf{P}^*(C) = \inf \{ \mathbf{P}(B) | B \in \mathcal{B}, B \supseteq C \}$: è evidente che \mathbf{P}^* ristretta a \mathcal{B} coincide con \mathbf{P} .

Proposizione 1.1.7. *La funzione d'insieme \mathbf{P}^* gode delle seguenti proprietà:*

- 1) per ogni C , si ha $0 \leq \mathbf{P}^*(C) \leq 1$;
- 2) se $C_1 \subseteq C_2$, allora $\mathbf{P}^*(C_1) \leq \mathbf{P}^*(C_2)$;
- 3) se $C_n \uparrow C$, allora $\mathbf{P}^*(C_n) \uparrow \mathbf{P}^*(C)$;
- 4) $\mathbf{P}^*(C_1 \cup C_2) + \mathbf{P}^*(C_1 \cap C_2) \leq \mathbf{P}^*(C_1) + \mathbf{P}^*(C_2)$.

Dimostrazione. Le proprietà 1) e 2) sono evidenti; proviamo ora 3). Da una parte è evidente che $\lim_{n \rightarrow \infty} \mathbf{P}^*(C_n) \leq \mathbf{P}^*(C)$.

Per provare la disuguaglianza opposta, dato $\varepsilon > 0$, sia $B_n \in \mathcal{B}$ con $B_n \supseteq C_n$ e $\mathbf{P}(B_n) \leq \mathbf{P}^*(C_n) + \frac{\varepsilon}{2^n}$, e poniamo $D_n = B_1 \cup \dots \cup B_n$ e $D = \bigcup_{n \geq 1} D_n$. Proviamo per induzione la disuguaglianza

$$\mathbf{P}(D_n) \leq \mathbf{P}^*(C_n) + \sum_{k=1}^n \frac{\varepsilon}{2^k}$$

Supponiamo che questa disuguaglianza (ovviamente verificata per $n = 1$) sia vera per n : si ha

$$\begin{aligned} \mathbf{P}(D_{n+1}) &= \mathbf{P}(D_n \cup B_{n+1}) = \mathbf{P}(D_n) + \mathbf{P}(B_{n+1}) - \mathbf{P}(D_n \cap B_{n+1}) \leq \\ &\leq \mathbf{P}^*(C_n) + \sum_{k=1}^n \frac{\varepsilon}{2^k} + \mathbf{P}^*(C_{n+1}) + \frac{\varepsilon}{2^{n+1}} - \mathbf{P}^*(C_n) = \mathbf{P}^*(C_{n+1}) + \sum_{k=1}^{n+1} \frac{\varepsilon}{2^k} \end{aligned}$$

Di conseguenza

$$\mathbf{P}^*(C) \leq \mathbf{P}(D) = \lim_{n \rightarrow \infty} \mathbf{P}(D_n) \leq \lim_{n \rightarrow \infty} \mathbf{P}^*(C_n) + \varepsilon$$

Ne segue la disuguaglianza voluta.

Proviamo ora 4): dato $\varepsilon > 0$, scegliamo $B_i \in \mathcal{B}$ con $B_i \supseteq C_i$ e $\mathbf{P}(B_i) \leq \mathbf{P}^*(C_i) + \varepsilon/2$. Si ha:

$$\begin{aligned} \mathbf{P}^*(C_1 \cup C_2) + \mathbf{P}^*(C_1 \cap C_2) &\leq \mathbf{P}(B_1 \cup B_2) + \mathbf{P}(B_1 \cap B_2) = \\ &= \mathbf{P}(B_1) + \mathbf{P}(B_2) \leq \mathbf{P}^*(C_1) + \mathbf{P}^*(C_2) + \varepsilon \end{aligned}$$

Poichè questa disuguaglianza è verificata per ogni ε positivo, segue la tesi. \square

Notiamo che, se $C \subseteq \Omega$, vale la disuguaglianza $\mathbf{P}^*(C) + \mathbf{P}^*(C^c) \geq 1$: poniamo allora

$$\mathcal{C} = \{C \subseteq \Omega \mid \mathbf{P}^*(C) + \mathbf{P}^*(C^c) = 1\}$$

e notiamo che $\mathcal{C} \supseteq \mathcal{B} \supseteq \mathcal{A}$.

Teorema 1.1.8. \mathcal{C} è una σ -algebra e \mathbf{P}^* ristretta a \mathcal{C} è σ -additiva.

Dimostrazione. È evidente che \mathcal{C} è stabile per passaggio al complementare: mostriamo che è stabile per unione e intersezione finita. Siano C_1 e C_2 elementi di \mathcal{C} : valgono le disuguaglianze

$$\mathbf{P}^*(C_1 \cup C_2) + \mathbf{P}^*(C_1 \cap C_2) \leq \mathbf{P}^*(C_1) + \mathbf{P}^*(C_2)$$

$$\mathbf{P}^*((C_1 \cup C_2)^c) + \mathbf{P}^*((C_1 \cap C_2)^c) \leq \mathbf{P}^*(C_1^c) + \mathbf{P}^*(C_2^c)$$

Sommando, a destra si ottiene 2: quindi tutte le disuguaglianze devono essere eguaglianze e \mathbf{P}^* è finitamente additiva. Tuttavia \mathbf{P}^* è allora σ -additiva poichè va al limite sulle successioni monotone d'insiemi.

Rimane da provare che \mathcal{C} è stabile per unione crescente: sia $(C_n)_{n \geq 1}$ una successione di elementi di \mathcal{C} con $C_n \uparrow C$ (e di conseguenza $C_n^c \downarrow C^c$). Per ogni n si ha

$$\mathbf{P}^*(C_n) + \mathbf{P}^*(C^c) \leq \mathbf{P}^*(C_n) + \mathbf{P}^*(C_n^c) = 1$$

e, al limite,

$$\mathbf{P}^*(C) + \mathbf{P}^*(C^c) = 1$$

e questo conclude la dimostrazione. \square

Il teorema 1.1.4 è una conseguenza del teorema 1.1.8: in generale $\mathcal{F} \subseteq \mathcal{C}$.

Il teorema 1.1.4 è valido anche per le misure σ -finite; questo fatto sarà precisato meglio nell'appendice.

1.2 Integrazione della variabili aleatorie reali.

L'esposizione di questo paragrafo è ristretta all'integrazione rispetto ad una probabilità, tuttavia praticamente tutti i risultati sono validi per l'integrazione rispetto a una misura σ -finita (con piccole modifiche nelle dimostrazioni): queste modifiche saranno precisate nell'appendice.

Supponiamo dunque assegnato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$: si chiama *variabile aleatoria reale* una funzione misurabile $X : \Omega \rightarrow \mathbb{R}$. Una tale v.a.r. è detta *semplice* se prende un numero finito di valori o, equivalentemente, se può essere scritta nella forma $X = \sum_{i=1}^n a_i I_{A_i}$, dove A_1, \dots, A_n sono elementi di \mathcal{F} ed I_A indica la *funzione indicatrice* dell'insieme A .

Per una v.a. semplice $X = \sum_{i=1}^n a_i I_{A_i}$ possiamo definire l'integrale

$$\int_{\Omega} X(\omega) d\mathbf{P}(\omega) = \int X d\mathbf{P} = \sum_{i=1}^n a_i \mathbf{P}(A_i)$$

È facile verificare che tale numero non dipende dalla particolare forma scelta per rappresentare X e che valgono le seguenti proprietà:

- $\int (aX + Y) d\mathbf{P} = a \int X d\mathbf{P} + \int Y d\mathbf{P}$
- se $X \leq Y$, allora $\int X d\mathbf{P} \leq \int Y d\mathbf{P}$.

Lemma 1.2.1. *Sia X_n una successione di v.a. semplici a valori positivi, supponiamo che X_n converga crescendo verso X e supponiamo che anche X sia semplice: allora $\int X_n d\mathbf{P} \uparrow \int X d\mathbf{P}$.*

È importante osservare che se la successione $(X_n)_{n \geq 1}$ converge (puntualmente), il limite X è sicuramente misurabile ma non è detto che sia semplice: in questo lemma *si suppone* che anche X sia semplice. Vediamo ora la facile dimostrazione:

Dimostrazione. Naturalmente $\int X d\mathbf{P} \geq \sup_n \int X_n d\mathbf{P} = \lim_{n \rightarrow \infty} \int X_n d\mathbf{P}$. Fissiamo ora $\varepsilon > 0$ e sia $A_\varepsilon^n = \{(X - X_n) \geq \varepsilon\}$, sia inoltre $c = \sup_{\omega} X(\omega)$: notiamo che $A_\varepsilon^n \downarrow \emptyset$ e di conseguenza $\mathbf{P}(A_\varepsilon^n) \downarrow 0$.

Dalla disuguaglianza $(X - X_n) \leq c I_{A_\varepsilon^n} + \varepsilon$, si ricava $\int X d\mathbf{P} \leq \int X_n d\mathbf{P} + c \mathbf{P}(A_\varepsilon^n) + \varepsilon$, e di conseguenza la tesi. \square

Lemma 1.2.2. *Siano $(X_n)_{n \geq 1}$ e $(X'_n)_{n \geq 1}$ due successioni crescenti di v.a. semplici a valori positivi e supponiamo che si abbia $\lim_n X_n \leq \lim_n X'_n$: allora*

$$\lim_n \int X_n d\mathbf{P} \leq \lim_n \int X'_n d\mathbf{P}$$

Dimostrazione. Per ogni n fissato, si ha $X_n = \sup_m (X_n \wedge X'_m) \leq \sup_m X'_m$, e, di conseguenza,

$$\int X_n d\mathbf{P} = \sup_m \int (X_n \wedge X'_m) d\mathbf{P} \leq \sup_m \int X'_m d\mathbf{P}$$

e da qui si ottiene immediatamente il risultato voluto. \square

Nella costruzione dell'integrale è fondamentale il risultato, dimostrato nel primo corso, che ricordo di seguito:

Teorema 1.2.3 (Approssimazione con variabili aleatorie semplici).

Sia X una variabile aleatoria a valori positivi: esiste una successione di v.a. semplici $(X_n)_{n \geq 1}$ (a valori positivi) tale che

$$X_n \uparrow X$$

Questo permette di dare la seguente

Definizione 1.2.4. Definiamo, per una v.a. X a valori positivi

$$\int X d\mathbf{P} = \lim_{n \rightarrow \infty} \int X_n d\mathbf{P}$$

dove $(X_n)_{n \geq 1}$ è una successione di v.a. semplici tale che $X_n \uparrow X$.

Il lemma 1.2.2 garantisce che questo numero non dipende dalla particolare successione approssimante scelta, inoltre in questa definizione si può supporre che la v.a. X sia a valori in $[0, +\infty]$.

Notiamo che, se $X \geq 0$, $\int X d\mathbf{P} \in [0, +\infty]$. Sono immediati i seguenti risultati, nei quali si suppone che X, Y siano v.a. a valori positivi:

- se $a \geq 0$, $\int (aX + Y) d\mathbf{P} = a \int X d\mathbf{P} + \int Y d\mathbf{P}$
- $X \geq 0$, $\int X d\mathbf{P} = 0 \Leftrightarrow X = 0$ q.c.
- se X e Y sono a valori reali (q.c.), vale l'eguaglianza $\int_A X d\mathbf{P} = \int_A Y d\mathbf{P}$ per ogni $A \in \mathcal{F} \Leftrightarrow X = Y$ q.c.

Lemma 1.2.5 (Proprietà di Beppo-Levi). *Siano $(X_n)_{n \geq 1}$ v.a. a valori positivi:*

$$X_n \uparrow X \quad \Rightarrow \quad \int X_n d\mathbf{P} \uparrow \int X d\mathbf{P}$$

Dimostrazione. Sia, per ogni n fissato, $(X_{n,m})_{m \geq 1}$ una successione crescente di v.a. semplici tale che $X_n = \sup_{m \geq 1} X_{n,m}$ e poniamo $Y_m = X_{1,m} \vee \dots \vee X_{m,m}$: $(Y_m)_{m \geq 1}$ è una successione crescente di v.a. semplici e $Y_m \uparrow X$.

Di conseguenza $\int Y_m d\mathbf{P} \uparrow \int X d\mathbf{P}$, però per ogni m fissato $\int Y_m d\mathbf{P} \leq \int X_m d\mathbf{P}$. Poichè ovviamente si ha, per m fissato, $\int X d\mathbf{P} \geq \int X_m d\mathbf{P}$, segue facilmente la tesi. \square

Lemma 1.2.6 (Proprietà di Fatou). *Sia $(X_n)_{n \geq 1}$ una successione di v.a. a valori positivi: vale la disuguaglianza*

$$\int \left(\liminf_{n \rightarrow \infty} X_n \right) d\mathbf{P} \leq \liminf_{n \rightarrow \infty} \int X_n d\mathbf{P}$$

Dimostrazione. Dalla definizione $\liminf_n X_n = \sup_{n \geq 1} (\inf_{k \geq n} X_k)$ e $Y_n = \inf_{k \geq n} X_k \uparrow \liminf_n X_n$.

Di conseguenza

$$\int \left(\liminf_{n \rightarrow \infty} X_n \right) d\mathbf{P} = \sup_{n \geq 1} \int \left(\inf_{k \geq n} X_k \right) d\mathbf{P} \leq \sup_{n \geq 1} \inf_{k \geq n} \int X_k d\mathbf{P} = \liminf_{n \rightarrow \infty} \int X_n d\mathbf{P}$$

\square

Consideriamo ora una v.a.r. X di segno qualsiasi, e siano $X^+ = X \vee 0$ e $X^- = -(X \wedge 0)$:

Definizione 1.2.7. X è detta *integrabile* se $\int |X| d\mathbf{P} = \int X^+ d\mathbf{P} + \int X^- d\mathbf{P} < +\infty$, e si chiama *integrale di X* il numero $\int X^+ d\mathbf{P} - \int X^- d\mathbf{P}$.

È facile verificare che se $X = U - V$, dove U e V sono entrambe a valori positivi e di integrale finito, si ha $\int X d\mathbf{P} = \int U d\mathbf{P} - \int V d\mathbf{P}$: questa osservazione è utile per provare, ad esempio, che se X e Y sono integrabili, allora $\int (X + Y) d\mathbf{P} = \int X d\mathbf{P} + \int Y d\mathbf{P}$ (mentre l'eguaglianza $\int (aX) d\mathbf{P} = a \int X d\mathbf{P}$ è immediata).

Supponiamo ora che si abbia $X \geq Y$ con Y integrabile (questo equivale a dire che $\int X^- d\mathbf{P} < +\infty$): allora ha senso il numero $\int X d\mathbf{P} = \int (X - Y) d\mathbf{P} + \int Y d\mathbf{P}$ (notiamo che in tal caso $\int X d\mathbf{P} \in]-\infty, +\infty]$). In questo caso si dice che X è *semi-integrabile inferiormente*.

Inoltre se $(X_n)_{n \geq 1}$ è una successione di v.a. con $X_n \geq Y$ qualunque sia n , ed Y integrabile, continuano a valere le *proprietà di Beppo-Levi e di Fatou*. Naturalmente analoghe proprietà valgono se $X \leq Y$ con Y integrabile.

Il teorema fondamentale di passaggio al limite sotto il segno di integrale è il seguente:

Teorema 1.2.8 (Convergenza dominata, o di Lebesgue). *Sia $(X_n)_{n \geq 1}$ una successione di v.a.r. e supponiamo che la successione $(X_n)_{n \geq 1}$ converga puntualmente ad X e che esista una v.a. integrabile Y tale che si abbia $|X_n| \leq Y$: allora*

$$\lim_{n \rightarrow \infty} \int X_n d\mathbf{P} = \int X d\mathbf{P}$$

Dimostrazione. Tenendo conto dell'osservazione precedente, poichè per ogni n si ha $-Y \leq X_n \leq Y$, la dimostrazione è una conseguenza immediata del lemma 1.2.6. Si ha infatti

$$\int (\liminf_{n \rightarrow \infty} X_n) d\mathbf{P} \leq \liminf_{n \rightarrow \infty} \int X_n d\mathbf{P} \leq \limsup_{n \rightarrow \infty} \int X_n d\mathbf{P} \leq \int (\limsup_{n \rightarrow \infty} X_n) d\mathbf{P}$$

e, poichè $X = \liminf_{n \rightarrow \infty} X_n = \limsup_{n \rightarrow \infty} X_n$, segue immediatamente la tesi. \square

Nel calcolo del valore atteso è di importanza fondamentale la nozione di *probabilità immagine* ed il teorema di integrazione rispetto ad una probabilità immagine: l'immagine di \mathbf{P} mediante la v.a.r. X (indicata $X(\mathbf{P})$ o anche \mathbf{P}_X , e chiamata *legge di probabilità* o *distribuzione di probabilità* della v.a. X), è definita da $\mathbf{P}_X(A) = \mathbf{P}(X^{-1}(A))$.

La **formula di integrazione rispetto ad una probabilità immagine** è la seguente: se $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ è boreliana, si ha

$$\int_{\Omega} \varphi(X(\omega)) d\mathbf{P}(\omega) = \int_{\mathbb{R}} \varphi(x) d\mathbf{P}_X(x)$$

(più precisamente $\varphi(X(\omega))$ è integrabile rispetto a \mathbf{P} se e solo se $\varphi(x)$ è integrabile rispetto a $X(\mathbf{P})$ e in tal caso vale la formula sopra scritta). Questa formula, di importanza fondamentale ma facile da dimostrare, è già stata provata nel primo corso.

Ricordiamo che $\mathbf{E}[X]$ è una notazione che indica (se esiste) l'integrale $\int_{\Omega} X(\omega) d\mathbf{P}(\omega)$: solitamente questo numero è chiamato *valore atteso* o *speranza matematica* della v.a. X .

Proposizione 1.2.9 (Disuguaglianza di Jensen). *Sia $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ una funzione convessa e supponiamo che X e $\varphi(X)$ siano integrabili: allora*

$$\varphi(\mathbf{E}[X]) \leq \mathbf{E}[\varphi(X)]$$

Dimostrazione. Ogni funzione convessa φ si può scrivere nella forma $\varphi(x) = \sup_n L_n(x)$, dove $L_n(x) = a_n x + b_n$ è una funzione lineare affine. Per ogni n fissato si ha

$$L_n(\mathbf{E}[X]) = \mathbf{E}[L_n(X)] \leq \mathbf{E}[\varphi(X)]$$

e, prendendo a sinistra l'estremo superiore al variare di n , si ottiene la disuguaglianza. \square

Corollario 1.2.10. *Se $1 \leq p < q < +\infty$ e se $\mathbf{E}[|X|^q] < +\infty$, allora anche $\mathbf{E}[|X|^p] < +\infty$.*

Osservazione 1.2.11. (Approssimazione uniforme con v.a. semplici)

In questo paragrafo abbiamo usato ripetutamente il risultato del Teorema 1.2.3, cioè ogni v.a. X a valori positivi è limite puntuale di una successione crescente di v.a. semplici.

Vale anche il seguente risultato, che sarà usato più avanti: se X è una v.a. limitata, esiste una successione $(X_n)_{n \geq 1}$ di v.a. semplici convergente uniformemente verso X . Lascio i dettagli per esercizio, ma come suggerimento (supponendo che si abbia $|X(\omega)| \leq M$ q.c.) per l'approssimazione n -sima si considera la variabile

$$X_n(\omega) = \sum_{k=-2^n}^{2^n-1} \frac{kM}{2^n} I_{\left\{ \frac{kM}{2^n} \leq X < \frac{(k+1)M}{2^n} \right\}}(\omega)$$

1.3 Misura definita da una densità e teorema di Radon–Nikodym.

L'esposizione in questo paragrafo si riferisce alla *densità* di una misura σ -finita rispetto ad un'altra, e non si limita alle probabilità: la situazione che ci interesserà in pratica sarà quella della densità di una probabilità rispetto ad una misura σ -finita.

Sia μ una misura σ -finita su (E, \mathcal{E}) e sia $f : E \rightarrow [0, +\infty]$ una funzione misurabile a valori positivi.

Definizione 1.3.1. Si chiama *misura definita dalla densità f* (rispetto a μ) (indicata $\nu = f \cdot \mu$) la misura

$$\nu(A) = \int_A f(x) d\mu(x)$$

È immediato constatare che

- ν è σ -finita se e solo se $f(x) < +\infty$ per quasi ogni x ;
- ν è finita se e solo se f è integrabile;
- ν è una probabilità se e solo se $\int_E f(x) d\mu(x) = 1$.

Teorema 1.3.2 (Formula di integrazione rispetto alla misura definita da una densità). *Supponiamo che ν sia σ -finita e sia $\varphi : E \rightarrow \mathbb{R}$ misurabile: vale la formula*

$$\int_E \varphi(s) d\nu(s) = \int_E \varphi(x) f(x) d\mu(x)$$

Naturalmente si intende che φ è integrabile rispetto a ν se e solo se φf è integrabile rispetto a μ , e in tal caso vale l'eguaglianza sopra scritta. La dimostrazione di questo risultato, sostanzialmente già fatta nel corso del II anno, è molto facile e come quasi tutte le dimostrazioni che riguardano formule con integrali si svolge secondo questi passi:

- si verifica direttamente che la formula è soddisfatta se $\varphi = I_A$, e a questo punto è immediata l'estensione al caso di φ semplice;
- si estende alle funzioni misurabili a valori positivi approssimando con una successione crescente di funzioni semplici ed utilizzando la *proprietà di Beppo-Levi*;
- si estende alle funzioni integrabili considerando la decomposizione $\varphi = \varphi^+ - \varphi^-$.

Definizione 1.3.3 (Assoluta continuità di misure). Siano μ, ν due misure σ -finite: si dice che ν è *assolutamente continua* rispetto a μ (e si scrive $\nu \ll \mu$) se ogni insieme μ -trascurabile è anche ν -trascurabile.

Osservazione 1.3.4. Se $\nu = f \cdot \mu$ è immediato constatare che $\nu \ll \mu$, inoltre la densità f è unica *a meno di equivalenza*.

Si usa la notazione $f = \frac{d\nu}{d\mu}$, tenendo però presente che $\frac{d\nu}{d\mu}$ è una *classe di equivalenza* di funzioni misurabili ed f un rappresentante di questa classe.

Teorema 1.3.5 (Radon-Nikodym). *Siano μ, ν σ -finite e supponiamo che si abbia $\nu \ll \mu$: allora esiste una densità f tale che si abbia*

$$\nu(A) = \int_A f(x) d\mu(x)$$

Osserviamo subito che nella dimostrazione di questo risultato ci si può ridurre al caso in cui sia μ che ν siano finite: infatti $(\mu + \nu)$ è σ -finita e, considerata una successione $(A_n)_{n \geq 1}$ di insiemi misurabili (che possiamo supporre disgiunti) tali che $(\mu + \nu)(A_n) < +\infty$, se è individuata la densità di ν rispetto a μ relativamente ai sottoinsiemi di A_n , è individuata la densità su tutto lo spazio.

La dimostrazione presentata qui è facile, ma si appoggia su alcuni risultati validi per gli spazi di Hilbert (in particolare il *teorema di rappresentazione di Riesz*): ci interessa in particolare lo spazio di Hilbert $L^2(E, \mathcal{E}, \mu)$ munito del prodotto scalare $\langle f|g \rangle_{L^2} = \int f g \, d\mu$. Le definizioni e proprietà essenziali degli spazi di Hilbert, insieme al fatto che lo spazio L^2 così definito è effettivamente uno spazio di Hilbert, sono richiamate nell'Appendice.

Dimostrazione. Osserviamo subito che la funzione $L(\varphi) = \int_E \varphi \, d\mu$ è lineare continua su $L^2(\mu + \nu)$: si ha infatti

$$\int_E |\varphi| \, d\mu \leq \sqrt{\int_E \varphi^2 \, d\mu} \cdot \sqrt{\mu(E)} \leq \sqrt{\int_E \varphi^2 \, d(\mu + \nu)} \cdot \sqrt{\mu(E)}$$

Pertanto esiste $g \in L^2(\mu + \nu)$ tale che si abbia, per ogni $\varphi \in L^2(\mu + \nu)$

$$\int_E \varphi \, d\mu = \int_E \varphi g \, d(\mu + \nu)$$

Proviamo ora che si ha, per quasi ogni x , $0 < g(x) \leq 1$ (q.o. rispetto a μ , e quindi anche rispetto a ν).

Sia infatti $A = \{g > 1\}$: prendendo $\varphi = I_A$, si ottiene

$$\mu(A) = \int_A g \, d(\mu + \nu) \geq \int_A g \, d\mu$$

ma, se A fosse non trascurabile, avremmo $\mu(A) < \int_A g \, d\mu$, e l'eguaglianza scritta sopra sarebbe pertanto impossibile.

In modo analogo, se $B = \{g \leq 0\}$ si ha $\mu(B) = \int_B g \, d(\mu + \nu) \leq 0$, e ne segue che $\mu(B) = 0$.

Dunque, se φ è misurabile limitata (e quindi in particolare è un elemento di $L^2(\mu + \nu)$), si ottiene

$$\int \varphi(1 - g) \, d\mu = \int \varphi g \, d\nu$$

Questa eguaglianza si estende poi (grazie alla proprietà di *Beppo Levi*) alle funzioni φ misurabili a valori positivi. Infine (sostituendo φ con φ/g), si ha

$$\int \varphi \, d\nu = \int \varphi \left(\frac{1 - g}{g} \right) \, d\mu$$

Si ottiene cioè come densità la funzione (misurabile a valori positivi) $f = (1 - g)/g$. \square

Si dice che μ è **equivalente** a ν (e si scrive $\mu \sim \nu$) se contemporaneamente $\mu \ll \nu$ e $\nu \ll \mu$ (cioè se μ e ν hanno gli stessi insiemi trascurabili). Si possono verificare per esercizio le seguenti affermazioni:

- supponiamo $\nu \ll \mu$: allora $\nu \sim \mu$ se e solo se $f = \frac{d\nu}{d\mu}$ è strettamente positiva q.o. e una versione della densità $\frac{d\mu}{d\nu}$ è fornita da $\frac{1}{f}$;
- supponiamo $\nu \ll \mu$ e $\mu \ll \sigma$: naturalmente $\nu \ll \sigma$ e vale la formula

$$\frac{d\nu}{d\sigma} = \frac{d\nu}{d\mu} \cdot \frac{d\mu}{d\sigma}$$

Con la notazione sopra scritta si intende che se f è una versione della densità $\frac{d\nu}{d\mu}$ e g della densità $\frac{d\mu}{d\sigma}$, allora il prodotto $f.g$ è una versione della densità $\frac{d\nu}{d\sigma}$.

Proposizione 1.3.6. *Se \mathbf{P} è una probabilità e μ una misura σ -finita, la condizione $\mathbf{P} \ll \mu$ è equivalente alla seguente:*

$$\forall \varepsilon > 0, \exists \delta > 0 : \mu(A) < \delta \implies \mathbf{P}(A) < \varepsilon$$

Dimostrazione. Naturalmente la condizione sopra scritta è più forte dell'assoluta continuità. Supponiamo viceversa $\mathbf{P} \ll \mu$: se la condizione sopra scritta non fosse vera, esisterebbe $\varepsilon > 0$ e, per ogni n , un insieme misurabile A_n con $\mu(A_n) \leq 2^{-n}$ e $\mathbf{P}(A_n) \geq \varepsilon$. Sia ora $B_n = A_n \cup A_{n+1} \cup A_{n+2} \dots$ e notiamo che $\mu(B_n) \leq 2^{-(n-1)}$: $(B_n)_{n \geq 1}$ è una successione decrescente di insiemi e $B_n \downarrow B$ che è μ -trascurabile. Però, per ogni n , $\mathbf{P}(B_n) \geq \varepsilon$ e di conseguenza $\mathbf{P}(B) \geq \varepsilon$. \square

Osservazione 1.3.7. La proposizione 1.3.6 non è soddisfatta se invece di una probabilità si considera una misura σ -finita.

La prova è lasciata per esercizio, suggerendo di considerare come ν la misura di Lebesgue su \mathbb{R} (usualmente indicata λ) e come μ una misura definita da una densità (rispetto alla misura λ) strettamente positiva che converge velocemente a 0 per $|x| \rightarrow \infty$ (ad esempio la densità $\frac{1}{1+x^2}$).

Proposizione 1.3.8. *Una misura σ -finita su (E, \mathcal{E}) è equivalente a una probabilità.*

Questa proprietà è molto importante; per provarla si deve costruire una funzione misurabile strettamente positiva ovunque con integrale eguale a 1.

Osservazione 1.3.9. Sia (E, \mathcal{E}, μ) uno spazio di misura e supponiamo che per ogni $x \in E$ l'insieme $\{x\}$ appartenga ad \mathcal{E} : provare che se μ è σ -finita l'insieme degli x tali che $\mu(\{x\}) > 0$ è al più numerabile. Indicando questo insieme con $\{x_1, x_2, \dots\}$ e indicando con ν la misura $\nu = \sum_{i \geq 1} c_i \cdot \delta_{x_i}$ (dove $c_i = \mu(\{x_i\})$ e δ_x è la misura concentrata nel punto x cioè $\delta_x(A) = I_A(x)$), la misura $(\mu - \nu)$ è *diffusa* (cioè ogni punto è trascurabile). Se $\mu = \nu$, la misura è detta *discreta* o anche *atomica*.

Esempio 1.3.10 (Una misura non σ -finita). Consideriamo su \mathbb{R} la misura “che conta i punti” cioè

$$\mu(A) = \begin{cases} \#(A) & \text{se } A \text{ è un insieme finito} \\ +\infty & \text{altrimenti} \end{cases}$$

È immediato constatare che μ non è σ -finita, inoltre ogni altra misura su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ è assolutamente continua rispetto a μ (poichè solo \emptyset è trascurabile rispetto a μ).

Lo scopo di questo esempio è mostrare che il Teorema 1.3.5 non è valido in questo contesto e a tal scopo è opportuno caratterizzare le funzioni integrabili secondo questa misura.

Dato un insieme di cardinalità qualsiasi $(a_i)_{i \in I}$ di numeri positivi si chiama $\sum_{i \in I} a_i$ l'estremo superiore delle somme $\sum_{i \in J} a_i$ al variare di tutti i sottinsiemi finiti $J \subseteq I$, e se $\sum_{i \in I} a_i < +\infty$ segue che solo un sottinsieme al più numerabile dei numeri a_i è diverso da 0. Presa una famiglia $(c_i)_{i \in I}$ di numeri di segno qualsiasi, si dice che questa è sommabile se $\sum_{i \in I} |c_i| < +\infty$ e in tal caso si definisce $\sum_{i \in I} c_i = \sum_{i \in I} c_i^+ - \sum_{i \in I} c_i^-$ (indicando come al solito $c_i^+ = (c_i \vee 0)$ e $c_i^- = -(c_i \wedge 0)$).

Si mostra facilmente che una funzione f è integrabile rispetto a μ se e solo se l'insieme dei numeri $f(x)$ al variare di $x \in \mathbb{R}$ è sommabile e che in tal caso si ha $\int f \, d\mu = \sum_{x \in \mathbb{R}} f(x)$.

A questo punto è facile constatare che la misura di Lebesgue non ha una densità rispetto a μ (pur essendo assolutamente continua) e più in generale che solo le misure discrete sono dotate di densità rispetto a μ (esplicitare in quest'ultimo caso la densità).

Da qui fino alla fine del capitolo tutte le misure sono tacitamente supposte σ -finite. Si dice che la misura μ è **portata da** A ($A \in \mathcal{E}$) se il complementare di A è μ -trascurabile: questo equivale a dire che per ogni $B \in \mathcal{E}$ si ha $\mu(B) = \mu(A \cap B)$.

Questo equivale anche a dire che $\mu = I_A \cdot \mu$: infatti la misura che ha densità I_A rispetto a μ è tale che $(I_A \cdot \mu)(B) = \int_B I_A \, d\mu = \int I_B \cdot I_A \, d\mu = \mu(A \cap B)$.

Definizione 1.3.11. Si dice che due misure μ e ν sono **singolari** tra di loro (o anche **ortogonali**) se esiste $A \in \mathcal{E}$ tale che μ sia portata da A e ν dal suo complementare A^c (o, ciò che è lo stesso, se μ e ν sono portate da due insiemi misurabili disgiunti).

Teorema 1.3.12 (Decomposizione di Lebesgue). *Siano μ e ν due misure su (E, \mathcal{E}) : ν si può decomporre in un sol modo nella forma $\nu = f \cdot \mu + \gamma$ dove $f \cdot \mu$ è assolutamente continua e γ è singolare rispetto a μ .*

Dimostrazione. Naturalmente μ è assolutamente continua rispetto alla misura $(\mu + \nu)$: scegliamo allora una versione h della densità $h = \frac{d\mu}{d(\mu + \nu)}$ e sia B l'insieme $B = \{h = 0\}$ che è μ -trascurabile.

Consideriamo allora la decomposizione $\nu = I_{B^c} \cdot \nu + I_B \cdot \nu$: poiché $\gamma = I_B \cdot \nu$ è singolare rispetto a μ , rimane da provare che $I_{B^c} \cdot \nu$ è assolutamente continua, cioè che se $\mu(A) = 0$, allora $\nu(B^c \cap A) = 0$.

Supponiamo dunque di avere

$$0 = \mu(A) = \int_A h \, d(\mu + \nu) = \int_{A \cap B^c} h \, d(\mu + \nu)$$

Poiché h è strettamente positiva su B^c e l'integrale di una funzione positiva è 0 solo se la funzione è nulla q.c., ne segue $(\mu + \nu)(B^c \cap A) = 0$ e di conseguenza $\nu(B^c \cap A) = 0$.

Vediamo ora l'unicità: supponiamo di avere due decomposizioni $\nu = f \cdot \mu + \gamma = g \cdot \mu + \delta$.

Sia γ che δ sono portate da un insieme B che è μ -trascurabile e possiamo pertanto supporre che f e g siano eguali a 0 su B : vogliamo provare che, per ogni $A \in \mathcal{E}$, si ha $\int_A f \, d\mu = \int_A g \, d\mu$.

Poiché f e g sono nulle su B , valgono le eguaglianze

$$\int_A f \, d\mu = \int_{A \cap B^c} f \, d\mu = \nu(A \cap B^c) = \int_{A \cap B^c} g \, d\mu = \int_A g \, d\mu$$

Ne segue che $f = g$ μ -quasi ovunque, cioè che le due misure $f \cdot \mu$ e $g \cdot \mu$ sono eguali e di conseguenza si ha l'eguaglianza di γ e δ . □

Osservazione 1.3.13. Ogni misura μ su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ si decompone (in un sol modo) nella forma $\mu = \mu_1 + \mu_2 + \mu_3$, dove μ_1 è discreta, μ_2 è definita da una densità rispetto alla misura di Lebesgue e μ_3 è diffusa ma *singolare* rispetto alla misura di Lebesgue. La misura μ_3 è talvolta chiamata la *parte Cantoriana* di μ .

La misura μ_1 si ottiene come nell'osservazione 1.3.9, si applica poi la decomposizione di Lebesgue (rispetto alla misura di Lebesgue) alla misura diffusa $(\mu - \mu_1)$.

1.4 Completamento di una Probabilità.

In questa sezione esponiamo il *completamento* di una σ -algebra rispetto ad una probabilità, tuttavia siccome come vedremo il completamento dipende dagli insiemi trascurabili e quindi dalla classe di equivalenza di una misura ed ogni misura σ -finita è equivalente ad una probabilità (vedi proposizione 1.3.8), questa costruzione si estende senza modifiche alle misure σ -finite.

