Capitolo 1

Introduzione ai Processi Stocastici

1.1 Prime definizioni

1.1.1 Processi stocastici

Ricordiamo che uno *spazio di probabilità* è una terna $(\Omega, \mathcal{F}, P)$ dove Ω è un insieme, $\mathcal{F}$ è una σ-algebra di parti di Ω, P è una misura di probabilità su $(\Omega, \mathcal{F})$; la coppia $(\Omega, \mathcal{F})$ è chiamata *spazio misurabile*. In uno spazio topologico E indicheremo con $\mathcal{B}(E)$ la σ-algebra dei boreliani (la più piccola σ-algebra che contiene gli aperti). Una *variabile aleatoria* Y su $(\Omega, \mathcal{F})$ a valori in uno spazio misurabile $(E, \mathcal{E})$ è una funzione $Y : \Omega \to E$ misurabile da $(\Omega, \mathcal{F})$ in $(E, \mathcal{E})$, cioè tale che $\{\omega \in \Omega : Y(\omega) \in B\} \in \mathcal{F}$ per ogni $B \in \mathcal{E}$. Una *variabile aleatoria reale* Y su $(\Omega, \mathcal{F})$ è una variabile aleatoria su $(\Omega, \mathcal{F})$ a valori in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Sia T un insieme non vuoto. L'insieme T sarà l'*insieme dei parametri* del processo stocastico. A livello interpretativo, può essere ad esempio un insieme di tempi, oppure di posizioni spaziali o di spazio-tempo. In alcuni momenti sarà necessario supporre di avere uno spazio misurabile $(T, \mathcal{T})$ ed in altri uno spazio topologico T, con $\mathcal{T} = \mathcal{B}(T)$; altrimenti T è del tutto arbitrario. Tutto ciò fino al momento in cui introdurremo le filtrazioni; con l'introduzione del concetto di filtrazione, restringeremo l'attenzione al caso in sia $T \subset [0, \infty)$, $\mathcal{T} = \mathcal{B}(T)$ ed interpreteremo T come insieme dei tempi. Tutti gli sviluppi avanzati del corso riguarderanno il caso $T = [0, \infty)$ (o $T = [0, t_0]$) per cui avrebbe senso restringersi fin da ora a quel caso. Però può essere concettualmente interessante osservare che questi primi paragrafi hanno carattere più generale, per cui per ora supporremo solo che T sia un insieme, non vuoto.

Siano dati, oltre a T, uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ ed uno spazio misurabile $(E, \mathcal{E})$ (detto spazio degli stati). Un *processo stocastico* $X = (X_t)_{t \in T}$, definito su $(\Omega, \mathcal{F}, P)$ a valori in $(E, \mathcal{E})$ è una funzione X definita su $\Omega \times T$ a valori in E, tale che per ogni $t \in T$ la funzione $\omega \longmapsto X_t(\omega)$ sia misurabile da $(\Omega, \mathcal{F})$ in $(E, \mathcal{E})$. Ne segue che, dati $t_1 < ... < t_n \in T$, la funzione

$$\omega \longmapsto (X_{t_1}(\omega), ..., X_{t_n}(\omega))$$

è misurabile da $(\Omega, \mathcal{F})$ in $(E^n, \otimes^n \mathcal{E})$. Le leggi di probabilità di questi vettori (leggi immagine di P rispetto a queste applicazioni) si chiamano *distribuzioni di dimensione finita* del processo.

Le funzioni $t \longmapsto X_t(\omega)$ da T in E (ad ω fissato) si dicono *realizzazioni* o *traiettorie* del processo stocastico. Il termine traiettoria va forse riservato al caso in cui sia $T \subset [0, \infty)$.

Due processi stocastici $X = (X_t)_{t \in T}$ e $Y = (Y_t)_{t \in T}$ si dicono *equivalenti* se hanno le stesse distribuzioni di dimensione finita. Si dicono *modificazione* (o *versione*) uno dell'altro se per ogni $t \in T$ vale

$$P(X_t = Y_t) = 1.$$

Si dicono *indistinguibili* se

$$P(X_t = Y_t \text{ per ogni } t \in T) = 1.$$

Due processi indistinguibili sono modificazione uno dell'altro. Due processi che sono modificazione uno dell'altro sono equivalenti. Si veda l'esercizio 1. Si noti che, a differenza delle altre due, la definizione di processi equivalenti non richiede che essi siano definiti sullo stesso spazio $(\Omega, \mathcal{F}, P)$.

Supponiamo che T ed E siano spazi topologici e siano $\mathcal{T} = \mathcal{B}(T)$, $\mathcal{E} = \mathcal{B}(E)$. Un processo si dice *continuo* se le sue realizzazioni sono funzioni continue; si dice *q.c. continuo* se ciò avviene per P-quasi ogni realizzazione.

Quando $T \subset [0, \infty)$, le definizioni di continuo a destra e/o a sinistra, costante a tratti ed altre della stessa natura sono simili. Un processo si dice *càdlàg* (continue à droite, limite à gauche) se le sue traiettorie sono continue a destra ed hanno limite finito a sinistra.

Se $(T, \mathcal{T})$ è uno spazio misurabile, possiamo dare la seguente definizione. Un processo si dice *misurabile* se l'applicazione $(t, \omega) \longmapsto X_t(\omega)$ è misurabile da $(T \times \Omega, \mathcal{T} \otimes \mathcal{F})$ in $(E, \mathcal{E})$.

Elenchiamo anche alcune nozioni che si usano soprattutto nella cosiddetta teoria della correlazione dei processi stazionari, per noi marginale ma toccata almeno nel caso dei processi gaussiani. Supponiamo che sia $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Indichiamo con $E[\cdot]$ la speranza matematica su $(\Omega, \mathcal{F}, P)$ ($E[Y] = \int_\Omega Y dP$, per ogni Y v.a. reale $(\Omega, \mathcal{F}, P)$ integrabile), con $Var[\cdot]$ la varianza ($Var[Y] = E\left[(Y - E[Y])^2\right]$ quando Y è di quadrato integrabile), con $Cov(\cdot, \cdot)$ la covarianza ($Cov(Y, Z) = E[(Y - E[Y])(Z - E[Z])]$ se Y e Z sono v.a. di quadrato integrabile). Sia X un processo stocastico a valori in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, su $(\Omega, \mathcal{F}, P)$. Se $E[|X_t|] < \infty$ per ogni $t \in T$, chiamiamo $m(t) = E[X_t]$, $t \in T$, *funzione valor medio* del processo X. Se $E[X_t^2] < \infty$ per ogni $t \in T$, chiamiamo $\sigma^2(t) = Var[X_t]$, $t \in T$, *funzione varianza* di X e chiamiamo $C(t, s) = Cov(X_t, X_s)$, $t, s \in T$, *funzione covarianza* di X.

Esercizio 1 *Costruire due processi equivalenti che non siano modificazione uno dell'altro e due processi modificazione uno dell'altro che non siano indistinguibili.*

1.1.2 Legge di un processo

Sia E^T lo spazio di tutte le funzioni $f : T \to E$. Indichiamo con $\mathcal{S}$ l'insieme di tutte le n-ple ordinate $\tau = (t_1, ..., t_n)$ di elementi di T tali che $t_i \neq t_j$ per ogni $i \neq j$, $i, j = 1, ..., n$. Indichiamo con $\mathcal{E}^n$ la σ-algebra prodotto di n copie di $\mathcal{E}$.

Presa $\tau = (t_1, ..., t_n) \in \mathcal{S}$ e preso $B \in \mathcal{E}^n$ consideriamo l'insieme

$$C(\tau, B) \subset E^T$$

di tutte le funzioni $f : T \to E$ tali che

$$(f(t_1), ..., f(t_n)) \in B.$$

Lo chiameremo *insieme cilindrico* (con base B e coordinate $t_1, ..., t_n$). Sia $\mathcal{A}$ la famiglia di tutti gli insiemi cilindrici $C(\tau, B)$, con $n \in \mathbb{N}$, $\tau = (t_1, ..., t_n) \in \mathcal{S}$ e $B \in \mathcal{E}^n$. La famiglia $\mathcal{A}$ è un'algebra di parti di E^T: contiene E^T; il complementare di un insieme della forma $\{(f(t_1), ..., f(t_n)) \in B\}$ è l'insieme cilindrico $\{(f(t_1), ..., f(t_n)) \in B^c\}$; è chiusa per intersezione finita. Quest'ultima proprietà è noiosa da scrivere in generale ma è ovvia se si pensa ad esempio al caso particolare dei due insiemi cilindrici $\{f(t_1) \in B_1\}$, $\{f(t_2) \in B_2\}$, con $\tau = (t_1, t_2)$, $B_1, B_2 \in \mathcal{E}$, la cui intersezione è l'insieme $\{(f(t_1), f(t_2)) \in B_1 \times B_2\}$ che appartiene ad $\mathcal{A}$.

Sia $\mathcal{E}^T$ la più piccola σ-algebra di parti di E^T che contiene gli insiemi cilindrici.

Dato un processo stocastico $X = (X_t)_{t \in T}$ definito su uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ a valori in uno spazio misurabile $(E, \mathcal{E})$ consideriamo l'applicazione

$$\omega \overset{\Lambda}{\longmapsto} X_\cdot(\omega)$$

da $(\Omega, \mathcal{F})$ in $\left(E^T, \mathcal{E}^T\right)$. Essa è misurabile. Infatti, preso un insieme cilindrico $C(\tau, B)$, $\tau = (t_1, ..., t_n) \in \mathcal{S}$, $B \in \mathcal{E}^n$, la sua controimmagine è l'insieme

$$\{(X_{t_1}, ..., X_{t_n}) \in B\}$$

e questo insieme sappiamo essere misurabile, un elemento di $\mathcal{F}$. Rammentiamo che per verificare la misurabilità di un'applicazione basta farlo su una famiglia generante la σ-algebra in arrivo (in simboli generali, se $X : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ soddisfa $X^{-1}(B) \in \mathcal{F}$ per ogni $B \in \mathcal{G}$, dove $\mathcal{G} \subset \mathcal{E}$ ed $\mathcal{E} = \sigma(\mathcal{G})$, allora X è misurabile; infatti, si prenda la famiglia $\mathcal{H} \subset \mathcal{P}(E)$ degli insiemi $B \subset E$ tali che $X^{-1}(B) \in \mathcal{F}$; si verifica che $\mathcal{H}$ è una σ-algebra e contiene $\mathcal{G}$, quindi contiene $\sigma(\mathcal{G})$).

Possiamo quindi considerare la legge dell'applicazione Λ (misura immagine di P attraverso Λ). Si chiamerà *legge del processo*. E' una misura di probabilità su $\left(E^T, \mathcal{E}^T\right)$. Indichiamola col simbolo P_X. Vale

$$P_X(C(\pi, B)) = P((X_{t_1}, ..., X_{t_n}) \in B)$$

per ogni $\tau = (t_1, ..., t_n) \in \mathcal{S}$ e $B \in \mathcal{E}^n$.

Proposizione 1 *Due processi X ed X', definiti rispettivamente su $(\Omega, \mathcal{F}, P)$ e $(\Omega', \mathcal{F}', P')$ a valori in $(E, \mathcal{E})$, sono equivalenti se e solo se hanno la stessa legge su $\left(E^T, \mathcal{E}^T\right)$.*

Proof. Se hanno la stessa legge allora, presi $\tau = (t_1, ..., t_n) \in \mathcal{S}$ e $B \in \mathcal{E}^n$, vale

$$\begin{aligned} P\left((X_{t_1}, ..., X_{t_n}) \in B\right) &= P_X\left(C(\tau, B)\right) = P'_{X'}\left(C(\tau, B)\right) \\ &= P'\left(\left(X'_{t_1}, ..., X'_{t_n}\right) \in B\right) \end{aligned}$$

quindi hanno le stesse distribuzioni di dimensione finita.

Viceversa, se hanno le stesse distribuzioni di dimensione finita, vale

$$\begin{aligned} P_X\left(C(\tau, B)\right) &= P\left((X_{t_1}, ..., X_{t_n}) \in B\right) = P'\left(\left(X'_{t_1}, ..., X'_{t_n}\right) \in B\right) \\ &= P'_{X'}\left(C(\tau, B)\right) \end{aligned}$$

quindi le misure di probabilità P_X e $P'_{X'}$, misure su $\left(E^T, \mathcal{E}^T\right)$, coincidono sugli insiemi cilindrici. La famiglia $\mathcal{A}$ di tali insiemi genera $\mathcal{E}^T$ ed è chiusa per intersezione finita e quindi, per un noto teorema di Caratheodory, le due misure coincidono su tutta $\mathcal{E}^T$. La dimostrazione è completa. ■

1.1.3 Legge nello spazio delle funzioni continue

Esaminiamo adesso un problema più delicato. Supponiamo che X sia continuo. Per capire bene il problema senza complicazioni topologiche collaterali, iniziamo discutendo il caso in cui

$$T = \mathbb{R}_+ := [0, \infty), \qquad E = \mathbb{R}^n \tag{1.1}$$

e le σ-algebre sono quelle dei boreliani. Indichiamo con C_0 l'insieme $C(\mathbb{R}_+; \mathbb{R}^n)$ delle funzioni continue da $\mathbb{R}_+$ in $\mathbb{R}^n$, munito della topologia della convergenza uniforme sui compatti.

L'insieme $\Lambda(\Omega)$ non è solo un sottoinsieme di $E^{\mathbb{R}_+}$ ma anche di C_0. E' naturale pensare che Λ definisca una misura immagine su $(C_0, \mathcal{B}(C_0))$. Purtroppo si può dimostrare che $C_0 \notin \mathcal{B}(E)^{\mathbb{R}_+}$ (omettiamo la dimostrazione) e questo non permette facilmente di "restringere" la legge del processo, precedentemente definita su $\left(E^{\mathbb{R}_+}, \mathcal{B}(E)^{\mathbb{R}_+}\right)$, ad una legge su $(C_0, \mathcal{B}(C_0))$ (in realtà anche questa strada è percorribile con opportune considerazioni basate sulla misura esterna, applicabili dopo aver verificato che la misura esterna dell'insieme non misurabile C_0 è pari ad 1).

Per aggirare questo ostacolo basta dimostrare il risultato seguente.

Proposizione 2 *Se $X = (X_t)_{t \geq 0}$ è un processo continuo a valori reali, allora Λ è misurabile da $(\Omega, \mathcal{F})$ in $(C_0, \mathcal{B}(C_0))$. Possiamo quindi considerare la misura immagine di P attraverso Λ, su $(C_0, \mathcal{B}(C_0))$, che chiameremo legge del processo X su $(C_0, \mathcal{B}(C_0))$.*

Proof. La σ-algebra $\mathcal{B}(C_0)$ è generata dagli insiemi della forma

$$B(f_0, n, R) = \left\{ f \in C_0 : \max_{t \in [0,n]} |f(t) - f_0(t)| \leq R \right\}$$

al variare di $f_0 \in C_0$, $n \in \mathbb{N}$, $R > 0$. Sia $\{t_j\} \subset \mathbb{R}_+$ una successione che conta i razionali non negativi. Indichiamo con $\left\{ t_j^{(n)} \right\}$ la sotto-successione composta dai t_j che appartengono a $[0, n]$; $\left\{ t_j^{(n)} \right\}$ è densa in $[0, n]$. Vale

$$B(f_0, n, R) = \left\{ f \in C_0 : \left| f\left(t_j^{(n)}\right) - f_0\left(t_j^{(n)}\right) \right| \leq R, \ \forall\, j \in \mathbb{N} \right\}.$$

Poniamo inoltre

$$B(f_0, n, N, R) = \left\{ f \in C_0 : \left| f\left(t_j^{(n)}\right) - f_0\left(t_j^{(n)}\right) \right| \leq R, \ \forall\, j \leq N \right\}$$

per ciascun $N \in \mathbb{N}$. La controimmagine di $B(f_0, n, N, R)$ attraverso Λ è l'insieme

$$\left\{ \omega \in \Omega : \left| X_{t_j^{(n)}}(\omega) - f_0\left(t_j^{(n)}\right) \right| \leq R, \ \forall\, j \leq N \right\}$$

che appartiene ad $\mathcal{F}$ (è anche la controimmagine di un insieme cilindrico). Vale inoltre

$$B(f_0, n, R) = \bigcap_{N \in \mathbb{N}} B(f_0, n, N, R).$$

La controimmagine di questo insieme è intersezione numerabile di elementi di $\mathcal{F}$, quindi è elemento di $\mathcal{F}$. Siccome questi insiemi, come detto sopra, generano $\mathcal{B}(C_0)$, Λ è misurabile tra le σ-algebre indicate. La dimostrazione è completa. ∎

Osservazione 1 *Con ragionamenti non molto diversi si dimostra che due processi continui con le stesse distribuzioni di dimensione finita, hanno la stessa legge su* $(C_0, \mathcal{B}(C_0))$.

Se un processo X è solamente q.c. continuo, definisce ugualmente una misura di probabilità su $(C_0, \mathcal{B}(C_0))$ del tutto analoga al caso precedente, che continueremo a chiamare misura immagine di P attraverso Λ su $(C_0, \mathcal{B}(C_0))$. Infatti, esiste un insieme $\Omega' \in \mathcal{F}$ di P-misura 1 tale che $\Lambda(\Omega') \subset C_0$. Consideriamo lo spazio probabilizzato $(\Omega', \mathcal{F}', P')$ dove $\mathcal{F}' = \mathcal{F} \cap \Omega'$ e P' è la restrizione di P ad $\mathcal{F}'$. Ora $\Lambda : \Omega' \to C_0$ è ben definita e induce una misura di probabilità su $(C_0, \mathcal{B}(C_0))$. Si verifica facilmente che essa non dipende dalla scelta di Ω' con le proprietà precedenti.

Le considerazioni illustrate sopra si estendono a varie situazioni più generali. Una molto simile è quella in cui T è uno spazio metrico, unione numerabile di compatti (spazio metrico σ-compatto) ed E è uno spazio metrico separabile.

1.2 Teorema di estensione di Kolmogorov

Abbiamo visto che, dato un processo $X = (X_t)_{t \in T}$, questo definisce una misura di probabilità su $\left(E^T, \mathcal{E}^T\right)$. Quando vale il viceversa, cioè per quali misure μ di probabilità su $\left(E^T, \mathcal{E}^T\right)$ esiste un processo che ha μ come legge? Per tutte: basta prendere il processo canonico, $\Omega = E^T$, $\mathcal{F} = \mathcal{E}^T$, $X_t(\omega) = \omega(t)$. Meno banale è il problema di inversione che ora descriveremo.

Data una misura di probabilità μ su $\left(E^T, \mathcal{E}^T\right)$, definiamo le sue distribuzioni di dimensione finita nel seguente modo. Presa $\tau = (t_1, ..., t_n) \in \mathcal{S}$, indichiamo con $\pi_\tau : E^T \to E^n$ l'applicazione $\pi_\tau(f) = (f(t_1), ..., f(t_n))$. E' misurabile, rispetto a $\mathcal{E}^T$ e $\mathcal{E}^n$ (preso $B \in \mathcal{E}^n$ la sua controimmagine è l'insieme cilindrico $\left\{f \in E^T : (f(t_1), ..., f(t_n)) \in B\right\}$, che appartiene ad $\mathcal{E}^T$). La distribuzione di dimensione n di μ relativa a π è la legge immagine μ_τ di μ attraverso π_τ:

$$\mu_\tau(B) = \mu\left(f \in E^T : (f(t_1), ..., f(t_n)) \in B\right), \qquad B \in \mathcal{E}^n.$$

Se $\mu = P_X$, essa è la legge del vettore aleatorio $(X_{t_1}, ..., X_{t_n})$ e $\{\mu_\tau; \tau \in \mathcal{S}\}$ è la famiglia delle distribuzione di dimensione finita del processo X.

Il problema di inversione è il seguente: data una famiglia di misure di probabilità $\{\mu_\tau; \tau \in \mathcal{S}\}$ (si sottintende che se $\tau = (t_1, ..., t_n) \in \mathcal{S}$, allora μ_τ è una misura di probabilità su $(E^n, \mathcal{E}^n)$), esiste una misura di probabilità μ su $\left(E^T, \mathcal{E}^T\right)$ di cui $\{\mu_\tau; \tau \in \mathcal{S}\}$ sia la famiglia delle distribuzioni di dimensione finita? Siccome l'esistenza di un processo con legge μ è ovvia, il problema è equivalente a: data una famiglia di misure di probabilità $\{\mu_\tau; \tau \in \mathcal{S}\}$, esiste uno spazio probabilizzato $(\Omega, \mathcal{F}, P)$ ed un processo X su $(\Omega, \mathcal{F}, P)$ che abbia $\{\mu_\tau; \tau \in \mathcal{S}\}$ come famiglia delle distribuzioni di dimensione finita?

Osservazione 2 *Alla base dei seguenti ragionamenti c'è il fatto che un insieme cilindrico non ha una sola rappresentazione. Ad esempio, dati* $t_1, ..., t_n, t_{n+1} \in T$, $B \in \mathcal{E}^n$, *vale*

$$\left\{f \in E^T : (f(t_1), ..., f(t_n)) \in B\right\} = \left\{f \in E^T : (f(t_1), ..., f(t_n), f(t_{n+1})) \in B \times E\right\}$$

oppure, dati $t_1, t_2 \in T$, $B_1, B_2 \in \mathcal{E}$, *vale*

$$\left\{f \in E^T : (f(t_1), f(t_2)) \in B_1 \times B_2\right\} = \left\{f \in E^T : (f(t_2), f(t_1)) \in B_2 \times B_1\right\}.$$

Servono due ingredienti: una proprietà di compatibilità tra le μ_τ ed un po' di regolarità dello spazio $(E, \mathcal{E})$. La proprietà di compatibilità tra le μ_τ si intuisce facilmente a posteriori, quando esse sono le distribuzioni di dimensione finita di μ. Si tratta di due condizioni, che a parole potremmo chiamare "invarianza sotto permutazioni degli indici" e "invarianza sotto contrazioni degli indici". Vediamo la prima. Se $\tau = (t_1, ..., t_n) \in \mathcal{S}$ e se $(i_1, ..., i_n)$ è una permutazione di $(1, ..., n)$, indicata con

$P_{(i_1,...,i_n)} : E^n \to E^n$ l'applicazione che manda la generica sequenza $(x_1, ..., x_n)$ nella $(x_{i_1}, ..., x_{i_n})$, deve valere (per le distribuzioni di dimensione finita di un processo)

$$\mu_{(t_1,...,t_n)}(B) = \mu_{\left(t_{i_1},...,t_{i_n}\right)}\left(P_{(i_1,...,i_n)}(B)\right). \tag{1.2}$$

Vediamo ora la seconda. Data $\tau = (t_1, ..., t_n) \in \mathcal{S}$, indichiamo con $\tau_{\widehat{n}}$ la sequenza $\tau_{\widehat{n}} = (t_1, ..., t_{n-1})$ (la sequenza ottenuta omettendo t_n dalla τ), le misure μ_τ e $\mu_{\tau_{\widehat{n}}}$ sono legate dalla proiezione $\pi_{\widehat{n}} : E^n \to E^{n-1}$ che manda una generica sequenza $(x_1, ..., x_n) \in E^n$ nella sequenza $(x_1, ..., x_{n-1}) \in E^{n-1}$ (la sequenza ottenuta omettendo l'ultima componente di $(x_1, ..., x_n)$). Vale

$$\mu_{\tau_{\widehat{n}}} = \pi_{\widehat{n}}(\mu_\tau) \tag{1.3}$$

nel senso che $\mu_{\tau_{\widehat{n}}}$ è la legge immagine di μ_τ attraverso $\pi_{\widehat{n}}$, ovvero esplicitamente $\mu_{\tau_{\widehat{n}}}(B) = \mu_\tau(\pi_{\widehat{n}} \in B)$ per ogni $B \in \mathcal{E}^{n-1}$.

Si può verificare facilmente che è equivalente richiedere queste condizioni per insiemi B di tipo rettangolare, rispetto a cui esse si scrivono più agevolmente. Possiamo richiedere che

$$\mu_{(t_1,...,t_n)}(B_1 \times \cdots \times B_n) = \mu_{\left(t_{i_1},...,t_{i_n}\right)}(B_{i_1} \times \cdots \times B_{i_n})$$

$$\mu_{(t_1,...,t_n)}(B_1 \times \cdots \times E) = \mu_{(t_1,...,t_{n-1})}(B_1 \times \cdots \times B_{n-1})$$

per ogni $n \in \mathbb{N}$, $(t_1, ..., t_n) \in \mathcal{S}$, $B_1, ..., B_n \in \mathcal{E}$.

Definizione 1 *Sia $\{\mu_\tau; \tau \in \mathcal{S}\}$ una famiglia di misure di probabilità (sempre sottintendendo che se $\tau = \{t_1, ..., t_n\} \in \mathcal{S}$, μ_τ sia una misura di probabilità su $(E^n, \mathcal{E}^n)$). Quando valgono (1.2)-(1.3) per ogni scelta di $n \in \mathbb{N}$, $\tau = (t_1, ..., t_n) \in \mathcal{S}$, $i = 1, ..., n$, diciamo che la famiglia $\{\mu_\tau; \tau \in \mathcal{S}\}$ è consistente.*

Ricordiamo che uno spazio metrico E si dice *σ-compatto* se è unione numerabile di compatti di E. Un risultato di teoria della misura dice che su uno spazio metrico E si dice σ-compatto, se ρ è una misura di probabilità definita sui boreliani $\mathcal{B}(E)$, allora per ogni $B \in \mathcal{B}(E)$ ed $\varepsilon > 0$ esiste un compatto $K \subset B$ tale che $\rho(B \backslash K) < \varepsilon$. Useremo questo risultato nella dimostrazione del seguente teorema.

Teorema 1 *Se $\{\mu_\tau; \tau \in \mathcal{S}\}$ è una famiglia consistente, E è uno spazio metrico σ-compatto, $\mathcal{E} = \mathcal{B}(E)$, allora esiste una ed una sola misura di probabilità μ su $\left(E^T, \mathcal{E}^T\right)$ di cui $\{\mu_\tau; \tau \in \mathcal{S}\}$ sia la famiglia delle distribuzioni di dimensione finita.*

Proof. Passo 1 (preparazione). L'unicità è del tutto analoga a quella della Proposizione 1: due misure con le stesse distribuzioni di dimensione finita coincidono su una classe chiusa per intersezione finita e generante la σ-algebra $\mathcal{E}^T$, quindi coincidono su tutta $\mathcal{E}^T$.

Per l'esistenza, ricordiamo il seguente teorema di Charateodory: dato uno spazio misurabile $(\Omega, \mathcal{G})$ ed un'algebra $\mathcal{A}$ che genera $\mathcal{G}$, se μ è una misura finitamente additiva su $(\Omega, \mathcal{A})$, continua in $\varnothing$ (cioè tale che, se $\{A_n\}$ è una successione di eventi di $\mathcal{A}$ decrescente con intersezione vuota allora $\lim_{n\to\infty} \mu(A_n) = 0$), allora μ si estende univocamente ad una misura numerabilmente additiva su $(\Omega, \mathcal{G})$.

Prendiamo l'algebra $\mathcal{A}$ di tutti gli insiemi cilindrici $C(\pi, B)$, con $n \in \mathbb{N}$, $\tau = \{t_1, ..., t_n\} \in \mathcal{S}$ e $B \in \mathcal{E}^n$. L'algebra $\mathcal{A}$ genera $\mathcal{E}^T$. Basta definire una misura μ su $\mathcal{A}$ con le proprietà del teorema di Charateodory ed avente $\{\mu_\tau; \tau \in \mathcal{S}\}$ come famiglia delle distribuzioni di dimensione finita, ed il teorema è dimostrato.

Preso un insieme cilindrico $C \in \mathcal{A}$, esistono infinite sue rappresentazioni nella forma $C = C(\tau, B)$ con $\tau = (t_1, ..., t_n)$ e $B \in \mathcal{E}^n$. Presa una di tali rappresentazioni, possiamo calcolare $\mu_\tau(B)$ e porre
$$\mu(C) = \mu_\tau(B).$$
Ma la definizione è ben data solo se non dipende dalla rappresentazione di C. Qui interviene l'ipotesi di consistenza della famiglia. Se $C(\tau', B')$ e $C(\tau'', B'')$ sono due rappresentazioni dello stesso insieme cilindrico C, abbiamo $\mu_{\tau'}(B') = \mu_{\tau''}(B'')$. La dimostrazione è elementare ma un po' laboriosa da scrivere, per cui la isoliamo nel Lemma 1.

La verifica che μ, così definita, è finitamente additiva su $\mathcal{A}$, si esegue nel seguente modo: presi degli insiemi disgiunti $C_1, ..., C_k \in \mathcal{A}$, c'è una sequenza $\tau = (t_1, ..., t_n)$ tale che tutti gli insiemi C_i possono essere rappresentati tramite τ, cioè esistono $B_1, ..., B_k \in \mathcal{E}^n$, oltretutto disgiunti, tali che $C_j = C(\tau, B_j)$, $j = 1, ..., k$. Vale $\bigcup_{j=1}^{k} C_j = C\left(\tau, \bigcup_{j=1}^{k} B_j\right)$ per cui
$$\mu\left(\bigcup_{j=1}^{k} C_j\right) = \mu_\tau\left(\bigcup_{j=1}^{k} B_j\right) = \sum_{j=1}^{k} \mu_\tau(B_j) = \sum_{j=1}^{k} \mu(C_j)$$
dove il passaggio intermedio si basa sull'additività di μ_τ su $\mathcal{E}^n$. L'additività su $\mathcal{A}$ si riconduce cioè a quella di un'opportuna distribuzione di dimensione finita.

Passo 2 (continuità della misura). Dobbiamo infine dimostrare la continuità in $\varnothing$. Sia $\{C_n\}$ è una successione di eventi di $\mathcal{A}$ decrescente con intersezione vuota. Dobbiamo dimostrare che $\lim_{n\to\infty} \mu(C_n) = 0$.

Facciamo una piccola digressione che può aiutare a capire la dimostrazione. Ci sono alcune famiglie di successioni $\{C_n\}$ per cui la dimostrazione è facile. Gli esercizi 2 e 3 illustrano esempi in cui ci si può ricondurre ad usare una singola distribuzione di dimensione finita, un po' come nella verifica fatta sopra dell'additività. In questi casi, la dimostrazione che $\lim_{n\to\infty} \mu(C_n) = 0$ è facile. Se ci si potesse restringere ad insiemi cilindrici come quelli descritti agli esercizi 2 e 3, non ci sarebbe bisogno dell'ipotesi di regolarità dello spazio E.

Ma le famiglie di insiemi cilindrici descritte da tali esercizi non sono algebre. Tra le successioni di insiemi cilindrici esistono esempi, come quello dell'esercizio 4, in cui non si vede come ricondursi ad usare una singola distribuzione di dimensione finita.

Facendo però riferimento a concetti topologici, precisamente alla compattezza, c'è un'altra classe in cui la dimostrazione che $\lim_{n\to\infty}\mu(C_n)=0$ si riesce a completare facilmente. Supponiamo che la successione $\{C_n\}$ di eventi di $\mathcal{A}$ decrescente con intersezione vuota abbia la forma $C_n = C(\tau_n, K_n)$ con K_n insieme compatto (per inciso, se E non è compatto, questa rappresentazione è unica, quando esiste). Allora gli insiemi K_n non possono essere tutti diversi dal vuoto. Rimandiamo la verifica al Lemma 2. Ma allora, se per un certo n_0 l'insieme K_{n_0} è vuoto, vale $\mu(C_{n_0})=0$, da cui discende (per monotonia) $\mu(C_n)=0$ per ogni $n\geq n_0$ e quindi anche $\lim_{n\to\infty}\mu(C_n)=0$.

L'idea della dimostrazione allora è la seguente: data una rappresentazione $C_n = C(\tau_n, B_n)$ degli insiemi cilindrici della successione, fissato $\varepsilon>0$, usando la proprietà di regolarità dello spazio E si può trovare una successione di compatti $\{K_n\}$, $K_n\subset B_n$ tali che, detto D_n l'insieme cilindrico di base K_n (invece che B_n) e coordinate τ_n, insieme che verifica $D_n\subset C_n$, vale

$$\mu(C_n)-\mu(D_n)=\mu(C_n\backslash D_n)\leq\varepsilon.$$

Se riusciamo a trovare $\{K_n\}$ in modo che $\{D_n\}$ sia anche decrescente, allora per il Lemma 2, $\lim_{n\to\infty}\mu(D_n)=0$. Questo implica che esiste $n_0\geq 0$ tale che per ogni $n\geq n_0$, $\mu(C_n)\leq\varepsilon$. Per l'arbitrarietà di ε si ottiene $\lim_{n\to\infty}\mu(C_n)=0$.

L'unico punto che richiede un attimo di lavoro è fare in modo che $\{D_n\}$ sia decrescente. Sia quindi, fissato $\varepsilon>0$, $\{K_1^0\}$ una successione di compatti, $K_n^0\subset B_n$, tali che $\mu_{\tau_n}(B_n\backslash K_n^0)<\frac{\varepsilon}{2^n}$ (essi esistono per la regolarità di E). Indichiamo con $\{D_n^0\}$ la successione degli insiemi cilindrici di base K_n^0 e coordinate τ_n; $D_n^0\subset C_n$ in quanto $K_n^0\subset B_n$, ma non sappiamo se $\{D_n^0\}$ è decrescente. Poniamo $D_n=D_1^0\cap...\cap D_n^0$. Sicuramente $\{D_n\}$ è una successione di insiemi cilindrici decrescente. Mostriamo che esistono dei compatti $K_n\subset B_n$, tali che D_n ha base K_n e coordinate τ_n; e $\mu(C_n\backslash D_n)\leq\varepsilon$.

Gli insiemi D_n hanno la forma

$$D_n=\left\{f|_{\tau_1}\in K_1^0,...,f|_{\tau_n}\in K_n^0\right\}.$$

Si immagini l'esempio $D_2=\left\{f|_{(t_1,t_2)}\in K_1^0, f|_{(t_2,t_3)}\in K_2^0\right\}$. Si può descrivere nella forma

$$\begin{aligned}D_2&=\left\{f|_{(t_1,t_2,t_3)}\in K_1^0\times E, f|_{(t_1,t_2,t_3)}\in E\times K_2^0\right\}\\&=\left\{f|_{(t_1,t_2,t_3)}\in\left(K_1^0\times E\right)\cap\left(E\times K_2^0\right)\right\}\end{aligned}$$

e l'insieme $(K_1^0\times E)\cap(E\times K_2^0)$ è compatto. Il caso generale si scrive con fatica ma è identico. Quindi esiste $K_n\subset B_n$, tali che D_n ha base K_n e coordinate τ_n.

Vale poi (si osservi che $D_k^0\subset C_k\subset C_n$; si scriva inoltre $C_n\backslash D_n=C_n\cap(D_1^0\cap...\cap D_n^0)^c$)

$$\begin{aligned}C_n\backslash D_n&=\left(C_n\backslash D_1^0\right)\cup\left(C_n\backslash D_2^0\right)\cup...\cup\left(C_n\backslash D_n^0\right)\\&\subset\left(C_1\backslash D_1^0\right)\cup\left(C_2\backslash D_2^0\right)\cup...\cup\left(C_n\backslash D_n^0\right)\end{aligned}$$

da cui

$$\mu\left(C_n \backslash D_n\right) \leq \sum_{k=1}^{n} \mu\left(C_k \backslash D_k^0\right) \leq \sum_{k=1}^{n} \frac{\varepsilon}{2^k} \leq \varepsilon.$$

La dimostrazione è completa. ■

Nella dimostrazione del seguente lemma usiamo la notazione $|\tau|$ per la cardinalità di $\tau \in \mathcal{S}$. Se $\tau = (t_1, ..., t_n)$, vale $|\tau| = n$.

Lemma 1 *Se $C(\tau', B')$ e $C(\tau'', B'')$ sono due rappresentazioni dello stesso insieme cilindrico C, abbiamo $\mu_{\tau'}(B') = \mu_{\tau''}(B'')$.*

Proof. Se vale $\tau' = \tau''$, si può riconoscere che vale anche $B' = B''$. In questo caso la tesi è ovvia. Se τ'' è ottenuta da τ' tramite una permutazione degli indici $(i_1, ..., i_n)$, allora $B'' = P_{(i_1,...,i_n)}(B')$ e l'invarianza della definizione di μ è garantita dalla proprietà (1.2).

Se, cosiderando τ' e τ'' come insiemi non oridinati, vale $\tau' \subset \tau''$, a meno di permutazione delle coordinate risulta B'' è della forma $B' \times E^{|\tau''|-|\tau'|}$. Quando $|\tau''| - |\tau'| = 1$ basta applicare la proprietà (1.3); quando $|\tau''| - |\tau'| > 1$ si agisce in $|\tau''| - |\tau'|$ passi sempre con la proprietà (1.3).

Se τ' e τ'', cosiderate come insiemi non oridinati, non sono contenute una nell'altra, si consideri $\tau = \tau' \cap \tau''$. Esiste B tale che $C(\tau, B)$ è una terza rappresentazione; B è la proiezione lungo le coordinate τ di B' o di B''. Per capire che è così, si pensi al caso $\tau' = (t_1, t_2)$, $\tau'' = (t_2, t_3)$ (il caso generale è solo notazionalmente più faticoso): vale

$$\left\{f \in E^T : (f(t_1), f(t_2)) \in B'\right\} = \left\{f \in E^T : (f(t_2), f(t_3)) \in B''\right\}$$

ovvero

$$\{(f(t_1), f(t_2), f(t_3)) \in B' \times E\} = \{(f(t_1), f(t_2), f(t_3)) \in E \times B''\}.$$

Questo implica che gli insiemi di E^3 dati da $B' \times E$ e $E \times B''$ coincidono. Questo è compatibile solo con la struttura $E \times B \times E$.

Vale allora $C(\tau', B') = C(\tau, B)$ ma $\tau \subset \tau'$, quindi $\mu_{\tau'}(B') = \mu_{\tau}(B)$. Lo stesso si può dire per $C(\tau'', B'')$ e quindi $\mu_{\tau'}(B') = \mu_{\tau''}(B'')$. La dimostrazione è completa. ■

Lemma 2 *Sia $\{C_n\} \subset \mathcal{A}$ decrescente con intersezione vuota, della forma $C_n = C(\tau_n, K_n)$ con K_n insieme compatto. Allora gli insiemi K_n non possono essere tutti diversi dal vuoto.*

Esercizio 2 *Sia $\{t_n\} \subset T$ una successione data e sia $\{C_n\} \subset \mathcal{A}$ della forma*

$$C_n = \{f(t_1) \in B_{1,n}, ..., f(t_n) \in B_{n,n}\}$$

dove la famiglia a due indici interi positivi $\{B_{k,n}\} \subset \mathcal{E}$ soddisfa

$$B_{k,n+1} \subset B_{k,n}$$

per ogni $k, n \in \mathbb{N}$. Quindi $\{C_n\}$ è decrescente. Mostrare in questo caso che, se $\{C_n\}$ ha intersezione vuota, allora $\lim_{n\to\infty} \mu(C_n) = 0$.

Esercizio 3 *Sia* $\{t_n\} \subset T$ *una successione data e sia* $\{C_n\} \subset \mathcal{A}$ *decrescente. Supponiamo che esista* $k_0 \in \mathbb{N}$ *tale che, per* $n \geq k_0$, $\{\pi_{k_0} C_n\} \subset \mathcal{E}$ *sia una successione decrescente con intersezione vuota. Qui* $\pi_{k_0} : E^T \to E$ *è la proiezione* $f \longmapsto \pi_{k_0}(f) = f(t_{k_0})$. *Allora* $\{C_n\}$ *ha intersezione vuota e* $\lim_{n\to\infty} \mu(C_n) = 0$.

Esercizio 4 *Sia* $\{t_n\} \subset T$ *una successione data. Verificare che gli insiemi*

$$C_n = \left\{ f(t_{k-1}) < f(t_k) < f(t_{k-1}) + \frac{1}{n}, k = 2, 3, ..., n \right\}$$

non rientrano nei casi trattati dagli esercizi precedenti ma formano una successione decrescente con intersezione vuota.

Esercizio 5 *Data una misura di probabilità* λ *sui boreliani di uno spazio metrico* (X, d), *diciamo che essa è tight se per ogni* $\varepsilon > 0$ *esiste un compatto* K_ε *tale che* $\lambda(K_\varepsilon) > 1 - \varepsilon$. *Mostrare che il teorema di costruzione dei processi di Kolmogorov continua a valere se, invece di supporre che lo spazio metrico* E *sia* σ*-compatto, si suppone che sia metrico e che ogni distribuzione di dimensione finita* μ_τ, $\tau \in \mathcal{S}$, *sia tight. [Presa* $\tau \in \mathcal{S}$ *e la corrispondente misura* μ_τ *su* $\mathcal{E}^n$, *esiste un boreliano* $X_n \subset E^n$ *che è uno spazio metrico* σ*-compatto e* μ_τ *può essere ristretta ad una misura di probabilità su* X_n. *Il resto della dimostrazione del teorema di Kolmogorov è inalterata.]*

Osservazione 3 *Ogni misura di probabilità* λ *sui boreliani uno spazio metrico completo e separabile (polacco) è tight. La proprietà di essere polacco passa al prodotto cartesiano finito. Allora il teorema di Kolmogorov vale se, invece di supporre che lo spazio metrico* E *sia* σ*-compatto, si suppone che sia polacco.*

1.2.1 Processi gaussiani

Ricordiamo che la densità gaussiana standard è la funzione $f(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$ e la densità gaussiana $N(\mu, \sigma^2)$ è la funzione

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Inoltre, un vettore aleatorio $Z = (Z_1, ..., Z_n)$ è *gaussiano standard* se ha densità congiunta

$$f(x_1, ..., x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_i^2}{2}\right) = (2\pi)^{-n/2} \exp\left(-\frac{x_1^2 + ... + x_n^2}{2}\right)$$

(questo equivale a chiedere che le componenti $Z_1, ..., Z_n$ siano gaussiane standard indipendenti), mentre un vettore $Y = (Y_1, ..., Y_m)$ si dice *gaussiano* se si può rappresentare nella forma $Y = AZ + b$ con A matrice $n \times m$, Z vettore gaussiano standard,

$b \in \mathbb{R}^m$. Tra le equivalenze con coi si può riscrivere questa definizione, ricordiamo la seguente: un vettore $Y = (Y_1, ..., Y_m)$ è gaussiano se e solo se la v.a. $\sum_{i=1}^m \lambda_i Y_i$ è gaussiana per ogni scelta di $(\lambda_1, ..., \lambda_m) \in \mathbb{R}^m$. Segue subito da queste definizioni che se $Y = (Y_1, ..., Y_m)$ è un vettore gaussiano, B è una matrice $m \times k$ e $c \in \mathbb{R}^k$, allora il vettore aleatorio $BY + c$ è un vettore gaussiano (in $\mathbb{R}^k$).

Ricordiamo inoltre che per *matrice di covarianza* di un vettore $Y = (Y_1, ..., Y_m)$ si intende la matrice $Q \in \mathbb{R}^{m \times m}$ definita da

$$Q_{ij} = Cov\,(Y_i, Y_j)\,, \qquad i, j = 1, ..., m.$$

E' simmetrica e semi-definita positiva. La media (o vettore dei valori medi) è il vettore di coordinate $E\,[Y_i]$, $i = 1, ..., m$. Un vettore gaussiano standard $Z = (Z_1, ..., Z_n)$ ha media nulla e covarianza pari all'identità di $\mathbb{R}^n$. Un vettore gaussiano della forma $Y = AZ + b$ come sopra, ha media b e matrice di covarianza $Q = AA^T$. Più in generale, ricordiamo la seguente proposizione:

Proposizione 3 *Se $Y = (Y_1, ..., Y_m)$ è un vettore gaussiano di media μ_Y e covarianza Q_Y, B è una matrice $m \times k$ e $c \in \mathbb{R}^k$, allora il vettore aleatorio $BY + c$ è un vettore gaussiano di media $B\mu_Y + c$ e covarianza*

$$Q = AQ_Y A^T.$$

Sia $Y = (Y_1, ..., Y_m)$ un vettore gaussiano di media μ e covarianza Q. Quando $\det Q \neq 0$, Y ha densità di probabilità congiunta (la sua legge è assolutamente continua rispetto alla misura di Lebesgue di $\mathbb{R}^m$), data da

$$f\,(x) = \frac{1}{\sqrt{(2\pi)^n \det Q}} \exp\left(-\frac{\langle Q^{-1}\,(x-\mu)\,, (x-\mu)\rangle}{2}\right)$$

dove $x = (x_1, ..., x_n)$. Altrimenti, se $\det Q = 0$, la legge di Y è singolare rispetto alla misura di Lebesgue di $\mathbb{R}^m$ ed è concentrata su un sottospazio proprio (precisamente una varietà affine, di codimensione maggiore di zero).

Chiameremo gaussiana ogni misura di probabilità su $(\mathbb{R}^n, \mathcal{B}\,(\mathbb{R}^n))$ che sia legge di una v.a. gaussiana. Equivalentemente, una misura di probabilità ρ su $(\mathbb{R}^n, \mathcal{B}\,(\mathbb{R}^n))$ è gaussiana se la sua legge immagine su $(\mathbb{R}, \mathcal{B}\,(\mathbb{R}))$ attraverso qualsiasi proiezione unidimensionale è una misura con densità gaussiana o una delta di Dirac. Ricordiamo che vale il seguente risultato:

Proposizione 4 *Dati un vettore $b \in \mathbb{R}^n$ ed una matrice $Q \in \mathbb{R}^{n \times n}$ simmetrica e semi-definita positiva, esiste una ed una sola misura gaussiana su $\left(\mathbb{R}^n, \mathcal{B}\,(\mathbb{R})^n\right)$ che ha b come vettore delle medie e Q come matrice di covarianza.*

Fatte queste premesse sui vettori gaussiani, possiamo definire ed analizzare i processi gaussiani. Osserviamo che anche in questo paragrafo l'insieme T è qualsiasi.

Definizione 2 *Un processo a valori reali* $X = (X_t)_{t \in T}$ *si dice gaussiano se tutte le sue marginali di dimensione finita* $(X_{t_1}, ..., X_{t_n})$ *sono vettori gaussiani. Analogamente, una misura di probabilità su* $\left(\mathbb{R}^T, \mathcal{B}(\mathbb{R})^T\right)$ *si dice gaussiana se tutte le sue distribuzioni di dimensione finita sono gaussiane.*

Un processo reale è quindi gaussiano se e solo se la sua legge su $\left(\mathbb{R}^T, \mathcal{B}(\mathbb{R})^T\right)$ è gaussiana.

Se T è un sottoinsieme di uno spazio euclideo, unione numerabile di compatti, diciamo che una misura su $(C(T,E), \mathcal{B}(C(T,E)))$ è gaussiana se tutte le sue distribuzioni di dimensione finita sono gaussiane.

Un processo gaussiano ha la legge caratterizzata da poche funzioni: la funzione valor medio $m(t) = E[X_t]$, $t \in T$ e la funzione di covarianza $C(t,s) = Cov(X_t, X_s)$, $t, s \in T$.

Proposizione 5 *Se due processi gaussiani hanno le stesse funzioni* $m(t)$ *e* $C(t,s)$, *allora hanno la stessa legge.*

Proof. La legge è identificata dalle distribuzioni di dimensione finita. Le leggi dei due processi sono misure gaussiane, con distribuzioni di dimensione finita gaussiane. Tali gaussiane, diciamo in $\mathbb{R}^n$, sono univocamente determinate dai loro vettori medi e dalle matrici di covarianza, che però a loro volta hanno come componenti le valutazioni delle funzioni $m(t)$ e $C(t,s)$ in opportuni punti, quindi coincidono. ■

Ancor più economica è la descrizione nel caso di processi stazionari. Supponiamo che sull'insieme T sia definita un'operazione di somma $+$, cioè $t + s \in T$ se $t, s \in T$. Ad esempio si possono considerare $T = \mathbb{R}^n$ o $T = [0, \infty)$ con l'usuale somma euclidea.

Definizione 3 *Un processo stocastico* $X = (X_t)_{t \in T}$ *a valori in* $(E, \mathcal{E})$ *si dice stazionario in senso stretto se, per ogni* $\tau = (t_1, ..., t_n) \in \mathcal{S}$ *le leggi di* $(X_{t_1}, ..., X_{t_n})$ *e* $(X_{t_1+h}, ..., X_{t_n+h})$ *coincidono per ogni* $h \in T$.

Definizione 4 *Un processo reale* $X = (X_t)_{t \in T}$, *con* $E[X_t^2] < \infty$ *per ogni* $t \in T$, *si dice stazionario in senso lato o debole se*

$$
\begin{aligned}
m(t+h) &= m(t) \\
C(t+h, s+h) &= C(t,s)
\end{aligned}
$$

per ogni $h, s, t \in T$.

Nel caso di un processo reale, la stazionarietà in senso stretto implica quella in senso lato, ma non viceversa (non possiamo risalire alle leggi dai momenti di ordine uno e due).

Proposizione 6 *Se un processo gaussiano è stazionario in senso lato allora è anche stazionario in senso stretto.*

Proof. Dati $\tau = (t_1, ..., t_n) \in \mathcal{S}$ e $h \in T$, le leggi di $(X_{t_1}, ..., X_{t_n})$ e $(X_{t_1+h}, ..., X_{t_n+h})$, essendo gaussiane, sono identificate dai vettori medi di componenti $E[X_{t_k}]$ e $E[X_{t_k+h}]$, che coincidono essendo $m(t_k + h) = m(t_k)$, $k = 1, ..., n$, e dalle matrici di covarianza di componenti $Cov\left(X_{t_i}, X_{t_j}\right)$ e $Cov\left(X_{t_i+h}, X_{t_j+h}\right)$, che coincidono essendo $C(t_i + h, t_j + h) = C(t_i, t_j)$. Quindi le leggi di $(X_{t_1}, ..., X_{t_n})$ e $(X_{t_1+h}, ..., X_{t_n+h})$ coincidono ed abbiamo la stazionarietà in senso stretto. ∎

Supponiamo che T sia un gruppo rispetto alla somma $+$ e sia 0 l'elemento neutro. In questo caso la stazionarietà in senso lato permette di descrivere la legge del processo gaussiano in modo estremamente economico.

Proposizione 7 *Se un processo gaussiano è stazionario in senso lato allora la sua legge è identificata dal numero $m := E[X_t]$ e dalla funzione di una variabile*

$$C(t) := Cov(X_t, X_0), \qquad t \in T.$$

Proof. Le distribuzioni di dimensione finita sono identificate dalle funzioni $m(t)$ e $C(t, s)$ ma queste, a loro volta, per la stazionarietà in senso lato sono l'una costante, $m(t) = m$, l'altra identificata dai suoi valori nei punti (t, s) della forma $(r, 0)$, in quanto

$$C(t - s, 0) = C(t, s)$$

($h = s$ nella definizione di stazionarietà). ∎

Concludiamo con un risultato di esistenza.

Proposizione 8 *Date due funzioni $m(t)$ e $C(t, s)$, $t, s \in T$, se $C(t, s) = C(s, t)$ e vale*

$$\sum_{i,j=1}^{n} C(t_i, t_j)\, \xi_i \xi_j \geq 0$$

per ogni $n \in \mathbb{N}$, $(t_1, ..., t_n) \in T^n$, $(\xi_1, ..., \xi_n) \in \mathbb{R}^n$, allora esiste un processo gaussiano che ha queste funzioni come media e covarianza.

Proof. Basta costruire una misura gaussiana su $\left(\mathbb{R}^T, \mathcal{B}(\mathbb{R})^T\right)$ e prendere il processo canonico. Per il teorema di costruzione di Kolmogorov ($(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ soddisfa l'ipotesi del teorema), basta costruire una famiglia consistente $\{\mu_\tau; \tau \in \mathcal{S}\}$ di distribuzioni di dimensione finita, che siano gaussiane (quindi la misura ed il processo saranno gaussiani) e tali che valga la seguente proprietà: presa $\tau = (t_1, ..., t_n) \in \mathcal{S}$, se $(X_{t_1}, ..., X_{t_n})$ è un vettore aleatorio di legge μ_τ, quindi gaussiano, valga $E[X_{t_k}] = m(t_k)$ e $Cov\left(X_{t_i}, X_{t_j}\right) = C(t_i, t_j)$ per ogni $k, i, j = 1, ..., n$.

Dato $\tau = (t_1, ..., t_n) \in \mathcal{S}$, sia μ_τ la misura gaussiana su $\mathbb{R}^n$ avente vettore medio di componenti $m(t_k)$ e matrice di covarianza di componenti $C(t_i, t_j)$, per ogni $k, i, j = 1, ..., n$. Un tale misura esiste ed è unica. Infatti la matrice di componenti $C(t_i, t_j)$ è semidefinita positiva, per ipotesi, ed abbiamo ricordato sopra che un vettore ed una matrice semidefinita positiva definiscono univocamente una misura gaussiana. La validità della proprietà detta poco sopra ($E[X_{t_k}] = m(t_k)$ e $Cov(X_{t_i}, X_{t_j}) = C(t_i, t_j)$, se $(X_{t_1}, ..., X_{t_n})$ ha legge μ_τ) è assicurata per definizione. Resta da verificare la consistenza. Omettiamo, per non appesantire la trattazione, la verifica della proprietà (1.2) e limitiamoci alla (1.3).

Con le notazioni usate in precedenza, $\tau = (t_1, ..., t_n) \in \mathcal{S}$, $\tau_{\widehat{n}} = (t_1, ..., t_{n-1})$, $\pi_{\widehat{n}} : E^n \to E^{n-1}$ che manda la generica sequenza $(x_1, ..., x_n) \in E^n$ nella sequenza $(x_1, ..., x_{n-1}) \in E^{n-1}$, dobbiamo dimostrare che

$$\mu_{\tau_{\widehat{n}}} = \pi_{\widehat{n}}(\mu_\tau).$$

La trasformazione $\pi_{\widehat{n}}$ è lineare e quindi i vettori delle medie $m_{\tau_{\widehat{n}}}$ e m_τ di $\mu_{\tau_{\widehat{n}}}$ e μ_τ rispettivamente sono legati dalla relazione $m_{\tau_{\widehat{n}}} = \pi_{\widehat{n}} m_\tau$, le matrici di covarianza $Q_{\tau_{\widehat{n}}}$ e Q_τ di $\mu_{\tau_{\widehat{n}}}$ e μ_τ rispettivamente sono legate dalla relazione $Q_{\tau_{\widehat{n}}} = \pi_{\widehat{n}} Q_\tau \pi_{\widehat{n}}^T$ (usando la notazione $\pi_{\widehat{n}}$ anche per la matrice associata alla trasformazione nella base canonica). Il vettore m_τ ha componenti $m(t_k)$, $k = 1, ..., n$, quindi $\pi_{\widehat{n}} m_\tau$ è il vettore $(m(t_1), ..., m(t_{n-1})) \in E^{n-1}$, che è proprio il vettore delle medie di $\mu_{\tau_{\widehat{n}}}$.

La verifica della proprietà $Q_{\tau_{\widehat{n}}} = \pi_{\widehat{n}} Q_\tau \pi_{\widehat{n}}^T$ è elementare ma noiosa da scrivere. Per completezza la riportiamo. La matrice $\pi_{\widehat{n}}$ ha componenti $(\pi_{\widehat{n}})_{j,\alpha} = \delta_{j,\alpha}$, per $j = 0, ..., n-1$, $\alpha = 1, ..., n$. Quindi la matrice $\pi_{n,i}^T$ ha componenti $\left(\pi_{\widehat{n}}^T\right)_{\beta,k} = (\pi_{\widehat{n}})_{k,\beta} = \delta_{k,\beta}$, per $k = 0, ..., n-1$, $\beta = 1, ..., n$. Quindi, dall'identità

$$\left(\pi_{\widehat{n}} Q_\tau \pi_{\widehat{n}}^T\right)_{jk} = \sum_{\alpha,\beta=1}^{n} (\pi_{\widehat{n}})_{j,\alpha} (Q_\tau)_{\alpha,\beta} \left(\pi_{\widehat{n}}^T\right)_{\beta,k}$$

si deduce, per $j, k = 0, ..., n-1$,

$$= \sum_{\alpha,\beta=1}^{n} \delta_{j,\alpha} C(t_\alpha, t_\beta) \delta_{k,\beta} = C(t_j, t_k).$$

Questa è la matrice $Q_{\tau_{\widehat{n}}}$. La dimostrazione è completa. ■

Esercizio 6 *Costruire un processo X con $T = [0, 1]$ che si annulli in $t = 0$ e $t = 1$ q.c., ed invece X_t abbia densità di probabilità strettamente positiva per ogni $t \in (0, 1)$ (un "ponte" stocastico).*

1.2.2 Filtrazioni

A partire da questo paragrafo supponiamo che T sia un intervallo di $\mathbb{R}$, o più precisamente per fissare le idee

$$T = [0, \infty).$$

Si intuisce che si possa svolgere una teoria più generale ma gli esempi che tratteremo nel corso non la motivano.

Chiamiamo *filtrazione* su $(\Omega, \mathcal{F}, P)$ una famiglia $(\mathcal{F}_t)_{t \in T}$ di σ-algebre di insiemi di Ω, $\mathcal{F}_t \subset \mathcal{F}$ per ogni $t \in T$, che sia crescente: $\mathcal{F}_s \subset \mathcal{F}_t$ se $s < t \in T$.

Un processo $X = (X_t)_{t \in T}$ definito su $(\Omega, \mathcal{F}, P)$ a valori in $(E, \mathcal{E})$ si dice *adattato* alla filtrazione $(\mathcal{F}_t)_{t \in T}$ se per ogni $t \in T$ la funzione X_t è misurabile da $(\Omega, \mathcal{F}_t)$ in $(E, \mathcal{E})$.

Supponiamo $T = [0, \infty)$; se T è un intervallo che contiene 0 la definizione è analoga. Il processo X si dice *progressivamente misurabile* se per ogni $t \geq 0$ l'applicazione $(s, \omega) \longmapsto X_s(\omega)$ da $([0, t] \times \Omega, \mathcal{B}([0, t]) \otimes \mathcal{F}_t)$ in $(E, \mathcal{E})$ è misurabile.

Se X è progressivamente misurabile, allora è misurabile ed adattato. Viceversa, vale ad esempio il seguente risultato.

Proposizione 9 *Sia E uno spazio topologico, $\mathcal{E} = \mathcal{B}(E)$. Se X è adattato e q.c. continuo a destra (oppure q.c. continuo a sinistra) allora è progressivamente misurabile.*

Proof. Sia $\mathcal{G}$ una σ-algebra su Ω. Il seguente criterio di misurabilità è noto: se una funzione $f : [a, b] \times \Omega \to E$ è continua a destra ed $\omega \longmapsto f(t, \omega)$ è misurabile da $(\Omega, \mathcal{G})$ in $(E, \mathcal{E})$ per ogni $t \in [a, b]$, allora $(t, \omega) \longmapsto f(t, \omega)$ è misurabile da $([a, b] \times \Omega, \mathcal{B}([a, b]) \otimes \mathcal{G})$ in $(E, \mathcal{E})$. La dimostrazione si fa ad esempio approssimando f con funzioni continue a destra e costanti a tratti in t.

Basta allora applicare questo criterio ad ogni restrizione di X ad insiemi della forma $[0, t] \times \Omega$. La dimostrazione è completa. ■

Un insieme $N \subset \Omega$ è trascurabile rispetto a $(\Omega, \mathcal{F}, P)$ se $P^*(N) = 0$ dove $P^*(A) = \inf\{P(B); B \in \mathcal{F}, A \subset B\}$. Quindi $N \subset \Omega$ è trascurabile se $\inf\{P(B); B \in \mathcal{F}, N \subset B\} = 0$. Indichiamo con $\mathcal{N}$ l'insieme degli insiemi trascurabili rispetto a $(\Omega, \mathcal{F}, P)$. Una filtrazione $(\mathcal{F}_t)_{t \in T}$ è *completa* se ogni $\mathcal{F}_t$ contiene $\mathcal{N}$. E' equivalente che $\mathcal{F}_0$ contenga $\mathcal{N}$. [Una σ-algebra $\mathcal{G}$ si dice completa quando contiene gli insiemi trascurabili rispetto a $(\Omega, \mathcal{G}, P)$. Quindi il chiedere che $\mathcal{F}_0$ contenga $\mathcal{N}$ - famiglia degli insiemi trascurabili rispetto a $(\Omega, \mathcal{F}, P)$ - o che sia completa sono affermazioni differenti.]

