Appunti del corso "Introduzione ai Processi Aleatori"

Indice

1 Probabilità condizionata a $\sigma-$algebre

1.1 Il caso discreto

Sia $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità:

- Ω spazio campionario;
- $\mathcal{F}$ $\sigma-$algebra di eventi di Ω:
  - $*$ $\emptyset, \Omega \in \mathcal{F}$,
  - $*$ se $A \in \mathcal{F}$ allora $A^c \in \mathcal{F}$ $[A^c = \Omega - A]$,
  - $*$ se $\{A_n\}_n \subset \mathcal{F}$ allora $\cup_{n=1}^{\infty} A_n \in \mathcal{F}$;
- P misura di probabilità su $(\Omega, \mathcal{F})$: $P : \mathcal{F} \to [0, 1]$ tale che
  - $**$ $P(\Omega) = 1$,
  - $**$ $P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$, se $A_n \cap A_m = \emptyset$ per ogni $n \neq m$ ($\sigma-$*additività*).

Formula di Bayes: se $B \in \mathcal{F}$, $P(B) > 0$,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

rappresenta la probabilità del verificarsi dell'evento A nota l'informazione data da B. Si noti che, al variare di $A \in \mathcal{F}$, $P(A|B)$ è una misura di probabilità.

Ovviamente, la probabilità di A condizionata al non verificarsi di B è

$$P(A|B^c) = \frac{P(A \cap B^c)}{P(B^c)}$$

purché $P(B^c) > 0$. Ciò significa che la probabilità condizionata cambia a seconda del verificarsi o meno di B. In altre parole, essa dipende dall'osservazione effettuata, cioè da $\omega \in \Omega$, ed il suo valore può essere calcolato tramite la formula di Bayes: se $0 < P(B) < 1$, essa vale

$$f(\omega) = \begin{cases} P(A|B) & \text{se } \omega \in B \\ P(A|B^c) & \text{se } \omega \in B^c \end{cases} \tag{1}$$

Osservazione 1.1 Ci si potrebbe chiedere quale valore debba essere attribuito alla funzione f quando $P(B) = 0$ oppure $P(B) = 1$. In realtà, in tal caso la definizione di f non è "ben posta". Infatti, quando $P(B) = 0$ l'aver osservato B non dà alcuna informazione riguardo la probabilità del verificarsi di A, ovvero $P(A|B)$. Se invece $P(B) = 1$, allora $P(B^c) = 0$ e quindi in tal caso è $P(A|B^c)$ a non essere ben definita. Ne segue che la formula (1) è sempre ben definita a meno di insiemi di probabilità nulla, cioè quasi certamente (q.c.). Qualora l'insieme su cui f non è definita è non vuoto (cioè $P(B) = 0$ oppure $P(B) = 1$, con $B \neq \emptyset$ o Ω), su tale insieme il valore di f sarà preso in modo arbitrario.

Come vedremo in seguito, la (1) definisce la **probabilità di A condizionata alla $\sigma-$algebra $\mathcal{G} = \{\emptyset, \Omega, B, B^c\}$**. Si noti che tale probabilità è una variabile aleatoria (v.a.) reale, ovvero una funzione reale $\mathcal{F}-$misurabile: se $\mathcal{B}(\mathbb{R})$ denota la $\sigma-$ algebra di Borel su $\mathbb{R}$, allora

$$f : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$
$$\text{con } f^{-1}(H) = \{\omega \in \Omega : f(\omega) \in H\} \in \mathcal{F}, \text{per ogni } H \in \mathcal{B}(\mathbb{R})$$

La famiglia di insiemi $\{B, B^c\}$ è una partizione (banale) di Ω e l'idea utilizzata per costruire la funzione f definita in (1) consente di generalizzare il concetto di probabilità condizionata a $\sigma-$algebre generate[1] da partizioni qualsiasi nel modo seguente.

Sia $\{B_i\}_{i \in I}$ una partizione al più numerabile di Ω:

- $\text{card}(I) \leq \text{card}(\mathbb{N})$;
- $B_i \cap B_j = \emptyset$ per ogni $i \neq j$;
- $\cup_{i \in I} B_i = \Omega$.

Sia $\mathcal{G}$ la $\sigma-$algebra generata da $\{B_i\}_{i \in I}$. Si dimostra (esercizio!) che $\mathcal{G}$ è formata da tutte le possibili unioni degli elementi della partizione:

$$\mathcal{G} = \{G : G = \cup_{i \in I_G} B_i, I_G \subset I\}$$

[1]Data una famiglia $\mathcal{P}$ di insiemi di Ω, ricordiamo che la **$\sigma-$algebra generata da** $\mathcal{P}$, in simboli $\sigma(\mathcal{P})$, è la più piccola $\sigma-$algebra contenente $\mathcal{P}$. Se X denota una variabile aleatoria su $(\Omega, \mathcal{F})$, ovvero $X : (\Omega, \mathcal{F}) \to (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, dove $\mathcal{B}(\mathbb{R}^n)$ denota la $\sigma-$algebra di Borel su $\mathbb{R}^n$, la **$\sigma-$algebra generata da** X, in simboli $\sigma(X)$, è la più piccola $\sigma-$algebra di eventi di Ω rispetto alla quale X è misurabile. Si può dimostrare che $\sigma(X)$ è costituita dalle controimmagini tramite X dei boreliani di $\mathbb{R}^n$: $\sigma(X) = \{X^{-1}(H) : H \in \mathcal{B}(\mathbb{R}^n)\}$.

Fissato $A \in \mathcal{F}$, sia

$$f(\omega) = P(A|B_i) \qquad \text{se } \omega \in B_i \text{ e } P(B_i) > 0 \tag{2}$$

Posto $P(A|B_i) = c$, con c costante arbitraria, per ogni indice i tale che $P(B_i) = 0$, la funzione f diviene definita per ogni $\omega \in \Omega$ e si può scrivere in modo equivalente

$$f(\omega) = \sum_{i \in I} P(A|B_i) \mathbb{1}_{B_i}(\omega) \tag{3}$$

oppure

$$f(\omega) = P(A|B_{i^*(\omega)}) \tag{4}$$

dove $i^* : \Omega \to I$, $i^*(\omega)$ è quell'indice (unico!) $i \in I$ tale che $\omega \in B_i$ ($i^*(\omega) = i$ se e solo se $\omega \in B_i$).

Alla funzione f si può dare il seguente significato.

La partizione $\{B_i\}_{i \in I}$, o in modo equivalente la $\sigma-$algebra $\mathcal{G}$, si può interpretare come un possibile esperimento; se un osservatore sa quale elemento B_i della partizione contiene ω significa allora che conosce il risultato dell'esperimento e $f(\omega)$ dà quindi la nuova probabilità dell'evento A nota l'osservazione. Per tale ragione la v.a. f è detta **probabilità di A condizionata alla $\sigma-$algebra $\mathcal{G}$** ed è usualmente denotata con il simbolo $P(A|\mathcal{G})$:

$$f(\omega) = P(A|\mathcal{G})(\omega)$$

Ovviamente, $P(A|\mathcal{G})$ non è univocamente definita per quei valori di ω appartenenti ad insiemi B_i per i qual $P(B_i) = 0$ e per questo motivo la probabilità condizionata è definita q.c. (si veda l'Osservazione 1.1).

Esempio 1.2 Sia N_t un processo di Poisson di parametro λ. Fissato $t > 0$, sia

$$B_i = \{\omega : N_t(\omega) = i\}, \quad i = 0, 1, 2, \ldots$$

Ovviamente $\{B_i\}_{i \geq 0}$ è una partizione di Ω. Sia $\mathcal{G}$ la $\sigma-$algebra generata da $\{B_i\}_{i \geq 0}$:

$$\mathcal{G} = \sigma(B_0, B_1, B_2, \ldots) = \sigma(N_t)$$

Fissato $s < t$, calcoliamo $P(N_s = k|\mathcal{G})$ per $k = 0, 1, 2, \ldots$, ovvero la probabilità di $A = \{\omega : N_s(\omega) = k\}$ condizionata alla $\sigma-$algebra $\mathcal{G}$. Si noti che, interpretando N_t come il numero di telefonate arrivate ad un centralino fino all'istante t, $P(N_s = k|\mathcal{G})(\omega)$ è la probabilità che ad un istante (passato) s siano arrivate k telefonate condizionatamente al numero totale di chiamate $N_t(\omega)$ pervenute al tempo (presente) t. Infatti, dalla (2),

$$P(N_s = k|\mathcal{G})(\omega) = P(N_s = k|N_t = i) \qquad \text{quando } N_t(\omega) = i$$

