Alcuni appunti per il corso di Metodi Probabilistici per l'Economia e la Finanza

Giovanna Nappo

A.A. 2003/04

versione del 3 novembre 2003

1 ESEMPI DI SPAZI DI PROBABILITÀ

Come dovrebbe essere noto uno spazio di probabilità è una terna $(\Omega, \mathcal{F}, \mathbb{P})$, dove

$\mathcal{F}$ è una ***σ-algebra***, ovvero $\mathcal{F}$ è una famiglia di sottoinsiemi di Ω, cioè $\mathcal{F}$ è un sottoinsieme di $\mathcal{P}(\Omega)$, tale che

$$\Omega \in \mathcal{F}; \tag{1}$$

$$\text{se } A \in \mathcal{F}, \text{ allora } A^c \in \mathcal{F}; \tag{2}$$

$$\text{se } A_n \in \mathcal{F}, n \in \mathbb{N}, \text{ allora } \cup_{n\in\mathbb{N}} A_n \in \mathcal{F}; \tag{3}$$

$\mathbb{P}$ è una ***misura di probabilità***, ovvero

$$\mathbb{P} : \mathcal{F} \mapsto [0,1]; \quad A \mapsto \mathbb{P}(A)$$

con le proprietà che

$$\mathbb{P}(\Omega) = 1; \tag{4}$$

$$\text{se } A_n \in \mathcal{F}, n \in \mathbb{N}, \text{ con } A_n \cap A_m = \emptyset \text{ per } n \neq m, \tag{5}$$
$$\text{allora } \mathbb{P}\big(\bigcup_{n\in\mathbb{N}} A_n\big) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n).$$

La σ-algebra $\mathcal{F}$ rappresenta l'informazione disponibile, ovvero gli unici eventi di cui abbiamo la possibilità di verificare se si sono verificati oppure no sono gli eventi appartenenti a $\mathcal{F}$.

Vediamo ora alcuni esempi elementari di spazi di probabilità:

Esempio 1.1. *Qualunque sia* Ω, *la* σ*-algebra banale* $\mathcal{F} = \{\emptyset, \Omega\}$ *è una* σ*-algebra, e necessariamente* $\mathbb{P}(\Omega) = 1$ *e* $\mathbb{P}(\emptyset) = 0$.

Esempio 1.2. *Qualunque sia* Ω*, preso un sottoinsieme proprio* A *di* Ω *la* σ*-algebra* $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$ *è una* σ*-algebra, e necessariamente* $\mathbb{P}(\Omega) = 1$*,* $\mathbb{P}(\emptyset) = 0$*,* $\mathbb{P}(A) = p$*,* $\mathbb{P}(A^c) = 1 - p$*, per un* $p \in [0, 1]$.

Esempio 1.3. *Qualunque sia* Ω*, sia* $\{H_m,\ m = 1, 2, \ldots, N\}$ *una* **partizione finita** *di* Ω*, cioè se*

$$H_n \cap H_m = \emptyset \ \textit{per}\ n \neq m,\ n, m \in \{1, 2, \ldots, N\}$$
$$\bigcup_{m=1}^{N} H_m = \Omega,$$

allora la famiglia $\mathcal{M} = \{A = \bigcup_{m \in I} H_m,\ \textit{al variare di}\ I \subset \{1, 2, \ldots, N\}\}$*, (con la convenzione che* $\bigcup_{m\in\emptyset} H_m = \emptyset$*) è una* σ*-algebra. Inoltre se* $p_1, p_2, \ldots, p_N$ *sono numeri non negativi, somma* 1*, ovvero*

$$p_m \geq 0,\ m = 1, 2, \ldots, N, \qquad \sum_{m=1}^{N} p_m = 1,$$

allora $\mathbb{P} :\ \mathcal{M} \mapsto [0, 1];\ A \mapsto \mathbb{P}(A)$*, con*

$$\mathbb{P}(A) = \sum_{m \in I} p_m, \qquad \textit{per}\ A = \bigcup_{m \in I} H_m, \tag{6}$$

definisce una probabilità su $(\Omega, \mathcal{M})$.

Esempio 1.4. *Le proprietà dell'esempio precedente valgono anche nel caso di una* **partizione numerabile** $\{H_m,\ m \in \mathbb{N}\}$ *con i dovuti cambiamenti: cioè, se*

$$H_n \cap H_m = \emptyset \ \textit{per}\ n \neq m,\ n, m \in \mathbb{N},$$
$$\bigcup_{m \in \mathbb{N}} H_m = \Omega,$$

allora la famiglia

$$\mathcal{M} = \{A = \bigcup_{m \in I} H_m,\ \textit{al variare di}\ I \subset \mathbb{N}\},$$

(con la convenzione che $\bigcup_{m\in\emptyset} H_m = \emptyset$*), è una* σ*-algebra*[1].

Inoltre se $p_1, p_2, \ldots, p_m, \ldots$ *sono numeri non negativi, somma* 1*, ovvero*

$$p_m \geq 0,\ m \in \mathbb{N}, \qquad \sum_{m \in \mathbb{N}} p_m = 1,$$

[1] La verifica è banale:

$$\Omega = \bigcup_{m \in \mathbb{N}} H_m,\ \text{ovvero}\ I = \mathbb{N}$$
$$\text{se}\ A = \bigcup_{m \in I} H_m,\ \text{allora}\ A^c = \bigcup_{m \in I^c} H_m$$
$$\text{se}\ A_n = \bigcup_{m \in I_n} H_m,\ n \geq 1,\ \text{allora}\ \bigcup_{n=1}^{\infty} A = \bigcup_{m \in I} H_m,\ \text{per}\ I = \cup_{n=1}^{\infty} I_n.$$

allora $\mathbb{P}:\ \mathcal{M} \mapsto [0,1];\ A \mapsto \mathbb{P}(A)$, *con*

$$\mathbb{P}(A) = \sum_{m \in I} p_m, \qquad \textit{per } A = \bigcup_{m \in I} H_m, \tag{7}$$

definisce una probabilità su $(\Omega, \mathcal{F})$[2].

Elenchiamo adesso alcune ***proprietà e notazioni relative alle σ-algebre***:

1 ***l'intersezione di σ-algebre è una σ-algebra***
Sia $\{\mathcal{G}_\alpha, \alpha \in \Lambda\}$ una famiglia di σ-algebre, allora

$$\mathcal{F} := \bigcap_{\alpha \in \Lambda} \mathcal{G}_\alpha$$

è una σ-algebra[3].

2 ***l'unione di σ-algebre non è (in generale) una σ-algebra***
Basta mostrare con un controesempio che l'unione di due σ-algebre non è una σ-algebra: ad esempio se $\mathcal{G}_i = \{\emptyset, A_i, A_i^c, \Omega\}$, con $A_1 \cap A_2 \neq \emptyset, A_1, A_2$, allora $\mathcal{G}_1 \cup \mathcal{G}_2 = \{\emptyset, A_1, A_2, A_1^c, A_2^c, \Omega\}$ non è una σ-algebra.

3 ***la σ-algebra generata da una collezione di eventi***
Sia $\mathcal{K}$ un sottoinsieme di $\mathcal{P}(\Omega)$, l'insieme delle parti di Ω, allora

$$\sigma(\mathcal{K}) := \bigcap_{\mathcal{G}:\mathcal{K} \subseteq \mathcal{G}}$$

è la σ-algebra[4] generata da $\mathcal{K}$.

In particolare quindi la σ-algebra $\mathcal{M}$, generata dalla partizione $\{H_m;\ m \in \mathbb{N}\}$ come nell'esempio 1.4, coincide con $\sigma(\{H_m;\ m \in \mathbb{N}\})$, in quanto, come già visto $\mathcal{M}$ è una σ-algebra, e inoltre ogni σ-algebra che contenga $\{H_m;\ m \in \mathbb{N}\}$, deve necessariamente contenere tutte le unioni del tipo $\cup_{m \in I} H_m$.

---

[2] La funzione $\mathbb{P}: \mathcal{M} \mapsto [0,1]$ definita in (7) è una probabilità, infatti

$$\mathbb{P}(\Omega) = \sum_{m \in \mathbb{N}} p_m = 1,$$

$$\text{se } A_n = \bigcup_{m \in I_n} H_m \in \mathcal{M}, n \in \mathbb{N},\ \text{con } A_n \cap A_{n'} = \emptyset \text{ per } n \neq n',$$

$$\text{allora } \bigcup_{n \in \mathbb{N}} A_n = \bigcup_{m \in I} H_m \text{ con } I = \bigcup_{n \in \mathbb{N}} I_n,\ \text{e con } I_n \cap I_{n'} = \emptyset \text{ per } n \neq n',$$

$$\text{e quindi } \mathbb{P}\Big(\bigcup_{n \in \mathbb{N}} A_n\Big) = \mathbb{P}(A) = \sum_{\ell \in I} p_\ell = \sum_{n \in \mathbb{N}} \sum_{m \in I_n} p_m = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n),$$

[3] La verifica è banale:

$\Omega \in \mathcal{F},$ in quanto $\Omega \in \mathcal{G}_\alpha,$ per ogni $\alpha \in \Lambda;$

se $A \in \mathcal{F},$ cioè se $A \in \mathcal{G}_\alpha,$ per ogni $\alpha \in \Lambda,$ allora $A^c \in \mathcal{G}_\alpha,$ per ogni $\alpha \in \Lambda,$ e quindi $A^c \in \mathcal{F};$

se $A_n \in \mathcal{F}, n \in \mathbb{N}$ cioè se $A_n \in \mathcal{G}_\alpha,$ per ogni $\alpha \in \Lambda, n \in \mathbb{N}$ allora $\displaystyle\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{G}_\alpha,$ per ogni $\alpha \in \Lambda,$ e quindi $\displaystyle\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F};$

[4] Il fatto che $\bigcap_{\mathcal{G}:\mathcal{K} \subseteq \mathcal{G}}$ sia una σ-algebra, deriva dalla proprietà che l'intersezione di σ-algebre è una σ-algebra.

4 ***la σ-algebra generata da una collezione di σ-algebre***
Nel caso in cui $\mathcal{K} = \bigcup_{\alpha\in\Lambda} \mathcal{G}_\alpha$, dove $\mathcal{G}_\alpha$ sono σ-algebre, allora si pone

$$\bigvee_{\alpha\in\Lambda} \mathcal{G}_\alpha := \sigma(\bigcup_{\alpha\in\Lambda} \mathcal{G}_\alpha).$$

In particolare se $\mathcal{M} = \sigma(\{H_m;\ m \in \mathbb{N}\})$ e $\mathcal{N} = \sigma(\{K_\ell;\ \ell \in \mathbb{N}\})$, allora

$$\mathcal{M} \vee \mathcal{N} = \sigma(\{H_m \cap K_\ell;\ m \in \mathbb{N},\, \ell \in \mathbb{N}\}) = \big\{E = \bigcup_{(m,\ell)\in J} H_m \cap K_\ell;\ \text{con } J \subseteq \mathbb{N}\times\mathbb{N}\big\}.$$

5 ***la σ-algebra dei Boreliani*** Nel caso in cui $\mathcal{K} = \mathcal{A}$, la famiglia degli aperti di $\mathbb{R}^k$, allora

$$\mathcal{B}(\mathbb{R}^k) := \sigma(\mathcal{A})$$

è detta σ-algebra dei boreliani, o σ-algebra di Borel. (ogni elemento di $\mathcal{I}$ di $\mathcal{B}(\mathbb{R}^k)$ è detto ***boreliano***).

1.1 Variabili aleatorie

Definizione 1.1. *Dato uno spazio di probabilità* $(\Omega, \mathcal{F}, \mathbb{P})$[5]*, una* **variabile aleatoria reale** X *è una funzione* $\mathcal{F}$-**misurabile***, ovvero una funzione*

$$X : \Omega \mapsto \mathbb{R}; \quad \omega \mapsto X(\omega),$$

tale che la controimmagine di ogni aperto $\mathcal{O} \in \mathcal{A}$ sia un elemento di $\mathcal{F}$[6]*, cioè tale che $X^{-1}(\mathcal{O} := \{\omega \text{ tali che } X(\omega) \in \mathcal{O}\} \in \mathcal{F}$, per ogni aperto $\mathcal{O} \in \mathcal{A}$.*
Si dice anche che X è una **variabile aleatoria $\mathcal{F}$-misurabile**.
Una definizione analoga vale nel caso di variabili aleatorie multidimensionali $\boldsymbol{X} : \Omega \mapsto \mathbb{R}^k; \quad \omega \mapsto \boldsymbol{X}(\omega) = \big(X_1(\omega), \ldots, X_k(\omega)\big)$: basta sostituire $\mathbb{R}$ con $\mathbb{R}^k$.

Vediamo alcuni esempi di variabili aleatorie $\mathcal{F}$-misurabili, al variare della σ-algebra $\mathcal{F}$.

Esempio 1.5. *Se $\mathcal{F} = \{\emptyset, \Omega\}$, allora le uniche variabili aleatorie reali X $\mathcal{F}$-misurabili sono le costanti:*
Se $X : \Omega \mapsto \mathbb{R};\ \omega \mapsto X(\omega) = c$, allora $X^{-1}(\mathcal{O})$ è l'evento impossibile(=insieme vuoto $\emptyset$), se $c \notin \mathcal{O}$, oppure è l'insieme certo(=Ω), se $c \in \mathcal{O}$.
Viceversa se $X : \Omega \mapsto \mathbb{R};\ \omega \mapsto X(\omega)$ non è costante allora X assume almeno due valori c_1 e c_2 distinti (cioè esistono ω_i tale che $X(\omega_i = c_i$, per $i = 1, 2$). Quindi se $c_1 \in \mathcal{O}$, ma $c_2 \notin \mathcal{O}$, allora $\omega_1 \in X^{-1}(\mathcal{O})$, mentre $\omega_2 \notin X^{-1}(\mathcal{O})$, ovvero $\emptyset \subset X^{-1}(\mathcal{O}) \subset \Omega$ (dove le inclusioni sono in senso stretto), e quindi X non è $\mathcal{F}$-misurabile.

---

[5]In realtà basta che ci sia uno spazio ***probabilizzabile***, ovvero basta solo la coppia $(\Omega, \mathcal{F})$, mentre non è necessario specificare la misura di probabilità $\mathbb{P}$.

[6]Si noti l'analogia con la definizione di funzione continua $f : \mathbb{R}^k \mapsto \mathbb{R}^d$, come una funzione tale che le controimmagini di aperti sono aperti.

Si noti che l'esempio precedente mostra anche che tutte le variabili aleatorie costanti sono misurabili rispetto a qualunque σ-algebra ($\{\emptyset, \Omega\} \subset \mathcal{F}$, per ogni σ-algebra $\mathcal{F}$).

Esempio 1.6. *Sia $\{H_m,\, m \in \mathbb{N}\}$ una partizione numerabile, e sia $\mathcal{M}$ come nell'esempio 1.4. Allora $X : \Omega \mapsto \mathbb{R};\ \omega \mapsto X(\omega)$ è $\mathcal{M}$-misurabile, se e solo se esiste una successione di costanti $\{c_m,\, m \in \mathbb{N}\}$*[7]*, tale che*

$$X(\omega) = \sum_{m \in \mathbb{N}} c_m \mathbb{I}_{H_m}(\omega). \tag{8}$$

Se X è definita come in (8) allora X è $\mathcal{M}$-misurabile, infatti per ogni aperto $\mathcal{O}$,

$$X^{-1}(\mathcal{O}) = \bigcup_{m : c_m \in \mathcal{O}} H_m,$$

ovvero $X^{-1}(\mathcal{O}) = \bigcup_{m \in I} H_m \in \mathcal{M}$, per $I = \{m : c_m \in \mathcal{O}\}$.
Viceversa se X è $\mathcal{M}$-misurabile, cioè, per ogni aperto $\mathcal{O}$, esiste un $I \subseteq \mathbb{N}$ tale che

$$X^{-1}(\mathcal{O}) = \bigcup_{m \in I} H_m,$$

allora qualunque sia $c \in \mathbb{R}$, preso $\mathcal{O}^n$ l'intervallo aperto $(c - 1/n, c + 1/n)$ si ha che

$$X^{-1}(\{c\}) = \bigcap_n X^{-1}(\mathcal{O}^n) = \bigcap_n \bigcup_{m \in I^n} H_m = \bigcup_{m \in \bigcap_n I^n} H_m \in \mathcal{M},$$

Esempio 1.7. *Sia $X : \Omega \mapsto \mathbb{R};\ \omega \mapsto X(\omega)$, una funzione discreta, ovvero tale che l'immagine $X(\Omega) = \{x \in \mathbb{R},$ tali che esiste un ω con $X(\omega) = x\}$ di X sia un insieme numerabile (finito o infinito), cioè $X(\Omega) = \{x_m, m \in \mathbb{N}\}$, con $x_m \neq x_n$ per $m \neq m$. Allora*

$$X(\omega) = \sum_{m \in \mathbb{N}} x_m \mathbb{I}_{H_m}(\omega). \tag{9}$$

dove

$$H_m = X^{-1}(\{x_m\}).$$

Si noti che $\{H_m,\, m \in \mathbb{N}\}$ forma una partizione numerabile.

La funzione X è una variabile aleatoria **semplice** *o* **elementare**, **$\mathcal{F}$-misurabile**, *se e solo se*

$$H_m = X^{-1}(\{x_m\}) \in \mathcal{F}, \text{ per ogni } m \in \mathbb{N},$$

come è immediato da (9), osservando che, come nel caso precedente,

$$X^{-1}(\mathcal{O}) = \bigcup_{m : x_m \in \mathcal{O}} H_m.$$

Si può dimostrare che

[7]Si noti che non si assume che i valori di $\{c_m\}$ siano tutti distinti, ad esempio nel caso della successione costante, cioè $c_m = c$ per ogni $m \in \mathbb{N}$, si trova una variabile aleatoria costante.

1 se X è una variabile aleatoria $\mathcal{F}$-misurabile, allora la controimmagine $X^{-1}(\mathcal{I}) \in \mathcal{F}$, per ogni boreliano $\mathcal{I} \in \mathcal{B}(\mathbb{R})$,

2 la variabile aleatoria $\boldsymbol{X}$ è $\mathcal{F}$-misurabile, se e solo se ciascuna componente X_i è $\mathcal{F}$-misurabile[8], per ogni $i = 1, \ldots, k$. In particolare $X^{-1}(\{x\}) \in \mathcal{F}$, per ogni $x \in \mathbb{R}$, in quanto $\{x\} = bigcap_n(x - 1/n, x + 1/n)$.

Connessa con la precedente Definizione 1.1 è la seguente definizione:

Definizione 1.2. *Sia data una funzione* $\boldsymbol{X} : \Omega \mapsto \mathbb{R}^k; \quad \omega \mapsto \boldsymbol{X}(\omega) = \big(X_1(\omega), \ldots, X_k(\omega)\big)$. *Si dice* **$\boldsymbol{\sigma}$-algebra generata da $\boldsymbol{X}$**, *la* σ-*algebra*

$$\sigma(\boldsymbol{X}) = \bigcap_{\mathcal{G} \in \mathcal{R}^{\boldsymbol{X}}} \mathcal{G}$$

dove $\mathcal{R}^{\boldsymbol{X}}$ *è la famiglia delle* σ-*algebre, per le quali* $\boldsymbol{X}$ *è* $\mathcal{G}$-*misurabile*[9].

Si dimostra che

3 la funzione $\boldsymbol{X}$ è $\mathcal{F}$-misurabile, se e solo se $\sigma(\boldsymbol{X}) \subseteq \mathcal{F}$,

4 La σ-algebra generata da $\boldsymbol{X}$, si può caratterizzare come:

$$\sigma(\boldsymbol{X}) = \{A = \boldsymbol{X}^{-1}(\mathcal{I}), \text{ per } \mathcal{I} \in \mathcal{B}(\mathbb{R}^k)\},$$

5 le variabili aleatorie $\sigma(\boldsymbol{X})$-misurabili sono tutte e sole le variabili aleatorie Z per le quali esiste una funzione g boreliana[10] tale che

$$Z = g(\boldsymbol{X}).$$

Esempio 1.8. *Sia* X *una funzione semplice, come in Esempio 1.7, allora*

$$\sigma(X) = \sigma(\{H_m,\, m \in \mathbb{N}\}) = \{A = \bigcup_{m \in I} H_m;\ I \subseteq \mathbb{N}\},$$

dove $H_m = X^{-1}(\{x_m\})$.

Inoltre tutte e sole le variabili aleatorie $\sigma(X)$ misurabili sono le funzioni

$$Z : \Omega \mapsto \mathbb{R};\ \omega \mapsto Z(\omega) := \sum_m c_m \mathbb{I}_{H_m},$$

[8]Dimostriamo solo la necessità, che è immediata: basta prendere $\mathcal{O} = \underbrace{\mathbb{R} \times \cdots \times \mathbb{R}}_{i-1 \text{ volte}} \times \mathcal{O}_i \times \underbrace{\mathbb{R} \times \cdots \times \mathbb{R}}_{k-i \text{ volte}}$.

[9]La famiglia $\mathcal{R}^{\boldsymbol{X}}$ non è vuota, in quanto contiene almeno $\mathcal{G} = \mathcal{P}(\Omega)$, l'insieme delle parti di Ω.

[10]Una funzione $g : \mathbb{R}^k \mapsto \mathbb{R}^d$, si dice boreliana se è una funzione tale che le controimmagini di aperti sono boreliani.
Ovviamente le funzioni continue sono boreliane. Sono boreliane anche le funzioni continue a tratti, o meglio ancora costanti a tratti.
Per chi non avesse familiarità con i concetti di misurabilità può pensare a queste funzioni, o a funzioni che siano limite puntuale di funzioni di uno dei due tipi precedenti.

come discende immediatamente dall'Esempio 1.6. Di conseguenza se $g : \mathbb{R} \mapsto \mathbb{R}$ tale che $g(x_m) = c_m$, per ogni $m \in \mathbb{N}$, allora

$$Z(\omega) := \sum_m c_m \mathbb{I}_{H_m} = Z(\omega) = \sum_m g(x_m)\mathbb{I}_{X^{-1}(\{x_m\})}(\omega) = \sum_m g(x_m)\mathbb{I}_{\{x_m\}}(X(\omega) = g(X(\omega)).$$

Terminiamo questa sezione, ricordando che le operazioni di massimo, minimo, somma, prodotto, di due funzioni misurabili, danno luogo a funzioni misurabili: quindi se X ed Y sono variabili aleatorie $\mathcal{F}$-misurabili, lo sono anche $X \vee Y = \max(X, Y)$, $X \wedge Y = \min(X, Y)$, $X + Y$, XY. In particolare sono variabili aleatorie $X^+ := X \vee 0$ e $X^-+ := (-X) \vee 0$.

1.2 Valori attesi

Definizione 1.3 (Valore atteso per variabili semplici). *Sia X una variabile aleatoria in $(\Omega, \mathcal{F}, \mathbb{P})$, non negativa e semplice, cioè come in Esempio 1.7,*

$$X(\omega) = \sum_{m \in \mathbb{N}} x_m \mathbb{I}_{H_m}(\omega), \quad \textit{con } H_m \in \mathcal{F} \textit{ per ogni } m \in \mathbb{N},$$

allora si definisce

$$\mathbb{E}[X] = \sum_{m \in \mathbb{N}} x_m \mathbb{P}(H_m).$$

Osservazione 1.1. *Ogni variabile aleatoria X in $(\Omega, \mathcal{F}, \mathbb{P})$, non negativa, ammette una successione di variabili aleatorie X_n, semplici e non negative, tali che*

$$0 \le X_n(\omega) \le X_{n+1}(\omega), \quad \textit{e tali che } \lim_{n \to \infty} X_n(\omega) = X(\omega).$$

Infatti[11] *basta prendere*

$$X_n(\omega) = \sum_{m=0}^{n2^n - 1} \frac{m}{2^n} \mathbb{I}_{H_m}(\omega) + n\mathbb{I}_{H_{n2^n}} = \sum_{m=0}^{n2^n - 1} \frac{m}{2^n} \mathbb{I}_{[\frac{m}{2^n}, \frac{m+1}{2^n})}(X(\omega)) + n\mathbb{I}_{[n,\infty)}(X(\omega)), \tag{10}$$

ovvero con

$$H_m = X^{-1}\left(\left[\frac{m}{2^n}, \frac{m+1}{2^n}\right)\right) \in \mathcal{F} \textit{ per } 0 \le m \le n2^n - 1, \quad H_{n2^n} = X^{-1}\big([n, \infty)\big).$$

Definizione 1.4 (Valore atteso per variabili nonnegative). *Sia X una variabile aleatoria in $(\Omega, \mathcal{F}, \mathbb{P})$, non negativa, si definisce*

$$\mathbb{E}[X] = \lim_{n \to \infty} \mathbb{E}[X_n],$$

dove $\{X_n;\ n \in \mathbb{N}\}$ è la successione monotona definita come in (10) dell'Osservazione precedente. Il limite esiste ed è monotono, per la proprietà di monotonia del valore atteso, sulle variabili aleatorie semplici. Si noti bene che tale limite può valere anche $+\infty$, nel qual caso si dice che la variabile X ha valore atteso infinito.

