Appunti del corso Probabilità e Finanza
A.A. 2003/2004

Capitolo 2.
Probabilità: richiami e complementi

Paolo Baldi
Lucia Caramellino

http://www.mat.uniroma2.it/~caramell

Indice

Capitolo 2

Probabilità: richiami e complementi

2.1 Cenni di teoria della misura

2.1.1 σ-algebre

Sia S un insieme e indichiamo con $\mathscr{P}(S)$ l'insieme delle sue parti (cioè la collezione di tutti i suoi sottoinsiemi). Siano $\mathscr{A}$, $\mathscr{F}$ due collezioni di sottoinsiemi di S: $\mathscr{A}, \mathscr{F} \subset \mathscr{P}(S)$.

Definizione 2.1.1. Si dice che la classe $\mathscr{A}$ è un'*algebra* su S se valgono le seguenti proprietà:

1. $\emptyset, S \in \mathscr{A}$;
2. se $A \in \mathscr{A}$ allora $A^c \in \mathscr{A}$ ($A^c =$ il complementare dell'insieme A);
3. se $\{A_n\}_{\{n=1,\dots,N\}} \subset \mathscr{A}$ allora $\cup_{n=1}^N A_n \in \mathscr{A}$.

Definizione 2.1.2. $\mathscr{F}$ è detta *σ-algebra di S* se valgono le seguenti proprietà:

1. $\emptyset, S \in \mathscr{F}$;
2. se $A \in \mathscr{F}$ allora $A^c \in \mathscr{F}$;
3. se $\{A_n\}_n \subset \mathscr{F}$ allora $\cup_{n=1}^\infty A_n \in \mathscr{F}$.

In breve, una σ-algebra, così come un'algebra, contiene sempre gli insiemi $\emptyset$ e S ed inoltre, è chiusa sotto operazione di complementare. L'unica differenza sta nella terza proprietà: una σ-algebra è chiusa sotto unioni numerabili di insiemi, mentre in un'algebra tale proprietà è vera solo per unioni finite. Ovviamente, una σ-algebra è anche un'algebra ma non vale il viceversa.

Esercizio 2.1.3. Dimostrare che una σ-algebra $\mathscr{F}$ [risp. un'algebra $\mathscr{A}$] è chiusa sotto intersezioni numerabili [risp. finite] di insiemi di $\mathscr{F}$ [risp. di $\mathscr{A}$].

Naturalmente sono σ-algebre $\mathscr{F}_1 = \{\emptyset, S\}$ (la più piccola di tutte) e $\mathscr{F}_2 = \mathscr{P}(S)$ (la più grande). Vediamo ora alcuni esempi meno banali.

Esempio 2.1.4. Sia $A \subset S$ e poniamo $\mathscr{F} = \{\emptyset, A, A^c, S\}$. È facile vedere che $\mathscr{F}$ è una σ-algebra (esercizio!).

Esempio 2.1.5. Siano $\mathscr{F}_1$ e $\mathscr{F}_2$ due σ-algebre. Allora (esercizio!):

- $\mathscr{F}_1 \cup \mathscr{F}_2$ non è in generale una σ-algebra (né un'algebra)[1];
- $\mathscr{F}_1 \cap \mathscr{F}_2$ è una σ-algebra.

Per quanto riguarda la seconda proprietà, più in generale si può dire che:

- se $\{\mathscr{F}_\gamma\}_{\gamma\in\Gamma}$ è una famiglia (qualsiasi) di σ-algebre allora $\mathscr{F} = \cap_{\gamma\in\Gamma}\mathscr{F}_\gamma$ è ancora una σ-algebra.

Dunque, mentre l'unione di σ-algebre non è in generale una σ-algebra, l'intersezione di un numero qualsiasi di σ-algebre rimane una σ-algebra.

Esempio 2.1.6. *(Partizioni finite o numerabili)* Sia $\mathscr{C} = \{C_i\}_{i\in I}$ una partizione al più numerabile di S; cioè l'insieme di indici I è finito o numerabile, i sottoinsiemi C_i sono a due a due disgiunti e la loro riunione è S. Sia $\mathscr{F}$ la classe di tutti gli insiemi che sono riunioni (finite o numerabili) degli insiemi C_i, cioè

$$\mathscr{F} = \{A \subset S\,;\; A = \textstyle\bigcup_{i\in I_A} C_i,\ \text{per qualche } I_A \subset I\},$$

(con la convenzione di porre $A = \emptyset$ se $I_A = \emptyset$). Allora, $\mathscr{F}$ è una σ-algebra. Infatti, ovviamente $\emptyset, S \in \mathscr{F}$ (si prenda $I_\emptyset = \emptyset$ e $I_S = I$). Prendiamo ora un elemento $A \in \mathscr{F}$. Allora

$$A^c = \Big(\bigcup_{i\in I_A} C_i\Big)^c = S \setminus \Big(\bigcup_{i\in I_A} C_i\Big) = \Big(\bigcup_{i\in I} C_i\Big) \setminus \Big(\bigcup_{i\in I_A} C_i\Big) = \bigcup_{i\in I_A^c} C_i,$$

dunque $A^c \in \mathscr{F}$. Infine, se $A_1, A_2, \ldots \in \mathscr{F}$ allora, posto $A_n = \bigcup_{i\in I_{A_n}} C_i$,

$$\bigcup_n A_n = \bigcup_n \Big(\bigcup_{i\in I_{A_n}} C_i\Big) = \bigcup_{i\in\cup_n I_{A_n}} C_i,$$

e ancora $\cup_n A_n \in \mathscr{F}$. Osserviamo infine che i passaggi sopra scritti sono giustificati dal fatto che gli insiemi C_i sono a due a due disgiunti.

[1] Si prendano, ad esempio, $A, B \subset S$ e siano $\mathscr{F}_1 = \{\emptyset, A, A^c, S\}$ e $\mathscr{F}_2 = \{\emptyset, B, B^c, S\}$. $\mathscr{F}_1$ e $\mathscr{F}_2$ sono due σ-algebre e $\mathscr{F} = \mathscr{F}_1 \cup \mathscr{F}_2 = \{\emptyset, A, A^c, B, B^c, S\}$: se $A \neq B, B^c$, con $A, B \neq \emptyset, S$, allora ad esempio $A \cup B \notin \mathscr{F}$, dunque $\mathscr{F}$ non è una σ-algebra né un'algebra.

Definizione 2.1.7. Sia $\mathscr{C}$ una classe di sottoinsiemi di S. La *σ-algebra generata da $\mathscr{C}$*, in simboli $\sigma(\mathscr{C})$, è la più piccola σ-algebra di S che contiene la classe $\mathscr{C}$. In altre parole, $\sigma(\mathscr{C})$ è la σ-algebra generata da $\mathscr{C}$ se e solo se $\sigma(\mathscr{C})$ è una σ-algebra di S tale che $\sigma(\mathscr{C}) \supset \mathscr{C}$ e per ogni σ-algebra $\mathscr{F}_{\mathscr{C}}$ di S tale che $\mathscr{F}_{\mathscr{C}} \supset \mathscr{C}$ allora $\mathscr{F}_{\mathscr{C}} \supset \sigma(\mathscr{C})$.

Osserviamo che la Definizione 2.1.7 è ben posta, cioè la "più piccola σ-algebra contenente $\mathscr{C}$" esiste sempre. Infatti, consideriamo la classe $\mathscr{I}$ di tutte le σ-algebre contenenti $\mathscr{C}$. $\mathscr{I}$ è non vuota, dal momento che contiene la σ-algebra $\mathscr{P}(S)$. Inoltre, dall'Esempio 2.1.5 segue facilmente che

$$\mathscr{G} = \bigcap_{\mathscr{F} \in \mathscr{I}} \mathscr{F}$$

è una σ-algebra. Evidentemente $\mathscr{G}$ è contenuta in ogni σ-algebra contenente $\mathscr{C}$. Dunque $\mathscr{G} = \sigma(\mathscr{C})$ è la più piccola σ-algebra contenente $\mathscr{C}$ e la Definizione 2.1.7 è ben posta.

Esempio 2.1.8. *(σ-algebra generata da una partizione numerabile)* Sia $\mathscr{C} = \{C_i\}_{i\in I}$ una partizione al più numerabile di S. Allora, la σ-algebra $\mathscr{F}$ costruita nell'Esempio 2.1.6 è la σ-algebra $\sigma(\mathscr{C})$. Verifichiamolo. Intanto $\mathscr{F}$ è una σ-algebra che ovviamente contiene $\mathscr{C}$ (basta prendere $I_A = \{i\}$, per $i \in I$), dunque $\sigma(\mathscr{C}) \subset \mathscr{F}$. Mostriamo ora che $\mathscr{F} \subset \sigma(\mathscr{C})$. Preso $A \in \mathscr{F}$, allora

$$A = \cup_{i\in I_A} C_i \in \sigma(\mathscr{C})$$

perché $C_i \in \mathscr{C} \subset \sigma(\mathscr{C})$ per ogni $i \in I_A$ e $\mathrm{card}(I_A) \leq \mathrm{card}(\mathbb{N})$.

