

Dimostrazione. Vediamo una prima "rappresentazione" che dimostra Proposizione 5.10, sebbene questa rappresentazione possa avere un sapore un po' tautologico. Consideriamo la costruzione seguente: si consideri $\Omega = \mathbb{R}$ e $\mathcal{F} = \mathcal{B}(\mathbb{R})$. Data F , da Proposizione 5.8, sappiamo che esiste una mdp P_F su $\mathcal{B}(\mathbb{R})$ tale che $P_F(-\infty, x] = F(x)$ per ogni $x \in \mathbb{R}$. Definiamo $X : (\Omega, \mathcal{F}, P_F) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ come $X(\omega) = \omega$. Tale funzione è ovviamente misurabile (è l'identità) e quindi è una variabile aleatoria. Inoltre $F_X(x) = P_F\{\omega : X(\omega) \leq x\} = P_F(-\infty, x] = F(x)$. \blackbox

Mentre la P_F di Proposizione 5.8 è unica, non è unica la scelta di $(\Omega, \mathcal{F}, \mathbb{P})$ e X in Proposizione 5.10. In aggiunta alla costruzione data durante la dimostrazione vediamo ora un'altra "rappresentazione" di X in Proposizione 5.10.

Per fare ciò introduciamo una nuova definizione. Data una funzione di ripartizione F il suo quantile (o inversa generalizzata) F^- si definisce come

$$F^-(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}.$$

Se F è strettamente monotona allora F^- coincide con l'inversa di F , ossia $F^- = F^{-1}$. Si noti che la definizione non presuppone che F sia una funzione di ripartizione di una variabile aleatoria X .

NON IN PROGRAMMA:

Proposizione 5.11. *Sia F una funzione di ripartizione. Se $U \sim \mathcal{U}(0, 1)$ è definita su uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, allora $X := F^-(U)$ ha funzione di ripartizione F , ossia*

$$F_X(x) = P\{X \leq x\} = \mathbb{P}\{F^-(U) \leq x\} = F(x).$$

Dimostrazione. Si verifica che $\{u \in (0, 1) : F^-(u) \leq x\} = \{u \in (0, 1) : u \leq F(x)\}$ (farsi un disegno!). La tesi segue. \blackbox

La precedente proposizione mostra che, disponendo di una variabile aleatoria uniforme su uno spazio $(\Omega, \mathcal{F}, \mathbb{P})$, si può costruire sullo stesso spazio una variabile aleatoria X con funzione di ripartizione assegnata qualunque.

Si noti che scegliendo $\Omega = (0, 1)$, $\mathcal{F} = \mathcal{B}((0, 1))$ e $\mathbb{P} = P$ con P la misura uniforme su $(0, 1)$ (Esempio 5.1), possiamo definire la variabile aleatoria $U(\omega) = \omega$ e verificare che tale variabile aleatoria ha legge uniforme su $(0, 1)$. Infatti, $\mathbb{P}\{\omega : U(\omega) \leq x\} = x$ per ogni $x \in (0, 1)$. Combinando questa osservazione con la precedente proposizione otteniamo un'altra dimostrazione di Proposizione 5.10.

Il concetto di funzione di ripartizione di un numero aleatorio X può essere esteso anche ad un vettore aleatorio \mathbf{X} . In questo caso la definizione diventa:

Definizione 5.4. *Dati un vettore aleatorio $\mathbf{X} = (X_1, \dots, X_D)$ e la sua distribuzione $P_{\mathbf{X}}$, si dice funzione di ripartizione di \mathbf{X} la funzione $F_{\mathbf{X}} : \mathbb{R}^D \rightarrow [0, 1]$ definita da*

$$F_{\mathbf{X}}(x) := \mathbb{P}\{X_1 \leq x_1, \dots, X_D \leq x_D\} = P_{\mathbf{X}}\{(-\infty, x]\}.$$

Il caso di funzioni di ripartizione $D > 1$ è più complicato da trattare e ci limitiamo qui ad osservare che il Teorema 3.3 si può ri-enunciare come segue:

