

1.7 Densità e misure assolutamente continue

Possiamo ora dare la definizione generale di densità.

Prendiamo uno spazio misurato $(E, \mathcal{E}, \mu)$, dove μ è una misura (non necessariamente di probabilità). Se $f : E \rightarrow \mathbb{R}^+$ è una funzione misurabile tale che $\int_E f d\mu = 1$, possiamo definire una misura di probabilità $\mathbb{P}$ su $(E, \mathcal{E})$ in questo modo:

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

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

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

Questo concetto è molto usato nel calcolo delle probabilità.

Esempio 1.46 Una prima applicazione si ha quando E è un aperto di $\mathbb{R}^d$, $\mathcal{E} = \mathcal{B}(E)$ e μ è la misura di Lebesgue d -dimensionale. Allora $\mathbb{P}$ si dice probabilità (o legge) assolutamente continua (e si sottintende quasi sempre “rispetto alla misura di Lebesgue”) di densità f .

Esempio 1.47 Un'altra applicazione, a prima vista meno evidente, si ha con le leggi discrete. Possiamo definire infatti una probabilità discreta definendo sull'insieme E la cosiddetta **misura che conta i punti**, definita da

$$\mu(A) = \text{Card}(A) \quad \forall A \in \mathcal{E}$$

Ovviamente questa misura non è una misura di probabilità. Se $p : E \rightarrow [0, 1]$ è tale che $\sum_{x \in E} p(x) = 1$, possiamo definire

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

Per questo motivo, p viene detta **densità discreta** di $\mathbb{P}$.

Esempio 1.48 Sia $(\Omega, \mathcal{A}, \mathbb{P})$ uno spazio probabilizzato e H un evento non trascurabile di $\mathcal{A}$. Si nota facilmente che la probabilità condizionata $\mathbb{P}_H = \mathbb{P}(\cdot|H)$ può essere ottenuta da $\mathbb{P}$ tramite la densità $\mathbf{1}_H/\mathbb{P}(H)$: difatti per ogni $A \in \mathcal{A}$ si ha

$$\int_A \frac{\mathbf{1}_H}{\mathbb{P}(H)} d\mathbb{P} = \int_{A \cap H} \frac{1}{\mathbb{P}(H)} d\mathbb{P} = \frac{\mathbb{P}(A \cap H)}{\mathbb{P}(H)} = \mathbb{P}_H(A)$$

La conseguenza è quindi che $\mathbb{P}_H(\cdot) = \mathbb{P}(\cdot|H)$ è una nuova misura di probabilità su $(\Omega, \mathcal{A})$.

Diamo ora la seguente definizione.

Definizione 1.49 Date due misure μ, ν su $(E, \mathcal{E})$, diciamo che ν è **assolutamente continua rispetto a μ** (e indichiamo questo con $\nu \ll \mu$) se $\mu(A) = 0$ implica $\nu(A) = 0$. Diciamo poi che ν è **equivalente a μ** (e lo indichiamo con $\nu \sim \mu$) se $\nu \ll \mu$ e $\mu \ll \nu$.

Un caso che abbiamo appena visto è quello di misure definite da una densità. Infatti, se ν ha densità f rispetto a μ , allora $\nu \ll \mu$. Per provarlo, se $\mu(A) = 0$ si ha che

$$\nu(A) = \int_A f \, d\mu = 0$$

Questa situazione sembra solo un esempio, ma è in realtà molto significativo. Infatti l'esistenza di una densità non è solo una condizione sufficiente, ma anche necessaria per l'assoluta continuità, come mostra il seguente teorema, che citiamo omettendo la dimostrazione.

Teorema 1.50 (Radon-Nykodim) *Se μ e ν sono misure σ -finite su $(E, \mathcal{E})$ e $\nu \ll \mu$, allora esiste un'unica $f \in L^+(E, \mathcal{E}, \mu)$ tale che $\nu = f \cdot \mu$, o anche $f = \frac{d\nu}{d\mu}$.*

La funzione f si può chiamare, oltre che densità, anche **derivata di Radon-Nykodim** di ν rispetto a μ .

Una domanda naturale a questo punto è la seguente: che relazione hanno gli integrali calcolati rispetto a una misura ν , che abbia densità f rispetto a un'altra misura μ , rispetto agli integrali calcolati rispetto a μ ? La risposta è fornita nella seguente

Proposizione 1.51 *Se g è una funzione misurabile reale su $(E, \mathcal{E})$, e μ, ν sono due misure su $(E, \mathcal{E})$ tali che $\nu = f \cdot \mu$, allora*

$$\int_E g(x) \, \nu(dx) = \int_E g(x)f(x) \, \mu(dx)$$

valida ogni volta che uno dei membri abbia senso.

Dimostrazione. Si procede come al solito, supponendo all'inizio che g sia semplice ed estendendo il risultato prima a funzioni semplici, poi a funzioni positive ed infine a funzioni integrabili. In particolare, supponiamo all'inizio che $g = \sum_{i=1}^n a_i \mathbf{1}_{A_i}$, con $a_i \in \mathbb{R}$, $A_i \in \mathcal{E}$: allora

$$\begin{aligned} \int_E g \, d\nu &= \int_E \sum_{i=1}^n a_i \mathbf{1}_{\{X \in A_i\}} \, d\nu = \sum_{i=1}^n a_i \nu(A_i) = \sum_{i=1}^n a_i \int_{A_i} f \, d\mu = \\ &= \int_E \sum_{i=1}^n a_i \mathbf{1}_{A_i} f \, d\mu = \int_E g f \, d\mu \end{aligned}$$

