

Calcolo delle probabilità

Tiziano Vargiolu
Dipartimento di Matematica Pura ed Applicata
via Trieste, 63 - 35121 Padova
email: `vargiolu@math.unipd.it`

Anno Accademico 2011/2012

Indice

Introduzione	ii
1 Probabilità e misura	1
1.1 Prime definizioni	1
1.2 Funzioni misurabili e variabili aleatorie	4
1.3 Variabili aleatorie semplici	8
1.4 Valore atteso, o speranza matematica, e integrazione	10
1.5 Teoremi di limite	12
1.6 Speranza condizionale	13
1.7 Spazi L^p , varianza e covarianza	17
1.8 Misure di Borel sulla retta reale	21
1.9 Funzione di ripartizione, pseudoinversa e quantili	23
1.10 Densità e misure assolutamente continue	24
1.11 Legge di una variabile aleatoria	25
1.12 Probabilità condizionale ed indipendenza	27
1.13 Vettori aleatori	32
2 Catene di Markov	33
2.1 Speranza condizionale	33
2.2 Legami tra legge condizionale e speranza condizionale	38
2.3 Introduzione ai processi stocastici	38
2.4 Tempi di arresto	39
2.5 Martingale	40
2.6 Catene di Markov	43
2.7 Classificazione degli stati	46
2.8 Misure invarianti	48
2.9 Arresto ottimale	49

Introduzione

Il calcolo delle probabilità è una disciplina matematica che ha circa 500 anni di storia, e nel corso di tutto questo tempo ci sono stati vari approcci ad essa. In particolare nel XX secolo, sia per applicazioni che venivano per la maggior parte (ma non solo) dalla fisica, sia per la crescente esigenza di dimostrazioni rigorose di “fenomeni matematici” conosciuti da tempo, sono stati proposti più approcci per formalizzare il calcolo delle probabilità. Matematicamente parlando, questi approcci si possono ricondurre a due formalizzazioni, delle quali di fatto oggi si usa solo una. Questa formalizzazione “vincitrice” è quella di Kolmogorov, che negli anni '30 riconduce gli assiomi del calcolo delle probabilità a quelli della teoria della misura, formalizzata una decina di anni prima nella generalità che conosciamo ora. L'altra formalizzazione che citiamo è quello dell'italiano Bruno de Finetti, che di fatto rifiuta di usare la teoria della misura e propone una probabilità finitamente additiva, oltre ad un modo totalmente diverso di vedere le variabili aleatorie. Citiamo questo solo come nota storica, in quanto da ora in poi seguiremo strettamente l'approccio di Kolmogorov, che ormai è l'approccio standard al moderno calcolo delle probabilità.

Nel seguito supporremo che il lettore sia già familiare con la teoria della misura alla Lebesgue. Daremo quindi le definizioni solo “in chiave probabilistica”, e il lettore potrà trovare parecchi paralleli tra questa teoria e la teoria della misura.

Capitolo 1

Probabilità e misura

1.1 Prime definizioni

Definizione 1.1 Prendiamo un insieme Ω e una famiglia $\mathcal{A}$ di sottoinsiemi di Ω (nel seguito si indicherà questo con $\mathcal{A} \subseteq \mathcal{P}(\Omega)$). Diciamo che $\mathcal{A}$ è una σ -algebra (o anche una **tribù**) se:

- i) $\emptyset \in \mathcal{A}$
- ii) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$ ($\mathcal{A}$ è **chiusa per complemento**)
- iii) $(A_n)_n \subseteq \mathcal{A} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ ($\mathcal{A}$ è **chiusa per unione numerabile**)

Osserviamo che dalla definizione di σ -algebra segue che $\Omega \in \mathcal{A}$ e che $(A_n)_n \subseteq \mathcal{A} \Rightarrow \bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$, cioè che $\mathcal{A}$ è **chiusa per intersezione numerabile**. Infatti si può ottenere $\bigcap_{n=1}^{\infty} A_n$ in questo modo:

$$\bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^c \right)^c$$

Definizione 1.2 Prendiamo una σ -algebra $\mathcal{A} \subseteq \mathcal{P}(\Omega)$. Una funzione d'insieme $\mu : \mathcal{A} \rightarrow [0, +\infty]$ si dice **misura** se

- i) $\mu(\emptyset) = 0$
- ii) $(A_n)_n \subseteq \mathcal{A}$ disgiunti $\Rightarrow \mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$ (μ è **numerabilmente additiva**)

Se $\mathbb{P}$ è una misura su $\mathcal{A}$ tale che gode anche della proprietà

- iii) $\mathbb{P}(\Omega) = 1$

allora si dice **misura di probabilità**, o più semplicemente **probabilità**.

Diamo ora un po' di terminologia di calcolo delle probabilità.

Definizione 1.3 Si dice **spazio probabilizzato** una terna $(\Omega, \mathcal{A}, \mathbb{P})$, dove Ω è un insieme generico, $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ è una σ -algebra, e $\mathbb{P}$ è una misura di probabilità su Ω . La coppia $(\Omega, \mathcal{A})$ (senza specificare una misura) si dice **spazio misurabile**. Gli insiemi $A \in \mathcal{A}$ si dicono **eventi**. Se $A \in \mathcal{A}$, $\mu(A) = 0$, allora A si dice **evento trascurabile**, mentre invece se $\mu(A) = 1$ si dice **evento quasi certo**. Se una proprietà vale $\forall \omega \in \Omega \setminus A$, con A trascurabile, allora si dice che la proprietà vale **quasi certamente** (abbreviato q.c.).

Ricordiamo ora alcune proprietà delle misure di probabilità.

Teorema 1.4 Sia $\mathbb{P}$ una misura di probabilità. Allora:

- a) $A, B \in \mathcal{A}, A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$ (**monotonia**), e in particolare $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$
- b) se $(A_n)_n \subseteq \mathcal{A}$ (non necessariamente disgiunti), allora $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ (**subadditività numerabile**)
- c) se $(A_n)_n \subseteq \mathcal{A}$, $A_1 \subseteq A_2 \subseteq \dots$, allora $\mathbb{P}(\bigcup_{n=1}^{\infty} A_n) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n)$ (**continuità dal basso**)
- d) se $(A_n)_n \subseteq \mathcal{A}$, $A_1 \supseteq A_2 \supseteq \dots$, allora $\mathbb{P}(\bigcap_{n=1}^{\infty} A_n) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n)$ (**continuità dall'alto**)

Vediamo ora alcuni possibili esempi di spazi probabilizzati.

Esempio 1.5 Prendiamo un insieme Ω numerabile e $\mathcal{A} = \mathcal{P}(\Omega)$. Allora $\mathcal{A}$ è una σ -algebra. Se poi prendiamo una generica funzione $p : \Omega \rightarrow [0, 1]$ tale che $\sum_{\omega \in \Omega} p(\omega) = 1$, allora questa definisce una probabilità su Ω in questo modo:

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

Più in generale, se Ω è un insieme qualsiasi ma $p : \Omega \rightarrow [0, 1]$ è diversa da zero in un insieme al più numerabile di punti e $\sum_{\omega \in \Omega} p(\omega) = 1$, si può ugualmente definire una probabilità $\mathbb{P}$ nel modo visto sopra. $\mathbb{P}$ assume il nome di **probabilità discreta**.