[11] La monotonia della successione delle variabili aleatorie X_n è evidente: se $X_n(\omega) = m/2^n$, allora i soli casi possibili sono $X_{n+1}(\omega) = (2m)/2^{n+1} = m/2^n = X_n(\omega)$, oppure $X_{n+1}(\omega) = (2m+1)/2^{n+1} = m/2^n + 1/2^{n+1} > X_n(\omega)$.
Per la convergenza basta osservare che, qualunque sia ω, pur di prendere n sufficientemente grande e in modo che $X(\omega) < n$, si ha che $0 \le X(\omega) - X_n(\omega) \le 1/2^n$.

Arriviamo ora alla definizione generale del valore atteso:

Definizione 1.5 (Valore atteso per variabili generali). *Sia X una variabile aleatoria in $(\Omega, \mathcal{F}, \mathbb{P})$, Siano X^+ ed X^-, le variabili aleatorie non negative, definite alla fine della sezione precedente. Si noti che $X = X^+ - X^-$ e che invece $|X| = X^+ + X^-$. Si definisce allora, se ha senso* [12]

$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-],$$

2 MEDIE E PROBABILITÀ CONDIZIONALI

Sia $(\Omega, \mathcal{F}, \mathbb{P})$ uno spazio di probabilità e sia X una variabile aleatoria. Si supponga di avere una sotto σ-algebra $\mathcal{G} \subseteq \mathcal{F}$. Cerchiamo una variabile aleatoria $\widetilde{X}$ che sia $\mathcal{G}$-misurabile e che "in qualche senso" abbia un comportamento simile a X. Vale la pena di ricordare che una σ-algebra può essere interpretata come "informazione" disponibile, e quindi cercare una variabile $\widetilde{X}$ che sia $\mathcal{G}$-misurabile con un comportamento simile ad X, significa cercare una variabile aleatoria che, sulla base dell'informazione disponibile $\mathcal{G}$, sia simile. Un'altro modo di definire questa variabile $\widetilde{X}$ consiste, più banalmente, nel richiedere che sia una variabile aleatoria $\mathcal{G}$-misurabile "vicina" ad X. Naturalmente è necessario definire il senso di vicinanza, cioè quale metrica mettere sullo spazio delle variabili aleatorie.

Diamo ora due pre-definizioni, in cui però mancano le ipotesi da fare su X e delle precisazioni, affinché risultino definizioni ben poste.

PRE-DEFINIZIONE 1. *Si cerca una variabile aleatoria $\widetilde{X}$, $\mathcal{G}$-misurabile, per la quale valga*

$$\mathbb{E}[X|A] = \mathbb{E}[\widetilde{X}|A], \qquad \forall A \in \mathcal{G}, \text{ con } \mathbb{P}(A) > 0. \tag{11}$$

dove

$$\mathbb{E}[X|A] = \frac{\mathbb{E}[XI_A]}{\mathbb{P}(A)}.$$

Si noti che (11) equivale a

$$\mathbb{E}[XI_A] = \mathbb{E}[\widetilde{X}I_A], \qquad \forall A \in \mathcal{G}, \text{ con } \mathbb{P}(A) > 0. \tag{12}$$

e che quindi la richiesta che $\mathbb{P}(A) > 0$ si può omettere.

PRE-DEFINIZIONE 2. *Si cerca una variabile aleatoria $\widetilde{X}$, $\mathcal{G}$-misurabile, per la quale valga*

---

[12] Si considera che la somma $\mathbb{E}[X^+] - \mathbb{E}[X^-]$ ha senso

1 se $\mathbb{E}[X^+] < \infty$, $\mathbb{E}[X^-] < \infty$, nel qual caso $\mathbb{E}[X] \in \mathbb{R}$ e inoltre si ha anche $\mathbb{E}[|X|] = \mathbb{E}[X^+] + \mathcal{E}[X^-] < \infty$;

2 se $\mathbb{E}[X^+] < \infty$, $\mathbb{E}[X^-] = \infty$, nel qual caso $\mathbb{E}[X] = -\infty$;

3 se $\mathbb{E}[X^+] = \infty$, $\mathbb{E}[X^-] < \infty$, nel qual caso $\mathbb{E}[X] = +\infty$;

Il caso che rimane escluso è quindi il caso in cui $\mathbb{E}[X^+] = \infty$, $\mathbb{E}[X^-] = \infty$, del resto si avrebbe la forma indeterminata di $\infty - infty$.

$$\mathbb{E}[(X-Z)^2] \geq \mathbb{E}[(X-\widetilde{X})^2], \qquad \forall Z \ \mathcal{G}-\text{misurabile}. \tag{13}$$

Prima di tutto, dobbiamo trovare sotto quali condizioni le pre-definizioni siano ben poste, cioè, in questo caso, che esista una variabile $\widetilde{X}$ per cui valga la (11) (o equivalentemente la (12)) oppure valga la (13), ed in che senso ne viene individuata una sola. Notiamo che intanto ci sono delle condizioni necessarie da rispettare: chiaramente, per la pre-definizione 1, è necessario che X sia una variabile aleatoria integrabile[13], cioè $\mathbb{E}[|X|] < \infty$, mentre, per la pre-definizione 2, è necessario richiedere che X sia di quadrato integrabile, cioè $\mathbb{E}[|X|^2] < \infty$ [14], ed inoltre anche la (13) va modificata, nel senso che è necessario richiedere che anche Z sia di quadrato integrabile, oltre che $\mathcal{G}$-misurabile[15]. Inoltre è chiaro che se $\widetilde{X}'$ è una variabile aleatoria $\mathcal{G}$-misurabile, che differisce da $\widetilde{X}$ a meno di un insieme di misura nulla, anche $\widetilde{X}'$ gode della proprietà (12) o (13) rispettivamente e quindi non si individua una sola variabile aleatoria, ma una classe di variabili aleatorie. Queste modifiche in realtà sono sufficienti a garantire che le due pre-definizioni diventino due definizioni.

Definizione 2.1 (valore atteso condizionale 1). *Sia X una variabile aleatoria integrabile, cioè $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Sia data una variabile aleatoria $\widetilde{X}$, $\mathcal{G}$-misurabile, per la quale valga*

$$\mathbb{E}[XI_A] = \mathbb{E}[\widetilde{X}I_A], \qquad \forall A \in \mathcal{G}. \tag{14}$$

In questo modo si individua univocamente una classe di funzioni che si indica con $\mathbb{E}[X \mid \mathcal{G}]$ e che si chiama anche **media condizionale** *(o* **condizionata***) di X data $\mathcal{G}$. Si dice inoltre che $\widetilde{X}$ è una versione di $\mathbb{E}[X \mid \mathcal{G}]$.*

Definizione 2.2 (valore atteso condizionale 2). *Sia X una variabile aleatoria di quadrato integrabile, cioè $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. Sia data una variabile aleatoria $\widetilde{X}$, $\mathcal{G}-$ misurabile, per la quale valga*

$$\mathbb{E}[(X-Z)^2] \geq \mathbb{E}[(X-\widetilde{X})^2], \forall Z \quad \mathcal{G}-\textit{misurabile, con } \mathbb{E}[Z^2] < \infty. \tag{15}$$

[13] L'insieme delle variabili aleatorie X, che sono $\mathcal{F}$-misurabili ed integrabili, cioè per le quali $\mathbb{E}[|X|] < \infty$, forma uno spazio vettoriale reale: se $\mathbb{E}[|X_i|] < \infty$, per $i = 1, 2$, con X_i $\mathcal{F}$-misurabili, allora la variabile aleatoria $a_1X_1 + a_2X_2$ è ancora $\mathcal{F}$-misurabile e inoltre

$$\mathbb{E}[|a_1X_1 + a_2X_2|] \leq \mathbb{E}[|a_1||X_1| + |a_2||X_2|] \leq |a_1|\mathbb{E}[|X_1|] + |a_2|\mathbb{E}[|X_2|] < \infty.$$

L'insieme di tali variabili aleatorie è indicato con $L^1(\Omega, \mathcal{F}, \mathbb{P})$.

[14] Anche in questo caso l'insieme delle variabili aleatorie X, che sono $\mathcal{F}$-misurabili e quadrato integrabili, cioè per le quali $\mathbb{E}[|X|^2] < \infty$, forma uno spazio vettoriale reale che si indica con $L^2(\Omega, \mathcal{F}, \mathbb{P})$, che inoltre è uno spazio di Hilbert (modulo passare a classi di equivalenza $X \sim Y$ se e solo se $\mathbb{E}[|X-Y|^2] = 0$). Questo significa che è possibile introdurre un prodotto scalare

$$< X_1, X_2 >:= \mathbb{E}(X_1X_2).$$

Si noti che, per la disuguaglianza di Cauchy,

$$|\mathbb{E}(X_1X_2)| \leq \mathbb{E}(|X_1X_2|) \leq \mathbb{E}^{1/2}(|X_1|^2)\mathbb{E}^{1/2}(|X_2|^2),$$

e quindi $< X_1, X_2 >$ è finito se $\mathbb{E}[|X_i|^2] < \infty$, per $i = 1, 2$.

[15] Si noti che quindi deve essere $Z \in L^2(\Omega, \mathcal{G}, \widehat{\mathbb{P}})$ dove $\widehat{\mathbb{P}} = \mathbb{P}|_{\mathcal{G}}$, cioè la restrizione di $\mathbb{P}$ a $\mathcal{G}$

Anche in questo modo si individua univocamente una classe di funzioni che si indica con $\mathbb{E}[X \mid \mathcal{G}]$ *e che si chiama anche* **media condizionale** *(o* **condizionata***) di* X *data* $\mathcal{G}$*. Si dice inoltre che* $\widetilde{X}$ *è una versione di* $\mathbb{E}[X \mid \mathcal{G}]$.

Rimandiamo la verifica che effettivamente la Definizione 2.1 e la Definizione 2.2 sono ben poste a dopo aver trattato alcuni esempi, e anticipiamo che, se X è di quadrato integrabile, allora le Definizioni 2.1 e 2.2 sono equivalenti, e quindi non c'è ambiguità nello scegliere una definizione o l'altra e che in seguito, riferendoci a $\mathbb{E}[X \mid \mathcal{G}]$, intenderemo riferirci alla Definizione 2.1, che valendo per variabili aleatorie in L^1, è più generale. Non c'è quindi ambiguità nella seguente definizione che viene data in analogia con $\mathbb{P}(A) = \mathbb{E}[I_A]$.

Definizione 2.3 (probabilità condizionale di un evento). *Sia* $A \in \mathcal{F}$ *un evento e* $\mathcal{G}$ *una sotto* $\sigma-$*algebra di* $\mathcal{F}$*, allora*

$$\mathbb{P}(A \mid \mathcal{G}) := \mathbb{E}[I_A \mid \mathcal{G}]$$

è detta **probabilità condizionale** *di* A *data* $\mathcal{G}$.

Va inoltre sottolineato ***un abuso di notazione*** per cui si identifica la classe di equivalenza $\mathbb{E}[X \mid \mathcal{G}]$ e la variabile aleatoria $\widetilde{X}$ che ne è un rappresentante.

Esempio 2.1. *Caso in cui* $\mathcal{G} = \{\Omega, \emptyset\}$.
In questo caso $\mathbb{E}[X \mid \mathcal{G}]$ *si riduce al valore medio usuale, cioè*

$$\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X],$$

in quanto ovviamente $\mathbb{E}[XI_\Omega] = \mathbb{E}[\mathbb{E}[X]I_\Omega] = \mathbb{E}[X] \cdot 1$*, mentre banalmente* $\mathbb{E}[XI_\emptyset] = \mathbb{E}[\mathbb{E}[X]I_\emptyset] = \mathbb{E}[X] \cdot 0$.

Esempio 2.2. *Caso in cui* $\mathcal{G} = \mathcal{M} = \sigma\{H_m,\, m \in \mathbb{N}\}$ *ed* $\{H_m,\, m \in \mathbb{N}\}$ *è una partizione.*

Intanto ricordiamo che in questo caso le variabili aleatorie $\mathcal{G}$*-misurabili sono le funzioni del tipo*

$$Z(\omega) = \sum_{m=1}^{\infty} c_m \mathbb{I}_{H_m}(\omega),$$

dove c_m *sono costanti reali, e gli insiemi* $\mathcal{G}$*-misurabili sono le unioni di sottofamiglie numerabili di elementi della partizione. Basta quindi calcolare i valori* $\widetilde{c}_m$ *che caratterizzano* $\widetilde{X}$*, imponendo la condizione*[16] *che*

$$\mathbb{E}[X\mathbb{I}_{H_n}] = \mathbb{E}[\widetilde{X}\mathbb{I}_{H_n}] = \mathbb{E}[\sum_{m=1}^{\infty} \widetilde{c}_m \mathbb{I}_{H_m}\mathbb{I}_{H_n}] = \mathbb{E}[\widetilde{c}_n \mathbb{I}_{H_n}] = \widetilde{c}_n \mathbb{P}[H_n] \tag{16}$$

[16] E' chiaro che la condizione che $\mathbb{E}[X\mathbb{I}_A] = \mathbb{E}[\widetilde{X}\mathbb{I}_A$ per ogni $A = \bigcup_{n \in I} H_n$, implica la condizione (16): basta prendere $A = H_n$.
Tuttavia vale anche il viceversa, in quanto

$$\begin{aligned} \mathbb{E}[X\mathbb{I}_A] &= \mathbb{E}[X \sum_{n \in I} \mathbb{I}_{H_n}] = \sum_{n \in I} \mathbb{E}[X\mathbb{I}_{H_n}] \\ &= \sum_{n \in I} \mathbb{E}[\widetilde{X}\mathbb{I}_{H_n}] = \mathbb{E}[\widetilde{X} \sum_{n \in I} \mathbb{I}_{H_n}] = \mathbb{E}[\widetilde{X}\mathbb{I}_A]. \end{aligned}$$

Quindi

$$\widetilde{c}_m = \frac{\mathbb{E}[XI_{H_m}]}{\mathbb{P}(H_m)} \quad \textit{se } \mathbb{P}(H_m) > 0 \qquad e \qquad \widetilde{X} = \mathbb{E}[X \mid \mathcal{G}] = \sum_{n\geq 1}^{*} \frac{\mathbb{E}[X\mathbb{I}_{H_n}]}{\mathbb{P}(H_n)} \mathbb{I}_{H_n},$$

dove $\sum_{n\geq 1}^{*}$ *è la somma estesa agli indici* n *per cui* $\mathbb{P}(H_n) > 0$.
Si noti l'abuso di notazione: in realtà

$$\mathbb{E}[X \mid \mathcal{G}] = \{\xi \textit{ v.a. } \mathcal{G} - \textit{misurabili, t.c. } \xi = \sum_{n\geq 1}^{*} \frac{\mathbb{E}[X\mathbb{I}_{H_n}]}{\mathbb{P}(H_n)} \mathbb{I}_{H_n} + \sum_{m\geq 1}^{**} c_m \mathbb{I}_{H_m}, \textit{ per } c_m \in \mathbb{R}\}$$

dove $\sum_{m\geq 1}^{**}$ *è la somma estesa agli indici* m *per cui* $\mathbb{P}(H_m) = 0$.

In particolare se $X = \mathbb{I}_B$, *con* $B \in \mathcal{F}$, *e ponendo (come è usuale)*

$$\mathbb{P}(B \mid \mathcal{G})(\omega) = \mathbb{E}[\mathbb{I}_B \mid \mathcal{G}](\omega),$$

otteniamo che una versione di $\mathbb{P}(B \mid \mathcal{G})(\omega)$ *è data da*

$$\mathbb{P}(B \mid \mathcal{G}) = \sum_{n\geq 1}^{*} \frac{\mathbb{E}[\mathbb{I}_B\mathbb{I}_{H_n}]}{\mathbb{P}(H_n)} \mathbb{I}_{H_n} = \sum_{n\geq 1}^{*} \frac{\mathbb{P}(B \cap H_n)}{\mathbb{P}(H_n)} \mathbb{I}_{H_n} = \sum_{n\geq 1}^{*} \mathbb{P}(B \mid H_n) \mathbb{I}_{H_n}. \tag{17}$$

Ritroviamo quindi forse più chiaramente l'idea che se si verifica H_n cambiamo la probabilità prendendo $\mathbb{P}(B \mid H_n)$ al posto di $\mathbb{P}(B)$, che consideriamo se non abbiamo alcuna informazione (corrisponde al caso in cui la σ-algebra a nostra disposizione è quella banale).
Vale la pena di considerare il caso in cui B sia un evento della σ-algebra $\mathcal{G}$ a nostra disposizione: $B \in \mathcal{G}$, o equivalentemente $B = \cup_{n\in I} H_n$, per un insieme di indici I. A parole si tratta del caso in cui B è un evento completamente osservabile, ovverosia il caso in cui B sia un evento che possiamo conoscere perfettamente. In tale caso una versione di $\mathbb{P}(B \mid \mathcal{G})$ è proprio I_B. Infatti $\mathbb{P}(B \mid H_n) = 1$ per $n \in I$ e $\mathbb{P}(H_n) > 0$, mentre $\mathbb{P}(B \mid H_n) = 0$ per $n \notin I$ e $\mathbb{P}(H_n) > 0$.

È interessante osservare che se $\mathcal{F}$ è a sua volta generato da una partizione $\{K_\ell,\, \ell \in \mathbb{N}\}$, più fine[17] di $\{H_m,\, m \in \mathbb{N}\}$, allora, nel caso in cui $\mathbb{P}(H_n) > 0$ per ogni $n \in \mathbb{N}$, (17) definisce una probabilità su $\mathcal{F}$: se $B = \bigcup_{\ell\in I_B} K_\ell$, allora

$$\begin{aligned}
\mathbb{P}(B \mid \mathcal{G})(\omega) &= \sum_{n\geq 1}^{*} \frac{\mathbb{E}[\mathbb{I}_B\mathbb{I}_{H_n}(\omega)]}{\mathbb{P}(H_n)} \mathbb{I}_{H_n} = \sum_{n\geq 1}^{*} \frac{\mathbb{P}(B \cap H_n)}{\mathbb{P}(H_n)} \mathbb{I}_{H_n}(\omega) \\
&= \sum_{n\geq 1}^{*} \sum_{\ell\in I_B} \frac{\mathbb{P}(K_\ell \cap H_n)}{\mathbb{P}(H_n)} \mathbb{I}_{H_n}(\omega) = \sum_{\ell\in I_B} \sum_{n\geq 1}^{*} \frac{\mathbb{P}(K_\ell \cap H_n)}{\mathbb{P}(H_n)} \mathbb{I}_{H_n}(\omega) = \sum_{\ell\in I_B} \mathbb{P}(K_\ell \mid \mathcal{G})(\omega).
\end{aligned}$$

Nel caso in cui **non** si faccia l'ipotesi che $\mathbb{P}(H_n) > 0$ per ogni $n \in \mathbb{N}$, si può trovare una versione di $\mathbb{P}(B \mid \mathcal{G})(\omega)$ in modo che per ogni ω la funzione

$$\mathbb{P}(\cdot \mid \mathcal{G})(\omega) : \mathcal{F} \mapsto [0,1];\ B \mapsto \mathbb{P}(B \mid \mathcal{G})(\omega)$$

[17]La partizione $\{K_\ell,\, \ell \in \mathbb{N}\}$ è più fine della partizione $\{H_m,\, m \in \mathbb{N}\}$ se e solo se per ogni $m \in \mathbb{N}$ esiste un $I_m \subset \mathbb{N}$ tale che

$$H_m = \bigcup_{\ell\in I_m} K_\ell.$$

sia una probabilità:
fissata a piacere una probabilità $\mathbb{P}^0$ su $\mathcal{F}$[18], si definisce

$$\mathbb{P}(K_\ell \mid \mathcal{G})(\omega) = \sum_{n\geq 1}^{*} \frac{\mathbb{P}(K_\ell \cap H_n)}{\mathbb{P}(H_n)} \mathbb{I}_{H_n}(\omega) + \sum_{n\geq 1}^{**} \mathbb{P}^0(K_\ell)\mathbb{I}_{H_n}(\omega) = \tag{18}$$

$$= \sum_{n\geq 1}^{*} \frac{\mathbb{P}(K_\ell \cap H_n)}{\mathbb{P}(H_n)} \mathbb{I}_{H_n}(\omega) + \mathbb{P}^0(K_\ell)\mathbb{I}_{\{\cup_{n\geq 1}^{**} H_n\}}(\omega). \tag{19}$$

e

$$\mathbb{P}(B \mid \mathcal{G})(\omega) := \sum_{\ell \in I_B} \mathbb{P}(K_\ell \mid \mathcal{G})(\omega). \tag{20}$$

Si vede immediatamente che in questo modo, qualunque sia ω, $\mathbb{P}(\cdot \mid \mathcal{G})(\omega)$ definisce una probabilità, e si vede facilmente che la parte a destra di (20) è effettivamente una versione di $\mathbb{P}(B \mid \mathcal{G})$. Inoltre questa probabilità gode della proprietà che se X è $\mathcal{F}$-misurabile, cioè se

$$X = \sum_{\ell \in \mathbb{N}} c_\ell \mathbb{I}_{K_\ell},$$

allora

$$\mathbb{E}[X \mid \mathcal{G}](\omega) = \sum_{\ell \in \mathbb{N}} c_\ell \mathbb{P}(K_\ell \mid \mathcal{G})(\omega) = \int_\Omega X(\omega') d\mathbb{P}(d\omega' \mid \mathcal{G})(\omega).$$

Per σ-algebre $\mathcal{F}$ più generali del caso di σ-algebre generate da una partizione, non è detto che queste proprietà valgano (per approfondimenti vedere la Sezione 4).

Esempio 2.3. *Caso in cui $\mathcal{G} = \sigma\{Y\}$, con Y una variabile aleatoria a valori in $\mathbb{R}^d$, e (X, Y) ammette densità di probabilità congiunta $f_{X,Y}(x, y)$. Allora, posto*

$$f_{X|Y}(x|y) = I_{\{z: f_Y(z)>0\}}(y) \frac{f_{X,Y}(x, y)}{f_Y(y)} + I_{\{z: f_Y(z)=0\}}(y) f_0(x)$$

dove $f_0(x)$ è una densità di probabilità prefissata, si ha

$$\widetilde{X}(\omega) = \mathbb{E}[X \mid Y](\omega) = \int_\mathbb{R} x \, f_{X|Y}(x|y)\Big|_{y=Y(\omega)} \, dx$$

ossia

$$\widetilde{X}(\omega) = \int_\mathbb{R} x \left(I_{\{z: f_Y(z)>0\}}(y) \frac{f_{X,Y}(x, y)}{f_Y(y)} \right)\Bigg|_{y=Y(\omega)} dx + \int_\mathbb{R} x \, I_{\{z: f_Y(z)>0\}}(y)\Big|_{y=Y(\omega)} \, dx,$$

dove $\mathbb{E}[X \mid Y]$ è una abbreviazione per $\mathbb{E}[X \mid \sigma(Y)]$.