Se $g \in L^+$ ma non è necessariamente semplice, allora esistono $(g_n)_n \subset L^+$ e semplici tali che $g_n \nearrow g$. Allora, usando due volte il teorema di convergenza monotona,

$$\int_E g \, d\nu = \int_E \lim_{n \rightarrow \infty} g_n \, d\nu = \lim_{n \rightarrow \infty} \int_E g_n \, d\nu = \lim_{n \rightarrow \infty} \int_E g_n f \, d\mu = \int_E g f \, d\mu$$

Infine, se $g \in L^1$, allora scriviamo $g = g^+ - g^-$, e si ha

$$\int_E g \, d\nu = \int_E g^+ \, d\nu - \int_E g^- \, d\nu = \int_E g^+ f \, d\mu - \int_E g^- f \, d\mu = \int_E g f \, d\mu$$

□

La prima applicazione di questo risultato è il calcolo della speranza di una variabile aleatoria X rispetto ad una probabilità condizionata $\mathbb{P}_H$, con $H \in \mathcal{A}$ non trascurabile: abbiamo difatti visto che $\mathbb{P}_H$ ha densità rispetto a $\mathbb{P}$ uguale a $\frac{\mathbf{1}_H}{\mathbb{P}(H)}$. Allora, se X ha speranza finita, abbiamo

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

e resta quindi dimostrata la formula (1.1).

1.8 Legge di una variabile aleatoria

Nel paragrafo precedente abbiamo visto che la speranza matematica di una variabile aleatoria non è altro che il suo integrale sullo spazio probabilizzato su cui è definita. Per calcolare concretamente questo valore, tuttavia, questa definizione potrebbe risultare scomoda (ad esempio perchè lo spazio Ω ha una struttura molto complessa), e potrebbero esserci dei modi più semplici. In questo paragrafo vedremo che il calcolo del valore atteso di una variabile aleatoria (o più in generale di una sua funzione misurabile) può essere ricondotto all'integrale su un opportuno spazio misurabile.

Il primo risultato è il seguente. Se X è una variabile aleatoria da $(\Omega, \mathcal{A}, \mathbb{P})$ a valori in $(E, \mathcal{E})$, e g è una funzione misurabile reale su $(E, \mathcal{E})$, possiamo chiederci quanto vale $E[g(X)]$ (se questa quantità è ben definita).

Proposizione 1.52 (formula della speranza di una funzione composta) *Se X è una variabile aleatoria su $(\Omega, \mathcal{A}, \mathbb{P})$ a valori in $(E, \mathcal{E})$ di legge $\mathbb{P}_X$ e g è una funzione misurabile reale su $(E, \mathcal{E})$, allora*

$$E[g(X)] = \int_E g(x) \, d\mathbb{P}_X(x)$$

valida ogni volta che uno dei membri abbia senso.

Dimostrazione. Supponiamo all'inizio che $g = \sum_{i=1}^n a_i \mathbf{1}_{A_i}$, con $a_i \in \mathbb{R}$, $A_i \in \mathcal{E}$: allora

$$E[g(X)] = E \left[\sum_{i=1}^n a_i \mathbf{1}_{\{X \in A_i\}} \right] = \sum_{i=1}^n a_i \mathbb{P}\{X \in A_i\} = \sum_{i=1}^n a_i \mathbb{P}_X(A_i) = \int_E g(x) \, d\mathbb{P}_X(x)$$

Se $g \in L^+(E, \mathcal{E}, \mathbb{P}_X)$ ma non è necessariamente semplice, allora esistono $(g_n)_n \subset L^+(E, \mathcal{E}, \mathbb{P}_X)$ e semplici tali che $g_n \nearrow g$. Allora si ha anche che $g_n(X) \nearrow g(X)$, e usando due volte il teorema di convergenza monotona (la prima su $(\Omega, \mathcal{A}, \mathbb{P})$ e la seconda su $(E, \mathcal{E}, \mathbb{P}_X)$) si ha

$$E[g(X)] = \lim_{n \rightarrow \infty} E[g_n(X)] = \lim_{n \rightarrow \infty} \int_E g_n(x) \, d\mathbb{P}_X(x) = \int_E g(x) \, d\mathbb{P}_X(x)$$

Infine, se $g \in L^1(E, \mathcal{E}, \mathbb{P}_X)$, allora scriviamo $g = g^+ - g^-$, e si ha anche che $g(X) = g^+(X) - g^-(X) \in L^1(\Omega, \mathcal{A}, \mathbb{P})$; allora

$$E[g(X)] = E[g^+(X)] - E[g^-(X)] = \int_E g^+(x) d\mathbb{P}_X(x) - \int_E g^-(x) d\mathbb{P}_X(x) = \int_E g(x) d\mathbb{P}_X(x)$$

□

Una prima applicazione della formula della speranza di una funzione composta nel caso di variabili aleatorie reali è con la funzione $g(x) = x$. Se X è una variabile aleatoria reale, allora la formula della speranza di una funzione composta ci dà il risultato

$$E[X] = \int_{\mathbb{R}} x d\mathbb{P}_X(x)$$

che permette di ricondurre il calcolo della speranza di una variabile aleatoria definita sullo spazio Ω (che a priori può essere molto complicato) al calcolo di un integrale sulla retta reale.

Due applicazioni immediate di questo risultato sono le seguenti.