Esempio 1.6 (delta di Dirac) Prendiamo un insieme qualsiasi Ω e definiamo $\mathcal{A} = \mathcal{P}(\Omega)$. Prendiamo poi un $\omega_0 \in \Omega$ fissato e definiamo

$$\delta_{\omega_0}(A) = \begin{cases} 1 & \text{se } \omega_0 \in A \\ 0 & \text{se } \omega_0 \notin A \end{cases} \quad \forall A \subseteq \Omega$$

La δ_{ω_0} è una misura nota come **delta di Dirac nel punto ω_0** , o anche **legge (degenere) concentrata in ω_0** .

Con la terminologia dell'esempio precedente, δ_{ω_0} può anche essere definita tramite

$$p(\omega) = \begin{cases} 1 & \text{se } \omega = \omega_0 \\ 0 & \text{se } \omega \neq \omega_0 \end{cases}$$

Esempio 1.7 (misure definite da una densità) Prendiamo lo spazio misurato $(\Omega, \mathcal{A}, \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, dove λ è la misura di Lebesgue. Se $f : \Omega \rightarrow \mathbb{R}^+$ è una funzione misurabile tale che $\int_{\Omega} f \, d\mu = 1$, possiamo definire una misura di probabilità $\mathbb{P}$ in questo modo:

$$\mathbb{P}(A) = \int_A f \, d\mu \quad \forall A \in \mathcal{A}$$

La $\mathbb{P}$ si chiama **probabilità di densità f (rispetto a μ)** rispetto alla misura di Lebesgue, e si può indicare con le notazioni

$$\mathbb{P} = f \cdot \mu, \quad \text{oppure} \quad \frac{d\mathbb{P}}{d\mu} = f$$

Scopriremo che questo concetto è molto usato nel calcolo delle probabilità.

Come in teoria della misura, nel calcolo delle probabilità è spesso utile poter costruire σ -algebre a partire da una data classe di generatori.

Definizione 1.8 Sia $\mathcal{I} \subseteq \mathcal{P}(\Omega)$ una generica classe di sottoinsiemi di Ω . Allora esiste una σ -algebra $\mathcal{A}$ che è la più piccola σ -algebra contenente $\mathcal{I}$. $\mathcal{A}$ si chiama allora **σ -algebra generata da $\mathcal{I}$** e si indica con $\sigma(\mathcal{I})$.

Definizione 1.9 Sia $(\Omega, \mathcal{A})$ un generico spazio misurabile. Ogni famiglia $\mathcal{I}$ tale che $\mathcal{A} = \sigma(\mathcal{I})$ si chiama **sistema di generatori** di $\mathcal{A}$. Essa si dice invece **base** di $\mathcal{A}$ se gode delle due proprietà seguenti:

- i) per ogni $A, B \in \mathcal{I}$, si ha che $A \cap B \in \mathcal{I}$;
- ii) esiste $(A_n)_n \subseteq \mathcal{I}$ tale che $\Omega = \bigcup_{n=1}^{\infty} A_n$ (unione non necessariamente disgiunta).

Ad esempio, se Ω è numerabile e $\mathcal{A} = \mathcal{P}(\Omega)$, allora una possibile base per $\mathcal{A}$ può essere la famiglia $\mathcal{I}$ formata dall'insieme vuoto e da tutti i singoletti $\{\omega\}$, con $\omega \in \Omega$.

L'importanza del concetto di base di una σ -algebra è dovuta al seguente risultato.

Teorema 1.10 Siano $\mathbb{P}_1$ e $\mathbb{P}_2$ due misure di probabilità sullo stesso spazio probabilizzabile $(\Omega, \mathcal{A})$, e sia $\mathcal{I}$ una base di $\mathcal{A}$. Se $\mathbb{P}_1(A) = \mathbb{P}_2(A)$ per ogni $A \in \mathcal{I}$, allora $\mathbb{P}_1(A) = \mathbb{P}_2(A)$ per ogni $A \in \mathcal{A}$.

Definizione 1.11 Se (X, τ) è uno spazio topologico, indichiamo con $\mathcal{B}(X) = \sigma(\tau)$ la σ -algebra generata dalla topologia di X . Essa viene comunemente chiamata σ -algebra dei **boreliani** di X , e i suoi insiemi vengono chiamati **boreliani**, o **insiemi di Borel** di X . $\mathcal{B}(X)$ contiene gli aperti e i chiusi di X , e anche insiemi più complicati, come unioni infinite di chiusi o intersezioni infinite di aperti.

1.2 Funzioni misurabili e variabili aleatorie

Definizione 1.12 Se $(E, \mathcal{E})$ ed $(F, \mathcal{F})$ sono spazi misurabili, una funzione $f : E \rightarrow F$ si dice **misurabile** se per ogni $A \in \mathcal{F}$ si ha

$$f^{-1}(A) = \{x \in E \mid f(x) \in A\} \in \mathcal{E}$$

Se X è una funzione misurabile da uno spazio probabilizzato $(\Omega, \mathcal{A}, \mathbb{P})$ nello spazio $(E, \mathcal{E})$, si dice **variabile aleatoria a valori in E** .

Notiamo che le variabili aleatorie sono definite come funzioni misurabili su uno spazio probabilizzato, tanto che se $\mathcal{B} \subset \mathcal{A}$ è una σ -algebra e per ogni $A \in \mathcal{E}$ si ha che $X^{-1}(A) \in \mathcal{B}$, allora X si dice **variabile aleatoria $\mathcal{B}$ -misurabile**.

Notiamo inoltre che nel calcolo delle probabilità si tende ad omettere dalle notazioni, quando possibile, la dipendenza dallo spazio probabilizzato (e quindi ad omettere di scrivere esplicitamente la dipendenza da ω). Come primo esempio, l'insieme $X^{-1}(A)$ si può scrivere nei due modi equivalenti

$$\{\omega \in \Omega \mid X(\omega) \in A\} \text{ ed } \{X \in A\}$$

A seconda dello spazio di arrivo di X , essa può assumere diversi nomi:

- **variabile aleatoria discreta** se $(E, \mathcal{E})$ è discreto (o ci si può ricondurre ad uno spazio discreto);
- **variabile aleatoria reale** se $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (o anche $(\mathbb{R}, \mathcal{L})$);
- **vettore aleatorio n -dimensionale** se $(E, \mathcal{E}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$;
- **funzione aleatoria** se E è uno spazio di funzioni (in questo caso tipicamente $\mathcal{E} = \mathcal{B}(E)$).

Notiamo che queste notazioni non si autoescludono: se ad esempio E è un sottoinsieme discreto di $\mathbb{R}$, si parla di variabili aleatorie reali discrete.

Dato che la composizione di funzioni misurabili è misurabile, abbiamo che se $X : \Omega \rightarrow E$ è una variabile aleatoria e $g : E \rightarrow F$ è una funzione misurabile, allora $g \circ X = g(X) : \Omega \rightarrow F$ è una variabile aleatoria.

Se $X : \Omega \rightarrow E$ ed $(E, \mathcal{E})$ è uno spazio misurabile, allora si può rendere X variabile aleatoria costruendo la σ -algebra $\mathcal{A} = X^{-1}(\mathcal{E})$ su Ω . Questa σ -algebra si chiama **σ -algebra generata da X** , e si indica con il simbolo $\sigma(X)$. Si può allora vedere che la cardinalità di $\sigma(X)$ dipende dal livello di “complessità” di X . Il caso più semplice si ha quando X è costante, cioè $X \equiv a$. Allora $X^{-1}(A) = \{X \in A\}$ è uguale a $\emptyset$ o ad Ω a seconda che $a \notin A$ o $a \in A$, e quindi $\sigma(X) = \{\emptyset, \Omega\}$. Se X assume due valori $a \neq b$, allora $\sigma(X) = \{\emptyset, X^{-1}(\{a\}), X^{-1}(\{b\}), \Omega\}$. Più in generale, se X assume un numero finito n di valori, allora $\sigma(X)$ è finita (di cardinalità 2^n). Se X assume una quantità numerabile di valori, allora $\sigma(X)$ ha cardinalità più che numerabile, e così via.