Per la verifica è intanto importante notare che $\sigma\{Y\} = \{A = Y^{-1}(B),\ \text{per } B \in \mathcal{B}(\mathbb{R})\}$, quindi $I_A(\omega) = I_B(Y(\omega))$ e

[18] ad esempio fissando una successione $\{p_\ell;\ \ell \in \mathbb{N}\}$, con $p_\ell \geq 0$, per ogni $\ell \in \mathbb{N}$ e $\sum_{\ell \in \mathbb{N}} p_\ell = 1$, in modo che $\mathbb{P}^0(K_\ell) = p_\ell$.

$$\mathbb{E}[XI_A] = \mathbb{E}[XI_B(Y)] = \int_{\mathbb{R}\times\mathbb{R}^d} xI_B(y) f_{X,Y}(x,y)\ dx\ dy.$$

Cominciamo con il caso in cui $f_{X,Y}(x,y) > 0$ *per ogni* $(x,y) \in \mathbb{R}\times\mathbb{R}^d$, *così anche* $f_Y(y) > 0$ *per ogni* $y \in \mathbb{R}^d$, *e quindi*

$$\begin{aligned}
\mathbb{E}[\widetilde{X}(\omega)I_A] &= \mathbb{E}[\widetilde{X}(\omega)I_B(Y)] \\
&= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}} x \frac{f_{X,Y}(x,y)}{f_Y(y)}\ dx \right) I_B(y) f_Y(y)\ dy
\end{aligned}$$

ovvero, per il Teorema di Fubini[19],

$$\begin{aligned}
\mathbb{E}[\widetilde{X}(\omega)I_A] &= \int_{\mathbb{R}} \int_{\mathbb{R}^d} xI_B(y) \frac{f_{X,Y}(x,y)}{f_Y(y)} f_Y(y)\ dx\ dy \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}^d} xI_B(y) f_{X,Y}(x,y)\ dx\ dy.
\end{aligned}$$

D'altra parte, nel caso generale

$$\begin{aligned}
\mathbb{E}[\widetilde{X}(\omega)I_A] &= \\
&= \mathbb{E}[\widetilde{X}(\omega)I_B(Y)] \\
&= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}} xI_{\{z: f_Y(z)>0\}}(y) \frac{f_{X,Y}(x,y)}{f_Y(y)}\ dx \right) I_B(y) f_Y(y)\ dy \\
&\quad + \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}} xI_{\{z: f_Y(z)=0\}}(y) f_0(x)\ dx \right) I_B(y) f_Y(y)\ dy
\end{aligned}$$

ovvero, per il Teorema di Fubini,

$$\begin{aligned}
&= \int_{\mathbb{R}} \int_{\mathbb{R}^d} xI_B(y) I_{\{z: f_Y(z)>0\}}(y) \frac{f_{X,Y}(x,y)}{f_Y(y)} f_Y(y)\ dx\ dy \\
&\quad + \int_{\mathbb{R}} \int_{\mathbb{R}^d} xI_B(y) I_{\{z: f_Y(z)=0\}}(y) f_0(x) f_Y(y)\ dx\ dy \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}^d} xI_B(y) I_{\{z: f_Y(z)>0\}}(y) \frac{f_{X,Y}(x,y)}{f_Y(y)} f_Y(y)\ dx\ dy \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}^d} xI_B(y) I_{\{z: f_Y(z)>0\}}(y) f_{X,Y}(x,y)\ dx\ dy
\end{aligned}$$

[19] Una versione del Teorema di Fubini è la seguente: se $\psi : \mathbb{R}^{n_1}\times\mathbb{R}^{n_2} \mapsto \mathbb{R};\ (x,y) \in \mathbb{R}^{n_1}\times\mathbb{R}^{n_2} \mapsto \psi(x,y) \in \mathbb{R}$ è una funzione boreliana, allora le seguenti condizioni sono equivalenti

$$\int_{R^{n_1}} \left(\int_{R^{n_2}} |\psi(x,y)| dy \right) dx < \infty,$$

$$\int_{R^{n_2}} \left(\int_{R^{n_1}} |\psi(x,y)| dx \right) dy < \infty$$

$$\int_{R^{n_1}\times^{n_2}_R} |\psi(x,y)| dx dy < \infty.$$

Inoltre se vale una delle precedenti condizioni vale, allora tutti i valori dei precedenti integrali coincidono. Il Teorema di Fubini è quindi usato per scambiare l'ordine degli integrali.

Si tratta quindi solo di controllare che, qualunque sia $B \in \mathcal{B}(\mathbb{R}^d)$

$$\int_{\mathbb{R}} \int_{\mathbb{R}^d} x I_B(y) I_{\{z: f_Y(z)>0\}}(y) f_{X,Y}(x,y)\ dx\ dy = \int_{\mathbb{R}} \int_{\mathbb{R}^d} x I_B(y) f_{X,Y}(x,y)\ dx\ dy.$$

La verifica è immediata in quanto

$$\int_{\mathbb{R}} \int_{\mathbb{R}^d} x I_B(y) I_{\{z: f_Y(z)=0\}}(y) f_{X,Y}(x,y)\ dx\ dy = 0,$$

infatti, se $I_{\{z: f_Y(z)=0\}}(y) = 1$, ovvero se $f_Y(y) = 0 \big(= \int_{\mathbb{R}} f_{X,Y}(x,y)\ dx\big)$, allora l'insieme $\{x : f_{X,Y}(x,y) > 0\}$ ha misura di Lebesgue nulla, e quindi, per tali y

$$\int_{\mathbb{R}^d} x I_B(y) f_{X,Y}(x,y)\ dx = 0.$$

Si osservi che anche in questo caso, se $f_{X,Y}(x,y) > 0$ per ogni (x,y), allora $\frac{f_{X,Y}(x,y)}{f_Y(y)}$ è una densità di probabilità in x, qualunque sia y, e che

$$\mathbb{P}(X \in C \mid Y) = \mathbb{E}[\mathbb{I}_C(X) \mid \sigma(Y)] = \int_C \frac{f_{X,Y}(x,y)}{f_Y(y)} dx$$

definisce una probabilità sui boreliani di $\mathbb{R}$.

Esempio 2.4. *Caso in cui $\mathcal{G} = \sigma\{Y\}$ e (X,Y) è una variabile (congiuntamente) gaussiana bidimensionale, di media nulla. Come caso particolare dell'esempio precedente si ottiene*

$$\mathbb{E}[X \mid Y] = \frac{\sigma_{X,Y}}{\sigma_Y^2} Y.$$

A sua volta questo risultato si ottiene dal caso più in generale: $(X_1, \cdots, X_n)$ è un vettore aleatorio gaussiano di media nulla e densità congiunta

$$f(x_1, \cdots, x_n) = c \exp\{-\sum_{i,j}^{1,n} \alpha_{i,j} x_i x_j\} = c \exp\{-\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_{i,j} x_i x_j\} \qquad \textit{con } \alpha_{i,j} = \alpha_{j,i},\ \alpha_{n,n} > 0$$

e

$$\mathbb{E}[X_n \mid X_1, \cdots, X_{n-1}] = -\sum_{i=1}^{n-1} \frac{\alpha_{i,n}}{\alpha_{n,n}} X_i. \tag{21}$$

(Pur essendo possibile utilizzare la tecnica dell'esempio precedente si consiglia lo svolgimento dei calcoli dopo l'introduzione delle distribuzioni condizionali) Svolgimento: Si tratta del caso $d = n-1$ con $X = X_n$ e $Y = (X_1, \cdots, X_{n-1})$ e con $f_{X,Y}(x,y) > 0$. Per calcolare $\mathbb{E}[X_n \mid X_1, \cdots, X_{n-1}]$ dobbiamo innanzitutto

$$\frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_{X_n,X}(x,y)}{f_Y(y)}$$

Abbiamo quindi

$$f(x_1, \cdots, x_{n-1}, x_n) = c \exp\left\{-\sum_{i,j}^{1,n-1} \alpha_{i,j} x_i x_j - x_n \left(\sum_{j=1}^{n-1} \alpha_{n,j} x_j\right) - \left(\sum_{i=1}^{n-1} \alpha_{i,n} x_i\right) x_n - \alpha_{n,n} x_n^2\right\}$$

Tenendo presente che $\alpha_{i,j} = \alpha_{j,i}$, e che $\alpha_{n,n} > 0$ si ha che

$$m(y) = m(x_1, \dots, x_{n-1}) := \sum_{j=1}^{n-1} \frac{\alpha_{n,j}}{\alpha_{n,n}} x_j = \sum_{i=1}^{n-1} \frac{\alpha_{i,n}}{\alpha_{n,n}} x_i,$$

e quindi che

$$f(x_1, \cdots, x_{n-1}, x_n) = f(y, x_n) = c \exp\left\{ -\alpha_{n,n} \left[\sum_{i,j}^{1,n-1} \frac{\alpha_{i,j}}{\alpha_{n,n}} x_i x_j + 2x_n m((x_1, \dots, x_{n-1})) + x_n^2 \right] \right\}$$

$$= c \exp\left\{ -\alpha_{n,n} \left[\sum_{i,j}^{1,n-1} \frac{\alpha_{i,j}}{\alpha_{n,n}} x_i x_j + m^2(y) + 2x_n m(y) + x_n^2 - m^2(y) \right] \right\}$$

$$= c(y) \exp\left\{ -\alpha_{n,n} \left[m^2(y) + 2x_n m(y) + x_n^2 \right] \right\} == c(y) \exp\left\{ -\alpha_{n,n} \left[x_n + m(y) \right]^2 \right\}$$

Di conseguenza

$$\frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{c(y)}{f_Y(y)} \exp\left\{ -\alpha_{n,n} \left[x_n - (-m(y)) \right]^2 \right\} = K(y) \exp\left\{ -\frac{\left[x_n - (-m(y)) \right]^2}{2 \frac{1}{2\alpha_{n,n}}} \right\}.$$

Poiché la precedente espressione deve essere una densità di probabilità, come osservato precedentemente, è chiaro che deve coincidere con la densità di una variabile aleatoria gaussiana, $N(-m(y), \frac{1}{2\alpha_{n,n}})$, di media $-m(y)$ e di varianza $\frac{1}{2\alpha_{n,n}}$. Il valore atteso si calcola quindi come

$$\begin{aligned} \mathbb{E}[X_n \mid X_1, \cdots, X_{n-1}] &= \int_{\mathbb{R}} x K(y) \exp\left\{ -\frac{\left[x_n - (-m(y)) \right]^2}{2 \frac{1}{2\alpha_{n,n}}} \right\} dx \Bigg|_{y=(X_1,\dots,X_{n-1})} \\ &= \quad -m(y) \Bigg|_{y=(X_1,\dots,X_{n-1})}, \end{aligned}$$

da cui si ottiene (21).

Esempio 2.5. *L'Esempio 2.3 si generalizza facilmente al caso in cui X è una variabile aleatoria a valori in $\mathbb{R}^k$, e si vuole calcolare il valore atteso condizionale $\mathbb{E}[h(X) \mid Y]$, dove $h(\cdot)$ è una funzione misurabile, a valori reali, ripetendo tutti i passaggi con i dovuti cambiamenti:*

$$\widetilde{h(X)}(\omega) = \mathbb{E}[h(X) \mid Y](\omega) = \int_{\mathbb{R}^k} h(x)\ f_{X|Y}(x|y)\big|_{y=Y(\omega)}\ dx$$

Enunciamo ora (senza dimostrarle, per il momento), le proprietà fondamentali della media condizionale.

PROPRIETÀ DELLA MEDIA CONDIZIONALE (secondo la Definizione 2.1)

Siano X e Y variabili aleatorie integrabili in $(\Omega, \mathcal{F}, \mathbb{P})$ e siano $\mathcal{G}$ e $\mathcal{H}$ sotto σ-algebre di $\mathcal{F}$, allora valgono le seguenti proprietà:

1. **Linearità**

$$\mathbb{E}[aX + bY \mid \mathcal{G}] = a\mathbb{E}[X \mid \mathcal{G}] + b\mathbb{E}[Y \mid \mathcal{G}]$$

2. **Monotonia**

$$\mathbb{P}(X \leq Y) = 1 \text{ implica } \mathbb{P}(\mathbb{E}[X \mid \mathcal{G}] \leq \mathbb{E}[Y \mid \mathcal{G}]\ ) = 1$$

3. **Formula dei condizionamenti successivi** (caso particolare $\mathcal{G} = \{\Omega, \emptyset\}$)

$$\text{se } \mathcal{G} \subseteq \mathcal{H}, \text{ allora } \mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{H}] \mid \mathcal{G}]$$

   quindi, in particolare,

$$\text{se } \mathcal{G} = \{\Omega, \emptyset\}, \text{ allora } \mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{H}]]$$

4. **Fattorizzazione**

   Se Z è $\mathcal{G}$-misurabile e ZX è integrabile allora

$$\mathbb{E}[ZX \mid \mathcal{G}] = Z\mathbb{E}[X \mid \mathcal{G}]$$

5. **Condizionamento rispetto a σ-algebre indipendenti**

   Se X e $\mathcal{G}$ sono indipendenti allora

$$\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X]$$

6. **Condizionamento ridondante**, cioè rispetto ad allargamenti indipendenti di σ-algebre

   Se X e $\mathcal{G}$ sono indipendenti da $\mathcal{H}$, allora

$$\mathbb{E}[X \mid \mathcal{G} \vee \mathcal{H}] = \mathbb{E}[X \mid \mathcal{G}]$$

7. **Disuguaglianza di Jensen per funzioni convesse** (caso particolare $\phi(x) = x^2$)

   Se ϕ è una funzione convessa, e $\phi(X)$ è integrabile, allora

$$\phi(\mathbb{E}[X \mid \mathcal{G}]) \leq \mathbb{E}[\phi(X) \mid \mathcal{G}]$$

   OSSERVAZIONE In particolare per $\phi(x) = x^2$ ed X in $L^2(\Omega, \mathcal{F}, \mathbb{P})$, si ottiene che

$$(\mathbb{E}[X \mid \mathcal{G}])^2 \leq \mathbb{E}[X^2 \mid \mathcal{G}],$$

   e quindi, passando al valore atteso, che

$$\mathbb{E}[(\mathbb{E}[X \mid \mathcal{G}])^2] \leq \mathbb{E}[\mathbb{E}[X^2 \mid \mathcal{G}]] = \mathbb{E}[X^2].$$

8. Convergenza sotto il segno di media condizionale, monotona e dominata

Se $\{X_n\}_{n\geq 1}$ è una successione di variabili aleatorie non negative ed integrabili, convergente con probabilità 1 ad X, monotonamente, cioè $0 \leq X_n \leq X_{n+1}$, allora

$$\mathbb{E}[X_n \mid \mathcal{G}] \uparrow \mathbb{E}[X \mid \mathcal{G}], \text{ con probabilità 1},$$

Se invece la successione converge ad X dominatamente, cioè $\mid X_n \mid \leq Y$, allora

$$\mathbb{E}[X_n \mid \mathcal{G}] \to \mathbb{E}[X \mid \mathcal{G}], \text{ in } L^1.$$

La Definizione 2.1 è ben posta, in quanto esiste almeno una variabile aleatoria che verifica la (14).

Si definisca infatti la misura $\nu(\cdot)$ su $\mathcal{G}$ come $\nu(A) := \mathbb{E}[XI_A]$, per $A \in \mathcal{G}$. La misura $\nu(\cdot)$ risulta assolutamente continua[20] rispetto a $\widehat{\mathbb{P}} = \mathbb{P}|_{\mathcal{G}}$, in quanto se $\widehat{\mathbb{P}}$(A)=$\mathbb{P}$(A)=0, allora $\nu(A) = 0$. Esiste quindi la derivata di Radon-Nikodym di ν rispetto a $\widehat{\mathbb{P}}$, cioè una funzione $f(\omega) = \frac{d\nu}{d\widehat{\mathbb{P}}}(\omega)$, $\mathcal{G}-$ misurabile, per la quale valga, qualunque sia $A \in \mathcal{G}$

$$\nu(A) = \int_A f(\omega) d\widehat{\mathbb{P}}(\omega) = \int_A f(\omega) d\mathbb{P}(\omega)$$

ovvero

$$\mathbb{E}[XI_A] = \mathbb{E}[fI_A].$$

La derivata di Radon-Nikodym è definita a meno di insiemi di $\widehat{\mathbb{P}}$-misura nulla. Infatti se $g(\omega)$ è un'altra funzione $\mathcal{G}$-misurabile per la quale valga $\mathbb{E}[XI_A] = \mathbb{E}[gI_A]$ per ogni $A \in \mathcal{G}$, allora $\mathbb{E}[fI_A] = \mathbb{E}[gI_A]$ per ogni $A \in \mathcal{G}$, ed in particolare per $A = \{\omega : f(\omega) \geq g(\omega)\}$ e per $A = \{\omega : f(\omega) < g(\omega)\}$, per cui $\mathbb{E}[\mid f - g \mid] = 0$. Basta quindi prendere come $\widetilde{X}$ la derivata di Radon-Nikodym f o qualunque altra variabile aleatoria che differisca da $\widetilde{X}$ al più in un insieme ($\mathcal{G}$-misurabile) di probabilità nulla.

Anche la Definizione 2.2 è ben posta, per convincersene basta considerare che lo spazio $L^2(\Omega, \mathcal{F}, \mathbb{P})$, pensato come classi di equivalenza, è uno spazio di Hilbert con la norma $\parallel X \parallel^2 = \mathbb{E}[\mid X \mid^2]$ ed $L^2(\Omega, \mathcal{G}, \widehat{\mathbb{P}})$ è un suo sottospazio chiuso. Quindi la classe di equivalenza $\mathbb{E}[X \mid \mathcal{G}]$ della Definizione 2.2 è la proiezione di X su $L^2(\Omega, \mathcal{G}, \widehat{\mathbb{P}})$.

Si noti la (14) potrebbe essere modificata come

$$\mathbb{E}[XW] = \mathbb{E}[\widetilde{X}W], \quad \forall \text{ v.a. } W \quad \mathcal{G}-\text{misurabile e per cui } XW \text{è integrabile.} \tag{22}$$

[20] Date due misure ν e μ su una σ-algebra $\mathcal{G}$, si dice che ν è risulta assolutamente rispetto a μ se e solo se per ogni $A \in \mathcal{G}$ con $\mu(A) = 0$ risulta $\nu(A) = 0$. Questa proprietà si può anche esprimere dicendo che la famiglia $\mathcal{N}^\nu = \{A : \nu(A) = 0\}$ degli insiemi di ν-misura nulla contiene la famiglia $\mathcal{N}^\mu = \{A : \mu(A) = 0\}$ degli insiemi di μ-misura nulla. La derivata di Radon-Nikodym $\frac{d\nu}{d\mu}$ è una funzione h, che sia $\mathcal{G}$-misurabile, e per la quale valga

$$\nu(A) = \int_\Omega h(\omega)\mu(d\omega), \quad \text{per ogni } A \in \mathcal{G}.$$

Dimostrazione: [ATTENZIONE: la dimostrazione si basa sulle proprietà di linearità e di monotonia dei valori attesi condizionali, che dimostreremo più in là, basandoci solo sulla Definizione 2.1 e quindi sulla (11)] Ovviamente (12) implica (11). Per mostrare il viceversa, basta considerare il caso in cui $W \geq 0$. In tale caso esiste una successione $W_n \uparrow$W, con W_n funzioni elementari, cioè combinazioni lineari di funzioni indicatrici. D'altra parte è facile vedere (confronta le proprietà di linearità e di monotonia, successivamente) che, posto

$$X = X^+ - X^-, \qquad \text{con } X^+ = X \vee 0, \text{ e con } X^- = (-X) \vee 0,$$

si ha $\widetilde{X} = \widetilde{X^+} - \widetilde{X^-}$, per la proprietà di linearitaà, con $\widetilde{X^+}$ e $\widetilde{X^-} \geq 0$,per la proprietà di monotonia. Inoltre si ha che

$$\mathbb{E}[XW_n] = \mathbb{E}[X^+W_n] - \mathbb{E}[X^-W_n] = \mathbb{E}[\widetilde{X^+}W_n] - \mathbb{E}[\widetilde{X^-}W_n].$$

Per la proprietà di convergenza monotona dei valori attesi, (se $Z \geq 0$ allora $0 \leq ZW_n \leq ZW_{n+1} \uparrow ZW$), si ha

$$\mathbb{E}[X^+W_n] \uparrow \mathbb{E}[X^+W], \qquad \mathbb{E}[X^-W_n] \uparrow \mathbb{E}[X^-W],$$
$$\mathbb{E}[\widetilde{X^+}W_n] \uparrow \mathbb{E}[\widetilde{X^+}W], \qquad \mathbb{E}[\widetilde{X^-}W_n] \uparrow \mathbb{E}[\widetilde{X^-}W],$$

quindi $\mathbb{E}[XW] = \mathbb{E}[\widetilde{X}W]$, in quanto

$$\mathbb{E}[XW] \leftarrow \mathbb{E}[XW_n] = \mathbb{E}[\widetilde{X}W_n] \rightarrow \mathbb{E}[\widetilde{X^+}W] - \mathbb{E}[\widetilde{X^-}W] = \mathbb{E}[\widetilde{X}W]$$

2.1 Equivalenza tra Definizione 2.1 e Definizione 2.2 per variabili aleatorie di quadrato sommabile

Mostreremo ora che, se X è di quadrato sommabile, allora ogni variabile aleatoria $\widetilde{X}_2$ che soddisfa la Definizione 2.2, soddisfa anche la Definizione 2.1. Per l'implicazione inversa, sempre nel caso in cui X sia di quadrato integrabile, abbiamo bisogno di alcune delle proprietà della media condizionale secondo la Definizione 2.1, enunciate precedentemente e che dimostreremo in seguito.

1) Se $\widetilde{X}_2$ è la (o meglio un rappresentante della) media condizionale di X secondo la Definizione 2.2, allora lo è anche secondo la Definizione 2.1.

Se infatti $\widetilde{X}_2$ soddisfa la condizione (13) allora, qualunque sia $Z \in L^2(\Omega, \mathcal{G}, \widehat{\mathbb{P}})$ la funzione

$$\phi_2(t) := \mathbb{E}[(X - \{(1-t)\widetilde{X}_2 + tZ\})^2] = \mathbb{E}[(X - \widetilde{X}_2)^2 - 2t(X - \widetilde{X}_2)(Z - \widetilde{X}_2) + t^2(Z - \widetilde{X}_2)^2]$$

ammette un minimo in $t = 0$, e quindi

$$\phi_2'(t)|_{t=0} = \frac{d}{dt}\mathbb{E}[(X - \{(1-t)\widetilde{X}_2 + tZ\})^2]\bigg|_{t=0} = -2\mathbb{E}[(X - \widetilde{X}_2)(Z - \widetilde{X}_2)] = 0$$

Quindi posto $W = Z - \widetilde{X}_2$ deve valere, qualunque sia $W \in L^2(\Omega, \mathcal{G}, \widehat{\mathbb{P}})$

$$\mathbb{E}[XW] = \mathbb{E}[\widetilde{X}_2 W].$$

Considerando che tutte le funzioni indicatrici del tipo I_A, con $A \in \mathcal{G}$, sono in $L^2(\Omega, \mathcal{G}, \widehat{\mathbb{P}})$, otteniamo che se vale la (13) allora vale la (11).

2) Se $\widetilde{X}$ è la media condizionale di X secondo la Definizione 2.1 (o meglio ne è un rappresentante), allora lo è anche secondo la Definizione 2.2.

Infatti, allora $\widetilde{X}$ è di quadrato sommabile (confrontare l'osservazione alla disuguaglianza di Jensen, proprietà 7.) e quindi la seguente funzione è finita per ogni Z di quadrato sommabile

$$\phi(t) := \mathbb{E}[(X - \{(1-t)\widetilde{X} + tZ\})^2] = \mathbb{E}[(X - \widetilde{X})^2 - 2t(X - \widetilde{X})(Z - \widetilde{X}) + t^2(Z - \widetilde{X})^2]$$

Poiché si tratta di una parabola essa ammette un minimo nel punto $\bar{t}$, in cui la derivata si annulla:

$$\phi'(\bar{t}) = -2\mathbb{E}[(X - \widetilde{X})(Z - \widetilde{X})] + 2\bar{t}\,\mathbb{E}[(Z - \widetilde{X})^2] = 0$$

Posto $W = Z - \widetilde{X}$ e tenendo conto della (12), si ottiene $\phi'(\bar{t}) = 2\bar{t}\,\mathbb{E}[(Z - \widetilde{X})^2] = 0$, da cui, per l'arbitrarietà di Z, segue che $\bar{t} = 0$. Quindi $\widetilde{X}$ gode della proprietà che per ogni Z di quadrato sommabile $\phi(1) \geq \phi(0)$, ovvero la (13).

3 DIMOSTRAZIONI DELLE PROPRIETÀ DELLA MEDIA CONDIZIONALE

Per alcune dimostrazioni delle proprietà della media condizionale è utile il seguente lemma di teoria della misura:

Lemma 3.1 (Lemma di Dynkin, Billingsley 1984). *Sia $\mathcal{A}$ una famiglia di eventi che genera la σ-algebra $\mathcal{G}$ e che è chiusa rispetto alla intersezione finita. Se due misure finite ν e μ coincidono su $\mathcal{A}$, allora le due misure coincidono su $\mathcal{G}$.*

Faremo ora le dimostrazioni delle proprietà enunciate nella sezione precedente utilizzando solo la Definizione 2.1.