Esempio 1.53 (vettori aleatori continui) Un vettore aleatorio n -dimensionale X si dice **continuo** se la sua legge $\mathbb{P}_X$ ha densità f rispetto alla misura di Lebesgue λ_n . Allora, per quanto appena visto, per ogni funzione $g : E \rightarrow \mathbb{R}$ tale che una delle due espressioni seguenti abbia senso, vale

$$E[g(X)] = \int_{\mathbb{R}^n} g(x) d\mathbb{P}_X(x) = \int_{\mathbb{R}^n} g(x)f(x) d\lambda_n(x)$$

Esempio 1.54 (vettori aleatori discreti) Se invece X è una variabile aleatoria discreta a valori in un insieme E , con densità discreta $p_X : E \rightarrow [0, 1]$, allora abbiamo visto che possiamo considerare p_X la densità di $\mathbb{P}_X$ rispetto alla misura μ che conta i punti. Allora, per ogni funzione $g : E \rightarrow \mathbb{R}$ tale che una delle espressioni seguenti abbia senso, vale

$$E[g(X)] = \int_E g(x)p_X(x) d\mu(x) = \sum_{x \in E} g(x)p_X(x)$$

Considerando che la legge di X si può rappresentare come $\sum_{x \in E} p_X(x)\delta_x$, si può scrivere la formula precedente come

$$E[g(X)] = \sum_{x \in E} p_X(x) \int_E g(y) d\delta_x(y)$$

Questa è una proprietà generale: se infatti la legge di X è una combinazione lineare (anche infinita) di leggi, allora anche la speranza di funzioni di X si calcola allo stesso modo.

Teorema 1.55 Se $\mathbb{P}_X = \sum_{n=1}^{\infty} p_n P_n$, con $(p_n)_n$ numeri reali non negativi e di somma unitaria e P_n misure di probabilità su $(E, \mathcal{E})$, allora

$$E[g(X)] = \sum_{n=1}^{\infty} p_n \int_E g(y) dP_n(y)$$

Dimostrazione. La dimostrazione è simile a quella della formula della speranza della funzione composta: prima si considerano funzioni della forma $g(x) = \mathbf{1}_A$, con $A \in \mathcal{E}$. Per linearità, il teorema è vero per ogni g semplice; per monotonia si estende ad ogni $g \in L^+$, e considerando la decomposizione in parte positiva e parte negativa la relazione è valida per ogni g tale che $g(X)$ è integrabile. $\square$

1.9 Misure di Borel sulla retta reale

In questa sezione costruiremo misure su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Queste misure si chiamano **misure di Borel** su $\mathbb{R}$. Da qui in poi considereremo intervalli aperti a sinistra e chiusi a destra su $\mathbb{R}$, cioè della forma $(a, b]$, oppure $(a, +\infty)$, oppure $\emptyset$, con $-\infty \leq a \leq b < \infty$; chiamiamo questi intervalli **h-intervalli** (h sta per half-open).

Prendiamo ora una funzione $F : \mathbb{R} \rightarrow \mathbb{R}$ crescente (nel senso di non decrescente). Diciamo che F è **continua a destra** se per ogni $a \in \mathbb{R}$ si ha

$$F(a) = \lim_{x \rightarrow a^+} F(x)$$

Inoltre esistono i limiti $F(+\infty) = \lim_{x \rightarrow +\infty} F(x)$ e $F(-\infty) = \lim_{x \rightarrow -\infty} F(x)$ (e possono essere finiti o infiniti). Vogliamo ora definire una misura μ a partire da F in modo che

$$\mu((a, b]) = F(b) - F(a)$$

Questo è possibile in virtù del seguente risultato.

Teorema 1.56 *Se F è crescente e c.a.d., allora esiste un'unica misura μ_F su $\mathcal{B}(\mathbb{R})$ tale che $\mu_F((a, b]) = F(b) - F(a)$ per ogni $a, b \in \mathbb{R}$, $a < b$. Se G è un'altra funzione crescente e c.a.d., allora $\mu_F = \mu_G$ se e solo se $G - F$ è costante. Viceversa, se μ è una misura su $\mathcal{B}(\mathbb{R})$ finita su ogni boreliano limitato e definiamo*

$$F(x) = \begin{cases} \mu((0, x]) & \text{se } x > 0 \\ 0 & \text{se } x = 0 \\ -\mu((x, 0]) & \text{se } x < 0 \end{cases}$$

allora F è c.a.d. e crescente, e $\mu = \mu_F$. Infine, μ_F si può estendere a una misura completa $\bar{\mu}_F$ che è definita su una σ -algebra che in generale contiene $\mathcal{B}(\mathbb{R})$.

Queste misure vengono comunemente chiamate **misure di Lebesgue-Stieltjes**. Facciamo ora alcune note.

1. Questo risultato si sarebbe potuto ottenere anche considerando intervalli del tipo $[a, b)$ e funzioni F continue a sinistra.
2. Se μ è una misura di probabilità su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, e poniamo $F(x) = \mu((-\infty, x])$, allora $\mu = \mu_F$. La F si chiama allora **funzione di ripartizione cumulativa** di μ .

Esempio 1.57 (misura di Lebesgue) Se $F(x) = x$, allora otteniamo che $\mu((a, b]) = b - a$. Questa misura coincide con l'usuale misura di Riemann sulle unioni finite di intervalli (ma ricordiamo che ha come dominio $\mathcal{B}(\mathbb{R})$) e si chiama **misura di Lebesgue**.

