Università degli Studi di Roma "La Sapienza"
Anno Accademico 2007-2008

Facoltà di Scienze Matematiche Fisiche e Naturali
Dottorato in Matematica

Alcuni appunti per il corso di

METODI PROBABILISTICI PER LE EQUAZIONI ALLE DERIVATE PARZIALI

Giovanna Nappo

A.A. 2008/2009

versione del 9.04.2009

Indice

Introduzione **iii**

1 Richiami su spazi di probabilità **1**
1.1 Esempi di spazi di probabilità . . . 1
1.2 Variabili aleatorie . . . 4
1.3 Distribuzioni di variabili aleatorie . . . 6
1.4 Valori attesi . . . 10
1.4.1 Variabili aleatorie in spazi misurabili ** . . . 12
1.5 Misura indotta e Cambio di variabile ** . . . 12
1.6 Appendice: variabili Gaussiane . . . 16
1.7 Appendice: Teorema di rappresentazione di Skorohod . . . 20
1.8 Appendice: Spazi di variabili aleatorie . . . 23
1.8.1 Sottospazi dello spazio delle v.a.di quadrato integrabile † . . . 24
1.8.2 Regressione lineare . . . 24
1.9 Appendice: Approccio soggettivista alla probabilità . . . 25

2 Valori attesi e probabilità condizionali **28**
2.1 Definizioni . . . 28
2.2 Esempi . . . 30
2.3 Proprietà del valore atteso condizionale . . . 36
2.4 Equivalenza tra le definizioni di valore atteso condizionale per variabili aleatorie di quadrato sommabile 39
2.5 Dimostrazioni delle proprietà del valore atteso condizionale . . . 40
2.6 Probabilità condizionali regolari . . . 43
2.7 Esempi . . . 45

3 Martingale **48**
3.1 Esempi di martingale e di submartingale . . . 49
3.2 Decomposizione di Doob . . . 55
3.2.1 Applicazioni: variazione quadratica e integrale stocastico a tempo discreto . . . 56
3.2.2 Applicazioni: verso l'integrale stocastico a tempo continuo . . . 59
3.3 Martingale, submartingale e tempi d'arresto . . . 65
3.3.1 Tempo discreto . . . 65
3.3.2 Tempo continuo . . . 66
3.4 Alcune proprietà dei tempi d'arresto . . . 67
3.5 Caso a tempo discreto . . . 68
3.6 Applicazione: la rovina del giocatore con le martingale . . . 70
3.7 Disuguaglianza di Kolmogorov per submartingale non negative . . . 72
3.8 Convergenza di martingale . . . 72
3.9 Disuguaglianza di Doob . . . 74
3.10 Estensioni alle martingale a tempo continuo (cenni) . . . 75
3.11 Processi misurabili e tempi d'arresto . . . 76

4 Processi aleatori a tempo continuo **78**
4.1 Processi aleatori, definizioni ed esempi . . . 78
4.2 Osservazione sulla definizione di un processo solo attraverso le sue distribuzioni finito dimensionali . . 82
4.3 Esistenza di una versione continua: criterio di Chensov-Kolmogorov. . . . 84
4.4 Le traiettorie del processo di Wiener non sono a variazione limitata . . . 84
4.5 Costruzione del processo di Wiener e della misura di Wiener su $C[0,1]$. ** . . . 88
4.5.1 Costruzione come serie di funzioni . . . 88

4.5.2 Costruzione come legge limite, sullo spazio delle funzioni continue ** . . . 93
4.6 Processi ad incrementi indipendenti ed omogenei . . . 95
4.7 Esempi di martingale a tempo continuo . . . 97
4.8 Processi di Markov regolari . . . 101
4.8.1 Processo di Orstein-Ulhenbeck . . . 104
4.8.2 Moto browniano geometrico e modello di Black-Scholes . . . 106
4.9 Appendice: dimostrazione del Teorema di esistenza di Kolmogorov . . . 110
4.9.1 Caso a tempo discreto: metodo diretto . . . 115
4.9.2 Osservazione su $\mathcal{R}^I$ † . . . 117
4.9.3 Problemi con lo spazio canonico . . . 117
4.10 Appendice: dimostrazione del criterio di Chensov-Kolmogorov . . . 119

5 Proprietà del moto browniano 121
5.1 Trasformazioni del moto browniano . . . 121
5.2 Proprietà di Markov forte per il processo di Wiener . . . 121
5.3 Principio di riflessione . . . 123
5.3.1 Trasformata di Laplace . . . 126
5.4 Tempi di uscita da una striscia . . . 127
5.5 Proprietà delle traiettorie di un processo di Wiener **** . . . 130
5.6 Calcolo del risolvente del nucleo del calore **** . . . 132

6 Integrale stocastico: cenni 134
6.1 Integrale stocastico per processi integrabili, a partire dai processi elementari . . . 134
6.1.1 Esempi di calcolo di integrali stocastici . . . 138
6.1.2 Integrale stocastico per processi localmente di quadrato integrabile . . . 141
6.1.3 Osservazioni su alcune proprietà degli integrali stocastici . . . 142
6.2 Calcolo stocastico e formula di Itô . . . 143
6.2.1 Moto browniano geometrico e il suo differenziale stocastico . . . 147
6.2.2 Funzioni armoniche e tempi di uscita per il processo di Wiener . . . 149
6.2.3 Applicazione ai tempi di uscita da una striscia per il processo di Wiener con drift . . . 149
6.2.4 Applicazione: ricorrenza e transienza per il processo di Wiener . . . 153
6.3 Equazioni differenziali stocastiche . . . 154
6.3.1 Unicità per EDS: il caso Lipschitz globale . . . 154
6.3.2 Esempi di soluzioni di equazioni differenziali stocastiche: il processo di Orstein-Ulhenbeck . . . 156
6.3.3 Esistenza per EDS: il caso Lipschitz globale . . . 158
6.3.4 Dipendenza continua dal dato iniziale e proprietà di Markov per soluzioni di EDS . . . 163
6.3.5 Tempi di uscita da una striscia per una soluzione di una EDS . . . 166
6.3.6 Soluzioni deboli di equazioni differenziali stocastiche . . . 168
6.3.7 Trasformazione di Girsanov e soluzioni deboli . . . 169

7 Rappresentazione di soluzioni di equazioni alle derivate parziali 172
7.1 Rappresentazione di Feynman-Kac per il processo di Wiener . . . 172
7.1.1 Applicazione: Formula dell'arcoseno (di Lévy) . . . 173
7.1.2 Generalizzazioni . . . 174
7.1.3 Connessioni con il problema di Dirichlet . . . 176
7.1.4 Il processo di Bessel . . . 177

Bibliografia 182

Introduzione

**(DA RIVEDERE e arricchire) **

Lo scopo del corso è quello di introdurre i concetti basilari della teoria dei processi stocastici e del calcolo stocastico per capire il legame tra questi e alcune equazioni alle derivate parziali.

Questi appunti si basano su una raccolta di vari appunti, scritti durante vari anni di insegnamento di corsi universitari sui processi aleatori. E questo spiega la non completa coerenza della presentazione degli argomenti.

Inoltre il testo contiene alcune note che negli ultimi anni non sono state utilizzate e quindi non sono state riguardate: il corso di dottorato mi dà l'opportunità di riparare a questa mancanza.

Nella preparazione delle lezioni ho utilizzato diversi testi e i principali sono il testo di P. Baldi [1] e quello di A. N. Shiryaev [14]. Ma andrebbero citati anche altri testi, come ad esempio il testo di L. Breiman [5], quello di I. Karatzas e S. E. Shreve [7], quello di D. Revuz e M. Yor [11], i testi di S. Ross [12] e [13], il testo di P. Billingsley [3], quello di J. M. Steele [15], e altri ancora.

Infine va consigliato agli studenti la lettura delle note del Prof. Lorenzo Bertini [2], preparate per questo stesso corso negli anni accademici precedenti (le note sono reperibile al suo sito Web).

Infine gli studenti potranno scegliere tra alcuni argomenti,non completamente svolti a lezione, e che potranno esporre in forma di seminario:

Teorema di rappresentazione di Kolmogorov, (queste note, Billingsley [3])
Teorema di Chensov-Kolmogorov (Baldi [1] o queste note e/o per una dimostrazione differente le note del Prof. Bertini [2], che a loro volta sono basate sul libro di Stroock e Varadhan [16])
Approssimazione del moto browniano (Billingsley [4])
Problema di Dirichlet (Baldi [1] e/o I. Karatzas e S. E. Shreve [7]).
Formula di Feynman-Kac per il processo di Wiener e/o per soluzioni di equazioni differenziali stocastiche (I. Karatzas e S. E. Shreve [7])
Formula di Tanaka e tempo locale (Revuz e Yor [11])
Applicazioni a problemi di controllo (Baldi [1])
Approssimazione di equazioni differenziali stocastiche (D. Higham SIAM [??])**(da completare)

DA FINIRE.....

Capitolo 1

Richiami su spazi di probabilità

1.1 Esempi di spazi di probabilità

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

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

$$\Omega \in \mathcal{F}; \tag{1.1}$$
$$\text{se } A \in \mathcal{F}, \text{ allora } A^c \in \mathcal{F}; \tag{1.2}$$
$$\text{se } A_n \in \mathcal{F}, n \in \mathbb{N}, \text{ allora } \cup_{n\in\mathbb{N}} A_n \in \mathcal{F}; \tag{1.3}$$

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

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

con le proprietà che

$$\mathbb{P}(\Omega) = 1; \tag{1.4}$$
$$\text{se } A_n \in \mathcal{F}, n \in \mathbb{N}, \text{ con } A_n \cap A_m = \emptyset \text{ per } n \neq m, \tag{1.5}$$
$$\text{allora } \mathbb{P}\big(\bigcup_{n\in\mathbb{N}} A_n\big) = \sum_{n\in\mathbb{N}} \mathbb{P}\big(A_n\big).$$

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

Oltre alla misura di probabilità $\mathbb{P}$, per tutti gli eventi $A \in \mathcal{F}$ con $\mathbb{P}(A) > 0$, si possono definire le ***probabilità condizionate***[1] ***all'evento*** A, che rappresentano la valutazione della probabilità nel caso in cui si verificasse l'evento A:

$$\mathbb{P}(\cdot|A)\, \mathcal{F} : \to [0,1] \tag{1.6}$$
$$E \mapsto \mathbb{P}(E|A) := \frac{\mathbb{P}(E \cap A)}{\mathbb{P}(A)} \tag{1.7}$$

Vediamo ora alcuni esempi elementari di spazi di probabilità:

[1] È facile verificare che la funzione $\mathbb{P}(\cdot|A)$ definita in (1.6) è una probabilità, cioè soddisfa gli assiomi delle probabilità. Per mettere in evidenza tale fatto va detto che Kolmogorov aveva adottato la notazione $\mathbb{P}_A(\cdot)$, ovvero $\mathbb{P}_A(E)$ invece di $\mathbb{P}(E|A)$, anche per mettere meglio in evidenza questa proprietà.

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

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

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

$$H_n \cap H_m = \emptyset \text{ per } n \neq m,\ n, m \in \{1, 2, \ldots, N\}$$

ed **esaustivi***:*

$$\bigcup_{m=1}^{N} H_m = \Omega,$$

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

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

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

$$\mathbb{P}(A) = \sum_{m\in I} p_m, \qquad \text{per } A = \bigcup_{m\in I} H_m, \tag{1.8}$$

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

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

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

allora la famiglia

$$\mathcal{F} = \{A = \bigcup_{m\in I} H_m,\ \text{al variare di } I \subseteq \mathbb{N}\},$$

(con la convenzione che $\bigcup_{m\in\emptyset} H_m = \emptyset$), è una σ-algebra[2]*.*
Inoltre se $p_1, p_2, \ldots, p_m, \ldots$ sono numeri non negativi, somma 1, ovvero

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

[2]La verifica è banale:

$$\Omega = \bigcup_{m\in\mathbb{N}} H_m, \text{ ovvero } I = \mathbb{N}$$

$$\text{se } A = \bigcup_{m\in I} H_m, \text{ allora } A^c = \bigcup_{m\in I^c} H_m$$

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

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

$$\mathbb{P}(A) = \sum_{m\in I} p_m, \qquad \textit{per } A = \bigcup_{m\in I} H_m, \tag{1.9}$$

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

La verifica di quest'ultima proprietà è banale[3].

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

1 ***l'intersezione di σ-algebre è una σ-algebra***
Sia $\{\mathcal{G}_\alpha, \alpha \in \Lambda\}$ una famiglia di σ-algebre, allora $\mathcal{F} := \bigcap_{\alpha\in\Lambda} \mathcal{G}_\alpha$ è una σ-algebra[4].

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

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

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

è la σ-algebra[5] generata da $\mathcal{K}$.

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

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

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

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

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

---

[3] La funzione $\mathbb{P} : \mathcal{M} \mapsto [0,1]$ definita in (1.9) è una probabilità, infatti

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

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

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

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

[4] La verifica è banale:

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

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

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

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

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

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

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

1.2 Variabili aleatorie

Definizione 1.1. *Dato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$[6], una* **variabile aleatoria reale** *X è una funzione $\mathcal{F}$-***misurabile***, ovvero una funzione*

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

tale che la controimmagine di ogni aperto $\mathcal{O} \in \mathcal{A}$ sia un elemento di $\mathcal{F}$[7], cioè tale che

$$X^{-1}(\mathcal{O}) := \{\omega \text{ tali che } X(\omega) \in \mathcal{O}\} \in \mathcal{F}, \text{ per ogni aperto } \mathcal{O} \in \mathcal{A}.$$

Si dice anche che X è una **variabile aleatoria $\mathcal{F}$-misurabile**.
Una definizione analoga vale nel caso di variabili aleatorie multidimensionali

$$\boldsymbol{X} : \Omega \mapsto \mathbb{R}^k; \quad \omega \mapsto \boldsymbol{X}(\omega) = \big(X_1(\omega), \ldots, X_k(\omega)\big),$$

basta infatti sostituire $\mathbb{R}$ con $\mathbb{R}^k$.

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

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

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

Esempio 1.6. *Sia $\{H_m,\, m \in \mathbb{N}\}$ una partizione numerabile, e sia $\mathcal{M}$ come nell'esempio 1.4. Allora $X : \Omega \mapsto \mathbb{R}$; $\omega \mapsto X(\omega)$ è $\mathcal{M}$-misurabile, se e solo se esiste una successione di costanti $\{c_m,\, m \in \mathbb{N}\}$[8], tale che*

$$X(\omega) = \sum_{m \in \mathbb{N}} c_m \mathbb{I}_{H_m}(\omega). \tag{1.10}$$

Se X è definita come in (1.10) allora X è $\mathcal{M}$-misurabile, infatti per ogni aperto $\mathcal{O}$,

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

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

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

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

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

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

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

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

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

$$X(\omega) = \sum_{m \in \mathbb{N}} x_m \mathbb{I}_{H_m}(\omega), \tag{1.11}$$

dove

$$H_m = X^{-1}(\{x_m\}) = \{\omega \text{ tali che } X(\omega) = x_m\}.$$

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

Inoltre la funzione X è una variabile aleatoria **$\mathcal{F}$-misurabile***, se e solo se*

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

come è immediato da (1.11), osservando che, come nel caso precedente,

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

Infine la variabile aleatoria X si dice **semplice** *o* **elementare***, se l'insieme $X(\Omega)$ è un insieme finito.*

Si può dimostrare che

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

2 la variabile aleatoria $\boldsymbol{X}$ è $\mathcal{F}$-misurabile, se e solo se ciascuna componente X_i è $\mathcal{F}$-misurabile[9], per ogni $i = 1, \ldots, k$. In particolare $X^{-1}(\{x\}) \in \mathcal{F}$, per ogni $x \in \mathbb{R}$, in quanto $\{x\} = \bigcap_n (x - 1/n, x + 1/n)$.

Connessa con la precedente Definizione 1.1 è la seguente definizione:

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

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

dove $\mathcal{R}^{\boldsymbol{X}}$ è la famiglia delle σ-algebre, per le quali $\boldsymbol{X}$ è $\mathcal{G}$-misurabile[10].

Si dimostra che

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

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

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

5 le variabili aleatorie $\sigma(\boldsymbol{X})$-misurabili a valori in $\mathbb{R}^d$ sono tutte e sole le variabili aleatorie $\boldsymbol{Z}$ per le quali esiste una funzione g boreliana[11] tale che

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

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

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

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

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

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

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

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

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

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

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

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

1.3 Distribuzioni di variabili aleatorie

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

$$X : \Omega \mapsto \mathbb{R}^k;\ \omega \mapsto X(\omega)$$

una variabile aleatoria a valori in $\mathbb{R}^k$. Tramite X è possibile definire una misura di probabilità $\mathbf{P}_X$ sullo spazio misurabile $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ nel seguente modo:

$$\mathbf{P}_X\ : \mathcal{B}(\mathbb{R}^k) \mapsto [0, 1] \quad \mathcal{I} \mapsto \mathbf{P}_X(\mathcal{I}) := \mathbb{P}\big(X \in \mathcal{I}\big).$$

È facile verificare che effettivamente $\mathbf{P}_X$ definisce una probabilità sui boreliani $\mathcal{B}(\mathbb{R}^k)$. La misura di probabilità così definita è detta ***misura di probabilità indotta da*** X, o ***distribuzione di*** X.

A volte, per indicare la misura di probabilità indotta, si usa il simbolo $\mathbb{P}X^{-1}$, che nasce dall fatto che $\mathbf{P}_X(\mathcal{I}) := \mathbb{P}\big(X \in \mathcal{I}\big) = \mathbb{P}\big(X^{-1}(\mathcal{I})\big)$. Nel seguito, a volte useremo anche il simbolo μ_X per indicare la distribuzione di probabilità di X.

Come è noto, associata alla variabile aleatoria X c'è anche la funzione di distribuzione[12]

$$F_X(x) := \mathbb{P}(\omega \in \Omega :\ X(\omega) \le x) = \mathbf{P}_X\big((-\infty, x]\big), \qquad x \in \mathbb{R}^k. \tag{1.12}$$

La funzione di distribuzione gode di alcune proprietà caratterizzanti[13]:

Proprietà delle funzioni di distribuzione

0 $F_X(x) \in [0, 1]$

1 La funzione F_X è continua dall'alto[14], nel senso che, per ogni $x \in \mathbb{R}^k$ si ha

$$\lim_{y \searrow x} F_X(y_1, \cdots, y_i, \cdots, y_k) = F(x_1, \cdots, x_i, \cdots, x_k),$$

dove $y \searrow x$ significa $y_i \to x_i^+$, per ogni $i = 1, \cdots, k$.

[12] Si ricordi che, per $k \ge 1$, l'evento e l'insieme nella (1.12) sono rispettivamente

$$\{\omega \in \Omega :\ X(\omega) \le x\} = \{\omega \in \Omega :\ X_1(\omega) \le x_1, \cdots, X_k(\omega) \le x_k\} \quad \text{e} \quad (-\infty, x] = (-\infty, x_1] \times \cdots \times (-\infty, x_k].$$

[13] Si veda la sezione 1.7

[14] Nel caso $k = 1$ la proprietà **1** corrisponde alla continuità da destra.

2 La funzione $F_X(x)$ è monotona non decrescente.

3 Siano $a = (a_1, \cdots, a_k)$ e $b = (b_1, \cdots, b_k)$, si definisca

$$\Delta(a,b) = \{x \in \mathbb{R}^k : \forall i = 1, \cdots, k, \text{ si ha } x_i = a_i \text{ oppure } x_i = b_i\},$$

e si definisca $n_a(x)$ il numero di i tali che $x_i = a_i$, per $x \in \Delta(a,b)$.
Se $a_i \leq b_i$, per ogni $i = 1, \cdots, k$, allora[15]

$$\sum_{x \in \Delta(a,b)} (-1)^{n_a(x)} F_X(x) \geq 0.$$

4 Per ogni $x \in \mathbb{R}^k$ e per ogni $i = 1, \cdots, k$ si ha che

$$\lim_{y_i \to -\infty} F_X(x_1, \cdots, x_{i-1}, y_i, x_{i+1}, \cdots, x_k) = 0.$$

Inoltre

$$\lim_{|x| \to +\infty} F_X(x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_k) = 1.$$

È importante sottolineare che la funzione di distribuzione F_X individua la misura di probabilità indotta $\mathbf{P}_X$ sulla famiglia (di boreliani)

$$(-\infty, b] := \{x \in \mathbb{R}^k \ : \ x_i \leq b_i\} \qquad \text{con } b = (b_1, \cdots, b_k) \in \mathbb{R}^k.$$

Questa famiglia ha la proprietà di essere chiusa rispetto all'intersezione finita:

$$(-\infty, b] \cap (-\infty, b'] = (-\infty, b \wedge b'], \qquad \text{dove } b \wedge b' := (b_1 \wedge b'_1, \cdots, b_k \wedge b'_k).$$

Ciò è sufficiente a individuare la misura di probabilità indotta, grazie a un risultato molto utile di teoria della misura:

Lemma 1.1 (Lemma di Dynkin, Billingsley 1984 [3])**.** *Sia $\mathcal{A}$ una famiglia di eventi che genera la σ-algebra $\mathcal{G}$ e che è chiusa rispetto alla intersezione finita (cioè: A, $B \in \mathcal{A}$ implica $A \cap B \in \mathcal{A}$). Se due misure di probabilità ν e μ coincidono su $\mathcal{A}$, allora le due misure coincidono su $\mathcal{G} = \sigma(\mathcal{A})$.*

Definizione 1.3 (variabili aleatorie con densità discreta)**.** *Si dice che una variabile aleatoria elementare X ha densità discreta*

$$\begin{pmatrix} x_1 & x_2 & \cdots & \cdots & x_m \\ p_1 & p_2 & \cdots & \cdots & p_m \end{pmatrix}$$

dove $x_1, x_2, \cdots x_m$ sono elementi di $\mathbb{R}^k$ e $p_1, p_2, \cdots p_m$ sono numeri reali tali che

$$p_j \geq 0 \quad \textit{per ogni } j = 1, 2, \cdots, m, \qquad \sum_{j=1}^{m} p_j = 1,$$

[15] Nel caso $k = 1$, la proprietà **3** corrisponde alla proprietà di monotonia **2**:

$$\text{se } a \leq b \quad \text{allora} \quad F_X(a) \leq F_X(b).$$

Nel caso $k = 2$, invece la proprietà **3** diviene:

$$\text{se } a_1 \leq b_1 \text{ e } a_2 \leq b_2 \quad \text{allora} \quad F_X(b_1, b_2) - F_X(a_1, b_2) - F_X(b_1, a_2) + F_X(a_1, a_2) \geq 0.$$

Per $k \geq 2$ la proprietà **3** non si riduce alla proprietà di monotonia **2**, come mostra il seguente controesempio:

$$F(x_1, x_2) =] \begin{cases} 0 & \text{se } x < 0, \quad \text{oppure se } x + y < 1, \quad \text{oppure se } y < 0. \\ 1 & \text{se } x \geq 0, \ y \geq 0, \ \text{e } x + y \geq 1 \end{cases}$$

Si vede facilmente che F è una funzione monotona. Tuttavia F non soddisfa la proprietà $\nVdash$, infatti

$$F(1,1) - F(1,0) - F(0,1) + F(0,0) = 1 - 1 - 1 + 0 = -1.$$

se, per ogni boreliano $\mathcal{I}$, vale

$$\mathbf{P}_X(\mathcal{I}) := \mathbb{P}(X \in \mathcal{I}) = \sum_{\substack{j=1 \\ x_j \in \mathcal{I}}}^{m} p_j.$$

In particolare quindi il significato di p_j è chiaro, essendo

$$\mathbb{P}(X = x_j) = p_j.$$

La definizione è analoga nel caso di variabili aleatorie discrete, la cui distribuzione viene caratterizzata attraverso una densità discreta su un insieme numerabile $\{x_k,\ k \geq 1\}$

$$\begin{pmatrix} x_1 & x_2 & \cdots & \cdots & x_m & x_{m+1} & \cdots \\ p_1 & p_2 & \cdots & \cdots & p_m & p_{m+1} & \cdots \end{pmatrix}$$

Esempio 1.9 (variabili aleatorie con distribuzione binomiale)**.** *Ogni variabile aleatoria X per la quale*

$$\mathbf{P}_X(\mathcal{I}) := \sum_{\substack{h=0 \\ h \in \mathcal{I}}}^{n} \binom{n}{h} p^h \, (1-p)^{n-h}$$

viene detta una variabile aleatoria binomiale di parametri n e p e si scrive in breve $X \sim Bin(n,p)$.

Definizione 1.4 (variabili con densità)**.** *Sisupponga di avere una funzione $f : \mathbb{R}^k \mapsto \mathbb{R}$ con le proprietà:*

$$f(x) \geq 0 \quad \text{per ogni } x \in \mathbb{R}^k, \qquad \int_{\mathbb{R}^k} f(x)\, dx = 1,$$

si dice che X ha distribuzione con **densità (di probabilità)** *f se accade che, per ogni boreliano $\mathcal{I} \in \mathcal{B}(\mathbb{R}^k)$,*

$$\mathbf{P}_X(\mathcal{I}) := \int_{\mathcal{I}} f(x)\, dx.$$

Esempio 1.10 (distribuzione gaussiana)**.** *Come caso particolare si consideri il caso della variabile aleatoria unidimensionale con densità*

$$f(x) = \frac{1}{\sqrt{2\,\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

dove μ è un numero reale e σ è un numero (strettamente) positivo. Una variabile aleatoria con questa distribuzione è detta **gaussiana** *o* **normale** *di valore atteso (o valore medio) μ e varianza σ^2. Brevemente si indica $X \sim N(\mu, \sigma^2)$. Se $\mu = 0$ e $\sigma^2 = 1$ si dice che X è una variabile gaussiana (o normale)* **standard**.

Vediamo ora dei semplici esempi di calcolo della distribuzione indotta.

Esempio 1.11 (una variabile aleatoria binomiale)**.** *Sia*

$$\Omega = \{0,1\}^N = \{\omega = (\omega_1, \omega_2, \ldots, \omega_N),\ \text{con } \omega_i \in \{0,1\},\ \text{per } i = 1, 2, \ldots, N\},$$

sia

$$\mathcal{F} = \mathcal{P}(\Omega),$$

l'insieme delle parti di Ω, sia la probabilità definita attraverso la relazione

$$\mathbb{P}(\{\omega\}) := p^{\sum_{i=1}^N \omega_i} \, (1-p)^{N - \sum_{i=1}^N \omega_i},$$

dove p è un numero fissato con la condizione che $p \in (0,1)$. Sia infine X la variabile aleatoria definita da

$$X(\omega) := \sum_{i=1}^{N} \omega_i.$$

Si vede facilmente che

• 1 *la variabile aleatoria X assume solo i valori $\{0, 1, \dots N\}$,*

• 2 *per $h \in \{0, 1, \dots N\}$ si ha*[16]

$$\mathbf{P}_X(h) := \mathbb{P}(X = h) = \binom{N}{h} p^h (1-p)^{N-h},$$

• 3 *per ogni boreliano $\mathcal{I}$*

$$\mathbf{P}_X(\mathcal{I}) := \mathbb{P}(X \in \mathcal{I}) = \sum_{\substack{h=0 \\ h \in \mathcal{I}}}^{N} \binom{N}{h} p^h (1-p)^{N-h}$$

Esempio 1.12 (Variabili esponenziali)**.** *Sia $\Omega = (0,1)$ $\mathcal{F} = \mathcal{B}(0,1)$ e $\mathbb{P}$ la misura di Lebesgue su $(0,1)$. Sia $\lambda > 0$ e $X(\omega) := -\log(1-\omega)/\lambda$. Allora*

$$F_X(x) = \mathbb{P}(X \le x) = mis\{\omega \in (0,1) \,:\, -\log(1-\omega) \le \lambda\, x\} = mis\{\omega \in (0,1) \,:\, \omega \le 1 - e^{-\lambda\, x}\},$$

e quindi

$$F_X(x) = \begin{cases} 0 & per\ x \le 0, \\ 1 - e^{-\lambda\, x} & per\ x > 0. \end{cases}$$

Per il Lemma di Dynkin (Lemma 1.1) sappiamo che la funzione di distribuzione individua univocamente la distribuzione di X. È quindi facile convincersi che, tale distribuzione coincide con la distribuzione

$$\nu_\lambda(dx) = \mathbf{1}_{(0,\infty)}(x)\, \lambda\, e^{-\lambda\, x}\, dx,$$

che è nota come la distribuzione esponenziale di parametro λ.
Sempre nello stesso spazio si può definire la variabile aleatoria

$$Y(\omega) = -\frac{\log(\omega)}{\mu},$$

dove μ è una costante strettamente positiva. È facile vedere che Y ha distribuzione esponenziale, di parametro μ.

Esempio 1.13. *Sempre nello stesso ambito dell'esempio precedente, ci si può chiedere quale sia la distribuzione congiunta di X e Y, ossia la distribuzione del vettore aleatorio (X, Y).*

Chiaramente, si ha $X(\omega)$, $Y(\omega) > 0$ e inoltre

$$Y(\omega) = -\frac{\log\left(1 - e^{-\lambda\, X(\omega)}\right)}{\mu},$$

come si ottiene subito da

$$\omega = 1 - e^{-\lambda\, X(\omega)}.$$

Di conseguenza, se $G := \{(x,y) \,:\, x > 0, y > 0, ey = -\frac{\log\left(1-e^{-\lambda\, x}\right)}{\mu}\}$, è facile convincersi che

$$\mathbf{P}_{X,Y}(\mathcal{I}) = \nu_\lambda\big(\pi_x(G \cap \mathcal{I})\big)\Big(= \nu_\mu\big(\pi_y(G \cap \mathcal{I})\big)\Big)$$

dove π_x e π_y sono le proiezione sull'asse x e sull'asse y, rispettivamente.

[16] L'evento $A_h := \{X = h\}$ è rappresentato dall'insieme, di cardinalità $\binom{N}{h}$, i cui elementi $\omega = (\omega_1, \omega_2, \dots, \omega_N)$ hanno la proprietà che $\sum_{i=1}^N \omega_i = h$. La probabilità di ciascuno di questi ω vale quindi

$$\mathbb{P}(\omega) = p^{\sum_{i=1}^N \omega_i} (1-p)^{N - \sum_{i=1}^N \omega_i} = p^h (1-p)^{N-h}$$

e la probabilità dell'insieme vale

$$\mathbb{P}(X = h) = \mathbb{P}(A_h) = \sum_{\omega \in A_h} \mathbb{P}(\omega) = \sum_{\omega \in A_h} p^h (1-p)^{N-h} = |A_h| p^h (1-p)^{N-h} = \binom{N}{h} p^h (1-p)^{N-h}$$

Esempio 1.14 (trasformazione di Box-Müller)**.** *Sia $\Omega = (0,1)\times(0,1)$, con la misura di Lebesgue sui boreliani. Siano*

$$X(\omega_1,\omega_2) := \sqrt{-2\,\log\omega_1}\,\cos(2\,\pi\,\omega_2);$$
$$Y(\omega_1,\omega_2) := \sqrt{-2\,\log\omega_1}\,\sin(2\,\pi\,\omega_2);$$

Si può dimostrare che la distribuzione congiunta di (X,Y) ammette densità si probabilità

$$p_{X,Y}(x,y) = \frac{1}{2\,\pi}\,e^{-\frac{x^2+y^2}{2}} = \frac{1}{\sqrt{2\,\pi}}\,e^{-\frac{x^2}{2}}\,\frac{1}{\sqrt{2\,\pi}}\,e^{-\frac{x^2}{2}}$$

Tale densità caratterizza le variabili aleatorie gaussiane com media nulla e matrice di covarianza l'identità (si veda l'Appendice 1.6).

A volte, invece di definire lo spazio di probabilità e la variabile aleatoria X ed infine trovare la distribuzione di X, si può dare direttamente la distribuzione di X. Questo è il caso delle variabili aleatorie che vengono caratterizzate solo attraverso la densità discreta o con densità (di probabilità).

Più in generale, le distribuzioni si possono specificare solo attraverso la funzione di distribuzione.

Quando si specifica una variabile aleatoria attraverso la sua distribuzione, e ancor di più se invece si specifica solo una funzione che goda delle proprietà delle funzioni di distribuzione (si veda pag. 6), rimane il dubbio che una tale variabile aleatoria esista, ovvero che esista uno spazio di probabilità $(\Omega,\mathcal{F},\mathbb{P})$ e una variabile aleatoria X. A questo problema risponde il teorema di Skorohod (vedere Appendice 1.7).

1.4 Valori attesi

In questa sezione ricordiamo come si può definire il valore atteso per variabili aleatorie generali, a partire dalla sua definizione per variabili aleatorie semplici. Per maggiori approfondimenti si rimanda, ad esempio, al libro di Billingsley [3] o a quello di Williams [17].

Definizione 1.5 (Valore atteso per variabili semplici)**.** *Sia X una variabile aleatoria in $(\Omega,\mathcal{F},\mathbb{P})$, non negativa e semplice, cioè come in Esempio 1.7,*

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

allora si definisce

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

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

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

Infatti[17] *basta prendere*

$$X_n(\omega) = \sum_{m=0}^{n2^n-1} \frac{m}{2^n}\mathbb{I}_{H_m^{(n)}}(\omega) + n\mathbb{I}_{H_{n2^n}^{(n)}}(\omega) = \sum_{m=0}^{n2^n-1} \frac{m}{2^n}\mathbf{1}_{[\frac{m}{2^n},\frac{m+1}{2^n})}(X(\omega)) + n\mathbf{1}_{[n,\infty)}(X(\omega)), \tag{1.13}$$

[17] La monotonia della successione delle variabili aleatorie X_n è evidente:
- se $X_n(\omega) = m/2^n$, con $m < n2^n$, allora i soli casi possibili sono

$$X_{n+1}(\omega) = (2m)/2^{n+1} = m/2^n = X_n(\omega),$$

oppure

$$X_{n+1}(\omega) = (2m+1)/2^{n+1} = m/2^n + 1/2^{n+1} > X_n(\omega);$$

- se $X_n(\omega) = n$ allora $X_{n+1}(\omega)$ può assumere un valore compreso tra n ed $n+1$.

Per la convergenza basta osservare che, qualunque sia ω, pur di prendere n sufficientemente grande e in modo che $X(\omega) < n$, si ha che

$$0 \le X(\omega) - X_n(\omega) \le 1/2^n.$$

dove si è posto

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

e, per $A \in \mathcal{F}$,

$$\mathbb{I}_A(\omega) = 1 \quad se \quad \omega \in A \qquad e \qquad \mathbb{I}_A(\omega) = 0 \quad se \quad \omega \notin A,$$

ed infine, per $a < b$ *numeri reali,*

$$\mathbf{1}_{[a,b)}(x) = 1 \quad se \quad x \in [a,b) \qquad e \qquad \mathbf{1}_{[a,b)}(x) = 0 \quad se \quad x \notin [a,b).$$

È infine interessante notare che, posto $\lfloor x \rfloor$ *la parte intera inferiore*[18] *di* x*, si può riscrivere nel seguente modo*

$$X_n(\omega) = \frac{\lfloor 2^n X(\omega) \rfloor}{2^n} \wedge n.$$

Definizione 1.6 (Valore atteso per variabili nonnegative)**.** *Sia* X *una variabile aleatoria in* $(\Omega, \mathcal{F}, \mathbb{P})$*, non negativa, si definisce*

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

dove $\{X_n;\ n \in \mathbb{N}\}$ *è la successione monotona definita come in (1.13) dell'Osservazione precedente. Il limite esiste ed è monotono, per la proprietà di monotonia del valore atteso, sulle variabili aleatorie semplici. Si noti bene che tale limite può valere anche* $+\infty$*, nel qual caso si dice che la variabile* X *ha valore atteso infinito.*

Osservazione 1.2. ***Ovviamente se* $\{\widehat{X}_n;\ n \in \mathbb{N}\}$ *è un'altra successione di variabili aleatorie semplici che converge monotonamente ad* X*, anche la successione dei valori attesi* $\mathbb{E}[\widehat{X}_n]$ *è una successione che converge monotonamente. Si può dimostrare che il limite non dipende dalla successione scelta*[19] *ed in particolare coincide con il limite considerato nella precedente Definizione 1.6.*

Arriviamo ora alla definizione generale del valore atteso:

Definizione 1.7 (Valore atteso per variabili generali)**.** *Sia* X *una variabile aleatoria in* $(\Omega, \mathcal{F}, \mathbb{P})$*, Siano* $X^+ := X \vee 0$ *e* $X^- := (-X) \vee 0$*, le variabili aleatorie non negative, definite alla fine della sezione precedente. Si noti che* $X = X^+ - X^-$ *e che invece* $|X| = X^+ + X^-$*. Si definisce allora, se ha senso*[20]

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

---

[18] La parte intera inferiore $\lfloor x \rfloor$ di x è quel numero intero k tale che $k \le x < k+1$.

[19] **Per ottenere l'unicità del limite basta dimostrare che se $\{Y_n;\ n \in \mathbb{N}\}$ e $\{Z_n;\ n \in \mathbb{N}\}$ sono due successioni di variabili aleatorie semplici che convergono monotonamente ad X, allora per ogni k si ha

$$\mathbb{E}[Y_k] \le \lim_{n\to\infty} \mathbb{E}[Z_n],$$

da cui si deduce immediatamente che $\lim_{k\to\infty} \mathbb{E}[Y_k] \le \lim_{n\to\infty} \mathbb{E}[Z_n]$e quindi l'uguaglianza, scambiando il ruolo delle due successioni.
Si fissi quindi k e si consideri che, per ipotesi Y_k è semplice e che quindi si può scrivere $Y_k = \sum_{i=1}^{\ell} y_i \mathbb{I}_{A_i}$, dove $A_i = \{Y_k = y_i\}$ (ovviamente ℓ ed y_i dipendono da k, ma tralasciamo l'indice k per comodità di notazione e perché è inessenziale). Sia ora $\varepsilon > 0$ e $B_i^{(n)} = A_i \cap \{Z_n > y_i - \varepsilon\}$. Essendo $\{Z_n;\ n \in \mathbb{N}\}$ una successione monotona si ottiene che $B_i^{(n)} \subset B_i^{(n+1)}$. Inoltre $Y_k(\omega) \le X(\omega)$ e $Z_n(\omega) \nearrow X(\omega)$ e quindi se $\omega \in A_i$, ossia se $Y_k(\omega) = y_i$, allora per un n sufficientemente grande deve valere $Z_n(\omega) > y_i - \varepsilon$ e quindi $\bigcup_n B_i^{(n)} = A_i$. Per la continuità della probabilità deve valere allora che $\mathbb{P}(B_i^{(n)}) \nearrow \mathbb{P}(A_i)$. Ovviamente si ha

$$\mathbb{E}[Z_n] \ge \sum_{i=1}^{\ell} (y_i - \varepsilon)\mathbb{P}(B_i^{(n)}) \qquad \text{e quindi} \qquad \lim_{n\to\infty} \mathbb{E}[Z_n] \ge \sum_{i=1}^{\ell} (y_i - \varepsilon)\mathbb{P}(A_i).$$

Per l'arbitrarietà di ε si ha allora $\lim_{n\to\infty} \mathbb{E}[Z_n] \ge \sum_{i=1}^{\ell} y_i \mathbb{P}(A_i) = \mathbb{E}[Y_k]$.

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

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

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

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

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

Se invece di usare la probabilità $\mathbb{P}$ si usa la probabilità condizionata ad un evento A, ovvero $\mathbb{P}(\cdot|A)$, allora si parla di ***valore atteso di*** X ***condizionato all'evento*** A e si usa la notazione

$$\mathbb{E}[X|A].$$

Ciò significa che, nel caso di una variabile aleatoria semplice

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

si ha

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

**Terminiamo questa sezione ricordando che la definizione di valore atteso di una variabile aleatoria X corrisponde alla definizione dell'integrale della funzione misurabile X rispetto alla misura $\mathbb{P}$ e che per il valore atteso valgono i due famosi risultati di passaggio al limite sotto il segno di integrale:
Teorema della convergenza monotona: se X_n sono variabili aletaorie limitate dal basso e che convergono monotonamente ad X ($\mathbb{P} - q.c.$) allora la successione dei valori attesi $\mathbb{E}[X_n]$ converge monotonamente a $\mathbb{E}[X]$.

Teorema della convergenza dominata: se X_n sono variabili aletaorie che convergono ad X $\mathbb{P} - q.c.$ e se Y è una variabile aleatoria tale che $|X_n| \leq Y$, con $\mathbb{E}[Y] < \infty$, allora la successione dei valori attesi $\mathbb{E}[X_n]$ converge a $\mathbb{E}[X]$.

1.4.1 Variabili aleatorie in spazi misurabili **

Oltre a definire le variabili aleatorie reali o vettoriali si possono definire in modo naturale anche variabili aleatorie a valori in spazi misurabili.

Definizione 1.8 (variabile aleatoria (o ente alealorio) a valori in $(S, \mathcal{S})$)**.** *Siano $(\Omega, \mathcal{F})$ e $(S, \mathcal{S})$ due spazi misurabili. Una variabile aleatoria a valori in S è una funzione misurabile*

$$X: \ (\Omega, \mathcal{F}) \to (S, \mathcal{S}); \ \omega \mapsto X(\omega).$$

In altre parole una funzione da Ω in S e tale che per ogni $B \in \mathcal{S}$, la sua controimmagine tramite X appartiene a $\mathcal{F}$, ossia l'insieme $X^{-1}(B) \in \mathcal{F}$.

Se S è uno **spazio metrico** *(o più in generale uno spazio topologico, allora la sigma-algebra $\mathcal{S}$ coincide con la sigma-algebra dei boreliani, ossia la sigma-algebra generata dagli aperti.*

Esempi tipici nascono quando si vogliono trattare i processi aleatori come funzioni aleatorie, ed in particolare a funzioni aleatorie continue. In tale caso si può prendere, ad esempio, lo spazio delle funzioni continue su $[0, T]$ a valori reali. Prendendo poi come sigma-algebra la sigma-algebra dei boreliani, allora si può affermare che funzioni come il massimo o il minimo, sono variabili aleatorie.

Come si vede, nella definizione di variabile aleatoria non abbiamo neanche nominato la misura di probabilità su $(\Omega, \mathcal{F})$. Tuttavia rimane poi il problema di definire le misure di probabilità su tali spazi. Questo argomento è trattato in maggiore dettaglio nel Capitolo 4 sui processi sticastici.

1.5 Misura indotta e Cambio di variabile **

Da scrivere

Negli Esempi 1.12, 1.13, 1.14 abbiamo trovato le distribuzioni di alcune variabili aleatorie a valori reali o vettoriali.

In termini astratti quello che abbiamo fatto è caratterizzare la misura indotta.

Definizione 1.9 (Misura indotta)**.** *Siano $(A_1, \mathcal{A}_1)$ e $(A_2, \mathcal{A}_2)$ due spazi con le rispettive sigma-algebre, e sia $\psi : A_1 \to A_2$, $a_1 \mapsto \psi(a_1)$ una funzione misurabile (cioè per ogni $B_2 \in \mathcal{A}_2$ si ha che la controimmagine $\psi^{-1}(B_2) \in \mathcal{A}_1$). Supponiamo che su $(A_1, \mathcal{A}_1)$ sia definita una misura μ_1. Allora si definisce misura indotta (da ψ) la misura*

$$\mu_2(B_2) := \mu_1\big(\psi^{-1}(B_2)\big), \qquad B_2 \in \mathcal{A}_2.$$

Ovviamente perché la precedente definizione sia ben posta bisogna verificare che effettivamente definisca una misura (questo è un semplice esercizio ed è lasciato al lettore).

Tornando agli Esempi precedentemente citati ed in particolare agli Esempi 1.12, 1.13, in entrambi $(A_1, \mathcal{A}_\infty) = \big((0,1), \mathcal{B}(0,1)\big)$ e $\mu_1 = \mathbb{P}$, la misura di Lebesgue ristretta a $(0,1)$, mentre $A_2 = \mathbb{R}$ nel primo esempio e invece $A_2 = \mathbb{R}^2$, nel secondo esempio. Inoltre nel primo esempio sono state considerate due funzioni $\psi_1(\omega) = X(\omega) = -\frac{\log(1-\omega)}{\lambda}$ e $\psi_2(\omega) = Y(\omega) = -\frac{\log(\omega)}{\lambda}$, mentre nel secondo esempio è stata considerata la funzione $\psi := (\psi_1, \psi_2)$. Nell'Esempio 1.14, invece $(A_1, \mathcal{A}_\infty) = \big((0,1) \times (0,1), \mathcal{B}((0,1) \times (0,1)\big)$ e $A_2 = \mathbb{R}^2$ e μ_1 è la misura di Lebesgue ristretta a $(0,1) \times (0,1)$. Infine la funzione ψ è definita da $\psi(\omega_1, \omega_2) = \big(\sqrt{-2 \log \omega_1}\, \cos(2\,\pi\, \omega_2), \sqrt{-2 \log \omega_1}\, \sin(2\,\pi\, \omega_2)\big)$.

Più in generale, nel caso di variabili aleatorie X a valori in $(S; \mathcal{S})$, se nello spazio misurabile $(\Omega, \mathcal{F})$ è definita una misura di probabilità $\mathbb{P}$, si definisce legge di X o distribuzione di X, la probabilità $\mathbf{P}_X : \mathcal{S} \to [0,1]$ definita come la misura indotta da $(\Omega, \mathcal{F}, \mathbb{P})$ tramite X:

$$\mathbf{P}_X(B) := \mathbb{P}(X \in B), \quad B \in \mathcal{S}.$$

Quello che più ci interessa qui è la formula del cambio di variabile negli integrali, che, nell'ambito del calcolo delle probabilità, corrisponde alla possibilità di calcolare i valori attesi di funzioni di variabili aleatorie X a valori[21] in $(S, \mathcal{S})$ sia come integrali sullo spazio $(\Omega, \mathcal{F}, \mathbb{P})$ che come integrali sullo spazio $(S, \mathcal{S}, \mathbf{P}_X)$.

In tale caso si ottiene che i valori attesi di $f(X)$, per f funzioni misurabili e limitate, si possono calcolare sia come integrali sulo spazio degli eventi Ω

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

sia come integrale sullo spazio degli stati S

$$\mathbb{E}[f(X)] = \int_S f(x)\, \mathbf{P}_X(dx).$$

Riportiamo qui la dimostrazione nell'ambito astratto della Definizione 1.9 di misura indotta.

Lemma 1.2 (Cambio di variabile). *Sia $f \in \mathcal{M}_b(A_2)$, ossia una funzione misurabile da $\big(A_2, \mathcal{A}_2\big)$ in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ e limitata. Allora*

$$\int_{A_2} f(a_2)\mu_2(da_2) = \int_{A_1} f(\psi(a_1))\mu_1(da_1) \tag{1.14}$$

Dimostrazione. Iniziamo con il mostrare che, per definizione di μ_2, (1.14) è valida per $f = \mathbb{I}_{B_2}$, per ogni $B_2 \in \mathcal{A}_2$: da una parte

$$\int_{A_2} \mathbb{I}_{B_2}(a_2)\mu_2(da_2) = \mu_2(B_2) := \mu_1\big(\psi^{-1}(B_2)\big),$$

dall'altra, tenuto conto che $\mathbb{I}_{B_2}(\psi(a_1)) = \mathbb{I}_{\psi_{-1}(B_2)}(a_1)$, in quanto $\psi(a_1) \in B_2$ se e solo se $a_1 \in \psi^{-1}(B_2)$,

$$\int_{A_1} \mathbb{I}_{B_2}(\psi(a_1))\mu_1(da_1) = \int_{A_1} \mathbb{I}_{\psi^{-1}(B_2)}(a_1)\mu_1(da_1) = \mu_1\big(\psi^{-1}(B_2)\big).$$

La dimostrazione segue poi con una tecnica che è standard nell'ambito della teoria della misura.

Sia $\mathcal{H}$ l'insieme delle funzioni f per cui è valida l'uguaglianza (1.14).

L'insieme $\mathcal{H}$ verifica le seguenti proprietà:

(i) **linearità**, ovvero se f, $g \in \mathcal{H}$, allora, per ogni a,$b \in \mathbb{R}$ la funzione $a\, f + b\, g \in \mathcal{H}$

[21] Nel caso di variabili aleatorie vettoriali lo spazio $(S, \mathcal{S})$ coincide con $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$. Ma la formula vale anche per variabil aleatorie a valori in spazi più generali, come ad esempio gli spazi metrici, prendendo come sigma-algebra la sigma-algebra dei boreliani, ossia la sigma-algebra generata dagli aperti (in altre parole la più piccola sigma algebra contente gli aperti). Come già detto esempi di tale genere si incontrano quando ci si interessa di processi aleatori, pensati come variabili aletorie a valori in uno spazio di funzioni, ad esempio lo spazio delle funzioni continue su un intervallo $[0, T]$, con la metrica della norma uniforme.

(come segue dalla proprietà di linearità per gli integrali rispetto a μ_1)
(ii) la funzione $\mathbf{1}$, cioè la funzione costante uguale ad 1, appartiene a $\mathcal{H}$
(come segue dall'osservazione iniziale e notando che $\mathbf{1} = \mathbb{I}_{A_2}$,)
(iii) **monotonia**, ovvero se $f_n \in \mathcal{H}$ e $f_n \nearrow f$, $f \in \mathcal{M}_2(A_2)$ allora $f \in \mathcal{H}$
(come segue dalla proprietà della convergenza monotona degli integrali rispetto a μ_1)
(iv) per ogni $B_2 \in \mathcal{A}_2$, la funzione $\mathbb{I}_{B_2} \in \mathcal{H}$
(come segue immediatamente dalla osservazione iniziale; si noti inoltre che in realtà la (ii) segue da questa proprietà)

Le precedenti proprietà assicurano che $\mathcal{H}$ è una classe monotona. Basta allora applicare il teorema delle classi monotone, che per comodità del lettore riportiamo di seguito. □

Teorema 1.3 (Teorema delle classi monotone)**.** *Sia $(\Omega, \mathcal{F})$ uno spazio misurabile e sia $\mathcal{H}$ un insieme di funzioni reali misurabili e limitate, con le seguenti proprietà:*
(i) $\mathcal{H}$ è uno spazio vettoriale,
(ii) $\mathcal{H}$ contiene la funzione costante $\mathbf{1}$,
(iii) $f_n \in \mathcal{H}$, $f_n \nearrow f$, f limitata implicano $f \in \mathcal{H}$
cioè $\mathcal{H}$ è una classe monotona.

Se inoltre $\mathcal{H}$ soddisfa anche la seguente proprietà

(iv) $\mathcal{H}$ contiene le funzioni del tipo $\mathbb{I}_A$ per ogni $A \in \mathcal{A}$, dove $\mathcal{A} \subset \mathcal{F}$ è un π-sistema, cioè è chiuso per intersezione finita,

allora $\mathcal{H}$ contiene tutte le funzioni limitate e $\sigma(\mathcal{A})$-misurabili.

Il precedente Teorema 1.2 si applica anche quando vogliamo calcolare la distribuzione di una trasformazione di una variabile aleatoria: ad esempio, se Z è una variabile aleatoria con distribuzione $\mathbf{P}_Y$ ed $Z = \varphi(Y)$, allora $\mathbf{P}_Z(B) = \mathbf{P}_Y\big(\varphi^{-1}(B)\big)$, come è immediato verificare.

Nel caso di variabili aleatorie multivariate, e per funzioni φ sufficientemente regolari, si possono ottenere formule esplicite, utilizzando noti risultati di analisi: ad esempio, se Y ammette densità f_Y e φ è invertibile[22] e con derivate continue, allora anche Z ammette densità e si ha

$$f_Z(z) = f_Y(\varphi^{-1}(z)) \left| det \left(\frac{\partial \varphi^{-1}(z)}{\partial z} \right) \right| = f_Y(\varphi^{-1}(z)) \frac{1}{\left| det \left(\frac{\partial \varphi(y)}{\partial y} \right) \right|_{y=\varphi^{-1}(z)}} .$$

Particolarmente semplice è il caso di trasformazioni lineari (o affini) in cui lo Jacobiano è il determinante della matrice. Ad esempio se $Z = AY$, con A invertibile, allora $\varphi^{-1}(z) = A^{-1}z$ e la formula precedente diviene

$$f_Z(z) = f_Y(A^{-1}(z)) \frac{1}{|det(A)|}.$$

Esempio 1.15. *Un esempio di trasformazione che incontreremo spesso nel seguito è il caso in cui $Y = (Y_1, Y_2, \cdots, Y_m)$ e*

$$\begin{aligned} Z_1 &= Y_1, & & & z_1 &= y_1, \\ Z_2 &= Y_1 + Y_2, & & \textit{ossia } z = \varphi(y) = A\,y, \textit{ con} & z_2 &= y_1 + y_2, \\ &\cdots & & & &\cdots \\ Z_m &= Y_1 + Y_2 + \cdots + Y_m, & & & z_m &= y_1 + y_2 + \cdots + y_m. \end{aligned}$$

Allora la matrice A è la matrice triangolare $A = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix}$ *con determinante uguale ad* 1.

[22] In realtà basta che esista un aperto $\mathcal{O}$, tale che la densità $f_Y(y) = 0$ per $y \notin A$ e tale che φ sia invertibile da $\mathcal{O}$ a $\varphi(\mathcal{O}$,

La trasformazione inversa è

$$\begin{aligned} y_1 &= z_1, \\ y_2 &= z_2 - z_1, \\ &\cdots \\ y_m &= z_m - z_{m-1}, \end{aligned} \qquad \textit{ossia } y = \varphi^{-1}(z) = A^{-1}y \textit{ dove } A^{-1} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

per cui, se Y ammette densità di probabilità,

$$f_Z(z_1, z_2, \cdots, z_m) = f_Y(z_1, z_2 - z_1, \cdots, z_m - z_{m-1}).$$

Il caso $m = 2$ è particolarmente interessante in quanto permette di ricavare la densità della somma di due variabili aleatorie, semplicemente calcolando la densità marginale di $Z_2 = Y_1 + Y_2$: per $z \in \mathbb{R}$

$$f_{Y_1+Y_2}(x)\Big(= f_{Z_2}(x)\Big) = \int_{\mathbb{R}} f_{Z_1,Z_2}(x, x')\, dx' = \int_{\mathbb{R}} f_{Y_1,Y_2}(x, x' - x)\, dx'.$$

1.6 Appendice: variabili Gaussiane

Cominciamo con il definire una variabile aleatoria gaussiana standard unidimensionale:

Definizione 1.10. *Si dice che una variabile aleatoria reale Z è* ***gaussiana*** *di valore atteso μ e varianza σ^2, se ammette densità*

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

In questo caso si usa la notazione $Z \sim N(\mu, \sigma^2)$. Se $\mu = 0$ e $\sigma^2 = 1$ allora si dice che Z segue una ***legge normale o gaussiana standard****.*

Caso $n-$dimensionale: iniziamo con il caso di un vettore (colonna) aleatorio

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_k \\ \cdots \\ Y_n \end{pmatrix}$$

a componenti indipendenti e tutte gaussiane standard, ovvero il caso in cui

$$\begin{aligned} f_Y(\boldsymbol{y}) &= \prod_{i=1}^{n} f_{Y_i}(y_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}y_i^2\right\} \\ &= \frac{1}{(\sqrt{2\pi})^n} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} y_i^2\right\} = \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\boldsymbol{y}'\boldsymbol{y}\right\}. \end{aligned}$$

dove l'apice indica l'operazione di trasposizione, ovvero $\boldsymbol{y}'$ è il vettore riga $(y_1, y_2, \cdots, y_n)$.

È immediato verificare che $\mathbb{E}(Y_i) = 0$, $Var(Y_i) = 1$ e che $Cov(Y_i, Y_j) = 0$, per $i \neq j$.