Sia $(\Omega, \mathcal{F}, \mathbf{P})$ uno spazio di probabilità e chiamiamo \mathcal{N} la famiglia dei sottinsiemi $A \subseteq \Omega$ tali che esiste $B \in \mathcal{F}$ con $A \subseteq B$ e $\mathbf{P}(B) = 0$. Gli elementi di \mathcal{N} sono chiamati gli *insiemi trascurabili* (in inglese “*null-sets*”).

È immediato constatare che, se $A \in \mathcal{N}$ e $B \subseteq A$ anche $B \in \mathcal{N}$, inoltre \mathcal{N} è stabile per unione numerabile. Indichiamo con $\mathcal{F}^{\mathbf{P}}$ la σ -algebra generata da \mathcal{F} e da \mathcal{N} e questa è chiamata il *completamento* di \mathcal{F} (rispetto alla probabilità \mathbf{P}).

Vale la caratterizzazione seguente:

Teorema 1.4.1. *Un sottinsieme $A \subseteq \Omega$ appartiene a $\mathcal{F}^{\mathbf{P}}$ se e solo se esistono $B, C \in \mathcal{F}$ tali che $B \subseteq A \subseteq C$ e $\mathbf{P}(C \setminus B) = 0$. Inoltre \mathbf{P} si prolunga in uno ed in un sol modo a $\mathcal{F}^{\mathbf{P}}$ ponendo, per A, B, C come sopra $\mathbf{P}(A) = \mathbf{P}(B) = \mathbf{P}(C)$.*

Notiamo che la condizione $\mathbf{P}(C \setminus B) = 0$ equivale a $\mathbf{P}(B) = \mathbf{P}(C)$ (quest'ultima affermazione non è vera con una misura μ non necessariamente finita).

Dimostrazione. Se indichiamo con \mathcal{G} la famiglia degli insiemi che godono della proprietà sopra scritta, è evidente che questa contiene sia \mathcal{F} che \mathcal{N} e, se proviamo che è una σ -algebra, ne seguirà che è appunto quella generata da \mathcal{F} e \mathcal{N} .

È immediato constatare che se $A \in \mathcal{G}$, anche $A^c \in \mathcal{G}$: infatti se $B \subseteq A \subseteq C$, allora $C^c \subseteq A^c \subseteq B^c$ e $B^c \setminus C^c = C \setminus B$.

Vediamo ora la stabilità per unione numerabile: se, per ogni n , $B_n \subseteq A_n \subseteq C_n$ (con $\mathbf{P}(C_n \setminus B_n) = 0$) allora

$$\bigcup_{n \geq 1} B_n \subseteq \bigcup_{n \geq 1} A_n \subseteq \bigcup_{n \geq 1} C_n$$

e si ha

$$\left(\bigcup_{n \geq 1} C_n\right) \setminus \left(\bigcup_{n \geq 1} B_n\right) \subseteq \bigcup_{n \geq 1} (C_n \setminus B_n)$$

che è trascurabile.

Vediamo ora che il prolungamento di \mathbf{P} è univocamente determinato. Supponiamo infatti che si abbia $B \subseteq A \subseteq C$ e contemporaneamente $B^* \subseteq A \subseteq C^*$ (con $\mathbf{P}(C \setminus B) = \mathbf{P}(C^* \setminus B^*) = 0$): allora $\mathbf{P}(B^*) \leq \mathbf{P}(C)$ e contemporaneamente $\mathbf{P}(B) \leq \mathbf{P}(C^*)$. □

Il nome *completamento* deriva da questa proprietà: se $A \in \mathcal{F}^{\mathbf{P}}$, $\mathbf{P}(A) = 0$ e $B \subseteq A$, allora anche $B \in \mathcal{F}^{\mathbf{P}}$. Se $A \in \mathcal{F}^{\mathbf{P}}$, si dice che A è \mathbf{P} -misurabile.

Osservazione 1.4.2. Il completamento di una σ -algebra \mathcal{E} su insieme E rispetto ad una misura *discreta* è sempre la famiglia di tutti i sottinsiemi di E .

Sulla retta (o più in generale su \mathbb{R}^n) il completamento della σ -algebra di Borel rispetto alla misura di Lebesgue coincide con la famiglia degli insiemi misurabili secondo Lebesgue.

L'intersezione delle σ -algebre $\mathcal{B}(\mathbb{R}^n)^\mu$ al variare di tutte le misure σ -finite μ è chiamata la σ -algebra degli insiemi *universalmente misurabili*.

1.5 Appendice

1.5.1 Estensione dei risultati alle misure sigma-finite.

Indichiamo brevemente le piccole modifiche che occorre fare per estendere i risultati delle sezioni 1 e 2 alle misure σ -finite.

Il teorema 1.1.4 si estende ad una funzione μ σ -additiva definita su un'algebra \mathcal{A} di parti di Ω e la dimostrazione rimane quasi identica con una sola importante modifica.

La classe $\mathcal{C} = \{C \subseteq \Omega \mid \mathbf{P}^*(C) + \mathbf{P}^*(C^c) = 1\}$ deve essere sostituita dalla classe degli insiemi *misurabili secondo Caratheodory*: A è detto *misurabile* se per ogni altro sottinsieme B di Ω si ha

$$\mu^*(B) = \mu^*(A \cap B) + \mu^*(A^c \cap B)$$

dove μ^* è definita come \mathbf{P}^* .

Anche la costruzione dell'integrale (sezione 2) è sostanzialmente identica per misure σ -finite, l'unica vera differenza è nella definizione di *funzione*

semplice. Se (E, \mathcal{E}, μ) è un spazio di misura, si chiamano funzioni semplici quelle che possono essere scritte nella forma $f = \sum_{i=1}^n a_i I_{A_i}$, dove A_1, \dots, A_n sono elementi di \mathcal{F} con $\mu(A_i) < +\infty$.

Il teorema fondamentale 1.2.3 rimane valido con questa nuova definizione di funzione semplice.

È bene puntualizzare che la *diseguaglianza di Jensen* (proposizione 1.2.9) è vera solo con le probabilità.

Se μ è una misura σ -finita su \mathcal{E} e $f : E \rightarrow F$ una applicazione misurabile, si definisce allo stesso modo la misura immagine $\nu = f(\mu)$: nel linguaggio della teoria geometrica della misura, la misura immagine è chiamata usualmente “*pushforward*” e indicata $f^\#(\mu)$.

La misura immagine $\nu = f(\mu)$ non è necessariamente σ -finita, tuttavia se lo è vale di nuovo la formula di integrazione

$$\int_F \varphi(x) d\nu(x) = \int_E (\varphi \circ f)(t) d\mu(t)$$

1.5.2 Nozioni essenziali sugli spazi di Hilbert.

In questa appendice vengono enunciate senza dimostrazioni (ad uso esclusivo degli studenti che non abbiano già incontrato queste nozioni in altri corsi) le proprietà essenziali degli spazi Hilbert che vengono usate in questo corso.

Consideriamo uno spazio vettoriale \mathcal{H} sul quale è definito un *prodotto scalare*, cioè una applicazione $(x, y) \rightarrow \langle x, y \rangle$ definita su $\mathcal{H} \times \mathcal{H}$ a valori in \mathbb{R} che gode delle seguenti proprietà

- $\langle \lambda x + y, z \rangle = \lambda \langle x, z \rangle + \langle y, z \rangle$ per $\lambda \in \mathbb{R}$ e $x, y, z \in \mathcal{H}$
- $\langle x, y \rangle = \langle y, x \rangle$ (se \mathcal{H} è uno spazio vettoriale su \mathbb{C} si suppone invece che $\langle x, y \rangle$ sia a valori complessi e che si abbia $\langle x, y \rangle = \overline{\langle y, x \rangle}$)
- $\langle x, x \rangle \geq 0$ e $\langle x, x \rangle = 0$ se e solo se $x = 0$ (in \mathcal{H}).

Uno spazio vettoriale munito di un prodotto scalare è chiamato **spazio pre-hilbertiano**: su esso è definita una **norma** data da $\|x\| = \sqrt{\langle x, x \rangle}$.

Vale la **diseguaglianza** di Schwartz $|\langle x, y \rangle| \leq \sqrt{\langle x, x \rangle} \sqrt{\langle y, y \rangle} = \|x\| \cdot \|y\|$ e come conseguenza la **diseguaglianza triangolare** $\|x + y\| \leq \|x\| + \|y\|$. Ne segue che il numero $d(x, y) = \|x - y\|$ è una **distanza**: lo spazio è detto **spazio di Hilbert** se è **completo** rispetto alla topologia indotta da questa distanza (cioè se ogni successione di Cauchy converge).

In realtà sono interessanti solo gli spazi di Hilbert a *dimensione infinita*, perché gli spazi di Hilbert a dimensione finita sono unicamente gli spazi Euclidei \mathbb{R}^n con l'usuale prodotto scalare.

Un esempio importante di spazio di Hilbert è lo spazio $L^2(E, \mathcal{E}, \mu)$ delle funzioni di quadrato integrabile rispetto alla misura σ -finita μ munito del prodotto scalare $\langle f, g \rangle_{L^2} = \int fg d\mu$ (anzi questo è praticamente l'unico esempio che incontreremo in questo corso): è immediato constatare che questo è pre-hilbertiano ed il fatto che sia completo sarà dimostrato nel successivo Capitolo 4 (in verità nel Capitolo 4 si prova che L^2 è completo *rispetto ad una misura di probabilità*, tuttavia la dimostrazione è praticamente identica anche per le misure σ -finite).

Si chiamano **ortogonali** due vettori x, y di \mathcal{H} tali che $\langle x, y \rangle = 0$ e, se x, y sono ortogonali, vale il **teorema di Pitagora** $\|x + y\|^2 = \|x\|^2 + \|y\|^2$.

Se M è un sottinsieme convesso e chiuso (non vuoto) di \mathcal{H} , dato $x \in \mathcal{H}$ esiste uno e un solo punto di M di *minima distanza* da x , cioè esiste un unico $y \in M$ tale che si abbia $\|x - y\| \leq \|x - z\|$ per ogni $z \in M$: quando M è un **sottospazio chiuso** V questo punto di minima distanza è chiamata la **proiezione ortogonale** di x su V . Questo nome deriva dal fatto che, se Px è questo punto di minima distanza, $(x - Px)$ è ortogonale ad ogni altro punto del sottospazio V e la proiezione ortogonale può essere caratterizzata dalla proprietà $\langle x, z \rangle = \langle Px, z \rangle$ per ogni $z \in V$.

Sostanzialmente la proiezione ortogonale negli spazi di Hilbert si caratterizza esattamente come la proiezione ortogonale sui sottospazi dello spazio euclideo \mathbb{R}^n , c'è però una differenza fondamentale: in \mathbb{R}^n ogni sottospazio è chiuso *mentre questo non è vero negli spazi vettoriali a dimensione infinita*.

Si chiama **sistema ortonormale** una famiglia $(e_i)_{i \in I}$ di elementi di \mathcal{H} (con I di cardinalità qualsiasi) tale che si abbia

$$\langle e_i, e_j \rangle = \delta_{ij} = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j \end{cases}$$

Se $(e_i)_{i \in I}$ è un sistema ortonormale, dato $x \in \mathcal{H}$ vale la disuguaglianza, detta **disuguaglianza di Bessel**

$$\sum_{i \in I} |\langle x, e_i \rangle|^2 \leq \|x\|^2$$

(in particolare, se I è più che numerabile, al più un sottinsieme numerabile dei numeri $\langle x, e_i \rangle$ è diverso da 0).

Il sistema ortonormale è detto **massimale** o **completo** se non è contenuto propriamente in un sistema ortonormale o, equivalentemente, se il sottospazio generato dai vettori $(e_i)_{i \in I}$ è denso in \mathcal{H} . Se $(e_i)_{i \in I}$ è un sistema ortonormale completo, per ogni $x \in \mathcal{H}$ si ha

$$\|x\|^2 = \sum_{i \in I} |\langle x, e_i \rangle|^2$$

(questa eguaglianza è nota come **identità di Parseval**) e più in generale, dati $x, y \in \mathcal{H}$, si ha

$$\langle x, y \rangle = \sum_{i \in I} \langle x, e_i \rangle \cdot \langle y, e_i \rangle$$

Si dimostra (con l'*assioma della scelta*) che ogni spazio di Hilbert ammette un sistema ortonormale massimale, e questo sistema è numerabile se e solo se lo spazio \mathcal{H} è *separabile* (cioè possiede un sottinsieme denso numerabile). In particolare gli spazi $L^2([a, b], \mathcal{B}([a, b]), \lambda)$ sono separabili.

Un risultato importante (che abbiamo utilizzato nella dimostrazione del Teorema di Radon-Nikodym) è il seguente, noto come **Teorema di rappresentazione di Riesz**. Sia data $L : \mathcal{H} \rightarrow \mathbb{R}$ *lineare e continua* (una funzione L come sopra è chiamata usualmente un **funzionale lineare**): allora esiste $y \in \mathcal{H}$ tale che si abbia, per ogni $x \in \mathcal{H}$, $L(x) = \langle x, y \rangle$

Infine è interessante osservare che se si considera lo spazio delle funzioni definite sull'intervallo $[a, b]$ di quadrato integrabile *secondo Riemann*, munito del prodotto scalare $\langle f, g \rangle = \int_a^b f(x)g(x) dx$, questo è uno *spazio prehilbertiano* ma *non è completo*: uno dei motivi per i quali è stato introdotto l'integrale di Lebesgue è proprio ottenere la completezza degli spazi L^2 (e più in generale degli spazi L^p).

1.5.3 Complementi su misurabilità (e legami con la teoria degli insiemi).

In questa sezione richiamiamo alcune proprietà e curiosità sulla misura secondo Lebesgue, insistendo sui legami con gli assiomi della teoria degli insiemi.

Per usare il linguaggio della probabilità, consideriamo sull'intervallo $[0, 1]$ la σ -algebra di Borel \mathcal{B} e la misura di Lebesgue λ , sia poi \mathcal{L} la σ -algebra degli insiemi misurabili secondo Lebesgue (il *completamento* di \mathcal{B} rispetto a λ).

L'esistenza di un sottinsieme di $[0, 1]$ non misurabile secondo Lebesgue è stata provata per la prima volta da Vitali: è bene insistere sul fatto che

il suo esempio dipende dall'**assioma della scelta** (se non si accetta questo assioma non è possibile esibire esempio di insiemi non misurabili).

Sull'intervallo $[0, 1]$ si può costruire l'insieme di Cantor C (un insieme con la cardinalità del continuo trascurabile secondo Lebesgue) e la misura di Cantor \mathbf{m} (una probabilità *diffusa* concentrata sull'insieme C , quindi *singolare* rispetto a λ). A differenza dell'insieme di Vitali, l'insieme e la misura di Cantor sono il risultato di una costruzione esplicita e *non dipendono dall'assioma della scelta*.

Usualmente in analisi si chiama *misurabile* una funzione $f : \mathbb{R} \rightarrow \mathbb{R}$ che sia misurabile come funzione da $(\mathbb{R}, \mathcal{L})$ in $(\mathbb{R}, \mathcal{B})$ (cioè tale che, per ogni $B \in \mathcal{B}$, $f^{-1}(B) \in \mathcal{L}$). È ben noto che $\mathcal{B} \not\subseteq \mathcal{L}$ e che composizione di due funzioni misurabili non è necessariamente misurabile: esibiamo allora degli esempi espliciti.

In modo perfettamente analogo a quanto fatto per l'intervallo $[0, 1]$, si può costruire sull'intervallo $I = [a, b]$ un insieme di Cantor C_I ed una misura di Cantor \mathbf{m}_I . Consideriamo allora una successione di intervalli $(I_n)_{n \geq 1} = [a_n, b_n]_{n \geq 1}$ tali che gli intervalli aperti $]a_n, b_n[$ formino una *base di aperti* per l'intervallo $[0, 1]$ (siano cioè tali che ogni aperto non vuoto di $[0, 1]$ contenga almeno un $]a_n, b_n[$, ad esempio si possono prendere gli intervalli $[\frac{k}{2^n}, \frac{k+1}{2^n}]$ al variare di $n \geq 1$ e di $0 \leq k < 2^n$).

Per ogni I_n sia C_n il relativo insieme di Cantor e \mathbf{m}_n la relativa probabilità di Cantor, sia poi $\mu = \sum_{n=1}^{+\infty} 2^{-n} \cdot \mathbf{m}_n$ e $A = \bigcup_{n \geq 1} C_n$: A è un boreliano trascurabile secondo Lebesgue e μ è una probabilità diffusa concentrata sull'insieme A (si ha cioè $\lambda(A) = 0$ e $\mu(A) = 1$).

Sia F la funzione di ripartizione di μ ristretta a $[0, 1]$ (cioè $F(x) = \mu([0, x])$): F è continua (perché μ è diffusa), strettamente crescente (perché per ogni intervallo non vuoto I , $\mu(I) > 0$) e quindi è biunivoca da $[0, 1]$ in sé, con inversa $G = F^{-1}$ anche lei continua strettamente crescente.

Osserviamo preliminarmente che μ è l'immagine di λ mediante l'applicazione G : basta verificare questo sugli intervalli $[0, x]$ e si vede facilmente che si ha

$$\mu([0, x]) = F(x) = \lambda([0, F(x)]) = \lambda(F([0, x])) = \lambda(G^{-1}([0, x]))$$

Come conseguenza $\lambda(A) = 0$ e $\lambda(F(A)) = \lambda(G^{-1}(A)) = \mu(A) = 1$.

Sia ora V una parte di $[0, 1]$ non misurabile secondo Lebesgue (ad esempio l'insieme di Vitali) e sia $D = V \cap F(A)$: D non è misurabile (poiché differisce da V per un insieme trascurabile), tuttavia $D^* = F^{-1}(D) = F^{-1}(V) \cap A$ è trascurabile e quindi misurabile. D^* è *un esempio di un insieme misurabile che non è boreliano*: se infatti fosse boreliano lo sarebbe pure $D = G^{-1}(D^*)$.

Consideriamo ora la funzione $g = I_D$ che non è misurabile. Si può scrivere $g = (g \circ F) \circ F^{-1} = (g \circ F) \circ G$, e la funzione $h = g \circ F = I_D \circ F = I_{D^*}$ è misurabile: dunque h è misurabile, g è boreliana e la composizione $h \circ g$ non è misurabile.

L'esistenza di un sottoinsieme di $[0, 1]$ non misurabile secondo Lebesgue può essere rafforzata da un importante teorema dovuto a Banach–Kuratowski: notiamo preliminarmente che per ottenere questo risultato si assume l'*ipotesi del continuo*, senza questa ipotesi il teorema non è dimostrabile.

Cominciamo con una notazione: se $S = (n_1, n_2, \dots)$ e $T = (k_1, k_2, \dots)$ sono due successioni di interi strettamente positivi, scriveremo $S \preceq T$ se, per ogni i , $n_i \leq k_i$. Quando questa proprietà non è verificata scriveremo $S \not\preceq T$.

Lemma 1.5.1. *Esiste un insieme F che ha la cardinalità del continuo, i cui punti sono successioni di interi positivi, e tale che, per ogni successione S (anche non appartenente ad F) l'insieme $\{T \in F \mid T \preceq S\}$ è al più numerabile.*

Dimostrazione. Consideriamo l'insieme delle successioni di interi positivi (che ha cardinalità del continuo) e ordiniamo i suoi elementi $S_0, S_1, \dots, S_\omega, \dots, S_a, \dots$ $a < \bar{\omega}$ (dove $\bar{\omega}$ è il primo ordinale non numerabile).

Preso una successione T_1, T_2, \dots si può costruire una nuova successione S tale che $S \not\preceq T_n \forall n$.

Per ogni ordinale $a < \bar{\omega}$ scegliamo ξ_a tale che, se $b < a$, $S_{\xi_a} \not\preceq S_b$ e $S_{\xi_a} \neq S_{\xi_b}$. L'insieme F è l'insieme delle successioni S_{ξ_a} , $a < \bar{\omega}$ che ha la cardinalità del continuo.

L'insieme F ha la proprietà voluta poiché, se $S_{\xi_a} \preceq S_b$ necessariamente $a \leq b$ (quindi $\{T \in F \mid T \preceq S_b\}$ è al più numerabile). \square

Lemma 1.5.2. *Sia E un insieme con la cardinalità del continuo: si può decomporre*

$$\begin{aligned} E &= A_1^1 \cup A_2^1 \cup \dots \\ E &= A_1^2 \cup A_2^2 \cup \dots \\ &\dots\dots\dots \\ E &= A_1^i \cup A_2^i \cup \dots \end{aligned}$$

dove, per ogni i , A_1^i, A_2^i, \dots è una partizione di E e, presa una qualunque successione di interi positivi k_1, k_2, \dots

$$\bigcap_{i=1}^{+\infty} (A_1^i \cup A_2^i \cup \dots \cup A_{k_i}^i)$$

è al più numerabile.

Dimostrazione. Stabiliamo una corrispondenza biunivoca $x \leftrightarrow T_x$ tra i punti di E ed i punti di F : scriviamo $T_x = n_1^x, n_2^x, \dots$.

Definiamo gli insiemi A_k^i in questo modo: $x \in A_k^i$ se $n_i^x = k$. In questo modo gli insiemi A_1^i, A_2^i, \dots formano effettivamente una partizione di E .

Notiamo ancora che $x \in (A_1^i \cup \dots \cup A_{k_i}^i)$ se $n_i^x \leq k_i$, quindi posto $S = k_1, k_2, \dots$

$$\bigcap_{i=1}^{+\infty} (A_1^i \cup A_2^i \cup \dots \cup A_{k_i}^i) = \{x \in E \mid T_x \preceq S\}$$

è al più numerabile. □

Teorema 1.5.3 (Banach-Kuratowski). *Assumiamo l'ipotesi del continuo, e sia E un insieme con la cardinalità del continuo: non può esistere una probabilità diffusa \mathbf{m} definita su tutti i sottinsiemi di E .*

Dimostrazione. Supponiamo l'esistenza di questa \mathbf{m} e proviamo che si arriva a un assurdo. Per ogni $i \geq 1$, esiste k_i tale che, posto $R^i = A_{k_i+1}^i \cup A_{k_i+2}^i \cup A_{k_i+3}^i \cup \dots$ si abbia $\mathbf{m}(R^i) \leq 2^{-(i+1)}$.

Dunque $\mathbf{m}\left(\bigcup_{i=1}^{+\infty} R^i\right) \leq 1/2$ e di conseguenza $\mathbf{m}\left(\bigcup_{i=1}^{+\infty} R^i\right)^c \geq 1/2$. Questo però è impossibile poiché

$$\left(\bigcup_{i=1}^{+\infty} R^i\right)^c = \bigcap_{i=1}^{+\infty} (A_1^i \cup A_2^i \cup \dots \cup A_{k_i}^i)$$

è al più numerabile. □

Capitolo 2

Indipendenza di eventi e di variabili aleatorie.

2.1 Complementi su misurabilità di variabili aleatorie.

Dati due spazi misurabili (E, \mathcal{E}) ed (F, \mathcal{F}) , ricordiamo che una funzione $X : E \rightarrow F$ è detta *misurabile* se, $\forall A \in \mathcal{F}$, $X^{-1}(A) \in \mathcal{E}$. Questa condizione può essere scritta nel modo seguente: $X^{-1}(\mathcal{F}) \subseteq \mathcal{E}$, intendendo $X^{-1}(\mathcal{F}) = \{X^{-1}(A) \mid A \in \mathcal{F}\}$.

Notiamo che $X^{-1}(\mathcal{F})$ è una σ -algebra. Inoltre sappiamo che, se \mathcal{I} è una famiglia di parti di F che genera \mathcal{F} , affinché X sia misurabile è sufficiente che, per ogni $B \in \mathcal{I}$, $X^{-1}(B)$ appartenga a \mathcal{E} .

Data una funzione $X : E \rightarrow F$, si indica $\sigma(X) = X^{-1}(\mathcal{F})$ la *più piccola* σ -algebra di E che rende misurabile X (σ -algebra generata da X).

Lemma 2.1.1 (Criterio di misurabilità di Doob). *Siano (Ω, \mathcal{F}) e (E, \mathcal{E}) due spazi misurabili, $X : \Omega \rightarrow E$, e supponiamo che $\mathcal{F} = X^{-1}(\mathcal{E})$ sia la σ -algebra generata da X : ogni variabile aleatoria reale $Y : \Omega \rightarrow \mathbb{R}$ (\mathcal{F} -misurabile) si fattorizza nella forma $Y = g \circ X$ dove $g : E \rightarrow \mathbb{R}$ è una opportuna funzione misurabile.*

Dimostrazione. Il primo passo è verificare che l'affermazione è vera se Y è l'indicatrice di un elemento $B \in \mathcal{F}$: $B = X^{-1}(A)$ con $A \in \mathcal{E}$. Questo è una conseguenza immediata della eguaglianza $I_B = I_{X^{-1}(A)} = I_A \circ X$.

La proprietà è allora verificata se Y è semplice, e per Y misurabile a valori positivi consideriamo $(Y_n)_{n \geq 1}$ con $Y_n \uparrow Y$.

Per ogni n vale l'eguaglianza $Y_n = g_n(X)$: non è detto che anche la successione di funzioni $(g_n)_{n \geq 1}$ sia crescente, tuttavia si ha

$$Y = \sup_n Y_n = \sup_n (g_n \circ X) = \limsup_n (g_n \circ X) = \left(\limsup_n g_n \right) \circ X$$

e la proprietà è pertanto soddisfatta. Per una Y di segno qualsiasi, si decompone $Y = Y^+ - Y^-$. \square

Supponiamo assegnata ora una famiglia di spazi misurabili (E_i, \mathcal{E}_i) indicizzati da $i \in I$, (dove l'insieme degli indici I può essere di cardinalità qualsiasi), sia E il prodotto cartesiano $E = \prod_{i \in I} E_i$ e sia π_i la proiezione canonica di indice i da E su E_i .

Definizione 2.1.2. Si chiama σ -algebra prodotto su E (indicata $\bigotimes_{i \in I} \mathcal{E}_i$) la più piccola σ -algebra che rende misurabili le proiezioni canoniche π_i , ossia quella generata dagli insiemi della forma $\pi_i^{-1}(A_i)$, al variare di $i \in I$ e di $A_i \in \mathcal{E}_i$.

Una conseguenza immediata è che una funzione $X : \Omega \rightarrow E$ è una variabile aleatoria (cioè è misurabile) se e solo se, per ogni i , $X_i = \pi_i \circ X$ è misurabile (a valori in (E_i, \mathcal{E}_i)).

In particolare, assegnata una famiglia di v.a. $(X_i)_{i \in I}$ a valori ciascuna in uno spazio (E_i, \mathcal{E}_i) la funzione $\mathbf{X}_I = (X_i)_{i \in I}$ (a valori nel prodotto cartesiano) è una variabile aleatoria.

La σ -algebra prodotto è generata anche dagli insiemi della forma

$$\pi_{i_1}^{-1}(A_1) \cap \dots \cap \pi_{i_k}^{-1}(A_k)$$

al variare dei sottinsiemi finiti di indici i_1, \dots, i_k e di $A_h \in \mathcal{E}_{i_h}$: gli insiemi di questa forma (usualmente chiamati *rettangoli cilindrici*) sono *stabili per intersezione* e dunque due probabilità che coincidono su di essi sono eguali.

Se I è una famiglia finita $I = \{1, \dots, n\}$, la σ -algebra prodotto è generata anche dai *rettangoli misurabili*

$$A_1 \times \dots \times A_n = \pi_1^{-1}(A_1) \cap \dots \cap \pi_n^{-1}(A_n)$$

(con $A_i \in \mathcal{E}_i$). La stessa proprietà vale anche se I è numerabile, ma se la cardinalità di I è più che numerabile, in generale *non è vero* che, presi $A_i \in \mathcal{E}_i$, il prodotto cartesiano $\prod_{i \in I} A_i$ sia misurabile.

2.2 Indipendenza e prime proprietà.

D'ora innanzi supponiamo assegnato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$ e ricordiamo che si chiamano *eventi* gli elementi di \mathcal{F} .

Definizione 2.2.1. Due eventi A e B sono detti *indipendenti* se vale la relazione $\mathbf{P}(A \cap B) = \mathbf{P}(A) \cdot \mathbf{P}(B)$.

Consideriamo ora sullo spazio Ω delle σ -algebre $\mathcal{F}_1, \dots, \mathcal{F}_n$ tutte contenute nella σ -algebra \mathcal{F} sulla quale è definita la probabilità \mathbf{P} .

Definizione 2.2.2. Le σ -algebre $\mathcal{F}_1, \dots, \mathcal{F}_n$ sono dette *indipendenti* se, scelti comunque $A_1 \in \mathcal{F}_1, \dots, A_n \in \mathcal{F}_n$ si ha

$$\mathbf{P}(A_1 \cap \dots \cap A_n) = \mathbf{P}(A_1) \dots \mathbf{P}(A_n)$$

Osserviamo che la σ -algebra generata da un evento A (indicata anche $\sigma(A)$) è formata da $\{\emptyset, A, A^c, \Omega\}$ e di conseguenza A e B sono indipendenti se e solo se lo sono le σ -algebre $\sigma(A)$ e $\sigma(B)$: è naturale estendere la definizione di indipendenza a una famiglia finita di eventi A_1, \dots, A_n dicendo che sono indipendenti se lo sono le σ -algebre da essi generate e si arriva alla seguente caratterizzazione la cui facile dimostrazione è lasciata per esercizio:

Proposizione 2.2.3. *Gli n eventi A_1, \dots, A_n sono indipendenti se e solo se, per ogni sottoinsieme di indici $1 \leq i_1 < \dots < i_k \leq n$ vale l'eguaglianza*

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

Definizione 2.2.4. Le variabili aleatorie X_1, \dots, X_n (a valori anche in spazi diversi (E_i, \mathcal{E}_i)) sono dette indipendenti se lo sono le σ -algebre da esse generate $X_1^{-1}(\mathcal{E}_1), \dots, X_n^{-1}(\mathcal{E}_n)$.

Una conseguenza immediata è che, se X_1, \dots, X_n (a valori in $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$) sono indipendenti, e per ogni i , $f_i : E_i \rightarrow F_i$ è una funzione misurabile, anche $f_1 \circ X_1, \dots, f_n \circ X_n$ sono indipendenti.

Il seguente criterio, di facile dimostrazione, sarà tuttavia di grande importanza: viene enunciato per due v.a. ma vale con la stessa dimostrazione per un numero finito di variabili aleatorie.

Proposizione 2.2.5. *Siano X_1, X_2 due v.a. a valori rispettivamente in (E_1, \mathcal{E}_1) e (E_2, \mathcal{E}_2) e sia, per ogni i , \mathcal{I}_i una famiglia di parti stabile per l'intersezione che genera \mathcal{E}_i : affinché X_1 e X_2 siano indipendenti è sufficiente che valga l'eguaglianza*

$$\mathbf{P}(X_1^{-1}(B_1) \cap X_2^{-1}(B_2)) = \mathbf{P}(X_1^{-1}(B_1)) \cdot \mathbf{P}(X_2^{-1}(B_2))$$

per ogni scelta di $B_1 \in \mathcal{I}_1$ e $B_2 \in \mathcal{I}_2$.

Dimostrazione. Si fissa $B_2 \in \mathcal{I}_2$ e si estende l'eguaglianza sopra scritta dagli elementi $B_1 \in \mathcal{I}_1$ a tutti gli elementi di \mathcal{E}_1 utilizzando il *teorema delle classi monotone* (i dettagli sono facili), quindi si estende con lo stesso metodo agli elementi $B_2 \in \mathcal{E}_2$. \square

Definizione 2.2.6. Sia $(X_i)_{i \in I}$ una famiglia *infinita* di variabili aleatorie: queste sono dette indipendenti se, per ogni sottinsieme finito di indici i_1, \dots, i_k le variabili X_{i_1}, \dots, X_{i_k} sono indipendenti

Ricordiamo che, se $J \subseteq I$, indichiamo con \mathbf{X}_J la variabile aleatoria $(X_i)_{i \in J}$ (a valori nello spazio prodotto $(\prod_{i \in J} E_i, \otimes_{i \in J} \mathcal{E}_i)$).

Teorema 2.2.7 (Criterio di Indipendenza per una famiglia infinita). Sia $(X_i)_{i \in I}$ una famiglia infinita di variabili aleatorie. Sono equivalenti le seguenti affermazioni:

- a) per ogni sottinsieme finito di indici i_1, \dots, i_k le variabili X_{i_1}, \dots, X_{i_k} sono indipendenti
- b) per ogni scelta di sottinsiemi disgiunti J_1, \dots, J_k di I , le variabili $\mathbf{X}_{J_1}, \dots, \mathbf{X}_{J_k}$ sono indipendenti.

Dimostrazione. È evidente che la condizione b) è più forte della condizione a); vediamo ora che la condizione a) implica b) limitandoci per semplicità di esposizione al caso di due sottinsiemi disgiunti J e K dell'insieme degli indici I (anche infiniti). Dobbiamo provare che scelti comunque $B_1 \in \otimes_{i \in J} \mathcal{E}_i$ e $B_2 \in \otimes_{j \in K} \mathcal{E}_j$ vale l'eguaglianza

$$\mathbf{P}(\mathbf{X}_J^{-1}(B_1) \cap \mathbf{X}_K^{-1}(B_2)) = \mathbf{P}(\mathbf{X}_J^{-1}(B_1)) \cdot \mathbf{P}(\mathbf{X}_K^{-1}(B_2))$$

ma, in base al criterio della Proposizione 2.2.5, è sufficiente verificare questa eguaglianza se B_1 e B_2 appartengono a due sottinsiemi stabili per l'intersezione che generano rispettivamente $\otimes_{i \in J} \mathcal{E}_i$ e $\otimes_{j \in K} \mathcal{E}_j$.

Indicando con π_i la proiezione canonica sullo spazio E_i , consideriamo una sottofamiglia finita i_1, \dots, i_s di J e prendiamo al posto di B_1 un insieme della forma $\pi_{i_1}^{-1}(A_1) \cap \dots \cap \pi_{i_s}^{-1}(A_s)$ (con $A_h \in \mathcal{E}_{i_h}$), e analoga è la scelta per B_2 .

Nel verificare che è soddisfatta l'eguaglianza sopra scritta ci si riduce al caso di un numero finito di variabili aleatorie, cioè alla condizione a). \square

Esempio 2.2.8. Tenendo conto del fatto che *funzioni di variabili indipendenti sono indipendenti*, se $(X_n)_{n \geq 1}$ è una successione di v.a.r. indipendenti, sono indipendenti le due variabili $Y = \limsup_n X_{2n}$ e $Z = \liminf_n X_{2n+1}$.

2.3 Costruzione di una Probabilità Prodotto.

Siano $(E, \mathcal{E}, \mathbf{P})$ e $(F, \mathcal{F}, \mathbf{Q})$ due spazi di probabilità: il nostro scopo è costruire sullo spazio prodotto $E \times F$ munito della σ -algebra prodotto $\mathcal{E} \otimes \mathcal{F}$ una probabilità (indicata $\mathbf{P} \otimes \mathbf{Q}$) tale che si abbia, per ogni *rettangolo misurabile* $A \times B$, $\mathbf{P} \otimes \mathbf{Q}(A \times B) = \mathbf{P}(A) \cdot \mathbf{Q}(B)$. Una tale probabilità, se esiste, è naturalmente unica.

Proposizione 2.3.1. *Sia $f : E \times F \rightarrow \mathbb{R}$ misurabile a valori positivi. Allora:*

- per ogni $x \in E$ fissato, la funzione $y \rightarrow f(x, y)$ è \mathcal{F} -misurabile
- la funzione $x \rightarrow \int_F f(x, y) d\mathbf{Q}(y)$ è \mathcal{E} -misurabile.