Si noti che ciò permette di concludere che la dimostrazione dell'equivalenza delle Definizioni 2.1 e 2.2 è autocontenuta: in realtà andrebbero prima dimostrate le proprietà 1,2,...,7; successivamente la $(1')$ e infine l'equivalenza delle Definizioni 2.1 e 2.2.

1. Linearità: $\mathbb{E}[aX + bY \mid \mathcal{G}] = a\mathbb{E}[X \mid \mathcal{G}] + b\mathbb{E}[Y \mid \mathcal{G}]$

La dimostrazione è ovvia infatti se $\mathbb{E}[XI_A] = \mathbb{E}[\widetilde{X}I_A]$ e $\mathbb{E}[YI_A] = \mathbb{E}[\widetilde{Y}I_A]$ per ogni $A \in \mathcal{G}$ allora

$$\mathbb{E}[(aX + bY)I_A] = a\mathbb{E}[XI_A] + b\mathbb{E}[YI_A] = a\mathbb{E}[\widetilde{X}I_A] + b\mathbb{E}[\widetilde{Y}I_A] = \mathbb{E}[(a\widetilde{X} + b\widetilde{Y})I_A]$$

2. Monotonia: $\mathbb{P}(X \leq Y) = 1$ implica $\mathbb{P}(\mathbb{E}[X \mid \mathcal{G}] \leq \mathbb{E}[Y \mid \mathcal{G}]\ ) = 1$

Se $\mathbb{P}(X \leq Y) = 1$, allora $\mathbb{E}[(Y - X)I_A] = \mathbb{E}[(\widetilde{Y} - \widetilde{X})I_A] \geq 0$ per ogni $A \in \mathcal{G}$ e quindi in particolare per $A = \{\widetilde{X} < \widetilde{Y}\}$, si ottiene che $\mathbb{P}(\widetilde{X} \leq \widetilde{Y}) = 1$, ovvero $\mathbb{P}(\mathbb{E}[X \mid \mathcal{G}] \leq \mathbb{E}[Y \mid \mathcal{G}]) = 1$.

3. Formula dei condizionamenti successivi: se $\mathcal{G} \subseteq \mathcal{H}$, allora $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{H}] \mid \mathcal{G}]$

Sia $\mathcal{G} \subseteq \mathcal{H}$ e siano $\widetilde{X}$ e $\widehat{X}$ tali che $\mathbb{E}[XI_A] = \mathbb{E}[\widetilde{X}I_A]$ per ogni $A \in \mathcal{G} \subseteq \mathcal{H}$ e $\mathbb{E}[XI_B] = \mathbb{E}[\widehat{X}I_B]$ per ogni $B \in \mathcal{H}$, ed in particolare per ogni $A \in \mathcal{G}$. Allora

$$\mathbb{E}[\widehat{X}I_A] = \mathbb{E}[\widetilde{X}I_A] (= \mathbb{E}[XI_A]), \quad \text{per ogni } A \in \mathcal{G}$$

e quindi $\widetilde{X} = \mathbb{E}[\widehat{X} \mid \mathcal{G}]$. La seconda parte deriva dal fatto che per la σ-algebra banale la media e la media condizionale coincidono.

4. Fattorizzazione: e Z è $\mathcal{G}$-misurabile e ZX è integrabile allora $\mathbb{E}[ZX \mid \mathcal{G}] = Z\mathbb{E}[X \mid \mathcal{G}]$.

Infatti allora $\mathbb{E}[XZI_A] = \mathbb{E}[\widetilde{X}ZI_A]$, per ogni $A \in \mathcal{G}$ e la funzione $Z\widetilde{X}$ è ovviamente $\mathcal{G}$-misurabile.

5. Condizionamento rispetto a σ-algebre indipendenti: se X e $\mathcal{G}$ sono indipendenti allora $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X]$.

Infatti allora, per ogni $A \in \mathcal{G}$

$$\mathbb{E}[XZI_A] = \mathbb{E}[X]\mathbb{E}[ZI_A] = \mathbb{E}[X]\mathbb{E}[Z\mathbb{E}[X]I_A]$$

e la funzione $Z\mathbb{E}[X]$ è ovviamente $\mathcal{G}$-misurabile.

6. Condizionamento ridondante: se X e $\mathcal{G}$ sono indipendenti da $\mathcal{H}$ allora $\mathbb{E}[X \mid \mathcal{G} \vee \mathcal{H}] = \mathbb{E}[X \mid \mathcal{G}]$.

La dimostrazione si basa sul fatto che gli eventi del tipo $A \cap B$, con $A \in \mathcal{G}$ e $B \in \mathcal{H}$, formano un sistema chiuso rispetto all'intersezione e generano $\mathcal{G} \vee \mathcal{H}$, per il lemma precedente basterà quindi verificare che

$$\mathbb{E}[XI_{A\cap B}] = \mathbb{E}[XI_AI_B] = \mathbb{E}[XI_A]\mathbb{E}[I_B] = \mathbb{E}[\widetilde{X}I_A]\mathbb{E}[I_B] = \mathbb{E}[\widetilde{X}I_AI_B] = \mathbb{E}[\widetilde{X}I_{A\cap B}].$$

$\big($Alternativamente, senza bisogno del Lemma di Dynkin: le unioni finite di eventi di questo tipo formano un'algebra che genera $\mathcal{G} \vee \mathcal{H}$, le due misure $\mathbb{E}[XI_{A\cap B}]$ e $\mathbb{E}[\widetilde{X}I_{A\cap B}]$ coincidono sull'algebra e sono σ-additive, quindi coincidono su tutta la σ-algebra $\mathcal{G} \vee \mathcal{H}$.$\big)$[21]

[21] Va tuttavia osservato che comunque, la dimostrazione del risultato sull'unicità dell'estensione da un'algebra $\mathcal{A}$ alla σ-algebra generata da $\mathcal{A}$, si basa a sua volta sul Lemma di Dynkin.

7. Disuguaglianza di Jensen[22]: se ϕ è una funzione convessa, e $\phi(X)$ è integrabile, allora $\phi(\mathbb{E}[X \mid \mathcal{G}]) \leq \mathbb{E}[\phi(X) \mid \mathcal{G}]$.

Per risultati classici di analisi ϕ è l'inviluppo delle sue tangenti (o sotto tangenti) ovvero esistono una successione di rette $\phi_n(x) = \alpha_n x + \beta_n$ per cui $\phi(x) = \sup_n\{\phi_n(x)\}$, per ogni x in $\mathbb{R}$. Per le proprietà di linearità e di monotonia si ha allora che

$$\phi_n(\mathbb{E}[X \mid \mathcal{G}]) = \alpha_n \mathbb{E}[X \mid \mathcal{G}] + \beta_n = \mathbb{E}[\phi_n(X) \mid \mathcal{G}] \leq \mathbb{E}[\phi(X) \mid \mathcal{G}];$$

passando all'estremo superiore su n si ottiene

$$\sup_n\{\phi_n(\mathbb{E}[X \mid \mathcal{G}])\} = \phi(\mathbb{E}[X \mid \mathcal{G}]) \leq \mathbb{E}[\phi(X) \mid \mathcal{G}].$$

8. Convergenza sotto il segno di media condizionale, monotona e dominata

8i) Se $\{X_n\}_{n\geq 1}$ è una successione di variabili aleatorie non negative ed integrabili, convergente con probabilità 1 ad X, **monotonamente**, cioè $0 \leq X_n \leq X_{n+1}$, allora $\mathbb{E}[X_n \mid \mathcal{G}] \nearrow \mathbb{E}[X \mid \mathcal{G}]$, con probabilità 1.

Dalla proprietà di monotonia si ottiene che se $\{X_n\}_{n\geq 1}$ è una successione monotona allora anche la successione delle medie condizionate $\widetilde{X}_n = \mathbb{E}[X_n \mid \mathcal{G}]$ è una successione monotona, ed è quindi convergente ad una variabile aleatoria $\widetilde{Z}$, $\mathcal{G}$-misurabile. (È importante notare che per ogni n esiste un insieme $A_n \in \mathcal{G}$ di probabilità nulla nel cui complementare vale $\widetilde{X}_n \leq \widetilde{X}_{n+1}$, che queste disuguaglianze valgono contemporaneamente nel complementare di $A = \bigcup_{n\geq 1} A_n$, ed infine che A ha ancora misura nulla).
Per mostrare che la successione converge ad un rappresentante di $\mathbb{E}[X \mid \mathcal{G}]$, basta notare che $\mathbb{E}[\widetilde{X}_n I_A] = \mathbb{E}[X_n I_A] \uparrow \mathbb{E}[X I_A]$, per la convergenza monotona di $\{X_n\}$ a X (e quindi

---

[22] Si noti che nel caso particolare in cui $\mathcal{G}$ è la σ-algebra banale la disuguaglianza di Jensen per i valori attesi condizionali diviene l'usuale disuguaglianza di Jensen

$$\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)].$$

Infine va notato che la precedente disuguaglianza di Jensen, nel caso particolare di una variabile aleatoria X semplice, a valori in $\{x_1, \ldots, x_n\}$, con $\mathbb{P}(\{X = x_i\}) = \lambda_i$ si scrive come

$$\phi(\mathbb{E}[X]) = \phi(\sum_{i=1}^n \lambda_i x_i) \leq \sum_{i=1}^n \lambda_i \phi(x_i) = \mathbb{E}[\phi(X)], \quad ,\text{con } \lambda_i \geq 0,\ \sum_{i=1}^n \lambda_i = 1$$

La disuguaglianza interna è equivalente alla definizione di funzione convessa. Lo è esattamente nel caso $n = 2$

$$\phi(\lambda x_1 + (1-\lambda)x_2) \leq \lambda\phi(x_1) + (1-\lambda)\phi(x_2)$$

Il caso generale si ottiene per induzione su n: Se

$$\phi(\sum_{i=1}^n \lambda_i x_i) \leq \sum_{i=1}^n \lambda_i \phi(x_i), \text{ per ogni } \lambda_i \geq 0,\ \sum_{i=1}^n \lambda_i = 1, \text{ e per ogni } x_i$$

allora

$$\phi(\sum_{i=1}^{n+1} \mu_i x_i) \leq \sum_{i=1}^{n+1} \mu_i \phi(x_i), \text{ per ogni } \mu_i \geq 0,\ \sum_{i=1}^{n+1} \mu_i = 1, \text{e per ogni } y_i$$

di $\{X_n I_A\}$ a XI_A), e che $\mathbb{E}[\widetilde{X}_n I_A] \uparrow \mathbb{E}[\widetilde{Z} I_A]$, per la convergenza monotona di $\{\widetilde{X}_n\}$ a $\widetilde{Z}$ (e quindi di $\{\widetilde{X}_n I_A\}$ a $\widetilde{Z} I_A$). Di conseguenza $\mathbb{E}[XI_A] = \mathbb{E}[\widetilde{Z} I_A], \forall A \in \mathcal{G}$.

8ii) Se $\{X_n\}_{n\geq 1}$ è una successione di variabili aleatorie integrabili, convergente con probabilità 1 ad X, **dominatamente**, cioè $\mid X_n \mid \leq Y$, per una variabile aleatoria Y integrabile, allora $\mathbb{E}[X_n \mid \mathcal{G}] \to \mathbb{E}[X \mid \mathcal{G}]$, in L^1.

La dimostrazione relativa alla convergenza dominata si basa sulla seguente osservazione: se $X_n \to X$ q.c. e $\mid X_n \mid \leq Y$ allora $X_n \to X$ in L^1 (infatti allora $\mid X_n - X \mid \leq \mid Y \mid + \mid X \mid$ e quindi $\mid X_n - X \mid \to 0$ dominatamente e perciò $\mathbb{E}[\mid X_n - X \mid] \to 0$). Di conseguenza per le proprietà di linearità e per la disuguaglianza di Jensen applicata alla funzione $\phi(x) = \mid x \mid$ e alla variabile aleatoria $X_n - X$,

$$\mid \mathbb{E}[X_n \mid \mathcal{G}] - \mathbb{E}[X \mid \mathcal{G}] \mid = \mid \mathbb{E}[X_n - X \mid \mathcal{G}] \mid \leq \mathbb{E}[|X_n - X| \mid \mathcal{G}].$$

Passando ai valori medi

$$\mathbb{E}[\mid \mathbb{E}[X_n \mid \mathcal{G}] - \mathbb{E}[X \mid \mathcal{G}] \mid] = \mathbb{E}[\mid \mathbb{E}[X_n - X \mid \mathcal{G}] \mid] \leq \mathbb{E}[\mathbb{E}[\mid X_n - X \mid \mid \mathcal{G}]] = \mathbb{E}[\mid X_n - X \mid] \to 0.$$

OSSERVAZIONE. Dalla dimostrazione è chiaro che basta la convergenza $X_n \to X$ in L^1, per ottenere la convergenza delle rispettive medie condizionali in L^1. In realtà, nel caso di convergenza dominata, c'è anche la convergenza puntuale delle medie condizionali. La dimostrazione di questo fatto è rimandata a dopo aver mostrato l'esistenza di distribuzioni condizionali regolari.

4 PROBABILITÀ CONDIZIONALI REGOLARI

Il problema è il seguente:

È possibile trovare una versione di $\mathbb{P}(A \mid \mathcal{G})$ in modo che l'applicazione

$$\mathbb{P}(\cdot \mid \mathcal{G})(\omega) : \ \mathcal{F} \to [0,1]; A \to \mathbb{P}(A \mid \mathcal{G})(\omega)$$

sia una probabilità per ogni ω?

Ad un primo sguardo superficiale sembrerebbe di sì:
Ovviamente $\mathbb{P}(\Omega \mid \mathcal{G})(\omega)$=1. Dalla proprietà di monotonia si ha che $\mathbb{P}(C \mid \mathcal{G}) \in [0,1]$, per ogni evento $C \in \mathcal{F}$. Dalla proprietà di convergenza monotona, applicata alla successione $X_n = \sum_{k=1}^n I_{A_k}$, dove $\{A_n\}$ è una successione di eventi di $\mathcal{F}$, oppure di una σ-algebra $\mathcal{A} \subseteq \mathcal{F}$, si ottiene che, comunque scelta una successione di eventi di $\mathcal{A}$, disgiunti a due a due, vale

$$\mathbb{P}(\bigcup_{n\geq 1} A_n \mid \mathcal{G}) = \sum_{n\geq 1} \mathbb{P}(A_n \mid \mathcal{G})$$

nel senso che al più differiscono per un insieme di misura nulla N.

Sembrerebbe quindi tutto funzionare. Il problema sta nel fatto che l'insieme N può dipendere però, in generale, dalla particolare successione $\{A_n, n \geq 1\}$ scelta e l'unione su tutte le successioni possibili è un'unione non numerabile, quindi non è detto che sia un evento e, anche se lo fosse, non è detto che sia di probabilità nulla. È proprio questo che in generale impedisce di affermare che $A \mapsto \mathbb{P}(A \mid \mathcal{G})(\omega)$ è una probabilità.

Lo stesso tipo di problema si pone nel caso in cui invece si cerchi una versione di $\mathbb{P}(A \mid \mathcal{G})(\omega)$ per $A \in \mathcal{A} \subseteq \mathcal{F}$, mentre non c'è nessun problema del tipo precedente se $\mathcal{A}$ è un'algebra finita (e quindi una $\sigma-$algebra).
Comunque, nel caso in cui sia possibile trovare un **nucleo di misure di probabilità**, cioè una famiglia

$$\mathbb{Q}(\cdot,\cdot): \ \mathcal{A} \times \Omega \to [0,1]; (A,\omega) \mapsto \mathbb{Q}(A,\omega)$$

1) per ogni $\omega \in \Omega$, $\mathbb{Q}(\cdot,\omega)$ è una misura di probabilità su $(\Omega,\mathcal{A})$

2) per ogni $C \in \mathcal{A}$, $\mathbb{Q}(C,\cdot)$ è $\mathcal{G}$-misurabile ed è una versione di $\mathbb{P}(C \mid \mathcal{G})(\omega)$,

allora si dice che $\mathbb{Q}(\cdot,\omega)$ è una **versione regolare** di $\mathbb{P}(\cdot \mid \mathcal{G})(\omega)$, cioè **delle probabilità condizionali**.

L'interesse di tali versioni deriva dal fatto che allora, per ogni v.a. Z, $\mathcal{A}$-misurabile e integrabile, vale

$$\mathbb{E}(Z \mid \mathcal{G})(\omega) = \int_\Omega Z(\omega')\mathbb{Q}(d\omega',\omega). \tag{3}$$

Per convincersene basta capire che ciò è vero per $Z = I_C$, e quindi per ogni funzione elementare, cioè combinazione lineare di funzioni indicatrici. Quindi (3) vale per ogni v.a. non negativa integrabile, in quanto limite monotono di funzioni elementari, e quindi per ogni v.a. integrabile in quanto differenza di due v.a. non negative ed integrabili.

L'interesse per l'esistenza di una versione regolare delle probabilità condizionali sta anche nel fatto che tutte le dimostrazioni delle proprietà delle medie condizionali sarebbero immediate (in particolare la disuguaglianza di Jensen e la convergenza monotona e dominata, anche nella versione dell'osservazione relativa, con la convergenza puntuale).

Non sempre, purtroppo, si hanno versioni regolari delle probabilità condizionali. In generale dipende dalla σ-algebra $\mathcal{A}$, qui dimostreremo che ciò è vero se $\mathcal{A} = \sigma(X)$, dove X è una variabile aleatoria a valori in $\mathbb{R}$, cioè se $C = \{X \in H\}$, $H \in \mathcal{B}(\mathbb{R})$. In realtà si potrebbe vedere che ciò è vero anche se X è una variabile aleatoria $d$dimensionale, oppure se X è una variabile aleatoria a valori in uno spazio metrico S, completo e separabile (ovvero uno spazio polacco). In questi casi si ottiene anche una probabilità sullo spazio S degli stati di X e si parla di distribuzione condizionale invece che di probabilità condizionale, che invece è una misura di probabilità su Ω.

L'idea è molto semplice: si considera $F(t \mid \mathcal{G})(\omega) := \mathbb{P}(X \leq t \mid \mathcal{G})(\omega)$ per t razionale. Per la proprietà della monotonia, possiamo prendere una versione per cui, per ogni $t_1 \leq t_2$ razionali, $F(t_1 \mid \mathcal{G})(\omega) \leq F(t_2 \mid \mathcal{G})(\omega)$, in un evento Ω_0 di probabilità 1: in questo caso l'evento in cui ciò non si verifica è un'unione numerabile di eventi di probabilità nulla. Per $\omega \in \Omega_0^c$ si definisce $F(t \mid \mathcal{G})(\omega) = F_0(t)$, con F_0 una fissata funzione di distribuzione, ad esempio (sempre sui razionali). Su tale evento Ω_0 esiste, per monotonia, il limite di $F(t \mid \mathcal{G})(\omega)$ sia per $t \to +\infty$

che per $t \to -\infty$, sempre per t razionale. Tali limiti sono rispettivamente uguali ai limiti di $F(n \mid \mathcal{G})(\omega)$ e di $F(-n \mid \mathcal{G})(\omega)$, e, di nuovo a parte un insieme di probabilità nulla, coincidono rispettivamente con $1 = \mathbb{P}(X < +\infty \mid \mathcal{G})$ e $0 = \mathbb{P}(X < -\infty \mid \mathcal{G})$.(Di nuovo su tale insieme si definisca $F(t \mid \mathcal{G})(\omega) = F_0(t)$)

Per ogni valore s reale, ma non razionale, si definisce una successione $t_n \downarrow$s. Per la proprietà della convergenza monotona (applicato a $1 - I_{\{X \le t_n\}}$) si ottiene che il limite di $F(t_n \mid \mathcal{G})(\omega)$ esiste ed è una versione di $\mathbb{P}(X \le s \mid \mathcal{G})(\omega)$. In questo modo si è ottenuta, per ogni ω una funzione $F(s \mid \mathcal{G})(\omega)$, definita su tutti i reali, e che soddisfa tutte le proprietà di una funzione di ripartizione, come si può vedere facilmente. Ad ogni ω è quindi associata una misura $\mu_X(\cdot \mid \mathcal{G})(\omega)$ di probabilità, che definisce appunto la distribuzione condizionale di X data $\mathcal{G}$.

Si faccia attenzione: $\mu_X(\cdot \mid \mathcal{G})(\omega)$ è una probabilità su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, mentre $\mathbb{Q}(\cdot \mid \mathcal{G})(\omega)$ è una misura su $(\Omega, \sigma(X))$. Comunque accade che $\mathbb{P}(\{X \in H\} \mid \mathcal{G})(\omega) = \mu_X(H \mid \mathcal{G})(\omega)$. Inoltre poiché le v.a. $\sigma(X)-$misurabili sono le variabili aleatorie del tipo $Z = f(X)$ con f boreliana, si ha che

$$\mathbb{E}(f(X) \mid \mathcal{G})(\omega) = \mu_X(f \mid \mathcal{G})(\omega),$$

dove, in generale,

$$\mu(f) := \int f(x) d\mu(x).$$

Per la verità esiste una generalizzazione a tutti gli **spazi di Borel** $(S, \mathcal{S})$, cioè quegli spazi misurabili $(S, \mathcal{S})$ per cui esiste un insieme $E \in \mathcal{B}(\mathbb{R})$ e una funzione biunivoca $\phi : S \to E$ che sia $(S, \mathcal{S}) - (E, \mathcal{B}(\mathbb{R})|_E)$ misurabile, e la cui inversa sia $(E, \mathcal{B}(\mathbb{R})|_E) - (S, \mathcal{S})$ misurabile.

In particolare quindi il risultato di esistenza della versione regolare è valido per $(S, \mathcal{S}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ o $(S, \mathcal{S}) = (\mathbb{R}^{\mathbb{N}}, \mathcal{R}^{\mathbb{N}})$, che sono spazi di Borel (vedere ad esempio P. Billingsley, "Convergence of Probability measures" pag. 218 e seguenti).

Prima della dimostrazione va notato che questo fatto permette di applicare il risultato anche al caso in cui siano coinvolte le variabili aleatorie X_n, $n \ge 1$, ed X (confrontare di nuovo l'osservazione alla proprietà della convergenza dominata per i valori attesi condizionali).

La dimostrazione è basata sull'osservazione che se Y è una variabile aleatoria a valori in S, allora $X := \phi(Y)$ è una variabile aleatoria reale,e quindi esiste una distribuzione condizionale $\mu_X(\cdot \mid \mathcal{G})$. La distribuzione condizionale di Y su $(S, \mathcal{S})$ si ottiene, per $J \in \mathcal{S}$, come

$$\nu_Y(J \mid \mathcal{G})(\omega) = \mu_X(\phi(J) \mid \mathcal{G})(\omega)$$

in quanto la seconda è una versione di

$$\mathbb{P}(\{X \in \phi(J)\} \mid \mathcal{G})(\omega) \equiv \mathbb{P}(\{\phi^{-1}(X) \in J\} \mid \mathcal{G})(\omega) \equiv \mathbb{P}(\{Y \in J\} \mid \mathcal{G})(\omega).$$

Abbiamo già usato il fatto che ogni funzione $\sigma(X)-$misurabile si può esprimere come $f(X)$. Supponiamo ora che $\mathcal{G} = \sigma(Y)$ allora, necessariamente deve accadere che $\mu_X(H)(\omega)$ sia una funzione di $Y(\omega)$, ovvero che, per ogni boreliano H, esista una funzione $f(H, y)$, misurabile in y, per cui

1) $f(H, Y(\omega)) = \mu_X(H)(\omega)$

2) $f(\cdot, y)$ sia una misura di probabilità per ogni y

(questo è sicuramente vero se $y \in Y(\Omega)$, mentre se $y \notin Y(\Omega)$ basta definire $f(H, y) = \nu(H)$ per una fissata misura di probabilità ν).

Indicheremo $f(H, y)$ con la notazione più evocativa di $\mu_X(H \mid Y = y)$ o anche $\mu_{X|Y}(H \mid y)$. Analogamente indicheremo con $F_X(t \mid Y = y)$ o con $F_{X|Y}(t \mid y)$ la funzione di distribuzione condizionale di X data Y, "valutata in $Y = y$". Ovviamente tale espressione non va confusa con $\mathbb{P}(X \leq t \mid \{Y = y\})$, che tra l'altro potrebbe non avere senso nel caso in cui $\mathbb{P}(\{Y = y\}) = 0$.