E' comodo che tutte le σ-algebre di una filtrazione contengano $\mathcal{N}$. Altrimenti si creano tante piccole complicazioni un po' innaturali; ad esempio se Y è modificazione di un processo adattato X e la filtrazione non è completa, non si può concludere che anche Y sia adattato. Infatti, preso $B \in \mathcal{E}$ e $t \in T$, l'evento $\{Y_t \in B\}$ può differire da $\{X_t \in B\}$ per un insieme di $\mathcal{N}$, ma tale insieme potrebbe non appartenere a $\mathcal{F}_t$, quindi la proprietà $\{X_t \in B\} \in \mathcal{F}_t$ può non implicare $\{Y_t \in B\} \in \mathcal{F}_t$. Invece, vale:

Osservazione 4 *Se $(\mathcal{F}_t)_{t \in T}$ è completa, X è adattato ed Y è una modificazione di X, allora Y è adattato.*

Una filtrazione $(\mathcal{F}_t)_{t\in T}$ si dice *continua a destra* se per ogni $t \in T$

$$\mathcal{F}_t = \bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon}.$$

Questa condizione interviene ogni tanto nei teoremi successivi, e la completezza ancora di più, per cui spesso per unificare gli enunciati di una teoria si assume sin dall'inizio che la filtrazione di riferimento della teoria sia completa a continua a destra. Diremo che una filtrazione *soddisfa le condizioni abituali* se è completa e continua a destra.

Si ricordi che l'intersezione (arbitraria) di σ-algebre è una σ-algebra, quindi $\bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon}$ è sempre una σ-algebra. Invece l'unione no; indicheremo col simbolo $\mathcal{F}\vee\mathcal{G}$ la più piccola σ-algebra che contiene $\mathcal{F}\cup\mathcal{G}$; per cui scriveremo ad esempio $\bigvee_{t\geq 0} \mathcal{F}_t$ per la più piccola σ-algebra che contiene ogni $\mathcal{F}_t$, denotata con $\mathcal{F}_\infty$.

Dato un processo stocastico X, ad esso è associata la *filtrazione generata da* X definita da

$$\mathcal{F}_t^{00} = \sigma\left\{X_s; s \in T, s \leq t\right\}$$

A livello interpretativo, gli eventi di $\mathcal{F}_t^{00}$ sono gli eventi conoscibili al tempo t se osserviamo il processo X. La notazione $\mathcal{F}_t^{00}$ non è universale ed è usata qui solo per distinguere questa filtrazione dalle seguenti.

Essendo comodo che la filtrazione di riferimento sia completa, si introduce la filtrazione $(\mathcal{F}_t^0)_{t\in T}$ definita da $\mathcal{F}_t^0 = \sigma\left\{\mathcal{F}_t^{00} \cup \mathcal{N}\right\}$. La filtrazione $(\mathcal{F}_t^0)_{t\in T}$ è il completamento della filtrazione $(\mathcal{F}_t^{00})_{t\in T}$ (naturalmente questo procedimento si può applicare a qualsiasi filtrazione).

Volendo richiedere che la filtrazione sia anche continua a destra, poniamo

$$\mathcal{F}_t = \bigcap_{\varepsilon>0} \mathcal{F}^0_{t+\varepsilon}.$$

Osservazione 5 *Questa filtrazione è continua a destra:* $\mathcal{F}_t = \bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon}$. *Infatti,* $\mathcal{F}_t^0 \subset \mathcal{F}_t$ *(per ogni* $t \in T$*) per monotonia di* $(\mathcal{F}_t^0)_{t\in T}$ *e definizione di* $\mathcal{F}_t$*, quindi* $\bigcap_{\varepsilon>0} \mathcal{F}^0_{t+\varepsilon} \subset \bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon}$*, per cui* $\mathcal{F}_t \subset \bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon}$ *per definizione di* $\mathcal{F}_t$*. Viceversa,* $\mathcal{F}_{t+\varepsilon} \subset \mathcal{F}^0_{t+2\varepsilon}$ *per ogni* $t \in T$ *ed* $\varepsilon > 0$*, per definizione di* $\mathcal{F}_t$*, quindi* $\bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon} \subset \bigcap_{\varepsilon>0} \mathcal{F}^0_{t+2\varepsilon} = \mathcal{F}_t$.

La filtrazione così costruita è la più piccola che soddisfi le condizioni abituali e rispetto a cui il processo sia adattato.

1.3 Speranza condizionale e probabilità condizionale

1.3.1 Speranza condizionale

Teorema 2 *Data una v.a. X a valori reali, integrabile su $(\Omega, \mathcal{F}, P)$ ed una σ-algebra $\mathcal{G} \subset \mathcal{F}$, esiste una v.a. $\mathcal{G}$-misurabile X' tale che*

$$\int_B X dP = \int_B X' dP$$

per ogni $B \in \mathcal{G}$. Inoltre è unica a meno di P-equivalenze.

La dimostrazione dell'esistenza si basa sul teorema di Radon-Nikodym, l'unicità su un semplice argomento prendendo l'evento $B = \{X' > X''\}$, se X', X'' soddisfano le stesse condizioni.

Definizione 5 *Sia X una v.a. integrabile su $(\Omega, \mathcal{F}, P)$ e sia $\mathcal{G}$ una σ-algebra, $\mathcal{G} \subset \mathcal{F}$. Chiamiamo speranza condizionale di X rispetto a $\mathcal{G}$ ogni variabile aleatoria $\mathcal{G}$-misurabile X' tale che*

$$\int_B X dP = \int_B X' dP$$

per ogni $B \in \mathcal{G}$. Chiameremo con lo stesso nome anche la classe di P-equivalenza di tali variabili. La speranza condizionale di X rispetto a $\mathcal{G}$ viene indicata con $E[X|\mathcal{G}]$.

Quando scriveremo uguaglianze tra diverse speranze condizonali o tra una speranza condizionale ed una v.a., si intenderà sempre l'uguaglianza come classi di equivalenza, o P-q.c.

L'intuizione è che, avendo a disposizione il grado di informazione fornito da $\mathcal{G}$, la nostra attesa circa il valore di X è più precisa della semplice $E[X]$ (attesa incondizionata), dipende da ciò che si avvera nei limiti della finezza di $\mathcal{G}$, quindi è una v.a. $\mathcal{G}$-misurabile. Inoltre, se B è un atomo di $\mathcal{G}$ con $P(B) > 0$, per cui X' deve essere costante su B, l'identità della definizione dice che

$$X'|_B = \frac{1}{P(B)} \int_B X dP$$

cioè X' è una sorta di media locale di X; per un B generale l'identità stabilisce una generalizzazione di tale proprietà. Si risolva il seguente esercizio, per aiutare ulteriormente l'intuizione.

Esercizio 7 *Sia $\mathcal{G}$ generata da una partizione misurabile $\{B_1, ..., B_n\}$. Allora $E[X|\mathcal{G}] = \sum_{i=1}^n \left(\frac{1}{P(B_i)} \int_{B_i} X dP \right) \cdot 1_{B_i}$. In altre parole, $E[X|\mathcal{G}]$ è costante su ciascun B_i e lì vale la media di X su B_i, $\frac{1}{P(B_i)} \int_{B_i} X dP$.*

Proof. La v.a. $X' = \sum_{i=1}^{n} \left(\frac{1}{P(B_i)} \int_{B_i} X dP \right) \cdot 1_{B_i}$ è $\mathcal{G}$-misurabile. Prendiamo $Y = 1_{B_1}$. Vale

$$
\begin{aligned}
E[XY] &= \int_{B_1} X dP \\
E[X'Y] &= \sum_{i=1}^{n} \left(\frac{1}{P(B_i)} \int_{B_i} X dP \right) E[1_{B_i} 1_{B_1}] \\
&= \sum_{i=1}^{n} \left(\frac{1}{P(B_i)} \int_{B_i} X dP \right) \delta_{i1} P(B_1) = \int_{B_1} X dP
\end{aligned}
$$

e quindi sono uguali. ■

Osservazione 6 *La definizione data sopra equivale a chiedere che X' sia $\mathcal{G}$-misurabile e valga*

$$E[XY] = E[X'Y]$$

per ogni v.a. Y limitata $\mathcal{G}$-misurabile. Un'implicazione è ovvia (prendendo Y della forma 1_B con $B \in \mathcal{G}$). Per l'altra, dalla definizione, che si riscrive $E[X1_B] = E[X'1_B]$, discende che $E[XY] = E[X'Y]$ per Y della forma $Y = \sum y_i 1_{B_i}$, $y_i \in \mathbb{R}$, $B_i \in \mathcal{G}$. Con variabili di quel tipo possiamo approssimare dal basso puntualmente ogni Y limitata $\mathcal{G}$-misurabile e passare al limite per convergenza monotona.

Proposizione 10 *Siano X, Y, $\{X_n\}$ integrabili, $\mathcal{G}$ una sotto σ-algebra di $\mathcal{F}$. Valgono le seguenti affermazioni:*

i) Se $\mathcal{G}' \subset \mathcal{G} \subset \mathcal{F}$ allora $E[E[X|\mathcal{G}]|\mathcal{G}'] = E[X|\mathcal{G}']$; in particolare ($\mathcal{G}' = \{\varnothing, \Omega\}$), $E[E[X|\mathcal{G}]] = E[X]$

ii) Se X è $\mathcal{G}$-misurabile ed XY è integrabile, allora $E[XY|\mathcal{G}] = XE[Y|\mathcal{G}]$

iii) Se X è indipendente da $\mathcal{G}$ allora $E[X|\mathcal{G}] = E[X]$

iv) $E[aX + bY + c|\mathcal{G}] = aE[X|\mathcal{G}] + bE[Y|\mathcal{G}] + c$

v) Se $\{X_n\}$ è una successione di v.a. monotona non decrescente, con $X = \lim_{n\to\infty} X_n$ integrabile, allora $E[X_n|\mathcal{G}] \to E[X|\mathcal{G}]$ q.c.

La verifica di queste proprietà è un utile esercizio; rimandiamo comunque ai corsi di base di Probabilità. Utile tecnicamente è la seguente generalizzazione delle proprietà (ii)-(iii). Si noti che φ è $\mathcal{G}'$-misurabile nel suo secondo argomento, X è $\mathcal{G}$-misurabile, e $\mathcal{G}$, $\mathcal{G}'$ sono σ-algebre indipendenti.

Proposizione 11 *Dato uno spazio probabilizzato $(\Omega, \mathcal{F}, P)$ ed uno spazio misurabile $(E, \mathcal{E})$, siano $\mathcal{G} \subset \mathcal{F}$ e $\mathcal{G}' \subset \mathcal{F}$ due σ-algebre indipendenti. Sia $\varphi : (E \times \Omega, \mathcal{E} \otimes \mathcal{G}') \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ misurabile limitata e sia $X : (\Omega, \mathcal{G}) \to (E, \mathcal{E})$ misurabile. Allora*

$$E[\varphi(X, \cdot)|\mathcal{G}] = \Phi(X)$$

dove Φ è definita da

$$\Phi(x) := E[\varphi(x, \cdot)], \qquad x \in E.$$

Proof. Supponiamo φ a variabili separate, $\varphi(x,\omega) = \varphi_1(x)\varphi_2(\omega)$, con $\varphi_1 : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $\varphi_2 : (\Omega, \mathcal{G}') \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ misurabili limitate. Allora

$$\begin{aligned} E[\varphi(X,\cdot)|\mathcal{G}] &= E[\varphi_1(X)\varphi_2(\cdot)|\mathcal{G}] = \varphi_1(X)E[\varphi_2(\cdot)] \\ \Phi(X) &= E[\varphi(x,\cdot)]_{x=X} = E[\varphi_1(x)\varphi_2(\cdot)]_{x=X} = \varphi_1(X)E[\varphi_2(\cdot)] \end{aligned}$$

quindi la formula è verificata. Per linearità, vale per combinazioni lineari di funzioni φ della forma $\varphi(x,\omega) = \varphi_1(x)\varphi_2(\omega)$. Si passa al caso generale per convergenza monotona, usando la stabilità della speranza condizionale rispetto a tale convergenza. ■

Apparentemente potrebbe sembrare che, rimuovendo l'ipotesi che $\mathcal{G}'$ sia indipendente da $\mathcal{G}$, ovvero prendendo una qualsiasi funzione $\varphi : (E \times \Omega, \mathcal{E} \otimes \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ misurabile limitata, valga l'identità

$$E[\varphi(X,\cdot)|\mathcal{G}] = E[\varphi(x,\cdot)|\mathcal{G}]|_{x=X}$$

di cui quella della proposizione è un caso particolare. Qui però si pone un problema di versioni: per ogni x la speranza condizionale $E[\varphi(x,\cdot)|\mathcal{G}]$ è definita a meno di insiemi di misura nulla e quindi la sostituzione $E[\varphi(x,\cdot)|\mathcal{G}]|_{x=X}$ non ha un senso ovvio.

Tra i risultati rilevanti citiamo anche il seguente. Se X è di quadrato integrabile, $E[X|\mathcal{G}]$ è la proiezione ortogonale di X sul sottospazio chiuso $L^2(\Omega, \mathcal{G}, P)$ di $L^2(\Omega, \mathcal{F}, P)$ (funzioni di quadrato integrabile misurabili rispetto a $\mathcal{G}$ e $\mathcal{F}$ rispettivamente).

1.3.2 Probabilità condizionale

Sia $(\Omega, \mathcal{F}, P)$ uno spazio probabilizzato. Dati due eventi $A, B \in \mathcal{F}$ con $P(B) > 0$, chiamiamo probabilità condizionale di A sapendo B il numero

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

Data una partizione misurabile $\{B_1, ..., B_n\}$ di Ω, possiamo definire in numeri $P(A|B_i)$ per tutti gli $i = 1, ..., n$ tali che $P(B_i) > 0$. Potremmo dire che la famiglia di numeri $\{P(A|B_i)\}$ è la probabilità di A condizionata alla partizione $\{B_1, ..., B_n\}$. In analogia col caso della speranza condizionale, potremmo codificare questa informazione nella funzione

$$\sum_{i=1}^{n} P(A|B_i) \cdot 1_{B_i}$$

(se per un certo i vale $P(B_i) = 0$, la funzione 1_{B_i} è equivalente a quella nulla e quindi possiamo definire $P(A|B_i)$ arbitrariamente). C'è quindi una funzione $\mathcal{G}$-misurabile, $P(A|\mathcal{G}) := \sum_{i=1}^{n} P(A|B_i) \cdot 1_{B_i}$, che racchiude le informazioni utili circa la probabilità condizionale di A rispetto ai vari elementi della partizione $\{B_1, ..., B_n\}$ e quindi delle informazioni contenute nella σ-algebra $\mathcal{G}$.

Tra l'altro, si noti che, sempre nel caso particolare di $\mathcal{G}$ generata da $\{B_1, ..., B_n\}$, vale

$$E[1_A|\mathcal{G}] = P(A|\mathcal{G})$$

in quanto $\int_{B_i} 1_A dP = P(A \cap B_i)$. Inoltre,

$$P(A) = E[P(A|\mathcal{G})]$$

o più in generale

$$P(A \cap B) = \int_B P(A|\mathcal{G})\, dP$$

per ogni $B \in \mathcal{G}$ (si verifichino queste due identità). Queste ultime formule sono una riscrittura compatta dell'utilissima formula di fattorizzazione (o delle probabilità totali)

$$P(A) = \sum_{i=1}^{n} P(A|B_i)\, P(B_i)$$

e sua generalizzazione a $P(A \cap B)$.

Possiamo estendere queste definizioni e proprietà al caso di una σ-algebra $\mathcal{G} \subset \mathcal{F}$ più generale, raggiungendo due livelli di generalizzazione.

Innanzi tutto, data $\mathcal{G} \subset \mathcal{F}$ qualsiasi, poniamo

$$P(A|\mathcal{G}) := E[1_A|\mathcal{G}]$$

detta *probabilità condizionale di A rispetto alla σ-algebra $\mathcal{G}$*. Quindi $P(A|\mathcal{G})$, definita a meno di P-equivalenza (o come classe di equivalenza), è una v.a. $\mathcal{G}$-misurabile tale che

$$\int_B P(A|\mathcal{G})\, dP = \int_B 1_A dP = P(A \cap B)$$

per ogni $B \in \mathcal{G}$. Questa identità, data ora per definizione, è una versione generalizzata della formula di fattorizzazione. La probabilità condizionale di A rispetto ad una σ-algebra $\mathcal{G}$ è definita tramite la formula di fattorizzazione; o in altre parole, è quella v.a. che fa funzionare la formula di fattorizzazione anche nel caso non finito (cioè di una σ-algebra generale invece che generata da una partizione finita). Questo è il primo livello.

C'è poi un secondo livello, più complesso. Nasce dalla seguente domanda naturale. Per ogni $A \in \mathcal{F}$, $P(A|\mathcal{G})$ è una classe di equivalenza, o comunque è definita a meno di insiemi trascurabili. Non ha senso fissare $\omega \in \Omega$ e considerare la funzione $A \longmapsto P(A|\mathcal{G})(\omega)$. Possiamo scegliere un rappresentante $\widetilde{P(A|\mathcal{G})}$ da ciascuna classe di equivalenza in modo che la funzione d'insieme $A \longmapsto \widetilde{P(A|\mathcal{G})}(\omega)$ sia una misura di probabilità? Se prendiamo degli insiemi disgiunti $A_1, A_2, ... \in \mathcal{F}$, nel senso delle classi di equivalenza (oppure quasi certamente per ogni scelta di rappresentanti) vale

$$P\left(\bigcup_i A_i|\mathcal{G}\right) = \sum_i P\left(A_i|\mathcal{G}\right).$$

Ma non possiamo sperare che valga per $\widetilde{P\left(\cdot|\mathcal{G}\right)}(\omega)$, con ω fissato, senza operare una scelta molto oculata e non banale dei rappresentanti.

Definizione 6 *Dati* $(\Omega, \mathcal{F}, P)$ *e* $\mathcal{G} \subset \mathcal{F}$, *chiamiamo versione regolare della probabilità condizionale rispetto a* $\mathcal{G}$ *una funzione* $(A, \omega) \longmapsto P_{\mathcal{G}}(A, \omega)$, *definita su* $\mathcal{F} \times \Omega$, *con le seguenti proprietà:*

i) per P*-q.o.* ω, *la funzione d'insieme* $A \longmapsto P_{\mathcal{G}}(A, \omega)$ *è una misura di probabilità su* $(\Omega, \mathcal{F})$

ii) per ogni $A \in \mathcal{F}$, *la funzione* $\omega \longmapsto P_{\mathcal{G}}(A, \omega)$ *è misurabile ed appartiene alla classe di equivalenza* $P(A|\mathcal{G})$.

Vale il seguente teorema non banale:

Teorema 3 *Se* (Ω, d) *è uno spazio metrico completo e separabile (spazio polacco) ed* $\mathcal{F} = \mathcal{B}(\Omega)$, *esiste sempre versione regolare della probabilità condizionale rispetto ad ogni* $\mathcal{G} \subset \mathcal{B}(\Omega)$.

Per la versione regolare valgono, come sopra,

$$P_{\mathcal{G}}(A, \cdot) = E\left[1_A|\mathcal{G}\right]$$

$$\int_B P_{\mathcal{G}}(A, \cdot)\, dP = P(A \cap B)$$

per ogni $A \in \mathcal{F}$, $B \in \mathcal{G}$. In più però, possiamo definire integrali del tipo

$$\int_\Omega X(\omega')\, P_{\mathcal{G}}(d\omega', \omega)$$

con ω fissato ed X ad esempio limitata misurabile. Si può allora dimostrare che vale

$$\int_\Omega X(\omega')\, P_{\mathcal{G}}(d\omega', \cdot) = E\left[X|\mathcal{G}\right]$$

(nel senso che l'integrale a sinistra è un elemento della classe di equivalenza a destra). Questo inverte il procedimento visto sopra: a livello uno si può definire la probabilità condizionale a partire dalla speranza condizionale; a livello due si può definire la speranza condizionale a partire da una versione regolare della probabilità condizionale.

Concludiamo con alcune varianti dei concetti precedenti. Data una v.a. X su $(\Omega, \mathcal{F}, P)$ ed un evento $A \in \mathcal{F}$ indichiamo con $\sigma(X)$ la più piccola σ-algebra rispetto a cui X è misurabile e con $P(A|X)$ la v.a. $P(A|\sigma(X))$. Per un teorema di Doob, esiste una funzione misurabile $g_A : \mathbb{R} \to \mathbb{R}$ tale che $P(A|X) = g_A(X)$. A causa di questo, viene anche usata la notazione $P(A|X = x)$ per intendere $g_A(x)$:

$$P(A|X = x) = g_A(x).$$

Ovviamente in generale $P(X = x)$ non è positivo, per cui non sarebbe lecito definire $P(A|X = x)$ come $P(A \cap \{X = x\}) / P(X = x)$.

1.4 Proprietà di Markov

La proprietà di Markov si può esprimere a vari livelli, sia per un singolo processo stocastico, sia per una famiglia di processi, sia per una famiglia di misure di probabilità relativamente ad uno stesso processo. Inoltre si può riscrivere in innumerevoli modi. Infine, esiste la proprietà di Markov e quella di Markov forte. Illustriamo qui una possibilità tra le tante, riservandoci in seguito di descrivere altri casi ed altre formulazioni.

Sia $X = (X_t)_{t \geq 0}$ un processo stocastico su $(\Omega, \mathcal{F}, P)$ a valori nello spazio misurabile $(E, \mathcal{E})$. Sia $(\mathcal{F}_t)_{t \geq 0}$ una filtrazione (qualsiasi) in $(\Omega, \mathcal{F}, P)$.

Definizione 7 *Il processo X si dice di Markov se*

$$E[\varphi(X_{t+h}) | \mathcal{F}_t] = E[\varphi(X_{t+h}) | X_t] \tag{1.4}$$

per ogni funzione $\varphi : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ misurabile limitata, ed ogni $t \geq 0$, $h > 0$.

Proposizione 12 *Chiedere l'uguaglianza $E[\varphi(X_{t+h}) | \mathcal{F}_t] = E[\varphi(X_{t+h}) | X_t]$ equivale semplicemente a chiedere che $E[\varphi(X_{t+h}) | \mathcal{F}_t]$ sia $\sigma(X_t)$-misurabile. Inoltre, equivale a chiedere che per ogni $A \in \mathcal{E}$ valga*

$$P(X_{t+h} \in A | \mathcal{F}_t) = P(X_{t+h} \in A | X_t). \tag{1.5}$$

Oppure, che per ogni $A \in \mathcal{E}$ la v.a. $P(X_{t+h} \in A | \mathcal{F}_t)$ sia $\sigma(X_t)$-misurabile.

Proof. Se vale l'identità (1.4), allora $E[\varphi(X_{t+h}) | \mathcal{F}_t]$ è $\sigma(X_t)$-misurabile. Viceversa, supponiamo che $Y := E[\varphi(X_{t+h}) | \mathcal{F}_t]$ sia $\sigma(X_t)$-misurabile. Dal momento che Y soddisfa l'uguaglianza $E[YZ] = E[\varphi(X_{t+h}) Z]$ per ogni v.a. Z che sia $\mathcal{F}_t$-misurabile, la soddisfa anche per tutte le v.a. Z che siano $\sigma(X_t)$-misurabili. Quindi Y è una speranza condizionale di $\varphi(X_{t+h})$ sapendo X_t, ovvero vale (1.4).

Infine, se vale (1.4), allora prendendo $\varphi(x) = 1_{x \in A}$ vale (1.5). Viceversa, se vale (1.5) per ogni $A \in \mathcal{E}$, vale (1.4) per le indicatrici $\varphi(x) = 1_{x \in A}$ e poi per linearità per le loro combinazioni lineari. Si passa alle funzioni misurabili limitate o con un teorema generale o con un esplicito passaggio al limite monotono (ogni funzione misurabile limitata è limite crescente puntuale di funzioni semplici; poi si usa la convergenza monotona delle speranze condizionali). L'ultima affermazione si dimostra con ragionamenti simili. ■

La proprietà di Markov esprime il fatto che ciò che possiamo prevedere del futuro è influenzato in uguale misura dal presente come dal presente incluso il passato. Detto altrimenti, il passato non influisce sul futuro, noto il presente. Con un po' di pazienza si possono dimostrare varie formulazioni equivalenti, alcune delle quali ad esempio enfatizzano il ruolo simmetrico del passato e del futuro. Ad esempio, vale:

Proposizione 13 *Il processo X è di Markov se e solo se*

$$E\left[\Phi\left(X|_{[t,\infty)}\right)\Psi\left(X|_{[0,t]}\right)|X_t\right] = E\left[\Phi\left(X|_{[t,\infty)}\right)\right]E\left[\Psi\left(X|_{[0,t]}\right)\right]$$

per ogni coppia di funzioni $\Phi : \left(E^{[0,\infty)}, \mathcal{E}^{[0,\infty)}\right) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $\Psi : \left(E^{[0,t]}, \mathcal{E}^{[[0,t]}\right) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, misurabili limitate, ed ogni $t \geq 0$.

Il futuro è indipendente dal passato, noto il presente.

Una classe interessante di processi di Markov è quella dei processi a incrementi indipendenti, che contiene processi di fondamentale importanza.

Definizione 8 *Sia E uno spazio vettoriale. Diciamo che un processo stocastico $X = (X_t)_{t\geq 0}$, rispetto ad una filtrazione $(\mathcal{F}_t)_{t\geq 0}$, ha incrementi indipendenti, se per ogni $t, h \geq 0$ la v.a. $X_{t+h} - X_t$ è indipendente da $\mathcal{F}_t$.*

Proposizione 14 *Un processo a incrementi indipendenti è di Markov.*

Proof. Applichiamo la Proposizione 11. Dati $t, h \geq 0$, sia $\mathcal{G}'=\sigma\{X_{t+h} - X_t\}$, $\mathcal{G} = \mathcal{F}_t$. Le σ-algebre $\mathcal{G}$ e $\mathcal{G}'$ sono indipendenti. La v.a. $X_{t+h} - X_t$ è $\mathcal{G}$-misurabile. Presa una funzione misurabile limitata $\psi : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, vogliamo verificare che $E[\psi(X_{t+h})|\mathcal{F}_t]$ è $\sigma\{X_t\}$-misurabile. Scriviamo

$$E[\psi(X_{t+h})|\mathcal{F}_t] = E[\psi(X_t + X_{t+h} - X_t)|\mathcal{F}_t]$$

e consideriamo la funzione $\varphi(x, \omega) = \psi(x + X_{t+h}(\omega) - X_t(\omega))$. Essa è $\mathcal{E}\otimes\mathcal{G}'$-misurabile. Quindi sono soddisfatte tutte le ipotesi della Proposizione 11 (con $X = X_t$), che fornisce l'identità

$$E[\psi(X_t + X_{t+h} - X_t)|\mathcal{F}_t] = E[\varphi(x, \cdot)]|_{x=X_t}.$$

Quindi $E[\psi(X_{t+h})|\mathcal{F}_t]$ è $\sigma\{X_t\}$-misurabile. La dimostrazione è completa. ■

Introduciamo ora un concetto collegato alla proprietà di Markov.

Definizione 9 *Su uno spazio misurabile $(E, \mathcal{E})$, diciamo che una funzione $p(s,t,x,A)$, definita per $t > s \geq 0$, $x \in E$, $A \in \mathcal{E}$, è una funzione di transizione markoviana se:*
i) per ogni (s,t,A), $x \longmapsto p(s,t,x,A)$ è $\mathcal{E}$-misurabile
ii) per ogni (s,t,x), $A \longmapsto p(s,t,x,A)$ è una misura di probabilità su $(E, \mathcal{E})$
iii) vale l'equazione di Chapman-Kolmogorov

$$p(s,t,x,A) = \int_E p(u,t,y,A)\, p(s,u,x,dy)$$

per ogni $t > u > s \geq 0$, $x \in E$, $A \in \mathcal{E}$.

Definizione 10 *Data una funzione di transizione markoviana p ed una misura di probabilità μ_0 su $(E, \mathcal{E})$, diciamo che un processo stocastico $X = (X_t)_{t \geq 0}$, rispetto ad una filtrazione $(\mathcal{F}_t)_{t \geq 0}$ (su uno spazio probabilizzato $(\Omega, \mathcal{F}, P)$) è un processo di Markov associato a p con legge iniziale μ_0 se:*
i) X_0 ha legge μ_0
ii) per ogni $t \geq 0$, $h > 0$, $A \in \mathcal{E}$ vale

$$P(X_{t+h} \in A | \mathcal{F}_t) = p(t, t+h, X_t, A). \tag{1.6}$$

Si noti che se un processo soddisfa le condizioni di quest'ultima definizione, allora è di Markov, per la Proposizione 12. La definizione quindi arricchisce la proprietà di Markov.

Come sopra, la proprietà (1.6) è equivalente a

$$E[\varphi(X_{t+h}) | \mathcal{F}_t] = \int \varphi(y)\, p(t, t+h, X_t, dy) \tag{1.7}$$

per ogni φ misurabile limitata.

L'introduzione della funzione $p(s,t,x,A)$ è concettualmente (ed anche tecnicamente) un effettivo arricchimento. Essa corrisponde all'idea intuitiva che, ad ogni istante $s \geq 0$, a causa della proprietà di Markov il processo "riparta" dalla posizione x occupata in quell'istante, scordandosi del passato. La posizione x al tempo s identifica la probabilità $p(s,t,x,A)$ di occupare una posizione in A in un generico istante successivo t.

Nascono due domande naturali:

1. dato un processo di Markov, esiste una funzione di transizione markoviana p ad esso associata?

2. Data una funzione di transizione markoviana p ed una generica probabilità iniziale μ_0, esiste un processo di Markov X ad esse associato?

La prima domanda è delicata per cui ci accontenteremo di verificare l'esistenza in vari esempi o di supporla in certi teoremi.

Osservazione 7 *Mostriamo alcuni passi nella direzione di una costruzione di* p *partendo da un processo di Markov. Se il processo* X *è definito su uno spazio* $(\Omega, \mathcal{F}, P)$ *in cui* Ω *è metrico completo e separabile,* $\mathcal{F} = \mathcal{B}(\Omega)$*, presa la sotto-*$\sigma$*-algebra* $\mathcal{G} = \mathcal{F}_s$*, esiste una versione regolare della probabilità condizionale di* P *rispetto a* $\mathcal{G}$*, che indichiamo con* $P_s(A', \omega)$*,* $A' \in \mathcal{F}$*,* $\omega \in \Omega$*. Essa soddisfa, oltre alla misurabilità rispetto a* $\mathcal{F}_s$ *ed all'essere una misura di probabilità rispetto ad* A'*,*

$$P_s(A', \cdot) = P(A'|\mathcal{F}_s).$$

Preso A' *della forma* $A' = \{X_t \in A\}$ *con* $A \in \mathcal{E}$*, scriviamo*

$$\rho(s, t, \omega, A) := P_s(\{X_t \in A\}, \omega).$$

Questo soddisfa

$$\rho(s, t, \cdot, A) = P(X_t \in A|\mathcal{F}_s) = P(X_t \in A|X_s)$$

dove abbiamo usato l'ipotesi di markovianità. La funzione $A \longmapsto \rho(s, t, \omega, A)$ *è una misura di probabilità. Il problema è passare da* $\rho(s, t, \omega, A)$*,* $\omega \in \Omega$*, a* $p(s, t, x, A)$*,* $x \in E$*.*

La seconda domanda ha risposta affermativa in spazi degli stati σ-compatti. Premettiamo una proposizione che illustra come la conoscenza della funzione di transizione markoviana e della legge iniziale sono sufficienti a identificare la legge di un processo, se si suppone che esso sia di Markov.

Proposizione 15 *Se* X *è un processo di Markov con funzione di transizione markoviana* p*, allora, per ogni* $\tau = (t_1, ..., t_n) \in \mathcal{S}$*,* $A_1, ..., A_n \in \mathcal{E}$*, vale*

$$\begin{aligned} &P(X_{t_n} \in A_3, \cdots, X_{t_1} \in A_1) \\ &\int_{A_1} \mu_{t_1}(dy_1) \int_{A_2} p(t_1, t_2, y_1, dy_2) \cdots \int_{A_n} p(t_{n-1}, t_n, y_{n-1}, dy_n) \end{aligned}$$

dove μ_{t_1} *è la legge di* X_{t_1}*.*

Proof. Il caso $n = 3$ chiarisce completamente la dimostrazione, che va poi svolta per induzione. Chi avesse visto la dimostrazione perle catene di Markov ne riconoscerebbe la struttura, basata sulla disintegrazione rispetto all'ultimo istante di tempo. Vale

$$\begin{aligned} P(X_{t_3} \in A_3, X_{t_2} \in A_2, X_{t_1} \in A_1) &= E[1_{A_3}(X_{t_3}) 1_{A_2}(X_{t_2}) 1_{A_1}(X_{t_1})] \\ &= E[E[1_{A_3}(X_{t_3}) 1_{A_2}(X_{t_2}) 1_{A_1}(X_{t_1}) |\mathcal{F}_{t_2}]] \\ &= E[1_{A_2}(X_{t_2}) 1_{A_1}(X_{t_1}) E[1_{A_3}(X_{t_3}) |\mathcal{F}_{t_2}]] \\ &= E[1_{A_2}(X_{t_2}) 1_{A_1}(X_{t_1}) p(t_2, t_3, X_{t_2}, A_3)]. \end{aligned}$$

Analogamente, usando (1.6),

$$\begin{aligned}
&E\left[1_{A_2}\left(X_{t_2}\right)1_{A_1}\left(X_{t_1}\right)p\left(t_2,t_3,X_{t_2},A_3\right)\right]\\
&=E\left[E\left[1_{A_2}\left(X_{t_2}\right)1_{A_1}\left(X_{t_1}\right)p\left(t_2,t_3,X_{t_2},A_3\right)|\mathcal{F}_{t_1}\right]\right]\\
&=E\left[1_{A_1}\left(X_{t_1}\right)E\left[1_{A_2}\left(X_{t_2}\right)p\left(t_2,t_3,X_{t_2},A_3\right)|\mathcal{F}_{t_1}\right]\right]\\
&=E\left[1_{A_1}\left(X_{t_1}\right)\int 1_{A_2}\left(y\right)p\left(t_2,t_3,y,A_3\right)p\left(t_1,t_2,X_{t_1},dy\right)\right]\\
&E\left[1_{A_1}\left(X_{t_1}\right)\int_{A_2}p\left(t_2,t_3,y,A_3\right)p\left(t_1,t_2,X_{t_1},dy\right)\right].
\end{aligned}$$

Infine, quest'ultima espressione si può riscrivere nella forma

$$\begin{aligned}
&\int 1_{A_1}\left(y_1\right)\int_{A_2}p\left(t_2,t_3,y_2,A_3\right)p\left(t_1,t_2,y_1,dy_2\right)\mu_{t_1}\left(dy_1\right)\\
&=\int_{A_1}\mu_{t_1}\left(dy_1\right)\int_{A_2}p\left(t_1,t_2,y_1,dy_2\right)\int_{A_3}p\left(t_2,t_3,y_2,dy_3\right).
\end{aligned}$$

■

Proposizione 16 *Supponiamo E metrico σ-compatto. Date una funzione di transizione markoviana p ed una generica probabilità iniziale μ_0, esiste un processo di Markov X ad esse associato, unico in legge.*

Proof. Diamo solo la traccia. Cerchiamo di usare il Teorema di Kolmogorov di costruzione dei processi. Per far questo, dobbiamo costruire quelle che saranno le distribuzioni di dimensione finita del processo. Sulla base della proposizione precedente, per ogni $\tau=(t_1,...,t_n)\in\mathcal{S}$ indichiamo con μ_τ la misura di probabilità su $(E^n,\mathcal{E}^n)$ tale che

$$\mu_\tau\left(A_1\times...\times A_n\right)=\int_{A_1}\mu_{t_1}\left(dy_1\right)\int_{A_2}p\left(t_1,t_2,y_1,dy_2\right)\cdots\int_{A_n}p\left(t_{n-1},t_n,y_{n-1},dy_n\right).$$

La consistenza si verifica immediatamente usando l'equazione di Chapman-Kolmogorov (si deve prendere $A_i=\mathcal{E}$). Le altre proprietà richiedono un po' di lavoro che non riportiamo. ■

Capitolo 2

Il moto browniano

2.1 Definizione, esistenza e proprietà di Markov

Definizione 11 *Un moto browniano grossolano è un processo stocastico* $B = (B_t)_{t\geq 0}$ *tale che*

i) $P(B_t = 0) = 1$

ii) per ogni $t > s \geq 0$, $B_t - B_s$ *è una gaussiana* $N(0, t-s)$

iii) per ogni $n \in \mathbb{N}$ *ed ogni sequenza* $t_n > ... > t_1 \geq 0$, *le v.a.* $B_{t_n} - B_{t_{n-1}}$, $B_{t_{n-1}} - B_{t_{n-2}}, \ldots, B_{t_2} - B_{t_1}$ *sono indipendenti.*

Se aggiungiamo che sia continuo, lo chiameremo moto browniano perfezionato o semplicemente moto browniano. La sua legge su $(C[0,\infty), \mathcal{B}(C[0,\infty)))$ *sarà detta misura di Wiener.*

A causa della proprietà (i) si suole dire che è un moto browniano *standard*. Altrimenti si può considerare il caso in cui il processo esca da un altro punto iniziale, diverso dallo zero.

A volte è utile considerare il concetto di moto browniano rispetto ad una filtrazione data.

Definizione 12 *Un moto browniano, rispetto ad una filtrazione* $(\mathcal{F}_t)_{t\geq 0}$, *è un processo stocastico* $B = (B_t)_{t\geq 0}$ *tale che*

i) $P(B_t = 0) = 1$

ii) è adattato

iii) per ogni $t > s \geq 0$, $B_t - B_s$ *è una gaussiana* $N(0, t-s)$ *indipendente da* $\mathcal{F}_s$.

Se non imponiamo che sia continuo, lo diremo moto browniano grossolano, altrimenti lo chiameremo moto browniano perfezionato o semplicemente moto browniano.

Nel caso della definizione 11, sia $(\mathcal{F}_t^{00})_{t\geq 0}$ la filtrazione naturale: $\mathcal{F}_t^{00}$ è la più piccola σ-algebra che rende misurabili tutte le v.a. B_s, con $s \in [0, T]$.

Proposizione 17 *Un processo soddisfa la definizione 12 rispetto alla filtrazione naturale* $(\mathcal{F}_t^{00})_{t\geq 0}$, *se e solo se soddisfa la definizione 11.*

Proof. Supponiamo che valga la definizione 12 rispetto alla filtrazione naturale $(\mathcal{F}_t^{00})_{t\geq 0}$. Presi $t_n > ... > t_1 \geq 0$, osserviamo che gli incrementi $B_{t_{n-1}} - B_{t_{n-2}}, \ldots, B_{t_2} - B_{t_1}$ sono misurabili rispetto a $\mathcal{F}_{t_{n-1}}^{00}$, quindi $B_{t_n} - B_{t_{n-1}}$ è indipendente da essi:

$$\begin{aligned} & P\left(B_{t_n} - B_{t_{n-1}} \in A_n, \left(B_{t_{n-1}} - B_{t_{n-2}}, ..., B_{t_2} - B_{t_1}\right) \in A'\right) \\ = \; & P\left(B_{t_n} - B_{t_{n-1}} \in A_n\right) P\left(\left(B_{t_{n-1}} - B_{t_{n-2}}, ..., B_{t_2} - B_{t_1}\right) \in A'\right) \end{aligned}$$

per ogni $A_n \in \mathcal{B}(\mathbb{R})$ ed $A' \in \mathcal{B}(\mathbb{R}^{n-1})$. Questo vale in particolare per insiemi del tipo $A' = A_{n-1} \times \cdots \times A_2$ con $A_i \in \mathcal{B}(\mathbb{R})$. Basta poi iterare il ragionamento per ottenere

$$P\left(B_{t_n} - B_{t_{n-1}} \in A_n, ..., B_{t_2} - B_{t_1} \in A_2\right) = \prod_{i=2}^{n} P\left(B_{t_i} - B_{t_{i-1}} \in A_i\right)$$

che è la condizione (iii) della definizione 11. Le altre proprietà sono ovvie.

Viceversa, supponiamo che valga la definizione 11. Presi $t > s \geq 0$, l'unico problema è dimostrare che $B_t - B_s$ è indipendente da $\mathcal{F}_s^{00}$. Ricordando il criterio dell'esercizio 8, è sufficiente dimostrare l'indipendenza tra $\sigma(B_t - B_s)$ e gli elementi di una base di $\mathcal{F}_s^{00}$. Prendiamo la famiglia $\mathcal{A} \subset \mathcal{F}_s^{00}$ degli eventi della forma $\{B_{u_1} \in A_1, ..., B_{u_n} \in A_n\}$ al variare di $u_1 \leq ... \leq u_n \in [0, s]$, $A_1, ..., A_n \in \mathcal{B}(\mathbb{R})$. Questa è una base, in quanto genera $\mathcal{F}_s^{00}$, è chiusa per intersezione finita, contiene Ω. Basta allora dimostrare l'indipendenza tra $\sigma(B_t - B_s)$ e gli elementi di $\mathcal{A}$. Fissati $u_1 \leq ... \leq u_n \in [0, s]$, basta quindi dimostrare l'indipendenza tra $\sigma(B_t - B_s)$ e gli elementi di $\sigma(B_{u_1}, B_{u_2}, ..., B_{u_n})$. Le due σ-algebre $\sigma(B_{u_1}, B_{u_2}, ..., B_{u_n})$ e $\sigma\left(B_{u_1}, B_{u_2} - B_{u_1}, ..., B_{u_n} - B_{u_{n-1}}\right)$ coincidono (i vettori $(B_{u_1}, B_{u_2}, ..., B_{u_n})$ e $\left(B_{u_1}, B_{u_2} - B_{u_1}, ..., B_{u_n} - B_{u_{n-1}}\right)$ sono legati da una trasformazione biunivoca). Quindi basta mostrare l'indipendenza tra $\sigma(B_t - B_s)$ ed una base di $\sigma\left(B_{u_1}, B_{u_2} - B_{u_1}, ..., B_{u_n} - B_{u_{n-1}}\right)$, ad esempio quella formata dagli eventi della forma $\left\{B_{u_1} \in A_1, B_{u_2} - B_{u_1} \in A_2, ..., B_{u_n} - B_{u_{n-1}} \in A_n\right\}$ al variare di $A_1, ..., A_n \in \mathcal{B}(\mathbb{R})$. A questo punto la verifica è ovvia, basandoci sull'indipendenza delle v.a. $B_t - B_s$, $B_s - B_{u_n}$, $B_{u_n} - B_{u_{n-1}}$, ... , $B_{u_2} - B_{u_1}$, B_{u_1}. La dimostrazione è completa. ■

Esistono numerose formulazioni equivalenti della definizione di moto browniano (si veda anche la sezione 2.1.1). Eccone una utile per la costruzione successiva.

Proposizione 18 *Se* $B = (B_t)_{t\geq 0}$ *è un moto browniano standard grossolano, presa* $\tau = (t_1, ..., t_n) \in \mathcal{S}$, *la distribuzione marginale* μ_τ *su* $(\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n)$ *è una gaussiana di media nulla e matrice di covarianza* $Q^{(\tau)}$ *di componenti*

$$Q_{ij}^{(\tau)} = t_i \wedge t_j.$$

Quindi B *è un processo gaussiano, con funzione valor medio* $m(t) = 0$ *e funzione di covarianza*

$$C(t, s) = t \wedge s.$$

Viceversa, un processo gaussiano con tali funzioni media e covarianza, è un moto browniano grossolano.

Proof. Il vettore $\left(B_{t_n}-B_{t_{n-1}}, B_{t_{n-1}}-B_{t_{n-2}}, ..., B_{t_2}-B_{t_1}, B_{t_1}\right)$ è gaussiano (si noti che possiamo scrivere B_{t_1} come $B_{t_1}-B_0$). Infatti, ha componenti indipendenti e gaussiane (in generale non sarebbe vero se le componenti non fossero indipendenti), quindi la densità congiunta si può scrivere come prodotto delle marginali, che essendo gaussiane conducono con facili calcoli ad una densità congiunta gaussiana.

Il vettore $(B_{t_1}, ..., B_{t_n})$ si ottiene dal precedente con una trasformazione lineare, quindi è anch'esso gaussiano. La sua media è nulla (per verifica diretta). Calcoliamo la sua matrice di covarianza direttamente, piuttosto che con la regola di trasformazione citata a suo tempo. Se $j>i$, quindi $t_j>t_i$, vale

$$\begin{aligned} Q_{ij}^{(\mu)} &= Cov\left(B_{t_i}, B_{t_j}\right) = Cov\left(B_{t_i}, B_{t_j}-B_{t_i}\right) + Cov\left(B_{t_i}, B_{t_i}\right) \\ &= Cov\left(B_{t_i}, B_{t_i}\right) = Var\left[B_{t_i}\right] = t_i \end{aligned}$$

per linearità della covarianza nei suoi argomenti, indipendenza tra $B_{t_j}-B_{t_i}$ e B_{t_i} ed altri fatti ovvi. Quindi $Q_{ij}^{(\mu)} = t_i \wedge t_j$.

Viceversa, supponiamo che B sia un processo gaussiano con funzioni $m(t)=0$ e $C(t,s)=t\wedge s$. Al tempo $t=0$ abbiamo $E[B_0]=0$, $Var[B_0]=C(0,0)=0$, quindi B_0 è una v.a. che vale 0 quasi certamente (ad esempio si deduce dalla disuguaglianza di Chebyshev). Quindi la (i) è verificata. Il vettore $(B_{t_1}, ..., B_{t_n})$ è gaussiano. Usando la trasformazione inversa di quella usata nella prima parte della dimostrazione, si deduce che il vettore $\left(B_{t_n}-B_{t_{n-1}}, B_{t_{n-1}}-B_{t_{n-2}}, ..., B_{t_2}-B_{t_1}, B_{t_1}\right)$ è gaussiano. La sua media è nulla per verifica diretta. Gli elementi della sua matrice di covarianza valgono

$$\begin{aligned} & Cov\left(B_{t_{n-i}}-B_{t_{n-i-1}}, B_{t_{n-j}}-B_{t_{n-j-1}}\right) \\ = \; & Cov\left(B_{t_{n-i}}, B_{t_{n-j}}\right) - Cov\left(B_{t_{n-i}}, B_{t_{n-j-1}}\right) - Cov\left(B_{t_{n-i-1}}, B_{t_{n-j}}\right) + Cov\left(B_{t_{n-i-1}}, B_{t_{n-j-1}}\right) \end{aligned}$$

che, per $i>j$ diventano ($t_{n-i-1}<t_{n-i}\leq t_{n-j-1}<t_{n-j}$), usando la formula $Cov(B_t, B_s) = t\wedge s$,

$$= t_{n-i} - t_{n-i} - t_{n-i-1} + t_{n-i-1} = 0$$

mentre per $i=j$ diventano ($t_{n-i-1}=t_{n-j-1}<t_{n-i}=t_{n-j}$)

$$= t_{n-i} - t_{n-i-1} - t_{n-i-1} + t_{n-i-1} = t_{n-i} - t_{n-i-1}.$$

Quindi la covarianza è diagonale, cioè le componenti del vettore aleatorio sono scorrelate e questo implica che sono indipendenti, perché si tratta di un vettore gaussiano. Le componenti sono gli incrementi del moto browniano, che quindi sono indipendenti (proprietà (iii)), oltre che gaussiani. Vale infine la proprietà (ii): abbiamo già stabilito che il generico incremento è gaussiano e centrato; vale poi, con gli stessi calcoli appena eseguiti, per $t > s \geq 0$,

$$\begin{aligned} Var\left[B_t - B_s\right] &= Cov\left(B_t, B_t\right) - Cov\left(B_t, B_s\right) - Cov\left(B_s, B_t\right) + Cov\left(B_s, B_s\right) \\ &= t - s - s + s = t - s. \end{aligned}$$

La dimostrazione è completa. ■

Esistono numerose costruzioni di un moto browniano grossolano. Fatta la fatica della proposizione precedente, per noi la strada più breve è basarsi sul teorema di costruzione di Kolmogorov, nella sua applicazione generale ai processi gaussiani.

Proposizione 19 *Un moto browniano grossolano esiste. La legge su* $\left(\mathbb{R}^{[0,\infty)}, \mathcal{B}\left(\mathbb{R}\right)^{[0,\infty)}\right)$ *è univocamente determinata.*

Proof. La legge su $\left(\mathbb{R}^{[0,\infty)}, \mathcal{B}\left(\mathbb{R}\right)^{[0,\infty)}\right)$ è univocamente determinata dalle marginali, che sono univocamente note in base al lemma precedente. Veniamo alla costruzione.

Per l'esistenza, grazie alla Proposizione 18, basta che mostriamo l'esistenza di un processo gaussiano con $m\left(t\right) = 0$ e $C\left(t, s\right) = t \wedge s$. Possiamo usare il criterio di esistenza di processi gaussiani con funzioni $m\left(t\right)$ e $C\left(t, s\right)$ date. Per far questo, basta verificare che $C\left(t, s\right) = t \wedge s$ sia una funzione semidefinita positiva, nel senso che valga

$$\sum_{i,j=1}^{n} \left(t_i \wedge t_j\right) \xi_i \xi_j \geq 0 \tag{2.1}$$

per ogni $\left(t_1, ..., t_n\right) \in [0, \infty)^n$, $\left(\xi_1, ..., \xi_n\right) \in \mathbb{R}^n$. Il resto della dimostrazione si occupa di questa verifica.

Ispirandoci a quanto già scoperto sopra (Proposizione 18), possiamo mostrare che esiste un vettore aleatorio con matrice di covarianza avente componenti $t_i \wedge t_j$. Siccome una matrice di covarianza è semidefinita positiva, la proprietà (2.1) sarà dimostrata. Prendiamo delle v.a. $X_1, ..., X_n$ indipendenti, centrate e di varianze $t_1, t_2 - t_1, ..., t_n - t_{n-1}$. Definiamo delle v.a. $Y_1, ..., Y_n$ tramite le relazioni

$$\begin{aligned} X_1 &= Y_1 \\ X_2 &= Y_2 - Y_1 \\ &\ldots \\ X_n &= Y_n - Y_{n-1} \end{aligned}$$

(relazioni chiaramente invertibili). Per ogni $i = 2, ..., n$, vale

$$Y_i = Y_i - Y_{i-1} + ... + Y_2 - Y_1 + Y_1 = X_i + ... + X_2 + X_1$$

quindi

$$Var\left[Y_i\right] = Var\left[X_i\right] + ... + Var\left[X_1\right] = t_i - t_{i-1} + ... + t_2 - t_1 + t_1 = t_i$$

ovvero

$$Cov\,(Y_i, Y_i) = t_i = t_i \wedge t_i.$$

Inoltre, se $j > i$, per ragioni analoghe (essendo cioè $Y_i = X_i + ... + X_2 + X_1$), le v.a. $X_j, X_{j-1}..., X_{i+1}$ sono indipendenti da Y_i, quindi

$$\begin{aligned} Cov\,(Y_j, Y_i) &= Cov\,(Y_j - Y_{j-1} + ... + Y_{i+1} - Y_i + Y_i, Y_i) \\ &= Cov\,(X_j + ... + X_{i+1} + Y_i, Y_i) \\ &= Cov\,(Y_i, Y_i) = t_i = t_i \wedge t_j. \end{aligned}$$

Quindi la matrice di covarianza di $(Y_1, ..., Y_n)$ ha le componenti desiderate. Questo conclude la dimostrazione.

Volendo si possono perseguire dimostrazioni più algebriche, o ragionando sui minori della matrice, oppure verificando la condizione (2.1) per n basso e trovando poi una struttura iterativa. Per $n = 2$ vale

$$\sum_{i,j=1}^{2} (t_i \wedge t_j)\,\xi_i\xi_j = t_1\,(\xi_1 + \xi_2)^2 + (t_2 - t_1)\,\xi_2^2$$

mentre per $n = 3$ vale

$$\begin{aligned} \sum_{i,j=1}^{3} (t_i \wedge t_j)\,\xi_i\xi_j &= t_1\,(\xi_1 + \xi_2 + \xi_3)^2 \\ &\quad + (t_2 - t_1)\,\xi_2^2 + 2\,(t_2 - t_1)\,\xi_2\xi_3 + (t_3 - t_1)\,\xi_3^2 \end{aligned}$$

dove la somma degli ultimi tre termini è simile al punto di partenza del caso $n = 2$. Ci vuole però molta pazienza a completare il conto in questo modo. La dimostrazione è completa. ∎

Esaminiamo ora l'esistenza di un moto browniano continuo. Premettiamo il seguente teorema che non dimostriamo in quanto ha fatto parte di un corso precedente. E' detto teorema di regolarità di Kolmogorov.

Teorema 4 *Se un processo stocastico $X = (X_t)_{t\in D}$ definito in un aperto $D \subset \mathbb{R}^n$ soddisfa*

$$E\left[|X_t - X_s|^p\right] \leq C\,|t-s|^{n+\alpha}$$

per certe costanti $p, C, \alpha > 0$, allora esiste una modificazione X'_t di X_t tale che le traiettorie di X'_t sono γ-hölderiane per ogni $\gamma < \frac{\alpha}{p}$.

Proposizione 20 *Esiste un moto browniano continuo e le sue traiettorie sono α-hölderiane, per ogni $\alpha \in (0, 1/2)$. Precisamente, per ogni $T > 0$ ed ogni $\alpha \in (0, 1/2)$ esiste una funzione $C_{T,\alpha} : \Omega \to (0,\infty)$ tale che*

$$|B_t(\omega) - B_s(\omega)| \leq C_{T,\alpha}(\omega)\,|t-s|^\alpha \qquad \textit{per ogni } t, s \in [0,T] \textit{ ed ogni } \omega \in \Omega.$$

Proof. Sia B un moto browniano grossolano. La v.a. $\frac{B_t - B_s}{\sqrt{t-s}}$ (per $t > s$) è una normale standard. Fissiamo $n \in \mathbb{N}$, diciamo $n > 2$. Siccome $\int_{-\infty}^{+\infty} |x|^n e^{-x^2/2} dx < \infty$, esiste una costante $C_n > 0$ tale che

$$E\left[\left|\frac{B_t - B_s}{\sqrt{t-s}}\right|^n\right] \leq C_n$$

e quindi

$$E\left[|B_t - B_s|^n\right] \leq C_n (t-s)^{n/2} = C_n (t-s)^{1+\frac{n-2}{2}}.$$

La costante C_n non dipende da $t > s \geq 0$. In base al teorema di regolarità di Kolmogorov, esiste una modificazione $B_t^{(n)}$ di B_t tale che le traiettorie di $B_t^{(n)}$ sono γ-hölderiane per ogni $\gamma < \frac{\frac{n-2}{2}}{n} = \frac{n-2}{2n}$. Innanzi tutto, preso ad es. $n = 3$, questo dimostra l'esistenza di un moto browniano continuo: infatti $B_t^{(3)}$ è un processo continuo che soddisfa le proprietà (i) ed (ii) della definizione 11; anche la (iii) è facile, usando l'esercizio 9. Quindi $B_t^{(3)}$ è un moto browniano continuo. Prendiamo ora $B_t^{(n)}$, con $n > 3$ qualsiasi. E' una modificazione continua di B_t, quindi anche di $B_t^{(3)}$. Ma due processi X ed Y continui che sono modificazione uno dell'altro sono indistinguibili: numerati i tempi positivi razionali con la successione $\{t_n\}$, posto $N_n = \{X_{t_n} \neq Y_{t_n}\}$, l'insieme $N = \cup_n N_n$ ha misura nulla, per $\omega \in N^c$ vale $X_t(\omega) = Y_t(\omega)$ per ogni $t \in \mathbb{Q} \cap [0, \infty)$, e quindi per ogni $t \geq 0$ essendo processi continui.

Abbiamo così dimostrato che $B_t^{(n)}$ è indistinguibile da $B_t^{(3)}$. Sia N_n l'insieme dove differiscono, di misura nulla. Sia $N = \cup_n N_n$, di misura nulla. Per $\omega \in N^c$, $B_t^{(3)}$ ha la regolarità delle traiettorie di qualsiasi $B_t^{(n)}$, quindi γ-hölderiane per ogni $\gamma < \frac{n-2}{2n}$, con n qualsiasi, quindi γ-hölderiane per ogni $\gamma < \frac{1}{2}$. Abbiamo trovato un moto browniano che quasi certamente ha traiettorie γ-hölderiane per ogni $\gamma < \frac{1}{2}$. Sull'insieme di misura nulla in cui questo non accade possiamo porlo identicamente uguale a zero. Il nuovo processo soddisfa le condizioni della proposizione. ■

Proposizione 21 *Il moto browniano (anche grossolano) è un processo di Markov.*

Proof. Basta applicare una proposizione vista nel paragrafo dei processi di Markov che asserisce che ogni processo a incrementi indipendenti è di Markov. Il moto browniano ha infatti incrementi indipendenti. ■

Esercizio 8 *Sia $(\Omega, \mathcal{F}, P)$ uno spazio probabilizzato. Ricordiamo che un insieme di generatori di $\mathcal{F}$ è una famiglia $\mathcal{H} \subset \mathcal{P}(\Omega)$ tale che $\mathcal{F} = \sigma(\mathcal{H})$. Invece una base di $\mathcal{F}$ è un insieme di generatori chiuso per intersezione finita e contenente Ω (o almeno tale che l'unione numerabile di opportuni elementi di $\mathcal{H}$ sia Ω). Ricordiamo ad esempio che due misure coincidenti su una base coincidono su tutta la σ-algebra. Ricordiamo poi che due sotto σ-algebre $\mathcal{G}$, $\mathcal{G}'$ di $\mathcal{F}$ si dicono indipendenti se vale $P(A \cap B) = P(A) P(B)$ per ogni $A \in \mathcal{G}$, $B \in \mathcal{G}'$. Sia $\mathcal{H}$ una base di $\mathcal{G}'$. Mostrare che, se vale $P(A \cap B) = P(A) P(B)$ per ogni $A \in \mathcal{G}$, $B \in \mathcal{H}$, allora $\mathcal{G}$ e $\mathcal{G}'$, sono indipendenti.*

[Verificare che l'insieme di tutti gli elementi $B \in \mathcal{F}$ tali che $P(A \cap B) = P(A) P(B)$ per ogni $A \in \mathcal{G}$, è una σ-algebra.]

Esercizio 9 *Verificare che se $X_1, ..., X_n$ sono v.a. indipendenti, e $X_1', ..., X_n'$ sono P-equivalenti alle $X_1, ..., X_n$ rispettivamente, allora le v.a. $X_1', ..., X_n'$ sono indipendenti. [Preso $A_1 \in \mathcal{B}(\mathbb{R})$, esistono due insiemi N_1, N_1' di misura nulla tali che $\{X_1' \in A_1\} \cup N_1' = \{X_1 \in A_1\} \cup N_1$.]*

2.1.1 Una motivazione modellistica

Brown osservò che particelle molto leggere ma visibili (tipo polline) messe in sospensione in un liquido si muovono incessantemente e molto irregolarmente. Chiamiamole particelle browniane. La spiegazione del loro moto, così erratico, è che le particelle browniane risentono degli innumerevoli urti molecolari delle molecole del fluido e si spostano in conseguenza di essi.