Esempio 1.58 (misure assolutamente continue rispetto a Lebesgue) Consideriamo $f : \mathbb{R} \rightarrow \mathbb{R}^+$ misurabile e tale che $\int_{\mathbb{R}} f(x) dx = 1$, e definiamo

$$F(x) := \int_{-\infty}^x f(t) dt$$

Questa funzione di ripartizione è una funzione assolutamente continua. Allora F definisce una misura di probabilità μ_F su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ tale che

$$\mu_F([a, b)) = \mu_F((a, b)) = \mu_F((a, b]) = \mu_F([a, b]) := \int_a^b f(t) dt$$

μ_F ha quindi f come densità rispetto alla misura di Lebesgue. Le misure di questo tipo si dicono **misure assolutamente continue (rispetto alla misura di Lebesgue)**.

Esempio 1.59 (misure discrete su $\mathbb{R}$) Consideriamo $p : \mathbb{R} \rightarrow [0, 1]$ tale che $\sum_{x \in \mathbb{R}} p(x) = 1$, e definiamo

$$F(x) := \sum_{t \leq x} p(t)$$

Questa funzione di ripartizione è costante a tratti. Allora F definisce una misura di probabilità μ_F su $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ (e quindi in particolare su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$) tale che

$$\mu_F((a, b]) = \sum_{a < t \leq b} p(t)$$

Facciamo notare che il dominio di μ_F in generale contiene $\mathcal{B}(\mathbb{R})$. Per vedere questo in un caso semplice, basta considerare la funzione $F = \mathbf{1}_{\{x \geq 0\}}$. Allora $\mu_F = \delta_0$ è la delta di Dirac nel punto 0 e il completamento è definito su tutta la σ -algebra $\mathcal{P}(\mathbb{R})$. Infatti abbiamo che $\delta_0(\mathbb{R} \setminus \{0\}) = 0$, quindi è possibile estendere δ_0 ad ogni sottoinsieme di $\mathbb{R} \setminus \{0\}$ assegnandogli misura nulla. Un altro esempio è dato dalle misure discrete viste sopra. Questi fenomeni possono sembrare patologie tipiche delle misure discrete, ma in realtà riguardano tutte le misure usate comunemente, e in particolare la misura di Lebesgue m (e quindi tutte le misure assolutamente continue!). Infatti anche essa può essere estesa ad una σ -algebra più grande di $\mathcal{B}(\mathbb{R})$, che chiameremo $\mathcal{L}$.

1.10 Funzione di ripartizione, pseudoinversa e quantili

Analizziamo ora il caso particolare in cui X è una variabile aleatoria reale. In questo caso $\mathbb{P}_X$ è una misura di probabilità su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, e abbiamo visto che è caratterizzata da

$$F_X(t) := \mathbb{P}_X((-\infty, t]) = \mathbb{P}\{X \leq t\}$$

che prende il nome di **funzione cumulativa di distribuzione** (o **di ripartizione**) di X . Abbiamo visto che la corrispondenza tra funzioni crescenti e continue a destra e misure è biunivoca. Si può inoltre dimostrare il seguente risultato.

Teorema 1.60 *Se una funzione F è crescente e continua a destra e tale che $\lim_{t \rightarrow -\infty} F(t) = 0$, $\lim_{t \rightarrow +\infty} F(t) = 1$, allora esiste una variabile aleatoria reale X , definita sullo spazio $(\Omega, \mathcal{A}, \mathbb{P}) = ((0, 1), \mathcal{B}((0, 1)), \lambda)$ (dove λ è la misura di Lebesgue su $(0, 1)$), tale che la sua funzione di ripartizione è F .*

Dimostrazione. Definiamo, per ogni $\omega \in \Omega$, la funzione

$$X(\omega) := \min\{x \in \mathbb{R} \mid F(x) \geq \omega\} \quad (1.3)$$

Allora abbiamo che

$$\{X > t\} = \{\omega \mid X(\omega) > t\} = \{\omega \mid F(t) < \omega\} = (F(t), 1) \in \mathcal{B}((0, 1)) = \mathcal{A}$$

e quindi X è una variabile aleatoria. Inoltre, se definiamo $F_X(t) := \mathbb{P}\{X \leq t\}$, allora per ogni $t \in \mathbb{R}$

$$F_X(t) = \mathbb{P}\{X \leq t\} = 1 - \mathbb{P}\{X > t\} = 1 - \mathbb{P}(F(t), 1) = 1 - 1 + F(t) = F(t)$$

e quindi abbiamo la tesi. $\square$

La variabile aleatoria $X : (0, 1) \rightarrow \mathbb{R}$ definita nell'Equazione (1.3) prende anche il nome di **pseudoinversa** della funzione di ripartizione F , e si indica con F^{-1} . Se F è invertibile, allora F^{-1} è in effetti l'inversa di F . In generale, per ogni $u > 0$ e ogni coppia di numeri reali $a < b$, abbiamo poi l'equivalenza

$$F(a) < u \leq F(b) \iff a < F^{-1}(u) \leq b$$

(questo si può verificare sia tramite la definizione, sia grazie al precedente teorema). Grazie al teorema, si ha che la legge di F^{-1} , che non è altro che μ_F , è l'immagine della misura di Lebesgue tramite F^{-1} .

Grazie al teorema, si ha che la legge di F^{-1} , che non è altro che μ_F , è l'immagine della misura di Lebesgue tramite F^{-1} . Per la formula della speranza di una funzione composta, questo significa che per ogni funzione reale g integrabile rispetto a μ_F , si ha

$$\int_{\mathbb{R}} g \, d\mu_F = E[g(X)] = E[g(F^{-1})] = \int_0^1 g(F^{-1}(u)) \, du$$

quindi, ogni integrale rispetto ad una generica misura μ sulla retta può essere ricondotto ad un integrale rispetto alla misura di Lebesgue.