Sia ora A una matrice non singolare e sia $\boldsymbol{m}$ un vettore (colonna). Definiamo ora $Z = AY + \boldsymbol{m}$ e cerchiamo la sua densità. Sappiamo dai risultati generali che se Y ammette densità e $Z = \varphi(Y)$ con φ invertibile e con derivate continue, allora anche Z ammette densità:

$$f_Z(z) = f_Y(\varphi^{-1}(z)) \left| det\left(\frac{\partial \varphi^{-1}(z)}{\partial z}\right)\right| = f_Y(\varphi^{-1}(z)) \frac{1}{\left|det\left(\frac{\partial \varphi(y)}{\partial y}\right)\right|_{y=\varphi^{-1}(z)}}$$

di conseguenza, poiché nel nostro caso $\varphi(y) = Ay + m$ e $\varphi^{-1}(z) = A^{-1}(z-m)$

$$f_Z(z) = \frac{1}{\sqrt{2\pi}^n} \exp\left\{-\frac{1}{2}\left(A^{-1}(z-m)\right)' A^{-1}(z-m)\right\} \frac{1}{|det(A)|}.$$

Essendo

$$\begin{aligned} &(A^{-1}(z-m))'A^{-1}(z-m) = (z-m)'(A^{-1})'A^{-1}(z-m) \\ &= (z-m)'(A')^{-1}A^{-1}(z-m) = (z-m)'(AA')^{-1}(z-m) \end{aligned}$$

si ottiene

$$f_Z(z) = \frac{1}{(2\pi)^{n/2}} \frac{1}{|det(A)|} \exp\left\{-\frac{1}{2}(z-m)'\left(AA'\right)^{-1}(z-m)\right\}.$$

La precedente espressione si basa sulle seguenti proprietà:

(i) $\boxed{(A')^{-1} = (A^{-1})'}$

in quanto

$$A'z = w \Leftrightarrow z = (A')^{-1}w$$

e inoltre

$$A'z = w \Leftrightarrow (z'A)' = w \Leftrightarrow z'A = w' \Leftrightarrow z' = w'A^{-1}$$

$$\Leftrightarrow z = \left(w'A^{-1}\right)' \Leftrightarrow z = \left(A^{-1}\right)' w.$$

(ii) $\boxed{(AA')^{-1} = (A')^{-1} A^{-1}}$

in quanto

$$(AA')^{-1} z = w \Leftrightarrow z = AA'w \Leftrightarrow A^{-1}z = A'w \Leftrightarrow (A')^{-1} A^{-1}z = w.$$

È interessante notare che sia il vettore $\boldsymbol{m}$ che la matrice $AA' = A'A$ hanno una interpretazione probabilistica:

$$\mathbb{E}(Z_i) = \mathbb{E}(\sum_{k=1}^{n} a_{i,k}Y_k) + m_i = \sum_{k=1}^{n} a_{i,k}\mathbb{E}(Y_k) + m_i = m_i$$

$$Cov(Z_i, Z_j) = \mathbb{E}[(Z_i - m_i)(Z_j - m_j)] = \mathbb{E}[\sum_{k=1}^{n} a_{i,k}Y_k \sum_{h=1}^{n} a_{j,h}Y_h] = \sum_{k=1}^{n}\sum_{h=1}^{n} a_{i,k}a_{j,h}\mathbb{E}[Y_kY_h]$$

e quindi

$$Cov(Z_i, Z_j) = \sum_{k=1}^{n} a_{i,k}a_{j,k}\mathbb{E}[Y_kY_k] + \sum_{k=1}^{n}\sum_{h\neq k}^{1,n} a_{i,k}a_{j,h}\mathbb{E}[Y_kY_h] = \sum_{k=1}^{n} a_{i,k}a_{j,k} = (AA')_{i,j}$$

Si osservi che se $\boldsymbol{Z} = (Z_1, ..., Z_n)$ è un vettore gaussiano allora $(Z_1...Z_k)$ e $(Z_{k+1}, ..., Z_n)$ sono indipendenti, se e solo se $Cov(Z_i, Z_h) = 0$ per ogni $i = 1, \cdots, k$ e $h = k+1, \cdots, n$. In tale caso allora è ovvio che il vettore $(Z_1...Z_k)$ è un vettore gaussiano[23]

Terminiamo questo paragrafo con il ricordare quanto valgono i ***momenti di una variabile aleatoria gaussiana***. Sia Z una variabile aleatoria $\mathcal{N}(0, \sigma^2)$. Per quanto visto prima possiamo considerare $Z = \sigma Y$ con Y una variabile aleatoria $\mathcal{N}(0, 1)$. Da questa osservazione segue subito che

$$\mathbb{E}[Z^k] = \sigma^k\mathbb{E}[Y^k]$$

e

$$\mathbb{E}[|Z|^k] = |\sigma|^k\mathbb{E}[|Y|^k].$$

[23] Per ottenere lo stesso risultato nel caso generale, ovvero che se $\boldsymbol{Z} = (Z_1, ..., Z_n)$ è un vettore gaussiano allora $(Z_1...Z_k)$ è un vettore gaussiano, si può procedere nel seguente modo. Innanzitutto basta considerare il caso in cui i valori attesi sono nulli senza ledere in generalità. Inoltre si può pensare che $\boldsymbol{Z} = A\boldsymbol{Y}$. Se la matrice $A' = (a'_{ij})$ è definita in modo che $a'_{ij} = a_{ij}$ qualunque siano $i = 1, ...k$ e $j = 1, ..., n$, e il vettore aleatorio $\boldsymbol{Z}'$ è definito da

$$\boldsymbol{Z}' = A'\boldsymbol{Y},$$

allora, chiaramente,

$$Z'_i = (A'\boldsymbol{Y})_i = Z_i = (A\boldsymbol{Y})_i, \quad \text{per } i = 1, ...k.$$

Se inoltre a'_{hj} per $h = k+1, ...n$ e $j = 1, ..., n$ sono presi in modo che il vettore $(Z'_1, \cdots, Z'_k) = (Z_1, \cdot, Z_k)$ sia indipendente dal vettore $(Z'_{k+1}, \cdots, Z'_n)$, ovvero in modo che

$$0 = E[Z_iZ'_h] = Cov(Z_i, Z'_h) = \sum_{\ell=1}^{n} a_{i,\ell}a'_{h,\ell}$$

per $i = 1, ...k$ e $h = k+1, ...n$, allora si ottiene il risultato voluto.
Nel caso in cui la matrice A sia non singolare ciò è sempre possibile perché i vettori $\boldsymbol{a}_{(i)} = (a_{i1}, a_{i2}, \cdots, a_{in})$ sono linearmente indipendenti e quindi basta trovare $n-k$ vettori $\boldsymbol{a}'_{(h)} = (a'_{i1}, a'_{h2}, \cdots, a'_{hn})$ ortogonali allo spazio vettoriale k-dimensionale $span(\boldsymbol{a}_{(i)}, i = 1, \cdot, k)$.

Vale poi la pena di ricordare che

$$\mathbb{E}[Y^{2k+1}] = 0,$$

$$\mathbb{E}[Y^{2k}] = (2k-1)!! = (2k-1)(2k-3)\cdots 5\cdot 3\cdot 1,$$

mentre[24] infine

$$\mathbb{E}[|Y|^{2k+1}] = \sqrt{\frac{2}{\pi}}(2k)!! = \sqrt{\frac{2}{\pi}}(2k)(2k-2)\cdots 4\cdot 2 = \sqrt{\frac{2}{\pi}}2^k k!.$$

Prima di dimostrare queste tre uguaglianze si osservi che le ultime due si possono scrivere in modo sintetico come

$$\mathbb{E}[|Y|^n] = C_{((-1)^n)}\,(n-1)!! \qquad C_{(+1)} = 1 \qquad C_{(-1)} = \sqrt{\frac{2}{\pi}}.$$

La prima relazione è banale, per ragioni di simmetria, e permette di ricavare la seconda osservando che

$$\mathbb{E}[\mathrm{e}^{uY}] = \mathrm{e}^{\frac{u^2}{2}} = \sum_{h=0}^{\infty}\frac{1}{h!}\frac{u^2}{2}^h = \sum_{h=0}^{\infty}\frac{1}{h!}\frac{u^{2h}}{2^h}.$$

e d'altra parte, essendo appunto ovviamente $\mathbb{E}[Y^{2k+1}] = 0$,

$$\mathbb{E}[\mathrm{e}^{uY}] = \mathbb{E}[\sum_{k=0}^{\infty}\frac{1}{k!}u^kY^k] = \sum_{h=0}^{\infty}\frac{1}{(2h)!}u^{2h}\mathbb{E}[Y^{2h}]$$

si deve necessariamente avere che i coefficienti delle due serie devono coincidere:

$$\frac{1}{h!}\frac{1}{2^h} = \frac{1}{(2h)!}\mathbb{E}[Y^{2h}],$$

ovvero

$$\begin{aligned}\mathbb{E}[Y^{2h}] &= \frac{(2h)!}{h!2^h} = \frac{2h(2h-1)(2h-2)(2h-3)\cdots 3\cdot 2\cdot 1}{h(h-1)\cdots 3\cdot 2\cdot 1\cdot 2^h}\\ &= \frac{(2h)!!(2h-1)!!}{h!2^h} = \frac{2^h h!\cdot(2h-1)!!}{2^h h!} = (2h-1)!!.\end{aligned}$$

Infine la terza si ricava per integrazione per parti e calcolando a mano che $\mathbb{E}[|Y|] = \sqrt{\frac{2}{\pi}}$.

Concludiamo questo paragrafo con un lemma che riguarda il comportamento asintotico della funzione di sopravvivenza di una gaussiana standard e del modulo di una gaussiana standard.

Lemma 1.4. *Sia Y una gaussiana standard, allora, posto $f_Y(y) = \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}$, si ha, per $x > 0$,*

$$\left(x+\frac{1}{x}\right)^{-1} f_Y(x) \le \mathbb{P}(Y > x) \le \frac{1}{x}\,f_Y(x), \qquad x > 0, \tag{1.15}$$

$$\left(x+\frac{1}{x}\right)^{-1} f_{|Y|}(x) \le \mathbb{P}(|Y| > x) \le \frac{1}{x}\,f_{|Y|}(x), \qquad x > 0, \tag{1.16}$$

$$\mathbb{P}(|Y| > x) \le e^{-\frac{x^2}{2}}, \qquad x > 0. \tag{1.17}$$

[24] Si noti che dalle ultime due relazioni sui momenti si ottiene che

$$\mathbb{E}[|Y|^m] = (m-1)!!C_{(-1)^m}, \quad \text{con } C_{+1} = 1, C_{-1} = \sqrt{\frac{2}{\pi}}.$$

Dimostrazione. La disuguaglianza (1.16) discende immediatamente dalla prima disuguaglianza (1.15), la quale equivale a

$$\left(x+\frac{1}{x}\right)^{-1}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{x^2}{2}}\le\mathbb{P}(Y>x)\le\frac{1}{x}\,\frac{1}{\sqrt{2\pi}}\,e^{-\frac{x^2}{2}},$$

e discende dalla seguente relazione

$$\left(w+\frac{1}{w}\right)^{-1}e^{-\frac{w^2}{2}}\le\int_w^{+\infty}e^{-\frac{z^2}{2}}dz\le\frac{1}{w}\,e^{-\frac{w^2}{2}},\qquad w>0. \tag{1.18}$$

La disuguaglianza destra della (1.18) discende da

$$\int_w^{+\infty}e^{-\frac{z^2}{2}}dz\le\frac{1}{w}\int_w^{+\infty}z\,e^{-\frac{z^2}{2}}dz=\frac{1}{w}\,e^{-\frac{w^2}{2}}\quad,$$

Inoltre

$$\frac{d}{dw}\frac{1}{w}\,e^{-\frac{w^2}{2}}=-\left(1+\frac{1}{w^2}\right)\,e^{-\frac{w^2}{2}}$$

e quindi

$$\frac{1}{w}\,e^{-\frac{w^2}{2}}=\int_w^{+\infty}\left(1+\frac{1}{z^2}\right)\,e^{-\frac{z^2}{2}}dz\le\left(1+\frac{1}{w^2}\right)\int_w^{+\infty}e^{-\frac{z^2}{2}}dz\quad,$$

che prova l'altra disuguaglianza nella (1.18).

Infine, per provare la disuguaglianza (1.17), basta osservare che,

$$\begin{aligned}\mathbb{P}(|Y|>x)&=2\int_x^{+\infty}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{y^2}{2}}dy=2\,e^{-\frac{x^2}{2}}\int_x^{+\infty}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{y^2-x^2}{2}}dy\\&=2\,e^{-\frac{x^2}{2}}\int_x^{+\infty}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{(y+x)(y-x)}{2}}dy=2\,e^{-\frac{x^2}{2}}\int_0^{+\infty}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{(z+2x)z}{2}}dz\qquad\text{(essendo } x>0,)\\&\le 2\,e^{-\frac{x^2}{2}}\int_0^{+\infty}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{z^2}{2}}dz=e^{-\frac{x^2}{2}}.\end{aligned}$$

□

1.7 Appendice: Teorema di rappresentazione di Skorohod

In questa sezione affrontiamo il problema seguente:

Data una funzione F, esiste uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e una variabile aleatoria X, definita su questo spazio, per la quale F è la funzione di distribuzione, cioè $F = F_X$?

Chiaramente F deve soddisfare le proprietà delle funzioni di distribuzione, (ossia le proprietà **0** − **4** di pagina 6). Si può dimostrare che tali proprietà sono sufficienti a individuare una misura di probabilità $\mu = \mu_F$ sui boreliani di $\mathbb{R}^k$, per la quale $F(x) = \mu(-\infty, x]$. Di conseguenza si può prendere come spazio di probabilità $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), \mu_F)$ e come variabile aleatoria l'identità, ossia $X(x_1, \cdots, x_k) = (x_1, \cdots, x_k)$.

Tuttavia c'è un altro spazio in cui costruire tale variabile aleatoria, lo spazio $(0, 1)$ con la misura di Lebesgue sui boreliani di $(0, 1)$.

In questa sezione ci limitiamo al caso unidimensionale, il caso a più dimensioni (e addirittura per successioni di variabili aleatorie viene brevemente considerato nella sottosezione 4.9.1).

Teorema 1.5 (di rappresentazione di Skorohod)**.** *Sia data una funzione F, che verifica le seguenti proprietà:*

P0 *F è a valori in $[0, 1]$;*

P1 *F è non decrescente;*

P2 *F è continua a destra, cioè, per ogni $\varepsilon > 0$,* $\lim_{\varepsilon \to 0} F(t + \varepsilon) = F(t)$;

P3 *F è normalizzata, cioè* $\lim_{t \to -\infty} F(t) = 0$; $\lim_{t \to +\infty} F(t) = 1$.

Sia $\varphi : (0, 1) \to \mathbb{R}$ definita da

$$\varphi(u) := \inf\{y : \ F(y) \geq u\}.$$

Allora φ è boreliana, inoltre la variabile aleatoria

$$\begin{array}{rcl} X : (0,1) & \to & \mathbb{R} \\ \omega & \mapsto & X(\omega) = \varphi(\omega) \end{array}$$

definita nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P}) \equiv ((0, 1), \mathcal{B}(0, 1), \lambda)$ con λ la misura di Lebesgue, ha funzione di distribuzione F, ovvero $F_X(x) = F(x)$.

Dimostrazione. La dimostrazione è basata sul fatto che

$$\{u \in (0, 1) : \ \varphi(u) \leq x\} = \{u \in (0, 1) : \ u \leq F(x)\} \tag{1.19}$$

per ogni $x \in \mathbb{R}$. Dalla precedente affermazione segue infatti che:

(i) φ è misurabile rispetto a $\mathcal{B}(0, 1)$;

(ii) Per ogni $x \in \mathbb{R}$ risulta

$$\begin{aligned} F_X(x) &= \mathbb{P}(\omega \in \Omega : \ X(\omega) \in (-\infty, x]) = \lambda(\{u \in (0, 1) : \ \varphi(u) \leq x\}) && \text{(per definizione di } X \text{ e di } (\Omega, \mathcal{F}, P)) \\ &= \lambda(\{u \in (0, 1) : \ u \leq F(x)\}) && \text{(per l'affermazione (1.19))} \\ &= \lambda\left((0, 1) \cap (-\infty, F(x)]\right) = \begin{cases} \lambda\left((0, F(x)]\right), & \text{se } F(x) < 1, \\ \lambda(0, 1), & \text{se } F(x) = 1, \end{cases} \end{aligned}$$

ovvero

$$F_X(x) = F(x).$$

Si tratta dunque di provare l'affermazione (1.19). Dimostriamo innanzitutto che

$$\varphi(u) = \min\{y : F(y) \geq u\}, \qquad \text{cioè} \qquad F(\varphi(u)) \geq u. \tag{1.20}$$

Infatti, essendo $\varphi(u) = \inf\{y : F(y) \geq u\}$, esiste una successione $\{y_n\}_{n=0}^{\infty}$ tale che:
(a) $F(y_n) \geq u$ per ogni n, e quindi $y_n \geq \varphi(u)$,
(b) $\{y_n\}$ tende a $\varphi(u)$ per $n \to \infty$.

Allora, poiché F è continua a destra,

$$F(\varphi(u)) = \lim_{n\to\infty} F(y_n) \geq u.$$

Possiamo ora mostrare l'uguaglianza in (1.19)

- Prima facciamo vedere che $\{u \in (0,1) : \varphi(u) \leq x\} \subseteq \{u \in (0,1) : u \leq F(x)\}$, mostrando che, per ogni u in $(0,1)$

  $$\text{se } \varphi(u) \leq x \quad \text{allora} \quad u \leq F(x).$$

  E infatti, poiché F è non decrescente, se $\varphi(u) \leq x$, allora $F(\varphi(u)) \leq F(x)$ e quindi per la (1.20)

  $$u \leq F(\varphi(u)) \leq F(x).$$

- Proviamo ora l'inclusione opposta $\{u \in (0,1) : \varphi(u) \leq x\} \supseteq \{u \in (0,1) : u \leq F(x)\}$, mostrando che, per ogni u in $(0,1)$,

  $$\text{se } u \leq F(x), \quad \text{allora} \quad \varphi(u) \leq x.$$

  Infatti, se $u \leq F(x)$, allora $x \in \{y : F(y) \geq u\}$ e quindi $\inf\{y : F(y) \geq u\} \leq x$, cioè $\varphi(u) \leq x$.

Prima di terminare la dimostrazione lasciamo al lettore il compito di osservare che fino ad ora non abbiamo (esplicitamente) usato la proprietà **P3** di normalizzazione. Tuttavia tale proprietà serve per garantire che la funzione φ sia a valori reali. □

Dall'affermazione (1.19) segue anche il seguente Corollario.

Corollario 1.6. *Sia U una v.a. uniformemente distribuita*[25] *in* $(0,1)$. *Allora $X := \varphi(U)$ ha distribuzione F*

Dimostrazione. Infatti

$$\begin{aligned} \mathbb{P}(X \leq x) &= \mathbb{P}(\varphi(U) \leq x) = \mathbb{P}(U \in \{u \in (0,1) : \varphi(u) \leq x\}) && \text{(per l'affermazione (1.19))} \\ &= \mathbb{P}(U \in \{u \in (0,1) : u \leq F(x)\}) = F(x). && \text{poiché } U \sim \text{Unif}(0,1) \end{aligned}$$

□

Osservazione 1.3 (spazi completi)**.** *Terminiamo questa sezione con una osservazione importante, che riguarda la possibilità di considerare spazi di probabilità completi, cioè di spazi che contengono anche gli insiemi trascurabili*

$$\mathcal{N} = \{A \subset \Omega : \forall \epsilon \exists B_\epsilon \in \mathcal{F}, \text{ tale che } A \subseteq B_\epsilon, \text{ e } \mathbb{P}(B_\epsilon) \leq \epsilon\},$$

[25] Ricordiamo che $U \sim$Unif$(0,1)$ è una v.a. che ha densità

$$f(t) := \begin{cases} 1, & 0 < t < 1 \\ 0, & \text{altrove} \end{cases}$$

che è la derivata della funzione

$$F(t) \equiv F_U(t) = \begin{cases} 0, & t < 0 \\ t, & 0 \leq t \leq 1 \\ 1, & t > 1 \end{cases}$$

Per una funzione con questa densità vale

$$\mathbb{P}(U \in [a,b]) = \int_a^b dt = b - a$$

quantità che dipende solo dall'ampiezza dell'intervallo, il che spiega la dizione *uniforme*

ossia i sottoinsiemi degli insiemi di misura nulla: infatti, se A è trascurabile allora esiste una successione di eventi $B_{1/n}$ tali che, per ogni n, $B_{1/n}$ contiene A e con probabilità minore uguale a $1/n$. Non si lede in generalità a supporre che $B_{1/n}$ sia una successione monotona. Di conseguenza $A \subseteq B := \cap_{n=1}^{\infty} B_{1/n}$ e $\mathbb{P}(B) = 0$.

*È noto che la misura di Lebesgue si può costruire (estendendo la misura definita sull'algebra delle unioni finite di intervalli) su spazi completi, e sulla σ-algebra $\mathcal{L}(0,1)$ degli insiemi Lebesgue misurabili, che è appunto il completamento della σ-algebra $\mathcal{B}(0,1)$ dei boreliani. Tutte le funzioni boreliane sono ovviamente $\mathcal{L}$-misurabili (cioè misurabili secondo Lebesgue). Quindi la variabile aleatoria definita nel teorema di Skorohod è ancora misurabile se consideriamo $\big((0,1), \mathcal{L}(0,1), \lambda\big)$, invece di $\big((0,1), \mathcal{B}(0,1), \lambda\big)$, e quindi possiamo affermare che il teorema di Skorohod assicura che, data F che soddisfa le proprietà **P0 - P3** esiste uno spazio completo, dove è possibile definire una variabile aleatoria X con $F_X = F$.*

1.8 Appendice: Spazi di variabili aleatorie

Sia dato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$. L'insieme delle variabili aleatorie X, che sono $\mathcal{F}$-misurabili ed integrabili[26], cioè per le quali $\mathbb{E}[|X|] < \infty$, forma uno **spazio vettoriale reale**: se $\mathbb{E}[|X_i|] < \infty$, per $i = 1, 2$, con X_i $\mathcal{F}$-misurabili, allora la variabile aleatoria $a_1X_1 + a_2X_2$ è ancora $\mathcal{F}$-misurabile e inoltre

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

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

Questo spazio differisce dall'usuale spazio $L^1(\Omega, \mathcal{F}, \mathbb{P})$ dell'analisi, solo in quanto in quest'ultimo spazio si considerano equivalenti due variabili aleatorie X ed Y se e solo se $\mathbb{P}(X = Y) = 1$. Lo spazio $L^1(\Omega, \mathcal{F}, \mathbb{P})$ è uno **spazio metrico completo e separabile** rispetto alla distanza

$$d_{L^1}(X, Y) = \mathbb{E}[|X - Y|]$$

(cioè si passa alle classi di equivalenza modulo la relazione di equivalenza[27]: $X \sim Y$ se e solo se $\mathbb{E}[|X - Y|] = 0$, perché altrimenti $d(X, Y)$ non gode della proprietà delle metriche che $d(X, Y) = 0$ se e **solo se** $X = Y$).

Anche l'insieme delle variabili aleatorie X, che sono $\mathcal{F}$-misurabili e ***quadrato integrabili***, cioè per le quali $\mathbb{E}[|X|^2] < \infty$, forma uno **spazio vettoriale reale** che si indica con $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. Come prima passando alle classi di equivalenza[28] si ottiene l'usuale spazio $L^2(\Omega, \mathcal{F}, \mathbb{P})$, che inoltre è uno spazio metrico completo e separabile rispetto alla distanza

$$d^2_{L^2}(X, Y) = \mathbb{E}[|X - Y|^2]$$

(modulo passare a classi di equivalenza $X \sim Y$ se e solo se $\mathbb{E}[|X - Y|^2] = 0$). Inoltre $L^2(\Omega, \mathcal{F}, \mathbb{P})$ è **uno spazio di Hilbert**: questo significa che è possibile introdurre un **prodotto scalare**

$$< X_1, X_2 >:= \mathbb{E}[X_1X_2],$$

e definire[29] $d^2_{L^2}(X, Y) =< X - Y, X - Y >= \mathbb{E}[|X - Y|^2]$. Si noti che, per la **disuguaglianza di Cauchy**,

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

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

Ovviamente si ha

$$\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P}) \subseteq \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P}),$$

ed analogamente

$$L^2(\Omega, \mathcal{F}, \mathbb{P}) \subseteq L^1(\Omega, \mathcal{F}, \mathbb{P}),$$

Infatti se X è di quadrato integrabile allora se X è integrabile:

$$\mathbb{E}[X^2] < \infty \Rightarrow \mathbb{E}[|X|] < \infty,$$

come si vede immediatamente, ad esempio, con la disuguaglianza di Cauchy:

$$\mathbb{E}[|X|] = \mathbb{E}[1|X|] \leq \big(\mathbb{E}[1^2]\big)^{1/2}\big(\mathbb{E}[X^2]\big)^{1/2} = \big(\mathbb{E}[X^2]\big)^{1/2},$$

[26] Ovvero, in termini analitici,

$$\mathbb{E}[|X|] = \int_\Omega |X(\omega)|\,\mathbb{P}(d\omega) =< \infty.$$

[27] Se $\mathbb{E}[|X|] < \infty$ e $\mathbb{E}[|Y|] < \infty$, allora $\mathbb{P}(X = Y) = 1$ se e solo se $\mathbb{E}[|X - Y|] = 0$.

[28] Se $\mathbb{E}[|X|^2] < \infty$ e $\mathbb{E}[|Y|^2] < \infty$, allora $\mathbb{P}(X = Y) = 1$ se e solo se $\mathbb{E}[|X - Y|^2] = 0$.

[29] Va ricordato che è usuale indicare $d^2_{L^2}(X, Y) =< X - Y, X - Y >= \mathbb{E}[|X - Y|^2]$ con $d^2_{L^2}(X, Y) = \|X - Y\|^2_{L^2}$, o equivalentemente

$$\|X\| \,(= \|X\|_{L^2}) := \big(\mathbb{E}[|X|^2]\big)^{1/2}.$$

oppure direttamente considerando che $|x| \leq 1 + x^2$, e per la **proprietà di monotonia** del valore atteso

$$\mathbb{E}[|X|] \leq 1 + \mathbb{E}[X^2].$$

1.8.1 Sottospazi dello spazio delle v.a. di quadrato integrabile †

Sia $\mathcal{G}$ una sotto-σ-algebra di $\mathcal{F}$, ovvero $\mathcal{G} \subseteq \mathcal{F}$: se X è $\mathcal{G}$-misurabile, allora X è anche $\mathcal{F}$-misurabile[30].

Di conseguenza[31]

$$\mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P}) \subseteq \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P}),$$

ed analogamente[32]

$$L^2(\Omega, \mathcal{G}, \mathbb{P}) \subseteq L^2(\Omega, \mathcal{F}, \mathbb{P}),$$

Nel Capitolo 2 sui valori attesi condizionali, per ogni variabile aleatoria $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ verrà definita la variabile aleatoria $\mathbb{E}[X|\mathcal{G}]$, sostanzialmente come la proiezione di X sul sottospazio chiuso $L^2(\Omega, \mathcal{G}, \mathbb{P})$, ovvero come quella variabile aleatoria[33] $\tilde{X} \in L^2(\Omega, \mathcal{G}, \mathbb{P})$ per la quale

$$\mathbb{E}[(X - \tilde{X})^2] = \min_{Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})} \mathbb{E}[(X - Z)^2]. \tag{1.21}$$

In particolare la proiezione sullo spazio generato dalla variabile aleatoria costante $\mathbf{1}$, $\omega \mapsto \mathbf{1}(\omega) = 1$, cioè prendendo $\mathcal{G} = \{\emptyset, \Omega\}$ si ottiene il valore atteso.

Nel caso in cui $\mathcal{G} = \sigma(Y)$, tenendo presente che le variabili aleatorie $\sigma(Y)$-misurabili sono le variabili aleatorie $Z = g(Y)$, con g boreliana, allora il valore atteso condizionato $\mathbb{E}[X|\sigma(Y)]$, secondo la precedente definizione (1.21) (si veda anche la successiva Definizione 2.2), è quella variabile aleatoria $\tilde{X} = \tilde{g}(Y)$, tale che

$$\mathbb{E}\big[\big(X - \tilde{g}(Y)\big)^2\big] = \min_{g \text{ boreliane}} \mathbb{E}\big[\big(X - g(Y)\big)^2\big],$$

dove il minimo va preso rispetto alle funzioni g per le quali $\mathbb{E}\big[|g(Y)|^2\big]$ è finito. In tale caso la media condizionata viene indicata con $\mathbb{E}[X|Y] = \tilde{g}(Y)$, invece che con $\mathbb{E}[X|\sigma(Y)]$.

1.8.2 Regressione lineare

Siano X ed Y due variabili aleatorie di quadrato integrabile. Il problema di trovare la retta di regressione lineare di X rispetto ad Y corrisponde al problema di trovare il punto di minimo, tra tutte le funzioni affini $a + \boldsymbol{b} \cdot y$ del valore atteso del quadrato della differenza tra X, ovvero trovare α e $\boldsymbol{\beta}$ tali che

$$\mathbb{E}\big[\big(X - (\alpha + \boldsymbol{\beta} \cdot Y)\big)^2\big] = \min_{a, \boldsymbol{b}} \mathbb{E}\big[\big(X - (a + \boldsymbol{b} \cdot Y)\big)^2\big].$$

[30] Infatti se la controimmagine di ogni aperto è un insieme (evento) di $\mathcal{G}$, allora ovviamente la controimmagine di ogni aperto è un insieme (evento) di $\mathcal{F}$.

[31] Si noti che, per maggiore precisione, si dovrebbe scrivere $\mathcal{L}^2(\Omega, \mathcal{G}, \widehat{\mathbb{P}})$ oppure $L^2(\Omega, \mathcal{G}, \widehat{\mathbb{P}})$ dove $\widehat{\mathbb{P}} = \mathbb{P}|_{\mathcal{G}}$, cioè la restrizione di $\mathbb{P}$ a $\mathcal{G}$.

[32] Bisogna però fare una precisazione:
se lo spazio $L^2(\Omega, \mathcal{F}, \mathbb{P})$ viene considerato come spazio di Hilbert, siamo passati alle classi di equivalenza delle variabili aleatorie $\mathcal{F}$-misurabili e tali che $P(X \neq X') = 0$, ossia che differiscono su un insieme di $\mathcal{N}_{\mathcal{F}}$, la classe degli insiemi $\mathcal{F}$-misurabili di $\mathbb{P}$-probabilità nulla. Invece lo spazio $L^2(\Omega, \mathcal{G}, \mathbb{P})$, [o meglio $L^2(\Omega, \mathcal{G}, \widehat{\mathbb{P}})$, dove $\widehat{\mathbb{P}}$ è la misura $\mathbb{P}$ ristretta a $\mathcal{G}$,] è lo spazio delle classi di equivalenza delle variabili aleatorie $\mathcal{G}$-misurabili e tali che $P(X \neq X') = 0$, ossia che differiscono su un insieme di $\mathcal{N}_{\mathcal{G}}$, la classe degli insiemi $\mathcal{G}$-misurabili di $\mathbb{P}$-probabilità nulla. Quindi il passaggio agli spazi di Hilbert si può fare soltanto se gli insiemi $\mathcal{N}_{\mathcal{F}}$ e $\mathcal{N}_{\mathcal{G}}$ coincidono. Ciò è sicuramente vero se $\mathcal{N}_{\mathcal{F}} \subset \mathcal{G}$.

[33] Sarebbe più corretto dire quella classe di equivalenza di variabili aleatorie.

Si noti che anche qui sia ha la proiezione su un sottospazio di $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Questa volta però si tratta di un sottospazio di dimensione 2:

$$V = \{Z = a + bY;\ a, b \in \mathbb{R}\}$$

La soluzione è data in analogia con il caso delle proiezioni sui sottospazi vettoriali in $\mathbb{R}^d$. Si denota con $\hat{X} = \alpha + \boldsymbol{\beta} \cdot Y$ la variabile aleatoria che si cerca. Allora deve essere

$$\frac{\hat{X} - \mathbb{E}(X)}{\sigma_X} = \rho_{X,Y} \frac{Y - \mathbb{E}(Y)}{\sigma_Y} \tag{1.22}$$

dove $\rho_{X,Y}$ è il ***coefficiente di correlazione***, cioè

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X\, \sigma_Y}$$

L'analogia sta nel fatto che $\frac{Y - \mathbb{E}(Y)}{\sigma_Y}$ rappresenta il vettore unitario con direzione parallela ad Y, e invece $\rho_{X,Y}$ rappresenta il coseno tra l'angolo formato tra due vettori[34]:

$$\rho_{X,Y} = \frac{< X - \mathbb{E}(X), Y - \mathbb{E}(Y) >}{\|X - \mathbb{E}(X)\|\, \|Y - \mathbb{E}(Y)\|}$$

Dalla (1.22) si possono ottenere i valori di α e di $\boldsymbol{\beta}$. In particolare

$$\hat{X} = \frac{\rho_{X,Y}}{\sigma_Y} \sigma_X\, (Y - \mathbb{E}(Y)) + \mathbb{E}(X),$$

da cui immediatamente, tenendo conto della definizione di coefficiente di correlazione

$$\boldsymbol{\beta} = \frac{\rho_{X,Y}}{\sigma_Y} \sigma_X = \frac{Cov(X,Y)}{\sigma_Y^2}$$

Vedremo nel capitolo sui valori attesi condizionali che, per variabili aleatorie (X, Y) congiuntamente gaussiane il valore atteso condizionale $\mathbb{E}[X|Y]$ di X dato Y è una funzione affine di Y, e che quindi coincide anche con la retta di regressione.

1.9 Appendice: Approccio soggettivista alla probabilità

In questo paragrafo vogliamo illustrare come con l'approccio soggettivista, si possa arrivare agli assiomi della probabilità (con la additività semplice) attraverso una opportuna definizione di valore atteso e di probabilità come prezzo, e attraverso opportune regole di linearità e di coerenza.

Daremo due definizioni che sostanzialmente si equivalgono, una con un linguaggio più neutrale e generale ed un'altra con un linguaggio più economico.

Definizione 1.11 (valore atteso)**.** *Il valore atteso $\mathbb{E}(X)$ di una variabile aleatoria X è quel valore certo c che sono disposto a scambiare con il valore aleatorio dato proprio da X, nel senso che per me è indifferente ricevere c o X. La probabilità $\mathbb{P}(A)$ di un evento A è definito come il valore atteso della variabile aleatoria indicatrice di A, ovvero la variabile aleatoria $X = \mathbb{I}_A$ che vale 1 se si verifica A e vale 0 altrimenti.*

La seconda definizione vede X come valore aleatorio che si ottiene come esito di una scommessa, mentre c come prezzo da pagare per prendere parte alla scommessa (o al gioco).

[34] Si ricordi che se $\boldsymbol{u}$ e $\boldsymbol{v}$ sono due vettori non nulli, allora

$$\cos(\widehat{\boldsymbol{u}, \boldsymbol{v}}) = \frac{\boldsymbol{u} \cdot \boldsymbol{v}}{\|\boldsymbol{u}\| \cdot \boldsymbol{v}\|}$$

e che con la notazione $\|X\| := \left(\mathbb{E}[|X|^2]\right)^{1/2}$ si ha $Var(X) = \|X - \mathbb{E}(X)\|^{1/2}$.

Definizione 1.12 (valore atteso come prezzo)**.** *Il valore atteso di X è quel valore certo c che si è disposti a pagare per effettuare la scommessa in cui si ottiene il valore aleatorio X, con l'accordo che si è disposti indifferentemente a prendere il ruolo sia dello scommettitore che del banco.*

Regola di linearità[35]: Siano X ed Y due variabili aleatorie allora

i $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$

ii $\mathbb{E}(\alpha X + \beta Y) = \alpha\,\mathbb{E}(X) + \beta\,\mathbb{E}(Y)$, per ogni α e β costanti reali

La regola di coerenza seguente è basata sul fatto alla fine della scommessa ricevo X e pago c e quindi alla fine ho $X - c$, se ho il ruolo dello scommettitore, mentre ricevo c e pago X se ho il ruolo del banco (broker).

Regola di coerenza: Non è possibile che $X - c$ sia certamente positivo o certamente negativo[36].

La giustificazione della precedente regola di coerenza sta nel fatto che se $X - c$ ha segno costante, non si troverebbe nessuno disposto a prendere il ruolo dello scommettitore, se $X - c$ fosse certamente negativo, e analogamente non si troverebbe nessuno disposto a prendere il ruolo del banco se invece $X - c$ fosse certamente positivo.

A titolo di esempio mostriamo come dalla regola di coerenza si ottenga che se X è una funzione indicatrice di un evento A, ovvero X assume solo i valori 0 ed 1 (e allora $A = \{X = 1\}$) e, posto per definizione $\mathbb{P}(A) = \mathbb{E}(X)$, allora necessariamente deve valere $\mathbb{P}(A) \in [0, 1]$.

La seguente tavola illustra i ricavi possibili se $\mathbb{P}(A) = p$ è il prezzo da pagare per la scommessa in cui si vince 1 se si verifica A (e quindi si ottiene globalmente $1 - p$), mentre se si verifica A^c non si vince nulla (e quindi si ottiene $-p$)

evento	A	A^c
pagamento	$-p$	$-p$
vincita	1	0
ricavo	$1-p$	$-p$

Per la regola di coerenza si ottiene che ***non può essere***

$$\Big(1 - p < 0 \quad \text{e} \quad -p < 0\Big) \quad \text{oppure} \quad \Big(1 - p > 0 \quad \text{e} \quad -p > 0\Big),$$

Equivalentemente ***non può essere***

$$\Big(1 < p \quad \text{e} \quad 0 < p\Big) \quad \text{oppure} \quad \Big(1 > p \quad \text{e} \quad 0 > p\Big),$$

cioè ***non può essere***

$$1 < p \quad \text{oppure} \quad 0 > p,$$

[35] In realtà, per ottenere regola di linearità **[i]**, basterebbe la regola di coerenza (data immediatamente dopo la regola di linearità) e aggiungere l'ipotesi che sia sempre possibile trovare qualcuno disposto a scommettere su ciascuna singola scommessa:
se il prezzo di due scommesse insieme fosse maggiore della somma dei prezzi delle due scommesse separatamente, allora converrebbe accettare (comprare) due singole scommesse e prendere la posizione del banco (vendere) la scommessa relativa alla somma

se invece il prezzo della somma fosse minore della somma dei prezzi, allora converrebbe vendere le due scommesse separatamente e invece accettare la scommessa. Per capire meglio si veda la seguente tabella: se non ci possono essere arbitraggi, allora necessariamente deve accadere che $c - c_1 - c_2 = 0$

tipo di scommessa	solo la 1	solo la 2	1 e 2, ma con il ruolo opposto	le precedenti insieme
vincita	X_1	X_2	$-X = -(X_1 + X_2)$	$X_1 + X_2 - X = 0$
pagamento	$-c_1$	$-c_2$	$+c$	$c_1 - c_2 + c$
ricavo	$X_1 - c_1$	$X_2 - c_2$	$c - X$	$X_1 - c_1 + X_2 - c_2 + c - X = c - c_1 - c_2$

Per la regola di linearità **[ii]** si può invece pensare all'assenza di sconti, nel senso che se si scommette αX si paga esattamente α volte il prezzo della scommessa X, nessun *prendi tre paghi due*!!!

[36] Interpretando $X - c$ come il guadagno di chi ha pagato c per ottenere in cambio la cifra aleatoria X, e quindi $c - X$ come il guadagno del venditore, la Regola di coerenza afferma che non può esserci un arbitraggio (forte) ne' per il compratore, ne' per il venditore.

e quindi in altre parole ***deve essere necessariamente***

$$0 \leq p = \mathbb{P}(A) \leq 1,$$

che corrisponde alla richiesta degli assiomi che la probabilità sia a valori in $[0, 1]$.

Nel caso particolare in cui $A = \Omega$ è l'evento certo la tabella si precedente si riduce a

evento	Ω
pagamento	$-p$
vincita	1
ricavo	$1-p$

Non potendo essere ne' $1 - p < 0$ ne' $1 - p > 0$, ovvero non potendo essere ne' $1 < p$ ne' $1 > p$, deve necessariamente essere $p = \mathbb{P}(\Omega) = 1$, che è un altro degli assiomi delle probabilità.

Se inoltre ho n scommesse relative ad n eventi A_1, A_2, ..., A_n incompatibili (cioè $A_i \cap A_j = \emptyset$ per $i \neq j$) allora la scommessa su $A = \bigcup_{i=1}^n A_i$ equivale[37] alla somma delle singole n scommesse su A_i, in quanto grazie all'incompatibilità degli A_i in entrambe le scommesse ricevo 1 se e solo se si verifica uno degli A_i, mentre altrimenti ottengo 0. Quindi la linearità dei prezzi, più il fatto implicito che se due scommesse danno luogo alla stessa vincita, allora devono avere lo stesso prezzo, si ottiene che

$$\mathbb{P}(A) \left(= \mathbb{P}\Big(\cup_{i=1}^n A_i \Big) \right) = \sum_{i=1}^n \mathbb{P}(A_i),$$

che corrisponde all'assioma dell'additività finita.

Riassumendo abbiamo mostrato come si possano riottenere gli assiomi della probabilità (con l'esclusione della $\sigma-$additività) con la definizione di valore atteso come prezzo di scommesse aleatorie, con la ***regola di coerenza*** che corrisponde all'***assenza*** di ***opportunità di arbitraggio***.

Va notato che l'ipotesi di linearità corrisponde all'assenza di costi di transazione: quando ci sono costi di transazione allora comprare all'ingrosso (ovvero fare un'unica scommessa su $A = \cup_{i=1}^n A_i$) è in genere più conveniente che comprare al dettaglio (ovvero fare n scommesse separate su ciascun A_i, per $i = 1, \ldots, n$).

Terminiamo queste osservazioni sulle probabilità soggettive dando l'***interpretazione del valore atteso condizionato***[38] ***ad un evento*** A di una variabile aleatoria X come il valore certo c_A che si è disposti a scambiare con X, tenendo presente che lo scambio avviene ***solo se si verifica*** l'eventualità rappresentata dall'evento A, ovvero con l'intesa che se l'evento A non si verifica, allora non viene effettuato alcuno scambio[39].

[37] Ovvero se $A = \bigcup_{i=1}^n A_i$ e se $A_i \cap A_j = \emptyset$ per $i \neq j$, allora

$$\mathbb{I}_A = \sum_{i=1}^n \mathbb{I}_{A_i}.$$

[38] Vale un'interpretazione analoga per le probabilità condizionate ad un evento A, prendendo $X = \mathbb{I}_E$.

[39] Come a dire che il contratto o la scommessa non hanno validità se non si verifica la condizione A.

Capitolo 2

Valori attesi e probabilità condizionali

2.1 Definizioni

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

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

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

$$\mathbb{E}[X|A] = \mathbb{E}[\widetilde{X}|A], \qquad \forall A \in \mathcal{G}, \text{ con } \mathbb{P}(A) > 0. \tag{2.1}$$

dove

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

Si noti che (2.1) equivale a

$$\mathbb{E}[X\mathbb{I}_A] = \mathbb{E}[\widetilde{X}\mathbb{I}_A], \qquad \forall A \in \mathcal{G}, \text{ con } \mathbb{P}(A) > 0. \tag{2.2}$$

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

Si noti inoltre che la precedente (2.2) per $A = \Omega$, implica che, se una tale variabile aleatoria $\widetilde{X}$ esiste, allora $\mathbb{E}[\widetilde{X}] = \mathbb{E}[X]$.

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

$$\mathbb{E}[(X - Z)^2] \geq \mathbb{E}[(X - \widetilde{X})^2], \qquad \forall Z \; \mathcal{G} - \text{misurabile}. \tag{2.3}$$

Prima di tutto, dobbiamo trovare sotto quali condizioni le pre-definizioni siano ben poste, cioè, in questo caso, che esista una variabile $\widetilde{X}$ per cui valga la (2.1) (o equivalentemente la (2.2)) oppure valga la (2.3), ed in che senso ne viene individuata una sola. Notiamo che intanto ci sono delle condizioni necessarie da rispettare: chiaramente, per la pre-definizione 1, è necessario che X sia una variabile aleatoria integrabile[1], cioè $\mathbb{E}[|X|] < \infty$, mentre, per la

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

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

pre-definizione 2, è necessario richiedere che X sia di quadrato integrabile [2], cioè $\mathbb{E}[|X|^2] < \infty$, ed inoltre anche la (2.3) va modificata, nel senso che è necessario richiedere che anche Z sia di quadrato integrabile, oltre che $\mathcal{G}$-misurabile[3]. Inoltre è chiaro che se $\widetilde{X}'$ è una variabile aleatoria $\mathcal{G}$-misurabile, che differisce da $\widetilde{X}$ a meno di un insieme di misura nulla, anche $\widetilde{X}'$ gode della proprietà (2.2) o (2.3) rispettivamente e quindi non si individua una sola variabile aleatoria, ma una classe di variabili aleatorie. Queste modifiche in realtà sono sufficienti a garantire che le due pre-definizioni diventino due definizioni.

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

$$\mathbb{E}[X\mathbb{I}_A] = \mathbb{E}[\widetilde{X}\mathbb{I}_A], \qquad \forall A \in \mathcal{G}. \tag{2.4}$$

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

Definizione 2.2 (valore atteso condizionale 2)**.** *Sia X una variabile aleatoria di quadrato integrabile, cioè $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. Sia data una variabile aleatoria (di quadrato integrabile) $\widetilde{X}$, $\mathcal{G} - $ misurabile, per la quale valga*[4]

$$\mathbb{E}[(X - Z)^2] \geq \mathbb{E}[(X - \widetilde{X})^2], \forall Z \quad \mathcal{G} - \textit{misurabile, con } \mathbb{E}[Z^2] < \infty. \tag{2.7}$$

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

Rimandiamo la verifica che effettivamente la Definizione 2.1 e la Definizione 2.2 sono ben poste a dopo aver trattato alcuni esempi, e anticipiamo che, se X è di quadrato integrabile[5], allora le Definizioni 2.1 e 2.2 sono equivalenti, e quindi non c'è ambiguità nello scegliere una definizione o l'altra e che in seguito, riferendoci a $\mathbb{E}[X \mid \mathcal{G}]$, intenderemo riferirci alla Definizione 2.1, che valendo per variabili aleatorie integrabili, è più generale. Non c'è quindi ambiguità nella seguente definizione che viene data in analogia con $\mathbb{P}(A) = \mathbb{E}[\mathbb{I}_A]$.

L'insieme di tali variabili aleatorie è indicato con $L^1(\Omega, \mathcal{F}, \mathbb{P})$ che è uno spazio metrico completo e separabile rispetto alla distanza $d(X, Y) = \mathbb{E}[|X - Y|]$ (modulo passare a classi di equivalenza $X \sim Y$ se e solo se $\mathbb{E}[|X - Y|] = 0$, perché altrimenti $d(X, Y)$ non gode della proprietà delle metriche che $d(X, Y) = 0$ se e **solo se** $X = Y$).

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

$$< X_1, X_2 > := \mathbb{E}[X_1 X_2],$$

e definire $d^2_{L^2}(X, Y) =< X - Y, X - Y >= \mathbb{E}[|X - Y|^2]$. Si noti che, per la disuguaglianza di Cauchy,

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

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

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

[4]La proprietà (2.7) è equivalente alla proprietà

$$\mathbb{E}[(X - \widetilde{X})^2] = \min_{Z \in L^2(\Omega, \mathcal{G}, \widehat{\mathbb{P}})} \left(\mathbb{E}[(X - Z)^2]\right), \tag{2.5}$$

ovvero

$$d^2_{L^2}(X, \widetilde{X}) = \min_{Z \in L^2(\Omega, \mathcal{G}, \widehat{\mathbb{P}})} d^2_{L^2}(X, Z). \tag{2.6}$$

(vedere le note precedenti per la definizione di $L^2(\Omega, \mathcal{G}, \widehat{\mathbb{P}})$ e di $d^2_{L^2}$)

[5]Si ricordi che se X è di quadrato integrabile allora se X è integrabile:

$$\mathbb{E}[X^2] < \infty \Rightarrow \mathbb{E}[|X|] < \infty,$$

come si vede immediatamente, ad esempio, con la disuguaglianza di Cauchy:

$$\mathbb{E}[|X|] = \mathbb{E}[1|X|] \leq \left(\mathbb{E}[1^2]\right)^{1/2}\left(\mathbb{E}[X^2]\right)^{1/2} = \left(\mathbb{E}[X^2]\right)^{1/2},$$

oppure direttamente considerando che $|x| \leq 1 + x^2$, e per la proprietà di monotonia del valore atteso

$$\mathbb{E}[|X|] \leq 1 + \mathbb{E}[X^2].$$

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

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

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

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

2.2 Esempi

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

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

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

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

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

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

dove c_m sono costanti reali, e gli insiemi $\mathcal{G}$-misurabili sono le unioni di sottofamiglie numerabili di elementi della partizione. Basta quindi calcolare i valori $\widetilde{c}_m$ che caratterizzano $\widetilde{X}$, imponendo la condizione[6] *che*

$$\mathbb{E}[X\mathbb{I}_{H_n}] = \mathbb{E}[\widetilde{X}\mathbb{I}_{H_n}] = \mathbb{E}[\sum_{m=1}^{\infty} \widetilde{c}_m \mathbb{I}_{H_m}\mathbb{I}_{H_n}] = \mathbb{E}[\widetilde{c}_n \mathbb{I}_{H_n}] = \widetilde{c}_n \mathbb{P}[H_n] \tag{2.8}$$

Quindi

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

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

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

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

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

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

otteniamo che una versione[7] *di $\mathbb{P}(B \mid \mathcal{G})(\omega)$ è data da*

$$\mathbb{P}(B \mid \mathcal{G}) = \sum_{n\geq 1}^{*} \frac{\mathbb{E}[\mathbb{I}_B \mathbb{I}_{H_n}]}{\mathbb{P}(H_n)} \mathbb{I}_{H_n} = \sum_{n\geq 1}^{*} \frac{\mathbb{P}(B \cap H_n)}{\mathbb{P}(H_n)} \mathbb{I}_{H_n} = \sum_{n\geq 1}^{*} \mathbb{P}(B \mid H_n)\mathbb{I}_{H_n}. \tag{2.9}$$

[6] È chiaro che la condizione che $\mathbb{E}[X\mathbb{I}_A] = \mathbb{E}[\widetilde{X}\mathbb{I}_A]$ per ogni $A = \bigcup_{n\in I} H_n$, implica la condizione (2.8): basta prendere $A = H_n$. Tuttavia vale anche il viceversa, in quanto

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

[7] La versione di $\mathbb{P}(B \mid \mathcal{G})(\omega)$ data in (2.9) non è in generale una probabilità per ogni ω: ad esempio se $\omega \in H_m$ e $\mathbb{P}(H_m) = 0$, allora $\mathbb{P}(\Omega \mid \mathcal{G})(\omega) = 0$ invece di 1.

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

È interessante osservare che, se $\mathcal{F}$ è a sua volta generato da una partizione $\{K_\ell,\, \ell \in \mathbb{N}\}$, più fine[8] di $\{H_m,\, m \in \mathbb{N}\}$, e nel caso in cui $\mathbb{P}(H_n) > 0$ per ogni $n \in \mathbb{N}$, allora (2.9) definisce una probabilità su $\mathcal{F}$:
per iniziare

$$\mathbb{P}(\Omega \mid \mathcal{G})(\omega) = \sum_{n\geq 1}{}^{*} \frac{\mathbb{E}[\mathbb{I}_\Omega \mathbb{I}_{H_n}(\omega)]}{\mathbb{P}(H_n)} \mathbb{I}_{H_n} = \sum_{n\geq 1} \frac{\mathbb{E}[\mathbb{I}_\Omega \mathbb{I}_{H_n}(\omega)]}{\mathbb{P}(H_n)} \mathbb{I}_{H_n} = \sum_{n\geq 1} \mathbb{I}_{H_n} = 1,$$

inoltre, se $B = \bigcup_{\ell\in I_B} K_\ell$, allora

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

e da questa relazione immediatamente si ricava che, qualunque sia ω, l'applicazione $B \mapsto \mathbb{P}(B \mid \mathcal{G})(\omega)$ definisce una probabilità. Anche nel caso in cui **non** si faccia l'ipotesi che $\mathbb{P}(H_n) > 0$ per ogni $n \in \mathbb{N}$, si può trovare una versione di $\mathbb{P}(B \mid \mathcal{G})(\omega)$ in modo che per ogni ω la funzione

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

sia una probabilità:
fissata a piacere una probabilità[9] $\mathbb{P}^0$ su $\mathcal{F}$, si definisce

$$\mathbb{P}(K_\ell \mid \mathcal{G})(\omega) = \sum_{n\geq 1}{}^{*} \frac{\mathbb{P}(K_\ell \cap H_n)}{\mathbb{P}(H_n)} \mathbb{I}_{H_n}(\omega) + \sum_{n\geq 1}{}^{**} \mathbb{P}^0(K_\ell) \mathbb{I}_{H_n}(\omega) = \tag{2.10}$$

$$= \sum_{n\geq 1}{}^{*} \frac{\mathbb{P}(K_\ell \cap H_n)}{\mathbb{P}(H_n)} \mathbb{I}_{H_n}(\omega) + \mathbb{P}^0(K_\ell) \mathbb{I}_{\{\cup^{**}_{n\geq 1} H_n\}}(\omega). \tag{2.11}$$

e

$$\mathbb{P}(B \mid \mathcal{G})(\omega) := \sum_{\ell\in I_B} \mathbb{P}(K_\ell \mid \mathcal{G})(\omega). \tag{2.12}$$

Si vede immediatamente che la parte a destra di (2.12) è effettivamente una versione di $\mathbb{P}(B \mid \mathcal{G})$, e si vede facilmente che in questo modo, qualunque sia ω, $\mathbb{P}(\cdot \mid \mathcal{G})(\omega)$ definisce una probabilità[10].

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

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

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

[10] Infatti in questo caso è come se avessimo definito una successione $\{p_\ell(\omega); \ell \in \mathbb{N}\}$, con $\sum_{\ell\in\mathbb{N}} p_\ell(\omega) = 1$ per ogni ω

$$\begin{cases} p_\ell(\omega) = \mathbb{P}(K_\ell | H_m) & \text{se } \omega \in H_m, \text{ con } \mathbb{P}(H_m) > 0, \\ p_\ell(\omega) = \mathbb{P}^0(K_\ell) & \text{se } \omega \in H_m, \text{ con } \mathbb{P}(H_m) = 0, \end{cases}$$

e poi avessimo definito, per $B = \bigcup_{\ell\in I}$,

$$\mathbb{P}(B|\mathcal{G})(\omega) = \sum_{\ell\in I} p_\ell(\omega).$$

Inoltre questa probabilità gode della proprietà che se X è $\mathcal{F}$-misurabile, cioè se

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

allora

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

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

Esempio 2.3. *Ritorniamo nel caso dell'Esempio precedente, quando $\mathcal{G} = \sigma(Y)$, e Y è una variabile aleatoria discreta, a valori in $\{y_m;\, m \in \mathbb{N}\}$. Infatti allora*

$$\mathcal{G} = \sigma(\{H_m := Y^{-1}(\{y_m\});\ m \in \mathbb{N}\}),$$

e di conseguenza, se $\mathbb{P}(Y = y_m) > 0$ per ogni $m \in \mathbb{N}$, si ha

$$\mathbb{E}[X \mid \sigma(Y)](\omega) = \sum_{m \in \mathbb{N}} \mathbb{E}[X \mid \{Y = y_m\}] \mathbb{I}_{\{Y = y_m\}}(\omega) = \sum_{m \in \mathbb{N}} \mathbb{E}[X \mid \{Y = y_m\}] \mathbb{I}_{\{y_m\}}(Y(\omega)),$$

con il solito abuso di notazione (il secondo membro è un rappresentante della classe di equivalenza $\mathbb{E}[X \mid \sigma(Y)]$).

Quindi posto

$$\psi(y) = \sum_{m \in \mathbb{N}} \mathbb{E}[X \mid \{Y = y_m\}] \mathbb{I}_{\{y_m\}}(y),$$

e indicato con

$$\mathbb{E}[X \mid \{Y = y\}] = \psi(y),$$

cioè la funzione, che vale $\mathbb{E}[X \mid \{Y = y\}]$, se $y \in \{y_m;\, m \in \mathbb{N}\}$, e zero altrimenti, si ha:

$$\mathbb{E}[X \mid \sigma(Y)](\omega) = \quad \mathbb{E}[X \mid \{Y = y\}]\Big|_{y = Y(\omega)} = \psi(Y(\omega)).$$

Ciò giustifica il fatto che si usa scrivere

$$\mathbb{E}[X \mid \sigma(Y)] = \mathbb{E}[X \mid Y],$$

e ci fa ritrovare il concetto elementare di valore atteso condizionato di una variabile aleatoria discreta X rispetto a una variabile aleatoria discreta Y.