Alcuni studiosi (Ornstein, Uhlenbeck, Chandrasekar ecc.) hanno dato una descrizione tramite equazioni di Newton, molto realistica. Adottiamo invece una descrizione più fenomenologica, basata più su argomenti ragionevoli e semplici piuttosto che su un dettagliato bilancio di forze.

Sia X_t la posizione al tempo t della particella browniana. Per semplicità supponiamola uni-dimensionale (oppure descriviamo una delle due componenti). La traiettoria $t \longmapsto X_t$ è continua. Gli spostamenti $X_{t_n} - X_{t_{n-1}}, X_{t_{n-1}} - X_{t_{n-2}}, \ldots, X_{t_2} - X_{t_1}$ relativi a intervalli temporali disgiunti si possono considerare ragionevolmente indipendenti, visto che gli innumerevoli urti in ciascun intervallo $[t_{i-1}, t_i]$ sono relativi a molecole diverse. Lo spostamento $X_{t+h} - X_t$ ha ragionevolmente una legge statistica che dipende solo da h e non da t. Possiamo inoltre supporre che i momenti di $X_{t+h} - X_t$ siano finiti, ad esempio fino all'ordine 3 (questa è una condizione puramente tecnica), per h piccolo, senza violare così l'intuizione fisica (non ci aspettiamo spostamenti troppo grandi, mediamente); e che abbia media nulla. Infine, poniamo la particella borwniana nell'origine, al tempo $t = 0$.

Tra queste, l'unica ipotesi che è un'idealizzazione della realtà è quella di indipendenza, ma sembra molto vicina al reale. Si noti che non abbiamo ipotizzato nulla di quantitativo circa la legge (es. gaussiana) e la dipendenza della varianza da h.

Supponiamo quindi di avere un processo stocastico X_t tale che:

i) $P(X_t = 0) = 1$, X_t ha traiettorie continue

ii) $X_{t_n} - X_{t_{n-1}}, X_{t_{n-1}} - X_{t_{n-2}}, \ldots, X_{t_2} - X_{t_1}$ sono v.a. indipendenti, per ogni $t_n \geq t_{n-1} \geq \ldots \geq t_1 \geq 0$

iii) la legge di $X_{t+h} - X_t$ dipende solo da h, ha media zero, ed esiste una costante $C > 0$ tale che $E\left[|X_h|^3\right] \leq C$ per h sufficientemente piccolo.

Un simile processo è unico, è cioè univocamente identificato da questi requisiti? Si può precisare la distribuzione statistica degli incrementi, o si può calcolare quella della posizione X_t? La risposta è affermativa. Omettiamo la dimostrazione completa del teorema che richiede alcuni argomenti un po' tecnici sulle funzioni caratteristiche.

Teorema 5 *La varianza di X_t è finita per ogni t. Supponendo (per fissare un'unità di misura) che $Var[X_1] = 1$, X è una moto browniano.*

Proof. Fissato $t > 0$, preso $n \in \mathbb{N}$ e detto $h = \frac{t}{n}$, vale

$$X_t = \left(X_t - X_{\frac{n-1}{n}t}\right) + ... + \left(X_{\frac{2}{n}t} - X_{\frac{1}{n}t}\right) + \left(X_{\frac{1}{n}t} - X_0\right).$$

Le v.a. tra parentesi sono indipendenti ed hanno varianza finita (almeno se n è grande) per cui X_t ha varianza finita. Fissiamo la scala ponendo $Var[X_1] = 1$.

Mostriamo che $Var[X_t] = t$ per ogni t razionale positivo. Se $t = \frac{n}{m}$ con n, m interi positivi, vale $X_t = \left(X_{n/m} - X_{(n-1)/m}\right) + ... + \left(X_{1/m} - X_0\right)$, quindi $Var\left[X_{n/m}\right] = nVar\left[X_{1/m}\right]$. La relazione vale anche per $n = m$, quindi $1 = mVar\left[X_{1/m}\right]$. Vale quindi

$$Var\left[X_{\frac{n}{m}}\right] = \frac{n}{m}.$$

Per l'ipotesi sul momento terzo e la continuità delle traiettorie, possiamo applicare il criterio di passaggio al limite in valore atteso di Vitali e dedurre $Var[X_t] = t$ per ogni $t \geq 0$. Questa dimostrazione si può ripetere per il generico incremento $X_t - X_s$ provando che $Var[X_t - X_s] = t - s$ per ogni $t > s \geq 0$.

Manca solo la gaussianità di $X_t - X_s$, per aver verificato tutte le condizioni di un moto browniano. La gaussianità si dimostra nello stesso modo del teorema limite centrale, ma facendo entrare in gioco la continuità delle traiettorie. Questa parte è piuttosto tecnica e la omettiamo. ■

2.2 Regolarità del moto browniano

2.2.1 Difficoltà ad usare il MB come integratore

Nei capitoli successivi saremo interessati ad integrali del tipo

$$\int_0^t X_s dB_s$$

dove B è un moto browniano. Come potrebbe essere definito questo integrale? Elenchiamo alcune possibilità (nessuna funzionante), in ordine di difficoltà.

1. Se le traiettorie del MB fossero funzioni differenziabili, e la derivata $\frac{dB_s(\omega)}{ds}$ avesse un minimo di integrabilità, potremmo definire l'integrale precedente in modo ovvio, come $\int_0^t X_s(\omega) \frac{dB_s(\omega)}{ds} ds$. Invece il MB ha traiettorie che, q.c., non sono differenziabili in alcun punto. Questa proprietà è così speciale che la dimostreremo in dettaglio nel seguito del capitolo.

2. Un'apparente ancora di salvezza sembra venire da un ragionamento semplicissimo. Se X è sufficientemente regolare, possiamo definire

$$\int_0^t X_s(\omega)\,dB_s(\omega) = [X_s(\omega)\,B_s(\omega)]_{s=0}^{s=t} - \int_0^t B_s(\omega)\,dX_s(\omega)$$

cioè scaricare il problema della regolarità su X. Basta ad esempio che X abbia traiettorie a variazione limitata. Ma questo, pur vero, è troppo restrittivo per i nostri scopi. Infatti, il caso per noi di maggior interesse sarà quello di integrali del tipo

$$\int_0^t f(X_s)\,dB_s$$

dove f è regolare ed X risolve un'equazione differenziale contenente B, e fatta in modo che la (poca) regolarità delle traiettorie di X sia simile a quella di B (si immagini ad esempio un'equazione della forma $X_t = B_t + A_t$, dove A è un processo con traiettorie molto regolari). Allora X non ha traiettorie a variazione limitata, come non le ha B, e quindi il trucchetto di integrare per parti non si può applicare.

3. Se le traiettorie di B fossero funzioni a variazione limitata (ricorderemo tra poco cosa questo significhi), allora potremmo definire $\int_0^t X_s(\omega)\,dB_s(\omega)$ per q.o. ω, per processi X opportuni, come *integrale di Lebesgue-Stieltjes*. Ma questo non è vero, le traiettorie di B non sono a variazione limitata, come dimostreremo poco sotto nel capitolo. La derivata delle traiettorie browniane $dB_s(\omega)/ds$ non è una misura, è solo una distribuzione di classe meno regolare e non utile per effettuare integrazioni.

4. Esiste poi un concetto di *integrale secondo Young* che si applicherebbe se $B_\cdot(\omega)$ fosse di classe $C^\alpha([a,b])$ con $\alpha > \frac{1}{2}$ (funzioni hölderiane di ordine α), cosa non vera, come vedremo sotto. E' una teoria utile ad esempio per integrare rispetto a varianti del MB come il moto browniano frazionario con indice di Hurst $H > \frac{1}{2}$.

5. In conclusione, serve un concetto nuovo di integrale, detto *integrale secondo Itô*, di cui ci occuperemo in un capitolo successivo. Va detto che in anni recenti è stata sviluppata la teoria dell'integrazione dei *rough paths*, che, entro certi limiti e con molte precisazioni piuttosto complesse, permette di integrare rispetto a traiettorie browniane, senza bisogno della teoria di Itô. Sia per la complessità, sia per le limitazioni, non ha rimpiazzato la teoria di Itô, per gli usuali scopi di questa teoria. E' invece risultata utile per scopi particolari ed è attualmente in grande evoluzione.

2.2.2 Commenti sulle funzioni a variazione limitata

Consideriamo funzioni $f : [a,b] \to \mathbb{R}$. Usiamo le seguenti notazioni: indichiamo una generica partizione di $[a,b]$ con $\pi = \{a \leq t_1 < ... < t_{n_\pi} \leq b\}$ e scriviamo $|\pi| = \max_{i=1,...,n_\pi-1} |t_{i+1} - t_i|$.

Una funzione $f : [a, b] \to \mathbb{R}$ si dice a variazione limitata, e scriviamo $f \in BV[a, b]$, se

$$\sup_{\pi} \sum_{i=1}^{n_\pi - 1} |f(t_{i+1}) - f(t_i)| < \infty.$$

Se $f \in BV[a, b]$ e $g \in C([a, b])$ allora, presa una successione $\{\pi_k\}_{k\in\mathbb{N}}$ di partizioni di $[a, b]$, con $\lim_{k\to\infty} |\pi_k| = 0$, della forma $\pi_k = \left\{a \le t_1^k < ... < t_{n_{\pi_k}}^k \le b\right\}$, esiste il limite

$$\lim_{|\pi|\to 0} \sum_{i=1}^{n_{\pi_k}-1} g(\xi_i)\left(f\left(t_{i+1}^k\right) - f\left(t_i^k\right)\right)$$

per ogni scelta di $\xi_i \in \left[t_i^k, t_{i+1}^k\right]$, indicato con $\int_a^b g df$ e chiamato integrale di Riemann-Stieltjes. Tale integrale si estende poi ad altre funzioni, in particolare nella direzione dell'integrazione di Lebesgue, ma sostanzialmente serve sempre l'ipotesi che f sia $BV[a, b]$.

Useremo poco più tardi la seguente proposizione. A parole, potremmo dire che se la variazione quadratica è non nulla, allora la funzione non è a variazione limitata. Indichiamo inoltre con $C^\alpha([a, b])$ lo spazio delle funzioni hölderiane di ordine $\alpha \in (0, 1)$.

Proposizione 22 *Sia f una funzione continua e sia $\{\pi_k\}_{k\in\mathbb{N}}$ una successione di partizioni di $[a, b]$, con $\lim_{k\to\infty} |\pi_k| = 0$, della forma $\pi_k = \left\{a \le t_1^k < ... < t_{n_{\pi_k}}^k \le b\right\}$. Se*

$$\lim_{k\to\infty} \sum_{i=1}^{n_{\pi_k}-1} \left|f\left(t_{i+1}^k\right) - f\left(t_i^k\right)\right|^2 > 0$$

allora f non è di classe $BV[a, b]$. Non è nemmeno di classe $C^\alpha([a, b])$ per qualche $\alpha \in \left(\frac{1}{2}, 1\right)$. In altre parole, se f è una funzione continua BV, oppure $C^\alpha([a, b])$ per qualche $\alpha \in \left(\frac{1}{2}, 1\right)$, allora $\lim_{k\to\infty} \sum_{i=1}^{n_{\pi_k}-1} \left|f\left(t_{i+1}^k\right) - f\left(t_i^k\right)\right|^2 = 0$.

Proof. Osserviamo che vale

$$\begin{aligned}
\sum_{i=1}^{n_{\pi_k}-1} \left|f\left(t_{i+1}^k\right) - f\left(t_i^k\right)\right|^2 &\le \sum_{i=1}^{n_{\pi_k}-1} \left|f\left(t_{i+1}^k\right) - f\left(t_i^k\right)\right| \left(\max_{i=1,\dots,n_{\pi_k}-1} \left|f\left(t_{i+1}^k\right) - f\left(t_i^k\right)\right|\right) \\
&\le \omega(f, \pi_k) \cdot \sup_{\pi} \sum_{i=1}^{n_\pi - 1} |f(t_{i+1}) - f(t_i)|
\end{aligned}$$

dove $\omega(f, \pi_k)$, oscillazione di f su π_k, è definita da

$$\omega(f, \pi_k) = \max_{i=1,\dots,n_{\pi_k}-1} \left|f\left(t_{i+1}^k\right) - f\left(t_i^k\right)\right|.$$

Vale $\lim_{k\to\infty}\omega(f,\pi_k)=0$, essendo f uniformemente continua su $[a,b]$. Se per assurdo fosse $f\in BV[a,b]$, allora dedurremmo $\lim_{k\to\infty}\sum_{i=1}^{n_{\pi_k}-1}\left|f\left(t_{i+1}^k\right)-f\left(t_i^k\right)\right|^2=0$, in contraddizione con l'ipotesi. Quindi $f\notin BV[a,b]$.

Se per assurdo fosse $f\in C^\alpha([a,b])$ per un qualche $\alpha\in\left(\frac{1}{2},1\right)$, cioè se relativamente ad una costante $C>0$ valesse

$$|f(t)-f(s)|\leq C|t-s|^\alpha,\qquad t,s\in[a,b]$$

allora

$$\begin{aligned}\sum_{i=1}^{n_{\pi_k}-1}\left|f\left(t_{i+1}^k\right)-f\left(t_i^k\right)\right|^2 &\leq C\sum_{i=1}^{n_{\pi_k}-1}\left|t_{i+1}^k-t_i^k\right|^{2\alpha}=C\sum_{i=1}^{n_{\pi_k}-1}\left|t_{i+1}^k-t_i^k\right|\left|t_{i+1}^k-t_i^k\right|^{2\alpha-1}\\ &\leq C\sum_{i=1}^{n_{\pi_k}-1}\left|t_{i+1}^k-t_i^k\right||\pi_k|^{2\alpha-1}=C(b-a)|\pi_k|^{2\alpha-1}\end{aligned}$$

da cui dedurremmo dedurremmo $\lim_{k\to\infty}\sum_{i=1}^{n_{\pi_k}-1}\left|f\left(t_{i+1}^k\right)-f\left(t_i^k\right)\right|^2=0$, in contraddizione con l'ipotesi. La dimostrazione è completa. ■

2.2.3 Variazione quadratica del MB e variazione non limitata

Teorema 6 *Sia B un moto browniano (continuo) e sia $T>0$. Sia $\{\pi_k\}_{k\in\mathbb{N}}$ una successione di partizioni di $[0,T]$, con $\lim_{k\to\infty}|\pi_k|=0$, della forma $\pi_k=\left\{0=t_1^k<...<t_{n_{\pi_k}}^k=T\right\}$. Allora*

$$\lim_{k\to\infty}E\left[\left(\sum_{i=1}^{n_{\pi_k}-1}\left|B_{t_{i+1}^k}-B_{t_i^k}\right|^2-T\right)^2\right]=0$$

e $\sum_{i=1}^{n_{\pi_k}-1}\left|B_{t_{i+1}^k}-B_{t_i^k}\right|^2$ converge a T in probabilità. Ne discende che, q.c., le traiettorie browniane non sono di classe $BV[0,T]$ e non appartengono ad alcun $C^\alpha([a,b])$ per $\alpha\in\left(\frac{1}{2},1\right)$.

Proof. La v.a. $\frac{B_{t_{i+1}^k}-B_{t_i^k}}{\sqrt{t_{i+1}^k-t_i^k}}$ è $N(0,1)$, quindi la v.a. $Y_{k,i}=\frac{\left|B_{t_{i+1}^k}-B_{t_i^k}\right|^2}{t_{i+1}^k-t_i^k}$ ha legge indipendente da k ed i. Inoltre, fissato k, le v.a. $Y_{k,1},...,Y_{k,n_{\pi_k}-1}$ sono indipendenti. Pertanto

$$\begin{aligned}E\left[\left(\sum_{i=1}^{n_{\pi_k}-1}\left|B_{t_{i+1}^k}-B_{t_i^k}\right|^2-T\right)^2\right] &= E\left[\left(\sum_{i=1}^{n_{\pi_k}-1}Y_{k,i}\left(t_{i+1}^k-t_i^k\right)-T\right)^2\right]\\ &= E\left[\left(\sum_{i=1}^{n_{\pi_k}-1}(Y_{k,i}-1)\left(t_{i+1}^k-t_i^k\right)\right)^2\right]\end{aligned}$$

$$
\begin{aligned}
&= E\left[\sum_{i,j=1}^{n_{\pi_k}-1}(Y_{k,i}-1)(Y_{k,j}-1)\left(t_{i+1}^k-t_i^k\right)\left(t_{j+1}^k-t_j^k\right)\right] \\
&= \sum_{i,j=1}^{n_{\pi_k}-1}\left(t_{i+1}^k-t_i^k\right)\left(t_{j+1}^k-t_j^k\right)E\left[(Y_{k,i}-1)(Y_{k,j}-1)\right].
\end{aligned}
$$

I termini con $i \neq j$ sono nulli: per l'indipendenza

$$E\left[(Y_{k,i}-1)(Y_{k,j}-1)\right] = E\left[Y_{k,i}-1\right]E\left[Y_{k,j}-1\right]$$

e ciascun fattore è nullo:

$$E\left[Y_{k,i}-1\right] = E\left[Y_{k,i}\right]-1 = \frac{1}{t_{i+1}^k-t_i^k}E\left[\left|B_{t_{i+1}^k}-B_{t_i^k}\right|^2\right]-1=0.$$

Quindi

$$
\begin{aligned}
E\left[\left(\sum_{i=1}^{n_{\pi_k}-1}\left|B_{t_{i+1}^k}-B_{t_i^k}\right|^2-T\right)^2\right] &= \sum_{i,j=1}^{n_{\pi_k}-1}\left(t_{i+1}^k-t_i^k\right)^2E\left[(Y_{k,i}-1)^2\right] \\
&= C\sum_{i,j=1}^{n_{\pi_k}-1}\left(t_{i+1}^k-t_i^k\right)^2
\end{aligned}
$$

dove $C = E\left[(Y_{k,1}-1)^2\right]$ (finita perché le v.a. gaussiane hanno momenti di ogni ordine finiti)

$$\leq C\left|\pi_k\right|\sum_{i,j=1}^{n_{\pi_k}-1}\left(t_{i+1}^k-t_i^k\right) = C\left|\pi_k\right|T$$

da cui la convergenza in media quadratica, e da essa la convergenza in probabilità per il lemma di Chebyshev. Infine, da queste convergenze discende la convergenza quasi certa di una sottosuccessione; ma allora vale (per le partizioni di quella sottosuccessione) l'ipotesi della Proposizione 22, da cui il fatto che le traiettorie non sono $BV[0,T]$, con probabilità uno. ■

Esercizio 10 *Mostrare che lungo le partizioni diadiche vale la convergenza quasi certa di* $\sum_{i=1}^{n_{\pi_k}-1}\left|B_{t_{i+1}^k}-B_{t_i^k}\right|^2$ *a* T.

2.2.4 Non differenziabilità

Per capire la dimostrazione della non differenziabilità, affrontiamo prima un problema più semplice: la non differenziabilità in un punto prefissato.

Proposizione 23 *Dato $t_0 \geq 0$, q.c. le traiettorie di un moto browniano non sono differenziabili in t_0. Detto più formalmente,*

$$P\left(\omega \in \Omega : \lim_{h \to 0} \frac{B_{t_0+h}(\omega) - B_{t_0}(\omega)}{h} \text{ esiste finito}\right) = 0.$$

Proof. Per ogni $R > 0$, sia N_R l'evento

$$N_R = \left\{\omega \in \Omega : l(\omega) := \lim_{h \to 0} \frac{B_{t_0+h}(\omega) - B_{t_0}(\omega)}{h} \text{ esiste, } |l(\omega)| \leq R\right\}.$$

Vogliamo dimostrare che $P(N_R) = 0$ (dall'arbitrarietà di R segue la tesi della proposizione). Come osservazione generale, se per un evento N abbiamo $N \subset \cup_{k \geq 0} \cap_{n \geq k} A_n$ e vale $\lim_{n\to\infty} P(A_n) = 0$, allora $P(N) = 0$. Infatti, $N \subset \cup_{k\geq 0} N_k$ con $N_k = \cap_{n\geq k} A_n$; quindi $N_k \subset A_n$ per ogni $n \geq k$, da cui segue $P(N_k) \leq \inf_{n \geq k} P(A_n) = 0$; questo implica infine $P(N) = 0$.

Premettiamo un facile fatto di analisi. Se $l(\omega) := \lim_{h\to 0} \frac{B_{t_0+h}(\omega) - B_{t_0}(\omega)}{h}$ esiste, con $|l(\omega)| \leq R$, allora esiste $\delta(\omega) > 0$ tale che per ogni h con $|h| \leq \delta(\omega)$ vale

$$|B_{t_0+h}(\omega) - B_{t_0}(\omega)| \leq (R+1)|h|.$$

Infatti, per definizione di limite, (preso $\varepsilon = 1$) esiste $\delta(\omega) > 0$ tale che per ogni h con $|h| \leq \delta(\omega)$ vale $l(\omega) - 1 \leq \frac{B_{t_0+h}(\omega) - B_{t_0}(\omega)}{h} \leq l(\omega) + 1$, da cui $-R-1 \leq \frac{B_{t_0+h}(\omega) - B_{t_0}(\omega)}{h} \leq R+1$, da cui $|B_{t_0+h}(\omega) - B_{t_0}(\omega)| \leq (R+1)|h|$.

In particolare, esiste $k(\omega) > 0$ tale che per ogni n con $n \geq k(\omega)$ vale

$$\left|B_{t_0+\frac{1}{n}}(\omega) - B_{t_0}(\omega)\right| \leq \frac{R+1}{n}.$$

Quindi

$$\begin{aligned} N_R &\subset \cup_{k\geq 0} \cap_{n \geq k} A_{k,n}^{(R)} \\ A_n^{(R)} &: = \left\{\omega \in \Omega : \left|B_{t_0+\frac{1}{n}}(\omega) - B_{t_0}(\omega)\right| \leq \frac{R+1}{n}\right\}. \end{aligned}$$

Ma

$$\begin{aligned} P\left(A_n^{(R)}\right) &= P\left(\left|B_{t_0+\frac{1}{n}} - B_{t_0}\right| \leq \frac{R+1}{n}\right) \\ &= P\left(\left|\frac{B_{t_0+\frac{1}{n}} - B_{t_0}}{\sqrt{1/n}}\right| \leq \frac{R+1}{\sqrt{n}}\right) = P\left(|Z| \leq \frac{R+1}{\sqrt{n}}\right) \end{aligned}$$

dove Z è una $N(0,1)$. Si verifica facilmente che $\lim_{n\to\infty} P\left(A_n^{(R)}\right) = 0$, quindi $P(N_R) = 0$. La dimostrazione è completa. ■

Corollario 1 *Dato un insieme numerabile $\mathcal{D} \subset [0, \infty)$, q.c. le traiettorie di un moto browniano non sono differenziabili in alcun punto di $\mathcal{D}$:*

$$P\left(\omega \in \Omega : \textit{esiste } t_0 \in \mathcal{D} \textit{ tale che} \lim_{h \to 0} \frac{B_{t_0+h}(\omega) - B_{t_0}(\omega)}{h} \textit{ esiste finito}\right) = 0.$$

Proof. Detto N_{t_0} l'insieme di misura nulla della proposizione precedente e detto $N_{\mathcal{D}}$ quello di questo corollario, vale

$$N_{\mathcal{D}} = \cup_{t_0 \in \mathcal{D}} N_{t_0}$$

da cui discende subito la tesi. ■

Teorema 7 *Sia B un moto browniano. Allora*

$$P\left(\omega \in \Omega : \textit{esiste } t_0 \in [0, \infty) \textit{ tale che} \lim_{h \to 0} \frac{B_{t_0+h}(\omega) - B_{t_0}(\omega)}{h} \textit{ esiste finito}\right) = 0.$$

Proof. Dati $R > 0$, $k \in \mathbb{N}$, sia $N_{R,k}$ l'insieme degli $\omega \in \Omega$ tali che la funzione $t \longmapsto B_t(\omega)$ è differenziabile almeno in un punto $t_0 \in [k, k+1]$ con derivata in modulo non superiore ad R. Basta verificare che $P(N_{R,k}) = 0$ per ogni $R > 0$, $k \in \mathbb{N}$. Si capisce che se lo dimostriamo per $k = 0$, il caso generale sarà uguale. Quindi dimostriamo $P(N_R) = 0$ dove

$$N_R = \left\{\omega \in \Omega : \text{esiste } t_0 \in [0,1] \text{ tale che esiste } \lim_{h \to 0} \frac{B_{t_0+h}(\omega) - B_{t_0}(\omega)}{h} \in [-R, R]\right\}.$$

Volendo ricalcare la dimostrazione vista sopra, si deve cercare di esprimere N_R tramite unioni ed intersezioni numerabili riconducendosi ad insiemi più elementari aventi probabilità molto piccola. Ma qui la prima difficoltà è che la condizione "differenziabile almeno in un punto" coinvolge a priori l'unione più che numerabile di eventi. Cerchiamo allora di dedurre, dalla condizione che definisce N_R, una condizione esprimibile in modo numerabile.

Abbiamo già verificato nella dimostrazione della Proposizione 23 che, se $B_t(\omega)$ è differenziabile in t_0 con derivata in modulo non superiore ad R, allora esiste $\delta(\omega) > 0$ tale che $|B_{t_0+h}(\omega) - B_{t_0}(\omega)| \leq (R+1)h$ per ogni $h \in [0, \delta(\omega)]$. Allora, posto

$$A_{M,\delta} := \{\omega \in \Omega : \text{esiste } t_0 \in [0,1] \text{ tale che } |B_{t_0+h}(\omega) - B_{t_0}(\omega)| \leq Mh \text{ per ogni } h \in [0, \delta]\}$$

vale

$$N_R \subset \bigcup_{\delta > 0} A_{R+1,\delta}.$$

Ma vale anche $\bigcup_{\delta>0} A_{R+1,\delta} = \bigcup_{n \in \mathbb{N}} A_{R+1,1/n}$, quindi se dimostriamo che $P(A_{M,\delta}) = 0$ per ogni $M > 0$, $\delta > 0$, possiamo dedurre $P(N_R) = 0$ per ogni $R > 0$. Mostriamo quindi che $P(A_{M,\delta}) = 0$.

Per inquadrare i calcoli che seguono, può essere utile introdurre anche gli eventi (fissato $t_0 \in [0,1]$)

$$A_{M,\delta,t_0} := \{\omega \in \Omega : |B_{t_0+h}(\omega) - B_{t_0}(\omega)| \leq Mh \text{ per ogni } h \in [0,\delta]\}.$$

Vale $A_{M,\delta} = \bigcup_{t_0 \in [0,1]} A_{M,\delta,t_0}$ ma, come già osservato, si tratta di un'unione più che numerabile.

Tuttavia, per ogni n tale che $\frac{2}{2^n} < \delta$, e per ogni $t_0 \in [0,1]$, vale

$$A_{M,\delta,t_0} \subset D_{M,n,k(t_0)} \subset \bigcup_{k=0,\dots,2^n} D_{M,n,k}$$

dove

$$D_{M,n,k} := \left\{\omega \in \Omega : \left|B_{\frac{k+1}{2^n}}(\omega) - B_{\frac{k}{2^n}}(\omega)\right| \leq \frac{3M}{2^n}\right\}$$

e $k(t_0)$ è il valore in $\{0,\dots,2^n\}$ per cui $\frac{k(t_0)-1}{2^n} < t_0 \leq \frac{k(t_0)}{2^n}$. Infatti, se $\omega \in A_{M,\delta,t_0}$, vale $|B_{t_0+h}(\omega) - B_{t_0}(\omega)| \leq Mh$ per ogni $h \in [0,\delta]$; quindi, osservando che grazie alla condizione $\frac{2}{2^n} < \delta$ i numeri $\frac{k(t_0)}{2^n}$ e $\frac{k+1}{2^n}$ stanno in $[t_0, t_0+\delta]$, e che $\frac{k(t_0)}{2^n} - t_0 \leq \frac{1}{2^n}$ e $\frac{k(t_0)+1}{2^n} - t_0 \leq \frac{2}{2^n}$, vale

$$\left|B_{\frac{k(t_0)+1}{2^n}}(\omega) - B_{\frac{k(t_0)}{2^n}}(\omega)\right| \leq \left|B_{\frac{k(t_0)}{2^n}}(\omega) - B_{t_0}(\omega)\right| + \left|B_{\frac{k(t_0)+1}{2^n}}(\omega) - B_{t_0}(\omega)\right|$$

che per la condizione $\omega \in A_{M,\delta,t_0}$ risulta

$$\leq M\left(\left(\frac{k(t_0)}{2^n} - t_0\right) + \left(\frac{k(t_0)+1}{2^n} - t_0\right)\right) \leq \frac{3M}{2^n}.$$

L'inclusione $A_{M,\delta,t_0} \subset \bigcup_{k=0,\dots,2^n} D_{M,n,k}$ così dimostrata vale per ogni $t_0 \in [0,1]$ e l'evento di destra non dipende da t_0, quindi possiamo dire che $A_{M,\delta} \subset \bigcup_{k=0,\dots,2^n} D_{M,n,k}$. Siamo riusciti a tradurre l'unione più che numerabile $A_{M,\delta}$ in un'unione numerabile (con un'inclusione, non un'identità).

Purtroppo questa fatica non basta. Infatti

$$P\left(\bigcup_{k=0,\dots,2^n} D_{M,n,k}\right) \leq \sum_{k=0}^{2^n} P(D_{M,n,k}) \leq \sum_{k=0}^{2^n} C\sqrt{\frac{1}{2^n}}$$

che diverge, dove la disuguaglianza $P(D_{M,n,k}) \leq C\sqrt{\frac{1}{2^n}}$ si dimostra come descritto sotto. Bisogna allora introdurre l'ultimo elemento non banale della dimostrazione.

Per ogni n tale che $\frac{4}{2^n} < \delta$, e per ogni $t_0 \in [0,1]$, vale

$$A_{M,\delta,t_0} \subset \widetilde{D}_{M,n,k(t_0)} \subset \bigcup_{k=0,\dots,2^n} \widetilde{D}_{M,n,k}$$

dove

$$\widetilde{D}_{M,n,k} := \left\{ \left| B_{\frac{k+1}{2^n}} - B_{\frac{k}{2^n}} \right| \le \frac{3M}{2^n}, \left| B_{\frac{k+2}{2^n}} - B_{\frac{k+1}{2^n}} \right| \le \frac{5M}{2^n}, \left| B_{\frac{k+3}{2^n}} - B_{\frac{k+2}{2^n}} \right| \le \frac{7M}{2^n} \right\}.$$

Infatti, dalla condizione $|B_{t_0+h} - B_{t_0}| \le Mh$ per ogni $h \in [0,\delta]$, osservando che $\frac{k(t_0)}{2^n}, \frac{k(t_0)+1}{2^n}, \frac{k(t_0)+2}{2^n}, \frac{k(t_0)+3}{2^n}$ stanno in $[t_0, t_0+\delta]$, e che $\frac{k(t_0)+j}{2^n} - t_0 \le \frac{1+j}{2^n}$, vale $\left| B_{\frac{k(t_0)+1}{2^n}} - B_{\frac{k(t_0)}{2^n}} \right| \le \frac{3M}{2^n}$ (già dimostrato),

$$\begin{aligned} \left| B_{\frac{k(t_0)+2}{2^n}} - B_{\frac{k(t_0)+1}{2^n}} \right| &\le \left| B_{\frac{k(t_0)+1}{2^n}} - B_{t_0} \right| + \left| B_{\frac{k(t_0)+2}{2^n}} - B_{t_0} \right| \\ &\le M \left(\left(\frac{k(t_0)+1}{2^n} - t_0 \right) + \left(\frac{k(t_0)+2}{2^n} - t_0 \right) \right) \le \frac{5M}{2^n} \end{aligned}$$

ed analogamente $\left| B_{\frac{k(t_0)+3}{2^n}} - B_{\frac{k(t_0)+2}{2^n}} \right| \le \frac{7M}{2^n}$. Quindi $A_{M,\delta} \subset \bigcup_{k=0,\dots,2^n} \widetilde{D}_{M,n,k}$, per cui

$$P(A_{M,\delta}) \le \sum_{k=0}^{2^n} P\left(\widetilde{D}_{M,n,k}\right).$$

Ora, per l'indipendenza degli incrementi vale

$$\begin{aligned} & P\left(\widetilde{D}_{M,n,k}\right) \le P\left(\left| B_{\frac{k+1}{2^n}} - B_{\frac{k}{2^n}} \right| \le \frac{7M}{2^n} \right) P\left(\left| B_{\frac{k+2}{2^n}} - B_{\frac{k+1}{2^n}} \right| \le \frac{7M}{2^n} \right) P\left(\left| B_{\frac{k+3}{2^n}} - B_{\frac{k+2}{2^n}} \right| \le \frac{7M}{2^n} \right) \\ = \;& P\left(|Z| \le 7M \sqrt{\frac{1}{2^n}} \right)^3 \end{aligned}$$

dove Z è una $N(0,1)$. Quindi

$$P(A_{M,\delta}) \le \sum_{k=0}^{2^n} P\left(|Z| \le 7M\sqrt{\frac{1}{2^n}} \right)^3 = (2^n+1)\, P\left(|Z| \le 7M\sqrt{\frac{1}{2^n}} \right)^3.$$

Ricordiamo che questa disuguaglianza vale per ogni n tale che $\frac{4}{2^n} < \delta$. Usando la densità gaussiana si verifica subito che

$$\lim_{n\to\infty} (2^n+1)\, P\left(|Z| \le 7M\sqrt{\frac{1}{2^n}} \right)^3 \le C \lim_{n\to\infty} (2^n+1) \left(\frac{1}{2^n} \right)^{3/2} = 0$$

dove $C = \left(\frac{2}{\pi}\right)^{3/2}(7M)^3$, in quanto $P(|Z| \leq r) \leq \sqrt{\frac{2}{\pi}}r$ per ogni $r > 0$ (basta osservare che $P(|Z| \leq r) = \int_{-r}^{r} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \leq \int_{-r}^{r} \frac{1}{\sqrt{2\pi}} dx$). Quindi $P(A_{M,\delta}) = 0$. La dimostrazione è completa. ∎

Osservazione 8 *Col linguaggio della misura di Wiener, l'enunciato precedente fa un certo effetto: l'insieme delle funzioni continue che sono derivabili almeno in un punto, ha misura di Wiener nulla, è un insieme trascurabile. Se immaginassimo la misura di Wiener alla stregua della misura di Lebesgue negli spazi* $\mathbb{R}^n$*, questo enunciato direbbe che le funzioni differenziabili in almeno un punto sono un insieme davvero piccolissimo rispetto alla totalità delle funzioni continue!*

Capitolo 3

Martingale

3.1 Definizioni

Sia $M = (M_t)_{t\geq 0}$ un processo stocastico a valori reali su uno spazio $(\Omega, \mathcal{F}, P)$ ed $(\mathcal{F}_t)_{t\geq 0}$ una filtrazione. Supponiamo che sia $E[|M_t|] < \infty$ per ogni $t \geq 0$ e che sia adattato. Diciamo che M è una *martingala* rispetto alla filtrazione $(\mathcal{F}_t)_{t\geq 0}$ se per ogni $t > s \geq 0$ vale [1]

$$E[M_t|\mathcal{F}_s] = M_s.$$

Se invece vale $E[M_t|\mathcal{F}_s] \geq M_s$ parliamo di *submartingala*, mentre se vale $E[M_t|\mathcal{F}_s] \leq M_s$ parliamo di *supermartingala*[2].

Calcolando il valore atteso di ambo i membri della condizione di martingala (risp. submartigala) si vede che $E[M_t] = E[M_s]$ (risp. $E[M_t] \geq E[M_s]$) quindi la funzione valore atteso $m(t)$ è costante (risp. crescente). Una martingala in media e costante (risp. una supermartingala in media cresce).

Le stesse definizioni si applicano al caso di un processo a tempo discreto $M = (M_n)_{n\in\mathbb{N}}$. In questo caso è sufficiente che valga la condizione quando i tempi s e t sono consecutivi; vediamolo nel caso delle martingale. Se vale

$$E[M_{n+1}|\mathcal{F}_n] = M_n$$

per ogni $n \in \mathbb{N}$, allora vale

$$E[M_{n+2}|\mathcal{F}_n] = E[E[M_{n+2}|\mathcal{F}_{n+1}]|\mathcal{F}_n] = E[M_{n+1}|\mathcal{F}_n] = M_n$$

[1]Ricordiamo che ogni identità coinvolgente una speranza condizionale va intesa quasi certamente o come identità tra classi di equivalenza

[2]I termini sub e super martinagala sono difficili da interiorizzare perché sembrano opposti all'intuizione: il prefisso sub sembra richiamare la decrescenza, mentre la miglior previsione del futuro (X_t) al momento attuale $(\mathcal{F}_s)$ è maggiore del presente (X_s), cioè in un certo senso il processo ha una tendenza a crescere. Se però si associa la parola coda (temporalmente intesa come il passato) a quella di martingala, si può pensare che una submartingala abbia la coda in ribasso e questo può aiutare la memoria.

e quindi iterativamente si verifica che vale $E[M_m|\mathcal{F}_n] = M_n$ per ogni $m > n \in \mathbb{N}$.

La definizione di martingala equivale alla seguente: i) $(M_t)_{t\geq 0}$ è integrabile ed adattato a $(\mathcal{F}_t)_{t\geq 0}$, ii) per ogni $t > s \geq 0$ vale

$$E[M_t - M_s|\mathcal{F}_s] = 0.$$

Infatti, per l'adattamento, vale $E[M_s|\mathcal{F}_s] = X_s$. Forse da questo punto di vista non stupisce la seguente proposizione, che stabilisce la vicinanza del concetto di martingala col concetto di processo a incrementi indipendenti e centrati (valore atteso nullo). Premettiamo due definizioni.

Un processo $X = (X_t)_{t\geq 0}$ si dice a *incrementi indipendenti rispetto ad una filtrazione* $(\mathcal{F}_t)_{t\geq 0}$ se è adattato e $X_t - X_s$ è indipendente da $\mathcal{F}_s$ per ogni $t \geq s \geq 0$. [La condizione di indipendenza, quando $\mathcal{F}_t = \sigma(X_u; u \in [0,t])$, equivale alla seguente: per ogni $n \in \mathbb{N}$ ed ogni $t_n \geq t_{n-1} \geq ... \geq t_1 \geq 0$ le v.a. $X_{t_n} - X_{t_{n-1}}, ..., X_{t_2} - X_{t_1}, X_{t_1}$ sono indipendenti. La dimostrazione di questa equivalenza è identica a quella svolta ne caso del moto browniano, Proposizione 17]. Diciamo che un processo X, tale che $E[X_t^2] < \infty$ per ogni $t \geq 0$, ha *incrementi scorrelati* se

$$Cov(X_t - X_s, X_u - X_v) = 0$$

per ogni $t \geq s \geq u \geq v \geq 0$.

Proposizione 24 *Se X è un processo, con $E[|X_t|] < \infty$ per ogni $t \geq 0$, ha incrementi centrati e indipendenti rispetto a $(\mathcal{F}_t)_{t\geq 0}$, allora è una martingala rispetto a $(\mathcal{F}_t)_{t\geq 0}$.*

Se X è una martingala rispetto ad una filtrazione $(\mathcal{F}_t)_{t\geq 0}$, con $E[X_t^2] < \infty$ per ogni $t \geq 0$, allora è un processo a incrementi scorrelati e centrati. Se inoltre è un processo gaussiano, allora ha incrementi indipendenti rispetto alla filtrazione generata dal processo $(\mathcal{F}_t^{00})_{t\geq 0}$.

Proof. La prima affermazione è immediata: presi $t \geq s \geq 0$, per l'indipendenza da $\mathcal{F}_s$ ed il valore atteso nullo dell'incremento $X_t - X_s$, vale

$$E[X_t - X_s|\mathcal{F}_s] = E[X_t - X_s] = 0.$$

Vediamo la seconda. Il valore atteso nullo si vede così: per la proprietà $E[X_t - X_s|\mathcal{F}_s] = 0$ ed una proprietà dei valori attesi condizionali (analogo della formula di fattorizzazione), vale

$$0 = E[E[X_t - X_s|\mathcal{F}_s]] = E[X_t - X_s].$$

Mostriamo che gli incrementi sono scorrelati. Presi $t \geq s \geq u \geq v \geq 0$, ed avendo già stabilito che il valore atteso di ogni incremento è nullo, vale (usando varie proprietà della speranza condizionale)

$$\begin{aligned} Cov(X_t - X_s, X_u - X_v) &= E[(X_t - X_s)(X_u - X_v)] = E[E[(X_t - X_s)(X_u - X_v)|\mathcal{F}_s]] \\ &= E[(X_u - X_v) E[(X_t - X_s)|\mathcal{F}_s]] = 0, \end{aligned}$$

l'ultimo passaggio essendo valido in quanto $E\left[\left(X_t - X_s\right)|\mathcal{F}_s\right] = 0$. Infine, se il processo è pure gaussiano, presi $t_n > t_{n-1} > ... > t_1 \geq 0$, il vettore $\left(X_{t_n} - X_{t_{n-1}}, ..., X_{t_2} - X_{t_1}\right)$ è gaussiano ed ha matrice di covarianza diagonale, per cui ha componenti indipendenti. Da questo discende che, per ogni $t \geq s \geq 0$, $X_t - X_s$ è indipendente da $\mathcal{F}_s^{00}$ (lo abbiamo già dimostrato in occasione delle riformulazioni della definizione di moto browniano. La dimostrazione è completa. ■

Il secondo esempio illustrato nella prossima sezione è una martingala ma non ha incrementi indipendenti (solo scorrelati). Quindi c'è spazio tra queste due categorie apparentemente molto vicine (processi a incrementi indipendenti e scorrelati) e le martingale si collocano in tale spazio.

3.1.1 Esempi

Vediamo due esempi fondamentali. Per noi le martingale saranno rilevanti principalmente per due ragioni (collegate): il moto browniano è una martingala e lo sono (sotto opportune ipotesi) gli integrali stocastici rispetto al moto browniano. E' utilissimo vedere poi numerosi altri tipi di esempi, soprattutto a tempo discreto, anche più elementari di questi, ma li rimandiamo agli esercizi.

La teoria delle martingale è uno snodo unico nel calcolo stocastico. La proprietà di martingala è tutto sommato apparentemente scarna, ed ha invece incredibili conseguenze. Allora è importante sapere che i processi che a noi servono maggiormente sono martingale (o sono collegati a martingale, vedi il teorema di decomposizione di Doob). Questo motiva i due esempi seguenti, anche se il secondo può apparire un po' tecnico.

Esempio 1 *Sia $B = (B_t)_{t\geq 0}$ un moto browniano rispetto ad una filtrazione $(\mathcal{F}_t)_{t\geq 0}$. Allora B è una martingala rispetto a $(\mathcal{F}_t)_{t\geq 0}$, in quanto è un processo centrato a incrementi indipendenti (Proposizione 24).*

Veniamo al secondo esempio. Sia $B = (B_t)_{t\geq 0}$ un moto browniano rispetto ad una filtrazione $(\mathcal{F}_t)_{t\geq 0}$ e sia $X = (X_t)_{t\geq 0}$ un processo adattato a $(\mathcal{F}_t)_{t\geq 0}$ ma costante a tratti relativamente ad una sequenza di tempi $t_n \geq t_{n-1} \geq ... \geq t_1 = 0$, cioè della forma

$$X_t = \sum_{i=1}^{n-1} X_{t_i} 1_{[t_i, t_{i+1})}(t), \qquad X_{t_i} \text{ misurabile rispetto a } \mathcal{F}_{t_i},\ E\left[X_{t_i}^2\right] < \infty.$$

Definiamo, per questo particolare processo[3], per ogni $b \geq a \geq 0$,

$$\int_a^b X_s dB_s = \sum_{i=1}^{n-1} X_{t_i}\left(B_{(a\vee t_{i+1})\wedge b} - B_{(a\vee t_i)\wedge b}\right).$$

[3]Ora ragioniamo su un processo X fissato, di questo tipo. Se si considera invece la classe di tutti questi processi, detti *elementari*, introducendo il simbolo $\int_a^b X_s dB_s$ si deve verificare che la definizione non dipende dalla rappresentazione di X, che non è univoca.

Le espressioni $(a \vee t_{i+1}) \wedge b$ e $(a \vee t_i) \wedge b$ sono un po' laboriose da interpretare, per cui può convenire l'uso della seguente riscrittura. Osserviamo che, fissati $b \geq a \geq 0$, si può arricchire la sequenza di tempi considerando la nuova sequenza $t'_{n+2} \geq t'_{n-1} \geq ... \geq t'_1 = 0$ formata dai tempi t_i e da a, b; poi si può scrivere X_t nella forma $X_t = \sum_{i=1}^{n+1} X'_{t'_i} 1_{[t'_i, t'_{i+1})}(t)$ con $X'_{t'_i}$ misurabile rispetto a $\mathcal{F}_{t'_i}$ così definiti: se $t'_i = t_j \in \{t_1, ..., t_n\}$, poniamo $X'_{t'_i} = X_{t_j}$; se $t'_i = a$ (risp. $t'_i = b$) e $t_j \in \{t_1, ..., t_n\}$ è il più grande con la proprietà $t_j \leq a$ (risp. $t_j \leq b$), allora poniamo $X'_{t'_i} = X_{t_j}$. A questo punto

$$\int_a^b X_s dB_s = \sum_{a \leq t'_i \leq t'_{i+1} \leq b} X_{t'_i} \left(B_{t'_{i+1}} - B_{t'_i} \right).$$

Naturalmente i punti t'_i dipendono da a, b. Si può verificare che vale $\int_a^b X_s dB_s + \int_b^c X_s dB_s = \int_a^c X_s dB_s$ per ogni $c \geq b \geq a \geq 0$; se ci si raffigura la successione di tempi t_i ed i tempi $c \geq b \geq a \geq 0$, l'idea è evidente, anche se noiosa da scrivere.

Osservazione 9 *L'integrale stocastico ora introdotto ha una naturale interpretazione in finanza matematica. Supponiamo di considerare un titolo finanziario (azioni, titoli di stato,...) il cui prezzo cambia nel tempo. Indichiamo con S_t il prezzo al tempo t. Immaginiamo poi una partizione temporale $t_n \geq t_{n-1} \geq ... \geq t_1 = 0$ corrispondente agli istanti in cui noi eseguiamo operazioni finanziarie. All'istante t_i acquistiamo una quantità X_{t_i} del titolo finanziario suddetto. Paghiamo quindi $X_{t_i} \cdot S_{t_i}$. Al tempo $X_{t_{i+1}}$ vendiamo la quantità acquistata X_{t_i}, incassando $X_{t_i} \cdot S_{t_{i+1}}$. Complessivamente, con questa singola operazione abbiamo guadagnato (o perso, dipende dal segno) $X_{t_i} \cdot \left(S_{t_{i+1}} - S_{t_i}\right)$. Supponiamo di ripetere l'operazione su ogni intervallino. Il guadagno complessivo sarà*

$$\sum_{i=1}^{n} X_{t_i} \left(S_{t_{i+1}} - S_{t_i} \right).$$

Vediamo quindi che integrali del tipo visto sopra rappresentano il guadagno complessivo in un certo periodo di tempo. Naturalmente, sopra abbiamo definito questo integrale nel caso in cui sia $S = B$, ma tale operazione si potrà generalizzare ad altri casi, una volta capita nel caso browniano.

Poniamo $M_t = \int_0^t X_s dB_s$.

Proposizione 25 $M = (M_t)_{t \geq 0}$ *è una martingala rispetto a* $(\mathcal{F}_t)_{t \geq 0}$.

Proof. Intanto

$$E[|M_t|] \leq \sum_{i=1}^{n-1} E\left[|X_{t_i}| \left| B_{t_{i+1} \wedge t} - B_{t_i \wedge t} \right| \right]$$

e $E\left[|X_{t_i}|\left|B_{t_{i+1}\wedge t}-B_{t_i\wedge t}\right|\right]<\infty$ per ogni i, grazie alla disuguaglianza di Hölder. Poi, $M_t=\int_0^t X_u dB_u$ è $\mathcal{F}_t$-misurabile, essendo somma e prodotto di variabili tutte $\mathcal{F}_t$-misurabili. Resta allora da dimostrare che $E\left[M_t-M_s|\mathcal{F}_s\right]=0$, ovvero che $E\left[\int_s^t X_u dB_u|\mathcal{F}_s\right]=0$. Si noti che l'incremento $\int_s^t X_u dB_u$ non è indipendente da $\mathcal{F}_s$: contiene (attraverso X) alcune delle v.a. X_{t_i}, che non sono supposte indipendenti da $\mathcal{F}_s$. Riscriviamo $\int_s^t X_u dB_u$ come detto sopra nella forma

$$\int_s^t X_u dB_u = \sum_{s\leq t_i'\leq t_{i+1}'\leq t} X_{t_i'}\left(B_{t_{i+1}'}-B_{t_i'}\right).$$

Vale

$$E\left[\int_s^t X_u dB_u|\mathcal{F}_s\right] = \sum_{s\leq t_i'\leq t_{i+1}'\leq t} E\left[X_{t_i'}\left(B_{t_{i+1}'}-B_{t_i'}\right)|\mathcal{F}_s\right].$$

Mostriamo che ciascun termine $E\left[X_{t_i'}\left(B_{t_{i+1}'}-B_{t_i'}\right)|\mathcal{F}_s\right]$ è nullo, concludendo così la dimostrazione. Vale

$$E\left[X_{t_i'}\left(B_{t_{i+1}'}-B_{t_i'}\right)|\mathcal{F}_s\right] = E\left[E\left[X_{t_i'}\left(B_{t_{i+1}'}-B_{t_i'}\right)|\mathcal{F}_{t_i'}\right]|\mathcal{F}_s\right] = E\left[X_{t_i'}E\left[B_{t_{i+1}'}-B_{t_i'}\right]|\mathcal{F}_s\right]=0$$

avendo usato varie proprietà della speranza condizionale, l'inclusione $\mathcal{F}_s\subset\mathcal{F}_{t_i'}$ (ricordiamo che nella somma precedente prendiamo solo i $t_i'\geq s$), l'indipendenza di $B_{t_{i+1}'}-B_{t_i'}$ da $\mathcal{F}_{t_i'}$, il fatto che $X_{t_i'}$ è $\mathcal{F}_{t_i'}$-misurabile, ed infine la proprietà $E\left[B_{t_{i+1}'}-B_{t_i'}\right]=0$. La dimostrazione è completa. ■

Osservazione 10 *Come già osservato, abbiamo così costruito un'ampia classe di esempi che stanno nell'intercapedine (apparentemente molto piccola) tra "scorrelazione" e "indipendenza" (intesa per gli incrementi temporali). Gli integrali stocastici hanno incrementi scorrelati ma non necessariamente indipendenti (perché il processo X li lega).*

3.1.2 Altri esempi e preliminari

Esempio 2 *Ricordiamo che un processo stocastico $(N_t)_{t\geq 0}$ si dice di Poisson di intensità $\lambda>0$, rispetto ad una filtrazione $(\mathcal{F}_t)_{t\geq 0}$, se $P(N_0=0)=1$, è cadlag e per ogni $t\geq s\geq 0$ l'incremento N_t-N_s è indipendente da $\mathcal{F}_s$ ed ha distribuzione di Poisson di parametro $\lambda(t-s)$:*

$$P(N_t-N_s=k)=e^{-\lambda(t-s)}\frac{(\lambda(t-s))^k}{k!},\qquad k\in\mathbb{N}.$$

Vale $E[N_t-N_s]=\lambda(t-s)$. Posto

$$M_t=N_t-\lambda t$$

il processo $(M_t)_{t\geq 0}$ *è una martingala rispetto a* $(\mathcal{F}_t)_{t\geq 0}$ *(a causa di questo, la funzione* λt *viene detta compensatore). Infatti la proprietà di indipendenza degli incrementi resta vera (non si modifica per addizione di una costante) e gli incrementi ora sono centrati:* $E[M_t - M_s] = 0$*. Quindi è una martingala per la Proposizione 24.*

Esempio 3 *Sia* X *una v.a. integrabile e* $(\mathcal{F}_t)_{t\geq 0}$ *una filtrazione. Allora*

$$M_t = E[X|\mathcal{F}_t]$$

è una martingala. Preliminarmente, osserviamo che $E[|M_t|] \leq E[E[|X||\mathcal{F}_t]] = E[|X|]$*, quindi* M_t *è integrabile. Inoltre,* $E[X|\mathcal{F}_t]$ *è* $\mathcal{F}_t$*-misurabile, quindi* M_t *è adattato. Vale poi*

$$E[M_t|\mathcal{F}_s] = E[E[X|\mathcal{F}_t]|\mathcal{F}_s] = E[X|\mathcal{F}_s] = M_s$$

dove abbiamo usato la proprietà di doppia proiezione della speranza condizionale. Quindi M *è una martingala. A volte è detta "martingala chiusa da* X*". Il processo* $E[X|\mathcal{F}_t]$ *ha un chiaro significato in varie applicazioni, ad esempio in finanza matematica:* X *rappresenta una grandezza incognita del futuro (ammontare di vendite del prossimo mese di luglio, valore di un titolo finanziario tra sei mesi, ecc.),* $\mathcal{F}_t$ *descrive il grado di informazione disponibile al tempo* t*,* $E[X|\mathcal{F}_t]$ *rappresenta la miglior stima di* X *fattibile al momento* t*. La formula* $E[M_t|\mathcal{F}_s] = M_s$ *esprime una coerenza tra le stime.*

Proposizione 26 *Se* $(M_t)_{t\geq 0}$ *è una martingala,* $\varphi : \mathbb{R} \to \mathbb{R}$ *una funzione convessa e* $E[|\varphi(M_t)|] < \infty$ *per ogni* $t \geq 0$ *allora* $(\varphi(M_t))_{t\geq 0}$ *è una submartingala. In particolare, lo sono* $|M_t|$ *e* M_t^2 *(nel secondo caso se* M *è di quadrato integrabile).*

Se $(M_t)_{t\geq 0}$ *è una submartingala,* $\varphi : \mathbb{R} \to \mathbb{R}$ *una funzione non decrescente convessa e* $E[|\varphi(M_t)|] < \infty$ *per ogni* $t \geq 0$ *allora* $(\varphi(M_t))_{t\geq 0}$ *è una submartingala.*

Proof. Grazie alla convessità, per ogni punto $x_0 \in \mathbb{R}$ esiste $m(x_0) \in \mathbb{R}$ tale che

$$\varphi(x) \geq \varphi(x_0) + m(x_0)(x - x_0), \qquad x \in \mathbb{R}.$$

Per il seguito ci serve che la funzione m sia misurabile; una tale funzione esiste, come si può verificare prendendo ad es. la derivata destra in x_0. Presi $t \geq s \geq 0$, vale allora

$$\varphi(M_t) \geq \varphi(M_s) + m(M_s)(M_t - M_s) \tag{3.1}$$

e quindi (per monotonia e linearità della speranza condizionale)

$$\begin{aligned} E[\varphi(M_t)|\mathcal{F}_s] &\geq E[\varphi(M_s)|\mathcal{F}_s] + E[m(M_s)(M_t - M_s)|\mathcal{F}_s] \\ &= \varphi(M_s) + m(M_s)E[(M_t - M_s)|\mathcal{F}_s] \end{aligned}$$

(per la misurabilità di $\varphi(M_s)$ e $m(M_s)$ rispetto a $\mathcal{F}_s$). A questo punto, se M è una martingala, $E[(M_t - M_s)|\mathcal{F}_s] = 0$; se invece è una submartingala e φ è non decrescente (quindi $m \geq 0$ in ogni punto), $E\left[(M_t - M_s)|\mathcal{F}_s\right] \geq 0$ e quindi $m(M_s) E\left[(M_t - M_s)|\mathcal{F}_s\right] \geq 0$. In entrambi i casi,

$$E\left[\varphi(M_t)|\mathcal{F}_s\right] \geq \varphi(M_s)$$

ovvero $\varphi(M_t)$ è una submartingala.

La dimostrazione ora data è però completa solo se $m(M_s)(M_t - M_s)$ è integrabile. La funzione m è limitata su intervalli limitati ma M_s non assume necessariamente valori in un intervallo limitato, per cui servirebbe una stima della crescita di m all'infinito ed altro ancora. Si risolve tutto con un troncamento: si introduce l'evento $A_R = \{|M_s| \leq R\}$, si moltiplica la disuguaglianza (3.1) per 1_{A_R} e si prosegue come sopra (con le stesse motivazioni dei passaggi, più il fatto che ora la v.a. $1_{A_R} m(M_s)$ è limitata):

$$\begin{aligned} E\left[1_{A_R}\varphi(M_t)|\mathcal{F}_s\right] &\geq E\left[1_{A_R}\varphi(M_s)|\mathcal{F}_s\right] + E\left[1_{A_R} m(M_s)(M_t - M_s)|\mathcal{F}_s\right] \\ &= 1_{A_R}\varphi(M_s) + 1_{A_R} m(M_s) E\left[(M_t - M_s)|\mathcal{F}_s\right] \end{aligned}$$

da cui, come sopra,

$$E\left[1_{A_R}\varphi(M_t)|\mathcal{F}_s\right] \geq 1_{A_R}\varphi(M_s).$$

Se $R < R'$, $A_R \subset A_{R'}$, quindi $1_{A_R} \leq 1_{A_{R'}}$. Basta allora passare al limite, per $R \to \infty$ col teorema di convergenza monotona per la speranza condizionale, osservando che

$$\lim_{R \to \infty} 1_{A_R} = 1$$

quasi certamente (in quanto $\cup_R A_R = \{|M_s| < \infty\}$ che ha probabilità 1). La dimostrazione è completa. ■

Osservazione 11 *Se M è una submartingala, non è detto che lo siano $|M_t|$ e M_t^2, in quanto le funzioni $x \longmapsto |x|$ e $x \longmapsto x^2$ non sono crescenti. Se però M è a valori positivi, si può modificare la dimostrazione e verificare che $|M_t|$ e M_t^2 sono submartingale (nel secondo caso se M è di quadrato integrabile).*

3.2 Tempi d'arresto

Fondamentali nello studio delle martingale sono i tempi aleatori. Genericamente, potremmo chiamare tempo aleatorio (non è una definizione canonica) ogni v.a. a valori in $[0, \infty)$, o addirittura a valori in $[0, \infty]$. Un tempo aleatorio $\tau : \Omega \to [0, \infty]$ si dirà *tempo d'arresto*, rispetto ad una filtrazione data $(\mathcal{F}_t)_{t \geq 0}$, se per ogni $t \geq 0$ l'evento $\{\tau \leq t\}$ appartiene a $\mathcal{F}_t$:

$$\{\tau \leq t\} \in \mathcal{F}_t \qquad \forall t \geq 0.$$

Quindi anche $\{\tau > t\} \in \mathcal{F}_t$. Se $\tau : \Omega \to [0, \infty)$, diremo che è un *tempo d'arresto finito*. Se esiste $T > 0$ tale che $|\tau| \leq T$, diremo che è un *tempo d'arresto limitato*.

Quando tratteremo martingale a tempo discreto, relativamente a filtrazioni $(\mathcal{F}_n)_{n \in \mathbb{N}}$, considereremo sempre (e a volte tacitamente) tempi d'arresto a tempo discreto, cioè tempi aleatori $\tau : \Omega \to \mathbb{N} \cup \{+\infty\}$ tali che $\{\tau \leq n\} \in \mathcal{F}_n$ per ogni $n \in \mathbb{N}$. Si verifica subito che questa condizione equivale a $\{\tau = n\} \in \mathcal{F}_n$ per ogni $n \in \mathbb{N}$. Infatti, se vale la prima ($\{\tau \leq n\} \in \mathcal{F}_n$), allora

$$\{\tau = n\} = \{\tau \leq n\} \setminus \{\tau \leq n-1\} \in \mathcal{F}_n$$

perché $\{\tau \leq n\} \in \mathcal{F}_n$, $\{\tau \leq n-1\} \in \mathcal{F}_{n-1} \subset \mathcal{F}_n$ ed $\mathcal{F}_n$ è chiusa per differenza. Viceversa, se vale la seconda ($\{\tau = n\} \in \mathcal{F}_n$), allora

$$\{\tau \leq n\} = \bigcup_{k=0}^{n} \{\tau = k\} \in \mathcal{F}_n$$

perché $\{\tau = k\} \in \mathcal{F}_k \subset \mathcal{F}_n$ per ogni $k = 0, ..., n$ ed $\mathcal{F}_n$ è chiusa per unione finita.