Teorema 5.12. *Se $F_{\mathbf{X}}$ e $F_{\mathbf{Y}}$ sono uguali allora $P_{\mathbf{X}} = P_{\mathbf{Y}}$.*

6. VARIABILI ALEATORIE DISCRETE

6.1. Variabili aleatorie discrete, densità, marginali. Sia $(\mathbb{X}, \mathcal{X}) = (\mathbb{R}^D, \mathcal{B}(\mathbb{R}^D))$ o $(\mathbb{X}, \mathcal{X}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Definizione 6.1. *Una variabile aleatoria $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{X}, \mathcal{X})$ è detta discreta se esiste un insieme $C \in \mathcal{X}$ numerabile tale che $\mathbb{P}\{X \in C\} = 1$.*

Se X è discreta in realtà non è molto rilevante chi sia $(\mathbb{X}, \mathcal{X})$, infatti ciò che conta è che gli insiemi immagine tramite X sono (a meno di insiemi di probabilità nulla) insiemi al più numerabili. Se P_X è la misura immagine di X

$$P_X(A) = P_X(A \cap C) + P_X(A \cap C^c) = P_X(A \cap C) \quad \text{con } A' = A \cap C \in \mathcal{P}(C).$$

Quindi la probabilità immagine P_X può essere pensata come una misura di probabilità non su $(\mathbb{X}, \mathcal{X})$ ma semplicemente su $(C, \mathcal{P}(C))$.³ In effetti al posto di P_X si può considerare senza perdita di generalità la sua traccia su \mathcal{X}_C , infatti

$$P_X(A) = P_X(A \cap C) = \frac{P_X(A \cap C)}{P_X(C)}.$$

In breve una v.a. discreta X è completamente caratterizzata dalla *densità discreta*

$$p_X(x) = P\{X = x\} \quad x \in C.$$

Infatti per ogni A in $\mathcal{P}(C)$ si ha che A è al più numerabile e

$$P\{X \in A\} = \sum_{x \in A} p_X(x)$$

[Theorem 4.1. in [4]]

Un vettore aleatorio $\mathbf{X} = (X_1, \dots, X_D)$ a valori in \mathbb{R}^D discreto sarà caratterizzato da una densità discreta $p_{\mathbf{X}}$ definita a priori su un insieme discreto $C \subset \mathbb{R}^D$. Si può sempre considerare come C un insieme prodotto del tipo $C_1 \times \dots \times C_D$ (in caso basta aggiungere un numero al più numerabile di punti). Ad esempio se $D = 2$ e $C = \{(0, 1), (1, 0), (1, 1)\}$ potremo considerare come nuovo C l'insieme $\{0, 1\} \times \{0, 1\}$. Risulta ovvio che se \mathbf{X} è discreto sono discrete anche le sue componenti X_i , $i = 1, \dots, D$ e i possibili sottovettori $(X_{i_1}, \dots, X_{i_k})$. Ogni X_i assume con probabilità 1 valori in C_i (definito sopra). Come conseguenza sono ben definite le densità delle componenti X_i , che si dicono *densità marginali* di \mathbf{X} .

Si verifica immediatamente che per ogni $x_i \in C_i$

$$p_{X_i}(x_i) = \sum_{y \in C: y_i = x_i} p_{\mathbf{X}}(y_1, \dots, y_D) = \sum_{y_1 \in C_1, \dots, y_D \in C_D} p_{\mathbf{X}}(y_1, \dots, y_{i-1}, x_i, y_{i+1}, \dots, y_D).$$

In modo analogo si possono considerare sottovettori, per cui, ad esempio

$$p_{(X_1, X_3)}(x_1, x_3) = \sum_{y_2 \in C_2} p_{\mathbf{X}}(x_1, y_2, x_3).$$

6.2. Esempi.

6.2.1. *Distribuzione binomiale.* Siano n un intero positivo e p un elemento fissato dell'intervallo $[0, 1]$, X un numero aleatorio che prende valori in $C = \{0, 1, 2, \dots, n\}$. La distribuzione di X si dice *binomiale* con parametro (n, p) [in simboli $Bi(n, p)$] se

$$\mathbb{P}\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{per } k = 0, 1, 2, \dots, n.$$

Si verifica facilmente che $\mathbb{P}\{X = k\}$ coincide con la probabilità che si verifichino esattamente k eventi su n , nell'ipotesi che gli eventi siano indipendenti e abbiano tutti probabilità p . Si veda (13).