Analogamente a quanto fatto nell'esempio precedente si ottiene che in tale caso, cioè se anche X è una variabile aleatoria discreta a valori in $\{x_n;\, n \in \mathbb{N}\}$, allora

$$\mathbb{E}[X \mid Y](\omega) = \sum_{n \in \mathbb{N}} x_n \mathbb{P}(X = x_n \mid \{Y = y\})\Big|_{y = Y(\omega)}$$

Infine notiamo che è facile ripetere quanto sopra nel caso in cui al posto di X ci sia una variabile aleatoria $Z = h(X)$, integrabile, e ottenere che

$$\mathbb{E}[h(X) \mid Y](\omega) = \sum_{n \in \mathbb{N}} h(x_n) \mathbb{P}(X = x_n \mid \{Y = y\})\Big|_{y = Y(\omega)}$$

Così, ad esempio,

$$\mathbb{P}(\Omega \mid \mathcal{G})(\omega) := \sum_{\ell \in \mathbb{N}} \mathbb{P}(K_\ell \mid \mathcal{G})(\omega) = \begin{cases} \mathbb{P}(\Omega | H_m) = 1 & \text{se } \omega \in H_m, \text{ con } \mathbb{P}(H_m) > 0, \\ \mathbb{P}^0(\Omega) = 1 & \text{se } \omega \in H_m, \text{ con } \mathbb{P}(H_m) = 0. \end{cases}$$

Nel caso in cui ***non*** *si abbia che* $\mathbb{P}(Y = y_m) > 0$ *per ogni* $m \in \mathbb{N}$*, si ottiene, fissata una probabilità* $\mathbb{P}^0$*, come nell'esempio precedente,*

$$\begin{aligned}\mathbb{E}[X \mid \sigma(Y)](\omega) &= \sum_{m\in\mathbb{N}}{}^{*}\mathbb{E}[X \mid \{Y = y_m\}]\mathbb{I}_{\{Y=y_m\}}(\omega) + \sum_{m\in\mathbb{N}}{}^{**}\mathbb{E}^0[X]\mathbb{I}_{\{Y=y_m\}}(\omega)\\ &= \sum_{m\in\mathbb{N}}{}^{*}\mathbb{E}[X \mid \{Y = y_m\}]\mathbb{I}_{\{y_m\}}(Y(\omega)) + \sum_{m\in\mathbb{N}}{}^{**}\mathbb{E}^0[X]\mathbb{I}_{\{y_m\}}(Y(\omega)),\end{aligned}$$

con il solito abuso di notazione (il secondo membro è un rappresentante della classe di equivalenza $\mathbb{E}[X \mid \sigma(Y)]$*).*

Quindi posto

$$\psi(y) = \sum_{m\in\mathbb{N}}{}^{*}\mathbb{E}[X \mid \{Y = y_m\}]\mathbb{I}_{\{y_m\}}(y) + \sum_{m\in\mathbb{N}}{}^{**}\mathbb{E}^0[X]\mathbb{I}_{\{y_m\}}(y),$$

e indicato con

$$\mathbb{E}[X \mid \{Y = y\}] = \psi(y),$$

cioè la funzione, che vale $\mathbb{E}[X \mid \{Y = y\}]$*, se* $y \in \{y_m;\, m \in \mathbb{N}\}$ *e se* $\mathbb{P}(\{Y = y\}) > 0$*, e vale* $\mathbb{E}^0[X]$ *altrimenti, si ha:*

$$\mathbb{E}[X \mid Y](\omega) = \quad \mathbb{E}[X \mid \{Y = y\}]\Big|_{y=Y(\omega)} = \psi(Y(\omega)).$$

Più in generale si ha anche che posto

$$\psi_h(y) = \sum_{m\in\mathbb{N}}{}^{*}\mathbb{E}[h(X) \mid \{Y = y_m\}]\mathbb{I}_{\{y_m\}}(y) + \sum_{m\in\mathbb{N}}{}^{**}\mathbb{E}^0[h(X)]\mathbb{I}_{\{y_m\}}(y),$$

si ottiene che

$$\mathbb{E}[h(X) \mid Y](\omega) = \psi_h(Y(\omega)).$$

Esempio 2.4. *Caso in cui* $\mathcal{G} = \sigma(Y)$*, con* Y *una variabile aleatoria a valori in* $\mathbb{R}^d$*, e* (X, Y) *ammette densità di probabilità congiunta* $f_{X,Y}(x, y)$*. Allora, posto*[11]

$$f_{X|Y}(x|y) = \mathbf{1}_{\{z:f_Y(z)>0\}}(y)\frac{f_{X,Y}(x, y)}{f_Y(y)} + \mathbf{1}_{\{z:f_Y(z)=0\}}(y)f_0(x)$$

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

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

ossia[12]

$$\widetilde{X}(\omega) = \int_{\mathbb{R}} x\left(\mathbf{1}_{\{z:f_Y(z)>0\}}(y)\frac{f_{X,Y}(x, y)}{f_Y(y)}\right)\Big|_{y=Y(\omega)} dx + \int_{\mathbb{R}} x\; \mathbf{1}_{\{z:f_Y(z)=0\}}(y)\big|_{y=Y(\omega)}\; f_0(x)dx,$$

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

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

$$\mathbb{E}[X\mathbb{I}_A] = \mathbb{E}[X\mathbf{1}_B(Y)] = \int_{\mathbb{R}\times\mathbb{R}^d} x\mathbf{1}_B(y)f_{X,Y}(x, y)\; dx\; dy.$$

[11] Indichiamo con $\mathbf{1}_B : \mathbb{R}^d \to \mathbb{R}$ la funzione che assume solo i valori in $\{0, 1\}$, e che vale $\mathbf{1}_B(y) = 1$ se $y \in B$ e $\mathbf{1}_B(y) = 0$ se $y \notin B$. Se $B \in \mathcal{B}(\mathbb{R}^d)$ allora $\mathbf{1}_B$ è una funzione boreliana.

[12] In realtà per poter scrivere la formula esplicita per $\widetilde{X}(\omega)$ è necessario prendere $f_0(x)$ in modo che $\int_{\mathbb{R}} |x|\; f_0(x)dx < \infty$. Inoltre un altro rappresentante per il valore atteso condizionato è $\int_{\mathbb{R}} x\left(\mathbf{1}_{\{z:f_Y(z)>0\}}(y)\frac{f_{X,Y}(x,y)}{f_Y(y)}\right)\Big|_{y=Y(\omega)} dx$ in quanto $\mathbb{P}\big(\mathbf{1}_{\{z:f_Y(z)=0\}}(y)\big|_{y=Y(\omega)} = 0\big) = 1$. Quest'ultima uguaglianza dipende dal fatto che $\mathbf{1}_{\{z:f_Y(z)=0\}}(y)\big|_{y=Y(\omega)} = 0 \Leftrightarrow Y(\omega) \notin \{z : f_Y(z) = 0\}$ e per il suo complementare si ha $\mathbb{P}\big(Y(\omega) \in \{z : f_Y(z) = 0\}\big) = \int_{\{z:f_Y(z)=0\}} f_y(z)\, dz = \int_{\{z:f_Y(z)=0\}} 0\, dz = 0$.

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

$$\begin{aligned}
\mathbb{E}[\widetilde{X}(\omega)\mathbb{I}_A] &= \mathbb{E}[\widetilde{X}(\omega)\mathbf{1}_B(Y)] \\
&= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}} x \frac{f_{X,Y}(x,y)}{f_Y(y)}\ dx \right) \mathbf{1}_B(y) f_Y(y)\ dy
\end{aligned}$$

ovvero, per il Teorema di Fubini[13]*,*

$$\begin{aligned}
\mathbb{E}[\widetilde{X}(\omega)\mathbb{I}_A] &= \int_{\mathbb{R}} \int_{\mathbb{R}^d} x \mathbf{1}_B(y) \frac{f_{X,Y}(x,y)}{f_Y(y)} f_Y(y)\ dx\ dy \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}^d} x \mathbf{1}_B(y) f_{X,Y}(x,y)\ dx\ dy.
\end{aligned}$$

D'altra parte, nel caso generale

$$\begin{aligned}
\mathbb{E}[\widetilde{X}(\omega)\mathbb{I}_A] &= \mathbb{E}[\widetilde{X}(\omega)\mathbf{1}_B(Y)] \\
&= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}} x \mathbf{1}_{\{z: f_Y(z)>0\}}(y) \frac{f_{X,Y}(x,y)}{f_Y(y)}\ dx \right) \mathbf{1}_B(y) f_Y(y)\ dy \\
&\quad + \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}} x \mathbf{1}_{\{z: f_Y(z)=0\}}(y) f_0(x)\ dx \right) \mathbf{1}_B(y) f_Y(y)\ dy
\end{aligned}$$

ovvero, per il Teorema di Fubini,

$$\begin{aligned}
&= \int_{\mathbb{R}} \int_{\mathbb{R}^d} x \mathbf{1}_B(y) \mathbf{1}_{\{z: f_Y(z)>0\}}(y) \frac{f_{X,Y}(x,y)}{f_Y(y)} f_Y(y)\ dx\ dy \\
&\quad + \int_{\mathbb{R}} \int_{\mathbb{R}^d} x \mathbf{1}_B(y) \mathbf{1}_{\{z: f_Y(z)=0\}}(y) f_0(x) f_Y(y)\ dx\ dy \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}^d} x \mathbf{1}_B(y) \mathbf{1}_{\{z: f_Y(z)>0\}}(y) \frac{f_{X,Y}(x,y)}{f_Y(y)} f_Y(y)\ dx\ dy \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}^d} x \mathbf{1}_B(y) \mathbf{1}_{\{z: f_Y(z)>0\}}(y) f_{X,Y}(x,y)\ dx\ dy
\end{aligned}$$

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

$$\int_{\mathbb{R}} \int_{\mathbb{R}^d} x \mathbf{1}_B(y) \mathbf{1}_{\{z: f_Y(z)>0\}}(y) f_{X,Y}(x,y)\ dx\ dy = \int_{\mathbb{R}} \int_{\mathbb{R}^d} x \mathbf{1}_B(y) f_{X,Y}(x,y)\ dx\ dy.$$

La verifica è immediata in quanto

$$\int_{\mathbb{R}} \int_{\mathbb{R}^d} x \mathbf{1}_B(y) \mathbf{1}_{\{z: f_Y(z)=0\}}(y) f_{X,Y}(x,y)\ dx\ dy = 0,$$

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

$$\int_{\mathbb{R}^d} x \mathbf{1}_B(y) f_{X,Y}(x,y)\ dx = 0.$$

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

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

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

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

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

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

Esempio 2.5. *Caso in cui $\mathcal{G} = \sigma(Y)$ e (X,Y) è una variabile (congiuntamente) gaussiana bidimensionale, di media nulla. Come caso particolare dell'esempio precedente si ottiene*

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

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

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

e

$$\mathbb{E}[X_n \mid X_1, \cdots, X_{n-1}] = -\sum_{i=1}^{n-1} \frac{\alpha_{i,n}}{\alpha_{n,n}} X_i. \tag{2.13}$$

soluzione[14]*: Si tratta del caso $d = n-1$ con $X = X_n$ e $Y = (X_1, \cdots, X_{n-1})$ e con $f_{X,Y}(x,y) > 0$. Per calcolare $\mathbb{E}[X_n \mid X_1, \cdots, X_{n-1}]$ dobbiamo innanzitutto*

$$\frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_{Y,X}(y,x)}{f_Y(y)} = \frac{f_{X_1,\ldots,X_{n-1},X_n}(y,x)}{f_{X_1,\ldots,X_{n-1}}(y)}, \qquad \text{con } y = (x_1 \ldots, x_{n-1}).$$

Abbiamo quindi

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

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

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

e quindi che

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

Di conseguenza

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

[14] Pur essendo possibile utilizzare la tecnica dell'esempio precedente si consiglia lo svolgimento dei calcoli dopo l'introduzione delle distribuzioni condizionali

Come osservato nel caso generale dell'esempio precedente, qualunque sia y, $\frac{f_{X,Y}(x,y)}{f_Y(y)}$ deve essere una densità di probabilità: è quindi chiaro che deve coincidere con la densità di una variabile aleatoria gaussiana[15] $N(-m(y), \frac{1}{2\alpha_{n,n}})$, *di media $-m(y)$ e di varianza $\frac{1}{2\alpha_{n,n}}$. Il valore atteso si calcola quindi come*

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

da cui si ottiene (2.13).

È infine interessante notare che essendo l'espressione del valore condizionato $\mathbb{E}[X_n \mid X_1, \cdots, X_{n-1}]$, una funzione lineare di $X_1, \cdots, X_{n-1}$, si ottiene che coincide con la retta di regressione[16] *di X_n rispetto a $X_1, \cdots, X_{n-1}$.*

Esempio 2.6. *L'Esempio 2.4 si generalizza facilmente al caso in cui X è una variabile aleatoria a valori in $\mathbb{R}^k$, e si vuole calcolare il valore atteso condizionale $\mathbb{E}[h(X) \mid Y]$, dove $h(\cdot)$ è una funzione misurabile, a valori reali, ripetendo tutti i passaggi con i dovuti cambiamenti:*

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

2.3 Proprietà del valore atteso condizionale

Enunciamo ora (senza dimostrarle, per il momento), le proprietà fondamentali della media condizionale (secondo la Definizione 2.1).

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

1. Linearità

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

2. Monotonia

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

[15] La funzione $K(y)$ di y deve inoltre necessariamente valere:

$$K(y) = \frac{1}{\sqrt{2\pi\frac{1}{2\alpha_{n,n}}}} = \sqrt{\frac{\alpha_{n,n}}{\pi}}.$$

[16] Ricordiamo che il problema di trovare la retta di regressione lineare di X rispetto ad Y corrisponde al problema di trovare il punto di minimo, tra tutte le funzioni affini $\boldsymbol{a} \cdot y + b$ del valore atteso del quadrato della differenza tra X, ovvero trovare $\tilde{\boldsymbol{a}}$ e $\tilde{b}$ tali che

$$\mathbb{E}\big[\big(X - (\tilde{\boldsymbol{a}} \cdot Y + \tilde{b})\big)^2\big] = \min_{\boldsymbol{a},b} \mathbb{E}\big[\big(X - (\boldsymbol{a} \cdot Y + b)\big)^2\big].$$

Tenendo presente che le variabili aleatorie $\sigma(Y)$-misurabili sono le variabili aleatorie $Z = g(Y)$, con g boreliana, allora il valore atteso condizionato, secondo la Definizione 2.2, è quella variabile aleatoria $\mathbb{E}[X|Y] = \tilde{X} = \tilde{g}(Y)$ tale che

$$\mathbb{E}\big[\big(X - \tilde{g}(Y)\big)^2\big] = \min_{g \text{ boreliane}} \mathbb{E}\big[\big(X - g(Y)\big)^2\big],$$

di conseguenza si ha che se $\mathbb{E}[X|Y] = \tilde{X} = \tilde{\boldsymbol{a}} \cdot Y + \tilde{b}$ è una funzione affine, allora coincide anche con la retta di regressione.

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

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

quindi, in particolare,

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

4. Fattorizzazione

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

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

5. Condizionamento rispetto a σ-algebre indipendenti

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

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

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

Se X e $\mathcal{G}$ sono indipendenti da $\mathcal{H}$, nel senso che $\sigma(X) \vee \mathcal{G}$ è indipendente da $\mathcal{H}$, allora

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

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

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

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

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

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

e quindi, passando al valore atteso, che

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

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

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

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

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

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

A questo punto passiamo ad osservare che la Definizione 2.1 è ben posta, in quanto esiste almeno una variabile aleatoria che verifica la (2.4).

Dimostrazione. Si definisca infatti la misura $\nu(\cdot)$ su $\mathcal{G}$ come $\nu(A) := \mathbb{E}[X\mathbb{I}_A]$, per $A \in \mathcal{G}$. La misura $\nu(\cdot)$ risulta assolutamente continua[17] rispetto a $\widehat{\mathbb{P}} = \mathbb{P}|_{\mathcal{G}}$, in quanto se $\widehat{\mathbb{P}}$(A)=$\mathbb{P}$(A)=0, allora $\nu(A) = 0$. Esiste quindi ***la derivata di Radon-Nikodym*** di ν rispetto a $\widehat{\mathbb{P}}$, cioè una funzione $f(\omega) = \frac{d\nu}{d\widehat{\mathbb{P}}}(\omega)$, $\mathcal{G}-$misurabile, per la quale valga, qualunque sia $A \in \mathcal{G}$

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

ovvero

$$\mathbb{E}[X\mathbb{I}_A] = \mathbb{E}[f\mathbb{I}_A].$$

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

□

Anche la Definizione 2.2 è ben posta, per convincersene basta considerare che lo spazio $L^2(\Omega, \mathcal{F}, \mathbb{P})$, pensato come classi di equivalenza, è uno spazio di Hilbert con la norma $\| X \|^2 = \mathbb{E}[\mid X \mid^2]$ ed $L^2(\Omega, \mathcal{G}, \widehat{\mathbb{P}})$ è un suo sottospazio chiuso. Quindi la classe di equivalenza $\mathbb{E}[X \mid \mathcal{G}]$ della Definizione 2.2 è la proiezione di X su $L^2(\Omega, \mathcal{G}, \widehat{\mathbb{P}})$. A questo proposito va però ricordato quanto già detto in una nota della sezione 1.8.1: per passare agli spazi di Hilbert bisogna assicurarsi che $\mathcal{G}$ contenga gli insiemi $\mathcal{F}$-misurabili e di probabilità nulla.

Si noti la (2.4) potrebbe essere modificata come

$$\mathbb{E}[XW] = \mathbb{E}[\widetilde{X}W], \quad \forall \text{ v.a. } W \quad \mathcal{G} - \text{misurabile e per cui } XW \text{è integrabile.} \tag{2.14}$$

Dimostrazione. **Attenzione**: *la dimostrazione si basa sulle proprietà di linearità e di monotonia dei valori attesi condizionali, che dimostreremo più in là, basandoci solo sulla Definizione 2.1 e quindi sulla (2.4)*
Ovviamente (2.14) implica (2.4). Per mostrare il viceversa, basta considerare il caso in cui $W \geq 0$. In tale caso esiste una successione $W_n \uparrow$W, con W_n funzioni elementari, cioè combinazioni lineari di funzioni indicatrici. D'altra parte è facile vedere (confronta le proprietà di linearità e di monotonia, dimostrate nella sezione successiva) che, posto

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

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

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

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

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

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

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

□

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

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

2.4 Equivalenza tra le definizioni di valore atteso condizionale per variabili aleatorie di quadrato sommabile

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

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

Se infatti $\widetilde{X}_2$ soddisfa la condizione (2.7) allora, qualunque sia $Z \in L^2(\Omega, \mathcal{G}, \widehat{\mathbb{P}})$ la funzione

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

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

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

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

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

Considerando che tutte le funzioni indicatrici del tipo $\mathbb{I}_A$, con $A \in \mathcal{G}$, sono in $L^2(\Omega, \mathcal{G}, \widehat{\mathbb{P}})$, otteniamo che se vale la (2.7) allora vale la (2.4).

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

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

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

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

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

Posto $W = Z - \widetilde{X}$ e tenendo conto della (2.14), si ottiene $\phi'(\bar{t}) = 2\bar{t}\,\mathbb{E}[(Z - \widetilde{X})^2] = 0$, da cui, per l'arbitrarietà di Z, segue che $\bar{t} = 0$. Quindi $\widetilde{X}$ gode della proprietà che per ogni Z di quadrato sommabile $\phi(1) \geq \phi(0)$, ovvero la (2.7).

Osservazione 2.1. *Si noti l'analogia*[18] *con il problema classico di geometria euclidea del trovare la proiezione di un vettore $\boldsymbol{x} \in \mathbb{R}^n$ su un sottospazio vettoriale $V \subset \mathbb{R}^n$ come quel vettore $\tilde{\boldsymbol{x}} \in V$ che gode della proprietà di minimizzare la distanza da V, ovvero*

$$\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|^2 = \min_{\boldsymbol{z} \in V} \|\boldsymbol{x} - \boldsymbol{z}\|^2,$$

che è notoriamente equivalente alla condizione di ortogonalità

$$< \boldsymbol{x} - \tilde{\boldsymbol{x}}, \boldsymbol{z} >= 0 \text{ per ogni } \boldsymbol{z} \in V \quad \Leftrightarrow \quad < \boldsymbol{x}, \boldsymbol{z} >=< \tilde{\boldsymbol{x}}, \boldsymbol{z} > \text{ per ogni } \boldsymbol{z} \in V.$$

Esercizio 2.1 (Formula di Pitagora). ***Si verifichi che, per ogni variabile aleatoria X di quadrato integrabile, posto $\widetilde{X}$ un rappresentante di $\mathbb{E}[X|\mathcal{G}]$, vale la seguente uguaglianza:*

$$Var(X) = Var(\widetilde{X}) + \mathbb{E}\big[Var_{\mathcal{G}}(X)\big],$$

dove

$$Var_{\mathcal{G}}(X) = \mathbb{E}\big[\big(X - \widetilde{X}\big)^2|\mathcal{G}\big].$$

Il motivo della notazione diviene evidente se si identifica $\widetilde{X}$ con $\mathbb{E}[X|\mathcal{G}]$ e si riscrive la precedente formula come segue:

$$Var_{\mathcal{G}}(X) = \mathbb{E}\big[\big(X - \mathbb{E}[X|\mathcal{G}]\big)^2|\mathcal{G}\big].$$

[18] L'analogia si vede meglio se si sostituisce $\|\boldsymbol{x} - \boldsymbol{z}\|^2$ con $d_{L^2}(X, Z) = \mathbb{E}[|X - Z|^2]$ e $< \boldsymbol{x}, \boldsymbol{z} >$ con $\mathbb{E}[XZ]$.

2.5 Dimostrazioni delle proprietà del valore atteso condizionale

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

Si noti che ciò permette di concludere che la dimostrazione dell'equivalenza delle Definizioni 2.1 e 2.2 è autocontenuta: in realtà andrebbero prima dimostrate le proprietà 1,2, e 7; successivamente la (2.14) e infine l'equivalenza delle Definizioni 2.1 e 2.2.

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

La dimostrazione è ovvia, infatti se $\tilde{X}$ e $\tilde{Y}$ sono rappresentanti di $\mathbb{E}[X \mid \mathcal{G}]$ e di $\mathbb{E}[Y \mid \mathcal{G}]$, rispettivamente, ovvero se $\mathbb{E}[X\mathbb{I}_A] = \mathbb{E}[\widetilde{X}\mathbb{I}_A]$ e $\mathbb{E}[Y\mathbb{I}_A] = \mathbb{E}[\widetilde{Y}\mathbb{I}_A]$ per ogni $A \in \mathcal{G}$, allora basta verificare che $\mathbb{E}[(aX + bY)\mathbb{I}_A] = \mathbb{E}[(a\widetilde{X} + b\widetilde{Y})\mathbb{I}_A]$, e ciò segue immediatamente da

$$\mathbb{E}[(aX + bY)\mathbb{I}_A] = a\mathbb{E}[X\mathbb{I}_A] + b\mathbb{E}[Y\mathbb{I}_A] = a\mathbb{E}[\widetilde{X}\mathbb{I}_A] + b\mathbb{E}[\widetilde{Y}\mathbb{I}_A] = \mathbb{E}[(a\widetilde{X} + b\widetilde{Y})\mathbb{I}_A].$$

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

Se $\mathbb{P}(X \leq Y) = 1$, allora $\mathbb{E}[(Y - X)\mathbb{I}_A] = \mathbb{E}[(\widetilde{Y} - \widetilde{X})\mathbb{I}_A] \geq 0$ per ogni $A \in \mathcal{G}$ e quindi in particolare per $A = \{\widetilde{X} < \widetilde{Y}\}$, si ottiene[19] che $\mathbb{P}(\widetilde{X} \leq \widetilde{Y}) = 1$, ovvero $\mathbb{P}(\mathbb{E}[X \mid \mathcal{G}] \leq \mathbb{E}[Y \mid \mathcal{G}]) = 1$.

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

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

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

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

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

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

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

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

$$\mathbb{E}[XZ\mathbb{I}_A] = \mathbb{E}[X]\mathbb{E}[Z\mathbb{I}_A] = \mathbb{E}[Z\mathbb{E}[X]\mathbb{I}_A]$$

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

[19] ****$\mathbb{E}[(\widetilde{Y} - \widetilde{X})\mathbb{I}_{\{\widetilde{X} < \widetilde{Y}\}}] \geq 0$ se e solo se $\mathbb{P}(\widetilde{X} < \widetilde{Y}) = 0$: infatti se fosse $\mathbb{P}(\widetilde{X} < \widetilde{Y}) > 0$, allora si avrebbe $\mathbb{E}[(\widetilde{Y} - \widetilde{X})\mathbb{I}_{\widetilde{X} < \widetilde{Y}}] < 0$.

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

La dimostrazione segue da una leggera variante del Lemma di Dynkin[20] e si basa sul fatto che la famiglia $\mathcal{A}$ di eventi del tipo $C = A \cap B$, con $A \in \mathcal{G}$ e $B \in \mathcal{H}$, formano un sistema chiuso rispetto all'intersezione e generano $\mathcal{G} \vee \mathcal{H}$,per il lemma di Dynkin (Lemma 1.1) **basterà quindi verificare che le due misure $\overline{\nu}(C) := \mathbb{E}[X\mathbb{I}_C]$ e $\overline{\mu}(C) := \mathbb{E}[\widetilde{X}\mathbb{I}_C]$ coincidono per $C = A \cap B$, $A \in \mathcal{G}$ e $B \in \mathcal{H}$

$$\begin{aligned}\mathbb{E}[X\mathbb{I}_{A\cap B}] &= \mathbb{E}[X\mathbb{I}_A\mathbb{I}_B] \qquad \text{(per l'indipendenza di } \sigma(X)\vee\mathcal{G} \text{ da } \mathcal{H}) \\ &= \mathbb{E}[X\mathbb{I}_A]\mathbb{E}[\mathbb{I}_B] = \mathbb{E}[\widetilde{X}\mathbb{I}_A]\mathbb{E}[\mathbb{I}_B] = \mathbb{E}[\widetilde{X}\mathbb{I}_A\mathbb{I}_B] = \mathbb{E}[\widetilde{X}\mathbb{I}_{A\cap B}].\end{aligned} \tag{2.15}$$

e che $\mathcal{A}$ contiene Ω (essendo ovviamente $\Omega = \Omega \cap \Omega$).

Nella dimostrazione abbiamo utilizzato l'indipendenza della σ-algebra $\sigma(X) \vee \mathcal{G}$, generata da $\sigma(X)$ e $\mathcal{G}$ da $\mathcal{H}$, ossia: per ogni evento $F \in \sigma(X) \vee \mathcal{G}$ e per ogni evento $H \in \mathcal{H}$ si ha $\mathbb{P}(F \cap H) = \mathbb{P}(F)\mathbb{P}(H)$. Attenzione, la sola indipendenza di $\sigma(X)$ da $\mathcal{H}$ e di $\mathcal{G}$ da $\mathcal{H}$ non è sufficiente[21] nel passaggio (2.15).

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

Per risultati classici di analisi ϕ è l'inviluppo delle sue tangenti (o sotto tangenti) ovvero esistono una successione di rette $\phi_n(x) = \alpha_n x + \beta_n$ per cui $\phi(x) = \sup_n\{\phi_n(x)\}$, per ogni x in $\mathbb{R}$. Per le proprietà **1.** di linearità e **2.** di monotonia si ha allora che

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

passando all'estremo superiore su n si ottiene

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

---

[20] Ricordiamo per comodità del lettore l'enunciato del Lemma di Dynkin (Lemma 1.1), che tra l'altro è la base per molti risultati di unicità, ed è noto anche come teorema dell'unicità della misura.

Lemma [Lemma di Dynkin, Billingsley 1984 [3]] *Sia $\mathcal{A}$ una famiglia di eventi che genera la σ-algebra $\mathcal{G}$ e che è chiusa rispetto alla intersezione finita.*
Se due misure di probabilità ν e μ coincidono su $\mathcal{A}$, allora le due misure di probabilità coincidono su $\mathcal{G}$.
Di conseguenza, se $\overline{\nu}$ e $\overline{\mu}$ sono due misure non negative e finite che coincidono su $\mathcal{A} \cup \{\Omega\}$ (cioè con $\overline{\nu}(\Omega) = \overline{\mu}(\Omega)$) allora le due misure coincidono su $\mathcal{G}$.
(la dimostrazione della affermazione sulle misure non negative riconduce subito al caso delle misure di probabilità, considerando le misure di probabilità $\nu(A) := \overline{\nu}(A)/\overline{\nu}(\Omega)$ e $\nu(A) := \overline{\mu}(A)/\overline{\mu}(\Omega)$.)

[21] Il fatto che la sola indipendenza di $\sigma(X)$ da $\mathcal{H}$ e di $\mathcal{G}$ da $\mathcal{H}$ non è sufficiente si può mostrare con il seguente controesempio: $X = \mathbb{I}_E$, $\mathcal{G} = \{\emptyset,\, G\, G^c,\, \Omega\}$, $\mathcal{H} = \{\emptyset,\, H\, H^c,\, \Omega\}$, con E, G ed H tali che $\mathbb{P}(E\cap H) = \mathbb{P}(E)\mathbb{P}(H)$, $\mathbb{P}(G\cap H) = \mathbb{P}(G)\mathbb{P}(H)$, ma $\mathbb{P}(E\cap G\cap H) \neq \mathbb{P}(E\cap G)\mathbb{P}(H)$ (ad esempio si prenda $\Omega = (0,1)\times(0,1)$, $E = (0,\frac{1}{2})\times(0,1)$, $G = (0,\frac{1}{4})\times(0,1) \cup (\frac{1}{4},\frac{1}{2})\times(0,\frac{1}{2})$ e $H = (\frac{1}{4},\frac{3}{4})\times(0,1)$).

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

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

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

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

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

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

Il caso generale si ottiene per induzione su n: Se

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

allora

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

Si noti che nella dimostrazione della disuguaglianza di Jensen vengono utilizzate solo le proprietà di linearità e di monotonia del valore atteso condizionato.

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

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

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

8ii) Se $\{X_n\}_{n\geq 1}$ è una successione di variabili aleatorie integrabili, convergente con probabilità 1 ad X, **dominatamente**, cioè $\mid X_n \mid \leq Y$, per una variabile aleatoria Y integrabile, allora $\mathbb{E}[X_n \mid \mathcal{G}] \to \mathbb{E}[X \mid \mathcal{G}]$, in L^1.
La dimostrazione relativa alla convergenza dominata si basa sulla seguente osservazione: se $X_n \to X$ q.c. e $\mid X_n \mid \leq Y$ allora $X_n \to X$ in L^1 (infatti allora $\mid X_n - X \mid \leq \mid Y \mid + \mid X \mid$ e quindi $\mid X_n - X \mid \to 0$ dominatamente e perciò $\mathbb{E}[\mid X_n - X \mid] \to 0$). Di conseguenza per le proprietà di linearità e per la disuguaglianza di Jensen applicata alla funzione $\phi(x) = \mid x \mid$ e alla variabile aleatoria $X_n - X$,

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

Passando ai valori medi

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

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

2.6 Probabilità condizionali regolari

Il problema è il seguente:

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

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

sia una probabilità per ogni ω?

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Proposizione 2.1. *Sia X una variabile aleatoria reale in $(\Omega, \mathcal{F}, \mathbb{P})$, sia $\mathcal{G}$ una sotto σ-algebra di $\mathcal{F}$, e sia $\mathcal{A} = \sigma(X) = \big\{A = \{\omega \in \Omega,\ t.c.\, X(\omega) \in H\},\ H \in \mathcal{B}(\mathbb{R})\big\}$. Allora esiste una versione regolare $\mathbb{Q}(\cdot,\cdot)$ di $\mathbb{P}(\cdot|\mathcal{G})(\omega)$.*

Dimostrazione. L'idea è molto semplice:

1. Si costruisce una funzione $\hat{F}(\cdot,\cdot) : \mathbb{R} \times \Omega \mapsto [0,1]$ che, per ogni ω, $\hat{F}(\cdot,\omega) : \mathbb{R} \mapsto [0,1]$ soddisfa tutte le proprietà di una funzione di ripartizione, e che è una versione della probabilità condizionale $\mathbb{P}(X \leq s \mid \mathcal{G})(\omega)$. Indicheremo tale funzione con $F(s \mid \mathcal{G})(\omega)$, per ricordare questa proprietà.

dimostrazione di 1. Si inizia considerando $F(t \mid \mathcal{G})(\omega) := \mathbb{P}(X \leq t \mid \mathcal{G})(\omega)$ per t razionale. Per la proprietà della monotonia, possiamo prendere una versione per cui, per ogni $t_1 \leq t_2$ razionali, $F(t_1 \mid \mathcal{G})(\omega) \leq F(t_2 \mid \mathcal{G})(\omega)$, in un evento Ω_0 di probabilità 1: in questo caso l'evento in cui ciò non si verifica è un'unione numerabile di eventi di probabilità nulla. Per $\omega \in \Omega_0^c$ si definisce $F(t \mid \mathcal{G})(\omega) = F_0(t)$, con F_0 una fissata funzione di distribuzione, ad esempio (sempre sui razionali). Su tale evento Ω_0 esiste, per monotonia, il limite di $F(t \mid \mathcal{G})(\omega)$ sia per $t \to +\infty$ che per $t \to -\infty$, sempre per t razionale. Tali limiti sono rispettivamente uguali ai limiti di $F(n \mid \mathcal{G})(\omega)$ e di $F(-n \mid \mathcal{G})(\omega)$, e, di nuovo a parte un insieme di probabilità nulla, coincidono rispettivamente con $1 = \mathbb{P}(X < +\infty \mid \mathcal{G})$ e $0 = \mathbb{P}(X < -\infty \mid \mathcal{G})$.(Di nuovo su tale insieme si definisca $F(t \mid \mathcal{G})(\omega) = F_0(t)$)

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

2. Ad ogni ω è associata una misura $\mu_X(\cdot \mid \mathcal{G})(\omega)$ di probabilità, per cui $\mu_X((-\infty, t] \mid \mathcal{G})(\omega) = F(t \mid \mathcal{G})(\omega)$, per ogni t reale.
Come conseguenza del punto **2.**, per ogni $H \in \mathcal{B}(\mathbb{R})$, la misura $\mu_X(H \mid \mathcal{G})(\omega)$ è una versione di $\mathbb{P}(X \in H)$, oltre ad essere una misura.
La misura $\mu_X(\cdot \mid \mathcal{G})(\omega)$ è detta appunto ***la distribuzione condizionale di X data $\mathcal{G}$***.

3. Per $A \in \mathcal{A} = \sigma(X)$, con $A = \{X \in H\}$ $(H \in \mathcal{B}(\mathbb{R}))$ si definisce

$$\mathbb{Q}(A, \omega) := \mu_X(H \mid \mathcal{G})(\omega),$$

e risulta la versione regolare delle probabilità cercata. □

Si faccia attenzione: la distribuzione condizionale $\mu_X(\cdot \mid \mathcal{G})(\omega)$, definita nel punto **2.** della dimostrazione precedente, è una probabilità su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, mentre $\mathbb{Q}(\cdot, \omega)$ è una misura su $(\Omega, \sigma(X))$. Comunque accade che $\mathbb{P}(\{X \in H\} \mid \mathcal{G})(\omega) = \mu_X(H \mid \mathcal{G})(\omega)$. Inoltre poiché le v.a. $\sigma(X)-$misurabili sono le variabili aleatorie del tipo $Z = f(X)$, con f boreliana, si ha che

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

dove, in generale,

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

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

In particolare quindi il risultato di esistenza della versione regolare è valido per $(S, \mathcal{S}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ o $(S, \mathcal{S}) = (\mathbb{R}^{\mathbb{N}}, \mathcal{R}^{\mathbb{N}})$, che sono spazi di Borel (vedere ad esempio P. Billingsley [3] "Convergence of Probability measures" pag. 218 e seguenti).

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

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

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

in quanto la seconda è una versione di

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

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

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

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

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

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

2.7 Esempi

Esempio 2.7 (caso dominato)**.** *Si tratta della generalizzazione dell'Esempio 2.4. Siano X ed Y due variabili aleatorie con legge congiunta assolutamente continua rispetto ad una misura prodotto $\nu_1(dx) \times \nu_2(dy)$ cioè con legge congiunta data da*

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

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

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

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

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

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

per μ_Y-quasi ogni y.

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

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

dove

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

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

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

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

Un risultato analogo vale ovviamente anche per la legge di Y data X.
Casi particolari sono i casi in cui $\nu_1(dx) = dx$, e $\nu_2(dy) = dy$, cioè torniamo al caso della misura di Lebesgue esaminato nell'Esempio 2.4, oppure $\nu_1(dx) = \sum_{n=0}^{\infty} \delta_{x_n}(dx)$, e $\nu_2(dy) = \sum_{n=0}^{\infty} \delta_{y_n}(dy)$ e si trova allora che

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

Esempio 2.8 (caso non dominato)**.** *Siano X ed Y due variabili aleatorie con legge congiunta data da*

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

dove $p + q = 1$, $f_1(x)$ è una densità di probabilità su $\mathbb{R}$, $f_2(x, y)$ è una densità di probabilità su $\mathbb{R}^2$, $\delta_z(dx)$ è la misura concentrata in z (ovvero $\delta_z(A) = 1$ se $z \in A$, mentre $\delta_z(A) = 0$ se $z \notin A$), α è una funzione invertibile, C^1 e con $|\ \alpha'(x)\ |$ strettamente positiva.

Le distribuzioni marginali sono equivalenti alla misura di Lebesgue, essendo

$$\mu_X(dx) = \left[pf_1(x) + q\int f_2(x,y)dy\right] dx,$$

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

L'ultima uguaglianza deriva da

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

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

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

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

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

Se poi troviamo che

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

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

Ora, qualunque sia la funzione h

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

mentre invece

$$\mathbb{E}[\phi(Y)h(Y)] = p\int \phi(y)h(y)f_1(\alpha^{-1}(y))\frac{1}{|\alpha'(\alpha^{-1}(y))|}dy + q\int \phi(y)h(y)dy\int f_2(x,y)dx$$
$$= \int \phi(y)\left[pf_1(\alpha^{-1}(y))\frac{1}{|\alpha'(\alpha^{-1}(y))|}+q\Big(\int f_2(x,y)dx\Big)\right]h(y)dy.$$

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

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

ovvero che

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

Si noti ancora che

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

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

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

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

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

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

Capitolo 3

Martingale

In questo capitolo introduciamo il concetto di martingala, che in un certo senso si può considerare una formalizzazione del concetto di gioco equo. Per definire una martingala abbiamo però prima bisogno di dare la seguente definizione.

Definizione 3.1 (Filtrazione)**.** *In uno spazio* $(\Omega, \mathcal{F})$*, la famiglia* $\{\mathcal{F}_t, t \geq 0\}$ *si dice una* **filtrazione** *se è una famiglia crescente di* σ*-algebre di* $\mathcal{F}$*, cioè* $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$*, per* $0 \leq s \leq t$*. (La definizione ha senso anche nel caso in cui* $t = n \in \mathbb{N}$*, e nel caso in cui* $t \in I \subseteq \mathbb{R}$*, ad esempio* $t \in [0, T]$*).*

La $\sigma-$algebra $\mathcal{F}_t$ rappresenta l'informazione disponibile fino al tempo t, e più in generale la filtrazione $\{\mathcal{F}_t\}$ viene detta anche ***flusso di σ-algebre***, in quanto rappresenta il flusso di informazioni disponibili, al variare del tempo.

Definizione 3.2 (Martingala)**.** *Un processo aleatorio*[1] (X_t)*, definito in uno spazio di probabilità* $(\Omega, \mathcal{F}, \mathbb{P})$*, si dice una* **martingala** *rispetto ad una filtrazione* $\{\mathcal{F}_t\}$*, con* $\mathcal{F}_t \subseteq \mathcal{F}$*, se*

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

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

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

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

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

Si noti che la proprietà **0)** è sovrabbondante in quanto la **2)** implica che X_t sia $\mathcal{F}_t$-misurabile.

Ciò non è vero nella seguente definizione di submartingala (supermartingala), che in un certo senso si può considerare una formalizzazione del concetto di gioco favorevole (sfavorevole).

Definizione 3.3 (Submartingala (Supermartingala))**.** *Un processo aleatorio* X_t *si dice una* **submartingala** *(***supermartingala***) rispetto ad una filtrazione* $\{\mathcal{F}_t\}$ *se*

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

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

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

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

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

[1] Il lettore per il momento può pensare di considerare solo il caso tempo discreto, e sostituire la parola processo con la parola successione $\{X_n\}_n$ di variabili aleatorie. Inoltre può limitarsi a considerare il caso di una filtrazione $\{\mathcal{F}_n\}_n$ generata dal processo stesso, ovvero il caso in cui $\mathcal{F}_n$ è σ-algebra generata da $\{X_1, X_2, \cdots, X_n\}$. In altre parole

$$\mathcal{F}_n = \{A \subseteq \Omega, \text{ tali che esiste un boreliano } H_n \in \mathcal{B}(\mathbb{R}^n) \text{ per il quale } A = \{\omega \in \Omega : (X_1, X_2, \cdots, X_n) \in H_n\}\}.$$

Per la definizione formale di processo aleatorio si rimanda al Capitolo 4.

Osservazione 3.1. *Se X_t è una martingala (o submartingala) rispetto a una filtrazione $\{\mathcal{F}_t\}$ e se $\{\mathcal{G}_t\}$ è una filtrazione per cui $\sigma(X_t) \subseteq \mathcal{G}_t \subseteq \mathcal{F}_t$, per ogni $t \geq 0$, allora X_t lo è anche rispetto alla nuova filtrazione:*

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

In particolare per ogni martingala (o submartingala) si può prendere sempre la filtrazione minimale $\mathcal{F}_t^X = \sigma\big(\{X_u, u \leq t\}\big) = \sigma\big(\bigcup_{0\leq u\leq t} \sigma(X_u)\big)$, che per la proprietà **0)** *è sempre contenuta in $\mathcal{F}_t$. Quindi, in genere*[2]*, se la filtrazione non è specificata, si deve intendere che si tratti della filtrazione naturale.*

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

Osservazione 3.2. *Ogni martingala ha media costante e ogni submartingala ha media crescente (in senso lato): basta passare al valore medio nella proprietà* **3.** *delle medie condizionali (condizionamenti successivi) e tenere presente che il valore medio di Y coincide con il valore medio di $\mathbb{E}[Y \mid \mathcal{G}]$.*

Osservazione 3.3. *Data una submartingala X_t si ottiene una supermartingala considerando $-X_t$, e viceversa.*

3.1 Esempi di martingale e di submartingale

Esempio 3.1. *Sia data una variabile aleatoria Y integrabile ed una filtrazione $\{\mathcal{F}_t\}$. Si definisca $X_t := \mathbb{E}[Y \mid \mathcal{F}_t]$. Si dimostra facilmente che X_t è una martingala:*

$$\mathbb{E}[X_{t+s} \mid \mathcal{F}_t] \underset{(per\ def.)}{=} \mathbb{E}[\mathbb{E}[Y \mid \mathcal{F}_{t+s}] \mid \mathcal{F}_t] \underset{(condiz.\ \ successivi)}{=} \mathbb{E}[Y \mid \mathcal{F}_t]$$

L'esempio precedente aiuta a capire il nome di submartingala: sia ora X_t una submartingala, allora, fissato $v = t+s$, il processo

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

è una martingala per $t \in [0, v]$, quindi la proprietà **2)** per le submartingale diviene

$$Z_t := \mathbb{E}[X_v \mid \mathcal{F}_t] \geq X_t, \qquad \text{per } t \in [0, v],$$

ovvero che X_t sia *sotto* la martingala $Z_t := \mathbb{E}[X_v \mid \mathcal{F}_t]$, per $t \in [0, v]$.

*** Inoltre sempre in relazione all'Esempio 3.1, si può osservare che se due martingale M_t^1 ed M_t^2 coincidono al tempo T, ossia se $M_T^1 = M_T^2$, allora esse coincidono per ogni $t \leq T$, ossia $M_t^1 = M_t^2$ (con probabilità 1) in $[0, T]$, in quanto per tali valori di t si ha $M_t^i = \mathbb{E}[M_T^i \mid \mathcal{F}_t]$. ***

Esempio 3.2. *Sia data una successione di variabili aleatorie Y_n indipendenti, identicamente distribuite*[3]*, integrabili e a media nulla. Si definisca*

$$S_0 = 0, \qquad S_n = \sum_{k=1}^{n} Y_k, \quad n \geq 1,$$

o equivalentemente

$$S_n = \sum_{k=1}^{n} Y_k, \quad n \geq 0,$$

(con la convenzione che $\sum_{k=1}^{0} a_k = \sum_{1\leq k\leq 0} a_k = \sum_{k\in\emptyset} a_k = 0$).

[2] Attenzione: ovviamente la filtrazione potrebbe anche essere specificata anche all'inizio di una sezione, o di un capitolo.

[3] Non è necessario che le variabili aleatorie Y_n abbiano la stessa distribuzione, basta che siano integrabili e abbiano valore atteso nullo.

Si ottiene facilmente che S_n è una martingala, rispetto[4] *a*

$$\mathcal{F}_n^S = \begin{cases} \{\emptyset, \Omega\} & se\ n = 0, \\ \sigma(\{Y_1, \cdots, Y_n\}) & se\ n > 0. \end{cases}$$

Infatti

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

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

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

è una martingala[5].

Infine si osservi che, nel caso in cui $\mu > 0$, si ottiene che S_n è una submartingala, mentre, nel caso in cui $\mu < 0$, si ottiene che S_n è una supermartingala.

Esempio 3.3. *Se siamo nelle stesse ipotesi del precedente Esempio 3.2 ed inoltre le variabili aleatorie Y_k ammettono momento secondo finito, e quindi varianza σ^2, allora $M_n := S_n^2 - n\sigma^2$ è una martingala:*

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

Passando alle medie condizionali si ottiene che

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

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

Se le variabili aleatorie Y_k non sono a media nulla, allora

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

è una martingala[6].

Il seguente esempio permette di generare submartingale a partire da martingale e da submartingale.

Esempio 3.4. *Sia X_t una martingala con $\mathbb{E}[\mid X_t \mid^\alpha] < +\infty$, per un $\alpha \geq 1$. Allora il processo $\mid X_t \mid^\alpha$ è una submartingala: basta applicare la disuguaglianza di Jensen alla funzione $|x|^\alpha$, che è convessa per $\alpha \geq 1$*

$$\mid X_t \mid^\alpha =_{(X_t\ \text{è una MG})} \mid \mathbb{E}[X_{t+s} \mid \mathcal{F}_t] \mid^\alpha \leq_{(\text{dis. Jensen per } |x|^\alpha)} \mathbb{E}[\mid X_{t+s} \mid^\alpha \mid \mathcal{F}_t].$$

[4]La successione $\{Y_k\}$ si ottiene immediatamente dalla successione $\{S_k\}$:

$$Y_n = S_n - S_{n-1}, \qquad \text{per } n \geq 1,$$

di conseguenza $\mathcal{F}_n^S = \mathcal{F}_n^Y$, per $n \geq 1$.

[5]Sempre nel caso in cui le variabili aleatorie Y_k non abbiano la stessa legge si dovr1à sostituire $Y_k - \mu_k$, dove $\mu_k = \mathbb{E}[Y_k]$. In questo caso

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

[6]Di nuovo la stessa dimostrazione funziona anche nel caso in cui le variabili aleatorie Y_k non hanno la stessa legge, purché abbiano momento secondo finito e siano indipendenti. In tale caso

$$M_n := (S_n - \sum_{k=1}^{n} \mu_k)^2 - \sum_{k=1}^{n} \sigma_k^2$$

è una martingala. Si confronti questo risultato con il successivo Teorema di Decomposizione di Doob 3.1.

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

$$\phi(X_t) =_{(X_t \text{ è una } MG)} \phi(\mathbb{E}[X_{t+s} \mid \mathcal{F}_t]) \leq_{(\text{dis. Jensen per } \phi)} \mathbb{E}[\phi(X_{t+s}) \mid \mathcal{F}_t].$$

Per ottenere un risultato analogo nel caso in cui X_t sia una submartingala bisogna aggiungere l'ipotesi che ϕ sia una funzione crescente, in modo che, essendo

$$\begin{aligned} X_t &\leq_{(X_t \text{ è una subMG})} \mathbb{E}[X_{t+s} \mid \mathcal{F}_t], \\ &\Downarrow \\ \phi(X_t) &\leq_{(\phi \text{ è crescente})} \phi(\mathbb{E}[X_{t+s} \mid \mathcal{F}_t]) \leq_{(\text{dis. Jensen per } \phi)} \mathbb{E}[\phi(X_{t+s}) \mid \mathcal{F}_t]. \end{aligned}$$

Esempio 3.5. *Se le variabili aleatorie W_k sono indipendenti identicamente distribuite*[7]*, con $\mathbb{E}[W_1] = 1$, allora la successione di variabili aleatorie*

$$Z_0 = 1, \qquad Z_n := \prod_{k=1}^{n} W_k, \quad n \geq 1,$$

o equivalentemente

$$Z_n := \prod_{k=1}^{n} W_k, \quad n \geq 0,$$

(con la convenzione che $\prod_{k=1}^{0} a_k = 1$) definisce una martingala, rispetto [8] *a*

$$\mathcal{F}_n = \begin{cases} \{\emptyset, \Omega\} & \text{se } n = 0, \\ \sigma(\{W_1, \cdots, W_n\}) & \text{se } n > 0, \end{cases}$$

ovvero, dato che la condizione di misurabilità è ovvia, quella di integrabilità deriva dall'integrabilità di ciascuna delle W_k e dalla loro indipendenza, e infine

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

Come caso particolare si consideri la situazione dell'Esempio 3.2 con l'ulteriore ipotesi che per un $\theta \in \mathbb{R}$ valga

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

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

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

Come conseguenza, posto $W_k = \exp\{\theta Y_k - \psi(\theta)\}$, si ha che

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

è una martingala strettamente positiva di media 1.

Esempio 3.6 (Integrale stocastico a tempo discreto)**.** *Sia $\tilde{S}_n$ una martingala rispetto a $\mathcal{F}_n$, nello spazio $(\Omega, \mathcal{F}, \tilde{\mathbb{P}})$, e sia γ_n* **predicibile** *rispetto a $\mathcal{F}_n$, ovvero sia γ_n misurabile rispetto a $\mathcal{F}_{n-1}$ per ogni $n \geq 1$. Allora l'***integrale stocastico discreto**

$$(\gamma.\tilde{S})_n = I_n(\gamma) := \sum_{k=1}^{n} \gamma_k (\tilde{S}_k - \tilde{S}_{k-1}) \tag{3.1}$$

[7] L'ipotesi che abbiano la stessa legge è superflua, basta che le variabili aleatorie W_k abbiano valore atteso 1 e siano indipendenti.

[8] In questo caso $\mathcal{F}_n^Z$ può essere strettamente contenuta in $\mathcal{F}_n = \mathcal{F}_n^W$, infatti mentre $Z_n = g_n(W_1, \ldots, W_n)$, e quindi ogni funzione misurabile rispetto a $Z_1, \ldots, Z_n$ è funzione misurabile di $W_1, \ldots, W_n$, il viceversa in genere non è vero: se $Z_1, \ldots, Z_n$ sono tutti positivi, allora $W_k = Z_k/Z_{k-1}$, ma se $Z_n = 0$ allora $Z_{n+m} = 0$ per ogni $m \geq 0$ e quindi è impossibile ricavare i valori di W_{n+m} per $m > 0$. Se tuttavia le variabili aleatorie $W_k(\omega)$, assumono valori strettamente positivi per ogni k e per ogni ω, allora la corrispondenza tra $\{Z_k\}$ e $\{W_k\}$ è biunivoca e $\mathcal{F}_n^Z = \mathcal{F}_n^W$, per ogni $n \geq 1$.

definisce una $\mathcal{F}_n$-martingala, sotto una delle due seguenti condizioni:
a) *per ogni k esiste una costante c_k tale che $\tilde{\mathbb{P}}(\{\omega \text{ tali che } \gamma_k(\omega) \le c_k\}) = 1$*
b) *La martingala $\tilde{S}_k$ e il processo γ_k sono di quadrato integrabile, ovvero*

$$\tilde{\mathbb{E}}[|\tilde{S}_k|^2] < \infty, \text{ per ogni } k \ge 0 \quad \tilde{\mathbb{E}}[|\gamma_k|^2] < \infty, \text{ per ogni } k \ge 1$$

Verifica: *Per la misurabilità basta osservare che se $k \le n$ allora γ_k e $\tilde{S}_{k-1}$ sono $\mathcal{F}_{k-1}$-misurabile e quindi anche $\mathcal{F}_n$-misurabile, analogamente $\tilde{S}_k$ è $\mathcal{F}_k$-misurabile e quindi anche $\mathcal{F}_n$-misurabile, di conseguenza*

$$I_n(\gamma) = \sum_{k=1}^{n} \gamma_k(\tilde{S}_k - \tilde{S}_{k-1})$$

è $\mathcal{F}_n$-misurabile. Per l'integrabilità si osservi che

$$|I_n(\gamma)| \le \sum_{k=1}^{n} |\gamma_k|(|\tilde{S}_k| + |\tilde{S}_{k-1}|)$$

e che, se vale la condizione **a)**, *allora*

$$\tilde{\mathbb{E}}[|\gamma_k||\tilde{S}_k|] \le c_k\tilde{\mathbb{E}}[|\tilde{S}_k|] < \infty \qquad \tilde{\mathbb{E}}[|\gamma_k||\tilde{S}_{k-1}|] \le c_k\tilde{\mathbb{E}}[|\tilde{S}_{k-1}|] < \infty,$$

mentre se vale la condizione **b)**, *allora per la disuguaglianza di Cauchy*

$$\tilde{\mathbb{E}}[|\gamma_k||\tilde{S}_k|] \le \tilde{\mathbb{E}}^{1/2}[|\gamma_k|^2]\tilde{\mathbb{E}}^{1/2}[|\tilde{S}_k|^2] < \infty \qquad \tilde{\mathbb{E}}[|\gamma_k||\tilde{S}_{k-1}|] \le \tilde{\mathbb{E}}^{1/2}[|\gamma_k|^2]\tilde{\mathbb{E}}^{1/2}[|\tilde{S}_{k-1}|^2] < \infty.$$

Infine basta osservare che

$$\begin{aligned}
\tilde{\mathbb{E}}[I_{n+1}(\gamma) - I_n(\gamma) \mid \mathcal{F}_n] &= \tilde{\mathbb{E}}[\gamma_{n+1}(\tilde{S}_{n+1} - \tilde{S}_n) \mid \mathcal{F}_n] \\
&\underset{(\gamma_{n+1} \text{ è } \mathcal{F}_n-\text{mis.})}{=} \gamma_{n+1}\,\tilde{\mathbb{E}}[(\tilde{S}_{n+1} - \tilde{S}_n) \mid \mathcal{F}_n] \\
&\underset{(\tilde{S}_n \text{ è una } \mathcal{F}_n-\text{MG})}{=} \gamma_{n+1}\,0 = 0
\end{aligned}$$

***Per finire aggiungiamo due osservazioni:*
(i) *Per ottenere l'integrabilità di $I_n(\gamma)$, è sufficiente avere l'integrabilità di $\gamma_k(\tilde{S}_k - \tilde{S}_{k-1})$, per ogni k, e le condizioni* **a)** *e* **b)** *sono condizioni sufficienti affinché ciò accada.*
(ii) *Se invece la successione $\tilde{S}_k$ fosse una submartingala e il processo γ_k fosse non negativo, allora $I_n(\gamma)$ risulta una submartingala.* **

Esempio 3.7. [9] *Dato uno spazio $(\Omega, \mathcal{F})$ e su di esso una filtrazione $\{\mathcal{F}_t\}$ e due misure di probabilità $\mathbb{P}$ e $\mathbb{Q}$, con $\mathbb{P}$ assolutamente continua rispetto a $\mathbb{Q}$ (di conseguenza lo sono anche rispetto ad $\mathcal{F}_t$ per ogni t). Si definisca*[10] *la derivata*

[9] Questo esempio richiede la conoscenza del Teorema di Radon Nikodym, e può essere tralasciato in una prima lettura. In alternativa il lettore può considerare solo il caso a tempo discreto con $t = k \in \{1, \cdots n\}$, $\Omega = \mathbb{R}^n$, $\mathcal{F}_k = \{A = H_k \times \mathbb{R}^{n-k}$, con $H_k \in \mathcal{B}(\mathbb{R}^k)\}$, ed infine $\mathbb{P}(dx_1 \cdots dx_n) = p(x_1, \cdots, x_n)dx_1 \cdots dx_n$ e $\mathbb{Q}(dx_1 \cdots dx_n) = q(x_1, \cdots, x_n)dx_1 \cdots dx_n$ con p e q densità di probabilità. La assoluta continuità di $\mathbb{P}$ rispetto a $\mathbb{Q}$ diviene allora la condizione

$$\{(x_1, \cdots, x_n) : q(x_1, \cdots, x_n) = 0\} \subseteq \{(x_1, \cdots, x_n) : p(x_1, \cdots, x_n) = 0\},$$

e

$$\frac{d\mathbb{P}}{d\mathbb{Q}}(x_1, \cdots, x_n) = \frac{p(x_1, \cdots, x_n)}{q(x_1, \cdots, x_n)}.$$

[10] Nel caso a tempo discreto della nota precedente

$$**L_k(x_1, \cdots, x_n) := \frac{p_k(x_1, \cdots, x_k)}{q_k(x_1, \cdots, x_k)},$$

dove

$$p_k(x_1, \cdots, x_k) = ** \int_{\mathbb{R}^{n-k}} p(x_1, \cdots, x_k, y_{k+1}, \cdots, y_n)dy_{k+1}, \cdots, dy_n$$

e

$$q_k(x_1, \cdots, x_k) = ** \int_{\mathbb{R}^{n-k}} q(x_1, \cdots, x_k, y_{k+1}, \cdots, y_n)dy_{k+1}, \cdots, dy_n$$

sono le densità marginali su $\mathbb{R}^k$ di p e q, rispettivamente.

di Radon Nikodym di $\mathbb{P}$ rispetto a $\mathbb{Q}$, entrambe ristrette a $\mathcal{F}_t$, cioè

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

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

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

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

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

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

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

Esempio 3.8. *Sia X_n una* **catena di Markov omogenea**[11] *con spazio degli stati finito **(o numerabile)**e con matrice delle probabilità di transizione $(p_{i,j})_{i,j}$. **Sia P il seguente operatore*[12]

$$P: \ \mathcal{M}_b \to \mathcal{M}_b, \quad f \mapsto Pf, \ dove Pf(i) := \sum_j p_{i,j} h(j), \qquad per\ ogni\ i \tag{3.2}$$

**

Sia inoltre h una **funzione armonica** *rispetto alla matrice delle probabilità di transizione $P = (p_{i,j})_{i,j}$, cioè*

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

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

Il processo

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

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

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

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

---

[11] Si ricorda che la successione $\{X_n\}_n$ è una catena di Markov omogenea con matrice delle probabilità di transizione $(p_{i,j})_{i,j}$ significa che vale la proprietà di Markov, ovvero: qualunque siano n, j, i, $i_{n-1}, \cdots, i_0$

$$\mathbb{P}(X_{n+1} = j | X_n = i,\ X_{n-1} = i_{n-1}, \cdots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j | X_n = i) = p_{i,j},$$

purché $\mathbb{P}(X_n = i,\ X_{n-1} = i_{n-1}, \cdots, X_0 = i_0) > 0$.