Abbiamo visto che, se abbiamo un'applicazione $X : \Omega \rightarrow E$ ed E è dotato di una σ -algebra, allora X induce una σ -algebra $\sigma(X)$ su Ω . Viene naturale chiedersi cosa succede

con le misure di probabilità. Possiamo vedere che succede il contrario: se infatti X è una variabile aleatoria da $(\Omega, \mathcal{A}, \mathbb{P})$ a valori in $(E, \mathcal{E})$, allora induce una misura di probabilità su $(E, \mathcal{E})$, che possiamo chiamare $\mathbb{P}_X$, data da

$$\mathbb{P}_X(A) := \mathbb{P}(X^{-1}(A)) = \mathbb{P}\{X \in A\} \quad \forall A \in \mathcal{E}$$

Si può facilmente verificare che $\mathbb{P}_X$ è una misura, e viene chiamata **legge di X (secondo $\mathbb{P}$)** (mentre in Analisi si usa invece il nome di **misura immagine di $\mathbb{P}$ tramite X**).

Dato che la composizione di una variabile aleatoria con una funzione misurabile è ancora una variabile aleatoria (ovviamente a valori in uno spazio che può essere diverso), sorge spontaneo chiedersi qual è il legame della legge di questa nuova variabile aleatoria con quella di partenza. Supponiamo quindi che X sia una variabile aleatoria da $(\Omega, \mathcal{A}, \mathbb{P})$ a valori in $(E, \mathcal{E})$, e $g : (E, \mathcal{E}) \rightarrow (F, \mathcal{F})$ sia una funzione misurabile. Allora la legge della variabile aleatoria $g(X)$, che indichiamo con $\mathbb{P}_{g(X)}$, non è altro che l'immagine della legge di X mediante g . Infatti, per ogni $A \in \mathcal{F}$ si ha:

$$\mathbb{P}_{g(X)}(A) = \mathbb{P}\{g(X) \in A\} = \mathbb{P}\{X \in g^{-1}(A)\} = \mathbb{P}_X(g^{-1}(A))$$

cioè $\mathbb{P}_{g(X)}$ è la misura immagine di $\mathbb{P}_X$ tramite g .

La definizione di variabile aleatoria richiede di verificare che tutte le controimmagini degli insiemi misurabili dello spazio di arrivo siano eventi: questo può essere molto laborioso. La proposizione che segue permette di limitare questo controllo ad una generica classe di generatori della σ -algebra dello spazio di arrivo.

Proposizione 1.13 (criterio fondamentale di misurabilità) *Se $\mathcal{E} = \sigma(\mathcal{I})$, allora X è una variabile aleatoria se e solo se $\{X \in A\} \in \mathcal{A}$ per ogni $A \in \mathcal{I}$ (ovviamente il risultato vale anche sostituendo le parole “funzione misurabile” a “variabile aleatoria”).*

Dimostrazione. L'implicazione “solo se” è banale. Per l'inverso, osserviamo che la classe $\{A \subseteq E \mid f^{-1}(A) \in \mathcal{A}\}$ è una σ -algebra che contiene $\mathcal{I}$, quindi contiene anche $\mathcal{E}$. $\square$

Da questo risultato seguono immediatamente:

Corollario 1.14 *Se X e Y sono spazi topologici, allora ogni funzione continua $f : X \rightarrow Y$ è anche misurabile.*

Dimostrazione. f è continua se e solo se $f^{-1}(U)$ è aperto in X per ogni aperto $U \subseteq Y$; siccome $\mathcal{B}(Y)$ è generata dagli aperti di Y , la tesi segue dalla proposizione 1.13. $\square$

Proposizione 1.15 *$f : \Omega \rightarrow \mathbb{R}$ è una variabile aleatoria reale se e solo se una delle 4 condizioni equivalenti qui sotto è verificata:*

- i) $\{X > a\} \in \mathcal{A} \forall a \in \mathbb{R}$
- ii) $\{X \geq a\} \in \mathcal{A} \forall a \in \mathbb{R}$

iii) $\{X < a\} \in \mathcal{A} \forall a \in \mathbb{R}$

iv) $\{X \leq a\} \in \mathcal{A} \forall a \in \mathbb{R}$

Dimostrazione. Segue dalla proposizione 1.13 e dal fatto che gli insiemi (a, ∞) (così come quelli degli altri 3 tipi) generano $\mathcal{B}(\mathbb{R})$. $\square$

Definizione 1.16 Se Ω è un insieme, $(E_i, \mathcal{E}_i)$ sono spazi misurabili e $X_i : \Omega \rightarrow E_i$ sono funzioni, per $i \in I$, allora si dice **σ -algebra generata dalle $(X_i)_i$** la σ -algebra $\mathcal{A} = \sigma(\{X_i \in A_i\} \mid A_i \in \mathcal{E}_i, i \in I)$. Essa si indica anche con $\sigma(X_i \mid i \in I)$, ed è la più piccola σ -algebra tale che le $(X_i)_i$ siano tutte variabili aleatorie.

Un esempio molto importante di applicazione di questo criterio si ha per le cosiddette misure prodotto. Supponiamo di avere gli spazi probabilizzati $(E_i, \mathcal{E}_i, \mu_i)$, dove $i \in I$ è un parametro, e di volere costruire una struttura di spazio probabilizzato sullo spazio prodotto $\Omega = E := \prod_{i \in I} E_i$. Nel seguito chiamiamo $\pi_i : E \rightarrow E_i$ la proiezione sull' i -esima coordinata, $i \in I$.

Definizione 1.17 Si chiama **σ -algebra prodotto** la σ -algebra

$$\mathcal{E} = \bigotimes_{i \in I} \mathcal{E}_i := \sigma(\mathcal{I})$$

definita su $E = \prod_{i \in I} E_i$, dove

$$\mathcal{I} := \left\{ \prod_{i \in I} A_i \mid A_i \in \mathcal{E}_i, \text{ e } A_i \neq E_i \text{ per un n. finito di } i \in I \right\}$$

Se gli spazi $(E_i, \mathcal{E}_i)$ sono tutti uguali a un $(E, \mathcal{E})$, la σ -algebra prodotto si può anche indicare con $\mathcal{E}^{\otimes I}$. Se poi $I = \{1, \dots, n\}$, allora la σ -algebra prodotto si può anche indicare con $\mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n$ o con $\mathcal{E}^{\otimes n}$.

Una definizione equivalente è la seguente.

Definizione 1.18 Si chiama **σ -algebra prodotto** la σ -algebra

$$\mathcal{E} = \bigotimes_{i \in I} \mathcal{E}_i := \sigma(\mathcal{I}')$$

dove

$$\mathcal{I}' := \{\pi_i^{-1}(A_i) \mid A_i \in \mathcal{E}_i, i \in I\}$$

La σ -algebra prodotto è quindi uguale alla σ -algebra generata dalle proiezioni sulle coordinate. La differenza con la prima definizione è che in questa si impone che il numero finito di $i \in I$ per cui $A_i \neq E_i$ sia uguale a 1. Chiaramente gli insiemi in $\mathcal{I}$ sono ottenibili mediante intersezioni finite di insiemi in $\mathcal{I}'$, quindi $\sigma(\mathcal{I}) = \sigma(\mathcal{I}')$.

