Calcolo delle Probabilità

Appunti del corso 2019/20

David Barbato
barbato@math.unipd.it

http://www.math.unipd.it/~barbato

Dipartimento di Matematica
Università degli Studi di Padova
via Trieste 63, 35121 Padova, Italy

Versione 1.0

Ultima modifica: 11 dicembre 2019.

SOMMARIO. Queste note riassumono il contenuto del corso di calcolo delle probabilità tenuto da me nell'anno accademico 2019/20 per il corso di laurea triennale in matematica dell'Università degli Studi di Padova.

Segnalazioni di errori, osservazioni, suggerimenti e critiche sono molto graditi.

Indice

Introduzione

Capitolo 1

Che cosa è la probabilità?

In questa primo capitolo introdurremo i modelli di probabilità discreta. Si tratta di un'introduzione rapida poiché gli studendi del corso hanno già familiarità con tale argomento che è stato trattato nel corso *probabilità e statistica* del primo anno. Una referenza utile per approfondimenti è [1].

Il prossimo paragrafo utilizzerà gli spazi di probabilità discreti per illustrare la differenza tra il concetto di variabile aleatoria in senso astratto e il modello di probabilità che la realizza, mentre nel paragrafo successivo metteremo in chiaro quale è la maniera corretta di eseguire le somme infinite nel calcolo dei valori attesi.

1.1 Spazi di probabilità discreti

Una spazio di probabilità è un insieme Ω finito o numerabile e un'applicazione $p : \Omega \to \mathbb{R}^+$ detta densità discreta tale che

$$\sum_{\omega \in \Omega} p(\omega) = 1 \ .$$

La densità induce una probabilità P sui sottinsiemi di Ω, tale probabilità è la funzione $P : \mathcal{P}(\Omega) \to [0,1]$ definita da $P(A) = \sum_{\omega \in A} p(\omega)$ per ogni sottoinsieme A di Ω. Le funzioni X da Ω in un insieme E sono le variabili aleatorie. La variabile aleatoria X in un certo senso "trasporta" la probabilità da Ω su E nel seguente modo: per ogni sottoinsieme A di E si ha $P(X \in A) = P(\{\omega \in \Omega | X(\omega) \in A\})$. Questa nuova probabilità è detta distribuzione della variabile aleatoria X.

Esempio 1.1 (Lancio di un dado.)**.** *Supponiamo di voler costruire un modello in cui X sia il lancio di un dado regolare, cioè vogliamo che X sia una variabile reale con $P(X = i) = \frac{1}{6}$ per ogni $i \in \{1, 2, 3, 4, 5, 6\}$.*
Un modo semplice per costruire tale modello è quello di scegliere l'insieme degli esiti $\Omega = \{1, 2, 3, 4, 5, 6\}$ e la densità discreta $p_\omega = \frac{1}{6}$. La variabile aleatoria X da Ω in $\mathbb{R}$ data da $X(\omega) = \omega$ rappresenta il risultato del lancio.

Per esempio la probababilità che il risultato del lancio sia pari è data da $P(X \in 2\mathbb{Z}) = P(\{2,4,6\}) = \frac{3}{6}$.

Esempio 1.2 (Lancio di due dadi.). *Supponiamo ora di voler costruire un modello in cui* X *e* Y *rappresentino il lancio di due dadi regolari, cioè vogliamo che* X *e* Y *siano due variabili reali tali che per ogni* $i, j \in \{1,2,3,4,5,6\}$ *si abbia* $P(X = i) = \frac{1}{6}$, $P(Y = i) = \frac{1}{6}$, $P(X = i, Y = j) = \frac{1}{36}$. *(In realtà l'ultima condizione già implica le prime due.)*
Procedendo come nell'esempio precedente possiamo porre
$\Omega_2 = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}$
$p_\omega = \frac{1}{36}$ *per ogni* $\omega \in \Omega_2$.
Le variabili aleatoria X *e* Y *da* Ω *in* $\mathbb{R}$ *date da* $X(i,j) = i$ *e* $Y(i,j) = j$ *soddisfano le ipotesi richieste.*

La seguente osservazione chiarirà in parte una questione di solito lasciata in sospeso dai corsi di probabilità elementare del primo anno e su cui a tutt'ora non tutti i probabilisti sono d'accordo; cioè se sia necessario o meno specificare sempre lo spazio di probabilità su cui sono definite le variabili aleatorie.

Osservazione 1.1. *Riprendiamo i due esempi precedenti restringendo la nostra attenzione alla sola variabile aleatoria* X. *Essa rappresenta il lancio di un dado ideale, nei due esempi possiamo vedere che essa viene realizzata su due modelli diversi (*Ω *e* Ω_2*), la distribuzione di* X *nel primo e nel secondo esempio chiaramente è la stessa e tutte le proprietà probabilistiche che ne seguono, come per esempio le medie, sono uguali. Adesso il punto chiave è che è possibile considerare il risultato* X *del lancio di un dado ideale come un oggetto astratto e gli spazi* Ω *e* Ω_2 *come dei modelli diversi che realizzano* X. X *rappresenta il risulato del lancio del dado ideale e tutte le proprietà probabilisticamente interessanti non devono dipendere dal modello scelto. È un po' come succede in altri ambiti della matematica, per esempio in geometria, le proprietà di una varietà non devono dipendere dal sistema di carte scelto per rappresentarla, cioè se sto studiando le proprietà della bottiglia di Klein i risultati che mi aspetto non devono dipendere dal sistema di carte scelto per rappresentarla. La bottiglia di Klein è un oggetto astratto mentre un sistema di carte è un modello che realizza l'oggetto astratto. Questo approccio sebbene molto radicale permette di giustificare perché molto spesso si parli di variabili aleatorie e distribuzioni senza citare lo spazio di Probabilità su cui sono definite.*

Alla luce dell'osservazione fatta sopra, durante il resto del corso ci limiteremo a citare lo spazio di probabilità $(\Omega, \mathcal{H}, P)$ su cui stiamo lavorando solo quando può essere utile a rendere più chiaro un enunciato o una dimostrazione.

1.1.1 Valore atteso di una variabile aleatoria discreta

Il valore atteso di una variabile aleatoria X reale discreta è definito

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) P(\omega) \tag{1.1}$$

con la solita convenzione sulle eventuali somme infinite (vedi appendice A.1)

$$\sum_{\omega \in \Omega} X(\omega) P(\omega) := \sum_{\omega \in \Omega} X^+(\omega) P(\omega) - \sum_{\omega \in \Omega} X^-(\omega) P(\omega) \tag{1.2}$$

cioè diremo che X non ammette media solo nel caso in cui al secondo membro della (1.2) si realizzi $+\infty - \infty$.

Osservazione 1.2. *Per passare dall'equazione* (1.1) *all'equazione* (1.2) *è stata utilizzata per la somma la definizione A.2. Tale nuova definizione permette di eseguire le somme anche su insiemi di indici più che numerabili. Inoltre si verifica facilmente (vedi prop.A.1) che nel caso di indici numerabili la somma (definizione A.2) differisce dalla serie solo nei casi in cui l'esito della serie dipenda dall'ordine in cui sono sommati, in tal caso si dirà che l'insieme non ammette somma.*

Il seguente esempio sottilinea la differenza tra serie convergente e somma infinita mettendo in guardia da possibili errori:

Esempio 1.3. *Sia* $(\Omega, \mathcal{H}, P)$ *uno spazio di probabilità con* $\Omega = \mathbb{N}$, $\mathcal{H} = \mathcal{P}(\Omega)$ *e* $P(\{n\}) = \frac{1}{2^n}$. *Sia* X *variabile aleatoria definita da* $X(n) = \frac{(-2)^n}{n}$. *Proviamo a calcolare la media di* X*:*

$$\mathbb{E}[X] = \sum_{n \in \mathbb{N}} n \cdot X(n) = \sum_{n \in \mathbb{N}} \frac{(-1)^n}{n} = \sum_{\substack{n \in \mathbb{N} \\ n \text{ pari}}} \frac{1}{n} - \sum_{\substack{n \in \mathbb{N} \\ n \text{ dispari}}} \frac{1}{n} =$$

Per quanto detto sopra la variabile aleatoria X *non ammette media, nonstante la serie sia convergente*

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n} = l \in \mathbb{R}$$

Un altro strumento importante per operare con le somme è il principio di somma a blocchi (proposizione A.2). Il principio di somma a blocchi stabilisce sotto quali condizioni vale la proprietà associativa nel caso di somme infinite; la sua dimostrazione (lunga e noiosa) può essere trovata in [1].
Grazie al principio di somma a blocchi è possibile mostrare che il valore atteso di una variabile aleatoria X dipende solo dalla sua distribuzione (1.3) e non dallo spazio Ω scelto per realizzare tale distribuzione.

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) P(\omega) = \sum_{x \in \mathbb{R}} \sum_{\substack{\omega \in \Omega \\ X(\omega) = x}} X(\omega) P(\omega) = \sum_{x \in \mathbb{R}} x \sum_{\substack{\omega \in \Omega \\ X(\omega) = x}} P(\omega)$$

e infine

$$\mathbb{E}[X] = \sum_{x\in\mathbb{R}} xP(X = x) \tag{1.3}$$

Infine il seguente esercizio mostra una proprietà elementare delle somme infinite che è necessario conoscere.

Esercizio 1.1.
Sia $\{a_x\}_{x\in\Omega}$ *una famiglia di numeri reali. Sia* $I := \{x \in \Omega | a_x \neq 0\}$.
Se $\{a_x\}_{x\in\Omega}$ *ammette somma finita allora l'insieme* I *è al più numerabile.*

Capitolo 2

Gli spazi misurabili

In questo capitolo introdurremo le σ-algebre e gli spazi misurabili. Si tratta anche in questo caso di un'introduzione succinta poiché molti dei concetti esposti fanno parte di un corso standard di teoria della misura.

Perché introdurre spazi di probabilità non discreti? Supponiamo di essere interessati ad una variabile aleatoria X che scelga un punto a caso dell'intervallo $[0,1]$, e supponiamo che a caso per noi voglia dire che intervalli di lunghezza uguale abbiano la stessa probabilità. Cioè se (a,b) e (c,d) sono due intervalli della stessa lunghezza inclusi in $[0,1]$ allora $P(X \in (a,b)) = P(X \in (c,d))$. E' facile verificare che per una tale variabile X si ha $P(X = x) = 0$ per ogni x. Dunque una tale variabile non può essere una variabile aleatoria discreta. Per poter far si che variabili aleatorie con le proprietà sopra indicate esistano è necessario introdurre le nozioni di insieme misurabile e σ-algebra, inoltre è necessario fornire una definizione di spazio di probabilità che tenga conto di tali nozioni.
In questo capitolo non parlerò di probabilità, tuttavia per avere un'idea intuitiva dei concetti che seguiranno può essere utile pensare che gli insiemi misurabili siano quei sottoinsiemi di Ω a cui vogliamo assegnare una probabilità.

2.1 Spazi misurabili

Definizione 2.1. Dato un insieme Ω, una σ-algebra $\mathcal{H}$ in Ω è una collezione di sottoinsiemi di Ω che soddisfa le seguenti proprietà.

- $\Omega \in \mathcal{H}$ e $\varnothing \in \mathcal{H}$
- Se $A \in \mathcal{H}$ allora $A^c := \Omega \setminus A \in \mathcal{H}$
- Per ogni $\forall A, B \in \mathcal{H}$ si ha $A \cup B \in \mathcal{A}$ e $A \cap B \in \mathcal{A}$

- $\forall \{A_n\}_{n\in\mathbb{N}}$ se $A_n \in \mathcal{H}$ $\forall n \in \mathbb{N}$ allora $\bigcup_{n=1}^{\infty} A_n \in \mathcal{H}$ e $\bigcap_{n=1}^{\infty} A_n \in \mathcal{H}$

gli elementi di $\mathcal{H}$ sono detti insiemi misurabili.

La seconda condizione è detta chiusura rispetto al complementare, la terza invece chiusura rispetto all'unione ed intersezione finita, e la quarta chiusura rispetto all'unione numerabile e all'intersezione numerabile. Osserviamo anche che le ipotesi sono ridondanti per esempio la chiusura rispetto al complementare più la chiusura rispetto all'unione dà la chiusura rispetto all'intersezione. Una collezione di sottoinsiemi di Ω che soddisfa le prime 3 condizione si dirà un'algebra. La proposizione che segue indica delle condizioni sufficienti affinché una collezione di sottoinsiemi sia una σ-algebra.

Proposizione 2.1. *Dato un insieme Ω e una collezione $\mathcal{H}$ di sottoinsiemi di Ω consideriamo le seguenti condizioni:*

1 $\Omega \in \mathcal{H}$

1′ $\varnothing \in \mathcal{H}$

2 *Se $A \in \mathcal{H}$ allora $A^c := \Omega \setminus A \in \mathcal{H}$*

3 $\forall \{A_n\}_{n\in\mathbb{N}}$ *se $A_n \in \mathcal{H}$ $\forall n \in \mathbb{N}$ allora $\cup_{n=1}^{\infty} A_n \in \mathcal{H}$*

3′ $\forall \{A_n\}_{n\in\mathbb{N}}$ *se $A_n \in \mathcal{H}$ $\forall n \in \mathbb{N}$ allora $\cap_{n=1}^{\infty} A_n \in \mathcal{H}$*

Se vale la condizione 2, una condizione tra la 1 e la 1′ e una condizione tra la 3 e la 3′ allora $\mathcal{H}$ è una σ-algebra.

Definizione 2.2. La coppia $(\Omega, \mathcal{H})$ con Ω insieme e $\mathcal{H}$ sua σ-algebra si dirà spazio misurabile.

Esempio 2.1 (σ-algebre)**.** *Dato un insieme non vuoto Ω esso ha almeno due σ-algebre: $\mathcal{H}_1 := \{\Omega, \varnothing\}$ e $\mathcal{H}_2 := \mathcal{P}(\Omega)$. $\mathcal{H}_1$ e $\mathcal{H}_2$ sono rispettivamente la σ-algebra meno fine e la più fine di Ω.*

Esempio 2.2 (σ-algebre)**.** *Dato un insieme non vuoto Ω e una sua partizione $\{A_i\}_{i\in I}$. È possibile costruire una σ-algebra $\mathcal{H}$ in modo tale che sia la più fine tra le σ-algebre che non spezzano gli insiemi della partizione.*

$$\mathcal{H} = \cup_{J\subseteq I}\{\cup_{i\in J} A_i\}$$

Esercizio 2.1. *Dimostrare che se una σ-algebra è finita allora la sua cardinalità è 2^n per qualche $n \in \mathbb{N}$.*

Proprietà (σ-algebra)**.** *Dimostrare le seguenti affermazioni.*

1 *Se $\mathcal{H}_1$ e $\mathcal{H}_2$ sono σ-algebre di Ω allora anche $\mathcal{H}_1 \cap \mathcal{H}_2$ lo è.*

1′ *Se* $\{\mathcal{H}_i\}_{i\in I}$ *è una famiglia di* σ*-algebre di* Ω *allora anche* $\cap_{i\in I}\mathcal{H}_i$ *è una* σ*-algebra.*

2 *Non è vero che per ogni insieme* Ω *e per ogni* $\mathcal{H}_1$ *e* $\mathcal{H}_2$*,* σ*-algebre di* Ω *vale* $\mathcal{H}_1 \cup \mathcal{H}_2$ *è una* σ*-algebra di* Ω *.*

2′ *Date* $\mathcal{H}_1$ *e* $\mathcal{H}_2$*,* σ*-algebre di* Ω *allora*
$\mathcal{H}_1 \cup \mathcal{H}_2$ *è una* σ*-algebra se e solo se* ($\mathcal{H}_1 \subseteq \mathcal{H}_2$ *oppure* $\mathcal{H}_2 \subseteq \mathcal{H}_1$)

3 *Dato* $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ *collezione di sottoinsiemi di* Ω *esiste la più piccola* σ*-algebra di* Ω *contenente* $\mathcal{A}$*. Questa* σ*-algebra sarà detta* σ*-algebra generata da* $\mathcal{A}$ *e sarà indicata con* $\sigma(\mathcal{A})$*.*

3′ *In particolare se* $\mathcal{H}_1$ *e* $\mathcal{H}_2$ *sono* σ*-algebre di* Ω *allora esiste la più piccola* σ*-algebra contenente* $\mathcal{H}_1$ *e* $\mathcal{H}_2$*.*

Proposizione 2.2. *Dato* Ω *insieme non vuoto. Sia* $\mathcal{A}$ *una collezione di sottoinsiemi di* Ω *e sia* P *una proprietà dei sottoinsiemi di* Ω*, (cioè* $P(E) \in$ $\{$*"vero", "falso"*$\}$ *per ogni* $E \subseteq \Omega$*). Se valgono le seguenti condizioni:*

1. $P(\Omega) =$ *"vero" oppure* $P(\varnothing) =$ *"vero".*
2. $P(E) =$ *"vero" per ogni* $E \in \mathcal{A}$*.*
3. *Per ogni* $E \in \sigma(\mathcal{A})$*,* $P(E) =$ *"vero" se e solo se* $P(E^c) =$ *"vero"*
4. *Per ogni* $\{E_n\}_{n\in\mathbb{N}}$ *collezione di elementi di* $\sigma(\mathcal{A})$ *con* $P(E_n) =$ *"vero" per ogni* $n \in \mathbb{N}$ *si ha* $P(\cap_{n\in\mathbb{N}}E_n) =$ *"vero".*

allora per ogni $E \in \sigma(\mathcal{A})$ *si ha* $P(E) =$ *"vero".*

Osservazione 2.3. *La proposizione precedente resta vera se nel punto (4) sostituiamo l'intersezione con l'unione.*

Dimostrazione. Consideriamo la collezione $\mathcal{H}$ di elementi di $\sigma(\mathcal{A})$ per i quali la proprietà è vera

$$\mathcal{H} := \{E \in \sigma(\mathcal{A}) | P(E) = \text{"vero"}\} .$$

Le condizioni (1), (3) e (4) ci dicono che $\mathcal{H}$ soddisfa le ipotesi della proposizione 2.1 quindi $\mathcal{H}$ è una σ-algebra. La condizione (2) ci dice che $\mathcal{H}$ contiene $\mathcal{A}$. Poiché $\sigma(\mathcal{A})$ è la più piccola σ-algebra che contiene $\mathcal{A}$, e $\mathcal{H}$ contiene $\mathcal{A}$ allora si deve avere $\mathcal{H} \supseteq \sigma(\mathcal{A})$ e quindi $P(E) =$ "vero" per ogni $E \in \sigma(\mathcal{A})$. □

Definizione 2.3. Dati due spazi misurabili $(\Omega, \mathcal{H})$ e $(E, \mathcal{E})$ un'applicazione $X : \Omega \to E$ si dirà misurabile da $(\Omega, \mathcal{H})$ a $(E, \mathcal{E})$ se per ogni $A \in \mathcal{E}$ si ha $X^{-1}(A) \in \mathcal{H}$.

La nozione di applicazione misurabile ha molte similitudini con quella di applicazione continua; per le applicazioni misurabili si richiede che la controimmagine di un misurabile sia misurabile mentre per le applicazioni continue si richiede che la controimmagine di un aperto sia aperto. Vale la regola di composizione, se si compongono due applicazioni misurabili si ottiene ancora un'applicazione misurabile.

Esercizio 2.2. *Sia $(E, \mathcal{E})$ uno spazio misurabile. Sia $X : \Omega \to E$ un'applicazione. Dimostrare che la collezione $\mathcal{H}$ di sottoinsiemi di Ω data da*

$$\mathcal{H} := X^{-1}(\mathcal{E}) = \{A \subseteq \Omega | \exists E \in \mathcal{E} \text{ con } X^{-1}(E) = A\}$$

è una σ-algebra. $\mathcal{H}$ è la più piccola σ-algebra di Ω che rende misurabile X.

Enunceremo ora una proposizione che utilizzeremo spesso per dimostrare che una applicazione è misurabile.

Proposizione 2.4. *Siano $(\Omega, \mathcal{H})$ e $(E, \mathcal{E})$ due spazi misurabili. Sia $X : \Omega \to E$ un'applicazione. Sia inoltre $\mathcal{A} \subseteq \mathcal{P}(E)$ tale che $\sigma(\mathcal{A}) = \mathcal{E}$.*
Se per ogni $A \in \mathcal{A}$ vale $X^{-1}(A) \in \mathcal{H}$ allora $X : (\Omega, \mathcal{H}) \to (E, \mathcal{E})$ è misurabile.

La proposizione precedente ci dice che per mostrare che una applicazione X è misurabile è sufficiente verificare che gli elementi di un insieme di generatori di $\mathcal{E}$ abbia controimmagine in $\mathcal{H}$.

Dimostrazione. Per ogni $A \in \mathcal{E}$ consideriamo la proprietà P definita da:

$$P(A) = \begin{cases} \text{“vero''} & \text{se } X^{-1}(A) \in \mathcal{H} \\ \text{“falso''} & \text{altrimenti} \end{cases}$$

La proprietà P soddisfa le condizioni della proposizione 2.2 quindi per ogni $A \in \sigma(\mathcal{A}) = \mathcal{E}$ si ha $X^{-1}(A) \in \mathcal{H}$ e dunque $X : (\Omega, \mathcal{H}) \to (E, \mathcal{E})$ è misurabile. □

Definizione 2.4. Dato uno spazio topologico Ω si chiama σ-algebra dei boreliani, $\mathcal{B}(\Omega)$, la più piccola σ-algebra che contiene gli aperti.

Esercizio 2.3. *Dimostrare che se $X : \mathbb{R}^n \to \mathbb{R}^m$ è un'applicazione continua allora è anche misurabile come applicazione da $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ in $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$.*

Definizione 2.5. Dato un insieme Ω, un sottoinsieme delle parti $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ si dice un p-system se è chiuso per intersezione cioè

$$\forall A, B \in \mathcal{A} \text{ si ha } A \cap B \in \mathcal{A}$$

Importanti esempi di p-system in $\mathbb{R}$ sono i seguenti:

$$\begin{cases} \mathcal{A}_1 := \{\ (a,+\infty)\ |a \in \mathbb{R}\} \\ \mathcal{A}_2 := \{\ (a,b]\ |a,b \in \mathbb{R}\} \\ \mathcal{A}_3 := \{\ (a_1,b_1]\cup\ldots\cup(a_n,b_n]\ |n \in \mathbb{N}, a_1,b_1\ldots a_n,b_n \in \overline{\mathbb{R}}\} \\ \mathcal{A}_4 := \{\ (-\infty,a]\ |a \in \mathbb{R}\} \end{cases}$$

$\mathcal{A}_3$ in realtà è anche un'algebra.

Esercizio 2.4. *Dimostrare che:*

$$\sigma(\mathcal{A}_1) = \sigma(\mathcal{A}_2) = \sigma(\mathcal{A}_3) = \sigma(\mathcal{A}_4) = \mathcal{B}(\mathbb{R})$$

Enunceremo adesso per completezza alcune proprietà di misurabilità delle applicazioni a valori in $(\mathbb{R},\mathcal{B}(\mathbb{R})$. Tali proprietà dovrebbero essere già note agli studenti dai corsi di analisi.