Poiché $N_s \leq N_t$ per ogni $s < t$, $P(N_s = k|N_t = i) = 0$ per ogni $i < k$, dunque

$$P(N_s = k|\mathcal{G})(\omega) = 0 \qquad \text{se } N_t(\omega) < k$$

Supponiamo ora $i \geq k$. In tal caso, dalla formula di Bayes,

$$\begin{aligned} P(N_s = k|N_t = i) &= \frac{P(N_s = k, N_t = i)}{P(N_t = i)} = \frac{P(N_s = k, N_t - N_s = i - N_s)}{P(N_t = i)} \\ &= \frac{P(N_s = k, N_t - N_s = i - k)}{P(N_t = i)} \end{aligned}$$

Useremo ora alcune proprietà caratteristiche del processo di Poisson:

- gli incrementi sono indipendenti: $P(N_s = k, N_t - N_s = i - k) = P(N_s = k)P(N_t - N_s = i - k)$;
- gli incrementi $N_t - N_s$ sono v.a. di Poisson di parametro $\lambda(t-s)$, per ogni $s < t$:

$$P(N_t - N_s = j) = \frac{(\lambda(t-s))^j}{j!} e^{-\lambda(t-s)}, \qquad j = 0, 1, \dots;$$

- per $r > 0$, N_r è una v.a. di Posson di parametro λr:

$$P(N_r = j) = \frac{(\lambda r)^j}{j!} e^{-\lambda r}, \qquad j = 0, 1, \dots;$$

Allora,

$$\begin{aligned} P(N_s = k|N_t = i) &= \frac{(\lambda s)^k}{k!} e^{-\lambda s} \cdot \frac{(\lambda(t-s))^{(i-k)}}{(i-k)!} e^{-\lambda(t-s)} \cdot \frac{i!}{(\lambda t)^i} e^{\lambda t} \\ &= \binom{i}{k} \left(\frac{s}{t}\right)^k \left(1 - \frac{s}{t}\right)^{i-k} \end{aligned}$$

che, osserviamo, è una distribuzione binomiale di parametri i e $\frac{s}{t}$. Possiamo dunque scrivere che

$$P(N_s = k|\mathcal{G})(\omega) = \begin{cases} \binom{N_t(\omega)}{k} \left(\frac{s}{t}\right)^k \left(1 - \frac{s}{t}\right)^{N_t(\omega)-k} & \text{se } k \in \{0, 1, \dots N_t(\omega)\} \\ 0 & \text{altrimenti} \end{cases} \tag{5}$$

1.2 Il caso generale

Diamo qui la definizione formale di probabilità condizionata ad una $\sigma-$algebra e presentiamo alcuni esempi e proprietà che ne conseguono.

Definizione 1.3 Sia $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità, A un evento fissato in $\mathcal{F}$ e $\mathcal{G}$ una sotto $\sigma-$algebra[2] di $\mathcal{F}$. Una versione $P(A|\mathcal{G})$ della *probabilità di A condizionata a $\mathcal{G}$* è una v.a. tale che:

[2]Ovvero, $\mathcal{G}$ è una $\sigma-$algebra di eventi di Ω e $\mathcal{G} \subset \mathcal{F}$.

(i) $P(A|\mathcal{G})$ è $\mathcal{G}-$misurabile[3] e integrabile;

(ii) Per ogni $G \in \mathcal{G}$,

$$\int_G P(A|\mathcal{G})(\omega) dP(\omega) = P(A \cap G)$$

o, equivalentemente,

$$\mathbb{E}[P(A|\mathcal{G})\, 1\!\!1_G] = P(A \cap G)$$

Occorre subito mostrare l'esistenza di una v.a. che soddisfi le condizioni richieste e giustificare il termine "versione" introdotto nella definizione precedente. A tale scopo, sia μ la misura su $\mathcal{G}$ definita da

$$\mu(G) = P(A \cap G), \qquad G \in \mathcal{G}$$

Se $P(G) = 0$ allora $\mu(G) = 0$, cioè μ è assolutamente continua rispetto a P. Il teorema di Radon–Nicodym[4] assicura allora l'esistenza di una funzione f, $\mathcal{G}-$misurabile e integrabile, tale che per ogni $G \in \mathcal{G}$,

$$\int_G f(\omega) dP(\omega) = \mu(G) \tag{6}$$

Basterà dunque porre $P(A|\mathcal{G})(\omega) = f(\omega)$. L'unicità della funzione f che verifica la (6) non è però assicurata[5]. Ciò significa che la probabilità condizionata $P(A|\mathcal{G})$ è unica q.c., cioè a meno di insiemi di $\mathcal{G}$ di probabilità 0, e per tale ragione si usa il termine "versione" della probabilità condizionata, termine che d'ora in poi verrà omesso.

Esempio 1.4 Se $\mathcal{G} = \{\emptyset, \Omega\}$ allora $P(A|\mathcal{G})(\omega) = P(A)$ (verificare (i) e (ii)!).

Esempio 1.5 Sia A un evento indipendente da $\mathcal{G}$, ovvero $P(A \cap G) = P(A)P(G)$ per ogni $G \in \mathcal{G}$. In tal caso, $P(A|\mathcal{G})(\omega) = P(A)$. Infatti, ogni funzione costante è misurabile rispetto alla $\sigma-$algebra banale $\{\emptyset, \Omega\}$, dunque rispetto ad ogni $\sigma-$algebra. Inoltre,

$$P(A \cap G) = P(A)P(G) = \int_G P(A) dP(\omega)$$

sicché anche (ii) è verificata.

Esempio 1.6 Se in particolare $A \in \mathcal{G}$, allora $P(A|\mathcal{G})(\omega) = 1\!\!1_A(\omega)$. Infatti, (i) è immediata. Riguardo (ii), basta osservare che

$$P(A \cap G) = \int_{A \cap G} dP(\omega) = \int_G 1\!\!1_A(\omega) dP(\omega).$$

[3] Ovvero, per ogni boreliano H di $\mathbb{R}$, $P(A|\mathcal{G})^{-1}(H) = \{\omega : P(A|\mathcal{G})(\omega) \in H\} \in \mathcal{G}$. Si noti che, poiché $P(A|\mathcal{G})$ è una v.a., in particolare $P(A|\mathcal{G})^{-1}(H) \in \mathcal{F}$. La $\mathcal{G}-$misurabilità esprime dunque la necessità che la probabilità condizionata a $\mathcal{G}$ debba dipendere solo dagli eventi di $\mathcal{G}$.

[4] (Radon–Nicodym) *Sia ν_1 una misura su $(\Omega, \mathcal{G})$ finita e $\sigma-$additiva. Sia ν_2 una misura (o, più in generale, una carica o misura con segno) definita sullo stesso spazio, assolutamente continua rispetto a ν_1. Allora esiste su Ω una funzione f integrabile rispetto a ν_1 tale che $\nu_2(G) = \int_G f(\omega) d\nu_1(\omega)$ per ogni $G \in \mathcal{G}$.*

[5] Se g è una funzione $\mathcal{G}-$misurabile tale che $\{x : g(x) \neq 0\}$ è un insieme di $P-$misura nulla, $f + g$ rimane ancora $\mathcal{G}-$misurabile, integrabile e verifica la (6).

Esempio 1.7 Sia $\mathcal{G}$ la $\sigma-$algebra generata da una partizione $\{B_i\}_{i\in I}$ al più numerabile di Ω e sia A un evento fissato di $\mathcal{F}$. Mostriamo che la formula (2), o equivalentemente (3) e (4), danno effettivamente la probabilità di A condizionata a $\mathcal{G}$ come richiesto dalla Definizione 1.3.

Mostriamo dapprima la $\mathcal{G}-$misurabilità. Sia H un boreliano di $\mathbb{R}$. Dalla (2) segue che

$$\omega \in f^{-1}(H) \text{ se e solo se esiste } i \in I \text{ t.c. } P(A|B_i) \in H$$

Dunque,

$$f^{-1}(H) = \cup_{i\in J} B_i, \text{ dove } J = \{i \in I : P(A|B_i) \in H\}.$$

Ricordando che $\mathcal{G}$ è fatta di tutte le possibili unioni degli elementi della partizione, la proprietà (i) segue immediatamente.