ESEMPIO 1 (caso dominato)
Si tratta della generalizzazione dell'ESEMPIO 2.3. Siano X ed Y due variabili aleatorie con legge congiunta assolutamente continua rispetto ad una misura prodotto $\nu_1(dx) \times \nu_2(dy)$ cioè con legge congiunta data da

$$\mu_{X,Y}(dx, dy) = f(x, y)\nu_1(dx) \times \nu_2(dy).$$

È facile vedere che la legge di X è assolutamente continua rispetto a $\nu_1(dx)$ e la legge di Y lo è rispetto a $\nu_2(dy)$, o meglio

$$\mu_X(dx) = \left[\int f(x, y)\nu_2(dy)\right]\nu_1(dx) = f_X(x)\nu_1(dx),$$

$$\mu_Y(dy) = \left[\int f(x, y)\nu_1(dx)\right]\nu_2(dy) = f_Y(y)\nu_2(dy).$$

In questo caso anche la legge condizionale di X data Y è assolutamente continua rispetto a $\nu_1(dx)$ e risulta

$$\mu_{X|Y}(dx|y) = \frac{f(x, y)}{f_Y(y)}\nu_1(dx),$$

per μ_Y-quasi ogni y.

Infatti è facile verificare che, per ogni funzione $g : \mathbb{R} \to \mathbb{R}$, misurabile,

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[\phi(Y)h(Y)], \quad \text{per ogni funzione h misurabile} \qquad (*)$$

dove

$$\phi(y) = \int g(x)\mu_{X|Y}(dx|y) = \int g(x)\frac{f(x, y)}{f_Y(y)}\nu_1(dx);$$

(ovviamente g ed h devono soddisfare ipotesi che garantiscano l'integrabilità delle v.a. $g(X)$ ed $h(Y)$)

Tutte le v.a. W che siano $\sigma(Y)-$misurabili sono del tipo $W = h(Y)$ per h boreliana, di conseguenza

$$\mathbb{E}[g(X) \mid Y] = \phi(Y) = \int g(x)\; \mu_{X|Y}(dx|y)\Big|_{y=Y(\omega)} = \int g(x)\; \frac{f(x, y)}{f_Y(y)}\nu_1(dx)\Big|_{y=Y(\omega)}$$

Un risultato analogo vale ovviamente anche per la legge di Y data X.
Casi particolari sono i casi in cui $\nu_1(dx) = dx$, e $\nu_2(dy) = dy$, cioè torniamo al caso della misura di Lebesgue esaminato nell'Esempio 2.3, oppure $\nu_1(dx) = \sum_{n=0}^{\infty} \delta_{x_n}(dx)$, e $\nu_2(dy) = \sum_{n=0}^{\infty} \delta_{y_n}(dy)$ e si trova allora che

$$\mathbb{E}[g(X) \mid Y] = \sum_n g(x_n)\, \mu_{X|Y}(\{x_n\}|y)\big|_{y=Y(\omega)} = \sum_n g(x_n)\, \mathbb{P}(X = x_n \mid Y = y)\big|_{y=Y(\omega)}.$$

ESEMPIO 2 (caso non dominato)

Siano X ed Y due variabili aleatorie con legge congiunta data da

$$\mu_{X,Y}(dx, dy) = p\, f_1(x)\, dx\, \delta_{\alpha(x)}(dy) + q\, f_2(x, y)\, dx\, dy,$$

dove

$\delta_z(dx)$ è la misura concentrata in z (ovvero $\delta_z(A) = 1$ se $z \in A$, mentre $\delta_z(A) = 0$ se $z \notin A$),

$p + q = 1$,

$f_1(x)$ è una densità di probabilità su $\mathbb{R}$

$f_2(x, y)$ è una densità di probabilità su $\mathbb{R}^2$,

α è una funzione invertibile, C^1 e con $\mid \alpha'(x) \mid$ strettamente positiva.

Le distribuzioni marginali sono equivalenti alla misura di Lebesgue, essendo

$$\mu_X(dx) = \left[p f_1(x) + q \int f_2(x, y) dy\right] dx,$$

$$\mu_Y(dy) = \left[p f_1(\alpha^{-1}(y)) \frac{1}{|\alpha'(\alpha^{-1}(y))|} + q \int f_2(x, y) dx\right] dy.$$

L'ultima uguaglianza deriva da

$$\mathbb{E}[h(Y)] = p \int h(\alpha(x)) f_1(x) dx + q \int h(y) dy \int f_2(x, y) dx =$$

$$= p \int h(y) f_1(\alpha^{-1}(y)) \frac{1}{|\alpha'(\alpha^{-1}(y))|} dy + q \int h(y) dy \int f_2(x, y) dx = \int h(y) \mu_Y(dy)$$

Come si procede per calcolare la media condizionata di $g(X)$ data Y?

Come nell'esempio precedente si deve trovare una v.a. $\sigma(Y)$-misurabile, cioè una funzione $\phi(Y)$ per cui valga

$$\mathbb{E}[g(X) h(Y)] = \mathbb{E}[\phi(Y) h(Y)], \qquad \text{per ogni funzione } h \text{ misurabile} \qquad (*)$$

Se poi troviamo che

$$\phi(y) = \int g(x) \nu(dx; y)$$

per una misura di probabilità $\nu(\cdot; y)$, allora potremo affermare che $\nu(\cdot; y)$ è una versione regolare della distribuzione condizionale di X data $Y = y$.

Ora, qualunque sia la funzione h

$$\begin{aligned}
\mathbb{E}[g(X)h(Y)] &= p \iint g(x)h(y)f_1(x)dx\delta_{\alpha(x)}(dy) + q \iint g(x)h(y)f_2(x,y)dxdy \\
&= p \int g(x)h(\alpha(x))f_1(x)dx + q \int \Big(\int g(x)f_2(x,y)dx\Big)h(y)dy \\
&= p \int g(\alpha^{-1}(y))f_1(\alpha^{-1}(y))\frac{1}{|\alpha'(\alpha^{-1}(y))|}h(y)dy + q \int \Big(\int g(x)f_2(x,y)dx\Big)h(y)dy \\
&= \int \left[pg(\alpha^{-1}(y))f_1(\alpha^{-1}(y))\frac{1}{|\alpha'(\alpha^{-1}(y))|} + q\Big(\int g(x)f_2(x,y)dx\Big)\right] h(y)dy,
\end{aligned}$$

mentre invece

$$\mathbb{E}[\phi(Y)h(Y)] = p \int \phi(y)h(y)f_1(\alpha^{-1}(y))\frac{1}{|\alpha'(\alpha^{-1}(y))|}dy + q \int \phi(y)h(y)dy \int f_2(x,y)dx$$

$$= \int \phi(y) \left[pf_1(\alpha^{-1}(y))\frac{1}{|\alpha'(\alpha^{-1}(y))|} + q\Big(\int f_2(x,y)dx\Big)\right] h(y)dy.$$

Quindi affinché valga l'uguaglianza (*), qualunque sia h, è necessario e sufficiente che

$$\left[pg(\alpha^{-1}(y))f_1(\alpha^{-1}(y))\frac{1}{|\alpha'(\alpha^{-1}(y))|} + q\Big(\int g(x)f_2(x,y)dx\Big)\right] =$$

$$= \phi(y) \left[pf_1(\alpha^{-1}(y))\frac{1}{|\alpha'(\alpha^{-1}(y))|} + q\Big(\int f_2(x,y)dx\Big)\right]$$

ovvero che

$$\phi(y) = \frac{pg(\alpha^{-1}(y))f_1(\alpha^{-1}(y))\frac{1}{|\alpha'(\alpha^{-1}(y))|} + q\Big(\int g(x)f_2(x,y)dx\Big)}{pf_1(\alpha^{-1}(y))\frac{1}{|\alpha'(\alpha^{-1}(y))|} + q\Big(\int f_2(x,y)dx\Big)}.$$

Si noti ancora che

$$g(\alpha^{-1}(y)) = \int g(x)\delta_{\alpha^{-1}(y)}(dx).$$

Per questo motivo, qualunque sia g, si può riscrivere il numeratore come

$$pf_1(\alpha^{-1}(y))\frac{1}{|\alpha'(\alpha^{-1}(y))|} \int g(x)\delta_{\alpha^{-1}(y)}(dx) + q\Big(\int g(x)f_2(x,y)dx\Big)$$

e quindi la legge $\mu_{X|Y}(dx|y)$ di X, condizionata ad $Y = y$, è proporzionale a

$$pf_1(\alpha^{-1}(y))\frac{1}{|\alpha'(\alpha^{-1}(y))|}\delta_{\alpha^{-1}(y)}(dx) + qf_2(x,y)dx.$$

Si noti che quindi la legge di X condizionata a $Y = y$ non è assolutamente continua rispetto alla distribuzione iniziale di X, che ha invece una densità rispetto alla misura di Lebesgue.

5 MARTINGALE

Per definire una martingala abbiamo bisogno di dare prima la seguente

Definizione 5.1 (FILTRAZIONE). *In uno spazio $(\Omega, \mathcal{F})$, la famiglia $\{\mathcal{F}_t, t \geq 0\}$ si dice una filtrazione se è una famiglia crescente di σ-algebre di $\mathcal{F}$, cioè $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$, per $0 \leq s \leq t$. (La definizione ha senso anche nel caso in cui $t = n \in \mathbb{N}$, e nel caso in cui $t \in I \subseteq \mathbb{R}$, ad esempio $t \in [0, T]$).*

La $\sigma-$algebra rappresenta l'informazione disponibile fino al tempo t.

Definizione 5.2 (MARTINGALA). *Un processo aleatorio X_t si dice una martingala rispetto ad una filtrazione $\{\mathcal{F}_t\}$ se*

0) *X_t è $\mathcal{F}_t$-misurabile per ogni $t \geq 0$. (o più rapidamente il processo X_t è adattato ad $\mathcal{F}_t$)*

1) *X_t è integrabile per ogni $t \geq 0$, cioè $\mathbb{E}[|X_t|] < +\infty$ per ogni $t \geq 0$.*

2) *$\mathbb{E}[X_{t+s} \mid \mathcal{F}_t] = X_t$, per ogni $t,s \geq 0$*

Nel caso in cui $t=n \in \mathbb{N}$ la 2) può essere sostituita con la richiesta che

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n, \text{ per ogni } n \in \mathbb{N}.$$

Si noti che la proprietà 0) è sovrabbondante in quanto la 2) implica che X_t sia $\mathcal{F}_t$-misurabile. Ciò non è vero nella seguente definizione:

Definizione 5.3 (SUBMARTINGALA (SUPERMARTINGALA)). *Un processo aleatorio X_t si dice una* ***submartingala (supermartingala)*** *rispetto ad una filtrazione $\{\mathcal{F}_t\}$ se*

0) *X_t è $\mathcal{F}_t$-misurabile per ogni $t \geq 0$. (o più rapidamente il processo X_t è adattato ad $\mathcal{F}_t$)*

1) *X_t è integrabile per ogni $t \geq 0$, cioè $\mathbb{E}[| X_t |] < +\infty$ per ogni $t \geq 0$.*

2) *$\mathbb{E}[X_{t+s} \mid \mathcal{F}_t] \geq X_t (\mathbb{E}[X_{t+s} \mid \mathcal{F}_t] \leq X_t)$, per ogni $t, s \geq 0$,*

Di nuovo, nel caso in cui $t = n \in \mathbb{N}$ la 2) può essere sostituita con la richiesta che

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \geq X_n (\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \leq X_n), \text{ per ogni } n \in \mathbb{N}.$$

Osservazione 5.1. *OSSERVAZIONE 1 Se X_t è una martingala (o submartingala) rispetto a una filtrazione $\{\mathcal{F}_t\}$ e se $\{\mathcal{G}_t\}$ è una filtrazione per cui $\sigma(X_t) \subseteq \mathcal{G}_t \subseteq \mathcal{F}_t$, allora X_t lo è anche rispetto alla nuova filtrazione:*

$$\mathbb{E}[X_{t+s} \mid \mathcal{G}_t] = \mathbb{E}[\mathbb{E}[X_{t+s} \mid \mathcal{F}_t] \mid \mathcal{G}_t] = \mathbb{E}[X_t \mid \mathcal{G}_t] = X_t.$$

In particolare per ogni martingala (o submartingala) si può prendere sempre la filtrazione minimale $\mathcal{F}_t^X = \sigma\{X_u, u \leq t\}$, che per la proprietà 0) è sempre contenuta in $\mathcal{F}_t$. Quindi, in genere, se la filtrazione non è specificata, si deve intendere che si tratti della filtrazione naturale.

Se invece $\{\mathcal{H}_t\}$ è una filtrazione per cui $\mathcal{H}_t$ è indipendente da $\mathcal{F}_t$ (e quindi da X_t), allora X_t è una martingala (o submartingala) anche rispetto alla filtrazione $\mathcal{G}_t := \mathcal{F}_t \vee \mathcal{H}_t$ (confrontare la proprietà 6 delle medie condizionali: condizionamento ridondante).

Osservazione 5.2. *OSSERVAZIONE 2 Ogni martingala ha media costante e ogni submartingala ha media crescente (in senso lato): basta passare al valore medio nella proprietà 3 delle medie condizionali (condizionamenti successivi) e tenere presente che il valore medio di Y coincide con il valore medio di $\mathbb{E}[Y \mid \mathcal{G}]$.*

Osservazione 5.3. *OSSERVAZIONE 3 Data una submartingala X_t si ottiene una supermartingala considerando $-X_t$, e viceversa.*

ESEMPI DI MARTINGALE E DI SUBMARTINGALE

Esempio 5.1. *Sia data una variabile aleatoria Y integrabile ed una filtrazione $\{\mathcal{F}_t\}$. Si definisca $X_t := \mathbb{E}[Y \mid \mathcal{F}_t]$. Si dimostra facilmente che X_t è una martingala:*

$$\mathbb{E}[X_{t+s} \mid \mathcal{F}_t] = \mathbb{E}[\mathbb{E}[Y \mid \mathcal{F}_{t+s}] \mid \mathcal{F}_t] = \mathbb{E}[Y \mid \mathcal{F}_t]$$

Questo esempio aiuta a capire il nome di submartingala: sia ora X_t una submartingala, allora, fissato $v = t + s$, il processo

$$Z_t := \mathbb{E}[X_{t+s} \mid \mathcal{F}_t] \equiv \mathbb{E}[X_v \mid \mathcal{F}_t],$$

è una martingala per $t \in [0, v]$, quindi la proprietà 2) per le submartingale diviene $Z_t := \mathbb{E}[X_v \mid \mathcal{F}_t] \geq X_t$, ovvero che X_t sia sotto *la martingala*

$$Z_t := \mathbb{E}[X_v \mid \mathcal{F}_t].$$

Esempio 5.2. *Sia data una successione di variabili aleatorie Y_n indipendenti, identicamente distribuite, integrabili e a media nulla. Si definisca*

$$S_n = \sum_{k=1}^{n} Y_k$$

Si ottiene facilmente che S_n è una martingala (rispetto a $\mathcal{F}_n^S = \sigma\{Y_1, \cdots, Y_n\}$):

$$\mathbb{E}[S_{n+1} \mid \mathcal{F}_n] = \mathbb{E}[S_n + Y_{n+1} \mid \mathcal{F}_n] = S_n + \mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] = S_n + \mathbb{E}[Y_{n+1}] = S_n.$$

Se le v.a. Y_k non sono a media nulla, basta sostituire $Y_k - \mu$, se $\mu = \mathbb{E}[Y_1]$, ed ottenere che

$$\widetilde{S}_n = \sum_{k=1}^{n}(Y_k - \mu) = \sum_{k=1}^{n} Y_k - n\mu = S_n - n\mu$$

è una martingala.

Esempio 5.3. *Se siamo nelle stesse ipotesi del precedente Esempio 5.2*

ed inoltre le v.a. Y_k ammettono momento secondo finito, e quindi varianza σ^2, allora $M_n := S_n^2 - n\sigma^2$ è una martingala:

$$M_{n+1} - M_n = S_{n+1}^2 - (n+1)\sigma^2 - (S_n^2 - n\sigma^2) = (S_n^2 + 2Y_{n+1}S_n + Y_{n+1}^2) - S_n^2 - \sigma^2 = 2Y_{n+1}S_n + Y_{n+1}^2 - \sigma^2$$

Passando alle medie condizionali si ottiene che

$$\mathbb{E}[M_{n+1} - M_n \mid \mathcal{F}_n] = \mathbb{E}[2Y_{n+1}S_n + Y_{n+1}^2 - \sigma^2 \mid \mathcal{F}_n]$$

$$= 2S_n\mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] + \mathbb{E}[Y_{n+1}^2 \mid \mathcal{F}_n] - \sigma^2 = 2S_n \cdot 0 + \sigma^2 - \sigma^2 = 0$$

in quanto Y_{n+1} *è indipendente da* $\mathcal{F}_n$ *e quindi i valori medi condizionali* $\mathbb{E}[Y_{n+1} \mid \mathcal{F}_n]$ *ed* $\mathbb{E}[Y_{n+1}^2 \mid \mathcal{F}_n]$ *coincidono con i rispettivi valori medi.*

Se le v.a. Y_k *non sono a media nulla, allora*

$$M_n := (S_n - n\mu)^2 - n\sigma^2$$

è una martingala.

Esempio 5.4. *Sia* X_t *una martingala con* $\mathbb{E}[\mid X_t \mid^\alpha] < +\infty$*, per un* $\alpha \geq 1$*. Allora il processo* $\mid X_t \mid^\alpha$ *è una submartingala: basta applicare la disuguaglianza di Jensen alla funzione* $|x|^\alpha$*, che è convessa per* $\alpha \geq 1$

$$\mid X_t \mid^\alpha = \mid \mathbb{E}[X_{t+s} \mid \mathcal{F}_t] \mid^\alpha \leq \mathbb{E}[\mid X_{t+s} \mid^\alpha \mid \mathcal{F}_t].$$

Questo esempio si generalizza immediatamente al caso di ogni funzione convessa ϕ*, purché, ovviamente,* $\mathbb{E}[\phi(X_t)] < +\infty$*.*

Per ottenere un risultato analogo nel caso in cui X_t *sia una submartingala bisogna aggiungere l'ipotesi che* ϕ *sia una funzione crescente, cosi' che, essendo* $X_t \leq \mathbb{E}[X_{t+s} \mid \mathcal{F}_t]$*,*

$$\phi(X_t) \leq \phi(\mathbb{E}[X_{t+s} \mid \mathcal{F}_t]) \leq \mathbb{E}[\phi(X_{t+s}) \mid \mathcal{F}_t].$$

Esempio 5.5. *Se le variabili aleatorie* W_k *sono i.i.d. con* $\mathbb{E}[W_1] = 1$*, allora la successione di v.a.*

$$Z_n := \prod_{k=1}^{n} W_k$$

definisce una martingala:

$$\mathbb{E}[Z_{n+1} \mid \mathcal{F}_n] = \mathbb{E}[Z_n W_{n+1} \mid \mathcal{F}_n] = Z_n\mathbb{E}[W_{n+1} \mid \mathcal{F}_n] = Z_n\mathbb{E}[W_{n+1}] = Z_n.$$

Come caso particolare si consideri la situazione dell'Esempio 5.2 con l'ulteriore ipotesi che per un $\theta \in \mathbb{R}$ *valga*

$$\mathbb{E}[\exp\{\theta Y_1\}] = \exp\{\psi(\theta)\} < +\infty$$

(si noti che non è necessario supporre $\mathbb{E}[Y_1] = 0$*). Allora ovviamente*

$$\mathbb{E}[\exp\{\theta Y_1\} \exp\{-\psi(\theta)\}] = \mathbb{E}[\exp\{\theta Y_1 - \psi(\theta)\}] = 1.$$

Come conseguenza

$$Z_n = \exp\{\theta S_n - n\psi(\theta)\}$$

è una martingala positiva di media 1*.*

Esempio 5.6. *Dato uno spazio* $(\Omega, \mathcal{F})$ *e su di esso una filtrazione* $\mathcal{F}_t$ *e due misure di probabilità* $\mathbb{P}$ *e* $\mathbb{Q}$*, con* $\mathbb{P}$ *assolutamente continua rispetto a* $\mathbb{Q}$ *(di conseguenza lo sono anche rispetto ad* $\mathcal{F}_t$ *per ogni* t*). Si definisca la derivata di Radon Nikodym di* $\mathbb{P}$ *rispetto a* $\mathbb{Q}$*, entrambe ristrette a* $\mathcal{F}_t$*, cioè*

$$L_t := \left.\frac{d\mathbb{P}}{d\mathbb{Q}}\right|_{\mathcal{F}_t}.$$

Il processo L_t è una $\mathcal{F}_t$-martingala nello spazio $(\Omega, \mathcal{F}, \mathbb{Q})$, anzi più in generale risulta che se X_t è un processo adattato ad $\mathcal{F}_t$, allora X_t è una $\mathcal{F}_t$-martingala nello spazio $(\Omega, \mathcal{F}, \mathbb{P})$ se e solo se X_tL_t è una $\mathcal{F}_t$-martingala nello spazio $(\Omega, \mathcal{F}, \mathbb{Q})$ (quindi il caso precedente deriva prendendo banalmente $X_t \equiv 1$).

Infatti X_t è una martingala in $(\Omega, \mathcal{F}, \mathbb{P})$, se e solo è integrabile rispetto a $\mathbb{P}$ e se per $0 \leq s \leq t$ e $A \in \mathcal{F}_s$

$$\mathbb{E}^{\mathbb{P}}[I_A X_t] = \mathbb{E}^{\mathbb{P}}[I_A X_s],$$

mentre X_tL_t lo è in $(\Omega, \mathcal{F}, \mathbb{Q})$ se e solo se è integrabile rispetto a $\mathbb{Q}$ e se per $0 \leq s \leq t$ e $A \in \mathcal{F}_s$

$$\mathbb{E}^{\mathbb{Q}}[I_A X_t L_t] = \mathbb{E}^{\mathbb{Q}}[I_A X_s L_s].$$

Ovviamente, essendo $I_A X_s$ una v.a. $\mathcal{F}_s$-misurabile, si ha $\mathbb{E}^{\mathbb{P}}[I_A X_s] = \mathbb{E}^{\mathbb{Q}}[I_A X_s L_s]$ e, essendo $I_A X_t$ una v.a. $\mathcal{F}_t$-misurabile, in quanto $A \in \mathcal{F}_s \subseteq \mathcal{F}_t$, si ha $\mathbb{E}^{\mathbb{P}}[I_A X_t] = \mathbb{E}^{\mathbb{Q}}[I_A X_t L_t]$. La verifica dell'integrabilità è banale.

Esempio 5.7. *Sia X_n una catena di Markov con spazio degli stati finito e con probabilità di transizione $p_{i,j}$. Sia inoltre h una* ***funzione armonica*** *rispetto alla matrice delle probabilitaà di transizione $P = (p_{i,j})_{i,j}$, cioè*

$$h(i) = (Ph)(i) := \sum_j p_{i,j} h(j), \qquad \textit{per ogni } i$$

(si noti che ciò corrisponde a chiedere che h sia la soluzione di $(P - I)h = 0$).

Il processo

$$M_n^h := h(X_n)$$

è una martingala (rispetto a $\mathcal{F}_n^X = \sigma\{X_k,\ k = 0, \cdots, n\}$).
Questo risultato deriva da un caso più generale: qualunque sia f

$$M_n^f := f(X_n) - \sum_{k=0}^{n-1} (P - I) f(X_k)$$

è una martingala rispetto a $\mathcal{F}_n^X$.