[12] Nel caso in cui lo spazio degli stati della catena sia finito (con cardinalità d), $\mathcal{M}_b$ coincide con lo spazio dei vettori $\mathbb{R}^d$, l'operatore P non è altro che una matrice e Pf non è altro che il prodotto righe per colonne della matrice P per il vettore f.

**La dimostrazione si basa sull'osservazione che, essendo $(X_n, n \geq 0)$ una catena di Markov si ha*[13]

$$\mathbb{E}[f(X_{n+1}) \mid \mathcal{F}_n^X] = \mathbb{E}[f(X_{n+1}) \mid X_n, X_{n-1}, \cdots, X_0] = (Pf)(X_n) \tag{3.3}$$

**

Cominciamo con il caso h armonica. Basta controllare che

$$\mathbb{E}[M_{n+1}^h \mid \mathcal{F}_n^X] = \mathbb{E}[M_{n+1}^h \mid X_n, X_{n-1}, \cdots, X_0] = M_n^h,$$

ovvero che

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

**tenendo conto del fatto che $Ph = h$ e che*

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

**il caso h armonica è immediato. Il caso generale deriva dall'osservare che*

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

$$M_{n+1}^f - M_n^f = f(X_{n+1}) - (P-I)f(X_n) - f(X_n) = f(X_{n+1}) - (Pf)(X_n),$$

e quindi

$$\mathbb{E}[M_{n+1}^f - M_n^f \mid \mathcal{F}_n^X] = \mathbb{E}[f(X_{n+1}) - (Pf)(X_n) \mid \mathcal{F}_n^X] = \mathbb{E}[f(X_{n+1}) \mid \mathcal{F}_n^X] - (Pf)(X_n) = 0.$$

[13] Nel caso di variabili aleatorie discrete possiamo applicare i risultati sulle densità condizionali (vedere l'Esempio 2.3): posto $\mathbf{X} = (X_n, X_{n-1}, \cdots, X_0)$, sappiamo che

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

Per la proprietà di Markov $\mathbb{P}(X_{n+1} = j \mid \mathbf{X} = \mathbf{x}) = \mathbb{P}(X_{n+1} = j \mid X_n = x_n) = p_{x_n,j}$, quindi

$$\mathbb{P}(X_{n+1} = j \mid \mathbf{X}) = \mathbb{P}(X_{n+1} = j \mid X_n) = p_{X_n,j}$$

e

$$\sum_j f(j)\mathbb{P}(X_{n+1} = j \mid \mathbf{X} = \mathbf{x}) \left(= \sum_j f(j)\, p_{x_n,j} \right) = (Pf)(x_n),$$

e perciò

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

3.2 Decomposizione di Doob

Tenendo conto dell'Esempio 3.4, possiamo affermare che, se sono soddisfatte delle condizioni di integrabilità, il quadrato della martingala S_n dell'Esempio 3.2 è una submartingala. Nell'Esempio 3.3, si ottiene che $S_n^2 - n\mu$ è una martingala, quindi si può scrivere come la somma di due processi

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

dove M_n è una martingala, ed $A_n = n\sigma^2$ è un processo deterministico crescente.

Il seguente teorema generalizza tale esempio a tutte le submartingale.

Teorema 3.1 (Decomposizione di Doob)**.** *: Sia $(\Omega, \mathcal{F}, \mathbb{P})$ uno spazio di probabilità, e sia $(\mathcal{F}_n)_n \in \mathbb{N}$ una filtrazione con $\mathcal{F}_n \subseteq \mathcal{F}$. Data una $\mathcal{F}_n$-submartingala X_n a tempo discreto, essa si può sempre scrivere in modo unico (a meno di insiemi di misura nulla) come*

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

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

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

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

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

$$\begin{aligned} A_{n+1} - A_n &= \mathbb{E}[X_{n+1} - X_n - (M_{n+1} - M_n) \mid \mathcal{F}_n] \\ &= \mathbb{E}[X_{n+1} - X_n \mid \mathcal{F}_n] - \mathbb{E}[M_{n+1} - M_n \mid \mathcal{F}_n] = \underset{(M_n \text{ è una } \mathcal{F}_n\text{-martingala})}{} \\ &= \mathbb{E}[X_{n+1} - X_n \mid \mathcal{F}_n] = \mathbb{E}[X_{n+1} \mid \mathcal{F}_n] - X_n. \end{aligned}$$

Quindi l'unico modo per definire[14] A_{n+1}, con $A_0 = 0$, è il seguente

$$A_{n+1} := A_{n+1} - A_0 = \sum_{k=0}^{n} (A_{k+1} - A_k) = \sum_{k=0}^{n} (\mathbb{E}[X_{k+1} \mid \mathcal{F}_k] - X_k). \tag{3.4}$$

È immediato verificare[15] che con questa definizione il processo $A_n := \sum_{\ell=1}^{n} (\mathbb{E}[X_\ell \mid \mathcal{F}_{\ell-1}] - X_{\ell-1})$ è integrabile, predicibile e crescente, con $A_0 = 0$.

Si tratta ora solo di verificare che con questa definizione di $\{A_n\}_{n\geq 0}$ il processo $M_n := X_n - X_0 - A_n$ è una martingala, con $M_0 = 0$. Ma ovviamente

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

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

---

[14] Da cui segue l'unicità della decomposizione, in quanto A_n deve essere definito come in (3.4) e poi si dovrà definire necessariamente $M_n := X_n - X_0 - A_n$.

[15] Per l'integrabilità basta osservare che

$$|A_n| \leq \sum_{\ell=1}^{n-1} \big(|\mathbb{E}[X_\ell \mid \mathcal{F}_{\ell-1}]| + |X_{\ell-1}|\big) \underset{\text{(per dis. Jensen)}}{\leq} \sum_{\ell=1}^{n} \big(\mathbb{E}[|X_\ell| \mid \mathcal{F}_{\ell-1}] + |X_{\ell-1}|\big),$$

ed utilizzare il fatto che X_n sono tutte integrabili.
Per la predicibilità, basta osservare che, per ogni $\ell \leq n$, $\mathbb{E}[X_\ell \mid \mathcal{F}_{\ell-1}]$ ed $X_{\ell-1}$ sono $\mathcal{F}_{\ell-1}$-misurabili e quindi $\mathcal{F}_{n-1}$-misurabili.
Per la crescenza basta osservare che, per definizione

$$A_{n+1} - A_n = \mathbb{E}[X_{n+1} \mid \mathcal{F}_n] - X_n \geq 0,$$

dove la disuguaglianza vale in quanto X_n è una submartingala.
Infine il fatto che $A_0 = 0$, è vero per definizione.

Osservazione 3.4. *Ogni processo crescente, integrabile ed adattato è una submartingala, quindi si può applicare il Teorema di Doob e riscriverlo come la somma di un processo predicibile e di una martingala. Più in generale, la somma di un processo crescente, integrabile ed adattato e di una martingala è una submartingala, e quindi, per il Teorema di Doob, si può riscrivere ancora come la somma di un'altra martingala e di un processo crescente che inoltre è predicibile.*

Osservazione 3.5. *Si noti che **dalla dimostrazione del teorema di decomposizione di Doob si deduce che la decomposizione di X_n è data da*

$$\begin{aligned} X_n &= X_0 + \sum_{\ell=1}^{n} (\mathbb{E}[X_\ell \mid \mathcal{F}_{\ell-1}] - X_{\ell-1}) + \sum_{\ell=1}^{n} \big[X_\ell - X_{\ell-1} - (\mathbb{E}[X_\ell \mid \mathcal{F}_{\ell-1}] - X_{\ell-1})\big] \\ &= X_0 + \sum_{\ell=1}^{n} (\mathbb{E}[X_\ell \mid \mathcal{F}_{\ell-1}] - X_{\ell-1}) + \sum_{\ell=1}^{n} \big(X_\ell - \mathbb{E}[X_\ell \mid \mathcal{F}_{\ell-1}]\big). \end{aligned}$$

****

Inoltre nella dimostrazione del teorema di decomposizione di Doob, il fatto che X_n sia una submartingala è servito solo nei seguenti punti:
(i) *ha senso calcolare la media condizionata di $\mathbb{E}[X_{n+1} - X_n \mid \mathcal{F}_n]$, in quanto X_k è integrabile per ogni k,*
(ii) *il valore atteso condizionato $\mathbb{E}[X_{n+1} - X_n \mid \mathcal{F}_n]$ coincide con $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] - X_n$ in quanto X_n è adattato alla filtrazione $\{\mathcal{F}_n\}$,*
(iii) *il processo A_n risulta crescente (in senso lato), in quanto $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] - X_n \geq 0$.*

Si vede quindi che il procedimento si applica a qualunque processo che sia integrabile ed $\mathcal{F}_n$-adattato, si ottiene però una decomposizione nella somma di una martingala e di un processo predicibile (che, in generale, non è crescente).

*Sarebbe interessante rivedere, sotto questa luce, l'Esempio 3.8 **sulle catene di Markov: in fondo applicando il procedimento della decomposizione di Doob al processo $\mathcal{F}_n^X$-adattato e integrabile $f(X_n)$, si ottiene che*

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

***ricordando l'espressione del valore condizionato $\mathbb{E}[f(X_{k+1}) \mid \mathcal{F}_k$ (data in (3.3)) e la definizione dell'operatore P (data in (3.2))*

$$\begin{aligned} &= f(X_n) - \sum_{k=0}^{n-1} \big(Pf(X_k) - f(X_k)\big) + \sum_{k=0}^{n-1} \big(Pf(X_k) - f(X_k)\big) \\ &= f(X_n) - \sum_{k=0}^{n-1} \big(P - I\big) f(X_k) + \sum_{k=0}^{n-1} \big(P - I\big) f(X_k) = M_n^f + \sum_{k=0}^{n-1} \big(P - I\big) f(X_k) \\ &= f(X_0) + [M_n^f - f(X_0)] + \sum_{k=0}^{n-1} \big(P - I\big) f(X_k). \end{aligned}$$

In effetti il processo

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

è un processo $\mathcal{F}_n^X$-predicibile, in quanto chiaramente A_n^f è $\mathcal{F}_{n-1}^X$-misurabile.

3.2.1 Applicazioni: variazione quadratica e integrale stocastico a tempo discreto

Esempio 3.9 (Variazione quadratica predicibile di una martingala)**.** *Se M_n è una martingala di quadrato integrabile, allora M_n^2 è una submartingala (come sappiamo dall'Esempio 3.4).*

Se $M_0 = 0$, allora la decomposizione di Doob in questo caso diviene:

$$M_n^2 = \sum_{k=1}^{n} \mathbb{E}[\Delta(M^2)_k | \mathcal{F}_{k-1}] + \sum_{k=1}^{n} \big(\Delta(M^2)_k - \mathbb{E}[\Delta(M^2)_k | \mathcal{F}_{k-1}]\big)\,,$$

dove $\Delta(M^2)_k := M_k^2 - M_{k-1}^2$.

Così, detto **variazione quadratica predicibile**, *o* **caratteristica quadratica**, *il processo predicibile definito da*

$$< M >_n := \sum_{k=1}^{n} \mathbb{E}[\Delta(M^2)_k | \mathcal{F}_{k-1}] = \sum_{k=1}^{n} \mathbb{E}[M_k^2 - M_{k-1}^2 | \mathcal{F}_{k-1}], \tag{3.5}$$

e definita

$$m_n := \sum_{k=1}^{n} \left(\Delta(M^2)_k - \mathbb{E}[\Delta(M^2)_k | \mathcal{F}_{k-1}] \right),$$

si ha la decomposizione di Doob

$$M_n^2 = < M >_n + m_n.$$

Va menzionato il fatto che, essendo M_n *una martingala,*

$$< M >_k - < M >_{k-1} := \mathbb{E}[\Delta(M^2)_k | \mathcal{F}_{k-1}] = \mathbb{E}[(\Delta M_k)^2 | \mathcal{F}_{k-1}], \tag{3.6}$$

ossia $\mathbb{E}[M_k^2 - M_{k-1}^2 | \mathcal{F}_{k-1}] = \mathbb{E}[(M_k - M_{k-1})^2 | \mathcal{F}_{k-1}]$, *come si vede facilmente*[16], *e quindi*

$$< M >_n := \sum_{k=1}^{n} \mathbb{E}[\Delta(M^2)_k | \mathcal{F}_{k-1}] = \sum_{k=1}^{n} \mathbb{E}\left[(M_k - M_{k-1})^2 | \mathcal{F}_{k-1} \right]. \tag{3.7}$$

Va infine menzionato anche il fatto che il processo

$$[M]_n := \sum_{k=1}^{n} (\Delta M_k)^2 = \sum_{k=1}^{n} (M_k - M_{k-1})^2$$

viene detto **variazione quadratica (opzionale)**. *Si noti inoltre che* $[M]_n$, *essendo un processo adattato e crescente è una submartingala e che* $[M]_n - < M >_n$ *è una martingala, ossia* $< M >_n$ *è il processo crescente e predicibile della decomposizione di Doob, relativo alla submartingala* $[M]_n$.

Esempio 3.10 (Decomposizione dell'integrale stocastico a tempo discreto)**.** *Ci mettiamo nelle stesse ipotesi e notazioni dell'Esempio 3.9 precedente:* M_n *è una martingala di quadrato integrabile. Se* γ_n *è un processo predicibile, di quadrato integrabile sappiamo (si veda l'Esempio 3.6) che l'integrale stocastico a tempo discreto*

$$I_n(\gamma) := \sum_{k=1}^{n} \gamma_k (M_k - M_{k-1})$$

è una martingala, con

$$\Delta I_k(\gamma) \big(:= I_k(\gamma) - I_{k-1}(\gamma) \big) = \gamma_k (M_k - M_{k-1}).$$

Si consideri ora il caso in cui γ_k *è limitato, ovvero esiste un* $L \in \mathbb{R}^+$ *tale che* $|\gamma_k(\omega)| \le L$. *Si osservi che*

$$\begin{aligned} I_n^2(\gamma) &= \sum_{k=1}^{n} \gamma_k (M_k - M_{k-1}) \sum_{h=1}^{n} \gamma_h (M_h - M_{h-1}) \\ &= \sum_{k=1}^{n} \gamma_k^2 (M_k - M_{k-1})^2 + 2 \sum_{k=1}^{n-1} \gamma_k (M_k - M_{k-1}) \sum_{h=k+1}^{n} \gamma_h (M_h - M_{h-1}), \end{aligned}$$

[16] Infatti

$$\begin{aligned} \Delta(M^2)_k &= M_k^2 - M_{k-1}^2 = (M_{k-1} + \Delta M_k)^2 - M_{k-1}^2 = M_{k-1}^2 + 2\, M_{k-1}\, \Delta M_k + (\Delta M_k)^2 - M_{k-1}^2 \\ &= 2\, M_{k-1}\, \Delta M_k + (\Delta M_k)^2, \end{aligned}$$

da cui, per la $\mathcal{F}_{k-1}$-misurabilità di M_{k-1},

$$\begin{aligned} \mathbb{E}[\Delta(M^2)_k | \mathcal{F}_{k-1}] &= \mathbb{E}[2\, M_{k-1}\, \Delta M_k | \mathcal{F}_{k-1}] + \mathbb{E}[(\Delta M_k)^2 | \mathcal{F}_{k-1}] \\ &= 2\, M_{k-1}\, \mathbb{E}[\Delta M_k | \mathcal{F}_{k-1}] + \mathbb{E}[(\Delta M_k)^2 | \mathcal{F}_{k-1}], \end{aligned}$$

e quindi, poiché $\mathbb{E}[\Delta M_k | \mathcal{F}_{k-1}] = 0$, in quanto M_n è una martingala,

$$\mathbb{E}[\Delta(M^2)_k | \mathcal{F}_{k-1}] = \mathbb{E}[(\Delta M_k)^2 | \mathcal{F}_{k-1}].$$

da cui

$$\mathbb{E}\left[I_n^2(\gamma)\right] = \mathbb{E}\left[\sum_{k=1}^{n}\gamma_k^2\left(M_k - M_{k-1}\right)^2 + 2\sum_{k=1}^{n-1}\gamma_k\left(M_k - M_{k-1}\right)\sum_{h=k+1}^{n}\gamma_h\left(M_h - M_{h-1}\right)\right]$$
$$= \sum_{k=1}^{n}\mathbb{E}\left[\gamma_k^2\left(M_k - M_{k-1}\right)^2\right] + 2\sum_{k=1}^{n-1}\sum_{h=k+1}^{n}\mathbb{E}\left[\gamma_k\left(M_k - M_{k-1}\right)\gamma_h\left(M_h - M_{h-1}\right)\right].$$

È chiaro che se $|\gamma_k(\omega)| \le L$, *allora* $\mathbb{E}\left[I_n^2(\gamma)\right]$ *risulta finita*[17].

Ora si può vedere direttamente[18] *che*

$$\mathbb{E}\left[I_n^2(\gamma)\right] = \sum_{k=1}^{n}\mathbb{E}\left[\gamma_k^2(< M >_k - < M >_{k-1})\right]$$
$$= \sum_{k=1}^{n}\mathbb{E}\left[\gamma_k^2\left(M_k - M_{k-1}\right)^2\right] = \mathbb{E}\left[\sum_{k=1}^{n}\gamma_k^2\left(M_k - M_{k-1}\right)^2\right],$$

tuttavia si può procedere anche in un altro modo.

Come visto, se γ_k *è limitato, allora* $I_n(\gamma)$ *è una martingala di quadrato integrabile. Per le formule (3.5) e (3.7) dell'Esempio precedente, applicate alla martingala* $I_n(\gamma)$, *si ha*

$$I_n^2(\gamma) =< I(\gamma) >_n + \mathcal{M}_n,$$

con

$$< I(\gamma) >_n := \sum_{k=1}^{n}\mathbb{E}[I_k^2(\gamma) - I_{k-1}^2(\gamma)|\mathcal{F}_{k-1}] = \sum_{k=1}^{n}\mathbb{E}\left[\left(I_k(\gamma) - I_{k-1}(\gamma)\right)^2|\mathcal{F}_{k-1}\right],$$

ed $\mathcal{M}_n$ *una martingala a media nulla.*

Ovviamente

$$\left(I_k(\gamma) - I_{k-1}(\gamma)\right)^2 = \gamma_k^2(M_k - M_{k-1})^2,$$

da cui

$$\mathbb{E}[\left(I_k(\gamma) - I_{k-1}(\gamma)\right)^2|\mathcal{F}_{k-1}] = \mathbb{E}[\gamma_k^2\left(M_k - M_{k-1}\right)^2|\mathcal{F}_{k-1}]$$
$$= \gamma_k^2\,\mathbb{E}[(M_k - M_{k-1})^2|\mathcal{F}_{k-1}] = \gamma_k^2\left(< M >_k - < M >_{k-1}\right),$$

[17] Se $|\gamma_k(\omega)| \le L$, ed M_n è quadrato integrabile, allora

$$\mathbb{E}\left[\gamma_k^2\left(M_k - M_{k-1}\right)^2\right] \le L^2\,\mathbb{E}\left[\left(M_k - M_{k-1}\right)^2\right] < \infty$$

e, per la disuguaglianza di Cauchy,

$$\mathbb{E}\left[|\gamma_k|\,|M_k - M_{k-1}|\,|\gamma_h|\,|M_h - M_{h-1}|\right] \le L^2\mathbb{E}\left[\left(M_k - M_{k-1}\right)^2\right]^{1/2}\mathbb{E}\left[\left(M_h - M_{h-1}\right)^2\right]^{1/2} < \infty.$$

[18] Si osservi che

$$\mathbb{E}\left[\gamma_k^2\left(M_k - M_{k-1}\right)^2\right] = \mathbb{E}\left[\mathbb{E}\left[\gamma_k^2\left(M_k - M_{k-1}\right)^2|\mathcal{F}_{k-1}\right]\right] = \mathbb{E}\left[\gamma_k^2\mathbb{E}\left[\left(M_k - M_{k-1}\right)^2|\mathcal{F}_{k-1}\right]\right]$$
$$=\mathbb{E}\left[\gamma_k^2(< M >_k - < M >_{k-1})\right],$$

mentre, per $k < h$,

$$\mathbb{E}\left[\gamma_k\left(M_k - M_{k-1}\right)\gamma_h\left(M_h - M_{h-1}\right)\right] = \mathbb{E}\left[\mathbb{E}\left[\gamma_k\left(M_k - M_{k-1}\right)\gamma_h\left(M_h - M_{h-1}\right)|\mathcal{F}_{h-1}\right]\right]$$
$$= \mathbb{E}\left[\gamma_k\left(M_k - M_{k-1}\right)\gamma_h\mathbb{E}\left[\left(M_h - M_{h-1}\right)|\mathcal{F}_{h-1}\right]\right] = \mathbb{E}\left[\gamma_k\left(M_k - M_{k-1}\right)\gamma_h\,0\right] = 0.$$

Quindi, tenendo conto dell'espressione trovata per $\mathbb{E}\left[I_n^2(\gamma)\right]$ si ottiene

$$\mathbb{E}\left[I_n^2(\gamma)\right] = \sum_{k=1}^{n}\mathbb{E}\left[\gamma_k^2(< M >_k - < M >_{k-1})\right].$$

ovvero

$$< I(\gamma) >_n = \sum_{k=1}^{n} \gamma_k^2 (< M >_k - < M >_{k-1}).$$

In altre parole

$$\begin{aligned} \mathcal{M}_n &:= I_n^2(\gamma) - < I(\gamma) >_n \\ &= I_n^2(\gamma) - \sum_{k=1}^{n} \gamma_k^2 (< M >_k - < M >_{k-1}) \end{aligned}$$

è una martingala a media nulla. E ciò implica che

$$\mathbb{E}\left[I_n^2(\gamma)\right] = \sum_{k=1}^{n} \mathbb{E}\left[\gamma_k^2 (< M >_k - < M >_{k-1})\right].$$

Esercizio 3.1. *Si supponga che, nel precedente Esempio 3.10, γ_k goda della proprietà che*

$$\mathbb{E}\left[\sum_{k=1}^{n} \gamma_k^2 (< M >_k - < M >_{k-1})\right] < \infty,$$

senza assumere la condizione di limitatezza.

*Si dimostri che allora il processo $I_n(\gamma)$ è **una martingala di quadrato integrabile, ossia che*

- *$I_n(\gamma)$ è una martingala*
- $\mathbb{E}\left[I_n^2(\gamma)\right] < \infty$

e che

- $< I^2(\gamma) >_n = \sum_{k=1}^{n} \gamma_k^2 (< M >_k - < M >_{k-1})$.

Si noti che di conseguenza $\mathbb{E}\left[I_n^2(\gamma)\right] = \mathbb{E}\left[\sum_{k=1}^{n} \gamma_k^2 (< M >_k - < M >_{k-1})\right] < \infty$.

soluzione: *Si definisca $\gamma_k^{(L)} = \gamma_k \wedge L$. Si cominci col dimostrare che*

$$\begin{aligned} \mathbb{E}[\gamma_k^2 (M_k - M_{k-1})^2] &= \lim_{L \nearrow \infty} \mathbb{E}[(\gamma^{(L)})_k^2 (M_k - M_{k-1})^2] \\ &= \lim_{L \nearrow \infty} \mathbb{E}\big[\mathbb{E}[(\gamma^{(L)})_k^2 (M_k - M_{k-1})^2 | \mathcal{F}_{k-1}]\big] = \lim_{L \nearrow \infty} \mathbb{E}\big[(\gamma^{(L)})_k^2 \, \mathbb{E}[(M_k - M_{k-1})^2 | \mathcal{F}_{k-1}]\big] \\ &= \lim_{L \nearrow \infty} \mathbb{E}\big[(\gamma^{(L)})_k^2 \, \mathbb{E}[< M >_k - < M >_{k-1} | \mathcal{F}_{k-1}]\big] = \mathbb{E}\big[\gamma_k^2 (< M >_k - < M >_{k-1})\big] < \infty, \end{aligned}$$

e di conseguenza anche $\mathbb{E}[|\gamma_k |M_k - M_{k-1}|] < \infty$. Ricordando l'osservazione **(i)** dell'Esempio 3.6 sull'integrale stocastico a tempo discreto possiamo allora affermare che $I_n(\gamma)$ è una martingala.
Per la disuguaglianza di Cauchy, si deduce che anche i prodotti $\gamma_k (M_k - M_{k-1}) \gamma_j (M_j - M_{j-1})$ sono integrabili, e quindi che la martingala $I_n(\gamma)$ è di quadrato integrabile. La parte rimanente, ed in particolare l'eguaglianza $\mathbb{E}[\gamma_k (M_k - M_{k-1}) \gamma_h (M_h - M_{h-1})] = 0$, per $k \neq h$, si dimostra esattamente con lo stesso procedimento usato nell'Esempio 3.10, nel caso in cui γ_k è limitato (si veda la nota 18).

3.2.2 Applicazioni: verso l'integrale stocastico a tempo continuo

Questa sezione richiede la conoscenza dei processi a tempo continuo. Se il lettore non è familiare con tali processi può tranquillamente saltare questa parte e rimandarne la lettura.

Esempio 3.11 (Primi passi verso l'integrale stocastico a tempo continuo)**.** *Sia $(X_t)_{t\in[0,T]}$ una $\mathcal{G}_t$-martingala di quadrato integrabile.*

Sia inoltre $(f(s))_{s\in[0,T]}$ un ***processo elementare $\mathcal{G}_t$-predicibile***[19]*, di quadrato integrabile, ovvero un processo definito da*

$$f(s,\omega) := \sum_{k=1}^{N} H_k(\omega)\mathbf{1}_{(t_{k-1},t_k]}(s),$$

per una partizione $\{t_k, k = 0,\cdots,N\}$ di $(0,T]$, con $0 = t_0 < t_1 < \cdots < t_N = T$, dove H_k sono variabili aleatorie $\mathcal{G}_{t_{k-1}}$-misurabili e di quadrato integrabile.

Si definiscano

$$\int_\alpha^\beta f(s,\omega)dX(s,\omega) := \sum_{k=1}^{N} H_k(\omega)\big(X_{\beta\wedge t_k\vee\alpha} - X_{\beta\wedge t_{k-1}\vee\alpha}\big), \tag{3.8}$$

e, nel caso $\alpha = 0$ e $\beta = t$

$$\mathcal{I}_t^X(f)(\omega) := \int_0^t f(s,\omega)dX(s,\omega) = \sum_{k=1}^{N} H_k(\omega)\big(X_{t_k\wedge t} - X_{t_{k-1}\wedge t}\big), \tag{3.9}$$

$$= \sum_{k=1}^{n} H_k(\omega)\big(X_{t_k} - X_{t_{k-1}}\big) + H_{n+1}(\omega)\big(X_t - X_{t_n}\big) \qquad \text{per } t_n \le t < t_{n+1} \tag{3.10}$$

che è indicato anche, più brevemente come

$$\mathcal{I}_t^X(f) := \int_0^t f(s)dX(s).$$

Si ha che il processo $(\mathcal{I}_t^X(f))_{t\in[0,T]}$ è una martingala. Si osservi che, dalla (3.10)*, si deduce immediatamente che $\mathcal{I}_t^X(f)$ ha traiettorie continue, se $(X_t)_t$ ha traiettorie continue). Se inoltre, se le variabili H_k sono limitate, allora il processo $(\mathcal{I}_t^X(f))_{t\in[0,T]}$ è di quadrato integrabile. Inoltre, il processo definito da*

$$\mathcal{M}_t^f := \big(\mathcal{I}_t^X(f)\big)^2 - \left(\sum_{k=1}^{n} H_k^2(\omega)\mathbb{E}[(X_{t_k} - X_{t_{k-1}})^2|\mathcal{G}_{t_{k-1}}] + H_{n+1}^2(\omega)\mathbb{E}[(X_t - X_{t_n})^2|\mathcal{G}_{t_n}]\right), \qquad \text{per } t_n \le t < t_{n+1} \tag{3.11}$$

è una $\mathcal{G}_t$-martingala, ed in particolare si ha

$$\mathbb{E}\left[\big(\mathcal{I}_t^X(f)\big)^2\right] = \mathbb{E}\left[\sum_{k=1}^{n} H_k^2(\omega)\mathbb{E}\left[(X_{t_k} - X_{t_{k-1}})^2|\mathcal{G}_{t_{k-1}}\right] + H_{n+1}^2(\omega)\mathbb{E}[(X_t - X_{t_n})^2|\mathcal{G}_{t_n}]\right], \qquad \text{per } t_n \le t < t_{n+1} \tag{3.12}$$

Le precedenti proprietà si dimostrano utilizzando i risultati dei precedenti Esempi 3.6 e 3.10, notando che nella partizione $\{t_k, k = 0,\cdots,N\}$ che definisce $f(s)$ si può sempre supporre[20] *che ci sia un indice h tale che $t = t_h$. In realtà nella dimostrazione della proprietà di martingala servono due tempi $t' \le t''$ e si deve mostrare che*

[19] Per la definizione generale di processo predicibile rispetto ad una filtrazione, si veda ???

[20] Infatti se $t \notin \{t_k : k = 0,\cdots,N\}$, allora esiste un ℓ tale che $t \in (t_{\ell-1}, t_\ell)$ e quindi

$$H_\ell(\omega)\mathbf{1}_{(t_{\ell-1},t]}(s) = H_\ell(\omega)\mathbf{1}_{(t_{\ell-1},t_\ell]}(s) + H_\ell(\omega)\mathbf{1}_{(t,t_\ell]}(s).$$

Ovviamente H_ℓ è $\mathcal{G}_t$-misurabile, in quanto Sia $\{t'_k, k = 0,\cdots,N+1\}$ la partizione ottenuta dalla partizione $\{t_k, k = 0,\cdots,N\}$ inserendo il punto t al posto $\ell+1$ (ovvero $t'_k = t_k$ per $k \le \ell$, $t'_{\ell+1} = t$, e $t'_k = t_{k-1}$ per i rimanenti valori di k) e sia $\{H'_k, k = 0,\cdots,N+1\}$ la famiglia di variabili aleatorie ottenuta in modo analogo dalla famiglia $\{H_k, k = 0,\cdots,N\}$ inserendo la variabile aleatoria H_ℓ al posto $\ell+1$ (ovvero $H'_k = H_k$ per $k \le \ell$, $H'_{\ell+1} = t$, e $H'_k = H_{k-1}$ per i rimanenti valori di k). Allora ovviamente

$$f(s,\omega) := \sum_{k=1}^{N} H_k(\omega)\mathbf{1}_{(t_{k-1},t_k]}(s) = \sum_{k=1}^{N+1} H'_k(\omega)\mathbf{1}_{(t'_{k-1},t'_k]}(s).$$

$\mathbb{E}\left[\mathcal{I}^X_{t''}(f)|\mathcal{G}_{t'}\right] = \mathcal{I}^X_{t'}(f)$ *e che* $\mathbb{E}[\mathcal{M}^f_{t''}|\mathcal{G}_{t'}] = \mathcal{M}^f_{t'}$. *Basta pensare che entrambi facciano parte della partizione* $\{t_k, k = 0, \cdots, N\}$.

A titolo di esempio si consideri la proprietà di martingala dell'integrale stocastico $\mathcal{I}^X_t(f)$ *(tralasciando le proprietà di misurabilità e di integrabilità: si tratta di dimostrare che per ogni* $t' \le t''$

$$\mathbb{E}\left[\mathcal{I}^X_{t''}(f)|\mathcal{G}_{t'}\right] = \mathcal{I}^X_{t'}(f). \tag{3.13}$$

Se nella partizione $\{t_k, k = 0, \cdots, N\}$, *si ha* $t' = t_m \le t'' = t_n$ *(con* $m \le n$*), posto*

$$M_k = X_{t_k}, \quad \mathcal{F}_k = \mathcal{G}_{t_k}, \quad \gamma_k = f(t_k) = H_k,$$

si ottiene che $(M_n)_n$ *è una martingala rispetto alla filtrazione* $(\mathcal{F}_n)_n$, *e, con le notazione dell'Esempio 3.10, che*

$$\mathcal{I}^X_{t'}(f) = I_m(\gamma), \quad \mathcal{I}^X_{t''}(f) = I_n(\gamma)$$

e quindi la (3.13) diviene

$$\mathbb{E}\left[I_n(\gamma)|\mathcal{F}_m\right] = I_m(\gamma),$$

che è esattamente la proprietà di martingala dell'integrale stocastico a tempo discreto dell'Esempio 3.6.
Infine si noti che, la (3.10)*, si può riscrivere, per* $t \in [t_k, t_{k+1}]$, *come*

$$\mathcal{I}^X_t(f) = \mathcal{I}^X_{t_k}(f) + H_{k+1}\left(X_t - X_{t_k}\right),$$

e che, entrambe queste espressioni hanno come conseguenza che, se la martingala $(X_t)_t$ *è una martingala a traiettorie continue, allora anche* $(\mathcal{I}^X_t(f))_t$ *è una martingala a traiettorie continue.*

Supponiamo ora che esista un processo crescente e adattato[21] $< X >= (< X >_t)_{t\in[0,T]}$, *tale che* $X_t^2 - < X >_t$ *è una martingala. In altre parole il processo* $< X >$ *è la variazione quadratica della martingala* X. *Allora si può dimostrare facilmente*[22] *che, per ogni* $0 \le s \le t \le T$,

$$\mathbb{E}\big[(X_t - X_s)^2\big|\mathcal{G}_s\big] = \mathbb{E}\big[< X >_t - < X >_s \big|\mathcal{G}_s\big], \tag{3.14}$$

che a sua volta è la versione a tempo continuo della (3.6).

Allora le (3.11) *e la* (3.12) *si possono riscrivere, come*

$$\mathcal{M}^f_t := \left(\mathcal{I}^X_t(f)\right)^2 - \int_0^t f^2(s)\, d < X >_s, \qquad \text{è una martingala, a media nulla} \tag{3.15}$$

$$\mathbb{E}\left[\left(\mathcal{I}^X_t(f)\right)^2\right] = \mathbb{E}\left[\int_0^t f^2(s)\, d < X >_s\right], \tag{3.16}$$

Si noti che, in entrambe le rappresentazioni, si ha che $H_k = f(t_k)$, con $f(t_k)$ che risulta $\mathcal{G}_{t_{k-1}}$ misurabile, e $H'_k = f(t'_k)$, con $f(t'_k)$ che risulta $\mathcal{G}_{t'_{k-1}}$ misurabile.

Si noti infine che anche l'integrale stocastico relativo non cambia cambiando rappresentazione, ovvero se f elementare e predicibile ammette due diverse rappresentazioni, l'integrale stocastico è sempre definito dallo stesso processo, che è come dire che la definizione è ben posta.

[21] Per essere la variazione quadratica (predicibile) deve avere anche la proprietà di essere predicibile. Per comodità del lettore riportiamo la definizione di processi $(\mathcal{G}_t)_{t\in[0,T]}$-predicibili: i processi $(\mathcal{G}_t)_{t\in[0,T]}$-predicibili sono i processi da $[0,T]\times\Omega \to \mathbb{R}$ e misurabili rispetto alla σ-algebra su $[0,T]\times\Omega$, generata dai processi elementari e $(\mathcal{G}_t)_{t\in[0,T]}$-predicibili, o equivalentemente dai processi $(\mathcal{G}_t)$-adattati e continui a sinistra. Più precisamente si definisce $\mathcal{P} = \mathcal{P}\big((\mathcal{G}_t)_t\big)$ la σ-algebra su $[0,T]\times\Omega$ generata dai processi elementari predicibili. La definizione di processo predicibile viene data dopo....????

[22] La dimostrazione della (3.14), si basa sul fatto che

$$(X_t - X_s)^2 - [< X >_t - < X >_s] = X_t^2 - X_s^2 - 2\,X_s\,(X_t - X_s) - < X >_t + < X >_s = X_t^2 - < X >_t - [X_s^2 - < X >_s] - 2\,X_s\,(X_t - X_s).$$

Infatti, passando ai valori attesi condizionali si ha che, essendo X_s una variabile aleatoria $\mathcal{G}_s$ misurabile,

$$\mathbb{E}\big[(X_t - X_s)^2 - [< X >_t - < X >_s]\big|\mathcal{G}_s\big] = \mathbb{E}\big[X_t^2 - < X >_t - [X_s^2 - < X >_s]\big|\mathcal{G}_s\big] - 2\,X_s\,\mathbb{E}\big[(X_t - X_s)\big|\mathcal{G}_s\big] = 0,$$

dove l'ultima uguaglianza dipende dal fatto che $X_t^2 - < X >_t$ e X_t sono martingale.

tenendo conto del fatto che

$$\begin{aligned}
&\sum_{k=1}^{n} H_k^2(\omega)\mathbb{E}\left[(X_{t_k}-X_{t_{k-1}})^2|\mathcal{G}_{t_{k-1}}\right]+H_{n+1}^2(\omega)\mathbb{E}\left[(X_t-X_{t_n})^2|\mathcal{G}_{t_n}\right]\\
&=\sum_{k=1}^{n} H_k^2(\omega)\mathbb{E}\left[<X>_{t_k}-<X>_{t_{k-1}}|\mathcal{G}_{t_{k-1}}\right]+H_{n+1}^2(\omega)\mathbb{E}\left[<X>_t-<X>_{t_n}|\mathcal{G}_{t_n}\right]\\
&=\sum_{k=1}^{n} H_k^2(\omega)\left[<X>_{t_k}-<X>_{t_{k-1}}\right]+H_{n+1}^2(\omega)\left[<X>_t-<X>_{t_n}\right]\\
&=\int_0^t f^2(s)\,d<X>_s
\end{aligned}$$

Questa identità è la base per poi definire l'integrale stocastico rispetto ad una martingala di quadrato integrabile, che ammetta come variazione quadratica (predicibile) il processo $<X>$. *Infine, si noti che, a sua volta, la relazione* (3.15) *si può esprimere dicendo che la variazione quadratica (predicibile) di* $\mathcal{I}_t^X(f)$ *coincide con l'integrale di* f *rispetto a* $d<X>_s$, *ossia*

$$<\mathcal{I}^X(f)>_t=\int_0^t f^2(s)\,d<X>_s\,.$$

Esercizio 3.2. *Dare la dimostrazione diretta del fatto che* $^{**}(\mathcal{I}_t^X(f))_t$ *e* $(\mathcal{M}_t^f)_t$ *sono martingale.*

soluzione: *La misurabilità e l'integrabilità di* $\mathcal{I}_t^X(f)$ *(se le variabili* H_k *sono di quadrato integrabile) e di* $\mathcal{M}_t^f$ *(se le variabili* H_k *sono limitate) sono banali da verificare.*

Per ***verificare la proprietà di martingala di* $\mathcal{I}_{t'}^X(f)$ *si osservi che, se* $^{**}t_{i-1}<t'\le t_i\le t_j\le t''<t_{j+1}$, *posto* $t'_0=t'\le {}^{**}t'_1=t_i\le\cdots\le t'_{m-1}=t'_j\le t'_m=t''$, *e posto* $H'_{k+1}=f(t'_{k+1})$, *allora* H'_{k+1} *è* $^{**}\mathcal{G}_{t'_k}$*-misurabile, per* $k=0,\cdots,m-1$, *e*

$$\begin{aligned}
\tilde{\mathbb{E}}[I_{t''}^X(f)-I_{t'}^X(f)\mid\mathcal{G}_{t'}]&=\sum_{k=0}^{m}\mathbb{E}[H'_{k+1}(X_{t'_{k+1}}-X_{t'_k})\mid\mathcal{G}_{t'_0}]=\sum_{k=0}^{m}\mathbb{E}\big[\mathbb{E}[H'_{k+1}(X_{t'_{k+1}}-X_{t'_k})\mid\mathcal{G}_{t'_k}]\mid\mathcal{G}_{t'_0}\big]\\
&\underset{(H'_{k+1}\ \text{è}\ ^{**}\mathcal{G}_{t'_k}-mis.)}{=}\sum_{k=0}^{m}\mathbb{E}\big[H'_{k+1}\,\mathbb{E}[(X_{t'_{k+1}}-X_{t'_k})\mid\mathcal{G}_{t'_k}]\mid\mathcal{G}_{t'_0}\big]\underset{(X_t\ \text{è una}\ \mathcal{G}_t-MG)}{=}\sum_{k=0}^{m}\mathbb{E}\big[H'_{k+1}\ 0\big]=0
\end{aligned}$$

Per quanto riguarda $\mathcal{M}_t^f$, *con la stessa tecnica usata precedentemente, ci si convince subito che basta dimostrare che* $\mathbb{E}[\mathcal{M}_{t''}^f|\mathcal{G}_{t\prime}]=\mathcal{M}_{t'}^f$ *per due tempi* t' *e* t'', *che sono due tempi consecutivi di una partizione, ossia, per cui*

$$\mathcal{I}_{t''}^X(f)=\mathcal{I}_{t'}^X(f)+H'\left(X_{t''}-X_{t'}\right),\quad \text{con } H'=f(t''),\ \text{che è } \mathcal{G}_{t'}\text{-misurabile.}$$

A questo punto basta osservare che

$$\begin{aligned}
\mathcal{M}_{t''}^f-\mathcal{M}_{t'}^f&=\left(\mathcal{I}_{t'}^X(f)+H'\left(X_{t''}-X_{t'}\right)\right)^2-\left(\mathcal{I}_{t'}^X(f)\right)^2-(H')^2\,\mathbb{E}\left[(X_{t''}-X_{t'})^2\mid\mathcal{G}_{t'}\right]\\
&=\left(\mathcal{I}_{t'}^X(f)\right)^2+\left(H'\left(X_{t''}-X_{t'}\right)\right)^2-2\,\mathcal{I}_{t'}^X(f)\,H'\,(X_{t''}-X_{t'})-\left(\mathcal{I}_{t'}^X(f)\right)^2-(H')^2\,\mathbb{E}\left[(X_{t''}-X_{t'})^2\mid\mathcal{G}_{t'}\right]\\
&=(H')^2\,(X_{t''}-X_{t'})^2-2\,\mathcal{I}_{t'}^X(f)\,H'\,(X_{t''}-X_{t'})-(H')^2\,\mathbb{E}\left[(X_{t''}-X_{t'})^2\mid\mathcal{G}_{t'}\right]
\end{aligned}$$

da cui, passando ai valori attesi condizionali, e sfruttando il fatto che H' *e* $\mathcal{I}_{t'}^X(f)$ *sono* $G_{t'}$*-misurabili, si ottiene*

$$\begin{aligned}
\mathbb{E}\left[\mathcal{M}_{t''}^f-\mathcal{M}_{t'}^f\mid\mathcal{G}_{t'}\right]&=\mathbb{E}\left[(H')^2\,(X_{t''}-X_{t'})^2-2\,\mathcal{I}_{t'}^X(f)\,H'\,(X_{t''}-X_{t'})-(H')^2\,\mathbb{E}\left[(X_{t''}-X_{t'})^2\mid\mathcal{G}_{t'}\right]\mid\mathcal{G}_{t'}\right]\\
&=(H')^2\,\mathbb{E}\left[(X_{t''}-X_{t'})^2\mid\mathcal{G}_{t'}\right]-2\,\mathcal{I}_{t'}^X(f)\,H'\,\mathbb{E}\left[X_{t''}-X_{t'}\mid\mathcal{G}_{t'}\right]-(H')^2\,\mathbb{E}\left[(X_{t''}-X_{t'})^2\mid\mathcal{G}_{t'}\right]=0
\end{aligned}$$

.

Esercizio 3.3. *Si dimostri che non è necessario fare l'ulteriore ipotesi di limitatezza di* f *(ossia di* H_k*) per avere l'integrabilità della martingala* $\mathcal{M}_t^f$, *ma basta supporre che*

$$\mathbb{E}\left[\int_0^T f^2(s)\,d<X>_s\right]=\mathbb{E}\left[\sum_{k=1}^{N}H_k^2(\omega)\ \left[<X>_{t_k}-<X>_{t_{k-1}}\right]\right]<\infty$$

soluzione: *Posto* $t'_k = t_k$ *per* $k \le n$ *e* $t'_{n+1} = t$ *si ha*

$$\left(\sum_{k=1}^{n} H_k(\omega)\big(X_{t_k} - X_{t_{k-1}}\big) + H_{n+1}(\omega)\big(X_t - X_{t_n}\big)\right)^2 = \left(\sum_{k=1}^{n+1} H_k(\omega)\big(X_{t'_k} - X_{t'_{k-1}}\big)\right)^2$$
$$= \sum_{k=1}^{n+1}\Big(H_k(\omega)\big(X_{t'_k} - X_{t'_{k-1}}\big)\Big)^2 + \sum_{\substack{k=1\\ **k\neq j**}}^{n+1}\sum_{j=1}^{n+1} H_k(\omega)\big(X_{t'_k} - X_{t'_{k-1}}\big)\, H_j(\omega)\big(X_{t'_j} - X_{t'_{j-1}}\big)$$

Ciascun termine del tipo

$$\big(H_k(\omega)\big)^2\,\big(X_{t'_k} - X_{t'_{k-1}}\big)^2$$

ammette valore atteso (finito o no, a priori) che è dato dal limite

$$\lim_{M\nearrow\infty} \mathbb{E}\left[(H_k(\omega)\wedge M)^2\,\big(X_{t'_k} - X_{t'_{k-1}}\big)^2\right] = \lim_{M\nearrow\infty} \mathbb{E}\left[(H_k(\omega)\wedge M)^2\,\mathbb{E}\big[\big(X_{t'_k} - X_{t'_{k-1}}\big)^2|\mathcal{G}_{t'_{k-1}}\big]\right]$$
$$= \lim_{M\nearrow\infty} \mathbb{E}\left[(H_k(\omega)\wedge M)^2\,\mathbb{E}\big[< X >_{t'_k} - < X >_{t'_{k-1}}|\mathcal{G}_{t'_{k-1}}\big]\right] = \lim_{M\nearrow\infty} \mathbb{E}\left[(H_k(\omega)\wedge M)^2\,\big(< X >_{t'_k} - < X >_{t'_{k-1}}\big)\right]$$
$$= \mathbb{E}\left[\big(H_k(\omega)\big)^2\,\big(< X >_{t'_k} - < X >_{t'_{k-1}}\big)\right]$$

Per l'integrabilità dei termini del tipo $H_k(\omega)\big(X_{t'_k} - X_{t'_{k-1}}\big)\, H_j(\omega)\big(X_{t'_j} - X_{t'_{j-1}}\big)$*, basta poi usare la disuguaglianza di Cauchy* ***(con il procedimento della nota 18, la dimostrazione che il valore atteso di ciascuno di questi termini è nullo).****

L'esempio che segue è in realtà un'anticipazione, in quanto richiede la conoscenza del processo di Wiener, che in queste note si trova nei capitoli successivi. La lettura di questo esempio va quindi rinviata e deve essere effettuata dopo aver introdotto tale processo.

Esempio 3.12. *Se nell'Esempio 3.11 si prende* $X_t = W_t$*, il processo di Wiener standard, e* $\mathcal{G}_t = \mathcal{F}_t^W$ *(oppure* $\mathcal{G}_t = \mathcal{F}_t$*, se si tratta di un processo di Wiener rispetto alla filtrazione* $(\mathcal{F}_t)_t$*)si ottiene*[23] *che*

$$\mathbb{E}[(X_{t_k} - X_{t_{k-1}})^2|\mathcal{G}_{t_{k-1}}] = \mathbb{E}[(W_{t_k} - W_{t_{k-1}})^2|\mathcal{F}^W_{t_{k-1}}] = t_k - t_{k-1}$$

e quindi che

$$\mathcal{M}_t^f := \mathcal{I}_t^2(f) - \sum_k H_k^2(\omega)(t_k\wedge t - t_{k-1}\wedge t) = \Big(\int_0^t f(s)dW(s)\Big)^2 - \int_0^t f^2(s)ds \tag{3.17}$$

è una martingala, ed infine che

$$\mathbb{E}[\mathcal{I}_t^2(f)] = \mathbb{E}\left[\Big(\int_0^t f(s)dW(s)\Big)^2\right] = \mathbb{E}\left[\int_0^t f^2(s)ds\right]. \tag{3.18}$$

Si noti che saremmo arrivati alla stessa conclusione anche nel caso di una martingala $(X_t)_{t\ge 0}$ *di quadrato integrabile, con variazione quadratica (predicibile)* $< X >_t = t$*. Se inoltre la martingala ha traiettorie continue, allora anche l'integrale stocastico* $\mathcal{I}_t^X(f)$*, sempre per processi elementari, ha traiettorie continue.*

[23] Si ricordi che il processo di Wiener W_t è un processo ad incrementi indipendenti, e quindi $W_{t_k} - W_{t_{k-1}}$ è una variabile aleatoria gaussiana $N(0, t_k - t_{k-1})$ indipendente dalla σ-algebra $\mathcal{F}^W_{t_{k-1}}$. Di conseguenza

$$\mathbb{E}[(W_{t_k} - W_{t_{k-1}})^2|\mathcal{F}^W_{t_{k-1}}] = \mathbb{E}[(W_{t_k} - W_{t_{k-1}})^2] = t_k - t_{k-1}$$

in quanto $\mathbb{E}[(W_{t_k} - W_{t_{k-1}})^2] = Var(W_{t_k} - W_{t_{k-1}})$.

Conclusione In questa sezione abbiamo mostrato come si può definire l'integrale stocastico di $(f(s))_{s\in[0,T]}$ rispetto ad una $\mathcal{G}_t$-martingala X_t, con $(f(s))_{s\in[0,T]}$ **processo elementare $\mathcal{G}_t$-predicibile**, ovvero[24]

$$f(s,\omega) := \sum_{k=1}^{N} H_k(\omega)\mathbf{1}_{(t_{k-1},t_k]}(s),$$

per una partizione $\{t_k, k = 0,\cdots,N\}$ di $(0,T]$, con $0 = t_0 < t_1 < \cdots < t_N = T$, dove H_k sono variabili aleatorie $\mathcal{G}_{t_{k-1}}$-misurabili:

$$\mathcal{I}_t^X(f)(\omega) := \int_0^t f(s,\omega)dX(s,\omega) = \sum_{k=1}^{N} H_k(\omega)\big(X_{t_k\wedge t} - X_{t_{k-1}\wedge t}\big),$$

Se inoltre la martingala $(X_t)_{t\geq 0}$ è di quadrato integrabile, con variazione quadratica (predicibile) $< X >_t$ (ossia con $X_t^2 - < X >_t$ una martingala), allora, nell'ipotesi che H_k siano variabili aleatorie limitate, oppure (si veda l'Esercizio 3.3) nell'ipotesi che

$$\mathbb{E}\left[\int_0^T f^2(s)\, d < X >_s\right] = \mathbb{E}\left[\sum_{k=1}^{N} H_k^2(\omega)\ \left[< X >_{t_k} - < X >_{t_{k-1}}\right]\right] < \infty,$$

si ottiene che

$$\big(\mathcal{I}_t^X(f)\big)^2 - \int_0^t f^2(s)\, d < X >_s, \qquad \text{è una martingala, a media nulla,}$$

ossia

$$< \mathcal{I}^X(f) >_t = \int_0^t f^2(s)\, d < X >_s .$$

Infine se $(X_t)_{t\geq 0}$ è una martingala di quadrato integrabile, con variazione quadratica (predicibile) $< X >_t = t$, ed è a traiettorie continue, allora il processo $\mathcal{I}_t^X(f)$ è a traiettorie continue e

$$< \mathcal{I}^X(f) >_t = \int_0^t f^2(s)\, ds,$$

per ogni funzione elementare con $\mathbb{E}\left[\int_0^t f^2(s)\, ds\right] < \infty$.

[24] **Attenzione:** nella parte relativa all'integrale stocastico (Sezione 6), al posto di H_k, che è $\mathcal{F}_{t_{k-1}}$-misurabile, si scrive C_{k-1}, che è $\mathcal{F}_{t_{k-1}}$-misurabile, ossia si considerano i processi del tipo

$$\sum_{k=1}^{N} C_{k-1}(\omega)\mathbf{1}_{(t_{k-1},t_k]}(s),$$

ponendo cioè $C_{k-1} = H_k$. Ciò può ingenerare qualche confusione, ma confidiamo nell'intelligenza del lettore....

3.3 Martingale, submartingale e tempi d'arresto

Definizione 3.4. . *Sia $\{\mathcal{F}_t\}$ una filtrazione e sia $\tau : \Omega \to \mathbb{R}^+ \cup \{\infty\}$, una variabile aleatori (la v.a. τ deve prendere valori negli indici del tempo preso in considerazione e quindi, ad esempio, in $\mathbb{N}$ se si tratta tempo discreto, ma può anche prendere il valore infinito). La v.a. τ si dice* ***tempo d'arresto*** *(o* ***stopping time****) rispetto ad $\{\mathcal{F}_t\}$, se per ogni t*

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

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

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

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

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

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

3.3.1 Tempo discreto

Nel caso in cui l'insieme dei tempi sia $\mathbb{N}$ è equivalente chiedere che $\{\tau = n\} \in \mathcal{F}_n$ per ogni n, come si vede facilmente[25].

Esempio 3.13. *Con le stesse notazioni dell'Esempio 3.3, e per ogni $a \in \mathbb{R}$,*

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

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

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

Definizione 3.5. *Dato un tempo d'arresto τ, si definisce* **la σ-algebra degli eventi fino al tempo** τ *come*

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

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

Si tratta cioè degli eventi, per i quali stabilire il loro verificarsi insieme al verificarsi di $\{\tau \leq t\}$ dipende solo dall'informazione disponibile fino al tempo t.
Nel caso in cui l'insieme dei tempi sia $\mathbb{N}$ è equivalente chiedere che $A \cap \{\tau = n\} \in \mathcal{F}_n$ per ogni n, come si vede facilmente[26].

---

[25] Ovviamente

$$\{\tau = n\} = \{\tau \leq n\} \backslash \{\tau \leq n-1\},$$

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

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

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

[26] Infatti: se $A \cap \{\tau = n\} \in \mathcal{F}_n$ per ogni n, allora

$$A \cap \{\tau \leq n\} = \cup_{k=0}^n A \cap \{\tau = k\}, \quad A \cap \{\tau = k\} \in \mathcal{F}_k \subseteq \mathcal{F}_n,\ k \leq n,$$

e quindi $A \cap \{\tau \leq n\} \in \mathcal{F}_n$.
Se invece $A \cap \{\tau \leq n\} \in \mathcal{F}_n$ per ogni n, allora

$$A \cap \{\tau = n\} = A \cap \{\tau \leq n\} \backslash A \cap \{\tau \leq n-1\}, \quad A \cap \{\tau \leq n-1\} \in \mathcal{F}_{n-1} \subseteq \mathcal{F}_n,$$

e quindi $A \cap \{\tau = n\} \in \mathcal{F}_n$.

Esercizio 3.4. *Controllare che $\mathcal{F}_\tau$ è una σ-algebra.*
(suggerimento: se $A \in \mathcal{F}_\tau$ allora l'evento $A^c \cap \{\tau \le t\} = \{\tau \le t\} \backslash \{A \cap \{\tau \le t\}\} \in \mathcal{F}_t$)

Esempio 3.14. *Se $\mathcal{F}_n = \sigma\{X_1, \cdots, X_n\}$ e $\tau = \inf\{n \text{ t.c. } X_n \in I\}$, con $I \in \mathcal{B}(\mathbb{R})$, allora τ è un $\mathcal{F}_n$ tempo d'arresto, infatti l'evento $\{\tau \le n\} = \{\exists k \le n \text{ t.c. } X_k \in I\} \in \mathcal{F}_n$.*

3.3.2 Tempo continuo

Per i tempi di uscita, nel caso a tempo continuo, le cose non sono così semplici come nel caso a tempo discreto, però qualcosa si può dire.