Esempio 2.1.9. *(σ-algebra di Borel ed insiemi boreliani)* Se S è uno spazio topologico, *la σ-algebra di Borel* su S è la σ-algebra $\mathscr{B}(S)$ generata dagli aperti di S, cioè, se poniamo $\mathscr{O}$=classe degli insiemi aperti di S, $\mathscr{B}(S) = \sigma(\mathscr{O})$. Si tratta di una σ-algebra che interviene molto spesso in probabilità. In questo esempio sviluppiamo un po' i metodi per lavorarci. Intanto osserviamo che ci sono altre classi di insiemi che generano la σ-algebra di Borel. Ad esempio, se poniamo $\mathscr{C}$=classe degli insiemi chiusi di S, allora si ha anche $\mathscr{B}(S) = \sigma(\mathscr{C})$. Infatti se chiamiamo $\mathscr{F}$ la σ-algebra generata da $\mathscr{C}$, allora certamente $\mathscr{F}$ contiene $\mathscr{O}$, dato che gli aperti sono i complementari dei chiusi, e dunque anche $\mathscr{B}(S)$, dato che quest'ultima è la più piccola σ-algebra contenente $\mathscr{O}$. Con lo stesso ragionamento si mostra che $\mathscr{B}(S) \supset \mathscr{F}$.
Il caso $S = \mathbb{R}$ è chiaramente da particolare interesse. Non è facile immaginare come siano fatti i boreliani di $\mathbb{R}$. È però di solito facile mostrare che un dato insieme è un boreliano. Ad esempio sono boreliani gli intervalli semiaperti, cioè della forma

$$]a, b]$$

Infatti è facile verificare che

$$]a,b] = \bigcap_{n=1}^{\infty}]a, b + \tfrac{1}{n}[$$

e dunque l'intervallo semiaperto si può scrivere come intersezione numerabile di aperti.
Si può dimostrare, ma non è proprio semplice, che non tutti i sottoinsiemi di $\mathbb{R}$ sono boreliani. In genere, per gli insiemi che si incontrano di solito, è facile dimostrare che sono boreliani con ragionamenti del tipo appena visto.
Altre classi di sottoinsiemi di $\mathbb{R}$ che generano i boreliani sono:

- *Gli intervalli aperti.* Infatti ogni insieme aperto di $\mathbb{R}$ si può scrivere come riunione numerabile di intervalli aperti($\mathbb{R}$ è uno spazio metrico separabile...).

- *Gli intervalli semi aperti.* Infatti ogni intervallo aperto si può scrivere come riunione numerabile di intervalli semi aperti:

$$]a,b[= \bigcup_{n=1}^{\infty}]a, b - \tfrac{1}{n}]$$

- *La classe delle semirette chiuse a destra* Indichiamola con $\pi(\mathbb{R})$, cioè

$$\pi(\mathbb{R}) = \{]-\infty, x]\,;\, x \in \mathbb{R}\}.$$

Infatti, mostriamo dapprima che $\mathscr{B}(\mathbb{R}) \supset \sigma(\pi(\mathbb{R}))$. Poniamo $\mathscr{O} = \{\text{insiemi aperti di } \mathbb{R}\}$. Preso $x \in \mathbb{R}$, allora

$$]-\infty, x] = \bigcap_{n=1}^{\infty} \underbrace{]-\infty, x + \tfrac{1}{n}[}_{\in \mathscr{O} \subset \mathscr{B}(\mathbb{R}), \forall n} \in \mathscr{B}(\mathbb{R}).$$

Ciò prova che $\pi(\mathbb{R}) \subset \mathscr{B}(\mathbb{R})$ e quindi $\sigma(\pi(\mathbb{R})) \subset \mathscr{B}(\mathbb{R})$.

Viceversa, proviamo che $\mathscr{B}(\mathbb{R}) \subset \sigma(\pi(\mathbb{R}))$. Per la definizione di σ-algebra generata, basta mostrare che tutti gli intervalli aperti sono contenuti in $\sigma(\pi(\mathbb{R}))$. Ma questo è facile. Infatti $\sigma(\pi(\mathbb{R}))$ contiene gli intervalli semi aperti: presi $a < b$, allora

$$]a,b] =]-\infty, b] \cap]-\infty, a]^c$$

da cui segue che $]a,b] \in \sigma(\pi(\mathbb{R}))$ per ogni $a < b$. Poiché sappiamo che $\mathscr{B}(\mathbb{R})$ è generata dagli intervalli semiaperti, segue che $\mathscr{B}(\mathbb{R}) \subset \sigma(\pi(\mathbb{R}))$.

2.1.2 Spazi e funzioni misurabili

Definizione 2.1.10. Siano S un insieme e $\mathscr{F}$ una σ-algebra di S. La coppia $(S, \mathscr{F})$ è detta uno *spazio misurabile*.

Definizione 2.1.11. Siano $(S, \mathscr{S})$ e $(E, \mathscr{E})$ due spazi misurabili e sia $f : S \to E$ una funzione su S a valori in E. f è detta una *funzione misurabile* se la controimmagine tramite f di un qualsiasi insieme $A \in \mathscr{E}$ è un sottoinsieme che appartiene alla σ-algebra $\mathscr{S}$. Ovvero se

$$\text{per ogni } A \in \mathscr{E} \text{ allora } A' = f^{-1}(A) = \{s; f(s) \in A\} \in \mathscr{S},$$

Nel seguito useremo, per indicare la controimmagine di un insieme A una qualunque delle notazioni equivalenti

$$f^{-1}(A) \quad \text{oppure} \quad \{s; f(s) \in A\} \quad \text{oppure} \quad \{f \in A\}$$

Talvolta ci troveremo in situazioni in cui sull'insieme S ci sono più di una σ-algebra. In questo caso, per precisare rispetto a quale di esse la funzione f è misurabile, diremo che f è $\mathscr{S}$-misurabile. Parleremo quindi di funzioni $\mathscr{S}$-misurabili se sull'insieme S si considera la σ-algebra $\mathscr{S}$.
Altra specifica notazionale. Anche la σ-algebra di riferimento su E può variare. Normalmente, qualora $E = \mathbb{R}$ o anche $E = \mathbb{R}^d$, se non verrà fatto riferimento alla σ-algebra $\mathscr{E}$ significherà che $\mathscr{E} = \mathscr{B}(E)$.
Le funzioni misurabili rivestono un'importanza fondamentale in particolare nel calcolo delle probabilità (più in generale, nella teoria della misura). È chiaro dalla definizione che la proprietà di misurabilità di una funzione dipende dalle σ-algebre $\mathscr{S}$ e $\mathscr{E}$. Ad esempio, se sostituiamo $\mathscr{S}$ con un'altra σ-algebra $\mathscr{S}'$ più piccola, f potrebbe non risultare più misurabile come applicazione tra gli spazi misurabili $(S, \mathscr{S}')$ e $(E, \mathscr{E})$. Lo stesso potrebbe succedere se sostituissimo $\mathscr{E}$ con una σ-algebra più grande.

Definizione 2.1.12. *(σ-algebra generata da una funzione misurabile)* Siano $(S, \mathscr{S})$ e $(E, \mathscr{E})$ due spazi misurabili e sia $f : S \to E$ una funzione $\mathscr{S}$-misurabile. La *σ-algebra generata da f*, in simboli $\sigma(f)$, è la più piccola σ-algebra di S rispetto alla quale f risulti misurabile.

In altre parole, $\sigma(f)$ è la σ-algebra tale che f è $\sigma(f)$ misurabile e per ogni σ-algebra $\mathscr{F}_f$ di S tale che f è $\mathscr{F}_f$ misurabile, allora $\mathscr{F}_f \supset \sigma(f)$.

Osservazione 2.1.13. È facile vedere che $\sigma(f)$ è esplicitamente data dagli insiemi che sono controimmagine tramite f di insiemi di $\mathscr{E}$, cioè

$$\sigma(f) = \{f^{-1}(A); A \in \mathscr{E}\}.$$

Vediamo perché. Intanto, poniamo per il momento $\mathscr{F}_f = \{f^{-1}(A); A \in \mathscr{E}\}$ (si noti che, poiché $f^{-1}(A) \subset S$, $\mathscr{F}_f$ è una classe di sottoinsiemi di S) e mostriamo che $\mathscr{F}_f$ è una σ-algebra di S. Innanzitutto essa contiene $\emptyset$ e S, poiché

$$\emptyset = f^{-1}(\underbrace{\emptyset}_{\in \mathscr{E}}) \in \mathscr{F}_f \quad \text{e} \quad S = f^{-1}(\underbrace{E}_{\in \mathscr{E}}) \in \mathscr{F}_f.$$

Inoltre, preso $A' \in \mathscr{F}_f$ allora $A' = f^{-1}(A)$ per qualche $A \in \mathscr{E}$, quindi

$$A'^c = \big(f^{-1}(A)\big)^c = f^{-1}(\underbrace{A^c}_{\in \mathscr{E}}) \in \mathscr{F}_f.$$

Infine, presi $A'_n \in \mathscr{F}_f$ per $n = 1, 2, \ldots$, allora $A'_n = f^{-1}(A_n)$ per qualche $A_n \in \mathscr{E}$ e si vede che

$$\bigcup_n A'_n = \bigcup_n f^{-1}(A_n) = f^{-1}\Big(\underbrace{\bigcup_n A_n}_{\in \mathscr{E}}\Big) \in \mathscr{F}_f.$$

Abbiamo usato qui due formule che sono sempre valide:

$$\big(f^{-1}(A)\big)^c = f^{-1}(A^c) \quad \text{e} \quad \bigcup_n f^{-1}(A_n) = f^{-1}\Big(\bigcup_n A_n\Big) \tag{2.1}$$

La seconda di queste vale anche se alle riunioni si sostituiscono le intersezioni. Attenzione, le stesse formule scritte con f al posto di f^{-1} (e con A, A_n sottoinsiemi di S) *non* sono vere.
Mostriamo ora che $\sigma(f) = \mathscr{F}_f$. Ovviamente, f è $\mathscr{F}_f$ misurabile per costruzione, dunque $\mathscr{F}_f \supset \sigma(f)$. Per verificare l'inclusione opposta, osserviamo che, per ipotesi, f è $\sigma(f)$-misurabile, quindi $f^{-1}(A)$ deve appartenere a $\sigma(f)$ per ogni $A \in \mathscr{E}$, dunque, poiché tutti gli insiemi di $\mathscr{F}_f$ sono di questa forma, $\mathscr{F}_f \subset \sigma(f)$.