Cominiciamo con il caso h armonica. Basta controllare che

$$\mathbb{E}[M_{n+1}^h \mid \mathcal{F}_n^X] = M_n^h,$$

ovvero che

$$\mathbb{E}[h(X_{n+1}) \mid X_n, X_{n-1}, \cdots, X_0] = h(X_n).$$

Essendo $(X_n, n \geq 0)$ *una catena di Markov si ha*[1]

$$\mathbb{E}[f(X_{n+1}) \mid X_n, X_{n-1}, \cdots, X_0] = (Pf)(X_n)$$

ed il caso $f = h$ *armonica è immediato. Il caso generale deriva dall'osservare che*

$$M^f_{n+1} - M^f_n = f(X_{n+1}) - \sum_{k=0}^{n}(P-I)f(X_k) - f(X_n) + \sum_{k=0}^{n-1}(P-I)f(X_k)$$

$$M^f_{n+1} - M^f_n = f(X_{n+1}) - (P-I)f(X_n) - f(X_n) = f(X_{n+1}) - (Pf)(X_n),$$

e quindi

$$\mathbb{E}[M^f_{n+1} - M^f_n \mid \mathcal{F}^X_n] = \mathbb{E}[f(X_{n+1}) - (Pf)(X_n) \mid \mathcal{F}^X_n] = \mathbb{E}[f(X_{n+1}) \mid \mathcal{F}^X_n] - (Pf)(X_n) = 0.$$

6 DECOMPOSIZIONE DI DOOB

Teorema 6.1 (Decomposizione di Doob). *: Data una submartingala* X_n *a tempo discreto, essa si può sempre scrivere in modo unico (a meno di insiemi di misura nulla) come*

$$X_n = X_0 + M_n + A_n,$$

dove M_n *è una martingala ed* A_n *è un processo predicibile (cioè, per ogni* $n \in \mathbb{N}$, A_n *è* $\mathcal{F}_{n-1}$*-misurabile) crescente in senso lato, con* $A_0 = 0$.

Dimostrazione

Se una tale decomposizione esiste necessariamente deve accadere che $M_0 = 0$, ed inoltre, essendo $A_n = X_n - X_0 - M_n$, deve accadere che

$$A_{n+1} - A_n = X_{n+1} - X_0 - M_{n+1} - (X_n - X_0 - M_n) = X_{n+1} - X_n - (M_{n+1} - M_n).$$

Essendo $A_{n+1} - A_n$ una v.a. $\mathcal{F}_n$-misurabile, passando alla media condizionale rispetto ad $\mathcal{F}_n$, essa non cambia, per cui deve necessariamente accadere che

$$A_{n+1} - A_n = \mathbb{E}[X_{n+1} - X_n - (M_{n+1} - M_n) \mid \mathcal{F}_n] = \mathbb{E}[X_{n+1} \mid \mathcal{F}_n] - X_n.$$

Quindi l'unico modo per definire A_{n+1} è il seguente

$$A_{n+1} = \sum_{k=0}^{n}(A_k + 1 - A_k) = \sum_{k=0}^{n}(\mathbb{E}[X_{k+1} \mid \mathcal{F}_k] - X_k).$$

[1] Nel caso di variabili aleatorie discrete possiamo applicare i risultati sulle densità condizionali: posto $\mathbf{X} = (X_n, X_{n-1}, \cdots, X_0)$, sappiamo che

$$\mathbb{E}[f(X_{n+1}) \mid \mathbf{X}](\omega) = \sum_j f(j)\, \mathbb{P}(X_{n+1} = j \mid \{\mathbf{X} = \mathbf{x}\})|_{\mathbf{x}=\mathbf{X}(\omega)}$$

Per la proprietà di Markov $\mathbb{P}(X_{n+1} = j \mid \mathbf{X} = \mathbf{x}) = \mathbb{P}(X_{n+1} = j \mid X_n = x_n) = p_{x_n,j}$, quindi $\mathbb{P}(X_{n+1} = j \mid \mathbf{X}) = \mathbb{P}(X_{n+1} = j \mid X_n) = p_{X_n,j}$ e $\sum_j f(j)\mathbb{P}(X_{n+1} = j \mid \mathbf{X} = \mathbf{x}) = (Pf)(x_n)$ e perciò $\mathbb{E}[f(X_{n+1}) \mid X_n, X_{n-1}, \cdots, X_0] = (Pf)(X_n)$

Si tratta ora solo di verificare che con questa definizione di $\{A_n\}_{n\geq 0}$ il processo $M_n := X_n - X_0 - A_n$ è una martingala. Ma ovviamente

$$M_{n+1} - M_n = X_{n+1} - X_0 - A_{n+1} - (X_n - X_0 - A_n) = X_{n+1} - X_n - (A_{n+1} - A_n),$$

per cui, tenendo conto che per definizione $A_n + 1 - A_n = \mathbb{E}[X_{n+1} \mid \mathcal{F}_n] - X_n$, e passando alla media condizionale, si ottiene la tesi.

7 MARTINGALE, SUBMARTINGALE E TEMPI D'ARRESTO

Definizione 7.1. *. Sia $\{\mathcal{F}_t\}$ una filtrazione e sia $\tau : \Omega \to \mathbb{R}^+ \cup \{\infty\}$, una variabile aleatori (la v.a. τ deve prendere valori negli indici del tempo preso in considerazione e quindi, ad esempio, in $\mathbb{N}$ se si tratta tempo discreto, ma può anche prendere il valore infinito). La v.a. τ si dice* ***tempo d'arresto*** *(o* ***stopping time****) rispetto ad $\{\mathcal{F}_t\}$, se per ogni t*

$$\{\tau \leq t\} \in \mathcal{F}_t$$

Si dice inoltre che τ è un ***tempo d'arresto finito*** *se*

$$\mathbb{P}(\tau < +\infty) = 1$$

e che τ è un ***tempo d'arresto limitato*** *se esiste un numero $L< +\infty$ per cui*

$$\mathbb{P}(\tau < L) = 1$$

(Si noti che a volte la filtrazione in considerazione è ovvia, e quindi non viene specificata.)

Nel caso in cui l'insieme dei tempi sia $\mathbb{N}$ è equivalente chiedere che $\{\tau = n\} \in \mathcal{F}_n$ per ogni n in quanto ovviamente

$$\{\tau = n\} = \{\tau \leq n\} \setminus \{\tau \leq n-1\}$$

e quindi se $\{\tau \leq n\} \in \mathcal{F}_n$ e $\{\tau \leq n-1\} \in \mathcal{F}_{n-1} \subseteq \mathcal{F}_n$ allora $\{\tau = n\} \in \mathcal{F}_n$ per ogni n, mentre

$$\{\tau \leq n\} = \cup_{k=0}^{n} \{\tau = k\}$$

e quindi se $\{\tau = k\} \in \mathcal{F}_k$ per ogni k, essendo $\mathcal{F}_k \subseteq \mathcal{F}_n$ per $k \leq n$, si ha che $\{\tau \leq n\} \in \mathcal{F}_n$ per ogni n.

Esempio 7.1. *Con le stesse notazioni dell'Esempio 5.3, e per ogni $a \in \mathbb{R}$,*

$$\tau(\omega) = \inf\{n : S_n \geq a\}, \ \textit{(con la convenzione che } \inf\{\phi\} = +\infty)$$

è un tempo d'arresto rispetto ad $\mathcal{F}_n^S$, in quanto per decidere se l'evento $\{\tau \leq k\}$ si è verificato, basta esaminare le prime k v.a. $S_1, \cdots, S_k$.

Invece $\sigma(\omega) = \sup\{n \le 10 : S_n \ge a\}$, *se un tale* n *esiste e* 10 *altrimenti, non è un tempo d'arresto, in quanto, ad esempio, per decidere se l'evento* $\{\sigma \le 3\}$ *si è verificato, bisogna esaminare tutte le v.a.* $S_1, \cdots, S_{10}$ *e non solo* S_1, S_2, S_3, *perciò* $\{\sigma \le 3\}$ *non è misurabile rispetto a* $\mathcal{F}_3^S$.

Definizione 7.2. . *Dato un tempo d'arresto* τ, *si definisce* **la $\boldsymbol{\sigma}$-algebra degli eventi fino al tempo** τ *come*

$$\mathcal{F}_\tau = \{A \in \mathcal{F}_\infty, \text{ per cui } A \cap \{\tau \le t\} \in \mathcal{F}_t \text{ per ogni } t\}$$

dove $\mathcal{F}_\infty = \bigvee_t \mathcal{F}_t$.

Si tratta cioè degli eventi, per i quali stabilire il loro verificarsi insieme al verificarsi di $\{\tau \le t\}$ dipende solo dall'informazione disponibile fino al tempo t.

ESERCIZIO:
Controllare che $\mathcal{F}_\tau$ è una σ-algebra.
(suggerimento: se $A \in \mathcal{F}_\tau$ allora l'evento $A^c \cap \{\tau \le t\} = \{\tau \le t\} \setminus \{A \cap \{\tau \le t\}\} \in \mathcal{F}_t$)

Esempio 7.2. *Se* $\mathcal{F}_n = \sigma\{X_1, \cdots, X_n\}$ *e* $\tau = \inf\{n \text{ t.c. } X_n \in I\}$, *con* $I \in \mathcal{B}(\mathbb{R})$, *allora* τ *è un* $\mathcal{F}_n$ *tempo d'arresto, infatti l'evento* $\{\tau \le n\} = \{\exists k \le n \text{ t.c. } X_k \in I\} \in \mathcal{F}_n$.

Nel caso tempo continuo non è cosi' semplice, però qualcosa si può dire.

Definizione 7.3. *Sia*

$$\tau_A = \inf\{t > 0 \text{ t.c. } X_t \notin A\},$$

con la convenzione che l'estremo inferiore dell'insieme vuoto è uguale a $+\infty$.

La v. a. τ_A *è detta* ***tempo di prima uscita da*** A.

Lemma 7.1 (LEMMA 1). *Se* X_t *è un processo a traiettorie continue e* A *è aperto, allora* τ_A *è un tempo d'arresto.*

Dimostrazione Si tratta di notare che, essendo A^c chiuso la funzione $x \mapsto \operatorname{dist}(x, A^c)$ è continua, e di conseguenza, essendo X_t a traiettorie continue, si ha che la funzione $s \mapsto \operatorname{dist}(X_s, A^c)$ è continua. Perciò

$$\inf_{0 \le s \le t} \operatorname{dist}(X_s, A^c) = \min_{0 \le s \le t} \operatorname{dist}(X_s, A^c)$$

e quindi

$$\{\tau_A > t\} = \{\inf_{0 \le s \le t} \operatorname{dist}(X_s, A^c) > 0\} = \bigcup_n \{\inf_{0 \le s \le t} \operatorname{dist}(X_s, A^c) > \frac{1}{n}\} =$$

$$= \bigcup_n \{\inf_{\substack{0 \le s \le t \\ s \in \mathbb{Q}}} \operatorname{dist}(X_s, A^c) \ge \frac{1}{n}\} = \bigcup_n \bigcap_{\substack{0 \le s \le t \\ s \in \mathbb{Q}}} \{\operatorname{dist}(X_s, A^c) \ge \frac{1}{n}\} \in \mathcal{F}_t$$

Si noti che se il $\inf_{0 \le s \le t} \operatorname{dist}(X_s, A^c)$ non fosse un minimo, allora potrebbero verificarsi contemporaneamente gli eventi

$$\{\tau_A > t\} \text{ e } \{\inf_{0 \le s \le t} \operatorname{dist}(X_s, A^c) = 0\},$$

e la prima delle precedenti uguaglianze non sarebbe valida.

Nel caso in cui l'insieme A non sia aperto non è detto che τ_A sia un tempo d'arresto. Se A è un **insieme chiuso** allora τ_A è un **tempo d'arresto in senso debole**, ovvero

$$\{\tau_A < t\} \in \mathcal{F}_t \text{ per ogni } t \ge 0.$$

OSSERVAZIONE

Affermare che τ è un tempo d'arresto debole è equivalente ad affermare che è un tempo di arresto rispetto alla filtrazione

$$\mathcal{F}_{t^+} := \bigcap_{s>t} \mathcal{F}_s.$$

Infatti in generale, se $\{\tau < t\} \in \mathcal{F}_t$ per ogni $t \ge 0$, allora, qualunque sia $m \ge 1$

$$\{\tau \le t\} = \bigcap_{n \ge m} \{\tau < t + \frac{1}{n}\} \in \mathcal{F}_{t+\frac{1}{m}},$$

e quindi $\{\tau \le t\} \in \mathcal{F}_s$ per ogni $s > t$, ovvero [23] $\{\tau \le t\} \in \mathcal{F}_{t^+}$ per ogni t.

LEMMA 2. Se X_t è un processo con traiettorie continue a destra con limiti a sinistra (*cadlag* acronimo dal francese *continue a droite limite a gauche*), la filtrazione è continua a destra (cioè $\mathcal{F}_t = \mathcal{F}_{t^+} := \bigcap_{s>t} \mathcal{F}_s$) ed F è un chiuso allora τ_F è un tempo d'arresto.

Dimostrazione Basta dimostrare che τ_F è un tempo d'arresto in senso debole, e difatti

$$\{\tau_F \ge t\} = \bigcap_{0 \le s < t} \{X_s \in F\},$$

ed essendo il processo X_t a traiettorie *cadlag* ed F chiuso si ha

$$\bigcap_{0 \le s < t} \{X_s \in F\} = \bigcap_{\substack{0 \le s < t \\ s \in \mathbb{Q}}} \{X_s \in F\} \in \mathcal{F}_t.$$

(infatti se $X_s \in F$ per ogni $s \in \mathbb{Q} \cap [0, t)$, allora $\forall r < t$ $X_r = \lim_{\substack{s \to r \\ s \in \mathbb{Q} \cap (r,t)}} X_s \in F$)

7.1 ALCUNE PROPRIETÀ DEI TEMPI D'ARRESTO

1) Se τ e σ sono tempi d'arresto allora $\tau \wedge \sigma$ e $\tau \vee \sigma$ sono tempi d'arresto.

Infatti, per ogni t,

$$\{\tau \vee \sigma \le t\} = \{\tau \le t\} \cap \{\sigma \le t\} \in \mathcal{F}_t e \{\tau \wedge \sigma \le t\} = \{\tau \le t\} \cup \{\sigma \le t\} \in \mathcal{F}_t$$

[23] Alternativamente: per ogni $s > t$ si ha $\{\tau \le t\} = \bigcap_{n \ge 1} \{\tau < t + \frac{s-t}{n}\} \in \mathcal{F}_{t+(s-t)} = \mathcal{F}_s$

2) Se τ è un tempo d'arresto allora la successione $\tau_n = \frac{\lceil \tau n \rceil}{n}$ è una successione di tempi d'arresto per cui $\tau_n \to \tau$, con $\tau_n \geq \tau$ per ogni n. (qui $\lceil x \rceil$ denota la parte intera superiore)

Infatti, $\forall t$, posto $t_n = \frac{\lfloor nt \rfloor}{n}$, risulta

$$\{\tau_n \leq t\} = \{\lceil \tau n \rceil \leq nt\} = \{\tau \leq \frac{\lfloor nt \rfloor}{n}\} \in \mathcal{F}_{t_n} \subseteq \mathcal{F}_t,$$

in quanto, posto $k = \lceil \tau n \rceil$, cioè $k - 1 < \tau n \leq k$, e $k \leq nt$, allora $\lfloor nt \rfloor \geq k \geq \tau n$, ovvero $\tau \leq \frac{\lfloor nt \rfloor}{n}$, ed ovviamente risulta $t_n \leq t$.
Per la convergenza di τ_n a τ, basta osservare che qualunque sia $x \in \mathbb{R}$ si ha

$$\frac{\lceil nx \rceil}{n} - \frac{1}{n} = \frac{\lceil nx \rceil - 1}{n} < x \leq \frac{\lceil nx \rceil}{n}$$

Si osservi che, se di prende $\widetilde{\tau}_n = \tau_{2^n}$, allora $\widetilde{\tau}_n \searrow \tau$, cioè $\widetilde{\tau}_n$ è anche una successione monotona non crescente, e che inoltre $\{\widetilde{\tau}_n\} \subseteq D$, dove D è l'insieme dei diadici. Infine va notato che tale risultato è interessante solo nel caso di tempi d'arresto che assumono valori in un insieme non discreto.

3) La v.a. τ è $\mathcal{F}_\tau$-misurabile.

Infatti, per ogni s, $\{\tau \leq s\} \in \mathcal{F}_\tau$ in quanto, qualunque sia t,

$$\{\tau \leq s\} \cap \{\tau \leq t\} = \{\tau \leq s \wedge t\} \in \mathcal{F}_{s \wedge t} \subseteq \mathcal{F}_t$$

4) Se τ e σ sono tempi d'arresto e $\mathbb{P}\{\sigma \leq \tau\} = 1$, ed $\mathcal{F}_0$ contiene tutti gli eventi trascurabili (cioè gli insiemi contenuti in insiemi di probabilità nulla), allora $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$.

Infatti se $A \in \mathcal{F}_\infty$ e $A \cap \{\sigma \leq t\} \in \mathcal{F}_t$ per ogni t, allora

$$A \cap \{\tau \leq t\} = (A \cap \{\tau \leq t\} \cap \{\sigma \leq t\}) \cup (A \cap \{\tau \leq t\} \cap \{\sigma > t\}) =$$

$$= ((A \cap \{\sigma \leq t\}) \cap \{\tau \leq t\}) \cup C \in \mathcal{F}_t$$

in quanto $C := (A \cap \{\tau \leq t\} \cap \{\sigma > t\}) \in \mathcal{F}_0 \subseteq \mathcal{F}_t$, essendo un evento trascurabile, e $\{\tau \leq t\}$ e $A \cap \{\sigma \leq t\}$ sono in $\mathcal{F}_t$ per ipotesi.

Si osservi che se invece $\sigma \leq \tau$ certamente, allora la condizione che $\mathcal{F}_0$ contenga tutti gli eventi trascurabili non è necessaria.

Conviene qui osservare che se una filtrazione soddisfa le **condizioni "ABITUALI"**, cioè

i) $\mathcal{F}_0$ contiene tutti gli eventi trascurabili non è necessaria.

ii) la filtrazione è continua a destra (cioè $\mathcal{F}_t = \mathcal{F}_{t^+} := \bigcap_{s>t} \mathcal{F}_s$)

si possono applicare sia il Lemma 2 (e ottenere dalla **ii)** che i tempi di uscita da un chiuso sono tempi d'arresto) sia la proprietà 4 (e ottenere dalla **i)** che $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$).

7.2 CASO A TEMPO DISCRETO

PROPOSIZIONE(Martingale arrestate) Data una $\mathcal{F}_n$-martingala (o submartingala) X_n ed un tempo d'arresto τ, il processo

$$Y_n := X_{n\wedge\tau}$$

è una $\mathcal{F}_n$-martingala (o submartingala).

Dimostrazione. Cominciamo con il dimostrare che Y_n è $\mathcal{F}_n$-misurabile e che è integrabile. Si noti che

$$Y_n = X_\tau I_{\{\tau\leq n\}} + X_n I_{\{\tau>n\}},$$

e quindi

$$Y_n = X_0 I_{\{\tau=0\}} + X_1 I_{\{\tau=1\}} + X_2 I_{\{\tau=2\}} + \cdots + X_n I_{\{\tau=n\}} + X_n I_{\{\tau>n\}}, \tag{23}$$
$$Y_n = X_0 I_{\{\tau=0\}} + X_1 I_{\{\tau=1\}} + X_2 I_{\{\tau=2\}} + \cdots + X_n I_{\{\tau\geq n\}} \tag{24}$$

e che $\{\tau = k\} \in \mathcal{F}_k \subseteq \mathcal{F}_n, \{\tau > n\} \in \mathcal{F}_n$ e che X_k sono $\mathcal{F}_n$-misurabili per $k \leq n$, da cui la $\mathcal{F}_n$-misurabilità di Y_n. Per ottenere l'integrabilità basta notare che dalla (24)

$$| Y_n |\leq| X_0 | + | X_1 | + | X_2 | + \cdots + | X_n |$$

e sfruttare l'integrabilità delle X_k.
Per ottenere il resto dobbiamo notare che dalla (24), con $n + 1$ al posto di n, si ha

$$Y_{n+1} = X_0 I_{\{\tau=0\}} + X_1 I_{\{\tau=1\}} + X_2 I_{\{\tau=2\}} + \cdots + X_n I_{\{\tau=n\}} + X_{n+1} I_{\{\tau>n\}}, \tag{25}$$

e quindi, confrontando (23) e (25)

$$Y_{n+1} - Y_n = (X_{n+1} - X_n) I_{\{\tau>n\}}$$

Di conseguenza, essendo $I_{\{\tau>n\}}$ una v.a. $\mathcal{F}_n$-misurabile, ed X_n una martingala,

$$\mathbb{E}[Y_{n+1} - Y_n \mid \mathcal{F}_n] = \mathbb{E}[X_{n+1} - X_n \mid \mathcal{F}_n] I_{\{\tau>n\}} = 0.$$

Nel caso in cui X_n sia una submartingala risulta ovviamente $\mathbb{E}[Y_{n+1} - Y_n \mid \mathcal{F}_n] \geq 0$.

PROPOSIZIONE Data una $\mathcal{F}_n$-submartingala (o martingala) X_n ed un tempo d'arresto τ, limitato quasi certamente, ovvero per cui

$$\mathbb{P}(1 \leq \tau \leq n) = 1$$

allora

$$\mathbb{E}[X_1] \leq \mathbb{E}[X_\tau] \leq \mathbb{E}[X_n].$$

(Nel caso di martingale $\mathbb{E}[X_1] = \mathbb{E}[X_\tau] = \mathbb{E}[X_n]$.)

Dimostrazione: Per la proposizione precedente si ha che $\{X_{\tau\wedge k}\}_k$ è una submartingala, e quindi in particolare il valore medio è crescente (in senso lato). Poiché $X_{\tau\wedge 1} = X_1$, la prima disuguaglianza segue immediatamente.

Per la seconda disuguaglianza basta mostrare che $\mathbb{E}[X_n - X_\tau] \geq 0$ e infatti

$$X_n - X_\tau = \sum_{i=1}^{n} (X_n - X_\tau) I_{\{\tau=i\}} = \sum_{i=1}^{n} (X_n - X_i) I_{\{\tau=i\}}$$

e quindi

$$\mathbb{E}[X_n - X_\tau] = \sum_{i=1}^{n} \mathbb{E}[(X_n - X_i) I_{\{\tau=i\}}] = \sum_{i=1}^{n} \mathbb{E}[\mathbb{E}[(X_n - X_i) I_{\{\tau=i\}} \mid \mathcal{F}_i]]$$

La tesi segue in quanto

$$\mathbb{E}[(X_n - X_i) I_{\{\tau=i\}} \mid \mathcal{F}_i] = I_{\{\tau=i\}} \mathbb{E}[(X_n - X_i) \mid \mathcal{F}_i] \geq 0,$$

nel caso delle submartingale ($= 0$ nel caso delle martingale).

7.3 Applicazioni: il problema della rovina del giocatore con le martingale

1) caso simmetrico

Sia Y_k una successione di v.a. indipendenti, con

$$\mathbb{P}(Y_k = 1) = \mathbb{P}(Y_k = -1) = \frac{1}{2},$$

e sia

$$S_n = \sum_{k=1}^{n} Y_k.$$

Sappiamo (vedi Esempio 5.2 che S_n è una martingala. Siano a e b numeri naturali non nulli e sia

$$\tau = \tau(a,b) := \inf\{n \text{ t.c. } S_n \notin (-a,b)\} = \inf\{n \text{ t.c. } S_n = -a \text{ o } S_n = b\}.$$

La variabile aleatoria τ è finita [24] con probabilità 1, cioè $\mathbb{P}(\tau < +\infty) = 1$.