Il primo di questi due enunciati si estende poi (per differenza) ad f misurabile di segno qualsiasi, ed il secondo ad f misurabile *limitata* (deve infatti avere senso l'integrale).

Dimostrazione. Le due proprietà sono una verifica diretta ed immediata se f è la funzione indicatrice di un *rettangolo misurabile* $A \times B$, e si estendono poi alla funzione indicatrice di un insieme misurabile $C \in \mathcal{E} \otimes \mathcal{F}$ utilizzando il *teorema delle classi monotone* (tralasciamo la facile verifica dei dettagli).

Di conseguenza i due enunciati sono soddisfatti se f è semplice; nel caso di f misurabile positiva si considera una successione $(f_n)_{n \geq 1}$ di funzioni semplici tale che $f_n \uparrow f$ e, per provare la seconda proprietà, si utilizza la proprietà di Beppo-Levi (Lemma 1.2.5). \square

Se C è un sottinsieme di $E \times F$, ricordiamo che si chiama *sezione* (nel punto $x \in E$), l'insieme $C_x = \{y \in F \mid (x, y) \in C\}$.

Teorema 2.3.2 (Fubini–Tonelli). *Definiamo, per $C \in \mathcal{E} \otimes \mathcal{F}$,*

$$\mathbf{R}(C) = \int_E \mathbf{Q}(C_x) d\mathbf{P}(x)$$

La funzione d'insieme \mathbf{R} è una probabilità, ed è proprio la probabilità cercata $\mathbf{P} \otimes \mathbf{Q}$. Inoltre per ogni funzione f definita su $E \times F$ misurabile a valori positivi, vale la formula

$$\int \int_{E \times F} f(x, y) d\mathbf{P} \otimes \mathbf{Q}(x, y) = \int_E \left[\int_F f(x, y) d\mathbf{Q}(y) \right] d\mathbf{P}(x)$$

Dimostrazione. La funzione d'insieme (definita su $\mathcal{E} \otimes \mathcal{F}$) $C \rightarrow \mathbf{R}(C)$ è additiva, va al limite sulle successioni crescenti d'insiemi (quindi è σ -additiva) e $\mathbf{R}(E \times F) = 1$: è dunque una probabilità e una verifica diretta mostra che sui *rettangoli misurabili* $A \times B$ vale $\mathbf{P}(A) \cdot \mathbf{Q}(B)$. Dunque \mathbf{R} è la probabilità prodotto.

La formula per il calcolo dell'integrale rispetto a $\mathbf{P} \otimes \mathbf{Q}$ è soddisfatta se f è l'indicatrice di un insieme $C \in \mathcal{E} \otimes \mathcal{F}$: secondo il procedimento ormai usuale, si estende alle funzioni semplici e quindi alle funzioni misurabili positive utilizzando la proprietà di Beppo-Levi. \square

A una funzione misurabile f di segno qualsiasi si può applicare la formula di Fubini-Tonelli *dopo aver verificato che è integrabile*, cioè che si ha

$$\int \int |f(x, y)| d\mathbf{P} \otimes \mathbf{Q} = \int \left[\int |f(x, y)| d\mathbf{Q}(y) \right] d\mathbf{P}(x) < +\infty$$

Naturalmente si può invertire l'ordine rispetto al quale si integra, inoltre il Teorema 2.3.2 si estende con piccole modifiche di notazione al prodotto di un numero finito di Probabilità.

Osservazione 2.3.3. La dimostrazione precedente rimane valida, praticamente senza modifiche, se invece di due probabilità si considerano due misure σ -finite μ e ν rispettivamente su (E, \mathcal{E}) e (F, \mathcal{F}) .

Osservazione 2.3.4. La dimostrazione del teorema 2.3.2 così enunciata è facile perchè si limita a considerare la probabilità prodotto sulla σ -algebra prodotto $\mathcal{E} \otimes \mathcal{F}$ e non sul *completamento* di questa σ -algebra: se si considera il completamento, infatti, sia l'enunciato che la dimostrazione diventano più delicati. Tuttavia per gli scopi del Calcolo delle Probabilità questa versione è sufficiente.

La nozione di probabilità prodotto è importante per enunciare la nozione di indipendenza in termini di *leggi di probabilità*: è facile infatti provare i seguenti enunciati (per semplicità ci limitiamo al caso di due variabili, ma sono veri per una famiglia finita di v.a.):

- due v.a. X e Y sono indipendenti se e solo se si ha $\mathbf{P}_{(X,Y)} = \mathbf{P}_X \otimes \mathbf{P}_Y$;
- X e Y sono indipendenti se e solo se, scelte comunque f e g misurabili limitate, si ha

$$\mathbf{E}[f(X)g(Y)] = \mathbf{E}[f(X)] \mathbf{E}[g(Y)].$$

Il teorema di Fubini–Tonelli ha una estensione a un prodotto infinito di Probabilità: la dimostrazione (limitatamente al caso di un prodotto numerabile) è rinviata all'Appendice.

Consideriamo una famiglia infinita (non necessariamente numerabile) di spazi di probabilità $(E_i, \mathcal{E}_i, \mathbf{P}_i)$, e indichiamo con π_i la proiezione canonica dal prodotto $\prod_{j \in I} E_j$ su E_i : l'enunciato seguente è un caso particolare di un teorema di A. e C. Ionescu-Tulcea.

Teorema 2.3.5. *Esiste sullo spazio prodotto $\prod_{i \in I} E_i$ munito della σ -algebra prodotto $\bigotimes_{i \in I} \mathcal{E}_i$ una ed una sola probabilità \mathbf{R} (indicata $\bigotimes_{i \in I} \mathbf{P}_i$) tale che, per ogni scelta di un numero finito di indici i_1, \dots, i_n e di $A_h \in \mathcal{E}_{i_h}$ valga l'eguaglianza*

$$\mathbf{R}(\pi_{i_1}^{-1}(A_1) \cap \dots \cap \pi_{i_n}^{-1}(A_{i_n})) = \mathbf{P}_{i_1}(A_1) \dots \mathbf{P}_{i_n}(A_n).$$

Questo risultato è importante, ad esempio, per costruire una successione di v.a.r. indipendenti X_1, X_2, \dots con leggi di probabilità rispettivamente $\mathbf{P}_1, \mathbf{P}_2, \dots$. Si considera infatti sullo spazio prodotti $\mathbb{R}^{\mathbb{N}} = \prod_{n \geq 1} \mathbb{R}_n$ la probabilità prodotto $\mathbf{P}_1 \otimes \mathbf{P}_2 \otimes \dots$ e si indica con X_n la proiezione canonica di indice n : è facile verificare che queste variabili sono indipendenti ed hanno la legge voluta.

2.4 Legge 0–1 di Kolmogorov e lemmi di Borel–Cantelli

Consideriamo una successione $(X_n)_{n \geq 1}$ di variabili aleatorie reali *indipendenti* ed indichiamo con $\mathcal{F}_n = \sigma(X_n, X_{n+1}, X_{n+2}, \dots)$ la σ -algebra generata dalle variabili X_i con $i \geq n$ (il *futuro* rispetto all'istante n): chiamiamo poi σ -algebra **terminale** $\mathcal{F}_\infty = \bigcap_{n \geq 1} \mathcal{F}_n$. Una variabile aleatoria Z è *terminale* (cioè misurabile rispetto a \mathcal{F}_∞) se e solo se, per ogni n , è \mathcal{F}_n -misurabile, ossia (per il criterio di misurabilità di Doob) può essere scritta nella forma $Z = g_n(X_n, X_{n+1}, \dots)$.

Ad esempio, è terminale l'evento $\{\omega \in \Omega \mid \sum_{n \geq 1} |X_n(\omega)| < +\infty\}$, mentre non è terminale l'evento $\{\omega \in \Omega \mid \sum_{n \geq 1} |X_n(\omega)| \leq 1\}$.

Teorema 2.4.1 (Legge 0–1 di Kolmogorov). *Sia $A \in \mathcal{F}_\infty$: allora $\mathbf{P}(A)$ può assumere solo i valori 0 o 1.*

Quindi la σ -algebra terminale contiene solo gli eventi trascurabili oppure i complementari dei trascurabili.

Dimostrazione. Chiamiamo $X_\infty = I_A$ e proviamo che le variabili $X_\infty, X_1, X_2, \dots$ sono indipendenti: ci si riduce al caso di una sottofamiglia finita (poniamo $X_\infty, X_1, \dots, X_k$) e la conclusione segue dal fatto che si può scrivere $X_\infty = g_{k+1}(X_{k+1}, X_{k+2}, \dots)$ (e dal fatto che funzioni di variabili indipendenti sono indipendenti).

Ma poichè si può scrivere $X_\infty = g_1(X_1, X_2, \dots)$, segue che X_∞ è indipendente da sé stessa: un evento A è indipendente da sé stesso se si ha $\mathbf{P}(A) = \mathbf{P}(A)^2$ (e gli unici numeri che verificano questa condizione sono 0 e 1). \square

Sia ora $(A_n)_{n \geq 1}$ una successione di eventi: definiamo **limite superiore** di questa successione di eventi l'insieme

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n \geq 1} \bigcup_{k \geq n} A_k = \{\omega \in \Omega \mid \omega \text{ appartiene infinite volte all'insieme } A_n\}$$

Notiamo che, se $A = \limsup_n A_n$, si ha $I_A = \limsup_n I_{A_n}$, e di conseguenza per il Lemma di Fatou si ha

$$\limsup_{n \rightarrow \infty} \mathbf{P}(A_n) \leq \mathbf{P}(\limsup_{n \rightarrow \infty} A_n)$$

(Naturalmente analoga è la definizione di *limite inferiore* di una successione di eventi).

Sia dunque $(A_n)_{n \geq 1}$ una successione di eventi, e sia $A = \limsup_n A_n$.

Lemma 2.4.2 (Borel–Cantelli, I parte). *Se $\sum_{n \geq 1} \mathbf{P}(A_n) < \infty$, l'insieme A è trascurabile.*

Dimostrazione. Notiamo che la v.a. a valori positivi $Z = \sum_{n \geq 1} I_{A_n}$ è integrabile, in quanto $\mathbf{E}[Z] = \sum_{n \geq 1} \mathbf{P}(A_n)$ e di conseguenza $\{Z = +\infty\}$ è trascurabile.

Ora è facile constatare che si ha $A = \{\omega \mid \sum_{n \geq 1} I_{A_n}(\omega) = +\infty\}$. \square

Lemma 2.4.3 (Borel–Cantelli, II parte). *Supponiamo che gli eventi $(A_n)_{n \geq 1}$ siano indipendenti e che $\sum_{n \geq 1} \mathbf{P}(A_n) = +\infty$: allora $\mathbf{P}(A) = 1$.*

Osserviamo preliminarmente che A è un evento *terminale* e di conseguenza la sua probabilità può essere solo 0 o 1: questo lemma specifica sotto quale condizione $\mathbf{P}(A) = 1$.

Dimostrazione. È più comodo considerare la probabilità dell'evento complementare

$$A^c = \bigcup_{n \geq 1} \bigcap_{k \geq n} A_k^c$$

e si tratta di provare che, per ogni n , $\bigcap_{k \geq n} A_k^c$ è trascurabile.

$$\mathbf{P}\left(\bigcap_{k \geq n} A_k^c\right) = \lim_{m \rightarrow \infty} \mathbf{P}\left(\bigcap_{n \leq k \leq m} A_k^c\right) = \lim_{m \rightarrow \infty} \prod_{k=n}^m \left(1 - \mathbf{P}(A_k)\right)$$

Passando ai logaritmi, ed utilizzando la disuguaglianza (valida per $0 \leq x \leq 1$) $\log(1 - x) \leq -x$, si ottiene

$$\sum_{k=n}^m \log\left(1 - \mathbf{P}(A_k)\right) \leq \sum_{k=n}^m -\mathbf{P}(A_k) \downarrow -\infty$$

□

Osservazione 2.4.4. In realtà il Lemma 2.4.3 è valido sotto l'ipotesi più debole che gli eventi $(A_n)_{n \geq 1}$ siano *a due a due indipendenti*: la dimostrazione è meno immediata e viene qui riportata schematicamente.

Sia $Z_n = \sum_{k=1}^n I_{A_k}$ (che converge crescendo a $Z = \sum_{n \geq 1} I_{A_n}$) e notiamo preliminarmente che $\mathbf{E}[Z_n] \geq \text{Var}[Z_n]$ e che $\lim_{n \rightarrow \infty} \mathbf{E}[Z_n] = +\infty$.

Prendiamo $\lambda > 0$ ed n tale che si abbia $\lambda \leq \mathbf{E}[Z_n]$: per la disuguaglianza di Chebishev

$$\begin{aligned} \mathbf{P}\{Z_n \leq \lambda\} &= \mathbf{P}\{Z_n - \mathbf{E}[Z_n] \leq \lambda - \mathbf{E}[Z_n]\} \leq \\ &\leq \mathbf{P}\{|Z_n - \mathbf{E}[Z_n]| \geq \mathbf{E}[Z_n] - \lambda\} \leq \frac{\text{Var}(Z_n)}{(\mathbf{E}[Z_n] - \lambda)^2} \end{aligned}$$

e da qui segue che $\mathbf{P}\{Z \leq \lambda\} = \lim_{n \rightarrow \infty} \mathbf{P}\{Z_n \leq \lambda\} = 0$ e quindi Z è q.c. eguale a $+\infty$.

2.5 Appendice

Vediamo ora la dimostrazione del Teorema 2.3.5 limitandoci al caso del prodotto di una *famiglia numerabile di Probabilità*: come è stato anticipato, questo risultato è valido anche con un prodotto di *cardinalità qualsiasi*, ma in tal caso la dimostrazione è ulteriormente più tecnica (e basata su risultati di A. e C. Ionescu-Tulcea).

Supponiamo dunque assegnata una successione di spazi di Probabilità $(E_n, \mathcal{E}_n, \mathbf{P}_n)$, $n = 1, 2, \dots$ e chiamiamo \mathcal{A} la famiglia dei cosiddetti *rettangoli cilindrici* su $\prod_{n \geq 1} E_n$, cioè gli insiemi della forma

$$\pi_1^{-1}(A_1) \cap \dots \cap \pi_n^{-1}(A_n) = A_1 \times \dots \times A_n \times E_{n+1} \times E_{n+2} \dots$$

al variare di $n \in \mathbb{N}$ e di $A_i \in \mathcal{E}_i$.

È facile verificare che \mathcal{A} genera la σ -algebra prodotto $\otimes_{n \geq 1} \mathcal{E}_n$ ed è stabile per intersezione, ma *non è un'algebra*: pertanto non si può applicare direttamente il Teorema 1.1.4 per prolungare a tutta la σ -algebra prodotto la funzione d'insieme definita su \mathcal{A} dalla formula

$$\mathbf{R}(\pi_1^{-1}(A_1) \cap \dots \cap \pi_n^{-1}(A_n)) = \mathbf{P}_1(A_1) \dots \mathbf{P}_n(A_n)$$

Occorre introdurre una nuova classe d'insiemi, gli *insiemi cilindrici*: si chiama **insieme cilindrico** un insieme della forma $C = \pi_J^{-1}(A)$ dove $J = \{i_1, \dots, i_n\} \subset \mathbb{N}$ è un insieme finito di indici ed $A \subseteq E_{i_1} \times \dots \times E_{i_n}$ appartiene alla σ -algebra prodotto $\mathcal{E}_{i_1} \otimes \dots \otimes \mathcal{E}_{i_n}$.

Si pone allora

$$\mathbf{R}(\pi_J^{-1}(A)) = \mathbf{P}_{i_1} \otimes \dots \otimes \mathbf{P}_{i_n}(A)$$

ed è immediato constatare che \mathbf{R} così definita è finitamente additiva sulla famiglia \mathcal{C} degli insiemi cilindrici e vale 1 sull'intero spazio prodotto.

Teorema 2.5.1. *La funzione d'insieme \mathbf{R} si prolunga (in un sol modo) ad una probabilità definita sulla σ -algebra $\otimes_{n \geq 1} \mathcal{E}_n$.*

Dimostrazione. Grazie al Teorema di prolungamento 1.1.4 è sufficiente provare che, presa una successione $(C_n)_{n \geq 1}$ di elementi di \mathcal{C} (cioè di *insiemi cilindrici*) tale che $C_n \downarrow \emptyset$, allora $\mathbf{R}(C_n) \downarrow 0$. Viene più agevole provare che, se esiste $\varepsilon > 0$ tale che si abbia $\mathbf{R}(C_n) \geq \varepsilon$ per ogni n , allora $C = \lim_n C_n = \bigcap_{n \geq 1} C_n$ non è vuoto.

Introduciamo alcune notazioni: chiamiamo $E = \prod_{i \geq 1} E_i = E_1 \times E_2 \times \dots$ e, per $n \geq 1$, poniamo $E^{(n)} = \prod_{i > n} E_i = E_{n+1} \times E_{n+2} \times \dots$

Definiamo su $E^{(n)}$ la funzione d'insieme $\mathbf{R}^{(n)}$ definita in modo analogo sugli insiemi cilindrici di $E^{(n)}$ (come prodotto delle probabilità $\mathbf{P}_{n+1}, \mathbf{P}_{n+2}, \dots$)

Se C è un sottinsieme di E e $x \in E_1$, indichiamo con $C^{(1)}(x)$ la *sezione* di C in $E^{(1)}$ cioè $C^{(1)}(x) = \{y \in E^{(1)} \mid (x, y) \in C\}$.

Notiamo ancora che se C è un insieme cilindrico di E , le sezioni $C^{(1)}(x)$ sono insiemi cilindrici in $E^{(1)}$ e vale la formula

$$\mathbf{R}(C) = \int_{E_1} \mathbf{R}^{(1)}(C^{(1)}(x)) \, d\mathbf{P}_1(x)$$

Questa formula è dovuta al fatto che, ristretta agli insiemi della forma $C = \pi_J^{-1}(A)$ la funzione \mathbf{R} coincide con $\mathbf{P}_{i_1} \otimes \dots \otimes \mathbf{P}_{i_n}$ e si può pertanto applicare il Teorema di Fubini-Tonelli.

Prendiamo dunque una successione decrescente $(C_n)_{n \geq 1}$ di insiemi cilindrici, sia $C = \lim_n C_n = \bigcap_{n \geq 1} C_n$ e supponiamo che esista $\varepsilon > 0$ tale che si abbia, per ogni n , $\mathbf{R}(C_n) \geq \varepsilon$.

La successione di funzioni $x \rightarrow \mathbf{R}(C_n^{(1)}(x))$ è decrescente (rispetto ad n) e poiché

$$\lim_{n \rightarrow \infty} \int_{E_1} \mathbf{R}^{(1)}(C_n^{(1)}(x)) \, d\mathbf{P}_1(x) \geq \varepsilon$$

è evidente che esiste $x_1 \in E_1$ tale che si abbia $\mathbf{R}^{(1)}(C_n^{(1)}(x_1)) \geq \varepsilon$ per ogni n , in particolare le sezioni $C_n^{(1)}(x_1)$ non sono vuote.

Si fissa allora questo punto x_1 e si ripete il ragionamento con le sezioni $C_n^{(1)}(x_1)$ e così via ... in questo modo si determina un elemento di E (cioè (x_1, x_2, \dots)) che appartiene ad ognuno dei cilindri C_n cioè $\bigcap_{n \geq 1} C_n \neq \emptyset$. \square

Capitolo 3

Le funzioni caratteristiche.

3.1 Definizione e prime proprietà.

In questo capitolo useremo ripetutamente l'integrale di una funzione a valori complessi: se (E, \mathcal{E}, μ) è uno spazio munito di una misura σ -finita, una funzione definita su E a valori complessi $f = f_1 + i f_2$ è detta *misurabile* se lo sono entrambe le componenti f_1 ed f_2 , è detta *integrabile* se lo sono f_1 ed f_2 ed in tal caso l'integrale di f è il numero complesso $\int f d\mu = \int f_1 d\mu + i \int f_2 d\mu$. Valgono tutte le proprietà usuali dell'integrale e vale anche la disuguaglianza $\left| \int f d\mu \right| \leq \int |f| d\mu$: quest'ultima proprietà è di verifica immediata tenendo conto del fatto che, per un numero complesso z , si ha $|z| = \sup \operatorname{Re}(\alpha \cdot z)$ dove il *sup* è preso su tutti i numeri complessi α con $|\alpha| = 1$.

L'esposizione si limita, per semplicità di notazioni, al caso delle v.a. a valori reali, ma tutti i risultati di questo capitolo rimangono validi per variabili vettoriali a valori in \mathbb{R}^n (con piccole modifiche nelle dimostrazioni). Sia dunque X una v.a.r., $\mathbf{P}_X = X(\mathbf{P})$ la sua *legge di probabilità* ed $F_X(\cdot)$ la sua funzione di ripartizione (se non c'è pericolo di confusione, scriveremo semplicemente F).

Definizione 3.1.1. Si chiama **funzione caratteristica** della v.a. la funzione $\varphi_X(\cdot)$ definita da

$$\varphi_X(t) = \mathbf{E}[e^{itX}] = \int_{\mathbb{R}} e^{itx} d\mathbf{P}_X(x)$$

Come si vede, la funzione caratteristica dipende solo dalla *legge di probabilità* della v.a. X . Nel caso di una variabile vettoriale $\mathbf{X} = (X_1, \dots, X_n)$, la sua funzione caratteristica è definita, per $\mathbf{u} \in \mathbb{R}^n$, da $\varphi_{\mathbf{X}}(\mathbf{u}) = \mathbf{E}[e^{i\langle \mathbf{u}, \mathbf{X} \rangle}]$.

Sono di dimostrazione immediata queste prime proprietà:

- per ogni t , $|\varphi_X(t)| \leq 1$ e $\varphi_X(0) = 1$; inoltre $t \rightarrow \varphi_X(t)$ è una funzione continua (anzi, uniformemente continua);
- se a, b sono costanti, $\varphi_{aX+b}(t) = \varphi_X(at) e^{itb}$;
- se X e Y sono *indipendenti*, $\varphi_{X+Y}(t) = \varphi_X(t) \varphi_Y(t)$.

Osservazione 3.1.2 (Legami con la trasformata di Fourier). Supponiamo che la v.a. X abbia una densità f : allora la sua f.c. è $\varphi_X(t) = \int_{\mathbb{R}} e^{itx} f(x) dx$ mentre la *trasformata di Fourier* della funzione f (quando esiste) è definita da $\hat{f}(t) = \int_{\mathbb{R}} e^{-itx} f(x) dx$.

Questa apparente diversità di definizione è dovuta al fatto che la f.c. è in realtà la *trasformata di Fourier della misura che ha densità f rispetto alla misura di Lebesgue*.

L'operatore *trasformata di Fourier* definito sulle misure è l'*operatore aggiunto* dell'operatore *trasformata di Fourier* definito sulle funzioni.

Teorema 3.1.3 (Derivabilità della funzione caratteristica). *Supponiamo che la v.a. X abbia momento di ordine k (k intero positivo): allora la sua funzione caratteristica è derivabile k volte e vale la formula*

$$\varphi_X^{(k)}(t) = i^k \mathbf{E}[X^k e^{itX}]$$

Nel caso vettoriale la formula diventa

$$\frac{\partial^{a_1+\dots+a_n} \varphi_X(u_1, \dots, u_n)}{\partial u_1^{a_1} \dots \partial u_n^{a_n}} = i^{(a_1+\dots+a_n)} \mathbf{E}[X_1^{a_1} \dots X_n^{a_n} e^{i\langle \mathbf{u}, \mathbf{X} \rangle}]$$

(a condizione che si abbia $\mathbf{E}[|X_1|^{a_1} \dots |X_n|^{a_n}] < +\infty$).

Dimostrazione. Per la dimostrazione ci limitiamo per semplicità al caso scalare con $k = 1$. Sia h_n una successione infinitesima:

$$\frac{\varphi_X(t + h_n) - \varphi_X(t)}{h_n} = \int_{\mathbb{R}} e^{itx} \left(\frac{e^{ih_n x} - 1}{h_n} \right) d\mathbf{P}_X(x)$$

È facile verificare che si ha $\left| \frac{e^{ih_n x} - 1}{h_n} \right| \leq 2|x|$ e che $\lim_{n \rightarrow \infty} \frac{e^{ih_n x} - 1}{h_n} = ix$: poichè la funzione $|x|$ è integrabile rispetto a \mathbf{P}_X (la v.a. ha momento primo) si può passare al limite sotto il segno d'integrale. \square

Abbastanza curiosamente, il risultato appena provato può essere in un certo senso invertito (a dimensione 1):

Teorema 3.1.4. *Supponiamo che $\varphi_X(\cdot)$ sia derivabile k volte in 0, con k pari: allora la v.a. X ha momento di ordine k .*

Dimostrazione. Dimostriamo il risultato solo per $k = 2$ (nel caso generale la dimostrazione si fa per induzione ed è un poco più delicata). Partiamo dall'osservazione che

$$\int_{\mathbb{R}} e^{itx} d\mathbf{P}_X(x) = \int_{\mathbb{R}} \cos(tx) d\mathbf{P}_X(x) + i \int_{\mathbb{R}} \sin(tx) d\mathbf{P}_X(x) = \alpha(t) + i\beta(t)$$

La funzione $\alpha(t)$ è pari (e $\beta(t)$ dispari): allora $\alpha'(0) = 0$ e $\varphi_X''(0) = \alpha''(0)$. Inoltre si ha $\alpha''(0) = \lim_{t \rightarrow 0} 2 \frac{\alpha(t) - 1}{t^2}$.

Sia $(t_n)_{n \geq 1}$ una successione infinitesima: le funzioni $2 \frac{\cos(t_n x) - 1}{t_n^2}$ sono negative e convergono a $-x^2$. Si può così applicare il *lemma di Fatou* e si ottiene

$$\begin{aligned} - \int_{\mathbb{R}} x^2 d\mathbf{P}_X(x) &\geq \limsup_{n \rightarrow \infty} \int_{\mathbb{R}} 2 \frac{\cos(t_n x) - 1}{t_n^2} d\mathbf{P}_X(x) = \\ &= \limsup_{n \rightarrow \infty} 2 \frac{\alpha(t_n) - 1}{t_n^2} = \alpha''(0) \end{aligned}$$

□

Se la variabile possiede momento fino all'ordine k , si può applicare alla funzione caratteristica lo sviluppo di Taylor fino a tale ordine in un intorno di 0; se la variabile possiede tutti i momenti, la sua funzione caratteristica ammette derivate di ogni ordine *ma non è detto che sia sviluppabile in serie di potenze*.

Tuttavia è facile fornire condizioni sufficienti affinché questo sviluppo sia possibile. Consideriamo una v.a. X , supponiamo che abbia momenti di ogni ordine e sia $m_n = \mathbf{E}[X^n]$. Per ogni n , $\varphi_X(\cdot)$ si può sviluppare con polinomio di Taylor di ordine n mediante la formula

$$\varphi_X(t) = \sum_{k=0}^n \frac{i^k m_k}{k!} t^k + R_{n+1}(t)$$

dove il *resto* $R_{n+1}(t)$ soddisfa la maggiorazione

$$|R_{n+1}(t)| \leq \frac{2|t|^{n+1}}{(n+1)!} \mathbf{E}[|X|^{n+1}]$$

La maggiorazione appena scritta si giustifica nel modo seguente: si parte dalla formula del resto per lo sviluppo di Taylor (intorno al punto 0) di una funzione f derivabile $(n+1)$ volte, e precisamente

$$f(t) = \sum_{k=0}^n \frac{f^{(k)}(0)}{k!} t^k + \frac{f^{(n+1)}(\tau)}{(n+1)!} t^{n+1}$$

dove τ è un punto *intermedio* tra 0 e t . Tuttavia non si può applicare direttamente questa formula alla funzione $\varphi_X(t)$ che è a valori complessi, bensì bisogna applicarla *separatamente* alla sua parte reale $\mathbf{E}[\cos(tX)]$ ed alla sua parte immaginaria $\mathbf{E}[\sin(tX)]$: i dettagli sono un po' noiosi ma per nulla difficili, e lasciati per esercizio.

A questo punto è facile trovare dei criteri che garantiscano che la f.c. φ sia sviluppabile in *serie di Taylor*: ad esempio si ha il risultato seguente, la cui facile dimostrazione è lasciata per esercizio

Corollario 3.1.5. *Supponiamo che esistano $M > 0$ e $r > 0$ tali che per ogni $n \in \mathbb{N}$ valga la maggiorazione*

$$\mathbf{E}[|X|^n] \leq \frac{M n!}{r^n}$$

Allora per $|t| < r$ vale lo sviluppo in serie di Taylor

$$\varphi_X(t) = \sum_{n=0}^{+\infty} \frac{i^n m_n}{n!} t^n$$

Si può andare un poco più in là: supponendo che la v.a. X abbia tutti i momenti, a partire da qualsiasi punto t_0 si può fare lo sviluppo di Taylor

$$\varphi_X(t) = \sum_{k=0}^n \frac{i^k \varphi_X^{(k)}(t_0)}{k!} (t - t_0)^k + R_{n+1}(t)$$

dove di nuovo $R_{n+1}(t)$ verifica la maggiorazione

$$|R_{n+1}(t)| \leq \frac{2|t - t_0|^{n+1}}{(n+1)!} \mathbf{E}[|X|^{n+1}]$$

Pertanto se è soddisfatta la condizione del Corollario 3.1.5 $\varphi_X(t)$ può essere sviluppata in *serie di Taylor* nell'intervallo $]t_0 - r, t_0 + r[$.

Si ottiene così il seguente risultato:

Corollario 3.1.6. *Siano X e Y due v.a. reali dotate di ogni momento, supponiamo che si abbia $\mathbf{E}[X^n] = \mathbf{E}[Y^n]$ per ogni n e che sia soddisfatta la maggiorazione del Corollario 3.1.5: allora $\varphi_X(t) = \varphi_Y(t)$ per ogni $t \in \mathbb{R}$.*

Dimostrazione. La dimostrazione è molto semplice: infatti grazie allo sviluppo in serie di Taylor φ_X e φ_Y coincidono sull'intervallo $] -r, r[$; a questo punto si prende un punto t_0 interno a questo intervallo e si ripete lo sviluppo ottenendo l'eguaglianza di φ_X e φ_Y su $]t_0 - r, t_0 + r[$... In questo modo si arriva a coprire tutto \mathbb{R} ottenendo il risultato voluto. \square

Vediamo adesso alcuni esempi di funzioni caratteristiche.

Esempio 3.1.7 (Variabili discrete).

Il calcolo della funzione caratteristica di una variabile discreta è di solito un esercizio elementare. Riportiamo alcuni risultati di facile verifica.

Se X è **Binomiale** di parametri n e p , la sua funzione caratteristica è $(pe^{it} + (1-p))^n$.

Se X è **Geometrica** di parametro p , la sua funzione caratteristica è $\frac{p}{p-1+e^{-it}}$.

Se X è di **Poisson** di parametro λ , la sua funzione caratteristica è $e^{\lambda(e^{it}-1)}$.

Esempio 3.1.8 (Variabili Gaussiane). Calcoliamo la funzione caratteristica di una variabile X con densità $N(0,1)$ (indicando per semplicità φ anziché φ_X).

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} \int \cos(tx) e^{-x^2/2} dx + \frac{i}{\sqrt{2\pi}} \int \sin(tx) e^{-x^2/2} dx$$

e notiamo subito che il secondo termine (la parte immaginaria) si annulla poichè è l'integrale su tutta la retta di una funzione dispari. Proviamo ora che la funzione verifica l'equazione differenziale $\varphi'(t) = -t\varphi(t)$:

$$\begin{aligned} \varphi'(t) &= \frac{1}{\sqrt{2\pi}} \int \sin(tx) (-xe^{-x^2/2}) dx = \\ &= \frac{e^{-x^2/2} \sin(tx)}{\sqrt{2\pi}} \Big|_{-\infty}^{+\infty} - \int t \cos(tx) \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = -t\varphi(t) \end{aligned}$$

Poichè $\varphi(0) = 1$, si ha $\varphi(t) = e^{-t^2/2}$. Ne segue che la funzione caratteristica di una variabile gaussiana $N(m, \sigma^2)$ è $\varphi(t) = \exp\left(imt - \frac{\sigma^2 t^2}{2}\right)$.

Esempio 3.1.9 (Variabili Gamma).

Se α è un numero complesso con parte reale strettamente positiva e $r > 0$ è reale, un esercizio sull'integrazione circolare di funzioni analitiche di variabile complessa mostra che vale l'eguaglianza

$$\int_0^{+\infty} \alpha^r x^{r-1} e^{-\alpha x} dx = \Gamma(r)$$

A questo riguardo osserviamo che, se $b \in \mathbb{R}$, la funzione z^b non può essere definita con continuità sull'intero piano complesso, ma può esserlo sul piano complesso *privato della semiretta dei numeri reali negativi*; inoltre sul semipiano dei numeri z con *parte reale strettamente positiva* la funzione $z \rightarrow z^b$ è analitica.

Dal conto sopra riportato segue facilmente che la funzione caratteristica di una variabile con densità $\Gamma(r, \lambda)$ è $\left(\frac{\lambda}{\lambda - it}\right)^r$.

3.2 Inversione delle funzioni caratteristiche.

In realtà raramente si usa una vera *inversione* delle funzioni caratteristiche: il risultato fondamentale di questo paragrafo è che due variabili che hanno la stessa funzione caratteristica sono *equidistribuite*.

Supponiamo assegnata, in questo paragrafo, una v.a.r. X con funzione di ripartizione F e funzione caratteristica φ .

Ricordiamo che con la notazione $\int g(x) dF(x)$ si indica (se esiste) l'integrale della funzione g rispetto alla probabilità associata alla funzione di ripartizione F , quindi $\int g(x) dF(x) = \int g(x) d\mathbf{P}_X(x)$.

In particolare, ad esempio, $\varphi(t) = \int e^{itx} dF(x)$.

Proposizione 3.2.1 (Formula di inversione). *Se a, b sono punti di continuità per la funzione di ripartizione F ed $a < b$, si ha*

$$F(b) - F(a) = \lim_{T \rightarrow +\infty} \frac{1}{2\pi} \int_{-T}^T \varphi(t) \left(\frac{e^{-ita} - e^{-itb}}{it} \right) dt$$

Rimandiamo per il momento la dimostrazione. Una immediata conseguenza della Proposizione 3.2.1 è il Teorema seguente, che è il risultato più importante di questo capitolo.

Teorema 3.2.2. *Siano X, Y due v.a.r. con funzione di ripartizione rispettivamente F e G e supponiamo che si abbia $\varphi_X \equiv \varphi_Y$: allora $F \equiv G$.*

Dimostrazione. Al di fuori di un insieme al più numerabile (i punti di discontinuità di F e di G) si ha $F(b) - F(a) = G(b) - G(a)$: tenendo conto del fatto che $F(-\infty) = G(-\infty) = 0$ e che F e G sono continue a destra, si conclude immediatamente che sono eguali. \square

Un'altra conseguenza della Proposizione 3.2.1 è il risultato seguente.

Corollario 3.2.3 (Formula per la densità). *Supponiamo che la funzione caratteristica φ della v.a. X sia tale che $\int_{-\infty}^{+\infty} |\varphi(t)| dt < +\infty$: allora X ha densità f data da*

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \varphi(t) dt$$

Dimostrazione. Tendendo conto del fatto che $\varphi(t)$ è integrabile su \mathbb{R} , fuori di un insieme numerabile si ha

$$\begin{aligned} F(b) - F(a) &= \lim_{T \rightarrow +\infty} \frac{1}{2\pi} \int_{-T}^{+T} \varphi(t) \left[\int_a^b e^{-itx} dx \right] dt = \\ &= \lim_{T \rightarrow +\infty} \frac{1}{2\pi} \int_a^b \left[\int_{-T}^{+T} \varphi(t) e^{-itx} dt \right] dx = \int_a^b \left(\frac{1}{2\pi} \int_{-\infty}^{+\infty} \varphi(t) e^{-itx} dt \right) dx \end{aligned}$$

È immediato concludere che F ha densità, data dall'espressione sopra scritta. \square

È opportuno ribadire che non per tutte le v.a. che hanno densità, questa può essere ottenuta attraverso la formula 3.2.3: in particolare si può osservare che la densità fornita dalla formula del Corollario 3.2.3 è una funzione continua. Se dunque una v.a. X ha una densità che non è continua, sicuramente il Corollario 3.2.3 non è applicabile.