Può succedere che una σ-algebra $\mathscr{F}$ sia definita in un modo poco esplicito (si pensi ad esempio al caso $\mathscr{S} = \mathscr{B}(\mathbb{R})$). Per questo verificare la misurabilità di una funzione usando direttamente la Definizione 2.1.11 può risultare complicato. La proposizione che segue fornisce un criterio pratico molto utile per la verifica della misurabilità:

Proposizione 2.1.14. *1. Siano $(S, \mathscr{S})$ e $(E, \mathscr{E})$ due spazi misurabili e sia $f : S \to E$ una funzione. Sia $\mathscr{C} \subset \mathscr{E}$ tale che $\sigma(\mathscr{C}) = \mathscr{E}$. Se $f^{-1}(\Gamma) \in \mathscr{S}$ per ogni $\Gamma \in \mathscr{C}$, allora f è $\mathscr{S}$-misurabile.*

2. Supponiamo $E = \mathbb{R}$, $\mathscr{E} = \mathscr{B}(\mathbb{R})$. Se $f^{-1}((-\infty, c]) = \{s : f(s) \leq c\} = \{f \leq c\} \in \mathscr{F}$ per ogni $c \in \mathbb{R}$, allora f è $\mathscr{S}$-misurabile.

Dimostrazione. 1. Poniamo

$$\mathscr{G} = \{A \in \mathscr{E}; f^{-1}(A) \in \mathscr{S}\}$$

$\mathscr{G}$ è una classe d'insiemi contenuta un $\mathscr{E}$, evidentemente. È facile verificare che $\mathscr{G}$ è una σ-algebra. Infatti, $\emptyset \in \mathscr{G}$, $(f^{-1}(\emptyset) = \emptyset \in \mathscr{S})$ e $E \in \mathscr{G}$ $(f^{-1}(E) = S \in \mathscr{S})$. Inoltre, se $A \in \mathscr{G}$ allora $f^{-1}(A) \in \mathscr{S}$, dunque $f^{-1}(A^c) = (f^{-1}(A))^c \in \mathscr{S}$ per la prima delle (2.1) e dunque $A^c \in \mathscr{G}$. Infine, se $A_1, A_2, \dots \in \mathscr{G}$, allora $f^{-1}(A_n) \in \mathscr{F}$ per ogni n, dunque $f^{-1}(\bigcup_n A_n) = \bigcup_n f^{-1}(A_n) \in \mathscr{S}$, per la seconda delle (2.1) e dunque $\bigcup_n A_n \in \mathscr{G}$.
L'ipotesi assicura che $\mathscr{C} \subset \mathscr{G} \subset \mathscr{E}$, da cui segue $\sigma(\mathscr{C}) \subset \mathscr{G} \subset \mathscr{E}$. Poiché per ipotesi $\sigma(\mathscr{C}) = \mathscr{E}$, deve essere $\mathscr{G} = \mathscr{E}$, ovvero: per ogni $A \in \mathscr{E}$ allora $f^{-1}(A) \in \mathscr{S}$, che non è altro che la definizione di $\mathscr{S}$-misurabilità.
2. Supponiamo $E = \mathbb{R}$, $\mathscr{E} = \mathscr{B}(\mathbb{R})$. La tesi segue dal punto precedente, prendendo $\mathscr{C} = \pi(\mathbb{R})$ e ricordando che $\sigma(\pi(\mathbb{R})) = \mathscr{B}(\mathbb{R})$ (si veda l'Esempio 2.1.9).

□

Esempio 2.1.15. *(Funzioni discrete: misurabilità)* Siano $(S, \mathscr{S})$ e $(E, \mathscr{E})$ due spazi misurabili e sia $f : S \to E$ una funzione. Supponiamo che l'insieme $E = \{e_1, e_2, \dots\}$ sia discreto e che $\mathscr{E} = \mathscr{P}(E)$. Allora, f è misurabile se e solo se

$$\{f = e_i\} \in \mathscr{S} \quad \text{per ogni} \quad e_i \in E.$$

Infatti, fissato $e_i \in E$, ovviamente $\{f = e_i\} = f^{-1}(\{e_i\}) \in \mathscr{S}$ perché tutti gli insiemi della forma $\{e_i\}$ appartengono a $\mathscr{E}$ per ogni $e_i \in E$. Viceversa, supponiamo che $\{f = e_i\} \in \mathscr{S}$ per ogni $e_i \in E$. Basta poi osservare che $\sigma(E) = \mathscr{E}$ in questo caso ed applicare la Proposizione 2.1.14.

Esempio 2.1.16. *(Funzioni discrete: σ-algebra generata)* Consideriamo la stessa situazione dell'esempio precedente. Vogliamo vedere come è fatta la σ-algebra $\sigma(f)$ generata da f. Più precisamente, poniamo $C_i = \{f = e_i\}$. Evidentemente $C_i \in \mathscr{S}$. Mostreremo che $\mathscr{C} = (C_i)_i$ è una partizione di S e che $\sigma(f)$ coincide con la σ-algebra generata dalla partizione $\mathscr{C}$, che abbiamo visto nell'Esempio 2.1.8
Mostriamo che $\mathscr{C}$ è una partizione. Se $i \neq j$, allora se esistesse un elemento $s \in S$ appartenente a $C_i \cap C_j$, dovrebbe essere simultaneamente $f(s) = e_i$ e $f(s) = e_j$, che è impossibile. Dunque C_i e C_j devono essere disgiunti. Inoltre se $s \in S$, allora esiste $e_i \in E$ tale che $f(s) = e_i$ e dunque $s \in C_i$. Ne segue che la riunione degli insiemi C_i è uguale a S.
Mostriamo ora che $\sigma(f) = \sigma(\mathscr{C})$. Al solito useremo il metodo della doppia inclusione.

1) $\sigma(f) \subset \sigma(\mathscr{C})$. Se $A' \in \sigma(f)$ allora (cfr. Esempio 2.1.12) si ha, per qualche $A \in \mathscr{E}$, $A' = f^{-1}(A)$. Possiamo allora scrivere $A = \bigcup_{e_i \in A} \{e_i\}$. Dunque, sempre usando la seconda delle (2.1)

$$A' = f^{-1}(A) = \bigcup_{e_i \in A} f^{-1}(\{x_i\}) = \bigcup_{e_i \in A} C_i \in \sigma(\mathscr{C}),$$

2) $\sigma(\mathscr{C}) \subset \sigma(f)$. Se $A \in \sigma(\mathscr{C})$ allora per qualche $I_A \subset I$ si ha (vedi l'Esempio 2.1.12)

$$A = \bigcup_{i \in I_A} \{f = e_i\} = f^{-1}\Big(\bigcup_{i \in I_A} \{e_i\}\Big) \in \sigma(f)$$

perché $\bigcup_{i \in I_A} \{e_i\} \in \mathscr{E}$.

Riassumiamo nella proposizione che segue alcune proprietà basilari delle funzioni misurabili.

Proposizione 2.1.17. *1. Se f, f_1, f_2 sono funzioni $\mathscr{S}$-misurabili e se $\alpha \in \mathbb{R}$ allora*[2] *$f_1 + f_2$, $f_1 \cdot f_2$, αf sono $\mathscr{S}$-misurabili.*

2. Se $f : S \to \mathbb{R}^d$ è $\mathscr{S}$-misurabile e $h : \mathbb{R}^d \to \mathbb{R}^d$ è $\mathscr{B}(\mathbb{R}^d)$-misurabile allora la funzione composta $h \circ f : S \to \mathbb{R}^d$ è $\mathscr{S}$-misurabile.

3. Se $\{f_n\}_n$ è una successione di funzioni $\mathscr{S}$-misurabili a valori in $\mathbb{R}$, allora $inf_n f_n$, $\sup_n f_n$, $\liminf_n f_n$, $\limsup_n f_n$ sono[3] *$\mathscr{S}$-misurabili. Inoltre,*

$$\{s \,:\, \textit{esiste } \lim_n f_n(s)\} \in \mathscr{S}.$$

Dimostrazione. *1.* Per verificare la prima proprietà, basta osservare che

$$\{f_1 + f_2 > c\} = \cup_{q \in \mathbb{Q}}(\{f_1 > q\} \cap \{f_2 > c - q\}).$$

Usando (opportunamente) la *2.* della Proposizione 2.1.14, si ottiene la tesi. Per le altre, si ragiona in modo analogo.

2. Preso $\Lambda \in \mathscr{B}(\mathbb{R}^d)$, si ha

$$(h \circ f)^{-1}(\Lambda) = \underbrace{f^{-1}(\underbrace{h^{-1}(\Lambda)}_{\substack{\in \mathscr{B}(\mathbb{R}^d)\ [h \text{ è} \\ \mathscr{B}(\mathbb{R}^d)\text{-mis.}]}})}_{\in \mathscr{S}[f \text{ è } \mathscr{F}\text{-mis.}])}) \in \mathscr{S},$$

quindi $h \circ f$ è $\mathscr{S}$-misurabile.