Occorre ora verificare (ii). Calcoliamo dunque $\mathbb{E}[f\, 1\!\!1_G]$, dove $G = \cup_{i\in I_G} B_i \in \mathcal{G}$. Dalla rappresentazione (3) possiamo scrivere $f(\omega)\, 1\!\!1_G(\omega) = \sum_{i\in I_G} P(A|B_i) 1\!\!1_{B_i}(\omega)$, da cui

$$\mathbb{E}[f\, 1\!\!1_G] = \sum_{i\in I_G} P(A|B_i)P(B_i) = \sum_{i\in I_G} P(A\cap B_i) = P(A\cap \cup_{i\in I_G} B_i) = P(A\cap G)$$

Sia $Y = (Y_1, \ldots, Y_n)$ un vettore di v.a. reali e sia $\sigma(Y)$ (o $\sigma(Y_1, \ldots, Y_n)$) la $\sigma-$algebra generata da Y:

$$\sigma(Y) = \sigma(Y_1, \ldots, Y_n) = \{Y^{-1}(H)\, ;\, H \in \mathcal{B}(\mathbb{R}^n)\}$$

dove $\mathcal{B}(\mathbb{R}^n)$ denota la $\sigma-$algebra di Borel su $\mathbb{R}^n$. Fissato $A \in \mathcal{F}$, si consideri la probabilità $P(A|\sigma(Y))(\omega)$ di A condizionata alla $\sigma-$algebra $\sigma(Y)$. In tal caso, vengono usualmente utilizzate le notazioni

$$P(A|Y)(\omega) \quad \text{oppure} \quad P(A|Y_1, \ldots, Y_n)(\omega)$$

per sottolineare il fatto che la probabilità condizionata dipende da ω solo tramite $Y(\omega) = (Y_1(\omega), \ldots, Y_n(\omega))$. Infatti, dalla proprietà (i) della Definizione 1.3, la v.a. $P(A|\sigma(Y))(\omega)$ è $\sigma(Y)$ misurabile: esiste quindi[6] una funzione boreliana[7] Ψ tale che

$$P(A|Y)(\omega) = \Psi(Y(\omega))$$

o in modo equivalente,

$$P(A|Y_1, \ldots, Y_n)(\omega) = \Psi(Y_1(\omega), \ldots, Y_n(\omega)).$$

La funzione Ψ viene denotata

$$\Psi(y) = P(A|Y = y), \quad y \in \mathbb{R}^n$$

ma attenzione a non confondere $P(A|Y = y)$ con la probabilità di A condizionata all'evento $\{Y = y\}$ (che, osserviamo, in molti casi di interesse è un evento di probabilità nulla, per il quale quindi la probabilità condizionata non è ben definita, si veda l'Esempio 1.9).

[6]Se Z è una v.a. $\sigma(Y)-$misurabile allora esiste una funzione (boreliana) ψ tale che $Z = \psi(Y)$.

[7]Una funzione $\psi : \mathbb{R}^k \to \mathbb{R}^m$ è boreliana se è misurabile rispetto alla $\sigma-$algebra di Borel: $\psi^{-1}(H) \in \mathcal{B}(\mathbb{R}^k)$ per ogni $H \in \mathcal{B}(\mathbb{R}^m)$. In questo caso, $k = n$ e $m = 1$.

Esempio 1.8 Sia N un processo di Poisson e definiamo $\mathcal{G} = \sigma(N_t)$, per $t > 0$ fissato. Dall'Esempio 1.2 segue allora che, per $s < t$,

$$P(N_s = k|\mathcal{G})(\omega) = P(N_s = k|N_t)(\omega) = \Psi(N_t(\omega)), \quad k \in \mathbb{N}$$

dove $\Psi(y) = P(N_s = k|N_t = y)$ è data dalla formula (5), i.e.

$$P(N_s = k|N_t = y) = \begin{cases} \binom{y}{k} \left(\frac{s}{t}\right)^k \left(1 - \frac{s}{t}\right)^{y-k} & \text{se } k \in \{0, 1, \dots y\} \\ 0 & \text{altrimenti} \end{cases}$$

Esempio 1.9 Siano X e Y due v.a. reali assolutamente continue: esiste una funzione di densità di probabilità $p_{X,Y}(x,y) : \mathbb{R}^2 \to [0, +\infty)$ (misurabile) tale che

$$P((X,Y) \in H) = \int_H p_{X,Y}(x,y)dxdy \tag{7}$$

Posto $\mathcal{G} = \sigma(Y)$, cerchiamo la probabilità dell'evento $A = \{X \in \Gamma\}$ condizionata a Y, cioè $P(X \in \Gamma|Y)(\omega) = P(X \in \Gamma|\mathcal{G})(\omega)$. Definiamo

$$p_{X|Y}(x|y) = \begin{cases} \dfrac{p_{X,Y}(x,y)}{p_Y(y)} & \text{se } p_Y(y) > 0 \\ 0 & \text{altrimenti} \end{cases}$$

dove $p_Y(y)$ è la densità marginale della v.a. Y: $p_Y(y) = \int_{\mathbb{R}} p_{X,Y}(x,y)dx$. La funzione $p_{X|Y}(\cdot|y) : \mathbb{R} \to [0, +\infty)$ prende il nome di **densità di X condizionata a Y**, perché fissato $\Gamma \in \mathcal{B}(\mathbb{R})$ e posto

$$\Psi(y) = \int_\Gamma p_{X|Y}(x|y)dx \tag{8}$$

si ha che

$$P(X \in \Gamma|Y)(\omega) = \Psi(Y(\omega)).$$

Verifichiamo la validità delle affermazioni contenute nella Definizione 1.3.

La $Y-$misurabilità è immediata. Infatti $\Psi(y) : \mathbb{R} \to [0, +\infty)$ è una funzione misurabile e quindi $\Psi(Y(\omega))$ è $\mathcal{G} \equiv \sigma(Y)-$misurabile perché funzione composta di funzioni misurabili. Per quanto riguarda (ii), occorre mostrare che

$$P(\{X \in \Gamma\} \cap G) = \int_G \Psi(Y(\omega))dP(\omega) \quad \text{per ogni } G \in \sigma(Y)$$

Poiché gli insiemi $G \in \sigma(Y)$ si possono rappresentare nella forma $G = \{\omega : Y(\omega) \in \Lambda\} \equiv \{Y \in \Lambda\}$ al variare di $\Lambda \in \mathcal{B}(\mathbb{R})$, utilizzando la legge indotta su $\mathbb{R}$ da Y la condizione su scritta equivale a

$$P(\{X \in \Gamma\} \cap \{Y \in \Lambda\}) \equiv P((X,Y) \in \Gamma \times \Lambda\}) = \int_\Lambda \Psi(y)p_Y(y)dy \quad \text{per ogni } \Lambda \in \mathcal{B}(\mathbb{R})$$

Infatti, dalla (7) (con $H = \Gamma \times \Lambda$), si ha

$$P((X,Y) \in \Gamma \times \Lambda) = \int_{\Gamma \times \Lambda} p_{X,Y}(x,y)dxdy = \int_{\Gamma \times \Lambda} \frac{p_{X,Y}(x,y)}{p_Y(y)} p_Y(y)dxdy =$$
$$= \int_\Lambda \left(\int_\Gamma \frac{p_{X,Y}(x,y)}{p_Y(y)} dx \right) p_Y(y)dy = \int_\Lambda \Psi(y) p_Y(y)dy$$

In tal caso dunque $P(X \in \Gamma | Y = y) = \int_\Gamma p_{X|Y}(x|y)dx$ (e si noti che $P(Y = y) = 0$, per ogni y!).

Riportiamo ora alcune proprietà principali della probabilità condizionata a $\sigma-$algebre, che raccoglieremo in un unico

Teorema 1.10 *Per ogni A,*

$$0 \leq P(A|\mathcal{G}) \leq 1 \quad q.c. \tag{9}$$

Per ogni successione $A_1, A_2, \ldots$ numerabile di insiemi disgiunti

$$P\left(\cup_n A_n | \mathcal{G}\right) = \sum_n P(A_n|\mathcal{G}) \quad q.c. \tag{10}$$

Se $A \subset B$,

$$P(B - A|\mathcal{G}) = P(B|\mathcal{G}) - P(A|\mathcal{G}) \quad q.c. \tag{11}$$

Vale la formula di inclusione–esclusione:

$$P\left(\cup_n A_n | \mathcal{G}\right) = \sum_n P(A_n|\mathcal{G}) - \sum_{n<m} P(A_n \cap A_m|\mathcal{G}) + \ldots \quad q.c. \tag{12}$$

Per ogni successione $A_1, A_2, \ldots$ numerabile di insiemi tali che, per ogni n, $A_n \subset A_{n+1}$

$$\lim_{n\to\infty} P(A_n|\mathcal{G}) = P\left(\cup_n A_n | \mathcal{G}\right) \quad q.c. \tag{13}$$

Per ogni successione $A_1, A_2, \ldots$ numerabile di insiemi tali che, per ogni n, $A_{n+1} \subset A_n$

$$\lim_{n\to\infty} P(A_n|\mathcal{G}) = P\left(\cap_n A_n | \mathcal{G}\right) \quad q.c. \tag{14}$$