Prepariamo la dimostrazione della Proposizione 3.2.1 con questo calcolo preliminare:

Lemma 3.2.4. *Vale il seguente risultato:*

$$\lim_{T \rightarrow \infty} \int_0^T \frac{\sin ax}{x} dx = \begin{cases} \frac{\pi}{2} & a > 0 \\ 0 & a = 0 \\ -\frac{\pi}{2} & a < 0 \end{cases}$$

inoltre la funzione $T \rightarrow \int_0^T \frac{\sin(ax)}{x} dx$ è limitata su \mathbb{R}^+ .

Dimostrazione. Ci si può ricondurre al caso $a = 1$. Osserviamo che la funzione $\frac{\sin x}{x}$ non è integrabile su \mathbb{R}^+ , tuttavia esiste il limite sopra scritto (integrale improprio). Cominceremo a provare che si ha $\lim_{n \rightarrow \infty} \int_0^{2\pi n} \frac{\sin x}{x} dx = \frac{\pi}{2}$; la conclusione è una conseguenza del fatto che, se $2\pi n < T < 2\pi(n+1)$, si ha $\left| \int_{2\pi n}^T \frac{\sin x}{x} dx \right| \leq \frac{1}{n}$.

$$\begin{aligned} \lim_{n \rightarrow \infty} \int_0^{2\pi n} \frac{\sin x}{x} dx &= \lim_{n \rightarrow \infty} \int_0^{2\pi n} \sin(x) \left[\int_0^{+\infty} e^{-xu} du \right] dx = \\ &= \lim_{n \rightarrow \infty} \int_0^{+\infty} \left[\int_0^{2\pi n} e^{-xu} \sin(x) dx \right] du \end{aligned}$$

Un conto facile (una integrazione per parti ripetuta due volte) mostra che

$$\int_0^{2\pi n} e^{-xu} \sin(x) dx = \frac{1 - e^{-2\pi nu}}{1 + u^2}.$$

Si ottiene di conseguenza

$$\lim_{n \rightarrow \infty} \int_0^{2\pi n} \frac{\sin x}{x} dx = \int_0^{+\infty} \frac{1}{1 + u^2} du = \frac{\pi}{2}$$

□

Passiamo ora alla dimostrazione della Proposizione 3.2.1.

Dimostrazione. Prendiamo $a < b$:

$$\begin{aligned} \frac{1}{2\pi} \int_{-T}^{+T} \varphi(t) \left(\frac{e^{-ita} - e^{-itb}}{it} \right) dt &= \frac{1}{2\pi} \int_{-T}^{+T} \left[\int_{-\infty}^{+\infty} e^{itx} dF(x) \right] \frac{e^{-ita} - e^{-itb}}{it} dt = \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left[\int_{-T}^{+T} \frac{e^{it(x-a)} - e^{it(x-b)}}{it} dt \right] dF(x) \end{aligned}$$

La funzione $\frac{e^{it(x-a)} - e^{it(x-b)}}{it}$ è somma della funzione $\frac{\cos(t(x-a)) - \cos(t(x-b))}{it}$ (che è dispari, e dunque il suo integrale su $[-T, T]$ è 0), e della funzione $\frac{\sin(t(x-a)) - \sin(t(x-b))}{t}$ (che è pari, e dunque il suo integrale su $[-T, T]$ è eguale a due volte l'integrale su $[0, T]$).

Inoltre, utilizzando il Lemma 3.2.4, un facile calcolo prova che

$$\lim_{T \rightarrow \infty} \int_0^T \frac{\sin(t(x-a)) - \sin(t(x-b))}{t} dt = \begin{cases} 0 & x < a \\ \pi & a < x < b \\ 0 & x > b \end{cases}$$

È un pò più noioso calcolare il valore del limite sopra scritto proprio nei punti a e b , ma non ne abbiamo bisogno: se a, b sono punti di continuità per F (e dunque trascurabili per la misura generata da F), si ottiene

$$\lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^{+T} \varphi(t) \left(\frac{e^{-ita} - e^{-itb}}{it} \right) dt = \frac{1}{\pi} \int_{-\infty}^{+\infty} \pi I_{[a,b]}(x) dF(x) = F(b) - F(a)$$

□

La Proposizione 3.2.1 ed il conseguente Teorema 3.2.2 sono validi anche per una variabile vettoriale $\mathbf{X} = (X_1, \dots, X_n)$ (con una dimostrazione sostanzialmente analoga ma più tediosa): riportiamo l'enunciato dell'estensione a più dimensioni della Proposizione 3.2.2. Se $I = \prod_{k=1}^n]a_k, b_k]$ ed i punti $\mathbf{a} = (a_1, \dots, a_n)$ e $\mathbf{b} = (b_1, \dots, b_n)$ sono trascurabili per la legge di \mathbf{X} , si ha

$$\mathbf{P}\{\mathbf{X} \in I\} = \lim_{T \rightarrow \infty} \frac{1}{(2\pi)^n} \int_{-T}^T \dots \int_{-T}^T \prod_{k=1}^n \frac{e^{-it_k a_k} - e^{-it_k b_k}}{it_k} \varphi(t_1, \dots, t_n) dt_1 \dots dt_n$$

Poichè gli insiemi I della forma sopra descritta (con \mathbf{a} e \mathbf{b} trascurabili per la legge di \mathbf{X}), formano una famiglia di insiemi stabile per intersezione che genera la σ -algebra di Borel su \mathbb{R}^n , segue che due variabili vettoriali che hanno la stessa funzione caratteristica sono equidistribuite.

Vediamo ora alcune conseguenze del Teorema 3.2.2

Corollario 3.2.5. *Supponiamo che le v.a. reali X e Y abbiano momento di ogni ordine, che per ogni n si abbia $\mathbf{E}[X^n] = \mathbf{E}[Y^n]$ e che sia verificata la maggiorazione del Corollario 3.1.5*

$$\mathbf{E}[|X|^n] \leq \frac{M n!}{r^n}$$

Allora X e Y sono equidistribuite.

Infatti grazie al Corollario 3.1.6 X e Y hanno la stessa funzione caratteristica.

Osservazione 3.2.6. Stabilire se due variabili X e Y che hanno tutti i momenti eguali siano equidistribuite è il cosiddetto **problema dei momenti**: abbiamo visto che se è soddisfatta la condizione del Corollario 3.2.5 la risposta è affermativa. È possibile tuttavia che due variabili X e Y abbiano momenti eguali (sia cioè verificata l'eguaglianza $\mathbf{E}[X^n] = \mathbf{E}[Y^n]$ per ogni intero positivo n) ma che *non siano equidistribuite*; due controesempi sono esposti nell'Appendice.

Corollario 3.2.7. *Due variabili aleatorie reali X e Y sono indipendenti se e solo se vale l'eguaglianza $\varphi_{(X,Y)}(u, v) = \varphi_X(u) \cdot \varphi_Y(v)$ (per ogni $(u, v) \in \mathbb{R}^2$).*

Dimostrazione. La dimostrazione è immediata: se X e Y sono indipendenti, un calcolo evidente prova che vale l'eguaglianza sopra scritta. Se vale l'eguaglianza sopra scritta, vuole dire che la coppia (X, Y) ha la stessa legge della coppia di due variabili indipendenti distribuite rispettivamente come X e come Y e quindi X e Y sono indipendenti. \square

3.3 La funzione generatrice dei momenti.

La **Funzione generatrice dei momenti** di una v.a. reale X è la funzione $\psi_X : \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\} =]-\infty, +\infty]$ definita da

$$\psi_X(t) = \mathbf{E}[e^{tX}]$$

In realtà si considera la *restrizione* di ψ_X al sottinsieme di \mathbb{R} sul quale ψ_X assume valori finiti e chiamiamo questo *insieme di definizione* della funzione generatrice dei momenti. Poichè la funzione $t \rightarrow e^{tx}$ è convessa (qualunque sia $x \in \mathbb{R}$), segue facilmente che anche la funzione ψ_X è convessa: di conseguenza l'insieme di definizione è un sottinsieme convesso di \mathbb{R} (cioè un intervallo in senso lato) e la funzione ψ_X è continua all'interno di questo intervallo.

A volte l'insieme di definizione è ridotto al solo punto 0, come accade ad esempio per la variabile che ha densità di Cauchy $f(x) = \frac{1}{\pi(1+x^2)}$.

Poiché vale lo sviluppo in serie $e^{tX} = \sum_{n=0}^{+\infty} \frac{t^n X^n}{n!}$, se la variabile X possiede momento di ogni ordine siamo tentati di scrivere

$$\psi_X(t) = \sum_{n=0}^{+\infty} \frac{t^n \mathbf{E}[X^n]}{n!} = \sum_{n=0}^{+\infty} \frac{t^n m_n}{n!}$$

Lo sviluppo in serie appena scritto (che non è sempre valido) deve essere giustificato: ad esempio una condizione sufficiente affinché si possa fare quello sviluppo (cioè si possa *integrare per serie*) è che si abbia

$$\sum_{n=0}^{+\infty} \frac{t^n \mathbf{E}[|X|^n]}{n!} < +\infty$$

La diseguaglianza appena scritta, se è verificata per un numero $t \neq 0$, è soddisfatta su tutto l'intervallo $] -|t|, |t| [$, la funzione ψ ammette derivate di ogni ordine in quell'intervallo e vale l'eguaglianza

$$\mathbf{E}[X^n] = \left. \frac{d^n \psi_X(t)}{dt^n} \right|_{t=0}$$

e da questa deriva appunto il nome di *funzione generatrice dei momenti*.

Ad esempio, se una v.a. X possiede tutti i momenti ed è soddisfatta la maggiorazione del Corollario 3.1.5 $\mathbf{E}[|X|^n] \leq \frac{M n!}{r^n}$, allora la f.g.m. $\psi_X(t)$ è sicuramente definita in $] -r, r[$.

Questa condizione può in un certo senso essere invertita come afferma il risultato seguente.

Teorema 3.3.1. *Sia X una v.a. e supponiamo che la sua f.g.m. $\psi_X(t)$ sia definita in un intorno di 0 : allora X possiede i momenti di ogni ordine.*

Dimostrazione. Partiamo da una ovvia diseguaglianza: qualunque sia $y \in \mathbb{R}$ e $n \in \mathbb{N}$, si ha $\frac{|y|^n}{n!} \leq (e^y + e^{-y})$

Di conseguenza, se $\psi_X(t)$ è definita per $t = \varepsilon$ e $t = -\varepsilon$ (con $\varepsilon > 0$), si ha

$$\mathbf{E}[|X|^n] \leq n! \frac{\psi_X(\varepsilon) + \psi_X(-\varepsilon)}{\varepsilon^n}$$

ed è evidente che esistono tutti i momenti di X . □

Osservazione 3.3.2. Nell'enunciato precedente è *essenziale* che l'intervallo considerato contenga al suo interno il punto 0 : ad esempio se si prende una v.a. X avente densità

$$f(x) = \begin{cases} \frac{2}{\pi(1+x^2)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

è facile constatare che ψ_X è definita su tutta la semiretta $] -\infty, 0]$, ma per ogni $n \in \mathbb{N}$ si ha $\mathbf{E}[X^n] = +\infty$.

Il risultato più importante che riguarda le f.g.m. è il seguente:

Teorema 3.3.3. *Siano X e Y due variabili e supponiamo che le loro f.g.m. siano definite e siano eguali in un intorno di 0 : allora X e Y sono equidistribuite.*

Dimostrazione. La dimostrazione è immediata: se vale l'eguaglianza $\psi_X(t) = \psi_Y(t)$ in un intervallo che contiene $[-\varepsilon, \varepsilon]$, X e Y hanno tutti i momenti eguali ed è soddisfatta la maggiorazione del Corollario 3.1.5 che permette di concludere grazie al Corollario 3.2.5. □

Osservazione 3.3.4. A differenza di quanto è stato detto nell'osservazione 3.3.2, si può dimostrare che se si ha l'eguaglianza $\psi_X(t) = \psi_Y(t)$ su tutto un intervallo (non necessariamente contenente al suo interno il punto 0), allora X e Y sono equidistribuite. La dimostrazione di questa affermazione è però molto più impegnativa ed è conseguenza di un risultato più generale sulla *trasformata di Laplace* delle misure.

3.4 Vettori aleatori gaussiani.

L'argomento dei vettori aleatori gaussiani viene inserito in questo capitolo perchè lo strumento delle *funzioni caratteristiche* ne rende lo studio semplice.

Definizione 3.4.1. Un vettore aleatorio $\mathbf{X} = (X_1, \dots, X_d)$ (a valori in \mathbb{R}^d) è detto **vettore gaussiano** se, per ogni $\mathbf{u} \in \mathbb{R}^d$, $\langle \mathbf{u}, \mathbf{X} \rangle = \sum_{j=1}^d u_j X_j$ è una variabile aleatoria gaussiana.

È conveniente considerare anche le *costanti* delle v.a. gaussiane: in particolare la legge della costante a è indicata $N(a, 0)$.

Notiamo che ogni componente $X_i = \langle \mathbf{e}_i, \mathbf{X} \rangle$ (dove \mathbf{e}_i è il vettore i -simo della base canonica) è gaussiana (e quindi ha momento di ogni ordine): indichiamo allora con $\mathbf{m} = (\mathbf{E}[X_1], \dots, \mathbf{E}[X_n])$ il vettore dei valori attesi e con \mathbf{Q} la matrice delle covarianze ($\mathbf{Q}_{ij} = \text{Cov}(X_i, X_j)$). \mathbf{Q} è una matrice $d \times d$ simmetrica, semidefinita positiva.

Proposizione 3.4.2. Il vettore \mathbf{m} e la matrice \mathbf{Q} identificano la legge di X , che è indicata $N_d(\mathbf{m}, \mathbf{Q})$.

Dimostrazione. Grazie al Teorema 3.2.1, è sufficiente verificare che \mathbf{m} e \mathbf{Q} determinano la funzione caratteristica del vettore aleatorio \mathbf{X} .

Osservando che, preso $\mathbf{u} \in \mathbb{R}^d$, $\langle \mathbf{u}, \mathbf{X} \rangle$ è una variabile gaussiana reale con media $\langle \mathbf{u}, \mathbf{m} \rangle$ e varianza $\langle \mathbf{Q}\mathbf{u}, \mathbf{u} \rangle$, si ottiene

$$\varphi_{\mathbf{X}}(\mathbf{u}) = \mathbf{E}[e^{i\langle \mathbf{u}, \mathbf{X} \rangle}] = e^{i\langle \mathbf{m}, \mathbf{u} \rangle - 1/2\langle \mathbf{Q}\mathbf{u}, \mathbf{u} \rangle}$$

□

Una conseguenza importante è il Corollario che segue: sappiamo che variabili indipendenti sono incorrelate, ma che in generale *non è vero il viceversa*. Tuttavia per variabili gaussiane vale questo risultato:

Corollario 3.4.3. Sia \mathbf{X} un vettore gaussiano: due componenti X_i e X_j sono indipendenti se e solo se sono incorrelate.

Dimostrazione. È sufficiente ridursi al caso di una variabile a dimensione 2 (X, Y) e si può supporre che X e Y siano *centrate*: se sono incorrelate, la matrice \mathbf{Q} delle covarianze è diagonale ed in tal caso (sfruttando la Proposizione 3.4.2) si ha $\varphi_{(X,Y)}(u, v) = \varphi_X(u) \varphi_Y(v)$ e questa identità equivale all'indipendenza (vedi Corollario 3.2.7). □

Osservazione 3.4.4. Sia \mathbf{X} di legge $N_d(\mathbf{m}, \mathbf{Q})$, A una matrice $k \times d$ e $\mathbf{n} \in \mathbb{R}^k$: la variabile vettoriale $\mathbf{Y} = A\mathbf{X} + \mathbf{n}$ è gaussiana, e la sua legge è $N_k(A\mathbf{m} + \mathbf{n}, A\mathbf{Q}A^*)$.

Infatti \mathbf{Y} è gaussiana perchè ogni combinazione lineare delle variabili Y_j si riduce ad una combinazione lineare delle variabili X_i (più una costante); il calcolo dei valori attesi e delle covarianze degli elementi Y_j è immediato.

Esempio 3.4.5. Affinchè \mathbf{X} sia un vettore gaussiano, non è sufficiente che le componenti siano gaussiane: consideriamo infatti questo esempio. Sia X con densità $N(0, 1)$, Z indipendente da X che prende i valori 1 e -1 ciascuno con probabilità $1/2$, e sia $Y = ZX$. La variabile Y è gaussiana (la verifica è semplice), ma la coppia (X, Z) non è gaussiana (si può constatare, ad esempio, che $X + Y$ non può essere gaussiana).

Inoltre si può constatare facilmente che X e Y sono *incorrelate* ma che *non sono indipendenti*.

Esempio 3.4.6. Siano X_1, \dots, X_d indipendenti con densità $N(0, 1)$: il vettore aleatorio $\mathbf{X} = (X_1, \dots, X_d)$ ha legge $N_d(\mathbf{0}, I_d)$, dove $\mathbf{0}$ è lo 0 in \mathbb{R}^d e I_d è la matrice identità su \mathbb{R}^d .

Teorema 3.4.7 (Esistenza di variabili gaussiane vettoriali). *Assegnati un vettore $\mathbf{m} \in \mathbb{R}^d$ ed una matrice $d \times d$ simmetrica, semidefinita positiva \mathbf{Q} , esiste un vettore aleatorio di legge $N_d(\mathbf{m}, \mathbf{Q})$.*

Dimostrazione. La dimostrazione si riduce a un esercizio di algebra lineare: si tratta infatti di costruire una matrice quadrata d -dimensionale A tale che si abbia $\mathbf{Q} = AA^*$: presa poi \mathbf{X} di legge $N_d(\mathbf{0}, I_d)$ (vedi Esempio 3.4.6), il vettore $\mathbf{Y} = A\mathbf{X} + \mathbf{m}$ ha (grazie all'Osservazione 3.4.4) la legge cercata.

La matrice A non è univocamente determinata. Una possibile scelta (probabilmente la più naturale) si ottiene in questo modo: poichè \mathbf{Q} è simmetrica e semidefinita positiva, gli autovettori sono positivi (indichiamoli $\lambda_1, \dots, \lambda_d$, contati con la loro molteplicità) ed esiste una base ortonormale formata di autovettori corrispondenti. Sia Λ la matrice diagonale che ha sulla diagonale i numeri $\lambda_1, \dots, \lambda_d$ e sia $\sqrt{\Lambda}$ la matrice diagonale che ha sulla diagonale i numeri $\sqrt{\lambda_1}, \dots, \sqrt{\lambda_d}$; sia infine C la matrice ortogonale del passaggio dalla base canonica alla base formata dagli autovettori.

È facile constatare che vale l'identità $\mathbf{Q} = C^*\Lambda C$ e definendo $A = C^*\sqrt{\Lambda}C$, è immediato constatare che $A = A^*$ e che $AA^* = \mathbf{Q}$. \square

Corollario 3.4.8 (Densità di variabili gaussiane vettoriali). *Sia \mathbf{X} di legge $N_d(\mathbf{m}, \mathbf{Q})$ e supponiamo che \mathbf{Q} sia invertibile: allora \mathbf{X} ha densità e la forma della densità è*

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} \sqrt{\det \mathbf{Q}}} \exp \left(-\frac{1}{2} \langle \mathbf{Q}^{-1}(\mathbf{x} - \mathbf{m}), (\mathbf{x} - \mathbf{m}) \rangle \right)$$

Dimostrazione. Partiamo da \mathbf{Y} di legge $N_d(\mathbf{0}, I_d)$, la cui densità, come si verifica facilmente, è data da $(2\pi)^{-n/2} \exp(-1/2 \|\mathbf{x}\|^2)$ e, sfruttando la rappresentazione $\mathbf{X} = A\mathbf{Y} + \mathbf{m}$ fornita dal Teorema 3.4.7, applichiamo la formula (vista nel corso “Elementi di Probabilità e Statistica”) che esprime come si trasforma la densità di una v.a. alla quale si applica un diffeomorfismo (applicazione biunivoca differenziabile con inversa differenziabile).

Tenendo conto del fatto che $\mathbf{Y} = A^{-1}(\mathbf{X} - \mathbf{m})$, che $\det(A^{-1}) = \frac{1}{\sqrt{\det \mathbf{Q}}}$ e che

$$\|A^{-1}(\mathbf{x} - \mathbf{m})\|^2 = \langle A^{-1}(\mathbf{x} - \mathbf{m}), A^{-1}(\mathbf{x} - \mathbf{m}) \rangle = \langle \mathbf{Q}^{-1}(\mathbf{x} - \mathbf{m}), (\mathbf{x} - \mathbf{m}) \rangle$$

si concludono agevolmente i conti. \square

3.5 Appendice

Vengono riportati di seguito due controesempi al **problema dei momenti**: in entrambi i casi i conti non compaiono con tutti i dettagli.

Nel primo esempio consideriamo una variabile X gaussiana standard, e sia $Y = e^X$ (variabile *lognormale*): non è necessario calcolare la densità di questa variabile ma un facile calcolo prova che, per ogni n , $\mathbf{E}[Y^n] = \exp(n^2/2)$.

Consideriamo poi una variabile discreta Z che prende i valori e^k (al variare di k intero *relativo*) e tale che $\mathbf{P}\{Z = e^k\} = c e^{-k^2/2}$. Non è necessario calcolare la costante di normalizzazione c , ma di nuovo un conto elementare prova che $\mathbf{E}[Z^n] = \exp(n^2/2)$.

Il secondo esempio è più elaborato e ad esso premettiamo un conto: se X ha densità $N(0, 1)$ e z è complesso, si ha $\mathbf{E}[e^{zX}] = e^{z^2/2}$ (questo conto è già stato fatto nel corso del capitolo per z reale oppure immaginario puro). Prendiamo ora n e p interi positivi, e sia $\rho = p\pi$: si ottiene che

$$\mathbf{E}[e^{(n+i\rho)X}] = \mathbf{E}[Y^n (\cos(p\pi X) + i \sin(p\pi X))]$$

è un numero reale e di conseguenza $\mathbf{E}[Y^n (\sin(p\pi X))] = 0$.

Fissiamo ora un intero $p \neq 0$ e consideriamo uno spazio $(\Omega, \mathcal{F}, \mathbf{P})$ sul quale è definita la variabile X , sia poi \mathbf{Q} la probabilità che ha rispetto a \mathbf{P} la densità $(1 + \frac{1}{2} \sin p\pi X)$, consideriamo infine due v.a. aventi rispettivamente le leggi $Y(\mathbf{P})$ e $Y(\mathbf{Q})$: queste variabili hanno tutti i momenti uguali ma non sono equidistribuite (quest’ultima affermazione è facile ma non immediata, sforzarsi di dimostrarla).

Capitolo 4

Convergenza di variabili aleatorie.

4.1 Convergenza in probabilità e quasi certa.

Tutti i risultati di questo paragrafo sono enunciati per variabili aleatorie reali, ma sono validi per v.a. a valori in \mathbb{R}^d ; inoltre più che a variabili aleatorie bisogna pensare a *classi di equivalenza* di variabili aleatorie (identificando due variabili che coincidono q.c.).

Supponiamo fissato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$ e ricordiamo che (per $1 \leq p < +\infty$) si indica con $L^p(\Omega, \mathcal{F}, \mathbf{P})$ lo spazio delle (classi di equivalenza delle) v.a. X tali che $\mathbf{E}[|X|^p] < +\infty$: grazie alla *diseguaglianza di Minkowski* il numero $\|X\|_p = (\mathbf{E}[|X|^p])^{1/p}$ è una *norma* su L^p , e di conseguenza $d(X, Y) = \|X - Y\|_p$ è una *distanza* su L^p . (La diseguaglianza di Minkowski sarà dimostrata in appendice).

Viceversa lo spazio $L^\infty(\Omega, \mathcal{F}, \mathbf{P})$ è lo spazio delle v.a. X per le quali esiste una costante M tale che si abbia $|X(\omega)| \leq M$ q.c., e la norma $\|X\|_\infty$ è la *minima* delle costanti maggioranti (anche l'esistenza di questa costante minima sarà provata in appendice).

Sia dunque assegnata una successione di v.a.r. $(X_n)_{n \geq 1}$ (naturalmente i limiti si intendono per $n \rightarrow \infty$).

Definizione 4.1.1. Si dice che la successione $(X_n)_{n \geq 1}$ converge a X

- **in probabilità** (e si scrive $X_n \xrightarrow{\mathbf{P}} X$) se, $\forall \varepsilon > 0$,
 $\lim_n \mathbf{P}\{|X_n - X| > \varepsilon\} = 0$;
- **quasi certamente** (e si scrive $X_n \xrightarrow{\text{q.c.}} X$) se, per quasi ogni ω , la successione di numeri $X_n(\omega)$ converge a $X(\omega)$;

- in L^p se ogni X_n (ed X) appartiene a L^p e $\lim_n \|X_n - X\|_p = 0$.

Proposizione 4.1.2. *Tra i tipi di convergenza sopra enunciati, valgono le seguenti relazioni:*

- se X_n converge a X in L^p , allora $X_n \xrightarrow{\mathbf{P}} X$;
- se $X_n \xrightarrow{\text{q.c.}} X$, allora $X_n \xrightarrow{\mathbf{P}} X$;
- se $X_n \xrightarrow{\mathbf{P}} X$, esiste una sottosuccessione X_{n_k} tale che $X_{n_k} \xrightarrow{\text{q.c.}} X$.

Dimostrazione. Proviamo a) : se $1 \leq p < +\infty$, l'affermazione è una conseguenza immediata della disuguaglianza, valida per $\varepsilon > 0$,

$$\varepsilon^p \mathbf{P}\{|X_n - X| > \varepsilon\} \leq \mathbf{E}[|X_n - X|^p]$$

Viceversa, nel caso $p = +\infty$, segue dal fatto che la convergenza in L^∞ è evidentemente più forte della convergenza q.c.

Per quanto riguarda le altre affermazioni, ricordiamo che si definisce $\limsup_n A_n = \{\omega \mid \omega \text{ appartiene a infiniti } A_n\}$, e che vale la disuguaglianza $\limsup_n \mathbf{P}(A_n) \leq \mathbf{P}(\limsup_n A_n)$.

Vediamo dunque b): dato $\varepsilon > 0$, l'insieme $\limsup_n \{|X_n - X| > \varepsilon\}$ è trascurabile e pertanto

$$\limsup_n \mathbf{P}\{|X_n - X| > \varepsilon\} \leq \mathbf{P}\left(\limsup_n \{|X_n - X| > \varepsilon\}\right) = 0$$

Per provare c), possiamo estrarre una sottosuccessione $n_1 < n_2 < \dots$ tale che si abbia

$$\mathbf{P}\left\{|X_{n_k} - X| > \frac{1}{k}\right\} \leq 2^{-k}$$

Dal Lemma di Borel-Cantelli 2.4.2 segue che, per quasi ogni ω , definitivamente $|X_{n_k}(\omega) - X(\omega)| \leq 1/k$ (e questo implica la convergenza quasi certa della sottosuccessione). \square

Esempio 4.1.3. Sullo spazio $[0, 1]$ munito della misura di Lebesgue, consideriamo la seguente successione di v.a. : $X_1 = I_{[0,1]}$, $X_2 = I_{[0,1/2]}$, $X_3 = I_{[1/2,1]}$, $X_4 = I_{[0,1/4]}$ e così via \dots . Questa successione di v.a. converge in probabilità a 0, ma non converge *in nessun punto* a 0.

La convergenza in probabilità può essere caratterizzata tramite la convergenza quasi certa nel modo seguente:

Proposizione 4.1.4. *Sia $(X_n)_{n \geq 1}$ una successione di v.a.. Sono equivalenti le affermazioni seguenti:*

- a) $X_n \xrightarrow{\mathbf{P}} X$;
- b) da ogni sottosuccessione si può estrarre una ulteriore sottosuccessione convergente ad X q.c.

Dimostrazione. È evidente che a) implica b); per quanto riguarda il viceversa, supponiamo che la successione $(X_n)_{n \geq 1}$ non converga ad X in probabilità. Ne segue che esistono $\varepsilon > 0$, $\delta > 0$ ed una sottosuccessione tale che si abbia per ogni k , $\mathbf{P}\{|X_{n_k} - X| > \varepsilon\} > \delta$: ogni ulteriore sottosuccessione verifica quella diseuguaglianza e pertanto non può convergere in probabilità né tantomeno q.c. \square

La caratterizzazione della convergenza in probabilità fornita dalla Proposizione 4.1.4 rende immediate alcune conseguenze (ovvie per la convergenza q.c.) tra le quali ad esempio (la dimostrazione è lasciata per esercizio):

- il limite della convergenza in probabilità è unico;
- la convergenza quasi certa non proviene da una distanza (più in generale da una topologia);
- se $X_n \xrightarrow{\mathbf{P}} X$ ed f è continua, $f(X_n) \xrightarrow{\mathbf{P}} f(X)$;
- se $X_n \xrightarrow{\mathbf{P}} X$ e $Y_n \xrightarrow{\mathbf{P}} Y$, anche $(X_n + Y_n) \xrightarrow{\mathbf{P}} (X + Y)$ e $X_n Y_n \xrightarrow{\mathbf{P}} XY$.

Proposizione 4.1.5 (Un rafforzamento del Teorema di Lebesgue).

Supponiamo che $X_n \xrightarrow{\mathbf{P}} X$ e che esista Y integrabile tale che si abbia, per ogni n , $|X_n| \leq Y$: allora $\mathbf{E}[X_n] \rightarrow \mathbf{E}[X]$.

La dimostrazione è lasciata per esercizio.

Diciamo che una successione $(X_n)_{n \geq 1}$ è **di Cauchy in probabilità** se, per ogni $\varepsilon > 0$ e $\delta > 0$, esiste \bar{n} tale che, presi $n, m > \bar{n}$, si abbia $\mathbf{P}\{|X_n - X_m| > \varepsilon\} \leq \delta$.

Teorema 4.1.6. *Ogni successione di v.a. di Cauchy in probabilità converge (in probabilità).*

Dimostrazione. Si può determinare una sottosuccessione $n_1 < n_2 < \dots$ tale che

$$\mathbf{P}\{|X_{n_{k+1}} - X_{n_k}| > 2^{-k}\} \leq 2^{-k}$$

Per il Lemma di Borel–Cantelli, segue che per quasi ogni ω , definitivamente $|X_{n_{k+1}}(\omega) - X_{n_k}(\omega)| \leq 2^{-k}$ e quindi esiste una v.a. X tale che $X_{n_k} \xrightarrow{\text{q.c.}} X$: occorre ora provare che tutta la successione converge a X (in probabilità).

Prendiamo due indici $m > n_k$: dalla relazione insiemistica

$$\{|X_m - X| > \varepsilon\} \subseteq \{|X_m - X_{n_k}| > \varepsilon/2\} \cup \{|X_{n_k} - X| > \varepsilon/2\}$$

segue la disuguaglianza

$$\mathbf{P}\{|X_m - X| > \varepsilon\} \leq \mathbf{P}\{|X_m - X_{n_k}| > \varepsilon/2\} + \mathbf{P}\{|X_{n_k} - X| > \varepsilon/2\}$$

È facile a questo punto concludere che, dato $\delta > 0$, si può determinare un intero $\bar{m} = n_k$ (con k abbastanza grande) tale che, se $m \geq \bar{m}$ si abbia $\mathbf{P}\{|X_m - X| > \varepsilon\} \leq \delta$. \square

Esaminiamo ora gli spazi L^p : è immediato verificare che lo spazio L^∞ è completo. Per quanto riguarda gli spazi L^p con $1 \leq p < +\infty$, la completezza di questi spazi segue rapidamente dai risultati che abbiamo appena visto.

Corollario 4.1.7. *Lo spazio $L^p(\Omega, \mathcal{F}, \mathbf{P})$ è completo.*

Dimostrazione. Se la successione $(X_n)_{n \geq 1}$ è di Cauchy in L^p , lo è pure in probabilità e quindi esiste per il Teorema 4.1.6 una sottosuccessione ed una v.a. X tali che $X_{n_k} \xrightarrow{\text{q.c.}} X$. Poichè si ha, per il Lemma di Fatou, la disuguaglianza

$$\int_{\Omega} |X|^p d\mathbf{P} \leq \liminf_{k \rightarrow \infty} \int_{\Omega} |X_{n_k}|^p d\mathbf{P}$$

segue che la v.a. X appartiene a L^p . Allo stesso modo dalla disuguaglianza

$$\int_{\Omega} |X_{n_k} - X|^p d\mathbf{P} \leq \liminf_{h \rightarrow \infty} \int_{\Omega} |X_{n_k} - X_{n_h}|^p d\mathbf{P}$$

si conclude facilmente che la sottosuccessione $(X_{n_k})_{k \geq 1}$ converge ad X in L^p . Tuttavia da ogni sottosuccessione si può estrarre una ulteriore sottosuccessione convergente in L^p allo stesso limite X , e dunque tutta la successione converge ad X . \square

Anche la convergenza in probabilità (come quella in L^p) è indotta da una distanza: possibili distanze sono ad esempio $d_1(X, Y) = \mathbf{E}[(|X - Y| \wedge 1)]$ oppure $d_2(X, Y) = \mathbf{E}[\arctan(|X - Y|)]$.

In generale si può considerare la distanza $d(X, Y) = \mathbf{E}[h(|X - Y|)]$ dove $h : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ è crescente, continua, limitata, con $h(0) = 0$, strettamente crescente in un intorno di 0 e tale che $h(s + t) \leq h(s) + h(t)$.

Se la funzione h gode di queste proprietà è immediato verificare che la funzione $d(X, Y)$ verifica la *diseguaglianza triangolare* e quindi è effettivamente una distanza: verifichiamo ora che la convergenza secondo questa distanza coincide con la convergenza in probabilità.

Da una parte infatti, se $X_n \xrightarrow{\mathbf{P}} X$, si ottiene che $\mathbf{E}[h(|X_n - X|)] \rightarrow 0$ (vedi la Proposizione 4.1.5).

Viceversa, se $d(X_n, X) = \mathbf{E}[h(|X_n - X|)] \rightarrow 0$, la successione $h(|X_n - X|)$ converge a 0 in probabilità e quindi esiste una sottosuccessione tale che $h(|X_{n_k} - X|)$ converga a 0 q.c.: poiché la funzione h è strettamente crescente in 0, segue immediatamente che la sottosuccessione $X_{n_k} \rightarrow X$ q.c.

Poiché questo si può applicare anche ad ogni sottosuccessione della successione originale, la Proposizione 4.1.4 permette di concludere.

Indichiamo con $L^0(\Omega, \mathcal{F}, \mathbf{P})$ lo spazio di tutte le variabili aleatorie reali, munito della convergenza in probabilità: tenendo conto del Teorema 4.1.6, abbiamo così provato il seguente risultato

Proposizione 4.1.8. *Lo spazio $L^0(\Omega, \mathcal{F}, \mathbf{P})$ è metrico completo.*

Osservazione 4.1.9. Notiamo che gli spazi L^0 e L^∞ sono *invarianti per il passaggio ad una probabilità equivalente* (infatti la loro caratterizzazione dipende solo dagli *insiemi trascurabili*); questa proprietà naturalmente non è soddisfatta dagli spazi L^p con $1 \leq p < +\infty$.