Si ha in realtà un risultato leggermente più generale, che citiamo per completezza.

Proposizione 1.19 *Se I è finito o numerabile, allora $\otimes_{i \in I} \mathcal{E}_i$ è generata da $\{\Pi_{i \in I} A_i \mid A_i \in \mathcal{E}_i\}$*

Dimostrazione. Abbiamo che $\pi_i^{-1}(A_i) = \Pi_{j \in I} A_j$, con $A_j = E_j$ per $j \neq i$, quindi $\otimes_{i \in I} \mathcal{E}_i$ è contenuta in $\sigma\{\Pi_{i \in I} A_i \mid A_i \in \mathcal{E}_i\}$. D'altra parte, $\Pi_{i \in I} A_i = \bigcap_{i \in I} \pi_i^{-1}(A_i)$. Siccome questa è un'intersezione numerabile, allora $\Pi_{i \in I} A_i \in \otimes_{i \in I} \mathcal{E}_i$. $\square$

Corollario 1.20 $\mathcal{B}(\mathbb{R}^n) = \mathcal{B}(\mathbb{R})^{\otimes n}$.

Definizione 1.21 Si chiama **misura prodotto** l'unica misura di probabilità $\mathbb{P} = \otimes_{i \in I} \mu_i$ definita su $(\Omega, \mathcal{A}) = (\prod_{i \in I} E_i, \otimes_{i \in I} \mathcal{E}_i)$ tale che

$$\mathbb{P} \left(\prod_{i \in I} A_i \right) = \prod_{i \in I} \mu_i(A_i) \quad \forall \prod_{i \in I} A_i \in \mathcal{I}$$

Se le misure μ_i sono tutte uguali a una stessa μ , la misura prodotto si può anche indicare con $\mu^{\otimes I}$. Se poi $I = \{1, \dots, n\}$, allora la misura prodotto si può anche indicare con $\mu_1 \otimes \dots \otimes \mu_n$ o con $\mu^{\otimes n}$.

Non dimostreremo qui l'esistenza della misura prodotto. Notiamo solo che l'unicità segue dal criterio di uguaglianza di due misure su una base, poichè è facile vedere che la famiglia $\mathcal{I}$ è una base per la σ -algebra prodotto.

Esempio 1.22 Un esempio di applicazione di questa definizione è il seguente. Supponiamo di voler modellizzare un fenomeno che si ripete n volte nel tempo con gli stessi possibili risultati a valori in uno spazio misurabile $(E, \mathcal{E})$. Per rappresentare gli n risultati finali, poniamo $\Omega := E^n$, $\mathcal{A} := \mathcal{E}^{\otimes n}$, e quindi $\mathcal{A}$ conterrà tutti gli eventi relativi agli n esperimenti elementari. Se invece fossimo interessati agli eventi relativi solo ai primi k esperimenti elementari (ad esempio perchè abbiamo osservato solo quelli), potremmo definire altre σ -algebre

$$\mathcal{F}_k := \sigma(A_1 \times \dots \times A_k \times E^{n-k} \mid A_i \in \mathcal{E})$$

Con questa definizione abbiamo $\mathcal{F}_0 = \{\emptyset, \Omega\} \subseteq \mathcal{F}_1 \subseteq \dots \subseteq \mathcal{F}_n = \mathcal{A}$. La più comune interpretazione della famiglia di σ -algebre $(\mathcal{F}_k)_k$ è che esse rappresentano l'informazione disponibile man mano che passa il tempo.

Possiamo ora rispondere a due importanti domande: se abbiamo il vettore aleatorio $X = (X_1, \dots, X_n)$, possiamo dire che $X_1, \dots, X_n$ sono variabili aleatorie? Vale anche il viceversa? La risposta ad entrambe queste domande è affermativa, in virtù del seguente risultato (più generale).

Proposizione 1.23 *Siano $(\Omega, \mathcal{A}, \mathbb{P})$ uno spazio probabilizzato ed $(E_i, \mathcal{E}_i)$, $i \in I$ spazi misurabili, e poniamo $E = \prod_i E_i$, $\mathcal{E} = \otimes_i \mathcal{E}_i$, e $\pi_i : E \rightarrow E_i$ la proiezione canonica; allora X è una variabile aleatoria a valori in $(E, \mathcal{E})$ se e solo se $X_i = \pi_i \circ X$ è una variabile aleatoria su $(E_i, \mathcal{E}_i)$ per ogni $i \in I$.*

Dimostrazione. Se X è una variabile aleatoria, allora anche le $X_i = \pi_i \circ X$ lo sono, perché composizioni di funzioni misurabili. Viceversa, se le X_i sono variabili aleatorie, questo significa che $X^{-1}(\pi_i^{-1}(A_i)) \in \mathcal{A}$ per ogni $A_i \in \mathcal{E}_i$; siccome $\mathcal{E}$ è generata da $\pi_i^{-1}(A_i)$, la tesi segue dalla Proposizione 1.13. $\square$

Corollario 1.24 *Se $X, Y : \Omega \rightarrow \mathbb{R}$ sono variabili aleatorie, allora anche $X + Y$ e XY lo sono.*

Dimostrazione. Definiamo $F : \Omega \rightarrow \mathbb{R} \times \mathbb{R}$, $\varphi, \psi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ come $F = (X, Y)$, $\varphi(x, y) = x + y$ e $\psi(x, y) = xy$. Allora, siccome $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R})^{\otimes 2}$, F è misurabile; inoltre φ e ψ sono continue, quindi sono misurabili. Siccome infine $X + Y = \varphi \circ F$ e $XY = \psi \circ F$ sono composizioni di funzioni misurabili, allora sono misurabili. $\square$

La Proposizione 1.24 ha quindi la conseguenza che le funzioni misurabili su $\mathbb{R}$ formano un'algebra. Vediamo ora come le variabili aleatorie conservino anche la struttura di reticolo su $\mathbb{R}$. Per vedere questo, potrebbe servire considerare variabili aleatorie “generalizzate” a valori sulla retta reale estesa $\bar{\mathbb{R}} = [-\infty, \infty]$. Questo ha una sua topologia (ad esempio considerando $\bar{\mathbb{R}}$ come spazio metrico, dotato della metrica $\rho(x, y) = |\arctg y - \arctg x|$), e quindi si può definire la σ -algebra $\mathcal{B}(\bar{\mathbb{R}})$ (si può verificare che $\mathcal{B}(\bar{\mathbb{R}})$ è generata dagli insiemi $(a, \infty]$ o dagli insiemi $[-\infty, a)$). Ovviamente, nel fare operazioni con tali variabili aleatorie, occorre fare attenzione alle espressioni indeterminate come $\infty - \infty$ (l'espressione $0 \cdot \infty$ di solito in teoria della misura si pone uguale a 0 per convenzione).

Proposizione 1.25 *Se $(X_n)_n$ è una successione di variabili aleatorie a valori in $(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$, allora*

$$Y_1 = \sup_n X_n, \quad Y_2 = \inf_n X_n, \quad Y_3 = \limsup_{n \rightarrow \infty} X_n, \quad Y_4 = \liminf_{n \rightarrow \infty} X_n$$

sono variabili aleatorie. Se poi $Y = \lim_{n \rightarrow \infty} X_n$ esiste per ogni $\omega \in \Omega$, allora è una variabile aleatoria.