Definizione 3.6. *Sia*

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

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

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

Lemma 3.2. *Se X_t è un processo a traiettorie continue e A è aperto, allora τ_A è un tempo d'arresto.*

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

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

e quindi

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

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

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

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

e la prima delle precedenti uguaglianze non sarebbe valida.

□

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

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

Osservazione 3.6. *Affermare che τ è un tempo d'arresto debole è equivalente ad affermare che è un tempo di arresto rispetto alla filtrazione*

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

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

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

e quindi $\{\tau \le t\} \in \mathcal{F}_s$ per ogni $s > t$, ovvero [27] *$\{\tau \le t\} \in \mathcal{F}_{t^+}$ per ogni t.*

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

Lemma 3.3. *Se X_t è un processo con traiettorie continue a destra con limiti a sinistra (*cadlag *acronimo dal francese* continue à droite limite à gauche*), la filtrazione è continua a destra (cioè $\mathcal{F}_t = \mathcal{F}_{t^+} := \bigcap_{s>t} \mathcal{F}_s$) ed F è un chiuso allora τ_F è un tempo d'arresto.*

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

$$\{\tau_F \geq t\} = \bigcap_{0 \leq s < t} \{X_s \in F\},$$

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

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

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

□

3.4 Alcune proprietà dei tempi d'arresto

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

Infatti, per ogni t,

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Per terminare questa sezione, osserviamo che, se una filtrazione soddisfa le cosiddette ***condizioni "abituali"***, cioè

i) $\mathcal{F}_0$ contiene tutti gli eventi trascurabili,

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

allora si possono applicare sia il Lemma 2, ed ottenere così dalla **ii)** che i tempi di uscita da un chiuso sono tempi d'arresto (e non solo tempi d'arresto in senso debole), sia la proprietà 4, e ottenere così dalla **i)** che $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$ ogni volta che $\mathbb{P}(\sigma \le \tau) = 1$.

3.5 Caso a tempo discreto

Proposizione 3.4 (Martingale arrestate)**.** *Data una $\mathcal{F}_n$-martingala (o submartingala) X_n ed un tempo d'arresto τ, il processo*

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

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

Nota bene: si usa anche la notazione

$$X_n^\tau := X_{n\wedge\tau},$$

e il processo $(X_n^\tau)_n$ è detto **martingala arrestata al tempo $\boldsymbol{\tau}$**.

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

$$Y_n = X_\tau \mathbb{I}_{\{\tau \le n\}} + X_n \mathbb{I}_{\{\tau > n\}},$$

e quindi

$$Y_n = X_0 \mathbb{I}_{\{\tau=0\}} + X_1 \mathbb{I}_{\{\tau=1\}} + X_2 \mathbb{I}_{\{\tau=2\}} + \cdots + X_n \mathbb{I}_{\{\tau=n\}} + X_n \mathbb{I}_{\{\tau>n\}}, \tag{3.19}$$

$$Y_n = X_0 \mathbb{I}_{\{\tau=0\}} + X_1 \mathbb{I}_{\{\tau=1\}} + X_2 \mathbb{I}_{\{\tau=2\}} + \cdots + X_n \mathbb{I}_{\{\tau\ge n\}} \tag{3.20}$$

e che $\{\tau = k\} \in \mathcal{F}_k \subseteq \mathcal{F}_n, \{\tau > n\} \in \mathcal{F}_n$ e che X_k sono $\mathcal{F}_n$-misurabili per $k \le n$, da cui la $\mathcal{F}_n$-misurabilità di Y_n. Per ottenere l'integrabilità basta notare che dalla (3.20)

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

e sfruttare l'integrabilità delle X_k.
Per ottenere il resto dobbiamo notare che dalla (3.20), con $n+1$ al posto di n, si ha

$$Y_{n+1} = X_0\mathbb{I}_{\{\tau=0\}} + X_1\mathbb{I}_{\{\tau=1\}} + X_2\mathbb{I}_{\{\tau=2\}} + \cdots + X_n\mathbb{I}_{\{\tau=n\}} + X_{n+1}\mathbb{I}_{\{\tau>n\}}, \tag{3.21}$$

e quindi, confrontando (3.19) e (3.21)

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

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

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

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

**

Esercizio 3.5 (Dimostrazione alternativa della Proposizione 3.4). *Utilizzare l'integrale stocastico a tempo discreto per dimostrare la precedente Proposizione 3.4.*
soluzione: *Si noti che*

$$Y_n = X_{n\wedge\tau} = \sum_{k=1}^{n\wedge\tau}(X_k - X_{k-1}) + X_0 = \sum_{k=1}^{n}\mathbb{I}_{\{\tau\geq k\}}\,(X_k - X_{k-1}) + X_0 = I_n(\gamma) + X_0,$$

*dove $\gamma_k = \mathbb{I}_{\{\tau\geq k\}} = \mathbb{I}_{\{\tau>k-1\}} = \mathbb{I}_{\{\tau\leq k-1\}}$ è $\mathcal{F}_{k-1}$-misurabile. Basta poi osservare che γ_k, oltre che predicibile, è limitato e ricordare l'osservazione **(ii)** dell'Esempio 3.6 sull'integrale stocastico a tempo discreto possiamo allora affermare che se X_n è una submartingala allora $I_n(\gamma)$ è una submartingala.*

**

Proposizione 3.5. *Data una $\mathcal{F}_n$-submartingala (o martingala) X_n ed un tempo d'arresto τ, limitato quasi certamente, ovvero per cui esiste un $n \in \mathbb{N}$, tale che*

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

allora

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

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

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

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

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

e quindi

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

La tesi segue in quanto

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

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

Questa sezione termina con una versione del famoso teorema del campionamento opzionale.

Teorema 3.6 (Optional Sampling Theorem o Teorema del campionamento opzionale). *Sia X_n una submartingala. Siano σ e τ due tempi d'arresto limitati, e tali che*

$$\sigma \leq \tau,$$

allora

$$\mathbb{E}[X_\tau|\mathcal{F}_\sigma] \geq X_\sigma.$$

Dimostrazione. Sia n tale che $\sigma \le \tau \le n$. Allora chiaramente

$$X_\tau = X_0 \mathbb{I}_{\{\tau=0\}} + X_1 \mathbb{I}_{\{\tau=1\}} + X_2 \mathbb{I}_{\{\tau=2\}} + \cdots + X_n \mathbb{I}_{\{\tau=n\}},$$
$$X_\sigma = X_0 \mathbb{I}_{\{\sigma=0\}} + X_1 \mathbb{I}_{\{\sigma=1\}} + X_2 \mathbb{I}_{\{\sigma=2\}} + \cdots + X_n \mathbb{I}_{\{\sigma=n\}}.$$

Ciò mostra che

$$|X_\tau| \le |X_0| + |X_1| + |X_2| + \cdots + |X_n|,$$
$$|X_\sigma| \le |X_0| + |X_1| + |X_2| + \cdots + |X_n|$$

e quindi in particolare che X_τ è integrabile, per cui ha senso calcolare la sua media condizionale.

Inoltre X_σ è $\mathcal{F}_\sigma$-misurabile, infatti, qualunque sia x, l'evento $\{X_\sigma \le x\} \in \mathcal{F}_\sigma$, in quanto per ogni h

$$\{X_\sigma \le x\} \cap \{\sigma = h\} = \{X_h \le x\} \cap \{\sigma = h\} \in \mathcal{F}_h.$$

Infine per mostrare che $\mathbb{E}[X_\tau | \mathcal{F}_\sigma] \ge X_\sigma$, basta mostrare che

$$\mathbb{E}[X_\tau \mathbb{I}_A] \ge \mathbb{E}[X_\sigma \mathbb{I}_A], \quad \forall A \in \mathcal{F}_\sigma,$$

o equivalentemente (essendo $\tau = \tau \wedge n$ e $\sigma = \sigma \wedge n$) che

$$\mathbb{E}[X_{\tau\wedge n} \mathbb{I}_A] \ge \mathbb{E}[X_{\sigma\wedge n} \mathbb{I}_A], \quad \forall A \in \mathcal{F}_\sigma.$$

Infatti

$$\mathbb{E}[X_{\tau\wedge n} \mathbb{I}_A] = \sum_{h\le n} \mathbb{E}[X_n^\tau \mathbb{I}_A \mathbb{I}_{\{\sigma=h\}}] = \sum_{h\le n} \mathbb{E}[X_n^\tau \mathbb{I}_{A\cap\{\sigma=h\}}].$$

Essendo A un evento $\mathcal{F}_\sigma$-misurabile, si ha che $A \cap \{\sigma = h\}$ è un evento di $\mathcal{F}_h$; inoltre, su tale insieme $\tau \wedge h = h$, in quanto $h = \sigma$ e $\sigma \le \tau$; infine, il processo $(X_n^\tau)_n$ è una submartingala. Di conseguenza,

$$\begin{aligned}\mathbb{E}[X_n^\tau \mathbb{I}_{A\cap\{\sigma=h\}}] &\ge \mathbb{E}[X_h^\tau \mathbb{I}_{A\cap\{\sigma=h\}}] \\ &= \mathbb{E}[X_{\tau\wedge h} \mathbb{I}_A \mathbb{I}_{\{\sigma=h\}}] = \mathbb{E}[X_h \mathbb{I}_A \mathbb{I}_{\{\sigma=h\}}].\end{aligned}$$

Per ottenere la tesi basta osservare che, sommando su $h \le n$, si ottiene

$$\mathbb{E}[X_\sigma \mathbb{I}_A].$$

□

3.6 Applicazione: la rovina del giocatore con le martingale

1) caso simmetrico
Sia Y_k una successione di v.a. indipendenti, con

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

e sia

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

Sappiamo (vedi Esempio 3.2) che S_n è una martingala, in quanto se $p = 1/2$ allora $\mathbb{E}[Y_k] = 0$. Siano a e b numeri naturali non nulli e sia

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

La variabile aleatoria τ è finita [28] con probabilità 1, cioè $\mathbb{P}(\tau < +\infty) = 1$, e quindi il gioco finisce in un tempo finito.

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

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

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

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

Di conseguenza

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

da cui immediatamente

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

2) caso generale

Come nell'applicazione **1)** ma con $\mathbb{P}(Y_k = 1) = p$ e $\mathbb{P}(Y_k = -1) = q = 1 - p$.

Si procede[29] in modo analogo al caso precedente, ma questa volta si prende come martingala

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

dove

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

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

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

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

---

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

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

Infatti, si ha $\{\tau = +\infty\} = \bigcap_{n\ge1}\{\tau > n(a+b)\}$ e $\mathbb{P}(\tau > n(a+b)) \le \alpha^n$ per un $\alpha < 1$. La prima uguaglianza è ovvia, mentre la seconda si può vedere facilmente osservando che

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

e che quindi $\mathbb{P}(\bigcup_{k<n}\{Y_{k(a+b)+1} = Y_{k(a+b)+2} = \cdots = Y_{k(a+b)+a+b} = 1\}) \le \mathbb{P}(\tau \le n(a+b))$

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

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

La dimostrazione è finita in quanto α^n tende a zero.

Si noti che in sostanza la precedente dimostrazione si riduce a dimostrare che $\tau(\omega) \le T(\omega)(a+b)$, dove T è la variabile aleatoria geometrica di parametro $\beta = p^{a+b} = 1 - \alpha$, definita come

$$T(\omega) = k + 1 \quad \Leftrightarrow \quad \omega \in B_k \cap \left(\cap_{h<k} B_h^c\right)$$

dove $B_k = \{Y_{k(a+b)+1} = \cdots = Y_{k(a+b)+a+b} = 1\}^c$. Allora la v.a. τ è finita, essendo la v.a. T finita.

[29] In questo caso $\mathbb{E}[Y_k] = 2p - 1 \ne 0$. Se si prendesse la martingala a media nulla $M_n := S_n - (2p-1)n$ ci sarebbero due problemi:
il primo è che la corrispondente martingala arrestata $M_{n\wedge\tau}$ non è limitata, per cui pur convergendo a $M_\tau = S_\tau - (2p-1)\tau$, **non** possiamo immediatamente dire che anche i valori attesi di $M_{n\wedge\tau}$ convergono al valore atteso di M_τ,
il secondo è che, anche se avessimo dimostrato che i valori attesi convergono, e quindi si avesse che $0 = \mathbb{E}[M_\tau] = \mathbb{E}[S_\tau - (2p-1)\tau]$, non avremmo finito, in quanto dovremmo conoscere anche il valore atteso di τ.

Di nuovo, sempre per convergenza dominata,

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

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

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

Si noti che

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

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

Per concludere questa sezione proponiamo un esercizio: mostrare che, per $p \neq \frac{1}{2}$, la funzione $h(x) := \left(\frac{q}{p}\right)^x$ è armonica rispetto alla matrice P delle probabilità di transizione della passeggiata aleatoria. Nel caso per $p = \frac{1}{2}$, mostrare che lo stesso vale per la funzione $h(x) = x$.

3.7 Disuguaglianza di Kolmogorov per submartingale non negative

Proposizione 3.7 (Disuguaglianza di Kolmogorov)**.** *Sia X_n una submartingala non negativa allora*

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

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

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

Dimostrazione. Cominciamo con il primo caso. Si definisca

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

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

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

Basta quindi mostrare che $\mathbb{E}[X_\tau] \leq \mathbb{E}[X_n]$, ma ciò discende immediatamente dalla precedente Proposizione 3.5, in quanto τ è a valori in $1, 2, \cdots, n$.

Per il caso delle martingale, la tesi segue osservando che

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

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

3.8 Convergenza di martingale

Nell'esempio della rovina del giocatore abbiamo trovato che delle martingale limitate per le quali esisteva il limite per n che converge ad infinito: le martingale $S_{n\wedge\tau}$ nel caso simmetrico e $\left(\frac{q}{p}\right)^{S_{n\wedge\tau}}$, nel caso generale. Questo fatto è un caso particolare di un risultato più generale.

Proposizione 3.8. *Sia X_n una martingala uniformemente limitata in L^1, cioè tale che*

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

Allora

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

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

$$\sup_n \mathbb{E}[|X_n|^2] \leq {**}M^{1/2} < +\infty.$$

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

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

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

ovvero

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

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

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

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

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

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

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

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

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

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

e che

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

Di conseguenza

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

e quindi

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

Non rimane che dimostrare che

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

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

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

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

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

Infatti

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

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

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

3.9 Disuguaglianza di Doob

Proposizione 3.9 (Disuguaglianza di Doob)**.** *Sia data una submartingala non negativa X_n, con $\mathbb{E}[(X_n)^p] < +\infty$, per un $p > 1$. Posto $X^*_n = \max_{k\leq n}(X_k)$ vale la seguente disuguaglianza:*

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

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

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

Dimostrazione. Si definisca

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

Ovviamente τ_γ è un tempo d'arresto e l'evento $\{X^*_n > \gamma\}$ coincide con $\bigcup_{k=1}^{n}\{\tau_\gamma = k\}$.
Quindi per ogni $\beta > 0$

$$(X^*_n)^\beta = \int_0^{X^*_n}\beta\gamma^{\beta-1}d\gamma = \int_0^\infty \beta\gamma^{\beta-1}\mathbb{I}_{\{X^*_n>\gamma\}}d\gamma = \int_0^\infty \beta\gamma^{\beta-1}\sum_{k=1}^{n}\mathbb{I}_{\{\tau_\gamma=k\}}d\gamma$$

Passando al valore medio[30], per $\beta > 0$,

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

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

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

e quindi

$$\begin{aligned}\mathbb{E}[(X^*_n)^\beta] &\leq \sum_{k=1}^{n}\int_0^\infty \beta\gamma^{\beta-2}\mathbb{E}[X_n\mathbb{I}_{\{\tau_\gamma=k\}}]d\gamma = \int_0^\infty \beta\gamma^{\beta-2}\mathbb{E}[X_n\sum_{k=1}^{n}\mathbb{I}_{\{\tau_\gamma=k\}}]d\gamma \\ &= \mathbb{E}[\int_0^\infty \beta\gamma^{\beta-2}X_n\mathbb{I}_{\{X^*_n>\gamma\}}d\gamma],\end{aligned}$$

[30] ****In realtà stiamo scambiando gli integrali in $d\mathbb{P}$ e in $d\gamma$ e bisognerebbe dimostrare che la variabile aleatoria $\mathbb{I}_{\{\tau_\gamma=k\}}$ è una funzione congiuntamente misurabile in (ω, γ) rispetto alla sigma algebra prodotto $\mathcal{F} \times \mathcal{B}$.

ovvero, se $\beta - 1 > 0$,

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

Inoltre, prendendo $\beta = p$, e usando la disuguaglianza di Hölder [31] si ottiene la tesi osservando che

$$\begin{aligned}\mathbb{E}[(X_n^*)^p] &\le \frac{p}{p-1} \mathbb{E}[(X_n)^p]^{1/p} \mathbb{E}\left[\left((X_n^*)^{p-1}\right)^{p/(p-1)}\right]^{(p-1)/p} \\ &= \frac{p}{p-1} \mathbb{E}[(X_n)^p]^{1/p} \mathbb{E}[(X_n^*)^p]^{(p-1)/p}.\end{aligned}$$

Infatti questa disuguaglianza non è banale (ovvero non è del tipo $+\infty \le +\infty$), in quanto si ha che $\mathbb{E}[(X_n^*)^p] = \mathbb{E}[\max_{k=1,\cdots,n}(X_k)^p]] \le \mathbb{E}[\sum_{k=1}^n (X_k)^p] < +\infty$ (il caso banale in cui $\mathbb{E}[(X_n^*)^p] = 0$ si può escludere, in questo caso la tesi del teorema è banalmente vera, ed il caso è poco interessante), e non rimane che dividere ambo i membri della precedente disuguaglianza per $\mathbb{E}[(X_n^*)^p]^{(p-1)/p}$.

□

3.10 Estensioni alle martingale a tempo continuo (cenni)

Quasi tutti i risultati delle sezioni precedenti si estendono al caso di martingale (o submartingale) continue o *cadlag* (cioè continue a destra con limite a sinistra).

Ad esempio, la disuguaglianza di Doob, così come quella di Kolmogorov, si estendono al caso di submartingale X_t a tempo continuo e con traiettorie continue, considerando che

$$\max_{t \in [0,T]} |X_t| = \lim_{n \to \infty} \max_{t = (i/2^n)T, i \le 2^n} |X_t|.$$

Per quanto riguarda le submartingale *cadlag* basta passare dal massimo all'estremo superiore:

$$\sup_{t \in [0,T]} |X_t| = \lim_{n \to \infty} \max_{t = (i/2^n)T, i \le 2^n} |X_t|.$$

Per ottenere le disuguaglianze va anche osservato, che la successione $X^{*,n} = \max_{t=(i/2^n)T, i \le 2^n} |X_t|$ è una successione monotona, e la monotonia permette di passare al limite sotto il segno di valore atteso (o di integrale rispetto a $\mathbb{P}$).

Un poco più delicato è il problema dell'estensione dei risultati che coinvolgono X_τ, quando τ è un tempo d'arresto, come ad esempio il fatto che X_τ è una variabile aleatoria, e addirittura una variabile aleatoria $\mathcal{F}_\tau$-misurabile. A questo problema dedichiamo una sezione alla fine di questo capitolo (Sezione 3.11).

Ad esempio, vale il seguente risultato:

Lemma 3.10. *Un processo stocastico X_t, con traiettorie* cadlag, *e $\mathcal{F}_t$-adattato è una martingala se e solo se*

$$\mathbb{E}[X_\tau] = \mathbb{E}[X_0], \qquad \textit{per ogni tempo d'arresto limitato } \tau. \tag{3.22}$$

Dimostrazione. La parte più semplice è quella relativa alla condizione sufficiente: se vale la (3.22), allora, prima di tutto X_t è una variabile aleatoria integrabile, per ogni t, dato che t è un tempo d'arresto rispetto a qualunque filtrazione. Inoltre per ogni $s \le t$, e per ogni $A \in \mathcal{F}_s$, posto

$$\tau = t\,\mathbb{I}_A + s\,\mathbb{I}_{A^c},$$

allora τ è un $\mathcal{F}_t$-tempo d'arresto e quindi

$$\mathbb{E}[X_\tau] = \mathbb{E}[X_t\,\mathbb{I}_A] + \mathbb{E}[X_s\,\mathbb{I}_{A^c}] = \mathbb{E}[X_0]$$

[31] Si ricordi che, per $p > 1$ si ha

$$\frac{1}{p} + \frac{1}{q} = 1 \quad \Leftrightarrow \quad q = \frac{p}{p-1},$$

e che la disuguaglianza di Hölder garantisce che per tutte le variabili aleatorie

$$\mathbb{E}[|XY|] \le \mathbb{E}[|X|^p]^{\frac{1}{p}} \mathbb{E}[|Y|^q]^{\frac{1}{q}},$$

con la convenzione che se almeno uno dei due valori attesi $\mathbb{E}[|X|^p]$ o $\mathbb{E}[|Y|^q]$ non è finito, allora il secondo membro vale infinito, e la disuguaglianza è banale. Si tratta di una generalizzazione della disuguaglianza di Cauchy, che corrisponde al caso $p = q = 2$.

Ovviamente anche

$$\mathbb{E}[X_s] = \mathbb{E}[X_s \, \mathbb{I}_A] + \mathbb{E}[X_s \, \mathbb{I}_{A^c}] = \mathbb{E}[X_0],$$

e quindi, per ogni $s \leq t$, e per ogni $A \in \mathcal{F}_s$, si ha

$$\mathbb{E}[X_t \, \mathbb{I}_A] = \mathbb{E}[X_s \, \mathbb{I}_A]$$

ovvero X_t è una $\mathcal{F}_t$-martingala.

Per quanto riguarda l'altra implicazione, spieghiamo solo l'idea: se τ è un tempo d'arresto limitato da L, allora anche $\tau_m = \frac{\lceil m\tau \rceil}{m}$ è una successione di tempi d'arresto limitata e a valori in $\{0, \frac{1}{m}, \cdots \frac{\lceil mL \rceil}{m}\}$. Il processo a tempo discreto $Y_k = X_{k/m}$ è una martingala a tempo discreto rispetto alla filtrazione $\mathcal{G}_k = \mathcal{F}_{k/m}$. Allora sappiamo che

$$\mathbb{E}[X_{\tau_m}] = \mathbb{E}[Y_{\lceil m\tau \rceil}] = \mathbb{E}[Y_0] = \mathbb{E}[X_0]$$

Inoltre, essendo X_t a traiettorie continue a destra si ha che la successione X_{τ_m} converge a X_τ. Per ottenere la tesi basta osservare che

$$\lim_{m \to \infty} \mathbb{E}[X_{\tau_m}] = \mathbb{E}[X_\tau],$$

in quanto la successione X_{τ_m} è una successione uniformemente integrabile. L'ultima affermazione segue dall'optional sampling theorem, che assicura che $X_{\tau_m} = \mathbb{E}[X_{L+1} | \mathcal{F}_{\tau_m}]$, e dal fatto, per ogni variabile aleatoria integrabile Z in uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, la famiglia dei valori attesi condizionali $\{\mathbb{E}[Z|\mathcal{G}], \, \mathcal{G}$ sotto σ-algebre di $\mathcal{F}\}$ è una famiglia uniformemente integrabile.

□

L'optional sampling theorem si generalizza al caso di martingale a tempo continuo e *cadlag*. Può essere interessante però sapere che esiste una generalizzazione, anche al caso di martingale uniformemente integrabili e di tempi d'arresto generali: per tali martingale esiste $X_\infty = \lim_{t \to \infty} X_t$ e si ha che per ogni coppia di tempo d'arresto $\sigma \leq \tau$, non necessariamente finiti, si ha

$$X_\sigma = \mathbb{E}[X_\tau | \mathcal{F}_\sigma] = \mathbb{E}[X_\infty | \mathcal{F}_\sigma].$$

3.11 Processi misurabili e tempi d'arresto

Lo scopo principale di questa sezione è affrontare il seguente problema: sotto quali condizioni sul processo X_t possiamo affermare che, se τ è un tempo d'arresto, allora X_τ è una variabile aleatoria, e/o una variabile aleatoria $\mathcal{F}_\tau$-misurabile?

A questo scopo ricordiamo la seguente definizione

Definizione 3.7 (Processi (congiuntamente) misurabili)**.** *Sia $X : [0,T] \times \Omega :\mapsto \mathbb{R}$, $(t,\omega) \mapsto X(t,\omega)\big(= X_t(\omega)\big)$. Sia $\mathcal{B}([0,T]) \times \mathcal{F}$ la σ-algebra prodotto su $[0,T] \times \Omega$, ovvero la σ-algebra generata da $\{J \times A \subseteq [0,T] \times \Omega; \; J \in \mathcal{B}([0,T]), A \in \mathcal{F}\}$. Il processo $(X_t)_t$ si dice* **congiuntamente misurabile***, o più semplicemente* **misurabile***, se la funzione X è $\mathcal{B}([0,T]) \times \mathcal{F}$-misurabile. La definizione è analoga per i processi definiti per tutti i tempi $t \geq 0$: basta sostituire $[0,T]$ con $[0,\infty)$.*

Cominciamo con l'osservare che se τ è una variabile aleatoria a valori in $[0,T]$ (non necessariamente un tempo d'arresto) e $(X_t)_{t \in [0,T]}$ è un processo misurabile, allora X_τ è una variabile aleatoria. Infatti basta considerare che la funzione

$$\big(\Omega, \mathcal{F}\big) \to \big([0,T] \times \Omega, \mathcal{B}([0,T]) \times \mathcal{F}\big) \qquad \omega \mapsto \big(\tau(\omega), \omega\big)$$

è misurabile, e che $\omega \mapsto X_\tau(\omega) = X(\tau(\omega), \omega)$ è misurabile, in quanto composizione di due funzioni misurabili.

Nel caso in cui τ sia un tempo d'arresto (rispetto alla filtrazione $\{\mathcal{F}_t\}$), ciò dimostra solo che X_τ è una variabile aleatoria $\mathcal{F}$-misurabile, ma non ci assicura che sia una variabile aleatoria $\mathcal{F}_\tau$-misurabile. Per ottenere ciò dobbiamo assumere un'ulteriore proprietà sul processo X_t.

Definizione 3.8 (Processi $\{\mathcal{F}_t\}$-progressivamente misurabili)**.** *Sia $X : [0,T] \times \Omega :\mapsto \mathbb{R}$, $(t,\omega) \mapsto X(t,\omega)\big(= X_t(\omega)\big)$. Sia $\mathcal{B}([0,t]) \times \mathcal{F}_t$ la σ-algebra prodotto su $[0,t] \times \Omega$, ovvero la σ-algebra generata da $\{J \times A \subseteq [0,t] \times \Omega; \; J \in \mathcal{B}([0,t]), A \in \mathcal{F}_t\}$. Il processo $(X_t)_t$ si dice* **progressivamente misurabile** *se, per ogni $t \in [0,T]$, la funzione X ristretta a $[0,t] \times \Omega$ è $\mathcal{B}([0,t]) \times \mathcal{F}_t$-misurabile. La definizione è analoga per i processi definiti per tutti i tempi $t \geq 0$: basta sostituire $[0,T]$ con $[0,\infty)$.*

Lemma 3.11. *Se X_t è un processo $\{\mathcal{F}_t\}$-progressivamente misurabile, e τ è un $\{\mathcal{F}_t\}$-tempo d'arresto a valori in $[0,T]$, allora la variabile aleatoria X_τ, è una variabile aleatoria $\mathcal{F}_\tau$-misurabile.*

Dimostrazione. La dimostrazione è banale, infatti[32] basta dimostrare che, per ogni $t \in [0,T]$ la variabile aleatoria $X_\tau \mathbb{I}_{\{\tau \le t\}}$ è $\mathcal{F}_t$-misurabile. Ma ciò è banalmente vero in quanto, posta $\overline{X}^t$ la funzione X ristretta a $[0,t] \times \Omega$, cioè $\overline{X}^t(s,\omega) = X(s,\omega)$, per $s \in [0,t]$ e $\omega \in \Omega$, allora $X_\tau \mathbb{I}_{\{\tau \le t\}} = \overline{X}^t_{\tau \wedge t} \mathbb{I}_{\{\tau \le t\}}$. La tesi segue osservando che $\tau \wedge t$ è una variabile $\mathcal{F}_t$- misurabile e $\overline{X}^t$ è un processo congiuntamente misurabile rispetto a $\mathcal{B}([0,t]) \times \mathcal{F}_t$. □

Per finire basta osservare che ogni processo $\mathcal{F}_t$-adattato e *cadlag* (oppure continuo a sinistra) è progressivamente misurabile. Infatti si ha che, per ogni $s \le t$ $X(s,\omega) = \lim_{n\to\infty} X^{(n,t)}(s,\omega)$, dove, posto $s_k^{(n,t)} = \frac{k}{2^n}\, t$, si ha

$$X^{(n,t)}(s,\omega) := X(0,\omega)\, \mathbf{1}_{\{0\}}(s) + \sum_{k=1}^{2^n} X(s_k^{(n,t)},\omega)\, \mathbf{1}_{(s_{k-1}^{(n,t)}, s_k^{(n,t)}]}(s), \qquad s \le t$$

che, chiaramente, sono processi $\mathcal{B}([0,t]) \times \mathcal{F}_t$-misurabili[33]. Ciò mostra che **il processo $\overline{X}^t$, ossia il processo X ristretto a $[0,t] \times \Omega$,**è anch'esso $\mathcal{B}([0,t]) \times \mathcal{F}_t$-misurabile. Per l'arbitrarietà di t, si ottiene la progressiva misurabilità.

Il caso dei processi continui a sinistra si può trattare in modo simile, prendendo però la successione

$$X^{(n,t)}(s,\omega) := X(0,\omega)\, \mathbf{1}_{\{0\}}(s) + \sum_{k=1}^{2^n} X(s_{k-1}^{(n,t)},\omega)\, \mathbf{1}_{[s_{k-1}^{(n,t)}, s_k^{(n,t)})}(s), \qquad s \le t.$$

Tuttavia in questo caso c'è un metodo molto più semplice: infatti $X(t,\omega) = \lim_{n\to\infty} X(\frac{\lfloor 2^n\, t\rfloor}{2^n},\omega)$, in quanto $\frac{\lfloor 2^n\, t\rfloor}{2^n} \le t$ e $\lim_{n\to\infty} \frac{\lfloor 2^n\, t\rfloor}{2^n} = t$. I processi $X^{(n)} : (t,\omega) \mapsto X(\frac{\lfloor 2^n\, t\rfloor}{2^n},\omega)$ sono chiaramente progressivamente misurabili, come si vede subito da:

$$X^{(n)}(t,\omega) := \sum_{k\ge 1} X(t_{k-1}^{(n)},\omega)\, \mathbf{1}_{[t_{k-1}^{(n)}, t_k^{(n)})}(t),$$

dove $t_k^{(n)} = k/2^n$.

**Vale poi la pena di notare che anche il processo

$$Y^n(t,\omega) := X(0,\omega)\, \mathbf{1}_{\{0\}}(t) + X^{(n)}(t^-,\omega) = X(0,\omega)\, \mathbf{1}_{\{0\}}(t) + \sum_{k\ge 1} X(t_{k-1}^{(n)},\omega)\, \mathbf{1}_{(t_{k-1}^{(n)}, t_k^{(n)}]}(t),$$

converge a $X(t,\omega)$, ed inoltre è continuo a sinistra ed adattato. **

Per maggiori dettagli e ulteriori generalizzazioni si consiglia, ad esempio, il testo di P. Baldi [1].

****Per finire riportiamo qui la definizione di processi opzionali e predicibili, rispetto ad una filtrazione.

Definizione 3.9 (Processi Opzionali e Predicibili)**.** *Sia $\mathcal{O} = \mathcal{O}\big((\mathcal{F}_t)_t\big)$ la σ-algebra su $[0,T] \times \Omega$ generata dai processi $(\mathcal{F}_t)$-adattati e continui a destra. La σ-algebra $\mathcal{O}$ è detta σ-algebra opzionale. I processi $X : [0,T] \times \Omega \to \mathbb{R}$ e che sono $\mathcal{O}$-misurabili sono detti processi opzionali (o $(\mathcal{F}_t)$-opzionali).*

Sia $\mathcal{P} = \mathcal{P}\big((\mathcal{F}_t)_t\big)$ la σ-algebra su $[0,T] \times \Omega$ generata dai processi $(\mathcal{F}_t)$-adattati e continui a sinistra. La σ-algebra $\mathcal{P}$ è detta σ-algebra predicibile. I processi $X : [0,T] \times \Omega \to \mathbb{R}$ e che sono $\mathcal{P}$-misurabili sono detti processi predicibili(o $(\mathcal{F}_t)$-predicibili).

---

[32] **La condizione $X_\tau \mathbb{I}_{\{\tau \le t\}}$ è $\mathcal{F}_t$-misurabile, per ogni $t \in [0,T]$, è sufficiente per la $\mathcal{F}_t$-misurabilità di X_τ. Infatti, se $X_\tau \mathbb{I}_{\{\tau \le t\}}$ è $\mathcal{F}_t$-misurabile, per ogni $t \in [0,T]$, allora, qualunque sia x gli eventi del tipo $\{X_\tau \le x\} \cap \{\tau \le t\} = \{X_\tau \mathbb{I}_{\{\tau \le t\}} \le x\} \cap \{\tau \le t\}$, per ogni $t \in [0,T]$, e quindi, per ogni t (tenendo conto che $\tau \le T$). Tale uguaglianza è ovvia per $x < 0$; per $x \ge 0$ conviene passare al complementare, ossia usare il fatto che $\{X_\tau > x\} \cap \{\tau \le t\} = \{X_\tau \mathbb{I}_{\{\tau \le t\}} > x\} \cap \{\tau \le t\}$

[33] **Un'altra scelta (nel caso di X_t processo cadlag e adattato) potrebbe essere

$$X^{n,t}(s,\omega) = X(\tfrac{\lceil 2^n\, s\rceil}{2^n} \wedge t),$$

la cui convergenza a $X(s,\omega)$ dipende dal fatto che, per $s = t$, si ha $X^{n,t}(t,\omega) = X(t,\omega)$, mentre per $s < t$, si ha definitivamente $\frac{\lceil 2^n\, s\rceil}{2^n} \wedge t = \frac{\lceil 2^n\, s\rceil}{2^n} \ge s$ e quindi $\frac{\lceil 2^n\, s\rceil}{2^n} \wedge t \searrow s$.

Capitolo 4

Processi aleatori a tempo continuo

4.1 Processi aleatori, definizioni ed esempi

Esistono diversi modi di definire un processo stocastico. Il primo consiste nel considerare semplicemente una famiglia di variabili aleatorie.

Definizione 4.1 (Processo stocastico come famiglia di v.a.)**.** *Dato uno spazio* $(\Omega, \mathcal{F}, \mathbb{P})$ *ed un insieme* I *di indici*[1]*, allora una famiglia di variabili aleatorie* $\{X_t : \Omega \mapsto \mathbb{R}(o\,\mathbb{R}^d)$*, per* $t \in I\}$ *è detta* **processo stocastico***, e* $\mathbb{R}$ *(o* $\mathbb{R}^d$*) è detto lo* **spazio degli stati** *del processo.*

In alcuni casi si vuole mettere in evidenza l'evoluzione rispetto al tempo e si preferisce dare una definizione di processo stocastico come funzione aleatoria.

Definizione 4.2 (Processo stocastico come funzione aleatoria)**.** *Dato uno spazio* $(\Omega, \mathcal{F}, \mathbb{P})$ *ed un insieme* I *di indici, si definisce* **processo stocastico (come funzione aleatoria)** *la funzione (misurabile)*[2]

$$X : \Omega \mapsto \mathbb{R}^I; \omega \mapsto \big(t \mapsto X(t, \omega)\big).$$

Una definizione analoga vale nel caso di processi con spazio degli stati $\mathbb{R}^d$*.*

Altre volte si è interessati anche ad alcune proprietà delle traiettorie $t \mapsto X(t, \omega)$, quali ad esempio la continuità. Viene allora naturale dare la definizione di processo a traiettorie continue.

Definizione 4.3 (Processo stocastico come funzione aleatoria continua)**.** *Dato uno spazio* $(\Omega, \mathcal{F}, \mathbb{P})$ *ed un insieme* I *di indici, si definisce* **processo stocastico (come funzione aleatoria continua)** *la funzione (misurabile)*[3]

$$X : \Omega \mapsto C(I; \mathbb{R}); \omega \mapsto \big(t \mapsto X(t, \omega)\big),$$

[1] Tipicamente l'insieme degli indici è $I = [0, \infty)$, o $I = [0, T]$, o ancora $I = \mathbb{N}$ (e in questo caso si parla più propriamente di successioni aleatorie), ma è possibile anche che l'insieme degli indici sia multidimensionale, ad esempio $I = \mathbb{R}^2$ (e in questo caso si parla più propriamente di campi aleatori).

[2] Per rendere la definizione completa andrebbe precisata la σ-algebra che si mette sullo spazio di tutte le funzioni $\mathbb{R}^I$. Di solito si tratta in realtà di richiedere almeno che tutte le funzioni

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

siano variabili aleatorie $\mathcal{F}$-misurabili. La σ-algebra considerata su $\mathbb{R}^I$ è di solito $\mathcal{R}^I$, la σ-algebra del Teorema di Kolmogorov 4.1 e la nota 4.1 corrispondente, per maggiori dettagli si veda l'Appendice 4.9 .

[3] Nel caso $I = [0, T]$ lo spazio $C(I; \mathbb{R})$ è uno spazio metrico con la metrica uniforme:

$$d(x(\cdot), y(\cdot)) := \sup_{t \in [0,T]} |x(t) - y(t)|,$$

e la σ-algebra su $C(I; \mathbb{R})$ è quella generata dagli aperti. Se invece $I = [0, \infty)$ si può, ad esempio, usare la metrica

$$d(x(\cdot), y(\cdot)) = \sum_{N=1}^{\infty} \frac{1}{2^N} \sup_{t \in [0,N]} |x(t) - y(t)| \wedge 1,$$

e di nuovo la σ-algebra su $C([0, \infty); \mathbb{R})$ è quella generata dagli aperti.

dove $C(I;\mathbb{R})$ è lo spazio metrico delle traiettorie continue. Una definizione analoga vale nel caso di processi con spazio degli stati $\mathbb{R}^d$.

Dato un processo si definiscono le funzioni di distribuzione finito-dimensionali.

Definizione 4.4 (Famiglia delle funzioni di distribuzione finito-dimensionali)**.** *Dato un processo stocastico (secondo una delle precedenti definizioni) la famiglia delle funzioni di distribuzione $F_{t_1,\cdots,t_n}$, definite per $n \geq 1$, e $t_1, \cdots t_n$, con $t_i \in I$, come le funzioni di distribuzione congiunte delle variabili $(X_{t_1}, \cdots, X_{t_n})$, ovvero*

$$F_{t_1,\cdots,t_n}(x_1, \cdots, x_n) = \mathbb{P}(X_{t_1} \leq x_1, \cdots, X_{t_n} \leq x_n),$$

è detta famiglia delle funzioni di distribuzione finito-dimensionali del processo $X = (X_t)_{t\in I}$.

Va detto che, così come una variabile aleatoria viene spesso individuata attraverso la sua distribuzione, senza specificare quale sia lo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sul quale è definita, spesso anche un processo viene individuato attraverso le sue distribuzioni finito-dimensionali. Si pone quindi il problema di individuare quali sono le famiglie di distribuzioni che sono effettivamente famiglie di distribuzioni finito-dimensionali di un processo.

Si individuano facilmente due condizioni necessarie (**(C1)** e **(C2)**), dette ***condizioni di consistenza***:

(C1) Sia $k > 1$, sia π una permutazione di $\{1, \cdots, k\}$, si ponga

$$\Phi_\pi : \mathbb{R}^k \to \mathbb{R}^k, (x_1, \cdots, x_k) \to \Phi_\pi(x_1, \cdots, x_k) := (x_{\pi_1}, \cdots, x_{\pi_k}).$$

Si richiede che per ogni $k > 1$, π, $t_1, \cdots, t_k$ e $(x_1, \cdots, x_k)$ valga

$$F_{t_1,\cdots,t_k}(x_1, \cdots, x_k) = F_{t_{\pi_1},\cdots,t_{\pi_k}}(\Phi_\pi(x_1, \cdots, x_k)) = F_{t_{\pi_1},\cdots,t_{\pi_k}}(x_{\pi_1}, \cdots, x_{\pi_k}). \tag{4.1}$$

La precedente condizione **(C1)** è chiaramente necessaria, infatti

$$\begin{aligned} &F_{t_1,\cdots,t_k}(x_1, \cdots, x_k) = \mathbb{P}\{(X_{t_1} \leq x_1, \cdots, X_{t_k} \leq x_k)\} \\ &= \mathbb{P}\{X_{t_{\pi_1}} \leq x_{\pi_1}, \cdots, X_{t_{\pi_k}} \leq x_{\pi_k}\} = F_{t_{\pi_1},\cdots,t_{\pi_k}}(x_{\pi_1}, \cdots, x_{\pi_k}). \end{aligned}$$

(C2) Si richiede che per ogni $k \geq 1$, $t_1, \cdots, t_k, t_{k+1}$ e $(x_1, \cdots, x_k)$ valga

$$\lim_{x_{k+1}\to\infty} F_{t_1,\cdots,t_k,t_{k+1}}(x_1, \cdots, x_k, x_{k+1}) = F_{t_1,\cdots,t_k}(x_1, \cdots, x_k). \tag{4.2}$$

La precedente condizione **(C1)** è chiaramente necessaria, infatti

$$\begin{aligned} &\lim_{x_{k+1}\to\infty} F_{t_1,\cdots,t_k,t_{k+1}}(x_1, \cdots, x_k, x_{k+1}) \\ &= \lim_{x_{k+1}\to\infty} \mathbb{P}(X_{t_1} \leq x_1, \cdots, X_{t_k} \leq x_k, X_{t_{k+1}} \leq x_{k+1}) \\ &= \mathbb{P}(X_{t_1} \leq x_1, \cdots, X_{t_k} \leq x_k) = F_{t_1,\cdots,t_k}(x_1, \cdots, x_k). \end{aligned}$$

È interessante notare che la prima condizione potrebbe essere automaticamente soddisfatta dando le funzioni di distribuzione solo per $t_1 < t_2 < \cdots, < t_k$ (e definendole negli altri casi in modo che la condizione **(C1)** sia soddisfatta). Allora, però, la condizione **(C2)**, va modificata:

(C2′) Sia $k \geq 1$, siano $t_1 < t_2 < \cdots < t_k < t_{k+1}$, allora

$$\lim_{x\to\infty} F_{t_1,\cdots,t_k,t_{k+1}}(x_1, \cdots, x_k, x) = F_{t_1,\cdots,t_k}(x_1, \cdots, x_k), \tag{4.3}$$

$$\begin{aligned} &\lim_{x\to\infty} F_{t_1,\cdots,t_{i-1},t_i,t_{i+1},\cdots,t_{k+1}}(x_1, \cdots, x_{i-1}, x, x_{i+1}, \cdots, x_{k+1}) \\ &= F_{t_1,\cdots,t_{i-1},t_{i+1},\cdots,t_{k+1}}(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_{k+1}), \quad \text{per } i = 1, \cdots, k. \end{aligned} \tag{4.4}$$

Il seguente teorema garantisce che le precedenti condizioni necessarie, sono anche sufficienti.

Teorema 4.1 (di Kolmogorov). *Sia data una famiglia $F_{t_1,\cdots,t_k}$ di funzioni di distribuzione finito-dimensionali consistente, cioè che verifica le condizioni di consistenza* **(C1)** *e* **(C2)** *(ovvero (4.1) e (4.2)), allora esiste uno spazio di probabilità ed un processo aleatorio che ammette $F_{t_1,\cdots,t_k}$ come funzioni di distribuzione finito-dimensionali.*[4]

La tesi rimane valida se valgono le condizioni equivalenti **(C1)** *e* **(C2′)** *(ovvero (4.1), (4.3) e (4.3).*

Vediamo subito delle applicazioni del precedente teorema, mentre per la dimostrazione rimandiamo all'Appendice 4.9.

Esempio 4.1 (Processi a coordinate indipendenti). *Data una famiglia di funzioni di distribuzione $\{F_t, t \in I\}$ ad un tempo, si consideri la famiglia delle funzioni di distribuzione finito-dimensionali $F_{t_1,\cdots,t_k}(x_1,\cdots,x_k) := F_{t_1}(x_1) \times \cdots \times F_{t_k}(x_k)$. Tale famiglia è chiaramente una famiglia consistente e quindi esiste un processo con tali funzioni di distribuzione finito-dimensionali.*
Si noti che non si richiede che I sia numerabile, e che questo esempio, nel caso numerabile, garantisce l'esistenza di successioni di v.a. indipendenti[5] *.*

Esempio 4.2 (Processi gaussiani). *Siano date una funzione $m : I \to \mathbb{R},\ t \to m(t)$ e una funzione, detta* **funzione di correlazione**, *$K : I \times I \to \mathbb{R},\ (t,s) \to K(t,s)$* **definita non negativa**, *cioè tale che comunque scelti $n \geq 1$, $t_1,\cdots,t_n \in I$, $\eta_1,\cdots,\eta_n \in \mathbb{C}$ valga*

$$\sum_{h,k}^{1,n} \eta_h \overline{\eta}_k K(t_h,t_k) \geq 0.$$

Si noti che quindi necessariamente $K(t,s) = K(s,t)$ e $K(t,t) \geq 0$ e che la condizione che la funzione di correlazione sia definita non negativa corrisponde alla richiesta che la matrice $\Gamma = (\Gamma_{i,j} := K(t_i,t_j))_{i,j=1,\cdots,n}$ sia definita non negativa (ovvero positiva in senso lato) qualunque siano t_j, per $j = 1,\cdots,n$.

Sia $F_{t_1,\cdots,t_k}$ la funzione di distribuzione congiunta gaussiana di media $(m(t_1),\cdots,m(t_k))$ e matrice di covarianza Γ definita da $\Gamma_{i,j} = K(t_i,t_j)$ per $i,j = 1,\cdots,k$, e sia $f_{t_1,\cdots,t_k}$ la sua densità.

Per convincersi dell'esistenza di un processo con tali distribuzioni finito-dimensionali si consideri il caso in cui $K(t,s)$ sia strettamente definita positiva e si noti che se $(Y_1,\cdots,Y_k)$ è un vettore aleatorio con componenti indipendenti e ciascuna con distribuzione normale $N(0,1)$, allora il vettore definito da $Z = (Z_1,\cdots,Z_k) = A(Y_1,\cdots,Y_k) + (m(t_1),\cdots,m(t_k)) = AY + m$, dove $A = \Gamma^{1/2}$ (cioè $\Gamma = A^tA = AA^t$), è un vettore con la distribuzione cercata[6].

Per la consistenza della famiglia $F_{t_1,\cdots,t_k}$ così definita, si noti che l'operatore Φ_π di (4.1) è una trasformazione lineare e che trasformazioni lineari di vettori gaussiani sono ancora gaussiani. Inoltre, indicando ancora con Φ_π la matrice associata all'operatore di (4.1), allora $\Phi_\pi Z = \Phi_\pi AY + \Phi_\pi m$, segue una legge gaussiana di media $\Phi_\pi m = (m(t_{\pi_1}),\cdots,m(t_{\pi_k}))$ e con matrice di covarianza $(\Phi_\pi A)(\Phi_\pi A)^t = \Phi_\pi AA^t\Phi_\pi^t = \Phi_\pi\Gamma\Phi_\pi^t = (\Gamma_{\pi_i,\pi_j})_{i,j=1,\cdots,k}$.
Da queste osservazioni si deduce immediatamente che per la densità vale

$$f_{t_1,\cdots,t_k}(x_1,\cdots,x_k) = f_{t_{\pi_1},\cdots,t_{\pi_k}}(\Phi_\pi(x_1,\cdots,x_k)),$$

ovvero la (4.1). Per quanto riguarda la (4.2), basta ricordare che ogni sottovettore di un vettore gaussiano è ancora un vettore gaussiano.

Esempio 4.3 (Processo di Wiener standard). *Come caso particolare dell'Esempio precedente possiamo stabilire l'esistenza del* **processo di Wiener standard**, *detto anche* **moto browniano**, *cioè del processo gaussiano con $m(t) = 0$ e $K(t,s) = t \wedge s$.*

---

[4] Inoltre è sempre possibile prendere come spazio di probabilità lo spazio canonico $\mathbb{R}^I$, come σ-algebra la σ-algebra generata dai cilindri, ovvero dagli insiemi del tipo

$$\mathcal{C} = \{x(\cdot) \in \mathbb{R}^I \mid (x(t_1),x(t_2),\cdots,x(t_k)) \in H\}, \text{ dove } H \in \mathcal{B}(\mathbb{R}^k)$$

ed infine come processo X_t il processo canonico $X_t(x(\cdot)) = x(t)$.

[5] L'esistenza di successioni indipendenti è di solito sottintesa nei teoremi fondamentali del *Calcolo delle Probabilità*, come ad esempio la Legge dei grandi numeri, o il Teorema centrale del limite.

[6] Si veda l'Appendice 1.6

Bisogna ovviamente controllare che la funzione $K(\cdot,\cdot)$ sia definita non negativa. Ciò può essere fatto direttamente, ma con una certa fatica[7].

Un metodo decisamente più probabilistico[8] *è il seguente: consiste nell'osservare che la funzione di correlazione*

$$K(t,s) := Cov(N_t, N_s)$$

di un processo N_t di Poisson standard (cioè con $\lambda = 1$) è proprio $t \wedge s$, ciò è sufficiente[9] *a garantire la sua non negatività. Infatti*

$$\begin{aligned}
Cov(N_t, N_s) &= Cov(N_{t\wedge s}, N_{t\vee s}) \\
&= \mathbb{E}[N_{t\wedge s}(N_{t\wedge s} + N_{t\vee s} - N_{t\wedge s})] - \mathbb{E}[N_{t\wedge s}]\mathbb{E}[N_{t\wedge s} + N_{t\vee s} - N_{t\wedge s}] \\
&= \mathbb{E}[N_{t\wedge s}N_{t\wedge s}] + \mathbb{E}[N_{t\wedge s}(N_{t\vee s} - N_{t\wedge s})] - \mathbb{E}[N_{t\wedge s}](\mathbb{E}[N_{t\wedge s}] + \mathbb{E}[N_{t\vee s} - N_{t\wedge s}]) \\
&= Var(N_{t\wedge s}) + \mathbb{E}[N_{t\wedge s}(N_{t\vee s} - N_{t\wedge s})] - \mathbb{E}[N_{t\wedge s}]\mathbb{E}[N_{t\vee s} - N_{t\wedge s}] = Var(N_{t\wedge s}) \\
&= t \wedge s
\end{aligned}$$

Si noti che la proprietà del processo di Wiener di avere la stessa funzione di correlazione del processo di Poisson standard, implica che gli incrementi sono non correlati. Poiché gli incrementi del processo di Wiener $(W_t)_{t\geq 0}$ hanno

[7] Dati $t_1 < t_2 < \cdots < t_k$ la matrice $(t_i \wedge t_j)_{i,j}$ si può scrivere come

$$\begin{pmatrix} t_1 & t_1 & \cdots & t_1 & t_1 \\ t_1 & t_2 & \cdots & t_2 & t_2 \\ \vdots & & \ddots & & \vdots \\ t_1 & t_2 & \cdot & t_{k-1} & t_{k-1} \\ t_1 & t_2 & \cdot & t_{k-1} & t_k \end{pmatrix}$$

Si può dimostrare che il determinante di questa matrice è $t_1(t_2 - t_1)(t_3 - t_2)\cdots(t_k - t_{k-1}) > 0$, e ciò dimostra subito il fatto che $K(t,s) = t \wedge s$ è definita positiva. Il primo passo per dimostrare tale identità è osservare che

$$det \begin{pmatrix} t_1 & t_1 & \cdots & t_1 & t_1 \\ t_1 & t_2 & \cdots & t_2 & t_2 \\ \vdots & & \ddots & & \vdots \\ t_1 & t_2 & \cdot & t_{k-1} & t_{k-1} \\ t_1 & t_2 & \cdot & t_{k-1} & t_k \end{pmatrix} = det \begin{pmatrix} t_1 & 0 & \cdots & 0 & 0 \\ t_1 & t_2 - t_1 & \cdots & t_2 - t_1 & t_2 - t_1 \\ \vdots & & \ddots & & \vdots \\ t_1 & t_2 - t_1 & \cdot & t_{k-1} - t_1 & t_{k-1} - t_1 \\ t_1 & t_2 - t_1 & \cdot & t_{k-1} - t_1 & t_k - t_1 \end{pmatrix}$$

da cui

$$det \begin{pmatrix} t_1 & t_1 & \cdots & t_1 & t_1 \\ t_1 & t_2 & \cdots & t_2 & t_2 \\ \vdots & & \ddots & & \vdots \\ t_1 & t_2 & \cdot & t_{k-1} & t_{k-1} \\ t_1 & t_2 & \cdot & t_{k-1} & t_k \end{pmatrix} = t_1\, det \begin{pmatrix} t_2 - t_1 & \cdots & t_2 - t_1 & t_2 - t_1 \\ \vdots & \ddots & & \vdots \\ t_2 - t_1 & \cdot & t_{k-1} - t_1 & t_{k-1} - t_1 \\ t_2 - t_1 & \cdot & t_{k-1} - t_1 & t_k - t_1 \end{pmatrix},$$

poi basta procedere per induzione, notando che $t_i - t_1 - (t_2 - t_1) = t_i - t_2$.

[8] Tuttavia questo metodo presuppone la conoscenza del processo di Poisson, che in queste note viene definito come processo ad incrementi indipendenti in Sezione 4.6.

[9] La matrice di covarianza di un vettore aleatorio è sempre definita non negativa, e di conseguenza la funzione di correlazione $K(s,t) = Cov(X_s, X_t)$ di un processo X_t qualsiasi è sempre definita non negativa:

$$\begin{aligned}
&\sum_{h,k}^{1,n} \eta_h \overline{\eta}_k K(t_h, t_k) = \sum_{h,k}^{1,n} \eta_h \overline{\eta}_k \mathbb{E}\left[(X_{t_h} - \mathbb{E}[X_{t_h}])(X_{t_k} - \mathbb{E}[X_{t_k}])\right] \\
&= \mathbb{E}\left[\sum_{h,k}^{1,n} \eta_h \overline{\eta}_k (X_{t_h} - \mathbb{E}[X_{t_h}])(X_{t_k} - \mathbb{E}[X_{t_k}])\right] \\
&= \mathbb{E}\left[\sum_{h=1}^{n} \eta_h (X_{t_h} - \mathbb{E}[X_{t_h}]) \sum_{k=1}^{n} \overline{\eta}_k (X_{t_k} - \mathbb{E}[X_{t_k}])\right] \\
&= \mathbb{E}\left[\left|\sum_{h=1}^{n} \eta_h (X_{t_h} - \mathbb{E}[X_{t_h}])\right|^2\right] \geq 0.
\end{aligned}$$

distribuzione gaussiana, sono quindi anche indipendenti. Inoltre per ogni $s < t$, l'incremento $W_t - W_s$ deve essere una variabile aleatoria gaussiana (in quanto differenza di due v.a. congiuntamente gaussiane), deve avere media nulla e varianza[10] *uguale alla varianza di $N_t - N_s$, ovvero uguale a $t - s$. Infine deve essere $\mathbb{P}(W_0 = 0) = 1$, in quanto la media deve essere nulla e la varianza deve essere uguale a $K(0,0) = 0$.*

Questa proprietà di indipendenza degli incrementi è fondamentale per ottenere la densità congiunta di $(W_{t_1}, \cdots, W_{t_n})$. Infatti si può pensare che il vettore $(W_{t_1}, \cdots, W_{t_n})$ si ottiene dal vettore degli incrementi

$$(\Delta W_{t_1}, \Delta W_{t_2}, \cdots, \Delta W_{t_n}) = (W_{t_1} - W_0, W_{t_2} - W_{t_1} \cdots, W_{t_n} - W_{t_{n-1}})$$

con la seguente trasformazione lineare

$$W_{t_h} = W_{t_h} - W_0 = \sum_{i=1}^{h} (W_{t_i} - W_{t_{i-1}}) = \sum_{i=1}^{h} \Delta W_{t_i},$$

(dove si è posto $t_0 = 0$) che corrisponde ad una matrice B la cui inversa B^{-1} è la matrice la cui diagonale ha tutti gli elementi uguali ad 1, la cui sottodiagonale ha gli elementi tutti uguali a -1 e tutti i rimanenti elementi uguali a 0, in quanto banalmente

$$\Delta W_{t_h} = W_{t_h} - W_{t_{h-1}}.$$

Da questa osservazione, dalla formula di trasformazione della densità e dall'osservazione che gli incrementi sono indipendenti, ovvero che

$$f_{\Delta W_{t_1}, \cdots, \Delta W_{t_n}}(y_1, \cdots, y_n) = f_{\Delta W_{t_1}}(y_1) \cdots f_{\Delta W_{t_n}}(y_n) = \prod_{h=1}^{n} g_{t_h - t_{h-1}}(y_h),$$

dove

$$g_s(y) = \frac{1}{\sqrt{2\pi s}} \exp\{-\frac{y^2}{2s}\}$$

si ottiene la densità congiunta[11] *di $(W_{t_1}, \cdots, W_{t_n})$*

$$f_{W_{t_1}, \cdots, W_{t_n}}(x_1, x_2, \cdots, x_n) = f_{\Delta W_{t_1}, \cdots, \Delta W_{t_n}}(x_1, x_2 - x_1 \cdots, x_n - x_{n-1}) \tag{4.5}$$

$$= \prod_{h=1}^{n} g_{t_h - t_{h-1}}(x_h - x_{h-1}) \tag{4.6}$$

$$= \frac{1}{\sqrt{2\pi t_1}} e^{-\frac{x_1^2}{2t_1}} \frac{1}{\sqrt{2\pi(t_2 - t_1)}} e^{-\frac{(x_2 - x_1)^2}{2(t_2 - t_1)}} \cdots \frac{1}{\sqrt{2\pi(t_n - t_{n-1})}} e^{-\frac{(x_n - x_{n-1})^2}{2(t_n - t_{n-1})}} \tag{4.7}$$

4.2 Osservazione sulla definizione di un processo solo attraverso le sue distribuzioni finito dimensionali

La descrizione di un processo attraverso la famiglia delle distribuzioni finito dimensionali non è sufficiente a stabilire le proprietà delle sue traiettorie. In questa sezione e nella successiva ci si pone il problema di spiegare meglio questa affermazione, senza alcuna pretesa di essere esaurienti. Il primo concetto che ci serve è l'equivalenza stocastica per due processi aleatori, che, come si vedrà immediatamente dopo la definizione, implica che i due processi hanno le stesse distribuzioni finito dimensionali.