Notiamo poi che la funzione F^{-1} risolve il seguente problema: data una variabile aleatoria X con funzione di ripartizione F e una soglia di probabilità $\alpha \in (0, 1)$, qual è il minimo numero reale t tale che $\mathbb{P}\{X \leq t\} = F(t) \geq \alpha$? Tale numero t si chiama **quantile** di livello α di X , e si indica genericamente con q_α . Per come abbiamo costruito F^{-1} , abbiamo semplicemente che $q_\alpha = F^{-1}(\alpha)$.

1.11 Indipendenza

Innanzitutto richiamiamo la nozione di indipendenza tra due eventi.

Definizione 1.61 Sia $(\Omega, \mathcal{A}, \mathbb{P})$ uno spazio probabilizzato e $A, B \in \mathcal{A}$. Diciamo che A e B sono **indipendenti** se $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

Ovviamente, se almeno uno dei due eventi è trascurabile, la relazione si riduce a $0 = 0$.

Vediamo ora come la nozione di indipendenza si può trasferire a variabili aleatorie e a σ -algebre.

Definizione 1.62 Sia $(\Omega, \mathcal{A}, \mathbb{P})$ uno spazio probabilizzato e $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (E, \mathcal{E})$ e $Y : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (F, \mathcal{F})$ due variabili aleatorie. Diciamo che X e Y sono **indipendenti** se, per ogni $A \in \mathcal{E}$ e $B \in \mathcal{F}$, gli eventi $\{X \in A\}$ e $\{Y \in B\}$ sono indipendenti, cioè se si ha:

$$\mathbb{P}\{X \in A\}\mathbb{P}\{Y \in B\} = \mathbb{P}\{X \in A, Y \in B\} \quad (1.4)$$

Nota 1.63 Due eventi $A, B \in \mathcal{A}$ sono indipendenti se e solo se $\mathbf{1}_A$ e $\mathbf{1}_B$ lo sono.

Dalla definizione, per verificare che X e Y sono indipendenti bisognerebbe effettuare il controllo su ogni possibile coppia di eventi delle loro σ -algebre di arrivo, cosa che a volte può risultare scomoda. Vediamo ora un risultato che permette di limitare questo controllo ad un numero minore di casi.

Teorema 1.64 (criterio di indipendenza fondato su due basi) Siano $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (E, \mathcal{E})$ e $Y : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (F, \mathcal{F})$ due variabili aleatorie, e siano $\mathcal{H}$ una base di $\mathcal{E}$ e $\mathcal{K}$ una base di $\mathcal{F}$. Allora X e Y sono indipendenti se e solo se l'Equazione (1.4) vale per ogni $A \in \mathcal{H}, B \in \mathcal{K}$.

Dimostrazione. L'implicazione “solo se” è banale. Per dimostrare l'implicazione “se”, procederemo in due passi: nel primo dimostreremo che, se (1.4) vale per ogni $A \in \mathcal{H}, B \in \mathcal{K}$, allora vale per ogni $A \in \mathcal{H}, B \in \mathcal{F}$. Fissiamo infatti $A \in \mathcal{H}$. Se A è trascurabile, allora (1.4) è banalmente verificata, altrimenti (1.4) è equivalente a

$$\mathbb{P}\{Y \in B\} = \frac{\mathbb{P}\{X \in A, Y \in B\}}{\mathbb{P}\{X \in A\}} = \mathbb{P}(\{Y \in B\} | \{X \in A\}) \quad \forall B \in \mathcal{K}$$

In questa uguaglianza compaiono due misure di probabilità su $(F, \mathcal{F})$: una è la legge di Y , e l'altra è la legge di Y condizionata all'evento $\{X \in A\}$. Siccome $\mathcal{K}$ è una base di $\mathcal{F}$, da questo si deduce che le due leggi coincidono su $\mathcal{F}$. Abbiamo quindi ottenuto che (1.4) vale per ogni $A \in \mathcal{H}, B \in \mathcal{F}$. Con un ragionamento analogo si riesce a dimostrare che (1.4) vale per ogni $A \in \mathcal{E}, B \in \mathcal{F}$. $\square$

Presentiamo ora la definizione di indipendenza di σ -algebre.

Definizione 1.65 Due σ -algebre $\mathcal{A}$ e $\mathcal{B}$ su Ω si dicono **indipendenti** se e solo se per ogni $A \in \mathcal{A}$, $B \in \mathcal{B}$, gli eventi A e B sono indipendenti.

Ovviamente il criterio di indipendenza relativo a due basi si può estendere a questo caso: se due σ -algebre sono generate da due basi, per verificare la loro indipendenza occorre e basta verificare l'indipendenza su ogni coppia di eventi delle due basi.

Notiamo infine che due variabili aleatorie X e Y sono indipendenti se e solo se le loro σ -algebre generate $\sigma(X)$ e $\sigma(Y)$ sono indipendenti.

A causa di tutte queste equivalenze, non è difficile interpretare espressioni come indipendenza di variabili aleatorie e/o di eventi da eventi e/o da σ -algebre.

Vediamo ora come il concetto di indipendenza si possa estendere ad un insieme di più di due oggetti. Presentiamo la definizione solo per le variabili aleatorie poichè, a causa dell'equivalenza appena vista, la cosa si può facilmente estendere al caso di σ -algebre e di eventi.