Proposizione 2.5. *Siano X, Y e $\{X_n\}_{n\in\mathbb{N}}$ applicazioni misurabili da uno spazio misurabile $(\Omega,\mathcal{H})$ in $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ sono applicazioni misurabili:*

1. X^+ *e* X^-
2. $X+Y$
3. $X\cdot Y$
4. X/Y *nell'ipotesi* $Y(\omega)\neq 0$ *per ogni* $\omega\in\Omega$
5. $\sup_n X_n$
6. $\inf_n X_n$
7. $\limsup_n X_n$
8. $\liminf_n X_n$

dove $\limsup_n X_n := \inf_n\{\sup_{m>n} X_m\} = \lim_n\{\sup_{m>n} X_m\}$
e $\liminf_n X_n := \sup_n\{\inf_{m>n} X_m\} = \lim_n\{\inf_{m>n} X_m\}$

Dimostrazione.
1) Sia $\mathcal{A}_1$ il p-system definito a pagina 15. Mostreremo che per ogni $A\in\mathcal{A}_1$ vale $X_+^{-1}(A)\in\mathcal{H}$, cosicché grazie all'esercizio 2.4 e alla proposizione 2.4 si avrà X_+ misurabile. Per ipotesi per ogni $a\in\mathbb{R}$ si ha $X^{-1}(a,\infty)\in\mathcal{H}$, per X_+ vale $(X_+)^{-1}(a,\infty) = X^{-1}(a,\infty)$ se $a\geq 0$ e $X_+^{-1}(a,\infty)=\Omega$ se $a<0$ in entrambi i casi $X_+^{-1}(a,\infty)\in\mathcal{H}$. La dimostrazione per X_- è analoga.
2) procediamo come nel caso precedente e mostriamo che $(X+Y)^{-1}((a,\infty))\in\mathcal{H}$. Si verifica facilmente che vale la seguente identità

$$(X+Y)^{-1}((a,\infty)) = \bigcup_{q\in\mathbb{Q}}\left(X^{-1}((q,\infty))\cap Y^{-1}((q-a,\infty))\right)$$

dunque $(X+Y)^{-1}((a,\infty))$ è misurabile. Questo punto potrà essere risolto in maniera più elegante utilizzando l'esercizio 2.6.

6) Sia $X := \sup_{n\in\mathbb{N}} X_n$ procediamo ancora come nel caso 1) e mostriamo che $X^{-1}((a,\infty)) \in \mathcal{H}$. Per $\omega \in \Omega$ vale $X(\omega) > a$ se e solo se esiste $n \in \mathbb{N}$ tale che $X_n(\omega > a$ dunque

$$X^{-1}((a,\infty)) = \bigcup_{n\in\mathbb{N}} X_n^{-1}((a,\infty))$$

per ipotesi $X_n^{-1}((a,\infty))$ appartiene ad $\mathcal{H}$ quindi anche $X^{-1}((a,\infty))$ appartiene ad $\mathcal{H}$

I restanti casi sono lasciati per esercizio. □

2.1.1 σ-algebra ed informazione

Così come negli spazi topologici la topologia è associata ad un concetto intuitivo di forma, in modo simile, negli spazi di probabilità la σ- algebra è associata all'informazione del modello. Capire in che se senso la σ- algebra contenga l'informazione del sistema è cruciale per poter sviluppare la corretta intuizione sugli argomenti che saranno svolti nella seconda parte del corso. Purtroppo però quando ci si avvicina per la prima volta alla probabilità è molto difficile capire in che senso una σ- algebra è associata all'informazione del modello. Il prossimo teorema ci spiega in che senso la σ-algebra associata ad una variabile aleatoria X contenga la sua informazione.

Teorema 2.6. *Sia Ω un insieme, sia $(E,\mathcal{E})$ uno spazio misurabile, sia X un'applicazione da Ω in E e sia $\sigma(X) := X^{-1}(\mathcal{E})$ la più piccola σ-algebra che rende misurabile X. Sia infine Y un'applicazione da $(\Omega,\sigma(X))$ in $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ allora Y è misurabile se e solo se esiste un'applicazione f da $(E,\mathcal{E})$ in $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ misurabile tale che $Y = f\circ X$.*

$$\begin{tikzcd}
(\Omega,\sigma(X)) \arrow[r, "X"] \arrow[dr, "Y"'] & (E,\mathcal{E}) \arrow[d, "f"] \\
& (\mathbb{R},\mathcal{B}(\mathbb{R}))
\end{tikzcd}$$

Dimostrazione. La dimostrazine del *se* è ovvia: Y è misurabile perché è composizione di due applicazioni misurabili. Per dimostrare il *solo se* occorre costruire una funzione f misurabile che fa commutare il diagramma. (La dimostrazione è non banale perchè l'applicazione X potrebbe non essere suriettiva e l'insieme degli elementi di E che non appartengono all'immagine di X potrebbe non essere misurabile.) Osserviamo per cominciare che si può supporre $Y \geq 0$ senza perdere in generalità. Infatti se così non fosse potremmo porre $Y = Y^+ - Y^-$, risolvendo il problema per Y^+ e Y^- otterrei $Y^+ = f^+\circ X$ e $Y^- = f^-\circ X$ e quindi posto $f := f^+ - f^-$ avrei $Y = f\circ X$.

Consideriamo quindi il caso $Y \geq 0$.
Per $q \geq 0$ razionale consideriamo l'applicazione $Y_q := q \cdot \mathbf{1}_{Y \geq q}$,

$$Y_q = \begin{cases} q & \text{se } Y \geq q \\ 0 & \text{altrimenti} \end{cases}$$

tali funzioni soddisfano l'uguaglianza:

$$Y = \sup_{q \in \mathbb{Q}_+} Y_q$$

Facciamo vedere ora che esiste f_q misurabile tale che $Y_q = f_q \circ X$. Poiché Y è $\sigma(X)$ misurabile allora per costruzione lo è anche Y_q, sia $A_q := Y_q^{-1}(\{q\})$, poiché A_q appartiene a $\sigma(X)$ allora esiste B_q in $\mathcal{E}$ tale che $A_q = X^{-1}(B_q)$,

$$\text{ponendo ora} \qquad f_q(e) := \begin{cases} q & \text{se } e \in B_q \\ 0 & \text{altrimenti} \end{cases} \qquad \text{si ha} \qquad Y_q = f_q \circ X$$

ora ricordando che il sup di applicazioni misurabili è misurabile allora:

$$\text{definendo} \qquad f := \begin{cases} \sup_{q \in \mathbb{Q}_+} f_q & \text{se il sup è finito} \\ 0 & \text{altrimenti} \end{cases} \qquad \text{si ha} \qquad Y = f \circ X$$

□

Due osservazioni sulla dimostrazione precedente:
1) se l'applicazione X non è suriettiva, la scelta degli insiemi B_q potrebbe non essere unica e quindi anche l'applicazione f potrebbe non essere unica.
2) Il $\sup_{q \in \mathbb{Q}_+} f_q(e)$ può effettivamente non essere finito, ma tale evenienza può verificarsi solo nel caso in cui e non appartenga all'immagine di X.

2.1.2 d-system e lemma di Dynkin

Definizione 2.6. Dato un insieme Ω, un sottoinsieme delle parti $\mathcal{D} \in \mathcal{P}(\Omega)$ si dice un d-system se soddisfa le seguenti proprietà:

- $\Omega \in \mathcal{D}$
- se $A, B \in \mathcal{D}$ e $A \subseteq B$ allora $B \setminus A \in \mathcal{D}$
- se per ogni $n \in \mathbb{N}$ si ha $A_n \in \mathcal{D}$ e $A_n \subseteq A_{n+1}$ allora $\bigcup A_n \in \mathcal{D}$.

Esercizio 2.5. *Sia Ω un insieme e sia $\mathcal{G} \subseteq \mathcal{P}(\Omega)$ consideriamo le seguenti ipotesi su $\mathcal{G}$:*

(i) $\Omega \in \mathcal{G}$

(ii) Se $A \in \mathcal{G}$ *allora* $A^c \in \mathcal{G}$

(iii) Se $A_n \in \mathcal{G}$ *e* $A_n \cap A_m = \varnothing$ *per ogni* $n \neq m$ *allora* $\cup_{n\in\mathbb{N}} A_n \in \mathcal{G}$

(a) $\Omega \in \mathcal{G}$

(b) Se $A, B \in \mathcal{G}$ *e* $A \subseteq B$ *allora* $B/A \in \mathcal{G}$

(c) Se $A_n \in \mathcal{G}$ *e* $A_n \uparrow A$ *allora* $A \in \mathcal{G}$

Dimostrare che le ipotesi (i), (ii) e (iii) implicano le ipotesi (a), (b) e (c) e viceversa le ipotesi (a), (b) e (c) implicano le ipotesi (i), (ii) e (iii).

Grazie all'esercizio precedente possiamo dare una definizine equivalente di d-system:

Definizione 2.7. Dato un insieme Ω, un sottoinsieme delle parti $\mathcal{D} \in \mathcal{P}(\Omega)$ si dice un d-system se soddisfa le seguenti proprietà:

- $\Omega \in \mathcal{D}$
- se $A \in \mathcal{D}$ allora $B^c \in \mathcal{D}$
- se per ogni $n, m \in \mathbb{N}$ con $n \neq m$ si ha $A_n \in \mathcal{D}$ e $A_n \cap A_m = \varnothing$ allora $\bigcup A_n \in \mathcal{D}$.

Le due definizioni 2.6 e 2.7 sono perfettamente equivalenti. La doppia definizione ci sarà molto utile nelle applicazioni; quando sappiamo che un insieme $\mathcal{D}$ è un d-system potremo utilizzare tutte e sei le condizioni dell'esercizio 2.5 mente quando dovremo dimostrare che $\mathcal{D}$ è un d-system potremo dimostrare solo le tre pi semplici.

Proposizione 2.7. *Una collezione di insiemi* $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ *è una* σ*-algebra se e solo se è sia un p-system che un d-system.*

Dimostrazione. Dimostrare per esercizio. □

Lemma 2.8 (Dynkin)**.** *Se una collezione di insiemi* $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ *è un p-system allora*

$$\sigma(\mathcal{A}) = d(\mathcal{A})$$

Dove $d(\mathcal{A})$ è il più piccolo d-system che contiene $\mathcal{A}$.

Dimostrazione. Ovviamente vale $\sigma(\mathcal{A}) \supseteq d(\mathcal{A})$. Grazie alla proposizione precedente per concludere la dimostrazione basterà mostrare che l'insieme $\mathcal{G} := d(\mathcal{A})$ è chiuso per intersezione finita. Occorre mostrare che se A e B appartengono a $\mathcal{G}$ allora anche $A \cap B$ appartiene a $\mathcal{G}$, il caso in cui sia A che B appartengo anche ad $\mathcal{A}$ è ovvio perchè $\mathcal{A}$ è un p-system. Il trucco per completare la dimostrazione è di studiare prima il caso in cui uno A appartiene a $\mathcal{A}$ e B no, e poi il caso generale con A e B in a $\mathcal{G}$.

Caso 1) Sia A in $\mathcal{A}$ fissato.
Sia $\mathcal{F}_{\mathcal{A}}$ la collezione di insiemi B di $\mathcal{G}$ che soddisfano la condizione $A \cap B \in \mathcal{G}$:

$$\mathcal{F}_{\mathcal{A}} := \{B \in \mathcal{G} | A \cap B \in \mathcal{G}\}$$

sappiamo per ipotesi che $\mathcal{G}$ soddisfa tutte le condizioni dell'esercizio 2.5, mostreremo che $\mathcal{F}_{\mathcal{A}}$ soddisfa le condizioni della definizione 2.7 e contiene $\mathcal{A}$ dunque $\mathcal{F}_A = d(\mathcal{A}) = \mathcal{G}$.
Se $B \in \mathcal{A}$ allora poichè $\mathcal{A}$ è p-system si ha $A \cap B \in \mathcal{A} \subseteq \mathcal{G}$ quindi $\mathcal{A} \subseteq \mathcal{F}_A$
Se $B = \Omega$ allora $A \cap B = A$ e quindi $\mathcal{F}_A$ soddisfa la condizione (i) di 2.5.
Se $B \in \mathcal{F}_A$ allora $A \cap B \in \mathcal{G}$, $A \cap B^c = A \backslash (A \cap B) \in \mathcal{G}$ dunque $\mathcal{F}_A$ soddisfa la condizione (ii) di 2.5.
Se $B_n \in \mathcal{F}_A$ per ogni n e $B_n \cap B_m = \emptyset$ per $n \neq m$ allora per ogni n si ha $B_n \cap A \in \mathcal{G}$, inoltre $A \cap (\bigcup_n B_n) = \bigcup_n (A \cap B_n) \in \mathcal{G}$ quindi $\mathcal{F}_A$ soddisfa anche la condizione (iii) di 2.5.
Questo conclude la dimostrazione del caso 1) ovvero se $A \in \mathcal{A}$ e $B \in \mathcal{G}$ allora $A \cap B \in \mathcal{G}$.

Caso 2) Sia A in $\mathcal{G}$ fissato. La dimostrazione da qui in poi procede come nel caso 1) ed è lasciata per esercizio. □

2.1.3 Prodotto di σ-algebre

Consideriamo una famiglia di spazi misurabili $\{(\Omega_i, \mathcal{F}_i)\}_{i \in I}$. È possibile definire il prodotto di tali spazi misurabili procedendo in maniera analogo a quanto gli studenti hanno già visto in topologia con il prodotto di spazi topologici.

$$(\Omega, \mathcal{F}) = \left(\bigtimes_{i \in I} \Omega_i, \bigotimes_{i \in I} \mathcal{F}_i \right)$$

dove $\times_{i \in I}$ indica il prodotto cartesiano degli insiemi Ω_i, e $\otimes_{i \in I} \mathcal{F}_i$ indica la più piccola σ-algebra di Ω che renda le proiezioni $\pi_i : \Omega \to \Omega_i$ misurabili.

Proposizione 2.9. *Siano $\{(\Omega_i, \mathcal{F}_i)\}_{i \in I}$ spazi misurabili, sia $(\Omega, \mathcal{F}) = (\times_{i \in I} \Omega_i, \otimes_{i \in I} \mathcal{F}_i)$ e siano $\{\mathcal{A}_i\}_{i \in I}$ collezioni di insiemi tali che per ogni $i \in I$ vale $\sigma(\mathcal{A}_i) = \mathcal{F}_i$ allora*

$$\mathcal{F} = \sigma \left(\left\{ A = \bigtimes_{i \in I} A_i \middle| k \in I, A_k \in \mathcal{A}_k, A_i = \Omega_i \textit{ per ogni } i \neq k \right\} \right) \tag{2.1}$$

$$\mathcal{F} = \sigma \left(\left\{ A = \bigtimes_{i \in I} A_i \middle| J \subseteq I \textit{ finito }, A_i \in \mathcal{A}_i \textit{ per ogni } i \in J, A_i = \Omega_i \textit{ per ogni } i \notin J \right\} \right) \tag{2.2}$$

$$\mathcal{F} = \sigma \left(\left\{ A = \bigtimes_{i \in I} A_i \middle| J \subseteq I \textit{ al più numerabile }, A_i \in \mathcal{A}_i \textit{ per ogni } i \in J, A_i = \Omega_i \textit{ per ogni } i \notin J \right\} \right) \tag{2.3}$$

Notare che l'uguaglianza (2.3) in generale non è più vera se si toglie l'ipotesi al più numerabile.

Dimostrazione. L'equivalenza tra le tre equazioni è immediata perchè i generatori della seconda e della terza sono le intersezioni finite o al più numerabili degli insiemi che generano la prima σ algebra. Basta che dimostriamo la prima. Per la proposizione 2.4 $\pi_k : \Omega \to \Omega_k$ è misurabile se e solo se sono misurabili le controimmaggini degli elementi di $\mathcal{A}_k$. Se $A_k \in \mathcal{A}_k$ allora $\pi_k^{-1}(A_k) = \times_{i\in I} A_i$ con $A_i = \Omega_i$ per ogni $i \neq k$. Questi sono esattamente gli insiemi presenti in (2.1). □

Proposizione 2.10. *Sia I un insieme più che numerabile. Siano $\{(\Omega_i, \mathcal{F}_i)\}_{i\in I}$ spazi misurabili e sia $(\Omega, \mathcal{F}) = (\times_{i\in I} \Omega_i, \bigotimes_{i\in I} \mathcal{F}_i)$ allora per ogni $A \in \mathcal{F}$ esiste $J \subseteq I$ al più numerabile tale che:*

$$A \in \sigma\left\{\{\pi_j\}_{j\in J}\right\}$$

Dimostrazione. Dimostrazione per esercizio. □

Proposizione 2.11. *Siano $(\Omega, \mathcal{H})$ e $\{(E_i, \mathcal{E}_i)\}_{i\in I}$ spazi misurabili.*
Sia $(E, \mathcal{E}) := (\times_{i\in I} E_i, \bigotimes_{i\in I} \mathcal{E}_i)$ supponiamo che $\{X_i\}_{i\in I}$ sia una famiglia di applicazioni $X_i : \Omega \to E_i$, e sia $X := (X_i)_{i\in I}$ il blocco delle variabili X_i. $X : \Omega \to E$. Allora sono cose equivalenti:

- *$X : (\Omega, \mathcal{H}) \to (E, \mathcal{E})$ è misurabile*
- *$\forall i \in I \quad X_i : (\Omega, \mathcal{H}) \to (E_i, \mathcal{E}_i)$ è misurabile*

Dimostrazione. Dimostrazione per esercizio. □

Esercizio 2.6. *Dimostrare che vale:*

$$\mathcal{B}\left(\mathbb{R}^2\right) = \mathcal{B}\left(\mathbb{R}\right) \bigotimes \mathcal{B}\left(\mathbb{R}\right)$$

L'esercizio precedente permette di dimostrare in maniera immediata che la somma o il prodotto di applicazioni misurabili è un'applicazione misurabile. In generale se X e Y sono applicazioni misurabili reali e $\Phi : (\mathbb{R}^2, \mathcal{B}(R^2)) \to (E, \mathcal{E})$ è un'applicazione misurabili allora anche $\Phi(X, Y)$ è misurabile.

Esercizio 2.7. *(***) Sia $\{\Omega_i\}_{i\in I}$ una famiglia di spazi metrici completi e separabili, sia $\{\mathcal{T}_i\}_{i\in I}$ la famiglia delle corrispondenti topologie e supponiamo infine che l'insieme degli indici I sia al più numerabile. Dimostrare che posto $\Omega = \times_{i\in I} \Omega_i$ e $\mathcal{T} = \bigotimes_{i\in I} \mathcal{T}_i$ allora vale*

$$\mathcal{B}\left(\Omega\right) = \bigotimes_{i\in I} \mathcal{B}(\Omega_i) \tag{2.4}$$

dove $\mathcal{B}(\Omega)$ e $\mathcal{B}(\Omega_i)$ indicano la σ-algebra dei boreliani ovvero $\mathcal{B}(\Omega) := \sigma(\mathcal{T})$ e $\mathcal{B}(\Omega_i) = \sigma(\mathcal{T}_i)$.

Capitolo 3

Gli spazi di misura

Le misure sono speciali applicazioni che associano agli elementi di una σ-algebra un elemento di $\overline{\mathbb{R}}$. Esistono molti tipi di misure: misure con segno, misure positive, misure σ-finite, misure finite e misure di probabilità. Il nostro corso chiaramente si concentrerà sulle misure di probabilità. Talvolta sarà necessario fare riferimento alla misura di Lebesgue su $\mathbb{R}$, misura σ-finita. Non verranno mai utilizzate le misure con segno.

3.1 Introduzione agli spazi di misura

3.1.1 Funzioni additive e σ-additive

Sia $\mathcal{A}$ un'algebra, sia μ_0 una funzione $\mu_0 : \mathcal{A} \to [0, +\infty]$.
La funzione μ_0 si dice additiva se per ogni A e B in $\mathcal{A}$ vale

$$A \cap B = \varnothing \quad \implies \quad \mu_0(A \cup B) = \mu_0(A) + \mu_0(B)$$

Definizione 3.1. Data $\mathcal{A}$ algebra. Una funzione $\mu_0 : \mathcal{A} \to [0, +\infty]$ si dice σ-additiva se per ogni successione $\{A_n\}_{n\in\mathbb{N}}$ di elementi di $\mathcal{A}$ disgiunti ($\forall j \neq k$ vale $A_j \cap B_k = \varnothing$) tali che $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$ vale:

$$\mu_0\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mu_0(A_n)$$

3.1.2 Definizione di misura

Dato uno spazio misurabile $(\Omega, \mathcal{H})$. Una applicazione $\mu : \Omega \to [0, +\infty]$ è detta misura se è σ-additiva. La tripletta $(\Omega, \mathcal{H}, \mu)$ è detto spazio di misura. Inoltre:

- Una misura μ è detta di probabilità se $\mu(\Omega) = 1$

- Una misura μ è detta finita se $\mu(\Omega) < \infty$

- Una misura μ è detta σ-finita se esiste una successione $\{A_n\}_{n\in\mathbb{N}}$ di elementi di $\mathcal{H}$ tali che per ogni n si ha $\mu(A_n) < \infty$ e inoltre $\bigcup_n (A_n) = \Omega$.

Nel caso di misure σ-finite poiché $\mathcal{H}$ è una σ-algebra allora esiste anche una successione di insiemi misurabili disgiunti (oppure crescenti) $\{B_n\}_{n\in\mathbb{N}}$ di misura finita tali che $\bigcup_n (B_n) = \Omega$.

3.1.3 Quasi sempre, quasi ovunque, quasi certamente, quasi

Sia $(\Omega, \mathcal{H}, \mu)$ uno spazio di misura. una proprietà P degli elementi di Ω si dice vera quasi ovunque, quasi sempre o quasi certamente se:

$$\exists A \in \mathcal{H} \text{ con } \mu(A) = 0 \text{ tale che } \forall \omega \notin A \text{ si abbia } \mathsf{P}(\omega) = \text{“vero”}$$

in questo caso si dirà che per quasi ogni $\omega \in \Omega$ vale $\mathsf{P}(\omega)$.
Sugli spazi di probabilità si può passare al complementare e dire che P è quasi certamente vera se esiste $B \in \mathcal{H}$ con $\mu(B) = 1$ tale che per ogni $\omega \in B$ è vera $\mathsf{P}(\omega)$.

Osservazione 3.1. *Il fatto che una proprietà P sia vera quasi certamente non implica che l'insieme degli ω in cui P è vera sia un insieme misurabile.*

Osservazione 3.2. *A differenza degli operatori 'per ogni' che commutano tra di loro ($\forall x\ \forall y\ \mathsf{P}(x, y)$ è equivalente a $\forall y\ \forall x\ \mathsf{P}(x, y)$) gli operatori 'per quasi ogni' non commutano tra di loro né con gli operatori 'per ogni'. Se $(\Omega, \mathcal{H}, \mu)$ è uno spazio di misura allora in generale le seguenti sono due affermazioni diverse:*

- *per quasi ogni $x \in \Omega$ per quasi ogni $y \in \Omega$ vale $\mathsf{P}(x, y)$*
- *per quasi ogni $y \in \Omega$ per quasi ogni $x \in \Omega$ vale $\mathsf{P}(x, y)$*

L'operatore 'per quasi ogni' in generale non commuta nemmeno con 'per ogni' ma vale il seguente esercizio:

Esercizio 3.1. *Sia $(\Omega, \mathcal{H}, \mu)$ è uno spazio di misura sia E un insieme numerabile (di solito $E = \mathbb{N}$) allora sono cose equivalenti:*

- *per quasi ogni $\omega \in \Omega$, per ogni $n \in \mathbb{E}$ vale $\mathsf{P}(\omega, n)$*
- *per ogni $n \in \mathbb{E}$, per quasi ogni $\omega \in \Omega$ vale $\mathsf{P}(\omega, n)$*

3.2 Un teorema di unicità della misura

Il prossimo teorema è una delle applicazioni principali del lemma di Dynkin.

Teorema 3.3. *Sia $(\Omega, \mathcal{H})$ uno spazio misurabile sia $\mathcal{A}$ un p-system tale che $\sigma(\mathcal{A}) = \mathcal{H}$ e $\Omega \in \mathcal{A}$. Se μ_1 e μ_2 sono due misure finite su $(\Omega, \mathcal{H})$ che coincidono su $\mathcal{A}$ allora μ_1 e μ_2 sono uguali su tutto $\mathcal{H}$.*

Una delle applicazioni più comuni di questo teorema sarà il seguente corollario.