3. Basta osservare che, per ogni $c \in \mathbb{R}$,

$$\{\inf_n f_n > c\} = \cap_n \{f_n > c\}$$

ed usare *2.* della proposizione 2.1.14. Analogamente si mostra la $\mathscr{S}$-misurabilità di $\sup_n f_n$, $\liminf_n f_n$ e $\limsup_n f_n$. Infine

$$\{s \,:\, \text{esiste } \lim_n f_n(s)\}$$

[2] Qui, la notazione "$\cdot$" ha senso per $d = 1$.

[3] Tali funzioni sono, in generale, a valori in $[-\infty, +\infty]$: ad essere precisi, occorrerebbe definire la σ-algebra $\mathscr{B}([-\infty, +\infty])$. Essa si definisce come la σ-algebra generata dagli aperti di $[-\infty, +\infty]$, o anche come la σ-algebra generata dalla classe $\mathscr{C}$ formata da $\mathscr{B}(\mathbb{R})$ e dai singleton $\{-\infty\}$ e $\{+\infty\}$. Ma non vogliamo entrare più in dettaglio...

$$= \{\liminf_n f_n > -\infty\} \cap \{\limsup_n f_n < +\infty\} \cap \{\liminf_n f_n = \limsup_n f_n\}$$
$$= \{\liminf_n f_n > -\infty\} \cap \{\limsup_n f_n < +\infty\} \cap \{\liminf_n f_n - \limsup_n f_n = 0\}$$

che è un elemento di $\mathscr{S}$ perché $\liminf_n f_n$, $\limsup_n f_n$ sono $\mathscr{S}$-misurabili e dunque anche $\liminf_n f_n - \limsup_n f_n$ lo è.

□

Nella proposizione che segue, vediamo che quando $\mathscr{S}$ è una σ-algebra generata da una partizione allora la $\mathscr{S}$-misurabilità dà informazioni molto precise su com'è fatta f. Infatti,

Proposizione 2.1.18. *Sia $f : S \to \mathbb{R}^d$ una funzione misurabile. Se f è $\sigma(\mathscr{C})$-misurabile, con $\mathscr{C}$ partizione al più numerabile di S, allora f è costante su ciascun elemento C di $\mathscr{C}$.*

Dimostrazione. Supponiamo, per assurdo, che esista $C \in \mathscr{C}$ tale che $s_1, s_2 \in C$, con $s_1 \neq s_2$, e $a_1 = f(s_1) \neq f(s_2)$. Poniamo $A_1 = f^{-1}(\{a_1\}$. Poiché $\{a_1\} \in \mathscr{B}(\mathbb{R})$, si ha $A_1 \in \mathscr{S} = \sigma(\mathscr{C})$. Ricordando che $s_1 \in A_1$ e che $\sigma(\mathscr{C})$ è costituita da tutte le possibili unioni al più numerabili di elementi di $\mathscr{C}$ (cfr. Esempio 2.1.8), necessariamente dev'essere $A_1 \supseteq C$, il che è assurdo perché $s_2 \in C$ ma $s_2 \notin A_1$.

□

Esercizio 2.1.19. Sia $f : S \to \mathbb{R}^d$ una funzione misurabile. Dimostrare che f è costante se e solo se $\sigma(f) = \{\emptyset, \Omega\}$ (la σ-algebra banale).

Presentiamo un ulteriore risultato, di cui omettiamo la dimostrazione (per la quale si rimanda, ad esempio, al Paragrafo A.3.2 di Williams [3]).

Proposizione 2.1.20. *Siano $f : S \to \mathbb{R}^d$, e $g : S \to \mathbb{R}^m$ due funzioni $\mathscr{S}$-misurabili. Supponiamo che g sia in particolare $\sigma(f)$-misurabile. Allora, esiste una funzione $h : \mathbb{R}^d \to \mathbb{R}^m$ che sia $\mathscr{B}(\mathbb{R}^d)$-misurabile tale che $g = h \circ f$.*

2.2 Spazi di probabilità

Definizione 2.2.1. *Uno spazio di probabilità* è una tripla $(\Omega, \mathscr{F}, \mathbb{P})$ costituita da:

- uno spazio (astratto) Ω, detto *spazio campionario*;
- una σ-algebra $\mathscr{F}$ su Ω, che rappresenta l'*insieme di tutti i possibili eventi*; un elemento $A \in \mathscr{F}$ prende quindi il nome di *evento*;

- una *misura di probabilità*, o semplicemente una probabilità, $\mathbb{P}$ su $(\Omega, \mathscr{F})$, finita e di massa 1:

  $\mathbb{P} : \mathscr{F} \to [0,1]$ è tale che

  $*$ $\mathbb{P}(\Omega) = 1$,

  $*$ $\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \displaystyle\sum_{n=1}^{\infty} \mathbb{P}(A_n)$, se $A_n \cap A_m = \emptyset$ per ogni $n \neq m$.

  La seconda proprietà è nota col nome di *σ-additività*.

Osserviamo che poiché $A_1, A_2, \ldots \in \mathscr{F}$ e $\mathscr{F}$ è una σ-algebra, allora $A = \bigcup_n A_n \in \mathscr{F}$, dunque $\mathbb{P}$ è effettivamente definita su A.
Osserviamo inoltre che, alla luce di quanto visto in precedenza, la coppia $(\Omega, \mathscr{F})$ è uno spazio di misura.

Esempio 2.2.2. Supponiamo $\mathscr{F} = \sigma(\mathscr{C})$, dove $\mathscr{C}$ è una partizione al più numerabile di Ω. Allora, una qualsiasi probabilità $\mathbb{P}$ su $(\Omega, \mathscr{F}) = (\Omega, \sigma(\mathscr{C}))$ è univocamente individuata dal valore che assume sugli elementi della partizione $\mathscr{C}$. Cioè, posto $\mathscr{C} = \{C_i\,;\, i \in I\}$, allora $\mathbb{P}$ è perfettamente individuata dai numeri

$$0 \leq p_i = \mathbb{P}(C_i), \quad i \in I, \quad \text{tali che } \sum_{i \in I} p_i = 1.$$

Questo perché ogni $A \in \sigma(\mathscr{C})$ è della forma $A = \bigcup_{i \in I_A} C_i$, con $I_A \subseteq I$ finito o numerabile, dunque la σ-additività di $\mathbb{P}$ garantisce che

$$\mathbb{P}(A) = \mathbb{P}\Big(\bigcup_{i \in I_A} C_i\Big) = \sum_{i \in I_A} \mathbb{P}(C_i) = \sum_{i \in I_A} p_i.$$

2.2.1 Prime proprietà

Vediamo alcune proprietà elementari della probabilità.

Proposizione 2.2.3. *Sia $(\Omega, \mathscr{F}, \mathbb{P})$ uno spazio di probabilità.*

1. *Se $A, B \in \mathscr{F}$ sono tali che $A \subset B$ allora $\mathbb{P}(A) \leq \mathbb{P}(B)$.*

2. *(Sub-additività) Per ogni $A_1, A_2 \in \mathscr{F}$ allora $\mathbb{P}(A_1 \cup A_2) \leq \mathbb{P}(A_1) + \mathbb{P}(A_2)$. Più in generale, per ogni $A_1, A_2, \ldots \in \mathscr{F}$,*

$$\mathbb{P}\Big(\bigcup_n A_n\Big) \leq \sum_n \mathbb{P}(A_n).$$

3. *Per ogni $A_1, A_2, \ldots \in \mathscr{F}$ tali che $\mathbb{P}(A_n) = 0$ per ogni n, allora*

$$\mathbb{P}\Big(\bigcup_n A_n\Big) = 0.$$

4. Per ogni $A \in \mathscr{F}$, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

5. (Formula di inclusione-esclusione) Per ogni $A_1, A_2 \in \mathscr{F}$ allora $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2)$. Più in generale, per ogni $A_1, \dots, A_N \in \mathscr{F}$,

$$\mathbb{P}\Big(\bigcup_{n=1}^{N} A_n\Big) = \sum_{n=1}^{N} \mathbb{P}(A_n) - \sum_{n<m\leq N} \mathbb{P}\Big(A_n \bigcap A_m\Big)$$
$$+ \sum_{n<m<\ell\leq N} \mathbb{P}\Big(A_n \bigcap A_m \bigcap A_\ell\Big) - \dots + (-1)^{N-1}\mathbb{P}\Big(A_1 \bigcap \dots \bigcap A_N\Big).$$

Dimostrazione. *1.* Poiché $B = A \cup (A^c \cap B)$ e quest'ultima unione è disgiunta,

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(A^c \cap B) \geq \mathbb{P}(A).$$

2. Poniamo $B_1 = A_1$ e per $n > 1$, $B_n = A_n \cap (\cup_{k=1}^{n-1} A_k)^c$. B_n sono a due a due disgiunti, $\cup_n B_n = \cup_n A_n$ ed inoltre $B_n \subseteq A_n$ per ogni n. Quindi,

$$\mathbb{P}(\cup_n A_n) = \mathbb{P}(\cup_n B_n) = \sum_n \mathbb{P}(B_n) \leq \sum_n \mathbb{P}(A_n).$$

3. Segue immediatamente da *2.*

4. Poiché $\Omega = A \cup A^c$, con A, A^c ovviamente disgiunti, si ha $1 = \mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A^c)$, quindi $\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$.