Il significato intuitivo di tempo d'arresto è il seguente: fissato $t_0 \geq 0$ (ad esempio l'orario di chiusura di un negozio) ci chiediamo se l'istante aleatorio τ (ad esempio l'arrivo di un cliente che deve ritirare della merce) avvenga prima o dopo t_0. Tale evento, $\{\tau \leq t_0\}$, è conoscibile al tempo t_0, se per "conoscibili" intendiamo gli eventi della filtrazione data (nell'esempio del negozio, se la filtrazione è associata alle informazioni note a chi sta lavorando nel negozio, vale $\{\tau \leq t_0\} \in \mathcal{F}_{t_0}$: lo vedono se il cliente è arrivato o no in tempo; se invece è una filtrazione associata alle informazioni note ad un gestore a distanza che viene informato il giorno successivo di ciò che è accaduto in negozio, $\{\tau \leq t_0\} \notin \mathcal{F}_{t_0}$).

Abbiamo descritto questa nozione nel caso a tempo continuo, ma è identica a tempo discreto.

Un tempo deterministico $\tau = T$ è un tempo d'arresto, perché gli eventi della forma $\{\tau \leq t\}$ appartengono a $\{\varnothing, \Omega\}$.

Tra le numerose proprietà generali a volte utili segnaliamo la seguente, usata spesso per ridurre un tempo d'arresto qualsiasi ad uno limitato:

Proposizione 27 *Se τ_1, τ_2 sono tempi d'arresto, allora lo è anche $\tau_1 \wedge \tau_2$. In particolare, lo è $\tau_1 \wedge T$ per ogni tempo determinsitico $T > 0$.*

Proof. Dato $t \geq 0$,

$$\{\tau_1 \wedge \tau_2 \leq t\} = \{\tau_1 \leq t\} \cup \{\tau_2 \leq t\}$$

e ciascuno di questi eventi appartiene a $\mathcal{F}_t$. ■

Sia $X = (X_t)_{t \geq 0}$ un processo stocastico a valori in uno spazio metrico E ($\mathcal{E} = \mathcal{B}(E)$), adattato alla filtrazione $(\mathcal{F}_t)_{t \geq 0}$. Sia $A \in \mathcal{B}(E)$. Introduciamo un tempo aleatorio $\tau_A : \Omega \to [0, \infty]$ ponendo $\tau_A(\omega) = \inf\{t \geq 0 : X_t(\omega) \in A\}$ per ogni $\omega \in \Omega$ tale che l'insieme $\{t \geq 0 : X_t(\omega) \in A\}$ è non vuoto, $\tau_A(\omega) = \infty$ altrimenti. Useremo scrivere più compattamente

$$\tau_A = \inf\{t \geq 0 : X_t \in A\}$$

intendendo tutte le frasi precedenti. Diciamo che τ_A è il *tempo (o istante) di ingresso* del processo nell'insieme A. Chiediamoci se sia un tempo d'arresto.

Intanto vediamo il caso a tempo discreto. Definiamo cioè

$$\tau_A = \inf \{n \geq 0 : X_n \in A\}$$

se l'insieme $\{n \geq 0 : X_n \in A\}$ è non vuoto, $\tau_A = \infty$ altrimenti. Fissato $n_0 \in \mathbb{N}$, ci chiediamo se $\{\tau_A \leq n_0\} \in \mathcal{F}_{n_0}$. Vale

$$\{\tau_A \leq n_0\} = \{\exists n \leq n_0 : X_n \in A\} = \bigcup_{n \leq n_0} \{X_n \in A\} \in \mathcal{F}_{n_0}$$

dove l'ultima disuguaglianza deriva dal fatto che $\{X_n \in A\} \in \mathcal{F}_n$ e che $\mathcal{F}_n \subset \mathcal{F}_{n_0}$ se $n \leq n_0$. Quindi è vero: τ_A è un tempo di arresto. Non serve alcuna altra ipotesi.

A tempo continuo non è così. Proviamo a ripetere i passaggi precedenti, relativamente ad un tempo fissato $t_0 \geq 0$:

$$\begin{aligned} \{\tau_A \leq t_0\} &= \{\forall \varepsilon > 0, \exists t \leq t_0 + \varepsilon : X_t \in A\} \\ \left\{\forall n \in \mathbb{N}, \exists t \leq t_0 + \frac{1}{n} : X_t \in A\right\} &= \bigcap_{n \in \mathbb{N}} \bigcup_{t \leq t_0 + \frac{1}{n}} \{X_t \in A\}. \end{aligned}$$

L'evento $\{X_t \in A\}$ appartiene a $\mathcal{F}_t$ e quindi a $\mathcal{F}_{t_0 + \frac{1}{n}}$ se $t \leq t_0 + \frac{1}{n}$. Ma prima di tutto l'unione $\bigcup_{t \leq t_0 + \frac{1}{n}}$ è (a priori) più che numerabile. Quand'anche potessimo riesprimerla come unione numerabile, quindi potessimo asserire che $\bigcup_{t \leq t_0 + \frac{1}{n}} \{X_t \in A\} \in \mathcal{F}_{t_0 + \frac{1}{n}}$, poi l'intersezione $\bigcap_{n \in \mathbb{N}}$ starebbe in $\mathcal{F}_{t_0}$ se la filtrazione fosse continua a destra, altrimenti non potremmo concluderlo.

Proposizione 28 *Se X è un processo continuo, A è un aperto e $(\mathcal{F}_t)_{t \geq 0}$ è continua a destra, allora τ_A è un tempo di arresto. Lo stesso vale se X è continuo a destra o continuo a sinistra.*

Proof. Se X è continuo e A è aperto allora

$$\left\{\forall n \in \mathbb{N}, \exists t \leq t_0 + \frac{1}{n} : X_t \in A\right\} = \left\{\forall n \in \mathbb{N}, \exists t \leq t_0 + \frac{1}{n}, t \in \mathbb{Q} : X_t \in A\right\}.$$

Basta verificare l'inclusione $\subset$. Sia $\omega \in \Omega$ tale che $\forall n \in \mathbb{N}$, $\exists t \leq t_0 + \frac{1}{n}$ tale che $X_t \in A$. Se, relativamente ad n e t che verificano questa condizione, prendiamo una successione

$t_n \to t$, $t_n \in \mathbb{Q}$, $t_n \leq t_0 + \frac{1}{n}$, abbiamo $X_{t_n}(\omega) \to X_t(\omega)$, ma quindi $X_{t_n}(\omega) \in A$ per n abbastanza grande, in quanto A è aperto.

L'identità precedente permette di scrivere

$$\{\tau_A \leq t_0\} = \bigcap_{n\in\mathbb{N}} \bigcup_{t\leq t_0+\frac{1}{n}, t\in\mathbb{Q}} \{X_t \in A\}$$

da cui discende la tesi per i ragionamenti già fatti sopra. Le verifiche nel caso continuo a destra o continuo a sinistra sono simili. ■

Quando A, che ora chiameremo C, è chiuso ed X è continuo, si può modificare la dimostrazione precedente dall'inizio e concludere che τ_C è un tempo d'arresto. Non serve la continuità a destra della filtrazione. Il criterio è quindi anche più semplice del precedente, ma la dimostrazione è più complicata. Si noti anche la riscrittura di τ_C come minimo corrisponde ancor più all'idea di primo istante di ingresso in C.

Proposizione 29 *Se X è un processo continuo e C è un chiuso, allora τ_C è un tempo di arresto e vale*

$$\tau_C = \min\{t \geq 0 : X_t \in C\}.$$

Proof. Se X è continuo e C è chiuso allora, preso $\omega \in \Omega$, se l'insieme $\{t \geq 0 : X_t(\omega) \in C\} \neq \emptyset$, vale

$$\tau_A(\omega) = \min\{t \geq 0 : X_t(\omega) \in C\}.$$

Infatti, per definizione di $\tau_C(\omega)$ come $\inf\{t \geq 0 : X_t(\omega) \in C\}$, esiste una successione $t_n(\omega) \downarrow \tau_C(\omega)$ tale che $X_{t_n(\omega)}(\omega) \in C$; ma per continuità delle traiettorie $X_{t_n(\omega)}(\omega) \to X_{\tau_C(\omega)}(\omega)$ e per chiusura di C vale $X_{\tau_A(\omega)}(\omega) \in C$. Quindi l'insieme $\{t \geq 0 : X_t(\omega) \in C\}$ possiede minimo. Abbiamo dimostrato l'ultima affermazione della proposizione. Vediamo la prima.

Ora possiamo dire che

$$\{\tau_C \leq t_0\} = \{\exists t \leq t_0 : X_t \in C\} = \bigcup_{t\leq t_0} \{X_t \in C\}.$$

Non possiamo però ridurre questa unione ad una numerabile perché dobbiamo immaginare ad esempio il caso in cui C sia un singolo punto, ed il processo lo tocchi a certi istanti (non necessariamente membri di un insieme numerabile prefissato) ma non permanga in C.

Il caso $t_0 = 0$ però si risolve: $\{\tau_C \leq 0\} = \{X_0 \in C\} \in \mathcal{F}_0$. Possiamo restringerci a $t_0 > 0$.

Introduciamo allora l'intorno aperto di raggio $\varepsilon > 0$ di C:

$$C_\varepsilon = \{x \in E : d(x, C) < \varepsilon\}.$$

Vale

$$\bigcup_{t\leq t_0}\{X_t\in C\}=\bigcap_{n\in\mathbb{N}}\bigcup_{t\leq t_0,t\in\mathbb{Q}}\left\{X_t\in C_{\frac{1}{n}}\right\}\in\mathcal{F}_{t_0}.$$

L'appartenenza a $\mathcal{F}_{t_0}$ è ovvia. Verifichiamo l'identità tra i due insiemi. L'inclusione $\subset$ è concettualmente più facile. Sia ω nell'insieme di sinistra. Se per un certo $t\leq t_0$ vale $X_t(\omega)\in C$, allora per ogni $n\in\mathbb{N}$ esiste un numero razionale $t_n\leq t_0$ tale che $X_{t_n}(\omega)\in C_{\frac{1}{n}}$, per continuità della traiettoria $X_\cdot(\omega)$. Quindi l'inclusione $\subset$ è dimostrata.

Vediamo l'inclusione $\supset$. Sia ω nell'insieme di destra. Per ogni $n\in\mathbb{N}$ esiste un numero razionale $t_n\leq t_0$ tale che $X_{t_n}(\omega)\in C_{\frac{1}{n}}$. Dalla successione limitata $\{t_n\}$ si può estrarre una sottosuccessione convergente $\{t_{n_k}\}$; sia t il suo limite; vale $t\leq t_0$. Per continuità di $X_\cdot(\omega)$ vale $\lim_{k\to\infty}X_{t_{n_k}}(\omega)=X_t(\omega)$. Se dimostriamo che $X_t(\omega)\in C$, allora ω appartiene anche all'insieme di sinistra.

Ma $X_{t_{n_k}}(\omega)\in C_{\frac{1}{n_k}}$. Questo implica che esiste un elemento $e_k(\omega)\in C$ tale che $d\left(X_{t_{n_k}}(\omega),e_k(\omega)\right)\leq\frac{2}{n_k}$. Siccome $\lim_{k\to\infty}d\left(X_{t_{n_k}}(\omega),X_t(\omega)\right)=0$,

$$\lim_{k\to\infty}d\left(X_t(\omega),e_k(\omega)\right)\leq\lim_{k\to\infty}d\left(X_{t_{n_k}}(\omega),X_t(\omega)\right)+\lim_{k\to\infty}d\left(X_{t_{n_k}}(\omega),e_k(\omega)\right)=0.$$

Questo significa $e_k(\omega)\to X_t(\omega)$, quindi $X_t(\omega)\in C$ perché C è chiuso. La dimostrazione è completa. ■

3.2.1 La σ-algebra associata ad un tempo d'arresto

Sia $(\Omega,\mathcal{F},P)$ uno spazio probabilizzato e $\tau:\Omega\to[0,\infty]$ un tempo d'arresto rispetto ad una filtrazione $(\mathcal{F}_t)_{t\geq0}$. Sappiamo che se per ogni $t\geq0$ l'evento $\{\tau\leq t\}$ appartiene a $\mathcal{F}_t$. Prendiamo un evento $A\in\mathcal{F}_\infty=\vee_t\mathcal{F}_t$. Esaminiamo la condizione

$$A\cap\{\tau\leq t\}\in\mathcal{F}_t\qquad\text{per ogni } t\geq0.$$

La famiglia di tutti gli insiemi $A\in\mathcal{F}_\infty$ che soddisfano questa condizione è una σ-algebra, che indichiamo con $\mathcal{F}_\tau$. Infatti $\Omega\in\mathcal{F}_\tau$ perché $\Omega\cap\{\tau\leq t\}=\{\tau\leq t\}\in\mathcal{F}_t$; se $A\in\mathcal{F}_\tau$ allora $A^c\in\mathcal{F}_\tau$ perché

$$A^c\cap\{\tau\leq t\}=\{\tau\leq t\}\setminus(A\cap\{\tau\leq t\});$$

ed infine, se $\{A_n\}\subset\mathcal{F}_\tau$ allora $\cup_nA_n\in\mathcal{F}_\tau$ per la proprietà distributiva di unione e intersezione.

Si noti che se $A\in\mathcal{F}_{t_0}$ per qualche $t_0\geq0$, non è detto che sia $A\in\mathcal{F}_\tau$, perché la condizione $A\in\mathcal{F}_\tau$ richiede che sia $A\cap\{\tau\leq t\}\in\mathcal{F}_t$ *per ogni* $t\geq0$, non solo al tempo t_0. Quindi può apparire non è ovvio esibire eventi di $\mathcal{F}_\tau$. Invece, ad esempio un evento del tipo

$$\{\tau\in[0,T]\}$$

con $T > 0$ ci sta. Infatti

$$\{\tau \in [0, T]\} \cap \{\tau \leq t\} = \{\tau \in [0, t \wedge T]\} \in \mathcal{F}_{t \wedge T} \subset \mathcal{F}_t.$$

Quindi τ è $\mathcal{F}_\tau$-misurabile. Pertanto, gli eventi della forma $\{\tau \in B\}$, con $B \in \mathcal{B}(\mathbb{R})$, sono esempi di eventi di $\mathcal{F}_\tau$.

Non si deve però pensare che $\mathcal{F}_\tau$ sia la σ-algebra generata da τ. La seguente proposizione fornisce ulteriori eventi di $\mathcal{F}_\tau$.

Proposizione 30 *Sia τ un tempo d'arresto finito. Sia $(X_t)_{t \geq 0}$ un processo progressivamente misurabile. Allora X_τ (definito da $\omega \longmapsto X_{\tau(\omega)}(\omega)$) è una v.a. $\mathcal{F}_\tau$-misurabile.*

Proof. Consideriamo l'applicazione misurabile $\omega \longmapsto (\tau(\omega), \omega)$ da $(\Omega, \mathcal{F}_\infty)$ in $([0, \infty) \times \Omega, \mathcal{B}([0, \infty)) \times$. e l'applicazione misurabile $(s, \omega) \longmapsto X_s(\omega)$ da $([0, \infty) \times \Omega, \mathcal{B}([0, \infty)) \times \mathcal{F}_\infty)$ in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. La loro composizione è l'applicazione $\omega \longmapsto X_{\tau(\omega)}(\omega)$ che quindi risulta misurabile da $(\Omega, \mathcal{F}_\infty)$ in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Questo chiarisce che ogni evento della forma $\{X_\tau \in B\}$ con $B \in \mathcal{B}(\mathbb{R})$ è $\mathcal{F}_\infty$-misurabile. Preso poi $t \geq 0$, esaminiamo la condizione $\{X_\tau \in B\} \cap \{\tau \leq t\} \in \mathcal{F}_t$; se vale, significa che $\{X_\tau \in B\} \in \mathcal{F}_\tau$, quindi X_τ è $\mathcal{F}_\tau$-misurabile. Osserviamo che

$$\{X_\tau \in B\} \cap \{\tau \leq t\} = \{X_{\tau \wedge t} \in B\} \cap \{\tau \leq t\}$$

quindi basta verificare che $\{X_{\tau \wedge t} \in B\} \in \mathcal{F}_t$.

Consideriamo allora l'applicazione $\omega \longmapsto (\tau(\omega) \wedge t, \omega)$ da $(\Omega, \mathcal{F}_t)$ in $([0, t] \times \Omega, \mathcal{B}([0, t]) \times \mathcal{F}_t)$ e l'applicazione $(s, \omega) \longmapsto X_s(\omega)$ da $([0, t] \times \Omega, \mathcal{B}([0, t]) \times \mathcal{F}_t)$ in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. La loro composizione è l'applicazione $\omega \longmapsto X_{\tau(\omega) \wedge t}(\omega)$. Se le due applicazioni che componiamo sono misurabili, allora lo è la loro composizione, da $(\Omega, \mathcal{F}_t)$ in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ e questo implica che $\{X_{\tau \wedge t} \in B\} \in \mathcal{F}_t$ per ogni $B \in \mathcal{B}(\mathbb{R})$. La seconda delle due applicazioni è misurabile per l'ipotesi di progressiva misurabilità. Vediamo la prima.

L'applicazione $\omega \longmapsto (\tau(\omega) \wedge t, \omega)$ da $(\Omega, \mathcal{F}_t)$ in $([0, t] \times \Omega, \mathcal{B}([0, t]) \times \mathcal{F}_t)$ è misurabile. Basta verificarlo su un insieme di generatori. Prendiamo un insieme della forma $[0, t_0] \times B$ con $t_0 \in (0, t)$ e $B \in \mathcal{F}_t$. La sua controimmagine è l'insieme

$$\{\omega \in \Omega : \tau(\omega) \wedge t \in [0, t_0]\} \cap \{\omega \in B\} = \{\tau \leq t_0\} \cap B$$

che appartiene a $\mathcal{F}_t$ perché $B \in \mathcal{F}_t$ e $\{\tau \leq t_0\} \in \mathcal{F}_{t_0} \subset \mathcal{F}_t$. La dimostrazione è completa. ■

3.3 Teoremi d'arresto e disuguaglianze

La strategia generale che seguiremo sarà di dimostrare prima i teoremi fondamentali nel caso a tempo discreto, poi di dedurre da essi il caso continuo con un passaggio al limite, da tempi diadici a tempi qualsasi (che coinvolgerà qualche ipotesi di continuità del processo).

3.3.1 Il caso a tempo discreto

Sia $M = (M_n)_{n\in\mathbb{N}}$ una martingala a tempo discreto. I tempi d'arresto di cui si parlerà nel seguito sono sempre a valori nei numeri naturali, eventualmente con l'aggiunta di $+\infty$. Un tempo d'arresto si dirà limitato superiormente se esiste una costante (intera nel caso a tempo discreto) $N > 0$ tale che $\tau(\omega) \leq N$ per ogni $\omega \in \Omega$. Se M è una martingala (quindi integrabile) e τ è un tempo d'arresto limitato superiormente dalla costante intera $N > 0$, allora le v.a. M_τ[4] e $\max_{n=1,\dots,N}|M_n|$ sono integrabili:

$$E[|M_\tau|] \leq E\left[\max_{n=1,\dots,N}|M_n|\right] \leq E\left[\sum_{n=1}^{N}|M_n|\right] = \sum_{n=1}^{N} E[|M_n|] < \infty.$$

Lo stesso vale per sub e super martingale o per qualsiasi processo discreto integrabile. Questa premessa interviene in molte delle dimostrazioni seguenti e non verrà ripetuta.

Nel seguente teorema l'affermazione $E[M_N] = E[M_0]$ è ovvia, come abbiamo osservato subito dopo la definizione di martingala.

Teorema 8 (d'arresto) *Se τ è un tempo d'arresto limitato superiormente da un intero N, allora*

$$E[M_\tau] = E[M_N] = E[M_0].$$

Se M è una submartingala, allora vale $E[M_\tau] \leq E[M_N]$.

Proof. Con le notazioni precedenti vale

$$E[M_\tau] = \sum_{n=0}^{N} E[M_\tau 1_{\tau=n}] = \sum_{n=0}^{N} E[M_n 1_{\tau=n}].$$

Per la proprietà di martingala $E[M_n 1_{\tau=n}] = E[M_N 1_{\tau=n}]$, quindi

$$E[M_\tau] = \sum_{n=0}^{N} E[M_N 1_{\tau=n}] = E[M_N].$$

Ma $E[M_N] = E[M_0]$. La proprietà è dimostrata. Il caso della submartingala è identico, senza il passaggio finale. ■

Corollario 2 *Se $\tau_1 \leq \tau_2$ sono due tempi d'arresto limitati superiormente, allora vale*

$$E[M_{\tau_2}|\mathcal{F}_{\tau_1}] = M_{\tau_1}.$$

Se M è una submartingala e $\tau_2 = N$ è costante, allora vale

$$E[M_N|\mathcal{F}_{\tau_1}] \geq M_{\tau_1}.$$

[4] Definita da $(M_\tau)(\omega) = M_{\tau(\omega)}(\omega)$

Proof. Preso $A \in \mathcal{F}_{\tau_1}$ dobbiamo dimostrare che $\int_A M_{\tau_2} dP = \int_A M_{\tau_1} dP$ (risp. $\int_A M_N dP = \int_A M_{\tau_1} dP$ nel caso della submartingala). Consideriamo la funzione $\tau_3 : \Omega \to [0, \infty)$ definita da

$$\tau_3(\omega) = \begin{cases} \tau_1(\omega) & \text{se} \quad \omega \in A \\ \tau_2(\omega) & \text{se} \quad \omega \notin A \end{cases}.$$

Verifichiamo che sia un tempo d'arresto. Preso $t \geq 0$, vale

$$\{\tau_3 \leq t\} = (A \cap \{\tau_1 \leq t\}) \cup (A^c \cap \{\tau_2 \leq t\}).$$

L'insieme $A \cap \{\tau_1 \leq t\}$ appartiene a $\mathcal{F}_t$ per l'ipotesi su A. $A^c \in \mathcal{F}_{\tau_1}$, quindi $A^c \in \mathcal{F}_{\tau_2}$ (si veda l'esercizio 11), pertanto $A^c \cap \{\tau_2 \leq t\} \in \mathcal{F}_t$. Quindi $\{\tau_3 \leq t\} \in \mathcal{F}_t$, cioè τ_3 è un tempo d'arresto. Ovviamente è anch'esso limitato.

Per il teorema d'arresto vale allora

$$E[M_{\tau_3}] = E[M_N] = E[M_{\tau_2}]$$

(risp. $E[M_{\tau_3}] \leq E[M_N]$ nel caso della submartingala) cioè, ricordando la definizione di τ_3,

$$E[M_{\tau_1} 1_A + M_{\tau_2} 1_{A^c}] = E[M_{\tau_2}]$$

da cui $E[M_{\tau_1} 1_A] = E[M_{\tau_2} 1_A]$, che è ciò che volevamo dimostrare (risp. $E[M_{\tau_1} 1_A + M_N 1_{A^c}] \leq E[M_N]$ da cui $E[M_{\tau_1} 1_A] \leq E[M_N 1_A]$ nel caso della submartingala). La dimostrazione è completa. ■

Osservazione 12 *I due risultati precedenti sono equivalenti. Abbiamo dimostrato il corollario a partire dal teorema. Se invece avessimo dimostrato prima il corollario, basterebbe prendere la speranza di ambo i membri delle sue formule per ottenere quelle del teorema.*

Osservazione 13 *Vedremo più avanti che si può rimuovere l'ipotesi $\tau_2 = N$, nel corollario.*

Teorema 9 (disuguaglianza massimale) *Sia M una martingala oppure una submartingala positiva. Allora, per ogni N intero positivo e per ogni $\lambda > 0$, vale*

$$P\left(\max_{n=1,\dots,N} |M_n| \geq \lambda\right) \leq \frac{1}{\lambda} E\left[|M_N| 1_{\max_{n=1,\dots,N} |M_n| \geq \lambda}\right] \leq \frac{1}{\lambda} E[|M_N|].$$

Proof. Per capire la naturalezza del primo passaggio della dimostrazione è utile ricordare la dimostrazione della disuguaglianza di Chebyshev: essa asserisce che $P(|X| \geq \lambda) \leq \frac{1}{\lambda} E[|X|]$, per ogni $\lambda > 0$ ed ogni v.a. X, e la dimostrazione si svolge semplicemente applicando la monotonia del valore atteso alla disuguaglianza

$$\lambda 1_{|X| \geq \lambda} \leq |X| 1_{|X| \geq \lambda}.$$

Veniamo al caso delle submartingale. Data la submartingala originaria $M = (M_n)_{n\in\mathbb{N}}$ e dato N intero positivo, introduciamo una nuova submartingala $M^N = \left(M_n^N\right)_{n\in\mathbb{N}}$ definita da $M_n^N = M_n$ per $n \leq N$, $M_n^N = M_N$ per $n > N$. E' equivalente dimostrare la tesi del teorema per M^N. Quindi possiamo supporre sin dall'inizio, senza restrizione, che M sia costante dal tempo N in poi; facciamo questa ipotesi ed usiamo la notazione M per la nostra submartingala.

Introduciamo la funzione $\tau_N : \Omega \to \{0, ..., N\}$ definita da

$$\tau_N = \min\{n \leq N : |M_n| \geq \lambda\}$$

se tale insieme non è vuoto, $\tau_N = N + 1$ altrimenti. E' un tempo d'arresto: se $n_0 \leq N$,

$$\{\tau_N \leq n_0\} = \bigcup_{n \leq n_0} \{|M_n| \geq \lambda\} \in \mathcal{F}_{n_0}$$

mentre, se $n_0 > N$,

$$\{\tau_N \leq n_0\} = \{\tau_N = N + 1\} = \bigcap_{n \leq N} \{|M_n| < \lambda\} \in \mathcal{F}_N \subset \mathcal{F}_{n_0}.$$

Vale

$$\lambda 1_{\max_{n=1,\ldots,N}|M_n|\geq\lambda} \leq |M_{\tau_N}|\, 1_{\max_{n=1,\ldots,N}|M_n|\geq\lambda}.$$

Infatti, sull'evento $\{\max_{n=1,\ldots,N} |M_n| \geq \lambda\}$ vale $\{\tau_N \leq N\}$ e $\{M_{\tau_N} \geq \lambda\}$.

Integrando,

$$P\left(\max_{n=1,\ldots,N} |M_n| \geq \lambda\right) \leq \frac{1}{\lambda} E\left[|M_{\tau_N}|\, 1_{\max_{n=1,\ldots,N}|M_n|\geq\lambda}\right].$$

Ora, l'evento $\{\max_{n=1,\ldots,N} |M_n| \geq \lambda\} = \{\tau_N \leq N\}$ appartiene a $\mathcal{F}_{\tau_N}$ ed il processo $|M_n|$ è una submartingala (se M_n è una martingala, allora $|M_n|$ è una submartingala come trasformazione convessa; se M_n è una martingala positiva, allora $|M_n|$ è una submartingala come trasformazione convessa crescente), quindi vale, per il Corollario 2,

$$E\left[|M_{\tau_N}|\, 1_{\max_{n=1,\ldots,N}|M_n|\geq\lambda}\right] \leq E\left[|M_{N+1}|\, 1_{\max_{n=1,\ldots,N}|M_n|\geq\lambda}\right] = E\left[|M_N|\, 1_{\max_{n=1,\ldots,N}|M_n|\geq\lambda}\right]$$

dove abbiamo usato il fatto che $\tau_N \leq N + 1$ e che $M_{N+1} = M_N$. La dimostrazione è completa. ■

Corollario 3 *Dato $p \geq 1$, se $E[|M_n|^p] < \infty$ per ogni n, allora*

$$P\left(\max_{n=1,\ldots,N} |M_n| \geq \lambda\right) \leq \frac{1}{\lambda^p} E[|M_N|^p].$$

Dato $\theta > 0$, se $E\left[e^{\theta|M_n|}\right] < \infty$ per ogni n, allora

$$P\left(\max_{n=1,\ldots,N} |M_n| \geq \lambda\right) \leq e^{-\theta\lambda} E\left[e^{\theta|M_N|}\right].$$

Proof. Il processo $|M_n|^p$ è una submartingala, per la Proposizione 26. Per la disuguaglianza massimale applicata a $|M_n|^p$ e λ^p

$$P\left(\max_{n=1,\dots,N}|M_n|\geq\lambda\right)=P\left(\max_{n=1,\dots,N}|M_n|^p\geq\lambda^p\right)\leq\frac{1}{\lambda^p}E\left[|M_N|^p\right].$$

La dimostrazione della seconda disuguaglianza è identica, usando la funzione convessa e crescente $x\longmapsto e^{\theta x}$. ■

Le stime delle code di una v.a. implicano stime sui valori medi. Ricordiamo:

Lemma 3 *Se X è una v.a. integrabile non negativa, allora*

$$E\left[X^p\right]=\int_0^\infty pa^{p-1}P\left(X\geq a\right)da$$

per ogni $p\geq 1$.

Proof.

$$\int_0^\infty pa^{p-1}P\left(X\geq a\right)da=\int_0^\infty pa^{p-1}E\left[1_{X\geq a}\right]da=E\left[\int_0^\infty pa^{p-1}1_{X\geq a}da\right]=E\left[\int_0^X pa^{p-1}da\right]=E\left[X^p\right].$$

■

Nell'ambito delle equazioni differenziali stocastiche, il seguente risultato è molto utile.

Teorema 10 (disuguaglianza di Doob) *Se M è una martingala oppure una submartingala positiva, per ogni $p>1$ vale*

$$E\left[\max_{n=1,\dots,N}|M_n|^p\right]\leq\left(\frac{p}{p-1}\right)^pE\left[|M_N|^p\right].$$

Precisamente, se $E\left[|M_N|^p\right]=+\infty$, la disuguaglianza è ovvia; se $E\left[|M_N|^p\right]<\infty$ allora anche $\max_{n=1,\dots,N}|M_n|^p$ è integrabile e vale la disuguaglianza.

Proof. Poniamo $M_N^*=\max_{n=1,\dots,N}|M_n|$. Allora

$$\begin{aligned}E\left[\max_{n=1,\dots,N}|M_n|^p\right] &= \int_0^\infty pa^{p-1}P\left(M_N^*\geq a\right)da\leq\int_0^\infty pa^{p-1}\frac{1}{a}E\left[|M_N|\,1_{M_N^*\geq a}\right]da\\ &= E\left[|M_N|\int_0^{M_N^*}pa^{p-2}da\right]=\frac{p}{p-1}E\left[|M_N|\left(M_N^*\right)^{p-1}\right].\end{aligned}$$

Per la disuguaglianza di Hölder, ricordando che $\frac{1}{p}+\frac{p-1}{p}=1$,

$$E\left[|M_N|\left(M_N^*\right)^{p-1}\right]\leq E\left[|M_N|^p\right]^{\frac{1}{p}}E\left[\left(M_N^*\right)^p\right]^{\frac{p-1}{p}}$$

quindi, dividendo per $E\left[\left(M_N^*\right)^p\right]^{\frac{p-1}{p}}$,

$$E\left[\left(M_N^*\right)^p\right]^{\frac{1}{p}}\leq\frac{p}{p-1}E\left[|M_N|^p\right]^{\frac{1}{p}}$$

da cui la tesi. ■

Esercizio 11 *Se $\tau_1\leq\tau_2$ sono due tempi d'arresto, allora $\mathcal{F}_{\tau_1}\subset\mathcal{F}_{\tau_2}$.*

3.3.2 Il caso a tempo continuo

Le dimostrazioni che seguono sono basate tutte sulla stessa strategia. Data una martingala (resp. submartingala) $M = (M_t)_{t\geq 0}$, dato un tempo $T > 0$, si considerano gli istanti "diadici" della forma $\frac{kT}{2^n}$, con $n \in \mathbb{N}$ e $k = 0, 1, ..., 2^n$ e si introduce, per ogni $n \in \mathbb{N}$, il processo a tempo discreto

$$M_k^{(n,T)} := M_{\frac{kT}{2^n}}, \qquad k \in \mathbb{N}.$$

Esso è una martingala (resp. submartingala) a tempo discreto, rispetto alla filtrazione $\left(\mathcal{F}_{\frac{kT}{2^n}}\right)_{k\in\mathbb{N}}$. Ad essa si applicano quindi i risultati precedenti. Bisogna poi eseguire un passaggio al limite per $n \to \infty$ e per controllare il limite di alcune quantità servono delle ipotesi di continuità di M. I primi due teoremi necessitano solo della seguente ipotesi:

$$\sup_{t\in[0,T]} |M_t| = \lim_{n\to\infty} \max_{k=1,...,2^n} \left| M_k^{(n,T)} \right|. \tag{3.2}$$

Questa vale ad esempio se ogni (o q.o.) traiettoria di M è continua a destra o a sinistra in ogni punto.

Teorema 11 (disuguaglianza massimale) *Sia M una martingala, oppure una submartingala positiva, che soddisfa (3.2). Allora, per ogni $T > 0$, $\lambda > 0$, vale*

$$P\left(\sup_{t\in[0,T]} |M_t| \geq \lambda\right) \leq \frac{1}{\lambda} E\left[|M_T|\right].$$

Proof. Per ogni $n \in \mathbb{N}$, introduciamo l'evento

$$A_n = \left\{ \max_{k=1,...,2^n} \left| M_k^{(n,T)} \right| \geq \lambda \right\}.$$

Per la disuguaglianza massimale discreta applicata alla submartingala $M_k^{(n,T)}$, vale

$$P(A_n) \leq \frac{1}{\lambda} E\left[\left| M_{2^n}^{(n,T)} \right|\right] = \frac{1}{\lambda} E\left[|M_T|\right]$$

osservando che $M_{2^n}^{(n,T)} = M_T$. Gli eventi A_n sono non decrescenti, quindi $P(\cup_n A_n) = \lim_{n\to\infty} P(A_n)$. Preso $\varepsilon > 0$, grazie all'ipotesi (3.2), se $\sup_{t\in[0,T]} |M_t| = \eta$ (o ancor meglio se $\sup_{t\in[0,T]} |M_t| \geq \eta$) allora esiste n tale che $\max_{k=1,...,2^n} \left| M_k^{(n,T)} \right| > \eta - \varepsilon$. Preso $\eta = \lambda + \varepsilon$, questo dice che

$$\left\{ \sup_{t\in[0,T]} |M_t| \geq \lambda + \varepsilon \right\} \subset \bigcup_{n\in\mathbb{N}} A_n$$

quindi

$$P\left(\sup_{t\in[0,T]}|M_t|\geq\lambda+\varepsilon\right)\leq\lim_{n\to\infty}P\left(A_n\right)\leq\frac{1}{\lambda}E\left[|M_T|\right].$$

I numeri $\lambda,\varepsilon>0$ sono arbitrari, quindi abbiamo anche dimostrato che, per $\varepsilon<\lambda$,

$$P\left(\sup_{t\in[0,T]}|M_t|\geq\lambda\right)\leq\frac{1}{\lambda-\varepsilon}E\left[|M_T|\right].$$

Pendendo il limite per $\varepsilon\to 0$ si ottiene la tesi. La dimostrazione è completa. ■

Corollario 4 *Dato $p\geq 1$, se $E\left[|M_t|^p\right]<\infty$ per ogni t, allora*

$$P\left(\sup_{t\in[0,T]}|M_t|\geq\lambda\right)\leq\frac{1}{\lambda^p}E\left[|M_T|^p\right].$$

Dato $\theta>0$, se $E\left[e^{\theta|M_t|}\right]<\infty$ per ogni t, allora

$$P\left(\sup_{t\in[0,T]}|M_t|\geq\lambda\right)\leq e^{-\theta\lambda}E\left[e^{\theta|M_T|}\right].$$

La dimostrazione è identica a quella del Corollario 3.

Teorema 12 (disuguaglianza di Doob) *Sia dato $p>1$. Se M è una martingala, oppure una submartingala positiva, che soddisfa (3.2) e $E\left[|M_t|^p\right]<\infty$ per ogni t, allora, per ogni $T>0$, la v.a. $\sup_{t\in[0,T]}|M_t|^p$ è integrabile e vale*

$$E\left[\sup_{t\in[0,T]}|M_t|^p\right]\leq\left(\frac{p}{p-1}\right)^p E\left[|M_T|^p\right].$$

Proof. Per ogni $n\in\mathbb{N}$, introduciamo la v.a. $X_n=\max_{k=1,\ldots,2^n}\left|M_k^{(n,T)}\right|^p$. Per la disuguaglianza di Doob discreta applicata alla submartingala $M_k^{(n,T)}$, vale

$$E\left[X_n\right]\leq\left(\frac{p}{p-1}\right)^p E\left[|M_T|^p\right].$$

La successione di v.a. $\{X_n\}_{n\in\mathbb{N}}$ è non descrescente e non negativa, quindi, per il teorema di convergenza monotona,

$$E\left[\lim_{n\to\infty}X_n\right]\leq\left(\frac{p}{p-1}\right)^p E\left[|M_T|^p\right].$$

Ma la v.a. $\lim_{n\to\infty}X_n$ coincide con $\sup_{t\in[0,T]}|M_t|^p$ per l'ipotesi (3.2). Questo completa la dimostrazione. ■

Teorema 13 (d'arresto) *Supponiamo che esista $p > 1$ tale che $E[|M_t|^p] < \infty$ per ogni t. Se M è una martingala continua a destra e τ è un tempo d'arresto limitato, allora*

$$E[M_\tau] = E[M_0].$$

Proof. Per ogni $n \in \mathbb{N}$, introduciamo il tempo aleatorio $\tau_n : \Omega \to [0, \infty)$ così definito:

$$\tau_n = \frac{k+1}{2^n} \text{ se } \frac{k}{2^n} < \tau \leq \frac{k+1}{2^n}.$$

E' un tempo aleatorio:

$$\{\tau_n \leq t\} = \bigcup_{k, \frac{k+1}{2^n} \leq t} \left\{\tau_n = \frac{k+1}{2^n}\right\} = \bigcup_{k, \frac{k+1}{2^n} \leq t} \left\{\frac{k}{2^n} < \tau \leq \frac{k+1}{2^n}\right\} \in \mathcal{F}_t.$$

Inoltre, $\lim_{n\to\infty} \tau_n = \tau$. Infine, esiste un intero $N > 0$ tale che $\tau_n \leq N$ per ogni $n \in \mathbb{N}$.

Per il teorema di arresto discreto,

$$E[M_{\tau_n}] = E[M_N] = E[M_0].$$

Per la continuità a destra, $\lim_{n\to\infty} M_{\tau_n} = M_\tau$. Inoltre,

$$|M_{\tau_n}| \leq \sup_{t\in[0,N]} |M_t|$$

che è integrabile per la disuguaglianza di Doob, e quindi, per il teorema di convergenza dominata, $\lim_{n\to\infty} E[M_{\tau_n}] = E[M_\tau]$. La dimostrazione è completa. ■

Esistono tante varianti del teorema d'arresto. Ne enunciamo una d'uso pratico nel caso di processi continui.

Corollario 5 *Supponiamo che esista $p > 1$ tale che $E[|M_t|^p] < \infty$ per ogni t. Se M è una martingala continua, τ è finito e la successione $M_{\tau\wedge n}$ è uniformemente integrabile (oppure limitata in modulo da una funzione integrabile), allora*

$$E[M_\tau] = E[M_0].$$

Proof. Consideriamo i tempi d'arresto limitati $\tau \wedge n$ al variare di $n \in \mathbb{N}$. Vale (per il teorema di arresto)

$$E[M_{\tau\wedge n}] = E[M_0]$$

e $\lim_{n\to\infty} M_{\tau\wedge n} = M_\tau$ per la continuità a sinistra. La tesi discende all'uniforme integrabilità ed il relativo teorema di passaggio al limite in valore atteso (o dal teorema di convergenza dominata di Lebesgue, se la successione $M_{\tau\wedge n}$ è limitata in modulo da una funzione integrabile). La dimostrazione è completa. ■

3.4 Applicazioni

3.4.1 Disuguaglianza esponenziale per il moto browniano

Si possono dimostrare le seguenti disuguaglianze, per $\lambda > 0$:

$$\left(\lambda + \frac{1}{\lambda}\right)^{-1} e^{-\frac{\lambda^2}{2}} \leq \int_\lambda^\infty e^{-\frac{x^2}{2}} dx \leq \frac{1}{\lambda} e^{-\frac{\lambda^2}{2}}.$$

Da quella di destra si ottiene, per ogni $T > 0$,

$$P\left(B_T \geq \lambda\right) \leq$$

Infatti

$$P\left(B_T \geq \lambda\right) = P\left(B_T/\sqrt{T} \geq \lambda/\sqrt{T}\right) \leq \frac{1}{2\pi} \frac{\sqrt{T}}{\lambda} e^{-\frac{\lambda^2}{2T}}.$$

Trascurando il fattore di fronte all'esponenziale, possiamo dire che la coda della v.a. B_T decade come $e^{-\frac{\lambda^2}{2T}}$ (in realtà decade un po' più velocemente, grazie al fattore che moltiplica l'esponenziale). La seguente disuguaglianza esponenziale rafforza questo risultato dal punto di vista dell'uniformità nel tempo, o se si vuole stabilisce un risultato simile per le traiettorie del moto browniano.

Proposizione 31 *Sia B un moto browniano. Allora, per ogni $\lambda > 0$, vale*

$$P\left(\sup_{t\in[0,T]} B_t \geq \lambda\right) \leq e^{-\lambda^2/2T}.$$

$$P\left(\sup_{t\in[0,T]} |B_t| \geq \lambda\right) \leq 2e^{-\lambda^2/2t}.$$

Proof. Per la Proposizione 26, preso $\theta > 0$, il processo $M_t := e^{\theta B_t}$ è una submartingala e quindi, per la disuguaglianza massimale,

$$P\left(\sup_{t\in[0,T]} B_t \geq \lambda\right) = P\left(\sup_{t\in[0,T]} e^{\theta B_t} \geq e^{\theta\lambda}\right) \leq e^{-\theta\lambda} E\left[e^{\theta B_t}\right].$$

Se X è una v.a. $N\left(0, \sigma^2\right)$ si dimostra con un calcolo elementare che $E\left[e^{\theta X}\right] = e^{\frac{\theta^2\sigma^2}{2}}$. Pertanto $E\left[e^{\theta B_t}\right] = e^{\frac{\theta^2 t}{2}}$. Otteniamo quindi

$$P\left(\sup_{t\in[0,T]} B_t \geq \lambda\right) \leq \inf_{\theta>0} e^{-\theta\lambda + \frac{\theta^2 t}{2}} = e^{\inf_{\theta>0}\left(\frac{\theta^2 t}{2} - \theta\lambda\right)}.$$

Per concludere basta osservare che il punto di minimo della parabola $\frac{\theta^2 t}{2} - \theta\lambda$ si trova in $\theta = \frac{\lambda}{t} > 0$ e vale $-\frac{\lambda^2}{2t}$.

Infine, osserviamo che $-B$ è ancora un moto browniano, quindi vale

$$P\left(\sup_{t\in[0,T]} (-B_t) \geq \lambda\right) \leq e^{-\lambda^2/2t}.$$

D'altra parte

$$\left\{\sup_{t\in[0,T]} |B_t| \geq \lambda\right\} \subset \left\{\sup_{t\in[0,T]} B_t \geq \lambda\right\} \bigcup \left\{\sup_{t\in[0,T]} (-B_t) \geq \lambda\right\}$$

per cui passando alle probabilità si ottiene la seconda disuguaglianza della proposizione.
■

3.4.2 Problema della rovina

Sia B un moto browniano (continuo) e, presi $a, b > 0$, consideriamo il problema dell'uscita di B dall'intervallo $[-a, b]$.

Osservazione 14 *Possiamo immaginare, per aiutare l'intuizione, che $B_t + a$ rappresenti il patrimonio di un giocatore al tempo t, patrimonio che ammonta al valore $a > 0$ al tempo $t = 0$, e che b sia la quantità che il giocatore si prefissa di vincere: smetterà di giocare quanto il suo patrimonio sarà pari ad $a + b$, ovvero quando $B_t = b$. Se però, prima di allora, il suo patrimonio si sarà ridotto a zero ($B_t = -a$), il giocatore sarà caduto in rovina prima di raggiungere il suo scopo. Questo problema serve a mostrare che la strategia di continuare a giocare fino al momento in cui si raggiunge la vincita desiderata (che in teoria prima o poi accadrà) si scontra col fatto che si dispone di un patrimonio iniziale limitato e quindi si può cadere in rovina prima di raggiungere il successo.*

Introduciamo il tempo di ingresso in $(-a, b)^c$, detto anche tempo d'uscita da $(-a, b)$, o primo istante in cui B raggiunge il valore a oppure b:

$$\tau_{a,b} = \min\left\{t \geq 0 : B_t \in (-a, b)^c\right\}.$$

Essendo B continuo ed $(-a, b)^c$ chiuso, $\tau_{a,b}$ è un tempo d'arresto.

Supponiamo di poter stabilire che $\tau_{a,b}$ sia finito. Siccome B è continuo e $B_t \in [-a, b]$ per $t \leq \tau_{a,b}$ (quindi la successione di v.a. $\left\{\left|B_{\tau_{a,b}\wedge n}\right|\right\}_{n\in\mathbb{N}}$ è equilimitata), possiamo applicare la versione del teorema d'arresto data dal Corollario 5 ed ottenere

$$E\left[B_{\tau_{a,b}}\right] = E\left[B_0\right] = 0.$$

D'altra parte, sempre nell'ipotesi che $\tau_{a,b}$ sia finito, avremmo

$$\begin{aligned}
E\left[B_{\tau_{a,b}}\right] &= E\left[B_{\tau_{a,b}} 1_{B_{\tau_{a,b}}=a}\right] + E\left[B_{\tau_{a,b}} 1_{B_{\tau_{a,b}}=b}\right] \\
&= E\left[a 1_{B_{\tau_{a,b}}=a}\right] + E\left[b 1_{B_{\tau_{a,b}}=b}\right] \\
&= aP\left(B_{\tau_{a,b}} = a\right) + bP\left(B_{\tau_{a,b}} = b\right)
\end{aligned}$$

ed anche

$$P\left(B_{\tau_{a,b}} = a\right) + P\left(B_{\tau_{a,b}} = b\right) = 1.$$

Quindi avremmo le due relazioni

$$\begin{aligned}
aP\left(B_{\tau_{a,b}} = a\right) + bP\left(B_{\tau_{a,b}} = b\right) &= 0 \\
P\left(B_{\tau_{a,b}} = a\right) + P\left(B_{\tau_{a,b}} = b\right) &= 1
\end{aligned}$$

da cui dedurremmo

$$P\left(B_{\tau_{a,b}} = a\right) = \frac{b}{a-b}, \qquad P\left(B_{\tau_{a,b}} = b\right) = \frac{a}{a-b}. \tag{3.3}$$

Proposizione 32 *Vale (3.3). Inoltre,*

$$E\left[\tau_{a,b}\right] = ab\frac{a+b}{a-b}.$$

Proof. Consideriamo il processo

$$M_t := B_t^2 - t.$$

E' una martingala (inquadreremo questo fatto più in generale nel seguito):

$$\begin{aligned}
E\left[M_t|\mathcal{F}_s\right] &= E\left[B_t^2|\mathcal{F}_s\right] - t = E\left[\left(B_t - B_s\right)^2|\mathcal{F}_s\right] + 2E\left[B_t B_s|\mathcal{F}_s\right] - E\left[B_s^2|\mathcal{F}_s\right] - t \\
&= E\left[\left(B_t - B_s\right)^2\right] + 2B_s E\left[B_t|\mathcal{F}_s\right] - B_s^2 - t \\
&= t - s + 2B_s^2 - B_s^2 - t = B_s^2 - s = M_s.
\end{aligned}$$

Preso $T > 0$, introduciamo il tempo d'arresto limitato $\tau_{a,b} \wedge T$. Vale

$$E\left[M_{\tau_{a,b}\wedge T}\right] = E\left[M_0\right] = 0$$

ovvero

$$E\left[\tau_{a,b} \wedge T\right] = E\left[B^2_{\tau_{a,b}\wedge T}\right].$$

Esiste una costante (il ragionamento è simile a quello fatto sopra) $C > 0$ tale che $B^2_{\tau_{a,b}\wedge T} \leq C$, indipendentemente da T, quindi

$$E\left[\tau_{a,b} \wedge T\right] \leq C.$$

Per convergenza monotona, si ottiene $E[\tau_{a,b}] \leq C$ e quindi $\tau_{a,b} < \infty$ quasi certamente. Valgono quindi i calcoli sviluppati sopra, che portano alle formule (3.3).

Inoltre, possiamo applicare il Corollario 5 ad M ed ottenere

$$E\left[M_{\tau_{a,b}}\right] = E[M_0] = 0.$$

D'altra parte,

$$\begin{aligned} E\left[M_{\tau_{a,b}}\right] &= E\left[M_{\tau_{a,b}} 1_{B_{\tau_{a,b}}=a}\right] + E\left[M_{\tau_{a,b}} 1_{B_{\tau_{a,b}}=b}\right] \\ &= E\left[\left(a^2 - \tau_{a,b}\right) 1_{B_{\tau_{a,b}}=a}\right] + E\left[\left(b^2 - \tau_{a,b}\right) 1_{B_{\tau_{a,b}}=b}\right] \\ &= a^2 P\left(B_{\tau_{a,b}} = a\right) + b^2 P\left(B_{\tau_{a,b}} = b\right) - E[\tau_{a,b}]. \end{aligned}$$

da cui

$$E[\tau_{a,b}] = a^2 P\left(B_{\tau_{a,b}} = a\right) + b^2 P\left(B_{\tau_{a,b}} = b\right) = a^2 \frac{b}{a-b} + b^2 \frac{a}{a-b} = ab\frac{a+b}{a-b}.$$

La dimostrazione è completa. ■

3.5 Teorema di decomposizione di Doob per submartingale discrete

Teorema 14 *Se X è una submartingala rispetto ad una filtrazione $(\mathcal{F}_n)_{n\geq 0}$, allora esistono una martingala $M = (M_n)_{n\geq 0}$ ed un processo crescente $A = (A_n)_{n\geq 0}$ tali che*

$$X_n = M_n + A_n.$$

Inoltre, A_n è $\mathcal{F}_{n-1}$ misurabile, per ogni n.

Proof. Definiamo M_n iterativamente in modo che valga

$$M_{n+1} - M_n = X_{n+1} - E[X_{n+1}|\mathcal{F}_n].$$

Poniamo ad esempio $M_0 = 0$, $M_1 = X_1 - E[X_1|\mathcal{F}_0]$, e così via $M_{n+1} = M_n + X_{n+1} - E[X_{n+1}|\mathcal{F}_n]$.

Vale

$$\begin{aligned} E[M_{n+1} - M_n|\mathcal{F}_n] &= E[X_{n+1} - E[X_{n+1}|\mathcal{F}_n]|\mathcal{F}_n] \\ &= E[X_{n+1}|\mathcal{F}_n] - E[X_{n+1}|\mathcal{F}_n] = 0. \end{aligned}$$

Quindi M è una martingala. Il processo

$$A_n := X_n - M_n$$

soddisfa

$$A_{n+1} = X_{n+1} - M_{n+1} = E\left[X_{n+1}|\mathcal{F}_n\right] - M_n \geq X_n - M_n = A_n$$

quindi è crescente. Inoltre, sempre da qui si vede che A_{n+1} è $\mathcal{F}_n$ misurabile. La dimostrazione è completa. ■

Possiamo finalmente tornare al teorema d'arresto e svincolarci da una limitazione antipatica per le submartingale.

Corollario 6 *Siano X una submartingala discreta e $\tau_1 \leq \tau_2$ due tempi d'arresto limitati. Allora*

$$E\left[X_{\tau_2}|\mathcal{F}_{\tau_1}\right] \geq X_{\tau_1}$$

ed in particolare

$$E\left[X_{\tau_2}\right] \geq E\left[X_{\tau_1}\right].$$

Proof. Scriviamo X nella forma

$$X_n = M_n + A_n$$

come nel teorema di decomposizione. Allora

$$E\left[X_{\tau_2}|\mathcal{F}_{\tau_1}\right] = E\left[M_{\tau_2}|\mathcal{F}_{\tau_1}\right] + E\left[A_{\tau_2}|\mathcal{F}_{\tau_1}\right].$$

Siccome A è crescente, $A_{\tau_2} \geq A_{\tau_1}$ (q.c.) e quindi, per la monotonia della speranza condizionale, $E\left[A_{\tau_2}|\mathcal{F}_{\tau_1}\right] \geq E\left[A_{\tau_1}|\mathcal{F}_{\tau_1}\right]$. Essendo poi A_{τ_1} misurabile rispetto a $\mathcal{F}_{\tau_1}$ (esercizio 12), vale

$$E\left[A_{\tau_2}|\mathcal{F}_{\tau_1}\right] \geq A_{\tau_1}.$$

Inoltre,

$$E\left[M_{\tau_2}|\mathcal{F}_{\tau_1}\right] = M_{\tau_1}$$

per il risultato d'arresto delle martingale. Quindi

$$E\left[X_{\tau_2}|\mathcal{F}_{\tau_1}\right] \geq M_{\tau_1} + A_{\tau_1} = X_{\tau_1}.$$

L'affermazione $E\left[X_{\tau_2}\right] \geq E\left[X_{\tau_1}\right]$ discende dalla precedente calcolando il valore atteso di ambo i membri. La dimostrazione è completa. ■

Esercizio 12 *Se X è un processo a tempo discreto e τ un tempo d'arresto finito, allora X_τ misurabile rispetto a $\mathcal{F}_\tau$.*

3.6 Teorema di convergenza per super-martingale discrete

Proviamo ad apprezzare intuitivamente il problema ragionando sulle traiettorie di una martingala. Abbiamo visto all'inizio del capitolo che gli incrementi sono indipendenti o quasi, e centrati. L'immagine che ci facciamo è quindi di traiettorie altamente oscillanti, con oscillazioni ugualmente distribuite verso l'alto e verso il basso (qui usiamo le parole in modo molto intuitivo); si pensi ad esempio al moto browniano. Il teorema che vedremo in questo paragrafo afferma che, sotto certe ipotesi, esiste finito il limite per $t \to \infty$, delle traiettorie (q.c.). Come può questo essere compatibile con le oscillazioni suddette? Le traiettorie browniane sono illimitate sia dall'alto sia dal basso, per $t \to \infty$.

Ciò che forse ci inganna pensando al moto browniano è la stazionarietà dei suoi incrementi. Una martingala, invece, può avere incrementi sempre più "piccoli", al crescere del tempo. Essi continueranno ad essere quasi indipendenti dal passato, ma se sono sempre più piccoli non è assurdo immaginare che le traiettorie abbiano limite; un po' come per la funzione $f(t) = \frac{1}{t} \sin t$.

3.6.1 Tentativo di euristica

Il teorema che stiamo per esporre è davvero inaspettato, rispetto alla media dei nostri teoremi e la sua dimostrazione non è affatto concettualmente semplice, pur molto sintetica. Proviamo allora ad avvicinarci per gradi con ragionamenti il più possibile plausibili (pur costruiti a posteriori).

Prendiamo una submartingala a tempo discreto $M = (M_n)_{n\geq 0}$, rispetto ad una filtrazione $(\mathcal{F}_n)_{n\geq 0}$:

$$E[M_{n+1} - M_n | \mathcal{F}_n] \geq 0$$

per ogni $n \in \mathbb{N}$. L'immagine intuitiva, almeno nel caso delle martingale (che vogliamo includere nel teorema), è che le sue traiettorie oscillino continuamente, per $n \to \infty$. Un modo di esplorare se abbia limite o meno, per $n \to \infty$, è quello di prendere due generiche soglie $a < b$ ed esplorare se le traiettorie "attraversano" l'intervallo $[a, b]$ infinite volte oppure no.

Osservazione 15 *Spieghiamo cosa si intende per attraversamento di $[a, b]$ dal basso verso l'alto, per una successione numerica $(x_n)_{n\geq 0}$. Analogamente a quanto vedremo sotto, poniamo*

$$\sigma_1 = \inf\{n \geq 0 : x_n < a\}, \qquad \tau_1 = \inf\{n > \sigma_1 : x_n > b\}$$

$$\sigma_{i+1} = \inf\{n > \tau_n : x_n < a\}, \qquad \tau_{i+1} = \inf\{n > \sigma_{n+1} : x_n > b\}, \qquad i \geq 1$$

con la convenzione che questi numeri sono infiniti se il corrispondente insieme è vuoto. Se $\tau_1 < \infty$, diremo che tra gli istanti σ_1 e τ_1 c'è un attraversamento di $[a, b]$ dal basso *verso l'alto; e così via se $\tau_{i+1} < \infty$. La quantità (eventualmente infinita)*

$$\gamma_{a,b} = Card\left\{i \geq 1 : \tau_i < \infty\right\}$$

rappresenta il numero di attraversamenti di $[a, b]$ dal basso verso l'alto.

Osservazione 16 *Se il numero di questi attraversamenti è finito, qualsiasi siano $a < b$, allora esiste $\lim_{n\to\infty} x_n$ (finito o infinito). Infatti, il limite $\lim_{n\to\infty} x_n$ esiste se e solo se i limiti superiore e inferiore coincidono (eventualmente infiniti); se questo non è verificato, allora si possono trovare (partendo da due sottosuccessioni che tendono ai limiti estremi) due numeri reali $a < b$ (basta prenderli strettamente contenuti tra i due limiti estremi) e due sottosuccessioni $\{n_k^a\}$, $\left\{n_k^b\right\}$ tali che $x_{n_k^a} < a$, $x_{n_k^b} > a$, $n_k^a < n_k^b < n_{k+1}^a$ per ogni k.*

Torniamo alla submartingala. Vogliamo capire se gli attraversamenti di un generico intervallo $[a, b]$, dal basso verso l'alto, delle sue traiettorie, sono finiti o infiniti. Poniamo

$$\begin{aligned}
\sigma_1(\omega) &= \inf\left\{n \geq 0 : M_n(\omega) < a\right\} \\
\tau_1(\omega) &= \inf\left\{n > \sigma_1(\omega) : M_n(\omega) > b\right\}
\end{aligned}$$

ponendo $\sigma_1(\omega) = \infty$ o $\tau_1(\omega) = \infty$ se i relativi insiemi sono vuoti. Poi iterativamente introduciamo, per $i \geq 1$,

$$\begin{aligned}
\sigma_{i+1}(\omega) &= \inf\left\{n > \tau_n(\omega) : M_n(\omega) < a\right\} \\
\tau_{i+1}(\omega) &= \inf\left\{n > \sigma_{n+1}(\omega) : M_n(\omega) > b\right\}
\end{aligned}$$

sempre con la convenzione suddetta se gli insiemi sono vuoti. La quantità aleatoria, a valori interi non negativi o $+\infty$,

$$\gamma_{a,b}(\omega) = Card\left\{i \geq 1 : \tau_i(\omega) < \infty\right\}$$

è il numero di attraversamenti di $[a, b]$ in modo ascendente. Come detto sopra, se è finita, allora c'è limite, per la realizzazione $\left(M_n(\omega)\right)_{n\geq 0}$.

Gli istanti σ_i e τ_i sono tempi d'arresto (sono simili agli istanti di primo ingresso, solo con un vincolo dal basso (es. $n > \tau_n$ nella definizione di σ_{i+1}, ma la verifica è identica, per ricorrenza).