Corollario 3.4. *Due variabili aleatorie reali X e Y hanno la stessa distribuzione se e solo se per ogni intervallo (a,b) vale $P(X \in (a,b)) = P(Y \in (a,b))$.*

Il corollario 3.4 resta valido se si considerano gli intervalli $[a,b]$ chiusi oppure se ci si limita a considerare solo gli intervalli con estremi a e b in $\mathbb{Q}$.

Dimostrazione del teorema 3.3. Cominciamo dalla tesi. Devo mostrare che per ogni $A \in \mathcal{H}$ vale $\mu_1(A) = \mu_2(A)$; sia $\mathcal{G}$ la collezione degli insiemi che soddisfano tale proprietà.

$$\mathcal{G} := \{A \in \mathcal{H} | \mu_1(A) = \mu_2(A)\}$$

Per costruzione $\mathcal{G}$ contiene il p-system $\mathcal{A}$, se mostriamo che $\mathcal{G}$ è un d-system allora con il lemma di Dynkin possiamo concludere $\mathcal{H} \supseteq \mathcal{G} \supseteq d(\mathcal{A}) = \sigma(\mathcal{A}) = \mathcal{H}$. Resta solo da mostrare che $\mathcal{G}$ è un d-system.

(i) $\Omega \in \mathcal{A} \subseteq \mathcal{G}$

(ii) Se $A \in \mathcal{G}$ allora $\mu_1(A) = \mu_2(A)$
quindi $\mu_1(A^c) - \mu_2(A^c) = \mu_1(\Omega) - \mu_1(A) - (\mu_2(\Omega) - \mu_2(A)) = 0$

(iii) Sia $\{A_n\}_{n \in \mathbb{N}}$ una successione di elementi di $\mathcal{G}$ disgiunti, per ogni n vale $\mu_1(A_n) = \mu_2(A_n)$ e quindi

$$\mu_1\left(\bigcup_n A_n\right) = \sum_n \mu_1(A_n) = \sum_n \mu_2(A_n) = \mu_2\left(\bigcup_n A_n\right)$$

□

Esercizio 3.2. *Mostrare con un controesempio che nel teorema 3.3 l'ipotesi $\Omega \in \mathcal{A}$ non può essere rimossa.*

Il teorema 3.3 può essere utilizzato anche per dimostrare che la misura di Lebesgue su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ è unica. La misura di Lebesgue non è finita quindi il teorema 3.3 non può essere applicato in maniera diretta. Affinché la dimostrazione funzioni, è sufficiente spezzare la retta $\mathbb{R}$ in unione di intervalli $[n, n+1)$ e mostrare l'unicità su ciascun intervallo.

3.3 Un teorema di esistenza della misura

Il prossimo teorema è il teorema di estensione di Carathéodory. Si tratta di un teorema che mostra l'esistenza di misure con certe proprietà. L'uso principale per noi è quello di mostrare l'esistenza della misura di Lebesgue. La sua trattazione dettagliata avviene nel corso di Analisi Reale. La sua dimostrazione classica fa uso del metodo di Carathéodory , chi non frequenta il corso di Analisi reale può trovare tale metodo esposto nell'appendice del Williams. Nell'appendice di queste dispense invece ho inserito una dimostrazione alternativa che fa uso delle lemma di Dynkin piuttosto che del metodo di Carathéodory. In tutti i casi la dimostrazione del teorema di estensione è solo per i più curiosi, per questo corso è sufficiente conoscerne l'enunciato.

Teorema 3.5 (Estensione di Carathéodory)**.** *Sia* $(\Omega, \mathcal{H})$ *uno spazio misurabile,* $\mathcal{A}$ *un'algebra tale che* $\sigma(\mathcal{A}) = \mathcal{H}$*. Se* $\mu_0 : \mathcal{A} \to [0, +\infty]$ *è una funzione* σ*-additiva allora esiste una misura* μ *su* $(\Omega, \mathcal{H})$ *che coincide con* μ_0 *su* $\mathcal{A}$*. Inoltre se* $\mu_0(\Omega) < \infty$ *allora tale misura è unica.*

Dal teorema precedente si può dedurre il seguente teorema di esistenza della misura di Lebesgue su $[0, 1]$.

Teorema 3.6. *Esiste una misura* μ *su* $([0, 1], \mathcal{B}([0, 1]))$ *tale che per ogni* $0 \le a \le b \le 1$ *vale* $\mu(a, b) = b - a$*.*

Anche per questo teorema vale quanto detto per quello precedente, la dimostrazine di questo teorema non è richiesta per l'esame, in appendice per i più curiosi può essere trovata una dimostrazione che in parte completa e in parte è alternativa a quella presente sul williams.

3.4 Proprietà degli spazi di probabilità

La seguente proposizione riassume alcune proprietà delle misure di probabilità che dovrebbero essere note.

Proposizione 3.7. *Sia* $(\omega, \mathcal{H}, \mathbb{P})$ *uno spazio di probabilità. Siano* A*,* B *e* $\{A_n\}_{n \in \mathbb{N}}$ *elementi di* $\mathcal{H}$*. Valgolo le seguenti affermazioni:*

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

(b) $\mathbb{P}(\bigcup_{n=1}^{\infty} A_n) \le \sum_{n=1}^{\infty} \mathbb{P}(A_n)$

(c) Se $A_n \uparrow A$ *allora* $\mathbb{P}(A_n) \uparrow \mathbb{P}(A)$

(d) Se $A_n \downarrow A$ *allora* $\mathbb{P}(A_n) \downarrow \mathbb{P}(A)$

Ad eccezione della proprietà (d) le precedenti proprieà valgono anche nel caso di misure infinite.

Dimostrazione. La proprietà (a) si deduce facilmente dalla additività di $\mathbb{P}$ sugli insiemi disgiunti.
Dimostriamo (b) (proprietà subadditiva della misura). Sia $\{B_n\}_{n\in\mathbb{N}}$ la famiglia di eventi definita da: $B_1 := A_1$, $B_2 = A_2 \setminus A_1$, $B_n := A_n \setminus (A_1 \cup \ldots \cup A_{n-1})$.
Gli eventi B_n soddisfano le seguenti proprietà:

1. $\{B_n\}_{n\in\mathbb{N}}$ sono una famiglia di eventi disgiunti
2. $B_n \subseteq A_n$ per ogni n
3. $\cup_{n\in\mathbb{N}} B_n = \cup_{n\in\mathbb{N}} A_n$

dunque

$$P(\cup_{n\in\mathbb{N}} A_n) = P(\cup_{n\in\mathbb{N}} B_n) = \sum_{n\in\mathbb{N}} P(B_n) \leq \sum_{n\in\mathbb{N}} P(A_n)$$

dove le prime due uguaglianze sono dedotte dalle proprietà 3 e 1 mentre la disuguaglianza segue dalla proprietà 2.
Dimostriamo ora (c).
Sia $B_1 = A_1$ e per $n > 1$ sia $B_n = A_n \backslash A_{n-1}$ allora gli insiemi $\{B_n\}_{n\in\mathbb{N}}$ sono disgiunti, $\bigcup_{k=1}^n B_k = A_n$ e $\bigcup_n B_n = A$. Quindi

$$\mathbb{P}(A) = \sum_n \mathbb{P}(B_n) = \lim_n \sum_{k=1}^n \mathbb{P}(B_k) = \lim_n \mathbb{P}(A_n)$$

Infine la dimostrazione di (d) si ottiene passando al complementare ed utilizzando (c). □

Indichiamo la funzione indicatrice di un insieme con la notazione: $\mathbf{1}_A$.
Valgono le seguenti definizione per i limiti sugli insiemi:

$$\begin{aligned}
\limsup_{n\to\infty} A_n = A \quad &\iff \quad \limsup_{n\to\infty} \mathbf{1}_{A_n} = \mathbf{1}_A \\
\liminf_{n\to\infty} A_n = A \quad &\iff \quad \liminf_{n\to\infty} \mathbf{1}_{A_n} = \mathbf{1}_A \\
\lim_{n\to\infty} A_n = A \quad &\iff \quad \lim_{n\to\infty} \mathbf{1}_{A_n} = \mathbf{1}_A
\end{aligned}$$

È necessario però aggiungere qualche spiegazione per capire meglio l'intuizione che c'è sotto le definizioni lim sup e lim inf di una famiglia di eventi. Se non si comprende tale intuizione sarà difficile poi capire come e quando possono essere utilizzati. Sia $(\Omega, \mathcal{H}, \mathbb{P})$ uno spazio di probabilità, sia $\{A_n\}_{n\in\mathbb{N}}$ una famiglia di eventi. Allora:

$$\begin{aligned}
A = \limsup_n A_n = &\bigcap_n \bigcup_{m\geq n} A_n \\
=&\{\omega \in \Omega | \forall m \ \exists n > m \text{ tale che } \omega \in A_n\} \\
=&\{\omega \in \Omega | \omega \in A_n \text{ per infiniti } n\} \\
=&\{\omega \in \Omega | \ \#\{n \in \mathbb{N} | \omega \in A_n\} = \infty\}
\end{aligned}$$

A parole l'evento A è l'evento "si verificano infiniti A_n".
Sia ancora $\{A_n\}_{n\in\mathbb{N}}$ una famiglia di eventi. Allora:

$$\begin{aligned}A = \liminf_n A_n = &\bigcup_n \bigcap_{m\geq n} A_n \\ =&\{\omega \in \Omega | \exists m \text{ tale che } \forall n \geq m \text{ si ha } \omega \in A_n\} \\ =&\{\omega \in \Omega | \ \#\{n \in \mathbb{N} | \omega \notin A_n\} < \infty \ \}\end{aligned}$$

A parole l'evento A è l'evento "da un certo punto in poi A_n si verifica sempre", o anche "A_n si verifica sempre a meno di un insieme finito di indici". Valgono i seguenti lemmi:

Lemma 3.8 (lemma di Fatou per eventi)**.** *Sia $(\Omega, \mathcal{H}, \mathbb{P})$ spazio di probabilità, per ogni famiglia di eventi $\{A_n\}_{n\in\mathbb{N}}$ vale*

$$\mathbb{P}\left(\liminf_n A_n\right) \leq \liminf_n \mathbb{P}(A_n)$$

Dimostrazione. Sia $A := \liminf_n A_n$.

$$\mathbb{P}(A) = \mathbb{P}\left(\liminf_n A_n\right) = \mathbb{P}\left(\bigcup_n \bigcap_{m\geq n} A_n\right)$$

Sia $B_n := \bigcap_{m\geq n} A_n$ allora si ha $B_n \subseteq A_n$ e $B_n \uparrow A$ quindi

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_n B_n\right) = \lim_n \mathbb{P}(B_n) \leq \liminf_n \mathbb{P}(A_n)$$

□

Lemma 3.9 (lemma di Fatou inverso per eventi)**.** *Sia $(\Omega, \mathcal{H}, \mathbb{P})$ spazio di probabilità, per ogni famiglia di eventi $\{A_n\}_{n\in\mathbb{N}}$ vale*

$$\mathbb{P}\left(\limsup_n A_n\right) \geq \limsup_n \mathbb{P}(A_n)$$

Dimostrazione. È sufficiente passare al complementare e applicare il lemma precedente. □

Esercizio 3.3. *Trovare un esempio in cui le disuguaglianze dei due precedenti lemmi di Fatou sono disuguaglianze strette.*

Lemma 3.10 (Primo lemma di Borel-Cantelli)**.** *Sia $(\Omega, \mathcal{H}, \mathbb{P})$ spazio di probabilità, sia $\{A_n\}_{n\in\mathbb{N}}$ una famiglia di eventi. Se vale $\sum_n \mathbb{P}(A_n) < \infty$ allora*

$$\mathbb{P}(\limsup_n (A_n)) = 0$$

Dimostrazione. Sia $A := \limsup_n A_n$.

$$\mathbb{P}(A) = \mathbb{P}\left(\limsup_n A_n\right) = \mathbb{P}\left(\bigcap_n \bigcup_{m\geq n} A_n\right)$$

Sia $B_n := \bigcup_{m\geq n} A_n$ allora si ha $B_n \supseteq A_n$ e $B_n \downarrow A$ quindi

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcap_n B_n\right) =\downarrow \lim_n \mathbb{P}(B_n) \leq \mathbb{P}(B_m) \quad \forall m$$

per la subattività si ha:

$$\mathbb{P}(A) \leq \mathbb{P}(B_m) \leq \sum_{n=m}^{\infty} \mathbb{P}(A_n) \qquad \forall m$$

poichè la serie $\sum_n \mathbb{P}(A_n)$ converge allora si ha

$$\lim_{m\to\infty} \sum_{n=m}^{\infty} \mathbb{P}(A_n) = 0$$

□

Il primo lemma di Borel-Cantelli verrà utilizzato molto spesso durante il corso quindi è importante comprendere bene il significa della sua tesi $P(\limsup_n (A_n)) = 0$.

$$A := \limsup A_n = \{\omega \in \Omega | \#\{n | \omega \in A_n\} = \infty\}$$
$$A^c = \{\omega \in \Omega | \#\{n | \omega \in A_n\} < \infty\}$$

Quindi

$$P(\limsup_n (A_n)) = 0 \qquad \Longleftrightarrow \qquad \text{per quasi ogni } \omega \in \Omega \quad \#\{n|\omega \in A_n\} < \infty \tag{3.1}$$

3.4.1 Legge di una variabile aleatoria

Sia $(\Omega, \mathcal{H}, P)$ uno spazio di probabilita, sia $(E, \mathcal{E})$ uno spazio misurabile e sia $X : (\Omega, \mathcal{H}, P) \to (E, \mathcal{E})$. X induce una misura μ su $(E, \mathcal{E})$ che di solita si denota con $\mu = P(X^{-1}) = X(P)$ ed è definita da:

$$\mu(A) = P(X^{-1}(A)) \qquad \forall A \in \mathcal{E}$$

μ è detta legge di X oppure distribuzione di X.
Sia ora $(F, \mathcal{F})$ un ulteriore spazio misurabile e sia $Y : (E, \mathcal{E}) \to (F, \mathcal{F})$ misurabile. Allora posto $Z = Y \circ X$ e $\mu_2 = \mu(Y^{-1})$ vale:

$$\mu_2 = P(Z^{-1}) \quad \text{ovvero} \qquad (Y \circ X)(P) = Y(X(P))$$

3.4.2 Funzione di ripartizione

Sia X una variabile aleatoria da $(\Omega, \mathcal{H}, P)$ in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ si definisce funzione di ripartizione la seguente funzione $F_X : \mathbb{R} \to \mathbb{R}$:

$$F_X(t) = P(X \leq t)$$

La funzione caratteristica gode delle seguenti quattro proprietà caratterizzanti:

(i) F_X è non decrescente

(ii) F_X è continua a destra

(iii) $\lim_{t\to-\infty} F_X(t) = 0$

(iv) $\lim_{t\to+\infty} F_X(t) = 1$

Verificare che la funzione caratteristica di una v.a. reale soddisfa le condizioni precedenti è immediato. Dimostreremo che vale anche il viceversa se F è una funzione che soddisfa le quattro proprietà enunciate sopra allora esiste una v.a. che ha F come funzione di ripartizione.

Proposizione 3.11. *Sia $F : \mathbb{R} \to \mathbb{R}$ una funzione, se F soddisfa le proprietà* (i),...,(iv) *allora esiste una variabile aleatoria X tale che la sua funzione di ripartizione è uguale ad F.*

Prima di tutto spieghiamo l'idea nel caso speciale in cui F sia strettamente crescente e continua. In questo caso la funzione $F : \mathbb{R} \to (0,1)$ è invertibile, sia $H : (0,1) \to \mathbb{R}$ la sua inversa. La funzione H è a sua volta continua e strettamente crescente. Sia $(\Omega, \mathcal{H}, P) = ((0,1), \mathcal{B}(0,1), m)$ dove m è la misura du Lebesgue. Sia $X := H$, X è non decrescente quindi misurabile inoltre

$$\begin{aligned} F_X(t) = P(X \leq t) =& P(\{\omega \in \Omega | X(\omega) \leq t\}) = && (3.2)\\ =& P(\{\omega \in \Omega | \omega \leq F(t)\}) = && (3.3)\\ =& m((-\infty, F(t)) \cap (0,1)) && (3.4)\\ =& m(0, F(t)) = F(t) && (3.5) \end{aligned}$$

In generale però una funzione non crescente e continua a destra non è invertibile. Cerchiamo allora di costruire una funzione che assomigli più possibile all'inversa di F e che sia monotona. Osserviamo la figura 3.1. Nella parte sinistra della figura è rappresentata il grafico di una funzione di distribuzione F che non è continua e non è strettamente crescente. A sinistra il possibile grafico dell'inversa di F, il grafico dell'inversa si ottiene scambiando le variabili t e y cioè scambiando gli assi e ribaltando il grafico di sinistra. Si nota subito che ci sono due difficoltà nel costruire l'inversa: la prima difficoltà è

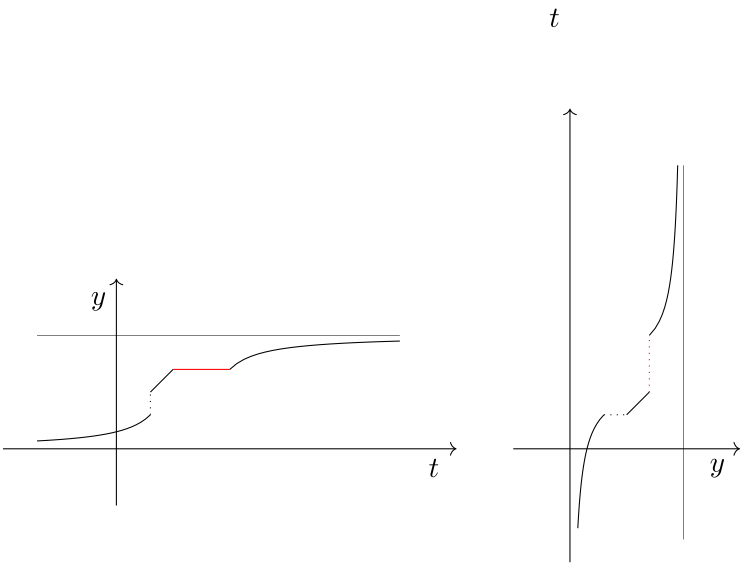

Figura 3.1: Sulla sinistra la funzione F sulla destra la sua inversa.

quella rappresentata dalla linea tratteggiata in nero dove non c'è un'inversa; la seconda difficoltà è costituita dalla linea tratteggiata in rosso dove l'inversa non unica. La prima difficoltà si supera imponendo la monotonia. La seconda difficoltà invece resta ci sono più funzioni monotone che coindono con l'inversa dove questa è unica. Tra queste ce ne sono due priviligiate la funzione q detta quantile che nei casi come quello rappresentato dati puntini in rosso seleziona l'estremo superiore e la funzione H detta pseudoinversa che nei casi in cui l'inversa non sia unica selezione il minimo.

$$q(y) := \inf\{t \in \mathbb{R} | F(t) > y\} = \sup\{t \in \mathbb{R} | F(t) \leq y\}$$
$$H(y) := \inf\{t \in \mathbb{R} | F(t) \geq y\} = \sup\{t \in \mathbb{R} | F(t) < y\}$$

si verifica facilmente che le funzioni q e H sono non decrescenti, la funzione q è continua a destra mentre la funzione H è continua a sinistra, vale $q \geq H$. Le funzione q ed H sono diverse solo nei punti di discontinuità dove la prima è continua a destra mentre la seconda è continua a sinistra. Inoltre poiché i punti di discontinuità di una funzione monotona sono al più numerabili allora anche l'insieme dei punti in cui q e H sono diverse è al più numerabile. Un po' più difficile è verificare che vale la seguente relazione:

$$H(y) \leq \bar{t} \qquad \Longleftrightarrow \qquad y \leq F(\bar{t}) \tag{3.6}$$

Dimostriamo la(3.6).
Se $y \leq F(\bar{t})$ allora $\bar{t} \in \{t \in \mathbb{R} | F(t) > y\}$ quindi $H(y) \leq \bar{t}$.

Se invece $y > F(\bar{t})$ allora poiché F è continua a destra esiste $\varepsilon > 0$ tale che $y > F(\bar{t} + \varepsilon)$ allora $\bar{t} + \varepsilon \notin \{t \in \mathbb{R} | F(t) > y\}$ quindi $H(y) \geq \bar{t} + \varepsilon > \bar{t}$. □

A questo punto grazie alla (3.6) posto $X := H$ vale quanto detto in (3.2).

Funzione di ripartizione per variabili aleatorie in $\overline{\mathbb{R}}$

Talvolta per questioni di compattezza può essere utile lavorare con le variabili aleatorie a valori in $\overline{\mathbb{R}}$. La funzione di ripartizione si definisce ancora come $F_X(t) = P(X \leq t)$ ed è caratterizzata dalle seguenti proprietà:

(i) F_X è non decrescente

(ii) F_X è continua a destra

(iii) $F_X(-\infty) \geq 0$

(iv) $F_X(+\infty) = 1$

Capitolo 4

Integrazione

4.1 Valore atteso di una variabile aleatoria

La definizione di integrale di un'applicazione misurabile di solito è sviluppata durante i corsi di analisi. In queste dispense presenterò per completezza le nozioni di base sensa dimostrazioni. Mi limiterò ad enunciare i teoremi e le definizione che verranno usate durante il corso; per i dettagli rimando gli studenti ad un buon corso di analisi. Sia $(\Omega, \mathcal{H}, \mu)$ uno spazio di misura. Sia $\mathcal{S}$ lo spazio delle funzioni semplici da $(\Omega, \mathcal{H})$ in $(\mathbb{R}, \mathcal{B}(\mathbb{R})$, (funzioni misurabili che assumono solo un numero finito di valori). L'integrale di una funzione semplice si definisce nel seguente modo:

$$\forall Y \in \mathcal{S} \qquad \int_\Omega Y(\omega) d\mu(\omega) = \sum_{x \in \mathbb{R}} x \cdot \mu(\{Y = x\}) \tag{4.1}$$

dove come al solito con $\{S = x\}$ si intende l'insieme: $\{\omega \in \Omega | S(\omega) = x\}$. In generale se X è un'applicazione misurabile non negativa si definisce l'integrale:

$$\int_\Omega X(\omega) d\mu(\omega) := \sup_{\substack{Y \in \mathcal{S} \\ 0 \leq Y \leq X}} \int_\Omega Y(\omega) d\mu(\omega) \tag{4.2}$$

Si dimostra che la definizione (4.2) coincide con la definizione (4.1) quando X è semplice. In pratica l'integrale di X si ottiene approssimando X dal basso con funzioni semplici. Si dimostra inoltre che non è necessario calcolare il *sup* della (4.2) su tutte le Y ma è sufficiente farlo su una successione che tende ad X dal basso vale la seguente proposizione:

Proposizione 4.1. *Sia X da $(\Omega, \mathcal{H}, \mu)$ in $(\mathbb{R}, \mathcal{B}(\mathbb{R})$ misurabile e non negativa, siano Y_n semplici tali che $0 \leq Y_n \leq X$ e $Y_n \uparrow X$ allora*

$$\int_\Omega X(\omega) d\mu(\omega) := \sup_n \int_\Omega Y_n(\omega) d\mu(\omega) \tag{4.3}$$

Data X non negativa è sempre possibile costruire una successione $\{Y_n\}_{n \in \mathbb{N}}$ che soddisfa le condizioni della proposizione precedente, se per esempio

$\{q_n\}_{n\in\mathbb{N}}$ è una successione dei numeri razionali positivi con $q_0 = 0$ allora si può porre

$$Y_n := \sup\{q_k | 0 \le k \le n, q_k \le X\} \tag{4.4}$$

La definizione generale dell'integrale di una funzione non necessariamente positiva è data da:

$$\int_\Omega X(\omega) d\mu(\omega) = \int_\Omega X^+(\omega) d\mu(\omega) - \int_\Omega X^-(\omega) d\mu(\omega) \tag{4.5}$$

con la solita convenzione sulla somma degli infiniti. A partire dalla definizione (4.5) e (4.2) è possibile ottenere varie proprietà: monotonia:

$$X \le Y \qquad \Longrightarrow \int_\Omega X(\omega) d\mu(\omega) \le \int_\Omega Y(\omega) d\mu(\omega)$$

linearità (nell'ipotesi che X e Y ammettano integrali finiti):

$$\int_\Omega \lambda_1 X(\omega) + \lambda_2 Y(\omega) d\mu(\omega) = \lambda_1 \int_\Omega X(\omega) d\mu(\omega) + \lambda_2 \int_\Omega Y(\omega) d\mu(\omega)$$

Se $\mu = P$ è una misura di probabilità allora scriveremo:

$$\mathbb{E}[X] = \int_\Omega X(\omega) dP(\omega)$$

Teorema 4.2 (Regola di composizione)**.** *$(\Omega, \mathcal{H}, \mu)$ uno spazio di misura. Sia ϕ misurabile da $(\Omega, \mathcal{H})$ nello spazio misurabile $(E, \mathcal{E})$. Sia $\nu := \phi(\mu)$. Sia infine $X : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ misurabile e $Y = X(\phi)$. Se X è integrabile vale:*

$$\int_E X d\nu = \int_\Omega Y d\mu \tag{4.6}$$

Dimostrazione. La dimostrazione procede con un metodo standard, si dimostra prima la tesi per il caso X funzione indicatrice poi si studia il caso X funzione semplice e infine si passa al caso generale.
Consideriamo per cominciare il caso $X = \mathbf{1}_A$ con $A \in \mathcal{E}$.

$$\int_E X d\nu = \nu(A) = \mu(\phi^{-1}(A)) = \int_\Omega \mathbf{1}_{\phi^{-1}(A)} d\mu = \int_\Omega Y d\mu$$

Caso X funzione semplice. Se X è una funzione semplice allora è combinazione lineare di funzioni indicatrici. Allora dalla linearità degli integrale della (4.6) si deduce la tesi.
Caso X funzione misurabile positiva. Se X_n sono funzioni semplici positive con $X_n \uparrow X$ allora posto $Y_n := X_n(\phi)$ si ha $Y_n \uparrow Y$. La tesi segue dalla proposizione 4.1. □

4.2 Teoremi limiti

Ora enunceremo tre importanti teoremi limiti di analisi e per completezza inserirò anche la dimostrazione.