Se $P(A) = 1$ $[P(A = 0)]$,

$$P(A|\mathcal{G}) = 1 \quad [P(A|\mathcal{G}) = 0] \quad q.c. \tag{15}$$

1.3 Distribuzione condizionata

Nel paragrafo precedente abbiamo definito la probabilità *di un evento fissato* $A \in \mathcal{F}$ condizionata ad una $\sigma-$algebra $\mathcal{G}$. In questa sezione discuteremo l'esistenza di *misure di probabilità* condizionate a $\sigma-$algebre, l'esistenza cioè di

$$P(A|\mathcal{G})(\omega), \quad \text{per } \omega \in \Omega \text{ fissato, al variare di } A \in \mathcal{F}$$

Supponiamo ad esempio che $\mathcal{G}$ sia la $\sigma-$algebra generata da una partizione $\{B_1, B_2, \ldots\}$ di Ω e facciamo l'ipotesi che $B_1 \neq \emptyset$, $P(B_1) = 0$ e $P(B_i) > 0$, per ogni $i \geq 2$. Fissato $A \in \mathcal{F}$, una versione di $P(A|\mathcal{G})$ è

$$P(A|\mathcal{G})(\omega) = \begin{cases} 5 & \text{se } \omega \in B_1 \\ P(A|B_i) & \text{se } \omega \in B_i, i \geq 2 \end{cases}$$

È facile verificare che, come funzione di $A \in \mathcal{F}$, $P(A|\mathcal{G})(\omega)$ è una misura di probabilità su $\mathcal{F}$ *solo se* $\omega \notin B_1$. Non è però questo un gran danno: $P(\cdot|\mathcal{G})(\omega)$ è una misura di probabilità su $\mathcal{F}$ per ogni $\omega \in \cup_{i\geq 2} B_i$, un insieme che pur non essendo Ω ha almeno probabilità 1 (in altre parole, in questo esempio $P(\cdot|\mathcal{G})(\omega)$ esiste a meno di un insieme di probabilità nulla). Inoltre, non è difficile modificare la probabilità su scritta per ottenere una misura di probabilità su $\mathcal{F}$ per ogni ω: basterà porre

$$P(A|\mathcal{G})(\omega) = \begin{cases} P(A) & \text{se } \omega \in B_1 \\ P(A|B_i) & \text{se } \omega \in B_i, i \geq 2 \end{cases}$$

In generale però non è possibile fare un discorso di questo tipo. Infatti, abbiamo visto che $P(A|\mathcal{G})(\omega)$ è ben posta (è cioè l'unica funzione di ω che verifica le condizioni della Definizione 1.3) a meno di insiemi di probabilità nulla: fissato $A \in \mathcal{F}$, esiste un insieme $N_A \in \mathcal{F}$ (in generale non vuoto) tale che $P(A|\mathcal{G})(\omega)$ è ben posta per ogni $\omega \notin N_A$. Ora, volendo far variare $A \in F$, otterremmo che $P(A|\mathcal{G})(\omega)$ è ben posta per ogni $\omega \notin N \equiv \cup_{A\in\mathcal{F}} N_A$, ovvero $P(A|\mathcal{G})(\omega)$ è ben definita per $\omega \in \Omega - N$. Il problema però è che in questo modo la probabilità condizionata potrebbe non essere più, come funzione di ω, un variabile aleatoria (cosa sulla quale non si transige!). Difatti, l'insieme N defiinto sopra potrebbe non appartenere più alla $\sigma-$algebra di riferimento $\mathcal{F}$: sappiamo infatti che $\mathcal{F}$ è chiusa sotto unioni numerabili e la cardinalità di $\mathcal{F}$ potrebbe non esserlo, così che $N \notin \mathcal{F}$. Inoltre, qualora $N \in \mathcal{F}$ niente assicura che $P(N) = 0$[8], e dunque $P(A|\mathcal{G})(\omega)$ potrebbe non essere definita su insiemi troppo grandi (di misura non nulla). Potrebbe anche essere $N = \Omega$, nel qual caso $P(A|\mathcal{G})(\omega)$ non sarebbe definita per nessun ω![9]

Il problema appena esposto diviene più semplice, e con soluzione, quando si vuole studiare la distribuzione di una v.a. X (che in questa trattazione supporremo a valori in $\mathbb{R}$) condizionata ad una $\sigma-$algebra $\mathcal{G}$, ovvero quando la misura di probabilità che si cerca è data da

$$\mu(H, \omega) = P(X \in H|\mathcal{G})(\omega), \quad \text{per } \omega \in \Omega \text{ fissato, al variare di } H \in \mathcal{B}(\mathbb{R})$$

La misura di probabilità $\mu(\cdot, \omega)$ su $(\mathbb{R}, \mathcal{B}(\mathbb{R})$ è detta **distribuzione di X condizionata a $\mathcal{G}$** e il teorema che segue ne assicura l'esistenza:

[8] L'unione numerabile di insiemi di probabilità nulla ha probabilità nulla. Come già osservato, $N = \cup_{A\in\mathcal{F}} N_A$ è unione di una quantità di eventi che potrebbe essere non numerabile, dunque potrebbe avere probabilità positiva.

[9] Esistono esempi in letteratura sull'impossibilità di costruire misure di probabilità condizionate.

Teorema 1.11 *Esiste una funzione $\mu(H, \omega)$, per $\omega \in \Omega$ e $H \in \mathcal{B}(\mathbb{R})$, tale che*

(D1) *per ogni $\omega \in \Omega$ fissato, $\mu(H, \omega)$ è, al variare di $H \in \mathcal{B}(\mathbb{R})$, una misura di probabilità;*

(D2) *per ogni $H \in \mathcal{B}(\mathbb{R})$ fissato, $\mu(H, \omega)$ è, al variare di $\omega \in \Omega$, una versione di $P(X \in H|\mathcal{G})(\omega)$.*

Esempio 1.12 Sia N un processo di Poisson e $t > s > 0$ due tempi fissati. Dall'Esempio 1.8 segue allora che la distribuzione di N_s condizionata alla $\sigma-$algebra generata da N_t è discreta e coincide con una distribuzione binomiale di parametri $N_t(\omega)$ e $\frac{s}{t}$:

$$\mu(\{k\}, \omega) = P(N_s = k|N_t)(\omega) = \binom{N_t(\omega)}{k} \left(\frac{s}{t}\right)^k \left(1 - \frac{s}{t}\right)^{N_t(\omega)-k}, \quad k = 0, 1, \ldots N_t(\omega)$$

Esempio 1.13 Sia $\mathcal{G} = \sigma(Y)$ e supponiamo che le v.a. X e Y soddisfino le ipotesi dell'Esempio 1.9. In tal caso,

$$\mu(H, \omega) = P(X \in H|Y)(\omega) = \int_H p_{X|Y}(x|y)dx\Big|_{y=Y(\omega)} = \int_H p_{X|Y}(x|Y(\omega))dx$$

2 Media condizionata a $\sigma-$algebre

2.1 Definizione e proprietà

In questo paragrafo, condidereremo per semplicità v.a. a valori reali.

Definizione 2.1 Sia $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità, X una v.a. integrabile[10] e $\mathcal{G}$ una sotto $\sigma-$algebra di $\mathcal{F}$. Una versione $\mathbb{E}[X|\mathcal{G}]$ della *media di X condizionata a $\mathcal{G}$* è una v.a. tale che:

(i) $\mathbb{E}[X|\mathcal{G}]$ è $\mathcal{G}-$misurabile e integrabile;

(ii) per ogni $G \in \mathcal{G}$,
$$\int_G \mathbb{E}[X|\mathcal{G}](\omega)dP(\omega) = \int_G X(\omega)dP(\omega)$$
o, equivalentemente,
$$\mathbb{E}\left[\,\mathbb{E}[X|\mathcal{G}\,]\, \mathbb{1}_G\right] = \mathbb{E}[X\, \mathbb{1}_G]$$

La definizione di media condizionata è assai simile a quella di probabilità condizionata. Ancora una volta occorre provare l'esistenza di una v.a. che verifichi le condizioni richieste nella definizione precedente. Utilizzeremo nuovamente il teorema di Radon–Nicodym. Definiamo

$$Q(G) = \int_G X(\omega)dP(\omega), \quad G \in \mathcal{G}$$

[10]Ovvero, $E[|X|] < \infty$.

Q definisce una carica su $\mathcal{G}$[11] assolutamente continua rispetto a P[12]. Per il teorema di Radon–Nicodym[13] esiste una funzione $f(\omega)$, unica a meno di insiemi di $P-$misura nulla, $\mathcal{G}-$misurabile e integrabile rispetto alla misura P, tale che per ogni $G \in \mathcal{G}$,

$$Q(G) = \int_G f(\omega) dP(\omega)$$

Posto $\mathbb{E}[X|\mathcal{G}](\omega) = f(\omega)$, il problema dell'esistenza della media condizionata è risolto. Così come nel caso della probabilità condizionata, l'unicità della media condizionata è garantita a meno di insiemi di $P-$misura nulla e tale considerazione giustifica l'uso del termine "versione" che viene fatto nella Definizione 2.1.