Dimostrazione. Abbiamo $\{Y_1 > a\} = \bigcup_{n=1}^{\infty} \{X_n > a\}$, e $\{Y_2 < a\} = \bigcap_{n=1}^{\infty} \{X_n < a\}$. Siccome queste unioni e intersezioni sono numerabili, allora Y_1 e Y_2 sono variabili aleatorie. Allo stesso modo abbiamo che Y_3 e Y_4 sono variabili aleatorie perché $Y_3 = \inf_n \sup_{j \geq n} X_j$ e $Y_4 = \sup_n \inf_{j \geq n} X_j$. Infine, se Y esiste, allora si ha che $Y = Y_3 = Y_4$, e quindi Y è una variabile aleatoria. $\square$

Corollario 1.26 *Se X, Y sono variabili aleatorie reali, allora anche $\max(X, Y)$ e $\min(X, Y)$ lo sono.*

1.3 Variabili aleatorie semplici

A questo punto possiamo presentare una utile decomposizione di funzioni: se X è una variabile aleatoria reale, definiamo **parte positiva** e **parte negativa di X** le variabili

aleatorie

$$X^+ = \max(X, 0), \quad X^- = -\min(X, 0)$$

Allora $X = X^+ - X^-$. Grazie a questa decomposizione, possiamo nel seguito limitarci a lavorare con variabili aleatorie reali a valori positivi.

Presentiamo ora degli esempi molto importanti di variabili aleatorie. Se $(\Omega, \mathcal{A})$ è uno spazio misurabile, e $A \in \mathcal{A}$, definiamo **funzione indicatrice** dell'evento A (nel calcolo delle probabilità il nome “funzione caratteristica” si usa per altri tipi di funzione) la funzione

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

Chiamiamo poi **variabile aleatoria semplice** una combinazione lineare finita (a coefficienti reali) di funzioni indicatrici. Allora una variabile aleatoria semplice è della forma $X = \sum_{i=1}^n a_i \mathbf{1}_{A_i}$. Tuttavia X può essere rappresentata in vari modi, e in particolare possiamo avere che $a_i = a_j$ per $i \neq j$. Se però abbiamo $X(\Omega) = \{a_1, \dots, a_n\}$ e poniamo $A_i := \{X = a_i\}$, allora possiamo rappresentare X come $X = \sum_{i=1}^n a_i \mathbf{1}_{A_i}$, con $a_i \neq a_j$ per $i \neq j$. Questa si chiama **rappresentazione standard** di X .

È chiaro che se X e Y sono variabili aleatorie semplici, allora anche $X + Y$ e XY lo sono. Vediamo ora come le variabili aleatorie non negative possono essere approssimate da variabili aleatorie semplici.

Teorema 1.27 *Se $(\Omega, \mathcal{A})$ è uno spazio misurabile, e X è una variabile aleatoria reale non negativa, allora esiste una successione di variabili aleatorie semplici $(X_n)_n$ tali che $0 \leq X_1 \leq \dots \leq X_n \leq \dots \leq X$, $X_n \rightarrow X$ puntualmente, e $X_n \rightarrow X$ uniformemente su ogni insieme su cui X è limitata.*

Dimostrazione. Prendiamo $n \in \mathbb{N}$ e definiamo per ogni k tra 0 e $2^{2^n} - 1$,

$$A_n^k = \{k2^{-n} < X \leq (k+1)2^{-n}\}, \quad B_n = \{X > 2^n\},$$

Definiamo poi la variabile aleatoria semplice

$$X_n = \sum_{k=0}^{2^{2^n}-1} k2^{-n} \mathbf{1}_{A_n^k} + 2^n \mathbf{1}_{B_n}$$

Allora si ha facilmente che $X_n \leq X_{n+1}$ per ogni n , e che $0 \leq X - X_n \leq 2^{-n}$ sugli insiemi in cui $X \leq 2^n$. Il risultato segue facilmente. $\square$

Se $(\Omega, \mathcal{A}, \mathbb{P})$ è uno spazio probabilizzato, a volte potrebbe farci comodo studiare le variabili aleatorie trascurando gli insiemi di misura nulla. Vediamo nel seguito (senza dimostrazione) che se $\mathcal{A}$ gode di una proprietà chiamata completezza allora la cosa si può fare facilmente.

Definizione 1.28 *Una σ -algebra $\mathcal{A}$ si dice **completa** rispetto alla misura $\mathbb{P}$ se contiene tutti i sottoinsiemi degli insiemi trascurabili, cioè se $\mathbb{P}(A) = 0$ implica che $B \in \mathcal{A} \forall B \subseteq A$.*

Si ha il seguente

Teorema 1.29 *Se $(\Omega, \mathcal{A}, \mu)$ è uno spazio misurato, e chiamiamo*

$$\mathcal{N} = \{N \in \mathcal{A} \mid \mu(N) = 0\}$$

la classe degli insiemi trascurabili, e

$$\bar{\mathcal{A}} = \{E \cup F \mid E \in \mathcal{A}, F \subseteq N \text{ per qualche } N \in \mathcal{N}\}$$

*allora $\bar{\mathcal{A}}$ è una σ -algebra completa rispetto a $\bar{\mu}$, che è l'unica estensione di μ a $\bar{\mathcal{A}}$. $\bar{\mathcal{A}}$ si chiama **completamento** di $\mathcal{A}$ rispetto a μ .*

Come annunciato, la completezza permette di trascurare gli eventi di misura nulla.

Proposizione 1.30 *Se $\mathcal{A}$ è completa rispetto a $\mathbb{P}$, allora:*

- a) se X è una variabile aleatoria e $X = Y$ q.c., allora anche Y è una variabile aleatoria;*
- b) se le $(X_n)_n$ sono variabili aleatorie e $X_n \rightarrow X$ q.c., allora X è una variabile aleatoria.*

Un esempio di applicazione può essere questo: avevamo visto che se le $(X_n)_n$ sono variabili aleatorie e convergono a X ovunque su Ω , allora anche X è una variabile aleatoria. La proposizione che abbiamo appena visto permette di dire che il risultato vale anche se invece di una convergenza certa si ha una convergenza quasi certa. Nella prossima proposizione vediamo che, anche se $\mathbb{P}$ non è completa, non ci si deve preoccupare troppo.

Proposizione 1.31 *Se $(\Omega, \mathcal{A}, \mathbb{P})$ è uno spazio probabilizzato, e $(\Omega, \bar{\mathcal{A}}, \bar{\mathbb{P}})$ è il suo completamento, e se X è una variabile aleatoria $\bar{\mathcal{A}}$ -misurabile, allora esiste una variabile aleatoria $\mathcal{A}$ -misurabile Y tale che $Y = X$ $\bar{\mathbb{P}}$ -quasi certamente.*

1.4 Valore atteso, o speranza matematica, e integrazione

In questa sezione fissiamo uno spazio probabilizzato $(\Omega, \mathcal{A}, \mathbb{P})$ e chiamiamo $L^+ = L^+(\Omega, \mathcal{A}, \mathbb{P})$ lo spazio delle variabili aleatorie reali non negative.