Definizione 1.66 Sia $(X_i)_{i \in I}$ una famiglia di variabili aleatorie definite sullo stesso spazio probabilizzato (a valori in spazi misurabili eventualmente diversi). Diciamo che le X_i sono **indipendenti** (o che $(X_i)_{i \in I}$ è una **famiglia di variabili aleatorie indipendenti**) se, comunque si prendano $H, K \subset I$ disgiunti e non vuoti, si abbia che $\sigma(X_i | i \in H)$ è indipendente da $\sigma(X_i | i \in K)$.

Possiamo vedere una famiglia di variabili aleatorie $(X_i)_{i \in I}$ come un'unica variabile aleatoria X a valori nello spazio prodotto $(\prod_{i \in I} E_i, \otimes_{i \in I} \mathcal{E}_i)$, definita in questo modo:

$$X(\omega) := (X_i(\omega))_{i \in I}$$

Con una tale identificazione, la legge della variabile aleatoria X si dice **legge congiunta della famiglia** $(X_i)_{i \in I}$, mentre per ogni indice $i \in I$ la legge di X_i si dice **legge marginale i -esima** della famiglia $(X_i)_{i \in I}$. Inoltre, risulta che $(X_i)_{i \in I}$ è una famiglia di variabili aleatorie indipendenti se e solo se, comunque si prendano $H, K \subset I$ disgiunti e non vuoti, si abbia che $X_H := (X_h)_{h \in H}$ è indipendente da $X_K := (X_k)_{k \in K}$.

Teorema 1.67 (criterio fondamentale d'indipendenza) Sia $(X_i)_{i \in I}$ una famiglia di variabili aleatorie definite sullo stesso spazio probabilizzato a valori rispettivamente negli spazi $(E_i, \mathcal{E}_i)$. Le due condizioni sono equivalenti:

1. le X_i sono indipendenti;
2. la legge congiunta di $X = (X_i)_{i \in I}$ è uguale alla legge prodotto delle leggi marginali, cioè si ha che

$$\mathbb{P} \left\{ X \in \prod_{i \in I} A_i \right\} = \prod_{i \in I} \mathbb{P} \{ X_i \in A_i \}$$

per ogni famiglia $(A_i)_{i \in I}$ con $A_i \in \mathcal{E}_i$ e $A_i \neq E_i$ per $i \in J$ con $J \subseteq I$ finito.

Dimostrazione.

1 $\Rightarrow$ 2) Se $|J| < 2$ non c'è niente da dimostrare. Se $|J| \geq 2$, allora possiamo dividere $J = H \cup K$, con H e K disgiunti e non vuoti: per definizione di indipendenza si ha allora

$$\begin{aligned}\mathbb{P}\left\{X \in \prod_{i \in I} A_i\right\} &= \mathbb{P}\left(\bigcap_{j \in J} \{X_j \in A_j\}\right) = \mathbb{P}\left(\bigcap_{h \in H} \{X_h \in A_h\} \cap \bigcap_{k \in K} \{X_k \in A_k\}\right) = \\ &= \mathbb{P}\left(\bigcap_{h \in H} \{X_h \in A_h\}\right) \mathbb{P}\left(\bigcap_{k \in K} \{X_k \in A_k\}\right)\end{aligned}$$

Ora, H e K hanno cardinalità minore di J , e quindi la tesi si ottiene per induzione sulla cardinalità di J .

2 $\Rightarrow$ 1) Per ogni $H, K \subset I$ non vuoti e disgiunti, le variabili aleatorie $X_H := (X_h)_{h \in H}$ e $X_K := (X_k)_{k \in K}$ sono indipendenti poichè soddisfano il criterio 1.64, con $\mathcal{H}$ e $\mathcal{K}$ date rispettivamente da

$$\begin{aligned}\mathcal{H} &:= \left\{ \prod_{h \in H} A_h \mid A_h \in \mathcal{E}_h, A_h \neq E_h \text{ per un n. finito di } h \in H \right\}, \\ \mathcal{K} &:= \left\{ \prod_{k \in K} A_k \mid A_k \in \mathcal{E}_k, A_k \neq E_k \text{ per un n. finito di } k \in K \right\}.\end{aligned}$$

Difatti per ogni $\prod_{h \in H} A_h \in \mathcal{H}$, $\prod_{k \in K} A_k \in \mathcal{K}$ si ha

$$\begin{aligned}\mathbb{P}\left\{X_H \in \prod_{h \in H} A_h, X_K \in \prod_{k \in K} A_k\right\} &= \prod_{h \in H} \mathbb{P}\{X_h \in A_h\} \prod_{k \in K} \mathbb{P}\{X_k \in A_k\} = \\ &= \mathbb{P}\left\{X_H \in \prod_{h \in H} A_h\right\} \mathbb{P}\left\{X_K \in \prod_{k \in K} A_k\right\}\end{aligned}$$

□

Vediamo infine un utile criterio di calcolo relativo all'indipendenza: se $(X_i)_{i \in I}$ è una famiglia di variabili aleatorie indipendenti a valori rispettivamente in $(E_i, \mathcal{E}_i)$, e g_i sono funzioni misurabili definite su $(E_i, \mathcal{E}_i)$, allora anche le variabili aleatorie $g_i(X_i)$ sono indipendenti. Inoltre, se le g_i sono funzioni misurabili reali, per ogni $J \subseteq I$ finito si ha che

$$E\left[\prod_{i \in J} g_i(X_i)\right] = \prod_{i \in J} E[g_i(X_i)]$$

ogniquale volta uno dei due membri sia definito.

Questo risultato è parente stretto del seguente risultato di teoria della misura, che permette di calcolare gli integrali rispetto a una misura prodotto (come quello al primo membro) in termini di integrali delle singole componenti.