6.2.2. *Distribuzione ipergeometrica.* Un numero aleatorio X a valori in $C = \{0, 1, 2, \dots, n\}$ si dice con distribuzione *ipergeometrica* se

$$\mathbb{P}\{X = k\} = \begin{cases} \binom{n}{k} \frac{N\theta(N\theta-1)\dots(N\theta-k+1)(N-N\theta)(N-N\theta-1)\dots(N-N\theta-n+k+1)}{N(N-1)\dots(N-n+1)} & \text{se } n \leq N, N\theta + n - N \leq k \leq N\theta \\ 0 & \text{altrove.} \end{cases}$$

Si suppone $n \leq N$ e $\theta \in (0, 1)$. In seguito indicheremo tale distribuzione con $\mathcal{H}(k; \theta, N, n)$. Si vede facilmente, che è la probabilità di avere k palline bianche in n estrazioni senza restituzione da un'urna che contiene N palline di cui $N\theta = R$ bianche, quando tutte le n -uple estraibili siano ritenute ugualmente probabili.

³Se consideriamo un $(\mathbb{X}, \mathcal{X})$ generico dobbiamo supporre che ogni $x \in C$ sia tale per cui $\{x\} \in \mathcal{X}$, cosa chiaramente soddisfatta se $(\mathbb{X}, \mathcal{X}) = (\mathbb{R}^D, \mathcal{B}(\mathbb{R}^D))$.

6.2.3. *Distribuzione di Poisson.* Sia X una variabile aleatoria tale che

$$\mathbb{P}\{X = k\} = \frac{\lambda^k e^{-\lambda}}{k!}$$

valga per ogni k intero non negativo con $\lambda > 0$ parametro assegnato. Tale distribuzione si dice di *Poisson*. Essa viene spesso utilizzata come legge del numero aleatorio degli arrivi in una coda, o di un numero aleatorio di nascite in un'unità di tempo. È interessante osservare che la si può leggere come limite di una successione di distribuzioni binomiali. Più precisamente, per ogni $n \geq 1$ si definisca la distribuzione binomiale $Bi(n, \theta_n)$ con $\theta_n = \frac{\lambda}{n} + o(1/n)$, $n \rightarrow +\infty$. Per $k = 0, 1, \dots, n$,

$$\begin{aligned} \binom{n}{k} \theta_n^k (1 - \theta_n)^{n-k} &= \frac{1}{k!} n(n-1) \cdots (n-k+1) \left(\frac{\lambda}{n} + o\left(\frac{1}{n}\right) \right)^k \left(1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right) \right)^{n-k} = \\ &= \frac{1}{k!} 1 \left(1 - \frac{1}{n} \right) \cdots \left(1 - \frac{k-1}{n} \right) \left(\lambda + n \cdot o\left(\frac{1}{n}\right) \right)^k \left(1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right) \right)^{n-k} \xrightarrow{n \rightarrow +\infty} \frac{1}{k!} \lambda^k e^{-\lambda}. \end{aligned}$$

6.2.4. *Distribuzione binomiale negativa.* I numeri

$$\binom{n+r-1}{r} p^n (1-p)^r, \quad r = 0, 1, \dots$$

sono ovviamente strettamente positivi quando n è un intero fissato maggiore di zero e p è un numero qualunque in $(0, 1)$. Inoltre, sotto queste medesime condizioni,

$$\begin{aligned} \sum_{r \geq 0} \binom{n+r-1}{r} p^n (1-p)^r &= \sum_{r \geq 0} \frac{(n+r-1) \cdots n}{r!} p^n (1-p)^r \\ &= \sum_{r \geq 0} (-1)^r \frac{(-n)(-n-1) \cdots (-n-r+1)}{r!} p^n (1-p)^r \\ &= \sum_{r \geq 0} (-1)^r \binom{-n}{r} p^n (1-p)^r \\ &= p^n [1 - (1-p)]^{-n} = 1. \end{aligned}$$