5. $A_1 \cup A_2 = (A_1 \cap A_2^c) \cup (A_1^c \cap A_2) \cup (A_1 \cap A_2)$. Poiché si tratta di elementi di $\mathscr{F}$ a due a due disgiunti, otteniamo

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1 \cap A_2^c) + \mathbb{P}(A_1^c \cap A_2) + \mathbb{P}(A_1 \cap A_2).$$

Ora, $A_1 = (A_1 \cap A_2^c) \cup (A_1 \cap A_2)$ ed essendo disgiunti, $\mathbb{P}(A_1) = \mathbb{P}(A_1 \cap A_2^c) + \mathbb{P}(A_1 \cap A_2)$. Analogamente, $\mathbb{P}(A_2) = \mathbb{P}(A_1^c \cap A_2) + \mathbb{P}(A_1 \cap A_2)$. Sostituendo, si ottiene la tesi.

La formula generale di inclusione-esclusione è lasciata per esercizio.

Osserviamo che, poiché $\mathscr{F}$ è una σ-algebra, tutti gli insiemi presi in questa dimostrazione di cui viene valutata la probabilità $\mathbb{P}$ appartengono alla σ-algebra $\mathscr{F}$, cioè appartengono effettivamente al dominio di definizione dell'applicazione $\mathbb{P}$.

$\square$

Nella teoria della probabilità (e più in generale in teoria della misura), particolare rilevanza hanno gli insiemi di probabilità nulla, per i quali sottolineiamo vale la proprietà *3.* della Proposizione 2.2.3: l'unione al più numerabile di eventi di probabilità nulla ha ancora probabilità nulla.

Presentiamo inoltre le seguenti proprietà di convergenza monotona della probabilità. A tale scopo, ricordiamo che, prese due successioni $\{a_n\}_n$ e $\{b_n\}_n$ di numeri reali, si scrive:

- $a_n \uparrow a$ se $a_n \leq a_{n+1}$ per ogni n e $a = \sup_n a_n = \lim_n a_n$;
- $b_n \downarrow b$ se $b_n \geq b_{n+1}$ per ogni n e $b = \inf_n b_n = \lim_n b_n$.

In qualche modo si può generalizzare a successioni $\{A_n\}_n$ e $\{B_n\}_n$ di insiemi, nel modo seguente:

- $A_n \uparrow A$ se $A_n \subset A_{n+1}$ per ogni n e $A = \cup_n A_n$;
- $B_n \downarrow B$ se $B_n \supset B_{n+1}$ per ogni n e $B = \cap_n B_n$.

Allora,

Proposizione 2.2.4. *Sia $(\Omega, \mathscr{F}, \mathbb{P})$ uno spazio di probabilità e siano $\{A_n\}_n$, $\{B_n\}_n \subset \mathscr{F}$.*

1. *Se $A_n \uparrow A$ allora $\mathbb{P}(A_n) \uparrow \mathbb{P}(A)$.*
2. *Se $B_n \downarrow B$ allora $\mathbb{P}(B_n) \downarrow \mathbb{P}(B)$.*

Dimostrazione. *1.* Ovviamente $\mathbb{P}(A_n)$ è una successione (numerica) crescente, perché $A_n \subset A_{n+1}$. Poniamo

$$A'_1 = A_1 \quad \text{e per } n > 1, \quad A'_n = A_n \setminus A_{n-1}.$$

Allora: gli A'_n sono a due a due disgiunti, tali che $A = \cup_n A_n = \cup_n A'_n$ ed inoltre $A_n = \cup_{k=1}^n A'_n$. Ma allora,

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(\cup_n A'_n) = \sum_n \mathbb{P}(A'_n) = \lim_{n\to\infty} \sum_{k=1}^n \mathbb{P}(A'_n) \\ &= \lim_{n\to\infty} \mathbb{P}(\cup_{k=1}^n A'_n) = \lim_{n\to\infty} \mathbb{P}(A_n) = \sup_n \mathbb{P}(A_n).\end{aligned}$$

2. Poniamo $A_n = B_n^c$: $A_n \uparrow A = \cup_n A_n = \cup_n B_n^c = (\cap_n B_n)^c = B^c$, quindi da *1.* segue che $\mathbb{P}(A_n) \uparrow \mathbb{P}(B^c)$. Allora, $\mathbb{P}(B_n) = 1 - \mathbb{P}(A_n) \downarrow 1 - \mathbb{P}(B^c) = \mathbb{P}(B)$.

□

2.2.2 Indipendenza

Sia $(\Omega, \mathscr{F}, \mathbb{P})$ uno spazio di probabilià. Ricordiamo che n eventi $A_1, \ldots, A_n \in \mathscr{F}$ si dicono indipendenti se per ogni sottoinsieme $\{i_1, \ldots, i_k\}$ di $\{1, \ldots, n\}$ si ha

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k}).$$

Ad esempio, limitiamoci al caso $n = 2$ e prendiamo due eventi indipendenti A e B: $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Allora,

$$\mathbb{P}(A \cap B^c) = \mathbb{P}(A) - \mathbb{P}(A \cap B) = \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B)$$

$$= \mathbb{P}(A)\Big(1 - \mathbb{P}(B)\Big) = \mathbb{P}(A)\mathbb{P}(B^c).$$

Analogamente, si fa vedere che $\mathbb{P}(A^c \cap B) = \mathbb{P}(A^c)\mathbb{P}(B)$ e $\mathbb{P}(A^c \cap B^c) = \mathbb{P}(A^c)\mathbb{P}(B^c)$. Ciò significa che, preso comunque un elemento $G_1 \in \mathscr{G}_1 = \sigma(\{A\}) = \{\emptyset, \Omega, A, A^c\}$ e un elemento $G_2 \in \mathscr{G}_2 = \sigma(\{B\}) = \{\emptyset, \Omega, B, B^c\}$, allora $\mathbb{P}(G_1 \cap G_2) = \mathbb{P}(G_1,)\mathbb{P}(G_2)$, cioè G_1 e G_2 sono indipendenti.

Ciò giustifica le seguente definizione, più generale, di indipendenza.

Definizione 2.2.5. Siano $\mathscr{G}_1, \dots, \mathscr{G}_n$ sotto σ-algebre i $\mathscr{F}$, cioè $\mathscr{G}_i \subset \mathscr{F}$ è una σ-algebra, $i = 1, \dots, n$. $\mathscr{G}_1, \dots, \mathscr{G}_n$ si diconno *indipendenti* se

$$\mathbb{P}(G_1 \cap \dots \cap G_n) = \mathbb{P}(G_1) \cdots \mathbb{P}(G_n), \quad \text{per ogni } G_1 \in \mathscr{G}_1, \dots, G_n \in \mathscr{G}_n.$$

Pensando a $\mathscr{G}_1, \dots, \mathscr{G}_n$ come a n possibili esperimenti, la Definizione 2.2.5 dice che gli n esperimenti sono indipendenti se presi comunque n possibili risultati $G_1 \in \mathscr{G}_1, \dots, G_n \in \mathscr{G}_n$ allora gli eventi $G_1, \dots, G_n$ sono indipendenti.

Esempio 2.2.6. Supponiamo $\mathscr{G}_i = \sigma(\mathscr{C}_i)$, con C_i partizione al più numerabile di Ω, per $i = 1, \dots, n$. Allora, $\mathscr{G}_1, \dots, \mathscr{G}_n$ sono indipendenti se e solo se

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n), \quad \text{per ogni } A_1 \in \mathscr{C}_1, \dots, A_n \in \mathscr{C}_n$$

In altre parole, quando le σ-algebre sono generate da una partizione al più numerabile, è sufficiente provare l'indipendenza tra gli atomi che compongono le partizioni.
Per semplicità di notazione, dimostriamolo sono nel caso $n = 2$. Siano quindi $G_1 \in \mathscr{G}_1$ e $G_2 \in \mathscr{G}_2$. Allora

$$\begin{aligned} G_1 &= \cup_{j \in I_1} A_j, \quad A_j \in \mathscr{C}_1, j \in I_1 \\ G_2 &= \cup_{k \in I_2} B_k, \quad B_k \in \mathscr{C}_2, k \in I_2 \end{aligned}$$

e si ha:

$$\mathbb{P}\Big(G_1 \cap G_2\Big) = \mathbb{P}\Big(\cup_{j \in I_1, k \in I_2} (A_j \cap B_k)\Big) = \sum_{j \in I_1, k \in I_2} \mathbb{P}(A_j \cap B_k)$$
$$= \sum_{j \in I_1} \sum_{k \in I_2} \mathbb{P}(A_j)\mathbb{P}(B_k) = \mathbb{P}\Big(\cup_{j \in I_1} A_j\Big)\mathbb{P}\Big(\cup_{k \in I_2} B_k\Big) = \mathbb{P}(G_1)\mathbb{P}(G_2).$$

A volte parleremo di un evento indipendente da una σ-algebra, intendendo valida la seguente

Definizione 2.2.7. Sia $A \in \mathscr{F}$ e $\mathscr{G}$ una sotto σ-algebra di $\mathscr{F}$. Si dice che A è indipendente da $\mathscr{G}$ se

$$\mathbb{P}(A \cap G) = \mathbb{P}(A) \cdot \mathbb{P}(G), \quad \text{per ogni } G \in \mathscr{G}.$$

2.3 Variabili aleatorie

Abbiamo visto che se $(\Omega, \mathscr{F}, \mathbb{P})$ è uno spazio di probabilità allora in particolare $(\Omega, \mathscr{F})$ è uno spazio misurabile. Le variabili aleatorie sono delle funzioni definite su Ω, e in particolare misurabili. Infatti,