Definizione 1.32 *Se X è una variabile aleatoria semplice in L^+ con rappresentazione standard $\sum_{i=1}^n a_i \mathbf{1}_{E_i}$, definiamo il **valore atteso** di X come*

$$E[X] = \int_{\Omega} X \, d\mathbb{P} =: \sum_{i=1}^n a_i \mathbb{P}(E_i)$$

Sinonimi di valore atteso sono **valore medio**, **media**, **speranza matematica** o più semplicemente **speranza**.

La speranza si può definire per una generica variabile aleatoria $X \in L^+$ come

$$E[X] = \sup \{E[Y] \mid Y \text{ v. al. semplice, } 0 \leq Y \leq X\}$$

Notiamo che in questo caso $E[X]$ può essere uguale a $+\infty$. Se $E[X] < \infty$, diciamo che X ha **speranza finita**, o anche che è **integrabile**. Infine, grazie alla decomposizione di una generica variabile aleatoria X in $X = X^+ - X^-$, si definisce

$$E[X] = E[X^+] - E[X^-]$$

Affinchè questa sia una buona definizione occorre che almeno una tra X^+ , X^- sia integrabile. In questo caso possiamo comunque avere che $E[X]$ può essere anche uguale a $+\infty$ o a $-\infty$. Se però $E[X]$ è un numero finito, allora diciamo che X ha **speranza finita**, o anche che è **integrabile**. Notiamo che $X^+, X^- \leq |X| \leq X^+ + X^-$; da questo si deduce che X è integrabile se e solo se $E[|X|] < +\infty$.

Proposizione 1.33 *Se X e Y sono variabili aleatorie con speranza finita, allora:*

- a) se $c \in \mathbb{R}$, allora $E[cX] = cE[X]$;
- b) $E[X + Y] = E[X] + E[Y]$;
- c) se $X \leq Y$ q.c., allora $E[X] \leq E[Y]$ (**monotonia**);
- d) $|E[X]| \leq E[|X|]$.

Le due caratteristiche principali dell'integrazione secondo Lebesgue sono che si possono definire facilmente degli spazi normati completi di funzioni integrabili, e che si hanno dei buoni teoremi di convergenza. Ovviamente questo si trasferisce integralmente al calcolo delle probabilità.

Iniziamo a vedere come possiamo definire degli spazi normati completi di variabili aleatorie. Abbiamo appena visto che le variabili aleatorie con speranza finita formano uno spazio vettoriale. Presentiamo ora due risultati intermedi, che ci serviranno nel seguito.

Proposizione 1.34 *Se $X \in L^+$, allora $E[X] = 0$ se e solo se $X = 0$ q.c. Inoltre, se $E[X] < +\infty$, allora l'insieme $A := \{X = +\infty\}$ è trascurabile.*

Dimostrazione. Se X è semplice, questo è ovvio: se $X = \sum_{i=1}^n a_i \mathbf{1}_{A_i}$, con $a_i \geq 0$, allora si ha che $E[X] = 0$ se e solo se per ogni i si ha che $a_i = 0$ oppure che $\mathbb{P}(A_i) = 0$. In generale, se $X = 0$ q.c. e φ è una variabile aleatoria semplice tale che $0 \leq \varphi \leq X$, allora $\varphi = 0$ q.c., quindi $E[X] = \sup_n E[\varphi] = 0$. Viceversa, abbiamo che $\{X > 0\} = \bigcup_{n=1}^{\infty} A_n$, dove $A_n = \{X > 1/n\}$, quindi se fosse falso che $X = 0$ q.c., dovremmo avere che $\mathbb{P}(A_n) > 0$ per qualche n . Ma siccome $X > \mathbf{1}_{A_n}/n$, allora $E[X] > \mathbb{P}(A_n)/n > 0$, e abbiamo una contraddizione. Per la seconda parte, basta notare che $X > +\infty \mathbf{1}_A$, quindi $E[X] > E[+\infty \mathbf{1}_A] = +\infty \mathbb{P}(A)$. Ma siccome $E[X] < +\infty$, allora bisogna avere per forza che $\mathbb{P}(A) = 0$. $\square$

Proposizione 1.35 *Se X e Y sono integrabili, allora sono equivalenti:*

- i) $E[X \mathbf{1}_A] = E[Y \mathbf{1}_A]$ per ogni $A \in \mathcal{A}$

$$ii) \ E[|X - Y|] = 0$$

$$iii) \ X = Y \text{ q.c.}$$

Dimostrazione. L'equivalenza $(ii) \iff (iii)$ segue dal fatto che $|X - Y|$ è una variabile aleatoria positiva e dalla Proposizione 1.34. Inoltre, se $E[|X - Y|] = 0$, allora per ogni $A \in \mathcal{A}$ si ha che

$$|E[X\mathbf{1}_A] - E[Y\mathbf{1}_A]| \leq E[\mathbf{1}_A|X - Y|] \leq E[|X - Y|] = 0$$

e quindi $(ii) \Rightarrow (i)$. D'altra parte, se è falso che $X = Y$ q.c., e chiamiamo $Z = X - Y$, allora Z^+ o Z^- deve essere diversa da zero in un insieme non trascurabile. Se ad esempio $A = \{Z^+ > 0\}$ è non trascurabile, allora $E[X\mathbf{1}_A] - E[Y\mathbf{1}_A] = E[Z^+] > 0$, poiché $Z^- = 0$ su A , e abbiamo una contraddizione. Questo prova che $(i) \Rightarrow (iii)$. $\square$

Questa proposizione mostra che la speranza non cambia se cambiamo la variabile aleatoria su insiemi di misura nulla. Per questo, ad esempio, possiamo trattare, per quanto riguarda l'integrazione, le variabili aleatorie a valori in $\bar{\mathbb{R}}$ che siano finite q.c. semplicemente come variabili aleatorie reali.

Definizione 1.36 Definiamo $L^1 = L^1(\Omega, \mathcal{A}, \mathbb{P})$ l'insieme delle classi di equivalenza delle variabili aleatorie su Ω con speranza finita rispetto a $\mathbb{P}$, dove $X \equiv Y$ se e solo se $X = Y$ $\mathbb{P}$ -q.o.

Questo spazio è uno spazio vettoriale reale, ed è anche uno spazio di Banach rispetto alla norma $\|X - Y\|_{L^1} = E[|X - Y|]$. Pur essendo L^1 uno spazio definito come classe di equivalenza di variabili aleatorie, come in analisi si usa la notazione $X \in L^1$ per indicare che X è una variabile aleatoria integrabile definita q.o. Questo leggero abuso di notazione è universalmente accettato e non causa quasi mai ambiguità.

1.5 Teoremi di limite

Vediamo ora i tre teoremi di convergenza citati prima.

Teorema 1.37 (della convergenza monotona) Se $(X_n)_n$ è una successione di variabili aleatorie in L^+ tale che $X_n \leq X_{n+1}$ per ogni n , e $X = \lim_{n \rightarrow \infty} X_n (= \sup_n X_n)$ q.c., allora

$$E[X] = \lim_{n \rightarrow \infty} E[X_n]$$

Lemma 1.38 (di Fatou) Se $(X_n)_n$ è una successione in L^+ , allora

$$E\left[\liminf_{n \rightarrow \infty} X_n\right] \leq \liminf_{n \rightarrow \infty} E[X_n]$$

Teorema 1.39 (della convergenza dominata di Lebesgue) Se $(X_n)_n \subset L^1$ tale che $X_n \rightarrow X$ q.o. ed esiste $Y \in L^1$ tale che $|X_n| \leq Y$ per ogni n , allora $X \in L^1$ e

$$\lim_{n \rightarrow \infty} E[X_n] = E[X]$$

Due immediate conseguenze di questi teoremi sono i seguenti teoremi di integrazione per serie.