Osservazione 17 *Se fossero limitati, avremmo (per il teorema d'arresto)*

$$E\left[M_{\tau_i}\right] \leq E\left[M_{\sigma_{i+1}}\right].$$

Questo è impossibile: $M_{\sigma_{i+1}} < a$ e $M_{\tau_i} > b$ (se σ_i e τ_i sono finiti) quindi

$$E\left[M_{\sigma_{i+1}}\right] \leq a < b \leq E\left[M_{\tau_i}\right].$$

Quindi σ_{i+1} *e* τ_i *non sono limitati. Per avere il primo attraversamento, per certi* ω *si deve aspettare moltissimo. Questo però non è (come invece si potrebbe pensare) un primo indizio di convergenza: vale anche per il moto browniano, dove la convergenza sicuramente non vale (non basta la proprietà di martingala per garantire la convergenza e per ora abbiamo usato solo questa).*

Tronchiamo questi tempi aleatori per poter applicare i risultati d'arresto. Dato un intero positivo T (lo si immagini molto grande), introduciamo i tempi aleatori $\sigma_i \wedge T$, $\tau_i \wedge T$. Fissato T, essi sono tempi d'arresto e sono limitati: i tempi deterministici sono tempi d'arresto, il minimo di due tempi d'arresto è un tempo d'arresto, e sono ovviamente limitati da T. Allora, per il Corollario 6,

$$E\left[M_{\tau_i \wedge T}\right] \leq E\left[M_{\sigma_{i+1} \wedge T}\right] \tag{3.4}$$

per ogni i. Questo è il primo ingrediente cruciale della dimostrazione.

Il secondo parte da una stima per $\gamma_{a,b}$ in termini della martingala. La disuguaglianza più facile da capire è (non la dimostriamo perché non verrà usata)

$$\gamma_{a,b}(b-a) \leq \sum_{i=1}^{\infty} \left(M_{\tau_i} - M_{\sigma_i}\right) 1_{\tau_i < \infty}.$$

Tuttavia, come sopra, dobbiamo lavorare con tempi troncati. Introduciamo allora il numero di attraversamenti precedenti a T

$$\gamma_{a,b}^T(\omega) = Card\left\{i \geq 1 : \tau_i(\omega) < T\right\}$$

ed osserviamo che

$$\gamma_{a,b}^T(b-a) \leq \sum_{i=1}^{T} \left(M_{\tau_i \wedge T} - M_{\sigma_i \wedge T}\right) 1_{\tau_i < T}.$$

Infatti, se (relativamente ad un certo ω) ci sono esattamente k attraversamenti precedenti a T, cioè $\gamma_{a,b}^T = k$, per un qualche $k = 0, 1, ..., T$ (si noti che il numero di attraversamenti prima di T è limitato almeno da T, ed in realtà da un numero minore o uguale a $T/2$), significa che $\sigma_1 < \tau_1 < \sigma_2 < ... < \tau_k < T \leq \tau_{k+1}$ (non sappiamo però dove sta σ_{k+1} rispetto a T), e quindi (il caso $k = 0$ va fatto separatamente; lo omettiamo)

$$\begin{aligned}
\sum_{i=1}^{T} \left(M_{\tau_i \wedge T} - M_{\sigma_i \wedge T}\right) 1_{\tau_i < T} &= \sum_{i=1}^{k} \left(M_{\tau_i} - M_{\sigma_i}\right) + \left(M_T - M_{\sigma_i \wedge T}\right) 1_{\tau_{k+1} < T} \\
&\geq k(b-a) = \gamma_{a,b}^T(b-a).
\end{aligned}$$

Purtroppo però il termine $1_{\tau_i < T}$ produce delle complicazioni più tardi (si pensi al fatto che volgiamo usare la (3.4)). Eppure questo termine è cruciale per cancellare l'addendo

$(M_T - M_{\sigma_i \wedge T}) 1_{\tau_{k+1}<T}$. Un'alternativa è la disuguaglianza

$$\gamma_{a,b}^T (b-a) \leq \sum_{i=1}^T (M_{\tau_i \wedge T} - M_{\sigma_i \wedge T}) - \inf_{i=1,\ldots,T} (M_T - M_{\sigma_i \wedge T})$$

che si dimostra esattamente come la precedente. Si noti che, se sapessimo che $M_n \geq a$ per ogni n, allora sarebbe

$$\inf_{i=1,\ldots,T} (M_T - M_{\sigma_i \wedge T}) \geq 0$$

perché, se $\sigma_i < T$ allora $M_{\sigma_i \wedge T} < a \leq M_T$, mentre se $\sigma_i \geq T$ allora $M_{\sigma_i \wedge T} = M_T$. In questo caso allora avremmo

$$\gamma_{a,b}^T (b-a) \leq \sum_{i=1}^T (M_{\tau_i \wedge T} - M_{\sigma_i \wedge T}).$$

Pertanto, se riusciamo a ricondurci al caso di una submartingala tale che $M_n \geq a$ per ogni n, allora una delle disuguaglianze cruciali si semplifica enormemente. Ma questo è sempre possibile: data una submartingala M e dati $a < b$, basta introdurre

$$M_n^a = (M_n - a)^+ + a.$$

Il processo $M_n - a$ è ovviamente una submartingala e quindi lo è anche $(M_n - a)^+$ per il teorema sulle trasformazioni convesse crescenti e quindi anche M_n^a. la submartingala M_n^a soddisfa $M_n^a \geq a$ per ogni n. Inoltre, la quantità $\gamma_{a,b}^T$ non varia con queste trasformazioni. Vale quindi

$$\gamma_{a,b}^T (b-a) \leq \sum_{i=1}^T \left(M_{\tau_i \wedge T}^a - M_{\sigma_i \wedge T}^a\right). \tag{3.5}$$

Il terzo ingrediente, che ora forse si può immaginare mettendo insieme i due precedenti, è la semplice identità

$$\begin{aligned} M_T - M_{\sigma_1 \wedge T} &= \sum_{i=1}^T \left(M_{\sigma_{i+1} \wedge T} - M_{\tau_i \wedge T} + M_{\tau_i \wedge T} - M_{\sigma_i \wedge T}\right) \\ &= \sum_{i=1}^T \left(M_{\sigma_{i+1} \wedge T} - M_{\tau_i \wedge T}\right) + \sum_{i=1}^T (M_{\tau_i \wedge T} - M_{\sigma_i \wedge T}). \end{aligned}$$

Osserviamo solo, circa la sua validità, che $M_{\sigma_{T+1} \wedge T} = M_T$ perché non ci possono essere T attraversamenti entro il tempo T. In realtà, vista la (3.5), utilizzeremo l'analoga identità per il processo M^a.

Mettendo insieme i tre ingredienti, abbiamo

$$M_T^a - M_{\sigma_1 \wedge T}^a \geq \sum_{i=1}^{T} \left(M_{\sigma_{i+1} \wedge T}^a - M_{\tau_i \wedge T}^a \right) + \gamma_{a,b}^T (b-a)$$

da cui, ricordando (3.4) (che vale ovviamente anche per M^a),

$$E\left[M_T^a\right] - E\left[M_{\sigma_1 \wedge T}^a\right] \geq E\left[\gamma_{a,b}^T\right](b-a).$$

Non si scordi che M^a non è una martingala, altrimenti avremmo $E\left[M_T^a\right] - E\left[M_{\sigma_1 \wedge T}^a\right] = 0$ che implicherebbe $E\left[\gamma_{a,b}^T\right] = 0$ (assurdo negli esempi).

Possiamo proseguire la dimostrazione così: vale $E\left[M_{\sigma_1 \wedge T}^a\right] \geq E\left[M_0^a\right]$ per il teorema di arresto per submartingale, quindi

$$\begin{aligned} E\left[\gamma_{a,b}^T\right] &\leq \frac{1}{b-a}\left(E\left[M_T^a\right] - E\left[M_0^a\right]\right) = \frac{1}{b-a}\left(E\left[(M_T-a)^+\right] - E\left[(M_0-a)^+\right]\right) \\ &\leq \frac{1}{b-a} E\left[(M_T-a)^+\right]. \end{aligned}$$

Si arriva alla stessa conclusione senza il teorema di arresto, con la seguente osservazione:

$$M_T^a - M_{\sigma_1 \wedge T}^a \leq (M_T - a)^+.$$

Questa disuguaglianza è ovvia se $\sigma_1 \geq T$, mentre se $\sigma_1 < T$ vale perché

$$M_T^a - M_{\sigma_1 \wedge T}^a = M_T^a - M_{\sigma_1}^a = (M_T-a)^+ + a - (M_{\sigma_1} - a)^+ + a \leq (M_T - a)^+.$$

A questo punto abbiamo dimostrato una stima per il numero medio di attraversamenti, prima di un tempo T qualsiasi. La stima non dipende da T e quindi è intuitivamente plausibile che si deduca che il numero di attraversamenti è finito. Bisogna poi escludere la possibilità di limite infinito, se vogliamo un teorema sulla convergenza. Nel prossimo paragrafo sistemiamo tutti questi i dettagli.

3.6.2 Risultati rigorosi

Siano $\gamma_{a,b}$ e $\gamma_{a,b}^T$ le grandezze aleatorie definite nella sottosezione precedente: numero di attraversamenti in salita di $[a,b]$ e numero prima del tempo T.

Lemma 4 *Se $M = (M_n)_{n \geq 0}$ è una submartingala a tempo discreto, allora per ogni $a < b$ e $T > 0$ intero vale*

$$E\left[\gamma_{a,b}^T\right] \leq \frac{1}{b-a} E\left[(M_T-a)^+\right]$$

$$E\left[\gamma_{a,b}\right] \leq \frac{1}{b-a} \sup_{n \in N} E\left[(M_n-a)^+\right].$$

Proof. La dimostrazione della prima disuguaglianza illustrata al paragrafo precedente è completa. Volendo sintetizzare a posteriori i suoi passi, senza le discussioni intermedie:

- si introduce la submartingala ausiliaria $M^a = (M_n - a)^+ + a$ e si verifica che vale
$$\gamma_{a,b}^T (b-a) \leq \sum_{i=1}^T \left(M^a_{\tau_i \wedge T} - M^a_{\sigma_i \wedge T} \right)$$
- si osserva che vale la semplice l'identità
$$M^a_T - M^a_{\sigma_1 \wedge T} = \sum_{i=1}^T \left(M^a_{\sigma_{i+1} \wedge T} - M^a_{\tau_i \wedge T} \right) + \sum_{i=1}^T \left(M^a_{\tau_i \wedge T} - M^a_{\sigma_i \wedge T} \right)$$
da cui si deduce
$$M^a_T - M^a_{\sigma_1 \wedge T} \geq \sum_{i=1}^T \left(M^a_{\sigma_{i+1} \wedge T} - M^a_{\tau_i \wedge T} \right) + \gamma_{a,b}^T (b-a)$$
- usando il teorema d'arresto si deduce
$$E\left[\gamma_{a,b}^T\right] \leq \frac{1}{b-a} \left(E\left[M^a_T\right] - E\left[M^a_{\sigma_1 \wedge T}\right] \right)$$
- ed infine si conclude osservando che $M^a_T - M^a_{\sigma_1 \wedge T} \leq (M_T - a)^+$.

La stima per $E\left[\gamma_{a,b}\right]$ discende dalla precedente e dal teorema di convergenza monotona, visto che $\gamma_{a,b} = \lim_{n\to\infty} \gamma_{a,b}^n$, e che $E\left[\gamma_{a,b}^T\right] \leq \frac{1}{b-a} \sup_{n\in N} E\left[(M_n - a)^+\right]$ per ogni $T > 0$. La dimostrazione è completa. ■

Teorema 15 *Se $M = (M_n)_{n\geq 0}$ è una submartingala a tempo discreto tale che*
$$\sup_{n\in\mathbb{N}} E\left[M_n^+\right] < \infty$$
allora la successione $(M_n(\omega))_{n\geq 0}$ converge q.c., per $n \to \infty$.

Proof. Per il lemma, $\gamma_{a,b} < \infty$ q.c., per ogni $a < b$ prefissati. Ne discende subito che
$$P\left(\gamma_{a,b} < \infty, \forall a < b, a, b \in \mathbb{Q}\right) = 1.$$
E' un facile esercizio verificare che l'affermazione $\gamma_{a,b} < \infty$, $\forall a < b$, $a, b \in \mathbb{Q}$ implica l'affermazione $\gamma_{a,b} < \infty$, $\forall a < b$, $a, b \in \mathbb{R}$. Quindi
$$P\left(\gamma_{a,b} < \infty, \forall a < b\right) = 1.$$

Abbiamo già dimostrato (osservazione 16) che se vale $\gamma_{a,b}(\omega) < \infty$, $\forall a < b$, allora esiste $\lim_{n\to\infty} M_n(\omega)$. Resta quindi solo da dimostrare che tale limite è finito, q.c.

Ricordiamo il lemma di Fatou: $E\left[\liminf_{n\to\infty} X_n\right] \leq \liminf_{n\to\infty} E[X_n]$, ad esempio se le v.a. X_n sono non negative (si accetta il caso di valori infiniti, ad esempio di $\liminf_{n\to\infty} X_n$). Allora

$$E\left[\lim_{n\to\infty} |M_n|\right] \leq \liminf_{n\to\infty} E[|M_n|] \leq \sup_{n\in\mathbb{N}} E[|M_n|].$$

Se avessimo ipotizzato $\sup_{n\in\mathbb{N}} E[|M_n|] < \infty$, concluderemmo subito: dedurremmo che $\lim_{n\to\infty} |M_n| < \infty$ q.c., come volevamo. Questa condizione però discende dall'ipotesi del teorema, usando nuovamente la proprietà di submartingala: siccome $|x| = 2x^+ - x$ per ogni numero reale x, abbiamo

$$E[|M_n|] = 2E\left[M_n^+\right] - E[M_n] \leq 2E\left[M_n^+\right] - E[M_0] \leq 2\sup_{n\in\mathbb{N}} E\left[M_n^+\right] - E[M_0]$$

da cui $\sup_{n\in\mathbb{N}} E[|M_n|] < \infty$. La dimostrazione è completa. ■

I seguenti corollari sono immediati.

Corollario 7 *Se per una martingala (o sub o super martingala) M esiste una costante $C > 0$ tale che $E[|M_n|] \leq C$ per ogni n, allora M_n converge.*

Corollario 8 *Una martingala (o super martingala) positiva converge.*

Vediamo un teorema sulle martingale chiuse a tempo discreto.

Corollario 9 ***Teorema 16*** *Data una v.a. X integrabile ed una filtrazione $(\mathcal{F}_n)_{n\geq 0}$, il processo $M_n = E[X|\mathcal{F}_n]$ è una martingala, vale $E[|M_n|] \leq E[|X|]$ e quindi M_n converge. Inoltre, il limite è $E[X|\mathcal{F}_\infty]$ dove $\mathcal{F}_\infty = \vee_n \mathcal{F}_n$. In particolare $E[X|\mathcal{F}_n]$ converge ad X se X è $\mathcal{F}_\infty$-misurabile.*

Proof. Abbiamo già dimostrato che M è una martingala, esempio 3. Lì avevamo già osservato che

$$E[|M_n|] \leq E[E[|X|\,|\mathcal{F}_n]] = E[|X|]$$

quindi l'ipotesi del Corollario 7 è soddisfatta. Ne discende che M_n converge q.c. Indichiamo con Z il suo limite.

Per dimostrare in totale generalità che $Z = E[X|\mathcal{F}_\infty]$ dovremmo usare il concetto di uniforme integrabilità ed alcune verifiche supplementari, che omettiamo. Per semplicità, limitiamoci a dimostrare l'ultima parte del teorema nel caso in cui X sia di quadrato integrabile: $E[X^2] < \infty$. Vale come sopra

$$E\left[M_n^2\right] \leq E\left[E\left[X^2|\mathcal{F}_n\right]\right] = E\left[X^2\right]$$

fatto che useremo tra un momento.

Dobbiamo dimostrare che $Z = E[X|\mathcal{F}_\infty]$. Usiamo la definizione di speranza condizionale. Innanzi tutto, Z è $\mathcal{F}_\infty$-misurabile. Infatti, M_n è $\mathcal{F}_n$ e quindi $\mathcal{F}_\infty$-misurabile, e Z è limite quasi certo di M_n. Ora basta verificare che $E[Z1_A] = E[X1_A]$ per ogni $A \in \mathcal{F}_\infty$. E' sufficiente dimostrarlo per ogni A appartenente ad una base di $\mathcal{F}_\infty$. Il fatto che basti verificarlo per una base è molto simile al ragionamento fatto in un'altra occasione, per l'indipendenza tra σ-algebre, che non riportiamo (l'idea è considerare l'insieme degli eventi $A \in \mathcal{F}$ tali che $E[Z1_A] = E[X1_A]$ e mostrare che è una σ-algebra).

Sia $\mathcal{A}$ la famiglia $\bigcup_n \mathcal{F}_n$, che non è una σ-algebra (non è $\mathcal{F}_\infty$). E' però una base di $\mathcal{F}_\infty$. Prendiamo $A \in \mathcal{A}$, precisamente $A \in \mathcal{F}_{n_0}$ per un certo n_0. Vale $E\left[(M_n 1_A)^2\right] \leq E[M_n^2] \leq E[X^2]$, e $M_n 1_A \to Z1_A$ q.c., quindi per un teorema di Vitali possiamo passare al limite sotto il segno di valore atteso: $E[M_n 1_A] \to E[Z1_A]$. Pertanto

$$\begin{aligned} E[Z1_A] &= \lim_{n\to\infty} E[M_n 1_A] = \lim_{n\to\infty} E[E[X|\mathcal{F}_n] 1_A] = \lim_{n\to\infty} E[E[1_A X|\mathcal{F}_n]] \\ &= \lim_{n\to\infty} E[1_A X] = E[1_A X]. \end{aligned}$$

Abbiamo usato, al terzo passaggio, il fatto che nel calcolo del limite bastano i valori di n maggiori di n_0 e per essi 1_A è $\mathcal{F}_n$-misurabile (in quanto $\mathcal{F}_{n_0}$-misurabile). La dimostrazione è completa. ■

3.7 Altri risultati per martingale a tempo continuo

Avendo dimostrato in grande dettaglio le affermazioni precedenti, per non appesantire la trattazione ci limitiamo qui ad un elenco di enunciati. Alcuni sono facili esercizi, altri meno. Supponiamo completo lo spazio di probabilità $(\Omega, \mathcal{F}, P)$.

Teorema 17 *Sia $D \subset [0,\infty)$ un insieme numerabile denso. Sia $M = (M_t)_{t\in D}$ una submartingala rispetto ad una filtrazione $(\mathcal{F}_t)_{t\in D}$. Supponiamo*

$$\sup_{t\in D} E[|M_t|] < \infty.$$

Allora, su un insieme di probabilità uno, esistono finiti il limite per $t \to \infty$ ed i limiti destro e sinistro per $t \to t_0$ in ogni punto t_0 ($t \in D$).

Teorema 18 *Sia $M = (M_t)_{t\geq 0}$ una submartingala continua a destra. Supponiamo*

$$\sup_{t\in D} E[|M_t|] < \infty.$$

Allora, q.c., esistono finiti il limite per $t \to \infty$ ed i limiti sinistri in ogni punto t_0.

Teorema 19 *Sia $M = (M_t)_{t\geq 0}$ una martingala rispetto ad una filtrazione $(\mathcal{F}_t)_{t\geq 0}$ che soddisfi le condizioni abituali. Allora esiste una modificazione càdlàg di M, che è ancora una martingala.*

Teorema 20 *Sia $M = (M_t)_{t\geq 0}$ una martingala continua a destra, rispetto ad una filtrazione $(\mathcal{F}_t)_{t\geq 0}$, e $\tau_1 \leq \tau_2$ due tempi d'arresto limitati. Allora*

$$E\left[M_{\tau_2}|\mathcal{F}_{\tau_1}\right] = M_{\tau_1}.$$

Se τ è un tempo d'arresto q.c. finito (non necessariamente limitato), allora $(M_{t\wedge\tau})_{t\geq 0}$ è una martingala rispetto a $(\mathcal{F}_t)_{t\geq 0}$.

Ricordiamo che, se $(M_t)_{t\geq 0}$ una martingala continua e di quadrato integrabile, allora $(M_t^2)_{t\geq 0}$ è una submartingala. Nell'ottica del teorema di decomposizione di Doob visto sopra nel caso discreto, ci chiediamo se M_t^2 si decomponga. Il processo crescente A_t che ne emerge, è di fondamentale importanza, che ricorda il concetto di variazione quadratica.

Teorema 21 *Sia $(M_t)_{t\geq 0}$ una martingala continua e di quadrato integrabile, rispetto a $(\mathcal{F}_t)_{t\geq 0}$; si supponga che $(\mathcal{F}_t)_{t\geq 0}$ sia completa. Allora vale la decomposizione*

$$M_t^2 = N_t + A_t$$

dove $(N_t)_{t\geq 0}$ è una martingala continua ed $(A_t)_{t\geq 0}$ è un processo crescente continuo, con $A_0 = 0$. La decomposizione è unica e vale, in probabilità,

$$A_t = \lim_{|\pi_k|\to 0} \sum_{i=0}^{n_k-1} \left|M_{t_{i+1}} - M_{t_i}\right|^2$$

dove $\{\pi_k\}_{k\in\mathbb{N}}$ è una successione di partizioni di $[0,t]$ della forma $\pi_k = \{0 = t_0 < t_1 < ... < t_{n_k} = t\}$.

3.8 Martingale locali e semimartingale

Abbiamo visto sopra che se $M = (M_t)_{t\geq 0}$ è una martingala continua a destra, rispetto ad una filtrazione $(\mathcal{F}_t)_{t\geq 0}$, e τ è un tempo d'arresto q.c. finito, allora $(M_{t\wedge\tau})_{t\geq 0}$ è una martingala rispetto a $(\mathcal{F}_t)_{t\geq 0}$. Questo risultato dà lo spunto per una nuova definizione.

Diciamo che un processo $(M_t)_{t\geq 0}$ è una *martingala locale* rispetto a $(\mathcal{F}_t)_{t\geq 0}$ se esiste una successione crescente $\{\tau_n\}_{n\in\mathbb{N}}$ di tempi d'arresto, con $\lim_{n\to\infty} \tau_n = +\infty$, tale che $(M_{t\wedge\tau_n})_{t\geq 0}$ è una martingala rispetto a $(\mathcal{F}_t)_{t\geq 0}$, per ogni $n \in \mathbb{N}$. Diciamo che i tempi τ_n *riducono* M.

Una martingala è una martingala locale ($\tau_n = +\infty$ per ogni n). Per noi l'interesse maggiore per questo concetto risiederà nel fatto che gli integrali stocastici di processi abbastanza generali sono martingale locali, ma non sempre martingale. Un esempio molto particolare di questa situazione è il seguente, che comunque illustra un'idea generale, quella di troncamento o localizzazione.

Esempio 4 *Sia B un moto browniano rispetto a $(\mathcal{F}_t)_{t\geq 0}$ e sia X_0 una v.a. $\mathcal{F}_0$-misurabile. Poniamo*

$$M_t = X_0 W_t.$$

Se ipotizziamo che X_0 sia di quadrato integrabile, o in questo semplice esempio anche solo integrabile, allora M_t è integrabile perché prodotto di v.a. indipendenti integrabili, e vale

$$E[X_0 W_t | \mathcal{F}_s] = X_0 E[W_t | \mathcal{F}_s] = X_0 W_s$$

cioè M_t è una martingala. Se invece non supponiamo alcuna integrabilità di X_0, M_t non sarà in generale integrabile e quindi non possiamo porci il problema se sia o meno una martingala. Introduciamo però il tempo aleatorio

$$\tau_n = \begin{cases} \infty & se \quad |X_0| \leq n \\ 0 & se \quad |X_0| > n \end{cases}.$$

E' un tempo d'arresto:

$$\begin{aligned} \{\tau_n \leq t\} &= (\{\tau_n \leq t\} \cap \{|X_0| \leq n\}) \cup (\{\tau_n \leq t\} \cap \{|X_0| > n\}) \\ &= \varnothing \cup \{|X_0| > n\} \in \mathcal{F}_0 \subset \mathcal{F}_t \end{aligned}$$

per ogni $t \geq 0$. Inoltre, $\tau_n \to \infty$ q.c. Infine,

$$M_{t\wedge\tau_n} = X_0 W_{t\wedge\tau_n} = \begin{cases} X_0 W_t & se \quad |X_0| \leq n \\ 0 & se \quad |X_0| > n \end{cases}$$

quindi

$$M_{t\wedge\tau_n} = X_0 1_{|X_0|\leq n} W_t.$$

Il processo $M_{t\wedge\tau_n}$ è quindi una martingala, perché della forma $Y_0 W_t$ con Y_0 integrabile.

Diciamo per brevità che un processo è a variazione limitata se q.c. le sue traiettorie hanno variazione limitata su ogni intervallo finito.

Diciamo infine che un processo $(X_t)_{t\geq 0}$ è una *semimartingala* rispetto a $(\mathcal{F}_t)_{t\geq 0}$ se è la somma di una martingala locale e di un processo adattato a variazione limitata. In base ad uno dei risultati precedenti, un importante esempio di semimartingala è M_t^2, dove M è una martingala continua. Vedremo più in generale nel seguito che anche $f(M_t)$ con $f \in C^2$ è una semimartingala, ed ancor più in generale vedremo che se X è una semimartingala continua, allora lo è anche $f(X_t)$ con $f \in C^2$. Per questa ed altre ragioni, la classe delle semimartingale è in un certo senso chiusa per il cosiddetto calcolo stocastico, contiene gli oggetti di maggior interesse (moto browniano, integrali stocastici) e quindi è la classe privilegiata in cui svolgere il calcolo.

Si può dimostrare il seguente risultato. Diciamo che una semimartingala è continua se è somma di una martingala locale continua e di un processo continuo a variazione limitata.

Teorema 22 *Sia* $(X_t)_{t\geq 0}$ *una semimartingala continua. La decomposizione in martingala locale e processo a variazione limitata è unica. Dato* t *e data una successione* $\{\pi_k\}_{k\in\mathbb{N}}$ *di partizioni di* $[0,t]$ *della forma* $\pi_k = \{0 = t_0 < t_1 < ... < t_{n_k} = t\}$, *esistono in probabilità i seguenti limiti e coincidono*

$$\lim_{|\pi_k|\to 0} \sum_{i=0}^{n_k-1} \left|X_{t_{i+1}} - X_{t_i}\right|^2 = \lim_{|\pi_k|\to 0} \sum_{i=0}^{n_k-1} \left|M_{t_{i+1}} - M_{t_i}\right|^2 .$$

La grandezza

$$\langle X\rangle_t = P - \lim_{|\pi_k|\to 0} \sum_{i=0}^{n_k-1} \left|X_{t_{i+1}} - X_{t_i}\right|^2$$

è detta variazione quadratica di X *su* $[0,t]$.

Non vedremo la dimostrazione nel suo complesso ed in totale generalità, ma sicuramente esamineremo alcune classi di processi (definiti tramite integrali stocastici) per cui alcune di queste affermazioni saranno dimostrate. Già conosciamo una prima classe: i processi della forma

$$X_t = B_t + V_t$$

dove B è un moto browniano e V è un processo continuo a variazione limitata. Infatti, per B abbiamo dimostrato che esiste la variazione quadratica (anche come limite in media quadratica) e, sempre nel capitolo del moto browniano, abbiamo visto che una funzione f continua a variazione limitata ha la proprietà $\lim_{|\pi_k|\to 0} \sum_{i=0}^{n_k-1} \left|f(t_{i+1}) - f(t_i)\right|^2 = 0$, quindi per V esiste $\lim_{|\pi_k|\to 0} \sum_{i=0}^{n_k-1} \left|V_{t_{i+1}} - V_{t_i}\right|^2$ quasi certamente (quindi anche in probabilità), e vale zero.

Per la stessa ragione, la coincidenza dei due limiti nel teorema è un fatto ovvio, non appena uno dei due esiste. Quello che è un po' laborioso è mostrare che esiste $\langle M\rangle_t$ per ogni martingala locale continua. Ci accontentiamo di saperlo per il moto browniano e, come vedremo, per gli integrali stocastici.

La grandezza $\langle X\rangle_t$ caratterizza molte proprietà del processo X. Si presti attenzione a questo fatto quando sarà il momento.

Capitolo 4

L'integrale stocastico

4.1 Integrale di processi elementari

Lo scopo di questo capitolo è quello di definire ed esaminare un integrale del tipo

$$\int_a^b X_t dB_t$$

dove B è un moto browniano ed X un processo, soggetto ad opportune ipotesi (ma non più regolare di B, altrimenti la teoria non sarebbe applicabile alle equazioni stocastiche).

Nel capitolo sul moto browniano abbiamo già discusso le difficoltà che si incontrano a tentare di definire tale integrale usando concetti classici, come l'integrale di Lebesgue-Stieltjes. Vediamo come K. Itô, nei primi anni quaranta, ha risolto questo problema.

Usiamo le notazioni già introdotte nel capitolo delle martingale. Sia $B = (B_t)_{t\geq 0}$ un moto browniano rispetto ad una filtrazione $(\mathcal{F}_t)_{t\geq 0}$. Chiamiamo *processo elementare* ogni processo stocastico $X = (X_t)_{t\geq 0}$ adattato a $(\mathcal{F}_t)_{t\geq 0}$ ma costante a tratti relativamente ad una sequenza di tempi $t_n \geq t_{n-1} \geq ... \geq t_1 = 0$, cioè ogni processo della forma

$$X_t = \sum_{i=1}^{n-1} X_{t_i} 1_{[t_i,t_{i+1})}(t), \qquad X_{t_i} \text{ misurabile rispetto a } \mathcal{F}_{t_i}.$$

Se in più richiediamo

$$E\left[X_{t_i}^2\right] < \infty \text{ per ogni } i = 1, ..., n$$

allora lo diremo *di quadrato integrabile.*

Per ogni $b \geq a \geq 0$, poniamo

$$\int_a^b X_s dB_s = \sum_{i=1}^{n-1} X_{t_i} \left(B_{(a\vee t_{i+1})\wedge b} - B_{(a\vee t_i)\wedge b}\right).$$

Come già osservato in quel capitolo, siccome le espressioni $(a \vee t_{i+1}) \wedge b$ e $(a \vee t_i) \wedge b$ sono troppo laboriose, si può adottare la seguente convenzione. Fissati $b \geq a \geq 0$, si può arricchire la sequenza di tempi considerando la nuova sequenza $t'_{n+2} \geq t'_{n-1} \geq ... \geq t'_1 = 0$ formata dai tempi t_i e da a, b, e riscrivendo X_t nella forma $X_t = \sum_{i=1}^{n+1} X'_{t'_i} 1_{[t'_i, t'_{i+1})}(t)$ con le v.a. $X'_{t'_i}$ misurabili rispetto a $\mathcal{F}_{t'_i}$ così definite: se $t'_i = t_j \in \{t_1, ..., t_n\}$, poniamo $X'_{t'_i} = X_{t_j}$; se $t'_i = a$ (risp. $t'_i = b$) e $t_j \in \{t_1, ..., t_n\}$ è il più grande con la proprietà $t_j \leq a$ (risp. $t_j \leq b$), allora poniamo $X'_{t'_i} = X_{t_j}$.

Qui, per semplicità di notazione, indichiamo la nuova sequenza ancora con $t_n \geq t_{n-1} \geq ... \geq t_1 = 0$; oppure, se si vuole, fissati $b \geq a \geq 0$, si considerano solo quelle sequenze $t_n \geq t_{n-1} \geq ... \geq t_1 = 0$ che contengono a e b (una terza strategia è di considerare suddivisioni dell'intervallo $[a, b]$ solamente, invece che processi definiti per tutti i $t \geq 0$).

Poniamo allora

$$\int_a^b X_s dB_s = \sum_{a \leq t_i \leq t_{i+1} \leq b} X_{t_i} \left(B_{t_{i+1}} - B_{t_i} \right).$$

Con un po' di pazienza si può riconoscere che le due definizioni ora scritte di $\int_a^b X_s dB_s$ coincidono. Possiamo anche introdurre l'insieme di indici

$$J_{a,b} = \{i = 1, ..., n : a \leq t_i \leq t_{i+1} \leq b\}$$

e scrivere

$$\int_a^b X_s dB_s = \sum_{i \in J_{a,b}} X_{t_i} \left(B_{t_{i+1}} - B_{t_i} \right).$$

Dati $c \geq b \geq a$, si può verificare con pazienza che vale

$$\int_a^c X_s dB_s = \int_a^b X_s dB_s + \int_b^c X_s dB_s$$

(omettiamo i dettagli). Il risultato forse più importante di questo inizio di teoria è il seguente. La seconda identità è detta a volte *formula di isometria.*

Proposizione 33 *Sia X un processo elementare di quadrato integrabile. Allora la v.a. $\int_a^b X_s dB_s$ ha media e varianza finite e vale*

$$\begin{aligned} E\left[\int_a^b X_s dB_s\right] &= 0 \\ E\left[\left(\int_a^b X_s dB_s\right)^2\right] &= \int_a^b E\left[X_s^2\right] ds. \end{aligned}$$

Proof. Ricordiamo che, date due v.a. X, Y indipendenti e integrabili, il loro prodotto XY è integrabile e vale $E[XY] = E[X]E[Y]$. Ciascun addendo $X_{t_i}\left(B_{t_{i+1}} - B_{t_i}\right)$ che compone $\int_a^b X_s dB_s$ è di tale forma, perché $B_{t_{i+1}} - B_{t_i}$ è indipendente da $\mathcal{F}_{t_i}$ ed X_{t_i} è $\mathcal{F}_{t_i}$-misurabile, quindi $\int_a^b X_s dB_s$ è integrabile, e vale

$$E\left[\int_a^b X_s dB_s\right] = \sum_{i \in J_{a,b}} E\left[X_{t_i}\left(B_{t_{i+1}} - B_{t_i}\right)\right] = \sum_{i \in J_{a,b}} E\left[X_{t_i}\right] E\left[B_{t_{i+1}} - B_{t_i}\right] = 0$$

in quanto gli incrementi browniani hanno attesa nulla.

Vale poi, con la notazione $\Delta_i B = B_{t_{i+1}} - B_{t_i}$,

$$\left(\int_a^b X_s dB_s\right)^2 = \sum_{i \in J_{a,b}} \sum_{j \in J_{a,b}} X_{t_i} X_{t_j} \Delta_i B \Delta_j B = \sum_{i \in J_{a,b}} X_{t_i}^2 \left(\Delta_i B\right)^2 + 2 \sum_{i,j \in J_{a,b}, i<j} X_{t_i} X_{t_j} \Delta_i B \Delta_j B.$$

Gli addendi $X_{t_i}^2 \left(\Delta_i B\right)^2$ sono integrabili perché $X_{t_i}^2$ e $\left(\Delta_i B\right)^2$ sono indipendenti (trasformazioni di v.a. indipendenti) ed integrabili ($X_{t_i}^2$ per ipotesi, $\left(\Delta_i B\right)^2$ perché gli incrementi browniani hanno tutti i momenti finiti). Esaminiamo l'integrabilità degli addendi $X_{t_i} X_{t_j} \Delta_i B \Delta_j B$. Le v.a. X_{t_i} e $\Delta_i B$ sono indipendenti, quindi anche $X_{t_i}^2$ e $\left(\Delta_i B\right)^2$, che sono integrabili, quindi $X_{t_i} \Delta_i B$ è di quadrato integrabile. Anche X_{t_j} è di quadrato integrabile, quindi $X_{t_i} X_{t_j} \Delta_i B$ è integrabile, essendo prodotto di v.a. di quadrato integrabile. Siccome $X_{t_i} X_{t_j} \Delta_i B$ è indipendente da $\Delta_j B$ (come spiegheremo tra un attimo), allora $X_{t_i} X_{t_j} \Delta_i B \Delta_j B$ è integrabile. Questo completa la verifica che $\left(\int_a^b X_s dB_s\right)^2$ è integrabile. L'indipendenza tra $X_{t_i} X_{t_j} \Delta_i B$ e $\Delta_j B$ si spiega così: la v.a. $\Delta_j B$ è indipendente da $\mathcal{F}_{t_j}$ ed $X_{t_i} X_{t_j} \Delta_i B$ è $\mathcal{F}_{t_j}$-misurabile (X_{t_i} è $\mathcal{F}_{t_i}$-misurabile, quindi $\mathcal{F}_{t_j}$-misurabile perché $i < j$, $\Delta_i B$ è $\mathcal{F}_{t_{i+1}}$-misurabile, quindi $\mathcal{F}_{t_j}$-misurabile perché $i + 1 \leq j$).

Usando questi stessi fatti, vale

$$\begin{aligned} E\left[\left(\int_a^b X_s dB_s\right)^2\right] &= \sum_{i \in J_{a,b}} E\left[X_{t_i}^2\right] E\left[\left(\Delta_i B\right)^2\right] + 2 \sum_{i,j \in J_{a,b}, i<j} E\left[X_{t_i} X_{t_j} \Delta_i B\right] E\left[\Delta_j B\right] \\ &= \sum_{i \in J_{a,b}} E\left[X_{t_i}^2\right] \left(t_{i+1} - t_i\right) = \int_a^b E\left[X_s^2\right] ds. \end{aligned}$$

Abbiamo usato anche le proprietà $E\left[\left(\Delta_i B\right)^2\right] = t_{i+1} - t_i$, $E\left[\Delta_j B\right] = 0$. Si osservi che

$$\int_a^b E\left[X_s^2\right] ds = \sum_{i \in J_{a,b}} \int_{t_i}^{t_{i+1}} E\left[X_s^2\right] ds = \sum_{i \in J_{a,b}} E\left[X_{t_i}^2\right] \left(t_{i+1} - t_i\right).$$

La dimostrazione è completa. ■

Vale in realtà una versione rafforzata del risultato precedente:

Proposizione 34 *Sia X un processo elementare di quadrato integrabile. Allora*

$$E\left[\int_s^t X_u dB_u | \mathcal{F}_s\right] = 0$$

$$E\left[\left(\int_s^t X_u dB_u\right)^2 \middle| \mathcal{F}_s\right] = E\left[\int_s^t X_u^2 du \middle| \mathcal{F}_s\right].$$

Proof. Sappiamo già che $\int_s^t X_u dB_u$ e $\left(\int_s^t X_u dB_u\right)^2$ sono integrabili. Vale

$$E\left[\int_s^t X_u dB_u | \mathcal{F}_s\right] = \sum_{i \in J_{s,t}} E\left[X_{t_i} \Delta_i B | \mathcal{F}_s\right]$$

ma, contrariamente a varie situazioni apparentemente simili, non possiamo ad esempio portare X_{t_i} fuori dalla speranza condizionale, perché t_i è più grande di s. Se però riscriviamo

$$E\left[X_{t_i} \Delta_i B | \mathcal{F}_s\right] = E\left[E\left[X_{t_i} \Delta_i B | \mathcal{F}_{t_i}\right] | \mathcal{F}_s\right]$$

allora vale

$$E\left[X_{t_i} \Delta_i B | \mathcal{F}_s\right] = E\left[X_{t_i} E\left[\Delta_i B | \mathcal{F}_{t_i}\right] | \mathcal{F}_s\right] = E\left[X_{t_i} E\left[\Delta_i B\right] | \mathcal{F}_s\right] = 0$$

dove abbiamo usato il fatto che $\Delta_i B$ è indipendente da $\mathcal{F}_{t_i}$ ed è centrato. Quindi $E\left[\int_s^t X_u dB_u | \mathcal{F}_s\right] = 0$.

Sappiamo poi già che vale

$$\left(\int_s^t X_u dB_u\right)^2 = \sum_{i \in J_{s,t}} X_{t_i}^2 \left(\Delta_i B\right)^2 + 2 \sum_{i,j \in J_{s,t}, i<j} X_{t_i} X_{t_j} \Delta_i B \Delta_j B.$$

Allora, usando lo stesso trucco e le solite proprietà si ottiene

$$\begin{aligned}
E\left[\left(\int_s^t X_u dB_u\right)^2 \middle| \mathcal{F}_s\right] &= \sum_{i \in J_{s,t}} E\left[X_{t_i}^2 \left(\Delta_i B\right)^2 | \mathcal{F}_s\right] + 2 \sum_{i,j \in J_{s,t}, i<j} E\left[X_{t_i} X_{t_j} \Delta_i B \Delta_j B | \mathcal{F}_s\right] \\
&= \sum_{i \in J_{s,t}} E\left[E\left[X_{t_i}^2 \left(\Delta_i B\right)^2 | \mathcal{F}_{t_i}\right] | \mathcal{F}_s\right] + 2 \sum_{i,j \in J_{s,t}, i<j} E\big[E\big[X_{t_i} X_{t_j} \Delta_i B \Delta_j B | \mathcal{F}_{t_j}\big] \,|\mathcal{F} \\
&= \sum_{i \in J_{s,t}} E\left[X_{t_i}^2 E\left[\left(\Delta_i B\right)^2\right] | \mathcal{F}_s\right] + 2 \sum_{i,j \in J_{s,t}, i<j} E\left[X_{t_i} X_{t_j} \Delta_i B E\left[\Delta_j B\right] | \mathcal{F}_s\right] \\
&= \sum_{i \in J_{s,t}} E\left[X_{t_i}^2 | \mathcal{F}_s\right] \left(t_{i+1} - t_i\right) = E\left[\sum_{i \in J_{s,t}} X_{t_i}^2 \left(t_{i+1} - t_i\right) | \mathcal{F}_s\right]
\end{aligned}$$

da cui la tesi. La dimostrazione è completa. ■

La prima parte del seguente corollario era già stata dimostrata, in modo un po' diverso, nel capitolo delle martingale.

Corollario 10 *Sia X un processo elementare di quadrato integrabile. Allora $M_t = \int_0^t X_s dB_s$ è una martingala rispetto a $(\mathcal{F}_t)_{t\geq 0}$. Inoltre, $M_t^2 - \int_0^t X_u^2 du$ è una martingala rispetto a $(\mathcal{F}_t)_{t\geq 0}$.*

Proof. E' adattato perché, ad ogni istante $t \geq 0$, $\int_0^t X_s dB_s$ è somma di v.a. $\mathcal{F}_t$-misurabili. E' integrabile. Vale poi, ricordando che per $t \geq s \geq 0$ abbiamo $\int_0^t X_u dB_u = \int_0^s X_u dB_u + \int_s^t X_u dB_u$,

$$E\left[M_t - M_s | \mathcal{F}_s\right] = E\left[\int_s^t X_u dB_u, |\mathcal{F}_s\right] = 0.$$

Quindi M è una martingala.

Per ragioni simili, anche $M_t^2 - \int_0^t X_u^2 du$ è adattato. M_t^2 è integrabile. $\int_0^t X_u^2 du$ è integrabile perché (per il teorema di Fubini)

$$E\left[\int_0^t X_u^2 du\right] = \int_0^t E\left[X_u^2\right] du < \infty.$$

Vale poi

$$\begin{aligned} E\left[M_t^2 - \int_0^t X_u^2 du | \mathcal{F}_s\right] &= E\left[(M_t - M_s)^2 - \int_s^t X_u^2 du | \mathcal{F}_s\right] \\ &\quad + E\left[2 M_t M_s - M_s^2 - \int_0^s X_u^2 du | \mathcal{F}_s\right] \\ &= 2 M_s E\left[M_t | \mathcal{F}_s\right] - M_s^2 - \int_0^s X_u^2 du = M_s^2 - \int_0^s X_u^2 du \end{aligned}$$

quindi $M_t^2 - \int_0^t X_u^2 du$ è una martingala. La dimostrazione è completa. ■

Con un linguaggio introdotto al termine del capitolo sulle martingale, si può dire che $\int_0^t X_u^2 du$ è il processo crescente associato alla submartigala M_t^2. Elenchiamo a questo proposito alcune situazioni viste fino ad ora.

1. Se M_n è una martingala a tempo discreto, di quadrato integrabile, allora per la sub-martigala M_n^2 vale la decomposizione di Doob $M_n^2 = N_n + A_n$, dove A_n è crescente. Quindi esiste un processo crescente A_n tale che $M_n^2 - A_n$ è una martingala.

2. Se B è un moto browniano, allora $B_t^2 - t$ è una martingala (facile esercizio oppure caso particolare del Corollario 10). Il processo $A_t = t$ è crescente.

3. Se X è un processo elementare di quadrato integrabile ed introduciamo la martingala $M_t = \int_0^t X_s dB_s$, allora $M_t^2 - \int_0^t X_u^2 du$ è una martingala (e $A_t = \int_0^t X_u^2 du$ è un processo crescente).

4. In generale abbiamo enunciato un teorema che afferma che, se M è una martingala continua, di quadrato integrabile, allora esiste un processo crescente A_t tale che $M_t^2 - A_t$ è una martingala.

4.1.1 Estensione a integratori martingale

Proviamo ad estendere la teoria precedente al caso dell'integrale del tipo $\int_a^b X_t dM_t$ dove M è una martingala rispetto ad una filtrazione $(\mathcal{F}_t)_{t\geq 0}$.

Supponiamo che X sia una *processo elementare limitato*, cioè della forma $X_t = \sum_{i=1}^{n-1} X_{t_i} 1_{[t_i,t_{i+1})}(t)$ dove le v.a. X_{t_i}, $\mathcal{F}_{t_i}$-misurabili, sono limitate. Supponiamo che la sequenza $t_n \geq t_{n-1} \geq ... \geq t_1 = 0$ contenga $b \geq a \geq 0$ e poniamo

$$\int_a^b X_t dM_t = \sum_{i\in J_{a,b}} X_{t_i} \left(M_{t_{i+1}} - M_{t_i}\right).$$

Supponiamo inoltre che la martingala sia continua e di quadrato integrabile. Pertanto esiste uno ed un solo processo crescente A_t tale che $M_t^2 - A_t$ sia una martingala (se non si vuole usare tale teorema, dal momento che non è stato dimostrato, si prenda l'esistenza di A_t come un'ipotesi sulla martingala M). Vale quindi

$$\begin{aligned} E\left[(M_t - M_s)^2 - (A_t - A_s)|\mathcal{F}_s\right] &= E\left[M_t^2 - A_t|\mathcal{F}_s\right] - 2M_s E\left[M_t|\mathcal{F}_s\right] + M_s^2 + A_s \\ &= M_s^2 - A_s - 2M_s^2 + M_s^2 + A_s = 0 \end{aligned}$$

ovvero

$$E\left[(M_t - M_s)^2|\mathcal{F}_s\right] = E\left[(A_t - A_s)|\mathcal{F}_s\right].$$

Proposizione 35 *La v.a. $\int_s^t X_u dM_u$ è di quadrato integrabile e vale*

$$E\left[\left.\int_s^t X_u dM_u\right|\mathcal{F}_s\right] = 0$$

$$E\left[\left.\left(\int_s^t X_u dM_u\right)^2\right|\mathcal{F}_s\right] = E\left[\left.\int_s^t X_u^2 dA_u\right|\mathcal{F}_s\right].$$

Proof. Dalla formula

$$\left(\int_s^t X_u dM_u\right)^2 = \sum_{i\in J_{s,t}} X_{t_i}^2 (\Delta_i M)^2 + 2\sum_{i,j\in J_{s,t}, i<j} X_{t_i} X_{t_j} \Delta_i M \Delta_j M$$

si vede che $\left(\int_s^t X_u dM_u\right)^2$ è integrabile (le X_{t_i} sono limitate). Allora, usando le solite proprietà della speranza condizionale e la proprietà di martingala per M, vale

$$\begin{aligned} E\left[\left.\int_s^t X_u dM_u\right|\mathcal{F}_s\right] &= \sum_{i\in J_{s,t}} E\left[X_{t_i}\Delta_i M|\mathcal{F}_s\right] = \sum_{i\in J_{s,t}} E\left[E\left[X_{t_i}\Delta_i M|\mathcal{F}_{t_i}\right]|\mathcal{F}_s\right] \\ &= \sum_{i\in J_{s,t}} E\left[X_{t_i} E\left[\Delta_i M|\mathcal{F}_{t_i}\right]|\mathcal{F}_s\right] = 0 \end{aligned}$$

(in quanto $E\left[\Delta_i M|\mathcal{F}_{t_i}\right]=0$) ed analogamente

$$\begin{aligned}
E\left[\left(\int_s^t X_u dM_u\right)^2\middle|\mathcal{F}_s\right] &= \sum_{i\in J_{s,t}} E\left[X_{t_i}^2\left(\Delta_i M\right)^2|\mathcal{F}_s\right]+2\sum_{i,j\in J_{s,t},i<j} E\left[X_{t_i}X_{t_j}\Delta_i M\Delta_j M|\mathcal{F}_s\right] \\
&= \sum_{i\in J_{s,t}} E\left[E\left[X_{t_i}^2\left(\Delta_i M\right)^2|\mathcal{F}_{t_i}\right]|\mathcal{F}_s\right]+2\sum_{i,j\in J_{s,t},i<j} E\left[E\left[X_{t_i}X_{t_j}\Delta_i M\Delta_j M|\mathcal{F}_{t_j}\right.\right. \\
&= \sum_{i\in J_{s,t}} E\left[X_{t_i}^2 E\left[\left(\Delta_i M\right)^2|\mathcal{F}_{t_i}\right]|\mathcal{F}_s\right]+2\sum_{i,j\in J_{s,t},i<j} E\left[X_{t_i}X_{t_j}\Delta_i M E\left[\Delta_j M|\mathcal{F}_{t_j}\right]\right. \\
&= \sum_{i\in J_{s,t}} E\left[X_{t_i}^2|\mathcal{F}_s\right]\left(A_{t_{i+1}}-A_{t_i}\right)=E\left[\sum_{i\in J_{s,t}} X_{t_i}^2\left(A_{t_{i+1}}-A_{t_i}\right)|\mathcal{F}_s\right]
\end{aligned}$$

da cui la tesi. La dimostrazione è completa. ■

4.2 Integrale di processi progressivamente misurabili, di quadrato integrabile

Sia $B=(B_t)_{t\geq 0}$ un moto browniano rispetto ad una filtrazione $(\mathcal{F}_t)_{t\geq 0}$. Dati $b\geq a\geq 0$, diciamo che un processo $X=(X_t)_{t\geq 0}$ è di classe $M_B^2(a,b)$ se è progressivamente misurabile e vale

$$E\left[\int_a^b X_t^2 dt\right]<\infty.$$

Per dare questa definizione basterebbe che il processo fosse definito e progressivamente misurabile su $[a,b]$. I processi elementari, cioè della forma $X_t=\sum_{i=1}^{n-1} X_{t_i}1_{[t_i,t_{i+1})}(t)$ con X_{t_i} misurabile rispetto a $\mathcal{F}_{t_i}$, sono progressivamente misurabili; sono di classe $M_B^2(a,b)$ se e solo se sono di quadrato integrabile su $[a,b]$, cioè $E\left[X_{t_i}^2\right]<\infty$ per ogni $i\in J_{a,b}$, in quanto

$$E\left[\int_a^b X_t^2 dt\right]=\sum_{i\in J_{a,b}} E\left[X_{t_i}^2\right](t_{i+1}-t_i).$$

Si potrebbe anche dire che $M_B^2(a,b)$ è la famiglia delle classi di equivalenza dei processi suddetti, identificando processi X,X' tali che $E\left[\int_a^b (X_t-X_t')^2 dt\right]=0$; lo spazio $M_B^2(a,b)$ con il prodotto scalare $\langle X,X'\rangle=E\left[\int_a^b X_t X_t' dt\right]$ risulterebbe uno spazio di Hilbert. Non useremo esplicitamente queste convenzioni e questi fatti.

Teorema 23 *Se X è di classe $M_B^2(a,b)$, allora esiste una successione di processi elementari $X^{(n)}$, anch'essi di classe $M_B^2(a,b)$, tale che*

$$\lim_{n\to\infty} E\left[\int_a^b \left(X_t-X_t^{(n)}\right)^2 dt\right]=0. \tag{4.1}$$

Esiste anche una successione di processi continui $X^{(n)}$, di classe $M_B^2(a,b)$, che converge ad X nel senso (4.1).

La dimostrazione è piuttosto laboriosa anche se sostanzialmente elementare. Bisogna prima dimostrare che X si può approssimare con processi continui adattati, poi che un processo continuo e adattato si può approssimare con processi costanti a tratti adattati. La prima cosa si fa tramite convoluzione con mollificatori unilaterali (cioè che integrano solo la parte di processo precedente all'istante in cui si calcola la convoluzione). La seconda si fa semplicemente prendendo il valore del processo continuo nell'estremo sinistro di ogni intervallino di una partizione, sempre più fine. Vanno però verificate numerose proprietà, anche un po' laboriose, per cui omettiamo i dettagli, anche perché simili a quelli di tante dimostrazioni di Analisi relative ai più svariati spazi di funzioni. Vediamo invece le conseguenze di questo teorema.

Proposizione 36 *Se $X^{(n)}$ è una successione di processi elementari di classe $M_B^2(a,b)$ che converge ad un processo X di classe $M_B^2(a,b)$ nel senso (4.1), allora la successione di v.a.*

$$\int_a^b X_t^{(n)} dB_t$$

è di Cauchy in $L^2(\Omega,\mathcal{F},P)$. Il suo limite, elemento di $L^2(\Omega,\mathcal{F},P)$, dipende solo da X (non dalla successione scelta) e verrà indicato con $\int_a^b X_t dB_t$, detto integrale di Itô di X.

Proof. Per la formula di isometria dimostrata per i processi elementari, vale

$$E\left[\left(\int_a^b \left(X_t^{(n)} - X_t^{(m)}\right) dB_t\right)^2\right] = \int_a^b E\left[\left(X_t^{(n)} - X_t^{(m)}\right)^2\right] dt$$

da cui discende subito che $\int_a^b X_t^{(n)} dB_t$ è una successione di Cauchy in $L^2(\Omega,\mathcal{F},P)$, quindi ha limite, essendo questo spazio completo. Se $\widetilde{X}^{(n)}$ è un'altra successione con le stesse caratteristiche convergente ad X, allora la successione $Y^{(n)}$ data da $Y^{(2n)} = X^{(n)}$, $Y^{(2n+1)} = \widetilde{X}^{(n)}$ ha le stesse caratteristiche e converge ad X, nel senso (4.1). Quindi $\int_a^b Y_t^{(n)} dB_t$ è una successione di Cauchy in $L^2(\Omega,\mathcal{F},P)$. Il suo limite deve essere lo stesso delle due sottosuccessioni di indici pari e dispari, limiti che quindi coincidono. La dimostrazione è completa. ■

Osservazione 18 *L'integrale stocastico $\int_a^b X_t dB_t$ è, per definizione, una classe di equivalenza, o se si preferisce è una v.a. definita a meno di modificazioni su insiemi di misura nulla. Non ha senso, fissato un certo $\omega \in \Omega$, chiedersi quanto valga $\left(\int_a^b X_t dB_t\right)(\omega)$. In particolare, non ha senso scrivere cose del tipo $\left(\int_a^b X_t dB_t\right)(\omega) = \int_a^b X_t(\omega) dB_t(\omega)$ (per i processi elementari invece ha senso e vale l'identità). L'unica* *teoria che, al momento attuale, è in grado di dare un significato a questo tipo di scritture, per particolari processi X, è quella dei rough paths, però molto più laboriosa di quella ora sviluppata.*

Teorema 24 *Sia X di classe $M_B^2(0,T)$, per un certo $T>0$. Allora*

$$\begin{aligned} E\left[\int_0^T X_s dB_s\right] &= 0 \\ E\left[\left(\int_0^T X_s dB_s\right)^2\right] &= \int_0^T E\left[X_s^2\right] ds. \end{aligned}$$

Più in generale, per ogni $0\le s\le t\le T$ vale

$$E\left[\int_s^t X_u dB_u|\mathcal{F}_s\right]=0$$

$$E\left[\left(\int_s^t X_u dB_u\right)^2\middle|\mathcal{F}_s\right]=E\left[\int_s^t X_u^2 du\middle|\mathcal{F}_s\right].$$

Inoltre, $M_t=\int_0^t X_s dB_s$ e $M_t^2-\int_0^t X_u^2 du$ sono martingale rispetto a $(\mathcal{F}_t)_{t\ge 0}$, quindi $\int_0^t X_u^2 du$ è il processo crescente associato alla submartigala M_t^2.

Proof. Sia $X^{(n)}$ è una successione di processi elementari di classe $M_B^2(0,T)$ che converge ad X nel senso (4.1) (con $a=0$, $b=T$). Allora $\int_0^T X_s^{(n)} dB_s$ converge a $\int_0^T X_s dB_s$ in $L^2(\Omega,\mathcal{F},P)$, quindi (tramite semplici disuguaglianze) $E\left[\int_0^T X_s^{(n)} dB_s\right]$ converge a $E\left[\int_0^T X_s dB_s\right]$, $E\left[\left(\int_0^T X_s^{(n)} dB_s\right)^2\right]$ converge a $E\left[\left(\int_0^T X_s dB_s\right)^2\right]$ e $\int_0^T E\left[\left(X_s^{(n)}\right)^2\right] ds$ converge a $\int_0^T E\left[X_s^2\right] ds$. Da questo si deducono le prime due identità.

Per mostrare le altre, si prendano (senza restrizione) i processi $X^{(n)}$ aventi t ed s come nodi, e convergenti ad X su $[0,T]$ nel senso (4.1); ne segue che convergono ad X anche su $[s,t]$, nello stesso senso. Quindi $\int_s^t X_u^{(n)} dB_u$ converge a $\int_s^t X_u dB_u$ in $L^2(\Omega,\mathcal{F},P)$. Inoltre, a causa di (4.1), $\int_s^t \left(X_u^{(n)}\right)^2 du$ converge a $\int_s^t X_u^2 du$ in $L^2(\Omega,\mathcal{F},P)$. Grazie a questi fatti si può passare al limite nelle speranze condizionali e verificare le due identità. Infatti, ricordiamo che se $Z_n\to Z$ in $L^1(\Omega,\mathcal{F},P)$, allora $E[Z_n|\mathcal{G}]\to E[Z|\mathcal{G}]$ in $L^1(\Omega,\mathcal{F},P)$:

$$E\left[|E[Z_n|\mathcal{G}]-E[Z|\mathcal{G}]|\right]=E\left[|E[Z_n-Z|\mathcal{G}]|\right]\le E\left[E\left[|Z_n-Z|\,|\mathcal{G}\right]\right]=E\left[|Z_n-Z|\right]\to 0.$$

Le affermazioni finali sulle proprietà di martingala sono ovvie conseguenze delle identità precedenti, come nel caso dei processi elementari. La dimostrazione è completa. ■

Nel seguito diremo che X è di classe M_B^2 se è di classe $M_B^2(0,T)$ per ogni $T>0$.

Teorema 25 *Sia X di classe M_B^2. Allora il processo $M_t = \int_0^t X_s dB_s$, $t \geq 0$, ha una versione continua.*

Proof. Sia $X^{(n)}$ è una successione di processi elementari di classe $M_B^2(0,T)$ che converge ad X nel senso (4.1) (con $a = 0$, $b = T$). I processi $M_t^{(n)} = \int_0^t X_s^{(n)} dB_s$ sono martingale continue. La continuità si vede subito dalla definizione che abbiamo usato meno spesso: se $X_t = \sum_{i=1}^{n-1} X_{t_i} 1_{[t_i,t_{i+1})}(t)$, $\int_0^t X_s dB_s = \sum_{i=1}^{n-1} X_{t_i} B_{t_{i+1}\wedge t} - B_{t_i \wedge t}$. Quindi vale la disuguaglianza massimale per la submartingala $\left|M_t^{(n)} - M_t^{(m)}\right|^2$:

$$\begin{aligned} P\left(\sup_{t\in[0,T]}\left|M_t^{(n)} - M_t^{(m)}\right| > \varepsilon\right) &= P\left(\sup_{t\in[0,T]}\left|M_t^{(n)} - M_t^{(m)}\right|^2 > \varepsilon^2\right) \\ &\leq \frac{1}{\varepsilon^2} E\left[\left|M_T^{(n)} - M_T^{(m)}\right|^2\right] \\ &= \frac{1}{\varepsilon^2}\int_0^T E\left[\left|X_s^{(n)} - X_s^{(m)}\right|^2\right] ds \end{aligned}$$

dove all'ultimo passaggio abbiamo usato la proprietà di isometria. Da questa disuguaglianza, deduciamo la seguente affermazione: esiste una sottosuccessione $M_t^{(n_k)}$ che converge q.c. ad un processo continuo N_t, uniformemente in t:

$$P\left(\lim_{k\to\infty}\sup_{t\in[0,T]}\left|M_t^{(n_k)} - N_t\right| = 0\right) = 1. \tag{4.2}$$

L'idea intuitiva (formalizzabile con l'uso di opportuni spazi funzionali) è: la disuguaglianza precedente indica che $M^{(n)}_{\cdot}$ è di Cauchy in probabilità nello spazio $C([0,T];\mathbb{R})$ munito della topologia della convergenza uniforme; dall'essere di Cauchy in probabilità discende che converge in probabilità e quindi ha una sottosuccessione che converge quasi certamente.