Quindi,

$$\mathbb{P}\{X = r\} = \binom{n+r-1}{r} p^n (1-p)^r \quad r = 0, 1, \dots$$

definisce una distribuzione di probabilità discreta che è nota come *legge binomiale negativa*. Nello schema di eventi indipendenti con probabilità fissa p , $\binom{n+r-1}{r} p^n (1-p)^r$ fornisce la probabilità che l' n -esimo successo si verifichi nella $(n+r)$ -esima prova. Quando $n = 1$, la distribuzione è detta *geometrica* o di *Pascal* ed è legge del tempo $r+1$ in cui si verifica il *primo* successo:

$$\mathbb{P}\{X = r\} = p(1-p)^{r-1} \quad r = 1, 2, \dots$$

Si noti che $r = 0$ vuol dire che il primo successo è al tempo 1, per questo motivo spesso si considera $\tilde{X} = X + 1$, e si dice che \tilde{X} ha legge geometrica di parametro p . In questo caso

$$(15) \quad \mathbb{P}\{\tilde{X} = r\} = p(1-p)^{r-1} \quad r = 1, 2, \dots$$

Lo studente è invitato a fare molta attenzione a quale convenzione viene usata ogni volta che consulta un libro o svolge un esercizio.

6.2.5. *Distribuzione geometrica e assenza di memoria.* Sia X una v.a. *geometrica* di parametro p con densità (15), allora per ogni s e t interi non negativi

$$\mathbb{P}\{X > t+s | X > t\} = \mathbb{P}\{X > s\}.$$

Per dimostrarlo si noti che

$$\mathbb{P}\{T > u\} = 1 - p \sum_{k=1}^u (1-p)^{k-1} = 1 - p \frac{1 - (1-p)^u}{1 - (1-p)} = (1-p)^u$$

e

$$\begin{aligned}\mathbb{P}\{T > t + s | T > t\} &= \frac{\mathbb{P}\{T > t + s, T > t\}}{\mathbb{P}\{T > t\}} = \frac{\mathbb{P}\{T > t + s\}}{\mathbb{P}\{T > t\}} \\ &= \frac{(1-p)^{t+s}}{(1-p)^t} = (1-p)^s = \mathbb{P}\{T > s\}.\end{aligned}$$

6.3. Valori attesi di numeri aleatori discreti. Se X è un numero aleatorio discreto a valori in $C \subset \mathbb{R}$, tale che

$$\sum_{x \in C} |x| \mathbb{P}\{\omega : X(\omega) = x\} < +\infty,$$

si dice che X ammette *valore atteso* (finito) e si pone

$$\mathbb{E}[X] := \sum_{x \in C} x \mathbb{P}\{\omega : X(\omega) = x\}.$$

Il valore atteso è anche detto *speranza matematica* o *media*. Si noti che la condizione di assoluta convergenza della serie $\sum_{x \in C} |x| \mathbb{P}\{\omega : X(\omega) = x\}$ implica che anche la serie con segni $\sum_{x \in C} x \mathbb{P}\{\omega : X(\omega) = x\}$ è convergente e non ci sono problemi nel sommarla secondo qualunque ordine in C .