Teorema 4.3 (Convergenza monotona)**.** *Sia $(\Omega, \mathcal{H}, \mu)$ uno spazio di misura. Sia X_n e X variabili aleatorie quasi certamente non negative. Se X_n è non decrescente in n e tende ad X quasi certamente allora*

$$\int_\Omega X_n d\mu \uparrow \int_\Omega X d\mu$$

se μ è una misura di probabilità il teorema precedente diventa

$$\begin{cases} 0 \le X_n \le X \\ X_n \uparrow X \quad \text{q.c.} \end{cases} \implies \mathbb{E}[X_n] \uparrow \mathbb{E}[X]$$

Osservazione 4.4. *L'ipotesi $X_n \geq 0$ del teorema di convergenza monotona può essere eventualemtne indebolita in $X_n \geq Y$ per ogni n e Y v.a. che ammette media finita.*

Esercizio 4.1. *Dimostrare l'osservazione precedente. Sia Y v.a. tale che $\mathbb{E}[Y^-] < \infty$ allora vale:*

$$\begin{cases} Y \le X_n \le X \\ X_n \uparrow X \quad q.c. \end{cases} \implies \mathbb{E}[X_n] \uparrow \mathbb{E}[X]$$

Adesso dimostreremo il teorema di convergenza monotona, la sua dimostrazione sarà molto semplice. Ma sia chiaro, la sua dimostrazione è semplice solo perchè abbiamo nascosto gli aspetti più noiosi all'interno della dimostrazione della proposizione 4.1.

Dimostrazione teorema 4.3.
Sia $\{q_n\}_{n\in\mathbb{N}}$ una numerazione dei numeri razionali positivi con $q_0 = 0$. Definiamo ora le seguenti v.a. $Y_{n,m}$ come in (4.4)

$$Y_{n,m} := \sup\{q_k | 0 \le k \le m, q_k \le X_n\}$$

Le funzioni $Y_{n,m}$ sono semplici e (debolmente) crescenti sia in n che in m. Vale

$$Y_{n,m} \xrightarrow[m\to\infty]{} X_n \quad \text{ e } \quad Y_{n,n} \xrightarrow[n\to\infty]{} X$$

Per la proprietà monotona anche i loro integrali sono crescenti sia in n che in m. Questo vuol dire che posso effettuare il limite per n e m che vanno all'infinito e l'esito non dipende da come faccio tendere n ed m all'infinito.

$$\lim_{n\to\infty} \int_\Omega X_n d\mu = \lim_{n\to\infty} \lim_{m\to\infty} \int_\Omega Y_{n,m} d\mu = \lim_{n\to\infty} \int_\Omega Y_{n,n} d\mu = \int_\Omega X d\mu$$

Dove per la prima e terza uguaglianza ho utilizzato la proposizione 4.1. □

Teorema 4.5 (Lemma di Fatou)**.**
Sia X_n una successione di funzioni misurabili quasi certamente non negative. Vale:

$$\int_\Omega \liminf_{n\to\infty} X_n d\mu \leq \liminf_{n\to\infty} \int_\Omega X_n d\mu$$

Dimostrazione.

$$\int_\Omega \liminf_{n\to\infty} X_n d\mu = \int_\Omega \sup_n \inf_{k\geq n} X_k d\mu = \sup_n \int_\Omega \inf_{k\geq n} X_k d\mu$$

$$\leq \sup_n \inf_{k\geq n} \int_\Omega X_k d\mu = \liminf_n \int_\Omega X_k d\mu$$

dove il primo e l'ultimo uguale sono la definizione di $\liminf$, il secondo uguale segue dal teorema di convergenza monotona mentre la disuguaglianza è dovuta alla proprietà monotona per cui per ogni $k \geq n$ vale $\int_\Omega X_k d\mu \geq \int_\Omega \inf_{k\geq n} X_k d\mu$. □

L'ipotesi $X_n \geq 0$ quasi certamente può in realtà essere indebolita.

Teorema 4.6 (Lemma di Fatou bis)**.**
Sia Y funzione misurabile tale che $\int_\Omega Y^- d\mu < \infty$. Sia $\{X_n\}_{n\in\mathbb{N}}$ una successione di applicazioni misurabili. Se per ogni n vale $X_n \geq Y$ quasi certamente allora

$$\int_\Omega \liminf_{n\to\infty} X_n d\mu \leq \liminf_{n\to\infty} \int_\Omega X_n d\mu$$

La dimostrazione del precedente teorema è immediata, basta definire le variabili $\widetilde{X_n} := X_n - Y$ ed applicare ad esse il teorema 4.5.

Teorema 4.7 (Lemma di Fatou inverso)**.**
Sia Y funzione misurabile tale che $\int_\Omega Y^+ d\mu < \infty$. Sia $\{X_n\}_{n\in\mathbb{N}}$ una successione di applicazioni misurabili. Se per ogni n vale $X_n \leq Y$ quasi certamente allora

$$\int_\Omega \limsup_{n\to\infty} X_n d\mu \geq \limsup_{n\to\infty} \int_\Omega X_n d\mu$$

Anche in questo caso la dimostrazione è immediata basta definire le applicazioni $\widetilde{X_n} := -X_n$ e $\widetilde{Y} := -Y$ ed applicare il teorema 4.6.

Teorema 4.8 (Convergenza dominata)**.**
Sia Y funzione misurabile tale che ammette integrale finito (ovvero vale $\int_\Omega |Y| d\mu < \infty$). Siano X e $\{X_n\}_{n\in\mathbb{N}}$ applicazioni misurabili. Se per ogni n vale $|X_n| \leq Y$ quasi certamente allora

$$X_n \xrightarrow[n\to\infty]{} X \quad q.c. \qquad \Longrightarrow \qquad \int_\Omega X_n d\mu \xrightarrow[n\to\infty]{} \int_\Omega X d\mu$$

Dimostrazione. È sufficiente applicare i due lemmi precedenti 4.7 e 4.6

$$\limsup_{n\to\infty} \int_\Omega X_n d\mu \leq \int_\Omega X d\mu \leq \liminf_{n\to\infty} \int_\Omega X_n d\mu$$

□

Esercizio 4.2. *Siano X, Y, $\{X_n\}_{n\in\mathbb{N}}$ applicazioni misurabili tali che per ogni n vale $Y \le X_n \le X$. Se $\int_\Omega Y^- d\mu < \infty$ allora vale*

$$X_n \xrightarrow[n\to\infty]{q.c.} X \qquad \Longrightarrow \qquad \int_\Omega X_n d\mu \xrightarrow[n\to\infty]{} \int_\Omega X d\mu$$

Esercizio 4.3. *Costruire un esempio di variabili aleatorie $\{X_n\}_{n\in\mathbb{N}}$ e X quasi certamente non-negative, tali che:*

- $X_n, \; X \in L^1$
- $X_n \xrightarrow[n\to\infty]{q.c.} X$
- $\lim_n \mathbb{E}[X_n] \neq \mathbb{E}[X]$

Esercizio 4.4. *Siano $\{X_n\}_{n\in\mathbb{N}}$ e X variabili aleatorie in L^1 quasi certamente non-negative tali che $X_n \xrightarrow[n\to\infty]{q.c.} X$. Dimostrare che*

$$\liminf_n \mathbb{E}[X_n] \ge \mathbb{E}[X]$$

4.3 Variabili aleatorie discrete e assolutamente continue

Questa parte è una sintesi di quanto dovrebbe essere stato acquisito dal corso di probabilità del primo anno.

Variabili aleatorie discrete

Una variabile aleatoria X da uno spazio di probabilità $(\Omega, \mathcal{H}, P)$ in $(E, \mathcal{E})$ si dice discreta se esiste un insieme $A \in \mathcal{E}$ al più numerabile tale che $P(X \in A) = 1$. In questo caso si ha:

$$\sum_{e\in E} P(X = e) = 1$$

Data $g : (E, \mathcal{E}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ vale

$$\mathbb{E}[g(X)] = \sum_{e\in E} P(X = e) \cdot g(e) = \sum_{t\in\overline{\mathbb{R}}} P(g(X) = t) \cdot t$$

con le solite regole sulla somma degli infiniti.

Variabili aleatorie assolutamente continue

Introdurremo ora la definizione di variabile aleatoria assolutamente continua a valori in uno spazio di misura genirico $(E, \mathcal{E}, \mu)$. Sia chiaro che in tutte le nostre applicazioni sarà $(E, \mathcal{E}, \mu) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), m)$ e la maggior parte delle

volte sarà $n = 1$.
Sia $(\Omega, \mathcal{H}, P)$ uno spazio di probabilità e $(E, \mathcal{E}, \mu)$ uno spazio di misura. Una variabile aleatoria X da $(\Omega, \mathcal{H})$ in $(E, \mathcal{E})$ si dice assolutamente continua (rispetto a μ) se esiste una funzione misurabile $f_X : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ detta densità di X tale che per ogni $A \in \mathcal{E}$ vale

$$\int_A f_X(x) d\mu = P(X \in A)$$

Vale la seguente caratterizzazione delle funzioni di densità

$$f \geq 0 \quad q.c. \qquad \int_E f(x) d\mu = 1 \tag{4.7}$$

La necessità di (4.7) è immediata, la seconda condizione segue da $\int_E f(x) d\mu = P(X \in E) = 1$ la prima segue dal seguente esercizio.

Esercizio 4.5. *Sia f è una funzione misurabile quasi certamente maggiore o uguale a zero (o quasi certamente minore o uguale). Allora vale:*

$$\int_E f(x) dx = 0 \qquad \Longleftrightarrow \qquad f = 0 \quad q.c.$$

La versione probabilistica dell'esercizio precedente è:

Esercizio 4.6. *Sia X è una variabile aleatoria quasi certamente maggiore o uguale a zero (o quasi certamente minore o uguale a zero). Allora vale:*

$$\mathbb{E}[X] = 0 \qquad \Longleftrightarrow \qquad X = 0 \quad q.c.$$

Svolgimento Esercizio 4.5. Consideriamo il caso $f \geq 0$. L'implicazione $\Longleftarrow$ è ovvia. Vediamo ora la seconda implicazione. Sia $\varepsilon > 0$ allora $f \geq \varepsilon \mathbf{1}_{f > \varepsilon}$ quindi

$$\forall \varepsilon > 0 \qquad 0 = \int_{\mathbb{R}^n} f(x) dx \geq \int_{\mathbb{R}^n} \varepsilon \mathbf{1}_{f > \varepsilon} dx = \varepsilon \cdot m(\{f > \varepsilon\}) \geq 0$$

Quindi per ogni $\varepsilon > 0$ vale $m(\{f > \varepsilon\}) = 0$.

Dimostriamo ora la caratterizzazione(4.7). Dobbiamo dimostrare che se f soddisfa (4.7) allora esiste uno spazio di probabilità $(\Omega, \mathcal{H}, P)$ ed una variabile aleatoria X che ammette f come funzione di densità. Si procede con un trucco si pone $\Omega := E$, $\mathcal{H} = \mathcal{E}$ e $P(A) := \int_A f(x) d\mu$ si verifica facilmente che $(\Omega, \mathcal{H}, P)$ è una spazio di probabilità e $X(\omega) := \omega$ ammette banalmente f come densità.

Proposizione 4.9. *Sia X una varibile aleatoria da $(\Omega, \mathcal{H}, P)$ in $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), m)$ assolutamente continua con densità f_X. Sia g misurabile da $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Se $g(X)$ ammette media allora vale:*

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}^n} f_X(x) g(x) dx$$

Dimostrazione. Sia $Y := g(X)$, la tesi diventa:

$$\int_\Omega Y(\omega)dP(\omega) = \int_{\mathbb{R}^n} f_X(x)g(x)dx$$

Sia $\mu := P(X^{-1})$ per la regola di composizione si ha:

$$\int_\Omega Y(\omega)dP(\omega) = \int_{\mathbb{R}^n} g(x)d\mu(x)$$

la tesi diventa:

$$\int_{\mathbb{R}^n} g(x)d\mu(x) = \int_{\mathbb{R}^n} f_X(x)g(x)dx \tag{4.8}$$

Per dimostrare la (4.8) si procede applicando il metodo standard su g prima si considera il caso g funzione indicatrice poi g semplice infine g positiva.
CASO g semplice: sia $g := \mathbf{1}_A$ con A boreliano di $\mathbb{R}^n$ allora di ha

$$\int_{\mathbb{R}^n} g(x)d\mu(x) = \mu(A) = P(X \in A) = \int_{\mathbb{R}^n} f_X(x)\mathbf{1}_A dx = \int_{\mathbb{R}^n} f_X(x)g(x)dx$$

dove la seconda uguaglianza segue dalla definizione di funzione di densità.
CASO g semplice: segue immediatamente dal caso precedente e dal fatto che i due integrali della (4.8) sono lineari in g.
CASO g positiva: poiché entrambi gli argomenti degli integrali dell'equazione (4.8) sono monotoni in g possiamo applicare il teorema di convergenza monotona a destra e sinistra ad una successione monotona di funzioni semplici che approssimano g. □

4.4 Disuguaglianza di Markov

Sia X una variabile aleatoria non negativa che ammette media finita. Siamo interessati a stimare la probabilità che X sia maggiore di una costante $c > 0$, conoscendo il valore medio $\mathbb{E}[X]$

$$\mathbb{E}[X] \geq \mathbb{E}[X\mathbf{1}_{X\geq c}] \geq \mathbb{E}[c\mathbf{1}_{X\geq c}] = cP(X \geq c) \tag{4.9}$$

quindi si ottiene la seguente disuguaglianza di Markov:

$$P(X > c) \leq \frac{\mathbb{E}[X]}{c}$$

se g è una funzione positiva crescente con $\mathbb{E}[g(X)] < \infty$ allora l'argomento utilizzato in (4.9) può essere ripetuto su $g(X)$:

$$\mathbb{E}[g(X)] \geq \mathbb{E}[g(X)\mathbf{1}_{g(X)\geq g(c)}] \geq \mathbb{E}[g(c)\mathbf{1}_{g(X)\geq g(c)}] = g(c)P(X \geq c)$$

da cui si ottiene:

$$P(X > c) \leq \frac{\mathbb{E}[g(X)]}{g(c)} \tag{4.10}$$

la formula (4.10) ha il vantaggio di essere valida anche nei casi in cui la v.a. X non sia positiva o non ammetta media finita.

4.5 Disuguaglianza di Jensen

La disuguaglianza di Jensen è una disuguaglianza che riguarda i valori attesi di funzioni convesse.

Definizione 4.1. Un insieme $C \subseteq \mathbb{R}^n$ si dice convesso se per ogni $x, y \in C$ e per ogni $\mu, \lambda \in (0,1)$ con $\mu + \lambda = 1$ si ha $z := \mu x + \lambda y \in C$.

L'intersezione di insiemi convessi è ancora un insieme convesso. Dato un insieme $A \in \mathbb{R}^n$ il più piccolo insieme convesso che contiene A si dice l'inviluppo convesso di A.

Definizione 4.2. Dato C un sottoinsieme convesso di $\mathbb{R}^n$. Una funzione $\Phi : C \to \mathbb{R}$ si dice convessa se per ogni $x, y \in C$ e per ogni $\mu,\ \lambda \in (0,1)$ con $\mu + \lambda = 1$ si ha:

$$\Phi(\mu x + \lambda y) \leq \mu\Phi(x) + \lambda\Phi(y)$$

Esercizio 4.7. *Una funzione convessa Φ su un insieme aperto C è continua.*

Valore atteso in $\mathbb{R}^n$

Sia $X = (X_1, X_2, \ldots, X_n)$ una variabile aleatoria a valori in $\mathbb{R}^n$ tale che $\mathbb{E}[|X|] < \infty$. Il valore atteso di X si definisce nel seguente modo.

$$\mathbb{E}[X] := (\mathbb{E}[X_1], \mathbb{E}[X_2], \ldots, \mathbb{E}[X_n])$$

Il valore atteso definito in questo modo è lineare, se A è una matrice possiamo definire il prodotto matrice per vettore nel modo usuale e si verifica facilmente che vale:

$$\mathbb{E}[AX^t] = A\mathbb{E}[X^t]$$

questo ci assicura che il valore atteso è tensoriale ovvero se O è la matrice associata ad un cambio di base allora vale

$$O^{-1}\mathbb{E}[OX^t] = \mathbb{E}[X^t]$$

Valore atteso su un insieme convesso di R^n

Sia $C \subseteq \mathbb{R}^n$ convesso. Allora valgono i seguenti lemmi:

Lemma 4.10 (versione 1)**.** *Sia C un chiuso convesso di $\mathbb{R}^n$, sia P un punto di $\mathbb{R}^n$ che non appartiene a C. Esiste un piano che separa P e C ovvero esiste un'applicazione affine Φ con $\Phi(x) = a \cdot x + b$, $a \in \mathbb{R}^n$ e $b \in \mathbb{R}$ tale che:*

$$\Phi(P) < 0 \qquad e \qquad \Phi(x) > 0 \quad \forall x \in C \tag{4.11}$$

Lemma 4.11 (versione 2)**.** *Sia C un convesso di $\mathbb{R}^n$, sia P un punto di $\mathbb{R}^n$ che non appartiene a C. Esiste un piano che separa P e C ovvero esiste un'applicazione affine Φ con $\Phi(x) = a \cdot x + b$, $a \in \mathbb{R}^n$ e $b \in \mathbb{R}$ tale che:*

$$\Phi(P) \leq 0 \qquad e \qquad \Phi(x) \geq 0 \quad \forall x \in C \tag{4.12}$$

Nella prima versione l'ipotesi C insieme chiuso permette di escludere il caso P appartiene alla frontiera di C e quindi da luogo alle disuguaglianze strette. Il lemma 4.11 vale per C non chiuso, in questo caso se P appartiene alla frontiera di C le disuguaglianze non possono essere strette.

Teorema 4.12. *Sia $C \subseteq \mathbb{R}^n$ un insieme chiuso e convesso. Sia X una variabile aleatoria a valori in C di media finita. Allora vale $\mathbb{E}[X] \in C$.*

Dimostrazione. Sia $P := \mathbb{E}[X]$. Supponiamo per assurdo che P non appartenga a C. Allora per il lemma 4.10 esistono $a \in \mathbb{R}^n$ e $b \in \mathbb{R}$ tali che

$$\Phi(P) < 0 \qquad \text{e} \qquad \Phi(x) > 0 \quad \forall x \in C$$

allora vale:

$$\Phi(P) = a \cdot \mathbb{E}[X] + b = \mathbb{E}[a \cdot X] + b = \mathbb{E}[a \cdot X + b] = \mathbb{E}[\Phi(X)]$$

La condizione $\Phi > 0$ su C implica $\mathbb{E}[\Phi(X)] > 0$ che contraddice $\Phi(P) < 0$. □

Teorema 4.13 (Disuguaglianza di Jensen)**.** *Sia X una v.a. in R_n, sia Φ un'applicazione convessa su $\mathbb{R}^n$. Se X e $\Phi(X)$ ammettono media finita allora vale:*

$$\Phi(\mathbb{E}[X]) \leq \mathbb{E}[\Phi(X)]$$

Dimostrazione. Sia $Y := (X, \Phi(X))$ variabile aleatoria a valori in R^{n+1} e sia $C \subseteq \mathbb{R}^{n+1}$ l'insieme seguente:

$$C := \left\{(x_1, \ldots, x_n, x_{n+1}) \in \mathbb{R}^{n+1} | x_{n+1} \geq \Phi((x_1, \ldots, x_n))\right\}$$

L'insieme C è chiuso perché la funzione Φ è continua mentre dalla convessità di Φ si deduce la convessità di C. La tesi segue immediatamente dal teorema 4.12. □

4.6 Spazi L^p.

In questo paragrafo vedremo due applicazioni della disuguaglianza di Jensen: la monotonia degli spazi L^p (4.13) e il teorema di Holder.

Definizione 4.3. Sia $(\Omega, \mathcal{H}, \mu)$ uno spazio di misura. Sia X misurabile da $(\Omega, \mathcal{H})$ in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ e $p \geq 1$ diremo che X è in $\mathcal{L}^p$ se:

$$\int_\Omega |X|^p \, d\mu < \infty$$

Indicheremo inoltre con $L^p := \mathcal{L}^p / \mathcal{R}$ lo spazio che si ottiene quozientando per la relazione di equivalenza:

$$X \sim_{\mathcal{R}} Y \qquad \Longleftrightarrow \qquad X = Y \; q.c.$$

Poiché $x+y<2\max(x,y)$ allora $|X+Y|^p \leq 2^p \max(|X|^p,|X|^p) \leq 2^p(|X|^p+|X|^p)$ da cui si deduce che L^p è uno spazio vettoriale. Definiamo ora la seguente funzione che (in seguito grazie alla disuguaglianza di Minkowski) sarà una norma per gli spazi L^p:

$$\|X\|_p := \mathbb{E}[|X|^p]^{\frac{1}{p}}$$

4.6.1 Monotonia degli spazi L^p

Proposizione 4.14. *Sia $(\Omega,\mathcal{H},P)$ uno spazio di probabilità. Se $1<p<q$ allora*

$$L^p(\Omega,\mathcal{H},P) \supseteq L^q(\Omega,\mathcal{H},P) \tag{4.13}$$

$$\|X\|_p \leq \|X\|_q \tag{4.14}$$

Si può dimostrare che l'inclusione (4.13) vale anche se lo spazio di misura non è uno spazio di probabilità ma solo uno spazio di misura finito. se μ è una misura finita la (4.14) diventa:

$$\|X\|_{L^p(\Omega,\mathcal{H},\mu)} \leq \|X\|_{L^q(\Omega,\mathcal{H},\mu)}$$

ORa dimostreremo la proposizione precedente nel caso $\|X\|_p, \|X\|_q < \infty$ lasciando per esercizio il caso $\|X\|_p = \infty$.