Verifichiamolo "a mano". Lasciamo al lettore il fatto elementare di costruire una successione $\{n_k\}$ tale che

$$P\left(\sup_{t\in[0,T]}\left|M_t^{(n_k)} - M_t^{(n_{k+1})}\right| > \frac{1}{2^k}\right) \leq \frac{1}{k^2}.$$

Da questo discende che

$$\sum_{k=1}^{\infty} P\left(\sup_{t\in[0,T]}\left|M_t^{(n_k)} - M_t^{(n_{k+1})}\right| > \frac{1}{2^k}\right) < \infty$$

e quindi, per il lemma di Borel-Cantelli (prima parte), esiste un evento $\Omega_0 \in \mathcal{F}$ di probabilità uno tale che, per ogni $\omega \in \Omega_0$, vale

$$\sup_{t\in[0,T]}\left|M_t^{(n_k)}(\omega) - M_t^{(n_{k+1})}(\omega)\right| \leq \frac{1}{2^k}$$

definitivamente in k. Quindi $\left\{M_t^{(n_k)}(\omega)\right\}_{k\in\mathbb{N}}$ è di Cauchy in $C([0,T];\mathbb{R})$ munito della topologia della convergenza uniforme; quindi converge ad una funzione continua $N_t(\omega)$. Fissato, t, la convergenza q.c. di v.a. implica che il limite è misurabile, è una v.a., quindi N_t è un processo stocastico; continuo (q.c.; in realtà va definito per $\omega \notin \Omega_0$ e possiamo definirlo identicamente nullo, quindi è continuo per ogni ω). Abbiamo dimostrato l'asserzione (4.2).

Basta ora ricordare che, fissato $t \in [0,T]$, la v.a. $M_t^{(n_k)}$ converge in $L^2(\Omega,\mathcal{F},P)$ a M_t. Siccome converge q.c. a N_t, vale $P(N_t = M_t) = 1$ (entrambe le convergenze implicano quella in probabilità, ed il limite in probabilità è unico, come classe di equivalenza). Quindi N, che è un processo continuo, è una modificazione di M. La dimostrazione è completa. ■

4.3 Integrale di processi progressivamente misurabili più generali

Sia $B = (B_t)_{t\geq 0}$ un moto browniano rispetto ad una filtrazione $(\mathcal{F}_t)_{t\geq 0}$. Dati $b \geq a \geq 0$, diciamo che un processo $X = (X_t)_{t\geq 0}$ è di classe $\Lambda_B^2(a,b)$ se è progressivamente misurabile e vale

$$P\left(\int_a^b X_t^2 dt < \infty\right) = 1$$

I processi elementari, cioè della forma $X_t = \sum_{i=1}^{n-1} X_{t_i} 1_{[t_i,t_{i+1})}(t)$ con X_{t_i} misurabile rispetto a $\mathcal{F}_{t_i}$, sono progressivamente misurabili e di classe $\Lambda_B^2(a,b)$, senza bisogno di alcuna ipotesi di integrabilità sulle X_{t_i}.

Si potrebbe anche dire che $\Lambda_B^2(a,b)$ è la famiglia delle classi di equivalenza dei processi suddetti, identificando processi X, X' tali che $P\left(\int_a^b (X_t - X_t')^2 dt < \infty\right) = 1$. Non useremo esplicitamente questa convenzione.

Teorema 26 *Se X è di classe $\Lambda_B^2(a,b)$, allora esiste una successione di processi elementari $X^{(n)}$ tale che*

$$\lim_{n\to\infty} \int_a^b \left(X_t - X_t^{(n)}\right)^2 dt = 0 \tag{4.3}$$

quasi certamente. Esiste anche una successione di processi continui $X^{(n)}$, di classe $\Lambda_B^2(a,b)$, che converge ad X nel senso (4.3).

La dimostrazione è simile a quella dei processi di classe $M_B^2(a,b)$, forse un po' più facile perché non bisogna tenere sotto controllo certe integrabilità rispetto ad ω. E' comunque molto lunga e di moderato interesse probabilistico, per cui la omettiamo.

Vogliamo ora definire l'integrale stocastico $\int_a^b X_t dB_t$, usando questo teorema. Un modo è quello di basarsi sul seguente lemma molto interessante.

Lemma 5 *Sia X un processo elementare. Allora, per ogni $\varepsilon, \rho > 0$, vale*

$$P\left(\left|\int_a^b X_t dB_t\right| > \varepsilon\right) \leq P\left(\int_a^b X_t^2 dt > \rho\right) + \frac{\rho}{\varepsilon^2}.$$

Proof. Separiamo Ω nei due eventi $A = \left\{\int_a^b X_t^2 dt > \rho\right\}$ ed $A^c = \left\{\int_a^b X_t^2 dt \leq \rho\right\}$:

$$P\left(\left|\int_a^b X_t dB_t\right| > \varepsilon\right) = P\left(\left\{\left|\int_a^b X_t dB_t\right| > \varepsilon\right\} \cap A\right) + P\left(\left\{\left|\int_a^b X_t dB_t\right| > \varepsilon\right\} \cap A^c\right)$$

e maggioriamo il primo termine con $P(A)$:

$$P\left(\left|\int_a^b X_t dB_t\right| > \varepsilon\right) = P\left(\int_a^b X_t^2 dt > \rho\right) + P\left(\left\{\left|\int_a^b X_t dB_t\right| > \varepsilon\right\} \cap A^c\right).$$

Resta da verificare che

$$P\left(\left|\int_a^b X_t dB_t\right| > \varepsilon, \int_a^b X_t^2 dt \leq \rho\right) \leq \frac{\rho}{\varepsilon^2}.$$

Introduciamo il processo Y_t definito così (con la solita notazione $X_t = \sum_{i=1}^{n-1} X_{t_i} 1_{[t_i, t_{i+1})}(t)$):

$$Y_t = \begin{cases} X_t & \text{per} \quad t \leq t_{i+1} \in [t_i, t_{i+1}) \quad \text{se} \quad \int_a^{t_{i+1}} X_t^2 dt \leq \rho \\ 0 & \text{altrimenti} \end{cases}$$

(per $i \in J_{a,b}$). Vale $\int_a^b Y_t^2 dt \leq \rho$. Inoltre, se

$$\left\{\int_a^b X_t^2 dt \leq \rho\right\} \subset \{Y = X \text{ su } [a,b]\} \subset \left\{\int_a^b X_t dB_t = \int_a^b Y_t dB_t\right\}$$

(per la verifica dell'ultima inclusione si pensi alla definizione di integrale stocastico per processi elementari). Quindi

$$\begin{aligned} P\left(\left|\int_a^b X_t dB_t\right| > \varepsilon, \int_a^b X_t^2 dt \leq \rho\right) &\leq P\left(\left|\int_a^b Y_t dB_t\right| > \varepsilon, \int_a^b X_t^2 dt \leq \rho\right) \\ &\leq P\left(\left|\int_a^b Y_t dB_t\right| > \varepsilon\right) \leq \frac{E\left[\left|\int_a^b Y_t dB_t\right|^2\right]}{\varepsilon^2} \\ &= \frac{E\left[\int_a^b Y_t^2 dt\right]}{\varepsilon^2} \leq \frac{\rho}{\varepsilon^2}. \end{aligned}$$

La dimostrazione è completa. ■

Corollario 11 *Se $X^{(n)}$ è una successione di processi elementari che converge ad un processo X di classe $\Lambda_B^2(a,b)$ nel senso (4.3), allora la successione di v.a.*

$$\int_a^b X_t^{(n)} dB_t$$

è di Cauchy in probabilità. Il suo limite in probabilità dipende solo da X (non dalla successione scelta) e verrà indicato con $\int_a^b X_t dB_t$, detto integrale di Itô di X.

Proof. Abbiamo

$$P\left(\left|\int_a^b \left(X_t^{(n)} - X_t^{(m)}\right) dB_t\right| > \varepsilon\right) \leq P\left(\int_a^b \left(X_t^{(n)} - X_t^{(m)}\right)^2 dt > \rho\right) + \frac{\rho}{\varepsilon^2}.$$

La convergenza quasi certa (4.3), implica che, preso $\rho > 0$, esiste $n_\rho \in \mathbb{N}$ tale che

$$P\left(\int_a^b \left(X_t^{(n)} - X_t^{(m)}\right)^2 dt > \rho\right) \leq \rho$$

per ogni $n, m \geq n_\rho$ (si pensi ad una delle varie ragioni). Quindi, fissato $\varepsilon > 0$,

$$P\left(\left|\int_a^b \left(X_t^{(n)} - X_t^{(m)}\right) dB_t\right| > \varepsilon\right) \leq \rho + \frac{\rho}{\varepsilon^2}$$

per ogni $n, m \geq n_\rho$. Relativamente al valore scelto di ε, prendiamo $\rho_\varepsilon = \varepsilon^3$:

$$P\left(\left|\int_a^b \left(X_t^{(n)} - X_t^{(m)}\right) dB_t\right| > \varepsilon\right) \leq \varepsilon^3 + \varepsilon$$

per ogni $n, m \geq n_{\rho_\varepsilon}$. Questo significa che $\int_a^b X_t^{(n)} dB_t$ è di Cauchy in probabilità. La dimostrazione è completa. ■

Si può dimostrare facilmente (omettiamo i dettagli):

Proposizione 37 *Sia X un processo di classe $\Lambda_B^2(a,b)$,. Allora, per ogni $\varepsilon, \rho > 0$, vale*

$$P\left(\left|\int_a^b X_t dB_t\right| > \varepsilon\right) \leq P\left(\int_a^b X_t^2 dt > \rho\right) + \frac{\rho}{\varepsilon^2}.$$

In particolare, se X_t^n è una successione di processi di classe $\Lambda_B^2(a,b)$ tale che $\int_a^b (X_t^n - X_t)^2 dt$ tende a zero in probabilità, allora $\int_a^b X_t^n dB_t$ converge a $\int_a^b X_t dB_t$ in probabilità.

Nel seguito diremo che X è di classe Λ_B^2 se è di classe $\Lambda_B^2(0,T)$ per ogni $T > 0$. Ricordiamo che un processo $M = (M_t)_{t\geq 0}$ è una martingala locale se c'è una successione di tempi d'arresto τ_n, crescente a $+\infty$, tale che $M_{t\wedge\tau_n}$ è una martingala per ogni n; diremo che una tale successione di tempi d'arresto riduce la martingala locale.

Teorema 27 *Sia X di classe Λ_B^2. Allora il processo $M_t = \int_0^t X_s dB_s$, $t \geq 0$, ha una versione continua ed è una martingala locale. Una successione di tempi d'arresto che riduce la martingala locale è data da*

$$\tau_n = \inf \left\{ t > 0 : \int_0^t X_s^2 ds \geq n \right\}$$

se l'insieme non è vuoto, altrimenti $\tau_n = +\infty$.

Proof. Diamo solo alcuni elementi della dimostrazione. Le v.a. τ_n sono tempi di arresto (basta considerare il processo $Y_t = \int_0^t X_s^2 ds$ e il suo istante di ingresso in $[n, \infty)$), sono crescenti (Y_t è crescente) e tendono a $+\infty$. Infatti, per ogni $T > 0$, esiste $n_0(\omega)$ tale che per ogni $n \geq n_0(\omega)$ vale $\tau_n(\omega) \geq T$, quasi certamente. Questo perché

$$A_{n,T} := \left\{ \int_0^T X_s^2 ds \leq n \right\} = \{\tau_n \geq T\}$$

e l'evento

$$A_T := \bigcup_{n \in \mathbb{N}} A_{n,T} = \left\{ \int_0^T X_s^2 ds < \infty \right\}$$

ha probabilità uno per la proprietà Λ_B^2.

Il processo $Y_{n,t} = X_t 1_{t \leq \tau_n}$ (simile a quello usato nella dimostrazione del Lemma 5) è di classe M_B^2:

$$\int_0^T Y_{n,t}^2 ds = \int_0^T X_s^2 1_{s \leq \tau_n} ds = \int_0^{T \wedge \tau_n} X_s^2 ds \leq n$$

quindi $E\left[\int_0^T Y_{n,t}^2 ds\right] \leq n$ (si verifica anche la progressiva misurabilità di Y_t). Allora $M_{n,t} = \int_0^t Y_{n,s} dB_s$ è una martingala continua (applichiamo la convenzione che il simbolo indichi la versione continua).

Su $A_{n,T}$, $Y_{n,t} = X_t$ per ogni $t \in [0, T]$. Per il Lemma 6 esposto poco sotto, applicato su ogni intervallo $[0, t]$, vale

$$M_t = \int_0^t X_s dB_s = \int_0^t Y_{n,s} dB_s = M_{n,t}$$

su $A_{n,T}$ quasi certamente, per ogni $t \in [0, T]$. Si noti che in questa affermazione, il "quasi certamente" dipende da t. Se $A_{n,T}$ avesse probabilità uno, avremmo finito: $M_{n,t}$ sarebbe una versione continua di M_t.

Se $m > n$, siccome $Y_{m,t} = Y_{n,t}$ per ogni $t \in [0, T]$ su $A_{n,T}$, per il Lemma 6, applicato su ogni intervallo $[0, t]$, vale

$$M_{m,t} = \int_0^t Y_{m,s} dB_s = \int_0^t Y_{n,s} dB_s = M_{n,t}$$

su $A_{n,T}$ quasi certamente, per ogni $t \in [0,T]$; con un ragionamento basato sui razionali, si vede che vale

$$M_{m,t} = M_{n,t} \text{ per ogni } t \in [0,T]$$

quasi certamente su $A_{n,T}$. E che la proprietà si può rendere uniforme in m. Pertanto, intersecando su ambo gli indici n ed m, esiste un insieme misurabile N di probabilità nulla tale che per ogni n, per ogni $\omega \in A_{n,T} \backslash N$,

$$M_{m,t}(\omega) = M_{n,t}(\omega) \text{ per ogni } t \in [0,T] \text{ ed } m \geq n.$$

Su $A_T \backslash N$ definiamo un processo $(M_{\infty,t})_{t\in[0,T]}$ nel seguente modo: preso $\omega \in A_T \backslash N$, poniamo

$$M_{\infty,t}(\omega) = M_{n,t}(\omega)$$

dove n è un qualsiasi indice tale che $\omega \in A_{n,T} \backslash N$ (per quanto visto poco sopra, questa definizione non dipende da n). Per le cose osservate precedentemente, $M_{\infty,t}$ è un processo continuo. Inoltre, è una versione di M_t. Infatti, fissato $t \in [0,T]$, vale

$$\begin{aligned} P(M_{\infty,t} = M_t) &= \lim_{n\to\infty} P(\{M_{\infty,t} = M_t\} \cap A_{n,T} \backslash N) \\ &= \lim_{n\to\infty} P(\{M_{n,t} = M_t\} \cap A_{n,T} \backslash N) = \lim_{n\to\infty} P(A_{n,T} \backslash N) = 1. \end{aligned}$$

Questo dimostra la prima affermazione del teorema, cioè che M ha una versione continua.

Sapere che la famiglia di classi di equivalenza $(M_t^0)_{t\in[0,T]}$, $M_t^0 := \int_0^t X_s dB_s$, ha una versione continua M_t, aiuta a dare un significato univoco al processo stoppato $M_{t\wedge\tau_n}$. Infatti, M_t, continuo, diventa ora definito univocamente a meno di indistinguibilità, quindi la posizione $(M_{t\wedge\tau_n})(\omega) := M_{t\wedge\tau_n(\omega)}(\omega)$ non pone interrogativi di significato.

Usando allora la versione continua, vale (omettiamo la verifica dettagliata della seconda uguaglianza)

$$M_{t\wedge\tau_n} = \int_0^{t\wedge\tau_n} X_s dB_s = \int_0^t X_s 1_{s\leq\tau_n} dB_s = \int_0^t Y_{n,s} dB_s$$

e questa è una martingala in quanto $Y_{n,t}$ è un processo di classe M_B^2. La dimostrazione è completa. ■

Abbiamo usato il seguente fatto, di interesse indipendente.

Lemma 6 *Siano X, Y di classe Λ_B^2 ed $A \in \mathcal{F}$ tali che $X_t(\omega) = Y_t(\omega)$ per ogni $t \in [0,T]$ e $\omega \in A$. Allora $\int_0^T X_s dB_s = \int_0^T Y_s dB_s$ su A quasi certamente (cioè esiste un insieme di misura nulla N tale che $\int_0^T X_s dB_s = \int_0^T Y_s dB_s$ per ogni $\omega \in A \backslash N$).*

Non possiamo dare la dimostrazione completa perché poggia sul metodo di approssimazione di processi di classe Λ_B^2 con processi elementari, ma almeno vediamo l'idea.

Vale, a proposito di tale approssimazione, che si possano prendere due successioni di processi elementari X^n, Y^n che convergono a X, Y nel modo richiesto, tali che $X_t^n(\omega) = Y_t^n(\omega)$ per ogni $t \in [0,T]$ e $\omega \in A$. Da ciò discende che $\int_0^T X_s^n dB_s = \int_0^T Y_s^n dB_s$ su A. La convergenza in probabilità implica quella quasi certa, quindi dall'identità precedente e dalla convergenza quasi certa di $\int_0^T X_s^n dB_s$ e $\int_0^T Y_s^n dB_s$ a $\int_0^T X_s dB_s$ e $\int_0^T Y_s dB_s$ rispettivamente, si ottiene $\int_0^T X_s dB_s = \int_0^T Y_s dB_s$ su A.

Il lemma dice che, se pur l'integrale stocastico non è definito ω per ω ma solo con un procedimento globale in ω, tuttavia rispetta una certa forma di dipendenza locale da ω.

Concludiamo con un risultato nello stile delle *somme di Riemann.*

Proposizione 38 *Se X è un processo continuo di classe Λ_B^2 e (π_n) è una successione di partizioni di $[0,T]$, con $|\pi_n| \to 0$, con $\pi_n = \{t_i^n, i = 0, ..., N_n\}$ allora vale in probabilità*

$$\lim_{n\to\infty} \sum_{i=0}^{N_n-1} X_{t_i^n}\left(B_{t_{i+1}^n} - B_{t_i^n}\right) = \int_0^T X_s dB_s.$$

Proof. Si denoti con $X_t^{(n)}$ il processo elementare $\sum X_{t_i^n} 1_{[t_i^n, t_{i+1}^n)}$. Vale

$$\sum_{i=0}^{N_n-1} X_{t_i^n}\left(B_{t_{i+1}^n} - B_{t_i^n}\right) = \int_0^T X_s^n dB_s.$$

Inoltre, per la continuità di X, quasi certamente accade che $X_t^{(n)}$ converge ad X_t uniformemente in t, quindi vale l'ipotesi dell'affermazione di convergenza della Proposizione 37, che garantisce pertanto che $\int_0^T X_s^n dB_s \to \int_0^T X_s dB_s$ in probabilità. La dimostrazione è completa. ■

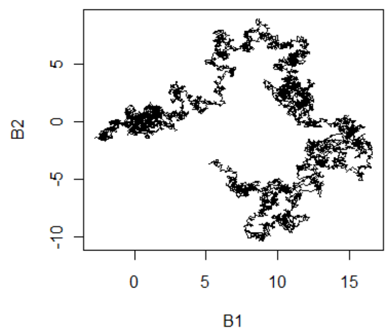
4.4 Il caso multidimensionale

4.4.1 Moto browniano in $\mathbb{R}^d$

Definizione 13 *Un moto browniano in $\mathbb{R}^d$, definito su uno spazio probabilizzato $(\Omega, \mathcal{F}, P)$, rispetto ad una filtrazione $(\mathcal{F}_t)_{t\geq 0}$ è un vettore di processi stocastici $B_t = \left(B_t^{(1)}, ..., B_t^{(d)}\right)$ tale che ciascuna componente $B_t^{(i)}$ è un moto browniano in $\mathbb{R}$ rispetto ad $(\mathcal{F}_t)_{t\geq 0}$ e le componenti sono indipendenti.*

Si possono dare varie formulazioni equivalenti, ad esempio:

Definizione 14 *Un moto browniano in $\mathbb{R}^d$, definito su uno spazio probabilizzato $(\Omega, \mathcal{F}, P)$, rispetto ad una filtrazione $(\mathcal{F}_t)_{t\geq 0}$ è un processo stocastico $(B_t)_{t\geq 0}$ su $(\Omega, \mathcal{F}, P)$ a valori in $\left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$ tale che:*



i) $P(B_0 = 0) = 1$

ii) è adattato a $(\mathcal{F}_t)_{t \geq 0}$ *e continuo*

iii) per ogni $t \geq s \geq 0$, *l'incremento* $B_t - B_s$ *è un vettore gaussiano in* $\mathbb{R}^d$ *a media nulla e matrice di covarianza pari a* $(t-s) Id$, *indipendente da* $\mathcal{F}_s$.

L'indipendenza di $B_t - B_s$ da $\mathcal{F}_s$, con $t \geq s \geq 0$ arbitrari, si può riformulare dicendo che per ogni $t_n \geq t_{n-1} \geq ... \geq t_1 \geq 0$ i vettori aleatori $B_{t_n} - B_{t_{n-1}}, \ldots , B_{t_2} - B_{t_1}, B_{t_1}$ sono indipendenti.

Essendo particolarmente elegante, mostriamo una realizzazione (approssimata) di un moto browniano in $\mathbb{R}^2$, ottenuta col software R, basata sulla formula

$$B^{(i)}_{nh} = B^{(i)}_h + \left(B^{(i)}_{2h} - B^{(i)}_h\right) + ... + \left(B^{(i)}_{nh} - B^{(i)}_{(n-1)h}\right), \qquad i = 1, 2$$

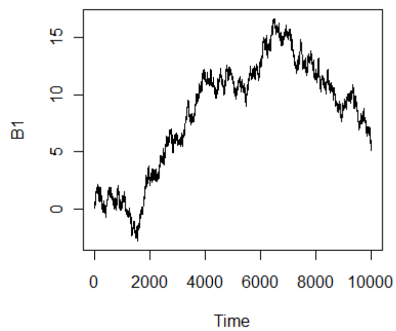
dove gli incrementi sono indipendenti (anche da una componente all'altra) e sono tutte gaussiane a media nulla e varianza h. Basta chiedere al software di generare tanti numeri gaussiani centrati aventi deviazione standard $\sqrt{h}$, tramite i comandi:

```
h=0.01
W1=rnorm(10000,0,sqrt(h))
W2=rnorm(10000,0,sqrt(h))
B1=cumsum(W1)
B2=cumsum(W2)
plot(B1,B2,type=lines)
```

A titolo di confronto, disegnamo la sua prima componente:

```
ts.plot(B1)
```

Osservazione 19 *Se* U *è una matrice ortogonale di* $\mathbb{R}^d$, *allora* UB_t *è ancora un moto browniano in* $\mathbb{R}^d$. *Lo si può vedere usando la seconda definizione.*



4.4.2 Integrale stocastico rispetto ad un moto browniano in $\mathbb{R}^d$

Dato un moto browniano $B_t = \left(B_t^{(1)}, ..., B_t^{(d)}\right)$ in $\mathbb{R}^d$ rispetto ad una filtrazione $(\mathcal{F}_t)_{t\geq 0}$ ed un processo stocastico X_t di classe Λ^2 (qui omettiamo l'indicazione B in quanto fuorviante), progressivamente misurabile rispetto a $(\mathcal{F}_t)_{t\geq 0}$, sono ben definiti tutti gli integrali di Itô

$$\int_0^t X_s dB_s^{(i)}$$

per $i = 1, ..., d$. [Si noti che qui è importante essere svincolati dalla filtrazione naturale del moto browniano rispetto a cui si integra.] Se consideriamo un processo stocastico $X_t = \left(X_t^{(1)}, ..., X_t^{(d)}\right)$ a valori in $\mathbb{R}^d$ avente tutte le componenti di classe Λ^2 (diremo che il processo è di classe Λ^2; analogamente per la classe M^2), è ben definita la somma di integrali stocastici

$$\int_0^t X_s \cdot dB_s := \sum_{i=1}^{d} \int_0^t X_s^{(i)} dB_s^{(i)}.$$

E' una martingala locale continua e, se X è di classe M^2, è anche una martingala, centrata. Vediamo che forma assume la formula di isometria. Intanto si noti il seguente fatto, forse apparentemente innaturale: gli integrali $\int_0^t X_s^{(i)} dB_s^{(i)}$ e $\int_0^t X_s^{(j)} dB_s^{(j)}$ sono scorrelati, nonostante contengano processi $X_s^{(i)}$ e $X_s^{(j)}$ che possono anche coincidere (gli integrali non sono però indipendenti).

Lemma 7 *Se X è di classe M^2, allora*

$$E\left[\int_0^t X_s^{(i)} dB_s^{(i)} \int_0^t X_s^{(j)} dB_s^{(j)}\right] = 0$$

per $i \neq j$.

Proof. Per evitare troppi indici, supponiamo $i = 1$, $j = 2$. Supponiamo $X_s^{(1)}$, $X_s^{(2)}$ processi elementari e, senza restrizione, basati sulla stessa partizione $t_n \geq t_{n-1} \geq ... \geq t_1 \geq t_0 = 0$:

$$X_s^{(i)} = \sum_{k=0}^{n-1} X_k^{(i)} 1_{[t_k, t_{k+1})} \qquad \text{per } i = 1, 2.$$

Allora, con la notazione $\Delta_k B^{(i)} = B_{t_{k+1}}^{(i)} - B_{t_k}^{(i)}$, abbiamo

$$E\left[\int_0^t X_s^{(i)} dB_s^{(i)} \int_0^t X_s^{(j)} dB_s^{(j)}\right] = \sum_{k,h=0}^{n-1} E\left[X_k^{(1)} X_h^{(2)} \Delta_k B^{(1)} \Delta_h B^{(2)}\right].$$

Se $h = k$ abbiamo

$$E\left[X_k^{(1)} X_k^{(2)} \Delta_k B^{(1)} \Delta_k B^{(2)}\right] = E\left[X_k^{(1)} X_k^{(2)}\right] E\left[\Delta_k B^{(1)}\right] E\left[\Delta_k B^{(2)}\right] = 0$$

perché $X_k^{(1)} X_k^{(2)}$ è $\mathcal{F}_{t_k}$ misurabile, e gli incrementi $\Delta_k B^{(i)}$ sono indipendenti da $\mathcal{F}_{t_k}$ e quindi da $X_k^{(1)} X_k^{(2)}$ ed anche tra loro perché componenti diverse del moto browniano in $\mathbb{R}^d$.

Se $h > k$ (il caso $h < k$ è identico) abbiamo

$$E\left[X_k^{(1)} X_h^{(2)} \Delta_k B^{(1)} \Delta_h B^{(2)}\right] = E\left[X_k^{(1)} X_h^{(2)} \Delta_k B^{(1)}\right] E\left[\Delta_h B^{(2)}\right] = 0$$

perché $X_k^{(1)} X_h^{(2)} \Delta_k B^{(1)}$ è $\mathcal{F}_{t_h}$ misurabile, $\Delta_h B^{(2)}$ è indipendente da $\mathcal{F}_{t_h}$ e quindi da $X_k^{(1)} X_h^{(2)} \Delta_k B^{(1)}$. La dimostrazione nel caso di processi elementari è completa e da essi si passa al caso generale con argomenti di approssimazione simili a quelli usati già altre volte, che omettiamo. ■

Proposizione 39 *Se X è di classe M^2, allora*

$$E\left[\left(\int_0^t X_s \cdot dB_s\right)^2\right] = E\left[\int_0^t |X_s|^2 \, ds\right]$$

dove $|x|$ indica la norma euclidea del vettore $x \in \mathbb{R}^d$.

Proof. Abbiamo

$$\begin{aligned} E\left[\left(\int_0^t X_s \cdot dB_s\right)^2\right] &= E\left[\left(\sum_{i=1}^d \int_0^t X_s^{(i)} dB_s^{(i)}\right)^2\right] \\ &= \sum_{i,j=1}^d E\left[\int_0^t X_s^{(i)} dB_s^{(i)} \int_0^t X_s^{(j)} dB_s^{(j)}\right]. \end{aligned}$$

I termini con $i \neq j$ sono nulli per il lemma precedente. La somma per $i = j$ produce il risultato della formula, in quanto

$$E\left[\left(\int_0^t X_s^{(i)} dB_s^{(i)}\right)^2\right] = E\left[\int_0^t \left(X_s^{(i)}\right)^2 ds\right].$$

La dimostrazione è completa. ■

Ragionando componente per componente, sono definiti anche gli integrali stocastici della forma $\int_0^t X_s dB_s$ dove X è un processo a valori matrici $d \times k$, con componenti X^{ij} tutte di classe almeno Λ^2. Il risultato $Y_t = \int_0^t X_s dB_s$ è un processo a valori in $\mathbb{R}^k$, di componenti

$$Y_t^i = \sum_{j=1}^d \int_0^t X_s^{ij} dB_s^j.$$

Se le componenti X^{ij} sono di classe M^2, la formula di isometria in questo caso assume la seguente forma. Ora con $|\cdot|$ indicheremo la norma euclidea in $\mathbb{R}^k$ (o altra dimensione) e con $\|A\|_{HS}$ la norma Hilbert-Schmidt di una matrice $A \in \mathbb{R}^{d\times k}$, ovvero la norma euclidea in $\mathbb{R}^{d\times k}$, definita da

$$\|A\|_{HS}^2 = \sum_{i=1}^k \sum_{j=1}^d A_{ij}^2.$$

Proposizione 40 *Se le componenti X^{ij} sono di classe M^2, allora*

$$E\left[\left|\int_0^T X_s dB_s\right|^2\right] = E\left[\int_0^T \|X_s\|_{HS}^2 ds\right].$$

Proof. Infatti, essendo $E\left[|Y_T|^2\right] = \sum_{i=1}^k E\left[(Y_T^i)^2\right]$,

$$\begin{aligned} E\left[\left|\int_0^T X_s dB_s\right|^2\right] &= \sum_{i=1}^k E\left[\left(\sum_{j=1}^d \int_0^T X_s^{ij} dB_s^j\right)^2\right] \\ &= \sum_{i=1}^k E\left[\sum_{j=1}^d \int_0^T \left(X_s^{ij}\right)^2 ds\right] = E\left[\int_0^T \|X_s\|_{HS}^2 ds\right] \end{aligned}$$

dove abbiamo usato la Proposizione 39. ■

Anche le disuguaglianze per le martingale si trasportano al caso vettoriale. A titolo di esempio, vediamo la disuguaglianza di Doob per $p = 2$.

Proposizione 41 *Se le componenti X^{ij} sono di classe M^2, allora*

$$E\left[\sup_{t\in[0,T]}\left|\int_0^t X_s dB_s\right|^2\right] \leq 4E\left[\int_0^T \|X_s\|_{HS}^2 ds\right].$$

Proof.

$$\begin{aligned} E\left[\sup_{t\in[0,T]}\left|\int_0^t X_s dB_s\right|^2\right] &= E\left[\sup_{t\in[0,T]}\sum_{i=1}^k\left(\sum_{j=1}^d\int_0^t X_s^{ij}dB_s^j\right)^2\right] \\ &\leq \sum_{i=1}^k E\left[\sup_{t\in[0,T]}\left(\sum_{j=1}^d\int_0^t X_s^{ij}dB_s^j\right)^2\right]. \end{aligned}$$

Ciascun processo $\sum_{j=1}^d \int_0^t X_s^{ij} dB_s^j$ è una martingala di quadrato integrabile e quindi, per la disuguaglianza di Doob,

$$\begin{aligned} &\leq 4\sum_{i=1}^k E\left[\left(\sum_{j=1}^d\int_0^T X_s^{ij}dB_s^j\right)^2\right] \\ &= 4\sum_{i=1}^k E\left[\sum_{j=1}^d\int_0^T \left(X_s^{ij}\right)^2 ds\right] = 4E\left[\int_0^T \|X_s\|_{HS}^2 ds\right] \end{aligned}$$

avendo usato nuovamente la Proposizione 39. ■

Valgono tanti altri fatti simili al caso uni-dimensionale, che non descriviamo in dettaglio.

Capitolo 5

La formula di Itô

Seguiamo un approccio a questo tema non completamente tradizionale ma estremamente istruttivo, dovuto a Hans Föllmer. In sintesi, esso sviluppa prima una formula di Itô per funzioni deterministiche, che poi viene applicata traiettoria per traiettoria ad alcune classi di processi stocastici.

5.1 La "chain rule" per funzioni poco regolari

Vogliamo dimostrare una formula della catena (chain rule) per una composizione del tipo $f(x_t)$ dove $f : \mathbb{R} \to \mathbb{R}$ è regolare, anzi di classe C^2, ma $(x_t)_{t\in[0,T]}$ è una funzione non necessariamente differenziabile (e nemmeno a variazione limitata), con una proprietà speciale di "variazione quadratica". Per non approfondire alcuni dettagli tecnici, supponiamo T fissato e finito, ma si potrebbe riscrivere tutto per funzioni $(x_t)_{t\geq 0}$. Senza restrizione, prendiamo $T = 1$.

Indichiamo con π una generica partizione di $[0,1]$ della forma $0 = t_0 \leq t_1 \leq ... \leq t_n = 1$, con $|\pi|$ il numero $\max_{i=0,...,n-1} |t_{i+1} - t_i|$, con $[x,x]_t^\pi$ il numero

$$[x,x]_t^\pi = \sum_{\pi \ni t_i \leq t} \left(x_{t_{i+1}} - x_{t_i}\right)^2, \qquad t \in [0,1]$$

con la convenzione che $t_{n+1} = t_n$, quindi $x_{t_{n+1}} - x_{t_n} = 0$. Questa convenzione, che serve semplicemente a cancellare un ultimo termine spurio nella somma precedente, verrà sempre adottata tacitamente in questi paragrafi. Oppure, quando questo semplifichi la scrittura, useremo la notazione

$$J_t^\pi = \{i = 0, ..., n-1 : t_i^n \leq t\}$$

con la quale scriveremo

$$[x,x]_t^\pi = \sum_{i \in J_t^\pi} \left(x_{t_{i+1}} - x_{t_i}\right)^2, \qquad t \in [0,1]$$

Introduciamo inoltre la misura su $([0,1], \mathcal{B}([0,1]))$

$$\mu^{\pi} = \sum_{i=0}^{n-1} \left(x_{t_{i+1}} - x_{t_i}\right)^2 \delta_{x_{t_i}}$$

di cui $t \mapsto [x,x]_t^{\pi}$ è la funzione di ripartizione.

Definizione 15 *Sia $(\pi_n)_{n\in\mathbb{N}}$ una successione di partizioni con $|\pi_n| \to 0$ per $n \to \infty$ ed $x : [0,1] \to \mathbb{R}$ una funzione continua. Diciamo che x ha variazione quadratica lungo la successione π_n se esiste una funzione continua, indicata nel seguito con $[x,x]_t$, $t \in [0,1]$, tale che*

$$\lim_{n\to\infty} [x,x]_t^{\pi_n} = [x,x]_t \tag{5.1}$$

per ogni $t \in [0,1]$.

Osservazione 20 *Siccome $[x,x]_t^{\pi_n}$ è non decrescente, anche $[x,x]_t$ è non decrescente.*

Osservazione 21 *Il simbolo $[x,x]_t$ non è completamente appropriato perché dipende anche dalla successione $(\pi_n)_{n\in\mathbb{N}}$, oltre che da x. In tutti i nostri esempi, però, dipenderà solo da x, per cui non ci sembra utile appesantire la notazione tramite una dipendenza esplicita dalla successione.*

Osservazione 22 *Se x è di classe C^{α} con $\alpha > \frac{1}{2}$ allora la variazione quadrata (esiste ed) è pari a zero. Per avere una variazione quadratica non nulla, x dev'essere meno regolare.*

La funzione continua non decrescente $[x,x]_t$, $t \in [0,1]$, definisce una misura finita $d[x,x]_t$ su $([0,1], \mathcal{B}([0,1]))$. Vale

$$\int_a^b d[x,x]_s = [x,x]_b - [x,x]_a$$

per ogni $0 \le a \le b \le 1$. La funzione $[x,x]_t$ è la funzione di ripartizione di questa misura. Gli integrali del tipo $\int_0^t \varphi(s)\, d[x,x]_s$ che scriveremo nel seguito, con $\varphi : [0,t] \to \mathbb{R}$ continua, sono da intendersi rispetto a tale misura.

Proposizione 42 *i) Sia $D \subset [0,1]$ un insieme denso tale che $1 \in D$. Se esiste una funzione uniformemente continua $t \longmapsto [x,x]_t$, definita su D, tale che vale (5.1) per ogni $t \in D$, allora x ha variazione quadratica lungo la successione π_n.*

ii) Se x ha variazione quadratica lungo la successione π_n, allora per ogni $t \in [0,1]$

$$\lim_{n\to\infty} \sum_{i\in J_t^{\pi_n}} \varphi(t_i^n) \left(x_{t_{i+1}^n} - x_{t_i^n}\right)^2 = \int_0^t \varphi(s)\, d[x,x]_s \tag{5.2}$$

per ogni funzione $\varphi : [0,t] \to \mathbb{R}$ continua. Abbiamo indicato con t_i^n, $i = 0,1,...,N_n$, i punti della partizione π_n.

Proof. Ricordiamo preliminarmente alcuni fatti sulla convergenza di misure sulla retta reale. Sia (μ_n) una successione di misure finite, μ una misura finita, $F_n(x) = \mu_n((-\infty, x])$ e $F(x) = \mu((-\infty, x])$ le loro funzioni di ripartizione. Ricordiamo che: a) la convergenza vaga (cioè contro le funzioni continue a supporto compatto) implica la stretta (cioè contro le funzioni continue limitate) se $\mu_n(\mathbb{R}) \to \mu(\mathbb{R})$; b) se vale la convergenza stretta allora $F_n(x) \to F(x)$ in tutti i punti in cui F è continua; c) se $F_n(x) \to F(x)$ in tutti i punti di un insieme denso allora vale la convergenza vaga.

Siano ora (μ_n) e μ a supporto in $[0,1]$. Particolarizzando le affermazioni precedenti vediamo che: a) la convergenza vaga e la stretta coincidono; b) se vale la convergenza stretta allora $F_n(x) \to F(x)$ in tutti i punti in cui F è continua; c) se esiste un insieme denso $D \subset [0,1]$ con $1 \in D$, per cui vale $F_n(x) \to F(x)$ per ogni $x \in D$, allora vale la convergenza vaga. In particolare se ne deduce il seguente criterio: d) se esiste un insieme denso $D \subset [0,1]$ con $1 \in D$, per cui vale $F_n(x) \to F(x)$ per ogni $x \in D$, allora $F_n(x) \to F(x)$ in tutti i punti in cui F è continua.

La dimostrazione dell'affermazione (i) ora è ovvia. La funzione $t \mapsto [x,x]_t$ ha estensione unica ad una funzione continua non decrescente (perché limite su un denso di funzioni non decrescenti) a tutto $[0,1]$, che continuiamo a indicare con $t \mapsto [x,x]_t$. Introduciamo la misura μ su $([0,1], \mathcal{B}([0,1]))$ avente $t \mapsto [x,x]_t$ come funzione di ripartizione e ricordiamo che $t \mapsto [x,x]_t^{\pi_n}$ è la funzione di ripartizione di μ^{π_n}. Poniamo per chiarezza espositiva $F_n(t) = [x,x]_t^{\pi_n}$, $F(t) = [x,x]_t$, funzioni di ripartizione di μ^{π_n} e μ rispettivamente. Si tratta di misure a supporto in $[0,1]$. Siccome $F_n(t) \to F(t)$ per t in un insieme denso di $[0,1]$ con $1 \in D$, $F_n(t) \to F(t)$ per tutti i t in cui $F(t)$ è continua, quindi per ogni $t \in [0,1]$.

Dimostriamo (ii). Per $t = 1$, (5.2) è stata appena dimostrata, perché equivale alla convergenza stretta $\mu^{\pi_n} \to \mu$ cioè a

$$\lim_{n\to\infty} \int_0^1 \varphi(t)\, \mu^{\pi_n}(dt) = \int_0^1 \varphi(t)\, \mu(dt) \tag{5.3}$$

per ogni funzione continua $\varphi : [0,1] \to \mathbb{R}$. Per $t \in [0,1]$, dobbiamo mostrare che la successione di misure $\mu^{\pi_n}|_{[0,t]}$ converge a $\mu|_{[0,t]}$, dove le restrizioni sono definite nel seguente modo: preso $A \in \mathcal{B}([0,t])$, lo si vede come elemento di $\mathcal{B}([0,1])$, e si applica la misura. Le funzioni di ripartizione di queste restrizioni sono le restrizioni delle funzioni di ripartizione originarie. Per esse vale la convergenza in ogni punto (per ipotesi del punto (ii) di $[0,t]$, quindi vale la convergenza stretta di $\mu^{\pi_n}|_{[0,t]}$ a $\mu|_{[0,t]}$. La dimostrazione è completa. ∎

Teorema 28 *Sia* $x : [0,1] \to \mathbb{R}$ *una funzione continua avente variazione quadratica* $[x,x]_t$ *lungo una successione* π_n*. Sia* $f : \mathbb{R} \to \mathbb{R}$ *di classe* C^2*. Allora esiste finito il limite* $\lim_{n\to\infty} \sum_{i \in J_t^{\pi_n}} f'\left(x_{t_i^n}\right)\left(x_{t_{i+1}^n} - x_{t_i^n}\right)$*, che indichiamo con* $\int_0^t f'(x_s)\, dx_s$*, e vale*

$$f(x_t) = f(x_0) + \int_0^t f'(x_s)\, dx_s + \frac{1}{2}\int_0^t f''(x_s)\, d[x,x]_s$$

Abbiamo indicato con t_i^n, $i = 0, 1, ..., N_n$, i punti della partizione π_n.

Proof. Abbreviamo $J_t^n = J_t^{\pi_n}$. Vale

$$f(x_t) - f(x_0) = \lim_{n\to\infty} \sum_{i\in J_t^n} \left(f\left(x_{t_{i+1}^n}\right) - f\left(x_{t_i^n}\right)\right)$$

essendo x continua. Cerchiamo di esprimere $f\left(x_{t_{i+1}^n}\right) - f\left(x_{t_i^n}\right)$ opportunamente.

Ricordiamo la formula di Taylor, in un generico punto x_0:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''\left(\xi_{x,x_0}\right)(x - x_0)^2$$

dove ξ_{x,x_0} è un punto intermedio tra x ed x_0. Possiamo riscriverla nella forma

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2 + r(x, x_0)$$

dove $r(x, x_0) = \frac{1}{2}\left(f''\left(\xi_{x,x_0}\right) - f''(x_0)\right)(x - x_0)^2$. Per x, x_0 appartenenti all'immagine di $t \mapsto x_t$ (o a qualsiasi compatto, da cui dipenderà la funzione a), $r(x, x_0)$ soddisfa

$$|r(x, x_0)| \le a(|x - x_0|)(x - x_0)^2$$

dove $r \longmapsto a(r)$, $r \ge 0$, è una funzione crescente e infinitesima nell'origine, con $a(0) = 0$. Basta porre, relativamente ad un compatto K fissato,

$$a(r) = \sup_{\substack{x,x_0\in K\\ |x-x_0|\le r}} \frac{1}{2}\left|f''\left(\xi_{x,x_0}\right) - f''(x_0)\right|$$

che è finito per ogni r grazie alla continuità di f'', e verifica $\lim_{r\to 0} a(r) = 0$ per l'uniforme continuità di f''.

Usiamo questa versione della formula di Taylor:

$$\begin{aligned}\sum_{i\in J_t^n}\left(f\left(x_{t_{i+1}^n}\right) - f\left(x_{t_i^n}\right)\right) &= \sum_{i\in J_t^n} f'\left(x_{t_i^n}\right)\left(x_{t_{i+1}^n} - x_{t_i^n}\right) \\ &+ \frac{1}{2}\sum_{i\in J_t^n} f''\left(x_{t_i^n}\right)\left(x_{t_{i+1}^n} - x_{t_i^n}\right)^2 + \sum_{i\in J_t^n} r\left(x_{t_{i+1}^n}, x_{t_i^n}\right).\end{aligned}$$

Il secondo termine a destra dell'uguale converge a $\frac{1}{2}\int_0^t f''(x_s)\, d[x,x]_s$. Il terzo termine tende a zero, essendo

$$\begin{aligned}\sum_{i\in J_t^n} r\left(x_{t_{i+1}^n}, x_{t_i^n}\right) &\le \sum_{i\in J_t^n} a\left(\left|x_{t_{i+1}^n} - x_{t_i^n}\right|\right)\left(x_{t_{i+1}^n} - x_{t_i^n}\right)^2 \le \max_{i\in J_t^n} a\left(\left|x_{t_{i+1}^n} - x_{t_i^n}\right|\right)\sum_{i\in J_t^n}\left(x_{t_{i+1}^n} - x_{t_i^n}\right)^2 \\ &\le a\left(\max_{i\in J_t^n}\left|x_{t_{i+1}^n} - x_{t_i^n}\right|\right)[x,x]_t^{\pi_n}\end{aligned}$$

ed usando il fatto che $\max_{i\in J_t^n}\left|x_{t_{i+1}^n}-x_{t_i^n}\right|\to 0$ per $n\to\infty$ (essendo x uniformemente continua e $|\pi_n|\to 0$), a è infinitesima e $[x,x]_t^{\pi_n}$ è limitato perché converge. Il termine a sinistra dell'uguale converge a $f(x_t)-f(x_0)$. Quindi anche il primo termine a destra dell'uguale converge e vale la formula enunciata dal teorema. La dimostrazione è completa. ■

Osservazione 23 *Se x è differenziabile con continuità, vale*

$$f(x_t)=f(x_0)+\int_0^t f'(x_s)\,dx_s$$

dove $\int_0^t f'(x_s)\,dx_s$ va inteso come $\int_0^t f'(x_s)\,x_s' ds$. Il correttore

$$\frac{1}{2}\int_0^t f''(x_s)\,d[x,x]_s$$

è la novità di questa formula, nel caso di x meno regolare ma avente variazione quadratica.

5.1.1 Caso multidimensionale e dipendente dal tempo

Definizione 16 *Date due funzione continue $x^1,x^2:[0,1]\to\mathbb{R}$ ed una successione di partizioni $(\pi_n)_{n\in\mathbb{N}}$ con $|\pi_n|\to 0$ per $n\to\infty$, diciamo che x^1 e x^2 hanno variazione mutua $[x^1,x^2]_t$, $t\in[0,1]$, lungo π_n, se per ogni $t\in[0,1]$ esiste finito il limite*

$$\lim_{n\to\infty}\sum_{k\in J_t^{\pi_n}}\left(x^1_{t_{k+1}^n}-x^1_{t_k^n}\right)\left(x^2_{t_{k+1}^n}-x^2_{t_k^n}\right)$$

e definisce una funzione continua, che indichiamo appunto con $[x^i,x^j]_t$, $t\in[0,1]$. Se $x^1=x^2$, il concetto si riduce a quello di variazione quadratica lungo π_n.

Consideriamo ora una funzione $x:[0,1]\to\mathbb{R}^d$ di componenti x^i, $i=1,...,d$.

Definizione 17 *Sia $(\pi_n)_{n\in\mathbb{N}}$ una successione di partizioni con $|\pi_n|\to 0$ per $n\to\infty$. Diciamo che una funzione continua $x:[0,1]\to\mathbb{R}^d$ ha variazione quadratica congiunta lungo la successione π_n se per ogni $i,j=1,...,d$ (anche $i=j$) le componenti x^i e x^j hanno variazione mutua $[x^i,x^j]_t$, $t\in[0,1]$, lungo π_n, ovvero*

$$\left[x^i,x^j\right]_t=\lim_{n\to\infty}\sum_{k\in J_t^{\pi_n}}\left(x^i_{t_{k+1}^n}-x^i_{t_k^n}\right)\left(x^j_{t_{k+1}^n}-x^j_{t_k^n}\right). \tag{5.4}$$

Lemma 8 *La funzione continua $x : [0,1] \to \mathbb{R}^d$ ha variazione quadratica congiunta lungo π_n se e solo se per ogni $i, j = 1, ..., d$ (anche $i = j$) le funzioni $x^i + x^j$ hanno variazione quadratica lungo π_n. Le variazioni suddette sono legate dalle formule*

$$\left[x^i, x^j\right]_t = \frac{1}{2}\left(\left[x^i + x^j, x^i + x^j\right]_t - \left[x^i, x^i\right]_t - \left[x^j, x^j\right]_t\right)$$

$$\left[x^i + x^j, x^i + x^j\right]_t = \left[x^i, x^i\right]_t + \left[x^j, x^j\right]_t + 2\left[x^i, x^j\right]_t.$$

Proof. Supponiamo che valga la condizione della definizione. Dall'indentità

$$\begin{aligned} & \left(\left(x^i_{t^n_{k+1}} + x^j_{t^n_{k+1}}\right) - \left(x^i_{t^n_k} + x^j_{t^n_k}\right)\right)^2 \\ = & \left(x^i_{t^n_{k+1}} - x^i_{t^n_k}\right)^2 + \left(x^j_{t^n_{k+1}} - x^j_{t^n_k}\right)^2 + 2\left(x^i_{t^n_{k+1}} - x^i_{t^n_k}\right)\left(x^j_{t^n_{k+1}} - x^j_{t^n_k}\right) \end{aligned}$$

deduciamo che $x^i + x^j$ ha variazione quadratica lungo π_n. Viceversa, se vale la condizione del lemma, dall'indentità

$$\begin{aligned} & \left(x^i_{t^n_{k+1}} - x^i_{t^n_k}\right)\left(x^j_{t^n_{k+1}} - x^j_{t^n_k}\right) \\ = & \frac{1}{2}\left(\left(\left(x^i_{t^n_{k+1}} + x^j_{t^n_{k+1}}\right) - \left(x^i_{t^n_k} + x^j_{t^n_k}\right)\right)^2 - \left(x^i_{t^n_{k+1}} - x^i_{t^n_k}\right)^2 - \left(x^j_{t^n_{k+1}} - x^j_{t^n_k}\right)^2\right) \end{aligned}$$

deduciamo che x^i e x^j hanno variazione mutua lungo π_n. ■

L'utilità del lemma è quella di riportare il problema della variazione mutua a quello della variazione quadratica, a cui possiamo applicare la Proposizione 42. Ne discende subito la variante in dimensione d:

Proposizione 43 *i) Sia $D \subset [0,1]$ un insieme denso tale che $1 \in D$. Se, per ogni $i, j = 1, ..., d$, esiste una funzione uniformemente continua $t \longmapsto \left[x^i, x^j\right]_t$, definita su D, tale che vale (5.4) per ogni $t \in D$, allora x ha variazione quadratica congiunta lungo π_n.*

ii) Se $x : [0,1] \to \mathbb{R}^d$ ha variazione quadratica lungo π_n allora, per ogni $i, j = 1, ..., d$ ed ogni $t \in [0,1]$ esiste il seguente limite

$$\lim_{n\to\infty} \sum_{k \in J_t^{\pi n}} \varphi\left(t_i^n\right)\left(x^i_{t^n_{k+1}} - x^i_{t^n_k}\right)\left(x^j_{t^n_{k+1}} - x^j_{t^n_k}\right) = \int_0^t \varphi(s)\, d\left[x^i, x^j\right]_s$$

per ogni funzione $\varphi : [0,t] \to \mathbb{R}$ continua e limitata. L'integrale $\int_0^t \varphi(s)\, d\left[x^i, x^j\right]_s$ può essere inteso per differenza, secondo la definizione di $\left[x^i, x^j\right]_t$.

Teorema 29 *Sia $x : [0,1] \to \mathbb{R}^d$ una funzione continua avente variazione quadratica congiunta lungo una successione π_n. Sia $f : [0,1] \times \mathbb{R}^d \to \mathbb{R}$ una funzione continua* *avente le derivate* $\frac{\partial f}{\partial t}$, $\frac{\partial f}{\partial x^i}$, $\frac{\partial^2 f}{\partial x^i \partial x^j}$ *continue (diremo di classe* $C^{1,2}$*). Allora esiste finito il limite*

$$\lim_{n\to\infty} \sum_{k\in J_t^{\pi n}} \sum_{i=1}^{d} \frac{\partial f}{\partial x^i}\left(t_k^n, x_{t_k^n}\right)\left(x^i_{t^n_{k+1}} - x^i_{t^n_k}\right)$$

che indichiamo con

$$\int_0^t \nabla f\left(s, x_s\right) \cdot dx_s$$

e vale

$$\begin{aligned} f\left(t, x_t\right) &= f\left(0, x_0\right) + \int_0^t \frac{\partial f}{\partial t}\left(s, x_s\right) ds + \int_0^t \nabla f\left(s, x_s\right) \cdot dx_s \\ &\quad + \frac{1}{2}\sum_{i,j=1}^{d} \int_0^t \frac{\partial^2 f}{\partial x^i \partial x^j}\left(s, x_s\right) d\left[x^i, x^j\right]_s . \end{aligned}$$

Osservazione 24 *Gli integrali* $\int_0^t \frac{\partial f}{\partial x^i}\left(s, x_s\right) dx_s^i$ *non sono autonomamente definiti, solo la loro somma. Questo è il difetto principale di questa teoria "deterministica" del calcolo di Itô. Il problema sparisce nel caso di processi stocastici, in quanto questi integrali saranno definiti nel senso di Itô.*

Proof. La dimostrazione è identica al caso $d = 1$; la diamo per completezza. Scriviamo J_t^n per $J_t^{\pi_n}$. Diamo la dimostrazione solo nel caso di f indipendente dal tempo. Vale

$$f\left(x_t\right) - f\left(x_0\right) = \lim_{n\to\infty} \sum_{k\in J_t^n} \left(f\left(x_{t^n_{k+1}}\right) - f\left(x_{t^n_k}\right)\right)$$

essendo x continua. Cerchiamo di esprimere $f\left(x_{t^n_{k+1}}\right) - f\left(x_{t^n_k}\right)$ opportunamente.

Ricordiamo la formula di Taylor, in un generico punto x_0:

$$f(x) = f\left(x_0\right) + \sum_{i=1}^{d} \frac{\partial f}{\partial x^i}\left(x_0\right)\left(x^i - x_0^i\right) + \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2 f}{\partial x^i \partial x^j}\left(\xi_{x,x_0}\right)\left(x^i - x_0^i\right)\left(x^j - x_0^j\right)$$

dove ξ_{x,x_0} è un punto intermedio tra x ed x_0. Possiamo riscriverla nella forma

$$\begin{aligned} f(x) &= f\left(x_0\right) + \sum_{i=1}^{d} \frac{\partial f}{\partial x^i}\left(x_0\right)\left(x^i - x_0^i\right) \\ &\quad + \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2 f}{\partial x^i \partial x^j}\left(x_0\right)\left(x^i - x_0^i\right)\left(x^j - x_0^j\right) + r\left(x, x_0\right) \end{aligned}$$

dove

$$r(x,x_0)=\frac{1}{2}\sum_{i,j=1}^{d}\left(\frac{\partial^2 f}{\partial x^i \partial x^j}\left(\xi_{x,x_0}\right)-\frac{\partial^2 f}{\partial x^i \partial x^j}\left(x_0\right)\right)\left(x^i-x_0^i\right)\left(x^j-x_0^j\right).$$

Per x, x_0 appartenenti alla chiusura convessa dell'immagine di $t \mapsto x_t$, $r(x,x_0)$ soddisfa

$$\left|r(x,x_0)\right| \leq a\left(\left|x-x_0\right|\right)\left|x-x_0\right|^2$$

dove a è una funzione crescente e infinitesima nell'origine.

Usiamo questa versione della formula di Taylor:

$$\begin{aligned}
\sum_{k\in J_t^n}\left(f\left(x_{t_{k+1}^n}\right)-f\left(x_{t_k^n}\right)\right) &= \sum_{k\in J_t^n}\sum_{i=1}^{d}\frac{\partial f}{\partial x^i}\left(x_{t_k^n}\right)\left(x^i_{t_{k+1}^n}-x^i_{t_k^n}\right)\\
&+\frac{1}{2}\sum_{k\in J_t^n}\sum_{i,j=1}^{d}\frac{\partial^2 f}{\partial x^i \partial x^j}\left(x_{t_k^n}\right)\left(x^i_{t_{k+1}^n}-x^i_{t_k^n}\right)\left(x^j_{t_{k+1}^n}-x^j_{t_k^n}\right)\\
&+\sum_{k\in J_t^n} r\left(x_{t_{k+1}^n},x_{t_k^n}\right).
\end{aligned}$$

Il secondo termine a destra dell'uguale converge a $\frac{1}{2}\sum_{i,j=1}^{d}\int_0^t \frac{\partial^2 f}{\partial x^i \partial x^j}\left(x_s\right)d\left[x^i,x^j\right]_s$, grazie al lemma. Il terzo termine tende a zero, essendo

$$\begin{aligned}
\sum_{k\in J_t^n} r\left(x_{t_{k+1}^n},x_{t_k^n}\right) &\leq \sum_{k\in J_t^n} a\left(\left|x_{t_{k+1}^n}-x_{t_k^n}\right|\right)\left|x_{t_{k+1}^n}-x_{t_k^n}\right|^2\\
&\leq \max_k a\left(\left|x_{t_{k+1}^n}-x_{t_k^n}\right|\right)\sum_{k\in J_t^n}\sum_{i=1}^{d}\left(x^i_{t_{k+1}^n}-x^i_{t_k^n}\right)^2
\end{aligned}$$

e ragionando come nel caso unidimensionale. Il termine a sinistra dell'uguale converge a $f(x_t)-f(x_0)$. Quindi anche il primo termine a destra dell'uguale converge e vale la formula enunciata dal teorema. La dimostrazione è completa. ■

5.1.2 Integrale secondo Stratonovich

Nella teoria precedente il termine del secondo ordine della formula di Taylor gioca un ruolo essenziale. Esiste però un modo alternativo di definire gli integrali che permette di ottenere formule più somiglianti al caso classico dell'analisi matematica. Sono detti *integrali secondo Stratonovich*. Hanno un innegabile vantaggio in contesti dove le composizioni sono frequentissime, come nel caso di processi stocastici su varietà in cui tutto è sempre definito tramite composizione con carte. Hanno però anche dei difetti, che descriveremo al momento buono.