Definizione 2.3.1. Sia $(E, \mathscr{E})$ uno spazio misurabile. Una *variabile aleatoria* (v.a.) X su $(\Omega, \mathscr{F}, \mathbb{P})$ è una funzione

$$X\ :\ (\Omega, \mathscr{F}) \longrightarrow (E, \mathscr{E}) \quad \text{misurabile.}$$

Una v.a. X è detta *discreta* se l'insieme dei valori che assume $E_X = X(\Omega) \equiv \{x \in \mathbb{R} : x = X(\omega),\ \omega \in \Omega\}$ è un insieme discreto, cioè finito o numerabile. In particolare, diremo che una v.a. discreta X è *finita* se E_X è un insieme finito. Senza perdere in generalità, in tal caso sceglieremo $\mathscr{E} = \mathscr{P}(E_X)$.
Quando non tratteremo v.a. discrete, saranno per lo più continue e a valori in $E = \mathbb{R}$ oppure $E = \mathbb{R}^d$, con la scelta di $\mathscr{E} = \mathscr{B}(E)$. Dunque, una v.a. X è semplicemente una funzione su Ω a valori in $\mathbb{R}$ o in $\mathbb{R}^d$ che sia $\mathscr{F}$-misurabile.
Anche nel caso di v.a. si definisce l'indipendenza:

Definizione 2.3.2. n v.a. $X_1, \ldots, X_n$ si dicono indipendenti se

$$\mathbb{P}(X_1 \in A_1, \cdots X_n \in A_n) = \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_n \in A_n)$$

per ogni scelta di $A_1, \ldots, A_n$ tali che le probabilità sopra scritte abbiano senso, e cioè $A_1 \in \mathscr{E}_1, \ldots, A_n \in \mathscr{E}_n$ (si sta infatti supponendo che X_i sia a valori nello spazio misurabile $(E_i, \mathscr{E}_i)$).

Ricordando che $\sigma(X_i) = \{X_i^{-1}(A)\,;\, A \in \mathscr{E}_i\}$, la Definizione 2.3.2 si può riscrivere in modo equivalente come

$$\begin{aligned} &X_1, \ldots, X_n \text{ sono indipendenti se e solo se} \\ &\sigma(X_1), \ldots, \sigma(X_n) \text{ sono indipendenti.} \end{aligned}$$

A volte, ci troveremo a parlare di una v.a. indipendente da una σ-algebra. Sulla base delle definizioni viste, il modo più naturale di descrivere matematicamente tale situazione è il seguente:

Definizione 2.3.3. Sia X una v.a. e $\mathscr{G}$ una sotto σ-algebra di $\mathscr{F}$. Diremo che X e $\mathscr{G}$ sono indipendenti se $\sigma(X)$ e $\mathscr{G}$ sono indipendenti, o in modo equivalente se

$$\mathbb{P}((X \in A) \cap G) = \mathbb{P}(X \in A)\mathbb{P}(G) \quad \text{per ogni} \quad A \in \mathscr{E}, G \in \mathscr{G}.$$

Esercizio 2.3.4. Sia $A \in \mathscr{F}$ e $\mathscr{G}$ una sotto σ-algebra di $\mathscr{F}$. Verificare che A è indipendente da $\mathscr{G}$ (cfr. Definizione 2.2.7) se e solo se $X = 1_A$ è indipendente da $\mathscr{G}$.

Nei paragrafi che seguono, ricordiamo brevemente le proprietà principali e gli strumenti comunemente utilizzati nella trattazione delle variabili aleatorie.

2.3.1 Proprietà

Poiché una v.a. X è semplicemente una funzione misurabile, valgono tutte le proprietà viste nel Paragrafo 2.1.2, che riassumiamo qui di seguito.

- La σ-algebra generata da X, in simboli $\sigma(X)$, ovvero la più piccola σ-algebra contenuta in $\mathscr{F}$ rispetto alla quale X rimane misurabile, è data da
$$\sigma(X) = \{X^{-1}(A)\,;\, A \in \mathscr{E}\}.$$
In particolare, se X è discreta allora
$$\sigma(X) = \sigma(\mathscr{C}_X), \quad \text{dove } \mathscr{C}_X = \{X^{-1}(\{x\})\,;\, x \in E_X = X(\Omega)\},$$
e ricordiamo che $\mathscr{C}_X$ è una partizione al più numerabile di Ω (cfr. Osservazione 2.1.13 ed Esempio 2.1.16).
- Sia $X : \Omega \to \mathbb{R}$. X è una v.a. se e solo se $X^{-1}((-\infty, c]) \in \mathscr{F}$ per ogni $c \in \mathbb{R}$ (cfr. Proposizione 2.1.14).
- In particolare, se X è discreta allora X è una v.a. se e solo se $X^{-1}(\{x\}) \in \mathscr{F}$ per ogni $x \in E_X = X(\Omega)$ (cfr. Esempio 2.1.15).
- Se X, Y, $\{X_n\}_n$ sono v.a., $\alpha, \beta \in \mathbb{R}$ e $h : E_X \to \mathbb{R}^m$ è misurabile allora (cfr. Proposizione 2.1.17): $\alpha X + \beta Y$, $X \cdot Y$ sono v.a.; $h \circ X$ è una v.a.; $\inf_n X_n$, $\sup_n X_n$, $\liminf_n X_n$, $\limsup_n X_n$ sono v.a. ed inoltre $\{\omega : \text{esiste } \lim_n X_n(\omega)\} \in \mathscr{F}$.
- Se $\sigma(X) = \sigma(\mathscr{C})$ con $\mathscr{C}$ partizione al più numerabile di Ω allora X è discreta ed assume valore costante su ogni elemento C di $\mathscr{C}$ (cfr. Proposizione 2.1.18).
- Se X è una v.a. a valori in $\mathbb{R}^d$ e Y è una v.a. a valori in $\mathbb{R}^m$ e $\sigma(X)$-misurabile allora esiste $h : \mathbb{R}^d \to \mathbb{R}^m$ misurabile e tale che $Y = h(X)$ (cfr. Proposizione 2.1.20).

2.3.2 Legge e distribuzione

Com'è noto, nella pratica il formalismo introdotto per le v.a. non viene usato. Infatti, quando si lavora con variabili aleatorie in realtà spesso si perde di vista qual è l'originario spazio di probabilità $(\Omega, \mathscr{F}, \mathbb{P})$ e X vista come *applicazione* (misurabile) da Ω su E, anzi tali ingredienti neanche si conoscono. Piuttosto, si lavora con la distribuzione di X e con la sua densità (discreta o continua). Oppure, con la legge di X, che dà il legame tra variabili aleatorie e probabilità. Ricordiamo infatti che nei precedenti corsi di probabilità si è lavorato con v.a. bernoulliane, binomiali, geometriche oppure gaussiane, esponenziali etc. Ovvero, si sta specificando qual è la legge della v.a. presa in considerazione. In generale,

Definizione 2.3.5. La *legge* della v.a. $X = (X_1, \ldots, X_d)$ a valori in $E \subset \mathbb{R}^d$, o equivalentemente la *legge congiunta* di $X_1, \ldots, X_d$, è la misura di probabilità $\mathbb{P}_X : \mathscr{E} \to [0, 1]$ definita da

$$A \mapsto \mathbb{P}_X(A) = \mathbb{P}(X \in A), \quad A \in \mathscr{E}.$$

È immediato verificare che $\mathbb{P}_X$ è effettivamente una probabilità (esercizio!), non più su $(\Omega, \mathscr{F})$ bensì su $(E, \mathscr{E})$.

Nei due paragrafi che seguono, studiamo o meglio rivediamo un po' più in dettaglio due esempi particolarmente rilevanti: il caso in cui X è una v.a. discreta ed il caso in cui X è assolutamente continua.

Variabili aleatorie discrete

In questo caso, X è una v.a. a valori in $E = E_X = X(\Omega) = \{x^1, x^2, \ldots\}$, con $x^i = (x_1^i, \ldots, x_d^i)$, per ogni i, e la σ-algebra di riferimento su E si prende, senza perdere in generalità, come $\mathscr{E} = \mathscr{P}(E)$.

Quando si lavora con v.a. discrete, particolarmente utile è la *densità discreta* di X, o equivalentemente la *densità discreta congiunta* di $X_1, \ldots, X_d$. Essa è data da $p_X : E_X \to [0, 1]$, definita come

$$p_X(x) = \mathbb{P}(X = x) = \mathbb{P}(X = x_1, \ldots, X_d = x_d), \quad x = (x_1, \ldots, x_d) \in E_X.$$

In generale, per comodità è bene definire p_X anche al di fuori di E_X, ponendo ovviamente $p_X(x) = 0$ se $x \notin E_X$.

Ovviamente,

$$p_X(x) > 0 \text{ per ogni } x \in E_X \text{ e } \sum_{x \in E_X} p_X(x) = 1$$

e la legge di X è

$$\mathbb{P}_X(A) = \sum_{x \in A} p_X(x), \quad A \in \mathscr{P}(E).$$

Nota p_X, è completamente noto il comportamento (probabilistico) di X. Inoltre, se p è una funzione su un insieme discreto E e tale che

$$p(x) > 0 \text{ per ogni } x \in E \text{ e } \sum_{x \in E} p(x) = 1$$

allora (Teorema di Skorohod) si può dimostrare che esiste uno spazio di probabilità $(\Omega, \mathscr{F}, \mathbb{P})$ su cui è definita una v.a. X tale che $E_X = E$ e $p_X = p$.

Riassumiamo brevemente alcune delle proprietà principali della densità discreta p_X.