Dimostrazione. Sia X misurabile da $(\Omega,\mathcal{H},P)$ in $(\mathbb{R},\mathcal{B}(\mathbb{R}))$. La tesi è:

$$\mathbb{E}[|X|^p]^{\frac{1}{p}} \leq \mathbb{E}[|X|^q]^{\frac{1}{q}}$$

$$\Updownarrow$$

$$\mathbb{E}[|X|^p]^{\frac{1}{p}} \leq \mathbb{E}[(|X|^p)^{\frac{q}{p}}]^{\frac{1}{q}}$$

Applicando ora la disuguglianza di Jensen alla funzione convessa $\Phi(x)=|x|^{\frac{q}{p}}$ si ottiene $\Phi(\mathbb{E}[|X|^p]) \leq \mathbb{E}[\Phi(|X|)]$ quindi

$$\mathbb{E}[|X|^p]^{\frac{1}{p}} \leq \mathbb{E}[(|X|^p)^{\frac{q}{p}}]^{\frac{1}{q}}$$

$$\Uparrow$$

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

□

4.6.2 Completezza

Una successione di variabili aleatorie $\{X_n\}_{n\in\mathbb{N}}$ si dice di Cauchy rispetto alla norma $\|\cdot\|_{L^p}$ se per ogni $\epsilon>0$ esiste $\overline{n}$ tale che per ogni $n,m \geq \overline{n}$ vale $\|X_n - X_m\|_p < \epsilon$.

Proposizione 4.15. *Le successioni di Cauchy in L^p convergono. Ovvero se $\{X_n\}_{n\in\mathbb{N}}$ è una successione di Cauchy in L^p allora esiste $X \in L^p$ tale che $\|X_n - X\|_p \xrightarrow[n\to\infty]{} 0$.*

Dimostrazione. La dimostrazine consta di due parti prima occorre determinare la v.a. X e poi mostrare che $\|X_n - X\|_p \xrightarrow[n\to\infty]{} 0$.

Poiché la successione $\{X_n\}_{n\in\mathbb{N}}$ è una successione di Cauchy allora per ogni $k \in \mathbb{N}$ esiste n_k tale che

$$\forall n, m \geq n_k \qquad \mathbb{E}[|X_n - X_m|^p] < \frac{1}{2^{2pk}}$$

Mostreremo ora che $\{X_{n_k}\}_{k\in\mathbb{N}}$ converge quasi certamente e indicheremo con X tale limite. Sia

$$A_k := \left\{\omega \in \Omega : |X_{n_k}(\omega) - X_{n_{k-1}}(\omega)| > \frac{1}{2^k}\right\}$$

applicando la disuguaglianza di Markov si ottiene

$$P(A_k) = P\left(|X_{n_k} - X_{n_{k-1}}|^p > \frac{1}{2^{pk}}\right) \leq \frac{\mathbb{E}[|X_n - X_m|^p]}{\frac{1}{2^{pk}}} \leq \frac{\frac{1}{2^{2pk}}}{\frac{1}{2^{pk}}} = \frac{1}{2^{pk}}$$

Applicando il primo lemma di Borel-Cantelli si ottiene che per quasi ogni $\omega \in \Omega$ esiste $\overline{k}$ tale che per ogni $k \geq \overline{k}$ si ha $\omega \notin A_k$. Poiché

$$X_{n_k}(\omega) = X_0(\omega) + \sum_{i=1}^{k}(X_{n_k}(\omega) - X_{n_{k-1}}(\omega))$$

e la serie è assolutamente convergente ovvero: $\sum_{i=1}^{k} |X_{n_k}(\omega) - X_{n_{k-1}}(\omega)| \leq \sum_{i=1}^{k} \frac{1}{2^k} < 1$ allora $\{X_{n_k}\}_{k\in\mathbb{N}}$ ammette limite quasi certo, sia X tale limite. Dobbiamo ora mostrare che $\mathbb{E}[|X_n - X|^p] \xrightarrow[n\to\infty]{} 0$.

$$\mathbb{E}[|X_n - X|^p] = \mathbb{E}[\liminf_k |X_n - X_{n_k}|^p] \leq \liminf_k \mathbb{E}[|X_n - X_{n_k}|^p]$$

per $n > n_{\overline{k}}$ e $k > \overline{k}$ vale $\mathbb{E}[|X_n - X_{n_k}|^p] < \frac{1}{2^{2p\overline{k}}}$ quindi

$$\forall n > n_{\overline{k}} \qquad \mathbb{E}[|X_n - X|^p] < \frac{1}{2^{2p\overline{k}}}$$

□

4.6.3 Disuguaglianze di Holder e Minkowski

Enunciamo ora le disuguaglianza di Holder e Minkowski. Per completezza ci sono le dimostrazioni.

Teorema 4.16. *Siano X e Y variabili aleatorie. Siano $p \geq 1$ e $q \geq 1$ tali che $\frac{1}{p} + \frac{1}{q} = 1$. Se $X \in L^p$ e $Y \in L^q$ allora:*

$$\mathbb{E}[XY] \leq \mathbb{E}[|XY|] \leq \|X\|_p \|Y\|_q$$

Dimostrazione. Senza perdere in generalità possiamo supporre X e Y non negative. Questo semplificherà le notazioni. Sia $c := \mathbb{E}[X^p]$,

$$\mathbb{E}[XY] = c\mathbb{E}\left[\frac{X^p}{c} \cdot \frac{Y}{X^{p-1}}\right] = c\mathbb{E}\left[\frac{X^p}{c} \cdot \left(\frac{Y^q}{X^p}\right)^{\frac{1}{q}}\right] \tag{4.15}$$

Se poniamo $f := \frac{X^p}{c}$ allora f è non negativa e vale $\mathbb{E}[f] = 1$ quindi f è la funzione di densità di una misura di probabilità Q, sia $\mathbb{E}_Q$ il valore atteso associato a Q (ovvero $\mathbb{E}_Q[Z] = \mathbb{E}[f \cdot Z]$). Applicando la disuguaglianza di Jensen alla funzione concava $\Phi(x) = x^{\frac{1}{q}}$ la (4.15) diventa

$$\mathbb{E}[XY] = c\mathbb{E}_Q\left[\left(\frac{Y^q}{X^p}\right)^{\frac{1}{q}}\right] \leq c\mathbb{E}_Q\left[\left(\frac{Y^q}{X^p}\right)\right]^{\frac{1}{q}} = c\mathbb{E}\left[\frac{X^p}{c} \cdot \left(\frac{Y^q}{X^p}\right)\right]^{\frac{1}{q}} = \|X\|_p \|Y\|_q$$

□

Teorema 4.17. *Siano X e Y variabili aleatorie. Sia $p \geq 1$. Se $X, Y \in L^p$ allora:*

$$\|X\|_p + \|Y\|_p \geq \|X + Y\|_p$$

Dimostrazione. Se $p = 1$ la tesi è immediata poiché vale $|x + y| \leq |x| + |y|$. Sia $p > 1$, e q tale che $\frac{1}{p} + \frac{1}{q} = 1$. Senza perdere in generalità supponiamo X e Y non negative.

$$\begin{aligned}
\mathbb{E}[|X + Y|^p] &= \mathbb{E}[X|X + Y|^{p-1}] + \mathbb{E}[Y|X + Y|^{p-1}] \\
&\leq \|X\|_p \|(X + Y)^{p-1}\|_q + \|Y\|_p \|(X + Y)^{p-1}\|_q \\
&\leq (\|X\|_p + \|Y\|_p)\mathbb{E}[(X + Y)^p]^{\frac{1}{q}}
\end{aligned}$$

$$\mathbb{E}[|X + Y|^p]^{\frac{1}{p}} \leq \|X\|_p + \|Y\|_p$$

□

Capitolo 5

Indipendenza

In questo capitolo tratteremo l'indipendenza, grosso modo il capitolo 4 del libro [2]. A meno che non sia specificato diversamente, indicheremo con $(\Omega, \mathcal{H}, P)$ lo spazio di probabilità su cui si trovano gli eventi o sono definite le variabili aleatorie.

5.1 Indipendenza tra eventi

Introdurremo dapprima l'indipendenza di due eventi e poi quella di una famiglia genirica di eventi.

5.1.1 Indipendenza di due eventi

Intuitivamente due eventi E ed F sono indipendenti se il verificarsi di uno non modifica il nostro grado di fiducia enll'altro. Questo vuol dire che se $P(F) > 0$ allora E ed F indipendenti vuol dire

$$P(E|F) = P(E) \tag{5.1}$$

ovvero il fatto che si sia verificato F non ha modificato la nostra fiducia nell'evento E. Per $P(F) > 0$ le equazioni (5.1) e (5.2) sono equivalenti tuttavia per la definizione di indipendenza è preferibile la seconda perché mette in evidenza la simmetria tra gli eventi E ed F e soprattutto perché ha senso anche quando $P(F) = 0$.

Definizione 5.1. Diciamo che due eventi E ed F sono indipendenti se e solo se vale

$$P(E \cap F) = P(E)P(F) \ . \tag{5.2}$$

Una conseguenza di (5.2) è che gli eventi di misura nulla sono indipendenti da qualunque altro evento. Notare che gli eventi di misura zero e uno sono tutti e soli gli eventi indipendenti da tutti gli altri. Si può verificare che E ed F sono indipendenti se e solo se lo sono E ed F^c.

Esercizio 5.1. *Dimostrare che un evento A è indipendente da tutti gli altri eventi se e solo vale $P(A) \in \{0,1\}$.*

5.1.2 Indipendenza di una famiglia generica di eventi

Definizione 5.2. Diciamo che una famiglia di eventi $\{E_i\}_{i\in I}$ si dice una famiglia di eventi indipendenti se e solo se per ogni $J \subseteq I$ con $|J| < \infty$ vale

$$P\left(\bigcap_{j\in J} E_j\right) = \prod_{j\in J} P(E_j) \ . \tag{5.3}$$

Dunque una famiglia di eventi $\{E_i\}_{i\in I}$ si dice una famiglia di eventi indipendenti se e solo se per tutti i sottoinsiemi finiti vale la formula (5.3). Vediamo ora un esercizio che può essere utile per le applicazioni.

Esercizio 5.2. *Siano $E_1, E_2, \dots, E_n$ eventi indipendenti. Se $F_1, F_2, \dots, F_n$ sono eventi tali che per ogni i o $F_i = E_i$ oppure $F_i = E_i^c$ allora anche $F_1, F_2, \dots, F_n$ sono eventi indipendenti.*

Svolgimento. È chiaro che se si riesce a fare il passaggio al complementare su un singolo evento, poi uno alla volta per induzione si può passare al complementare su qualunque sottoinsieme di eventi. La proprietà (5.3) va dimostrata su qualunque sottoinsieme di eventi, se tale sottoinsieme è costituito da m elementi con $m \in \{2,3,\dots,n\}$ (i casi $m = 0$ e $m = 1$ sono sempre veri) allora a patto di permutare gli indici possiamo supporre di dover dimostrare che:

$$P(E_1 \cap \dots \cap E_{m-1} \cap E_m^c) = P(E_1)\dots P(E_{m-1})P(E_m^c) \tag{5.4}$$

il punto chiave adesso sta nell'esprimere l'evento della prima probabilità della (5.4) in termini di sole intersezioni di eventi E_i per i quali possiamo utilizzare l'indipendenza, osserviamo che:

$$E_1 \cap \dots \cap E_{m-1} \cap E_m^c = (E_1 \cap \dots \cap E_{m-1}) \setminus (E_1 \cap \dots \cap E_{m-1} \cap E_m)$$

dunque

$$P(E_1 \cap \dots \cap E_{m-1} \cap E_m^c) = P(E_1 \cap \dots \cap E_{m-1}) - P(E_1 \cap \dots \cap E_{m-1} \cap E_m)$$

poiché stiamo supponendo gli E_i indipendenti si ha:

$$P(E_1 \cap \dots \cap E_{m-1} \cap E_m^c) = P(E_1)\dots P(E_{m-1}) - P(E_1)\dots P(E_{m-1})P(E_m)$$

da cui segue immediatamente la (5.4) e dunque la tesi. □

Con il seguente esercizio mostreremo che indipendenza a due a due non implica indipendenza.

Esercizio 5.3. *Consideriamo il lancio di due dadi. Consideriamo i seguenti eventi:*
$A :=$ *"il risultato del primo lancio è pari"*
$B :=$ *"il risultato del secondo lancio è pari"*
$C :=$ *"il risultato della somma è pari"*
Dimostrare che gli eventi $\{A, B\}$ *sono indipendenti, gli eventi* $\{A, C\}$ *sono indipendenti, gli eventi* $\{B, C\}$ *sono indipendenti, ma gli eventi* $\{A, B, C\}$ *non sono indipendenti*

Svolgimento. Calcolando le probabilità degli eventi A, B e C e delle loro intersezioni si ottiene:
$P(A) = 1/2$, $P(B) = 1/2$, $P(C) = 1/2$, $P(A \cap B) = 1/4$, $P(A \cap C) = 1/4$, $P(B \cap C) = 1/4$, $P(A \cap B \cap C) = 1/4$ dunque

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

$$P(B \cap C) = P(B)P(C) \qquad P(A \cap B \cap C) \neq P(A)P(B)P(C)$$

□

5.2 Indipendenza di σ-algebre

Definizione 5.3. Diciamo che una famiglia di σ-algebre $\{\mathcal{F}_i\}_{i\in I}$ si dice una famiglia di σ-algebre indipendenti se e solo se per ogni $J \subseteq I$ con $|J| < \infty$ vale

$$P\left(\bigcap_{j\in J} A_j\right) = \prod_{j\in J} P(A_j) \qquad \forall\ (A_j)_{j\in J} \in (\mathcal{F}_j)_{j\in J} \tag{5.5}$$

Dunque una famiglia di σ-algebre $\{\mathcal{F}_i\}_{i\in I}$ si dice una famiglia di σ-algebre indipendenti se e solo se per tutti i sottoinsiemi finiti J vale la formula (5.5).

Proposizione 5.1. *Sia* $\{E_i\}_{i\in I}$ *una famiglia di eventi e siano* $\mathcal{F}_i := \sigma(E_i)$ *le* σ*-algebre generate. Sono cose equivalenti:*

(a) $\{E_i\}_{i\in I}$ *è una famiglia di eventi indipendenti*

(b) $\{\mathcal{F}\}_{i\in I}$ *è una famiglia di* σ*-algebre indipendenti*

Dimostrazione. Osserviamo che $\mathcal{F}_i = \{\varnothing, E_i, E_i^c, \Omega\}$. Allora un'implicazione è ovvia e l'altra è una conseguenza degli esercizi 5.1 e 5.2. □

5.3 Indipendenza di variabili aleatorie

Definizione 5.4. Sia $\{X_i\}_{i\in I}$ una famiglia di variabili, e per ogni $i \in I$ sia $\mathcal{F}_i := \sigma(X_i)$. Diciamo che una famiglia di variabili aleatorie $\{X_i\}_{i\in I}$ è una famiglia di variabili aleatorie indipendenti se $\{\mathcal{F}_i\}_{i\in I}$ è una famiglia di σ-algebre indipendenti.

Proposizione 5.2. *Per ogni $i \in I$ sia X_i variabile aleatoria a valori in $(E_i, \mathcal{E}_i)$. Le variabili aleatorie $\{X_i\}_{i\in I}$ sono indipendenti se e solo se per ogni $J \subseteq I$ finito e per ogni $(A_j)_{j\in J} \in (\mathcal{E}_j)_{j\in J}$ vale:*

$$P\left(\bigcap_{j\in J}\{X_j \in A_j\}\right) = \prod_{j\in J} P(X_j \in A_j)$$

Proposizione 5.3. *Sia $\{X_i\}_{i\in I}$ una famiglia di v.a. a valori in $(E_i, \mathcal{E}_i)_{i\in I}$. Siano ϕ_i applicazioni misurabili da $(E_i, \mathcal{E}_i)$ in $(F_i, \mathcal{F}_i)$. Sia infine $Y_i := \phi_i(X_i)$ per ogni $i \in I$. Se $\{X_i\}_{i\in I}$ è una famiglia di v.a. indipendenti allora anche le v.a $\{Y_i\}_{i\in I}$ costituiscono una famiglia di v.a. indipendenti*

Dimostrazione. La dimostrazione segue immediatamente da $\sigma(Y_i) \subseteq \sigma(X_i)$. □

Esercizio 5.4. *Con le notazioni della proposizione precedente ($Y_i = \phi_i(X_i)$). Costruire un esempio in cui le v.a. $\{Y_i\}_{i\in I}$ siano indipendenti mentre le v.a. $\{X_i\}_{i\in I}$ non siano indipendenti.*

5.4 Indipendenza e p-system

Procederemo per gradi, analizzeremo prima l'indipendenza tra un evento ed una σ-algebra generata da un p-system, poi studieremo l'indipendenza tra due σ-algebre ed infine studieremo il problema per una famiglia qualsiasi di σ-algebre.

Proposizione 5.4. *Sia A un evento, $\mathcal{A}$ un p-system e $\mathcal{F} := \sigma(\mathcal{A})$. Se A è indipendente da tutti gli eventi di $\mathcal{A}$ allora A è indipendente anche da tutti gli eventi di $\mathcal{F}$.*

Dimostrazione. Per cominciare poiché Ω è indipendente da tutti gli eventi possiamo assumere $\Omega \in \mathcal{A}$. Per il lemma di Dynkin vale:

$$\mathcal{F} := \sigma(\mathcal{A}) = d(\mathcal{A})$$

procederemo ora in maniera standard, dimostrando che gli eventi che soddisfano la proprietà richiesta sono un d-system e contengono $\mathcal{A}$. Sia

$$\mathcal{G} := \{B \in \mathcal{F} | P(A \cap B) = P(A) \cdot P(B)\}$$

Poiché A è indipendente dagli eventi di $\mathcal{A}$ allora $\mathcal{G} \supseteq \mathcal{A}$. Resta da dimostrare che $\mathcal{G}$ è un d-system.
(i) $\Omega \in \mathcal{A} \supseteq \mathcal{G}$
(ii) $B \in \mathcal{G}$ implica $P(A \cap B) = P(A) \cdot P(B)$.
$P(A \cap B^c) = P(A) - P(A \cap B) = P(A)(1 - P(B)) = P(A) \cdot P(B^c)$

quindi anche B^c appartiene a $\mathcal{G}$
(iii) Siano $\{B_n\}_{n\in\mathbb{N}}$ disgiunti e in $\mathcal{G}$, quindi $P(A \cap B_n) = P(A) \cdot P(B_n)$

$$\begin{aligned} P\left(A \cap \left(\bigcup_n B_n\right)\right) &= P\left(\bigcup_n (A \cap B_n)\right) = \sum_n P(A \cap B_n) \\ &= \sum_n P(A)P(B_n) = P(A) \sum_n P(B_n) = \\ &= P(A)P\left(\bigcup_n B_n\right) \end{aligned}$$

dunque anche $\bigcup_n B_n$ appartiene a $\mathcal{G}$. □

Proposizione 5.5. *Siano $\mathcal{F}_1$ e $\mathcal{F}_2$ due σ-algebre e siano $\mathcal{A}_1$ e $\mathcal{A}_2$ due p-system tali che $\mathcal{F}_i := \sigma(\mathcal{A}_i)$. Se gli eventi di $\mathcal{A}_1$ sono indipendenti dagli eventi di $\mathcal{A}_2$ allora le σ-algebre $\mathcal{F}_1$ e $\mathcal{F}_2$ sono indipendenti.*

Dimostrazione. Applichiamo dapprima la proposizione 5.4 ad $A \in \mathcal{A}_1$ e $\mathcal{A} = \mathcal{A}_2$ cosicché si ottiene che gli elementi di $\mathcal{A}_1$ sono indipendenti da quelli di $\mathcal{F}_2$. Applicchiamo ora la proposizione 5.4 ad $A \in \mathcal{F}_2$ e $\mathcal{A} = \mathcal{A}_1$ cosicché si ottiene che gli elementi di $\mathcal{F}_1$ sono indipendenti da quelli di $\mathcal{F}_2$. □

Proposizione 5.6. *Sia $\{\mathcal{F}_i\}_{i\in I}$ una famiglia di σ-algebre, sia $\{\mathcal{A}_i\}_{i\in I}$ una famiglia di p-system tali che $\mathcal{F}_i := \sigma(\mathcal{A}_i)$. Se per ogni $J \subseteq I$ finito e per ogni $(A_j)_{j\in J} \in \bigtimes_{j\in J} \mathcal{A}_j$ vale*

$$P\left(\bigcap_{j\in J} A_j\right) = \prod_{j\in J} P(A_j)$$

allora $\{\mathcal{F}_i\}_{i\in I}$ è una famiglia di σ-algebre indipendenti.

Dimostrazione. Innanzitutto l'insieme Ω è indipendente da qualunque altro evento, quindi possiamo assumere senza perdere in generalità che Ω appartenga ai p-system $\mathcal{A}_i$. La definizione di famiglia di σ-algebre indipendenti si basa sulla verifica della condizione (5.5) su insiemi finiti di indici quindi senza perdere in generalità possiamo assumere $I = \{1, 2, \dots, n\}$ finito. Inoltre per la proposizione precedente il caso $n = 2$ è vero e per induzione possiamo assumere che sia vero per $n - 1$.
Sia dunque $n \geq 3$.

`Ipotesi` $\forall\, (A_i)_{i\in I} \in \bigtimes_{i\in I} (\mathcal{A}_i)$ vale $P(A_1 \cap \ldots \cap A_n) = P(A_1) \cdot \ldots \cdot P(A_n)$

`Tesi` $\forall\, (B_i)_{i\in I} \in \bigtimes_{i\in I} (\mathcal{F}_i)$ vale $P(B_1 \cap \ldots \cap B_n) = P(B_1) \cdot \ldots \cdot P(B_n)$

Per induzione sappiamo che $P(B_1 \cap \ldots \cap B_{n-1}) = P(B_1) \cdot \ldots \cdot P(B_{n-1})$ quindi la tesi diventa:

$$P((B_1 \cap \ldots \cap B_{n-1}) \cap B_n) = P(B_1 \cap \ldots \cap B_{n-1}) \cdot P(B_n) \tag{5.6}$$

A questo punto l'idea è di dimostrare l'uguaglianza precedente utilizzando la proposizione 5.5. Definiremo ora le σ-algebre $\mathcal{G}_1$ e $\mathcal{G}_2$ e i p-system $\mathcal{B}_1$ e $\mathcal{B}_2$ a cui applicare la proposizione 5.5.

$$\mathcal{G}_1 := \sigma(\mathcal{F}_1, \dots, \mathcal{F}_{n-1}) \qquad \mathcal{G}_2 := \mathcal{F}_n$$

$$\mathcal{B}_1 := \{A_1 \cap \dots \cap A_{n-1} | \forall i \ A_i \in \mathcal{A}_i\} \qquad \mathcal{B}_2 = \mathcal{A}_n$$

Chiaramente $\mathcal{B}_1$ e $\mathcal{B}_2$ sono p-system e $\sigma(\mathcal{B}_2) = \mathcal{G}_2$.
Mentre $\sigma(\mathcal{B}_1) = \sigma(\mathcal{A}_1, \dots, \mathcal{A}_{n-1}) = \sigma(\mathcal{F}_1, \dots, \mathcal{F}_{n-1}) = \mathcal{G}_1$ Quindi per la proposizione 5.5 vale la (5.6). □

5.5 Indipendenza a blocchi

Teorema 5.7. *Sia $\{\mathcal{H}_i\}_{i\in I}$ una famiglia di σ-algebre indipendenti. Sia $\{I_j\}_{j\in J}$ una partizione di I. Se poniamo per ogni $j \in J$:*

$$\mathcal{G}_j := \sigma(\{\mathcal{H}_i | i \in I_j\})$$

allora $\{\mathcal{G}_j\}_{j\in J}$ è una famiglia di σ-algebre indipendenti.