Introduciamo il seguente simbolo, ove definito:

$$\int_0^t y_s \circ dx_s = \int_0^t y_s dx_s + \frac{1}{2} [y, x]_t$$

dove $[y, x]_t$ è la variazione mutua sopra introdotta e $\int_0^t y_s dx_s$ indica, se esiste, il limite

$$\int_0^t y_s dx_s = \lim_{n\to\infty} \sum_{i \in J_t^{\pi n}} y_{t_i^n} \left(x_{t_{i+1}^n} - x_{t_i^n} \right).$$

Vale (quando esiste il limite)

$$\int_0^t y_s \circ dx_s = \lim_{n\to\infty} \sum_{i \in J_t^{\pi n}} \frac{y_{t_{i+1}^n} + y_{t_i^n}}{2} \left(x_{t_{i+1}^n} - x_{t_i^n} \right)$$

in quanto

$$[y, x]_t = \lim_{n\to\infty} \sum_{i \in J_t^{\pi n}} \left(y_{t_{i+1}^n} - y_{t_i^n} \right) \left(x_{t_{i+1}^n} - x_{t_i^n} \right).$$

Teorema 30 *Sia $x : [0, 1] \to \mathbb{R}$ una funzione continua avente variazione quadratica lungo una successione π_n. Sia $f : \mathbb{R} \to \mathbb{R}$ una funzione di classe C^2. Allora esiste la variazione mutua $[f'(x), x]_t$ e vale*

$$f(x_t) = f(x_0) + \int_0^t f'(x_s) \circ dx_s.$$

Proof. Sappiamo che vale

$$f(x_t) = f(x_0) + \int_0^t f'(x_s) dx_s + \frac{1}{2} \int_0^t f''(x_s) d[x, x]_s.$$

Dobbiamo dimostrare che $[f'(x), x]_t$ esiste e vale

$$[f'(x), x]_t = \int_0^t f''(x_s) d[x, x]_s.$$

Abbiamo, se esiste,

$$[f'(x), x]_t = \lim_{n\to\infty} \sum_{i \in J_t^{\pi n}} \left(f'\left(x_{t_{i+1}^n} \right) - f'\left(x_{t_i^n} \right) \right) \left(x_{t_{i+1}^n} - x_{t_i^n} \right)$$

mentre sappiamo che

$$\int_0^t f''(x_s) d[x, x]_s = \lim_{n\to\infty} \sum_{i \in J_t^{\pi n}} f''\left(x_{t_i^n} \right) \left(x_{t_{i+1}^n} - x_{t_i^n} \right)^2.$$

Inoltre sappiamo che

$$f'\left(x_{t_{i+1}^n}\right) - f'\left(x_{t_i^n}\right) = f''\left(x_{t_i^n}\right)\left(x_{t_{i+1}^n} - x_{t_i^n}\right) + r\left(x_{t_i^n}, x_{t_{i+1}^n}\right)$$

con

$$|r(x, x_0)| \leq a(|x - x_0|)(x - x_0)^2$$

dove a è una funzione crescente e infinitesima nell'origine. Quindi basta dimostrare che

$$\lim_{n\to\infty} \sum_{i \in J_t^{\pi_n}} r\left(x_{t_i^n}, x_{t_{i+1}^n}\right) = 0.$$

Ma questo segue facilmente dalla proprietà suddetta di r. ■

5.2 Applicazione ai processi stocastici

In questa sezione indichiamo la linea generale con cui si può applicare la chain rule nell'ambito dei processi stocastici. Da un lato, dobbiamo superare una dfficoltà: che per gli usuali processi stocastici non si riesce ad esaminare direttamente la variazione quadratica delle loro singole traiettorie, ma solo secondo una definizione con la convergenza in probabilità. Dall'altro però, nel caso dei processi stocastici, gli integrali stocastici sono autonomamente definiti, non solo come sottoprodotto della chain rule stessa; dobbiamo però verificare che la definizione autonoma coincide col la formula data dalla chain rule. Iniziamo discutendo il problema della variazione quadratica. Formuliamo il risultato su $[0,1]$ ma vale su ogni $[0,T]$ oppure per $t \geq 0$ (intendendo che vale su $[0,T]$ per ogni $T > 0$, con il risultato che non dipende da T). Sia (π_n) una successione di partizioni di $[0,1]$ con $|\pi_n| \to 0$.

Definizione 18 *Diciamo che un processo stocastico continuo $X = (X_t)_{t\in[0,1]}$ in $\mathbb{R}^d$ ha variazione quadratica congiunta $[X^i, X^j]_t$, $i,j = 1,...,d$, lungo π_n se per ogni $i,j = 1,...,d$, esiste un processo stocastico continuo, che indichiamo con $[X^i, X^j]_t$, tale che per ogni $t \in [0,1]$ sia*

$$\left[X^i, X^j\right]_t = P - \lim_{n\to\infty} \sum_{i \in J_t^{\pi_n}} \left(X^i_{t_{i+1}^n} - X^i_{t_i^n}\right)\left(X^j_{t_{i+1}^n} - X^j_{t_i^n}\right).$$

Chiamiamo $[X^i, X^j]_t$ variazione mutua tra X^i e X^j, o variazione quadratica quando $i = j$.

E' equivalente a chiedere che, per ogni $i,j = 1,...,d$ ed ogni $t \in [0,1]$, il processo $X^i + X^j$ abbia variazione quadratica.

Proposizione 44 *Sia $X = (X_t)_{t\in[0,1]}$ un processo stocastico continuo in $\mathbb{R}^d$ che ha variazione quadratica congiunta $[X^i, X^j]_t$, $i, j = 1, ..., d$, lungo π_n Allora esiste una sottosuccessione (π'_n) ed un evento $\Omega_0 \in \mathcal{F}$ con $P(\Omega_0) = 1$ tali che per ogni $\omega \in \Omega_0$ la funzione vettoriale $t \mapsto X_t(\omega)$ ha variazione quadratica congiunta su $[0,1]$ lungo π_n, e vale $[X^i(\omega), X^j(\omega)]_t = [X^i, X^i]_t(\omega)$ per ogni $i, j = 1, ..., d$. Quando tutto questo accade, diremo semplicemente che il processo ha variazione quadratica congiunta $[X^i, X^j]_t$.*

Proof. Scriviamo la dimostrazione in dimensione uno; il caso generale è identico. Sia $t \in \mathbb{Q}\cap[0,1]$. Esiste una sottosuccessione (π_n^t) per cui valga la convergenza quasi certa:

$$P\left(\lim_{n\to\infty}\sum_{i\in J_t^{\pi_n^t}}\left(X_{t_{i+1}^n} - X_{t_i^n}\right)^2 = [X,X]_t\right) = 1.$$

Con un procedimento diagonale (numerando gli elementi di $\mathbb{Q}\cap[0,1]$), si può trovare un'unica sottosuccessione, che indichiamo con (π'_n), tale che

$$P\left(\lim_{n\to\infty}\sum_{i\in J_t^{\pi_n^t}}\left(X_{t_{i+1}^n} - X_{t_i^n}\right)^2 = [X,X]_t\right) = 1 \qquad \text{per ogni } t\in\mathbb{Q}\cap[0,T].$$

Siccome l'intersezione numerabile di insiemi di probabilità uno ha probabilità uno, l'insieme

$$A = \left\{\omega\in\Omega : \lim_{n\to\infty}\sum_{i\in J_t^{\pi_n^t}}\left(X_{t_{i+1}^n}(\omega) - X_{t_i^n}(\omega)\right)^2 = [X,X]_t(\omega), \forall t\in\mathbb{Q}\cap[0,1]\right\}$$

ha la proprietà $P(A) = 1$. Per la Proposizione 42 (parte (i)),

$$A = \left\{\omega\in\Omega : \lim_{n\to\infty}\sum_{i\in J_t^{\pi_n^t}}\left(X_{t_{i+1}^n}(\omega) - X_{t_i^n}(\omega)\right)^2 = [X,X]_t(\omega), \forall t\in[0,1]\right\}.$$

La dimostrazione è completa. ■

Corollario 12 *Sia $X = (X_t)_{t\in[0,1]}$ un processo stocastico continuo in $\mathbb{R}^d$ avente variazione quadratica congiunta lungo una successione di partizioni; indichiamo poi con (π_n) una successione per cui sia vera la tesi della proposizione precedente. Sia $f : [0,1]\times\mathbb{R}\to\mathbb{R}$ una funzione di classe $C^{1,2}$. Allora, per ogni $t\in[0,1]$, esiste il limite in probabilità e q.c.*

$$\lim_{n\to\infty}\sum_{k\in J_t^{\pi_n}}\sum_{i=1}^{d}\frac{\partial f}{\partial x^i}\left(t_k^n, X_{t_k^n}\right)\left(X^i_{t_{k+1}^n} - X^i_{t_k^n}\right)$$

che indichiamo con

$$\int_0^t \nabla f(s, X_s) \cdot dX_s$$

e vale q.c.

$$\begin{aligned} f(t, X_t) &= f(0, X_0) + \int_0^t \frac{\partial f}{\partial t}(s, X_s)\, ds + \int_0^t \nabla f(s, X_s) \cdot dX_s \\ &+ \frac{1}{2} \sum_{i,j=1}^d \int_0^t \frac{\partial^2 f}{\partial x^i \partial x^j}(s, X_s)\, d\left[X^i, X^j\right]_s . \end{aligned}$$

Per rendere ancor più operativa questa formula servono due ingredienti. Da un lato, avere un'ampia classe di processi X aventi variazione quadratica congiunta $[X^i, X^j]_t$ e che questa sia calcolabile. Dall'altro, sapere che la v.a. $\int_0^t \nabla f(s, X_s) \cdot dX_s$ si può riscrivere tramite integrali noti (di Itô o di Lebesgue). Affrontiamo i due problemi nelle prossime due sezioni.

5.2.1 Variazione quadratica

Discutiamo l'esistenza della variazione quadratica congiunta per diversi processi.

ll moto browniano unidimensionale

Sia B un moto browniano in $\mathbb{R}$. Il Teorema 6 afferma che, data una successione (π_n) con $|\pi_n| \to 0$, vale

$$\lim_{n\to\infty} E\left[\left|[B, B]_t^{\pi_n} - t\right|^2\right] = 0 \tag{5.5}$$

per ogni $t \geq 0$. In verità, a essere puntigliosi, in tale teorema, π_n sono partizioni dello specifico intervallo $[0, t]$ trattato, quindi cambiando t si devono prendere partizioni diverse; ma correggendo di poco quella dimostrazione di può prendere la successione di partizioni a priori, e farla valere per ogni t. Scriviamo tutto per $[0, 1]$ ma vale per qualsiasi intervallo.

Proposizione 45 *Data una successione di partizioni (π_n) di $[0, 1]$, con $|\pi_n| \to 0$, per ogni $t \in [0, 1]$ vale (5.5). Quindi il moto browniano ha variazione quadratica, e vale $[B, B]_t = t$.*

Proof. Fissato $t \in [0, 1]$, indichiamo con π_n^t la partizione π_n completata dal punto t. Chiamando t_i^n i nodi di π_n e s_i^n i nodi di π_n^t, vale

$$\begin{aligned} &[B, B]_t^{\pi_n^t} - [B, B]_t^{\pi_n} \\ &= \sum_{i \in J_t^{\pi_n^t}} \left(B_{s_{i+1}^n} - B_{s_i^n}\right)^2 - \sum_{i \in J_t^{\pi_n}} \left(B_{t_{i+1}^n} - B_{t_i^n}\right)^2 = \left(B_t - B_{t_{i_{\max}}^n}\right)^2 \end{aligned}$$

dove $i_{\max} = \max J_t^{\pi_n}$. Sappiamo già che

$$\lim_{n\to\infty} E\left[\left|[B,B]_t^{\pi_n^t} - t\right|^2\right] = 0.$$

Allora (5.5) discende dal fatto che

$$E\left[\left|[B,B]_t^{\pi_n^t} - [B,B]_t^{\pi_n}\right|^2\right] = E\left[\left(B_t - B_{t^n_{i_{\max}}}\right)^4\right] \leq C\left(t - t^n_{i_{\max}}\right)^2$$

che tende a zero per $n \to \infty$. ■

Il moto browniano in $\mathbb{R}^d$

Sia $B = \left(B^1, ..., B^d\right)$ un moto browniano in $\mathbb{R}^d$. Ogni componente B^i ha variazione quadratica, pari a t. Invece la variazione mutua tra le componenti è nulla:

Proposizione 46 *Data una successione di partizioni (π_n) di $[0,1]$, con $|\pi_n| \to 0$, per $i \neq j$,*

$$\lim_{n\to\infty} E\left[\left|\left[B^i, B^j\right]_t^{\pi_n}\right|^2\right] = 0.$$

Quindi esiste la variazione mutua $\left[B^i, B^j\right]_t$ e vale zero.

Proof. Prendiamo $i = 1$, $j = 2$. Abbiamo

$$\left[B^1, B^2\right]_t^{\pi_n} = \sum_{k\in J_t^n} \left(B^1_{t^n_{k+1}} - B^1_{t^n_k}\right)\left(B^2_{t^n_{k+1}} - B^2_{t^n_k}\right)$$

quindi, con la solita notazione $\Delta_k B^i = B^i_{t^n_{k+1}} - B^i_{t^n_k}$,

$$\begin{aligned} E\left[\left|\left[B^1, B^2\right]_t^{\pi_n}\right|^2\right] &= \sum_{k,h\in J_t^n} E\left[\Delta_k B^1 \Delta_k B^2 \Delta_h B^1 \Delta_h B^2\right] \\ &= \sum_{k,h\in J_t^n} E\left[\Delta_k B^1 \Delta_h B^1\right] E\left[\Delta_k B^2 \Delta_h B^2\right] \end{aligned}$$

per l'indipendenza delle componenti. Se $h \neq k$, $E\left[\Delta_k B^1 \Delta_h B^1\right] = 0$. Quindi

$$E\left[\left|\left[B^1, B^2\right]_t^{\pi_n}\right|^2\right] = \sum_{k\in J_t^n} E\left[\left(\Delta_k B^1\right)^2\right] E\left[\left(\Delta_k B^2\right)^2\right] = \sum_{k\in J_t^n} \left(t^n_{k+1} - t^n_k\right)^2$$

da cui è facile dedurre la tesi. ■

In conclusione, la variazione quadratica congiunta di un moto browniano $B = \left(B^1, ..., B^d\right)$ in $\mathbb{R}^d$ è $\left[B^i, B^j\right]_t = \delta_{ij} t$.

Gli integrali stocastici

Teorema 31 *Sia $B = (B^1, ..., B^d)$ un moto browniano in $\mathbb{R}^d$. Siano Z^1, Z^2 due processi di classe Λ^2. Sia (π_n) una successione di partizioni di $[0,1]$, con $|\pi_n| \to 0$. Allora, per ogni $i, j = 1, 2$ (anche $i = j$) e per ogni t, la successione di v.a. $\left[\int_0^\cdot Z_s^1 dB_s^i, \int_0^\cdot Z_s^2 dB_s^j\right]_t^{\pi_n}$ converge in probabilità ed il limite vale*

$$\left[\int_0^\cdot Z_s^1 dB_s^i, \int_0^\cdot Z_s^2 dB_s^j\right]_t = \delta_{ij} \int_0^t Z_s^1 Z_s^2 ds.$$

La dimostrazione completa è piuttosto lunga. Accontentiamoci di comprendere due casi particolari. Il primo è quando Z^1, Z^2 sono semplicemente due costanti, che senza restrizione prendiamo pari ad uno. In questo caso la formula si riduce a $[B^i, B^j]_t = t\delta_{ij}$, che abbiamo già dimostrato. Il secondo è il caso in cui Z^1, Z^2 sono processi continui. Non diamo la dimostrazione completa nemmeno in questo caso particolare, ma solo l'idea. Presa una successione di partizioni π_n, lungo essa le componenti del moto browniano hanno variazione quadratica e mutua, con $[B^i, B^j]_t = t\delta_{ij}$. Usando un analogo per processi stocastici della Proposizione 43, perte (ii), abbiamo

$$\lim_{n\to\infty} \sum_{k \in J_t^{\pi n}} Z_{t_i^n}^1 Z_{t_i^n}^2 \left(B_{t_{k+1}^n}^i - B_{t_k^n}^i\right)\left(B_{t_{k+1}^n}^j - B_{t_k^n}^j\right) = \int_0^t Z_s^1 Z_s^2 ds.$$

Ma

$$\sum_{k \in J_t^{\pi n}} Z_{t_i^n}^1 Z_{t_i^n}^2 \left(B_{t_{k+1}^n}^i - B_{t_k^n}^i\right)\left(B_{t_{k+1}^n}^j - B_{t_k^n}^j\right) = \sum_{i \in J_t^{\pi n}} \int_{t_i^n}^{t_{i+1}^n} Z_{t_i^n}^1 dB_s^i \int_{t_i^n}^{t_{i+1}^n} Z_{t_i^n}^2 dB_s^j$$

è circa uguale a

$$\left[\int_0^\cdot Z_s^1 dB_s^i, \int_0^\cdot Z_s^2 dB_s^j\right]_t^{\pi_n} = \sum_{i \in J_t^{\pi n}} \int_{t_i^n}^{t_{i+1}^n} Z_s^1 dB_s^i \int_{t_i^n}^{t_{i+1}^n} Z_s^2 dB_s^j.$$

Con un po' di fatica si può controllare il resto e completare la dimostrazione.

Il caso generale richiede ad esempio di sviluppare più teoria sulla variazione quadratica in sé, includendo un teorema di dipendenza continua dai processi dei quali si calcola. In tal modo, avendo il risultato per Z^i continui, lo si estende a Z^i di classe Λ^2. I dettagli sono troppo numerosi per cui li omettiamo.

I processi di Itô

Sia B un moto browniano in $\mathbb{R}^n$ rispetto ad una filtrazione $(\mathcal{F}_t)_{t\geq 0}$.

Definizione 19 *Chiamiamo processo di Itô in $\mathbb{R}^d$ un processo $X = (X_t)_{t\in[0,T]}$ a valori in $\mathbb{R}^d$ della forma*

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dB_s$$

con X_0 misurabile rispetto a $\mathcal{F}_0$, b un processo vettoriale in $\mathbb{R}^d$ di classe Λ_B^1 (cioè progressivamente misurabile e tale che $\int_0^T |b_s|\, ds < \infty$ q.c.), σ processo matriciale in $\mathbb{R}^{n\times d}$ di classe Λ_B^2. Per componenti,

$$X_t^i = X_0^i + \int_0^t b_s^i ds + \sum_{k=1}^n \int_0^t \sigma_s^{ik} dB_s^k.$$

Scriveremo anche

$$dX_t = b_t dt + \sigma_t dB_t.$$

Le componenti dei processi di Itô sono somma di processi continui a variazione limitata (il termine $X_0^i + \int_0^t b_s^i ds$) e di martingale locali. Per questo sono *semimartingale*. Per calcolare la loro variazione quadratica abbiamo bisogno dei seguenti lemmi. Sia (π_n) una successione di partizioni di $[0,1]$, con $|\pi_n| \to 0$.

Lemma 9 *Nelle ipotesi precedenti, preso $k = 1,...,d$, $V_t := \int_0^t b_s^k ds$ ha variazione quadratica nulla lungo (π_n) ($[V,V]_t^{\pi_n}$ converge in probabilità a zero, per ogni $t \in [0,1]$).*

Proof. Basta osservare che le traiettorie di V sono continue e BV ed applicare la Proposizione 22; va solo curato il fatto che, dato $t \in [0,T]$, le partizioni π_n non sono partizioni di $[0,t]$; ma la dimostrazione è la stessa. Per completezza svolgiamo il conto esplicitamente:

$$\begin{aligned}
[V,V]_t^{\pi_n} &= \sum_{i\in J_t^{\pi_n}} \left(V_{t_{i+1}^n} - V_{t_i^n}\right)^2 \le \max_{i\in J_t^{\pi_n}} \left|V_{t_{i+1}^n} - V_{t_i^n}\right| \sum_{i\in J_t^{\pi_n}} \left|V_{t_{i+1}^n} - V_{t_i^n}\right| \\
&\le \max_{i\in J_T^{\pi_n}} \left|V_{t_{i+1}^n} - V_{t_i^n}\right| \sum_{i\in J_T^{\pi_n}} \int_{t_i^n}^{t_{i+1}^n} b_s^k ds = \max_{i\in J_T^{\pi_n}} \left|V_{t_{i+1}^n} - V_{t_i^n}\right| \int_0^T b_s^k ds.
\end{aligned}$$

Ora basta osservare che $\max_{i\in J_T^{\pi_n}} \left|V_{t_{i+1}^n} - V_{t_i^n}\right|$ tende a zero q.c. perché le traiettorie di V sono uniformemente continue. Quindi $[V,V]_t^{\pi_n}$ converge in probabilità a zero. ∎

Lemma 10 *Siano V ed M due processi continui a valori reali, V avente variazione quadratica nulla lungo π_n, M variazione quadratica $[M,M]_t$. Allora $M+V$ ha variazione quadratica $[M,M]_t$.*

Proof. E' immediato verificare che

$$[M+V, M+V]_t^{\pi_n} = [M,M]_t^{\pi_n} + [V,V]_t^{\pi_n} + 2\,[M,V]_t^{\pi_n}$$

dove, come sopra, $[M,V]_t^{\pi_n} = \sum_{i \in J_t^{\pi_n}} \left(M_{t_{i+1}^n} - M_{t_i^n} \right) \left(V_{t_{i+1}^n} - V_{t_i^n} \right)$. Basta quindi dimostrare che $[M,V]_t^{\pi_n}$ tende a zero in probabilità. Ma si verifica immediatamente anche la disuguaglianza

$$\left| [M,V]_t^{\pi_n} \right| \leq \sqrt{[M,M]_t^{\pi_n}} \sqrt{[V,V]_t^{\pi_n}}$$

e quindi $[M,V]_t^{\pi_n} \to 0$ in probabilità (si verifichi che il prodotto di una successione convergente in probabilità ed una infinitesima in probabilità è infinitesimo in probabilità).
■

Teorema 32 *Sia $X = (X_t)_{t \in [0,T]}$ un processo di Itô in $\mathbb{R}^d$ della forma*

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dB_s$$

con b in $\mathbb{R}^d$ di classe Λ_B^1 e σ in $\mathbb{R}^{n \times d}$ di classe Λ_B^2. Allora X ha variazione quadratica congiunta; precisamente, per ogni $i, j = 1, ..., d$ e per ogni t, la successione di v.a. $[X^i, X^j]_t^{\pi_n}$ converge in probabilità ed il limite vale

$$\left[X^i, X^j \right]_t = \int_0^t \sigma_s^{ik} \sigma_s^{jk} ds = \int_0^t \left(\sigma_s \sigma_s^T \right)^{ij} ds.$$

Proof. L'operazione $[\cdot,\cdot]_t^{\pi_n}$ è lineare in ambo gli argomenti, quindi, posto $V_t^i = X_0^i + \int_0^t b_s^i ds$, $M_t^{ik} = \int_0^t \sigma_s^{ik} dB_s^k$, vale

$$\left[X^i, X^j \right]_t^{\pi_n} = \left[V^i, X^j \right]_t^{\pi_n} + \left[X^i, V^j \right]_t^{\pi_n} - \left[V^i, V^j \right]_t^{\pi_n} + \sum_{k,k'} \left[M^{ik}, M^{jk'} \right]_t^{\pi_n}.$$

I primi tre termini a destra tendono a zero per i due lemmi precedenti. Usando poi il teorema sulla variazione quadratica degli integrali di Itô,

$$P - \lim_{n \to \infty} \left[M^{ik}, M^{jk'} \right]_t^{\pi_n} = \delta_{kk'} \int_0^t \sigma_s^{ik} \sigma_s^{jk'} ds$$

quindi

$$P - \lim_{n \to \infty} \left[X^i, X^j \right]_t^{\pi_n} = \sum_{k=1}^n \int_0^t \sigma_s^{ik} \sigma_s^{jk} ds.$$

■

Osservazione 25 *Osserviamo che la traccia della matrice* $([X^i, X^j]_t)_{ij}$ *vale*

$$Tr\left(\left[X^i, X^j\right]_t\right)_{ij} = \int_0^t \|\sigma_s\|_{HS}^2 \, ds.$$

C'è una forte somiglianza tra le formule in valore atteso quadratico e quelle in variazione quadratica (si ricordi la formula di isometria $E\left[\left|\int_0^t \sigma_s dB_s\right|^2\right] = E\left[\int_0^t \|\sigma_s\|_{HS}^2 \, ds\right]$*).*

5.2.2 Integrale stocastico rispetto ad un processo di Itô

Sia $X = (X_t)_{t\in[0,T]}$ un processo di Itô in $\mathbb{R}^d$ della forma

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dB_s$$

con b in $\mathbb{R}^d$ di classe Λ_B^1 e σ in $\mathbb{R}^{n\times d}$ di classe Λ_B^2. Per componenti,

$$X_t^i = X_0^i + \int_0^t b_s^i ds + \sum_{k=1}^n \int_0^t \sigma_s^{ik} dB_s^k.$$

Sia Y un processo di classe Λ_B^2. Allora, per ogni $i = 1, ..., d$, poniamo

$$\int_0^t Y_s dX_s^i = \int_0^t Y_s b_s^i ds + \sum_{k=1}^n \int_0^t Y_s \sigma_s^{ik} dB_s^k.$$

Proposizione 47 *Se* Y *è continuo,*

$$\int_0^t Y_s dX_s^i = \lim_{n\to\infty} \sum_{k\in J_t^{\pi n}} Y_{t_k^n}\left(X_{t_{k+1}^n}^i - X_{t_k^n}^i\right).$$

Proof. Vale

$$\begin{aligned}
&\sum_{k\in J_t^{\pi n}} Y_{t_k^n}\left(X_{t_{k+1}^n}^i - X_{t_k^n}^i\right) \\
= &\sum_{k\in J_t^{\pi n}} Y_{t_k^n} \int_{t_k^n}^{t_{k+1}^n} b_s^i ds + \sum_{j=1}^n \sum_{k\in J_t^{\pi n}} Y_{t_k^n} \int_{t_k^n}^{t_{k+1}^n} \sigma_s^{ij} dB_s^j \\
= &\sum_{k\in J_t^{\pi n}} \int_{t_k^n}^{t_{k+1}^n} Y_{t_k^n} b_s^i ds + \sum_{j=1}^n \sum_{k\in J_t^{\pi n}} \int_{t_k^n}^{t_{k+1}^n} Y_{t_k^n} \sigma_s^{ij} dB_s^j
\end{aligned}$$

(l'ultimo passaggio richiederebbe una dimostrazione che omettiamo)

$$= \int_0^t Y_s^n b_s^i ds + \sum_{j=1}^n \int_0^t Y_s^n \sigma_s^{ij} dB_s^j$$

dove Y_s^n è il processo che vale $Y_{t_k^n}$ su $[t_k^n, t_{k+1}^n)$. Basta ora usare la continuità di Y ed alcuni argomenti di passaggio al limite tradizionali. ■

5.3 Formula di Itô

Teorema 33 *Sia $f : [0,1] \times \mathbb{R}^d \to \mathbb{R}$ una funzione di classe $C^{1,2}$ ed $X = (X_t)_{t\in[0,T]}$ un processo di Itô in $\mathbb{R}^d$, della forma*

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dB_s$$

con b in $\mathbb{R}^d$ di classe Λ_B^1 e σ in $\mathbb{R}^{n\times d}$ di classe Λ_B^2. Allora

$$\begin{aligned} f(t, X_t) &= f(0, X_0) + \int_0^t \frac{\partial f}{\partial t}(s, X_s)\, ds + \sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x_i}(s, X_s)\, b_s^i ds \\ &+ \sum_{i=1}^d \sum_{k=1}^n \int_0^t \frac{\partial f}{\partial x_i}(s, X_s)\, \sigma_s^{ik} dB_s^k \\ &+ \frac{1}{2} \sum_{i,j=1}^d \sum_{k=1}^n \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(s, X_s)\, \sigma_s^{ik} \sigma_s^{jk} ds. \end{aligned}$$

In particolare, $f(t, X_t)$ è un processo di Itô.

Osservazione 26 *Si noti che*

$$\sum_{i,j=1}^d \sum_{k=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s)\, \sigma_s^{ik} \sigma_s^{jk} = Trace\left(\sigma_s \sigma_s^T D^2 f(X_s)\right).$$

Proof. Partiamo dal risultato del Corollario 12. L'ipotesi, che $X = (X_t)_{t\in[0,1]}$ abbia variazione quadratica congiunta lungo una successione di partizioni, è verificata per il Teorema 32. Indichiamo poi con (π_n) una successione lungo cui abbiano variazione quadratica congiunta quasi tutte le traiettorie di X. Per il Corollario 12, sappiamo che per ogni $t \in [0,1]$, esiste il limite in probabilità e q.c.

$$\lim_{n\to\infty} \sum_{k\in J_t^{\pi_n}} \sum_{i=1}^d \frac{\partial f}{\partial x^i}\left(t_k^n, X_{t_k^n}\right)\left(X_{t_{k+1}^n}^i - X_{t_k^n}^i\right)$$

che abbiamo indicato con $\int_0^t \nabla f(s, X_s) \cdot dX_s$. Ma per la Proposizione 47, esiste il limite in probabilità di ciascun addendo $\sum_{k\in J_t^{\pi_n}} \frac{\partial f}{\partial x^i}\left(t_k^n, X_{t_k^n}\right)\left(X_{t_{k+1}^n}^i - X_{t_k^n}^i\right)$ della precedente somma e vale

$$\int_0^t \frac{\partial f}{\partial x^i}(s, X_s)\, dX_s^i = \int_0^t \frac{\partial f}{\partial x^i}(s, X_s)\, b_s^i ds + \sum_{k=1}^n \int_0^t \frac{\partial f}{\partial x^i}(s, X_s)\, \sigma_s^{ik} dB_s^k.$$

Pertanto, possiamo riscrivere la formula dimostrata nel Corollario 12 nella forma

$$\begin{aligned} f(t,X_t) &= f(0,X_0) + \int_0^t \frac{\partial f}{\partial t}(s,X_s)\,ds \\ &+ \sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x^i}(s,X_s)\,b_s^i ds + \sum_{i=1}^d \sum_{k=1}^n \int_0^t \frac{\partial f}{\partial x^i}(s,X_s)\,\sigma_s^{ik} dB_s^k \\ &+ \frac{1}{2}\sum_{i,j=1}^d \int_0^t \frac{\partial^2 f}{\partial x^i \partial x^j}(s,X_s)\,d\left[X^i,X^j\right]_s. \end{aligned}$$

Infine, dal Teorema 32 sappiamo che

$$\left[X^i,X^j\right]_t = \int_0^t \sigma_s^{ik}\sigma_s^{jk} ds$$

Basta allora usare il fatto che $\int_0^t \frac{\partial^2 f}{\partial x^i \partial x^j}(s,X_s)\,d\left[X^i,X^j\right]_s$

$$\int_0^t \frac{\partial^2 f}{\partial x^i \partial x^j}(s,X_s)\,d\left[X^i,X^j\right]_s = \int_0^t \frac{\partial^2 f}{\partial x^i \partial x^j}(s,X_s)\,\sigma_s^{ik}\sigma_s^{jk} ds$$

che si può dimostrare prima nel caso in cui σ è continuo (in tal caso $\left[X^i,X^j\right]_t$ è differenziabile e la misura $d\left[X^i,X^j\right]_s$ è $\frac{d\left[X^i,X^j\right]_s}{ds}ds$), poi nel caso generale per approssimazione. La dimostrazione è completa. ■

5.3.1 Notazione differenziale

E' uso scrivere molte delle equazioni precedenti in forma differenziale, pur intendendo con questo la forma integrale corrispondente. Ad esempio, un processo di Itô in $\mathbb{R}^d$ è un processo $X = (X_t)_{t\in[0,T]}$ a valori in $\mathbb{R}^d$ che ha differenziale stocastico della forma

$$dX_t = b_t dt + \sigma_t dB_t$$

e a formula di Itô corrispondente, per $f \in C^{1,2}$, è

$$\begin{aligned} df(t,X_t) &= \frac{\partial f}{\partial t}(t,X_t)\,dt + \sum_{i=1}^d \frac{\partial f}{\partial x_i}(t,X_t)\,b_t^i dt \\ &+ \sum_{i=1}^d \sum_{k=1}^n \frac{\partial f}{\partial x_i}(t,X_t)\,\sigma_t^{ik} dB_t^k \\ &+ \frac{1}{2}\sum_{i,j=1}^d \sum_{k=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(t,X_t)\,\sigma_t^{ik}\sigma_t^{jk} dt. \end{aligned}$$

Si tratta chiaramente di una notazione, ma è utile per sveltire i calcoli. Anche utile è contrarre il termine $b_t^i dt + \sum_{k=1}^n \sigma_t^{ik} dB_t^k$ in dX_t^i per cui scriveremo

$$\begin{aligned} df(t, X_t) &= \frac{\partial f}{\partial t}(t, X_t)\, dt + \sum_{i=1}^d \frac{\partial f}{\partial x_i}(t, X_t)\, dX_t^i \\ &+ \frac{1}{2}\sum_{i,j=1}^d \sum_{k=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(t, X_t)\, \sigma_t^{ik}\sigma_t^{jk} dt. \end{aligned}$$

Infine, può essere utile sostituire $\sum_{k=1}^n \sigma_t^{ik}\sigma_t^{jk} dt$ con $d[X^i, X^j]_t$ e scrivere

$$\begin{aligned} df(t, X_t) &= \frac{\partial f}{\partial t}(t, X_t)\, dt + \sum_{i=1}^d \frac{\partial f}{\partial x_i}(t, X_t)\, dX_t^i \\ &+ \frac{1}{2}\sum_{i,j=1}^d \frac{\partial^2 f}{\partial x_i \partial x_j}(t, X_t)\, d\left[X^i, X^j\right]_t. \end{aligned}$$

Se poi, infine (ma questo è un molto informale) scriviamo

$$d\left[X^i, X^j\right]_t = dX_t^i dX_t^j$$

la formula assume una forma facilissima da ricordare:

$$df(t, X_t) = \frac{\partial f}{\partial t}(t, X_t)\, dt + \sum_{i=1}^d \frac{\partial f}{\partial x_i}(t, X_t)\, dX_t^i + \frac{1}{2}\sum_{i,j=1}^d \frac{\partial^2 f}{\partial x_i \partial x_j}(t, X_t)\, dX_t^i dX_t^j$$

in quanto sembra la riscrittura della formula di Taylor. Possiamo anche scrivere

$$df(t, X_t) = \frac{\partial f}{\partial t}(t, X_t)\, dt + \nabla f(t, X_t) \cdot dX_t + \frac{1}{2}\left\langle D^2 f(t, X_t)\, dX_t, dX_t \right\rangle.$$

Abbiamo un po' rifatto il passaggio all'indietro, dai coefficienti espliciti b e σ del processo di Itô alla notazione con dX e $d[X^i, X^j]_t$. Ogni volta si può scegliere la più conveniente per svolgere i calcoli.

5.3.2 Esempi

Esempio 1. Se

$$dX_t^i = b_t^i dt + \sigma_t^i dB_t, \qquad i = 1, 2$$

allora

$$d\left(X_t^1 X_t^2\right) = X_t^1 dX_t^2 + X_t^2 dX_t^1 + d\left[X^1, X^2\right]_t$$

avendo usato la funzione

$$\begin{aligned} f\left(x^1, x^2\right) &= x^1 x^2 \\ \nabla f\left(x^1, x^2\right) &= \left(x^2, x^1\right) \\ D^2 f\left(x^1, x^2\right) &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \end{aligned}$$

Siccome $d\left[X^1, X^2\right]_t = \sigma_t^1 \sigma_t^2 dt$, troviamo

$$d\left(X_t^1 X_t^2\right) = X_t^1 dX_t^2 + X_t^2 dX_t^1 + \sigma_t^1 \sigma_t^2 dt.$$

Si può anche esplicitare nella forma

$$d\left(X_t^1 X_t^2\right) = X_t^1 b_t^2 dt + X_t^1 \sigma_t^2 dB_t + X_t^2 b_t^1 dt + X_t^2 \sigma_t^1 dB_t + \sigma_t^1 \sigma_t^2 dt$$

che però risulta meno interpretabile.

Esempio 2. Se (in $\mathbb{R}^d$)

$$dX_t = b_t dt + \sigma_t dB_t,$$

allora

$$d\left|X_t\right|^2 = 2X_t dX_t + \sum_{i=1}^{d} \sum_{k=1}^{n} \sigma_t^{ik} \sigma_t^{ik} dt$$

avendo usato la funzione

$$\begin{aligned} f(x) &= |x|^2 \\ \nabla f(x) &= 2x \\ D^2 f(x) &= 2\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

Capitolo 6

Applicazioni della formula di Itô

6.1 Caratterizzazione del moto browniano secondo Lévy

Ricordiamo che un moto browniano $B = \left(B^1, ..., B^d\right)$ in $\mathbb{R}^d$ è una martingala continua e vale $\left[B^i, B^j\right]_t = t \cdot \delta_{ij}$. Il teorema di questa sezione dice che vale il viceversa. Vale anche se si suppone solamente la proprietà di martingala locale. Premettiamo un criterio di indipendenza.

Lemma 11 *Dato $(\Omega, \mathcal{F}, P)$, sia $\mathcal{G}$ una sotto σ-algebra di $\mathcal{F}$. Se un vettore aleatorio X in $\mathbb{R}^d$ ha la proprietà che la v.a. $E\left[e^{i\lambda \cdot X} | \mathcal{G}\right]$ è costante, per ogni $\lambda \in \mathbb{R}^d$, allora X è indipendente da $\mathcal{G}$.*

Proof. Innanzi tutto, se $E\left[e^{i\lambda \cdot X} | \mathcal{G}\right]$ è costante, allora è pari a $E\left[e^{i\lambda \cdot X}\right]$. Quindi

$$E\left[1_A e^{i\lambda \cdot X}\right] = E\left[e^{i\lambda \cdot X}\right] P(A)$$

per ogni $A \in \mathcal{G}$. Usiamo ora alcuni fatti sulla trasformata di Fourier. Ricirdiamo che, data $\varphi \in L^2\left(\mathbb{R}^d\right)$, è ben definita la sua trasformata di Fourier

$$\widehat{\varphi}(\lambda) = \frac{1}{2\pi} \int_{\mathbb{R}^d} e^{-i\lambda \cdot x} \varphi(x)\, dx$$

come funzione di $L^2\left(\mathbb{R}^d\right)$, e che vale la formula di antitrasformata

$$\varphi(x) = \int_{\mathbb{R}^d} e^{i\lambda \cdot x} \widehat{\varphi}(\lambda)\, d\lambda.$$

Prendiamo allora $\varphi \in L^2\left(\mathbb{R}^d\right)$ e la sua trasformata di Fourier $\widehat{\varphi} \in L^2\left(\mathbb{R}^d\right)$, integriamo ambo i membri dell'identità iniziale,

$$\int_{\mathbb{R}^d} \widehat{\varphi}(\lambda)\, E\left[1_A e^{i\lambda \cdot X}\right] d\lambda = P(A) \int_{\mathbb{R}^d} \widehat{\varphi}(\lambda)\, E\left[e^{i\lambda \cdot X}\right] d\lambda$$

ed applichiamo Fubini

$$E\left[1_A \int_{\mathbb{R}^d} \widehat{\varphi}(\lambda) e^{i\lambda\cdot X} d\lambda\right] = P(A) E\left[\int_{\mathbb{R}^d} \widehat{\varphi}(\lambda) e^{i\lambda\cdot X} d\lambda\right].$$

Usando il fatto che $\int_{\mathbb{R}^d} \widehat{\varphi}(\lambda) e^{i\lambda\cdot X} d\lambda = \varphi(X)$ otteniamo

$$E\left[1_A \varphi(X)\right] = P(A) E\left[\varphi(X)\right].$$

Da qui l'indipendenza segue approssimando una generica indicatrice 1_Γ, $\Gamma \in \mathcal{B}\left(\mathbb{R}^d\right)$, con funzioni $\varphi \in L^2\left(\mathbb{R}^d\right)$. ■

Teorema 34 (caratterizzazione di P. Lévy) *Sia $M = \left(M^1, ..., M^d\right)$ una martingala continua vettoriale, rispetto alla filtrazione $(\mathcal{F}_t)_{t\geq 0}$, nulla in zero. Se $\left[M^i, M^j\right]_t = \delta_{ij} t$ allora M è un moto browniano vettoriale, rispetto a $(\mathcal{F}_t)_{t\geq 0}$.*

Proof. Nota. In un punto della dimostrazione useremo la teoria generale dell'integrazione per martingale continue e la corrispondente formula di Itô, elementi che non abbiamo completamente sviluppato. In realtà si tratta semplicemente della formula di Itô del Corollario 12; ma dovremmo dimostrare che una qualsiasi martingala continua vettoriale ha variazione quadratica congiunta. Lo abbiamo fatto però per i processi di Itô e, osserviamo, quando useremo questo teorema di caratterizzazione nell'ambito della formula di Girsanov, lì la martingala M sarà un processo di Itô. In conclusione, anche se la dimostrazione che svolgiamo ora non è del tutto completa, lo è nel caso che verrà utilizzato in seguito.

Basta verificare che, presi $t \geq s \geq 0$, il vettore $M_t - M_s$ è gaussiano di media nulla e covarianza $t \cdot Id$, ed indipendente da $\mathcal{F}_s$. Preso $\lambda \in \mathbb{R}^d$, se verifichiamo che $\lambda \cdot (M_t - M_s)$ è una v.a. gaussiana $N\left(0, |\lambda|^2 (t-s)\right)$, allora $M_t - M_s$ è gaussiano di media nulla e covarianza $t \cdot Id$. Usando il teorema di inversione delle funzioni caratteristiche, basta verificare che

$$E\left[e^{i\lambda\cdot(M_t - M_s)}\right] = e^{-\frac{1}{2}|\lambda|^2(t-s)}.$$

Cerchiamo di capire le proprietà di

$$Y_t^\lambda := e^{i\lambda\cdot M_t + \frac{1}{2}|\lambda|^2 t}$$

tramite la formula di Itô. Questa, per funzioni a valori complessi, segue le regole usuali, come si può verificare applicandola a parte reale e parte immaginaria. Usiamo, come già detto, la versione del Corollario 12.

Allora, presa $f(t,x) = e^{i\lambda\cdot x+\frac{1}{2}|\lambda|^2 t}$ ($\frac{\partial f}{\partial t} = \frac{1}{2}|\lambda|^2 f$, $\nabla f = i\lambda f$, $D^2 f = -\lambda \otimes \lambda f$), $Y_t^\lambda = f(t, M_t)$, abbiamo (qui usiamo in modo cruciale l'ipotesi $[M^i, M^j]_t = \delta_{ij} t$)

$$\begin{aligned} dY_t^\lambda &= \frac{1}{2}|\lambda|^2 Y_t^\lambda dt + Y_t^\lambda i\lambda \cdot dM_t - \frac{1}{2} Y_t^\lambda \sum_{i,j=1}^d \lambda_i \lambda_j d\left[M^i, M^j\right]_t \\ &= \frac{1}{2}|\lambda|^2 Y_t^\lambda dt + Y_t^\lambda i\lambda \cdot dM_t - \frac{1}{2} Y_t^\lambda \sum_{i=1}^d \lambda_i^2 dt \\ &= Y_t^\lambda i\lambda \cdot dM_t \end{aligned}$$

ovvero

$$Y_t^\lambda = \int_0^t Y_s^\lambda i\lambda \cdot dM_s$$

Non abbiamo sviluppato la teoria dell'integrazione rispetto a martingale continue qualsiasi, ma possiamo immaginare che l'integrale $\int_0^t Y_s^\lambda i\lambda \cdot dM_s$ sia una martingala, se l'integrando $Y_s^\lambda i\lambda$ soddisfa opportune ipotesi di sommabilità. Nel nostro caso $Y_s^\lambda i\lambda$ è addirittura limitato:

$$\left|Y_s^\lambda i\lambda\right| = \left|e^{i\lambda\cdot M_t} e^{\frac{1}{2}|\lambda|^2 t} i\lambda\right| = e^{\frac{1}{2}|\lambda|^2 t}|\lambda| \le e^{\frac{1}{2}|\lambda|^2 T}|\lambda|.$$

In tal caso si può dimostrare che $\int_0^t Y_s^\lambda i\lambda \cdot dM_s$, e quindi Y_t^λ, è una martingala.

Ma allora, per $t \ge s \ge 0$, $E\left[Y_t^\lambda | \mathcal{F}_s\right] = Y_s^\lambda$, ovvero

$$E\left[e^{i\lambda\cdot M_t} | \mathcal{F}_s\right] = e^{i\lambda\cdot M_s} e^{\frac{1}{2}|\lambda|^2 s} e^{-\frac{1}{2}|\lambda|^2 t}$$

da cui

$$\begin{aligned} E\left[e^{i\lambda\cdot(M_t-M_s)} | \mathcal{F}_s\right] &= E\left[e^{i\lambda\cdot M_t} e^{-i\lambda\cdot M_s} | \mathcal{F}_s\right] = e^{-i\lambda\cdot M_s} E\left[e^{i\lambda\cdot M_t} | \mathcal{F}_s\right] \\ &= e^{-\frac{1}{2}|\lambda|^2 (t-s)}. \end{aligned}$$

Ne segue $E\left[e^{i\lambda\cdot(M_t-M_s)}\right] = e^{-\frac{1}{2}|\lambda|^2(t-s)}$, come volevamo dimostrare. Inoltre la v.a. $E\left[e^{i\lambda\cdot(M_t-M_s)} | \mathcal{F}_s\right]$ è costante, per ogni $\lambda \in \mathbb{R}^d$ e questo implica, per il lemma precedente, che $M_t - M_s$ è indipendente da $\mathcal{F}_s$. ■

6.2 Il teorema di Girsanov

6.2.1 Cambi di misura

Ci accingiamo ad eseguire un'operazione non comune fino ad ora: modificare la probabilità P su $(\Omega, \mathcal{F})$ tramite una densità. Premettiamo quindi alcune semplici idee per inquadrare il problema.

Sia $(\Omega, \mathcal{F})$ uno spazio probabilizzabile. Fino ad ora abbiamo sempre fissato una misura di probabilità P su di esso ed indicato con E il valore atteso corrispondente. D'ora in poi, quando servirà, lo indicheremo con E^P per evidenziare che si tratta del valore atteso calcolato tramite la misura P.

Supponiamo di avere due diverse misure di probabilità P e Q, su $(\Omega, \mathcal{F})$. Ricordiamo che $Q << P$ (Q assolutamente continua rispetto a P) se $P(A) = 0$ implica $Q(A) = 0$, $A \in \mathcal{F}$, e questo equivale all'esistenza di una densità di probabilità $\rho = \frac{\partial Q}{\partial P}$, ovvero una funzione non negativa misurabile $\rho : \Omega \to [0, \infty)$ tale che

$$Q(A) = \int_A \rho dP$$

per ogni $A \in \mathcal{F}$. Oppure

$$E^Q[X] = E^P[\rho X]$$

per ogni v.a. X limitata (ed altre). Scriviamo anche $dQ = \rho dP$.

Le misure P e Q si dicono equivalenti se $Q << P$ e $P << Q$. Se $Q << P$ e $\rho = \frac{\partial Q}{\partial P}$ è strettamente positiva q.c., allora sono equivalenti: ρ^{-1} è $\frac{\partial P}{\partial Q}$:

$$E^Q\left[\rho^{-1}X\right] = E^P\left[\rho\rho^{-1}X\right] = E^P[X]$$

per ogni v.a. X limitata.

Il concetto di martingala cambia a seconda della misura: un processo può essere una martingala rispetto a P e non rispetto a Q (il valore atteso, che dipende dalla misura scelta, entra in gioco nella definizione di martingala).

Invece il concetto di convergenza in probabilità non dipende dalla misura, se si considerano misure equivalenti. Infatti, se $X_n \xrightarrow{P} X$ allora per ogni $\varepsilon > 0$

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0$$

e quindi

$$Q(|X_n - X| > \varepsilon) = \int_{|X_n - X| > \varepsilon} \rho dP \to 0$$

per l'assoluta continuità dell'integrale. E viceversa.

Pertanto il concetto di variazione quadratica di un processo stocastico non dipende dalla misura, se si considerano misure equivalenti, essendo legato alla misura solo attraverso una condizione di convergenza in probabilità.

Circa le classi per cui definiamo integrali stocastici, la classe M^2 dipende dalla misura P di riferimento, mentre la classe Λ^2 no se si considerano misure equivalenti.

6.2.2 Martingale esponenziali e misure equivalenti

Sia $(\Omega, \mathcal{F}, P)$ uno spazio probabilizzato, $(\mathcal{F}_t)_{t\geq 0}$ una filtrazione, B un moto browniano in $\mathbb{R}^d$. Sia $(X_t)_{t\in[0,T]}$ un processo in $\mathbb{R}^d$ di classe Λ^2. Consideriamo il processo stocastico

$$L_t = \int_0^t X_s \cdot dB_s$$

avente variazione quadratica

$$[L, L]_t = \int_0^t |X_s|^2 \, ds.$$

Consideriamo poi il processo stocastico

$$\rho_t = e^{L_t - \frac{1}{2}[L,L]_t} = e^{\int_0^t X_s \cdot dB_s - \frac{1}{2}\int_0^t |X_s|^2 ds}.$$

Vale:

Lemma 12

$$\rho_t = \int_0^t \rho_s X_s \cdot dB_s.$$

In generale,

$$E[\rho_T] \leq 1.$$

Se $E\left[\int_0^T \rho_s^2 |X_s|^2 \, ds\right] < \infty$*, allora* ρ_t *è una martingala, rispetto a* P*, ed in particolare*

$$E[\rho_T] = 1.$$

Proof. Abbiamo che $\rho_t = f(Y_t)$ dove $f(x) = e^x$ e

$$dY_t = X_t \cdot dB_t - \frac{1}{2}|X_t|^2 \, dt$$

quindi, per la formula di Itô,

$$\begin{aligned}
d\rho_t &= f'(Y_t)\, dY_t + \frac{1}{2} f''(Y_t)\, d[Y,Y]_t \\
&= e^{Y_t}\left(X_t \cdot dB_t - \frac{1}{2}|X_t|^2 \, dt\right) + \frac{1}{2} e^{Y_t} |X_t|^2 \, dt \\
&= e^{Y_t} X_t \cdot dB_t.
\end{aligned}$$

Prendiamo poi $\tau_n = \inf\{t > 0 : \rho_t \geq n\}$ (infinito altrimenti). Vale

$$\rho_{t\wedge\tau_n} = \int_0^t 1_{s\leq\tau_n} \rho_s X_s \cdot dB_s$$

ed ora $1_{s\leq\tau_n}\rho_s |X_s|$ è limitato (si ricordi che X è limitato) e quindi di classe M^2, quindi $\rho_{t\wedge\tau_n}$ è una martingala e vale $E\left[\rho_{t\wedge\tau_n}\right] = E[\rho_0] = 1$. Per Fatou, usando la continuità delle traiettorie di ρ_t,

$$E[\rho_t] = E\left[\lim_{n\to\infty}\rho_{t\wedge\tau_n}\right] \leq 1.$$

Le restanti affermazioni sono ovvie. ■

Esaminiamo delle condizioni sufficienti affinché ρ_t sia una martingala.

Proposizione 48 *Se X è limitato (esiste $C > 0$ tale che $|X_t| \leq C$ per ogni t q.c.), allora ρ_t è una martingala. Vale inoltre*

$$\sup_{t\in[0,T]} E\left[\rho_t^p\right] < \infty$$

per ogni $p \geq 1$.

Proof. Vogliamo usare la condizione sufficiente $E\left[\int_0^T \rho_s^2 |X_s|^2 ds\right] < \infty$. Per l'ipotesi su X, basta verificare che $E\left[\int_0^T \rho_s^2 ds\right] < \infty$. Abbiamo

$$\rho_t^2 = e^{2L_t - [L,L]_t} = e^{2L_t - \alpha[L,L]_t} e^{\beta[L,L]_t}$$

con $\alpha - \beta = 1$. Allora

$$E\left[\rho_t^2\right] \leq E\left[e^{2pL_t - \alpha p[L,L]_t}\right]^{1/p} E\left[e^{\beta p'[L,L]_t}\right]^{1/p'}$$

Scriviamo

$$e^{2pL_t - \alpha p[L,L]_t} = e^{2pL_t - \frac{\alpha}{4p}[2pL,2pL]_t}$$

e scegliamo α e p in modo che sia $\frac{\alpha}{4p} = \frac{1}{2}$. Applicando il lemma precedente al processo $2pL_t$ (cioè $2pX_t$) abbiamo $E\left[e^{2pL_t - \alpha p[L,L]_t}\right] \leq 1$. Quindi

$$E\left[\rho_t^2\right] \leq E\left[e^{\beta p'[L,L]_t}\right]^{1/p'}.$$

Ma X è limitato da C, quindi anche $[L,L]_t$ da $C^2 T$, da cui $\sup_{t\in[0,T]} E\left[\rho_t^2\right] < \infty$.

Con la stessa dimostrazione, solo simbolicamente più complicata, si verifica che $\sup_{t\in[0,T]} E\left[\rho_t^p\right] < \infty$ per ogni $p \geq 1$. La dimostrazione è completa. ■

Con un po' più di sforzo (ragionamenti simili ma non più basati sulla condizione sufficiente $E\left[\int_0^T \rho_s^2 |X_s|^2 ds\right] < \infty$), si può dimostrare il seguente criterio, detto *di Novikov*.

Teorema 35 *Se*

$$E\left[e^{\lambda \int_0^T |X_s|^2 ds}\right] < \infty$$

per qualche $\lambda > \frac{1}{2}$ allora vale $E\left[\int_0^T \rho_s^2 |X_s|^2 ds\right] < \infty$ e quindi ρ_t è una martingala.

La martingala esponenziale $\rho_t = e^{L_t - \frac{1}{2}[L,L]_t}$ (quando è appunto una martingala) permette di definire una nuova misura su $(\Omega, \mathcal{F})$. Per la proprietà di martingala, $E[\rho_T] = 1$. La funzione $\rho_T : \Omega \to \mathbb{R}$ è non negativa e tale che $\int_\Omega \rho_T dP = 1$, cioè è una densità di probabilità, per cui possiamo introdurre la nuova misura di probabilità Q su $(\Omega, \mathcal{F})$ definita da

$$dQ = \rho_T dP$$

ovvero

$$E^Q[X] = \int_\Omega X dQ = \int_\Omega X \rho_T dP = E^P[X \rho_T]$$

per ogni v.a. X limitata. Abbiamo $\rho_T > 0$, quindi P e Q sono equivalenti (sezione precedente).

Useremo il seguente fatto.

Lemma 13 *Se Y è una v.a. $\mathcal{F}_t$-misurabile e limitata, $t \in [0,T]$, allora*

$$E^Q[Y] = E^P[Y \rho_t]$$

ovvero $dQ = \rho_t dP$ limitatamente agli eventi $A \in \mathcal{F}_t$. Scriveremo

$$\rho_t = \left. \frac{dQ}{dP} \right|_{\mathcal{F}_t}.$$

Proof. Per la proprietà di martingala di ρ_t e per definizione di Q,

$$E^Q[Y] = E^P[Y \rho_T] = E^P[Y \rho_t].$$

■

Osservazione 27 *Il risultato vale anche per v.a. Y, $\mathcal{F}_t$-misurabili, tali che uno dei due valori attesi $E^Q[Y]$ o $E^P[Y \rho_t]$ sia finito. Basta scomporre Y in parte positiva e negativa ed approssimare dal basso con troncamenti limitati.*

Osservazione 28 *Se avessimo preso dall'inizio $T = +\infty$, avendo così una martingala ρ_t definita per tutti i tempi $t \geq 0$, avremmo potuto introdurre inizialmente una famiglia di misure $dQ_T = \rho_T dP$ indicizzata da $T \geq 0$, osservando poi però che, in base al lemma,*

$$\left. \frac{dQ_T}{dP} \right|_{\mathcal{F}_t} = \left. \frac{dQ_t}{dP} \right|_{\mathcal{F}_t}$$

per ogni $T \geq t \geq 0$. Possiamo quindi dire che esiste una funzione $Q : \bigvee_{t \geq 0} \mathcal{F}_t \to [0,1]$, σ-additiva su ogni $\mathcal{F}_t$, tale che $\left. \frac{dQ}{dP} \right|_{\mathcal{F}_t} = \rho_t$ per ogni $t \geq 0$. Anche per questo non abbiamo sottolineato la dipendenza da T nella notazione Q, precedentemente.

6.2.3 Teorema di Girsanov

Il risultato che stiamo per esporre vale più in generale ma lo formuliamo e dimostriamo in un caso particolare per evitare troppi dettagli tecnici.

Sia $(\Omega, \mathcal{F}, P)$ uno spazio probabilizzato, $(\mathcal{F}_t)_{t\geq 0}$ una filtrazione, B un moto browniano in $\mathbb{R}^d$. Sia $(X_t)_{t\in[0,T]}$ un processo in $\mathbb{R}^d$ di classe Λ^2. Supponiamo per semplicità che sia limitato: esiste $C > 0$ tale che $|X_t| \leq C$ per ogni t q.c.; quindi in realtà è di classe M^2. Consideriamo come sopra il processo stocastico

$$L_t = \int_0^t X_s \cdot dB_s$$

che, nelle attuali ipotesi, è una martingala e introduciamo la martingala esponenziale (Proposizione 48)

$$\rho_t = e^{L_t - \frac{1}{2}[L,L]_t}.$$

Come detto sopra, questa definisce la misura equivalente

$$dQ = \rho_T dP.$$

Lemma 14 *Sia M un processo stocastico adattato. Se $\rho_t M_t$ è una martingala rispetto a P, allora M_t è una martingala rispetto a Q.*

Proof. Intanto bisogna verificare che M sia integrabile rispetto a Q. Per ogni $n \in \mathbb{N}$ vale, in virtù del Lemma 13,

$$E^Q [|M_t| \wedge n] = E^P [\rho_t (|M_t| \wedge n)]$$

e quindi per convergenza monotona

$$E^Q [|M_t|] = E^P [\rho_t |M_t|]$$

ma $E^P [\rho_t |M_t|] = E^P [|\rho_{tMt}|]$ è finito perché ρ_{tMt} è integrabile rispetto a P (è una P-martingala). Quindi $E^Q [|M_t|]$ è finito.

Dobbiamo ora dimostrare che, presi $t \geq s \geq 0$,

$$E^Q [M_t Z] = E^Q [M_s Z]$$

per ogni v.a. $\mathcal{F}_s$-misurabile e limitata Z. Usando il Lemma 13 (si veda anche l'osservazione che lo segue) e l'ipotesi che $\rho_t M_t$ sia una P-martingala, vale

$$E^Q [M_t Z] = E^P [\rho_t M_t Z] = E^P [\rho_s M_s Z] = E^Q [M_s Z].$$

■

Lemma 15 *Il processo stocastico*

$$M_t := B_t - \int_0^t X_s ds.$$

è una martingala continua rispetto a Q ed $(\mathcal{F}_t)_{t\geq 0}$. Vale

$$\left[M^i, M^j\right]_t = \delta_{ij} t$$

(si noti che questa è la variazione mutua sia rispetto a P che Q).

Proof. La continuità di M è nota. Per dimostrare che M è una martingala rispetto a Q basta verificare, per il Lemma 14, che $\rho_t M_t$ è una martingala rispetto a P.

Sappiamo poi che $d\rho_t = \rho_t X_t \cdot dB_t$ e che $dM_t = dB_t - X_t ds$. Quindi per la formula di Itô

$$d\left(\rho_t M_t^i\right) = M_t^i \rho_t X_t \cdot dB_t + \rho_t \left(dB_t^i - X_t^i dt\right) + d\left[\rho, M^i\right]_t.$$

Ora, posto $Y_t^j = \int_0^t \rho_s X_s^j dB_s^j$, così che $\rho_t = \sum_j Y_t^j$, vale

$$\left[\rho, M^i\right]_t = \sum_j \left[Y^j, M^i\right]_t = \sum_j \delta_{ij} \int_0^t \rho_s X_s^j ds = \int_0^t \rho_s X_s^i ds.$$

Quindi

$$\begin{aligned} d\left(\rho_t M_t^i\right) &= M_t^i \rho_t X_t \cdot dB_t + \rho_t \left(dB_t^i - X_t^i dt\right) + \rho_t X_t^i dt \\ &= M_t^i \rho_t X_t \cdot dB_t + \rho_t dB_t^i \end{aligned}$$

ovvero

$$\rho_t M_t^i = \rho_0 M_0^i + \int_0^t M_s^i \rho_s X_s \cdot dB_s + \int_0^t \rho_s dB_s^i.$$

Per avere che $\rho_t M_t^i$ è una martingala, basta sapere che $M_t^i \rho_t X_t$ e ρ_t sono di classe M^2 rispetto a P. La seconda affermazione discende subito dalla stima della Proposizione 48. Per la prima è sufficiente, ricordando che X è limitato, verificare che

$$E\left[\int_0^T \left|M_t^i\right|^2 \rho_t^2 dt\right] < \infty.$$

Vale

$$E\left[\int_0^T \left|M_t^i\right|^2 \rho_t^2 dt\right] \leq E\left[\int_0^T \left|M_t^i\right|^4 dt\right]^{1/2} E\left[\int_0^T \rho_t^4 dt\right]^{1/2}.$$

Da un lato

$$E\left[\int_0^T \rho_t^4 dt\right] \leq \sup_{t\in[0,T]} E\left[\rho_t^4\right] < \infty$$

per quanto verificato poco sopra. Dall'altro,

$$\begin{aligned} E\left[\int_0^T \left|M_t^i\right|^4 dt\right] &\leq 4E\left[\int_0^T \left|B_t^i\right|^4 dt\right] + 4E\left[\int_0^T \left|\int_0^t X_s^i ds\right|^4 dt\right] \\ &\leq 4 \sup_{t\in[0,T]} E\left[\left|B_t^i\right|^4\right] + 4E\left[\int_0^T \left(\int_0^t \left|X_s^i\right| ds\right)^4 dt\right]. \end{aligned}$$

Il primo termine è limitato per note proprietà di integrabilità del moto browniano, il secondo perché X è limitato.