Per uno dei risultati precedenti sappiamo che $S_{n\wedge\tau}$ è una martingala e che quindi

$$\mathbb{E}[S_{n\wedge\tau}] = \mathbb{E}[S_{1\wedge\tau}] = \mathbb{E}[S_1] = 0$$

[24] Si può dimostrare direttamente, anche nel caso generale, con $\mathbb{P}(Y_h = 1) = p$, che

$$\mathbb{P}(\tau = +\infty) = \lim_{n\to\infty} \mathbb{P}(\tau > n(a+b)) = 0.$$

Infatti, si ha $\{\tau = +\infty\} = \bigcap_{n\geq 1} \{\tau > n(a+b)\}$ e $\mathbb{P}(\tau > n(a+b)) \leq \alpha^n$ per un $\alpha < 1$. La prima uguaglianza è ovvia, mentre la seconda si può vedere facilmente osservando che

$$\{\omega \text{ tali che esiste } k < n \text{ per cui } Y_{k(a+b)+1} = Y_{k(a+b)+2} = \cdots = Y_{k(a+b)+a+b} = 1\} \subset \{\tau \leq n(a+b)\},$$

e che quindi

$$\mathbb{P}\Big(\bigcup_{k<n} \{Y_{k(a+b)+1} = Y_{k(a+b)+2} = \cdots = Y_{k(a+b)+a+b} = 1\}\Big) \leq \mathbb{P}(\tau \leq n(a+b))$$

Inoltre $S_{n\wedge\tau} \to S_\tau$ per $n \to +\infty$, e $|S_{n\wedge\tau}| \leq \max(a,b)$ e quindi per il teorema della convergenza dominata

$$\mathbb{E}[S_{n\wedge\tau}] \to \mathbb{E}[S_\tau].$$

Di conseguenza

$$\mathbb{E}[S_\tau] = -a\mathbb{P}(S_\tau = -a) + b\,(1 - \mathbb{P}(S_\tau = -a)) = 0,$$

da cui immediatamente

$$\mathbb{P}(S_\tau = -a) = \frac{b}{a+b}.$$

2) caso generale
Come nell'applicazione **1)** ma con $\mathbb{P}(Y_k = 1) = p$ e $\mathbb{P}(Y_k = -1) = q = 1-p$. Si procede in modo analogo al caso precedente, ma questa volta si prende come martingala

$$Z_n = \exp\{\theta S_n - n\psi(\theta)\}$$

dove

$$\exp\{\psi(\theta)\} = \mathbb{E}[\exp\{\theta Y_1\}] = \exp\{\theta\}p + \exp\{-\theta\}q.$$

Si cerca, se esiste θ in modo che $\psi(\theta) = 0$ ovvero, posto $\exp\{\theta\} = \alpha$, si cerca $\alpha p + \alpha^{-1}q = 1$, ovvero $\alpha^2 p - \alpha + q = 0$. Ciò è possibile solo per

$$\alpha = \frac{1 \pm \sqrt{1-4pq}}{2p} = \frac{1 \pm \sqrt{1-4p+4p^2}}{2p} = \frac{1 \pm \sqrt{1-4p+4p^2}}{2p} = \frac{1 \pm |1-2p|}{2p} = \frac{1 \pm (1-2p)}{2p}$$

ovvero per $\alpha = 1$ o $\alpha = \frac{q}{p}$ (come del resto si può vedere anche subito). Il caso $\alpha = 1$ corrisponderebbe a $Z_n \equiv 1$ e non porterebbe ad alcun risultato, mentre $\exp\{\theta\} = \alpha = \frac{q}{p}$ corrisponde a $Z_n = \left(\frac{q}{p}\right)^{S_n}$. Di nuovo, sempre per convergenza dominata,

$$1 \equiv \mathbb{E}[Z_{n\wedge\tau}] \to \mathbb{E}[Z_\tau] = \left(\frac{q}{p}\right)^{-a}\mathbb{P}(S_\tau = -a) + \left(\frac{q}{p}\right)^{b}[1 - \mathbb{P}(S_\tau = -a)] = 1,$$

Da cui di nuovo si può ricavare, posto $\rho = \frac{q}{p}$

$$\mathbb{P}(S_\tau = -a) = \frac{1-\rho^b}{\rho^{-a}-\rho^b} = \frac{\rho^a - \rho^{b+a}}{1-\rho^{b+a}} = \frac{\rho^{b+a}-\rho^a}{\rho^{b+a}-1}.$$

---

ovvero, passando ai complementari, e utilizzando l'indipendenza delle v.a. Y_h,

$$\begin{aligned}
&\mathbb{P}\Big(\bigcap_{k<n}\{Y_{k(a+b)+1} = Y_{k(a+b)+2} = \cdots = Y_{k(a+b)+a+b} = 1\}^c\Big)\\
&= \prod_{k<n}\mathbb{P}(\{Y_{k(a+b)+1} = Y_{k(a+b)+2} = \cdots = Y_{k(a+b)+a+b} = 1\}^c)\\
&= \prod_{k<n}(1-p^{a+b}) = (1-p^{a+b})^n = \alpha^n \geq \mathbb{P}(\tau > n(a+b)).
\end{aligned}$$

Si noti che in fondo si è dimostrato che $\tau \leq T(a+b)$, dove T è una variabile aleatoria geometrica di parametro $\beta = p^{a+b} = 1-\alpha$.

Si noti che

$$\mathbb{P}(S_\tau = -a) \to \frac{b}{a+b}$$

per $\rho \to 1$, cioè per $p \to \frac{1}{2}$.

7.4 DISUGUAGLIANZA DI KOLMOGOROV PER SUBMARTINGALE NON NEGATIVE

Sia X_n una submartingala non negativa allora

$$\mathbb{P}(\max(X_1, X_2, \cdots, X_n) > \gamma) \leq \frac{\mathbb{E}[X_n]}{\gamma}(i)$$

Sia X_n una martingala con $\mathbb{E}[|X_n|^\alpha] < +\infty, \alpha \geq 1$, allora

$$\mathbb{P}(\max(|X_1|, |X_2|, \cdots, |X_n|) > \gamma) \leq \frac{\mathbb{E}[|X_n|^\alpha]}{\gamma^\alpha}(ii)$$

Dimostrazione
Cominciamo con il primo caso. Si definisca

$$\begin{cases} \tau := \inf\{k \text{ tali che } 1 \leq k \leq n, X_k > \gamma\}, & \text{se un tale } k \text{ esiste} \\ \tau := n, & \text{altrimenti.} \end{cases}$$

Ovviamente τ è un tempo d'arresto e $\{X_\tau > \gamma\} = \{\max(X_1, \cdots, X_n) > \gamma\}$ e quindi per la disuguaglianza di Markov

$$\mathbb{P}(\max(X_1, X_2, \cdots, X_n) > \gamma) = \mathbb{P}(X_\tau > \gamma) \leq \frac{\mathbb{E}[|X_\tau|]}{\gamma} = \frac{\mathbb{E}[X_\tau]}{\gamma}$$

Basta quindi mostrare che $\mathbb{E}[X_\tau] \leq \mathbb{E}[X_n]$, ma ciò discende immediatamente dalla proposizione precedente in quanto τ è a valori in $1, 2, \cdots, n$.

Per il caso di martingale, la tesi segue osservando che

$$\mathbb{P}(\max(|X_1|, |X_2|, \cdots, |X_n|) > \gamma) = \mathbb{P}(\max(|X_1|^\alpha, |X_2|^\alpha, \cdots, |X_n|^\alpha) > \gamma^\alpha),$$

la funzione $|x|^\alpha$, per $\alpha \geq 1$ è convessa e quindi $|\mathrm{X}_n|^\alpha$ è una submartingala non negativa (confrontare proprietà 4) e infine applicando la disuguaglianza precedente.

8 CONVERGENZA DI MARTINGALE

Sia X_n una martingala uniformemente limitata in L^1, cioè tale che

$$\sup_n \mathbb{E}[|X_n|] \leq M < +\infty.$$

Allora

$$\mathbb{P}(\{\omega \text{ t.c. } \exists \text{ finito } \lim_{n\to\infty} X_n(\omega)\}) = 1$$

Dimostrazione
Daremo la dimostrazione solo nel caso in cui valga una ipotesi più forte:

$$\sup_n \mathbb{E}[|X_n|^2] \leq M < +\infty.$$

(si noti infatti che in tale caso $\mathbb{E}[|X_n|] \leq \Big(\mathbb{E}[|X_n|^2]\Big)^{1/2} \leq M$)

Basta mostrare che la successione X_n è una successione di Cauchy con probabilità 1. Ciò significa che

$$\{\omega \text{ t.c. } \forall \epsilon > 0 \exists m \geq 1 \text{ t.c.} \forall k \geq 1 |X_{m+k}(\omega) - X_m(\omega)| \leq \epsilon\}$$

ovvero

$$\bigcap_{\epsilon>0} \bigcup_{m\geq 1} \bigcap_{k\geq 1} \{|X_{m+k} - X_m| \leq \epsilon\}$$

è un insieme misurabile di probabilità 1. La misurabilità segue osservando che è equivalente prendere ϵ razionale positivo. Per calcolare la probabilità, ricordiamo che in generale se B_h sono eventi $\mathbb{P}(\bigcap_{h\geq 1} B_h) = 1$ se e solo se $\mathbb{P}(B_h) = 1 \forall h \geq 1$.

Infatti se $\mathbb{P}(\bigcap_{h\geq 1} B_h) = 1$ allora, poiché $\bigcap_{h\geq 1} B_h \subseteq B_{\overline{h}}$, ne segue che $\mathbb{P}(B_{\overline{h}}) = 1$, comunque fissato $\overline{h} \geq 1$. Se viceversa $\mathbb{P}(B_h) = 1 \forall h \geq 1$, allora

$$\mathbb{P}(\bigcap_{h\geq 1} B_h) = 1 - \mathbb{P}(\bigcup_{h\geq 1} B_h^c), e \mathbb{P}(\bigcup_{h\geq 1} B_h^c) \leq \sum_{h\geq 1} \mathbb{P}(B_h^c) = 0.$$

Di conseguenza la tesi equivale a mostrare che, per ogni $\epsilon > 0$,

$$\mathbb{P}(\bigcup_{m\geq 1} \bigcap_{k\geq 1} \{|X_{m+k} - X_m| \leq \epsilon\}) = 1,$$

ovvero che, per ogni $\epsilon > 0$,

$$\mathbb{P}(\bigcap_{m\geq 1} \bigcup_{k\geq 1} \{|X_{m+k} - X_m| > \epsilon\}) = 0.$$

Si osservi che,$\forall \overline{m} \geq 1$,

$$\bigcap_{m\geq 1} \bigcup_{k\geq 1} \{|X_{m+k} - X_m| > \epsilon\} \subseteq \bigcup_{k\geq 1} \{|X_{\overline{m}+k} - X_{\overline{m}}| > \epsilon\}$$

e che

$$\mathbb{P}(\bigcup_{k\geq 1} \{|X_{\overline{m}+k} - X_{\overline{m}}| > \epsilon\}) = \lim_{n\to\infty} \mathbb{P}(\bigcup_{k=1}^{n} \{|X_{\overline{m}+k} - X_{\overline{m}}| > \epsilon\}).$$

Di conseguenza

$$\mathbb{P}(\bigcap_{m\geq 1} \bigcup_{k\geq 1} \{|X_{m+k} - X_m| > \epsilon\}) \leq \lim_{n\to\infty} \mathbb{P}(\bigcup_{k=1}^{n} \{|X_{\overline{m}+k} - X_{\overline{m}}| > \epsilon\}),$$

e quindi

$$\mathbb{P}(\bigcap_{m\geq 1} \bigcup_{k\geq 1} \{|X_{m+k} - X_m| > \epsilon\}) \leq \lim_{\overline{m}\to\infty} \lim_{n\to\infty} \mathbb{P}(\bigcup_{k=1}^{n} \{|X_{\overline{m}+k} - X_{\overline{m}}| > \epsilon\}).$$

Non rimane che dimostrare che

$$\lim_{\overline{m}\to\infty}\lim_{n\to\infty}\mathbb{P}(\bigcup_{k=1}^{n}\{|X_{\overline{m}+k}-X_{\overline{m}}|>\epsilon\})=0.$$

Si osservi ora che $\overline{X}_k := X_{\overline{m}+k}-X_{\overline{m}}$, è una martingala rispetto alla filtrazione $\mathcal{G}_k := \mathcal{F}_{\overline{m}+k}$ e che

$$\bigcup_{k=1}^{n}\{|X_{\overline{m}+k}-X_{\overline{m}}|>\epsilon\}=\{\max(|\bar{X}_1|,|\bar{X}_2|,\cdots,|\bar{X}_n|)>\epsilon\}.$$

Basta quindi applicare la disuguaglianza di Kolmogorov per $\alpha=2$ per ottenere che

$$\mathbb{P}(\bigcup_{k=1}^{n}\{|X_{\overline{m}+k}-X_{\overline{m}}|>\epsilon\})\leq\frac{\mathbb{E}[|X_{\overline{m}+n}-X_{\overline{m}}|^2]}{\epsilon^2}=\frac{\mathbb{E}[X^2_{\overline{m}+n}]-\mathbb{E}[X^2_{\overline{m}}]}{\epsilon^2}.$$

Infatti

$$\mathbb{E}[|X_{\overline{m}+n}-X_{\overline{m}}|^2]=\mathbb{E}[X^2_{\overline{m}+n}]+\mathbb{E}[X^2_{\overline{m}}]-2\mathbb{E}[X_{\overline{m}+n}X_{\overline{m}}]=$$
$$=\mathbb{E}[X^2_{\overline{m}+n}]+\mathbb{E}[X^2_{\overline{m}}]-2\mathbb{E}[\mathbb{E}[X_{\overline{m}+n}X_{\overline{m}}\mid\mathcal{F}_{\overline{m}}]=$$
$$=\mathbb{E}[X^2_{\overline{m}+n}]+\mathbb{E}[X^2_{\overline{m}}]-2\mathbb{E}[X_{\overline{m}}\mathbb{E}[X_{\overline{m}+n}\mid\mathcal{F}_{\overline{m}}]=\mathbb{E}[X^2_{\overline{m}+n}]+\mathbb{E}[X^2_{\overline{m}}]-2\mathbb{E}[X^2_{\overline{m}}]$$

A questo punto si noti che $\mathbb{E}[X^2_{\overline{m}}]$ è una successione convergente ad un numero $\mu\leq M$, in quanto è una successione limitata per ipotesi, e monotona non decrescente ($X^2_{\overline{m}}$ è una submartingala). Quindi

$$\lim_{\overline{m}\to\infty}\lim_{n\to\infty}\mathbb{P}(\bigcup_{k=1}^{n}\{|X_{\overline{m}+k}-X_{\overline{m}}|>\epsilon\})\leq\lim_{\overline{m}\to\infty}\lim_{n\to\infty}\frac{\mathbb{E}[X^2_{\overline{m}+n}]-\mathbb{E}[X^2_{\overline{m}}]}{\epsilon^2}=\lim_{\overline{m}\to\infty}\frac{\mu-\mathbb{E}[X^2_{\overline{m}}]}{\epsilon^2}=0.$$

9 DISUGUAGLIANZA DI DOOB

Sia data una submartingala non negativa X_n, con $\mathbb{E}[(X_n)^p]<+\infty$, per un $p>1$. Posto $X^*_n=\max_{k\leq n}(X_k)$ vale la seguente disuguaglianza:

$$\mathbb{E}[(X^*_n)^p]\leq\left(\frac{p}{p-1}\right)^p\mathbb{E}[(X_n)^p].$$

Nel caso di una martingala X_n, con $\mathbb{E}[|X_n|^p]<+\infty$, e posto $X^*_n=\max_{k\leq n}(|X_k|)$ vale la seguente disuguaglianza:

$$\mathbb{E}[|X^*_n|^p]\leq\left(\frac{p}{p-1}\right)^p\mathbb{E}[|X_n|^p])$$

DIMOSTRAZIONE: Si definisca

$$\tau_\gamma := \inf\{k \text{ tali che } 1\leq k\leq n, X_k>\gamma\},\qquad \text{se un tale } k \text{ esiste}$$

$$\tau_\gamma := n+1 \qquad \text{altrimenti.}$$

Ovviamente τ_γ è un tempo d'arresto e $\{X^*_n>\gamma\}$ coincide con $\bigcup_{k=1}^{n}\{\tau_\gamma=k\}$ e quindi per ogni

$$(X_n^*)^\beta = \int_0^{X_n^*} \beta\gamma^{\beta-1} d\gamma = \int_0^\infty \beta\gamma^{\beta-1} I_{\{X_n^*>\gamma\}} d\gamma = \int_0^\infty \beta\gamma^{\beta-1} \sum_{k=1}^n I_{\{\tau_\gamma=k\}} d\gamma$$

Passando al valore medio, per $\beta > 0$,

$$\mathbb{E}[(X_n^*)^\beta] = \sum_{k=1}^n \int_0^\infty \beta\gamma^{\beta-2}\mathbb{E}[\gamma I_{\{\tau_\gamma=k\}}]d\gamma \leq \sum_{k=1}^n \int_0^\infty \beta\gamma^{\beta-2}\mathbb{E}[X_k I_{\{\tau_\gamma=k\}}]d\gamma$$

in quanto $\gamma I_{\{\tau_\gamma=k\}} \leq X_k I_{\{\tau_\gamma=k\}}$. Inoltre, essendo X_n una submartingala $X_k \leq \mathbb{E}[X_n \mid \mathcal{F}_k]$, non negativa, ed essendo $\{\tau_\gamma = k\} \in \mathcal{F}_k$ si ha

$$\mathbb{E}[X_k I_{\{\tau_\gamma=k\}}] \leq \mathbb{E}[\mathbb{E}[X_n \mid \mathcal{F}_k] I_{\{\tau_\gamma=k\}}] = \mathbb{E}[\mathbb{E}[X_n I_{\{\tau_\gamma=k\}} \mid \mathcal{F}_k]] = \mathbb{E}[X_n I_{\{\tau_\gamma=k\}}]$$

e quindi

$$\mathbb{E}[(X_n^*)^\beta] \leq \sum_{k=1}^n \int_0^\infty \beta\gamma^{\beta-2}\mathbb{E}[X_n I_{\{\tau_\gamma=k\}}]d\gamma = \int_0^\infty \beta\gamma^{\beta-2}\mathbb{E}[X_n \sum_{k=1}^n I_{\{\tau_\gamma=k\}}]d\gamma = \mathbb{E}[\int_0^\infty \beta\gamma^{\beta-2} X_n I_{\{X_n^*>\gamma\}} d\gamma]$$

ovvero

$$\mathbb{E}[(X_n^*)^\beta] \leq \beta\mathbb{E}[X_n \int_0^\infty \gamma^{\beta-2} I_{\{X_n^*>\gamma\}} d\gamma] = \frac{\beta}{\beta-1}\mathbb{E}[X_n (X_n^*)^{\beta-1}].$$

Questa disuguaglianza non è banale (ovvero no è del tipo $+\infty \leq +\infty$) prendendo β=p, in quanto si ha che $\mathbb{E}[(X_n^*)^p] \leq \mathbb{E}[\sum_{k=1}^n (X_k)^p] < +\infty$. Inoltre, usando la disuguaglianza di Hölder si ottiene la tesi osservando che

$$\mathbb{E}[(X_n^*)^p] \leq \frac{p}{p-1}\mathbb{E}[(X_n)^p]^{1/p}\mathbb{E}[\Big((X_n^*)^{p-1}\Big)^{p/(p-1)}]^{(p-1)/p} = \frac{p}{p-1}\mathbb{E}[(X_n)^p]^{1/p}\mathbb{E}[(X_n^*)^p]^{(p-1)/p}$$

ESEMPI DI MARTINGALE A TEMPO CONTINUO

In modo molto simile a quanto fatto a tempo discreto per le somme di v.a. indipendenti, si può mostrare che se X_t è un processo ad incrementi indipendenti e omogenei, con $X_0 = 0$, e con media nulla allora X_t è una martingala. Inoltre è facile mostrare che se X_t è un processo ad incrementi indipendenti e omogenei, integrabile e con $X_0 = 0$, allora $\mathbb{E}[X_t] = mt$ per t nei razionali:

$$\mathbb{E}[X_1] = \sum_{k=1}^n \mathbb{E}[X_{k/n} - X_{(k-1)/n}] = n\mathbb{E}[X_{1/n}]$$

da cui $\mathbb{E}[X_{1/n}] = \frac{1}{n}\mathbb{E}[X_1]$ e analogamente $\mathbb{E}[X_{m/n}] = \sum_{k=1}^m \mathbb{E}[X_{k/n} - X_{(k-1)/n}] = \frac{m}{n}\mathbb{E}[X_1]$. Per ottenere che ciò valga anche per ogni t reale, si deve notare che comunque $\mathbb{E}[X_{t+s}] = \mathbb{E}[X_t] + \mathbb{E}[X_s]$ e aggiungere una piccola ulteriore ipotesi di regolarità: se $\mathbb{E}[X_t]$ è una funzione continua in t (o continua a destra) si otterrebbe immediatamente la tesi per continuità.

Il processo $X_t - \mathbb{E}[X_t] = X_t - mt$ è allora un processo ad incrementi indipendenti ed omogenei, a media nulla e quindi è una martingala.

Se ancora X_t ammette momento secondo finito, allora di nuovo (con una dimostrazione simile) si ha che $Var(X_t) = \sigma^2 t$, purché si possa affermare a priori che $Var(X_t)$ è una funzione continua. Di nuovo similmente al caso a tempo discreto, accade che $(X_t - \mathbb{E}[X_t])^2 - Var(X_t) = (X_t - mt)^2 - \sigma^2 t$ è una martingala.

Infine è possibile mostrare che, sotto opportune ipotesi di regolarità (continuità in probabilità), se $\mathbb{E}[\exp\{\theta(X_t - mt)\}] < +\infty$, allora

$$\mathbb{E}[\exp\{\theta(X_t - mt)\}] = \exp\{K(\theta)t\}$$

e che quindi

$$Z_t := \mathbb{E}[\exp\{\theta(X_t - mt) - K(\theta)t\}$$

è una martingala a media 1.
Il tutto vale anche per i processi $Y_t = Y_0 + X_t$, con dato iniziale Y_0 indipendente da $\{X_t\}$, tranne per i valori medi. Ad esempio

$$\widetilde{Z}_t := \mathbb{E}[\exp\{\theta(Y_t - mt) - K(\theta)t\}]$$

è ancora una martingala a media costante uguale a $\mathbb{E}[\exp\{\theta Y_0\}]$, purché ovviamente tale valore medio sia finito.

APPLICAZIONE

Il processo di Wiener standard o moto Browniano W_t è una martingala, anche $M_t = W_t^2 - t$ e infine, per ogni θ reale

$$\widetilde{Z}_t := \mathbb{E}[\exp\{\theta W_t - \frac{1}{2}\theta^2 t\}]$$

è una martingala a media 1.

Si noti che W_t^2 è una submartingala e che quindi $W_t^2 = M_t + t$. Anche il processo di Poisson N_t di parametro λ, essendo un processo crescente è una submartingala, e si può decomporre nella somma di una martingala $M_t := N_t - \lambda t$ e di un processo crescente $A_t = \lambda t$. Si tratta di casi particolari della decomposizione di Doob a tempo continuo. Altre applicazioni si potrebbero ottenere per i processi di Poisson composti.

In generale data una v.a. Z non negativa e a media 1 in uno spazio $(\Omega, \mathcal{F}, \mathbb{P})$, si può definire una nuova misura di probabilità $\mathbb{Q}$ definita da

$$\mathbb{Q}(C) := \mathbb{E}^{\mathbb{P}}[I_C Z]$$

ESERCIZI
Presa $Z = \widetilde{Z}_T = \mathbb{E}^{\mathbb{P}}[\exp\{\theta W_T - \frac{1}{2}\theta^2 T\}]$ nell'espressione precedente si trovi

1) la derivata di Radon-Nikodym

$$\left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_{\mathcal{F}_t},$$

2) la legge di W_t rispetto a $\mathbb{Q}$,

3) le distribuzioni finito dimensionali di $(W_t, t \geq 0)$ rispetto a $\mathbb{Q}$.

10 PROPRIETÀ DEL MOTO BROWNIANO

Abbiamo visto che il moto browniano è un particolare processo gaussiano a incrementi indipendenti con media nulla e varianza t. Abbiamo anche visto che per la proprietà delle v.a. gaussiane per cui non correlazione ed indipendenza sono equivalenti, un modo alternativo di definirlo è attraverso la funzione di correlazione $Cov(W_t, W_s) = t \wedge s$. Inoltre sappiamo che W_t è una martingala rispetto alla filtrazione naturale $\mathcal{F}_t^W$.

10.1 TRASFORMAZIONI DEL MOTO BROWNIANO

1) Per ogni $s \geq 0, \widetilde{W}_t := W_{s+t} - W_s$ è un moto Browniano ed è una martingala rispetto alla filtrazione $\mathcal{G}_t := \mathcal{F}_{s+t}^W$.

2) Per ogni $c \in \mathbb{R}$, $W_t^{(c)} := cW_{c^2t}$ è un moto Browniano ed è una martingala rispetto alla filtrazione $\mathcal{G}_t^{(c)} := \mathcal{F}_{c^2t}^W$ (in particolare per $c = -1$ si ha che $-W_t$ è ancora un moto Browniano rispetto a $\mathcal{F}_t^W$).

3) Il processo definito da $\widehat{W}_t := tW_{1/t}$, per $t > 0$, e $Z_0 := 0$, è ancora un moto Browniano, rispetto alla sua filtrazione naturale.

4) Per ogni tempo d'arresto limitato τ il processo $\widetilde{W}_t^{(\tau)} := W_{\tau+t} - W_\tau$ è ancora un moto Browniano rispetto alla filtrazione $\mathcal{F}_{\tau+t}^W$.