Una notazione alterantiva è la seguente

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) \mathbb{P}(d\omega).$$

Tale notazione mette in evidenza che una variabile aleatoria è definita come una funzione da uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ in \mathbb{R} e che il suo valore atteso (se esiste) dipende chiaramente da $(\Omega, \mathcal{F}, \mathbb{P})$ oltre che da X . In realtà è importante osservare che il valore atteso di una variabile aleatoria discreta non dipende realmente da "tutta" l'informazione contenuta in $(\Omega, \mathcal{F}, \mathbb{P}, X)$ ma "solo" da $(C, \{p_X(x) : x \in C\})$, dove per ogni $x \in C$ abbiamo posto $p_X(x) = \mathbb{P}\{\omega : X(\omega) = x\}$. In altri termini il valore atteso dipende solo dalla legge della variabile aleatoria e non dallo spazio di probabilità ove questa risulta definita. Se X e X' sono due variabili aleatorie discrete definite su due spazi di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e $(\Omega', \mathcal{F}', \mathbb{P}')$, tali che ammettano lo stesso insieme C di realizzazioni e tali per cui per ogni $x \in C$ si abbia $\mathbb{P}\{\omega : X(\omega) = x\} = \mathbb{P}'\{\omega' : X'(\omega') = x\}$, allora $\mathbb{E}[X] = \mathbb{E}[X']$ (posto che questi valori attesi esistano).

Dalla definizione segue che per ogni evento A in \mathcal{F} e $B \in \mathcal{B}(\mathbb{R})$ si ha

$$\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$$

e

$$\mathbb{E}(\mathbb{1}_{X \in B}) = P_X(B).$$

Inoltre se $\mathbb{P}\{X = a\} = 1$ allora $\mathbb{E}(X) = a \cdot \mathbb{P}\{X = a\} = a$.

Iniziamo da un'importante proposizione che ci sarà molto utile.

(P0) Cambio di variabili. Sia $\mathbf{X} = (X_1, \dots, X_D)$ è un vettore aleatorio discreto a valori in $C \subset \mathbb{R}^D$ e $g : C \rightarrow \mathbb{R}$ una funzione a valori reali. Allora se il numero aleatorio $Y = g(\mathbf{X})$ ammette valore atteso finito si ha

$$\mathbb{E}[Y] = \mathbb{E}[g(\mathbf{X})] = \sum_{x \in C} g(x) \mathbb{P}\{\omega : \mathbf{X}(\omega) = x\} = \sum_{x \in C} g(x) p_{\mathbf{X}}(x).$$

A proposito di (P0) si ricordi che per definizione si avrebbe

$$\mathbb{E}[Y] = \sum_{z \in g(C)} z p_Y(z)$$

con $p_Y(z) = \mathbb{P}\{Y = z\}$, quindi quella enunciata sopra non è la definizione di $\mathbb{E}[g(\mathbf{X})]$ ma una proprietà del valore atteso, che ora dimostriamo. Indicati con g_j i valori distinti di $g(\mathbf{X})$ e con x_i quelli distinti di \mathbf{X} , si ponga $p_j^* = \mathbb{P}\{g(\mathbf{X}) = g_j\}$ e $p_i = \mathbb{P}\{\mathbf{X} = x_i\}$. Si ricorra alla definizione di speranza matematica per ottenere

$$\mathbb{E}(g(\mathbf{X})) = \sum_j g_j p_j^* = \sum_j g_j \sum_{\{i: g(x_i)=g_j\}} p_i = \sum_i p_i \sum_{\{j: g_j=g(x_i)\}} g_j = \sum_i p_i g(x_i).$$

Enunciamo alcune proprietà fondamentali del valore atteso di variabili aleatorie discrete.

(P1) **Linearità.** Se X_1 e X_2 sono variabili aleatorie discrete con valore atteso finito, allora per ogni a e b numeri reali $aX_1 + bX_2$ ha valore atteso finito e vale

$$\mathbb{E}[aX_1 + bX_2] = a\mathbb{E}[X_1] + b\mathbb{E}[X_2].$$

In particolare se il valore atteso di X esiste, allora per ogni a e b in \mathbb{R} si ha

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b.$$

(P2) **Monotonia.** Se X_1 e X_2 sono variabili aleatorie discrete con valore atteso finito tali che $P\{X_1 \leq X_2\} = 1$ allora

$$\mathbb{E}[X_1] \leq \mathbb{E}[X_2].$$

In particolare se $\mathbb{P}\{a < X \leq b\} = 1$ si ha $a < \mathbb{E}(X) \leq b$.