Dimostrazione. Per ogni $j \in J$ sia:

$$\mathcal{A}_j := \left\{ \bigcap_{i\in\Gamma} A_i \middle| \Gamma \subseteq I_j \text{ finito e } \forall i \in \Gamma \quad A_i \in \mathcal{H}_i \right\}$$

Chiaramente $\mathcal{A}_j$ è un p-system e vale $\mathcal{G}_j = \sigma(\mathcal{A}_j)$. Se mostriamo che $\{\mathcal{A}_j\}_{j\in J}$ è una famiglia di p-system indipendenti allora per la proposizione 5.6 abbiamo finito. Mostriamo ora che $\{\mathcal{A}_j\}_{j\in J}$ è una famiglia di p-system indipendenti.
Siano $j_1, \dots, j_n \in J$ distinti e consideriamo gli eventi $B_k \in \mathcal{A}_{j_k}$ con $k \in \{1, \dots, n\}$. Dobbiamo mostrare che gli eventi $\{\mathcal{B}_k\}_{1\leq k\leq n}$ sono indipendenti.
Poiché $B_k \in \mathcal{A}_{j_k}$ allora esistono interi m_k ed eventi $A_{i_{k,h}} \in \mathcal{H}_{i_{k,h}}$ tali che:

$$B_k = A_{i_{k,1}} \cap \dots \cap A_{i_{k,m_k}}$$

dove gli indici $i_{h,k}$ sono tutti distinti al variare di k e h.

$$P(B_1 \cap \dots \cap B_n) = P\left(\bigcap_{\substack{k\in\{1,\dots,n\} \\ h\in\{1,\dots,m_k\}}} A_{i_{k,h}} \right) = \prod_{\substack{k\in\{1,\dots,n\} \\ h\in\{1,\dots,m_k\}}} P(A_{i_{k,h}})$$

$$= \prod_{k\in\{1,\dots,n\}} \left(\prod_{h\in\{1,\dots,m_k\}} P(A_{i_{k,h}}) \right) = P(B_1) \cdot \dots \cdot P(B_n)$$

□

5.6 Secondo Lemma di Borel-Cantelli

Lemma 5.8. *Sia* $\{A_n\}_{n\in\mathbb{N}}$ *una successione di eventi indipendenti. Se vale* $\sum_n P(A_n) = \infty$ *allora*

$$P\left(\limsup_n A_n\right) = 1$$

Dimostrazione.

$$P\left(\limsup_n A_n\right) = P\left(\bigcap_n \bigcup_{k\geq n} A_k\right) = P\left(\bigcap_n B_n\right) =$$

Dove abbiamo posto $B_k := \bigcup_{k\geq n} A_k$, gli eventi B_n sono noncrescenti in n quindi

$$P\left(\limsup_n A_n\right) = \lim_n P(B_n)$$

$P(B_n)$ è non crescente quindi se vogliamo che il limite al secondo membro sia uguale ad 1 allora l'unica possibilità è dimostrare che $P(B_n) = 1$.

$$\begin{aligned} P(B_n) =& P\left(\bigcup_{k\geq n} A_k\right) = 1 - P\left(\left(\bigcup_{k\geq n} A_k\right)^c\right) = 1 - P\left(\bigcap_{k\geq n} A_k^c\right) \\ =& 1 - \prod_{k\geq n} P\left(A_k^c\right) = 1 - \prod_{k\geq n}\left(1 - P\left(A_k\right)\right) = 1 \end{aligned}$$

dove l'ultima uguaglianza è dovuta al lemma 5.9. □

Lemma 5.9. *Data una successione una successione* $\{a_n\}_{n\in\mathbb{N}}$ *con* $a_n \in (0,1)$ *per ogni* n*. Vale:*

$$\prod_{n\in\mathbb{N}}(1-a_n) = 0 \qquad \Longleftrightarrow \qquad \sum_{n\in\mathbb{N}} a_n = \infty$$

Dimostrazione. Consideriamo dapprima il caso semplice a_n non tende a zero allora esiste $\varepsilon > 0$ e una sottosuccesione n_k tale che per ogni k si ha $a_{n_k > \varepsilon}$.
Quindi
$\prod_{n\in\mathbb{N}}(1-a_n) \leq \prod_{k\in\mathbb{N}}(1-a_{n_k}) = 0$
$\sum_{n\in\mathbb{N}} a_n \geq \sum_{k\in\mathbb{N}} a_{n_k} = \infty$
Consideriamo ora il caso a_n tende a zero.

$$\prod_{n\in\mathbb{N}}(1-a_n) = \prod_{n\in\mathbb{N}} e^{\log(1-a_n)}$$

Poiché gli esiti della produttoria e della sommatoria non dipendono da un numero finito di a_n allora possiamo considerare a_n piccolo cosicché

$$\begin{gathered} -2a_n \leq log(1-a_n) \leq -a_n \\ \prod_{n\in\mathbb{N}} e^{-a_n} \leq \prod_{n\in\mathbb{N}}(1-a_n) \leq \prod_{n\in\mathbb{N}} e^{-a_n} \\ e^{-2\sum_n a_n} \leq \prod_{n\in\mathbb{N}}(1-a_n) \leq e^{-\sum_n a_n} \end{gathered}$$

□

5.7 σ-algebra coda e legge 0-1 di Kolmogorov

Introdurremo ora la sigma algebra coda e la legge 0-1 di Kolmogorov.

Definizione 5.5. Data una successione di σ-algebre $\{\mathcal{G}_n\}_{n\in\mathbb{N}}$, sia $\tau_n := \sigma\{\mathcal{G}_k | k > n\}$, si dice σ-algebra coda la σ-algebra $\tau := \bigcap_n \tau_n$.

Intuitivamente gli eventi di τ sono quelli che possono essere determinati dalle informazioni presenti su una qualunque coda delle σ-algebre.

Teorema 5.10 (Legge 0-1 di Kolmogorov)**.**
Sia $\{\mathcal{G}_n\}_{n\in\mathbb{N}}$ una successione di σ-algebre, sia $\tau_n := \sigma\{\mathcal{G}_k | k > n\}$ e sia $\tau := \bigcap_n \tau_n$ la σ-algebra coda. Se $\{\mathcal{G}_n\}_{n\in\mathbb{N}}$ è una successione di σ-algebre indipendenti allora τ è una σ-algebra banale ovvero

$$\textit{per ogni} \;\; A \in \tau \qquad \textit{vale} \;\; P(A) \in \{0,1\}$$

Dimostrazione. Dimostreremo che la σ-algebra τ è indipendente da se stessa cosicché per ogni $A \in \tau$ si ha $P(A \cap A) = P(A) \cdot P(A)$ e quindi $P(A) \in \{0,1\}$. Sia

$$\mathcal{F}_n := \sigma(\mathcal{G}_1, \cdots, \mathcal{G}_n)$$

per il principio di indipendenza a blocchi $\mathcal{F}_n$ e τ_n sono indipendenti e poiché $\tau \subseteq \tau_n$ allora $\mathcal{F}_n$ e τ sono indipendenti. Sia

$$\mathcal{A} := \bigcup_n \mathcal{F}_n \qquad \text{e} \qquad \mathcal{G} := \sigma(\{\mathcal{G}_n\}_{n\in\mathbb{N}}) = \sigma(\mathcal{A})$$

chiaramente $\mathcal{A}$ è un p-system genera $\mathcal{G}$ e per quanto dimostrato prima è indipendente da τ. Per il teorema 5.6 si ha τ è indipendente da $\mathcal{G}$. Infine poiché $\tau \subseteq \mathcal{G}$ allora τ è indipendente da τ. □

La legge 0-1 di Kolmogorov ci dice che nelle ipotesi del teorema se un'evento appartiene a τ allora è quasi certamente vero o quasi certamente falso senza vie di mezzo. Se una variabile aleatoria X reale è τ misurabile allora è quasi certamente constante ovvero esiste x tale che $P(X = x) = 1$.

Esempio 5.1. *Siano $\{X_n\}_{n\in\mathbb{N}}$ indipendenti, sia $S_n := X_1 + \ldots + X_n$. Sia τ la σ-algebra coda ovvero $\tau = \bigcap_{k>n} \tau_k$ con $\tau_k := \sigma(X_{k+1}, X_{k+2}, \ldots)$. Per la legge* 0-1 *di Kolmogorov allora:*

- $\exists\, C \in [-\infty, +\infty]$ *tale che* $\limsup_n X_n = C$ *q.c.*
- *la probabilità che* $\frac{S_n}{n}$ *converga è zero oppure uno.*
- *se la probabilità che* $\frac{S_n}{n}$ *converga è maggiore stretto di zero allora esiste* C *tale che* $\frac{S_n}{n}$ *converge quasi certamente a* C.

5.8 Unicità della distribuzione congiunta di v.a. indipendenti

Molto spesso gli esercizi iniziano con affermazioni del tipo: "siano X e Y variabili aletorie indipendenti con distribuzione $X \sim Unif(a,b)$ e $Y \sim Esp(\lambda)$.", si assume per ipotesi che esistano e si dà per scontato che la distribuzione congiunta sia unica. In questo paragrafo vedremo che l'unicità è vera. Mentre l'esistenza verrà mostrata quando si introdurranno le misure prodotto.

Teorema 5.11. *Siano* $(\Omega_1, \mathcal{H}_1, P_1)$ *e* $(\Omega_2, \mathcal{H}_2, P_2)$ *spazi di probabilità. Sia* $\{(E_i, \mathcal{E}_i)\}_{i \in I}$ *una famiglia di spazi misurabili. Siano*

$$\begin{aligned} X_i &: (\Omega_1, \mathcal{H}_1) \to (E_i, \mathcal{E}_i) \\ Y_i &: (\Omega_2, \mathcal{H}_2) \to (E_i, \mathcal{E}_i) \end{aligned}$$

misurabili con eguale distribuzione, $X_i(P_1) = Y_i(P_2)$ *per ogni* i. *Se poniamo* $X := (X_i)_{i \in I}$ *e* $Y := (Y_i)_{i \in I}$ *allora* X *e* Y *sono applicazioni misurabili a valori in* $(E, \mathcal{E}) := (\times_{i \in I} E_i, \otimes_{i \in I} \mathcal{E}_i)$ *ed hanno la stessa legge,* $P_1(X^{-1}) = P_2(Y^{-1})$.

Dimostrazione. Per la proposizione 2.11 X e Y sono misurabili. Consideriamo il seguente p-system

$$\mathcal{A} = \left\{ A = \bigtimes_{i \in I} A_i \,\middle|\, J \subseteq I \text{ finito }, A_i \in \mathcal{E}_i \text{ per ogni } i \in J, A_i = E_i \text{ per ogni } i \notin J \right\} \tag{5.7}$$

per la propsizione 2.9 $\mathcal{A}$ genera $\mathcal{E}$, mostreremo che le due misure $P_1(X^{-1})$ e $P_2(Y^{-1})$ coincidono sul p-system $\mathcal{A}$ cosicché per il teorema 3.3 sono uguali. Sia $A = \bigtimes_{i \in I} A_i \in \mathcal{A}$ come in (5.7),

$$\begin{aligned} P_1(X^{-1}(A)) =& P_1\left(\bigcap_{i \in J} X_i^{-1}(A_i) \right) = \prod_{i \in J} P_1\left(X_i^{-1}(A_i) \right) \\ =& \prod_{i \in J} P_1\left(Y_i^{-1}(A_i) \right) = P_1\left(\bigcap_{i \in J} Y_i^{-1}(A_i) \right) \\ =& P_1(Y^{-1}(A)) \end{aligned}$$

dove per la seconda e quarta uguaglianza è stata utilizzata l'indipendenza. □

5.9 Cifre decimali di un numero a caso in $(0, 1)$

In questo paragrafo forniremo un metodo costruttivo per ottenere una famiglia numerabile di variabili aleatorie indipendenti. Consideriamo un numero U scelto a caso tra zero e uno ($U \sim Unif(0,1)$). Siano $\{X_n\}_{n\in\mathbb{N}}$ le cifre di U dopo la virgola e sia $X := (X_n)_{n\in\mathbb{N}}$. Quindi X v.a. a valori in

$$(E, \mathcal{E}) := \left(\bigtimes_{i\in\mathbb{N}} \{0, \dots, 9\}, \bigotimes_{i\in\mathbb{N}} \mathcal{P}(\{0, \dots, 9\}) \right)$$

Dimostreremo ora che le $\{X_n\}_{n\in\mathbb{N}}$ sono uniformi su $\{0, 1, \dots, 9\}$ e indipendenti.
METODO 1 Questo metodo richiede un argomento ancora non affrontato. È bene illustrarlo comunque perché è il metodo che dovrebbe essere utilizzato in genere per risolvere questo tipo di problema. Siano $\{Y_n\}_{n\in\mathbb{N}}$ variabili aleatorie uniformi su $\{0, 1, \dots, 9\}$ e indipendenti, sia $Y := (Y_n)_{n\in\mathbb{N}}$. Per ogni $(x_1, \dots, x_n)$ vale

$$P(X_1 = x_1, \dots, X_n = x_n) = 10^{-n} = P(Y_1 = x_1, \dots, Y_n = x_n)$$

gli eventi descritti sopra con l'aggiunta dell'insieme vuoto formano un p-system

$$\mathcal{A} := \varnothing \cup \bigcup_{\substack{n\in\mathbb{N} \\ x_1,..,x_n}} \{\{e \in E | e_1 = x_1, \dots, e_n = x_n\}\}$$

che genera $\mathcal{E}$ quindi per il teorema 3.3 X e Y hanno la stessa distribuzione.
METODO 2 Le $\{X_n\}_{n\in\mathbb{N}}$ sono indipendenti se i sottoinsiemi finiti sono indipendenti, quindi basta mostrare che per ogni n le v.a. $\{X_1, \dots, X_n\}$ sono indipendenti. Il problema si semplifica ulteriormente perché le variabili aleatorie sono discrete, quindi basta verificare l'indipendenza sui singoletti:

$$P(X_1 = x_1, \dots, X_n = x_n) = 10^{-n} = P(X_1 = x_1) \cdot \ldots \cdot P(X_n = x_n)$$

La prima disuguaglianza è ovvia
$P(X_1 = x_1, \dots, X_n = x_n) = P(U \in (0.x_1x_2 \dots x_n, 0.x_1x_2 \dots x_n + 10^{-n}))$
(per esempio $P(X_1 = 1, X_2 = 6, X_3 = 2) = P(U \in [0.162, 0.163)) = 10^{-3}$)
mostrare che $P(X_n = x_n) = \frac{1}{10}$ è più complesso

$$P(X_n = x_n) = P\left(\bigcup_{x_1,\dots,x_{n-1}} \{X_1 = x_1, \dots, X_n = x_n\} \right) = 10^{n-1} \cdot 10^{-n} = 10^{-1}$$

Abbiamo quindi mostrato che le $\{X_n\}_{n\in\mathbb{N}}$ sono i.i.d. Vedremo ora che vale anche il viceversa. Denotiamo con Φ l'applicazione descritta sopra che manda U in X: $X = \Phi(U)$. L'applicazione Φ è iniettiva, denotiamo che Φ^{-1} la sua inversa. Siano $\{Y_n\}_{n\in\mathbb{N}}$ variabili aleatorie indipendenti e uniformemente distribuite su $\{0, 1, \dots, 9\}$ e sia $Y = (Y_n)_{n\in\mathbb{N}}$ (per l'unicitá della distribuzione di una famiglia di v.a. indipendenti) le v.a. X e Y hanno la stessa distribuzione e quindi anche $\Phi^{-1}(X)$ e $\Phi^{-1}(Y)$ hanno la stessa distribuzione. Questo vuol dire che data una variabile aleatoria $U \sim Unif(0,1)$ è possibile costruire una successione di variabili aleatorie i.i.d. e uniformi su $\{0,\dots,9\}$ e viceversa data una successione di variabili aleatorie i.i.d. con distribuzione uniforme su $\{0,\dots,9\}$ è possibile costruire una variabile aleatoria uniforme su $(0,1)$.

Supponiamo ora $U \sim Unif(0,1)$ e $\{X_n\}_{n\in\mathbb{N}}$ siano le cifre di U allora posso porre $Y_n := X_{2n}$ e $Z_n := X_{2n+1}$ e ricavare due successioni di variabili indipendenti dalle quali posso ricavare due v.a. indipendenti con distribuzione uniforme su $(0,1)$. Allo stesso modo se poniamo $Z_{k,n} := X_{2^k 3^n}$ otteniamo una successione $\{Z_{k,n}\}_{n\in\mathbb{N}}$ di successioni di variabili indipendenti, alle quali possiamo far corrispondere una successione di variabili aleatorie uniformi su $(0,1)$ e indipendenti. Ovvero abbiamo costruito un'applicazione $\Psi : (0,1) \to (0,1)^{\mathbb{N}}$ tale che posto $T = (T_n)_{n\in\mathbb{N}} := \Psi(U)$ si abbia $\{T_n\}_{n\in\mathbb{N}}$ è una famiglia di variabili aleatorie indipendenti uniformi su $(0,1)$. Non è possibile fare di meglio ovvero vale il seguente esercizio.

Esercizio 5.5. *****
Sia $U \sim Unif(0,1)$ sia Ψ un'applicazione misurabile

$$\Psi : ((0,1), \mathcal{B}(0,1)) \to \left(\bigtimes_{i\in I}(0,1), \bigotimes_{i\in\mathbb{N}} \mathcal{B}((0,1)) \right)$$

Sia $X := (X_i)_{i\in I} := \Psi(U)$. Se le variabili aleatorie $\{X_i\}_{i\in I}$ sono i.i.d. con distribuzione uniforme su $(0,1)$ allora $\#(I) \leq \#(\mathbb{N})$.

`Suggerimento` può essere utile l'esercizio 3 del foglio 2.

Capitolo 6

Prodotto di spazi di probabilità

In questo paragrafo verrà esposto il teorema di Fubini-Tonelli e l'esistenza del prodotto degli spazi di probabilità. Nel capitolo 2 abbiamo visto che data una famiglia di spazi misurabili $\{(\Omega_i, \mathcal{H}_i)\}_{i \in I}$ è possibile definire il loro prodotto $(\Omega, \mathcal{H}) := (\times_{i \in I} \Omega_i, \bigotimes_{i \in I} \mathcal{H}_i)$ dove $\mathcal{H}$ è la più piccola σ-algebra che rende misurabili le proiezioni π_i. In questo capitolo mostreremo che se $\{(\Omega_i, \mathcal{H}_i, P_i)\}_{i \in I}$ è una famiglia di spazi di probabilità allora è possibile definire una probabilità P sullo spazio prodotto $(\Omega, \mathcal{H})$ tale che le proiezioni $\{\pi_i\}_{i \in I}$ siano indipendenti ed abbiano legge P_i tale misura di probabilità è unica. Mostreremo anche sotto quali condizioni dati due spazi di misura $(\Omega_i, \mathcal{H}_i, \mu_i)$ σ-finiti vale la seguente uguaglianza:

$$\int_{\Omega_1} \int_{\Omega_2} f(x,y) d\mu_2(y)\ d\mu_1(x) = \int_{\Omega_2} \int_{\Omega_1} f(x,y) d\mu_1(x)\ d\mu_2(y) \tag{6.1}$$

Di questo capitolo per l'esame occorre conoscere solo gli enunciati. Ma per completezza verranno illustrate anche le dimostrazioni.

Esempi notevoli

Mostreremo ora due esempi in cui non vale l'uguaglianza (6.1).

Esempio 1

Sia $(\Omega_1, \mathcal{H}_1, \mu_1) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_1)$ con μ_1 misura di Lebesgue e sia $(\Omega_2, \mathcal{H}_2, \mu_2) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_2)$ con μ_2 misura che conta i punti ($\mu_2(A) := \sharp(A)$) sia infine

$$f(x,y) = \begin{cases} 1 & \textbf{se } x = y \\ 0 & \textbf{se } x \neq y \end{cases}$$

allora

$$\int_{\Omega_1}\left(\int_{\Omega_2} f(x,y)d\mu_2(y)\right)\ d\mu_1(x) = \int_{\Omega_1}\left(\sum_y f(x,y)\right)\ d\mu_1(x)$$
$$= \int_{\Omega_1} 1\ d\mu_1(x) = +\infty$$

mentre

$$\int_{\Omega_2}\left(\int_{\Omega_1} f(x,y)d\mu_1(x)\right)\ d\mu_2(y) = \int_{\Omega_2}\left(\int_{-\infty}^{+\infty} f(x,y)dx\right)\ d\mu_2(y)$$
$$= \int_{\Omega_2} 0\ d\mu_2(y) = 0$$

Esempio 2

Sia $(\Omega_i,\mathcal{H}_i,\mu_i) = ((0,1),\mathcal{B}(0,1),m)$ con m misura di Lebesgue. sia infine

$$f(x,y) = \frac{1}{x} - \frac{1}{y}$$

allora

$$\int_{\Omega_1}\left(\int_{\Omega_2} f(x,y)d\mu_2(y)\right)\ d\mu_1(x) = \int_0^1\int_0^1\left(\frac{1}{x}-\frac{1}{y}\right)dy\ dx$$
$$= \int_0^1 (-\infty)dx = -\infty$$

mentre

$$\int_{\Omega_2}\left(\int_{\Omega_1} f(x,y)d\mu_1(x)\right)\ d\mu_2(y) = \int_0^1\int_0^1\left(\frac{1}{x}-\frac{1}{y}\right)dx\ dy$$
$$= \int_0^1 (+\infty)dy = +\infty$$

6.1 Prodotto di due spazi di probabilità

Affronteremo ora il caso del prodotto di due spazi di misura finita che include il prodotto di due spazi di probablità. Data f variabile aleatoria non negativa da $(\Omega,\mathcal{H})$ in $(\mathbb{R},\mathcal{B}(\mathbb{R}))$. Tenuto conto della (6.1) definiamo le seguenti notazioni:

$$I_1^f(x) := \int_{\Omega_2} f(x,y)d\mu_2(y) \qquad\qquad I_2^f(y) := \int_{\Omega_1} f(x,y)d\mu_1(x)$$

inoltre sarà utile la definizione:

$$\mathcal{R} := \{A_1\times A_2 | A_1\in\mathcal{H}_1, A_2\in\mathcal{H}_2\}$$

cosicché $\mathcal{R}$ è un p-system che contiene Ω e genera $\mathcal{H}$.