Osservazione 4.1.10. Se μ è una misura σ -finita su uno spazio (E, \mathcal{E}) , si può definire la *convergenza in misura* sostanzialmente con la stessa definizione, ma non tutti i risultati rimangono validi: la principale differenza è che, se f_n converge ad f quasi ovunque, *non è detto che converga in misura* (trovare un controesempio). E quindi non può essere vera, ad esempio, una caratterizzazione della convergenza in misura simile a quella fornita dalla Proposizione 4.1.4.

4.2 Convergenza di misure sulla retta reale.

In questo paragrafo consideriamo misure sulla retta reale (naturalmente munita della σ -algebra di Borel), ma tutti i risultati si estendono con piccole modifiche formali a misure definite su $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.

Una misura μ σ -finita su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ è chiamata **misura di Borel**; una misura tale che, per ogni compatto K , si abbia $\mu(K) < +\infty$ è chiamata

misura di Radon, mentre è chiamata **finita** se $\mu(\mathbb{R}) < +\infty$. La misura che ha densità $|x|^{-1}$ rispetto alla misura di Lebesgue è un esempio di misura di Borel che non è una misura di Radon.

Se f è una funzione a valori reali e se ha senso $\int_{\mathbb{R}} f d\mu$, useremo la notazione abbreviata $\mu(f)$ per indicare $\int_{\mathbb{R}} f d\mu$.

Definizione 4.2.1. Sia $(\mu_n)_{n \geq 1}$ una successione di misure limitate su \mathbb{R} e μ una misura limitata su \mathbb{R} . Si dice che:

- $\mu_n \longrightarrow \mu$ **vagamente** se, per ogni funzione f continua a supporto compatto, $\mu_n(f) \rightarrow \mu(f)$;
- $\mu_n \longrightarrow \mu$ **debolmente** se, per ogni funzione f continua e infinitesima all'infinito, $\mu_n(f) \rightarrow \mu(f)$;
- $\mu_n \longrightarrow \mu$ **strettamente** se, per ogni funzione f continua e limitata, $\mu_n(f) \rightarrow \mu(f)$.

In realtà la definizione di *convergenza vaga* ha senso per misure di Radon (mentre nelle altre due definizioni bisogna assumere che tutte le misure siano limitate), tuttavia noi ci limitiamo a considerare sempre misure limitate (e ci concentreremo in particolare sulle misure di probabilità).

Cominciamo ad osservare che nelle definizioni precedenti è sufficiente limitarsi a considerare funzioni a valori positivi (utilizzando la decomposizione $f = f^+ - f^-$), si può inoltre supporre che si abbia $0 \leq f \leq 1$.

Proposizione 4.2.2. *Sono equivalenti le seguenti affermazioni:*

- a) $\mu_n \longrightarrow \mu$ *strettamente*;
- b) $\mu_n \longrightarrow \mu$ *vagamente* e $\mu_n(\mathbb{R}) \rightarrow \mu(\mathbb{R})$.

Dimostrazione. L'implicazione a) \implies b) è evidente (osservare che $\mu_n(\mathbb{R}) = \int_{\mathbb{R}} 1 d\mu_n$); vediamo invece l'implicazione inversa.

Partiamo da una successione $(\mu_n)_{n \geq 1}$ e supponiamo $\mu_n \rightarrow \mu$ vagamente. Sia f continua, limitata a valori positivi, e consideriamo una successione $(f_k)_{k \geq 1}$ di funzioni continue a supporto compatto tali che $f_k \uparrow f$. Per ogni k si ha

$$\int f_k d\mu = \lim_{n \rightarrow \infty} \int f_k d\mu_n \leq \liminf_{n \rightarrow \infty} \int f d\mu_n$$

e, facendo tendere $k \rightarrow \infty$, si ottiene

$$\int f d\mu = \sup_k \int f_k d\mu \leq \liminf_{n \rightarrow \infty} \int f d\mu_n$$

Notiamo che in particolare $\mu(\mathbb{R}) \leq \liminf_{n \rightarrow \infty} \mu_n(\mathbb{R})$.

Supponiamo ora che si abbia $\mu(\mathbb{R}) = \lim_{n \rightarrow \infty} \mu_n(\mathbb{R})$, e prendiamo f continua con $0 \leq f(x) \leq 1$.

$$\begin{aligned} \int (1 - f) d\mu &= \mu(\mathbb{R}) - \int f d\mu \leq \liminf_{n \rightarrow \infty} \int (1 - f) d\mu_n = \\ &= \lim_{n \rightarrow \infty} \mu_n(\mathbb{R}) - \limsup_{n \rightarrow \infty} \int f d\mu_n \end{aligned}$$

Questa, unita alla disuguaglianza precedente, porta alla relazione

$$\int f d\mu = \lim_{n \rightarrow \infty} \int f d\mu_n$$

□

Quanto alla relazione tra convergenza vaga e debole, vale la seguente relazione:

Proposizione 4.2.3. *Se $\mu_n \rightarrow \mu$ vagamente e $\sup_n \mu_n(\mathbb{R}) < +\infty$, allora $\mu_n \rightarrow \mu$ debolmente.*

Lasciamo la dimostrazione completa per esercizio (osservare che, presa una funzione f continua e infinitesima all'infinito, dato $\varepsilon > 0$, esiste g continua a supporto compatto tale che si abbia $\sup_{x \in \mathbb{R}} |f(x) - g(x)| \leq \varepsilon$).

Notiamo che il risultato precedente può essere invertito nel senso che, se $\mu_n \rightarrow \mu$ debolmente, allora necessariamente $\sup_n \mu_n(\mathbb{R}) < +\infty$; questo risultato però è più delicato e conseguenza di risultati di analisi funzionale (teorema di *Banach Steinhaus*).

Se si considera poi la successione di misure $\mu_n = n \delta_n$ (dove δ_a è la misura di massa 1 concentrata nel punto a) questa fornisce un esempio di successione che converge *vagamente* alla misura 0 ma non converge *debolmente*.

Se μ è una misura finita su \mathbb{R} (non necessariamente di probabilità), è in corrispondenza biunivoca con la sua funzione di ripartizione F definita da $F(x) = \mu(]-\infty, x])$ (mediante la relazione $\mu(]a, b]) = F(b) - F(a)$). Notiamo che si ha $F(+\infty) = \lim_{x \rightarrow \infty} F(x) = \mu(\mathbb{R})$.

Inoltre se μ è associata ad F , si usa la notazione $\int_{\mathbb{R}} f(x) dF(x)$ per indicare $\int_{\mathbb{R}} f(x) d\mu(x)$.

Proposizione 4.2.4. *Sia $(\mu_n)_{n \geq 1}$ una successione di misure di probabilità su \mathbb{R} , μ una misura limitata e siano F_n ed F le relative funzioni di ripartizione. Si hanno i seguenti risultati:*

- a) se $\mu_n \rightarrow \mu$ strettamente, $F_n(x) \rightarrow F(x)$ in tutti i punti x nei quali F è continua;
- b) se $F_n(x) \rightarrow F(x)$ in tutti i punti x di un insieme denso in \mathbb{R} , $\mu_n \rightarrow \mu$ vagamente.

Dimostrazione. Proviamo la prima affermazione: fissiamo $x \in \mathbb{R}$ e $\delta > 0$: prendiamo ora una funzione continua decrescente f tale che $f(t) = 1$ per $t \leq x$ e $f(t) = 0$ per $t \geq (x + \delta)$: si ha

$$F(x + \delta) \geq \mu(f) = \lim_{n \rightarrow \infty} \mu_n(f) \geq \limsup_{n \rightarrow \infty} F_n(x)$$

In modo analogo si ottiene la disuguaglianza $F(x - \delta) \leq \liminf_{n \rightarrow \infty} F_n(x)$ e quindi, se F è continua nel punto x , $F(x) = \lim_{n \rightarrow \infty} F_n(x)$.

Per quanto riguarda la seconda affermazione, si procede in questo modo: su D l'insieme (denso) dei punti nei quali $F_n(x) \rightarrow F(x)$. Data f continua a supporto compatto e dato $\varepsilon > 0$, si può determinare una funzione φ costante a tratti della forma $\varphi(x) = \sum_{j=1}^k a_j I_{]x_j, x_{j+1}]}(x)$ (dove i punti x_j appartengono all'insieme D) e tale che si abbia $\sup_{x \in \mathbb{R}} |f(x) - \varphi(x)| \leq \varepsilon$. Per φ di quella forma, si ha $\int \varphi(x) dF_n(x) \rightarrow \int \varphi(x) dF(x)$, e a questo punto segue facilmente la conclusione. \square

Il risultato che segue è un risultato di *relativa compattezza* (per successioni) delle *sotto-probabilità* (misure μ tali che $\mu(\mathbb{R}) \leq 1$).

Teorema 4.2.5 (Teorema di Helly). *Sia $(\mu_n)_{n \geq 1}$ una successione di misure di probabilità su \mathbb{R} : esiste una sottosuccessione convergente debolmente ad una misura μ tale che $\mu(\mathbb{R}) \leq 1$.*

Dimostrazione. Sia F_n la successione delle relative funzioni di ripartizione e numeriamo l'insieme dei razionali $\mathbb{Q} = \{p_1, p_2, \dots\}$: esiste una sottosuccessione tale che $F_{n'}(p_1)$ converga a un numero $G(p_1)$, da questa si può estrarre una sottosuccessione tale che $F_{n''}(p_2)$ converga a $G(p_2)$ e così via ... con il criterio diagonale, si estrae una sottosuccessione tale che, per ogni razionale p , $F_{n_k}(p)$ converga a $G(p)$.

La funzione G , definita sui razionali, è crescente e a valori in $[0, 1]$. Definiamo

$$F(x) = \lim_{p > x, p \rightarrow x} G(p) = \inf_{p > x, p \in \mathbb{Q}} G(p)$$

La funzione F è crescente, continua a destra, però non necessariamente $F(-\infty) = 0$ e $F(+\infty) = 1$; ad essa è associata una ed una sola misura μ tale che, per $a < b$, valga l'eguaglianza $\mu([a, b]) = F(b) - F(a)$.

Si ha $\mu(\mathbb{R}) = F(+\infty) - F(-\infty) \leq 1$.

Proviamo che $F_n(x) \rightarrow F(x)$ in tutti i punti in cui F è continua. Sia $x \in \mathbb{R}$ ed $\varepsilon > 0$: scegliamo due razionali p e q tali che $x - \varepsilon < p < x < q < x + \varepsilon$. Si ha

$$F(x - \varepsilon) \leq G(p) = \lim_{k \rightarrow \infty} F_{n_k}(p) \leq \liminf_{k \rightarrow \infty} F_{n_k}(x)$$

In modo analogo si prova che $F(x + \varepsilon) \geq \limsup_{k \rightarrow \infty} F_{n_k}(x)$ e, se F è continua in x , si ottiene $F_{n_k}(x) \rightarrow F(x)$.

Ripetendo la dimostrazione del punto b) della Proposizione 4.2.4, si ottiene che $\mu_{n_k} \rightarrow \mu$ debolmente. \square

Purtroppo la misura limite, la cui esistenza è provata dal teorema precedente, non è necessariamente una *probabilità*: è importante allora fornire delle condizioni che garantiscano che il limite è ancora una probabilità. È fondamentale a questo riguardo il concetto che viene ora introdotto.

Definizione 4.2.6. Una famiglia \mathcal{M} di probabilità su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ è detta **tesa** se, per ogni $\varepsilon > 0$, esiste un compatto K tale che, qualunque sia $\mu \in \mathcal{M}$, valga la disuguaglianza $\mu(K) \geq 1 - \varepsilon$.

Naturalmente una famiglia finita μ_1, \dots, μ_n di probabilità è tesa: si considera per ogni i un compatto K_i tale che $\mu_i(K_i) \geq 1 - \varepsilon$ e si prende poi $K = K_1 \cup \dots \cup K_n$.

Teorema 4.2.7 (Teorema di Prohorov). *Sia \mathcal{M} una famiglia di probabilità su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Sono equivalenti le seguenti proprietà:*

- a) *la famiglia \mathcal{M} è relativamente compatto per successioni rispetto alla convergenza stretta;*
- b) *la famiglia \mathcal{M} è tesa.*

Dimostrazione. Cominciamo a provare che b) implica a): sia $(\mu_n)_{n \geq 1}$ una successione contenuta in \mathcal{M} . Per il Teorema 4.2.5, esiste una sottosuccessione $(\mu_{n_k})_{k \geq 1}$ convergente debolmente a una misura μ con $\mu(\mathbb{R}) \leq 1$: rimane da provare che si ha $\mu(\mathbb{R}) = 1$.

Per la condizione di tensione, dato $\varepsilon > 0$, esiste un compatto (che possiamo supporre della forma $[-M, M]$) tale che si abbia $\mu_n([-M, M]) \geq (1 - \varepsilon)$ qualunque sia n . Consideriamo la funzione (continua a supporto compatto) f che vale 1 sull'intervallo $[-M, M]$, che vale 0 fuori dall'intervallo $[-(M+1), (M+1)]$ e che si raccorda linearmente: valgono le disuguaglianze

$$\mu([- (M + 1), (M + 1)]) \geq \int f \, d\mu = \lim_{k \rightarrow \infty} \int f \, d\mu_{n_k} \geq$$

$$\geq \limsup_{k \rightarrow \infty} \mu_{n_k}([-M, M]) \geq (1 - \varepsilon)$$

e ne segue l'eguaglianza $\mu(\mathbb{R}) = 1$.

Per provare che b) implica a), cominciamo a vedere che se una successione $(\mu_n)_{n \geq 1}$ converge strettamente a μ , è tesa: prendiamo infatti $\varepsilon > 0$ ed M tali che $\mu([-M, M]) \geq 1 - (\varepsilon/2)$. Con un procedimento analogo a quanto fatto sopra, si prova la disequaglianza

$$\mu([-M, M]) \leq \liminf_{n \rightarrow \infty} \mu_n([-(M+1), (M+1)])$$

e quindi, da un certo \bar{n} in poi, si ha $\mu_n([-(M+1), (M+1)]) \geq (1 - \varepsilon)$.

Supponiamo ora che la famiglia \mathcal{M} non sia tesa: esiste $\varepsilon > 0$ tale che, per ogni n , si può scegliere una misura $\mu_n \in \mathcal{M}$ tale che $\mu_n([-n, n]) < (1 - \varepsilon)$. È evidente che nessuna sottosuccessione può essere tesa e pertanto non può convergere strettamente. \square

Segnaliamo il risultato seguente, la cui dimostrazione è rinviata all'Appendice, ma che spesso risulta comodo:

Teorema 4.2.8. *Supponiamo che $\mu_n \rightarrow \mu$ strettamente e sia f una funzione boreliana limitata tale che l'insieme dei suoi punti di discontinuità sia trascurabile per la misura limite μ : allora $\mu_n(f) \rightarrow \mu(f)$.*

Osservazione 4.2.9. Tutti i risultati di questo paragrafo sono validi, con piccole modifiche nelle dimostrazioni, per misure a valori nello spazio Euclideo \mathbb{R}^d .

Sono anche validi per misure a valori in uno spazio E metrizzabile, separabile e completo (detto *spazio polacco*), ma questa volta le dimostrazioni sono sensibilmente diverse e poggiano essenzialmente su risultati di analisi funzionale. In particolare l'estensione del Teorema di Helly 4.2.5 è una conseguenza di un teorema più generale di Analisi Funzionale (il teorema di Banach-Alouglu).

Inoltre la convergenza stretta di misure di probabilità è indotta da una distanza: anche questo fatto è conseguenza di teoremi più generali di Analisi Funzionale. Tuttavia nel caso delle probabilità su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ si può darne anche una dimostrazione elementare: una possibile distanza tra due probabilità su \mathbb{R} aventi funzione di ripartizione rispettivamente F e G è data da

$$d(F, G) = \int_{-\infty}^{+\infty} e^{-|x|} |F(x) - G(x)| dx$$

4.3 Convergenza in Legge di variabili aleatorie.

Anche questa volta ci limitiamo, per semplicità di esposizione, a variabili aleatorie reali, ma tutti i risultati sono validi per variabili aleatorie a valori nello spazio \mathbb{R}^d .

Definizione 4.3.1. Si dice che la successione $(X_n)_{n \geq 1}$ di v.a. reali converge **in legge** (o anche **in distribuzione**) alla variabile aleatoria X (e si scrive $X_n \xrightarrow{\mathcal{L}} X$) se le relative leggi di probabilità $\mathbf{P}_{X_n} = X_n(\mathbf{P})$ convergono strettamente alla legge $\mathbf{P}_X = X(\mathbf{P})$.

Questo equivale a dire che, presa f continua e limitata,

$$\mathbf{E}[f(X_n)] = \int_{\mathbb{R}} f(x) d\mathbf{P}_{X_n}(x) \longrightarrow \int_{\mathbb{R}} f(x) d\mathbf{P}_X(x) = \mathbf{E}[f(X)]$$

Naturalmente, poichè il limite è di nuovo una probabilità, è la stessa cosa imporre che la funzione sopra scritta sia continua a supporto compatto (cioè che la convergenza sia *vaga*, vedi 4.2.2).

Come conseguenza del Teorema 4.2.8, il limite $\mathbf{E}[f(X_n)] \rightarrow \mathbf{E}[f(X)]$ vale anche se f è boreliana, limitata e *l'insieme dei punti di discontinuità di f è trascurabile per la probabilità limite \mathbf{P}_X* .

Sono immediate le seguenti conseguenze:

- se $X_n \xrightarrow{\mathcal{L}} X$ e g è continua $g(X_n) \xrightarrow{\mathcal{L}} g(X)$;
- se $X_n \xrightarrow{\mathcal{L}} X$ ed f è continua a valori positivi, si ha $\mathbf{E}[f(X)] \leq \liminf_{n \rightarrow \infty} \mathbf{E}[f(X_n)]$ (vedi la prima parte della dimostrazione di 4.2.2).

Proposizione 4.3.2. a) se $X_n \xrightarrow{\mathbf{P}} X$, allora $X_n \xrightarrow{\mathcal{L}} X$;

b) se $X_n \xrightarrow{\mathcal{L}} c$, allora $X_n \xrightarrow{\mathbf{P}} c$.

La facile dimostrazione è lasciata per esercizio: per la prima parte si suggerisce di ricordare la Proposizione 4.1.5 e per la seconda di considerare la *distanza* della convergenza in probabilità $d(X, c) = \mathbf{E}[h(|X_n - c|)]$.

Osservazione 4.3.3. La **convergenza in legge** ha anche delle *cattive proprietà* dalle quali bisogna stare in guardia: soprattutto non è vero (a differenza di quanto succede con la convergenza in probabilità) che se $X_n \xrightarrow{\mathcal{L}} X$ e $Y_n \xrightarrow{\mathcal{L}} Y$, allora $(X_n + Y_n) \xrightarrow{\mathcal{L}} (X + Y)$.

Guardiamo questo esempio: sia X una v.a. $N(0, 1)$ e definiamo $X_n = X$ e $Y_n = -X$. Notiamo che $\mathbf{P}_{X_n} = \mathbf{P}_{Y_n} = \mathbf{P}_X$ e tuttavia $X_n + Y_n = 0$ non converge a $2X$.

In verità la convergenza in legge non dovrebbe essere vista come convergenza di variabili aleatorie ma come convergenza di misure (cioè delle loro leggi di probabilità).

È molto istruttiva la Proposizione seguente, la cui dimostrazione è lasciata per esercizio:

Esercizio 4.3.4. Siano X_n ed X variabili aleatorie definite sul medesimo spazio di probabilità. Sono equivalenti

- a) $X_n \xrightarrow{\mathbf{P}} X$;
- b) la coppia (X_n, X) converge in legge a (X, X) .

La Proposizione 4.2.4 si traduce immediatamente nel seguente risultato:

Teorema 4.3.5. Sia $(X_n)_{n \geq 1}$ una successione di variabili aleatorie, $(F_n)_{n \geq 1}$ le relative funzioni di ripartizione ed X una v.a. con funzione di ripartizione F . Sono equivalenti le seguenti affermazioni:

- a) $X_n \xrightarrow{\mathcal{L}} X$;
- b) $F_n(\cdot)$ converge ad $F(\cdot)$ in tutti i punti in cui F è continua;
- c) $F_n(\cdot)$ converge ad $F(\cdot)$ in tutti i punti di un insieme denso.

È di fondamentale importanza il risultato seguente, che lega la convergenza in legge alla convergenza delle funzioni caratteristiche: questo risultato è indubbiamente la proprietà più importante delle funzioni caratteristiche.

Teorema 4.3.6 (Teorema di Paul Lévy). Sia $(X_n)_{n \geq 1}$ una successione di variabili aleatorie, siano $\varphi_n(\cdot)$ le relative funzioni caratteristiche e sia X una v.a.

- a) se $X_n \xrightarrow{\mathcal{L}} X$, allora $\varphi_n(t)$ converge puntualmente a $\varphi_X(t)$;
- b) supponiamo che la successione delle funzioni caratteristiche converga puntualmente ad una funzione φ **continua nel punto 0**: allora φ è la funzione caratteristica di una v.a. X e $X_n \xrightarrow{\mathcal{L}} X$.

Dimostrazione. La parte a) è immediata: poichè la funzione $x \rightarrow e^{itx}$ è continua limitata, $\varphi_n(t) = \mathbf{E}[e^{itX_n}] \rightarrow \mathbf{E}[e^{itX}] = \varphi_X(t)$.

Per quanto riguarda la parte b), tutto si riduce a provare che dalla convergenza delle funzioni φ_n (a una funzione continua in 0) segue che la successione delle leggi \mathbf{P}_{X_n} è *tesa*.

Infatti (una volta dimostrata questa proprietà) si può estrarre una sottosuccessione \mathbf{P}_{n_k} convergente strettamente ad una probabilità \mathbf{P} su \mathbb{R} : presa una v.a. X con legge \mathbf{P} , dalla parte a) segue che \mathbf{P} è l'unica probabilità su \mathbb{R} che ha funzione caratteristica φ ; poichè da ogni sottosuccessione si può estrarre una ulteriore sottosuccessione convergente strettamente sempre a \mathbf{P} , si ottiene il risultato voluto.

Per dimostrare la *tensione* delle leggi $(\mathbf{P}_{X_n})_{n \geq 1}$, cominciamo con una disegualianza: se Y è una v.a. con funzione caratteristica φ_Y ed $u > 0$, si ha

$$\mathbf{P}\{|Y| > \frac{1}{u}\} \leq \frac{1}{2u} \int_{-2u}^{2u} (1 - \varphi_Y(t)) dt$$

Valgono le seguenti relazioni (con $u > 0$):

$$\begin{aligned} & \frac{1}{2u} \int_{-2u}^{2u} (1 - \varphi_Y(t)) dt = \int_{-\infty}^{+\infty} \left[\frac{1}{2u} \int_{-2u}^{+2u} (1 - e^{ity}) dt \right] d\mathbf{P}_Y(y) = \\ & = 2 \int_{-\infty}^{+\infty} \left(1 - \frac{\sin 2uy}{2uy}\right) d\mathbf{P}_Y(y) \geq 2 \int_{[-(1/u), (1/u)]^c} \left(1 - \frac{\sin 2uy}{2uy}\right) d\mathbf{P}_Y(y) \geq \\ & \geq 2 \int_{[-(1/u), (1/u)]^c} \left(1 - \frac{1}{2|uy|}\right) d\mathbf{P}_Y(y) \geq \mathbf{P}_Y\left(\left[-\frac{1}{u}, \frac{1}{u}\right]^c\right) \end{aligned}$$

Poichè la funzione φ è continua in 0 e $\varphi(0) = 1$, dato $\varepsilon > 0$ esiste u abbastanza piccolo tale che $\frac{1}{2u} \int_{-2u}^{2u} (1 - \varphi(t)) dt \leq \varepsilon$ e poichè (per il teorema di convergenza dominata)

$$\frac{1}{2u} \int_{-2u}^{2u} (1 - \varphi_n(t)) dt \longrightarrow \frac{1}{2u} \int_{-2u}^{2u} (1 - \varphi(t)) dt$$

ne segue che esiste \bar{n} tale che per $n \geq \bar{n}$ si abbia $\frac{1}{2u} \int_{-2u}^{2u} (1 - \varphi_n(t)) dt < 2\varepsilon$, cioè $\mathbf{P}_{X_n}\left(\left[-\frac{1}{u}, \frac{1}{u}\right]\right) \geq 1 - 2\varepsilon$.

Questo conclude la dimostrazione. □

Vediamo una conseguenza interessante:

Corollario 4.3.7. *Supponiamo che, per ogni n , X_n sia gaussiana e che $X_n \xrightarrow{\mathcal{L}} X$: anche X è gaussiana.*

I dettagli della *dimostrazione* sono lasciati per esercizio: si tratta di provare che, se $X_n \sim N(m_n, \sigma_n^2)$ e le funzioni caratteristiche convergono, esistono m e σ^2 tali che $m_n \rightarrow m$ e $\sigma_n^2 \rightarrow \sigma^2$ e di conseguenza, presa $X \sim N(m, \sigma^2)$, $X_n \xrightarrow{\mathcal{L}} X$.

Per quanto riguarda la convergenza delle varianze σ_n^2 , questa è facile: infatti se convergono le funzioni caratteristiche convergono anche i loro *moduli* (che sono eguali a $\exp(-\sigma^2 t^2/2)$) e tenendo ad esempio fisso $t = 1$ si vede subito il risultato voluto. È un poco più delicato provare che dalla convergenza di $\exp(im_n t)$ segue necessariamente la convergenza delle medie m_n ad un numero m : tutto questo però diventa ragionevolmente semplice se si prova che la successione delle medie $(m_n)_{n \geq 1}$ è limitata. Quest'ultima affermazione segue dal fatto che le varianze sono limitate e dalla condizione di *tensione*.

L'estensione del Teorema 4.3.6 a variabili aleatorie a valori in \mathbb{R}^d è facile: limitiamoci per semplicità al caso $d = 2$.

L'unica implicazione non ovvia è che, se la successione delle funzioni caratteristiche $\varphi_{X_n, Y_n}(u, v)$ converge puntualmente ad una funzione $\varphi(u, v)$ continua in 0, allora la successione delle leggi \mathbf{P}_{X_n, Y_n} è tesa. Poiché $\varphi_{X_n, Y_n}(u, 0) = \varphi_{X_n}(u)$ segue che la successione delle leggi \mathbf{P}_{X_n} è tesa (e analogamente quella delle leggi \mathbf{P}_{Y_n}): da questo segue che anche la successione delle leggi congiunte è tesa.

Infatti, fissato $\varepsilon > 0$, esistono due reali positivi k_1 e k_2 tali che, qualunque sia n , si abbia

$$\mathbf{P}\{|X_n| > k_1\} = \mathbf{P}_{X_n}\left([-k_1, k_1]^c\right) \leq \frac{\varepsilon}{2}$$

e analogamente

$$\mathbf{P}\{|Y_n| > k_2\} = \mathbf{P}_{Y_n}\left([-k_2, k_2]^c\right) \leq \frac{\varepsilon}{2}$$

Poiché il complementare del compatto (di \mathbb{R}^2) $[-k_1, k_1] \times [-k_2, k_2]$ è eguale a $[-k_1, k_1]^c \times \mathbb{R} \cup \mathbb{R} \times [-k_2, k_2]^c$ e

$$\mathbf{P}_{X_n, Y_n}\left([-k_1, k_1]^c \times \mathbb{R}\right) = \mathbf{P}_{X_n}\left([-k_1, k_1]^c\right)$$

$$\mathbf{P}_{X_n, Y_n}\left(\mathbb{R} \times [-k_2, k_2]^c\right) = \mathbf{P}_{Y_n}\left([-k_2, k_2]^c\right)$$

si ottiene immediatamente

$$\mathbf{P}_{X_n, Y_n}\left([-k_1, k_1] \times [-k_2, k_2]\right) \geq 1 - \varepsilon$$

Si può notare che in realtà è sufficiente un'ipotesi un poco più debole, è sufficiente cioè che le funzioni caratteristiche φ_{X_n, Y_n} convergano puntualmente ad una funzione φ tale che le due funzioni $u \rightarrow \varphi(u, 0)$ e $v \rightarrow \varphi(0, v)$ siano continue in 0.

Una facile conseguenza del Teorema 4.3.6 è il seguente risultato, noto talvolta come **Teorema di Slutsky**.

Corollario 4.3.8. *Siano X_n, Y_n, X tutte definite sul medesimo spazio di probabilità e supponiamo che $X_n \xrightarrow{\mathcal{L}} X$ e $Y_n \xrightarrow{\mathbf{P}} c$: allora la successione delle coppie (X_n, Y_n) converge in legge alla coppia (X, c) .*

La dimostrazione probabilmente più rapida consiste nel provare che la successione delle funzioni caratteristiche $\varphi_{(X_n, Y_n)}(u, v)$ converge puntualmente a $\varphi_X(u) \cdot e^{ivc}$: i dettagli sono lasciati per esercizio.

È importante rimarcare che, se $X_n \xrightarrow{\mathcal{L}} X$ e $Y_n \xrightarrow{\mathbf{P}} Y$, non è affatto detto che la successione (X_n, Y_n) converga in legge a (X, Y) .

È molto importante anche il seguente risultato:

Teorema 4.3.9 (Teorema di rappresentazione di Skorohod). *Sia $(X_n)_{n \geq 1}$ una successione di variabili aleatorie reali e supponiamo che $X_n \xrightarrow{\mathcal{L}} X$: si può costruire su uno spazio di probabilità $(\Omega', \mathcal{F}', \mathbf{P}')$ una successione di v.a. $(Y_n)_{n \geq 1}$ ed una v.a. Y tali che $X_n(\mathbf{P}) = Y_n(\mathbf{P}')$, $X(\mathbf{P}) = Y(\mathbf{P}')$ e $Y_n \xrightarrow{\text{q.c.}} Y$.*

Questo teorema è valido anche per variabili aleatorie a valori in \mathbb{R}^d o più in generale in uno spazio polacco; la dimostrazione però (anche solo nel caso d -dimensionale) è decisamente più complessa. Tuttavia questo risultato (nella sua formulazione più generale) è molto utile nello studio dei processi stocastici.

Solo nel caso di variabili a valori reali la dimostrazione è ragionevolmente semplice, ed è questa che affronteremo.

Dimostrazione. Data una funzione di ripartizione F , definiamo la sua *pseudo-inversa sinistra* nel modo seguente (per $0 < t < 1$):

$$G(t) = \inf \{x \in \mathbb{R} : F(x) \geq t\}$$

Non è difficile verificare che si ha

$$G(t) \leq x \iff t \leq F(x)$$

Consideriamo allora sullo spazio $\Omega' = [0, 1]$ (munito della σ -algebra di Borel) la probabilità $\mathbf{P}' = \lambda$ (misura di Lebesgue).

Se si considera la funzione $G(\cdot)$ come v.a. sullo spazio Ω' , si trova che

$$\lambda\{t : G(t) \leq x\} = \lambda\{t : t \leq F(x)\} = F(x)$$

cioè la legge di G è associata alla funzione di ripartizione F .

Assegnate dunque le v.a. X_n e X e le loro funzioni di ripartizione F_n ed F , chiamiamo Y_n ed Y le v.a. G_n e G corrispondenti alle pseudo-inverse sinistre di F_n ed F : si tratta di provare che se $F_n(\cdot) \rightarrow F(\cdot)$ in tutti i punti di continuità di F , $G_n(\cdot)$ converge q.o. a $G(\cdot)$.

Chiamiamo D l'insieme (al più numerabile) dei punti $t \in]0, 1[$ per i quali esiste un $y \in \mathbb{R}$ con $F(y) = t$ ed inoltre F è costante su un intorno del punto y : se $t \notin D$ è facile constatare che, dato $\varepsilon > 0$, esiste un punto x con $G(t) < x < G(t) + \varepsilon$ tale che $F(x) > t$ ed F sia continua nel punto x (osservare che $F(G(t)) \geq t$).

Poichè $F_n(x) \rightarrow F(x)$, si ha $F_n(x) \geq t$ per n abbastanza grande, e per tali n di conseguenza $G_n(t) \leq x < G(t) + \varepsilon$.

Si ottiene così la diseuguaglianza $\limsup_{n \rightarrow \infty} G_n(t) \leq G(t) + \varepsilon$.

In modo del tutto analogo si ottiene (sempre se $t \notin D$) la diseuguaglianza $\liminf_{n \rightarrow \infty} G_n(t) \geq (G(t) - \varepsilon)$ e, poichè questo è vero per ogni $\varepsilon > 0$, si ha appunto l'eguaglianza $\lim_{n \rightarrow \infty} G_n(t) = G(t)$. \square

Osservazione 4.3.10. Il precedente Teorema 4.3.9 permette di dimostrare, senza far uso del Teorema 4.2.8, il seguente risultato: se $X_n \xrightarrow{\mathcal{L}} X$ e se f è misurabile limitata, e continua eccetto che su un insieme trascurabile per \mathbf{P}_X , allora $\mathbf{E}[f(X_n)] \rightarrow \mathbf{E}[f(X)]$.

4.4 Appendice

4.4.1 Diseuguaglianze negli spazi L^p .

Le principali disuguaglianze che riguardano gli spazi L^p sono una conseguenza della *diseuguaglianza di Jensen* (vedi Lemma 1.2.9) che è stata dimostrata nel caso scalare ma ammette una estensione su \mathbb{R}^d che si dimostra con la stessa tecnica: se $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ è convessa, le variabili X_1, \dots, X_d sono integrabili e ha senso $\mathbf{E}[\varphi(X_1, \dots, X_d)]$ si ha

$$\varphi\left(\mathbf{E}[X_1], \dots, \mathbf{E}[X_d]\right) \leq \mathbf{E}\left[\varphi(X_1, \dots, X_d)\right]$$

Se invece si considera una funzione *concava* ψ , la diseuguaglianza è soddisfatta nell'altro senso. Inoltre la funzione φ (o ψ) non è necessario che sia definita su tutto \mathbb{R}^d , è sufficiente che sia definita su un sottinsieme convesso.

Prendiamo allora $0 < \alpha < 1$ e consideriamo la funzione (definita per $u \geq 0, v \geq 0$) $\varphi_\alpha(u, v) = u^\alpha v^{1-\alpha}$: è un esercizio verificare che è *concava*. Se si prendono due variabili aleatorie U, V a valori positivi si ottiene

$$\mathbf{E}[U^\alpha V^{1-\alpha}] \leq \mathbf{E}[U]^\alpha \mathbf{E}[V]^{1-\alpha}$$

Prendiamo allora $p > 1$, $\alpha = 1/p$ e $1-\alpha = 1/q$ (dove q è l'esponente *coniugato* di p): applicando la disuguaglianza sopra scritta alle variabili $U = |X|^p$ e $V = |Y|^q$ si ottiene la **disuguaglianza di Holder**

$$\mathbf{E}[|XY|] \leq \left(\mathbf{E}[|X|^p]\right)^{\frac{1}{p}} \left(\mathbf{E}[|Y|^q]\right)^{\frac{1}{q}}$$

che si scrive più spesso nella forma $\|XY\|_1 \leq \|X\|_p \|Y\|_q$.

Come si vede, la disuguaglianza di Holder generalizza la *disuguaglianza di Schwartz*. Inoltre essa si estende immediatamente a questa disuguaglianza: se $1/r = 1/p + 1/q$ (r, p, q maggiori di 1), si ha $\|XY\|_r \leq \|X\|_p \|Y\|_q$.