Definizione 4.5 (Equivalenza stocastica)**.** *Due processi aleatori $(X_t, t \in I)$ e $(Y_t, t \in I)$ sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ si dicono* **stocasticamente equivalenti** *se*

[10] Il lettore che non conosce il processo di Poisson, può procedere anche nel seguente semplicissimo modo

$$\begin{aligned} Var(W_t - W_s) &= Cov(W_t - W_s, W_t - W_s) = Cov(W_t, W_t) - Cov(W_t, W_s) - Cov(W_s, W_t) + Cov(W_s, W_s) \\ &= K(t,t) - 2K(s,t) + K(s,s) = t - 2s \wedge t + s = t - s. \end{aligned}$$

Si noti che il procedimento permette di calcolare la varianza degli incrementi per un qualunque processo, purché sia nota la funzione di covarianza $K(s,t)$.

[11] Per ottenere la desnità si procede come nell'Esempio 1.15.

$$\mathbb{P}(X_t(w) = Y_t(\omega)) = 1 \text{ per ogni } t \in I.$$

(Si dice anche che $(Y_t, t \in I)$ è una **versione** *di $(X_t, t \in I)$)*

Lemma 4.2. *Due processi stocasticamente equivalenti hanno le stesse distribuzioni finito-dimensionali.*

Dimostrazione. Basta osservare che

$$\begin{aligned}
&\mathbb{P}\big(\{(X_{t_1}, ..., X_{t_k}) \in H\}\big) \\
&= \mathbb{P}\Big(\{(X_{t_1}, ..., X_{t_k}) \in H\} \cap \bigcap_{i=1}^{k} \{X_{t_i} = Y_{t_i}\}\Big) + \mathbb{P}\Big(\{(X_{t_1}, ..., X_{t_k}) \in H\} \cap \Big(\bigcap_{i=1}^{k} \{X_{t_i} = Y_{t_i}\}\Big)^c\Big) \\
&= \mathbb{P}\Big(\{(Y_{t_1}, ..., Y_{t_k}) \in H\} \cap \bigcap_{i=1}^{k} \{X_{t_i} = Y_{t_i}\}\Big) = \mathbb{P}\Big(\{(Y_{t_1}, ..., Y_{t_k}) \in H\}\Big)
\end{aligned}$$

in quanto

$$\mathbb{P}\Big(\Big(\bigcap_{i=1}^{k} \{X_{t_i} = Y_{t_i}\}\Big)^c\Big) = \mathbb{P}\Big(\bigcup_{i=1}^{k} \{X_{t_i} = Y_{t_i}\}^c\Big) \leq \sum_{i=1}^{k} \mathbb{P}\big(\{X_{t_i} = Y_{t_i}\}^c\big) = 0.$$

□

Esempio 4.4. *Come primo esempio di processo si definisca, in uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$,*

$$N_t(\omega) := \mathbb{I}_{\{T \leq t\}}(\omega) = \mathbf{1}_{[T(\omega), \infty)}(t) \begin{cases} = 0 & \text{se } T(\omega) > t, \\ = 1 & \text{se } T(\omega) \leq t \end{cases}, \quad t \geq 0,$$

dove T è una variabile aleatoria a valori in $(0, \infty)$.
Si noti che, qualunque sia $\omega \in \Omega$ le traiettorie di questo processo, cioè le funzioni

$$t \mapsto \mathbb{I}_{\{T \leq t\}}(\omega)$$

sono funzioni crescenti (in senso lato) rispetto a t.

Se T ammette densità di probabilità, ad esempio se è una variabile aleatoria esponenziale, allora il processo definito da

$$M_t(\omega) := N_t(\omega) + f(t + T(\omega)),$$

dove

$$\begin{cases} f(s) = 1 & \text{per } s \in \mathbb{Q} \\ f(s) = 0 & \text{per } s \in \mathbb{R} \backslash \mathbb{Q}, \end{cases}$$

è stocasticamente equivalente ad N_t, e quindi ha le stesse distribuzioni finito-dimensionali di N_t, ma evidentemente non ha traiettorie crescenti (in senso lato).

Per controllare che $M_t(\omega)$ ed $N_t(\omega)$ sono stocasticamente equivalenti, si osservi che, $M_t(\omega) \neq N_t(\omega)$ se e solo se $f(t + T(\omega)) = 1$ ovvero se e solo se $t + T(\omega) \in \mathbb{Q}$ ed inoltre

$$\mathbb{P}(t + T(\omega) \in \mathbb{Q}) = \sum_{r \in \mathbb{Q}} \mathbb{P}(T(\omega) = r - t) = 0$$

per ogni t, in quanto T è una variabile aleatoria continua e si ha quindi che

$$\mathbb{P}(M_t(\omega) = N_t(\omega)) = 1 \text{ per ogni } t \geq 0.$$

4.3 Esistenza di una versione continua: criterio di Chensov-Kolmogorov.

Il seguente criterio sufficiente fornisce condizioni che assicurano l'esistenza di una versione hölderiana, non solo continua. Ricordiamo che una funzione $f(x)$ è **hölderiana** di esponente γ se per ogni x esistono un $\delta(x) > 0$ e un $L_\gamma(x)$ tali che, per ogni y per il quale $|y - x| \leq \delta(x)$, si abbia

$$|f(x) - f(y)| \leq L_\gamma(x)|x - y|^\gamma.$$

Nel caso in cui $\delta(x)$ e $L_\gamma(x)$ possano essere presi in modo indipendente da x, per $x \in I$, si dice che f è **uniformemente hölderiana** nell'insieme I.

Proposizione 4.3 (Criterio di Chensov-Kolmogorov)**.** *Sia X_t un processo per cui esistono α, β e C strettamente positivi, per cui*

$$\mathbb{E}[|X_t - X_s|^\beta] \leq C|t - s|^{1+\alpha}.$$

Allora esiste una versione $\widetilde{X}_t$ di X_t, a traiettorie continue. Inoltre le traiettorie sono uniformemente hölderiane di esponente γ, per ogni $\gamma < \frac{\alpha}{\beta}$, in ogni intervallo limitato.

Esempio 4.5 (Applicazione al processo di Wiener)**.**

Il processo W_t di Wiener standard, o moto Browniano, ammette sempre una versione continua, anzi hölderiana per ogni $\gamma < \frac{1}{2}$. Infatti, per $s \leq t$,

$$\mathbb{E}[|W_t - W_s|^k] = \mathbb{E}[|W_{t-s}|^k] = C(k)|t - s|^{k/2}.$$

Il criterio di Chensov-Kolmogorov si può quindi applicare per $k \geq 3$ (così $\alpha = (k/2) - 1 > 0$) e ci garantisce che esiste una versione di W_t le cui traiettorie sono hölderiane di esponente γ per $\gamma < \frac{(k/2)-1}{k} = \frac{1}{2} - \frac{1}{k}$, e quindi per ogni $\gamma < \frac{1}{2}$. È possibile dimostrare che, a parte un eventuale insieme di misura nulla, le traiettorie non sono hölderiane di esponente $\frac{1}{2}$, e tantomeno di esponente maggiore di $\frac{1}{2}$.

4.4 Le traiettorie del processo di Wiener non sono a variazione limitata

In questa sezione si dimostra una proprietà che non permette di definire nel modo usuale (Lebesgue-Stieltjes) l'integrale rispetto al processo di Wiener. Nonostante si abbia che le traiettorie della versione continua del processo di Wiener siano hölderiane, tuttavia esse non possono essere molto regolari, in quanto ad esempio sono anche a variazione non limitata con probabilità 1 su ogni intervallo $[s, t]$. Infatti, per ogni versione continua[12] di W_u, se esistesse un intervallo $[s, t]$ ed un insieme A con $\mathbb{P}(A) > 0$, su cui la variazione[13] di W_u, cioè se

$$\begin{aligned} V(\omega) &= V(s, t, W; \omega) \\ &:= \sup\Big\{ \sum_{k=0}^{n-1} |W_{t_{k+1}}(\omega) - W_{t_k}(\omega)|, \text{ al variare delle partizioni } \pi : t_0 = s \leq t_1 \leq \cdots \leq t_n = t \Big\} \end{aligned}$$

[12] Se il processo W_u ha le traiettorie continue, allora su ogni intervallo chiuso e limitato $[s, t]$, esse sono uniformemente continue, e quindi, per ogni ω

$$\sup_{u,v \in [s,t], |u-v| \leq \delta} |W_u(\omega) - W_v(\omega)| \to 0 \qquad \text{per } \delta \to 0.$$

[13] Ricordiamo che se f è una funzione definita su un intervallo $[a, b]$ la variazione di f su $[a, b]$ è appunto definita come

$$\sup\{\sum_{k=0}^{n} |f(x_{k+1}) - f(x_k)|\},$$

dove l'estremo superiore è preso su tutte le partizioni $\{x_0, x_1, \cdots, x_n\}$ di $[a, b]$ tali che $x_0 = a \leq x_1 \leq \cdots \leq x_n = b$.

fosse finita per ogni $\omega \in A$, allora si avrebbe che, se $|\pi| := \max_k |t_{k+1} - t_k| \to 0$,

$$\begin{aligned} S_\pi &:= \sum_{k=0}^{n-1} |W_{t_{k+1}} - W_{t_k}|^2 \le \sum_{k=0}^{n-1} \left(\sup_h |W_{t_{h+1}} - W_{t_h}| \right) |W_{t_{k+1}} - W_{t_k}| \\ &= \left(\sup_h |W_{t_{h+1}} - W_{t_h}| \right) \sum_{k=0}^{n-1} |W_{t_{k+1}} - W_{t_k}| \le \left(\sup_{u,v \in [s,t], |u-v| \le |\pi|} |W_u - W_v| \right) V(\omega) \to 0. \end{aligned}$$

Questo fatto è in contraddizione con la seguente proprietà del moto browniano (proprietà che è valida anche per versioni non continue del processo di Wiener):

Lemma 4.4. *Si definisca*

$$S_\pi := \sum_{k=0}^{n-1} |W_{t_{k+1}} - W_{t_k}|^2,$$

allora S_π converge in media quadratica a $t - s$, ovvero $S_\pi \xrightarrow{L^2} (t-s)$, al tendere a zero di $|\pi| := \max_k |t_{k+1} - t_k|$.

Dimostrazione. Si tratta di mostrare che $\mathbb{E}[(S_\pi - (t-s))^2] \to 0$ al tendere a zero di $|\pi| := \max_k |t_{k+1} - t_k|$, e infatti

$$\mathbb{E}\left[(S_\pi - (t-s))^2\right] = 2 \sum_{k=0}^{n-1} |t_{k+1} - t_k|^2,$$

come si vede facilmente[14]. Di conseguenza si ottiene che

$$\begin{aligned} \mathbb{E}\left[(S_\pi - (t-s))^2\right] &\le 2 \sum_{k=0}^{n-1} |t_{k+1} - t_k| (\max_h |t_{h+1} - t_h|) \\ &= 2(\max_h |t_{h+1} - t_h|) \sum_{k=0}^{n-1} |t_{k+1} - t_k| = 2(|\pi|(t-s) \to 0. \end{aligned}$$

[14] Basta osservare che

$$\begin{aligned} &\mathbb{E}[(S_\pi - (t-s))^2] = \mathbb{E}[S_\pi^2 + (t-s))^2 - 2S_\pi(t-s)] \\ &= \mathbb{E}\left[\Big(\sum_{k=0}^{n-1} |W_{t_{k+1}} - W_{t_k}|^2 \Big)^2 + (t-s)^2 - 2\Big(\sum_{k=0}^{n-1} |W_{t_{k+1}} - W_{t_k}|^2 \Big)(t-s) \right] \\ &= \mathbb{E}\left[\sum_{k=0}^{n-1} |W_{t_{k+1}} - W_{t_k}|^2 \sum_{h=0}^{n-1} |W_{t_{h+1}} - W_{t_h}|^2 + (t-s)^2 - 2\Big(\sum_{k=0}^{n-1} |W_{t_{k+1}} - W_{t_k}|^2 \Big)(t-s) \right] \\ &= \sum_{k=0}^{n-1} \mathbb{E}\left[|W_{t_{k+1}} - W_{t_k}|^4 \right] + \sum_{k=0}^{n-1} \sum_{\substack{h=0 \\ h \ne k}}^{n-1} \mathbb{E}\left[|W_{t_{k+1}} - W_{t_k}|^2 |W_{t_{h+1}} - W_{t_h}|^2 \right] + \\ &\quad + (t-s)^2 - 2(t-s) \sum_{k=0}^{n-1} \mathbb{E}\left[|W_{t_{k+1}} - W_{t_k}|^2 \right] \\ &= \sum_{k=0}^{n-1} 3|t_{k+1} - t_k|^2 + \sum_{k=0}^{n-1} \sum_{\substack{h=0 \\ h \ne k}}^{n-1} |t_{k+1} - t_k||t_{h+1} - t_h| + (t-s)^2 - 2(t-s) \sum_{k=0}^{n-1} |t_{k+1} - t_k| \end{aligned}$$

Di conseguenza si ottiene che

$$\begin{aligned} &\mathbb{E}\left[(S_\pi - (t-s))^2\right] \\ &= \sum_{k=0}^{n-1} 2|t_{k+1} - t_k|^2 + \sum_{k=0}^{n-1} \sum_{h=0}^{n-1} |t_{k+1} - t_k||t_{h+1} - t_h| + (t-s)^2 - 2(t-s)(t-s) \\ &= \sum_{k=0}^{n-1} 2|t_{k+1} - t_k|^2 + (t-s)^2 + (t-s)^2 - 2(t-s)(t-s) = 2 \sum_{k=0}^{n-1} |t_{k+1} - t_k|^2 \end{aligned}$$

La conseguenza di questo risultato è che le traiettorie di un processo di Wiener non sono a variazione limitata[15], o meglio[16],

$$\mathbb{P}\left(V(s,t,W;\omega) := \sup_{\pi} \sum_{k=0}^{n-1} |W_{t_{k+1}} - W_{t_k}| = \infty\right) = 1.$$

Questo fatto in particolare pone dei problemi nel tentare di definire $\int_0^t f(s,\omega)dW_s(\omega)$, in quanto non si può adottare la classica definizione di integrale di Lebesgue-Stieltjes, che avrebbe senso solo se W_s avesse traiettorie a variazione finita. La definizione che si dà è quindi diversa e si parla in questo caso di integrale stocastico, inoltre per ottenerla si ha bisogno del concetto di martingala **(vedere la Sezione 6, in cui viene introdotto l'integrale stocastico rispetto a Wiener, e la Sezione 3.2.2, che anticipa l'integrale stocastico a tempo continuo, ma in cui viene considerato l'integrale rispetto ad una martingala di quadrato integrabile, non necessariamente il processo di Wiener).

**A questo proposito facciamo notare che non si può definire l'integrale stocastico come si definisce l'usuale integrale di Riemann.

Esempio 4.6 (Come definire $\int_a^b 2W_s\, dW_s$?)**.** *Prima di tutto ricordiamo alcune formule che sono banalmente valide: sia $\{b_k\}$ una successione di numeri reali, allora*

$$\begin{aligned}
2 \sum_{k_{min}\le k\le k_{max}-1} b_k(b_{k+1}-b_k) &= -\sum_k [b_{k+1}-b_k]^2 + [b_{k_{max}}^2 - b_{k_{min}}^2],\\
2 \sum_{k_{min}\le k\le k_{max}-1} b_{k+1}(b_{k+1}-b_k) &= \sum_k [b_{k+1}-b_k]^2 + [b_{k_{max}}^2 - b_{k_{min}}^2],\\
2 \sum_{k_{min}\le k\le k_{max}-1} \frac{b_k+b_{k+1}}{2}(b_{k+1}-b_k) &= [b_{k_{max}}^2 - b_{k_{min}}^2],
\end{aligned}$$

la cui dimostrazione è molto semplice[17].

Posto $b_k = W_{t_k}$ con $t_{k_{min}} = a$ e $t_{k_{max}} = b$, le precedenti formule, insieme al Lemma 4.4, danno rispettivamente i seguenti limiti (in L^2):

$$\lim_{n\to\infty} 2 \sum_{k_{min}\le k\le k_{max}-1} W_{t_k}(W_{t_{k+1}} - W_{t_k}) = -\lim_{n\to\infty} \sum_k [W_{t_{k+1}} - W_{t_k}]^2 + W_b^2 - W_a^2 = -b - a + W_b^2 - W_a^2$$

[15] In realtà con la stessa tecnica del lemma si dimostra anche che il processo di Wiener non ha traiettorie di Hölder con esponente $\gamma > 1/2$: se così fosse allora, posto $L_\gamma(\omega)$ la costante di Hölder, si avrebbe necessariamente

$$S_\pi(\omega) \le \sum_{k=0}^{n-1} L_\gamma^2(\omega)|t_{k+1}-t_k|^{2\gamma} \le \sum_{k=0}^{n-1} L_\gamma^2(\omega)|t_{k+1}-t_k|\delta^{2\gamma-1} = L_\gamma^2(\omega)\delta^{2\gamma-1}(t-s) \to 0,\ q.c.$$

[16] **Si noti che, per $\pi : t_0 = s \le t_1 \le \cdots \le t_n = t$ posto $V_\pi(\omega) := \sum_{k=0}^{n-1} |W_{t_{k+1}}(\omega) - W_{t_k}(\omega)|$, per la continuità delle traiettorie $t \mapsto W_t(\omega)$, si ottiene che $V(\omega) = V(s,t,W;\omega) = \sup\{V_{\pi^m};\ \pi = \{t_0 = s,\ t_k = s + \frac{k}{2^m}(t-s),\ k = 1,\cdots,2^m\}\ m \in \mathbb{N}\}$, e che la successione V_{π^m} è una successione monotona non decrescente, quindi $V(\omega) = \lim_{m\to\infty} V_{\pi^m}(\omega)$, per ogni ω.

[17] **Per ogni successione $\{b_k\}$ si ha

$$\begin{aligned}
2 \sum_{k_{min}\le k\le k_{max}-1} b_k(b_{k+1}-b_k) &= \sum_k [2b_{k+1}b_k - 2b_k^2] = \sum_k [2b_{k+1}b_k - 2b_k^2 - b_{k+1}^2 + b_{k+1}^2]\\
&= \sum_k [(2b_{k+1}b_k - b_k^2 - b_{k+1}^2) + b_{k+1}^2 - b_k^2] = -\sum_k [b_{k+1}-b_k]^2 + \sum_k [b_{k+1}^2 - b_k^2]\\
&= -\sum_k [b_{k+1}-b_k]^2 + [b_{k_{max}}^2 - b_{k_{min}}^2].
\end{aligned}$$

Invece

$$\begin{aligned}
2 \sum_{k_{min}\le k\le k_{max}-1} b_{k+1}(b_{k+1}-b_k) &= \sum_k [2b_{k+1}^2 - 2b_{k+1}b_k] = \sum_k [2b_{k+1}^2 + b_k^2 - b_k^2 - 2b_{k+1}b_k]\\
** &= \sum_k [(b_{k+1}^2 - 2b_{k+1}b_k + b_k^2) + b_{k+1}^2 - b_k^2] = \sum_k [b_{k+1}-b_k]^2 + \sum_k [b_{k+1}^2 - b_k^2]\\
&= \sum_k [b_{k+1}-b_k]^2 + [b_{k_{max}}^2 - b_{k_{min}}^2].
\end{aligned}$$

$$\lim_{n\to\infty} 2 \sum_{k_{min}\le k\le k_{max}-1} W_{t_{k+1}}(W_{t_{k+1}} - W_{t_k}) = \lim_{n\to\infty} \sum_k [W_{t_{k+1}} - W_{t_k}]^2 + W_b^2 - W_a^2 = b - a + W_b^2 - W_a^2$$

e infine

$$\lim_{n\to\infty} 2 \sum_{k_{min}\le k\le k_{max}-1} \frac{W_{t_k} + W_{t_{k+1}}}{2} (W_{t_{k+1}} - W_{t_k}) = W_b^2 - W_a^2.$$

In conclusione, come del resto c'era da aspettarsi, anche se le somme di Riemann

$$2 \sum_{k_{min}\le k\le k_{max}-1} W_{t_k^*}(W_{t_{k+1}} - W_{t_k}) \qquad \textit{(con } t_k^* \in [t_k, t_{k+1}]\textit{)}$$

convergono per alcune scelte di t_k^*, ma i limiti dipendono dalla scelta del punto t_k^*.

E ancora

$$\begin{aligned} 2 \sum_{k_{min}\le k\le k_{max}-1} \frac{b_k + b_{k+1}}{2} (b_{k+1} - b_k) &= \frac{1}{2} \sum_k 2\, b_k\, [b_{k+1} - b_k] + \frac{1}{2} \sum_k 2 b_{k+1} [b_{k+1} - b_k] \\ &= \frac{1}{2} \left(- \sum_k [b_{k+1} - b_k]^2 + [b_{k_{max}}^2 - b_{k_{min}}^2] \right) + \frac{1}{2} \left(\sum_k [b_{k+1} - b_k]^2 + [b_{k_{max}}^2 - b_{k_{min}}^2] \right) \\ &= [b_{k_{max}}^2 - b_{k_{min}}^2] \end{aligned}$$

4.5 Costruzione del processo di Wiener e della misura di Wiener su $C[0,1]$. **

**In questa sezione viene data dapprima una costruzione diretta di un processo di Wiener con le traiettorie continue. La misura di Wiener è la sua distribuzione (ovvero la misura indotta) sullo spazio delle funzioni continue. Ci limitiamo al caso dell'intervallo di tempo $[0,1]$, ma indichiamo come costruire il processo su tutto $[0,+\infty)$. Successivamente, si accenna a come si può costruire la misura di Wiener come limite (in un senso opportuno) di una successione di misure di probabilità su $C([0,1];\mathbb{R})$.

4.5.1 Costruzione come serie di funzioni

In questa sezione esaminiamo una costruzione diretta del processo di Wiener, a partire da uno spazio in cui ci sia una successione di variabili aleatorie indipendenti $\{X_k\}_{k\geq 0}$, tutte con distribuzione gaussiana standard:

$$W(t,\omega) := \sum_{n=0}^{\infty} Z_n(\omega)\, s_n(t),$$

dove le funzioni (deterministiche) $s_n(\cdot)$ sono le funzioni di Schauder, ossia gli integrali delle funzioni di Haar.

Questa costruzione è nota in letteratura come la costruzione di Lévy-Ciesielski, e permette di ottenere direttamente l'esistenza di un processo di Wiener (continuo). Poiché, come è noto[18], nello spazio $\{(0,1),\mathcal{B}(0,1),\mathbb{P})$, con $\mathbb{P}$ la misura di Lebesgue ristretta a $(0,1)$ **è possibile costruire una succesione di variabili aleatorie indipendenti con distribuzioni marginali assegnate, **ciò significa che possiamo costruire il processo di Wiener direttamente su tale spazio.

Ci limiteremo a vedere la costruzione nel caso dell'intervallo temporale $[0,1]$. Per ottenere la costruzione del processo di Wiener su tutto $[0,\infty)$, bisognerebbe costruire un'infinità numerabile di copie e *incollarle* opportunamente. Ma qui ci occupiamo solo del caso di un singolo intervallo.

Per la dimostrazione della continuità ci basiamo su un articolo di Mark A. Pinsky [10], in cui l'autore presenta una dimostrazione molto semplice della continuità globale di Lévy, che permette di affermare che, per ogni T, esiste una variabile aleatoria M, che dipende da T, finita con probabilità 1, tale che

$$\big|W_t(\omega) - W_s(\omega)\big| \;\leq\; M(\omega)\,\sqrt{\,|t-s|\log(1/|t-s|)} \qquad \text{per } s,\, t\in[0,T] \tag{4.8}$$

Diamo solo la costruzione per $t\in[0,1]$, lasciando al lettore[19], il compito di estendere la costruzione per $t\in[0,\infty)$. Prima di tutto ricordiamo la definizione delle funzioni di Haar, che, per comodità, numeriamo con un doppio indice

$$\phi_{00}(t) \;=\; 1 \qquad \text{per } 0\leq t\leq 1 \tag{4.9}$$

$$\phi_{01}(t) \;=\; \begin{cases} 1 & \text{per } 0\leq t<\frac{1}{2} \\ -1 & \text{per } \frac{1}{2}\leq t\leq 1 \end{cases} \quad . \tag{4.10}$$

[18] Si veda la Sezione 4.9.1, **ed in particolare il Lemma 4.10. Inoltre si può osservare che invece di prendere i boreliani, si potrebbe prendere la *sigma*-algebra degli insiemi misurabili secondo Lebesgue, ed ottenere quindi il processo di Wiener in uno spazio completo.

[19] L'idea è la seguente: si costruisce una successione di processi $W^{(n)}(t)$ per $t\in(0,1)$, indipendenti, e si definisce

$$W(t) = W^{(1)}(t) \qquad t\in[0,1]$$
$$W(t) = W^{(1)}(1) + W^{(2)}(1) + \cdots + W^{(n-1)}(1) + W^{(n)}(t-n) \qquad t\in[n,n+1]$$

e, se $i \geq 1$ e $j = 1, 2, ..., 2^i$,

$$\phi_{ij}(t) = \begin{cases} 2^{\frac{i}{2}} & \text{se } \frac{(2j-2)}{2^{i+1}} \leq t < \frac{2j-1}{2^{i+1}} \\ -2^{\frac{i}{2}} & \text{se } \frac{(2j-1)}{2^{i+1}} \leq t < \frac{2j}{2^{i+1}} \\ 0 & \text{altrimenti} \end{cases} . \tag{4.11}$$

È importante ricordare che le $\phi_{ij}(t)$ formano un sistema ortonormale e completo in $L^2(0,1)$, che è uno spazio di Hilbert, con prodotto scalare $\langle f, g\rangle = \int_0^1 f(t)g(t)dt$.

Ricordiamo che per $f, g \in L^2(0,1)$ valgono **l'identità di Parseval**:

$$\langle f, f\rangle = \int_0^1 f(t)^2 dt = \sum_{i,j} \left(\int_0^1 f(t)\phi_{ij}(t)dt \right)^2 = \sum_{i,j} \langle f, \phi_{i,j}\rangle^2,$$

e

$$\langle f, g\rangle = \int_0^1 f(t)g(t)dt = \sum_{i,j} \int_0^1 f(t)\phi_{ij}(t)dt \int_0^1 g(t)\phi_{ij}(t)dt = \sum_{i,j} \langle f, \phi_{i,j}\rangle\langle g, \phi_{i,j}\rangle.$$

Per costruire il moto browniano consideriamo una sequenza, a doppio indice, Y_{ij} di variabili aleatorie indipendenti con distribuzione normale[20].

Il moto browniano è definito da

$$W_t(\omega) = Y_{00}(\omega) \int_0^t \phi_{00}(s)ds + \sum_{i=0}^{\infty} \sum_{j=1}^{2^i} Y_{ij}(\omega) \int_0^t \phi_{ij}(s)ds; \tag{4.12}$$

ed è immediato verificare che, essendo le ϕ_{ij} una base ortonormale e le variabili Y_{ij} indipendenti, con media 0 e varianza 1, per ogni $t \in [0,1]$, la serie (4.12) converge in $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Di conseguenza il processo $(W_t)_{t\in[0,1]}$ è dotato delle proprietà di un moto browniano relativamente alle distribuzioni finito dimensionali: essendo la somma della serie (4.12), W_tè una variabile aleatoria con distribuzione normale[21] per la quale esistono:

$$\mathbb{E}[W_t] = 0 \quad ;$$

$$\mathbb{E}[W_t^2] = \left(\int_0^t \phi_{00}(s)ds \right)^2 + \sum_{i=0}^{\infty} \sum_{j=1}^{2^i} \left(\int_0^t \phi_{ij}(s)ds \right)^2 \quad ,$$

dove riconosciamo i coefficienti di Fourier per la funzione indicatrice $\mathbf{1}_{[0,t]}$ nella forma:

$$\langle \mathbf{1}_{[0,t]}, \phi_{ij}\rangle = \int_0^1 \mathbf{1}_{[0,t]}(s)\, \phi_{ij}(s)\, ds = ** \int_0^t \phi_{ij}(s)\, ds,$$

possiamo allora applicare l'identità di Parseval per ottenere:

$$\mathbb{E}[W_t^2] = \langle \mathbf{1}_{[0,t]}, \phi_{00}\rangle^2 + \sum_{i=0}^{\infty} \sum_{j=1}^{2^i} \langle \mathbf{1}_{[0,t]}, \phi_{ij}\rangle^2 = \|\mathbf{1}_{[0,t]}\|^2 = t.$$

Allo stesso modo possiamo dimostrare che gli incrementi $W_t - W_s$ hanno distribuzione normale[22] con

$$\mathbb{E}[W_t - W_s] = 0 \quad ;$$

$$\mathbb{E}[(W_t - W_s)^2] = \langle \mathbf{1}_{[s,t]}, \phi_{00}\rangle^2 + \sum_{i=0}^{\infty} \sum_{j=1}^{2^i} \langle \mathbf{1}_{[s,t]}, \phi_{ij}\rangle^2 = \|\mathbf{1}_{[s,t]}\|^2 = t - s,$$

[20] Queste variabili aleatorie possono essere definite su uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, che a sua volta può essere assunto completo nel senso che i sottoinsiemi con probabilità nulla che vivono in $\mathcal{F}$ hanno misura di probabilità uguale a zero.

[21] Se una successione di variabili aleatorie congiuntamente gaussiane converge in distribuzione (e a maggior ragione se converge in probabilità) ad un vettore di variabili aleatorie, allora le variabili aleatorie limite sono ancora gaussiane

[22] **Il fatto che la distribuzione sia gaussiana dipende dal fatto che sono variabili aleatorie ottenute come limite di variabili aleaorie gaussiane.

e quindi possiamo riscrivere la proprietà di indipendenza degli incrementi (nel caso di due incrementi) attraverso la covarianza sviluppata tramite la formula di Parseval: *se $u < s < t$, allora*

$$\mathbb{E}[(W_t - W_s)W_u] = \sum_{ij} \langle \mathbf{1}_{[s,t]}, \phi_{ij}\rangle \langle \mathbf{1}_{[0,u]}, \phi_{ij}\rangle = \langle \mathbf{1}_{[0,u]}, \mathbf{1}_{[s,t]}\rangle = 0$$

dove gli intervalli $[0,u]$ e $[s,t]$ sono disgiunti. Ciò è sufficiente per dimostrare che si tratta di un processo gaussiano a media nulla e con funzione di correlazione $K(s,t) = s \wedge t$.

Rimane da verificare la continuità delle traiettorie.

Continuità globale

Affrontiamo lo studio della continuità tramite il modulo di continuità

$$w_x(\delta) := \sup_{s,t\in[0,1]:|t-s|\le\delta} |x(t) - x(s)|,$$

e dimostreremo (Teorema 4.7) che esiste una variabile aleatoria $^{**}M_1$, finita quasi certamente, per la quale

$$w_{W(\omega)}(\delta) \le {}^{**}M_1(\omega)\sqrt{\delta \log \tfrac{1}{\delta}}.$$

Prima di procedere, però, è opportuno presentare due proprietà fondamentali di $Y_{ij}(\omega)$.

Lemma 4.5. *Esiste una variabile aleatoria M, finita con probabilità 1, e tale che:*

$$\mathbb{P}\left(\omega : \sup_{i\ge 0, j=1,\dots,2^i} \frac{|Y_{ij}(\omega)|}{\sqrt{i+1}} \le M(\omega)\right) = 1$$

Dimostrazione. Sia $N(0,0) = 1$ e $N(i,j) = 2^i + j$ per $i \ge 0$ e $j = 1, 2, 3, ..., 2^i$, così che

$$^{**}2^i < N(i,j) \le 2^{i+1}.$$

Definiamo la successione (a indice singolo) $\{Z_n\}$ di variabili aleatorie definite in modo che $Y_{ij} = Z_{N(i,j)}$ $\big($ovvero $Z_n = Y_{ij}$ se e solo se $n = 2^i + j\big)$. Le variabili aleatorie Z_n hanno distribuzione gaussiana standard e quindi otteniamo (si veda la relazione (1.17)) la maggiorazione

$$\mathbb{P}(|Z_n| > x) \le e^{-\frac{x^2}{2}}.$$

Posto $x = 2\sqrt{\log n}$ **(in modo che $e^{-\frac{x^2}{2}} = e^{-\frac{4\log n}{2}} = e^{-2\log n} = 1/n^2$), otteniamo

$$\mathbb{P}(|Z_n| > 2\sqrt{\log n}) \le n^{-2},$$

quindi, per il Lemma di Borel-Cantelli, esiste $n_0(\omega)$ tale che

$$n \ge n_0(\omega) \qquad \Rightarrow \qquad |Z_n| \le 2\sqrt{\log n}.$$

Osservando che, per $N(i,j) \ge n_0(\omega)$

$$|Y_{ij}| = |Z_{N(i,j)}| \le 2\sqrt{\log N(i,j)} \le 2\sqrt{(i+1)\log 2} \tag{4.13}$$

otteniamo la tesi[23], ponendo

$$M(\omega) := \max\Big(\max_{i,j:2^i+j\le n_0(\omega)} \frac{|Y_{ij}|}{\sqrt{i+1}},\ 2\sqrt{\log 2}\Big).** \tag{4.14}$$

□

L'altra proprietà viene fornita dal seguente lemma (che permette anche di affermare che la successione dei processi definiti come somme finite, converge uniformemente nello spazio delle funzioni continue su [0,1]):

Lemma 4.6. *Esiste una variabile aleatoria M, finita con probabilità 1, e tale che, per $i \ge 1$,*

$$\left|\sum_{j=1}^{2^i} Y_{ij}(\omega)\int_0^t \phi_{ij}(s)ds\right| \le M(\omega)\cdot\sqrt{i+1}\cdot 2^{-\frac{i}{2}}.$$

Dimostrazione. Fissato i, le funzioni $\psi_{ij}(t) = \int_0^t \phi_{ij}(s)ds$, ovvero le funzioni di Schauder, sono funzioni poligonali non negative e strettamente positive su intervalli disgiunti di ampiezza 2^{-i}, con[24] $0 \le \psi_{ij}(t) \le 2^{-i-1}\times 2^{\frac{i}{2}} = 2^{-\frac{i}{2}-1} \le 2^{-\frac{i}{2}}$. Quindi

$$\sum_{j=1}^{2^i}\psi_{ij}(t) \le 2^{-\frac{i}{2}}$$

Quindi, **prendendo M come in (4.14):

$$\sup_{0\le t\le 1}\left|\sum_{j=1}^{2^i} Y_{ij}(\omega)\int_0^t \phi_{ij}(s)ds\right| \le \max_{1\le j\le 2^i}|Y_{ij}(\omega)|\sum_{j=1}^{2^i}\psi_{ij}(t) \le \sqrt{i+1}\,M(\omega)\,2^{-\frac{i}{2}}$$

come richiesto.

□

Osservazione 4.1. ***Si osservi che il precedente Lemma 4.6 permette già di affermare che la serie che definisce il processo di Wiener converge in norma uniforme: infatti la serie $\sum_i M(\omega)\sqrt{i+1}\ 2^{-\frac{i}{2}}$ è una serie convergente. Quindi, dalla definizione (4.12) di $W(t)$ possiamo già affermare che il processo $\{W(t)\}_{t\in[0,1]}$ è a traiettorie continue in quanto è il limite uniforme delle funzioni continue*

$$W_t^m(\omega) := Y_{00}(\omega)\int_0^t \phi_{00}(s)ds + \sum_{i=0}^{m}\sum_{j=1}^{2^i} Y_{ij}(\omega)\int_0^t \phi_{ij}(s)ds.$$

Resta tuttavia l'interesse di verificare che il modulo di continuità sia proprio quello già annunciato in precedenza

Possiamo ora concludere il nostro studio

[23] **Chiaramente

$$\sup_{i\ge 0, j=1,\dots,2^i}\frac{|Y_{ij}(\omega)|}{\sqrt{i+1}} = \max\Big(\max_{i,j:2^i+j\le n_0(\omega)}\frac{|Y_{ij}|}{\sqrt{i+1}},\quad \max_{i,j:2^i+j>n_0(\omega)}\frac{|Y_{ij}|}{\sqrt{i+1}}\Big)$$

e inoltre, per la (4.13)

$$\max_{i,j:2^i+j>n_0(\omega)}\frac{|Y_{ij}|}{\sqrt{i+1}} \le 2\sqrt{\log 2}.$$

[24] Infatti la funzione ψ_{ij} vale 0 per $t \le j/2^i$, cresce con derivata uguale a $2^{\frac{i}{2}}$ per un intervallo di ampiezza $2^{-(i+1)}$, fino a raggiungere l'altezza massima $2^{\frac{i}{2}}\cdot 2^{-(i+1)}$, e poi decresce con derivata uguale a $-2^{\frac{i}{2}}$ per un intervallo di ampiezza $2^{-(i+1)}$, fino a raggiungere nuovamente lo 0.

Teorema 4.7. *Esiste un $M_1 = M_1(\omega) < \infty$ tale che* ***per ogni $\delta \in (0,1)$, se s, $t \in [0,1]$ con $|t-s| \le \delta$, allora:*

$$|W_t(\omega) - W_s(\omega)| \le M_1(\omega)\sqrt{\delta \log \tfrac{1}{\delta}}.$$

Dimostrazione. Separiamo gli incrementi della (4.12) in due parti come segue:

$$W_t(\omega) - W_s(\omega) = Y_{00}(t-s) + \left(\sum_{i=0}^{L} + \sum_{i=L+1}^{\infty}\right)\sum_{j=1}^{2^i} Y_{ij}\int_s^t \phi_{ij}(u)du$$

dove $L = L(\delta)$, verrà scelto in modo opportuno. Da questa espressione, e dalle proprietà delle funzioni di Haar e di Schauder[25] otteniamo

$$\left|\sum_{i=0}^{L}\sum_{j=1}^{2^i} Y_{ij}\int_s^t \phi_{ij}(u)du\right| \le \sum_{i=0}^{L}\max_{j=1,\cdots,2^i}|Y_{ij}|\int_s^t\Big|\sum_{j=1}^{2^i}\phi_{ij}(u)\Big|du \le |s-t|M(\omega)\sum_{i=0}^{L}\sqrt{i+1}\cdot 2^{\frac{i}{2}}$$

e inoltre, posto $\psi_{ij}(t) = \int_0^t \phi_{ij}(u)du$,

$$\begin{aligned}\left|\sum_{i=L+1}^{\infty}\sum_{j=1}^{2^i} Y_{ij}\int_s^t \phi_{ij}(u)du\right| &\le \sum_{i=L+1}^{\infty}\max_{j=1,\cdots,2^i}|Y_{ij}|\left|\sum_{j=1}^{2^i}\psi_{ij}(t) - \sum_{j=1}^{2^i}\psi_{ij}(s)\right| \\ &\le M(\omega)\sum_{i=L+1}^{\infty}\sqrt{i+1}\cdot 2^{-\frac{i}{2}}.\end{aligned}$$

Le due sommatorie sono maggiorate da due stime elementari[26], rispettivamente:

$$\sum_{i=0}^{L}\sqrt{i+1}\cdot 2^{\frac{i}{2}} \le K_1\cdot\sqrt{L}\cdot 2^{\frac{L}{2}} \qquad \text{e} \qquad \sum_{i=L+1}^{\infty}\sqrt{i+1}\cdot 2^{-\frac{i}{2}} \le K_2\cdot\sqrt{L}\cdot 2^{-\frac{L}{2}}$$

tali che se $|t-s| \le \delta$, allora, posto $K = \max(K_1, K_2)$

$$|W_t(\omega) - W_s(\omega)| \le |Y_{00}(\omega)|\delta + K\,M(\omega)\,[\sqrt{L}\,(\delta\, 2^{\frac{L}{2}} + 2^{-\frac{L}{2}})].$$

Gli ultimi due termini si semplificano ponendo $\delta 2^L \sim 1$, o, più precisamente, scegliendo $L = L(\delta) = \lfloor \log_2(\frac{1}{\delta})\rfloor$, nel seguente modo[27]

$$|W_t(\omega) - W_s(\omega)| \le Y_{00}(\omega)\delta + K'\,M(\omega)\sqrt{\delta\log_2(\tfrac{1}{\delta})}.$$

Ciò termina la dimostrazione, dato che $\delta \le \sqrt{\delta}$ per $\delta \in [0,1]$. □

---

[25] Sia i fissato, allora, per ogni $u \in [0,1]$, il valore di $\phi_{ij}(u)$ è diverso da zero solo per un $j \in \{1,\cdots,2^i\}$ e, in tale caso, vale $\pm 2^{\frac{i}{2}}$. Di conseguenza $\left|\sum_{j=1}^{2^i}\phi_{ij}(u)\right| = 2^{\frac{i}{2}}$, qualunque sia $u \in [0,1]$.

Inoltre come già visto nella dimostrazione del Lemma 4.6, si ha, posto $\psi_{ij}(t) = \int_0^t \phi_{ij}(u)du$, la maggiorazione $0 \le \psi_{ij}(t) \le 2^{-\frac{i}{2}-1}$, e quindi

$$\Big|\int_s^t \phi_{ij}(u)du\Big| = \big|\psi_{ij}(t) - \psi_{ij}(s)\big| \le \psi_{ij}(t) + \psi_{ij}(s) \le 2\,2^{-\frac{i}{2}-1} = 2^{-\frac{i}{2}}$$

[26] **Banalmente

$$\sum_{i=0}^{L}\sqrt{i+1}\cdot 2^{\frac{i}{2}} \le \sqrt{L}\cdot 2^{\frac{L}{2}}\cdot\frac{\sqrt{L+1}}{\sqrt{L}}\sum_{i=0}^{L}2^{-\frac{L-i}{2}} \le 2\sum_{j=0}^{\infty}2^{-\frac{j}{2}}\cdot\sqrt{L}\cdot 2^{\frac{L}{2}} = 2\,\frac{1}{1-\frac{1}{4}}\cdot\sqrt{L}\cdot 2^{\frac{L}{2}} = \frac{8}{3}\cdot\sqrt{L}\cdot 2^{\frac{L}{2}}$$

$$\begin{aligned}\sum_{i=L+1}^{\infty}\sqrt{i+1}\cdot 2^{-\frac{i}{2}} &\le \sqrt{L}\cdot 2^{-\frac{L}{2}}\cdot\sum_{i=L+1}^{\infty}\frac{\sqrt{L+i-L+1}}{\sqrt{L}}\cdot 2^{-\frac{i-L}{2}} = \sqrt{L}\cdot 2^{-\frac{L}{2}}\cdot\sum_{j=1}^{\infty}\sqrt{1+\frac{j+1}{L}}\cdot 2^{-\frac{j}{2}} \\ &\le \sqrt{L}\cdot 2^{-\frac{L}{2}}\cdot\sum_{j=1}^{\infty}\sqrt{2+j}\cdot 2^{-\frac{j}{2}} = K_2\cdot\sqrt{L}\cdot 2^{-\frac{L}{2}}\end{aligned}$$

[27] Si osservi che $\sqrt{L(\delta)} \le \sqrt{\log_2(\frac{1}{\delta})}$ e che, essendo $2^{L(\delta)} \le \frac{1}{\delta} \le 2^{L(\delta)+1}$, dalla seconda disuguaglianza si ha che $2^{-\frac{L(\delta)}{2}} \le \sqrt{2}\sqrt{\delta}$.

4.5.2 Costruzione come legge limite, sullo spazio delle funzioni continue **

In questa (sotto)sezione seguiamo invece l'idea di costruire la misura di probabilità direttamente sullo spazio delle funzioni reali e continue sull'intervallo chiuso $[0,1]$, ossia $C([0,1],\mathbb{R})$, che per semplicità di notazione denoteremo qui semplicemente con $C[0,1]$. La definizione viene data attraverso un procedimento di approssimazione di funzioni lineari a tratti. Maggiori dettagli si possono trovare nel libro di Billinsgley [4] (qui ci basiamo anche sulle note del Prof. Bertini [2])

Prima di iniziare vogliamo chiarire un punto essenziale: come abbiamo già accennato un processo aleatorio si può pensare come elemento aleatorio a valori in uno spazio di funzioni. Se vogliamo che sia un elemento aleatorio in uno spazio metrico, come è lo spazio delle funzioni continue $C[0,1]$, dobbiamo per prima cosa definire la sigma-algebra su tale spazio. In questo caso la più naturale è la sigma-algebra dei boreliani, ossia

Definizione 4.6. (Boreliani di C[0, 1]) *La sigma-algebra dei boreliani di $C[0,1]$ è la sigma algebra generata dagli aperti di $C[0,1]$, ossia la più piccola sigma-algebra che contiene gli aperti di $C[0,1]$*

$$\mathcal{B}\big(C[0,1]\big) := \sigma\{A,\ A \textit{ aperto in } C[0,1]\}.$$

D'altra parte esiste un'altra sigma-algebra naturale, vista la definizione di processo come famiglia di variabili aleatorie misurabili, ossia la sigma-algebra dei cilindri di $C[0,1]$.

Definizione 4.7. (Cilindri di C[0, 1]) *La sigma-algebra dei cilindri di $C[0,1]$ è la sigma algebra generata dai cilindri di $C[0,1]$ con base gli intervalli, ossia la più piccola sigma-algebra che contiene i cilindri del tipo*[28]

$$\mathsf{C}(\underline{t},\underline{I}) = \{x(\cdot) \in C[0,1] :\ x(t_1) \in I_1, x(t_2) \in I_2, \dots, x(t_n) \in I_n, \},$$

dove $\underline{t} = (t_1, t_2, \dots, t_n)$, con $0 \le t_1 \le t_2 \le \dots \le t_n \le 1$, e dove $\underline{I} = (I_1, I_2, \dots, I_n)$, con $I_1, I_2, \dots, I_n$ intervalli aperti di $\mathbb{R}$. In breve

$$\mathcal{C}\big(C[0,1]\big) := \sigma\{\mathsf{C}(\underline{t},\underline{I}) \subset C[0,1],\ \textit{al variare di } n,\ \textit{di } \underline{t} \textit{ e di } \underline{I}\}.$$

Osservazione 4.2. *Si noti che sia se prendiamo come spazio $(\Omega,\mathcal{F})$ lo spazio $\big(C[0,1],\mathcal{B}(C[0,1])\big)$, sia se prendiamo $\big(C[0,1],\mathcal{C}(C[0,1])\big)$, per ogni $t \in [0,1]$, la funzione $x(\cdot) \in C[0,1] \mapsto x(t) \in \mathbb{R}$ è misurabile: la contorimmagine di un intervallo aperto I è $\mathsf{C}(t,I)$ che è un aperto, ma è anche un cilindro. In entrambi i casi, trattandosi di una funzione misurabile, possiamo quindi parlare della variabile aleatoria* coordinata $x(t)$.

Sembrerebbe quindi sorgere un problema: quale sigma-algebra scegliere? Si tratta in realtà di un falso problema in quanto le precedenti sigma-algebre coincidono.

Lemma 4.8. *La sigma-algebra dei boreliani di $C[0,1]$ e la sigma-algebra dei cilindri di $C[0,1]$ coincidono, ossia*

$$\mathcal{B}\big(C[0,1]\big) = \mathcal{C}\big(C[0,1]\big)$$

Dimostrazione.
(i) $\mathcal{C}\big(C[0,1]\big) \subset \mathcal{B}\big(C[0,1]\big)$.

Basta osservare che, avendo scelto di prendere $I_1, I_2, \dots I_n$ intervalli aperti si ha che $\mathsf{C}(\underline{t},\underline{I})$ è un aperto di $C[0,1]$.

(ii) $\mathcal{B}\big(C[0,1]\big) \subset \mathcal{C}\big(C[0,1]\big)$.

Sia $x(\cdot)$ una funzione in $C[0,1]$, ed $r > 0$. Basta dimostrare che la sfera chiusa di raggio r

$$\overline{B}_r(x(\cdot)) = \{y \in C[0,1] :\ \sup_{t\in[0,1]} |y(t) - x(t)| \le r\}$$

[28] Si noti che la definizione di cilindro vale anche per insiemi che non siano intervalli aperti, ma si può generalizzare anche ad insiemi, boreliani di $\mathbb{R}$ di tipo diverso, o più in generale a boreliani H di $\mathbb{R}^n$:

$$\mathsf{C}(\underline{t},H) = \{x(\cdot) \in C[0,1] :\ (x(t_1), x(t_2), \dots, x(t_n)) \in H\}.$$

Da notare infine che tali cilindri sono sottoinsiemi dello spazio delle funzioni continue, mentre il nome cilindro viene usato anche per insiemi dello spazio di tutte le funzioni, non necessariamente continue (si veda ad esempio la dimostrazione del Teorema di esistenza di Kolmogorov, nella successiva Appendice 4.9).

appartiene alla sigma-algebra generata dai cilindri. Infatti allora lo stesso vale per la sfera aperta

$$B_r(x(\cdot)) = \{y \in C[0,1] : \sup_{t\in[0,1]} |y(t) - x(t)| < r\} = \bigcup_{s<r, s\in\mathbb{Q}} \overline{B}_s(x(\cdot))$$

e di conseguenza per ogni aperto A di $C[0,1]$, grazie alla separabilità dello spazio metrico $C[0,1]$:

esiste un insieme numerabile e denso $D = \{\xi_n(\cdot),\, n \in \mathbb{N}\} \subset C[0,1]$, ossia tale che ogni funzione $x(\cdot) \in C[0,1]$ può essere approssimata uniformemente con una successione di funzioni in D, quindi, per A aperto di $C[0,1]$ esiste una successione r_n tale che

$$A = \bigcup_{n:\, \xi_n \in A} B_{r_n}(\xi_n(\cdot)).$$

Rimane quindi solo da dimostrare che le sfere chiuse sono in $\mathcal{C}\big(C[0,1]\big)$, e infatti

$$\begin{aligned}\overline{B}_r(x(\cdot)) &= \{y \in C[0,1] : \sup_{t\in[0,1]} |y(t) - x(t)| \le r\} = \{y \in C[0,1] : \sup_{t\in[0,1]\cap\mathbb{Q}} |y(t) - x(t)| \le r\} \\ &= \bigcap_{t\in[0,1]\cap\mathbb{Q}} \mathsf{C}\big(t, [x(t) - r, x(t) + r]\big) = \bigcap_{t\in[0,1]\cap\mathbb{Q}} \bigcap_{j\in\mathbb{N}} \mathsf{C}\big(t, (x(t) - r - 1/j, x(t) + r + 1/j)\big).\end{aligned}$$

□

Grazie al lemma precedente possiamo definire una misura di probabilità su $\big(C[0,1], \mathcal{B}(C[0,1])\big)$ dando solo le probabilità sui cilindri[29], o equivalentemente le distribuzioni marginali del processo $x(\cdot)$, ossia le distribuzioni delle variabili $x(t)$, e più in generale, le distribuzioni congiunte delle variabili $(x(t_1), x(t_2), \ldots, x(t_n))$. Tuttavia, in genere, non è ovvio che tali marginali diano luogo ad una misura di probabiltà su $C[0,1]$. In particolare la misura di Wiener su $C[0,1]$ viene definita come la misura di probabilità (previo dimostrarne l'esistenza) che ha le seguenti caratteristiche.

Definizione 4.8. (Misura di Wiener su C[0, 1]) *La misura di Wiener su $\big(C[0,1], \mathcal{B}(C[0,1])\big)$ è la misura di probabilità che gode delle seguenti proprietà*

(i) per ogni $t \in [0,1]$ la variabile aleatoria $x(t)$ ha distribuzione gaussiana $N(0,t)$, ossia, per $t \in (0,1]$

$$P(x(t) \in I) = \int_I \frac{1}{\sqrt{2\,\pi\, t}} \exp\{-\tfrac{x^2}{2\,t}\}\, dx,$$

e con la convenzione che la distribuzione $N(0,0)$ coincide con la distribuzione degenere concentrata in 0, ossia

$$P(x(0) \in I) = \delta_{\{0\}}(I) \qquad \Big(\delta_{\{0\}}(I) = 1 \text{ se } 0 \in I, \quad \delta_{\{0\}}(I) = 0 \text{ altrimenti}\Big)$$

*(ii) il processo $\{x(t),\, t \in [0,1]\}$ è a incrementi indipendenti **ed omogenei[30], ossia per $0 \le t_1 \le t_2 \le \ldots, \le t_n \le 1$ le variabili aleatorie $x(t_2) - x(t_1)$, $x(t_3) - x(t_2)$, $\ldots$, $x(t_n) - x(t_{n-1})$ sono indipendenti **e con distribuzione $N(0, t_{k+1} - t_k)$. In altre parole, posto*

$$g_t(x) := \frac{1}{\sqrt{2\,\pi\, t}} \exp\{-\tfrac{x^2}{2\,t}\},$$

$$P\big(\mathsf{C}(\underline{t}, \underline{I}\big) = \int_{I_1} g_{t_1}(x_1)\, dx_1 \int_{I_2} g_{t_2 - t_1}(x_2 - x_1)\, dx_2 \cdots \int_{I_n} g_{t_n - t_{n-1}}(x_n - x_{n-1})\, dx_n$$

****DA FINIRE ****

[29] Si noti che ciò è possibile perché la classe dei cilindri è una classe che genera la sigma-algebra e che è chiusa rispetto all'intersezione finita, e quindi, una volta specificate le probabilità sui cilindri, per il Lemma di Dynkin (Lemma 1.1) può esistere solo una misura di probabilità su tutto $\mathcal{B}\big(C(01)\big)$ con tali caratteristiche. Rimane però aperto il problema dell'esistenza.

[30] **Per la definizione generale di processo ad incrementi indipendenti, si veda la Sezione 4.6

4.6 Processi ad incrementi indipendenti ed omogenei

Il procedimento usato per ottenere le distribuzioni finito dimensionali del processo di Wiener, si può estendere ad una classe più generale:

Definizione 4.9 (Processi ad incrementi indipendenti ed omogenei)**.** *Un processo* $(X_t)_{t\geq 0}$ *si dice* **ad incrementi indipendenti ed omogenei** *se*

(0) $X_0 = 0$*;*

(1) *per* $0 = t_0 < t_1 < \cdots t_n$ *gli incrementi* $\Delta X_{t_i} = X_{t_i} - X_{t_{i-1}}$ *sono variabili aleatorie indipendenti;*

(2) *gli incrementi* $\Delta X_{t_i} = X_{t_i} - X_{t_{i-1}}$ *sono variabili aleatorie la cui distribuzione dipende solo dall'ampiezza dell'intervallo* $(t_i - t_{i-1})$*;*

Per fissare le idee e capire meglio la condizione **(2)**, si consideri la famiglia $(F_u)_{u\geq 0}$ di funzioni di distribuzione dipendente da un parametro, per cui $X_{t+u} - X_t \sim F_u$, qualunque siano t ed u in $[0, \infty)$.