Lemma 6.1. *Siano* $(\Omega_1, \mathcal{H}_1, \mu_1)$ *e* $(\Omega_2, \mathcal{H}_2, \mu_2)$ *due spazi di misura finita. Per ogni* $A \in \mathcal{H}$ *posto* $f := \mathbf{1}_A$ *si ha* I_1^f *è* $\mathcal{H}_1$ *misurabile,* I_2^f *è* $\mathcal{H}_2$ *misurabile e vale*

$$\int_{\Omega_1} I_1^f(x)\ d\mu_1(x) = \int_{\Omega_2} I_2^f(y)\ d\mu_2(y) \tag{6.2}$$

Dimostrazione. Sia $\mathcal{G}$ l'insieme degli elementi di $\mathcal{H}$ che soddisfano le richieste del teorema ovvero

$$\mathcal{G} := \{A \in \mathcal{H} | I_1^{\mathbf{1}_A} \text{ è } \mathcal{H}_1 \text{ mis.}, I_2^{\mathbf{1}_A} \text{ è } \mathcal{H}_2 \text{ mis. e vale l'equazione (6.2)}\}$$

procederemo dimostrando prima che $\mathcal{G} \supseteq \mathcal{R}$ e poi che $\mathcal{G}$ è un d-system cosicché per il lemma di Dyndin si abbia $\mathcal{G} = \mathcal{H}$.
Sia $A = A_1 \times A_2 \in \mathcal{R}$,

$$I_1^{\mathbf{1}_A}(x) = \begin{cases} \mu_2(A_2) & \text{se } x \in A_1 \\ 0 & \text{se } x \notin A_1 \end{cases} \qquad I_2^{\mathbf{1}_A}(y) = \begin{cases} \mu_1(A_1) & \text{se } y \in A_2 \\ 0 & \text{se } y \notin A_2 \end{cases}$$

quindi $I_i^{\mathbf{1}_A}$ è $\mathcal{H}_i$ misurabile e vale

$$\int_{\Omega_1} I_1^{\mathbf{1}_A}(x)\ d\mu_1(x) = \mu_1(A_1) \cdot \mu_2(A_2) = \int_{\Omega_2} I_2^{\mathbf{1}_A}(y)\ d\mu_2(y)$$

dunque $\mathcal{R} \subseteq \mathcal{G}$ per concludere basta mostrare che $\mathcal{G}$ è un d-system.
Poiché $\Omega \in \mathcal{R}$ allora $\Omega \in \mathcal{G}$. La verifica che $\mathcal{G}$ è chiuso rispetto al complementare è immediata perché $\mathbf{1}_{A^c} = 1 - \mathbf{1}_A$. Resta solo da dimostrare la chiusura di $\mathcal{G}$ per unioni numerabili di elementi disgiunti.
Sia $\{A_n\}_{n \in \mathbb{N}}$ una successione di elementi di $\mathcal{G}$ disgiunti e sia $A := \bigcup_n A_n$ allora per il teorema di convergenza monotona

$$I_i^{\mathbf{1}_A} = \limsup_n \sum_{k=1}^n I_i^{\mathbf{1}_{A_k}}$$

ancora per il teorema di convergenza monotona e l'ipotesi $A_n \in \mathcal{G}$ si ha:

$$\int_{\Omega_1} I_1^{\mathbf{1}_A}(x)\ d\mu_1(x) = \limsup_n \sum_{k=1}^n \int_{\Omega_1} I_1^{\mathbf{1}_{A_k}}(x)\ d\mu_1(x) =$$

$$= \limsup_n \sum_{k=1}^n \int_{\Omega_2} I_2^{\mathbf{1}_{A_k}}(y)\ d\mu_2(y) = \int_{\Omega_2} I_2^{\mathbf{1}_A}(y)\ d\mu_2(y)$$

□

Teorema 6.2. *Siano* $(\Omega_1, \mathcal{H}_1, \mu_1)$ *e* $(\Omega_2, \mathcal{H}_2, \mu_2)$ *due spazi di misura finita. Sia* $(\Omega, \mathcal{H}) := (\times_{i \in I} \Omega_i, \bigotimes_{i \in I} \mathcal{H}_i)$. *Per ogni* $A \in \mathcal{H}$ *definiamo la funzione* μ:

$$\mu(A) := \int_{\Omega_1} \int_{\Omega_2} I_A(x,y)\ d\mu_2(y)\ d\mu_1(x) = \int_{\Omega_2} \int_{\Omega_1} I_A(x,y)\ d\mu_1(x)\ d\mu_2(y) \tag{6.3}$$

allora la precedente definizione di μ *è una buona definizione, inoltre* μ *è l'unica misura su* $(\Omega, \mathcal{H})$ *tale che per ogni* $A_1 \in \mathcal{H}_1$ *e* $A_2 \in \mathcal{H}_2$ *si abbia* $\mu(A_1 \times A_2) = \mu_1(A_1) \cdot \mu_2(A_2)$.

Dimostrazione. La buona definizione è una conseguenza del lemma 6.1. La verifica che μ è una misura segue dalla linearità degli integrali e dal teorema di convergenza monotona. L'unicità segue dal fatto che $\mathcal{R}$ è un p-system che genera $\mathcal{H}$. □

Nelle ipotesi del teorema diremo che $\mu = \mu_1 \times \mu_2$
$(\Omega, \mathcal{H}, \mu) = (\Omega_1, \mathcal{H}_1, \mu_1) \times (\Omega_2, \mathcal{H}_2, \mu_2)$ e per induzione

$$\bigtimes_{i \in \{1,2,\dots,n\}} (\Omega_i, \mathcal{H}_i, \mu_i) = \left(\bigtimes_{i \in \{1,2,\dots,n-1\}} (\Omega_i, \mathcal{H}_i, \mu_i) \right) \times (\Omega_n, \mathcal{H}_n, \mu_n)$$

Osservazione 6.3. *Il teorema ed il lemma precedente valgono anche se si sostituisce l'ipotesi "μ_i misure finite" con l'ipotesi "μ_i misure σ-finite".*

Dimostrazione. La dimostrazione è semplice, basta spezzare gli spazi Ω_i in unione numerabile di eventi disgiunti di misura finita ed applicare a questi i teoremi e lemmi già dimostrati. □

6.2 Teorema di Fubini Tonelli

Una versione molto generale dei teoremi di Fubini e Tonelli è la seguente.

Teorema 6.4. *Siano $(\Omega_1, \mathcal{H}_1, \mu_1)$ e $(\Omega_2, \mathcal{H}_2, \mu_2)$ due spazi di misura σ-finiti. Sia $(\Omega, \mathcal{H}, \mu) := (\Omega_1, \mathcal{H}_1, \mu_1) \times (\Omega_2, \mathcal{H}_2, \mu_2)$. Sia f da $(\Omega, \mathcal{H})$ in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ misurabile. Se vale una delle seguenti 6 condizioni:*

$$\int_{\Omega_1} \int_{\Omega_2} f^+(x,y)\ d\mu_2(y)\ d\mu_1(x) < \infty \qquad \int_{\Omega_1} \int_{\Omega_2} f^-(x,y)\ d\mu_2(y)\ d\mu_1(x) < \infty$$
$$\int_{\Omega_2} \int_{\Omega_1} f^+(x,y)\ d\mu_1(x)\ d\mu_2(y) < \infty \qquad \int_{\Omega_2} \int_{\Omega_1} f^-(x,y)\ d\mu_1(x)\ d\mu_2(y) < \infty$$
$$\int_{\Omega} f^+(x,y)\ d\mu(x,y) < \infty \qquad \int_{\Omega} f^-(x,y)\ d\mu(x,y) < \infty$$

allora si ha

$$\int_{\Omega} f(x,y)\ d\mu(x,y) = \int_{\Omega_1} \int_{\Omega_2} f(x,y)\ d\mu_2(y)\ d\mu_1(x) \tag{6.4}$$
$$= \int_{\Omega_2} \int_{\Omega_1} f(x,y)\ d\mu_1(x)\ d\mu_2(y) \tag{6.5}$$

Dimostrazione. Si procede con lo schema di dimostrazione standard se $f = \mathbf{1}_A$ allora l'enunciato del teorema è vero grazie al teorema 6.2.
Se f è positiva semplice allora la tesi segue dalla linearità degli integrali.
Se f è positiva (e misurabile) allora f si approssima dal basso con funzioni semplici e la tesi segue dal teorema di convergenza monotona.
Il caso generale è un po' più delicato. Per quanto visto prima i tre integrali a destra (e a sinistra) sono uguali quindi o tutti e tre finiti o tutti e tre

infiniti. Per ipotesi allora almeno 3 degli integrali sono finiti, se sono tutti e sei finiti allora la tesi segue per la linearità degli integrali. Resta solo il caso tre integrali finiti e tre infiniti. Supponiamo che i 3 integrali di destra siano finiti e i 3 integrali di sinistra siano infiniti. L'idea è di approssimare dal basso la funzione f con delle fuzioni f_n tali che:

1. $f_n \leq f$
2. gli integrali di f_n^+ e f_n^- siano tutti finiti
3. $\int_\Omega f_n(x,y)\ d\mu(x,y) \uparrow +\infty$

La condizione 2 ci assicura che per f_n possiamo ricondurci al caso precedente, la condizione 1 ci dice che gli integrali della f sono maggiori di quelli delle f_n l'ultima condizione con la prima ci assicura che gli integrali della f devono essere infiniti(tutti e tre). Di solito la scelta più semplice per definire f_n è data da $f_n := \min(f, n)$, questa scelta funziona bene sugli spazi di probabilità e sugli spazi di misura finiti, ma in caso di spazi di misura σ-finiti per essere sicuri che la condizione 2 sia soddisfatta occorre richiedere un po' di più. Dato che μ_1 e μ_2 sono σ-finiti allora chiaramente anche μ lo sarà. Siano $\{A_n\}_{n\in\mathbb{N}}$ tali che $\mu(A_n) < \infty$ e $A_n \uparrow \Omega$ allora posto

$$f_n := f^- + \min(f^+, n) \cdot \mathbf{1}_{A_n}$$

si verifica facilmetne che le funzioni $\{f_n\}_{n\in\mathbb{N}}$ soddisfano le condizioni richieste. □

6.3 Prodotto di una famiglia di spazi di probabilità

In questa sezione studieremo il prodotto di una famiglia generica di spazi di probabilità $(\Omega_i, \mathcal{H}_i, P_i)_{i\in I}$. A differenza delle sezioni precedenti ora è cruciale che gli spazi siano di probabilità, i risultati non possono essere generalizzati al caso di spazi di misura finiti.

Teorema 6.5. *Siano $(\Omega_i, \mathcal{H}_i, P_i)_{i\in I}$ spazi di probabilità, e sia $(\Omega, \mathcal{H}) := (\times_{i\in I} \Omega_i, \otimes_{i\in I} \mathcal{H}_i)$. Esiste un'unica misura di probabilità P su $(\Omega, \mathcal{H})$ tale che le proiezioni $\{\pi_i\}_{i\in I}$ siano indipendenti ed abbiano legge P_i.*

Dimostrazione. Il caso $I = \{1, 2\}$ è stato visto nella sezione precedente.
Il caso finito $I = \{1, 2, \ldots, n\}$ si affronta per induzione su n. La definizione avviene per induzione:

$$(\overline{\Omega}_n, \overline{\mathcal{H}}_n, \overline{P}_n) = \bigtimes_{i\in\{1,\ldots,n\}} (\Omega_i, \mathcal{H}_i, P_i) := (\overline{\Omega}_{n-1}, \overline{\mathcal{H}}_{n-1}, \overline{P}_{n-1}) \times (\Omega_n, \mathcal{H}_n, P_n)$$

Vale $\overline{P}_n(A) = \int_{\Omega_1} \dots \int_{\Omega_n} \mathbf{1}_A(x_1, \dots, x_n) d\mu_n(x_n) \dots d\mu_1(x_1)$ e tutte le verifiche sono elementari.

Caso I numerabile. Il caso I Numerabile è quello più difficile. Sia $I = \mathbb{N}$. Consideriamo la successione di σ-algebre $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \dots \subseteq \mathcal{G}_n \subseteq \dots \subseteq \mathcal{H}$ e l'algebra $\mathcal{R}$ date da:

$$\mathcal{G}_n := \sigma(\pi_1, \dots, \pi_n) \qquad\qquad \mathcal{R} := \bigcup_{n \in \mathbb{N}} \mathcal{G}_n$$

$\mathcal{R}$ è un'algebra che genera $\mathcal{H}$. A parole $\mathcal{G}_n$ è costituito dagli elementi A di $\mathcal{H}$ che 'dipendono' solo dalle prime n coordinate, ovvero per ogni $A \in \mathcal{H}$ vale:

$$A \in \mathcal{G}_n \Leftrightarrow \left(\forall x = (x_n)_{n \in \mathbb{N}}, y = (y_n)_{n \in \mathbb{N}} \in \Omega \text{ tali che } x_1 = y_1, \dots, x_n = y_n \text{ vale } x \in A \Leftrightarrow y \in A \right)$$

Sia $\overline{\pi}_n : \Omega \to \overline{\Omega}_n$ la proiezione naturale ovvero $\overline{\pi}_n((x_i)_{i \in \mathbb{N}}) = (x_1, \dots, x_n)$. Vale

$$\mathcal{G}_n = \{A \in \mathcal{H} | \overline{\pi}_n^{-1}(\overline{\pi}_n(A)) = A\}$$

Definiamo ora la funzione P_0 da $\mathcal{R}$ in $[0,1]$ che poi sarà estesa con il teorema di estensione di Caratheodory a misura di probabilità su tutto $\mathcal{H}$:

$$\forall A \in \mathcal{G}_n \quad P_0(A) := \overline{P}_n(\overline{\pi}_n(A))$$

La buona definizione di P_0 è immediata perché se $A \in \mathcal{G}_n$ e $m > n$ allora $A \in \mathcal{G}_m$ e vale $\overline{P}_n(\overline{\pi}_n(A)) = \overline{P}_m(\overline{\pi}_m(A))$. $\mathcal{R}$ è un'algebra che genera $\mathcal{H}$, la dimostrazione del caso I numerabile si chiude applicando il teorema di estensione di Caratheodory, l'unica cosa difficile per poter applicare tale teorema è la σ-additività di P_0. Dedicheremo il resto della dimostrazione a provare la σ-additività di P_0. Sia $A = \bigcup_{n=1}^{\infty} A_n$ con $A, A_n \in \mathcal{R}$ e A_n disgiunti. Sia $A_0 := A^c \in \mathcal{R}$ cosicché $\Omega = \bigcup_{n=0}^{\infty} A_n$, per ottenere la subadditività di P_0 occorre dimostrare che $\sum_{n=0}^{\infty} P_0(A_n) = 1$.

Dimostreremo che data una successione $\{A_n\}_{n \geq 0}$ di eventi disgiunti con $A_n \in \mathcal{R}$ se $\sum_{n=0}^{\infty} P_0(A_n) = p < 1$ allora esiste $\overline{x} \in \Omega$ tale che $\overline{x} \notin \bigcup_{n=0}^{\infty} A_n =: B$.

Sia $B_n := \bigcup_{k=0}^{n} A_k$ quindi $P_0(B_n) \uparrow p$. Sia inoltre k_n una successione crescente tale che $B_n \in \mathcal{G}_{k_n}$. Ora determineremo $\overline{x} = (\overline{x}_n)_{n \geq 1}$ una componente per volta. Cominciamo dal cercare $\overline{x}_1$. Sia:

$$f_n(x_1) := \int_{\Omega_2} \dots \int_{\Omega_{k_n}} \mathbf{1}_{\overline{\pi}_{k_n}(B_n)}(x_1, \dots, x_{k_n}) d\mu_{k_n}(x_{k_n}) \dots d\mu_2(x_2) \tag{6.6}$$

cosicché $f_n(x_1)$ è non decrescente in n e vale $\int_{\Omega_1} f_n(x_1) d\mu_1(x_1) = P_0(B_n)$. consideriamo il limite in n di f

$$f(x_1) = \lim_n f_n(x_1)$$

Vale $f(x_1) \in [0,1]$ e per il teorema di convergenza monotona

$$\int_{\Omega_1} f(x_1) d\mu_1(x_1) = p$$

quindi esiste $\overline{x}_1$ tale che $f(\overline{x}_1) \leq p$. $\overline{x}_1$ è la prima componente di $\overline{x}$. Per determinare la seconda componente riprendiamo l'equazione (6.6) come funzione questa volta di x_2 fissando $x_1 = \overline{x}_1$ e integrendo sulle variabili di indice maggiore di 2:

$$f_n^{(\overline{x}_1)}(x_2) := \int_{\Omega_3} \dots \int_{\Omega_{k_n}} \mathbf{1}_{\overline{\pi}_{k_n}(B_n)}(\overline{x}_1, x_2, \dots, x_{k_n}) d\mu_{k_n}(x_{k_n}) \dots d\mu_3(x_3)$$

si ha:

$$\int_{\Omega_2} f_n^{(\overline{x}_1)}(x_2) d\mu_2(x_2) = f(\overline{x}_1) \leq p$$

se si pone $f^{(\overline{x}_1)}(x_2) = \lim_n f_n^{(\overline{x}_1)}(x_2)$ si ha

$$\int_{\Omega_2} f^{(\overline{x}_1)}(x_2) d\mu_2(x_2) \leq p$$

quindi esiste $\overline{x}_2$ tale che $f^{(\overline{x}_1)}(\overline{x}_2) \leq p$ procediamo ora per induzione su m

$$f_n^{(\overline{x}_1,\dots,\overline{x}_{m-1})}(x_m) := \int_{\Omega_{m+1}} \dots \int_{\Omega_{k_n}} \mathbf{1}_{\overline{\pi}_{k_n}(B_n)}(\overline{x}_1, \dots, \overline{x}_{m-1}, x_m, \dots, x_{k_n}) d\mu_{k_n}(x_{k_n}) \dots d\mu_{m+1}(x_{m+1})$$

per $n > m$ si ha $k_n > m$ e l'equazione precedente è ben posta. Inoltre in maniera analoga a quanto fatto per x_2 posto $f^{(\overline{x}_1,\dots,\overline{x}_{m-1})}(x_m) = \lim_n f_n^{(\overline{x}_1,\dots,\overline{x}_{m-1})}(x_m)$ esiste $\overline{x}_m$ tale che $f_n^{(\overline{x}_1,\dots,\overline{x}_{m-1})}(\overline{x}_m) \leq p$.
Abbiamo così costruito il punto $\overline{x}$, occorre ora mostrare che non appartiene a nessun insieme B_n. Procediamo per assurdo se $\overline{x} \in B_{\overline{n}}$ allora poiché $B_{\overline{n}}$ è in $\mathcal{G}_{k_{\overline{n}}}$ si ha che per ogni $\{x_i\}_{i > k_{\overline{n}}}$ il punto $(\overline{x}_1, \dots, \overline{x}_{k_{\overline{n}}}, x_{k_{\overline{n}}+1}, \dots)$ appartiene B_n per ogni $n > \overline{n}$ quindi

$$f_n^{(\overline{x}_1,\dots,\overline{x}_{k_{\overline{n}}-1})}(\overline{x}_{k_{\overline{n}}}) = 1 \qquad \forall n > \overline{n}$$

quindi

$$f^{(\overline{x}_1,\dots,\overline{x}_{k_{\overline{n}}-1})}(\overline{x}_{k_{\overline{n}}}) = 1$$

contraddicendo la condizione $f^{(\overline{x}_1,\dots,\overline{x}_{k_{\overline{n}}-1})}(\overline{x}_{k_{\overline{n}}}) \leq p$

Caso I più che numerabile. Il caso I più che numerabile si deduce dal caso precedente poiché per ogni $A \in \mathcal{H}$ esiste $J \subset I$ al più numerabile tale che $A \in \sigma(\{\pi_j\}_{j \in J})$. Le verifiche sono tutte più o meno semplici.

$\square$

Capitolo 7

Convergenza di successioni di variabili aleatorie

7.1 Convergenza in distribuzione

7.2 Convergenza in probabilità

7.3 Convergenza quasi certa

7.4 convergenza in L^p

7.5 Relazioni tra le convergenze

7.6 Caratterizzazione della convergenza in distribuzione

7.7 Criterio di convergenza in distribuzione

7.8 Teorema di Prokorov

7.9 Legge dei grandi numeri

7.10 Funzione caratteristica

7.11 Formula di inversione

7.12 Teorema di continuità di Lévy

7.13 Teorema del limite centrale

Appendice A

Somme infinite

In questo capitolo generalizzeremo il concetto classico di serie. Durante i corsi del primo anno la serie è stata definita come la somma in successione di numeri. Poteva succedere in alcuni casi speciale che riordinando gli elementi della serie il risultato della somma cambiasse. Per le applicazioni in probabilità l'esito di una somma non può dipendere dall'ordine in cui vengono sommati gli elementi, inoltre occorrerà introdurre una definizione più solida di somma che ci permetta di sommare gli elementi di una famiglia di numeri $\{a_x\}_{x\in\Omega}$ su un insieme di indici Ω che non ha un ordinamento ed eventualmente ha una cardinalità più che numerabile.

Possiamo giungere alla definizione corretta procedendo per gradi e ragionandoci su. Cominciamo con il caso delle successione positive, sia dunque $\{a_n\}_{n\in\mathbb{N}}$ una successione di numeri $a_n \geq 0$. Consideriamo le seguenti due definizioni

$$\text{def 1} \qquad \sum_{n=1}^{\infty} a_n := \lim_{N\to\infty} \sum_{n=1}^{N} a_n$$

$$\text{def 2} \qquad \sum_{n=1}^{\infty} a_n := \sup_{\substack{I\subseteq\mathbb{N}\\ I \text{ finito}}} \sum_{n\in I} a_n$$

È facile dimostrare che (per $a_n \geq 0$) le due definizioni sono equivalenti. Abbiamo così trovato una definizione (la seconda) che è equivalente alla definizione classica (quando $a_n \geq 0$) e che come vedremo può essere generalizzata per definire la somma su un insieme generico di indici.

Definizione A.1. Sia Ω un insieme (per esempio $\Omega = \mathbb{R}$) e sia $\{a_\omega\}_{\omega\in\Omega}$ una famiglia di numeri con $a_\omega \geq 0$. Definiamo la somma degli a_ω su Ω come:

$$\sum_{\omega\in\Omega} a_\omega := \sup_{\substack{I\subseteq\Omega\\ I \text{ finito}}} \sum_{\omega\in I} a_\omega$$

Con la definizione A.1 abbiamo definito la somma nel caso Ω generico e numeri positivi. Diamo adesso la definizione nel caso generale.

Definizione A.2. Sia Ω un insieme e sia $\{a_\omega\}_{\omega\in\Omega}$ una famiglia di numeri reali. Grazie alla definizione A.1 possiamo considerare le seguenti due somme a termini positivi:

$$\sum_{\omega\in\Omega} a_\omega^+ \qquad\qquad \sum_{\omega\in\Omega} a_\omega^-$$

se sono entrambe uguali a $+\infty$ diremo che $\{a_\omega\}_{\omega\in\Omega}$ non ammette somma altrimenti definiamo

$$\sum_{\omega\in\Omega} a_\omega := \sum_{\omega\in\Omega} a_\omega^+ - \sum_{\omega\in\Omega} a_\omega^-$$

con le solite convenzioni: $\infty - c = \infty$ e $c - \infty = -\infty$.

La seguente proposizione afferma che se una successione ammette somma allora la definizione classica e quella da noi introdotta sono equivalenti.

Proposizione A.1. *Sia $\{a_n\}_{n\in\mathbb{N}}$ una successione di numeri reali. Se $\{a_n\}_{n\in\mathbb{N}}$ ammette somma allora*

$$\lim_{N\to\infty} \sum_{n=1}^{N} a_n = \sup_{\substack{I\subseteq\mathbb{N} \\ I \text{ finito}}} \sum_{n=1}^{N} a_n^+ - \sup_{\substack{I\subseteq\mathbb{N} \\ I \text{ finito}}} \sum_{n=1}^{N} a_n^-$$

Può succedere tuttavia che una successione che non ammetta somma converga invece nel senso classico. Un esempio è il seguente.