Consideriamo ora $1 < p < +\infty$ e la funzione (anche questa definita per u e v positivi) $\psi_p(u, v) = (u^{1/p} + v^{1/p})^p$: anche questa funzione è *concava* e applicando la disuguaglianza alle due variabili $|X|^p$ e $|Y|^p$ si ottiene

$$\mathbf{E}\left[(|X| + |Y|)^p \right] \leq \left(\mathbf{E}[|X|^p]^{\frac{1}{p}} + \mathbf{E}[|Y|^p]^{\frac{1}{p}} \right)^p$$

e di conseguenza la **disuguaglianza di Minkovsky** $\|X+Y\|_p \leq \|X\|_p + \|Y\|_p$. Unita all'ovvia eguaglianza $\|\lambda X\|_p = |\lambda| \|X\|_p$, questa mostra che l'applicazione $X \rightarrow \|X\|_p$ è una *norma* su L^p .

Per quanto riguarda lo spazio $L^{+\infty}$, questo è lo spazio delle v.a. per le quali esiste una costante M tale che si abbia $|X(\omega)| \leq M$ q.c.: di tali costanti maggioranti (quasi ovunque) esiste un *elemento minimo*.

Ponendo infatti α l'estremo inferiore dell'insieme $\mathcal{M} = \{M : |X(\omega)| \leq M \text{ q.c.}\}$, esiste una successione M_n convergente decrescendo ad α e poiché si ha $\{|X| > \alpha\} = \bigcup_n \{|X| > M_n\}$ anche $\{|X| > \alpha\}$ è trascurabile: si definisce allora $\|X\|_\infty$ la **minima** costante M che maggiora $|X|$ q.c. e la disuguaglianza $\|X+Y\|_\infty \leq \|X\|_\infty + \|Y\|_\infty$ è evidente, così come è immediata la *completezza* dello spazio L^∞ .

4.4.2 Convergenza di misure e “funzioni semicontinue”.

Scopo di questa sezione è dimostrare il Teorema 4.2.8, del quale non forniamo però una dimostrazione dettagliata ma ci limitiamo ad esporre le *linee della dimostrazione completa*.

Per semplicità, supponiamo che μ_n e μ siano *misure di probabilità* su \mathbb{R} e che la successione $(\mu_n)_{n \geq 1}$ converga *strettamente* a μ .

Prendiamo un aperto A : è facile constatare che esiste una successione crescente di funzioni continue a valori in $[0, 1]$ convergente alla *funzione indicatrice* di A (ad esempio si può prendere $f_k(x) = (k \cdot \text{dist}(x, A^c)) \wedge 1$).

Si ha pertanto, per ogni k ,

$$\mu(f_k) = \lim_{n \rightarrow \infty} \mu_n(f_k) \leq \liminf_{n \rightarrow \infty} \mu_n(A)$$

e di conseguenza

$$\mu(A) = \sup_k \mu(f_k) \leq \liminf_{n \rightarrow \infty} \mu_n(A)$$

Naturalmente, per G chiuso, vale la diseuguaglianza opposta

$$\mu(G) \geq \limsup_{n \rightarrow \infty} \mu_n(G)$$

(si può anzi dimostrare facilmente che queste due diseuguaglianze, equivalenti tra di loro, *caratterizzano* la convergenza stretta).

Consideriamo ora una funzione (limitata, a valori positivi) **semicontinua inferiormente**: una funzione f definita su \mathbb{R} è detta *semicontinua inferiormente* se per ogni $c \in \mathbb{R}$ gli insiemi $\{f > c\}$ sono aperti, o equivalentemente se per ogni $x \in \mathbb{R}$ si ha $f(x) \leq \liminf_{y \rightarrow x} f(y)$ e si prova facilmente che se f è *semicontinua inferiormente* a valori in $[0, 1]$ esiste una successione crescente di funzioni continue convergente ad f (si può porre ad esempio $f_k(x) = \sum_{i=1}^{k-1} \frac{1}{k} I_{\{f > \frac{i}{k}\}}$ e a loro volta gli insiemi $\{f > \frac{i}{k}\}$ sono aperti e le loro funzioni indicatrici si possono approssimare con una successione crescente di funzioni continue).

Pertanto, esattamente come si era fatto per gli aperti, si prova che se f è *semicontinua inferiormente* (a valori in $[0, 1]$) si ha

$$\mu(f) \leq \liminf_{n \rightarrow \infty} \mu_n(f)$$

Analogamente è la definizione di funzione **semicontinua superiormente**: il modo più semplice di procedere è dire che f è *semicontinua superiormente* se $-f$ è *semicontinua inferiormente* e di conseguenza (per f s.c.s. a valori in $[0, 1]$) si ha

$$\mu(f) \geq \limsup_{n \rightarrow \infty} \mu_n(f)$$

Prendiamo ora f boreliana limitata: esiste la *minima funzione semicontinua superiormente maggiorante* della quale si può dare l'espressione esplicita

$$f^*(x) = \limsup_{y \rightarrow x} f(y) = \lim_{d \rightarrow 0^+} \sup_{|x-y| < d} f(y)$$

e analogamente esiste la *massima funzione semicontinua inferiormente minorante* f_* .

Prendiamo ora f boreliana limitata (a valori in $[0, 1]$) e notiamo che vale la catena di disequaglianze

$$\mu(f_*) \leq \liminf_{n \rightarrow \infty} \mu_n(f_*) \leq \liminf_{n \rightarrow \infty} \mu_n(f) \leq \limsup_{n \rightarrow \infty} \mu_n(f) \leq \limsup_{n \rightarrow \infty} \mu_n(f^*) \leq \mu(f^*)$$

Notiamo a questo punto che l'insieme $\{f_* \neq f^*\}$ coincide con *l'insieme dei punti di discontinuità di f* . Se tale insieme è *trascurabile* per μ si ha $f = f^* = f_*$ μ q.c. , le disequaglianze sopra scritte sono tutte eguaglianze e ne segue la dimostrazione del teorema 4.2.8.

Capitolo 5

Teoremi limite

5.1 Teoremi Limite Centrale.

In tutto questo capitolo, se X_1, X_2, \dots è una successione di v.a.r., si indica $S_n = X_1 + \dots + X_n$.

Il risultato che ora segue generalizza il Teorema Limite Centrale di De Moivre Laplace (per variabili binomiali) visto al primo corso:

Teorema 5.1.1 (Teorema Limite Centrale di Paul Lévy). *Sia $(X_n)_{n \geq 1}$ una successione di variabili i.i.d. dotate di momento secondo, e siano $\mu = \mathbf{E}[X_i]$ e $\sigma^2 = \text{Var}(X_i) > 0$: allora*

$$Z_n = \frac{X_1 + \dots + X_n - n \cdot \mu}{\sigma \sqrt{n}} \xrightarrow{\mathcal{L}} X$$

dove $X \sim N(0, 1)$.

Dimostrazione. Grazie al Teorema 4.3.6 la dimostrazione è molto semplice: è sufficiente infatti provare che la successione delle funzioni caratteristiche delle v.a. Z_n converge a $e^{-t^2/2}$.

Chiamiamo $Y_i = \frac{X_i - \mu}{\sigma}$ e notiamo che le v.a. sono i.i.d. con valore atteso 0 e varianza 1: si ha pertanto lo sviluppo $\varphi_{Y_i}(t) = 1 - t^2/2 + o(t^2)$. Notiamo che $Z_n = \frac{Y_1 + \dots + Y_n}{\sqrt{n}}$ e di conseguenza

$$\varphi_{Z_n}(t) = \varphi_{Y_1}\left(\frac{t}{\sqrt{n}}\right) \cdots \varphi_{Y_n}\left(\frac{t}{\sqrt{n}}\right) = \left[\varphi_{Y_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

Per provare che si ha $\lim_{n \rightarrow \infty} \varphi_{Z_n}(t) = e^{-t^2/2}$, conviene passare ai *logaritmi*: qui però è doverosa una precisazione, poichè le funzioni $\varphi_{Z_n}(t)$ sono

a valori complessi. Infatti non si può definire il logaritmo in modo univoco e con continuità su tutto il piano complesso, ma si può farlo sul piano complesso *privato della semiretta dei numeri reali negativi*: in tal caso infatti z si può scrivere nella forma $x = re^{i\theta}$ con $r > 0$ e $-\pi < \theta < \pi$. Si pone allora $\log z = \log r + i\theta$ (alcuni testi chiamano questo il *logaritmo principale*): inoltre se $|z| < 1$, vale lo sviluppo in serie $\log(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} + \dots$

A questo punto un facile calcolo permette di concludere. □

Il risultato precedente ammette una immediata estensione alle variabili aleatorie vettoriali:

Teorema 5.1.2. *Sia $(\mathbf{X}_n)_{n \geq 1}$ una successione di variabili aleatorie i.i.d. a valori in \mathbb{R}^d dotate di momenti del second'ordine, e siano $\mathbf{m} = \mathbf{E}[\mathbf{X}_i]$ e $\mathbf{Q} = \text{Cov}(\mathbf{X}_i)$: allora*

$$\frac{\mathbf{X}_1 + \dots + \mathbf{X}_n - n\mathbf{m}}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathbf{X}$$

dove \mathbf{X} ha legge $N_d(0, \mathbf{Q})$.

La dimostrazione segue facilmente dal seguente ben noto risultato: siano $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}$ variabili aleatorie a valori in \mathbb{R}^d . Sono equivalenti le seguenti affermazioni:

- a) $\mathbf{X}_n \xrightarrow{\mathcal{L}} \mathbf{X}$
- b) per ogni $\mathbf{u} \in \mathbb{R}^d$, $\langle \mathbf{u}, \mathbf{X}_n \rangle \xrightarrow{\mathcal{L}} \langle \mathbf{u}, \mathbf{X} \rangle$.

In generale data una successione X_1, X_2, \dots di variabili aleatorie reali, si dice che vale un **Teorema Limite Centrale** se la successione

$$\frac{S_n - \mathbf{E}[S_n]}{\sqrt{\text{Var}(S_n)}}$$

converge in legge alla legge $N(0, 1)$.

Il Teorema 5.1.1 ammette diverse estensioni, volte a indebolire le ipotesi di indipendenza o equidistribuzione delle variabili coinvolte: un risultato che si può dimostrare facilmente è il seguente (nel quale non si suppone che le variabili siano equidistribuite, tuttavia si richiede che la successione sia *limitata in L^3*).

Teorema 5.1.3. *Sia $(X_n)_{n \geq 1}$ una successione di v.a.r. indipendenti dotate di momento terzo e supponiamo che si abbia $\mathbf{E}[X_i] = 0$, $\mathbf{E}[X_i^2] = 1$ e $\mathbf{E}[|X_i|^3] \leq K$: allora vale il Teorema Limite Centrale per la successione $(X_n)_{n \geq 1}$.*

Lasciamo per esercizio la dimostrazione che si ottiene ricalcando le linee della dimostrazione di 5.1.1.

Nelle generalizzazioni di cui si diceva sopra, le dimostrazioni diventano usualmente piuttosto lunghe (anche se non veramente difficili).

Questo è il caso ad esempio del seguente notissimo *criterio di Lindeberg*: nell'enunciare questo risultato supponiamo che le variabili siano *centrate* (ci si può sempre ricondurre a questa situazione).

Teorema 5.1.4 (Condizione di Lindeberg). *Siano X_1, X_2, \dots indipendenti con funzione di ripartizione rispettivamente F_1, F_2, \dots , sia $\sigma_k^2 = \text{Var}(X_k)$ e poniamo $D_n^2 = \sigma_1^2 + \dots + \sigma_n^2$. Supponiamo che sia soddisfatta, per ogni $\varepsilon > 0$, la condizione seguente:*

$$\lim_{n \rightarrow \infty} \frac{1}{D_n^2} \cdot \sum_{k=1}^n \int_{\{x: |x| \geq \varepsilon D_n\}} x^2 dF_k(x) = 0$$

Allora vale il Teorema Limite Centrale per la successione $(X_n)_{n \geq 1}$.

Il risultato seguente è molto più preciso: in esso non si tratta solo di *convergenza in legge*, ma viene fornito un risultato di *convergenza uniforme* delle funzioni di ripartizione approssimanti alla funzione di ripartizione limite, con una precisa stima della velocità di convergenza.

Anche del teorema che segue viene tralasciata la dimostrazione, abbastanza complicata. Nell'enunciato di questo risultato si indica come d'uso con $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x \exp(-\frac{t^2}{2}) dt$ la funzione di ripartizione della legge $N(0, 1)$.

Teorema 5.1.5 (Diseguaglianze di Berry–Esseen). *Sia $(X_n)_{n \geq 1}$ una successione di variabili i.i.d. dotate di momento terzo, e sia $G_n(\cdot)$ la funzione di ripartizione di*

$$\frac{S_n - \mathbf{E}[S_n]}{\sqrt{\text{Var}(S_n)}} = \frac{S_n - n \mathbf{E}[X_1]}{\sigma \sqrt{n}}$$

Esiste una costante c per la quale vale la diseguaglianza

$$\sup_{x \in \mathbb{R}} |G_n(x) - \Phi(x)| \leq c \frac{\mathbf{E}[|X_1|^3]}{\sigma^3 \sqrt{n}}$$

5.2 Serie di variabili aleatorie indipendenti

In questa sezione esaminiamo la convergenza quasi certa della somma di una successione di v.a. indipendenti; cominciamo con due importanti diseguaglianze.

Proposizione 5.2.1 (Diseguaglianze di Kolmogorov). *Siano X_1, X_2, \dots centrate indipendenti di quadrato integrabile e sia $c > 0$: si ha*

$$\mathbf{P}\left\{\max_{1 \leq k \leq n} |S_k| \geq c\right\} \leq \frac{\mathbf{E}[S_n^2]}{c^2} \quad (5.2.1)$$

Se inoltre le variabili sono uniformemente limitate, cioè esiste λ tale che $|X_n(\omega)| \leq \lambda$, preso $c > 0$ si ha

$$\mathbf{P}\left\{\max_{1 \leq k \leq n} |S_k| < c\right\} \leq \frac{(c + \lambda)^2}{\mathbf{E}[S_n^2]} \quad (5.2.2)$$

Dimostrazione. Cominciamo ad osservare che, poiché le variabili sono indipendenti e centrate, si ha $\mathbf{E}[S_n^2] = \sum_{k=1}^n \mathbf{E}[X_k^2]$. Chiamiamo poi A_k l'insieme $A_k = \{|S_i| < c \text{ per } i = 1, \dots, k-1; |S_k| \geq c\}$ e notiamo che $A = \{\max_{1 \leq k \leq n} |S_k| \geq c\} = A_1 \cup \dots \cup A_n$ (e questi insiemi sono disgiunti).

Per ogni $k \leq n$ fissato si ha:

$$\mathbf{E}[S_n^2 \cdot I_{A_k}] = \mathbf{E}[S_k^2 \cdot I_{A_k}] + 2 \mathbf{E}[S_k \cdot I_{A_k} \cdot (X_{k+1} + \dots + X_n)] + \mathbf{E}[I_{A_k} (X_{k+1} + \dots + X_n)^2]$$

Notiamo che, poiché $S_k \cdot I_{A_k}$ è indipendente da $(X_{k+1} + \dots + X_n)$, si ha $\mathbf{E}[S_k \cdot I_{A_k} \cdot (X_{k+1} + \dots + X_n)] = \mathbf{E}[S_k \cdot I_{A_k}] \cdot \mathbf{E}[X_{k+1} + \dots + X_n] = 0$, inoltre $\mathbf{E}[I_{A_k} (X_{k+1} + \dots + X_n)^2] = \mathbf{P}(A_k) \cdot \mathbf{E}[(X_{k+1} + \dots + X_n)^2]$.

Si ottiene pertanto la diseguaglianza

$$\mathbf{E}[S_k^2 \cdot I_{A_k}] \leq \mathbf{E}[S_n^2 \cdot I_{A_k}] \leq \mathbf{E}[S_k^2 \cdot I_{A_k}] + \mathbf{P}(A_k) \cdot \mathbf{E}[S_n^2]$$

a questo punto

$$\mathbf{E}[S_n^2] \geq \sum_{k=1}^n \mathbf{E}[S_n^2 \cdot I_{A_k}] \geq \sum_{k=1}^n \mathbf{E}[S_k^2 \cdot I_{A_k}] \geq \sum_{k=1}^n c^2 \mathbf{P}(A_k) = c^2 \mathbf{P}(A)$$

cioè la 5.2.1. Per quanto riguarda la 5.2.2, osserviamo che, se $|X_n(\omega)| \leq \lambda$, sull'insieme A_k si ha $|S_k(\omega)| \leq (\lambda + c)$ e su $A^c = \{\max_{1 \leq k \leq n} |S_k| < c\}$, si ha $|S_n(\omega)| \leq c \leq (\lambda + c)$. Quindi

$$\begin{aligned} \mathbf{E}[S_n^2] &= \sum_{k=1}^n \mathbf{E}[S_n^2 \cdot I_{A_k}] + \mathbf{E}[S_n^2 \cdot I_{A^c}] \leq \\ &\leq \sum_{k=1}^n \left(\mathbf{E}[S_k^2 \cdot I_{A_k}] + \mathbf{P}(A_k) \cdot \mathbf{E}[S_n^2] \right) + \mathbf{P}(A^c) (\lambda + c)^2 \leq \end{aligned}$$

$$\leq \sum_{k=1}^n \left((\lambda + c)^2 \mathbf{P}(A_k) + \mathbf{P}(A_k) \cdot \mathbf{E}[S_n^2] \right) + \mathbf{P}(A^c)(\lambda + c)^2$$

e da questa si ottiene $\mathbf{E}[S_n^2] \cdot \mathbf{P}(A^c) \leq (\lambda + c)^2$.

□

Queste disuguaglianze sono lo strumento essenziale per provare il seguente

Teorema 5.2.2. *Sia X_1, X_2, \dots una successione di v.a.r. indipendenti centrate e supponiamo che si abbia $\sum_{n=1}^{+\infty} \mathbf{E}[X_n^2] < +\infty$: allora la serie $\sum_{n=1}^{+\infty} X_n$ converge q.c.*

Inoltre se le variabili sono uniformemente limitate e la serie converge q.c., allora $\sum_{n=1}^{+\infty} \mathbf{E}[X_n^2] < +\infty$.

Dimostrazione. Per quanto riguarda la prima parte, notiamo che poiché $\mathbf{E}[S_n^2] = \mathbf{E}[X_1^2] + \dots + \mathbf{E}[X_n^2]$, S_n converge in L^2 e pertanto in probabilità.

Dati $\varepsilon > 0$ e $\delta > 0$, si può determinare $\bar{n} = \bar{n}(\varepsilon, \delta)$ tale che, qualunque sia $m > \bar{n}$, si abbia $\mathbf{P}\left\{ \max_{\bar{n} \leq k \leq m} |S_k - S_{\bar{n}}| \geq \varepsilon \right\} < \delta$. Infatti, per la disuguaglianza 5.2.1, si ha

$$\mathbf{P}\left\{ \max_{\bar{n} \leq k \leq m} |S_k - S_{\bar{n}}| \geq \varepsilon \right\} \leq \frac{\mathbf{E}[X_{\bar{n}+1}^2] + \dots + \mathbf{E}[X_m^2]}{\varepsilon^2} \leq \frac{\sum_{k > \bar{n}} \mathbf{E}[X_k^2]}{\varepsilon^2}$$

Allora si può determinare una sottosuccessione $n_1 < n_2 < \dots$ tale che S_{n_k} converga q.c. e che si abbia

$$\mathbf{P}\left\{ \max_{n_k < j < n_{k+1}} |S_j - S_{n_k}| > \frac{1}{k} \right\} \leq 2^{-k}$$

Chiamando $D_k = \max_{n_k < j < n_{k+1}} |S_j - S_{n_k}|$, il Lemma di Borel–Cantelli mostra che $D_k \xrightarrow{\text{q.c.}} 0$.

Scrivendo, per $n_k \leq n < n_{k+1}$,

$$S_n = S_{n_k} + (S_n - S_{n_k})$$

è facile concludere che la successione $(S_n)_{n \geq 1}$ converge q.c.

La seconda parte segue facilmente dalla disuguaglianza 5.2.2: infatti se la serie $\sum_{n=1}^{+\infty} X_n$ converge q.c., la v.a. $\sup_n |S_n|$ è a valori finiti e pertanto esiste $c > 0$ con $\mathbf{P}\left\{ \sup_n |S_n| < c \right\} > 0$.

Si ha pertanto, per ogni n ,

$$\mathbf{P}\left\{\sup_n |S_n| < c\right\} \leq \mathbf{P}\left\{\max_{1 \leq k \leq n} |S_k| < c\right\} \leq \frac{(c + \lambda)^2}{\sum_{k=1}^n \mathbf{E}[X_k^2]}$$

e ne segue $\sum_{n=1}^{+\infty} \mathbf{E}[X_n^2] < +\infty$. \square

Vediamo ora un risultato che riguarda serie di v.a. indipendenti non necessariamente centrate.

Proposizione 5.2.3. *Sia $(X_n)_{n \geq 1}$ una successione di v.a. indipendenti uniformemente limitate: condizione necessaria e sufficiente affinché la serie $\sum_{n=1}^{+\infty} X_n$ converga q.c. è che convergano le due serie*

$$\sum_{n=1}^{+\infty} \mathbf{E}[X_n] \quad \sum_{n=1}^{+\infty} \text{Var}(X_n)$$

Dimostrazione. Se convergono le due serie, il risultato è facile. Poniamo infatti $\tilde{X}_n = X_n - \mathbf{E}[X_n]$: poiché $\mathbf{E}[\tilde{X}_n^2] = \text{Var}(X_n)$, per 5.2.2 converge q.c. $\sum_{n=1}^{+\infty} \tilde{X}_n$.

Supponiamo viceversa che la serie $\sum_{n=1}^{+\infty} X_n$ converga q.c. e consideriamo una successione di v.a. $(Y_n)_{n \geq 1}$ con la stessa legge delle (X_n) ed indipendenti dalle prime: anche la serie $\sum_{n=1}^{+\infty} Y_n$ converge q.c. (la convergenza dipende dalla legge della successione di v.a.) e di conseguenza converge q.c. anche la successione delle v.a. $(X_n - Y_n)$ che sono centrate.

Poiché $\text{Var}(X_n - Y_n) = 2 \text{Var}(X_n)$ segue $\sum_{n=1}^{+\infty} \text{Var}(X_n) < +\infty$, quindi converge q.c. la serie $\sum_{n=1}^{+\infty} (X_n - \mathbf{E}[X_n])$, e, poiché la serie $\sum_{n=1}^{+\infty} X_n$ converge per ipotesi, per differenza converge anche $\sum_{n=1}^{+\infty} \mathbf{E}[X_n]$. \square

Osservazione 5.2.4. Nella dimostrazione precedente viene data per scontata l'esistenza della successione di v.a. $(Y_n)_{n \geq 1}$: l'esistenza di questa successione non è difficile da provare, ma non è evidente. Dall'appendice del capitolo 2, oppure più avanti dal capitolo 6, si vede che è possibile costruire su un opportuno spazio di probabilità una successione di v.a. indipendenti aventi leggi prefissate; non è però detto che questo si possa fare sullo stesso spazio sul quale sono definite le v.a. $(X_n)_{n \geq 1}$ e che risultino indipendenti.

Si può allora ricorrere a uno *spazio prodotto*: più precisamente, dato $(\Omega, \mathcal{F}, \mathbf{P})$ sul quale sono definite le $(X_n)_{n \geq 1}$, si prende un nuovo spazio $(\Omega', \mathcal{F}', \mathbf{P}')$ sul quale sono costruite le variabili $(Y_n)_{n \geq 1}$ e si considera infine $\Omega \times \Omega'$ con la σ -algebra $\mathcal{F} \otimes \mathcal{F}'$ e la probabilità prodotto $\mathbf{P} \otimes \mathbf{P}'$. In questo modo si ottiene il risultato voluto.

Siamo ora in grado di provare facilmente il celebre risultato seguente:

Teorema 5.2.5 (Teorema delle tre serie di Kolmogorov). Sia $(X_n)_{n \geq 1}$ una successione di v.a. indipendenti, prendiamo $c > 0$ e sia $Y_n = X_n \cdot I_{\{|X_n| < c\}}$: condizione necessaria e sufficiente affinché converga q.c. la serie $\sum_{n=1}^{+\infty} X_n$ è che convergano le tre serie

$$\sum_{n=1}^{+\infty} \mathbf{E}[Y_n] \quad \sum_{n=1}^{+\infty} \text{Var}(Y_n) \quad \sum_{n=1}^{+\infty} \mathbf{P}\{X_n \neq Y_n\}$$

Dimostrazione. Notiamo che, grazie all'indipendenza delle v.a., la condizione

$$\sum_{n=1}^{+\infty} \mathbf{P}\{X_n \neq Y_n\} = \sum_{n=1}^{+\infty} \mathbf{P}\{|X_n| \geq c\} < +\infty$$

equivale a dire che, per quasi ogni ω , esiste $\bar{n}(\omega)$ tale che, se $n \geq \bar{n}(\omega)$, si abbia $|X_n(\omega)| \leq c$.

Essendo le variabili $(Y_n)_{n \geq 1}$ uniformemente limitate, la Proposizione 5.2.3 permette facilmente di concludere

□

Osservazione 5.2.6. È interessante sapere che le diseguaglianze che hanno portato alla dimostrazione della Proposizione 5.2.1 sono state modificate da Paul Lévy per ottenere il seguente risultato: *una serie di variabili aleatorie indipendenti converge in probabilità se e solo se converge q.c.* La dimostrazione di questo classico teorema è riportata in appendice.

5.3 Leggi dei Grandi Numeri.

Nel primo corso è già stata vista una prima versione della *Legge dei Grandi Numeri*: se X_1, X_2, \dots è una successione di v.a. indipendenti, equidistribuite, dotate di momento secondo e con $\mathbf{E}[X_i] = \mu$, posto $S_n = X_1 + \dots + X_n$, allora $\frac{S_n}{n}$ converge a μ in probabilità, o (ciò che è lo stesso) $\frac{S_n - n \cdot \mu}{n}$ converge a 0 in probabilità. Notiamo che in questo caso $n \cdot \mu = n \cdot \mathbf{E}[X_1] = \mathbf{E}[S_n]$: questo ha dato origine alla seguente

Definizione 5.3.1 (Legge dei Grandi Numeri). Sia $(X_n)_{n \geq 1}$ una successione di v.a.r. dotate di momento primo, e sia $S_n = X_1 + \dots + X_n$: si dice che vale la **legge debole dei Grandi Numeri** (rispettivamente **legge forte dei Grandi numeri**) se

$$\frac{S_n - \mathbf{E}[S_n]}{n} \longrightarrow 0$$

in probabilità (rispettivamente **quasi certamente**).

Per semplificare le notazioni, non è restrittivo supporre che le variabili $(X_n)_{n \geq 1}$ siano *centrate*. Premettiamo alcuni risultati preparatori:

Lemma 5.3.2 (Lemma di Kronecker). *Sia y_1, y_2, \dots una successione di numeri e supponiamo che la serie $\sum_{n=1}^{+\infty} \frac{y_n}{n}$ converga: allora $\lim_{n \rightarrow \infty} \frac{y_1 + \dots + y_n}{n} = 0$.*

Dimostrazione. Non si richiede che la serie indicata *converga assolutamente*, ma solo che esista il limite di $r_n = y_1 + \frac{y_2}{2} + \dots + \frac{y_n}{n}$: notiamo che $r_j - r_{j-1} = \frac{y_j}{j}$ e quindi $y_1 + \dots + y_n = r_1 + 2(r_2 - r_1) + \dots + n(r_n - r_{n-1})$.

Si ottiene allora $\frac{y_1 + \dots + y_n}{n} = r_n - \frac{r_1 + \dots + r_{n-1}}{n}$, e questa successione converge evidentemente a 0. \square

Possiamo ora facilmente provare il risultato seguente:

Teorema 5.3.3 (Legge Forte dei Grandi Numeri). *Supponiamo che le variabili X_1, X_2, \dots siano indipendenti e tali che $\sum_{n=1}^{+\infty} \frac{\text{Var}(X_n)}{n^2} < +\infty$: allora vale la legge Forte dei Grandi Numeri.*

Dimostrazione. Possiamo supporre che le variabili siano centrate: per il Teorema 5.2.2, la serie $\sum_{n \geq 1} \frac{X_n}{n}$ converge q.c., e di conseguenza, per il Lemma di Kronecker 5.3.2, $\frac{S_n}{n} \xrightarrow{\text{q.c.}} 0$. \square

Nel caso in cui le variabili, oltre ad essere indipendenti, siano anche *equidistribuite*, è stata provata da Kolmogorov una Legge dei Grandi Numeri per variabili dotate solo di momento primo. Premettiamo due disequaglianze.

Lemma 5.3.4. *Sia X una v.a.r. : valgono le disequaglianze*

$$\sum_{n=1}^{+\infty} \mathbf{P}\{|X| \geq n\} \leq \mathbf{E}[|X|] \leq 1 + \sum_{n=1}^{+\infty} \mathbf{P}\{|X| \geq n\} \quad (5.3.1)$$

$$\sum_{n=1}^{+\infty} \frac{1}{n^2} \mathbf{E}[X^2 \cdot I_{\{|X| < n\}}] \leq 2 \mathbf{E}[|X|] \quad (5.3.2)$$

Dimostrazione. Per quanto riguarda la prima diseguaglianza, utilizzando il Teorema di Fubini-Tonelli 2.3.2, si ha

$$\begin{aligned}\mathbf{E}[|X|] &= \int_{\Omega} |X(\omega)| d\mathbf{P}(\omega) = \int_{\Omega} d\mathbf{P}(\omega) \int_0^{|X(\omega)|} dx = \\ &= \int_0^{+\infty} \mathbf{P}\{|X| \geq x\} dx = \sum_{h=0}^{+\infty} \int_h^{h+1} \mathbf{P}\{|X| \geq x\} dx\end{aligned}$$

Poichè la funzione $x \rightarrow \mathbf{P}\{|X| \geq x\}$ è decrescente, si ha per ogni h

$$\mathbf{P}\{|X| \geq h+1\} \leq \int_h^{h+1} \mathbf{P}\{|X| \geq x\} dx \leq \mathbf{P}\{|X| \geq h\}$$

e sommando i vari termini si ottiene il risultato.

Per quanto riguarda la seconda diseguaglianza, si ha

$$\begin{aligned}\sum_{n=1}^{+\infty} \frac{1}{n^2} \mathbf{E}[X^2 \cdot I_{\{|X| < n\}}] &= \sum_{n=1}^{+\infty} \sum_{k \leq n} \frac{1}{n^2} \mathbf{E}[X^2 \cdot I_{\{k-1 \leq |X| < k\}}] = \\ &= \sum_{k=1}^{+\infty} \mathbf{E}[X^2 \cdot I_{\{k-1 \leq |X| < k\}}] \sum_{n=k}^{+\infty} \frac{1}{n^2} \leq 2 \sum_{k=1}^{+\infty} \frac{1}{k} \mathbf{E}[X^2 \cdot I_{\{k-1 \leq |X| < k\}}] \\ &\leq 2 \sum_{k=1}^{+\infty} \mathbf{E}[|X| \cdot I_{\{k-1 \leq |X| < k\}}] = 2 \mathbf{E}[|X|]\end{aligned}$$

□

Possiamo ora provare il risultato enunciato.

Teorema 5.3.5 (Legge Forte di Kolmogorov). *Sia X_1, X_2, \dots una successione di v.a.r. indipendenti ed equidistribuite, dotate di momento primo: allora vale la Legge Forte dei Grandi Numeri.*

Dimostrazione. Anche questa volta possiamo supporre che le variabili X_1, X_2, \dots

siano centrate. Tenendo conto della diseguaglianza 5.3.1, $\sum_{n=1}^{+\infty} \mathbf{P}\{|X_n| \geq n\} =$

$= \sum_{n=1}^{+\infty} \mathbf{P}\{|X_1| \geq n\} < +\infty$, e quindi (ricordando il Lemma di Borel–Cantelli),

per quasi ogni ω fissato, definitivamente $|X_n(\omega)| < n$.

Definiamo allora

$$\tilde{X}_n(\omega) = \begin{cases} X_n(\omega) & \text{se } |X_n(\omega)| < n \\ 0 & \text{se } |X_n(\omega)| \geq n \end{cases}$$

Le variabili \tilde{X}_n non sono necessariamente centrate, però $\mathbf{E}[\tilde{X}_n] = \mathbf{E}[X_n \cdot I_{\{|X_n| < n\}}] = \mathbf{E}[X_1 \cdot I_{\{|X_1| < n\}}] \rightarrow 0$.

Poichè

$$\frac{X_1 + \cdots + X_n}{n} = \frac{\tilde{X}_1 + \cdots + \tilde{X}_n}{n} + \text{“qualcosa di infinitesimo”}$$

basta provare la *legge dei Grandi Numeri* per le variabili \tilde{X}_n . Cominciamo a provare che si ha

$$\sum_{n=1}^{+\infty} \frac{\text{Var}(\tilde{X}_n)}{n^2} \leq \sum_{n=1}^{+\infty} \frac{\mathbf{E}[\tilde{X}_n^2]}{n^2} < +\infty$$

Infatti, utilizzando la disuguaglianza 5.3.2

$$\begin{aligned} \sum_{n=1}^{+\infty} \frac{\mathbf{E}[\tilde{X}_n^2]}{n^2} &= \sum_{n=1}^{+\infty} \frac{1}{n^2} \mathbf{E}[X_n^2 \cdot I_{\{|X_n| < n\}}] = \sum_{n=1}^{+\infty} \frac{1}{n^2} \mathbf{E}[X_1^2 \cdot I_{\{|X_1| < n\}}] \leq \\ &\leq 2 \mathbf{E}[|X_1|] < +\infty \end{aligned}$$

Ne segue che $\sum_{n \geq 1} \frac{\tilde{X}_n - \mathbf{E}[\tilde{X}_n]}{n}$ converge q.c., quindi

$\frac{(\tilde{X}_1 - \mathbf{E}[\tilde{X}_1]) + \cdots + (\tilde{X}_n - \mathbf{E}[\tilde{X}_n])}{n}$ converge q.c. a 0 e di conseguenza

$\frac{\tilde{X}_1 + \cdots + \tilde{X}_n}{n}$ converge q.c. a 0.

□

Osservazione 5.3.6. È interessante osservare che il risultato precedente può essere in un certo senso ribaltato, cioè se una successione $(X_n)_{n \geq 1}$ di v.a. indipendenti equidistribuite è tale che $\frac{X_1 + \cdots + X_n}{n}$ converge q.c. a una costante μ , necessariamente le variabili hanno momento primo.

La dimostrazione può essere fatta ad esempio provando che, se $\sum_{n=1}^{+\infty} \mathbf{P}\{|X| \geq n\} = +\infty$, la convergenza sopra scritta non può essere soddisfatta.

Vediamo una interessante applicazione del risultato precedente: sappiamo che ogni numero $x \in [0, 1[$ si può scrivere con rappresentazione diadica $x = 0, a_1 a_2 \dots$ (dove a_i è eguale a 0 oppure a 1) ed il numero è detto **normale** se $\lim_{n \rightarrow \infty} \frac{a_1 + \dots + a_n}{n} = \frac{1}{2}$.

Proposizione 5.3.7. *Quasi ogni numero dell'intervallo $[0, 1[$ è normale.*

Dimostrazione. Il “quasi” si riferisce alla misura di Lebesgue: consideriamo dunque sullo spazio $\Omega = [0, 1]$ munito della σ -algebra di Borel e della misura di Lebesgue la successione di v.a. X_1, X_2, \dots dove $X_i(\omega)$ è la i -ma cifra della rappresentazione diadica. Non è difficile verificare che le variabili sono indipendenti, equidistribuite, con valore atteso $1/2$: l'affermazione è dunque una conseguenza immediata del Teorema 5.3.5. \square

5.4 Appendice

In questa appendice verrà dimostrato il risultato annunciato nell'Osservazione 5.2.6. In particolare il risultato che segue sostituisce in un certo senso la disuguaglianza 5.2.1.