Teorema 1.40 *Se $(X_n)_n$ è una successione finita o infinita in L^+ e $X = \sum_n X_n$, allora $E[X] = \sum_n E[X_n]$.*

Teorema 1.41 (Beppo Levi) *Se $(X_n)_n$ è una successione in L^1 tale che $\sum_{n=1}^{\infty} E[|X_n|] < +\infty$, allora $\sum_{n=1}^{\infty} X_n$ converge q.c. a una variabile aleatoria in L^1 , e $\sum_{n=1}^{\infty} E[X_n] = E[\sum_{n=1}^{\infty} X_n]$.*

1.6 Speranza condizionale

Un problema generale del calcolo delle probabilità è il seguente: data una variabile aleatoria X definita su $(\Omega, \mathcal{A}, \mathbb{P})$ e una σ -algebra $\mathcal{B} \subset \mathcal{A}$, riusciamo ad approssimare la prima variabile aleatoria X con un'altra variabile aleatoria che sia solo $\mathcal{B}$ -misurabile?

Supponiamo all'inizio che $\mathcal{B} = \sigma(B_n, n = 1, \dots, N)$ con $(B_n)_n \subseteq \mathcal{A}$ eventi non trascurabili, disgiunti e tali che $\cup_{n=1}^N B_n = \Omega$. Una variabile aleatoria $\mathcal{B}$ -misurabile allora assume valori che dipendono solo da quale particolare evento B_n si verifichi. Quindi, se sappiamo che si è verificato un particolare B_n , sembra naturale approssimare una variabile aleatoria X , integrabile ma a priori non $\mathcal{B}$ -misurabile, con la sua speranza rispetto a una opportuna $\mathbb{P}_{B_n}$, invece che rispetto a $\mathbb{P}$: se X ha speranza finita, allora si ha

$$E_{\mathbb{P}_{B_n}}[X] = \int_{\Omega} X \, d\mathbb{P}_{B_n} = E_{\mathbb{P}} \left[X \frac{\mathbf{1}_{B_n}}{\mathbb{P}(B_n)} \right] = \frac{E_{\mathbb{P}}[X \mathbf{1}_{B_n}]}{\mathbb{P}(B_n)}$$

Il lettore può dimostrare per esercizio questa formula, che comunque è un caso particolare di un risultato che vedremo in seguito. Questa speranza si chiama anche **speranza di X condizionata all'evento H** , e si indica con $E[X|H]$.

Possiamo definire la variabile aleatoria cercata, che meglio approssimi X , come

$$E[X|\mathcal{B}](\omega) := E[X|B_n] \quad \text{per } \omega \in B_n$$

Siccome il secondo membro è chiaramente $\mathcal{B}$ -misurabile, quello che abbiamo ottenuto è una variabile aleatoria $\mathcal{B}$ -misurabile.

Vediamo ora in che senso la variabile aleatoria trovata è la migliore approssimazione per X .

Lemma 1.42 *L'unica variabile aleatoria $\mathcal{B}$ -misurabile Y ad avere la proprietà*

$$E[Y \mathbf{1}_A] = \int_A Y \, d\mathbb{P} = \int_A X \, d\mathbb{P} = E[X \mathbf{1}_A] \quad \forall A \in \mathcal{B} \quad (1.1)$$

è $E[X|\mathcal{B}]$.

Dimostrazione. Innanzitutto dimostriamo che $E[X|\mathcal{B}]$ gode della proprietà (1.1): se $A \in \mathcal{B}$, allora $A = \cup_{j \in J} B_j$ per un opportuno $J \subseteq \{1, \dots, N\}$. Allora

$$\begin{aligned} E[E[X|\mathcal{B}]\mathbf{1}_A] &= E\left[E[X|\mathcal{B}]\sum_{j \in J} \mathbf{1}_{B_j}\right] = E\left[\sum_{j \in J} E[X|\mathcal{B}]\mathbf{1}_{B_j}\right] = E\left[\sum_{j \in J} E[X|B_j]\mathbf{1}_{B_j}\right] = \\ &= E\left[\sum_{j \in J} \frac{E[X\mathbf{1}_{B_j}]}{\mathbb{P}(B_j)}\mathbf{1}_{B_j}\right] = \sum_{j \in J} E\left[\frac{E[X\mathbf{1}_{B_j}]}{\mathbb{P}(B_j)}\mathbf{1}_{B_j}\right] = \sum_{j \in J} E[X\mathbf{1}_{B_j}] = \\ &= E\left[\sum_{j \in J} X\mathbf{1}_{B_j}\right] = E[X\mathbf{1}_A] \end{aligned}$$

Dimostriamo ora l'unicità: se Y è una variabile aleatoria $\mathcal{B}$ -misurabile ad avere la proprietà (1.1), allora definiamo gli eventi

$$A^\pm := \{Y \gtrless E[X|\mathcal{B}]\}$$

Allora ovviamente $A^\pm \in \mathcal{B}$, e per (1.1) si ha $E[\mathbf{1}_{A^+}(Y - E[X|\mathcal{B}])] = 0$; ma $\mathbf{1}_{A^+}(Y - E[X|\mathcal{B}]) \in L^+$, quindi $\mathbf{1}_{A^+}(Y - E[X|\mathcal{B}]) = (Y - E[X|\mathcal{B}])^+ = 0$ q.c., cioè $\mathbb{P}\{Y \leq E[X|\mathcal{B}]\} = 1$, e $\mathbb{P}(A^+) = 0$; allo stesso modo si dimostra che $\mathbb{P}(A^-) = 0$, e quindi $Y = E[X|\mathcal{B}]$ q.c. $\square$

Un risultato notevole è che la proprietà (1.1) riesce a risolvere, anche in situazioni più generali di questa, il problema che ci eravamo posti all'inizio: difatti una tale variabile aleatoria Y è una variabile aleatoria $\mathcal{B}$ -misurabile che coincide con X ogni volta che viene integrata su eventi $A \in \mathcal{B}$. Inoltre, la proprietà (1.1) permette da sola di dimostrare alcune proprietà notevoli.

Definizione 1.43 Una variabile aleatoria Y che gode della proprietà (1.1), viene chiamata **speranza condizionale** (o **valore atteso condizionato**) di X rispetto a $\mathcal{B}$, e si indica con $Y = E[X|\mathcal{B}]$.

Nota 1.44 Se $\mathcal{B} = \sigma(T)$, con T variabile aleatoria, si indica più semplicemente $E[X|T]$ piuttosto che $E[X|\sigma(T)]$.

Vediamo ora alcune proprietà della speranza condizionale.

Teorema 1.45 Sia $(\Omega, \mathcal{A}, \mathbb{P})$ uno spazio probabilizzato, $\mathcal{B} \subseteq \mathcal{A}$ una σ -algebra ed X, Y variabili aleatorie reali.

1. **linearità:** se $X, Y \in L^1$ e $a, b \in \mathbb{R}$, si ha

$$E[aX + bY|\mathcal{B}] = aE[X|\mathcal{B}] + bE[Y|\mathcal{B}]$$

2. **monotonia:** se $X \leq Y$ q.c., allora $E[X|\mathcal{B}] \leq E[Y|\mathcal{B}]$ q.c..