Esempio A.1. *Sia $\{a_n\}_{n\in\mathbb{N}}$ la successione definita da $a_n := \frac{(-1)^n}{n}$, si dimostra che:*

- $\sum_{n\in\mathbb{N}} a_n^+ = \sum_{n\in\mathbb{N}} a_n^- = +\infty$
- $\lim_{N\to\infty} \sum_{n=1}^{N} a_n \in \mathbb{R}$

A.1 Somma a blocchi

Enunceremo ora una proprietà delle somme infinite detta somma a blocco. Si tratta di una generalizzazione della proprietà associativa al caso infinito.

Proposizione A.2. *Sia $\{a_\omega\}_{\omega\in\Omega}$ una famiglia di numeri reali che ammette somma, sia I un insieme non vuoto e sia $\{\Omega_i\}_{i\in I}$ una partizione di Ω allora vale*

$$\sum_{\omega\in\Omega} a_\omega := \sum_{i\in I}\left(\sum_{\omega\in\Omega_i} a_\omega\right)$$

La dimostrazione è molto articolata e non verrà trattata in queste note. Tuttavia per i curiosi, è possibile trovare una dimostrazione nell'appendice del libro [1].

Appendice B

Teorema di estensione di Carathéodory ed esistenza della misura di Lebesgue

Il teorema di estensione di Carathéodory è necessario per dimostrare l'esistenza della misura di Lebesgue in R^n. La sua dimostrazione non fa parte del programma del corso di calcolo delle probabilità. Tuttavia ho pensato di inserire in appendice una dimostrazione che a me piace molto e che fa uso del lemma di Dynkin piuttosto che del metodo di Carathéodory.

Teorema B.1 (Teorema di estensione di Carathéodory)**.** *Sia Ω un insieme non vuoto. Sia $\mathcal{A}$ un'algebra di Ω e sia $\mathcal{G} = \sigma(\mathcal{A})$. Se $\mu_0 : \mathcal{A} \to [0, \infty]$ è una funzione numerabilmente additiva allora esiste μ misura su $(\Omega, \mathcal{G})$ tale $\mu = \mu_0$ su $\mathcal{A}$.*

Numerabilmente additiva vuol dire che per ogni successione $\{A_n\}_{n\in\mathbb{N}}$ di elementi disgiunti di $\mathcal{A}$ vale

$$\bigcup_n A_n \in \mathcal{A} \qquad \Longrightarrow \qquad \mu_0\left(\bigcup_n A_n\right) = \sum_n \mu_0(A_n)$$

Dimostrazione teorema B.1

Per semplicità dimostreremo solo il caso $\mu_0(\Omega) = 1$.
Inizieremo studiando la chiusura $\mathcal{B}$ di $\mathcal{A}$ rispetto all'unione numerabile. Sono

cose equivalenti:

$$\mathcal{B} := \left\{ B \in \mathcal{P}(\Omega) \middle| \exists \{A_n\}_{n\in\mathbb{N}} \text{ con } A_n \in \mathcal{A} \text{ t.c. } B = \bigcup_n A_n \right\}$$
$$\mathcal{B} := \left\{ B \in \mathcal{P}(\Omega) \middle| \exists \{B_n\}_{n\in\mathbb{N}} \text{ disgiunti con } B_n \in \mathcal{A} \text{ t.c. } B = \bigcup_n B_n \right\}$$
$$\mathcal{B} := \left\{ B \in \mathcal{P}(\Omega) \middle| \exists \{C_n\}_{n\in\mathbb{N}} \text{ crescenti con } C_n \in \mathcal{A} \text{ t.c. } B = \bigcup_n C_n \right\}$$

La collezione $\mathcal{B}$ è la chiusura di $\mathcal{A}$ rispetto all'unione numerabile. L'equivalenza tra le 3 definizioni è dovuta al fatto che $\mathcal{A}$ è un'algebra, notiamo inoltre che tra le successioni disgiunte B_n e quelle crescenti C_n c'è una corrispondenza biunivoca data da $C_n = B_1 \cup \ldots \cup B_n$ e $B_n = C_n \backslash C_{n-1}$.
La collezione $\mathcal{B}$ contiene $\mathcal{A}$, è chiusa per unioni numerabili e intersezioni finite mentre in generale non è chiusa rispetto al complementare e all'intersezione numerabile. Estendiamo ora μ_0 a $\mathcal{B}$ (se vogliamo che alla fine sia una misura allora l'estensione possibile è unica ed è la seguente):

$$\forall B = \bigcup_{n\in\mathbb{N}} A_n \text{ con } A_n \text{ disgiunti in } \mathcal{A} \text{ poniamo } \overline{\mu}(B) := \sum_n \mu_0(A_n) \tag{B.1}$$

Claim *La definizione B.1 di $\overline{\mu} : \mathcal{B} \to [0, \infty]$ è una buona definizione.*

Dimostrazione. Occorre verificare che $\overline{\mu}(B)$ non dipende dalla scelta della successione A_n. Osserviamo innanzitutto che posto $B_n = A_1 \cup \ldots \cup A_n$ allora vale $\mu_0(B_n) \uparrow \sum_n \mu_0(A_n)$. Dunque data la corrispondenza biunivoca tra successioni disgiunte e successioni crescenti possiamo verificare la buona posizione sulle successioni crescenti. Siano B_n e C_n crescenti in $\mathcal{A}$ tali che $B_n \uparrow B$ e $C_n \uparrow B$, fissato m vale $\lim_n (B_n \cap C_m) = C_m$ quindi per ogni m vale $\mu_0(C_m) \leq \lim_n \mu_0(B_n)$ e in maniera analoga $\mu_0(B_m) \leq \lim_n \mu_0(C_n)$ dunque $\lim_n \mu_0(B_n) = \lim_n \mu_0(C_n)$. □

Vediamo ora che sebbene $\mathcal{B}$ non sia chiuso per intersezione numerabile e complementare, $\overline{\mu}$ soddisfa alcune tipiche proprietà che dipendono solo da intersezione finita e unione numerabile.
Claim *Sone vere le seguenti proprietà di $\overline{\mu}$ su $\mathcal{B}$:*

$$\textit{per ogni } B, C \in \mathcal{B} \textit{ vale } \overline{\mu}(B) + \overline{\mu}(C) = \overline{\mu}(B \cup C) + \overline{\mu}(B \cap C) \tag{B.2}$$
$$\textit{per ogni } \{B_n\}_{n\in\mathbb{N}} \textit{ crescente in } \mathcal{B} \textit{ vale } \overline{\mu}(\bigcup_n B_n) = \lim_n \overline{\mu}(B_n) \tag{B.3}$$
$$\textit{per ogni } \{C_n\}_{n\in\mathbb{N}} \textit{ disgiunti in } \mathcal{B} \textit{ vale } \overline{\mu}(\bigcup_n C_n) = \sum_n \overline{\mu}(C_n) \tag{B.4}$$
$$\textit{per ogni } \{D_n\}_{n\in\mathbb{N}} \textit{ in } \mathcal{B} \textit{ vale } \overline{\mu}(\bigcup_n D_n) \leq \sum_n \overline{\mu}(D_n) \tag{B.5}$$

Dimostrazione. Cominciamo da (B.2). Siano B e C due elementi di $\mathcal{B}$ siano B_n e C_n successioni di elementi di $\mathcal{A}$ tali che $B_n \uparrow B$ e $C_n \uparrow C$ allora vale $\mu_0(B_n)+\mu_0(C_n) = \mu_0(B_n \cup C_n)+\mu_0(B_n \cap C_n)$ passando al limite si ottiene la (B.2). Dimostrazione (B.3) e (B.4), in $\mathcal{B}$ non c'è corrispondenza biunivoca tra successioni crescenti e successioni disgiunte questo avviene perché $\mathcal{B}$ è chiuso rispetto all'unione ma non è chiuso rispetto alla differenza tra insiemi, la corrispondenza avviene solo in un verso, occorre dimostrare prima (B.3) e poi da (B.3) si può dedurre (B.4). Supponiamo che $\{B_n\}_{n\in\mathbb{N}}$ sia una successione crescente di elementi di $\mathcal{B}$, sia $B := \bigcup_{n\in\mathbb{N}} B_n$ e siano $A_{n,m}$ elementi di $\mathcal{A}$ tali che per ogni n vale $A_{n,m} \uparrow B_n$ allora posto $C_n := \bigcup_{n=1,\ldots,m} A_{n,m}$ si ha $C_n \uparrow B$ e $\overline{\mu}(B) = \lim_n \mu_0(C_n) = \lim_n \overline{\mu}(B_n)$. Dimostriamo ora (B.5), sia $B_n := D_1 \cup \ldots \cup D_n$ dalla (B.2) si ottiene $\overline{\mu}(B_n) \le \sum_{k=1}^{n} \overline{\mu}(D_k)$ e da (B.3) segue la tesi. □

A questo punto occorre estendere $\overline{\mu}$ a tutto $\mathcal{G}$. Procedere chiudendo $\mathcal{B}$ rispetto all'intersezione numerabile non porterebbe a nulla perchè la nuova collezione non sarebbe chiusa rispetto all'unione numerabile e così via. Procediamo approssimando gli insiemi di $\mathcal{G}$ dall'esterno con insiemi di $\mathcal{B}$, la nuova funzione μ (B.6) è detta misura esterna:

$$\forall C \in \mathcal{P}(\Omega) \qquad \mu(C) := \inf\{\overline{\mu}(B) | B \in \mathcal{B}, B \supseteq C\} \tag{B.6}$$

La misura esterna μ coincide con $\overline{\mu}$ su $\mathcal{B}$ inoltre dalla proprietà (B.2) si deduce che μ è monotona e subadditiva ovvero: se $B \subseteq C$ allora $\mu(B) \le \mu(C)$; e per ogni B e C vale $\mu(B \cup C) \le \mu(B) + \mu(C)$. Per completare la dimostrazione occorre mostrare che μ è numerabilmente additiva. Per capire l'idea che c'è sotto ai passaggi che seguiranno bisogna tenere conto che non è possibile richiedere che μ sia numerabilmente additiva su tutti i sottoinsiemi di Ω, questo avviene perché in $\mathcal{P}(\Omega)$ ci sono insiemi "cattivi" C il cui bordo è così frastagliato che la misura esterna μ non riesce a separare C dal suo complementare cosicché si verifica $\mu(C) + \mu(C^c) > 1$. L'idea è che C è "cattivo" quando ci sono delle zone in cui C e C^c sono talmente intreccati che la misura esterna μ le conta due volte, una volta in $\mu(C)$ e una in $\mu(C^c)$. Dimostreremo che in $\mathcal{G}$ non ci sono insiemi "cattivi", e sugli insiemi "buoni" la misura esterna μ è numerabilmente additiva.

Sia $\mathcal{D}$ la collezione di insiemi "buoni":

$$\mathcal{D} := \{B \in \mathcal{P}(\Omega) | \mu(B) + \mu(B^c) = 1\}$$

La definizione di μ e l'equazione (B.2) permettono di dedurre la seguente caretterizzazione di $\mathcal{D}$:

$$B \in \mathcal{D} \iff \forall \epsilon > 0\ \exists B_1, B_2 \in \mathcal{B} \text{ t.c. } B_1 \supseteq B, B_2 \supseteq B^c, \overline{\mu}(B_1 \cap B_2) < \epsilon \tag{B.7}$$

in pratica un insieme B appartiene a $\mathcal{D}$ se sia B che B^c possano essere approssimati da insiemi di $\mathcal{B}$ con intersezione piccola.

Claim *Vale l'inclusione:* $\mathcal{D} \supseteq \mathcal{B}$

Dimostrazione. Sia $B \in \mathcal{B}$, per la subadditività vale $\mu(B) + \mu(B^c) \geq 1$ occorrerà dimostrare che vale anche l'altra disuguaglianza. Sia $A_n \uparrow B$ con $A_n \in \mathcal{A}$ allora vale $\mu_0(A_n) \uparrow \overline{\mu}(B)$, $A_n^c \supseteq B^c$ quindi $\mu(B^c) \leq \mu_0(A_n^c) = 1 - \mu_0(A_n)$ e passando al limite

$$\mu(B) + \mu(B^c) \leq 1$$

□

Per il lemma di Dynkin vale:

$$\mathcal{G} = \sigma(\mathcal{A}) = \sigma(\mathcal{B}) = d(\mathcal{B}) = d(\mathcal{A})$$

Per dimostrare che $\mathcal{D}$ contiene $\mathcal{G}$ basterà dimostrare che $\mathcal{D}$ è un d-system. La collezione $\mathcal{D}$ è evidentemente chiusa rispetto al complementare quindi per dimostrare che è un d-system basterà dimostrare che è chiuso rispetto all'unione numerabile di insieme disgiunti, dimostreremo contemporaneamente che $\mathcal{D}$ è chiuso rispetto all'unione numerabile di insiemi disgiunti e che μ è additiva sulle successioni di insiemi disgiunti.

Cominciamo con l'unione finita. Siano B e C due elementi disgiunti di $\mathcal{D}$ allora per la proprietà (B.7) esistono in $\mathcal{B}$ quattro insiemi $B_1 \supseteq B$, $B_2 \supseteq B^c$, $C_1 \supseteq C$, $C_2 \supseteq C^c$ tali che:

$$\overline{\mu}(B_1 \cap B_2) \leq \epsilon \qquad \overline{\mu}(C_1 \cap C_2) \leq \epsilon$$

posto $D := B_1 \cup C_1$ e $E := B_2 \cap C_2$ si ha $D \supseteq (B \cup C)$ e $E \supseteq (B \cup C)^c$ e

$$\overline{\mu}(D \cap E) = \overline{\mu}((B_1 \cup C_1) \cap (B_2 \cap C_2)) \leq \overline{\mu}(B_1 \cap B_2) + \overline{\mu}(C_1 \cap C_2) \leq 2\epsilon$$

quindi per la per la proprietà (B.7) $B \cup C$ appartiene a $\mathcal{D}$.
Adesso dobbiamo mostrare che $\mu(B \cup C) = \mu(B) + \mu(C)$, dati B_1, B_2, C_1 e C_2 come sopra, per la (B.2) si ha

$$\begin{aligned} \overline{\mu}(B_1) &\leq \mu(B) + \epsilon \qquad & \overline{\mu}(C_1) &\leq \mu(C) + \epsilon \\ \overline{\mu}(B_2) &\leq \mu(B^c) + \epsilon \qquad & \overline{\mu}(C_2) &\leq \mu(C^c) + \epsilon \end{aligned}$$

sia inoltre $G \in \mathcal{B}$ tale che $G \supseteq (B \cup C)$ e $\overline{\mu}(G) \leq \mu(B \cup C) + \epsilon$.
Poniamo $\widetilde{B} := B_1 \cap C_2 \cap G$, $\widetilde{C} := C_1 \cap B_2 \cap G$ e $\widetilde{G} := \widetilde{B} \cup \widetilde{C}$ allora si ha:

$$B_1 \supseteq \widetilde{B} \supseteq B \qquad C_1 \supseteq \widetilde{C} \supseteq C \qquad G \supseteq \widetilde{G} \supseteq (B \cup C)$$

inoltre per costruzione $\overline{\mu}(\widetilde{B} \cap \widetilde{C}) < \epsilon$ da cui si ricava:

$$\overline{\mu}(G) \geq \overline{\mu}(\widetilde{G}) \geq \overline{\mu}(\widetilde{B}) + \overline{\mu}(\widetilde{C}) - \epsilon \geq \mu(B) + \mu(C) - \epsilon$$

quindi $\mu(B \cup C) \geq \mu(B) + \mu(C)$.
A questo punto manca solo l'unione numerabile di insieme disgiunti, siano

$\{C_n\}$ in $\mathcal{D}$ disgiunti, sia $C := \bigcup_n C_n$. Per quanto detto sull'unione finita per ogni n vale $\mu(\bigcup_{k=1}^n C_k) = \sum_{k=1}^n \mu(C_k)$ poiché $C_n \subseteq C$ allora vale

$$\mu(C) \geq \sum_n \mu(C_n)$$

Dato $\epsilon > 0$ sia $B_n \in \mathcal{B}$ tali che

$$\overline{\mu}(B_n) < \mu(C_n) + \frac{\epsilon}{2^n}$$

allora per la (B.5) vale:

$$\mu(C) \leq \overline{\mu}(\bigcup_n B_n) \leq \sum_n \overline{\mu}(B_n) < \sum_n \mu(C_n) + \epsilon$$

quindi vale $\mu(C) = \sum_n \mu(C_n)$. Resta da mostrare che C appartiene a $\mathcal{D}$.

$$\mu(C^c) \leq \overline{\mu}\left(\left(\bigcup_{k=1}^n C_k\right)^c\right) = 1 - \overline{\mu}\left(\bigcup_{k=1}^n C_k\right)$$

passando al limite si ottiene $\mu(C^c) + \mu(C) \leq 1$ poiché per la subadditività vale anche l'altra disuguaglianza allora si ha $\mu(C^c) + \mu(C) = 1$.

Esistenza della misura di Lebesgue

Teorema B.2. *Esiste una misura μ su $([0,1], \mathcal{B}([0,1]))$ tale che per ogni $0 \leq a \leq b \leq 1$ vale $\mu(a,b) = b - a$.*

La dimostrazione che segue consta di due parti nella prima si dimostra che la definizione (B.8) è una buona definizione e μ_0 è una funzione additiva. Nella seconda parte si dimostra che μ_0 è σ-additiva. Sul libro di riferimento il Williams, la dimostrazione della prima parte è lasciata al lettore e la seconda è svolta in maniera diversa.

Dimostrazione. Per cominciare possiamo considerare la seguente algebra $\mathcal{A}$:

$$\mathcal{A} := \{[a_1, b_1) \cup [a_2, b_2) \cup \ldots \cup [a_n, b_n) | n \in \mathbb{N}, 0 \leq a_1 < b_1 \leq a_2 < b_2 \ldots < b_n \leq 1\}$$

Abbiamo già visto che l'algebra $\mathcal{A}$ genera i boreliani. È chiaro anche che la misura richiesta dal teorema esiste se e solo se per ogni $a < b$ in $[0,1]$ vale $\mu([a,b)) = b - a$. Inoltre per ogni a deve valere $\mu(\{a\}) = 0$ quindi $\mu(\{1\}) = 0$. Senza perdere in generalità possiamo costruire la misura su $([0,1), \mathcal{B}([0,1)))$. Consideriamo la seguente funzione $\mu_0 : \mathcal{A} \to [0, \infty]$:

$$\forall A \in \mathcal{A} \quad A = [a_1, b_1) \cup [a_2, b_2) \cup \ldots \cup [a_n, b_n) \qquad \mu_0(A) := \sum_{k=1}^n (b_k - a_k) \quad \text{(B.8)}$$

Per completare la dimostrazione bisognerà dimostrare che la precedente definizione è una buona definizione e che la funzione μ_0 è σ-additiva. Dopo

di che la tesi seguirà in modo immediato dal teorema di estensione di Carathéodory.
La dimostrazione della buona definizione di (B.8) è semplice ma piuttosto noiosa. Il punto chiave è che un insieme $A \in \mathcal{A}$ può avere più rappresentazioni della forma (B.8), per esempio se $a < b < c$ allora $[a, c) = [a, b) \cup [b, c)$, bisogna verificare che qualunque sia la rappresentazione il risultato della somma non cambia. Nell'esempio precedente la verifica è immediata: $\mu_0([a, b) \cup [b, c)) = (b - a) + (c - b) = c - a = \mu_0([a, c))$. In generale un insieme $A \in \mathcal{A}$ ammette una sola rappresentazione con n minimo, tutte le altre rappresentazioni possono essere ricondotte a quella minima unendo uno alla volta gli intervalli consegutivi.
Dalla buona definizione segue immediatamente che μ_0 è additiva.
Dimostrare che μ_0 è σ-additiva è più difficile. Sappiamo che μ_0 è additiva e questo ci tornerà molto utile. Per cominciare μ_0 additiva sugli insiemi disgiunti ci dà la subadditività sugli insiemi non disgiunti e la monotonia rispetto all'inclusione. Sia ora $A \in \mathcal{A}$ e supponiamo $A = \bigcup_{n \in \mathbb{N}} A_n$ con gli $A_n \in \mathcal{A}$ disgiunti. Occorre mostrare che $\mu_0(A) = \sum_n \mu_0(A_n)$. Poiché $(\bigcup_{n \in \mathbb{N}} A_n) \cup A^c = [0, 1)$ e $\mu_0(A) + \mu_0(A^c) = \mu_0([0, 1)) = 1$ allora senza perdere in generalità possiamo supporre $A = [0, 1)$.
Indichiamo con $\{[a_n, b_n)\}_{n \in \mathbb{N}}$ la famiglia di tutti gli intervallini degli insiemi $\{A_n\}_{n \in \mathbb{N}}$. Abbiamo:

$$n \neq m \implies [a_n, b_n) \cap [a_m, b_m) = \varnothing \qquad \text{e} \qquad \bigcup_{n \in \mathbb{N}} [a_n, b_n) = [0, 1)$$

occorre mostrare che $\sum_n (b_n - a_n) = 1$. Una disuguaglianza è più semplice, poiché μ_0 è additiva allora

$$\sum_{k=1}^{n} (b_k - a_k) = \mu\left(\bigcup_{k=1}^{n} [a_n, b)\right) \leq 1$$

quindi vale

$$\sum_n (b_n - a_n) \leq 1$$

occorre ora mostrare che tale somma non può essere strettamente minore di 1.
La questione centrale è se sia possibile ricoprire $[0, 1)$ con una successione di intervallini $[a_n, b_n)$ di modo che la somma totale delle lunghezze sia strettamente minore di 1?
L'idea della dimostrazione che seguirà è di allargare un pochino gli intervalli $[a_n, b_n)$ e in modo da renderli aperti, a questo punto utilizzeremo il principio che permette di estrarre un ricoprimento finito da un ricoprimento di aperti, e sul ricoprimento finito potremo applicare l'additività di μ_0.
Sia $\varepsilon > 0$ piccolo. Per ogni n sia

$$c_n = a_n - \varepsilon(b_n - a_n) \qquad \text{e} \qquad d_n = b_n$$

cosicché per ogni n si ha

$$(c_n, d_n) \supset [a_n, b_n) \qquad \text{e} \qquad (d_n - c_n) = (1+\varepsilon)\cdot(b_n - a_n)$$

$$\sum_n (d_n - c_n) = (1+\varepsilon)\cdot \sum_n (b_n - a_n)$$

A questo punto poiché $\{(c_n, d_n)\}_{n\in\mathbb{N}}$ è un ricoprimento di aperti di $[0, 1-\varepsilon]$ esiste $\overline{n}$ tale che

$$\bigcup_{n=1}^{\overline{n}} (c_n, d_n) \supseteq [0, 1-\varepsilon]$$

per la subattività di μ_0

$$\sum_{n=1}^{\overline{n}} (b_n - a_n) = \frac{1}{1+\epsilon}\sum_{n=1}^{\overline{n}} (d_n - c_n) = \frac{1}{1+\epsilon}\sum_{n=1}^{\overline{n}} \mu_0([c_n - d_n)) \geq \frac{\mu_0([0, 1-\varepsilon))}{1+\varepsilon} \geq \frac{1-\varepsilon}{1+\varepsilon}$$

Per l'arbitrarietà di $\varepsilon > 0$ si ha:

$$\sum_n (b_n - a_n) \geq 1$$

$\square$

Bibliografia

[1] F. Caravenna, P. Dai Pra Probabilità, un introduzione attraverso modelli e applicazioni, *Springer*

[2] David Williams Probability with Martingales *Cambridge mathematical textbooks*