Lemma 5.4.1. *Sia $(X_n)_{n \geq 1}$ una successione di v.a.r. indipendenti, e supponiamo che esistano $\varepsilon > 0$, $\eta > 0$, $m \in \mathbb{N}$ tali che, preso $1 \leq n < m$ si abbia*

$$\mathbf{P}\{|X_{n+1} + \dots + X_m| > \varepsilon/2\} \leq \eta$$

Allora vale la disuguaglianza

$$\mathbf{P}\left\{\max_{1 \leq k \leq m} |S_k| > \varepsilon\right\} \leq \frac{\eta}{1 - \eta}$$

Dimostrazione. Con le stesse notazioni della Proposizione 5.2.1, chiamiamo A_k l'insieme $A_k = \{|S_i| < \varepsilon \text{ per } i = 1, \dots, k-1; |S_k| \geq \varepsilon\}$ e notiamo che $A = \{\max_{1 \leq k \leq m} |S_k| \geq \varepsilon\} = A_1 \cup \dots \cup A_m$.

Notiamo che

$$A_k \cap \{|X_{k+1} + \dots + X_m| \leq \varepsilon/2\} \subseteq A_k \cap \{|S_m| > \varepsilon/2\}$$

e poiché A_k e $\{|X_{k+1} + \dots + X_m| \leq \varepsilon/2\}$ sono indipendenti, si ha

$$(1 - \eta) \cdot \mathbf{P}(A_k) \leq \mathbf{P}(A_k \cap \{|S_m| > \varepsilon/2\})$$

e sommando

$$(1 - \eta) \cdot \mathbf{P}\left\{\max_{1 \leq k \leq m} |S_k| > \varepsilon\right\} \leq \mathbf{P}\{|S_m| > \varepsilon/2\} \leq \eta$$

\square

Teorema 5.4.2 (Paul Lévy). *Sia $(X_n)_{n \geq 1}$ una successione di v.a.r. indipendenti: se la serie $\sum_{n=1}^{+\infty} X_n$ converge in probabilità, converge anche q.c.*

Dimostrazione. Come nel teorema 5.2.2, è sufficiente provare che, dati $\varepsilon > 0$ e $\delta > 0$, si può determinare $\bar{n} = \bar{n}(\varepsilon, \delta)$ tale che, qualunque sia $m > \bar{n}$, si abbia $\mathbf{P}\left\{\max_{\bar{n} \leq k \leq m} |S_k - S_{\bar{n}}| \geq \varepsilon\right\} < \delta$.

Dato $\varepsilon > 0$, si sceglie infatti $\eta > 0$ tale che $\eta/(1 - \eta) < \delta$, e (poiché la successione $(S_n)_{n \geq 1}$ converge in probabilità), è possibile determinare \bar{n} tale che, presi comunque $\bar{n} \leq n < m$, si abbia

$$\mathbf{P}\{|S_m - S_n| > \varepsilon/2\} = \mathbf{P}\{|X_{n+1} + \cdots + X_m| > \varepsilon/2\} \leq \eta$$

Allora il Lemma 5.4.1 garantisce la disuguaglianza sopra indicata.

A questo punto la dimostrazione prosegue esattamente come in 5.2.2. \square

Capitolo 6

Speranza condizionale e alcuni complementi.

6.1 La speranza condizionale.

Consideriamo uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$ e sia $\mathcal{E} \subseteq \mathcal{F}$.

Proposizione 6.1.1. *Data $X \in L^1(\Omega, \mathcal{F}, \mathbf{P})$, esiste (unica a meno di equivalenza) una v.a. Y \mathcal{E} -misurabile tale che si abbia, per ogni $A \in \mathcal{E}$, $\int_A X \, d\mathbf{P} = \int_A Y \, d\mathbf{P}$. Questa Y è chiamata la **speranza condizionale** di X rispetto alla σ -algebra \mathcal{E} ed è indicata $Y = \mathbf{E}[X|\mathcal{E}]$.*

Dimostrazione. La dimostrazione è una semplice conseguenza del Teorema di Radon–Nikodym.

Supponiamo X a valori positivi e consideriamo la misura che ha densità X rispetto a \mathbf{P} ristretta alla σ -algebra \mathcal{E} (cioè $\mu(A) = \int_A X \, d\mathbf{P}$, $A \in \mathcal{E}$): questa ha rispetto a \mathbf{P} una densità Y che è \mathcal{E} -misurabile e soddisfa le proprietà richieste.

Si considera poi come d'uso la decomposizione $X = X^+ - X^-$.

□

Si deve fare attenzione al fatto che $\mathbf{E}[X|\mathcal{E}]$ è in realtà una *classe d'equivalenza* di variabili aleatorie: quando si scrive $Y = \mathbf{E}[X|\mathcal{E}]$ si intende che la v.a. Y è un rappresentante della classe di equivalenza $\mathbf{E}[X|\mathcal{E}]$.

Proposizione 6.1.2. *La speranza condizionale gode delle seguenti proprietà:*

- a) $\mathbf{E}[X] = \mathbf{E}[\mathbf{E}[X|\mathcal{E}]]$
- b) $\mathbf{E}[aX + Y|\mathcal{E}] = a\mathbf{E}[X|\mathcal{E}] + \mathbf{E}[Y|\mathcal{E}]$
- c) se $X \leq Y$ q.c., allora $\mathbf{E}[X|\mathcal{E}] \leq \mathbf{E}[Y|\mathcal{E}]$ q.c.

- d) se X è \mathcal{E} -misurabile, $\mathbf{E}[X|\mathcal{E}] = X$
 e) se X è indipendente da \mathcal{E} , $\mathbf{E}[X|\mathcal{E}] = \mathbf{E}[X]$
 f) se $\mathcal{G} \subseteq \mathcal{E} \subseteq \mathcal{F}$, $\mathbf{E}[X|\mathcal{E}|\mathcal{G}] = \mathbf{E}[X|\mathcal{G}]$
 g) se Y è \mathcal{E} -misurabile e limitata, $\mathbf{E}[XY|\mathcal{E}] = Y \mathbf{E}[X|\mathcal{E}]$
 h) se $X_n \uparrow X$ q.c., allora $\mathbf{E}[X_n|\mathcal{E}] \uparrow \mathbf{E}[X|\mathcal{E}]$ q.c.

Dimostrazione. Le proprietà a), b), c), d), f) sono immediate. Vediamo e): supponiamo X indipendente da \mathcal{E} e sia $A \in \mathcal{E}$.

$$\int_A X \, d\mathbf{P} = \mathbf{E}[X.I_A] = \mathbf{E}[X] \cdot \mathbf{E}[I_A] = \int_A \mathbf{E}[X] \, d\mathbf{P}$$

Per provare g), innanzi tutto osserviamo che si suppone Y limitata perchè XY deve essere integrabile; poi cominciamo a provare l'eguaglianza se $Y = I_B$ con $B \in \mathcal{E}$. Presa $A \in \mathcal{E}$:

$$\int_A I_B \cdot X \, d\mathbf{P} = \int_{A \cap B} X \, d\mathbf{P} = \int_{A \cap B} \mathbf{E}[X|\mathcal{E}] \, d\mathbf{P} = \int_A I_B \cdot \mathbf{E}[X|\mathcal{E}] \, d\mathbf{P}$$

Di conseguenza l'eguaglianza è soddisfatta se Y è una v.a. semplice; per estenderla a tutte le v.a. Y \mathcal{E} -misurabili limitate, si può utilizzare il fatto che ogni variabile aleatoria limitata è limite *per la convergenza uniforme* di una successione di v.a. semplici (vedi Osservazione 1.2.11), oppure utilizzare la proprietà *tipo Beppo-Levi* vista al successivo punto h).

Vediamo ora la proprietà h): notiamo che (per il punto c)) la successione di v.a. $(\mathbf{E}[X_n|\mathcal{E}])_{n \geq 1}$ è crescente ad una certa v.a. Y \mathcal{E} -misurabile che (come conseguenza del Lemma di Beppo-Levi) verifica, per ogni $A \in \mathcal{E}$, l'eguaglianza $\int_A X \, d\mathbf{P} = \int_A Y \, d\mathbf{P}$.

Questa ultima proprietà permette di definire, per una v.a. X a valori positivi, $\mathbf{E}[X|\mathcal{E}] = \lim_{n \rightarrow \infty} \mathbf{E}[X_n|\mathcal{E}]$ dove $(X_n)_{n \geq 1}$ è una successione di v.a. positive limitate tali che $X_n \uparrow X$: in base a quanto si è appena visto, il limite non dipende dalla particolare successione scelta. \square

Proposizione 6.1.3 (Diseguaglianza di Jensen). *Sia $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ una funzione convessa e supponiamo che X e $\varphi(X)$ siano integrabili: allora*

$$\varphi(\mathbf{E}[X|\mathcal{E}]) \leq \mathbf{E}[\varphi(X)|\mathcal{E}] \quad \text{q.c.}$$

Dimostrazione. La dimostrazione è sostanzialmente identica alla dimostrazione che è stata fatta per la diseguaglianza di Jensen *non condizionale* (vedi

Lemma 1.2.9): si parte dall'eguaglianza $\varphi(x) = \sup_n L_n(x)$, dove $L_n(x) = a_n x + b_n$ è una funzione lineare affine.

Per ogni n fissato si ha

$$L_n(\mathbf{E}[X|\mathcal{E}]) = \mathbf{E}[L_n(X)|\mathcal{E}] \leq \mathbf{E}[\varphi(X)|\mathcal{E}] \quad \text{q.c.}$$

e, prendendo a sinistra l'estremo superiore al variare di n , si ottiene la diseguaglianza. \square

Corollario 6.1.4. *Se $X \in L^p(\Omega, \mathcal{F}, \mathbf{P})$, allora $\mathbf{E}[X|\mathcal{E}] \in L^p(\Omega, \mathcal{E}, \mathbf{P})$.*

Dimostrazione. Per $1 \leq p < +\infty$ il risultato si ottiene considerando la funzione convessa $\varphi(x) = |x|^p$; per $p = +\infty$, il risultato è evidente. \square

Osservazione 6.1.5. Nella proprietà *g)* della Proposizione 6.1.2 non è necessario supporre che la v.a. Y sia *limitata*: si può ad esempio supporre che $X \in L^p$ e $Y \in L^q$ con p e q esponenti *coniugati* (cioè $1/p + 1/q = 1$).

La proprietà che segue, la cui dimostrazione è lasciata per esercizio, verrà utilizzata più volte.

Proposizione 6.1.6. *Siano X, Y due variabili aleatorie di quadrato integrabile (oppure con $X \in L^p$ e $Y \in L^q$ con $1/p + 1/q = 1$): vale la formula*

$$\mathbf{E}[X \mathbf{E}[Y|\mathcal{E}]] = \mathbf{E}[\mathbf{E}[X|\mathcal{E}] Y]$$

Proposizione 6.1.7. *Se $X \in L^2(\Omega, \mathcal{F}, \mathbf{P})$, $\mathbf{E}[X|\mathcal{E}]$ coincide con la proiezione ortogonale di X sul sottospazio chiuso $L^2(\Omega, \mathcal{E}, \mathbf{P})$.*

Dimostrazione. Cominciamo a provare che $L^2(\Omega, \mathcal{E}, \mathbf{P})$ è un sottospazio chiuso: se Y_n sono \mathcal{E} -misurabili e convergono a Y in L^2 , vi convergono anche in probabilità ed una sottosuccessione converge q.c. La convergenza q.c. conserva la misurabilità e quindi Y è \mathcal{E} -misurabile.

Dobbiamo ora provare che, preso $X \in L^2(\Omega, \mathcal{F}, \mathbf{P})$ e $Y \in L^2(\Omega, \mathcal{E}, \mathbf{P})$, il prodotto scalare (in L^2) di X con Y coincide col prodotto scalare di $\mathbf{E}[X|\mathcal{E}]$ con Y . Infatti

$$\mathbf{E}[X.Y] = \mathbf{E}[\mathbf{E}[XY|\mathcal{E}]] = \mathbf{E}[Y \mathbf{E}[X|\mathcal{E}]]$$

\square

Osservazione 6.1.8. La Proposizione 6.1.7 fornisce una *definizione alternativa* di speranza condizionale: per $X \in L^2(\Omega, \mathcal{F}, \mathbf{P})$, $\mathbf{E}[X|\mathcal{E}]$ è la proiezione ortogonale di X sul sottospazio chiuso $L^2(\Omega, \mathcal{E}, \mathbf{P})$ (ma questo la definisce solo per le v.a. di quadrato integrabile). Si prova quindi la *diseguaglianza di Jensen* e questo permette di provare la diseguaglianza $\mathbf{E}[|\mathbf{E}[X|\mathcal{E}]|] \leq \mathbf{E}[|X|]$ e quindi di estendere la speranza condizionale a tutte le variabili integrabili.

La formula seguente è talvolta chiamata (impropriamente) **formula di Bayes**:

Teorema 6.1.9. *Sia $\mathbf{Q} \ll \mathbf{P}$, sia $Z = \frac{d\mathbf{Q}}{d\mathbf{P}}$ una versione della densità e sia $X \in L^1(\Omega, \mathcal{F}, \mathbf{Q})$: vale la formula*

$$\mathbf{E}^{\mathbf{Q}}[X|\mathcal{E}] = \frac{\mathbf{E}^{\mathbf{P}}[XZ|\mathcal{E}]}{\mathbf{E}^{\mathbf{P}}[Z|\mathcal{E}]}$$

Dimostrazione. Innanzi tutto bisogna provare che l'insieme sul quale si annulla il denominatore è \mathbf{Q} -trascurabile: sia $A = \{\mathbf{E}^{\mathbf{P}}[Z|\mathcal{E}] = 0\}$.

$$\mathbf{Q}(A) = \int_A Z d\mathbf{P} = \int_A \mathbf{E}^{\mathbf{P}}[Z|\mathcal{E}] d\mathbf{P} = 0$$

Per provare la formula, proviamo che vale (q.c.) l'eguaglianza $\mathbf{E}^{\mathbf{P}}[XZ|\mathcal{E}] = \mathbf{E}^{\mathbf{Q}}[X|\mathcal{E}]\mathbf{E}^{\mathbf{P}}[Z|\mathcal{E}]$. Prendiamo $A \in \mathcal{E}$:

$$\begin{aligned} \int_A \mathbf{E}^{\mathbf{P}}[XZ|\mathcal{E}] d\mathbf{P} &= \int_A XZ d\mathbf{P} = \int_A X d\mathbf{Q} = \int_A \mathbf{E}^{\mathbf{Q}}[X|\mathcal{E}] d\mathbf{Q} = \\ &= \int_A \mathbf{E}^{\mathbf{Q}}[X|\mathcal{E}] Z d\mathbf{P} = \int_A \mathbf{E}^{\mathbf{Q}}[X|\mathcal{E}]\mathbf{E}^{\mathbf{P}}[Z|\mathcal{E}] d\mathbf{P} \end{aligned}$$

Poichè questo è vero per ogni $A \in \mathcal{E}$ la conseguenza è immediata. □

6.2 Esempi concreti di calcolo di speranza condizionale

I risultati esposti nel paragrafo precedente hanno una dimostrazione non difficile e formano un quadro teorico completo, tuttavia di fronte ad un esercizio concreto solitamente il calcolo della speranza condizionale risulta misterioso: lo scopo di questo paragrafo (i cui enunciati andrebbero piuttosto visti come esercizi) è proprio illustrare alcuni esempi. Quando non compare la dimostrazione vuol dire che in realtà è elementare ed è lasciata come esercizio.

Il facile esempio che segue in realtà è già stato incontrato nel corso di E.P.S.

Esempio 6.2.1. Supponiamo che la σ -algebra \mathcal{E} sia generata da una partizione numerabile A_1, A_2, \dots di insiemi non-trascurabili: una v.a. Y è \mathcal{E} -misurabile se e solo se è costante su ogni insieme A_i . Inoltre, presa X integrabile, $\mathbf{E}[X|\mathcal{E}]$ sull'insieme A_i è eguale a

$$\frac{1}{\mathbf{P}(A_i)} \int_{A_i} X d\mathbf{P}$$

L'esempio che segue sarà molto importante nel seguito:

Proposizione 6.2.2. *Sia \mathcal{G} una σ -algebra contenuta in \mathcal{F} , siano X e Y due v.a.r. e supponiamo che X sia indipendente da \mathcal{G} ed Y sia \mathcal{G} -misurabile: data $\varphi: \mathbb{R}^2 \rightarrow \mathbb{R}$ boreliana limitata, poniamo $G(y) = \mathbf{E}[\varphi(X, y)]$. L'applicazione $y \rightarrow G(y)$ è boreliana e vale la formula*

$$\mathbf{E}[\varphi(X, Y)|\mathcal{G}] = G(Y)$$

Cenno di dimostrazione: è immediato provare l'asserto se $\varphi(x, y) = I_A(x) \cdot I_B(y)$ con $A, B \in \mathcal{B}(\mathbb{R})$ e col Teorema delle Classi monotone si può estendere a $\varphi(x, y) = I_C(x, y)$ dove C è un boreliano di \mathbb{R}^2 . Di conseguenza l'eguaglianza è soddisfatta se φ è *semplice* e si passa quindi alle funzioni boreliane limitate nel modo usuale.

Osservazione 6.2.3. Se T è una variabile aleatoria (a valori in un generico spazio (E, \mathcal{E})) $\mathbf{E}[X|T]$ è una *notazione* che indica $\mathbf{E}[X|\sigma(T)]$: per il criterio di misurabilità di Doob, si può scrivere $\mathbf{E}[X|T] = g(T)$ con una opportuna $g: E \rightarrow \mathbb{R}$ misurabile.

Talvolta per la funzione $g(t)$ si usa la notazione (impropria) $\mathbf{E}[X|T = t]$.

Esempio 6.2.4. Supponiamo che la v.a. T sia *discreta*, chiamiamo t_1, t_2, \dots i suoi valori e supponiamo che gli insiemi $\{T = t_i\}$ siano non trascurabili: presa φ boreliana limitata, vogliamo calcolare la funzione g tale che si abbia $\mathbf{E}[\varphi(X)|T] = g(T)$. Notiamo che è sufficiente definire g nei punti t_1, t_2, \dots ed è facile verificare che si ha

$$g(t_i) = \frac{1}{\mathbf{P}\{T = t_i\}} \int_{\{T=t_i\}} \varphi(X) d\mathbf{P} = \int_{\Omega} \varphi(X) d\mathbf{P}^i$$

dove \mathbf{P}^i è la probabilità condizionale $\mathbf{P}^i(A) = \mathbf{P}(A|T = t_i)$.

Proposizione 6.2.5. *Supponiamo che la variabile doppia (X, Y) abbia una densità $f(x, y)$ (rispetto alla misura di Lebesgue su \mathbb{R}^2), sia φ boreliana limitata e poniamo*

$$g(y) = \int \varphi(x) \frac{f(x, y)}{f_Y(y)} dx$$

Allora si ha $\mathbf{E}[\varphi(X)|Y] = g(Y)$.

Dimostrazione. Si tratta infatti di provare che su ogni evento della forma $\{Y \in B\}$ con B boreliano (un elemento di $\sigma(Y)$) si ha $\int_{\{Y \in B\}} \varphi(X) d\mathbf{P} = \int_{\{Y \in B\}} g(Y) d\mathbf{P}$, e questa è una conseguenza dell'eguaglianza

$$\int_{\mathbb{R}} \int_B \varphi(x) f(x, y) dx dy = \int_B f_Y(y) dy \int_{\mathbb{R}} \varphi(x) \frac{f(x, y)}{f_Y(y)} dx = \int_B g(y) f_Y(y) dy$$

□

Per essere rigorosi, il numero $\frac{f(x,y)}{f_Y(y)}$ è definito se $f_Y(y) > 0$, mentre se $f_Y(y) = 0$ anche $f(x,y) = 0$ e in quei punti ci troviamo con una forma indeterminata. Questo però non è un problema perché poi si moltiplica per $f_Y(y)$ e quindi si integra.

Torniamo all'esempio precedente e definiamo $N(y, A) = \int_A \frac{f(x,y)}{f_Y(y)} dx$: fissato $A \in \mathcal{B}(\mathbb{R})$, la funzione $y \rightarrow N(y, A)$ è boreliana, e fissato $y \in \mathbb{R}$, la funzione d'insieme $A \rightarrow N(y, A)$ è una *probabilità*. Una funzione con queste proprietà è chiamata **probabilità di transizione** o anche **nucleo markoviano**. Si può riscrivere l'eguaglianza vista sopra nella forma

$$\mathbf{E}[\varphi(X)|Y = y] = \int \varphi(x) N(y, dx)$$

Quando per una coppia di variabili (X, Y) esiste una *probabilità di transizione* che verifica l'eguaglianza sopra scritta (per ogni funzione φ boreliana limitata), la probabilità di transizione $N(y, \cdot)$ è chiamata la **legge condizionale** di X data Y : la dimostrazione dell'esistenza della legge condizionale (sotto opportune condizioni) è riservata a corsi più avanzati.

6.3 Unione essenziale e un teorema di Halmos-Savage.

Introduciamo un nuovo concetto, quello di *unione essenziale* di una famiglia di insiemi misurabili: a questo scopo supponiamo assegnato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$, ma questa definizione si può dare benissimo con una misura σ -finita, in quanto è invariante per passaggio ad una misura equivalente.

Dati $A, B \in \mathcal{F}$, diciamo che $A \subseteq B$ q.c. se $A \cap B^c$ è trascurabile: se contemporaneamente $A \subseteq B$ q.c. e $B \subseteq A$ q.c. (o, ciò che è lo stesso, se $A \Delta B = (A \cap B^c) \cup (B \cap A^c)$ è trascurabile), diciamo che A e B sono *equivalenti* (e scriviamo $A \sim B$).

Definizione 6.3.1. Data una famiglia $(A_i)_{i \in I}$ di elementi di \mathcal{F} , diciamo che $B \in \mathcal{F}$ è l'**unione essenziale** degli insiemi A_i se si ha:

- $A_i \subseteq B$ q.c. per ogni $i \in I$;
- se $C \in \mathcal{F}$ è tale che $A_i \subseteq C$ q.c. per ogni i , allora anche $B \subseteq C$ q.c.

Dalla definizione appare evidente che l'unione essenziale, *se esiste*, è definita a meno di equivalenza di insiemi ed è unica: intuitivamente l'unione

6.3. UNIONE ESSENZIALE E UN TEOREMA DI HALMOS-SAVAGE.93

essenziale degli insiemi $(A_i)_{i \in I}$ è il più piccolo (a meno di equivalenza) insieme misurabile che contiene q.c. ogni A_i .

Notiamo ancora che l'unione (di cardinalità qualsiasi) di insiemi misurabili non è detto che sia misurabile, invece l'unione essenziale esiste sempre ed è misurabile come mostra il seguente risultato:

Proposizione 6.3.2. *Assegnata una famiglia qualsiasi $(A_i)_{i \in I}$ di elementi di \mathcal{F} , esiste un sottinsieme numerabile $\{i_1, i_2, \dots\} \subseteq I$ tale che l'insieme $B = \bigcup_{n \geq 1} A_{i_n}$ sia una versione dell'unione essenziale degli insiemi $(A_i)_{i \in I}$.*

Dimostrazione. Sia $\alpha = \sup_{J \subseteq I, J \text{ finito}} \mathbf{P}\left(\bigcup_{i \in J} A_i\right)$: esiste una successione J_1, J_2, \dots di sottinsiemi finiti di I tale che $\mathbf{P}\left(\bigcup_{i \in J_n} A_i\right) \uparrow \alpha$, e possiamo supporre $J_1 \subseteq J_2 \subseteq \dots$.

Poniamo $J = \bigcup_{n \geq 1} J_n$ (che è un sottinsieme numerabile di I) e poniamo $B = \bigcup_{i \in J} A_i$: mostriamo che questo insieme B è una versione dell'unione essenziale. Da una parte è evidente che, se $A_i \subseteq C$ q.c. per ogni i , anche $B \subseteq C$ q.c.

Rimane da provare che ogni $A_i \subseteq B$ q.c.: supponiamo che esista un indice $k \in I$ tale che si abbia $\mathbf{P}(A_k \cap B^c) = \delta > 0$ e consideriamo $\tilde{J}_n = J_n \cup \{k\}$. È immediato constatare che $\mathbf{P}\left(\bigcup_{i \in \tilde{J}_n} A_i\right) \geq \mathbf{P}\left(\bigcup_{i \in J_n} A_i\right) + \delta$, e questo è in contraddizione con la definizione del numero α . \square

Esempio 6.3.3. Se consideriamo, nell'intervallo $[0, 1]$ (munito della misura di Lebesgue) tutti gli insiemi formati da un solo punto $\{x\}$, $x \in [0, 1]$, l'unione di questi insiemi è naturalmente tutto $[0, 1]$, ma l'unione essenziale è l'insieme vuoto.

Osservazione 6.3.4 (Estremo superiore essenziale di una famiglia di variabili aleatorie). Simile alla Definizione 6.3.1 è quella di “sup essenziale” di una famiglia di v.a. reali $(X_i)_{i \in I}$: questa è una v.a. Y (a valori in $] -\infty, +\infty[$) definita da queste due proprietà:

- $X_i \leq Y$ q.c. per ogni $i \in I$;
- se Z è tale che $X_i \leq Z$ q.c. per ogni i , allora anche $Y \leq Z$ q.c.

La dimostrazione dell'esistenza del sup essenziale ricalca quella della Proposizione 6.3.2: innanzi tutto sostituendo X_i con $Z_i = \arctan(X_i)$ (e considerando la funzione “arctan” biunivoca strettamente crescente da $[-\infty, +\infty[$ a $[-\pi/2, \pi/2[$), si può supporre che le variabili siano uniformemente limitate.

Si considera poi il numero

$$\alpha = \sup_{J \subseteq I, J \text{ finito}} \mathbf{E} \left[\sup_{i \in J} X_i \right]$$

e si procede poi in maniera simile: i dettagli sono lasciati per esercizio

Naturalmente la *funzione indicatrice* dell'unione essenziale degli insiemi $(A_i)_{i \in I}$ è il *sup essenziale* delle funzioni $(I_{A_i})_{i \in I}$.

Il concetto di *unione essenziale* è alla base del risultato seguente:

Teorema 6.3.5 (Teorema di Halmos–Savage). *Sia $(\mathbf{P}^i, i \in I)$ una famiglia di probabilità su (Ω, \mathcal{F}) , tutte assolutamente continue rispetto ad una misura σ -finita μ : esiste una probabilità \mathbf{Q} della forma $\mathbf{Q} = \sum_{n \geq 1} a_n \mathbf{P}^{i_n}$ (con $a_n \geq 0$, $\sum_{n=1}^{+\infty} a_n = 1$), tale che ogni \mathbf{P}^i sia assolutamente continua rispetto a \mathbf{Q} .*

Dimostrazione. Scegliamo una versione f_i della densità $\frac{d\mathbf{P}^i}{d\mu}$ e sia A l'unione essenziale degli insiemi $\{f_i > 0\}$ (unione essenziale fatta rispetto alla misura μ). Si ha dunque $A = \bigcup_{n \geq 1} \{f_{i_n} > 0\}$, con una opportuna successione di indici i_1, i_2, \dots

Poichè ogni densità f_i è nulla μ -q.c. fuori dell'insieme A , è immediato verificare che, per ogni i , $\mathbf{P}^i \ll I_A \cdot \mu$ (la misura che ha densità l'indicatrice di A rispetto a μ).

Sia ora $\mathbf{Q} = \sum_{n \geq 1} 2^{-n} \mathbf{P}^{i_n}$: la densità di \mathbf{Q} rispetto a μ è $\sum_{n \geq 1} 2^{-n} f_{i_n}$ e quindi è strettamente positiva sull'insieme A , cioè $\mathbf{Q} \sim I_A \cdot \mu$. Quindi, per ogni i , $\mathbf{P}^i \ll \mathbf{Q}$.

□

Osservazione 6.3.6. Il Teorema appena visto è rilevante nei *modelli statistici*, nei quali si considera su uno spazio (Ω, \mathcal{F}) una famiglia di probabilità $(\mathbf{P}^\theta, \theta \in \Theta)$: quando esiste una misura σ -finita μ rispetto alla quale ogni \mathbf{P}^θ sia assolutamente continua, il modello è detto *dominato*. In tal caso, per il Teorema 6.3.5, si può considerare una *probabilità dominante* della forma

$$\mathbf{Q} = \sum_{n=1}^{+\infty} 2^{-n} \mathbf{P}^{\theta_n} \quad (\text{con indici } \theta_1, \theta_2, \dots \text{ opportunamente scelti}).$$

Si chiama poi *verosimiglianza* una funzione $L: \Theta \times \Omega$ tale che, fissato θ , $L(\theta, \omega)$ sia una versione della densità di \mathbf{P}^θ rispetto alla misura dominante scelta.

Una conseguenza importante della nozione di *speranza condizionale*, della caratterizzazione fornita dalla Proposizione 6.1.7 e della formula che compare nel Teorema 6.1.9 è il risultato seguente, la cui dimostrazione è lasciata per esercizio.

Teorema 6.3.7. *Supponiamo che il modello statistico abbia una verosimiglianza della forma $L(\theta, \omega) = h(\theta, T(\omega)) \cdot k(\omega)$ dove T è una v.a. a valori in uno spazio (E, \mathcal{E}) (in tal caso T è chiamata riassunto esaustivo): la speranza condizionale $\mathbf{E}^\theta[X|T]$ di una v.a. (limitata) X non dipende dalla probabilità \mathbf{P}^θ . Inoltre se U è una stima corretta di quadrato integrabile della funzione $g(\theta)$, allora $V = \mathbf{E}^*[U|T]$ (versione della speranza condizionale indipendente dal parametro) è una stima preferibile a U , strettamente preferibile a meno che U non sia già T -misurabile.*

6.4 Sugli spazi di probabilità “non atomici”.

Definizione 6.4.1. Assegnato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$, si chiama **atomo** un elemento non trascurabile $A \in \mathcal{F}$ tale che non esista un elemento $B \in \mathcal{F}$ che sia $B \subseteq A$ e per il quale che si abbia $0 < \mathbf{P}(B) < \mathbf{P}(A)$.

Sembra evidente che in uno spazio di probabilità *privo di atomi* sia possibile trovare un evento di probabilità qualsiasi; la dimostrazione di questo fatto però non è immediata e quella che segue è il risultato di una discussione con Giorgio Letta.

Teorema 6.4.2. *Su uno spazio di probabilità privo di atomi, assegnato $0 < \alpha < 1$, esiste $A \in \mathcal{F}$ con $\mathbf{P}(A) = \alpha$.*

Dimostrazione. Cominciamo ad osservare che, dato $\varepsilon > 0$, esiste $B \in \mathcal{F}$ con $0 < \mathbf{P}(B) < \varepsilon$. Prendiamo infatti un intero $n > 1/\varepsilon$, si può costruire una partizione di Ω in n eventi A_1, \dots, A_n di probabilità strettamente positiva: di questi almeno uno ha probabilità inferiore a $1/n$.

Allo stesso modo si vede che, dati $B \in \mathcal{F}$ non trascurabile ed $\varepsilon > 0$, esiste $A \in \mathcal{F}$ con $A \subseteq B$ e $0 < \mathbf{P}(A) < \varepsilon$.

Definiamo

$$\mathcal{H}_0 = \{B \in \mathcal{F} \mid \mathbf{P}(B) \leq \alpha\} \quad , \quad A_0 = \emptyset$$

e poniamo poi per induzione $\mathcal{H}_n = \{B \supseteq A_{n-1} \mid \mathbf{P}(B) \leq \alpha\}$ e scegliamo $A_n \in \mathcal{H}_n$ tale che si abbia

$$\mathbf{P}(A_n) \geq \sup \{\mathbf{P}(B) \mid B \in \mathcal{H}_n\} - \frac{1}{n}$$

Si è così determinata una successione $A_1 \subseteq A_2 \dots$ di elementi di \mathcal{H}_0 e poniamo $A = \bigcup_n A_n$: poichè $\mathbf{P}(A) = \lim_{n \rightarrow \infty} \mathbf{P}(A_n)$ anche $A \in \mathcal{H}_0$ (cioè $\mathbf{P}(A) \leq \alpha$). A è *massimale* in \mathcal{H}_0 nel senso che se $B \in \mathcal{H}_0$ e $B \supseteq A$, allora $\mathbf{P}(B) = \mathbf{P}(A)$: infatti B appartiene ad ogni \mathcal{H}_n e quindi $\mathbf{P}(B) \leq \mathbf{P}(A_n) + 1/n$ e al limite si ottiene $\mathbf{P}(B) \leq \mathbf{P}(A)$.

Ma se A è *massimale*, necessariamente $\mathbf{P}(A) = \alpha$: se infatti $\delta = \alpha - \mathbf{P}(A)$ fosse strettamente positivo, potrei trovare $C \in A^c$ con $0 < \mathbf{P}(C) < \delta$ e $A \cup C$ sarebbe ancora un elemento di \mathcal{H}_0 . □

Osservazione 6.4.3. Naturalmente con la stessa dimostrazione si prova che, dati due eventi A e B con $A \subset B$ ed un numero α con $\mathbf{P}(A) < \alpha < \mathbf{P}(B)$, esiste un evento C con $A \subset C \subset B$ e $\mathbf{P}(C) = \alpha$: nella Proposizione 6.4.2 $A = \emptyset$ e $B = \Omega$.

Osservazione 6.4.4. In verità in uno spazio non-atomico si ha un risultato più preciso: *è possibile determinare, per ogni t con $0 \leq t \leq 1$, un evento A_t tale che $\mathbf{P}(A_t) = t$ e che, se $s < t$, $A_s \subset A_t$* . La dimostrazione di quest'ultima affermazione (non necessaria agli scopi di questo corso ma importante in talune questioni di Processi Stocastici) è più delicata e segue dall'*assioma della scelta*, o meglio dal *lemma di Zorn*. Lo studente interessato può sforzarsi di provarlo come esercizio impegnativo.

Gli spazi di probabilità *privi di atomi* sono interessanti perché su di essi è possibile costruire una successione di v.a. *indipendenti con distribuzione di probabilità qualsiasi*. Vediamo meglio perché.

Cominciamo con l'intervallo $[0, 1]$ (munito della misura di Lebesgue). Su di esso è usuale costruire una successione X_1, X_2, \dots di variabili indipendenti con *legge di Bernoulli* di parametro $1/2$ nel modo seguente: la v.a. X_1 assume il valore 0 nell'intervallo $[0, 1/2[$ e 1 nell'intervallo $[1/2, 1]$; la variabile X_2 assume il valore 0 negli intervalli $[0, 1/4[$ e $[1/2, 3/4[$ e 1 nei restanti intervalli, e così via

Sostanzialmente la variabile $X_i(\omega)$ è il valore della i -ma cifra nello sviluppo diadico del numero $\omega \in [0, 1]$ e quello che è detto sopra ripete la dimostrazione della Proposizione 5.3.7.

Esattamente la stessa costruzione si può fare su uno spazio Ω privo di atomi: si *divide* Ω in due eventi disgiunti A_1 e A_2 ciascuno di probabilità $1/2$ e si considera la v.a. Z_1 che assume il valore 0 sul primo e 1 sul secondo. Quindi si *divide* A_1 in due sottinsiemi $A_{1,1}$ e $A_{1,2}$ ciascuno di probabilità $1/4$ (e si fa la stessa operazione con A_2) e si considera la v.a. Z_2 che assume il valore 0 su $A_{1,1}$ e 1 su $A_{1,2}$ (e analogamente su $A_{2,1}$ e $A_{2,2}$) e così si prosegue.