Esempio 2.2 Se $\mathcal{G} = \{\emptyset, \Omega\}$ allora $\mathbb{E}[X|\mathcal{G}](\omega) = \mathbb{E}[X]$ (verificare (i) e (ii)!).

Esempio 2.3 Sia X una v.a. indipendente da $\mathcal{G}$, ovvero $P(\{X \in H\} \cap G) = P(X \in H)P(G)$ per ogni $G \in \mathcal{G}$ e $H \in \mathcal{B}(\mathbb{R})$. In tal caso, $\mathbb{E}[X|\mathcal{G}](\omega) = \mathbb{E}[X]$. Infatti, tale funzione è costante e dunque $\mathcal{G}-$misurabile. Inoltre, poiché X e $1\!\!1_G$ sono variabili aleatorie indipendenti, si ha

$$\mathbb{E}\left[\, \mathbb{E}[X] 1\!\!1_G\right] = \mathbb{E}[X]\, \mathbb{E}[1\!\!1_G] = \mathbb{E}[X 1\!\!1_G]$$

sicché anche (ii) è verificata.

Esempio 2.4 Se X è una v.a. $\mathcal{G}-$misurabile, $\mathbb{E}[X|\mathcal{G}](\omega) = X$ (verificare (i) e (ii)!).

Esempio 2.5 Sia $\mathcal{G}$ la $\sigma-$algebra generata da una partizione $\{B_i\}_{i\in I}$ al più numerabile di Ω e sia X una v.a. integrabile. In tal caso,

$$\mathbb{E}[X|\mathcal{G}](\omega) = \frac{1}{P(B_i)} \int_{B_i} X(\omega) dP(\omega), \quad \text{quando } \omega \in B_i \text{ e } P(B_i) > 0 \tag{16}$$

e si definisca, ad esempio, $\mathbb{E}[X|\mathcal{G}](\omega) = c$, con c costante arbitraria, qualora $\omega \in B_i$, con $P(B_i) = 0$. Fissato $H \in \mathcal{B}(\mathbb{R})$, sia

$$J = \{i \in I : P(B_i) > 0 \text{ e } \frac{1}{P(B_i)} \int_{B_i} X(\omega) dP(\omega) \in H \text{ oppure } P(B_i) = 0 \text{ e } c \in H\}$$

Allora

$$\mathbb{E}[X|\mathcal{G}]^{-1}(H) = \{\omega : \mathbb{E}[X|\mathcal{G}](\omega) \in H\} = \cup_{i\in J} B_i \in \mathcal{G}$$

e dunque (i) è verificata. Sia ora $G = \cup_{i \in I_G} B_i$ un generico insieme di $\mathcal{G}$. La v.a. $\mathbb{E}[X|\mathcal{G}](\omega)\, 1\!\!1_G$ è discreta e assume i valori $\frac{1}{P(B_i)} \int_{B_i} X(\omega) dP(\omega)$ sugli insiemi B_i, con

[11]Una *carica* o anche *misura con segno* $Q : \mathcal{G} \to \mathbb{R}$ è una funzione di insieme, finita e $\sigma-$additiva.

[12]Ovvero, se $P(G) = 0$ allora $Q(G) = 0$.

[13]Nella versione più generale del teorema di Radon Nicodym, riassunto nella nota 4, la misura ν_2 si può sostituire con una carica.

$P(B_i) > 0$ e $i \in I_G$, così che

$$\begin{aligned}
\mathbb{E}\,[\,\mathbb{E}[X|\mathcal{G}]\ \mathbb{1}_G] &= \mathbb{E}\left[\sum_{i\in I_G}\left(\frac{1}{P(B_i)}\int_{B_i} X(\omega)dP(\omega)\right)\mathbb{1}_{B_i}\right] \\
&= \sum_{i\in I_G}\left(\frac{1}{P(B_i)}\int_{B_i} X(\omega)dP(\omega)\right)\mathbb{E}[\mathbb{1}_{B_i}] \\
&= \sum_{i\in I_G}\int_{B_i} X(\omega)dP(\omega) = \int_G X(\omega)dP(\omega) = \mathbb{E}[X\,\mathbb{1}_G]
\end{aligned}$$

Esempio 2.6 Per $A \in \mathcal{F}$, $\mathbb{E}[\mathbb{1}_A|\mathcal{G}](\omega) = P(A|\mathcal{G})(\omega)$. Più in generale, se $X = \sum_i a_i \mathbb{1}_{A_i}(\omega)$, con $a_i \in \mathbb{R}$, allora $\mathbb{E}[X|\mathcal{G}](\omega) = \sum_i a_i P(A_i|\mathcal{G})(\omega)$ (verificare (i) e (ii)!).

Nel teorema che segue sono riportate alcune proprietà della media condizionata.

Teorema 2.7 *Siano X, X_n, Y v.a. integrabili.*

1. *Se $X = c$ q.c. allora $\mathbb{E}[X|\mathcal{G}](\omega) = c$ q.c.*
2. *Per ogni $a, b \in \mathbb{R}$, $\mathbb{E}[aX + bY|\mathcal{G}](\omega) = a\mathbb{E}[X|\mathcal{G}](\omega) + b\mathbb{E}[Y|\mathcal{G}](\omega)$ q.c. (linearità).*
3. *Se $X \le Y$ q.c. allora $\mathbb{E}[X|\mathcal{G}](\omega) \le \mathbb{E}[Y|\mathcal{G}](\omega)$ q.c.*
4. *$|\mathbb{E}[X|\mathcal{G}](\omega)| \le \mathbb{E}[\,|X|\,|\mathcal{G}](\omega)$ q.c.*
5. *Se $\lim_{n\to\infty} X_n = X$ q.c. e $|X_n| \le Y$ per qualche v.a. integrabile Y, allora*
$$\lim_{n\to\infty}\mathbb{E}[X_n|\mathcal{G}](\omega) = \mathbb{E}[X|\mathcal{G}](\omega) \quad q.c.$$
6. *Se XY è integrabile e X è $\mathcal{G}-$misurabile allora $\mathbb{E}[XY|\mathcal{G}](\omega) = X\mathbb{E}[Y|\mathcal{G}](\omega)$.*
7. *Se $\mathcal{G}_1$ e $\mathcal{G}_2$ sono due sotto $\sigma-$algebre di $\mathcal{F}$ tali che $\mathcal{G}_1 \subset \mathcal{G}_2$ allora*
$$\mathbb{E}\,[\,\mathbb{E}[X|\mathcal{G}_2]\,|\mathcal{G}_1]\,(\omega) = \mathbb{E}\,[\,\mathbb{E}[X|\mathcal{G}_1]\,|\mathcal{G}_2]\,(\omega) = \mathbb{E}[X|\mathcal{G}_1](\omega) \quad q.c.$$

Qualora $\mathcal{G}$ fosse una $\sigma-$algebra generata da una v.a. Y, $\mathcal{G} = \sigma(Y)$, la proprietà (i) della Definizione 2.1 assicura che la media condizionata dipende da ω solo tramite $Y(\omega)$[14]: esiste una funzione boreliana Φ tale che

$$\mathbb{E}[X|\sigma(Y)](\omega) \equiv \mathbb{E}[X|Y](\omega) = \Phi(Y(\omega))$$

Per tale motivo la funzione $\Phi(y)$ viene solitamente denotata con $\mathbb{E}[X|Y = y]$.

[14]Ricordiamo che una v.a. se Z è $\sigma(Y)-$misurabile allora è una funzione di Y, si veda la nota 6.

2.2 Distribuzione e media condizionata

Sia X una v.a. integrabile e μ la legge indotta (distribuzione), i.e. $\mu(\Gamma) = P(X \in \Gamma)$. È ben noto che tale misura consente di determinare la media di una qualsiasi funzione della v.a. X:

$$\mathbb{E}[\varphi(X)] = \int \varphi(X(\omega))dP(\omega) = \int \varphi(x)d\mu(x)$$

per ogni funzione φ per la quale tali integrali esistono. In altre parole, la media di $\varphi(X)$ non è altro che l'integrale della funzione φ rispetto alla distribuzione indotta da X. Poiché il Teorema 1.11 assicura l'esistenza della distribuzione condizionata, ci si potrebbe chiedere se tale proprietà rimane vera anche nel caso condizionato: la media di $\varphi(X)$ condizionata a $\mathcal{G}$ è l'integrale di φ rispetto alla distribuzione di X condizionata a $\mathcal{G}$? La risposta, affermativa, è contenuta nel seguente

Teorema 2.8 *Sia $\mu(\cdot, \omega)$ la distribuzione di X condizionata a $\mathcal{G}$ (nel senso del Teorema 1.11). Se $\varphi : \mathbb{R} \to \mathbb{R}$ è una funzione boreliana tale che $\varphi(X)$ è una v.a. integrabile allora*

$$\mathbb{E}[\varphi(X)|\mathcal{G}](\omega) = \int_{\mathbb{R}} \varphi(x)\mu(dx, \omega) \quad q.c.$$

Vediamo subito alcune applicazioni di questo risultato.