1. Nota p_X, è nota la densità discreta di un qualsiasi sotto-vettore di $X = (X_1, \ldots, X_d)$, e in particolare quindi tutte le densità marginali. Ad esempio, supponiamo per semplicità $d = 3$: $X = (X_1, X_2, X_3)$. La v.a. (X_1, X_2) assume valori in $\{z = (z_1, z_2)$: esiste i tale che $z_1 = x_1^i, z_2 = x_2^i\}$ e la densità discreta (congiunta) di X_1 e X_2 è data da

$$p_{X_1,X_2}(z_1, z_2) = \sum_{i\,:\,x_1^i = z_1, x_2^i = z_2} p_X(x^i) = \sum_i p_X(z_1, z_2, x_3^i).$$

Procedendo analogamente, la densità marginale, ad esempio, di X_1 è invece data da

$$p_{X_1}(\xi) = \sum_{i\,:\,x_1^i = \xi} p_X(x^i) = \sum_i p_X(\xi, x_2^i, x_3^i).$$

2. Le v.a. (discrete) $X_1, \ldots, X_d$ sono indipendenti[4] se e solo se per ogni $x = (x_1, \ldots, x_d) \in E_X$,

$$p_X(x_1, \ldots, x_d) = p_{X_1}(x_1) \cdots p_{X_d}(x_d)$$

dove p_{X_i} denota l'i-esima densità marginale.

3. Supponiamo, per semplicità, $d = 2$. La densità di $Y = X_1 + X_2$ è data da

$$p_Y(y) = \sum_{i\,:\,x_1^i + x_2^i = y} p_X(x_1^i, x_2^i) = \sum_t p_X(t, y - t) = \sum_t p_X(y - t, t).$$

Se in particolare X_1 e X_2 sono indipendenti, allora

$$p_Y(y) = \sum_t p_{X_1}(t)\, p_{X_2}(y - t) = \sum_t p_{X_1}(y - t)\, p_{X_2}(t).$$

4. Supponiamo X e Y v.a. discrete, l'una a valori in $E_X \subset \mathbb{R}^d$ e l'altra in $E_Y \subset \mathbb{R}^m$. La *densità condizionale* di X dato Y è data da

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

purché ovviamente $p_Y(y) > 0$, dove $p_{X,Y}$ e p_Y denotano, rispettivamente, la densità di $(X, Y) = (X_1, \ldots, X_d, Y_1, \ldots, Y_m)$ e quella (marginale) di Y. Ricordiamo che, fissato $y \in E_Y$ tale che $p_Y(y) > 0$, $p_{X|Y}(x|y)$ è una densità discreta su E_X, cioè

$$p_{X|Y}(x|y) \geq 0 \text{ per ogni } x \in E_X \text{ e } \sum_{x \in E_X} p_{X|Y}(x|y) = 1$$

[4]Ricordiamo che, in generale, d v.a. $X_1, \ldots, X_d$ si dicono indipendenti se $\mathbb{P}(X_1 \in A_1, \ldots, X_d \in A_d) = \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_d \in A_d)$, per ogni possibile scelta di $A_1, \ldots, A_d$ $(\in \mathscr{B}(\mathbb{R}))$.

Variabili aleatorie assolutamente continue

In questo caso, X assume valori in $E = \mathbb{R}^d$ e la σ-algebra di riferimento su E si prende come $\mathscr{E} = \mathscr{B}(E)$. Inoltre, dire che X è assolutamente continua significa supporre l'esistenza della *densità di probabilità* p_X di $X = (X_1, \dots, X_d)$, o equivalente la *densità di probabilità congiunta* di $X_1, \dots, X_d$:

$$\begin{aligned} &p_X : \mathbb{R}^d \to [0, +\infty) \text{ è integrabile su } \mathbb{R}^d \text{ e tale che} \\ &\mathbb{P}(X \in A) = \int_A p_X(x)\, dx, \quad A \in \mathscr{E}. \end{aligned}$$

Dunque, una v.a. X assolutamente continue ha una legge $\mathbb{P}_X$ che soddisfa ad una rappresentazione integrale, essendo

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \int_A p_X(x)\, dx, \quad A \in \mathscr{E}.$$

Ovviamente, la densità p_X è una funzione integrabile tale che

$$p_X(x) \geq 0 \quad \text{e} \quad \int_{\mathbb{R}^d} p_X(x)\, dx = 1.$$

Nota p_X, è completamente noto il comportamento (probabilistico) di X. Inoltre, se p è una funzione integrabile su $\mathbb{R}^d$ tale che

$$p(x) \geq 0 \text{ per ogni } x \text{ e } \int_{\mathbb{R}^d} p(x)\, dx = 1$$

allora (Teorema di Skorohod) si può dimostrare che esiste uno spazio di probabilità $(\Omega, \mathscr{F}, \mathbb{P})$ su cui è definita una v.a. assolutamente continua X con densità di probabilità $p_X = p$.

Riassumiamo brevemente alcune delle proprietà principali della densità (continua) p_X. Osserviamo che si tratta più o meno delle stesse proprietà già ricordate per una densità discreta: l'unica accortezza è quella di sostituire le somme con degli integrali!

1. Nota p_X, allora qualsiasi sotto-vettore di $X = (X_1, \dots, X_d)$ è assolutamente continuo, con densità ricavabile da p_X. In particolare, ogni componente X_i ha densità. Ad esempio, supponiamo per semplicità $X = (X_1, X_2, X_3)$: la densità (congiunta) di X_1 e X_2 è data da

$$p_{X_1,X_2}(x_1, x_2) = \int_{\mathbb{R}} p_X(x_1, x_2, x_3)\, dx_3$$

e la densità marginale, ad esempio, di X_1 è invece data da

$$p_{X_1}(x_1) = \int_{\mathbb{R}^2} p_X(x_1, x_2, x_3)\, dx_2\, dx_3$$

Ricordiamo che non vale il viceversa: se esiste la densità (marginale) di X_1 e di X_2, non è detto che esista la densità (congiunta) di (X_1, X_2).

2. Le v.a. assolutamente continue $X_1, \dots, X_d$ sono indipendenti[5] se e solo se per ogni $x = (x_1, \dots, x_d) \in \mathbb{R}^d$,

$$p_X(x_1, \dots, x_d) = p_{X_1}(x_1) \cdots p_{X_d}(x_d)$$

dove p_{X_i} denota l'i-esima densità marginale.

3. Supponiamo, per semplicità, $d = 2$. Se $X = (X_1, X_2)$ è assolutamente continua, allora $Y = X_1 + X_2$ è assolutamente continua ed ha densità

$$p_Y(y) = \int_{\mathbb{R}} p_X(t, y-t)\, dt = \int_{\mathbb{R}} p_X(y-t, t)\, dt$$

e se in particolare X_1 e X_2 sono indipendenti, allora

$$p_Y(y) = \sum_{\mathbb{R}} p_{X_1}(t)\, p_{X_2}(y-t)\, dt = \int_{\mathbb{R}} p_{X_1}(y-t)\, p_{X_2}(t)\, dt.$$

4. Supponiamo X e Y v.a. congiuntamente assolutamente continue, l'una a valori in $\mathbb{R}^d$ e l'altra in $\mathbb{R}^m$. La *densità condizionale* di X dato Y è data da

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

purché ovviamente $p_Y(y) > 0$, dove $p_{X,Y}$ e p_Y denotano, rispettivamente, la densità di $(X, Y) = (X_1, \dots, X_d, Y_1, \dots, Y_m)$ e quella (marginale) di Y. Ricordiamo che, fissato y tale che $p_Y(y) > 0$, $p_{X|Y}(x|y)$ è una densità (continua) su $\mathbb{R}^d$, cioè

$$x \mapsto p_{X|Y}(x|y) \geq 0 \text{ è integrabile su } \mathbb{R}^d \text{ e } \int_{\mathbb{R}^d} p_{X|Y}(x|y)\, dx = 1$$

5. Nel caso di v.a. non discrete, particolare rilievo assume la *funzione di distribuzione o di ripartizione (congiunta)* $F_X : \mathbb{R}^d \to [0,1]$: per $x = (x_1, \dots, x_d) \in \mathbb{R}^d$,

$$F_X(x) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d) = \mathbb{P}_X\Big((-\infty, x_1] \times \cdots \times (-\infty, x_d]\Big).$$

Ricordiamo in particolare che X è *assolutamente continua se e solo se esiste*

$$g(x) = \frac{\partial^d}{\partial x_1 \dots \partial x_d} F_X(x),$$

con g integrabile su $\mathbb{R}^d$ tale che $\int_{\mathbb{R}^d} g(x)\, dx = 1$, e in tal caso

$$p_X(x) = g(x) \textit{ per quasi ogni } x,$$

dove la dicitura "per quasi ogni x" significa per tutti gli $x \in \mathbb{R}^d \setminus \mathcal{N}$, dove $\mathcal{N}$ è un insieme di $\mathbb{R}^d$ al più numerabile.

[5] Ricordiamo che, in generale, d v.a. $X_1, \dots, X_d$ si dicono indipendenti se $\mathbb{P}(X_1 \in A_1, \dots, X_d \in A_d) = \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_d \in A_d)$, per ogni possibile scelta di $A_1, \dots, A_d$ $(\in \mathscr{B}(\mathbb{R}))$.

2.3.3 Speranza matematica

In questo paragrafo riassumiamo la definizione e le proprietà principali della speranza matematica. Prendiamo quindi una v.a. X, che per il momento supponiamo essere 1-dimensionale (X a valori in $\mathbb{R}$).
Ricordiamo quanto sviluppato nei precedenti corsi di probabilità.