Teorema 1.68 (Fubini - Tonelli) Supponiamo che $(E, \mathcal{E}, \mu)$ e $(F, \mathcal{F}, \nu)$ siano spazi misurati, con μ e ν misure σ -finite. Allora:

Fubini: se $f \in L^1(E \times F, \mathcal{E} \otimes \mathcal{F}, \mu \otimes \nu)$, allora $f(x, \cdot) \in L^1(F, \mathcal{F}, \nu)$ per q.o. $x \in E$, $f(\cdot, y) \in L^1(E, \mathcal{E}, \mu)$ per q.o. $y \in F$, le funzioni (definite q.o.) $\int_F f(x, \cdot) d\nu$ e $\int_E f(\cdot, y) d\mu$ sono rispettivamente in $L^1(E, \mathcal{E}, \mu)$ e in $L^1(F, \mathcal{F}, \nu)$ e vale

$$\begin{aligned} \int_{E \times F} f(x, y) d(\mu \otimes \nu)(x, y) &= \int_E \left(\int_F f(x, y) d\nu(y) \right) d\mu(x) = \\ &= \int_F \left(\int_E f(x, y) d\mu(x) \right) d\nu(y) \end{aligned} \quad (1.5)$$

Tonelli: se $f \in L^+(E \times F, \mathcal{E} \otimes \mathcal{F}, \mu \otimes \nu)$, allora le funzioni $\int_F f(x, \cdot) d\nu$ e $\int_E f(\cdot, y) d\mu$ sono rispettivamente in $L^+(E, \mathcal{E}, \mu)$ e in $L^+(F, \mathcal{F}, \nu)$ e vale la (1.5).

Per induzione, è possibile calcolare l'integrale di una funzione su uno spazio prodotto n -dimensionale iterando gli integrali su ogni componente.

Il legame con la formula vista sopra è il seguente: se ad esempio le g_i sono limitate, allora (per facilità di notazione, supponiamo che $J = \{1, \dots, n\}$)

$$\begin{aligned} E \left[\prod_{i \in J} g_i(X_i) \right] &= \int_{E_1 \times \dots \times E_n} g_1(x_1) \cdots g_n(x_n) d\mathbb{P}_X = \\ &= \int_{E_1 \times \dots \times E_n} g_1(x_1) \cdots g_n(x_n) d(\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}) = \prod_{i=1}^n \int_{E_i} g_i d\mathbb{P}_{X_i} = \prod_{i=1}^n E[g_i(X_i)] \end{aligned}$$

Un'altra utile applicazione del teorema di Tonelli è la seguente. Supponiamo che le variabili aleatorie indipendenti $X_1, \dots, X_n$ assumano valori rispettivamente in $(E_i, \mathcal{E}_i)$, $i = 1, \dots, n$ con leggi μ_i aventi densità f_i rispetto a misure ν_i su $(E_i, \mathcal{E}_i)$. Allora la legge del vettore aleatorio $X = (X_1, \dots, X_n)$, che per quanto visto sopra è data dalla misura prodotto $\otimes_{i=1}^n \mu_i$, ha densità $\prod_{i=1}^n f_i$ rispetto alla misura $\otimes_{i=1}^n \nu_i$. Per verificarlo, basta considerare la base della σ -algebra prodotto costituita da $\mathcal{I} := \{\prod_{i=1}^n A_i | A_i \in \mathcal{E}_i\}$: per ogni $\prod_{i=1}^n A_i \in \mathcal{I}$ si ha

$$\mathbb{P} \left\{ X \in \prod_{i=1}^n A_i \right\} = \prod_{i=1}^n \mathbb{P}\{X_i \in A_i\} = \prod_{i=1}^n \mu_i(A_i) = \prod_{i=1}^n \int_{A_i} f_i d\nu_i$$

e, siccome $\prod_{i=1}^n f_i \in L^+(E_1 \times \dots \times E_n)$, per il teorema di Tonelli si ha

$$\int_{\prod_{i=1}^n A_i} \prod_{i=1}^n f_i d\nu_1 \otimes \dots \otimes \nu_n = \prod_{i=1}^n \int_{A_i} f_i d\nu_i$$

e quindi

$$\mu_1 \otimes \dots \otimes \mu_n = \left(\prod_{i=1}^n f_i \right) \cdot \nu_1 \otimes \dots \otimes \nu_n$$

1.12 Vettori aleatori

Si chiama **vettore aleatorio** a valori in $\mathbb{R}^d$ una variabile aleatoria $X = (X_1, \dots, X_d)$ a valori nello spazio misurabile $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Chiaramente questo significa che ogni componente X_i , $i = 1, \dots, d$ è una variabile aleatoria reale.

Dato un vettore aleatorio, la sua legge $\mathbb{P}_X$ su $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ si dice **legge congiunta** di X , mentre la legge $\mathbb{P}_{X_i}$ della componente X_i si dice i -esima **legge marginale** per ogni $i = 1, \dots, d$.