Esempio 2.9 Sia ancora N un processo di Poisson, $t > s > 0$ due istanti fissati. Calcoliamo $\mathbb{E}[N_s|N_t](\omega)$. Sappiamo (cfr. Esempio 1.12) che N_s dato N_t ha distribuzione di tipo binomiale di parametri N_t e $\frac{s}{t}$, dunque[15]

$$\mathbb{E}[N_s|N_t](\omega) = \frac{s}{t} N_t(\omega)$$

Più in generale,

$$\mathbb{E}[\varphi(N_s)|N_t](\omega) = \sum_{k=1}^{N_t(\omega)} \varphi(k) \binom{N_t(\omega)}{k} \left(\frac{s}{t}\right)^k \left(1 - \frac{s}{t}\right)^{N_t(\omega)-k}$$

Si noti che in tal caso,

$$\mathbb{E}[\varphi(N_s)|N_t = m] = \sum_{k=1}^{m} \varphi(k) \binom{m}{k} \left(\frac{s}{t}\right)^k \left(1 - \frac{s}{t}\right)^{m-k}$$

(si veda la fine del paragrafo precedente per il significato di $\mathbb{E}[\varphi(N_s)|N_t = m]$).

Esempio 2.10 Siano X e Y due variabili aleatorie reali, congiuntamente assolutamente continue (si vedano le ipotesi dell'Esempio 1.9) e poniamo $\mu(H, \omega) = P(X \in H|Y)(\omega)$. Dall'Esempio 1.9, segue che

$$\mu(dx, \omega) = p_{X|Y}(x|Y(\omega))dx$$

[15] Ricordiamo che se $Z \sim \mathrm{Bi}(m, p)$ allora $\mathbb{E}[Z] = mp$

Quindi,

$$\mathbb{E}[\varphi(X)|Y](\omega) = \int \varphi(x) p_{X|Y}(x|Y(\omega))dx$$

per ogni funzione φ tale che $\varphi(X)$ sia integrabile. In questo caso,

$$\mathbb{E}[\varphi(X)|Y = y] = \int \varphi(x) p_{X|Y}(x|y)dx$$

(si veda ancora la fine del paragrafo precedente per il significato di $\mathbb{E}[\varphi(X)|Y = y]$).

3 Processi stocastici

Un **processo stocastico** è una collezione di v.a. $(X_t)_{t\in\mathcal{T}}$ su uno spazio di probabilità $(\Omega, \mathcal{F}, P)$: per ogni $t \in \mathcal{T}$ fissato[16],

$$X_t : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

Un esempio è dato dalle successioni di v.a. (nel qual caso $\mathcal{T} = \mathbb{N}$), o anche dal processo di Poisson (in cui $\mathcal{T} = [0, +\infty)$). L'indice t ha il significato di tempo e l'insieme $\mathcal{T}$ in cui può variare il parametro t può essere discreto, ad esempio $\mathbb{N}$, oppure continuo, solitamente $\mathcal{T} = [0, +\infty)$ oppure $\mathcal{T} = [0, T]$.

3.1 Distribuzioni finito dimensionali e teorema di esistenza di Kolmogorov

Ci occuperemo qui del problema dell'esistenza dei processi stocastici.

Cominciamo dapprima con il definire la distribuzione indotta dal processo, o almeno dal processo osservato in un numero finito di tempi.

Sia $(X_t)_{t\in\mathcal{T}}$ un processo stocastico. Per k intero positivo, siano $t_1, \dots t_k$ k tempi fissati in $\mathcal{T}$. Il vettore aleatorio $(X_{t_1}, \dots X_{t_k})$ induce su $\mathbb{R}^k$ la distribuzione

$$\mu_{t_1,\dots t_k}(H) = P((X_{t_1}, \dots X_{t_k}) \in H), \quad H \in \mathcal{B}(\mathbb{R}^k) \tag{17}$$

dove $\mathcal{B}(\mathbb{R}^k)$ è la $\sigma-$algebra di Borel su $\mathbb{R}^k$.

Definizione 3.1 L'insieme delle misure di probabilità

$$\mathcal{D} = \{\mu_{t_1,\dots,t_k} : k \geq 1, t_1, \dots, t_k \in \mathcal{T}\}$$

è detto *insieme delle distribuzioni finito dimensionali del processo* $(X_t)_{t\in\mathcal{T}}$.

Le distribuzioni finito dimensionali soddisfano due **proprietà di consistenza**:

[16]Lo spazio $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ può essere sostituito da un generico spazio misurabile; la trattazione proposta in queste note si limiterà al caso di processi reali.

(C1) per ogni k, per ogni scelta dei tempi $t_1, \dots t_k$ in $\mathcal{T}$ e degli insiemi $H_1, \dots, H_k$ in $\mathcal{B}(\mathbb{R})$, per ogni permutazione $(\pi_1, \dots \pi_k)$ di $(1, \dots, k)$,

$$\mu_{t_1,\dots,t_k}(H_1 \times \cdots \times H_k) = \mu_{t_{\pi_1},\dots,t_{\pi_k}}(H_{\pi_1} \times \cdots \times H_{\pi_k});$$

(C2) per ogni k, per ogni scelta dei tempi $t_1, \dots t_k, t$ in $\mathcal{T}$ e degli insiemi $H_1, \dots, H_k$ in $\mathcal{B}(\mathbb{R})$,

$$\mu_{t_1,\dots,t_k}(H_1 \times \cdots \times H_k) = \mu_{t_1,\dots,t_k,t}(H_1 \times \cdots \times H_k \times \mathbb{R}).$$

La dimostrazione di (C1) e (C2) è immediata conseguenza della (17) (l'unico vero problema è la comprensione delle notazioni!). A titolo di esempio, supponiamo $k = 3$, $(\pi_1, \pi_2, \pi_3) = (2, 3, 1)$. Allora,

$$\begin{gathered}\mu_{t_1,t_2,t_3}(H_1 \times H_2 \times H_3) = P(X_{t_1} \in H_1, X_{t_2} \in H_2, X_{t_3} \in H_3) \\ = P(X_{t_2} \in H_2, X_{t_3} \in H_3, X_{t_1} \in H_1) = \mu_{t_2,t_3,t_1}(H_2 \times H_3 \times H_1)\end{gathered}$$

Inoltre, $P(X_t \in \mathbb{R}) = 1$ per ogni $t \in \mathcal{T}$, dunque

$$\begin{gathered}\mu_{t_1,t_2,t_3}(H_1 \times H_2 \times H_3) = P(X_{t_1} \in H_1, X_{t_2} \in H_2, X_{t_3} \in H_3) \\ = P(X_{t_1} \in H_1, X_{t_2} \in H_2, X_{t_3} \in H_3, X_t \in \mathbb{R}) = \mu_{t_1,t_2,t_3,t}(H_1 \times H_2 \times H_3 \times \mathbb{R})\end{gathered}$$

Affronteremo ora il problema della costruzione e dell'esistenza dei processi stocastici.

Le distribuzioni finito dimensionali consentono di studiare la legge indotta dal processo in un numero *finito* di istanti. Un processo però definisce il comportamento di un evento aleatorio che si sviluppa, in generale, su tempi infiniti. Per tale ragione il comportamento del processo al variare di $t \in \mathcal{T}$ e per ω fissato in Ω, ovvero

$$t \mapsto X_t(\omega) \in \mathbb{R}, \quad \omega \in \Omega \text{ fissato}$$

è detta **traiettoria**. I processi stocastici vengono quindi costruiti sullo **spazio delle traiettorie**

$$\mathbb{R}^{\mathcal{T}} = \{f : \mathcal{T} \to \mathbb{R}\}$$

l'insieme cioè di tutte le possibili funzioni da $\mathcal{T}$ a $\mathbb{R}$. Ad esempio, se $\mathcal{T} = \mathbb{N}$, $\mathbb{R}^{\mathcal{T}}$ è l'insieme di tutte le succesioni reali; se $\mathcal{T} = [0, +\infty)$ [$\mathcal{T} = [0, T]$], $\mathbb{R}^{\mathcal{T}}$ è l'insieme di tutte le funzioni definite su $\mathcal{T} = [0, +\infty)$ [su $\mathcal{T} = [0, T]$] a valori reali. Come vedremo, $\mathbb{R}^{\mathcal{T}}$ sarà lo spazio su cui costruiremo il processo. Denotiamo con x il generico elemento di $\mathbb{R}^{\mathcal{T}}$ e con x_t il valore di x all'istante t.