(P3) **Disuguaglianza del valore assoluto.** Se X ha valore atteso finito allora

$$(16) \quad |\mathbb{E}[X]| \leq \mathbb{E}[|X|].$$

(P4) **Finitezza.** Si ha

$$(17) \quad \mathbb{E}|X| < +\infty \rightarrow \mathbb{P}\{|X| < +\infty\} = 1.$$

Dimostrazione di (P1). Ponendo $p(x, y) = \mathbb{P}\{X_1 = x, X_2 = y\}$ si ha, supposto $b \neq 0$, usando (P3) che

$$\begin{aligned} \mathbb{E}(|aX_1 + bX_2|) &= \sum_{(x,y)} |ax + by|p(x, y) \\ &\leq |a| \sum_{(x,y)} |x|f(x, y) + |b| \sum_{(x,y)} |y|p(x, y) \\ &= |a| \sum_x |x| \sum_y f(x, y) + |b| \sum_y |y| \sum_x p(x, y) \\ &= |a| \sum_x |x|P\{X_1 = x\} + |b| \sum_y |y|P\{X_2 = y\} \\ &= |a|\mathbb{E}(|X_1|) + |b|\mathbb{E}(|X_2|) < +\infty \quad \text{per ipotesi.} \end{aligned}$$

Quindi esiste finita la speranza matematica di $aX_1 + bX_2$ e, con calcoli analoghi a quelli eseguiti, si ottiene $\mathbb{E}(aX_1 + bX_2) = a\mathbb{E}(X_1) + b\mathbb{E}(X_2)$.

Dimostrazione di (P2). Siano (x_i, y_i) le possibili coppie di valori di (X_1, X_2) , enumerate convenzionalmente con $i = 1, 2, \dots$. Si ponga $p_i = \mathbb{P}\{(X_1, X_2) = (x_i, y_i)\} = p(x_i, y_i)$. Si noti che per $i \neq j$, $(x_i, y_i) \neq (x_j, y_j)$ ma potrebbe accadere che $x_i = x_j$ (o analogamente $y_i = y_j$). Ad esempio $(x_1, y_1) = (1, 2)$, $(x_2, y_2) = (1, 3)$ e $(x_3, y_3) = (2, 3)$. Poiché $\mathbb{P}\{X_1 \leq X_2\} = 1$ ne segue che $x_i \leq y_i$ per ogni i e quindi usando (P3)

$$\mathbb{E}[X_1] = \sum_i x_i p_i \leq \sum_i y_i p_i = \mathbb{E}[X_2].$$

Si noti anche che se $\mathbb{E}|X| < +\infty$ allora

$$-\sum_i |x_i| \mathbb{P}\{X = x_i\} \leq \sum_i x_i \mathbb{P}\{X = x_i\} \leq \sum_i |x_i| \mathbb{P}\{X = x_i\}$$

ossia la (16).

Se per k intero positivo si ha $\sum_i |x_i|^k \mathbb{P}\{X = x_i\} < +\infty$, allora $\mathbb{E}(X^k)$ si dice *momento di ordine k* , o *momento k -esimo*, di X .

Il prossimo obiettivo sarà quello di estendere la definizione di valore atteso a variabili aleatorie reali generali, ossia non necessariamente discrete. Per fare questo faremo un nuovo excursus sulla teoria della misura. Come per la prima parte di teoria della misura affrontata nel primo Capitolo, anche in questa parte non vedremo nessuna dimostrazione.

7. CENNI DI INTEGRAZIONE ASTRATTA E INTEGRALE DI LEBESGUE

Nel paragrafo successivo introdurremo (senza dimostrazioni) il concetto di integrale astratto, ossia definiremo l'integrale astratto di una funzione misurabile

$$\xi : (E, \mathcal{E}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

rispetto ad una misura positiva σ -additiva μ su \mathcal{E} . Tale integrale verrà indicato (quando esiste) con

$$\int_E \xi(e) \mu(de).$$

Le dimostrazioni dei risultati qui riportati si possono trovare in un qualunque buon libro di teoria della misura. Si veda in alternativa [3].