Ci serve ora una proprietà la cui dimostrazione è lasciata per esercizio:

Proposizione 6.4.5. *Sia X_1, X_2, \dots una successione di v.a. indipendenti con legge di Bernoulli di parametro $1/2$: la variabile*

$$Y = \sum_{n=1}^{+\infty} \frac{X_n}{2^n}$$

ha legge uniforme su $[0, 1]$.

Abbiamo visto che, assegnata una v.a. X uniforme su $[0, 1]$ ed una funzione di ripartizione F , se si considera la *pseudo inversa sinistra* G di F (definita nel Teorema 4.3.9), la legge della v.a. $Y = G(X)$ è quella associata alla funzione di ripartizione F .

Possiamo riassumere queste affermazioni nel seguente risultato:

Teorema 6.4.6. *Assegnato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$, è equivalente poter costruire su di esso:*

- a) *una variabile aleatoria X con legge uniforme su $[0, 1]$;*
- b) *una successione X_1, X_2, \dots di variabili indipendenti di Bernoulli di parametro $1/2$;*
- c) *per ogni $h = 1, 2, \dots$ una successione X_1^h, X_2^h, \dots di v.a. di Bernoulli di parametro $1/2$, tutte indipendenti tra di loro*
- d) *assegnata una successione di leggi di probabilità su \mathbb{R} corrispondenti alle funzioni di ripartizione F_1, F_2, \dots una successione Z_1, Z_2, \dots di variabili indipendenti, tali che la funzione di ripartizione di Z_i sia F_i .*

Inoltre vale una delle proprietà equivalenti sopra elencate se e solo se lo spazio è non atomico.

Sostanzialmente abbiamo già dimostrato tutto.

Per quanto riguarda $a) \implies b)$, notiamo che se X è uniforme su $[0, 1]$ e R_1, R_2, \dots sono le variabili che rappresentano la i -ma cifra dopo la virgola nello sviluppo diadico, le variabili $R_n \circ X$ sono indipendenti di Bernoulli di parametro $1/2$; l'implicazione $b) \implies a)$ è la proposizione 6.4.5.

Mentre è ovvio che $c) \implies b)$, per provare l'implicazione opposta consideriamo una partizione di \mathbb{N} in una successione A_1, A_2, \dots di sottinsiemi infiniti (necessariamente di cardinalità \aleph_0) e *raccogliamo* le variabili $(X_h)_h$ con $h \in A_j$ nella successione X_1^j, X_2^j, \dots

L'equivalenza $c) \iff d)$ è sostanzialmente già vista.

L'ultima affermazione è conseguenza di questo fatto (la cui dimostrazione è lasciata per esercizio):

Supponiamo che sullo spazio $(\Omega, \mathcal{F}, \mathbf{P})$ sia definita una variabile aleatoria reale X con legge diffusa: necessariamente lo spazio Ω è non atomico.

Capitolo 7

Breve introduzioni ai Processi Stocastici.

7.1 Prime definizioni.

Supponiamo assegnato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$ ed un *insieme dei tempi* \mathcal{T} : si chiama **processo stocastico** (a valori reali) una famiglia di variabili aleatorie reali $(X_t)_{t \in \mathcal{T}}$ definite su Ω . Un processo stocastico si può identificare con una funzione $\mathbf{X} : \Omega \times \mathcal{T} \rightarrow \mathbb{R}$ tale che, fissato $t \in \mathcal{T}$, la funzione $\omega \rightarrow X(\omega, t)$ sia una variabile aleatoria.

Nelle applicazioni più frequenti l'insieme \mathcal{T} dei tempi è \mathbb{N} o un sottinsieme di \mathbb{N} (e si parla di processo a **tempi discreti**), oppure un intervallo di \mathbb{R} (e si parla di processo a **tempi continui**). In questa breve introduzione ci limitiamo al caso in cui l'insieme dei tempi è un intervallo limitato di \mathbb{R} , più precisamente $\mathcal{T} = [0, T]$.

Definizione 7.1.1. Si chiama **traiettoria** nel punto ω la funzione $t \rightarrow X(\omega, t)$. Si noti che in generale non è detto che le traiettorie siano funzioni misurabili.

Definizione 7.1.2. Dati due processi stocastici \mathbf{X} e \mathbf{Y} , si dice che \mathbf{X} è una **modificazione** di \mathbf{Y} se, per ogni t fissato, $X_t = Y_t$ q.c.

Definizione 7.1.3. Si dice che \mathbf{X} e \mathbf{Y} sono **indistinguibili** se esiste un sottinsieme trascurabile $N \subseteq \Omega$ tale che, se $\omega \notin N$, $X(\omega, t) = Y(\omega, t)$ per ogni $t \in \mathcal{T}$. Più precisamente, tutte le traiettorie sono eguali eccetto che su un insieme trascurabile.

Mentre due processi indistinguibili sono *a tutti gli effetti lo stesso processo stocastico*, due processi che siano una modificazione dell'altro possono essere molto diversi come mostra l'esempio che segue.

Esempio 7.1.4. Sia $\Omega = [0, 1]$ (munito della σ -algebra di Borel e della misura di Lebesgue) e sia

$$X(x, t) = \begin{cases} 1 & x = t \\ 0 & \text{altrimenti} \end{cases}$$

e sia invece $Y(x, t) \equiv 0$. Si verifica immediatamente che i due processi sono una modificazione l'uno dell'altro, tuttavia *tutte le traiettorie di \mathbf{Y} sono continue, mentre tutte le traiettorie di \mathbf{X} sono discontinue.*

Esattamente come per le variabili aleatorie, si può definire la **legge di probabilità** di un processo stocastico. Consideriamo l'applicazione

$$\mathbf{X} : \Omega \longrightarrow \left(\prod_{0 \leq t \leq T} \mathbb{R}_t, \otimes_{0 \leq t \leq T} \mathcal{B}(\mathbb{R}_t) \right)$$

che al punto $\omega \in \Omega$ associa la **traiettoria** $t \rightarrow X(\omega, t)$: vediamo perché questa applicazione è *misurabile*. Infatti la σ -algebra prodotto $\otimes_{0 \leq t \leq T} \mathcal{B}(\mathbb{R}_t)$ è generata dalle proiezioni coordinate, e quindi è sufficiente constatare che, per ogni t , $\pi_t \circ \mathbf{X} = X_t$ è misurabile.

Definizione 7.1.5. Si chiama **legge del processo stocastico** $(X_t)_{t \in [0, T]}$ l'immagine della probabilità \mathbf{P} mediante l'applicazione \mathbf{X} sopra definita.

Quando due processi stocastici hanno la stessa legge vengono talvolta detti *equivalenti*.

Osservazione 7.1.6. Se due processi \mathbf{X} e \mathbf{Y} sono *modificazione* uno dell'altro, hanno la stessa legge. Infatti le probabilità $\mathbf{X}(\mathbf{P})$ e $\mathbf{Y}(\mathbf{P})$ coincidono sugli insiemi della forma $\pi_{t_1}^{-1}(A_1) \cap \dots \cap \pi_{t_k}^{-1}(A_k)$ al variare di t_1, \dots, t_k e di A_1, \dots, A_k boreliani: questo perché

$$\mathbf{X}(\mathbf{P})\left(\pi_{t_1}^{-1}(A_1) \cap \dots \cap \pi_{t_k}^{-1}(A_k)\right) = \mathbf{P}\left(X_{t_1}^{-1}(A_1) \cap \dots \cap X_{t_k}^{-1}(A_k)\right)$$

e gli insiemi di quella forma sono *stabili per intersezione* e generano la σ -algebra prodotto. Si può applicare pertanto il Corollario 1.1.2.

Tuttavia lo spazio $\mathbb{R}^{[0, T]} = \prod_{0 \leq t \leq T} \mathbb{R}_t$ è uno spazio *troppo grande*: ad esempio si può provare che il sottoinsieme di $\mathbb{R}^{[0, T]}$ formato dalle funzioni continue non è misurabile.

Come vedremo più avanti, se le traiettorie del processo hanno una certa regolarità è più conveniente considerare la legge del processo su un opportuno spazio di funzioni.

7.2 Il Processo di Wiener (o Moto Browniano).

In questo paragrafo esaminiamo il *processo stocastico di Wiener*, chiamato usualmente anche *moto Browniano*: come insieme dei tempi scegliamo $[0, T]$, ma la stessa costruzione si può fare prendendo come insieme dei tempi $\mathbb{R}^+ = [0, +\infty[$.

Definizione 7.2.1. Si chiama **moto Browniano grossolano** un processo stocastico $(B_t)_{0 \leq t \leq T}$ che gode delle seguenti proprietà:

- a) $B_0 = 0$;
- b) presi $0 < t_1 < \dots < t_n$, le v.a. $(B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}})$ sono indipendenti;
- c) presi $0 \leq s < t$, la variabile $(B_t - B_s)$ ha legge gaussiana $N(0, t - s)$.

Quando è verificata la condizione b) sopra scritta si usa dire che il processo è *a incrementi indipendenti*.

Definizione 7.2.2. Si chiama **moto Browniano standard** o **processo di Wiener** un processo stocastico $(W_t)_{0 \leq t \leq T}$ che soddisfa le proprietà a), b) e c) della Definizione 7.2.1 ed inoltre è tale che:

- d) le traiettorie del processo $(W_t)_{0 \leq t \leq T}$ sono continue.

La strada che seguiremo consiste nel costruire dapprima un moto Browniano *grossolano* (seguendo sostanzialmente il metodo proposto da Wiener) e poi nell'utilizzare un Teorema di Kolmogorov che permette di costruire una *modificazione* con traiettorie continue, anzi γ -holderiane per ogni $0 < \gamma < 1/2$.

A tale scopo, consideriamo uno spazio $(\Omega, \mathcal{F}, \mathbf{P})$ sul quale è definita una successione X_1, X_2, \dots di variabili indipendenti gaussiane $N(0, 1)$ (abbiamo visto nel capitolo precedente che questo si può fare su un qualsiasi spazio di probabilità *non-atomico*), e prendiamo una successione g_1, g_2, \dots che sia un *sistema ortonormale completo* in $L^2(0, T) = L^2([0, T], \mathcal{B}([0, T]), \lambda)$.

Poniamo poi $G_i(t) = \int_0^t g_i(s) ds = \int_{[0, t]} g_i(s) I_{[0, t]}(s) ds = \langle g_i, I_{[0, t]} \rangle_{L^2(0, T)}$, e definiamo

$$B(\omega, t) = \sum_{n=1}^{+\infty} X_n(\omega) G_n(t)$$

Naturalmente occorre provare che la serie sopra scritta converge in qualche senso e che il processo stocastico così definito soddisfa le condizioni della definizione 7.2.1.

A tale scopo consideriamo le somme parziali $B^n(\omega, t) = \sum_{k=1}^n X_k(\omega)G_k(t)$: fissato t , ogni v.a. B_t^n è una variabile gaussiana con media 0 e varianza $\sum_{k=1}^n G_k(t)^2$. Osserviamo che si ha

$$\sum_{k=1}^{+\infty} G_k^2(t) = \sum_{k=1}^{+\infty} (\langle g_k, I_{[0,t]} \rangle_{L^2(0,T)})^2 = \|I_{[0,t]}\|_{L^2(0,T)}^2 = t$$

e di conseguenza la serie $\sum_{k=1}^{+\infty} X_k(\omega)G_k(t)$ converge in L^2 a una variabile gaussiana con media 0 e varianza t (ricordiamo che limite in legge di variabili gaussiane è ancora una variabile gaussiana). Allo stesso modo si prova che $(B_t - B_s)$ ha legge $N(0, t-s)$.

Vediamo ora la proprietà b) della Definizione 7.2.1, limitandoci per semplicità a due istanti $0 \leq s < t$: poichè la coppia $(B_s, B_t - B_s)$ forma un vettore gaussiano, è sufficiente provare che queste variabili sono incorrelate. Il fatto che questa coppia sia un vettore gaussiano segue dal fatto che, per ogni n , $(B_s^n, B_t^n - B_s^n)$ è un vettore gaussiano, e questa proprietà si conserva al limite; per ogni n , la covarianza di $(B_s^n, B_t^n - B_s^n)$ è eguale a

$$\begin{aligned} \mathbf{E}[B_s^n \cdot (B_t^n - B_s^n)] &= \sum_{k=1}^n G_k(s) \cdot (G_k(t) - G_k(s)) = \\ &= \sum_{k=1}^n \langle g_k, I_{[0,s]} \rangle_{L^2(0,T)} \cdot \langle g_k, I_{[s,t]} \rangle_{L^2(0,T)} \end{aligned}$$

e al limite si ottiene

$$\mathbf{E}[B_s \cdot (B_t - B_s)] = \sum_{k=1}^{+\infty} \langle g_k, I_{[0,s]} \rangle_{L^2(0,T)} \cdot \langle g_k, I_{[s,t]} \rangle_{L^2(0,T)} = \langle I_{[0,s]}, I_{[s,t]} \rangle_{L^2(0,T)} = 0$$

cioè il risultato voluto: abbiamo così ottenuto la costruzione del *moto Browniano grossolano*.

Per ottenere il *Moto Browniano Standard* o *Processo di Wiener*, utilizzeremo il seguente importante teorema dovuto a Kolmogorov

Teorema 7.2.3 (Criterio di continuità di Kolmogorov). *Sia $(X_t)_{0 \leq t \leq T}$ un processo stocastico e supponiamo che esistano tre costanti positive α, β, C tali che per ogni $0 \leq s < t \leq T$ si abbia*

$$\mathbf{E}[|X_t - X_s|^\alpha] \leq C |t-s|^{1+\beta}$$

Allora \mathbf{X} ha una modificazione con traiettorie γ -hölderiane, per ogni $0 < \gamma < \beta/\alpha$.

Ricordiamo che, preso $0 < \gamma < 1$, una funzione f definita su un sottinsieme di \mathbb{R} è detta γ -hölderiana se esiste una costante C tale che si abbia, per ogni coppia (x, y) , $|f(x) - f(y)| \leq C |x - y|^\gamma$.

Se nella disuguaglianza sopra scritta si prende $\gamma = 1$, la funzione è detta *lipschitziana*; è facile verificare che non si può prendere $\gamma > 1$ perché *solo le costanti* sono γ -hölderiane.

Per semplificare le notazioni, supponiamo che il processo abbia come insieme dei tempi $\mathcal{T} = [0, 1]$, chiamiamo D_n l'insieme dei numeri della forma $\{k \cdot 2^{-n} \mid 0 \leq k \leq 2^n\}$ e sia $D = \cup_n D_n$ l'insieme dei *numeri diadici* dell'intervallo $[0, 1]$. Cominciamo a separare un conto elementare:

Lemma 7.2.4. *Sia f una funzione definita su D a valori reali e supponiamo che esista n_0 tale che, per $n \geq n_0$ si abbia, per due successivi punti di D_n*

$$|f((k+1) \cdot 2^{-n}) - f(k \cdot 2^{-n})| \leq 2^{-n\gamma}$$

Allora la funzione f è γ -hölderiana su D (e di conseguenza si prolunga ad una funzione γ -hölderiana su $[0, 1]$).

Dimostrazione. Cominciamo ad osservare che esiste una costante C tale che, se $n \geq n_0$ e $x \in D_n$, preso $s \in D$ con $|s - x| < 2^{-n}$, si ha $|f(s) - f(x)| \leq C 2^{-n\gamma}$.

Infatti $x = k \cdot 2^{-n}$ e, se $x < s < x + 2^{-n}$, si ha $s = k \cdot 2^{-n} + \sum_{h=1}^{+\infty} \varepsilon_h \cdot 2^{-(n+h)}$, dove i numeri ε_h possono prendere i valori 0 oppure 1 (e da un certo indice in poi sono sempre 0). Se $x - 2^{-n} < s < x$ la rappresentazione è la stessa, con gli ε_h che possono prendere i valori 0 oppure -1 .

Usando ripetutamente la disuguaglianza dell'ipotesi, si ottiene

$$|f(s) - f(k \cdot 2^{-n})| \leq \sum_{h=1}^{+\infty} |\varepsilon_h| 2^{-\gamma(n+h)} \leq 2^{-n\gamma} \sum_{h=1}^{+\infty} 2^{-h\gamma} = C 2^{-n\gamma}$$

Proviamo ora che esiste una costante C^* tale che, se $x, y \in D$ e $|x - y| < 2^{-n_0}$, si ha $|f(x) - f(y)| \leq C^* |x - y|^\gamma$. Consideriamo infatti il numero $n \geq n_0$ tale che $2^{-(n+1)} < |x - y| \leq 2^{-n}$: tra x e y è compreso al più un solo punto di D_n ed usando due volte la disuguaglianza sopra ottenuta si ha

$$|f(x) - f(y)| \leq 2C \cdot 2^{-n\gamma} \leq 2^{1+\gamma} C \cdot |x - y|^\gamma$$

Poichè la funzione f è evidentemente limitata su D , posto $M = \sup_{x \in D} |f(x)|$, se $x, y \in D$ e $|x - y| > 2^{-n_0}$, si ha

$$|f(x) - f(y)| \leq \frac{2M}{|x - y|^\gamma} \cdot |x - y|^\gamma \leq (2M 2^{n_0\gamma}) \cdot |x - y|^\gamma$$

e questo completa la dimostrazione. □

Vediamo ora la dimostrazione del Teorema 7.2.3.

Dimostrazione. Consideriamo il processo stocastico \mathbf{X} con l'insieme dei tempi ristretto a D e, fissato $m \in \mathbb{N}$, poniamo

$$Z_m = \max_{0 \leq k < 2^m} |X_{(k+1) \cdot 2^{-m}} - X_{k \cdot 2^{-m}}|$$

Valgono le seguenti maggiorazioni

$$\begin{aligned} \mathbf{P}\{Z_m > 2^{-m\gamma}\} &\leq \sum_{k=0}^{2^m-1} \mathbf{P}\{|X_{(k+1) \cdot 2^{-m}} - X_{k \cdot 2^{-m}}| > 2^{-m\gamma}\} \leq \\ &\leq C 2^m 2^{m\gamma\alpha} 2^{-m(1+\beta)} = C 2^{m(\gamma\alpha-\beta)} \end{aligned}$$

Se $\gamma < \beta/\alpha$, la serie sopra scritta converge: di conseguenza per il Lemma di Borel-Cantelli 2.4.2, per quasi ogni ω esiste $\bar{m}(\omega)$ tale che, se $m \geq \bar{m}(\omega)$, si ha $Z_m(\omega) \leq 2^{-m\gamma}$ ed il Lemma 7.2.4 permette di affermare che la traiettoria corrispondente, ristretta a D , è γ -h\"olderiana.

L'insieme trascurabile sul quale questa propriet\`a non \`e soddisfatta dipende da γ , tuttavia si pu\`o considerare una successione di numeri γ_n convergente crescendo a β/α e l'unione (numerabile) degli insiemi trascurabili sui quali le traiettorie corrispondenti non sono γ_n -h\"olderiane: poich\`e una funzione γ -h\"olderiana \`e γ' -h\"olderiana se $\gamma' < \gamma$, ne segue che fuori di un insieme trascurabile le traiettorie del processo, ristrette a D , sono γ -h\"olderiane per ogni $\gamma < \beta/\alpha$.

Per ottenere la modificazione di \mathbf{X} con traiettorie h\"olderiane, si pone $\tilde{X}_t = \lim_{s \in D, s \rightarrow t} X_s$: \`e facile vedere che, per ogni t fissato, $\tilde{X}_t = X_t$ q.c. Infatti, presa una successione s_n di elementi di D convergente a t , $X_{s_n} \rightarrow X_t$ in probabilit\`a. \square

Torniamo ora al Moto Browniano: notiamo che $(B_t - B_s)$ \`e eguale in legge a $|t-s|^{1/2}X$ dove $X \sim N(0, 1)$ e di conseguenza, preso $1 \leq p < +\infty$, si ha

$$\mathbf{E}[|B_t - B_s|^p] = c_p |t-s|^{p/2}$$

(dove $c_p = \mathbf{E}[|X|^p]$). \`E facile verificare che il Teorema 7.2.3 permette di ottenere la modificazione con traiettorie γ -h\"olderiane per ogni $\gamma < 1/2$; questa regolarit\`a non pu\`o essere superata.

Vale infatti il seguente risultato, che ha importanza fondamentale nella costruzione dell'*integrale stocastico*: nell'enunciato di questo ci riferiamo al *processo di Wiener* (cio\`e alla modificazione con traiettorie continue).

Proposizione 7.2.5. *Fissato t , per quasi ogni ω vale l'eguaglianza*

$$\lim_{n \rightarrow \infty} \sum_{k=0}^{2^n-1} \left(W_{\frac{t(k+1)}{2^n}} - W_{\frac{tk}{2^n}} \right)^2 = t$$

Dimostrazione. Anche nella facile dimostrazione di questo risultato supponiamo per semplicità di notazioni $t = 1$.

Poniamo $S_n = \sum_{k=0}^{2^n-1} \left(W_{\frac{k+1}{2^n}} - W_{\frac{k}{2^n}} \right)^2$: è facile constatare che si ha $\mathbf{E}[S_n] = 1$.

Inoltre si ha

$$\mathbf{E} \left[\left(W_{\frac{k+1}{2^n}} - W_{\frac{k}{2^n}} \right)^2 \left(W_{\frac{h+1}{2^n}} - W_{\frac{h}{2^n}} \right)^2 \right] = \begin{cases} 2^{-2n} & k \neq h \\ 3 \cdot 2^{-2n} & k = h \end{cases}$$

Di conseguenza si ha $\mathbf{E}[S_n^2] = 3 \cdot 2^{-n} + 1 - 2^{-n} = 2 \cdot 2^{-n} + 1$, e quindi $\mathbf{E}[(S_n - 1)^2] = 2^{-n+1}$, cioè S_n converge a 1 in L^2 . Tuttavia si può dire di più: preso $\alpha_n^2 = 2^{-n/2}$, si ha

$$\mathbf{P}\{|S_n - 1| > \alpha_n\} \leq \frac{2^{-(n-1)}}{\alpha_n^2}$$

ed il Lemma di Borel–Cantelli permette facilmente di concludere che la convergenza in realtà è quasi certa. \square

Una facile conseguenza del risultato appena enunciato è che le traiettorie del processo di Wiener non possono essere hölderiane con un esponente maggiore di $1/2$; rimarrebbe aperto il caso $\gamma = 1/2$ ma una proprietà ancora più raffinata dovuta a Paul Lévy (la *Legge del logaritmo iterato*) mostra che le traiettorie non sono neppure $1/2$ -hölderiane. Enunciamo il risultato (la cui dimostrazione, piuttosto complessa) è lasciata a un corso più avanzato:

Teorema 7.2.6. *Sia $(W_t)_{t \geq 0}$ un processo di Wiener, e fissiamo $t \leq 0$: per quasi ogni ω si ha*

$$\limsup_{s \rightarrow t+} \frac{W_{t+s}(\omega) - W_t(\omega)}{\sqrt{2s \log \log \frac{1}{s}}} = 1 \qquad \liminf_{s \rightarrow t+} \frac{W_{t+s}(\omega) - W_t(\omega)}{\sqrt{2s \log \log \frac{1}{s}}} = -1$$

Nel paragrafo precedente abbiamo detto che lo *spazio canonico* per la legge dei processi stocastici $\mathbb{R}^{[0,T]} = \prod_{0 \leq t \leq T} \mathbb{R}_t$ è in realtà troppo grande per essere utile: nel caso del processo di Wiener si può considerare la legge sullo spazio delle funzioni continue $C(0, T)$ (che è uno spazio normato completo,

cioè di Banach) munito della norma $\|f\| = \sup_{0 \leq x \leq T} |f(x)|$ (o meglio ancora sul sottospazio $C_0(0, T)$ delle funzioni continue nulle in 0).

Su questo spazio si possono considerare due σ -algebre: quella di Borel cioè generata dagli aperti (che è quella naturale *dal punto di vista topologico*) e quella generata dalle proiezioni coordinate $\pi_t(f) = f(t)$ (che è quella naturale dal punto di vista *dei processi stocastici*): queste due σ -algebre *per fortuna coincidono*.

Da una parte infatti, poiché le proiezioni coordinate sono funzioni continue (quindi misurabili) sullo spazio $C_0(0, T)$, la σ -algebra generata dalle proiezioni coordinate è contenuta nella σ -algebra di Borel. Questa a sua volta è generata dalle sfere chiuse (poiché la topologia di $C_0(0, T)$ è a base numerabile) e la sfera di centro f e raggio δ si può descrivere come

$$\{g \in C_0 : |g(s) - f(s)| \leq \delta \forall s \in \mathbb{Q}\} = \bigcap_{0 \leq s \leq T, s \in \mathbb{Q}} \{|\pi_s(f) - \pi_s(g)| \leq \delta\}$$

e si ottiene quindi l'eguaglianza delle due σ -algebre.

La **misura di Wiener** è in un certo senso la *legge canonica* del processo di Wiener: questa è una probabilità su $(C_0(0, T), \mathcal{B}(C_0))$ e, in modo analogo a quanto fatto nel paragrafo precedente, è l'immagine di \mathbf{P} mediante l'applicazione

$$\mathbf{W} : \Omega \longrightarrow (C_0(0, T), \mathcal{B}(C_0(0, T)))$$

che al punto $\omega \in \Omega$ associa la **traiettoria** $t \rightarrow W(\omega, t)$. La misurabilità di questa applicazione segue dal fatto che la σ -algebra di Borel coincide con quella generata dalle applicazioni coordinate.

Inoltre *la misura di Wiener è unica*: la giustificazione di questa affermazione si fa esattamente come nell'Osservazione 7.1.6.

7.3 Il processo di Poisson.

Definizione 7.3.1 (Processo di Poisson). Si chiama processo di Poisson di intensità λ un processo stocastico $(N_t)_{t \geq 0}$ con traiettorie continue a destra definito dalle seguenti proprietà:

- a) $N_0 = 0$;
- b) presi $0 < t_1 < \dots < t_n$, le v.a. $(N_{t_1}, N_{t_2} - N_{t_1}, \dots, N_{t_n} - N_{t_{n-1}})$ sono indipendenti;
- c) presi $0 \leq s < t$, la variabile $(N_t - N_s)$ ha legge di Poisson di parametro $\lambda(t-s)$.

Appare subito evidente che le traiettorie sono *costanti a tratti*, a valori interi, con *salto* di ampiezza un numero intero: in realtà quasi certamente i salti *hanno sempre ampiezza 1*.

Per il processo di Poisson è più comodo considerare come insieme dei tempi $[0, +\infty[$ e la sua costruzione può essere fatta nel modo seguente: sia S_1, S_2, \dots una successione di v.a. indipendenti con densità *esponenziale* di parametro λ e poniamo $T_n = S_1 + \dots + S_n$: possiamo immaginare le variabili aleatorie T_1, T_2, \dots come i successivi *tempi d'arrivo* di certi fenomeni casuali, mentre le variabili S_1, S_2, \dots sono gli *intertempi d'arrivo*. La v.a. N_t conta quanti arrivi ci sono stati fino all'istante t , più precisamente $N_t = \sum_{n=1}^{+\infty} n \cdot I_{\{T_n \leq t < T_{n+1}\}}$.

Notiamo che per la *legge forte dei grandi numeri di Kolmogorov 5.3.5*, T_n/n converge q.c. a $1/\lambda$ e quindi *i salti non si possono addensare*.

Ricordiamo che la v.a. T_n ha densità Gamma di parametri n e λ e pertanto si ottiene

$$\begin{aligned} \mathbf{P}\{N_t = n\} &= \mathbf{P}\{T_n \leq t < T_{n+1}\} = \mathbf{P}\{T_n \leq t\} - \mathbf{P}\{T_{n+1} \leq t\} = \\ &= \frac{1}{(n-1)!} \int_0^t \lambda^n x^{n-1} e^{-\lambda x} dx - \frac{1}{n!} \int_0^t \lambda^{n+1} x^n e^{-\lambda x} dx = \frac{(\lambda t)^n e^{-\lambda t}}{n!} \end{aligned}$$

e dunque N_t ha legge di Poisson di parametro λt .

Per provare la proprietà b) della definizione, anche questa volta consideriamo due istanti $0 < s < t$ e dobbiamo provare che, presi due interi positivi h e k , si ha $\mathbf{P}\{N_t - N_s = h | N_s = k\} = \frac{e^{-\lambda(t-s)} \lambda^h (t-s)^h}{h!}$: in questo modo infatti si mostra contemporaneamente l'indipendenza di N_s e $(N_t - N_s)$ ed il fatto che $(N_t - N_s)$ è una v.a. di Poisson di parametro $\lambda(t-s)$.

Consideriamo allora il processo $(N_t - N_s)_{t \geq s}$ e consideriamo gli *intertempi d'arrivo* di questo processo $\tilde{S}_1, \tilde{S}_2, \dots$: condizionatamente a $\{N_s = k\}$, si ha $\tilde{S}_1 = S_{k+1} - (s - T_k) = (T_{k+1} - s)$ e poi $\tilde{S}_h = S_{k+h}$ per $h \geq 2$. Si tratta di provare che, condizionatamente a $\{N_s = k\}$ queste variabili sono ancora indipendenti e esponenziali di parametro λ .

Per $h \geq 2$ la proprietà è evidente perché l'evento $\{N_s = k\}$ si descrive con le v.a. S_1, \dots, S_{k+1} ed è indipendente da S_{k+2}, S_{k+3}, \dots .

Nell'esame della variabile \tilde{S}_1 gioca un ruolo essenziale la proprietà che ha una v.a. esponenziale *di essere senza memoria*: più precisamente, se Y ha densità esponenziale di parametro λ , si ha

$$\mathbf{P}\{(Y-s) \geq t | Y > s\} = \mathbf{P}\{Y \geq t+s | Y > s\} = \mathbf{P}\{Y \geq t\} = e^{-\lambda t}$$

(per s, t positivi), cioè *condizionatamente a* $\{Y > s\}$, $(Y - s)$ ha ancora la stessa legge esponenziale.

In realtà si ha una proprietà ancora più forte : se Z è una v.a. indipendente da Y , a valori positivi e con densità, *condizionatamente a* $\{Y > Z\}$, $(Y - Z)$ è ancora esponenziale di parametro λ , cioè si ha

$$\mathbf{P}\{Y - Z \geq t \mid Y > Z\} = \mathbf{P}\{Y \geq t\} = e^{-\lambda t}$$

e più in generale, se Z non è necessariamente a valori positivi ma si ha $\mathbf{P}\{Z > 0\} > 0$, vale l'eguaglianza

$$\mathbf{P}\{Y - Z \geq t \mid Y > Z, Z \geq 0\} = e^{-\lambda t}$$

Tornando al caso $h = 1$, la variabile \tilde{S}_1 è la variabile $(S_{k+1} - (s - T_k))$ *condizionata a*

$$\{N_s = k\} = \{T_k \leq s < T_{k+1}\} = \{S_{k+1} > (s - T_k), (s - T_k) \geq 0\}$$

Ponendo, con le notazioni sopra scritte, $S_{k+1} = Y$ e $(s - T_k) = Z$, si ottiene che \tilde{S}_1 che è ancora esponenziale di parametro λ .

La costruzione del Processo di Poisson a partire dai *tempi di arrivo* può essere invertita: abbiamo visto che dalla Definizione 7.3.1 segue che le traiettorie sono costanti a tratti e con salti di ampiezza 1. Si possono così definire i successivi tempi dei salti T_1, T_2, \dots e gli intertempi S_1, S_2, \dots : non è difficile provare che le v.a. S_i sono indipendenti esponenziali di parametro λ .

7.4 Due parole sui Processi di Markov.

Per semplicità consideriamo anche in questa sezione processi stocastici a valori reali con insieme dei tempi $\mathcal{T} = [0, T]$ (oppure $\mathcal{T} = [0, +\infty[$).

Assegnato il processo stocastico $(X_t)_{t \in \mathcal{T}}$ indichiamo con $\mathcal{F}_s = \sigma\{X_r, 0 \leq r \leq s\}$ la σ -algebra generata dalle variabili X_r con $r \leq s$.

Definizione 7.4.1. Si dice che il processo \mathbf{X} è un **Processo di Markov in senso lato** se presi $0 \leq s < t$ e φ boreliana limitata si ha

$$\mathbf{E}[\varphi(X_t) \mid \mathcal{F}_s] = \mathbf{E}[\varphi(X_t) \mid X_s]$$

Intuitivamente, il *passato* ed il *presente* forniscono le stesse informazioni per quanto riguarda il futuro.

Definizione 7.4.2. Si dice che il processo \mathbf{X} è un **Processo di Markov in senso stretto** se presi $0 \leq s < t$, esiste una *probabilità di transizione* (o *nucleo markoviano*) $P_{s,t}$ tale che per ogni φ boreliana limitata si abbia

$$\mathbf{E}[\varphi(X_t) \mid \mathcal{F}_s] = (P_{s,t}\varphi)(X_s)$$

Ricordiamo che la probabilità di transizione è stata definita brevemente dopo la Proposizione 6.2.5: su \mathbb{R} , $\mathcal{B}(\mathbb{R})$ (come in questo caso) è una funzione $N : \mathbb{R} \times \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ tale che, fissato $A \in \mathcal{B}(\mathbb{R})$, $x \rightarrow N(x, A)$ è boreliana e, fissato $x \in \mathbb{R}$, $A \rightarrow N(x, A)$ è una probabilità. Alla probabilità di transizione è associato un operatore definito sulle funzioni boreliane limitate (a valori nello stesso spazio) mediante la formula $(N\varphi)(x) = \int_{\mathbb{R}} \varphi(y) N(x, dy)$.

Quando $P_{s,t}$ dipende solo da $(t-s)$, si dice che il processo è *omogeneo nel tempo*.

Un esempio di processo di Markov omogeneo nel tempo si ha con un processo *a incrementi indipendenti e stazionari*: un processo è *a incrementi indipendenti* se è soddisfatta la proprietà b) delle Definizioni 7.2.1 e 7.3.1, ed è *stazionario* se la legge di $(X_t - X_s)$ dipende solo da $(t-s)$.

In tal caso infatti, poiché $(X_t - X_s)$ è indipendente da \mathcal{F}_s , si ha (vedi Proposizione 6.2.2)

$$\mathbf{E}[\varphi(X_t) \mid \mathcal{F}_s] = \mathbf{E}[\varphi(X_s + (X_t - X_s)) \mid \mathcal{F}_s] = G_{s,t}(X_s)$$

con $G_{s,t}(x) = \mathbf{E}[\varphi(x + (X_t - X_s))] = \int \varphi(y) P_r(x, dy)$ dove $r = t-s$ e $P_r(x, \cdot)$ è la legge di $(x + (X_t - X_s))$.

Ad esempio nel **processo di Wiener**, $P_r(x, \cdot)$ è la legge $N(x, r)$ e si ha pertanto

$$(P_r\varphi)(x) = \frac{1}{\sqrt{2\pi r}} \int_{\mathbb{R}} \varphi(y) \exp\left(-\frac{(x-y)^2}{2r}\right) dy$$

Nel caso del **processo di Poisson**, la legge di $(N_t - N_s)$ è la legge di Poisson di parametro $\lambda(t-s)$: è quindi concentrata su \mathbb{N} e si ha

$$(P_r\varphi)(h) = \sum_{k=0}^{+\infty} e^{-\lambda r} \frac{(\lambda r)^k}{k!} \varphi(h+k)$$