Per $t \in \mathcal{T}$, sia Z_t la **proiezione all'istante** t:

$$\begin{gathered}Z_t : \mathbb{R}^{\mathcal{T}} \to \mathbb{R} \\ Z_t(x) = x_t\end{gathered} \tag{18}$$

Costruiamo ora una $\sigma-$algebra, che denoteremo con $\mathcal{B}^{\mathcal{T}}$, in modo tale che $(Z_t)_{t\in\mathcal{T}}$ sia un processo stocastico su $(\mathbb{R}^{\mathcal{T}}, \mathcal{B}^{\mathcal{T}})$[17].

[17] Intendiamo qui che per ogni $t \in \mathcal{T}$ fissato, $Z_t : (\mathbb{R}^{\mathcal{T}}, \mathcal{B}^{\mathcal{T}}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ è misurabile: $Z_t^{-1}(H) = \{x : Z_t(x) \in H\} = \{x : x_t \in H\} \in \mathcal{B}^{\mathcal{T}}$, per ogni $H \in \mathcal{B}(\mathbb{R})$.

Definiamo gli **insiemi cilindrici**: fissati k, $t_1, \ldots, t_k \in \mathcal{T}$, $H \in \mathcal{B}(\mathbb{R}^k)$, il cilindro associato è il sottoinsieme dello spazio delle traiettorie

$$A = \{x \in \mathbb{R}^{\mathcal{T}} : (x_{t_1}, \ldots, x_{t_k}) \in H\}. \tag{19}$$

Per esempio, se $k = 2$, $t_1, t_2 \in \mathcal{T}$, $H = [a, b) \times (c, d)$, il cilindro associato è dato dall'insieme delle traiettorie x obbligate ad essere in $[a, b)$ all'istante t_1 e a trovarsi nell'intervallo (c, d) al tempo t_2.

Sia $\mathcal{B}_0^{\mathcal{T}}$ l'insieme di tutti i possibili cilindri e $\mathcal{B}^{\mathcal{T}}$ la $\sigma-$algebra generata da $\mathcal{B}_0^{\mathcal{T}}$[18]. Si dimostra allora che il processo $(Z_t)_{t \in \mathcal{T}}$ è misurabile rispetto alla $\sigma-$algebra $\mathcal{B}^{\mathcal{T}}$[19]. Inoltre,

Teorema 3.2 (Teorema di costruzione) *Sia $\mathcal{D}$ una famiglia di distribuzioni finito dimensionali che soddisfano le proprietà di consistenza (C1) e (C2). Esiste allora una misura di probabilità P su $\mathbb{R}^{\mathcal{T}}$ che rende $(Z_t)_{t \in \mathcal{T}}$ un processo stocastico sullo spazio $(\mathbb{R}^{\mathcal{T}}, \mathcal{B}^{\mathcal{T}}, P)$ e avente $\mathcal{D}$ come insieme delle sue distribuzioni finito dimensionali.*

Tale risultato può essere riformulato nei termini seguenti:

Teorema 3.3 (Teorema di esistenza di Kolmogorov) *Sia $\mathcal{D}$ una famiglia di distribuzioni finito dimensionali che soddisfano le proprietà di consistenza (C1) e (C2). Esiste allora uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ ed un processo stocastico $(X_t)_{t \in \mathcal{T}}$ che ha proprio $\mathcal{D}$ come famiglia delle sue distribuzioni finito dimensionali.*

Ovviamente, lo spazio di probabilità $(\Omega, \mathcal{F}, P)$ e il processo stocastico $(X_t)_{t \in \mathcal{T}}$ cui si riferisce il Teorema di esistenza sono, rispettivamente, $(\mathbb{R}^{\mathcal{T}}, \mathcal{B}^{\mathcal{T}}, P)$ e il processo–proiezione $(Z_t)_{t \in \mathcal{T}}$ del Teorema di costruzione.

Idea della dimostrazione del Teorema 3.2. La misura di probabilità P del Teorema 3.2 viene dapprima definita sulla classe degli insiemi cilindrici e poi estesa su tutta la $\sigma-$algebra da essa generata.

Più in particolare, se $A \in \mathcal{B}_0^{\mathcal{T}}$ allora A è della forma (19) e la misura di probabilità P agisce su tali insiemi nel modo seguente:

$$P(A) = \mu_{t_1, \ldots, t_k}(H)$$

Si noti che, una volta dimostrato che P è estendibile su tutta la $\sigma-$algebra $\mathcal{B}^{\mathcal{T}}$, tale uguaglianza assicura che $\mathcal{D}$ è la famiglia delle distribuzioni finito dimensionali del processo Z: si noti infatti che

$$A = \{x : (Z_{t_1}(x), \ldots, Z_{t_k}(x)) \in H\}$$

Utilizzando le proprietà di consistenza, si dimostra che P è una misura di probabilità sull'algebra[20] (anche questo va provato!) $\mathcal{B}_0^{\mathcal{T}}$. Applicando il Teorema di Carathéodory[21], segue che P è estendibile sulla $\sigma-$algebra $\mathcal{B}^{\mathcal{T}}$.

---

[18] Ovvero la più piccola $\sigma-$algebra di sottoinsiemi di $\mathbb{R}^{\mathcal{T}}$ che contiene $\mathcal{B}_0^{\mathcal{T}}$.

[19] Si veda la nota 17.

[20] Un'**algebra** $\mathcal{A}$ su uno spazio E è una classe di sottoinsiemi di E tale che $\emptyset \in \mathcal{A}$, chiusa rispetto al complementare ed all'unione *finita* di insiemi.

[21] (Teorema di Carathéodory) *Una misura ν $\sigma-$finita e $\sigma-$additiva su un'algebra A di insiemi di uno spazio E ha un'unica estensione ($\sigma-$additiva) sulla $\sigma-$algebra $\sigma(\mathcal{A})$ generata da $\mathcal{A}$.*

3.2 Un esempio: il moto browniano o processo di Wiener

Definizione 3.4 Il *moto browniano* o *processo di Wiener* è un processo stocastico $(B_t)_{t\geq 0}$[22], su uno spazio di probabilià $(\Omega, \mathcal{F}, P)$, che verifica le seguenti proprietà:

(B1) all'istante inziale vale 0: $B_0 = 0$;

(B2) è un processo ad incrementi indipendenti: per ogni k, $0 \leq t_0 < t_1 < \ldots < t_k$, gli incrementi aleatori $B_{t_0}, B_{t_1} - B_{t_0}, \ldots, B_{t_k} - B_{t_{k-1}}$ sono indipendenti[23];

(B3) per ogni $t > s \geq 0$, l'incremento $B_t - B_s$ ha distribuzione gaussiana di media 0 e varianza $t - s$, in simboli: $B_t - B_s \sim \mathrm{N}(0, t-s)$, cioè

$$P(B_t - B_s \in H) = \frac{1}{\sqrt{2\pi(t-s)}} \int_H e^{-\frac{\xi^2}{2(t-s)}} d\xi$$

Tale processo deve i suoi nomi a due studiosi, Brown, botanico, che ha studiato il moto caotico delle particelle, e Wiener, che ha formalizzato il lavoro di Brown da un punto di vista matematico (probabilistico). Infatti, la v.a. B_t rappresenta la posizione al tempo t di una particella sottoposta ad un moto casuale. La proprietà (B1) è solo una convenzione: all'istante iniziale la posizione della particella è nota e viene posta per definizione pari all'origine (qui stiamo studiano il caso reale, e dunque il moto avviene lungo la retta). La condizione (B2) esprime in qualche senso la proprietà di *perdita di memoria*: durante gli intervalli successivi $[t_0, t_1], \ldots, [t_{k-1}, t_k]$, la posizione iniziale B_{t_0} e gli incrementi successivi $B_{t_1} - B_{t_0}, \ldots, B_{t_k} - B_{t_{k-1}}$ non si influenzano. Il fatto poi che, come specificato in (B3), la media di $B_t - B_s$ sia zero evidenzia che non c'è una direzione privilegiata: in media la particella non si muove, a partire da qualsiasi istante (si veda la (20)). Inoltre, la varianza degli incrementi, ovvero dello spostamento della particella a partire da un istante prefissato, è direttamente proporzionale al tempo in cui viene osservato il moto.

La distribuzione dagli incrementi $B_t - B_s$ dipende solo dall'ampiezza $t-s$ dell'intervallo di osservazione; per tale ragione il moto browniano è un processo a **incrementi stazionari**.