L'affermazione sulla variazione quadratica è nota (se vista rispetto a P). Il fatto rilevante è che valga anche rispetto a Q, fatto generale essendo P e Q equivalenti. La dimostrazione è completa. ■

Mettendo insieme questi lemmi ed il criterio di Lévy abbiamo dimostrato il seguente:

Teorema 36 (Girsanov) *Il processo stocastico*

$$\widetilde{B}_t := B_t - \int_0^t X_s ds.$$

è un moto browniano rispetto a Q *ed* $(\mathcal{F}_t)_{t\geq 0}$.

Capitolo 7

Equazioni differenziali stocastiche

7.1 Richiami sulle equazioni differenziali ordinarie

Il classico problema di Cauchy in $\mathbb{R}^d$, su un intervallo di tempo limitato $[0,T]$, ha la forma

$$\begin{cases} \frac{dx_t}{dt} = f(t, x_t) & t \in [0,T] \\ x|_{t=0} = x_0 \end{cases} . \tag{7.1}$$

La sua forma integrale è

$$x_t = x_0 + \int_0^t f(s, x_s)\, ds \qquad t \in [0,T]\,. \tag{7.2}$$

Supponiamo $f : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ almeno continua. Una funzione $x_\cdot : [0,T] \to \mathbb{R}^d$ si dice soluzione di (7.1) se è di classe $C^1\left([0,T]; \mathbb{R}^d\right)$ e soddisfa le condizioni scritte in (7.1). Si dice soluzione di (7.2) se è di classe $C\left([0,T]; \mathbb{R}^d\right)$ e soddisfa l'equazione integrale (7.2). Si verifica senza difficoltà che una soluzione di (7.1) è anche soluzione di (7.2), e che una soluzione di (7.2) è (grazie a (7.2) stessa) di classe $C^1\left([0,T]; \mathbb{R}^d\right)$ ed è soluzione di (7.1). Quindi le due formulazioni sono equivalenti; ma la formulazione integrale permette di lavorare nello spazio più grande $C\left([0,T]; \mathbb{R}^d\right)$, o se si vuole, permette di lavorare con meno regolarità (un fatto chiave in vista del caso stocastico).

Ricordiamo alcuni risultati ed alcune tecniche. Introduciamo, la seguente ipotesi:

$$\begin{aligned} f \text{ continua, esiste } L &> 0 \text{ tale che} \\ |f(t,x) - f(t,y)| &\leq L\,|x-y| \text{ per ogni } t \in [0,T] \text{ e } x, y \in \mathbb{R}^d. \end{aligned} \tag{7.3}$$

Proposizione 49 *Supponiamo (7.3). Nella classe $C\left([0,T]; \mathbb{R}^d\right)$ c'è unicità per (7.2).*

Proof. Siano x^1 e x^2 due soluzioni e sia $y = x^1 - x^2$. Abbiamo

$$y_t = \int_0^t \left(f\left(s, x_s^1\right) - f\left(s, x_s^2\right)\right) ds$$

da cui

$$|y_t| \leq \int_0^t \left| f\left(s, x_s^1\right) - f\left(s, x_s^2\right) \right| ds \leq L \int_0^t \left| x_s^1 - x_s^2 \right| ds = L \int_0^t |y_s| \, ds.$$

Ricordiamo il Lemma di Gronwall: se v_t, a_t, $t \in [0, T]$, sono funzioni non negative che soddisfano

$$v_t \leq v_0 + \int_0^t a_s v_s ds, \qquad t \in [0, T]$$

allora

$$v_t \leq v_0 e^{\int_0^t a_s ds}, \qquad t \in [0, T].$$

Applicato al nostro caso produce $|y_t| \leq 0$, da cui $x_t^1 - x_t^2$, per ogni $t \in [0, T]$. La dimostrazione è completa. ■

Teorema 37 *Supponiamo (7.3). Nella classe $C\left([0, T]; \mathbb{R}^d\right)$ c'è esistenza per (7.2).*

Proof. Ricordiamo due dimostrazioni.

1. Dimostrazione col principio delle contrazioni. Per ogni $\tau \in [0, T]$, consideriamo lo spazio $X_\tau = C\left([0, \tau]; \mathbb{R}^d\right)$ e l'applicazione $\Gamma_\tau : X_\tau \to X_\tau$ definita da

$$\Gamma_\tau (y_\cdot)(t) = x_0 + \int_0^t f(s, y_s) \, ds.$$

Su X_τ consideriamo la metrica $d_\tau\left(y^1, y^2\right) = \max_{t \in [0, \tau]} \left|y_t^1 - y_t^2\right|$. Vale

$$\Gamma_\tau\left(y^1\right)(t) - \Gamma_\tau\left(y^2\right)(t) = \int_0^t \left(f\left(s, y_s^1\right) - f\left(s, y_s^2\right)\right) ds$$

da cui, come sopra,

$$\left|\Gamma_\tau\left(y^1\right)(t) - \Gamma_\tau\left(y^2\right)(t)\right| \leq L \int_0^t \left|y_s^1 - y_s^2\right| ds.$$

Possiamo ulteriormente maggiorare con

$$\leq Lt \max_{t \in [0, t]} \left|y_t^1 - y_t^2\right| \leq L\tau d_\tau\left(y^1, y^2\right).$$

Abbiamo dimostrato quindi che

$$d_\tau\left(\Gamma_\tau\left(y^1\right), \Gamma_\tau\left(y^2\right)\right) \leq L\tau d_\tau\left(y^1, y^2\right).$$

Se prendiamo τ in modo che $L\tau < 1$, l'applicazione Γ_τ è una contrazione e quindi ha un unico punto fisso in X_τ, ovvero c'è una ed una sola funzione $x \in X_\tau$ che risolve (7.2), limitatamente ai tempi $t \in [0, \tau]$.

Si deve poi ripetere il procedimento su un intervallo della forma $[\tau, \tau_2]$ con dato iniziale al tempo τ dato dal valore x_τ della soluzione trovata su $[0, \tau]$. Si scopre che, grazie all'ipotesi (7.3), si può prendere $\tau_2 = 2\tau$ e provare che esiste una sola funzione $x \in C\left([\tau, 2\tau]; \mathbb{R}^d\right)$ che risolve (7.2), per tempi $t \in [\tau, 2\tau]$, con dato iniziale x_τ. Incollandole, si trova una soluzione $x \in C\left([0, 2\tau]; \mathbb{R}^d\right)$ che risolve (7.2), per tempi $t \in [0, 2\tau]$, con dato iniziale x_0. Procedendo nello stesso modo su intervalli successivi, si ricopre $[0, T]$ in un numero finito di passi. Questa dimostrazione, per inciso, fornisce anche l'unicità.

2. Dimostrazione con le approssimazioni successive. Definiamo una successione di funzioni $x^n_\cdot \in C\left([0, T]; \mathbb{R}^d\right)$ per ricorrenza, ponendo $x^0_\cdot = x_0$ e, per $n \geq 0$,

$$x_t^{n+1} = x_0 + \int_0^t f\left(s, x_s^n\right) ds \qquad t \in [0, T].$$

Con gli stessi calcoli fatti al punto precedente, si verifica che

$$d_\tau\left(x^{n+1}, x^n\right) \leq L\tau d_\tau\left(x^n, x^{n-1}\right).$$

Quindi

$$d_\tau\left(x^{n+1}, x^n\right) \leq (L\tau)^n d_\tau\left(x^1, x^0\right).$$

Da qui è facile dedurre che (x^n) è di Cauchy in X_τ, se si prende, come sopra, τ in modo che sia $L\tau < 1$. Lo spazio X_τ è completo, quindi $x^n \to x \in X_\tau$ e si verifica facilmente che x risolve (7.2) su $[0, \tau]$. Il procedimento si deve poi ripetere su intervallini successivi come nella dimostrazione precedente. In realtà, la dimostrazione fatta ora non è altro che la dimostrazione del principio delle contrazioni.

3. Dimostrazione in un solo passo. Si può sviluppare una variante delle dimostrazioni precedenti, es. della 2, che agisce in un solo passo. Si consideri sullo spazio X_T la distanza $\delta_\lambda\left(y^1, y^2\right) = \max_{t \in [0,\tau]}\left(e^{-\lambda t}\left|y_t^1 - y_t^2\right|\right)$, per ogni $\lambda \geq 0$. Con le notazione del punto 2, abbiamo come sopra

$$\left|x_t^{n+1} - x_t^n\right| \leq \int_0^t \left|f\left(s, x_s^n\right) - f\left(s, x_s^{n-1}\right)\right| ds \leq L \int_0^t \left|x_s^n - x_s^{n-1}\right| ds$$

ma ora maggioriamo stimiamo

$$\begin{aligned} e^{-\lambda t}\left|x_t^{n+1} - x_t^n\right| &\leq L \int_0^t e^{-\lambda(t-s)} e^{-\lambda s}\left|x_s^n - x_s^{n-1}\right| ds \\ &\leq L \max_{s \in [0,t]}\left(e^{-\lambda s}\left|x_s^n - x_s^{n-1}\right|\right) \int_0^t e^{-\lambda(t-s)} ds \\ &\leq \lambda^{-1} L \delta_\lambda\left(x^n, x^{n-1}\right) \end{aligned}$$

in quanto $\int_0^t e^{-\lambda(t-s)} ds = \lambda^{-1}\left(1 - e^{-\lambda t}\right)$. Quindi

$$\delta_\lambda\left(x^{n+1}, x^n\right) \leq \left(\lambda^{-1} L\right)^n \delta_\lambda\left(x^1, x^0\right).$$

Ora basta prendere $\lambda = 2L$. Naturalmente, le norme δ_{2L} e d_T sono equivalenti. ■

7.2 Equazioni differenziali stocastiche con noise additivo

Sia $(\Omega, \mathcal{F}, P)$ uno spazio probabilizzato, $(\mathcal{F}_t)_{t\geq 0}$ una filtrazione, B un moto browniano in $\mathbb{R}^n$. Chiameremo *base stocastica* la famiglia $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P, (B_t)_{t\geq 0}\right)$.

Sia $\sigma \in \mathbb{R}^{n\times d}$ una matrice. Vorremmo dare senso ad un'equazione della forma

$$\frac{dX_t}{dt} = b(t, X_t) + \sigma \frac{dB_t}{dt}$$

dove $b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$.

Osservazione 29 *L'idea da cui nasce questo tipo di equazione è la seguente. Si immagina che il sistema fisico, economico, finanziario, biologico ecc. che stiamo esaminando sia descritto da alcune leggi riassunte dall'equazione* $\frac{dX_t}{dt} = f(t, X_t)$*; ma che questa descrizione non sia esaustiva, in quanto siano presenti delle perturbazioni che non sappiamo descrivere con lo stesso gradi di precisione e che consideriamo causali. La scelta particolare del termine* $\sigma\frac{dB_t}{dt}$ *è dettata da queste sue caratteristiche intuitive: gaussianità, media nulla, indipendenza istante per istante, invarianza temporale. Si suggerisce di accontentarsi di comprendere intuitivamente queste affermazioni, visto che la loro formalizzazione sarebbe difficile.*

In analogia con quanto visto sopra, scriviamo la forma integrale

$$X_t = X_0 + \int_0^t b(s, X_s)\, ds + \sigma B_t, \qquad t \in [0,T]. \tag{7.4}$$

Questa ha perfettamente senso (tutti i termini sono ben definiti). Supponiamo che b sia almeno continua e che X_0 sia un vettore aleatorio assegnato, definito su $(\Omega, \mathcal{F}, P)$ a valori in $\mathbb{R}^d$, $\mathcal{F}_0$-misurabile; e che $\sigma \in \mathbb{R}^{n\times d}$ sia una matrice data.

Possiamo anche fissare $\omega \in \Omega$ e considerare l'equazione completamente deterministica

$$x_t = X_0(\omega) + \int_0^t b(s, x_s)\, ds + \sigma B_t(\omega), \qquad t \in [0,T]. \tag{7.5}$$

Definizione 20 *Fissata una base stocastica* $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P, (B_t)_{t\geq 0}\right)$ *ed una v.a.* $\mathcal{F}_0$*-misurabile* X_0*, chiamiamo soluzione forte di (7.4) ogni processo stocastico* $X = (X_t)_{t\in[0,T]}$ *definito su* $(\Omega, \mathcal{F}, P)$*, continuo, adattato alla filtrazione* $(\mathcal{F}_t)_{t\geq 0}$*, che soddisfa (7.4) uniformemente in* $t \in [0,T]$ *con probabilità uno:*

$$P\left(X_t = X_0 + \int_0^t b(s, X_s)\, ds + \sigma B_t, \forall t \in [0,T]\right) = 1.$$

Diciamo che c'è esistenza forte quando questo vale a partire da qualsiasi base stocastica e qualsiasi v.a. $\mathcal{F}_0$*-misurabile* X_0*.*

Ora che abbiamo dato una definizione precisa di soluzione, permettiamoci di introdurre una notazione più leggera, a patto di ricordare sempre che si tratta solo di una notazione. Scriveremo l'equazione (7.4) nella forma

$$dX_t = b\left(t, X_t\right) dt + \sigma dB_t.$$

Veniamo ai concetti di unicità, rilevanti per questa equazione.

Definizione 21 *Diciamo che vale l'unicità forte (pathwise uniqueness, in lingua inglese), se, data una qualsiasi base stocastica* $\left(\Omega, \mathcal{F}, \left(\mathcal{F}_t\right)_{t\geq 0}, P, \left(B_t\right)_{t\geq 0}\right)$ *ed una qualsiasi v.a.* $\mathcal{F}_0$*-misurabile* X_0*, prese due soluzioni forti* X^1 *e* X^2*, esse sono indistinguibili.*

Definizione 22 *Diciamo che c'è unicità per quasi ogni* ω *se, data una qualsiasi base stocastica* $\left(\Omega, \mathcal{F}, \left(\mathcal{F}_t\right)_{t\geq 0}, P, \left(B_t\right)_{t\geq 0}\right)$ *ed una qualsiasi v.a.* $\mathcal{F}_0$*-misurabile* X_0*, esiste un insieme di misura uno* $\Omega_0 \in \mathcal{F}$ *tale che, per ogni* $\omega \in \Omega_0$*, l'equazione deterministica (7.5) ha unicità in* $C\left(\left[0, T\right]; \mathbb{R}^d\right)$*.*

Si può verificare che se c'è unicità per quasi ogni ω allora c'è unicità forte.

Teorema 38 *Se* b *soddisfa l'ipotesi (7.3) c'è esistenza forte ed unicità forte, anzi unicità per quasi ogni* ω*.*

Proof. Introduciamo un problema ausiliario:

$$Y_t = X_0 + \int_0^t b\left(s, Y_s + \sigma B_s\right) ds, \qquad t \in \left[0, T\right]. \tag{7.6}$$

Per questo problema definiamo il concetto di soluzione forte nello stesso modo. Se X è soluzione forte di (7.4) allora $Y_t := X_t - \sigma B_t$ è soluzione forte di (7.6). Viceversa, se Y è soluzione forte di (7.6) allora $X_t := Y_t + \sigma B_t$ è soluzione forte di (7.4). Se dimostriamo esistenza e unicità forte per il problema (7.6) allora abbiamo esistenza e unicità forte per il problema (7.4). Infatti, dall'esistenza per (7.6) discende l'esistenza per (7.4); se poi abbiamo due soluzioni di (7.4), queste generano due soluzioni di (7.6), che devono coincidere per unicità, dando l'unicità anche per (7.4). Con lo stesso ragionamento, l'unicità per quasi ogni ω per (7.6) implica quella per (7.4).

Sia $\Omega_0 \in \mathcal{F}$ un insieme di misura uno tale che, per ogni $\omega \in \Omega_0$, $t \mapsto B_t\left(\omega\right)$ sia continua. Consideriamo, per $\omega \in \Omega_0$ fissato, l'equazione deterministica

$$y_t = X_0\left(\omega\right) + \int_0^t b\left(s, y_s + \sigma B_s\left(\omega\right)\right) ds, \qquad t \in \left[0, T\right].$$

Consideriamo la funzione $g_\omega\left(t, z\right) = b\left(t, z + \sigma B_t\left(\omega\right)\right)$. E' una funzione continua e verifica

$$\begin{aligned} \left|g_\omega\left(t, z\right) - g_\omega\left(t, w\right)\right| &= \left|b\left(t, z + \sigma B_t\left(\omega\right)\right) - b\left(t, w + \sigma B_t\left(\omega\right)\right)\right| \\ &\leq L\left|z + \sigma B_t\left(\omega\right) - \left(w + \sigma B_t\left(\omega\right)\right)\right| = L\left|z - w\right| \end{aligned}$$

cioè vale per $g_\omega(t,z)$ l'ipotesi (7.3). Pertanto, per i risultati della sezione precedente, esiste una ed una sola $y \in C\left([0,T];\mathbb{R}^d\right)$ che risolve questa equazione. Indichiamo questa soluzione con $Y_t(\omega)$ per evidenziare la dipendenza dal punto $\omega \in \Omega_0$ fissato a priori.

In questo modo abbiamo definito una funzione $(t,\omega) \mapsto Y_t(\omega)$, da $[0,T] \times \Omega$ in $\mathbb{R}^d$. Essa è un processo adattato: per ogni $t \in [0,T]$, la funzione $\omega \mapsto Y_t(\omega)$ è $\mathcal{F}_t$-misurabile. La verifica dettagliata di questa proprietà è lunga da scrivere ma concettualmente è elementare. Infatti, si ispezioni la dimostrazione di esistenza del paragrafo precedente (specialmente la seconda). Le approssimanti definite per ricorrenza sono processi adattati, per ragioni elementari ed esplicite. Da questo si deduce che il loro limite è un processo adattato.

Il processo $Y_t(\omega)$ è anche continuo e risolve (7.6) uniformemente in t, con probabilità uno. Quindi è soluzione forte. Infine, vale l'unicità per quasi ogni ω, per (7.6) (lo abbiamo appena descritto), per cui c'è anche unicità forte. La dimostrazione è completa. $\blacksquare$

Mostriamo anche una proprietà quantitativa della soluzione.

Proposizione 50 *Sempre sotto l'potesi (7.3), se $E\left[X_0^2\right] < \infty$, allora X è di classe M^2.*

Proof. Che sia progressivamente misurabile, è già noto (continuo e adattato). Dobbiamo verificare che è di quadrato integrabile in entrambe le variabili. Da (7.6) ricaviamo (si applica la disuguaglianza di Hölder)

$$\begin{aligned} |Y_t|^2 &\leq 2|X_0|^2 + 2\left|\int_0^t b(s, Y_s + \sigma B_s)\,ds\right|^2 \\ &\leq 2|X_0|^2 + 2\left(\int_0^t |b(s, Y_s + \sigma B_s)|\,ds\right)^2 \\ &\leq 2|X_0|^2 + 2t\int_0^t |b(s, Y_s + \sigma B_s)|^2\,ds. \end{aligned}$$

Ora osserviamo che l'ipotesi (7.3) implica che esiste $C_0 > 0$ tale che

$$|b(t,x)| \leq C_0(1+|x|).$$

Infatti

$$|b(t,x)| \leq |b(t,x) - b(t,0)| + |b(t,0)| \leq L|x| + \max_{t\in[0,T]}|b(t,0)|$$

per cui basta prendere $C_0 = \max\left(L, \max_{t\in[0,T]}|b(t,0)|\right)$. Ma allora $|b(t,x)|^2 \leq 2C\left(1+|x|^2\right)$ cioè esiste $C > 0$ tale che

$$|b(t,x)|^2 \leq C\left(1+|x|^2\right) \tag{7.7}$$

da cui (indicheremo con C_i varie costanti che semplificano le scritture volta per volta)

$$\begin{aligned}
|Y_t|^2 &\leq 2|X_0|^2 + 4t\int_0^t C^2\left(1+|Y_s+\sigma B_s|^2\right)ds \\
&\leq 2|X_0|^2 + C_1 + C_2\int_0^t |Y_s|^2\,ds + C_3\int_0^t |\sigma B_s|^2\,ds
\end{aligned}$$

con opportune costanti che dipendono da T. Da qui si deduce

$$E\left[|Y_t|^2\right] \leq 2E\left[|X_0|^2\right] + C_1 + C_2\int_0^t E\left[|Y_s|^2\right]ds + C_3\int_0^t E\left[|\sigma B_s|^2\right]ds$$

che implicherebbe tramite il lemma di Gronwall una stima su $E\left[|Y_t|^2\right]$ se già sapessimo che $E\left[|Y_t|^2\right] < \infty$. Ma non lo sappiamo e quindi la disuguaglianza ora scritta potrebbe essere $\infty \leq \infty$. Introduciamo però il tempo di arresto

$$\tau_n = \inf\left\{t \geq 0 : |Y_t|^2 \geq n\right\}$$

(infinito se l'insieme è vuoto) ed osserviamo che

$$\begin{aligned}
|Y_{t\wedge\tau_n}|^2 &\leq 2|X_0|^2 + C_1 + C_2\int_0^{t\wedge\tau_n} |Y_s|^2\,ds + C_3\int_0^{t\wedge\tau_n} |\sigma B_s|^2\,ds \\
&\leq 2|X_0|^2 + C_1 + C_2\int_0^{t\wedge\tau_n} |Y_{s\wedge\tau_n}|^2\,ds + C_3\int_0^{T} |\sigma B_s|^2\,ds \\
&\leq 2|X_0|^2 + C_1 + C_2\int_0^{t} |Y_{s\wedge\tau_n}|^2\,ds + C_3\int_0^{T} |\sigma B_s|^2\,ds.
\end{aligned}$$

Quindi

$$E\left[|Y_{t\wedge\tau_n}|^2\right] \leq 2E\left[|X_0|^2\right] + C_1 + C_2\int_0^t E\left[|Y_{s\wedge\tau_n}|^2\right]ds + C_3\int_0^T E\left[|\sigma B_s|^2\right]ds$$

ed ora $E\left[|Y_{t\wedge\tau_n}|^2\right] \leq n$, quindi possiamo applicare il lemma di Gronwall e trovare

$$E\left[|Y_{t\wedge\tau_n}|^2\right] \leq \left(2E\left[|X_0|^2\right] + C_1 + C_3\int_0^T E\left[|\sigma B_s|^2\right]ds\right)e^{tC_2} \leq C_4.$$

Vale $\lim_{n\to\infty}\tau_n = +\infty$ ed Y è un processo continuo, quindi $\lim_{n\to\infty} Y_{t\wedge\tau_n} = Y_t$ q.c. e quindi, per il lemma di Fatou,

$$E\left[|Y_t|^2\right] = E\left[\lim_{n\to\infty}|Y_{t\wedge\tau_n}|^2\right] \leq \liminf_{n\to\infty} E\left[|Y_{t\wedge\tau_n}|^2\right] \leq C_4.$$

Questo vale per ogni $t \in [0,T]$. Ne discende

$$E\left[|X_t|^2\right] \leq C_5$$

per ogni $t \in [0,T]$ ed in particolare che X è di classe M^2. ■

7.2.1 Esempio di motivazione per il calcolo stocastico

Potrebbe sembrare dal paragrafo precedente che, almeno per equazioni stocastiche molto semplici, il calcolo stocastico (integrale di Itô, formula di Itô) non servano. Per il teorema di esistenza e unicità è così ma basta addentrarsi in altre proprietà per vedere che senza calcolo stocastico non si va avanti. Vediamo un esempio.

Si consideri l'equazione differenziale in $\mathbb{R}^2$

$$\frac{dx_t}{dt} = Ax_t$$

dove

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Posto $x_t = \begin{pmatrix} q_t \\ p_t \end{pmatrix}$, vale

$$\frac{dq_t}{dt} = p_t, \qquad \frac{dp_t}{dt} = -q_t$$

e quindi

$$\frac{d^2 q_t}{dt^2} = -q_t$$

che è l'equazione di un oscillatore armonico di massa unitaria. Le soluzioni x_t ruotano nel piano $\mathbb{R}^2$; in particolare conservano l'energia, se con questo nome intendiamo l'espressione

$$\mathcal{E}_t := \frac{1}{2} |x_t|^2 = \frac{1}{2} \left(q_t^2 + p_t^2 \right)$$

(conservano anche $|x_t|$, la distanza dall'origine). Infatti

$$\frac{d}{dt} |x_t|^2 = 2 \left\langle x_t, \frac{dx_t}{dt} \right\rangle = 2 \langle x_t, Ax_t \rangle = 0$$

ovvero $|x_t|^2$ è costante. Abbiamo usato le notazioni $|\cdot|$ e $\langle \cdot, \cdot \rangle$ per norma e prodotto scalare euclideo e la proprietà $\langle x, Ax \rangle = 0$ valida per matrici antisimmetriche.

Veniamo al caso stocastico. Consideriamo l'equazione

$$dX_t = AX_t dt + \sigma dB_t$$

dove B è un moto browniano unidimensionale e σ è il vettore ($\sigma_0 \in \mathbb{R}$)

$$\sigma = \begin{pmatrix} 0 \\ \sigma_0 \end{pmatrix}.$$

Per componenti, si tratta del sistema

$$\frac{dq_t}{dt} = p_t, \qquad dp_t = -q_t dt + \sigma_0 dB_t$$

o, più informalmente,

$$\frac{d^2 q_t}{dt^2} = -q_t + \sigma_0 \frac{dB_t}{dt}.$$

Chiediamoci: cosa accade ora all'energia? Il termine aggiuntivo $\sigma_0 \frac{dB_t}{dt}$ introduce, sottrae energia oppure è neutrale? Dobbiamo capire come cambia $\mathcal{E}_t = \frac{1}{2} |X_t|^2$. Ma per far questo serve la formula di Itô, ed in essa compaiono automaticamente degli integrali di Itô, anche se non c'erano nell'equazione originaria. Vediamo il calcolo. Vale $(f(x) = |x|^2$, $\nabla f(x) = x$, $D^2 f(x) = Id)$

$$\begin{aligned} d\frac{|X_t|^2}{2} &= \langle X_t, dX_t \rangle + \frac{\sigma_0^2}{2} dt \\ &= \langle X_t, AX_t \rangle dt + \langle X_t, \sigma dB_t \rangle + \frac{\sigma_0^2}{2} dt \\ &= \langle X_t, \sigma dB_t \rangle + \frac{\sigma_0^2}{2} dt \end{aligned}$$

perché $\langle X_t, AX_t \rangle = 0$. Non è chiaro cosa accada all'energia delle singole traiettorie, da questa identità, ma scrivendola in forma integrale

$$\mathcal{E}_t = \mathcal{E}_0 + \int_0^t \langle X_s, \sigma dB_s \rangle + \frac{\sigma_0^2}{2} t$$

ed usando il fatto che il processo X è di classe M^2 (Proposizione 50) possiamo usare il fatto che $E\left[\int_0^t \langle X_s, \sigma dB_s \rangle\right] = 0$ e concludere

$$E[\mathcal{E}_t] = E[\mathcal{E}_0] + \frac{\sigma_0^2}{2} t$$

cioè l'energia media aumenta linearmente nel tempo. Il disturbo casuale introduce sistematicamente energia nel sistema, mediamente parlando.

Osservazione 30 *Si noti che un disturbo deterministico ε_t può anche introdurre energia ma lo fa in modo meno chiaro e controllato. Consideriamo l'equazione deterministica*

$$\frac{dx_t}{dt} = Ax_t + \varepsilon_t.$$

Con calcoli identici ai precedenti si dimostra che

$$\mathcal{E}_t = \mathcal{E}_0 + \int_0^t \langle X_s, \varepsilon_s \rangle ds$$

Non è chiaro a priori se il termine $\int_0^t \langle X_s, \varepsilon_s \rangle ds$ sia positivo o negativo (dipende dalla soluzione, non solo direttamente dal disturbo assegnato) e come cresca nel tempo. Vediamo quindi che il modo di immettere energia in un sistema proprio del moto browniano è assai speciale e conveniente per avere sotto controllo con precisione il bilancio dell'energia del sistema.

7.3 SDE con coefficiente di diffusione non costante

L'acronimo SDE, che useremo ogni tanto per abbreviare, sta per stochastic differential equations. In questa sezione studiamo equazioni della forma

$$X_t = X_0 + \int_0^t b(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s, \qquad t \in [0, T] \tag{7.8}$$

che scriveremo anche nella forma

$$dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t, \qquad X|_{t=0} = X_0.$$

Qui B è un moto browniano in $\mathbb{R}^n$ definito su $(\Omega, \mathcal{F}, P)$ rispetto a $(\mathcal{F}_t)_{t\geq 0}$, X_0 è una v.a. in $\mathbb{R}^d$ misurabile rispetto a $\mathcal{F}_0$, $b : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ e $\sigma : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{n\times d}$ sono funzioni continue e lipschitziane in x uniformemente in t (si veda sotto). Le soluzioni che si cercano sono processi a valori in $\mathbb{R}^d$, continui e adattati, che risolvano l'equazione integrale uniformemente in t, quasi certamente. Le definizioni di esistenza forte ed unicità pathwise sono le stesse del caso con σ costante.

Definizione 23 *Fissata una base stocastica $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P, (B_t)_{t\geq 0}\right)$ ed una v.a. $\mathcal{F}_0$-misurabile X_0, chiamiamo soluzione forte di (7.8) ogni processo stocastico $X = (X_t)_{t\in[0,T]}$ definito su $(\Omega, \mathcal{F}, P)$, continuo, adattato alla filtrazione $(\mathcal{F}_t)_{t\geq 0}$, che soddisfa (7.8) uniformemente in $t \in [0, T]$ con probabilità uno:*

$$P\left(X_t = X_0 + \int_0^t b(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s, \forall t \in [0, T]\right) = 1.$$

Diciamo che c'è esistenza forte quando questo vale a partire da qualsiasi base stocastica e qualsiasi v.a. $\mathcal{F}_0$-misurabile X_0.

Definizione 24 *Diciamo che vale l'unicità forte se, data una qualsiasi base stocastica $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P, (B_t)_{t\geq 0}\right)$ ed una qualsiasi v.a. $\mathcal{F}_0$-misurabile X_0, prese due soluzioni forti X^1 e X^2, esse sono indistinguibili.*

Osservazione 31 *Se X è un processo continuo, allora $s \mapsto b(s, X_s)$ e $s \mapsto \sigma(s, X_s)$ sono processi continui (stiamo sempre supponendo che b e σ siano funzioni continue) e quindi di classe Λ^1 e Λ^2 rispettivamente; pertanto gli integrali che compaiono nell'equazione hanno senso.*

Osservazione 32 *Non possiamo dare la definizione di unicità per quasi ogni ω, invece, perché l'integrale stocastico $\int_0^t \sigma(s, X_s)\, B_s$ non ha significato puntuale in ω.*

Su b e σ imponiamo:

$$\begin{aligned} b \text{ continua, esiste } L &> 0 \text{ tale che} \\ \left|b(t,x) - b(t,y)\right| &\leq L\left|x-y\right| \text{ per ogni } t \in [0,T] \text{ e } x,y \in \mathbb{R}^d \end{aligned}$$

$$\begin{aligned} \sigma \text{ continua, esiste } L &> 0 \text{ tale che} \\ \left\|\sigma(t,x) - \sigma(t,y)\right\|_{HS} &\leq L\left|x-y\right| \text{ per ogni } t \in [0,T] \text{ e } x,y \in \mathbb{R}^d \end{aligned}$$

dove ricordiamo che $\|A\|_{HS}^2 = \sum_{ij} A_{ij}^2$. Osserviamo che esiste $C > 0$ tale che (si veda (7.7))

$$\left|b(t,x)\right|^2 \leq C\left(1 + |x|^2\right)$$

ed analogamente (eventualmente rinominando C)

$$\left\|\sigma(t,x)\right\|_{HS}^2 \leq C\left(1 + |x|^2\right)$$

in quanto

$$\begin{aligned} \left\|\sigma(t,x)\right\|_{HS}^2 &\leq 2\left\|\sigma(t,x) - \sigma(t,0)\right\|_{HS}^2 + 2\left\|\sigma(t,0)\right\|_{HS}^2 \\ &\leq 2L^2 |x|^2 + 2 \max_{t \in [0,T]} \left\|\sigma(t,0)\right\|_{HS}^2 . \end{aligned}$$

Teorema 39 *Sotto queste ipotesi, se $E\left[|X_0|^2\right] < \infty$, allora c'è esistenza forte per l'equazione (7.8), di una soluzione che verifica anche*

$$E\left[\sup_{t \in [0,T]} |X_t|^2\right] < \infty.$$

Nella classe dei processi che verificano anche questa condizione quantitativa (basta $E\left[\int_0^T |X_s|^2 ds\right] < \infty$) c'è anche unicità forte.

Osservazione 33 *Con opportuni argomenti (tempi di arresto ecc.) si può omettere la condizione $E\left[|X_0|^2\right] < \infty$, rinunciando così alla proprietà $E\left[\sup_{t \in [0,T]} |X_t|^2\right] < \infty$, ed avere tuttavia ancora esistenza ed unicità forte. Omettiamo i dettagli.*

Proof. Passo 1. Vediamo l'unicità pathwise. Data una base stocastica $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P, (B_t)_{t \geq 0}\right)$, una v.a. $\mathcal{F}_0$-misurabile X_0 con la proprietà $E\left[|X_0|^2\right] < \infty$, e due soluzioni X^1 e X^2 tali che $E\left[\sup_{t \in [0,T]} |X_t^i|^2\right] < \infty$, $i = 1,2$, dobbiamo dimostrare che esse sono indistinguibili. Vale

$$X_t^1 - X_t^2 = \int_0^t \left(b\left(s, X_s^1\right) - b\left(s, X_s^2\right)\right) ds + \int_0^t \left(\sigma\left(s, X_s^1\right) - \sigma\left(s, X_s^2\right)\right) dB_s$$

da cui

$$\left|X_t^1 - X_t^2\right|^2 \leq 2\left|\int_0^t b\left(s, X_s^1\right) - b\left(s, X_s^2\right) ds\right|^2 + 2\left|\int_0^t \left(\sigma\left(s, X_s^1\right) - \sigma\left(s, X_s^2\right)\right) dB_s\right|^2.$$

Da un lato, vale

$$\begin{aligned}\left|\int_0^t b\left(s, X_s^1\right) - b\left(s, X_s^2\right) ds\right|^2 &\leq \left(\int_0^t \left|b\left(s, X_s^1\right) - b\left(s, X_s^2\right)\right| ds\right)^2 \\ &\leq t\int_0^t \left|b\left(s, X_s^1\right) - b\left(s, X_s^2\right)\right|^2 ds\end{aligned}$$

usando la disuguaglianza di Hölder. Dall'altro, osservando che $s \mapsto \left(\sigma\left(s, X_s^1\right) - \sigma\left(s, X_s^2\right)\right)$ è un processo di classe M^2 (spiegato poco oltre), per la formula di isometria vettoriale, Proposizione 40, abbiamo

$$E\left[\left|\int_0^t \left(\sigma\left(s, X_s^1\right) - \sigma\left(s, X_s^2\right)\right) dB_s\right|^2\right] = E\left[\int_0^t \left\|\sigma\left(s, X_s^1\right) - \sigma\left(s, X_s^2\right)\right\|_{HS}^2 ds\right].$$

Pertanto,

$$\left|X_t^1 - X_t^2\right|^2 \leq 2T\int_0^t \left|b\left(s, X_s^1\right) - b\left(s, X_s^2\right)\right|^2 ds + 2\left|\int_0^t \left(\sigma\left(s, X_s^1\right) - \sigma\left(s, X_s^2\right)\right) dB_s\right|^2$$

$$\begin{aligned}E\left[\left|X_t^1 - X_t^2\right|^2\right] &\leq 2T\int_0^t E\left[\left|b\left(s, X_s^1\right) - b\left(s, X_s^2\right)\right|^2\right] ds + 2E\left[\int_0^t \left\|\sigma\left(s, X_s^1\right) - \sigma\left(s, X_s^2\right)\right\|_{HS}^2 ds\right] \\ &\leq 2TL^2\int_0^t E\left[\left|X_s^1 - X_s^2\right|^2\right] ds + 2L^2\int_0^t E\left[\left|X_s^1 - X_s^2\right|^2\right] ds \\ &= \left(2TL^2 + 2L^2\right)\int_0^t E\left[\left|X_s^1 - X_s^2\right|^2\right] ds\end{aligned}$$

e quindi, per il lemma di Gronwall applicato alla funzione $t \mapsto E\left[\left|X_t^1 - X_t^2\right|^2\right]$, abbiamo

$$E\left[\left|X_t^1 - X_t^2\right|^2\right] = 0$$

per ogni $t \in [0, T]$. Da questo discende che $X_t^1 = X_t^2$ q.c., per ogni $t \in [0, T]$, ovvero che i due processi sono modificazione uno dell'altro; ma essendo continui, sono indistinguibili.

Resta da verificare che $s \mapsto \sigma\left(s, X_s^i\right)$, $i = 1, 2$, è un processo di classe M^2. Questo deriva dalle disuguaglianze (ricordando che $\|\sigma(t, x)\|_{HS}^2 \leq C\left(1 + |x|^2\right)$)

$$E\left[\int_0^T \left\|\sigma\left(s, X_s^i\right)\right\|_{HS}^2 ds\right] \leq E\left[\int_0^T \left(C + \left|X_s^i\right|^2\right) ds\right] = CT + E\left[\int_0^T \left|X_s^i\right|^2 ds\right] < \infty$$

per ipotesi.

Passo 2. Per dimostrare l'esistenza, usiamo il metodo delle approssimazioni successive (si potrebbe fare benissimo col principio delle contrazioni). Poniamo $X_t^0 = X_0$, $t \in [0,T]$ e

$$X_t^{n+1} = X_0 + \int_0^t b(s, X_s^n)\, ds + \int_0^t \sigma(s, X_s^n)\, dB_s, \qquad t \in [0,T]$$

per ogni $n \geq 0$.

Vediamo innanzi tutto che le iterazioni siano ben definite e che processi definiscano. Supponiamo di sapere che X^n è un processo continuo che soddisfa $E\left[\sup_{t\in[0,T]} |X_t^n|^2\right] < \infty$. Allora $b(s, X_s^n)$ e $\sigma(s, X_s^n)$ sono processi continui e adattati, per cui, per lo meno, $\int_0^t b(s, X_s^n)\, ds$ e $\int_0^t \sigma(s, X_s^n)\, B_s$ sono processi ben definiti, continui e adattati. Ma, più precisamente,

$$|b(s, X_s^n)|^2 \leq C\left(1 + |X_s^n|^2\right)$$

quindi

$$\left|\int_0^t b(s, X_s^n)\, ds\right|^2 \leq t \int_0^t |b(s, X_s^n)|^2\, ds \leq TC \int_0^t \left(1 + |X_s^n|^2\right) ds$$

da cui discende che

$$E\left[\sup_{t\in[0,T]} \left|\int_0^t b(s, X_s^n)\, ds\right|^2\right] < \infty.$$

Inoltre vale

$$\|\sigma(s, X_s^n)\|_{HS}^2 \leq C\left(1 + |X_s^n|^2\right)$$

per cui $\sigma(s, X_s^n)$ è un processo (matriciale) di classe M^2, $\int_0^t \sigma(s, X_s^n)\, B_s$ è una martingala (vettoriale) di quadrato integrabile, e vale (disuguaglianza di Doob vettoriale, Proposizione 41)

$$E\left[\sup_{t\in[0,T]} \left|\int_0^t \sigma(s, X_s^n)\, B_s\right|^2\right] \leq 4E\left[\int_0^T \|\sigma(s, X_s^n)\|_{HS}^2\, ds\right].$$

Abbiamo verificato che la proprietà $E\left[\sup_{t\in[0,T]} |X_t^n|^2\right] < \infty$ si trasferisce da n a $n+1$ e siccome vale per $n = 0$, vale per ogni n.

Passo 3. Discutiamo ora la convergenza della successione X^n. Lo facciamo in due modi, in analogia col caso deterministico. Il primo è basato su una stima del tipo

$$E\left[\sup_{t\in[0,\tau]} \left|X_t^{n+1} - X_t^n\right|^2\right] \leq \alpha_\tau^n E\left[\sup_{t\in[0,\tau]} \left|X_t^1 - X_t^0\right|^2\right] \tag{7.9}$$

dove $\alpha_\tau < 1$ se τ è abbastanza piccolo. Da qui si deduce l'esistenza di una soluzione su $[0,\tau]$ e poi si deve iterare il procedimento su intervallini successivi di eguale ampiezza. Vediamo alcuni dettagli. Il secondo modo verrà illustrato al passo successivo.

Vale, come sopra, per $t \in [0, \tau]$ con $\tau \in [0, T]$ fissato,

$$\left|X_t^{n+1} - X_t^n\right|^2 \leq 2\tau \int_0^t \left|b\left(s, X_s^n\right) - b\left(s, X_s^{n-1}\right)\right|^2 ds + 2\left|\int_0^t \left(\sigma\left(s, X_s^n\right) - \sigma\left(s, X_s^{n-1}\right)\right) dB_s\right|^2$$

ed inoltre

$$\begin{aligned} 2\tau \int_0^t \left|b\left(s, X_s^n\right) - b\left(s, X_s^{n-1}\right)\right|^2 ds &\leq 2\tau L^2 \int_0^t \left|X_s^n - X_s^{n-1}\right|^2 ds \\ &\leq 2\tau^2 L^2 \sup_{s\in[0,\tau]} \left(\left|X_s^n - X_s^{n-1}\right|^2\right). \end{aligned}$$

Quindi

$$\begin{aligned} E\left[\sup_{t\in[0,\tau]} \left|X_t^{n+1} - X_t^n\right|^2\right] &\leq 2\tau^2 L^2 E\left[\sup_{s\in[0,\tau]} \left(\left|X_s^n - X_s^{n-1}\right|^2\right)\right] \\ &\quad + 2E\left[\sup_{t\in[0,\tau]} \left|\int_0^t \left(\sigma\left(s, X_s^n\right) - \sigma\left(s, X_s^{n-1}\right)\right) dB_s\right|^2\right]. \end{aligned}$$

Il secondo termine a destra si maggiora, tramite la disuguaglianza di Doob vettoriale, Proposizione 41, con

$$\begin{aligned} &2E\left[\sup_{t\in[0,\tau]} \left|\int_0^t \left(\sigma\left(s, X_s^n\right) - \sigma\left(s, X_s^{n-1}\right)\right) dB_s\right|^2\right] \\ \leq\ & 8E\left[\int_0^\tau \left\|\sigma\left(s, X_s^n\right) - \sigma\left(s, X_s^{n-1}\right)\right\|_{HS}^2 ds\right] \\ \leq\ & 8L^2 E\left[\int_0^\tau \left|X_s^n - X_s^{n-1}\right|^2 ds\right] \leq 8L^2 \tau E\left[\sup_{s\in[0,\tau]} \left(\left|X_s^n - X_s^{n-1}\right|^2\right)\right]. \end{aligned}$$

In definitiva,

$$E\left[\sup_{t\in[0,\tau]} \left|X_t^{n+1} - X_t^n\right|^2\right] \leq \left(2\tau^2 L^2 + 8L^2\tau\right) E\left[\sup_{s\in[0,\tau]} \left(\left|X_s^n - X_s^{n-1}\right|^2\right)\right].$$

Se prendiamo $\tau > 0$ così piccolo che sia $\alpha_\tau := 2\tau^2 L^2 + 8L^2\tau < 1$, abbiamo (7.9).

Da (7.9) si può dedurre che esiste un processo continuo adattato $(X_t)_{t\in[0,\tau]}$, con $E\left[\sup_{t\in[0,\tau]} |X_t|^2\right] < \infty$, tale che

$$\lim_{n\to\infty} E\left[\sup_{t\in[0,\tau]} \left|X_t^n - X_t\right|^2\right] = 0.$$

O lo si deduce usando lo spazio di Banach $L^2\left(\Omega; C\left([0,\tau]; \mathbb{R}^d\right)\right)$, in cui $\{X^n\}$ è una successione di Cauchy grazie a (7.9), oppure si sviluppa una serie di ragionamenti più elementari ma lunghi (ad esempio si ragiona prima a t fissato, usando la completezza dello spazio $L^2\left(\Omega; \mathbb{R}^d\right)$, e poi si riuniscono i tempi a posteriori).

Dalla convergenza delle X_t^n a X_t si deduce facilmente che X è soluzione, applicando calcoli simili a quelli appena visti, nella versione più facile del passo 1: per ogni $t \in [0,\tau]$

$$\begin{aligned}
E\left[\left|\int_0^t b(s, X_s^n)\,ds - \int_0^t b(s, X_s)\,ds\right|^2\right] &\leq \tau E\left[\int_0^t \left|b(s, X_s^n) - b(s, X_s)\right|^2 ds\right] \\
&\leq \tau L^2 E\left[\int_0^t \left|X_s^n - X_s\right|^2 ds\right] \to 0
\end{aligned}$$

$$\begin{aligned}
E\left[\left|\int_0^t \sigma(s, X_s^n)\,dB_s - \int_0^t \sigma(s, X_s)\,dB_s\right|^2\right] &= E\left[\int_0^t \left\|\sigma(s, X_s^n) - \sigma(s, X_s)\right\|_{HS}^2 ds\right] \\
&\leq L^2 E\left[\int_0^t \left|X_s^n - X_s\right|^2 ds\right] \to 0.
\end{aligned}$$

Bisogna infine ripartire dal tempo τ, con la condizione iniziale X_τ, ovvero il valore della soluzione appena trovata, al tempo τ. Le stime non cambiano per cui si può ripetere il ragionamento sull'intervallo $[\tau, 2\tau]$. In un numero finito di passi si ricopre $[0,T]$.

Passo 4. Se si vuole lavorare in un passo solo si può usare il trucco, illustrato nel caso deterministico, della penalizzazione esponenziale. Purtroppo esso impedisce di usare le disuguaglianze delle martingale, per cui le stime che inizialmente si dimostrano sono più deboli di (7.9):

$$\sup_{t\in[0,T]} E\left[e^{-2\lambda t}\left|X_t^{n+1} - X_t^n\right|^2\right] \leq \alpha_\lambda^n \sup_{t\in[0,T]} E\left[e^{-2\lambda t}\left|X_t^1 - X_t^0\right|^2\right] \qquad (7.10)$$

dove $\alpha_\lambda < 1$ se $\lambda > 0$ è abbastanza grande. Esse permettono comunque di trovare un processo $(X_t)_{t\in[0,T]}$, con $\sup_{t\in[0,T]} E\left[|X_t|^2\right] < \infty$, tale che

$$\lim_{n\to\infty} \sup_{t\in[0,T]} E\left[|X_t^n - X_t|^2\right] = 0.$$

Questo è sufficiente per passare al limite (con gli stessi conti del passo precedente) per cui X_t soddisfa l'equazione integrale (7.8). Però la soddisfa nel seguente senso: per ogni $t \in [0,T]$, l'identità in (7.8) vale quasi certamente (non quindi come identità uniforme in t). Da essa si deduce, a posteriori, che X ha una modificazione continua $\widetilde{X}$, perché il termine di destra dell'equazione è un processo continuo. Negli integrali si può a questo punto sostituire X con $\widetilde{X}$ e quindi $\widetilde{X}$ è soluzione continua e tradizionale dell'equazione (7.8).

Dimostriamo (7.10). Ripartiamo da

$$\left|X_t^{n+1}-X_t^n\right|^2 \leq 2T\int_0^t \left|b\left(s,X_s^n\right)-b\left(s,X_s^{n-1}\right)\right|^2 ds+2\left|\int_0^t \left(\sigma\left(s,X_s^n\right)-\sigma\left(s,X_s^{n-1}\right)\right)dB_s\right|^2$$

moltiplichiamo per $e^{-2\lambda t}$:

$$\begin{aligned} e^{-2\lambda t}\left|X_t^{n+1}-X_t^n\right|^2 &\leq 2T\int_0^t e^{-2\lambda(t-s)}e^{-2\lambda s}\left|b\left(s,X_s^n\right)-b\left(s,X_s^{n-1}\right)\right|^2 ds \\ &\quad +2\left|\int_0^t e^{-\lambda(t-s)}e^{-\lambda s}\left(\sigma\left(s,X_s^n\right)-\sigma\left(s,X_s^{n-1}\right)\right)dB_s\right|^2 \end{aligned}$$

e prendiamo il valore atteso (senza *sup*):

$$\begin{aligned} E\left[e^{-2\lambda t}\left|X_t^{n+1}-X_t^n\right|^2\right] &\leq 2T\int_0^t e^{-2\lambda(t-s)}E\left[e^{-2\lambda s}\left|b\left(s,X_s^n\right)-b\left(s,X_s^{n-1}\right)\right|^2\right] ds \\ &\quad +2E\left[\int_0^t e^{-2\lambda(t-s)}e^{-2\lambda s}\left\|\sigma\left(s,X_s^n\right)-\sigma\left(s,X_s^{n-1}\right)\right\|_{HS}^2 ds\right] \end{aligned}$$

$$\begin{aligned} &\leq \left(2TL^2+2L^2\right)\int_0^t e^{-2\lambda(t-s)}E\left[e^{-2\lambda s}\left|X_s^n-X_s^{n-1}\right|^2\right] ds \\ &\leq \left(2TL^2+2L^2\right)\sup_{s\in[0,T]}E\left[e^{-2\lambda s}\left|X_s^n-X_s^{n-1}\right|^2\right]\int_0^t e^{-2\lambda(t-s)}ds \\ &\leq \frac{TL^2+L^2}{\lambda}\sup_{s\in[0,T]}E\left[e^{-2\lambda s}\left|X_s^n-X_s^{n-1}\right|^2\right]. \end{aligned}$$

Quindi

$$\sup_{t\in[0,T]}E\left[e^{-2\lambda t}\left|X_t^{n+1}-X_t^n\right|^2\right] \leq \alpha_\lambda \sup_{s\in[0,T]}E\left[e^{-2\lambda s}\left|X_s^n-X_s^{n-1}\right|^2\right]$$

con $\alpha_\lambda = \frac{TL^2+L^2}{\lambda}$. Questo implica la (7.10) e la dimostrazione è completa. ■

Facendo un uso più accurato delle disuguaglianze per le martingale si può studiare il caso in cui $E\left[|X_0|^p\right] < \infty$ per un $p \geq 1$ qualsiasi; inoltre, riconducendosi al caso precedente tramite tempi d'arresto, si può studiare il caso in cui non si abbiano ipotesi di integrabilità su X_0. Preferiamo tralasciare tutti questi dettagli e rinunciare ad un po' di generalità. Il risultato comunque è il seguente.

Teorema 40 *Se X_0 è misurabile rispetto a $\mathcal{F}_0$, b e σ soddisfano le ipotesi precedenti, allora c'è esistenza forte e unicità pathwise per l'equazione (7.8). Se, per un certo $p \geq 1$, vale $E\left[|X_0|^p\right] < \infty$, allora*

$$E\left[\sup_{t\in[0,T]}|X_t|^p\right] < \infty.$$

7.4 Presentazione informale del legame con le PDE

L'equazione differenziale stocastica (SDE)

$$dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t$$

si lega a varie equazioni differenziali alle derivate parziali. Per semplicità di esposizione, limitiamoci al caso $\sigma(t, X_t) = Id$, quindi all'equazione con noise additivo

$$dX_t = b(t, X_t)\, dt + dB_t \tag{7.11}$$

dove supponiamo che B sia un moto browniano in $\mathbb{R}^d$, $b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ sia sufficientemente regolare (se si vuole fissare un'ipotesi unica che vada bene per tutti i teoremi, si prenda ad esempio per semplicità C^∞ con derivate limitate; ma caso per caso basta molto meno) e la soluzione sia un processo continuo adattato in $\mathbb{R}^d$.

Oltre ad alcune altre varianti, le tre equazioni di base a cui leghiamo la SDE sono:

1. l'equazione parabolica, scritta in forma retrograda nel tempo, detta *equazione di Kolmogorov*:
$$\frac{\partial u}{\partial t} + \frac{1}{2}\Delta u + b \cdot \nabla u = 0 \qquad \text{su } [0,T] \times \mathbb{R}^d$$
$$u|_{t=T} = u_T$$
dove la soluzione $u = u(t,x)$ è una funzione $u : [0,T] \times \mathbb{R}^d \to \mathbb{R}$,
$$\Delta u = \sum_{i=1}^d \frac{\partial^2 u}{\partial x_i^2}, \qquad b \cdot \nabla u = \sum_{i=1}^d b_i \frac{\partial u}{\partial x_i}$$
e si impone la condizione al tempo finale T, $u(T,x) = u_T(x)$;

2. l'equazione parabolica (in forma di equazione di continuità con viscosità), detta *equazione di Fokker-Planck*:
$$\frac{\partial p}{\partial t} = \frac{1}{2}\Delta p + \operatorname{div}(bp) \qquad \text{su } [0,T] \times \mathbb{R}^d$$
$$p|_{t=0} = p_0$$
dove la soluzione $p = pu(t,x)$ è una funzione $u : [0,T] \times \mathbb{R}^d \to \mathbb{R}^+$, in genere una densità di probabilità (ad ogni istante t fissato), e
$$\operatorname{div}(bp) = \sum_{i=1}^d \frac{\partial}{\partial x_i}(b_i p)$$
e si impone la condizione al tempo iniziale, $p(0,x) = p_0(x)$, anch'essa densità di probabilità;

3. il *problema di Dirichlet*

$$\frac{1}{2}\Delta u + b \cdot \nabla u = 0 \qquad \text{su } D$$

$$u|_{\partial D} = g$$

dove $D \subset \mathbb{R}^d$ è un insieme limitato regolare, e la funzione g assegnata al bordo è pure sufficientemente regolare.

I legami sono i seguenti. Fissato un qualsiasi $t_0 \in [0, T]$, risolviamo l'equazione (7.11) sull'intervallo $[t_0, T]$ con dato iniziale $x \in \mathbb{R}^d$ al tempo t_0:

$$\begin{aligned} dX_t &= b(t, X_t)\,dt + dB_t \qquad \text{per } t \in [t_0, T] \\ X_{t_0} &= x \end{aligned}$$

ovvero in forma integrale

$$X_t = x + \int_{t_0}^{t} b(s, X_s)\,ds + B_t \qquad \text{per } t \in [t_0, T].$$

Indichiamo con $X(t; t_0, x)$ la soluzione forte, unica. Allora vale:

1. per l'equazione di Kolmogorov:

$$u(t, x) = E[u_T(X(T; t, x))] \tag{7.12}$$

per ogni $(t, x) \in [0, T] \times \mathbb{R}^d$

2. per l'equazione di Fokker-Planck: se si prende la soluzione X_t dell'equazione (7.11) con dato iniziale aleatorio X_0 avente densità di probabilità $p_0(x)$, allora ad ogni istante $t \in [0, T]$ la v.a. X_t ha densità di probabilità, data da $p(t, x)$, soluzione dell'equazione di Fokker-Planck

3. per il problema di Dirichlet: detto τ_D il tempo di uscita da D (primo istante di ingresso in D^c), vale

$$u(x) = E[g(X_{\tau_D})].$$

Osservazione 34 *L'equazione di Fokker-Planck ha notevole utilità pratica nelle applicazioni perché offre un modo di calcolare (magari anche solo numericamente) la densità di probabilità della soluzione X_t, da cui si possono calcolare valori attesi e probabilità di interesse applicativo. Inoltre, è di un certo interesse fisico il legame tra le due equazioni (per questo se ne sono interessati Fokker e Planck).*

Osservazione 35 *I legami invece tra l'equazione (7.11) e le altre due equazioni alle derivate parziali vengono di solito usati per scopi interni alla matematica, o per studiare le SDE tramite risultati e metodi della teoria delle PDE (equazioni alle derivate parziali), o viceversa.*

Osservazione 36 *Circa il legame con l'equazione di Kolmogorov, si può notare che esso è una variante probabilistica del noto legame tra PDE del prim'ordine di tipo trasporto e le loro curve caratteristiche associate. Infatti, è noto che se si vuole studiare la PDE*

$$\frac{\partial u}{\partial t} + b \cdot \nabla u = 0 \qquad su \;\; [0,T] \times \mathbb{R}^d$$

$$u|_{t=T} = u_T$$

si può dare una formula esplicita della soluzione in termini di caratteristiche, cioè risolvendo l'equazione ordinaria associata

$$\begin{aligned} \frac{dX_t}{dt} &= b(t, X_t)\, dt \qquad per\ t \in [t_0, T] \\ X_{t_0} &= x \end{aligned}$$

ed il legame è (usando la notazione $X(t; t_0, x)$ come sopra)

$$u(t,x) = u_T(X(T; t, x)).$$

Concludiamo con un cenno alla dimostrazione della prima formula, quella per l'equazione di Kolmogorov. Se volessimo partire dalla (7.11) e mostrare che la funzione $u(t,x)$ definita dalla formula (7.12) è soluzione dell'equazione di Kolmogorov, dovremmo svolgere molti calcoli (es. dovremmo derivare due volte $X(t; t_0, x)$ rispetto ad x). Non esponiamo questo tipo di calcoli. Invece, è moto semplice mostrare il seguente risultato.

Proposizione 51 *Se u è una soluzione di classe $C^{1,2}$ dell'equazione di Kolmogorov, con derivate $\frac{\partial u}{\partial x_i}$ limitate, allora vale (7.12).*

Proof. Per la regolarità di u possiamo applicare la formula di Itô:

$$du(t, X(t; t_0, x)) = \frac{\partial u}{\partial t} dt + \nabla u \cdot (dX) + \frac{1}{2} \sum_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} d\left[X^i, X^j\right]$$

ma $d[X^i, X^j]_t = d[B^i, B^j]_t = \delta_{ij} dt$, quindi

$$du(t, X(t; t_0, x)) = \frac{\partial u}{\partial t} dt + \nabla u \cdot b dt + \nabla u \cdot dB + \frac{1}{2} \Delta u dt.$$

Siccome vale $\frac{\partial u}{\partial t} + \frac{1}{2}\Delta u + b \cdot \nabla u$, rimane

$$du\left(t, X\left(t; t_0, x\right)\right) = \nabla u\left(t, X\left(t; t_0, x\right)\right) \cdot dB_t$$

ovvero, scrivendo in forma integrale su $[t_0, T]$,

$$u\left(T, X\left(T; t_0, x\right)\right) = u\left(t_0, X\left(t_0; t_0, x\right)\right) + \int_{t_0}^{T} \nabla u\left(s, X\left(s; t_0, x\right)\right) \cdot dB_s.$$

Per l'ipotesi di limitatezza delle derivate di u, vale $E\left[\int_{t_0}^{T} \nabla u\left(s, X\left(s; t_0, x\right)\right) \cdot dB_s\right] = 0$. Inoltre, $X\left(t_0; t_0, x\right) = x$, $u\left(T, X\left(T; t_0, x\right)\right) = u_T\left(X\left(T; t_0, x\right)\right)$, quindi

$$E\left[u_T\left(X\left(T; t_0, x\right)\right)\right] = u\left(t_0, x\right).$$

La dimostrazione è completa. $\blacksquare$