- Sia X discreta, a valori in E_X, e sia p_X la densità discreta. Si dice che X ha *speranza matematica finita* se

  $$\sum_{x\in E_X} |x|\, p_X(x) < \infty$$

  e in tal caso, la *speranza matematica, o media o aspettazione* di X è

  $$\mathbb{E}(X) = \sum_{x\in E_X} x\, p_X(x).$$

- Sia X assolutamente continua e sia p_X la densità di probabilità (continua). Si dice che X ha *speranza matematica finita* se

  $$\int_{\mathbb{R}} |x|\, p_X(x)\, dx < \infty$$

  e in tal caso, la *speranza matematica, o media o aspettazione* di X è

  $$\mathbb{E}(X) = \int_{\mathbb{R}} x\, p_X(x)\, dx.$$

Osservazione 2.3.6. Supponiamo che X sia una v.a. discreta definita su uno spazio di probabilità discreto $(\Omega, \mathscr{F}, \mathbb{P})$, cioè Ω è finito o numerabile e $\mathscr{F} = \mathscr{P}(\Omega)$. In tal caso, per ogni $x \in E_X$, possiamo scrivere

$$p_X(x) = \mathbb{P}(X^{-1}(\{x\})) = \sum_{\omega\in X^{-1}(\{x\})} \mathbb{P}(\{\omega\}).$$

Allora,

$$\begin{aligned}
\textstyle\sum_{x\in E_X} |x|\, p_X(x) &= \textstyle\sum_{x\in E_X} |x| \sum_{\omega\in X^{-1}(\{x\})} \mathbb{P}(\{\omega\}) \\
&= \textstyle\sum_{x\in E_X} \sum_{\omega\in X^{-1}(\{x\})} |X(\omega)|\, \mathbb{P}(\{\omega\}) \\
&= \textstyle\sum_{\omega\in \cup_{x\in E_X} X^{-1}(\{x\})} |X(\omega)|\, \mathbb{P}(\{\omega\}) \\
&= \textstyle\sum_{\omega\in\Omega} |X(\omega)|\, \mathbb{P}(\{\omega\})
\end{aligned}$$

e, analogamente,

$$\sum_{x\in E_X} x\, p_X(x) = \sum_{\omega\in\Omega} X(\omega)\, \mathbb{P}(\{\omega\})$$

Allora, X ha speranza finita se e solo se $\sum_{\omega\in\Omega} |X(\omega)|\, \mathbb{P}(\{\omega\}) < \infty$ e in tal caso la speranza matematica è data da

$$\mathbb{E}(X) = \sum_{\omega\in\Omega} X(\omega)\, \mathbb{P}(\{\omega\}). \tag{2.2}$$

La quantità $\sum_{\omega\in\Omega} X(\omega)\, \mathbb{P}(\{\omega\})$ viene anche denotata con $\int_\Omega X(\omega)\, d\mathbb{P}$, così che

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

In realtà, *la quantità sopra scritta dà la definizione di speranza matematica della v.a.* X, a prescindere che X sia discreta o assolutamente continua. È evidente che, perché abbia senso, occorre definire l'integrale rispetto a $\mathbb{P}$. Ciò può essere fatto: poiché $\mathbb{P}$ è una misura in generale astratta, la teoria della misura consente di sviluppare l'integrale rispetto ad una misura qualsiasi, e la speranza matematica è semplicemente l'integrale rispetto a $\mathbb{P}$ della v.a. X. Questi argomenti saranno sviluppati nei corsi superiori di probabilità. Ciò nonostante, per le medie di v.a. discrete o assolutamente continue si ritroveranno le formule qui proposte, che rappresentano delle formule davvero operative: ricordiamo infatti che, nella pratica, spesso non si conoscono Ω, $\mathscr{F}$ e $\mathbb{P}$ ma si conosce la densità (discreta o continua) di X.

Riassumiamo qui di seguito alcune delle proprietà principali (di cui omettiamo le dimostrazioni rimandando ad un qualsiasi testo base in probabilità).

1. Sia X una v.a. discreta (risp. assolutamente continua), con densità discreta (risp. densità continua) p_X. Sia Y una v.a. su $\mathbb{R}$ della forma $Y = f(X)$. Allora Y ha speranza matematica finita se e solo se

$$\textstyle\sum_{x\in E_X} |f(x)|\, p_X(x)\, dx < \infty \qquad (\text{risp. } \int_\mathbb{R} |f(x)|\, p_X(x)\, dx < \infty)$$

e in tal caso, la speranza matematica di Y è

$$\textstyle\sum_{x\in E_X} f(x)\, p_X(x)\, dx < \infty \qquad (\text{risp. } \int_\mathbb{R} f(x)\, p_X(x)\, dx < \infty).$$

2. Nel seguito, X e Y denotano due v.a. con speranza finita. Si ha:

- (linearità) se $\alpha, \beta \in \mathbb{R}$ allora $\alpha X + \beta Y$ ha speranza finita e $\mathbb{E}(\alpha X + \beta Y) = \alpha\mathbb{E}(X) + \beta\mathbb{E}(Y)$;
- (positività) se $\mathbb{P}(X \geq 0) = 1$ allora $\mathbb{E}(X) \geq 0$; quindi, se $\mathbb{P}(X \geq Y) = 1$ allora $\mathbb{E}(X) \geq \mathbb{E}(Y)$;
- si ha sempre: $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$;
- se X e Y sono indipendenti allora $\mathbb{E}(X\,Y) = \mathbb{E}(X)\,\mathbb{E}(Y)$; più in generale, X e Y sono indipendenti se e solo se prse due funzioni f e g tali che esistono le speranze matematiche di $f(X)$ e $g(Y)$ allora esiste la speranza di $f(X)\, g(Y)$ ed inoltre $\mathbb{E}(f(X)\, g(Y)) = \mathbb{E}(f(X))\, \mathbb{E}(g(Y))$.

In particolare, si dice che X ha *momento k-esimo finito* se esiste $\mathbb{E}(|X|^k)$ ed in tal caso $\mathbb{E}(X^k)$ prende il nome di momento k-esimo. È facile vedere che se esiste il momento k-esimo allora esiste il momento j-esimo, per ogni $j \leq k$.
Particolarmente rilevante è il caso $k = 2$, in cui si può definire la *varianza*:

$$\mathrm{Var}(X) = \mathbb{E}\Big((X - \mathbb{E}(X))^2 \Big) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

Ricordiamo che la media dà la migliore approssimazione costante della v.a. nel senso del momento secondo, e la varianza dà l'errore che si commette sostituendo la v.a. con questa costante. In termini matematici:

$$\inf_{c \in \mathbb{R}} \mathbb{E}\Big((X - c)^2 \Big) = \min_{c \in \mathbb{R}} \mathbb{E}\Big((X - c)^2 \Big) = \mathbb{E}\Big((X - \mathbb{E}(X))^2 \Big) = \mathrm{Var}(X).$$

Ricordiamo che valgono anche le seguenti disuguaglianze:

- (Markov) Se $p > 0$ e $|X|^p$ ha speranza finita allora per ogni $c > 0$ si ha
$$\mathbb{P}(X > c) \leq \frac{\mathbb{E}(|X|^p)}{c^p}.$$
- (Chebycev) Se X ha momento secondo finito allora per ogni $c > 0$ si ha
$$\mathbb{P}(|X - \mathbb{E}(X)| > c) \leq \frac{\mathrm{Var}(X)}{c^2}.$$
- (Jensen) Se φ è una funzione convessa allora $\mathbb{E}(\varphi(X)) \geq \varphi(\mathbb{E}(X))$.
- (Schwarz) Se esistono le speranze di X^2 e Y^2 allora XY ha speranza finita e $|\mathbb{E}(X\,Y)| \leq \mathbb{E}(|X\,Y|) \leq \mathbb{E}(X^2)\,\mathbb{E}(Y^2)$.

Infine, se supponiamo che X sia una v.a. d-dimensionale, cioè $X = (X_1, \ldots, X_d)$, discreta oppure assolutamente continua, definiamo la speranza matematica di X come il vettore

$$\mathbb{E}(X) = (\mathbb{E}(X_1), \ldots, \mathbb{E}(X_d)).$$

Dunque, X ha speranza matematica finita se e solo se ciascuna componente X_i ha speranza matematica finita e la speranza di X è il vettore composto dalle speranze matematiche delle componenti. Ovviamente, tutte le proprietà viste continuano a valere (dando un senso opportuno alle quantità prima scalari e che ora diventano vettoriali). Ricordiamo, perché sarà spesso utilizzata in seguito, solo la seguente proprietà: presa $X = (X_1, \ldots, X_d)$ di densità discreta (risp. continua) p_X, allora $Y = f(X_1, \ldots, X_d)$ ha speranza finita se e solo se

$$\sum_{x=(x_1,\ldots,x_d) \in E_X} |f(x)|\, p_X(x) < \infty \quad \Big[\text{risp. } \int_{\mathbb{R}^d} |f(x)|\, p_X(x)\, dx < \infty\Big]$$

e in tal caso

$$\mathbb{E}(f(X)) = \sum_{x=(x_1,\dots,x_d)\in E_X} f(x)\, p_X(x) \quad \left[\text{risp. } \mathbb{E}(f(X)) = \int_{\mathbb{R}^d} f(x)\, p_X(x)\, dx\right].$$

Bibliografia

[1] P. Baldi (2000) *Equazioni differenziali stocastiche e applicazioni.* Pitagora Editrice.

[2] P. Billingsley (1989) *Probability and measure.* Whiley.

[3] D. Williams (1991) *Probability with martingales.* Cambridge University Press.