Vediamo ora alcune proprietà immediate del moto browniano. Per ogni t,

$$\mathbb{E}[B_t] = 0 \qquad \mathrm{Var}(B_t) = t \tag{20}$$

Infatti, dalla (B3), con $s = 0$, segue che $B_t \sim \mathrm{N}(0, t)$. Inoltre, le distribuzioni finito dimensionali sono gaussiane multivariate:

$$\mu_{t_1,\ldots t_k}(H) = \int_H f_{t_1,\ldots t_k}(\xi) d\xi$$

[22]Qui, $\mathcal{T} = [0, +\infty)$.

[23]Ovvero, per ogni $H_0, H_1, \ldots, H_k$, $P(B_{t_0} \in H_0, B_{t_1} - B_{t_0} \in H_1, \ldots, B_{t_k} - B_{t_{k-1}} \in H_k) = P(B_{t_0} \in H_0)P(B_{t_1} - B_{t_0} \in H_1) \cdots P(B_{t_k} - B_{t_{k-1}} \in H_k)$.

dove $\xi = (\xi_1, \ldots \xi_k)$ e

$$f_{t_1,\ldots t_k}(\xi) = \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} e^{-\frac{(\xi_i - \xi_{i-1})^2}{2(t_i - t_{i-1})}} \tag{21}$$

(si prenda, nella formula precedente, $t_0 = \xi_0 = 0$). Poichè le distribuzioni finito dimensionali sopra scritte verificano le proprietà di consistenza (C1) e (C2), i Teoremi 3.2 e 3.3 assicurano dunque l'esistenza del moto browniano ed inoltre la possibilità di costruire tale processo sullo spazio delle traiettorie $\mathbb{R}^{[0,+\infty)} = \{x : [0, +\infty) \to \mathbb{R}\}$.

Tali teoremi non assicurano però la continuità delle traiettorie del moto browniano: è infatti possibile dimostrare che il sottoinsieme $\mathcal{C}([0, +\infty))$ di $\mathbb{R}^{[0,+\infty)}$ delle traiettorie continue non appartiene alla $\sigma-$algebra di riferimento $\mathcal{B}^{[0,+\infty)}$. Ciò significa, ad esempio, che non è possibile calcolare la probabilità che il moto browniano sia continuo, cioè $P(B \in \mathcal{C}([0, +\infty)))$.

D'altro canto il moto browniano rappresenta il moto di una particella ed è naturale richiedere che le sue traiettorie (che danno lo spostamento, per ogni ω fissato) siano continue. Esistono però altre costruzioni che garantiscono la continuità del moto browniano. Utilizzeremo qui il seguente risultato:

Teorema 3.5 (Teorema di continuità di Kolmogorov) *Sia $(X_t)_{t\in\mathcal{T}}$ un processo stocastico tale che esistono delle costanti positive α, β, c per le quali*

$$\mathbb{E}\left[|X_t - X_s|^\beta\right] \le c\,|t - s|^{1+\alpha}$$

per ogni $t, s \in \mathcal{T}$. Allora esiste una versione di $(X_t)_{t\in\mathcal{T}}$ a traiettorie continue.

Poiché (è facile vedere che)

$$\mathbb{E}\left[|B_t - B_s|^4\right] = 3\,|t - s|^2$$

applicando il Teorema 3.5 (con $\beta = 4, \alpha = 1, c = 3$), segue immediatamente

Teorema 3.6 *Esiste una versione del moto browniano $(B_t)_{t\ge 0}$ a traiettorie continue.*

D'ora in poi faremo sempre riferimento a moti browniani continui.

È interessante però osservare che le traiettorie di un moto browniano, nonostante siano continue, sono molto irregolari:

Teorema 3.7 *Esiste un insieme N di probabilità nulla tale che per ogni $\omega \notin N$ la traiettoria associata*

$$t \mapsto B_t(\omega)$$

non è differenziabile in nessun istante t.

Faremo ora alcune considerazioni che mostrano tale risultato solo in grandi linee e che sono interessanti perché mostrano alcune ulteriori proprietà del moto browniamo.

Si noti che la non derivabilità di una funzione continua x implica che la convergenza di x_s a x_t per $s \to t$ è troppo lenta oppure molto irregolare.

Prima di tutto, introduciamo i processi $(\widetilde{B}_t)_{t\geq 0}$ e $(\widehat{B}_t)_{t\geq 0}$:

$$\widetilde{B}_t = \begin{cases} tB_{\frac{1}{t}} & t > 0 \\ 0 & t = 0 \end{cases} \tag{22}$$

e, per $s \geq 0$ fissato,

$$\widehat{B}_t = B_{t+s} - B_s \tag{23}$$

È facile vedere che $(\widetilde{B}_t)_{t\geq 0}$ e $(\widehat{B}_t)_{t\geq 0}$ sono ancora moti browniani continui. La verifica riguardante $(\widehat{B}_t)_{t\geq 0}$ è immediata (esercizio!). Per quanto riguarda $\widetilde{B}$, possiamo dire:

(B1) $\widetilde{B}_0 = 0$, per definizione;

(B3)

$$\widetilde{B}_t - \widetilde{B}_s = tB_{\frac{1}{t}} - sB_{\frac{1}{s}} = tB_{\frac{1}{t}} - sB_{\frac{1}{t}} + sB_{\frac{1}{t}} - sB_{\frac{1}{s}} = (t-s)B_{\frac{1}{t}} - s\left(B_{\frac{1}{s}} - B_{\frac{1}{t}}\right)$$

dunque[24] $\widetilde{B}_t - \widetilde{B}_s$ è una v.a. gaussiana, di media 0 e varianza $(t-s)^2\dfrac{1}{t} + s^2\left(\dfrac{1}{s} - \dfrac{1}{t}\right) = t-s$.

(B2) Segue da proprietà caratteristiche dei processi gaussiani.

Occorre ora far vedere che $(\widetilde{B}_t)_{t\geq 0}$ è un processo a traiettorie continue. Ovviamente, l'unica cosa da mostrare è che

$$\lim_{t\to 0} tB_{\frac{1}{t}} = 0, \quad \text{q.c.}$$

Tale proprietà è diretta conseguenza del

Teorema 3.8 (Legge del logaritmo iterato) *Sia B un moto browniano. Allora,*

$$\limsup_{u\to+\infty} \frac{B_u}{\sqrt{u \log\log u}} = \sqrt{2} \qquad \liminf_{u\to+\infty} \frac{B_u}{\sqrt{u \log\log u}} = -\sqrt{2} \qquad q.c.$$

Ciò significa che $B_u \sim \sqrt{u \log\log u}$ per $u \to +\infty$[25]. Dunque, posto $u = \frac{1}{t}$, allora $u \to +\infty$ per $t \to 0^+$ e

$$tB_{\frac{1}{t}} = \frac{1}{u}B_u \sim \sqrt{\frac{\log\log u}{u}} \to 0 \quad \text{se } u \to +\infty$$

Mostriamo ora che $t \mapsto B_t(\omega)$ non è derivabile in ogni t. Cominciamo a studiare il caso $t = 0$. Il rapporto incrementale è $\frac{B_t}{t}$. Posto $u = \frac{1}{t}$, $u \to \infty$ quando $t \to 0^+$ e quindi, usando il Teorema 3.8 applicato al moto browniano $\widetilde{B}$, otteniamo

$$\frac{B_t}{t} = uB_{\frac{1}{u}} = \widetilde{B}_u \sim \sqrt{u \log\log u} \to \infty$$

[24] Si sta usando il seguente risultato: se Z e W sono due v.a. gaussiane indipendenti, allora $aZ + bW$ è ancora gaussiana, di media $a\mathbb{E}[Z] + b\mathbb{E}[W]$ e di varianza $a^2\mathrm{Var}(Z) + b^2\mathrm{Var}(W)$. Qui, $Z = B_{\frac{1}{t}}$, $W = B_{\frac{1}{s}} - B_{\frac{1}{t}}$, $a = t - s$, $b = -s$.

[25] Ovvero, per quasi ogni ω fissato, cioè a meno di un insieme di probabilità nulla, esistono due costanti c_1, c_2 tali che $c_1\sqrt{u \log\log u} \leq B_u \leq c_2\sqrt{u \log\log u}$ per ogni u grande.

Per mostrare la non derivabilità in un generico istante $s > 0$ useremo il moto browniano $\widehat{B}$ definito in (23):

$$\frac{B_{t+s} - B_s}{t} = \frac{\widehat{B}_t}{t}$$

che, per $t \to 0$, non converge perché $\widehat{B}$ è un moto browniano che abbiamo appena dimostrato non essere derivabile in 0.