3. **disuguaglianza triangolare:** se $X \in L^1$, allora

$$|E[X|\mathcal{B}]| \leq E[|X| |\mathcal{B}]$$

4. se $X \in L^1$ e Y è $\mathcal{B}$ -misurabile e tale che $XY \in L^1$, allora

$$E[XY|\mathcal{B}] = YE[X|\mathcal{B}]$$

In particolare, se X è $\mathcal{B}$ -misurabile, allora $E[X|\mathcal{B}] = X$.

5. **proiezione (o tower property):** se $\mathcal{C} \subseteq \mathcal{B}$ è una terza σ -algebra, allora

$$E[E[X|\mathcal{B}] | \mathcal{C}] = E[X|\mathcal{C}]$$

6. **indipendenza:** se X è indipendente da $\mathcal{B}$, allora $E[X|\mathcal{B}] = E[X]$.

7. se $\mathcal{B}$ è costituita solo da eventi trascurabili e quasi certi, allora $E[X|\mathcal{B}] = E[X]$.

8. **teoremi di convergenza:** se $X_n \nearrow X$ e $(X_n)_n \subseteq L^+ \cap L^1$, allora $E[X_n|\mathcal{B}] \nearrow E[X|\mathcal{B}]$ (convergenza monotona);
 se $(X_n)_n \subseteq L^+ \cap L^1$, allora $E[\liminf_{n \rightarrow \infty} X_n|\mathcal{B}] \leq \liminf_{n \rightarrow \infty} E[X_n|\mathcal{B}]$ (Fatou);
 se $|X_n| \leq Y$, con $Y \in L^1$ e $X_n \rightarrow X$ q.c., allora $E[X_n|\mathcal{B}] \rightarrow E[X|\mathcal{B}]$ (Lebesgue).

Dimostrazione.

1. La variabile aleatoria $aE[X|\mathcal{B}] + bE[Y|\mathcal{B}]$ è ovviamente $\mathcal{B}$ -misurabile. Se prendiamo $A \in \mathcal{B}$, possiamo verificare che soddisfa la definizione di speranza condizionale:

$$\begin{aligned} \int_A (aE[X|\mathcal{B}] + bE[Y|\mathcal{B}]) d\mathbb{P} &= a \int_A E[X|\mathcal{B}] d\mathbb{P} + b \int_A E[Y|\mathcal{B}] d\mathbb{P} = \\ &= a \int_A X d\mathbb{P} + b \int_A Y d\mathbb{P} = \int_A (aX + bY) d\mathbb{P} \end{aligned}$$

2. L'evento $A := \{E[X|\mathcal{B}] > E[Y|\mathcal{B}]\}$ appartiene a $\mathcal{B}$, e siccome $Y \geq X$ q.c., possiamo scrivere

$$0 \leq E[(Y - X)\mathbf{1}_A] = E[E[Y - X|\mathcal{B}]\mathbf{1}_A] \leq 0$$

poichè su A si ha $E[Y - X|\mathcal{B}] < 0$. Ma questo significa che $E[E[Y - X|\mathcal{B}]\mathbf{1}_A] = 0$, e quindi $E[Y - X|\mathcal{B}]\mathbf{1}_A = 0$ q.c., cioè $\mathbb{P}(A) = 0$.

3. Segue immediatamente da $X = X^+ - X^-$, $|X| = X^+ + X^-$ e dalle proprietà 1. e 2.

4. Ovviamente $YE[X|\mathcal{B}]$ è $\mathcal{B}$ -misurabile. Supponiamo all'inizio che $Y = \mathbf{1}_A$, con $A \in \mathcal{B}$. Allora per ogni $B \in \mathcal{B}$ si ha che

$$E[\mathbf{1}_B E[XY|\mathcal{B}]] = E[\mathbf{1}_B XY] = E[\mathbf{1}_{A \cap B} X] = E[\mathbf{1}_{A \cap B} E[X|\mathcal{B}]] = E[\mathbf{1}_B Y E[X|\mathcal{B}]]$$

Quindi la relazione è vera se Y è una funzione indicatrice. Per linearità, è vera anche se Y è semplice. A questo punto, se $X, Y \in L^+$, allora esiste una successione $(Y_n)_n \subseteq L^+$ di variabili aleatorie semplici tali che $Y_n \nearrow Y$ q.c., e quindi, per ogni $B \in \mathcal{B}$,

$$E[\mathbf{1}_B E[XY|\mathcal{B}]] = E[\mathbf{1}_B XY] = \lim_n E[\mathbf{1}_B XY_n] = \lim_n E[\mathbf{1}_B Y_n E[X|\mathcal{B}]] = E[\mathbf{1}_B Y E[X|\mathcal{B}]]$$

Infine, se X e Y hanno segno qualsiasi, allora si applica quanto appena dimostrato ai 4 addendi di $XY = (X^+ - X^-)(Y^+ - Y^-)$, e per linearità si conclude.

5. Prendiamo $A \in \mathcal{C}$; allora

$$\int_A E[E[X|\mathcal{B}]|\mathcal{C}] d\mathbb{P} = \int_A [E[X|\mathcal{B}]] d\mathbb{P} = \int_A X d\mathbb{P} = \int_A E[X|\mathcal{C}] d\mathbb{P}$$

dove la prima e la terza uguaglianza seguono dalla definizione di $E[\cdot|\mathcal{C}]$ e la seconda dal fatto che $A \in \mathcal{C} \subseteq \mathcal{B}$.

6. Per ogni $A \in \mathcal{B}$ si ha

$$E[E[X|\mathcal{B}]\mathbf{1}_A] = E[X\mathbf{1}_A] = E[X]E[\mathbf{1}_A] = E[E[X]\mathbf{1}_A]$$

dove la prima uguaglianza segue dal punto 4. e la terza dall'indipendenza. La tesi segue dal fatto che la costante $E[X]$ è banalmente $\mathcal{B}$ -misurabile.

7. Per ogni $A \in \mathcal{A}$, $B \in \mathcal{B}$ si ha che

$$\mathbb{P}(A \cap B) = \begin{cases} 0 & \text{se } \mathbb{P}(B) = 0, \\ \mathbb{P}(A) & \text{se } \mathbb{P}(B) = 1, \end{cases}$$

e quindi $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Questo significa che $\mathcal{B}$ è indipendente da $\mathcal{A}$, e quindi da ogni σ -algebra in essa contenuta, come in particolare $\sigma(X)$. La tesi segue dunque dal punto precedente.

8. Dimostriamo solo la convergenza dominata (le altre due dimostrazioni sono analoghe). Chiamando $Z_n := |X_n - X|$, abbiamo che $Z_n \rightarrow 0$ q.c., e $Z_n \leq 2Y$ per ogni n ; inoltre

$$|E[X_n|\mathcal{B}] - E[X|\mathcal{B}]| \leq E[|X_n - X| |\mathcal{B}] = E[Z_n|\mathcal{B}]$$

per il punto 3. Ora, se chiamiamo $Z := \limsup_n E[Z_n|\mathcal{B}]$, abbiamo ovviamente che $Z \geq 0$ q.c., e

$$E[Z] \leq \limsup_n E[E[Z_n|\mathcal{B}]] = \limsup_n E[Z_n] = 0$$

dove l'ultima uguaglianza segue dal teorema di Lebesgue. Quindi $Z = 0$ q.c., che significa che $\lim_n E[Z_n|\mathcal{B}] = 0$, e quindi $|E[X_n|\mathcal{B}] - E[X|\mathcal{B}]| \rightarrow 0$, che implica la tesi.

□