La famiglia $(F_u)_{u\geq 0}$ non può essere presa a piacere, ma deve soddisfare la seguente condizione necessaria[31]:

$$F_u * F_v = F_{u+v}, \qquad \text{per ogni } u, v \geq 0, \tag{4.15}$$

dove $*$ corrisponde alla convoluzione.

Infatti ciò corrisponde alla condizione che

$$X_u = X_u - X_0 \sim F_u, \quad X_{u+v} - X_u \sim F_v, \quad X_{u+v} \sim F_{u+v},$$

e d'altra parte

$$X_{u+v} = (X_{u+v} - X_u) + (X_u - X_0) \sim F_v * F_u.$$

In realtà questa condizione risulta anche sufficiente, come si può verificare facilmente. Infatti le funzioni $F_{t_1,\cdots,t_k}(x_1, \cdots, x_k)$ di distribuzione finito dimensionale risultano definite[32], per $0 < t_1 < \cdots < t_k$, come, la funzione di distribuzione ottenuta da quella degli incrementi, cioè

$$\mathbb{P}\left(X_{t_1} - X_0 \leq z_1,\, X_{t_2} - X_{t_1} \leq z_2, \cdots X_{t_k} - X_{t_{k-1}} \leq z_k\right) = F_{t_1}(z_1) F_{t_2-t_1}(z_2) \cdots F_{t_k-t_{k-1}}(z_k),$$

attraverso la trasformazione[33] $x_1 = z_1, x_2 = z_1 + z_2, \cdots, x_k = z_1 + \cdots + z_k$. La condizione (4.15) implica immediatamente che la condizione di consistenza di Kolmogorov **(C2′)** sia soddisfatta.

Per rendere più concreta la verifica, si consideri, ad esempio, il caso con densità, ovvero il caso in cui

$$F_u(x) = \int_{-\infty}^{x} q_u(y) dy.$$

Procedendo come per il processo di Wiener, si ottiene che, per $0 < t_1 < \cdots < t_k$

$$F_{t_1,\cdots,t_k}(x_1, \cdots, x_k) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_k} q_{t_1}(y_1) \cdots q_{t_k-t_{k-1}}(y_k - y_{k-1}) dy_1 \cdots dy_k.$$

[31] Inoltre è necessario che $F_0(x) = 0$ per $x < 0$ ed $F_0(x) = 1$ per $x \geq 0$, ovvero che l'*incremento* $X_t - X_t$ sia concentrato nello 0.

[32] Nel caso $0 = t_0 < t_1 < \cdots < t_k$ si ottiene immediatamente che

$$F_{t_0,t_1,\cdots,t_k}(x_0, x_1, \cdots, x_k) = 0, \qquad \text{per } x_0 < 0,$$

$$F_{t_0,t_1,\cdots,t_k}(x_0, x_1, \cdots, x_k) = F_{t_1,\cdots,t_k}(x_1, \cdots, x_k), \qquad \text{per } x_0 \geq 0.$$

[33] Si tratta solo di notare che se $Z_i := X_{t_i} - X_{t_{i-1}}$ allora $(X_{t_1}, X_{t_2}, \cdots, X_{t_k}) = (Z_1, Z_1 + Z_2, \cdots, Z_1 + \cdots + Z_k)$

Per controllare la condizione di consistenza **(C2′)**, si prendano $k \geq 1$ e $t_1 < t_2 < \cdots < t_k < t_{k+1}$, allora la 4.3) è verificata:

$$\begin{aligned}
&\lim_{x\to\infty} F_{t_1,\cdots,t_k,t_{k+1}}(x_1,\cdots,x_k,x)\\
&= \lim_{x\to\infty} \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_k} \int_{-\infty}^{x} q_{t_1}(y_1)\cdots q_{t_k-t_{k-1}}(y_k-y_{k-1}) q_{t_{k+1}-t_k}(y-y_k) dy_1\cdots dy_k\, dy\\
&= \lim_{x\to\infty} \int_{-\infty}^{x_1} q_{t_1}(y_1)\, dy_1 \cdots \int_{-\infty}^{x_k} q_{t_k-t_{k-1}}(y_k-y_{k-1})\, dy_k \int_{-\infty}^{x} \cdots q_{t_{k+1}-t_k}(y-y_k)\, dy\\
&= \lim_{x\to\infty} \int_{-\infty}^{x_1} q_{t_1}(y_1)\, dy_1 \cdots \int_{-\infty}^{x_k} q_{t_k-t_{k-1}}(y_k-y_{k-1})\, dy_k \int_{-\infty}^{x-y_k} \cdots q_{t_{k+1}-t_k}(y')\, dy'\\
&= \int_{-\infty}^{x_1} q_{t_1}(y_1)\, dy_1 \cdots \int_{-\infty}^{x_k} q_{t_k-t_{k-1}}(y_k-y_{k-1})\, dy_k = F_{t_1,\cdots,t_k}(x_1,\cdots,x_k),
\end{aligned}$$

e la (4.4) anche:

$$\begin{aligned}
&\lim_{x\to\infty} F_{t_1,\cdots,t_{i-1},t_i,t_{i+1},\cdots,t_{k+1}}(x_1,\cdots,x_{i-1},x,x_{i+1},\cdots,x_{k+1})\\
&= \lim_{x\to\infty} \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_{i-1}} \int_{-\infty}^{x} \int_{-\infty}^{x_{i+1}} \cdots \int_{-\infty}^{x_{k+1}} q_{t_1}(y_1)\cdots q_{t_{i-1}-t_{i-2}}(y_{i-1}-y_{i-2})\\
&\qquad q_{t_i-t_{i-1}}(y_i-y_{i-1}) q_{t_{i+1}-t_i}(y_{i+1}-y_i)\cdots q_{t_{k+1}-t_k}(y_{k+1}-y_k) dy_1\cdots dy_{i-1}\, dy_i\, dy_{i+1}\cdots dy_{k+1}\\
&= \lim_{x\to\infty} \int_{-\infty}^{x_1} q_{t_1}(y_1)\, dy_1 \cdots \int_{-\infty}^{x_{i-1}} q_{t_{i-1}-t_{i-2}}(y_{i-1}-y_{i-2})\, dy_{i-1}\\
&\qquad \int_{-\infty}^{x_{i+1}} \left(\int_{-\infty}^{x} q_{t_i-t_{i-1}}(y_i-y_{i-1})\, q_{t_{i+1}-t_i}(y_{i+1}-y_i)\, dy_i \right) dy_{i+1}\cdots\\
&\qquad\qquad \cdots \int_{-\infty}^{x_{k+1}} q_{t_{k+1}-t_k}(y_{k+1}-y_k)\, dy_{k+1}\\
&= \int_{-\infty}^{x_1} q_{t_1}(y_1)\, dy_1 \cdots \int_{-\infty}^{x_{i-1}} q_{t_{i-1}-t_{i-2}}(y_{i-1}-y_{i-2})\, dy_{i-1}\\
&\qquad \int_{-\infty}^{x_{i+1}} q_{t_{i+1}-t_i}(y_{i+1}-y_i)\, dy_{i+1} \cdots \int_{-\infty}^{x_{k+1}} q_{t_{k+1}-t_k}(y_{k+1}-y_k)\, dy_{k+1}\\
&= F_{t_1,\cdots,t_{i-1},t_{i+1},\cdots,t_{k+1}}(x_1,\cdots,x_{i-1},x_{i+1},\cdots,x_{k+1}), \quad \text{per } i=1,\cdots,k.
\end{aligned}$$

Nella penultima uguaglianza si è tenuto conto del fatto che la condizione $F_u * F_v = F_{u+v}$ implica che, posto $y = y_i - y_{i-1}$, in modo che $y_{i+1} - y_i = (y_{i+1} - y_{i-1}) - y$, si abbia

$$\begin{aligned}
&\lim_{x\to\infty} \int_{-\infty}^{x} q_{t_i-t_{i-1}}(y_i-y_{i-1})\, q_{t_{i+1}-t_i}(y_{i+1}-y_i)\, dy_i\\
&= \lim_{x\to\infty} \int_{-\infty}^{x-y_{i-1}} q_{t_i-t_{i-1}}(y) q_{t_{i+1}-t_i}(y_{i+1}-y_i-y)\, dy\\
&= \int_{-\infty}^{\infty} q_{t_i-t_{i-1}}(y) q_{t_{i+1}-t_i}(y_{i+1}-y_i-y)\, dy\\
&= q_{t_{i+1}-t_i} * q_{t_i-t_{i-1}}(y_{i+1}-y_{i-1}) = q_{t_{i+1}-t_{i-1}}(y_{i+1}-y_{i-1}).
\end{aligned}$$

Nel caso in cui F_v sia discreta il discorso si ripete identico, mettendo le densità discrete al posto delle densità di probabilità e le somme al posto degli integrali.

Come esempi di famiglie ad un parametro di funzioni di distribuzione, oltre al caso della famiglia gaussiana $F_u \sim N(0,u)$, che dà luogo al processo di Wiener standard, si possono considerare

1 il ***processo di Wiener con drift*** (o deriva) μ e ***coefficiente di diffusione*** σ^2, ovvero

$$F_u \sim N(\mu u, \sigma^2 u);$$

2 il ***processo di Cauchy***, ovvero il caso in cui

$$F_u \sim Cauchy(u), \text{ ovvero } q_u(x) = \frac{u}{\pi}\frac{1}{u^2+x^2};$$

3 il ***processo di Poisson*** di ***parametro*** λ, ovvero il caso in cui

$$F_u \sim Poisson(\lambda u), \text{ ovvero } p_u(k) = F_u(k) - F_u(k-1) = \frac{(\lambda u)^k}{k!}\exp(-\lambda u), \quad k \in \mathbb{N};$$

4 i ***processi di Poisson composti***, ovvero i processi $(X_t)_{t\geq 0}$ ottenuti per mezzo di un processo di Poisson $(N_t)_{t\geq 0}$ ed una successione di variabili aleatorie indipendenti ed identicamente distribuite $\{\xi_n\}_{n\in\mathbb{N}}$, (tutte indipendenti dal processo di Poisson) tramite la seguente regola

$$X_t = 0, \text{ se } N_t = 0; \qquad X_t = \sum_{k=1}^{n} \xi_k, \text{ se } N_t = n.$$

Terminiamo questa sezione ricordando anche la definizione di processi ad incrementi indipendenti rispetto ad una filtrazione $(\mathcal{F}_t)_{t\geq 0}$.

Definizione 4.10 (Processi ad incrementi indipendenti ed omogenei rispetto ad una filtrazione)**.** *Un processo* $(X_t)_{t\geq 0}$ *si dice* **ad incrementi indipendenti ed omogenei rispetto alla filtrazione** $(\mathcal{F}_t)_{t\geq 0}$**, con** $(\mathcal{F}_t)_{t\geq 0} \supseteq \mathcal{F}_t^X = \sigma\{X_u;\ u \in [0,t]\}$ *se*

(0) $X_0 = 0$*;*

(1) *per* $s, t \geq 0$ *gli incrementi* $X_{t+s} - X_t$ *sono variabili aleatorie indipendenti da* $\mathcal{F}_t$*;*

(2) *gli incrementi* $X_{t+s} - X_t$ *sono variabili aleatorie la cui distribuzione dipende solo da* s*.*

Si vede facilmente che questa definizione implica l'altra considerando che $X_{t_n} - X_{t_{n-1}}$ è indipendente da $\mathcal{F}_{t_{n-1}} \supseteq \sigma\{X_{t_i} - X_{t_{i-1}};\ i = 1, \cdots, n-1\}$. Si può anche vedere che la prima definizione implica la seconda con $\mathcal{F}_t = \mathcal{F}_t^X = \sigma\{X_u;\ u \in [0,t]\} = \sigma\{X_u - X_v;\ u, v \in [0,t]\}$.

4.7 Esempi di martingale a tempo continuo

In modo molto simile a quanto fatto a tempo discreto per le somme di v.a. indipendenti, si può mostrare[34] che se X_t è un processo ad incrementi indipendenti e omogenei, rispetto ad una filtrazione $(\mathcal{F}_t)_{t\geq 0}$, con $X_0 = 0$, e con media nulla allora X_t è una martingala, purché sia integrabile.

Inoltre è facile mostrare che se X_t è un processo ad incrementi indipendenti (e omogenei), integrabile e con $X_0 = 0$, allora $\mathbb{E}[X_t] = \mu t$ per t nei razionali:

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

da cui $\mathbb{E}[X_{1/n}] = \frac{1}{n}\mathbb{E}[X_1]$ e analogamente $\mathbb{E}[X_{m/n}] = \sum_{k=1}^{m} \mathbb{E}[X_{k/n} - X_{(k-1)/n}] = \frac{m}{n}\mathbb{E}[X_1]$.
Per ottenere che ciò valga anche per ogni t reale, si deve notare che comunque $\mathbb{E}[X_{t+s}] = \mathbb{E}[X_t] + \mathbb{E}[X_s]$ e aggiungere

[34] La condizione di misurabilità dipende dal fatto che $\mathcal{F}_t \supseteq \mathcal{F}_t^X$. La condizione di integrabilità è verificata per ipotesi. Infine

$$\mathbb{E}[X_{t+s} - X_t|\mathcal{F}_t] = \mathbb{E}[X_{t+s} - X_t] = \mathbb{E}[X_{t+s}] - \mathbb{E}[X_t] = 0 - 0 = 0,$$

dove nella prima uguaglianza si usa l'indipendenza di $X_{t+s} - X_t$ da $\mathcal{F}_t$, e nell'ultima si usa il fatto che la media di X_u è nulla. Si noti che l'omogeneità degli incrementi qui non è necessaria.

una piccola ulteriore **ipotesi di regolarità**: la $\mathbb{E}[X_t]$ è una funzione continua in t (o continua a destra). Con questa ipotesi si ottiene immediatamente la tesi per continuità.

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

Se ancora X_t ammette momento secondo finito, allora, con una dimostrazione simile[35] si ha che $Var(X_t) = \sigma^2 t$, purché si possa affermare a priori che $Var(X_t)$ è una funzione continua. Di nuovo similmente al caso a tempo discreto, accade che $(X_t - \mathbb{E}[X_t])^2 - Var(X_t) = (X_t - \mu t)^2 - \sigma^2 t$ è una martingala (a media nulla) .

Infine è possibile mostrare[36] che, sotto opportune ipotesi di regolarità (continuità in probabilità), se $\mathbb{E}[\exp\{\theta(X_t - \mu t)\}] < +\infty$, allora

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

e che quindi

$$Z_t := \exp\{\theta(X_t - \mu t) - K(\theta)t\} \tag{4.16}$$

è una martingala[37] a media 1.
Tutte le proprietà precedenti valgono anche per i processi $Y_t = Y_0 + X_t$, con dato iniziale Y_0 indipendente da $\{X_t\}$ (tranne per i valori medi). Bisogna però che Y_0 sia $\mathcal{F}_0$-misurabile[38], e soddisfi alcuni requisiti di integrabilità. Ad

[35] Si tratta di osservare che la varianza della somma degli incrementi è la somma delle varianze degli incrementi e quindi si procede come nel caso del valore atteso, sostituendo Var a $\mathbb{E}$.

[36] In questo caso, posto $X'_t = X_t - \mu t$ si sfrutta il fatto che

$$\exp\{\theta X'_1\} = \prod_{k=1}^{n} \exp\{\theta \left(X'_{k/n} - X'_{(k-1)/n}\right)\},$$

da cui, passando al valore atteso, per l'indipendenza degli incrementi e per l'omogeneità

$$\mathbb{E}[\exp\{\theta X'_1\}] = \prod_{k=1}^{n} \mathbb{E}[\exp\{\theta \left(X'_{k/n} - X'_{(k-1)/n}\right)\}] = \left(\mathbb{E}[\exp\{\theta \left(X'_{1/n} - X'_0\right)\}]\right)^n.$$

Posto $\exp\{K(\theta)\} := \mathbb{E}[\exp\{\theta X'_1\}]$ si ottiene dunque che

$$\mathbb{E}[\exp\{\theta \left(X'_{k/n} - X'_{(k-1)/n}\right)\}] = \mathbb{E}[\exp\{\theta \left(X'_{1/n}\right)\}] = \exp\{K(\theta)(1/n)\},$$

da cui ancora la tesi per ogni t razionale.

[37] Di nuovo misurabilità e integrabilità sono banali. Osservando che

$$\begin{aligned} Z_{t+s} &= \exp\{\theta(X_{t+s} - \mu(t+s)) - K(\theta)(t+s)\} \\ &= \exp\{\theta(X_t - \mu t) - K(\theta)t\} \exp\{\theta(X_{t+s} - X_t - \mu s) - K(\theta)s\} \\ &= Z_t \exp\{\theta(X_{t+s} - X_t - \mu s) - K(\theta)s\} \end{aligned}$$

si ottiene subito

$$\begin{aligned} \mathbb{E}[Z_{t+s} - Z_t|\mathcal{F}_t] &= \mathbb{E}\left[Z_t \left(\exp\{\theta(X_{t+s} - X_t - \mu s) - K(\theta)s\} - 1\right)|\mathcal{F}_t\right] \\ &\qquad \text{(per la misurabilità di } Z_t) \\ &= Z_t \mathbb{E}\left[(\exp\{\theta(X_{t+s} - X_t - \mu s) - K(\theta)s\} - 1)\,|\mathcal{F}_t\right] \\ &\qquad \text{(per l'indipendenza degli incrementi da } \mathcal{F}_t) \\ &= Z_t \mathbb{E}\left[(\exp\{\theta(X_{t+s} - X_t - \mu s) - K(\theta)s\} - 1)\right] \\ &\qquad \text{(per l'omogeneitàdegli incrementi)} \\ &= Z_t \mathbb{E}\left[(\exp\{\theta(X_s - \mu s) - K(\theta)s\} - 1)\right] = Z_t\,(1-1) = 0 \end{aligned}$$

[38] In realtà la richiesta che Y_0 sia $\mathcal{F}_0$-misurabile non è strettamente necessaria, se vale la condizione di indipendenza di tra il processo (X_t) e la variabile aleatoria Y_0: nel caso in cui $\mathcal{F}_t = \mathcal{F}_t^X$ si potrebbe, in alternativa, cambiare la filtrazione e prendere la filtrazione definita da $\mathcal{F}_t \vee \sigma\{Y_0\}$. In questo caso infatti il processo $Y_0 + X_t$ viene automaticamente adattato alla nuova filtrazione. Inoltre $\{X_t\}$ è ancora una martingala rispetto alla nuova filtrazione, come si può vedere facilmente usando la proprietà dei condizionamenti ridondanti: infatti la σ-algebra $\mathcal{H} = \sigma\{Y_0\}$ è indipendente da $X = X_{t+s}$ e da $\mathcal{G} = \mathcal{F}_t$.

$$\mathbb{E}[X|\mathcal{G} \vee \mathcal{H}] = \mathbb{E}[X|\mathcal{G}] \quad \Longleftrightarrow \quad \mathbb{E}[X_{t+s}|\mathcal{F}_t \vee \sigma\{Y_0\}] = \mathbb{E}[X_{t+s}|\mathcal{F}_t] = X_t;$$

lo stesso discorso vale nel caso in cui si assuma Y_0 indipendente da $\mathcal{F}_t \supseteq F_t^X$ per ogni t (e quindi anche dal processo $\{X_t\}$.

esempio

$$\widetilde{Z}_t := \mathbb{E}[\exp\{\theta(Y_t - \mu t) - K(\theta)t\}] \tag{4.17}$$

è ancora una martingala a media costante uguale a $\mathbb{E}[\widetilde{Z}_0] = \mathbb{E}[\exp\{\theta Y_0\}]$, purché ovviamente il valore medio di $\exp\{\theta Y_0\}$ sia finito.

Esempio 4.7 (Decomposizione di Doob e martingala esponenziale per il processo di Wiener)**.** *Come applicazione si consideri che il processo di Wiener standard o moto browniano W_t è una martingala, anche $M_t = W_t^2 - t$ e infine, per ogni θ reale*

$$Z_t^\theta := \exp\{\theta W_t - \frac{1}{2}\theta^2 t\} \tag{4.18}$$

è una martingala[39] *a media* 1*, che viene detta* **martingala esponenziale**.

Si noti che W_t^2 è una submartingala (in quanto quadrato di una martingala) e che si può decomporre nella somma di una martingala (la martingala m_t) e di un processo crescente (il processo deterministico t), cioè $W_t^2 = m_t + t$. Si tratta di un caso particolare della decomposizione di Doob a tempo continuo. Abbiamo quindi che il processo di Wiener W_t è una martingala a media nulla e con $W_t^2 - t$ ancora una martingala a media nulla. Questo fatto non caratterizza il processo di Wiener come mostra il seguente esempio. Tuttavia, se si aggiunge che si tratta di una martingala a traiettorie continue, allora le precedenti proprietà caratterizzano il processo di Wiener. Tale caratterizzazione è nota come caratterizzazione di Levy del processo di Wiener.

Esempio 4.8 (Decomposizione di Doob e martingala esponenziale per il processo di Poisson)**.** *Anche il processo di Poisson N_t di parametro λ, essendo un processo crescente è una submartingala, e si può decomporre nella somma di una martingala $M_t := N_t - \lambda t$ e di un processo crescente $A_t = \lambda t$ (che è poi il valore atteso di N_t). Si tratta anche qui di un caso particolare della decomposizione di Doob a tempo continuo.*

Applicando il risultato generale relativo ai processi a incrementi indipendenti, si ottiene che, poiché la varianza di N_t è λt, anche il processo $(N_t - \lambda t)^2 - \lambda t = M_t^2 - \lambda t$ è una martingala.

Va osservato che, nel caso $\lambda = 1$, il processo $M_t := N_t - t$ è una martingala a media nulla e che $M_t^2 - t$ è anche una martingala a media nulla.

Infine, essendo

$$\mathbb{E}[\exp\{\theta N_1\}] = \sum_{k=0}^{\infty} e^{\theta\, k}\, \frac{\lambda^k}{k!}\, e^{-\lambda} = \exp\{-\lambda(e^\theta - 1)\}$$

si ottiene che $K(\theta) = \log\big(\mathbb{E}[\exp\{\theta(N_1 - \lambda\}]\big) = -\lambda\,(e^\theta - \theta - 1)$. Di conseguenza la martingala (4.16) *è*

$$\mathcal{Z}_t^\theta = \exp\{\theta(N_t - \lambda\, t) + (e^\theta - \theta - 1)\,\lambda\, t\}$$

Si osservi che in generale data una v.a. Z non negativa e a media 1 in uno spazio $(\Omega, \mathcal{F}, \mathbb{P})$,

$$\mathbb{Q}(C) := \mathbb{E}^{\mathbb{P}}[\mathbb{I}_C Z], \qquad C \in \mathcal{F}, \tag{4.19}$$

definisce una nuova misura di probabilità.[40] $\mathbb{Q}$ su $(\Omega, \mathcal{F})$.

Esercizio 4.1 (Un caso particolare del Teorema di Girsanov)**.** *Posto*

$$Z = Z_T^\theta \Big(= \exp\{\theta W_T - \frac{1}{2}\theta^2 T\}\Big)$$

[39] Basta ricordare che $K(\theta)$ è definito dal fatto che $\exp\{K(\theta)\} = \mathbb{E}[\exp\{\theta(X_1 - \mu 1)\}]$. In questo caso quindi $\exp\{K(\theta)\} = \mathbb{E}[\exp\{\theta W_1\}] = \exp\{\frac{1}{2}\theta^2\}$, da cui $K(\theta) = \frac{1}{2}\theta^2$.

[40] È ovvio che $\mathbb{Q}(C) \geq 0$, essendo Z non negativa, e che $\mathbb{Q}(\Omega) = 1$, in quanto $\mathbb{Q}(\Omega) = \mathbb{E}^{\mathbb{P}}[\mathbb{I}_\Omega Z] = \mathbb{E}^{\mathbb{P}}[Z] = 1$. La σ-additività segue dalle proprietà di σ-additività di $\mathbb{P}$ e dal fatto che

$$\mathbb{I}_{\{\cup_n A_n\}} = \sum_n \mathbb{I}_{A_n}, \qquad \text{se gli insiemi } A_n \text{ sono disgiunti a due a due.}$$

dove Z_t^θ è la martingala esponenziale (4.18) relativa al processo di Wiener, e $\mathcal{F} = \mathcal{F}_T$ nell'espressione precedente (4.19), si trovi

1) la derivata di Radon-Nikodym

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

ovvero[41] *quella variabile aleatoria $h_t(\omega)$, $\mathcal{F}_t$-misurabile, tale che*

$$\mathbb{Q}(A) = \int_A h_t(\omega)\mathbb{P}(d\omega), \quad \textit{per ogni } A \in \mathcal{F}_t;$$

suggerimento: *si veda l'Esempio 3.7, riguardante le martingale e le derivate di Radon-Nikodym.*

soluzione: $\frac{d\mathbb{Q}}{d\mathbb{P}}\big|_{\mathcal{F}_t} = Z_t^\theta$

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

suggerimento: *basta capire che $\mathbb{E}^{\mathbb{Q}}[g(W_t)]$ si calcola equivalentemente come*

$$\begin{aligned}
&\mathbb{E}^{\mathbb{Q}}[g(W_t)] = \mathbb{E}^{\mathbb{P}}[g(W_t)Z_t^\theta] = \mathbb{E}^{\mathbb{P}}[g(W_t)\exp\{\theta W_t - \frac{1}{2}\theta^2 t\}] \\
&= \int_{\mathbb{R}} g(w)\exp\{\theta w - \frac{1}{2}\theta^2 t\}\exp\{-\frac{1}{2t}w^2\}dw = \int_{\mathbb{R}} g(w)\exp\{-\frac{1}{2t}(w^2 - 2w\theta t + \theta^2 t^2)\}dw \\
&= \int_{\mathbb{R}} g(w)\exp\{-\frac{1}{2t}(w-\theta t)^2\}dw
\end{aligned}$$

soluzione: *la legge di W_t è $N(\theta t, t)$*

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

suggerimento: *si tratta di capire che $\mathbb{E}^{\mathbb{Q}}[g_1(W_{t_1})g_2(W_{t_2} - W_{t_1})\cdots g_n(W_{t_n} - W_{t_n-1})]$ si calcola come*

$$\begin{aligned}
&\mathbb{E}^{\mathbb{P}}[g_1(W_{t_1})g_2(W_{t_2} - W_{t_1})\cdots g_n(W_{t_n} - W_{t_n-1})Z_{t_n}^\theta] \\
&= \mathbb{E}^{\mathbb{P}}[g_1(W_{t_1})g_2(W_{t_2} - W_{t_1})\cdots g_n(W_{t_n} - W_{t_n-1})\prod_{k=1}^n \frac{Z_{t_k}^\theta}{Z_{t_{k-1}}^\theta}] \\
&= \mathbb{E}^{\mathbb{P}}[g_1(W_{t_1})g_2(W_{t_2} - W_{t_1})\cdots g_n(W_{t_n} - W_{t_{n-1}})\prod_{k=1}^n \exp\{\theta(W_{t_k} - W_{t_{k-1}}) - \frac{1}{2}\theta^2(t_k - t_{k-1})\}] \\
&= \mathbb{E}^{\mathbb{P}}[\prod_{k=1}^n g_k(W_{t_k} - W_{t_{k-1}})\exp\{\theta(W_{t_k} - W_{t_{k-1}}) - \frac{1}{2}\theta^2(t_k - t_{k-1})\}] \\
&= \prod_{k=1}^n \mathbb{E}^{\mathbb{P}}[g_k(W_{t_k} - W_{t_{k-1}})\exp\{\theta(W_{t_k} - W_{t_{k-1}}) - \frac{1}{2}\theta^2(t_k - t_{k-1})\}] \\
&= \prod_{i=1}^n \int_{\mathbb{R}} g(w_i - w_{i-1})\exp\{-\frac{(w_i - w_{i-1} - \theta(t_i - t_{i-1}))^2}{2(t_i - t_{i-1})}\}dw_i.
\end{aligned}$$

dove per convenzione si è posto $t_0 = 0$ e $w_0 = 0$.

soluzione: *il processo $\{W_t\}$ diviene sotto $\mathbb{Q}$ un processo di Wiener con drift θ e coefficiente di diffusione 1.*

[41] Si ricordi la nota 2.3.

4.8 Processi di Markov regolari

I processi di Markov regolari sono processi costruiti attraverso una ***famiglia di probabilità di transizione regolari*** $P(s,t,x,A)$, e attraverso una misura di probabilità μ_0, detta ***(misura delle) probabilità iniziali***.

Definizione 4.11 (Probabilità di transizione regolari)**.** *Una famiglia* $P(s,t,x,A)$

$$P : (\mathbb{R}^+ \times \mathbb{R}^+)_+ \times \mathbb{R} \times \mathcal{B}(\mathbb{R}) \to [0,1], (s,t,x,A) \to P(s,t,x,A),$$

dove

$$(\mathbb{R}^+ \times \mathbb{R}^+)_+ = \{(s,t) \in \mathbb{R}^+ \times \mathbb{R}^+, \text{ tali che } s \le t\},$$

che soddisfa le seguenti proprietà:

(i) *per ogni* $0 \le s \le t$ *ed* $x \in \mathbb{R}$,

$$P(s,t,x,\cdot) : \mathcal{B}(\mathbb{R}) \to [0,1], A \to P(s,t,x,A)$$

è una misura di probabilità,

(ii) *per ogni* $0 \le s \le t$ *ed* $A \in \mathcal{B}(\mathbb{R})$,

$$P(s,t,\cdot,A) : \mathbb{R} \to [0,1], x \to P(s,t,x,A)$$

è una funzione misurabile,

(iii) *la famiglia* $P(\cdot,\cdot,\cdot,\cdot)$ *soddisfa l'***equazione di Chapman-Kolmogorov**, *cioè per ogni* $0 \le r \le s \le t$, *per ogni* $x \in \mathbb{R}$ *e per ogni* $A \in \mathcal{B}(\mathbb{R})$ *vale*

$$P(r,t,x,A) = \int_{\mathbb{R}} P(r,s,x,dy)P(s,t,y,A),$$

viene detta **famiglia di probabilità di transizione regolari**.
Si noti che quindi necessariamente $P(t,t,x,A) = \delta_x(A)$, *cioè vale* 1 *se* $x \in A$ *e vale* 0 *altrimenti e che le proprietà (i) e (ii) permettono di dare senso all'integrale in (iii).*

Osservazione 4.3. *E' interessante notare che l'equazione di Chapman-Kolmogorov, nel caso in cui* $P(s,t,x,\cdot)$ *ammette densità, ovvero*

$$P(s,t,x,A) = \int_A p(s,t,x,y)\,dy,$$

diviene

$$\int_A p(r,t,x,z)\,dz = \int_{\mathbb{R}} p(r,s,x,y)\,dy \int_A p(s,t,y,z)\,dz = \int_A \left(\int_{\mathbb{R}} p(r,s,x,y)p(s,t,y,z)\,dy \right) dz,$$

ovvero

$$p(r,t,x,z) = \int_{\mathbb{R}} p(r,s,x,y)p(s,t,y,z)\,dy.$$

Attraverso le probabilità di transizioni regolari e la probabilità iniziale, si può definire una famiglia di funzioni di distribuzione finito dimensionale nel seguente modo: si definiscano, per $0 \le t_1 \le \cdots \le t_k$, e per $(z_0, z_1, \cdots, z_k) \in \mathbb{R}^{k+1}$,

$$\begin{aligned} F_{0,t_1,\cdots,t_k}(z_0,z_1,\cdots,z_k) \quad := \quad & \int_{-\infty}^{z_0} \int_{-\infty}^{z_1} \cdots \int_{-\infty}^{z_k} \mu_0(dx_0)P(0,t_1,x_0,dy_1)P(t_1,t_2,y_1,dy_2)\cdots \\ & \cdots P(t_{k-2},t_{k-1},y_{k-2},dy_{k-1})P(t_{k-1},t_k,y_{k-1},dy_k), \end{aligned}$$

ed

$$F_{t_1,\cdots,t_k}(z_1,\cdots,z_k) := \lim_{z\to\infty} F_{0,t_1,\cdots,t_k}(z,z_1,\cdots,z_k),$$

Se invece non vale $0 \le t_1 \le \cdots \le t_k$, si definiscono attraverso un'opportuna permutazione π per la quale $0 \le t_{\pi_1} \le \cdots \le t_{\pi_k}$ ed in modo che valga la condizione di consistenza **(C1)**.

La proprietà (iii), ovvero l'equazione di Chapman-Kolmogorov, permette di verificare immediatamente la condizione di consistenza **(C2′)**[42]. La famiglia di distribuzioni finito-dimensionali così ottenute soddisfa quindi le proprietà di consistenza del teorema di Kolmogorov, e pertanto esiste un processo con queste distribuzioni finito-dimensionali. Un processo le cui distribuzioni finito-dimensionali si possono ottenere attraverso una famiglia di probabilità di transizione regolare e una probabilità iniziale come sopra, viene detto ***processo di Markov (o processo markoviano) regolare***.

Esempio 4.9. *Il processo di Wiener ed il processo di Poisson sono processi markoviani regolari in questo senso con* $\mu_0 = \delta_{\{0\}}$ *per entrambi i processi e*

$$P(s,t,x,dy) = \frac{1}{\sqrt{2\pi(t-s)}} \exp\{-\frac{(y-x)^2}{2(t-s)}\} dy$$

per il processo di Wiener, mentre

$$P(s,t,n,\{m\}) = \frac{(\lambda(t-s))^{n-m}}{(n-m)!} \exp\{-\lambda(t-s)\},\ 0 \le m \le n$$

per il processo di Poisson.

Esempio 4.10 (Processi di Markov regolari omogenei e processi a incrementi indipendenti e omogenei)**.** *In entrambi i casi dell'esempio precedente* $P(s,t,x,A)$ *dipende solo dalla differenza* $t-s$. *Più in generale ogni volta che*

[42] Nel caso con densità la verifica è simile a quella del caso dei processi ad incrementi indipendenti ed omogenei: si prendano $k \ge 1$ e $0 < t_1 < t_2 < \cdots, < t_k < t_{k+1}$, allora

$$\begin{aligned}
&\lim_{x\to\infty} F_{0,t_1,\cdots,t_{i-1},t_i,t_{i+1},\cdots,t_{k+1}}(x_0,x_1,\cdots,x_{i-1},x,x_{i+1},\cdots,x_{k+1})\\
&= \lim_{x\to\infty} \int_{-\infty}^{x_0}\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_{i-1}}\int_{-\infty}^{x}\int_{-\infty}^{x_{i+1}}\cdots\int_{-\infty}^{x_{k+1}} \mu_0(dy_0)p(0,t_1,y_0,y_1)\cdots p(t_{i-2},t_{i-1},y_{i-2},y_{i-1})\\
&\qquad p(t_{i-1},t_i,y_{i-1},y_i)p(t_i,t_{i+1},y_i,y_{i+1})\cdots p(t_k,t_{k+1},y_k,y_{k+1})dy_1\cdots dy_{i-1}\,dy_i\,dy_{i+1}\cdots dy_{k+1}\\
&= \lim_{x\to\infty} \int_{-\infty}^{x_0}\mu_0(dy_0)\int_{-\infty}^{x_1} p(0,t_1,y_0,y_1)\,dy_1\cdots\int_{-\infty}^{x_{i-1}} p(t_{i-2},t_{i-1},y_{i-2},y_{i-1})\,dy_{i-1}\\
&\qquad \int_{-\infty}^{x_{i+1}}\left(\int_{-\infty}^{x} p(t_{i-1},t_i,y_{i-1},y_i)\,p(t_i,t_{i+1},y_i,y_{i+1})\,dy_i\right)\,dy_{i+1}\cdots\int_{-\infty}^{x_{k+1}} p(t_k,t_{k+1},y_k,y_{k+1})\,dy_{k+1} =\\
&= \int_{-\infty}^{x_0}\mu_0(dy_0)\int_{-\infty}^{x_1} p(0,t_1,y_0,y_1)\,dy_1\cdots\int_{-\infty}^{x_{i-1}} p(t_{i-2},t_{i-1},y_{i-2},y_{i-1})\,dy_{i-1}\\
&\qquad \int_{-\infty}^{x_{i+1}}\left(\int_{-\infty}^{\infty} p(t_{i-1},t_i,y_{i-1},y_i)\,p(t_i,t_{i+1},y_i,y_{i+1})\,dy_i\right)\,dy_{i+1}\cdots\int_{-\infty}^{x_{k+1}} p(t_k,t_{k+1},y_k,y_{k+1})\,dy_{k+1}\\
&= \int_{-\infty}^{x_0}\mu_0(dy_0)\int_{-\infty}^{x_1} p(0,t_1,y_0,y_1)\,dy_1\cdots\int_{-\infty}^{x_{i-1}} p(t_{i-2},t_{i-1},y_{i-2},y_{i-1})\,dy_{i-1}\\
&\qquad \int_{-\infty}^{x_{i+1}} p(t_{i-1},t_{i+1},y_{i-1},y_{i+1})\,dy_{i+1}\cdots\int_{-\infty}^{x_{k+1}} p(t_k,t_{k+1},y_k,y_{k+1})\,dy_{k+1}\\
&= F_{0,t_1,\cdots,t_{i-1},t_{i+1},\cdots,t_{k+1}}(x_0,x_1,\cdots,x_{i-1},x_{i+1},\cdots,x_{k+1}), \quad \text{per } i=1,\cdots,k.
\end{aligned}$$

Nella penultima uguaglianza si è tenuto conto dell'equazione di Chapman-Kolmogorov. Il caso $i=0$ è verificato per definizione. Infine

$$\begin{aligned}
&\lim_{x\to\infty} F_{0,t_1,\cdots,t_k,t_{k+1}}(x_0,x_1,\cdots,x_k,x)\\
&= \lim_{x\to\infty} \int_{-\infty}^{x_0}\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_k}\int_{-\infty}^{x} \mu_0(dy_0)p(0,t_1,y_0,y_1)\cdots p(t_{k-1},t_k,y_{k-1},y_k)p(t_k,t_{k+1},y_k,y)\,dy_0\,dy_1\cdots dy_k\,dy\\
&= \lim_{x\to\infty} \int_{-\infty}^{x_0}\mu_0(dy_0)\int_{-\infty}^{x_1} p(0,t_1,y_0,y_1)\,dy_1\cdots\int_{-\infty}^{x_k} p(t_{k-1},t_k,y_{k-1},y_k)\,dy_k\int_{-\infty}^{x}\cdots p(t_k,t_{k+1},y_k,y)\,dy\\
&\quad = \int_{-\infty}^{x_0}\mu_0(dy_0)\int_{-\infty}^{x_1} p(0,t_1,y_0,y_1)\,dy_1\cdots\int_{-\infty}^{x_k} p(t_{k-1},t_k,y_{k-1},y_k)\,dy_k\int_{-\infty}^{\infty}\cdots p(t_k,t_{k+1},y_k,y)\,dy\\
&\quad = \int_{-\infty}^{x_0}\mu_0(dy_0)\int_{-\infty}^{x_1} p(0,t_1,y_0,y_1)\,dy_1\cdots\int_{-\infty}^{x_k} p(t_{k-1},t_k,y_{k-1},y_k)\,dy_k = F_{0,t_1,\cdots,t_k}(x_0,x_1,\cdots,x_k),
\end{aligned}$$

$P(s,t,x,A) = P(s+h,t+h,x,A) = P(0,t-s,x,A)$ *per ogni* s,t,h,x,A, *si parla di* **processo di Markov omogeneo (nel tempo)**. *Inoltre se per ogni* s,t,h,x,z,y,

$$P(s,t,x,(-\infty,z]) = P(s+h,t+h,x+y,(-\infty,z+y]), \tag{4.20}$$

allora in realtà si tratta di esempi di **processi a incrementi indipendenti ed omogenei (nello spazio)** *(con* $X_0 = 0$, *se* μ_0 *è la misura concentrata in* 0*). Nel caso generale la famiglia delle distribuzioni finito-dimensionali così ottenuta coincide con la famiglia delle distribuzioni finito-dimensionali di un processo* $(X'_t)_{t\geq 0}$, *con*

$$X'_t = Y_0 + X_t,$$

dove $(X_t)_{t\geq 0}$ *è un processo ad incrementi indipendenti ed omogenei con* $X_0 = 0$, *ed* Y_0 *è una variabile aleatoria con funzione di distribuzione* $F_0(y) = \mu_0((-\infty,y])$, *indipendente dal processo* $(X_t)_{t\geq 0}$.

Per verificare le precedenti affermazioni si osservi che, prendendo $h = -s$ *ed* $y = -x$ *nella formula (4.20) si ottiene*

$$P(s,t,x,(-\infty,z]) = P(0,t-s,0,(-\infty,z-x]),$$

allora basta porre

$$F_u(z) = P(0,u,0,(-\infty,z]).$$

Sempre nel caso con densità, ovvero nel caso in cui $F_u(z) = \int_{-\infty}^{z} q_u(y)dy$, *l'equazione di Chapman-Kolmogorov diviene*

$$q_{t-r}(z-x) = \int_{\mathbb{R}} q_{s-r}(y-x)q_{t-s}(z-y)\,dy$$

ovvero, ponendo $s-r = u$, $t-s = v$, $\zeta = z-x$ *ed* $\eta = y-x$,

$$q_{u+v}(\zeta) = \int_{\mathbb{R}} q_u(\eta)q_v(\zeta-\eta)\,d\eta, \quad \textit{ovvero} \quad q_{u+v} = q_u * q_v,$$

che è esattamente la condizione di compatibilità già incontrata per i processi ad incrementi indipendenti ed omogenei, e ciò dimostra che la famiglia markoviana di distribuzioni finito-dimensionali definita con $\mu_0 = \delta_{\{0\}}$, *e quella definita per i processi* $(X_t)_{t\geq 0}$ *ad incrementi indipendenti ed omogenei (si vedano le definizioni 4.9 e 4.10), sono le stesse, ovvero che* $(X_t)_{t\geq 0}$ *è un processo di quest'ultimo tipo.*

Infine per mostrare che le distribuzioni finito-dimensionali del processo[43] $(X'_t)_{t\geq 0} = (Y_0 + X_t)_{t\geq 0}$, *con* Y_0 *indipendente da* $(X_t)_{t\geq 0}$,*sono quelle della famiglia markoviana, osserviamo che, nel caso in cui* $\mu_0(dy) = p_0(y)dy$, *si ottiene*

$$\begin{aligned}
&\mathbb{P}(X'_0 \leq z_0, X'_{t_1} \leq z_1, \cdots, X'_{t_k} \leq z_k)\\
&= \mathbb{P}(Y_0 \leq z_0, Y_0 + X_{t_1} \leq z_1, \cdots, Y_0 + X_{t_k} \leq z_k)\\
&= \int_{-\infty}^{z_0} \mathbb{P}(Y_0 \in dx_0)\mathbb{P}(Y_0 + X_{t_1} \leq z_1, \cdots, Y_0 + X_{t_k} \leq z_k | Y_0 = x_0)\\
&= \int_{-\infty}^{z_0} \mathbb{P}(Y_0 \in dx_0)\mathbb{P}(X_{t_1} \leq z_1 - x_0, \cdots, X_{t_k} \leq z_k - x_0 | Y_0 = x_0)\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \textit{(per l'indipendenza di } (X_t)_{t\geq 0} \textit{ ed } Y_0\textit{)}\\
&= \int_{-\infty}^{z_0} \mathbb{P}(Y_0 \in dx_0)\mathbb{P}(X_{t_1} \leq z_1 - x_0, \cdots, X_{t_k} \leq z_k - x_0)\\
&= \int_{-\infty}^{z_0} p_{Y_0}(x_0)dx_0 \int_{-\infty}^{z_1-x_0} \cdots \int_{-\infty}^{z_k-x_0} p_{X_{t_1},\cdots,X_{t_k}}(y'_1,\cdots,y'_k)\,dy'_1\cdots,\,dy'_k\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \textit{(posto } y_i = y'_i + x_0\textit{)}\\
&= \int_{-\infty}^{z_0} p_{Y_0}(x_0)dx_0 \int_{-\infty}^{z_1} \cdots \int_{-\infty}^{z_k} p_{X_{t_1},\cdots,X_{t_k}}(y_1 - x_0,\cdots,y_k - x_0)\,dy_1\cdots,\,dy_k\\
&= \int_{-\infty}^{z_0}\int_{-\infty}^{z_1} \cdots \int_{-\infty}^{z_k} p_0(x_0)\,dx_0\,q_{t_1}(y_1 - x_0)\,dy_1\,q_{t_2-t_1}(y_2 - y_1)\,dy_2\cdots\\
&\qquad\qquad \cdots q_{t_{k-1},t_{k-2}}(y_{k-1} - y_{k-2})\,dy_{k-1}\,q_{t_k-t_{k-1}}(y_k - y_{k-1})\,dy_k,\\
&=: F_{0,t_1,\cdots,t_k}(z_0,z_1,\cdots,z_k).
\end{aligned}$$

[43] Si veda la nota 38, dove il processo $Y_0 + X_t$ è denotato come Y_t.

4.8.1 Processo di Orstein-Ulhenbeck

Si consideri la famiglia delle probabilità di transizione con densità, definita da

$$P(s,t,x,dy) = \frac{1}{\sqrt{2\pi\sigma^2 \frac{1-\mathrm{e}^{-2\lambda(t-s)}}{2\lambda}}} \exp\{-\frac{1}{2}\frac{(y-\mathrm{e}^{-\lambda(t-s)}x)^2}{\sigma^2 \frac{1-\mathrm{e}^{-2\lambda(t-s)}}{2\lambda}}\}dy.$$

Si può dimostrare che $P(s,t,x,dy)$ definisce una famiglia di probabilità di transizione regolare (gaussiana). Inoltre si potrebbe dimostrare che, se Y_0 è una variabile aleatoria con distribuzione μ_0, indipendente da un processo di Wiener standard $(W_t)_{t\geq 0}$, allora la famiglia di distribuzioni finito dimensionali ottenute tramite $P(s,t,x,dy)$ e la distribuzione iniziale μ_0, ha le stesse distribuzioni finito dimensionali del processo

$$X_t := \mathrm{e}^{-\lambda t}Y_0 + \sigma W_t - \int_0^t \sigma\lambda \mathrm{e}^{-\lambda(t-s)}W_s ds.$$

Si noti che l'integrale ha senso purché si prenda la versione a traiettorie continue del processo di Wiener standard[44].

È inoltre interessante notare che, se $Y_0 = x$, cioè se Y_0 è una variabile aleatoria degenere, allora $X_t := \mathrm{e}^{-\lambda t}x + \sigma W_t - \int_0^t \sigma\lambda \mathrm{e}^{-\lambda(t-s)}W_s ds$ è un processo gaussiano[45], di valore atteso $m(t) = \mathrm{e}^{-\lambda t}x$ e funzione di covarianza

$$K(s,t) = \mathbb{E}\left[\left(\sigma W_t - \int_0^t \sigma\lambda \mathrm{e}^{-\lambda(t-u)}W_u du\right)\left(\sigma W_s - \int_0^s \sigma\lambda \mathrm{e}^{-\lambda(s-u)}W_u du\right)\right].$$

Con un po' di calcoli[46] si può controllare che

$$K(s,t) = \frac{\sigma^2}{2\lambda}(\mathrm{e}^{-\lambda|t-s|} - \mathrm{e}^{-\lambda(t+s)}),$$

[44] Vale la pena ricordare che di solito questo processo viene introdotto dopo aver parlato dell'integrale stocastico rispetto al processo di Wiener. Riscrivendo

$$\sigma W_t - \int_0^t \sigma\lambda \mathrm{e}^{-\lambda(t-s)}W_s ds = \sigma \mathrm{e}^{-\lambda t}\left(\mathrm{e}^{\lambda t}W_t - \int_0^t \lambda \mathrm{e}^{\lambda s}W_s ds\right),$$

e interpretando la formula tra parentesi come un'integrazione per parti

$$\mathrm{e}^{\lambda t}W_t - \int_0^t W_s d\mathrm{e}^{\lambda s} =: \int_0^t \mathrm{e}^{\lambda s}dW_s,$$

si può riscrivere

$$X_t := \mathrm{e}^{-\lambda t}\left(Y_0 + \sigma\int_0^t \mathrm{e}^{\lambda s}dW_s\right),$$

o in forma differenziale (stocastica)

$$dX_t = -\lambda X_t dt + \sigma dW_t.$$

Per poter dare un senso più preciso a questa espressione, tuttavia andrebbe prima visto il significato dell'integrale stocastico.
In effetti la ***prima definizione di integrale stocastico data da Wiener*** è stata la seguente: per ogni funzione deterministica $h(t) \in C^1$, ovvero con derivata prima continua, Wiener definiva

$$\int_\alpha^\beta h(s)dW_s := W_\beta h(\beta) - W_\alpha h(\alpha) - \int_\alpha^\beta W_s h'(s)ds.$$

Questa variabile aleatoria risulta gaussiana (vedere la nota successiva) di valore medio nullo e di varianza $\int_\alpha^\beta h^2(s)ds$. Se $h(t) \notin C^1$, ma è misurabile e con $\int_\alpha^\beta h^2(s)ds$ finito, allora Wiener definiva

$$\int_\alpha^\beta h(s)dW_s := \lim_{n\to\infty}\int_\alpha^\beta h_n(s)dW_s,$$

dove il limite è inteso in media quadratica, e dove $\{h_n\}_n$ è una successione di funzioni C^1 con la proprietà che

$$\int_\alpha^\beta \big(h_n(s) - h(s)\big)^2 ds \to 0.$$

[45] Il fatto che sia gaussiano dipende dal fatto che l'integrale $\int_0^t \mathrm{e}^{-\lambda(t-s)}W_s ds = \mathrm{e}^{-\lambda t}\int_0^t \mathrm{e}^{\lambda s}W_s ds$ si può ottenere come limite delle somme di Riemann

$$\mathrm{e}^{-\lambda t}\sum_{k=1}^n \mathrm{e}^{\lambda s_k}W_{s_k}(s_k - s_{k-1}),$$

che sono variabili aleatorie gaussiane, e tenendo conto del fatto che il limite (in distribuzione) di variabili aleatorie gaussiane è ancora una variabile aleatoria gaussiana, purché valore atteso e varianza convergano.

[46] Questo conto risulta più agevole dopo aver introdotto l'integrale stocastico di Itô (vedere la Sezione 6)**.

ed in particolare quindi

$$Var(X_t) = K(t,t) = \frac{\sigma^2}{2\lambda}(1 - \mathrm{e}^{-2\lambda\, t})$$

Nel caso in cui $\lambda > 0$, quando t va all'infinito, allora chiaramente $\mathbb{E}[X_t] = m(t) = \mathrm{e}^{-2\lambda t}x$ converge a zero e la varianza $Var(X_t) = K(t,t)$ converge a $\frac{\sigma^2}{2\lambda}$. Di conseguenza la legge unidimensionale del processo $(X_t)_{t\geq 0}$ converge in distribuzione ad una legge gaussiana $N(0, \frac{\sigma^2}{2\lambda})$.

Questa legge gode di una interessante proprietà: se la legge iniziale μ_0 è appunto una legge gaussiana $N(0, \frac{\sigma^2}{2\lambda})$, cioè $\mu_0(dx) = \frac{1}{\sqrt{2\pi\frac{\sigma^2}{2\lambda}}}\exp\{-\frac{x^2}{2\frac{\sigma^2}{2\lambda}}\} = \sqrt{\frac{\lambda}{\pi\sigma^2}}\exp\{-\frac{\lambda}{\sigma^2}x^2\}$, allora la legge di X_t è ancora una legge gaussiana $N(0, \frac{\sigma^2}{2\lambda})$, ovvero:

$$\begin{aligned}
\mathbb{P}(X_t \in A) &= \int_{\mathbb{R}} \mu_0(dx) \int_A p(0,t,x,y)dy \\
&= \int_{\mathbb{R}} \sqrt{\frac{\lambda}{\pi\sigma^2}}\exp\{-\frac{\lambda}{\sigma^2}x^2\}dx \int_A \frac{1}{\sqrt{2\pi\sigma^2\frac{1-\mathrm{e}^{-2\lambda t}}{2\lambda}}}\exp\{-\frac{1}{2}\frac{(y-\mathrm{e}^{-\lambda t}x)^2}{\sigma^2\frac{1-\mathrm{e}^{-2\lambda t}}{2\lambda}}\}dy \\
&= \int_A \left(\int_{\mathbb{R}} \sqrt{\frac{\lambda}{\pi\sigma^2}}\exp\{-\frac{\lambda}{\sigma^2}x^2\}\sqrt{\frac{\lambda}{\pi\sigma^2(1-\mathrm{e}^{-2\lambda t})}}\exp\{-\frac{\lambda(y-\mathrm{e}^{-\lambda t}x)^2}{\sigma^2(1-\mathrm{e}^{-2\lambda t})}\}dx\right) dy \\
&= \int_A \sqrt{\frac{\lambda}{\pi\sigma^2}}\exp\{-\frac{\lambda}{\sigma^2}y^2\}\, dy, .
\end{aligned}$$

per ogni A boreliano[47].

Questo esempio porta a dare la definizione di distribuzione stazionaria.

Definizione 4.12 (Distribuzione stazionaria)**.** *Sia μ una distribuzione tale che*

$$\int_{\mathbb{R}} \mu(dx)P(s,t,x,A) = \mu(A)$$

allora μ viene detta **distribuzione stazionaria** *o* **invariante** *del processo di Markov regolare determinato dalle probabilità di transizione $P(s,t,x,A)$.*

Si può dimostrare che se $\mu_0 = \mu$ è una distribuzione stazionaria, allora le distribuzioni finito-dimensionali godono della proprietà di stazionarietà

Definizione 4.13 (Processi stazionari)**.** *Se per ogni $h > 0$, k, $(t_1, \cdots, t_k)$ e $(z_1, \cdots, z_k)$ la famiglia di distribuzioni finito-dimensionali $F_{t_1,\cdots,t_k}(z_1, \cdots, z_k)$ gode della proprietà che*

$$F_{t_1,\cdots,t_k}(z_0, z_1, \cdots, z_k) = F_{t_1+h,\cdots,t_k+h}(z_0, z_1, \cdots, z_k)$$

allora $F_{0,t_1,\cdots,t_k}$ è detta una famiglia di distribuzioni finito-dimensionali stazionarie, e il processo con tale famiglia di distribuzioni finito-dimensionali è detto un **processo stazionario**.

[47] L'ultima uguaglianza deriva dal fatto che, la densità di X_t è data dall'espressione tra parentesi, ovvero da

$$\begin{aligned}
&\sqrt{\frac{\lambda}{\pi\sigma^2}}\int_{\mathbb{R}} \frac{\sqrt{\lambda}}{\sqrt{\pi\sigma^2(1-\mathrm{e}^{-2\lambda t})}}\exp\{-(\frac{\lambda(y-\mathrm{e}^{-\lambda t}x)^2}{\sigma^2(1-\mathrm{e}^{-2\lambda t})} + \frac{\lambda x^2(1-\mathrm{e}^{-2\lambda t})}{\sigma^2(1-\mathrm{e}^{-2\lambda t})})\}dx \\
&= \sqrt{\frac{\lambda}{\pi\sigma^2}}\int_{\mathbb{R}} \frac{\sqrt{\lambda}}{\sqrt{\pi\sigma^2(1-\mathrm{e}^{-2\lambda t})}}\exp\{-\frac{\lambda}{\sigma^2(1-\mathrm{e}^{-2\lambda t})}\big((y-\mathrm{e}^{-\lambda t}x)^2 + x^2(1-\mathrm{e}^{-2\lambda t})\big)\}dx \\
&= \sqrt{\frac{\lambda}{\pi\sigma^2}}\int_{\mathbb{R}} \frac{\sqrt{\lambda}}{\sqrt{\pi\sigma^2(1-\mathrm{e}^{-2\lambda t})}}\exp\{-\frac{\lambda\big(y^2 - 2y\mathrm{e}^{-\lambda t}x + \mathrm{e}^{-2\lambda t}x^2 + x^2 - x^2\mathrm{e}^{-2\lambda t}\big)}{\sigma^2(1-\mathrm{e}^{-2\lambda t})}\}dx \\
&= \sqrt{\frac{\lambda}{\pi\sigma^2}}\int_{\mathbb{R}} \frac{\sqrt{\lambda}}{\sqrt{\pi\sigma^2(1-\mathrm{e}^{-2\lambda t})}}\exp\{-\frac{\lambda\big(x^2 - 2xy\mathrm{e}^{-\lambda t} + y^2\big)}{\sigma^2(1-\mathrm{e}^{-2\lambda t})}\}dx \\
&= \sqrt{\frac{\lambda}{\pi\sigma^2}}\int_{\mathbb{R}} \frac{\sqrt{\lambda}}{\sqrt{\pi\sigma^2(1-\mathrm{e}^{-2\lambda t})}}\exp\{-\frac{\lambda\big(x^2 - 2xy\mathrm{e}^{-\lambda t} + y^2\mathrm{e}^{-2\lambda t}\big)}{\sigma^2(1-\mathrm{e}^{-2\lambda t})}\}\exp\{-\frac{\lambda(-y^2\mathrm{e}^{-2\lambda t} + y^2)}{\sigma^2(1-\mathrm{e}^{-2\lambda t})}\}dx \\
&= \sqrt{\frac{\lambda}{\pi\sigma^2}}\exp\{-\frac{\lambda}{\sigma^2}y^2\}\int_{\mathbb{R}} \frac{\sqrt{\lambda}}{\sqrt{\pi\sigma^2(1-\mathrm{e}^{-2\lambda t})}}\exp\{-\frac{\lambda\big(x-y\mathrm{e}^{-\lambda t}\big)^2}{\sigma^2(1-\mathrm{e}^{-2\lambda t})}\}dx = \sqrt{\frac{\lambda}{\pi\sigma^2}}\exp\{-\frac{\lambda}{\sigma^2}y^2\}.
\end{aligned}$$

4.8.2 Moto browniano geometrico e modello di Black-Scholes

Il prossimo esempio è di fondamentale importanza nell'ambito dei modelli finanziari. Il ***modello proposto da Black e Scholes*** è un modello a tempo continuo con una azione rischiosa (una azione di prezzo S_t all'istante t) e una azione non rischiosa (di prezzo B_t all'istante t). Si suppone che l'evoluzione di B_t sia descritta dall'equazione differenziale seguente

$$dB_t = rB_t dt,$$

dove r è una costante positiva. Questo significa che il tasso di interesse è costante e uguale a r. Se $B_0 = 1$, allora ovviamente $B_t = e^{rt}$, per $t \geq 0$.
Si suppone che il prezzo dell'azione sia descritto[48] dal processo

$$S_t = S_0 \exp\left\{\mu t - \frac{\sigma^2}{2}t + \sigma W_t\right\}.$$

dove μ e σ sono due costanti e $(W_t)_t$ è un moto browniano standard, ed S_0 è una variabile aleatoria indipendente da $(W_t)_t$.