7.1. Integrale astratto rispetto ad una misura positiva. In questa sezione consideriamo uno spazio di misure (E, \mathcal{E}, μ) , con μ misura σ -finita e una funzione misurabile

$$\xi : (E, \mathcal{E}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})).$$

Descriviamo la costruzione per uno spazio astratto (E, \mathcal{E}) e una generica funzione misurabile $\xi : (E, \mathcal{E}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. In questa sezione usiamo la lettera ξ per indicare una funzione misurabile e non la lettera X , e la coppia (E, \mathcal{E}) e non (Ω, \mathcal{F}) per indicare lo spazio di misura sottostante nella speranza che questa differente notazione ricordi allo studente che la costruzione non è limitata al caso di uno spazio di probabilità e ξ variabile aleatoria reale.

In seguito, infatti, useremo i risultati di questa sezione in tre casi distinti

- $(E, \mathcal{E}, \mu) = (\mathbb{R}^D, \mathcal{B}(\mathbb{R}^D), \text{Leb}_D)$ (Sezione 7.2);
- $(E, \mathcal{E}, \mu) = (\Omega, \mathcal{F}, \mathbb{P})$ e $\xi(e) = X(\omega)$ con \mathbb{P} misura di probabilità (Capitolo 9);
- $(E, \mathcal{E}, \mu) = (\mathbb{R}^D, \mathcal{B}(\mathbb{R}^D), P_{\mathbf{X}})$, $\xi(e) = g(x)$ dove $P_{\mathbf{X}}$ (o P_X) è la legge di un vettore aleatorio (Sezione 9.3).

Definizione 7.1. Una funzione misurabile $S : E \rightarrow \mathbb{R}$ si dice **semplice in forma canonica** se

$$(18) \quad S(e) = \sum_{i=1}^M c_i \mathbb{I}_{A_i}(e)$$

con A_1, \dots, A_M insiemi a due a due incompatibili di \mathcal{E} tali che $\cup_{i=1}^M A_i = E$, $c_i \in \mathbb{R}$ con $c_i \neq c_j$ per $i \neq j$. Una funzione S come in (18) dove A_1, \dots, A_M insiemi di \mathcal{E} (non necessariamente incompatibili) e $c_i \in \mathbb{R}$ non necessariamente distinti, si dice **semplice**. Una funzione semplice si dice **positiva** se $S(e) \geq 0$ per ogni e .

Data una funzione semplice S è sempre possibile riscriverla come una funzione semplice in forma canonica S' . Inoltre se S è una funzione semplice in forma canonica, essa è positiva se e solo se $c_i \geq 0$.

Definizione 7.2. Data una misura μ su \mathcal{E} e una funzione semplice positiva S , l'integrale di S rispetto μ , $\int_E S(e) \mu(de)$ in simboli, si definisce come il valore della somma a termini positivi

$$\int_E S(e) \mu(de) := \sum_i c_i \mu(A_i).$$

Tale somma può assumere il valore $+\infty$. L'integrale $\int_E S(e) \mu(de)$ si dice **finito** se la somma che lo definisce è finita.

Si noti che, come conseguenza dell'additività finita della misura μ , la definizione data non dipende dalla rappresentazione di S (a patto di prendere tutti i $c_i \geq 0$). In particolare in Definizione 7.2 si può assumere che S sia in forma canonica.

Proposizione 7.1. Sia $\xi : (E, \mathcal{E}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ una funzione misurabile a valori reali positivi, ossia $\xi(e) \geq 0$ per ogni e . Allora esiste una successione di funzioni semplici $S_n(e) = \sum_{i=1}^{M_n} c_{in} \mathbb{I}_{A_{i,n}}(e)$ tale che $\lim_n S_n = \xi$ e $S_n(e) \leq S_{n+1}(e)$ per ogni $n \geq 1$ e e in E . Inoltre se S_n e S'_n sono due successioni di funzioni misurabili semplici, positive e monotone crescenti tali che $\lim_n S_n = \lim_n S'_n$, allora $\lim_n \int S_n(e) \mu(de) = \lim_n \int S'_n(e) \mu(de)$ per qualunque misura μ .