Se X è un vettore aleatorio formato da componenti indipendenti, ovviamente la legge congiunta è il prodotto delle leggi marginali. Più in generale, il legame tra legge congiunta e leggi marginali è il seguente: per ogni $i = 1, \dots, d$, $A_i \in \mathcal{B}(\mathbb{R})$, abbiamo che

$$\mathbb{P}_{X_i}(A_i) = \mathbb{P}_X \left(\prod_{j=1}^d B_j \right), \quad \text{dove } B_j := \begin{cases} A_i & \text{se } j = i, \\ \mathbb{R} & \text{se } j \neq i. \end{cases}$$

Difatti, con le notazioni appena viste,

$$\mathbb{P}_{X_i}(A_i) = \mathbb{P}\{X_i \in A_i\} = \mathbb{P} \left\{ X \in \prod_{j=1}^d B_j \right\}$$

Esempio 1.69 (vettori aleatori discreti) Supponiamo che X sia un vettore aleatorio discreto, con densità discreta p_X . Allora la formula sopra permette di recuperare la nota formula per le densità discrete marginali

$$p_{X_i}(x_i) = \sum_{x_1, \dots, \hat{x}_i, \dots, x_d} p_X(x)$$

Se difatti per ogni $x_i \in \mathbb{R}$ tale che $\mathbb{P}\{X_i = x_i\} > 0$ poniamo $A_i := \{x_i\}$, con le notazioni sopra abbiamo che

$$p_{X_i}(x_i) = \mathbb{P}\{X_i \in A_i\} = \mathbb{P}_X \left(\prod_{j=1}^d B_j \right) = \sum_{x \in \prod_{j=1}^d B_j} p_X(x) = \sum_{x_1, \dots, \hat{x}_i, \dots, x_d} p_X(x)$$

Esempio 1.70 (vettori aleatori continui) Supponiamo che X sia un vettore aleatorio continuo, con densità f_X rispetto alla misura di Lebesgue d -dimensionale λ_d . Allora la formula sopra permette di recuperare una analoga della formula per le densità discrete marginali: difatti, per ogni $A \in \mathcal{B}(\mathbb{R})$ e per $i = 1, \dots, d$, applicando il teorema di Tonelli si ha

$$\begin{aligned} \mathbb{P}\{X_i \in A\} &= \int_{\mathbb{R} \times \dots \times A \times \dots \times \mathbb{R}} f_X \, d\lambda_d = \\ &= \int_A \left(\int_{\mathbb{R}^{d-1}} f_X(x_1, \dots, x_d) \, d\lambda_{d-1}(x_1, \dots, \hat{x}_i, \dots, x_d) \right) d\lambda_1(x_i) \end{aligned}$$

e quindi X_i ammette come densità rispetto alla misura di Lebesgue la funzione

$$f_{X_i}(t) := \int_{\mathbb{R}^{d-1}} f_X(x_1, \dots, t, \dots, x_d) \, d\lambda_{d-1}(x_1, \dots, \hat{x}_i, \dots, x_d)$$

La conseguenza del precedente esempio è che, nota la densità congiunta di un vettore aleatorio continuo, è possibile ricavarne le marginali. Il viceversa è però vero solo nel caso di vettori aleatori con componenti indipendenti, in cui la legge congiunta è uguale al prodotto delle leggi marginali e quindi (vedi risultato alla fine della sezione precedente) la densità congiunta è uguale al prodotto delle densità marginali.

Nel caso in cui un vettore aleatorio non abbia componenti indipendenti non è nemmeno detto che, anche se le componenti ammettono una densità marginale, il vettore aleatorio ammetta una densità congiunta.

Esempio 1.71 Consideriamo $Y \sim U(0, 1)$ e definiamo il vettore aleatorio $X = (X_1, X_2) := (Y, Y)$. Allora entrambe le componenti di X ammettono come densità $\mathbf{1}_{(0,1)}$ rispetto alla misura di Lebesgue, tuttavia X non ammette densità congiunta. Difatti, se per assurdo ammettesse una densità congiunta f_X , si avrebbe che, definito ad esempio $A := \{(x, x) \mid x \in (0, 1)\} \in \mathcal{B}(\mathbb{R}^2)$, si avrebbe

$$\mathbb{P}\{X \in A\} = \int_A f_X \, d\lambda_2 = 0$$

poichè $\lambda_2(A) = 0$; ma si avrebbe anche, per come è definito X ,

$$\mathbb{P}\{X \in A\} = \mathbb{P}\{(Y, Y) \in A\} = 1$$

e si arriverebbe ad un assurdo.

Nel caso di vettori aleatori continui, è possibile calcolare gli integrali di Lebesgue che appaiono nel calcolo di speranze, probabilità, ecc., usando tutte le tecniche dell'integrazione multipla. Tra le varie tecniche, citiamo qui per completezza il caso di cambiamento di variabili.

Proposizione 1.72 (composizione di un vettore aleatorio con diffeomorfismi)

Sia X vettore aleatorio continuo di densità f_X a valori in un aperto $B \subseteq \mathbb{R}^d$, e $Y := \varphi(X)$, con $\varphi : B \rightarrow \mathbb{R}^d$ diffeomorfismo di classe C^1 (quindi con immagine $\varphi(B) \subseteq \mathbb{R}^d$). Allora anche Y è continuo, e ha densità data da

$$f_Y(y) := f_X(\varphi^{-1}(y)) |\det J\varphi^{-1}(y)|$$

Dimostrazione. Per ogni $A \in \mathcal{B}(\varphi(B))$ si ha

$$\mathbb{P}\{Y \in A\} = \mathbb{P}\{\varphi(X) \in A\} = \mathbb{P}\{X \in \varphi^{-1}(A)\} = \int_{\varphi^{-1}(A)} f_X(x) \, d\lambda_d(x)$$

Facendo il cambio di variabili $y := \varphi(x)$, l'integrale diventa

$$\mathbb{P}\{Y \in A\} = \int_A f_X(\varphi^{-1}(y)) |\det J\varphi^{-1}(y)| \, d\lambda_d(y)$$

e si ha la tesi. □