Questa proprietà non è immediata, ed è una proprietà che, nei processi di Markov viene detta Proprietà di Markov Forte.

10.2 PROPRIETÀ DI MARKOV FORTE PER IL PROCESSO DI WIENER

Sia τ un tempo d'arresto finito con probabilità 1. Allora il processo $Y_t := W_{t+\tau} - W_\tau$ è un processo di Wiener standard ed è indipendente da $\mathcal{F}_\tau$. (qui si suppone di prendere una versione continua dei W_t)

Quindi per ogni funzionale Φ, che sia $\mathcal{R}^{[0,\infty)}$-misurabile delle traiettorie

$$\mathbb{E}[\Phi(W_s, s \geq \tau) \mid \mathcal{F}_\tau] = \mathbb{E}[\Phi(W_s, s \geq \tau) \mid W_\tau] = \mathbb{E}[\Phi(w + Y_u, u \geq 0)] \mid_{w=W_\tau} .$$

In questo senso è una generalizzazione della proprietà di Markov, che invece di valere solo per tempi deterministici t vale anche per tempi d'arresto τ.

Più precisamente la proprietà di Markov (rispetto ad una filtrazione $\mathcal{F}_t \supseteq \mathcal{F}_t^X$) per un generico processo X_t è la seguente:

$$\mathbb{P}(X_{t+s} \in A \mid \mathcal{F}_t) = \mathbb{P}(X_{t+s} \in A \mid X_t) \qquad \text{per ogni } s, t \geq 0.$$

(nel caso dei processi di Markov regolari

$$\mathbb{P}(X_{t+s} \in A \mid X_t) = p(t, t+s, x, A) \mid_{x=X_t},$$

dove $p(u, v, x, \cdot)$ sono le probabilità di transizione)

La proprietà di Markov forte (che non è sempre verificata per un generico processo di Markov) è invece

$$\mathbb{P}(X_{\tau+s} \in A \mid \mathcal{F}_\tau) = \mathbb{P}(X_{\tau+s} \in A \mid X_\tau) \qquad \text{per ogni tempo d'arresto finito } \tau \text{ ed } s \geq 0.$$

Si noti che è necessario che τ sia finito affinché abbia senso X_τ ed è necessario che X_τ sia una variabile aleatoria, e che sia $\mathcal{F}_\tau$-misurabile.

Questa proprietà di misurabilità è sempre verificata per processi X_t *cadlag*, in realtà potrebbe bastare un poco meno e precisamente la progressiva misurabilità)

Dimostrazione (della proprietà di Markov forte per il processo di Wiener standard)

La tesi equivale a richiedere che per ogni $Rk \geq 1, 0 \leq t_1 \leq \cdots \leq t_k, A_1, \cdots, A_k$ boreliani di $\mathbb{R}$, e per ogni $C \in \mathcal{F}_\tau$

$$\mathbb{P}(Y_{t_1} \in A_1, \cdots, Y_{t_k} \in A_k, C) = \mathbb{P}(W_{t_1} \in A_1, \cdots, W_{t_k} \in A_k)\mathbb{P}(C)$$

I CASO: τ assume al più un'infinità di valori $\{s_h\}_{h\geq 1}$.

$$\mathbb{P}(Y_{t_1} \in A_1, \cdots, Y_{t_k} \in A_k, C) = \sum_{h=1}^{\infty} \mathbb{P}(Y_{t_1} \in A_1, \cdots, Y_{t_k} \in A_k, C, \tau = s_h) =$$

$$= \sum_{h=1}^{\infty} \mathbb{P}(W_{t_1+s_h} - W_{s_h} \in A_1, \cdots, W_{t_k+s_h} - W_{s_h} \in A_k, C, \tau = s_h) =$$

$$(C \cap \{\tau = s_h\} \in \mathcal{F}_{s_h}, \text{ poiché } C \in \mathcal{F}_\tau \text{ e } W_t \text{ ha incrementi indipendenti})$$

$$= \sum_{h=1}^{\infty} \mathbb{P}(W_{t_1+s_h} - W_{s_h} \in A_1, \cdots, W_{t_k+s_h} - W_{s_h} \in A_k)\mathbb{P}(C, \tau = s_h) =$$

$$(W \text{ ha incrementi omogenei})$$

$$= \sum_{h=1}^{\infty} \mathbb{P}(W_{t_1} \in A_1, \cdots, W_{t_k} \in A_k)\mathbb{P}(C, \tau = s_h) = \mathbb{P}(W_{t_1} \in A_1, \cdots, W_{t_k} \in A_k)\mathbb{P}(C)$$

II CASO: τ generico.

Per individuare la distribuzione congiunta in realtà basta verificare che per ogni $k \geq 1$, per ogni funzione continua e limitata $\Phi(y_1, \cdots, y_k)$ e per ogni $C \in \mathcal{F}_\tau$

$$\mathbb{E}[\Phi(Y_{t_1}, \cdots, Y_{t_k}) I_C] = \mathbb{E}[\Phi(W_{t_1}, \cdots, W_{t_k})]\mathbb{P}(C)$$

in quanto le funzioni continue determinano univocamente la legge di una variabile aleatoria. Per il Caso I, presa $\{\tau_m\}$ una successione di tempi d'arresto, con $\tau_m \geq \tau$ e $\tau_m \to \tau$

$$\mathbb{E}[\Phi(W_{t_1+\tau_m} - W_{\tau_m}, \cdots, W_{t_k+\tau_m} - W_{\tau_m}) I_C] = \mathbb{E}[\Phi(W_{t_1}, \cdots, W_{t_k})]\mathbb{P}(C)$$

Essendo le traiettorie di W_t continue (ma basterebbe che fossero continue a destra), possiamo affermare che $(W_{t_{h+\tau_m}} - W_{\tau_m}) \to \ (W_{t_{h+\tau}} - W_\tau) = Y_{t_h}$ e quindi

$$\Phi(W_{t_1+\tau_m} - W_{\tau_m}, \cdots, W_{t_k+\tau_m} - W_{\tau_m}) \to \Phi(Y_{t_1}, \cdots, Y_{t_k}).$$

Essendo Φ limitata si può passare al limite sotto il segno di media.

OSSERVAZIONE
È evidente che questa dimostrazione (e quindi la proprietà di Markov forte) rimane valida se al posto del moto browniano si mette un qualsiasi processo ad incrementi indipendenti ed omogenei, con traiettorie *cadlag*.

10.3 PRINCIPIO DI RIFLESSIONE

Sotto questo nome vanno diverse proprietà ma la più nota è la seguente, per $a \geq 0$,

$$\mathbb{P}(\tau_a \leq t) = 2\mathbb{P}(W_t \geq a) = \frac{2}{\sqrt{2\pi t}} \int_A^\infty \exp\{-\frac{x^2}{2t}\}dx = \sqrt{\frac{2}{\pi}} \int_{A/\sqrt{t}}^\infty \exp\{-y^2/2\}dy.$$

dove $\tau_a = \inf\{t > 0 \text{ t.c. } W_t \geq a\}$, cioè il primo tempo di uscita da $(-\infty, a)$, che è un tempo d'arresto se come al solito prendiamo la versione continua di W_t.

Per $a < 0$ si definisce $\tau_a = \inf\{t > 0 \text{ t.c. } W_t \leq a\}$.

Notando che $\tau_a = \inf\{t > 0 \text{ t.c. } -W_t \geq -a\}$ e che $-W_t$ è ancora un moto browniano si ottiene che per ogni $a \in \mathbb{R}$

$$\mathbb{P}(\tau_a \leq t) = 2\mathbb{P}(W_t \geq |a|) = \frac{2}{\sqrt{2\pi t}} \int_{|a|}^\infty \exp\{-\frac{x^2}{2t}\}dx = \sqrt{\frac{2}{\pi}} \int_{|a|/\sqrt{t}}^\infty \exp\{-y^2/2\}dy.$$

e quindi la sua densità è

$$g_{\tau_a}(t) = \sqrt{\frac{1}{2\pi}} \frac{|a|}{t^{3/2}} \exp\{-a^2/2t\}.$$

e il suo valore atteso è infinito:

$$\mathbb{E}[\tau_a] = \int_0^\infty t\sqrt{\frac{1}{2\pi}}\frac{|a|}{t^{3/2}}\exp\{-a^2/2t\}\,dt = \int_0^\infty \sqrt{\frac{1}{2\pi}}\frac{|a|}{t^{1/2}}\exp\{-a^2/2t\}\,dt = \infty$$

$$\begin{aligned}\int_0^\infty \sqrt{\frac{1}{2\pi}}\frac{|a|}{t^{1/2}}\exp\{-a^2/2t\}\,dt &\geq \int_1^\infty \sqrt{\frac{1}{2\pi}}\frac{|a|}{t^{1/2}}\exp\{-a^2/2t\}\,dt\\ &\geq \int_1^\infty \sqrt{\frac{1}{2\pi}}\frac{|a|}{t^{1/2}}\exp\{-a^2/2\}\,dt = \infty\end{aligned}$$

Inoltre questa proprietà permette di calcolare la distribuzione del massimo di un moto browniano (per $a \geq 0$):

$$\mathbb{P}(\sup_{s\leq t}(W_s) \geq a) = \mathbb{P}(\tau_a \leq t) = 2\mathbb{P}(W_t \geq a)$$

DIMOSTRAZIONE INTUITIVA

$$\begin{aligned}\mathbb{P}(W_t \geq a) &= \mathbb{P}(W_t \geq a \mid \tau_a \leq t)\mathbb{P}(\tau_a \leq t) + \mathbb{P}(W_t \geq a \mid \tau_a > t)\mathbb{P}(\tau_a > t) =\\ &= \mathbb{P}(W_t \geq a \mid \tau_a \leq t)\mathbb{P}(\tau_a \leq t) = \frac{1}{2}\mathbb{P}(\tau_a \leq t)\end{aligned}$$

in quanto ovviamente

$$\mathbb{P}(W_t \geq a \mid \tau_a > t) = 0 \text{ e } \mathbb{P}(W_t \geq a \mid \tau_a \leq t) = \mathbb{P}(W_t \leq a \mid \tau_a \leq t) = \frac{1}{2}.$$

Nell'ultima uguaglianza è implicito l'uso del fatto che l'informazione contenuta nell'evento $\{\tau_a \leq t\}$ è riassumibile nel fatto che $\{W_{\tau_a} = a\}$ e che $W_{\tau_a+s} - W_{\tau_a}$ è ancora un moto browniano. In realtà formalmente c'è qualche problema in quanto pur essendo $\{\tau_a \leq t\} \in \mathcal{F}_\tau$, e quindi

$$\begin{aligned}\mathbb{P}(W_t \geq a \mid \tau_a \leq t) &= \mathbb{P}(W_t \geq W_{\tau_a} \mid \tau_a \leq t) = \mathbb{P}(W_t - W_{\tau_a} \geq 0 \mid \tau_a \leq t) =\\ &= \mathbb{P}(Y_{t-\tau_a} \geq 0 \mid \tau_a \leq t) = \mathbb{P}(Y_{t-\tau_a} \geq 0)\end{aligned}$$

essendoci di mezzo τ_a non è chiarissimo, anche se intuitivo che $\mathbb{P}(Y_{t-\tau_a} \geq 0) = \frac{1}{2}$.

DIMOSTRAZIONE FORMALE:

Faremo vedere che le trasformate di Laplace rispetto a t, di $\mathbb{P}(W_t \geq a)$ e di $\frac{1}{2}\mathbb{P}(\tau_a \leq t)$ sono uguali, utilizzando il fatto che $W_{\tau_a} = a$, per la continuità delle traiettorie di W_t, e che $W_{\tau+s} - W_\tau$ è indipendente da $\mathcal{F}_\tau$, e quindi da τ: qualunque sia $\alpha \geq 0$,

$$\int_0^\infty e^{-\alpha t}\mathbb{P}(W_t \geq a)dt = \text{(è necessario usare la congiunta misurabilità in } (t,\omega) \text{ di } W_t\text{, per scambiare gli integrali)}$$

$$= \mathbb{E}[\int_0^\infty e^{-\alpha t} I_{[a,+\infty)}(W_t)dt] =$$

$$(W_t < a \text{ per } t < \tau_a)$$

$$= \mathbb{E}[\int_{\tau_a}^{\infty} e^{-\alpha t} I_{[a,+\infty)}(W_t) dt] = \mathbb{E}[\int_{\tau_a}^{\infty} e^{-\alpha \tau_a} e^{-\alpha(t-\tau_a)} I_{[0,+\infty)}(W_t - W_{\tau_a}) dt] =$$

$$(W_t - W_{\tau_a} = Y_{t-\tau_a} \text{ e cambio di variabile})$$

$$= \mathbb{E}[e^{-\alpha\tau_a} \int_0^{\infty} e^{-\alpha s} I_{[0,+\infty)}(Y_s) ds] =$$

$$(\{Y_t, t \geq 0\} \text{ e } \tau_a \text{ sono indipendenti})$$

$$= \mathbb{E}[e^{-\alpha\tau_a}]\mathbb{E}[\int_0^{\infty} e^{-\alpha s} I_{[0,+\infty)}(Y_s) ds] = \mathbb{E}[e^{-\alpha\tau_a}] \int_0^{\infty} e^{-\alpha s} \mathbb{P}(Y_s \geq 0) ds] =$$

$$= \frac{1}{2\alpha} \mathbb{E}[e^{-\alpha\tau_a}].$$

In modo analogo

$$\int_0^{\infty} e^{-\alpha t} \mathbb{P}(\tau_a \leq t) dt = \mathbb{E}[\int_0^{\infty} e^{-\alpha t} I_{\{\tau_a \leq t\}} dt] = \mathbb{E}[\int_{\tau_a}^{\infty} e^{-\alpha t} dt] = \frac{1}{\alpha} \mathbb{E}[e^{-\alpha\tau_a}].$$

In alcuni testi viene dato come principio di riflessione la seguente proprietà (la cui dimostrazione è simile alla dimostrazione della proprietà di Markov forte):

PRINCIPIO DI RIFLESSIONE: seconda formulazione
Per ogni tempo d'arresto τ finito, il processo "riflesso"

$$\begin{cases} Y_t = W_t & \textit{per } t \leq \tau, \\ Y_t - Y_\tau = -(W_t - W_\tau) & \textit{per } t > \tau, \end{cases}$$

è ancora un moto browniano.

10.4 TEMPI DI USCITA DA UNA STRISCIA

Anche nel caso del moto Browniano si possono ottenere delle applicazioni simili a quelle della rovina del giocatore. Più in generale si può considerare il processo di Wiener con coefficiente di drift μ e coefficiente di diffusione σ, cioè

$$X_t := \sigma W_t + \mu t,$$

e il tempo τ di prima uscita da una striscia

$$\tau \equiv \tau(-a, b) := \inf\{t > 0 : X_t \notin (-a, b)\}, a, b > 0,$$

e cercare di calcolare la probabilità $p \equiv p(x, a, b) := \mathbb{P}_x(X_\tau = -a)$, e la trasformata di Laplace $\mathbb{E}_x[\exp\{-\alpha\tau\}]$, per $x \in (-a, b)$.

Sappiamo che, essendo $W_t = (X_t - \mu t)/\sigma$, il processo

$$Z_t := \exp\{\theta(X_t - \mu t)/\sigma - \theta^2 t/2\}$$

è una martingala con

$$\mathbb{E}_x[Z_t] = \exp\{\frac{\theta}{\sigma}x\}.$$

Anche $Z_{t\wedge\tau}$, è una martingala che inoltre converge a Z_τ. La convergenza è limitata:

$$Z_{t\wedge\tau} \leq \exp\{|\theta| \max(a,b)/\sigma\}, \text{ se } \frac{\theta\mu}{\sigma} + \frac{\theta^2}{2} \geq 0$$

e quindi anche

$$\mathbb{E}_x[\exp\{\frac{\theta}{\sigma}X_\tau\} \exp\{-\frac{\theta}{2}(2\frac{\mu}{\sigma} + \theta)\tau\}] = \exp\{\frac{\theta}{\sigma}x\}.$$

Scegliendo θ in modo che $2\frac{\mu}{\sigma} + \theta = 0$, ovvero $\theta{=}-2\frac{\mu}{\sigma}$, la precedente uguaglianza diviene

$$\mathbb{E}_x[\exp\{\frac{\theta}{\sigma}X_\tau\}] = \mathbb{P}_x[X_\tau = -a]\exp\{-\frac{\theta}{\sigma}a\} + \mathbb{P}_x[X_\tau = b]\exp\{\frac{\theta}{\sigma}b\} = \exp\{\frac{\theta}{\sigma}x\},$$

da cui subito, tenendo conto del valore di θ,

$$p\exp\{2\frac{\mu}{\sigma^2}a\} + (1-p)\exp\{-2\frac{\mu}{\sigma^2}b\} = \exp\{-2\frac{\mu}{\sigma^2}x\},$$

e quindi

$$p = \frac{\exp\{-2\frac{\mu}{\sigma^2}x\} - \exp\{-2\frac{\mu}{\sigma^2}b\}}{\exp\{2\frac{\mu}{\sigma^2}a\{-\exp\{-2\frac{\mu}{\sigma^2}b\}}.$$

Per ogni $\alpha > 0$, si pongano ora θ^+_α e θ^-_α i due valori per cui $\alpha := \frac{\theta}{2}(2\frac{\mu}{\sigma} + \theta)$, cioè le due soluzioni di $\theta^2 + 2\frac{\mu}{\sigma}\theta - 2\alpha = 0$, ovvero

$$\theta^+_\alpha, \theta^-_\alpha := -\frac{\mu}{\sigma} \pm \sqrt{\left(\frac{\mu}{\sigma}\right)^2 + 2\alpha}.$$

Come prima si ottiene che

$$\mathbb{E}_x[\exp\{\frac{\theta^+_\alpha}{\sigma}X_\tau\}\exp\{-\alpha\tau\}] = \exp\{\frac{\theta^+_\alpha}{\sigma}x\},\ \mathbb{E}_x[\exp\{\frac{\theta^-_\alpha}{\sigma}X_\tau\}\exp\{-\alpha\tau\}] = \exp\{\frac{\theta^-_\alpha}{\sigma}x\}.$$

Considerando che per $\theta = \theta^+_\alpha, \theta^-_\alpha$

$$\mathbb{E}_x[\exp\{\frac{\theta}{\sigma}X_\tau\}\exp\{-\alpha\tau\}] = \exp\{-\frac{\theta}{\sigma}a\}\mathbb{E}_x[I_{\{X_\tau=-a\}}\exp\{-\alpha\tau\}] + \exp\{\frac{\theta}{\sigma}b\}\mathbb{E}_x[I_{\{X_\tau=b\}}\exp\{-\alpha\tau\}]$$

si ottiene immediatamente il seguente sistema lineare nelle incognite

$$\begin{cases} y_{-a} & := & \mathbb{E}_x[I_{\{X_\tau=-a\}}\exp\{-\alpha\tau\}] \\ y_b & := & \mathbb{E}_x[I_{\{X_\tau=b\}}\exp\{-\alpha\tau\}] \end{cases}$$

$$\begin{cases} \exp\{-\frac{\theta^+_\alpha}{\sigma}a\}y_{-a} + \exp\{\frac{\theta^+_\alpha}{\sigma}b\}y_b & = & \exp\{\frac{\theta^+_\alpha}{\sigma}x\} \\ \exp\{-\frac{\theta^-_\alpha}{\sigma}a\}y_{-a} + \exp\{\frac{\theta^-_\alpha}{\sigma}b\}y_b & = & \exp\{\frac{\theta^-_\alpha}{\sigma}x\} \end{cases}$$

di cui si ricava la soluzione, ovvero

$$\begin{cases} y_{-a} = \frac{\exp\{\nu_\alpha^+ x\}\exp\{\nu_\alpha^- b\}-\exp\{\nu_\alpha^+ b\}\exp\{\nu_\alpha^- x\}}{\exp\{-\nu_\alpha^+ a\}\exp\{\nu_\alpha^- b\}-\exp\{\nu_\alpha^+ b\}\exp\{-\nu_\alpha^- a\}} \\ \\ y_b = \frac{\exp\{-\nu_\alpha^+ a\}\exp\{\nu_\alpha^- x\}-\exp\{\nu_\alpha^+ x\}\exp\{-\nu_\alpha^- a\}}{\exp\{-\nu_\alpha^+ a\}\exp\{\nu_\alpha^- b\}-\exp\{\nu_\alpha^+ b\}\exp\{-\nu_\alpha^- a\}}, \end{cases}$$

avendo posto

$$\nu_\alpha^\pm := \frac{\theta_\alpha^\pm}{\sigma} = -\frac{\mu}{\sigma^2} \pm \sqrt{\left(\frac{\mu}{\sigma^2}\right)^2 + 2\frac{\alpha}{\sigma^2}}.$$

Basta poi considerare la somma

$$\mathbb{E}_x[\exp\{-\alpha\tau\}] = y_{-a} + y_b =$$

$$= \frac{\exp\{\nu_\alpha^+ x\}\exp\{\nu_\alpha^- b\} - \exp\{\nu_\alpha^+ b\}\exp\{\nu_\alpha^- x\} + \exp\{-\nu_\alpha^+ a\}\exp\{\nu_\alpha^- x\} - \exp\{\nu_\alpha^+ x\}\exp\{-\nu_\alpha^- a\}}{\exp\{-\nu_\alpha^+ a\}\exp\{\nu_\alpha^- b\} - \exp\{\nu_\alpha^+ b\}\exp\{-\nu_\alpha^- a\}}.$$

ESERCIZI

1) Nel caso $x = 0$, $\mu > 0$ mandare a ad infinito in modo da ottenere la trasformata di Laplace di $\tau := \tau(-\infty, b)$

Soluzione:
Si noti che, essendo $\alpha, \mu > 0$, si ha $\nu_\alpha^+ > 0$ e $\nu_\alpha^- < 0$. Di conseguenza $\exp\{-\nu_\alpha^- a\} \to \infty$, mentre $\exp\{-\nu_\alpha^+ a\} \to 0$, e dalla precedente espressione per $\mathbb{E}_0[\exp\{-\alpha\tau(-a,b)\}]$ si ricava, per $a \to \infty$

$$\mathbb{E}_0[\exp\{-\alpha\tau(-a,b)\}] \to \mathbb{E}_0[\exp\{-\alpha\tau(-\infty,b)\}] = \frac{1}{\exp\{\nu_\alpha^+ b\}} = \exp\left[\left(\frac{\mu}{\sigma^2} - \sqrt{\left(\frac{\mu}{\sigma^2}\right)^2 + 2\frac{\alpha}{\sigma^2}}\right) b\right].$$

2) Si consideri il caso del moto browniano o Wiener standard, cioè con μ=0 e σ=1, e con $x = 0$ ed $a = b$.

Soluzione:

$$\mathbb{E}_0[\exp\{-\alpha\tau(-a,a)\}] = \frac{\exp\{\nu_\alpha^- a\} - \exp\{\nu_\alpha^+ a\} + \exp\{-\nu_\alpha^+ a\} - \exp\{-\nu_\alpha^- a\}}{\exp\{-\nu_\alpha^+ a\}\exp\{\nu_\alpha^- a\} - \exp\{\nu_\alpha^+ a\}\exp\{-\nu_\alpha^- a\}},$$

e quindi, poiché in questo caso $\nu_\alpha^- = -\nu_\alpha^+ = -\sqrt{2\alpha}$,

$$\begin{aligned} \mathbb{E}_0[\exp\{-\alpha\tau(-a,a)\}] &= \frac{\exp\{-\sqrt{2\alpha}a\} - \exp\{\sqrt{2\alpha}a\} + \exp\{-\sqrt{2\alpha}a\} - \exp\{\sqrt{2\alpha}a\}}{\exp\{-\sqrt{2\alpha}a\}\exp\{-\sqrt{2\alpha}a\} - \exp\{\sqrt{2\alpha}a\}\exp\{\sqrt{2\alpha}a\}} \\ &= 2\frac{\exp\{\sqrt{2\alpha}a\} - \exp\{-\sqrt{2\alpha}a\}}{\exp\{2\sqrt{2\alpha}a\} - \exp\{-2\sqrt{2\alpha}a\}} \\ &= 2\frac{\sinh(\sqrt{2\alpha}a)}{\sinh(2\sqrt{2\alpha}a)} = 2\frac{\sinh(\sqrt{2\alpha}a)}{2\sinh(\sqrt{2\alpha}a)\cosh(\sqrt{2\alpha}a)} \\ &= \frac{1}{\cosh(\sqrt{2\alpha}a)} \end{aligned}$$