Tale processo è detto ***moto browniano geometrico***.[49]

Il modello è studiato sull'intervallo $[0, T]$, dove T è la data di scadenza dell'opzione da studiare. In particolare, risulta che la legge di S_t è una legge log-normale, vale a dire che il suo logaritmo segue una legge normale.
Il processo $(S_t)_t$ è una trasformazione biunivoca di un processo di Markov, infatti

$$Y_t = \log(S_t) = \log(S_0) + \sigma W_t + (\mu - \frac{\sigma^2}{2})t$$

è un processo ad incrementi indipendenti ed omogenei, con

$$Y_0 = \log(S_0),$$

e con

$$X_t := \sigma W_t + (\mu - \frac{\sigma^2}{2})t$$

un processo di Wiener, non standard, con coefficiente di drift $\mu' = \mu - \frac{\sigma^2}{2}$ e coefficiente di diffusione σ^2. Tenendo conto di questo fatto si può affermare immediatamente che il processo $(S_t)_t$ verifica le seguenti proprietà:

- continuità delle traiettorie;
- se $u \leq t$, S_t/S_u è indipendente da $\mathcal{F}_u^S = \sigma(S_v, v \leq u)$;
- se $u \leq t$, la legge di $(S_t - S_u)/S_u$ è identica a quella di $(S_{t-u} - S_0)/S_0$.

La prima proprietà è banale. Per verificare le altre due proprietà basta osservare che

$$S_t/S_u = \exp\{Y_t - Y_u\} = \exp\{\sigma(W_t - W_u) + (\mu - \frac{\sigma^2}{2})(t-u)\},$$

che $\mathcal{F}_u = \mathcal{F}_u^S = \mathcal{F}_u^W \vee \sigma(S_0)$ e che $(W_t - W_u)$ è indipendente da $\mathcal{F}_u^W$ e da $\sigma(S_0)$ ed ha la stessa legge di $(W_{t-u} - W_0) = W_{t-u}$.

[48] Di solito si introduce il processo del prezzo S_t come la soluzione dell'equazione differenziale stocastica

$$dX_t = X_t(\mu dt + \sigma dW_t), \quad X_0 = S_0,$$

che si risolve esplicitamente:

$$X_t = S_0 \exp\left\{\mu t - \frac{\sigma^2}{2}t + \sigma W_t\right\}.$$

Di nuovo per capire bene il senso dell'equazione differenziale stocastica precedente andrebbe prima chiarito il significato dell'integrale stocastico di Itô e del differenziale stocastico (vedere la Sezione 6.

[49] Questo processo si può ottenere anche come limite del processo dei prezzi per il modello di Cox, Ross e Rubistein. Per un approccio elementare si veda il testo di S. Ross [13].

Inoltre il processo $(S_t)_t$ risulta esso stesso un processo di Markov regolare: per $x_i > 0$, $i = 0, 1, \cdots, k$

$$
\begin{aligned}
&F_{S_0,S_1,\cdots,S_k}(x_0, x_1, \cdots, x_k) = \mathbb{P}(S_0 \le x_0, S_{t_1} \le x_1, \cdots, S_{t_k} \le x_k) \\
&= \mathbb{P}(Y_0 \le \log(x_0),\, Y_{t_1} \le \log(x_1), \cdots, Y_{t_k} \le \log(x_k)) \\
&= \int_{-\infty}^{\log(x_0)} \int_{-\infty}^{\log(x_1)} \cdots \int_{-\infty}^{\log(x_k)} p_{Y_0,Y_{t_1},\cdots,Y_{t_k}}(y_0, y_1, \cdots, y_k)\, dy_0\, dy_1 \cdots dy_k \\
&= \int_{-\infty}^{\log(x_0)} \int_{-\infty}^{\log(x_1)} \cdots \int_{-\infty}^{\log(x_k)} p_{Y_0}(y_0) \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi(t_i - t_{i-1})\sigma^2}} \mathrm{e}^{-\frac{(y_i - y_{i-1} - \mu'(t_i - t_{i-1}))^2}{2(t_i - t_{i-1})\sigma^2}}\, dy_0\, dy_1 \cdots dy_k
\end{aligned}
$$

di conseguenza la densità congiunta di $(S_0, S_{t_1}, \cdots, S_{t_k})$ si ottiene derivando rispetto a x_0, x_1, $\cdots$, x_k la precedente espressione:

$$
\begin{aligned}
&p_{S_0,S_{t_1},\cdots,S_{t_k}}(x_0, x_1, \cdots, x_k) = \frac{\partial^n}{\partial x_0 \partial x_1 \cdots \partial x_k} F_{S_0,S_1,\cdots,S_k}(x_0, x_1, \cdots, x_k) \\
&= p_{Y_0,Y_{t_1},\cdots,Y_{t_k}}(\log(x_0), \log(x_1), \cdots, \log(x_k)) \frac{1}{x_0} \frac{1}{x_1} \cdots \frac{1}{x_k} \\
&= \frac{p_{Y_0}(\log(x_0))}{x_0} \prod_{i=1}^{k} \frac{1}{x_i \sqrt{2\pi(t_i - t_{i-1})\sigma^2}} \mathrm{e}^{-\frac{(\log(x_i) - \log(x_{i-1}) - \mu'(t_i - t_{i-1}))^2}{2(t_i - t_{i-1})\sigma^2}}.
\end{aligned}
$$

La densità risulta evidentemente nulla se una delle x_i non è strettamente positiva.

Da ciò risulta evidente che le probabilità di transizione ammettono la seguente densità

$$
p(s, t, x, y) = \frac{1}{y\sqrt{2\pi(t-s)\sigma^2}} \exp\Big\{ - \frac{\big(\log(y) - \log(x) - \mu'(t-s)\big)^2}{2(t-s)\sigma^2} \Big\}.
$$

E infine facile vedere che queste densità di probabilità di transizione verificano l'equazione di Chapman-Kolmogorov in quanto ci si riporta immediatamente attraverso un cambio di variable al caso delle densità di probabilità di transizione del processo Y_t.

La formula di Black-Scholes permette di calcolare in modo esplicito il prezzo di un'opzione call europea. Si indichi con $C_0(x)$ il prezzo dell'opzione emessa al tempo 0 con la condizione $S_0 = x$. Si ponga

$$
Z_t^\theta = \exp\left\{ -\frac{1}{2}\theta^2 t + \theta W_t \right\}
$$

che è la martingala esponenziale (già definita in (4.18)) con $\mathbb{E}[Z_t^\theta] = 1$.

Se si prende $\theta = \frac{r-\mu}{\sigma}$ per il Teorema di Girsanov (vedere l'Esercizio 4.1) il processo scontato dei prezzi $(\tilde{S}_t = \frac{S_t}{B_t} = \mathrm{e}^{-rt} S_t)_t$ è una martingala[50] rispetto alla misura

$$
\mathbb{Q}(A) = \mathbb{Q}^\theta(A) = \mathbb{E}^{\mathbb{P}}[\mathbb{I}_A Z_T^\theta], \quad A \in \mathcal{F}_T,
$$

[50] Posto

$$
\widetilde{W}_t = \frac{\mu - r}{\sigma} t + W_t,
$$

il processo

$$
\tilde{S}_t = x \exp\{(\mu - r - \frac{\sigma^2}{2})t + \sigma W_t\}
$$

si può riscrivere anche come

$$
\tilde{S}_t = x \exp\{-\frac{\sigma^2}{2} t + \sigma \widetilde{W}_t\}
$$

Rispetto alla misura di probabilità $\mathbb{Q}$ il processo W_t risulta un processo di Wiener con coefficiente di drift θ, cioè equivalentemente

$$
W_t = (W_t - \theta t) + \theta t = B_t + \theta t,
$$

con B_t un processo di Wiener standard rispetto a $\mathbb{Q} = \mathbb{Q}^\theta$. Di conseguenza

$$
\widetilde{W}_t = \frac{\mu - r}{\sigma} t + W_t = \frac{\mu - r}{\sigma} t + \theta t + B_t.
$$

ovvero con $d\mathbb{Q} = Z_T^\theta d\mathbb{P}$. Estendendo la definizione data nel caso a tempo discreto, si ha quindi che la misura $\mathbb{Q}$ è una ***misura martingala equivalente*** a $\mathbb{P}$.

Inoltre si ha che, per ogni t, la legge della variabile aleatoria $S_t = \mathrm{e}^{rt}\widetilde{S}_t$ rispetto a $\mathbb{Q}$ è la stessa[51] della variabile aleatoria $x\exp\{(r-\sigma^2/2)t + \sigma\sqrt{t}Z\}$, dove Z è una variabile aleatoria gaussiana standard $N(0,1)$.

Quindi

$$\begin{aligned} C_0(x) &= e^{-rT}\mathbb{E}^{\mathbb{Q}}\left[(S_T-K)^+\right] \\ &= e^{-rT}\mathbb{E}^{\mathbb{Q}}\left[\left(xe^{(r-\frac{\sigma^2}{2})T+\sigma\sqrt{T}Z}-K\right)^+\right]. \end{aligned}$$

La speranza matematica a destra vale

$$e^{-rT}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\left(xe^{(r-\frac{\sigma^2}{2})T+\sigma\sqrt{T}z}-K\right)^+ e^{-z^2/2}dz.$$

L'integrando si annulla per $z \le \zeta$, dove

$$\zeta = \frac{1}{\sigma\sqrt{T}}\left(\log\frac{K}{x} - \left(r-\frac{\sigma^2}{2}\right)T\right),$$

e quindi

$$C_0(x) = \frac{e^{-rT}}{\sqrt{2\pi}}\int_{\zeta}^{+\infty}\left(xe^{(r-\frac{\sigma^2}{2})T+\sigma\sqrt{T}z}-K\right)e^{-z^2/2}dz.$$

Se si indica con Φ la funzione di ripartizione della legge $\mathcal{N}(0,1)$,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-z^2/2}dz = \frac{1}{\sqrt{2\pi}}\int_{-x}^{+\infty} e^{-z^2/2}dz,$$

allora

$$\begin{aligned} C_0(x) &= \frac{x}{\sqrt{2\pi}}\int_{\zeta}^{+\infty} e^{-\frac{1}{2}(z-\sigma\sqrt{T})^2}dz - Ke^{-rT}\Phi(-\zeta) \\ &= \frac{x}{\sqrt{2\pi}}\int_{\zeta-\sigma\sqrt{T}}^{+\infty} e^{-z^2/2}dz - Ke^{-rT}\Phi(-\zeta) \\ &= x\Phi\left(-\zeta+\sigma\sqrt{T}\right) - Ke^{-rT}\Phi(-\zeta). \end{aligned}$$

In conclusione si ottiene la *formula di Black-Scholes*

$$C_0(x) = x\Phi\left(-\zeta+\sigma\sqrt{T}\right) - Ke^{-rT}\Phi(-\zeta).$$

Per ulteriori approfondimenti consultare il libro di P. Baldi [1], quello di D. Lamberton e B. Lapeyre [8], oppure di J.M. Steele [15] .

Prendendo quindi

$$\theta = -\frac{\mu-r}{\sigma} = \frac{r-\mu}{\sigma},$$

si ottiene che $\widetilde{W}_t = B_t$ è un processo di Wiener standard rispetto a $\mathbb{Q} = \mathbb{Q}^\theta$.
A questo punto si osserva che, proprio per questo motivo,

$$\tilde{S}_t = x\exp\{-\frac{\sigma^2}{2}t + \sigma\widetilde{W}_t\}$$

è rispetto alla misura $\mathbb{Q}$ una martingala esponenziale ovvero

$$\tilde{S}_t = x\widetilde{Z}_t^\sigma = x\exp\{-\frac{\sigma^2}{2}t + \sigma\widetilde{W}_t\}, \tag{4.21}$$

si tratta di applicare il risultato dell'Esercizio 4.1 con $\mathbb{Q}$ al posto di $\mathbb{P}$, $\widetilde{W}_t$ al posto di W_t ed infine σ al posto di θ.

[51] L'uguaglianza in legge si vede dall'espressione del prezzo scontato (4.21), e tenendo conto del fatto che la legge della variabile aleatoria $\widetilde{W}_t$ rispetto a $\mathbb{Q}$ è $N(0,t)$ e che anche la legge di $\sqrt{t}Z$ ha legge $N(0,t)$, se Z ha legge $N(0,1)$. È importante sottolineare che ovviamente ciò vale solo come variabili aleatorie e non vale come processi.

Mediante la formula di Black-Scholes, possiamo ricavare il prezzo equo di un'opzione call europea. Tuttavia, grazie alla formula di parità (si veda ad esempio il libro di S. Ross [13]), si ricava immediatamente anche il prezzo equo di una put europea P_t, dato da

$$P_t = C_t - S_t + Ke^{-r(T-t)}.$$

La formula di Black-Scholes ha il pregio di essere semplice e di dipendere da tre parametri: r, μ e σ. L'unico parametro difficile da stimare è la volatilità σ[52]

Infine va notata l'analogia della formula di Black e Scholes con il corrispondente risultato per il modello di Cox, Ross e Rubistein (CRR). A tale proposito si veda, sempre nel testo di S. Ross[53] [13], come sia possibile ottenere tale formula in modo elementare come limite del prezzo del modello CRR, mandando il numero di passi all'infinito, con opportuni riscalamenti nel modello stesso.

[52]La volatilità è un parametro che gioca un ruolo importante nelle applicazioni. Per questo motivo, negli ultimi anni, è stato molto studiato, in statistica, il problema di stimare il coefficiente di diffusione, a partire dall'osservazione di una traiettoria.

[53]Il Ross del modello CRR e l'autore di [13] sono due persone diverse.

4.9 Appendice: dimostrazione del Teorema di esistenza di Kolmogorov

Un problema noto nel caso unidimensionale è il seguente: data una distribuzione di probabilità μ su $\mathbb{R}$, esiste uno spazio di probabilità $(\Omega,\mathcal{F},\mathbb{P})$ ed una variabile aleatoria X che ammette come distribuzione μ?

Ci sono due possibili risposte "classiche" a tale quesito: la prima consiste nel prendere lo spazio canonico $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$, X uguale all'identità, cioè $X(x) = x$, e infine $\mathbb{P} = \mu$; la seconda (Teorema di rappresentazione di Skorohod) consiste nel prendere $\Omega = (0,1)$, $\mathcal{F} = \mathcal{B}(0,1)$, $\mathbb{P}$ la misura di Lebesgue ristretta a (0,1) ed $X(u) = F^{-1}(u) := \inf\{x \text{ tali che } u \leq F(x)\}$, dove $F(x) := \mu\big((-\infty, x]\big)$ è la funzione di distribuzione ed F^{-1} è la ***funzione inversa generalizzata*** [54] della funzione di distribuzione F.

Anche nel caso dei processi c'è qualcosa di simile. Il problema si esprime nel seguente modo: data una famiglia di distribuzioni finito-dimensionali $\mu_{t_1,\cdots,t_k}$, al variare di $k \geq 1$ e di $t_1, \cdots, t_k$ in I, esiste uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ ed un processo aleatorio $(X_t, t \geq 0)$ che ammette tali distribuzioni finito-dimensionali?

Lo spazio canonico, analogo ad $\mathbb{R}$, è $\mathbb{R}^I = \{x(\cdot) : \ I \to \mathbb{R}\}$, ed il processo canonico è il processo coordinata per il quale $X_t(x(\cdot)) := x(t)$. La scelta della σ-algebra viene fatta in modo che il processo canonico sia misurabile. Quindi necessariamente deve contenere gli insiemi del tipo

$$\{x(\cdot) \in \mathbb{R}^I \mid x(t) \in A\}, \text{ per } A \in \mathcal{B}(\mathbb{R}),$$

e delle intersezioni finite di insiemi di questo tipo, i rettangoli di base finita

$$**\mathcal{C}^R_{t_1,\cdots,t_k;A_1\times A_2\times\cdots\times A_k} = **\{x(\cdot) \in \mathbb{R}^I \mid x(t_1) \in A_1, \cdots, x(t_k) \in A_k,\}, \text{ per } A_h \in \mathcal{B}(\mathbb{R}) \text{ e } h = 1, \cdots, k.$$

Necessariamente, quindi, deve contenere anche i cilindri (o insiemi finito-dimensionali), ovvero degli insiemi del tipo

$$**\mathcal{C}_{t_1,\cdots,t_k;H} = **\mathcal{C} = \{x(\cdot) \in \mathbb{R}^I \mid (x(t_1), x(t_2), \cdots, x(t_k)) \in H\}, \text{ dove } H \in \mathcal{B}(\mathbb{R}^k) \tag{4.22}$$

Si sceglie quindi la σ-algebra $\mathcal{R}^I$ generata dall'algebra $\mathcal{R}^I_0$ dei cilindri, al variare di $k \geq 1$, degli indici $t_1, \cdots, t_k$ in I e di H in $\mathcal{B}(\mathbb{R}^k)$[55].

Infine come misura di probabilità $\mathbb{P}$, chiaramente, si vuole prendere una misura di probabilità su $\mathcal{R}^I$ per la quale valga

$$\mathbb{P}(\mathcal{C}) = \mu_{t_1,\cdots,t_k}(H),$$

per ogni cilindro $\mathcal{C} = **\mathcal{C}_{t_1,\cdots,t_k;H} = **$ di $\mathcal{R}^I_0$, definito come in (4.22).

A questo punto si pone il problema: esiste una tale probabilità $\mathbb{P}$? e più precisamente

(i) la definizione di $\mathbb{P}$ è ben posta su $\mathcal{R}^I_0$? (ii) $\mathbb{P}$ si può estendere a tutto $\mathcal{R}^I$?

La risposta è affermativa sotto alcune semplici condizioni di consistenza ed è il contenuto del Teorema di esistenza di Kolmogorov. Come si intuisce dal discorso precedente l'ingrediente essenziale della dimostrazione di tale teorema è il procedimento di estensione di una misura definita su un'algebra alla σ-algebra da essa generata (Teorema di Caratheodory).

Le ***condizioni di consistenza*** sono le seguenti:

Sia $k > 1$ e sia π una permutazione di $\{1, \cdots, k\}$ e sia

$$\Phi_\pi : \mathbb{R}^k \to \mathbb{R}^k, (x_1, \cdots, x_k) \to \Phi_\pi(x_1, \cdots, x_k) := (x_{\pi_1}, \cdots, x_{\pi_k}).$$

Chiaramente, essendo Φ_π biunivoca, $(x_1, \cdots, x_k) \in H$ se e solo se $\Phi_\pi(x_1, \cdots, x_k) \in \Phi_\pi(H)$, per $H \in \mathcal{B}(\mathbb{R}^k)$, e quindi

$$\mu_{t_1,\cdots,t_k}(H) = \mu_{t_{\pi_1},\cdots,t_{\pi_k}}(\Phi_\pi(H)), \tag{4.23}$$

[54] Se non si è familiari con il teorema di Skorohod basta considerare solo il caso in cui la funzione di distribuzione $F(\cdot)$ sia strettamente crescente e continua e quindi invertibile.

[55] Per la dimostrazione del fatto che $\mathcal{R}^I_0$ sia un'algebra vedere la dimostrazione del seguente teorema.

infatti

$$\mu_{t_1,\cdots,t_k}(H) = \mathbb{P}\{(X_{t_1},\cdots,X_{t_k}) \in H\} = \mathbb{P}\{\Phi_\pi(X_{t_1},\cdots,X_{t_k}) \in \Phi_\pi(H)\} =$$
$$= \mathbb{P}\{(X_{t_{\pi_1}},\cdots,X_{t_{\pi_k}}) \in \Phi_\pi(H)\} = \mu_{t_{\pi_1},\cdots,t_{\pi_k}}(\Phi_\pi(H)),$$

inoltre

$$\mu_{t_1,\cdots,t_k,t_{k+1}}(H \times \mathbb{R}) = \mu_{t_1,\cdots,t_k}(H). \tag{4.24}$$

È immediato constatare che le condizioni di consistenza (4.23) e (4.24) si possono riscrivere nel seguente modo:

Siano $k \geq h \geq 1$, sia π una permutazione di $\{1,\cdots,k\}$ e sia $\Psi_{\pi,h} : \mathbb{R}^k \to \mathbb{R}^h$, $(x_1,\cdots,x_k) \to \Psi_{\pi,h}(x_1,\cdots,x_k) := (x_{\pi_1},\cdots,x_{\pi_h})$, allora, per ogni $(t_1,\cdots,t_k) \in I^k$ e $H \in \mathcal{B}(\mathbb{R}^h)$

$$\mu_{t_1,\cdots,t_k}(\Psi^{-1}_{\pi,h}(H)) = \mu_{t_{\pi_1},\cdots,t_{\pi_h}}(H), \tag{4.25}$$

infatti, nel caso in cui H sia un rettangolo cioè $H = A_1 \times \cdots \times A_h$, posto π^{-1} la permutazione inversa di π, pensata quindi come funzione biunivoca di $\{1,\cdots,k\}$ in sé, e posto $A_m = \mathbb{R}$ per $m = h+1,\cdots,k$, la relazione (4.25) diviene

$$\begin{aligned}
&\mathbb{P}((X_{t_{\pi_1}},\cdots,X_{t_{\pi_h}}) \in H) = \mathbb{P}(X_{t_{\pi_1}} \in A_1,\cdots,X_{t_{\pi_h}} \in A_h) = \\
&= \mathbb{P}(X_{t_{\pi_1}} \in A_1,\cdots,X_{t_{\pi_h}} \in A_h, X_{t_{\pi_{h+1}}} \in A_{h+1},\cdots,X_{t_{\pi_k}} \in A_k) = \\
&= \mathbb{P}(X_{t_{\pi_1}} \in A_{\pi^{-1}(\pi_1)},\cdots,X_{t_{\pi_h}} \in A_{\pi^{-1}(\pi_h)}, X_{t_{\pi_{h+1}}} \in A_{\pi^{-1}(\pi_{h+1})},\cdots,X_{t_{\pi_k}} \in A_{\pi^{-1}(\pi_k)}) = \\
&\qquad\qquad \text{(riordinando opportunamente le condizioni richieste)} \\
&= \mathbb{P}(X_{t_1} \in A_{\pi^{-1}(1)},\cdots,X_{t_h} \in A_{\pi^{-1}(h)}, X_{t_{h+1}} \in A_{\pi^{-1}(h+1)},\cdots,X_{t_k} \in A_{\pi^{-1}(k)}) = \\
&= \mathbb{P}((X_{t_1},\cdots,X_{t_k}) \in \Psi^{-1}_{\pi,h}(H)),
\end{aligned}$$

e se vale per i rettangoli, poi vale per ogni cilindro.

Prima di enunciare e dimostrare il teorema di Kolmogorov, menzioniamo il fatto che esiste un'altra possibilità, più simile a quanto fatto nel teorema di rappresentazione di Skorohod, nel caso in cui I è numerabile, cioè $I = \mathbb{N}$, ed illustreremo questo caso in seguito. In tale caso lo spazio canonico è di nuovo (0,1) con la $\sigma -$ algebra dei boreliani $\mathcal{B}(0,1)$ e la misura di Lebesgue ristretta a $(0,1)$, e quindi, a partire dalle distribuzioni finito-dimensionali non deve essere definita la misura di probabilità, bensì il processo stesso. Ingrediente essenziale per la dimostrazione è il teorema di esistenza delle versioni regolari delle distribuzioni condizionali di una variabile aleatoria, dato una variabile aleatoria multidimensionale.

Cominciamo con l'enunciare il Teorema di Kolmogorov[56]

Teorema 4.9 (di Kolmogorov)**.** *Sia data una famiglia $\mu_{t_1,\cdots,t_k}$ di distribuzioni finito-dimensionali consistente, cioè che verifica le condizioni di consistenza (4.23) e (4.24), allora esiste uno spazio di probabilità ed un processo aleatorio che ammette $\mu_{t_1,\cdots,t_k}$ come distribuzioni finito-dimensionali. Inoltre è sempre possibile prendere come spazio di probabilità lo spazio canonico $\mathbb{R}^I$ e come processo il processo canonico $X_t(x(\cdot)) = x(t)$.*

Dimostrazione: Cominciamo con l'ultima parte del teorema. Sia $(\Omega, \mathcal{F}, \mathbb{Q})$ uno spazio di probabilità ed $(Y_t, t \in I)$ un processo su tale spazio che ammette come distribuzioni finito-dimensionali $\mu_{t_1,\cdots,t_k}$. Si definisca la funzione $\xi : \Omega \to \mathbb{R}^I, \omega \to \xi(\omega)(\cdot) = Y(\cdot,\omega)$, cioè $\xi(\omega)$ è la funzione $t \to \xi(\omega)(t) = Y_t(\omega)$.

Risulta che ξ è una funzione $\mathcal{F}/\mathcal{R}^I$ misurabile: infatti per ogni cilindro $\mathcal{C}$ come in (4.22), l'insieme

$$\xi^{-1}(\mathcal{C}) = \{\omega \text{ tali che } (Y_{t_1},\cdots,Y_{t_k}) \in H\} \in \mathcal{F},$$

e ciò è sufficiente a dimostrare la misurabilità di ξ.

A questo punto ponendo

$$\mathbb{P}(\mathcal{A}) = \mathbb{Q}(\xi^{-1}(\mathcal{A})), \text{ per } \mathcal{A} \in \mathcal{R}^I,$$

si ottiene la rappresentazione canonica.

[56] Le condizioni di questa formulazione del Teorema di Kolmogorov leggermentepiù forti di quelle richieste nell'enunciato del Teorema 4.1, ma sono equivalenti.

Continuiamo dando lo **schema della dimostrazione**

punto 1) $\mathcal{R}_0^I$ è un'algebra.

punto 2) La definizione $\mathbb{P}$ su $\mathcal{R}_0^I$ come $\mathbb{P}(\mathcal{C}) := \mu_{t_1,\cdots,t_k}(H)$, dove $\mathcal{C}$ è il cilindro definito in (4.22), è ben posta e risulta $\mathbb{P}$ una misura di probabilità finitamente additiva su $\mathcal{R}_0^I$.

punto 3) La misura $\mathbb{P}$ è numerabilmente additiva su $\mathcal{R}_0^I$, e quindi si può estendere in modo univoco ad una misura alla σ-algebra $\mathcal{R}^I$ generata da $\mathcal{R}_0^I$.

Sia il punto 1) che il punto 2) sono basati sulla osservazione che un cilindro $\mathcal{C}$ può avere più di una rappresentazione:

$$\mathcal{C} = \{x(\cdot) \in \mathbb{R}^I \mid (x(t_1), x(t_2), \cdots, x(t_k)) \in H\}, \text{ con } H \in \mathcal{B}(\mathbb{R}^k),$$

oppure

$$\mathcal{C} = \{x(\cdot) \in \mathbb{R}^I \mid (x(s_1), x(s_2), \cdots, x(s_m)) \in J\}, \text{ con } J \in \mathcal{B}(\mathbb{R}^m).$$

Notazione: per k e $t_1, \cdots, t_k$ prefissati, indicheremo con $\mathcal{R}^{\{t_1,\cdots,t_k\}}$ la famiglia dei cilindri del tipo precedente al variare di $H \in \mathcal{B}(\mathbb{R}^k)$.

Supponiamo che $k \le m$, e che $\{t_1, \cdots, t_k\} \subseteq \{s_1, \cdots s_m\}$, allora esiste una permutazione σ di $\{1, \cdots, m\}$ per cui $t_i = s_{\sigma_i}$, $i = 1, \cdots, k$, per cui $\Psi_{\sigma,k}(J) = H$ e per cui $J = \Psi^{-1}_{\pi,h}(H)$ o, equivalentemente $\Phi_\sigma(J) = H \times \mathbb{R}^{m-k}$.

Si noti allora che dalla (4.25) si ottiene che

$$\mu_{s_1,\cdots,s_m}(J) = \mu_{s_1,\cdots,s_m}(\Psi^{-1}_{\sigma,k}(H)) = \mu_{s_{\sigma_1},\cdots,s_{\sigma_k}}(H) = \mu_{t_1,\cdots,t_k}(H). \tag{4.26}$$

Inoltre, due cilindri $\mathcal{C}$ e $\mathcal{C}'$, con indici di tempo $\{t_1, \cdots, t_k\}$ e $\{t'_1, \cdots, t'_{k'}\}$ rispettivamente, si possono sempre rappresentare come cilindri con indici comuni $\{s_1, \cdots s_m\} = \{t_1, \cdots, t_k\} \cup \{t'_1, \cdots, t'_{k'}\}$.

Più precisamente se

$$\mathcal{C} = \{x(\cdot) \in \mathbb{R}^I \mid (x(t_1), x(t_2), \cdots, x(t_k)) \in H\} \in \mathcal{R}^{\{t_1,\cdots,t_k\}}$$

e

$$\mathcal{C}' = \{x(\cdot) \in \mathbb{R}^I \mid (x(t'_1), x(t'_2), \cdots, x(t'_{k'})) \in H'\} \in \mathcal{R}^{\{t'_1,\cdots,t'_{k'}\}},$$

allora, posto

$$\{s_1, \cdots, s_m\} = \{t_1, \cdots, t_k\} \cup \{t'_1, \cdots, t'_{k'}\},$$

si può supporre, senza ledere in generalità (si tratta eventualmente di ricorrere a permutazioni opportune), che

$$\{s_1, \cdots, s_h, s_{h+1}, \cdots, s_k\} = \{t_1, \cdots, t_k\}, \{s_{h+1}, \cdots, s_k\} = \{t_1, \cdots, t_k\} \cap \{t'_1, \cdots, t'_{k'}\}$$

ed infine che

$$\{s_{h+1}, \cdots, s_m\} = \{t'_1, \cdots, t'_{k'}\}.$$

Infatti, per due opportune permutazioni π di $\{1, \cdots, k\}$ e π' di $\{1, \cdots, k'\}$, si può riscrivere $(t_{\pi_1}, \cdots, t_{\pi_k}) = (s_1, \cdots, s_k)$ e $(t'_{\pi_1}, \cdots, t'_{\pi_{k'}}) = (s_{h+1}, \cdots, s_m)$ e quindi

$$\begin{aligned} \mathcal{C} &= \{x(\cdot) \in \mathbb{R}^I \mid (x(s_1), x(s_2), \cdots, x(s_k)) \in \Phi_\pi(H)\} \\ &= \{x(\cdot) \in \mathbb{R}^I \mid (x(s_1), \cdots, x(s_m)) \in \Phi_\pi(H) \times \mathbb{R}^{m-k}\} \in \mathcal{R}^{\{s_1,\cdots,s_m\}} \end{aligned} \tag{4.27}$$

e

$$\begin{aligned} \mathcal{C}' &= \{x(\cdot) \in \mathbb{R}^I \mid (x(s_{h+1}), x(s_{h+2}), \cdots, x(s_m)) \in \Phi_{\pi'}(H')\} \\ &= \{x(\cdot) \in \mathbb{R}^I \mid (x(s_1), \cdots, x(s_m)) \in \mathbb{R}^h \times \Phi_{\pi'}(H')\} \in \mathcal{R}^{\{s_1,\cdots,s_m\}} \end{aligned} \tag{4.28}$$

Dimostrazione in dettaglio dei punti 1), 2) e 3).

punto 1)

È immediato capire che, fissato $k \geq 1$ e $\{t_1, \cdots, t_k\}$, la famiglia $\mathcal{R}^{\{t_1,\cdots,t_k\}}$ dei cilindri del tipo

$$\mathcal{C} = \{x(\cdot) \in \mathbb{R}^I \mid (x(t_1), x(t_2), \cdots, x(t_k)) \in H\},$$

con $H \in \mathcal{B}(\mathbb{R}^k)$, forma un'algebra (anche se in realtà si tratta di una $\sigma -$ algebra):

Per ottenere $\mathbb{R}^I$, basta prendere $H = \mathbb{R}^k$.

Per ottenere $\mathcal{C}^c$, basta prendere H^c.

Per ottenere $\mathcal{C} \cup \mathcal{C}'$, con $\mathcal{C}' = \{x(\cdot) \in \mathbb{R}^I \mid (x(t_1), x(t_2), \cdots, x(t_k)) \in H'\}$, basta prendere $H \cup H'$.

Dalle osservazioni precedenti sappiamo che comunque presi due cilindri essi si possono esprimere come cilindri con indici comuni. Si ottiene che quindi $\mathcal{R}_0^I$ è un'algebra.

punto 2)

Le osservazioni iniziali e le proprietà di consistenza permettono di ottenere immediatamente che la definizione è ben posta. Infatti se i cilindri $\mathcal{C}$ e $\mathcal{C}'$ di (4.27) e (4.28) sono uguali allora necessariamente $\Phi_\pi(H) \times \mathbb{R}^{m-k} = \mathbb{R}^h \times \Phi_{\pi'}(H') = J$, e allora per la (4.26) 5

$$\mu_{s_1,\cdots,s_m}(J) = \mu_{t_1,\cdots,t_k}(H) = \mu_{t'_1,\cdots,t'_{k'}}(H').$$

La finita additività dipende dal fatto che ciascuna $\mu_{t_1,\cdots,t_k}$ è una misura di probabilità su $\mathcal{R}^{\{t_1,\cdots,t_k\}}$ e che, dato un numero finito di cilindri di $\mathcal{R}_0^I$, si può sempre pensare che tutti i cilindri siano appartenenti ad un'algebra del tipo $\mathcal{R}^{\{t_1,\cdots,t_k\}}$.

punto 3)

Per provare la σ-additività è sufficiente mostrare la continuità, e per questo, a sua volta, è sufficiente mostrare che se $\mathcal{A}_n \in \mathcal{R}_0^I$ ed $\mathcal{A}_n \downarrow \emptyset$ allora $\mathbb{P}(\mathcal{A}_n) \downarrow 0$. La prova procede per assurdo. Esista, per assurdo, un $\epsilon > 0$ tale che $\mathbb{P}(\mathcal{A}_n) \geq \epsilon$ per ogni n. Se mostriamo che allora $\bigcap_n \mathcal{A}_n$ non può essere l'insieme vuoto la dimostrazione è completa.

Poiché per rappresentare un cilindro si può sempre aumentare il numero degli indici, cioè degli istanti di tempo coinvolti nella sua definizione, possiamo supporre che esista una successione $\{t_n, n \geq 1\}$ e una successione di boreliani $H_n \in \mathcal{B}(\mathbb{R}^n)$ per cui

$$\mathcal{A}_n = \{x \in \mathbb{R}^I, (x(t_1), \cdots, x(t_n)) \in H_n\}$$

(eventualmente ripetendo qualche $\mathcal{A}_n$[57])

Ovviamente $\mathbb{P}(\mathcal{A}_n) = \mu_{t_1,\cdots,t_n}(H_n)$, ed essendo $\mu_{t_1,\cdots,t_n}$ una ***misura internamente regolare*** [58] è possibile

[57] In generale ciascun $\mathcal{A}_n$ coinvolgerà gli istanti $t_1, \cdots, t_{m_n}$; in tale caso si dovrà ripetere $\mathbb{R}^I$ per $m_1 - 1$ volte, con $H_n = \mathbb{R}_n$ per $n = 1, \cdots, m_1 - 1, \mathcal{A}_1$ per $m_2 - m_1$ volte e così via.

[58] Ricordiamo che, se S è uno spazio metrico, si dice che una misura μ sui boreliani di S è ***internamente regolare*** se per ogni insieme misurabile B si ha che $\mu(B) = \sup\{\mu(J), J \subseteq B, J \text{ compatto}\}$, e quindi esiste una successione di compatti J_h contenuti in B per cui la misura di B è il limite delle misure di J_h.

Comunque in $\mathbb{R}^n$ ciò significa che per ogni boreliano H è possibile trovare una successione monotona di iper-rettangoli chiusi e limitati $\{K_m\}_{m\geq 1}$ e convergente ad H. Per la proprietà di continuità delle probabilità vale allora $\mu(K_m) \uparrow \mu(H)$.

Questo è un punto cruciale nella dimostrazione e che vale per tutte le misure di probabilità su uno spazio metrico localmente compatto. Anzi ciò vale addirittura per spazi di Hausdorff sempre localmente compatto (confrontare, ad esempio, il Teorema di rappresentazione di Riesz). Questa osservazione permette di estendere il teorema di Kolmogorov non solo al caso di processi a valori reali, ma anche al caso di processi a valori in spazi metrici localmente compatti, ed in particolare a $\mathbb{R}^d$.

trovare un compatto $K_n \subseteq H_n$ per cui

$$\mu_{t_1,\cdots,t_n}(H_n \backslash K_n) \leq \frac{\epsilon}{2^{n+1}},$$

di modo che se $\mathcal{B}_n = \{x \in \mathbb{R}^I, (x(t_1), \cdots, x(t_n)) \in K_n\}$ allora

$$\mathbb{P}(\mathcal{A}_n \backslash \mathcal{B}_n) \leq \frac{\epsilon}{2^{n+1}}.$$

Posto $\mathcal{C}_n = \bigcap_{m=1}^n \mathcal{B}_m$, di modo che

$$\mathcal{C}_n = \{x \in \mathbb{R}^I, (x(t_1), \cdots, x(t_n)) \in \Gamma_n\}, \text{ con } \Gamma_n = \bigcap_{m=1}^n (K_m \times \mathbb{R}^{n-m}) \subseteq K_n \text{ compatto,}$$

allora

$$\mathcal{C}_n \subseteq \mathcal{B}_n \subseteq \mathcal{A}_n \text{ e } \mathbb{P}(\mathcal{A}_n \backslash \mathcal{C}_n) < \frac{\epsilon}{2}.$$

Infatti

$$\mathbb{P}(\mathcal{A}_n \backslash \mathcal{C}_n) = \mathbb{P}(\mathcal{A}_n \cap \Big(\bigcap_{m=1}^n \mathcal{B}_m\Big)^c) = \mathbb{P}(\bigcup_{m=1}^n \mathcal{A}_n \cap \mathcal{B}_m^c) \leq$$

$$\leq \sum_{m=1}^n \mathbb{P}(\mathcal{A}_m \cap \mathcal{B}_m^c) = \sum_{m=1}^n \mathbb{P}(\mathcal{A}_m \backslash \mathcal{B}_m) \leq \sum_{m=1}^n \frac{\epsilon}{2^{m+1}} < \frac{\epsilon}{2}.$$

Di conseguenza $\mathbb{P}(\mathcal{C}_n) \geq \frac{\epsilon}{2} > 0$ (essendo $\mathbb{P}(\mathcal{A}_n) \geq \epsilon$ e $\mathbb{P}(\mathcal{A}_n \backslash \mathcal{C}_n) < \frac{\epsilon}{2}$). Quindi $\mathcal{C}_n$ non è vuoto e lo stesso vale per $\Gamma_n \subseteq K_n$.

Si scelga per ogni n un punto appartenente al cilindro $\mathcal{C}_n$, ovvero una funzione, $x^{(n)}(\cdot) \in \mathcal{C}_n$. Se$n \geq k$ allora $x^{(n)}(\cdot) \in \mathcal{C}_n \subseteq \mathcal{C}_k \subseteq \mathcal{B}_k$ e quindi il punto $(x^{(n)}(t_1), \cdots, x^{(n)}(t_k)) \in K_k \subseteq J_k \times J_k \times \cdots \times J_k$ (k volte, con J_k un intervallo limitato). Di conseguenza, ciascuna successione del tipo $\{x^{(1)}(t_k), x^{(2)}(t_k), \cdots, x^{(n)}(t_k), \cdots\}$ è contenuta in J_k, e quindi è limitata.

In realtà la successione $\{x^{(1)}(t_k), x^{(2)}(t_k), \cdots, x^{(n)}(t_k), \cdots\}$ è contenuta in J_k solo definitivamente, anzi $x^{(n)}(t_k) \in J_k$ per $n \geq k$, comunque è una successione che ammette almeno una sottosuccessione convergente. Anche questo è un punto da tenere presente nel caso in cui si voglia fare una generalizzazione a processi a valori in spazi più generali di $\mathbb{R}$ o $\mathbb{R}^d$.

Con il metodo diagonale si può scegliere una successione n_h crescente in modo che, qualunque sia $k \geq 1$, la successione $\{x^{(n_h)}(t_k)\}_{h \geq 1}$ sia convergente ad un punto y_k. Si noti che per ogni $k \geq 1$ il punto $(y_1, \cdots, y_k) \in K_k$, in quanto definitivamente $(x^{(n)}(t_1), \cdots, x^{(n)}(t_k)) \in K_k$.

Richiamo sul metodo diagonale:
Sia data una successione a due indici

$$\begin{array}{cccc} x_{1,1}, & x_{1,2}, & \cdots & \cdots \\ x_{2,1}, & x_{2,2}, & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ x_{n,1}, & x_{n,2}, & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{array}$$

per cui, qualunque sia n, la successione $\{x_{n,h}\}_{h \geq 1}$ è limitata (o relativamente compatta). Allora esiste una successione $\{h_m\}_{m \geq 1}$ di interi per cui, qualunque sia n, la successione $\{x_{n,h_m}\}_{m \geq 1}$ è convergente. Si considera la sottosuccessione convergente $\{x_{1,h_{1,m}}\}_{m \geq 1}$ ad y_1. Poi si considera la successione $\{x_{2,h_{1,m}}\}_{m \geq 1}$, che essendo limitata ammette una sottosuccessione convergente $\{x_{2,h_{2,m}}\}_{m \geq 1}$ ad y_2.

Si noti che quindi anche $\{x_{1,h_{2,m}}\}_{m\geq 1}$ è ancora una successione convergente ad y_1 e che anche ogni sua sottosuccessione è convergente ad y_1. Si continua in questo modo fino ad ottenere che $\{x_{n,h_{n,m}}\}_{m\geq 1}$ è una successione convergente ad y_n e $\{h_{n,m}\}_{m\geq 1}$ è una sottosuccessione di $\{h_{n-1,m}\}_{m\geq 1}$, ed anche ogni sottosuccessione di $\{x_{n,h_{n,m}}\}_{m\geq 1}$ converge ad y_n. A questo punto abbiamo una nuova configurazione

$$\begin{array}{cccc} x_{1,h_{1,1}}, & x_{1,h_{1,2}}, & \cdots & \to y_1 \\ x_{2,h_{2,1}}, & x_{2,h_{2,2}}, & \cdots & \to y_2 \\ \cdots & \cdots & \cdots & \cdots \\ x_{n,h_{n,1}}, & x_{n,h_{n,n}}, & \cdots & \to y_n \end{array}$$

Basta prendere $h_m = h_{m,m}$, infatti, qualunque sia n, la successione $\{x_{n,h_{m,m}}\}_{m\geq 1}$ è (definitivamente) una sottosuccessione di $\{x_{n,h_{n,m}}\}_{m\geq 1}$ e quindi converge ad y_n, per m che tende ad infinito.

Perciò, ogni funzione $x(\cdot)$ di $\mathbb{R}^I$ tale che $x(t_h) = y_h$, per ogni $h \geq 1$, appartiene ad ogni $\mathcal{B}_k$, $k \geq 1$, e quindi ad ogni $\mathcal{A}_k$, $k \geq 1$, e quindi $\bigcap_n \mathcal{A}_n$ risulta un insieme non vuoto.

□

Si noti che l'argomento della dimostrazione è lo stesso della dimostrazione del teorema di Tikhonov: il prodotto infinito di compatti è un compatto. Una dimostrazione simile appare anche nella dimostrazione che l'intersezione di una successione di compatti non vuoti è non vuota.

4.9.1 Caso a tempo discreto: metodo diretto

Passiamo ora ad illustrare il preannunciato metodo costruttivo di dimostrazione del teorema di Kolmogorov, ma prima di tutto, proviamo ad ottenere una variabile aleatoria bidimensionale con funzione di distribuzione $F(x, y)$ data. Notiamo che si può riscrivere

$$F(x, y) = \int_{-\infty}^{x} F_{Y|X}(y \mid z) dF_X(z),$$

dove $F_{Y|X}(y \mid x)|_{x=X(\omega)}$ è una versione regolare delle probabilità $\mathbb{P}(Y \leq y|\sigma(X))$.

Daremo per il momento per scontato che tale scomposizione (o meglio disintegrazione) sia sempre possibile.

Si possono inoltre definire le inverse generalizzate $\Gamma_X(\cdot)$ di $F_X(\cdot)$ e $\Gamma_{Y|X}(\cdot \mid x)$ di $F_{Y|X}(\cdot \mid x)$, qualunque sia x.

Siano ora U e V due v.a. uniformi in $(0, 1)$ ed indipendenti, si definiscano

$$\widetilde{X}(\omega) = \Gamma_X(U(\omega))$$

ed

$$\widetilde{Y}(\omega) = \Gamma_{Y|X}(V(\omega) \mid \widetilde{X}(\omega)) = \Gamma_{Y|X}(V(\omega) \mid \Gamma_X(U(\omega))).$$

È facile verificare che

$$\mathbb{P}(\widetilde{X}(\omega) \leq x, \widetilde{Y}(\omega) \leq y) = F(x, y).$$

Infatti

$$\begin{aligned} \mathbb{P}(\widetilde{X}(\omega) \leq x, \widetilde{Y}(\omega) \leq y) &= \mathbb{P}(\Gamma_X(U(\omega)) \leq x, \Gamma_{Y|X}(V(\omega) \mid \Gamma_X(U(\omega)) \leq y) = \\ &= \mathbb{P}\{U(\omega) \leq F_X(x), V(\omega) \leq F_{Y|X}(y \mid \Gamma_X(U(\omega))\} = \\ &= \int_{(0,F_X(x)]} F_{Y|X}(y \mid \Gamma_X(u)) du = \\ &\qquad \text{(applicando il cambio di variabile } z = \Gamma_X(u)) \\ &\qquad \text{(e considerando che } u \leq F_X(x) \Leftrightarrow \Gamma_X(u) = z \leq x) \\ &= \int_{-\infty}^{x} F_{Y|X}(y \mid z) dF_X(z) = F(x, y). \end{aligned}$$

È facile generalizzare al caso in dimensione n e costruire una variabile aleatoria n-dimensionale $(X_1, \cdots, X_n)$ con funzione di distribuzione data, a partire da n variabili aleatorie uniformi in $(0,1)$ ed indipendenti.

A questo punto è chiaro come estendere il procedimento al caso di una successione di v.a. $\{X_n\}_{n\geq 1}$ con distribuzioni finito-dimensionali $\mu_{1,\cdots,n}$ assegnate (e relativa funzione di ripartizione $F(x_1, x_2, \cdots, x_n)$) in modo che

$$\mu_{1,\cdots,n+1}(H \times \mathbb{R}) = \mu_{1,\cdots,n}(H),$$

pur di avere a disposizione una successione di v.a. uniformi in $(0,1)$ ed indipendenti.

Infine, va notato che sostanzialmente le condizioni di compatibilità servono solo a garantire che le distribuzioni finito-dimensionali della successione aleatoria così ottenuta, e relative a tempi non consecutivi siano quelle volute.

L'affermazione che una successione di variabili aleatorie $\{X_n\,,\ n \in \mathbb{N}\}$ è una successione di v.a. indipendenti con $\mu_{X_n} = \mu_n$, è un'affermazione che riguarda le distribuzioni finito dimensionali del processo $\{X_n\,,\ n \in \mathbb{N}\}$. L'esistenza di una tale successione si potrebbe quindi dedurre dal teorema di rappresentazione di Kolmogorov, o magari da un risultato *ad hoc* la cui prova fosse la semplificazione del procedimento usato nel dimostrare tale teorema. Tuttavia l'esistenza di una tale successione tuttavia si può dedurre direttamente, pur di dare per scontato che esiste la misura di Lebesgue su $(0,1)$. Infatti su $(0,1)$ si possono definire delle variabili aleatorie indipendenti ed identicamente distribuite, a valori nell'insieme $\{0,1\}$, e che assumono il valore 0 con probabilità $1/2$ (lo stesso vale per il valore 1). A partire da questa successione di variabili aleatorie si può costruire una successione di variabili aleatorie $\{U_j\,,\ j \in \mathbb{N}\}$ indipendenti ed uniformi in $(0,1)$, come descritto qui di seguito. Infine, posto $F_n(x) = \mu_n\big((-\infty, x]\big)$, la successione cercata è data dalla successione delle v.a. $F_n^{-1}(U_n)$.

Lemma 4.10 (Successioni di v.a. indipendenti uniformi in $(0,1)$: esistenza)**.** *Nello spazio $\Omega = (0,1)$ con la misura di Lebesgue sui boreliani, è possibile avere una successione di v.a. uniformi in $(0,1)$ ed indipendenti.*

Per costruire tale successione si ricordi che scrivendo $\omega \in (0,1)$ in forma diadica $\omega = \sum_1^\infty W_i(\omega)\frac{1}{2^i}$, le v.a. W_i risultano indipendenti e $\mathbb{P}(W_i = 0) = \mathbb{P}(W_i = 1) = \frac{1}{2}$. La successione U_n di v.a. uniformi ed indipendenti si può costruire, a partire dalle v.a. $\{W_i\}$, riordinandole in modo che formino una sequenza a doppio indice $\{\widetilde{W}_{i,n}\}$ così da poter definire

$$U_n(\omega) = \sum_{i=0}^{\infty} \widetilde{W}_{i,n}(\omega)\,\frac{1}{2^i}.$$

Ad esempio si può prendere $\widetilde{W}_{i,n} = W_{2^{i-1}(2n+1)}$, che corrisponde a riordinare la successione $\{W_i\}$ in questo modo:

$$\begin{array}{cccccc}
W_1 & W_3 & W_5 & W_7 & W_9 & \cdots \\
W_2 & W_6 & W_{10} & W_{14} & \cdots & \\
W_4 & W_{12} & W_{20} & W_{28} & \cdots & \\
W_8 & W_{24} & W_{40} & \cdots & & \\
\vdots & \vdots & \vdots & & &
\end{array}$$

ottenendo da $\{W_i\}$ infinite sottosuccessioni (corrispondenti alle colonne di questa matrice) in modo tale che nessuna W_i venga tralasciata né ripetuta.

Osservazione 4.4. ***Ribadiamo che la precedente costruzione è riportata affinché sia chiaro che l'affermazione che esiste una successione di v.a. indipendenti ed uniformi in $(0,1)$, non dipende dal Teorema di Kolmogorov dimostrato precedentemente.*

*Inoltre, a partire da una successione di v.a. $\{U_n\}_{n\in\mathbb{N}}$ uniformi in $(0,1)$ ed indipendenti, e data una successione $\{F_n(x)\}_{n\in\mathbb{N}}$ di funzioni di distribuzione (ossia di funzioni che verificano le proprietà **P0-P3** (Teorema 1.5 di rappresentazione di Skorohod) si può costruire direttamente una successione di v.a. $\{X_n\}_{n\in\mathbb{N}}$ tali che $F_{X_n} = F_n$, per ogni $n \in \mathbb{N}$: basta prendere $X_n := \varphi_n(U_n)$, dove φ_n è la funzione inversa generalizzata della funzione F_n, costruita nel Teorema 1.5 di rappresentazione di Skorohod.*

4.9.2 Osservazione su $\mathcal{R}^I$ †

La σ-algebra $\mathcal{R}^I$ coincide con la famiglia degli insiemi del tipo

$$\mathcal{A} = \{x(\cdot) \mid \{x(t_k)\}_k \in \mathcal{N}\}, \tag{4.29}$$

al variare delle successioni $S = \{t_k\}_k \subseteq I$ ed $\mathcal{N} \in \mathcal{R}^{\mathbb{N}}$, cioè un σ-cilindro.

Anzi, più in generale, dato un processo, cioè una famiglia di variabili aleatorie $(Y_t, t \in I)$ in uno spazio $(\Omega, \mathcal{F}, \mathbb{P})$, e posto $\mathcal{F}_S = \sigma\{Y_s, \text{ per } s \in S\}$, valgono le seguenti affermazioni:

(i) Se $A \in \mathcal{F}_I, \omega \in A$ e $Y_t(\omega) = Y_t(\omega')$ per ogni $t \in I$, allora $\omega' \in A$.

(ii) Se $A \in \mathcal{F}_I$, allora esiste un $S \subseteq$I e numerabile per cui $A \in \mathcal{F}_S$, ovvero $\mathcal{F}_I = \bigcup_S \mathcal{F}_S$, dove l'unione è fatta su tutti i sottoinsiemi S numerabili di I.

Entrambe le due proprietà si basano sul fatto che, posto $\xi : \Omega \to \mathbb{R}^I, \omega \to \xi(\omega)(\cdot)$, dove $\xi(\omega)(t) = Y_t(\omega)$ per ogni $t \in I$, allora

$$\mathcal{F}_I = \{\xi^{-1}(M), \quad M \in \mathcal{R}^I\}.$$

Infatti allora $Y_t(\omega) = Y_t(\omega')$ per ogni $t \in I$ equivale a $\xi(\omega) = \xi(\omega')$ e quindi $\omega \in A = \xi^{-1}(M)$, cioè $\xi(\omega) \in M$ implica $\xi(\omega') \in M$ e quindi $\omega' \in A$. Infatti ogni $\sigma -$ algebra rispetto alla quale ξ è misurabile deve contenere le controimmagini dei cilindri e quindi deve contenere anche $\{\xi^{-1}(M), M \in \mathcal{R}^S\}$ per ogni S numerabile, ed inoltre la famiglia $\bigcup_S \mathcal{F}_S$ è una σ-algebra, come è facile verificare.

L'unico punto in cui è necessaria un po' di attenzione è il caso di unioni numerabili di eventi $E_n \in \bigcup_S \mathcal{F}_S$. Sia S_n tale che $E_n \in \mathcal{F}_{S_n}$. Poiché $\mathcal{F}_{S_n} \subseteq \mathcal{F}_{\cup_m S_m}$ ed $\mathcal{F}_{\cup_m S_m}$ è una $\sigma -$ algebra, ovviamente $\bigcup_n E_n \in \mathcal{F}_{\cup_m S_m} \subseteq \bigcup_S \mathcal{F}_S$.

In particolare $\mathcal{R}^I$=$\bigcup_S \mathcal{R}^S$, dove l'unione è fatta su tutti i sottoinsiemi S numerabili di I.

4.9.3 Problemi con lo spazio canonico

Come visto nella precedente sezione $\mathcal{R}^I$=$\bigcup_S \mathcal{R}^S$, dove l'unione è fatta su tutti i sottoinsiemi S numerabili di I. Consideriamo il caso in cui I è un insieme continuo e per fissare le idee $I = [0, T]$ oppure $[0, \infty)$. Allora non ha senso considerare la probabilità che le traiettorie abbiano particolari proprietà tipo siano crescenti, o continue, in quanto tali insiemi non sono misurabili nello spazio canonico non essendo esprimibili come in (4.29): non è possibile trovare una successione di tempi (ovvero un sottoinsieme numerabile di tempi) e condizioni relative solo a tali istanti per determinare se una funzione è continua oppure no. In altre parole: comunque sia data una successione di tempi esistono funzioni crescenti (o continue) su tale successione di tempi, ma non su tutto I.

Tipicamente, al massimo si può sperare di ottenere un processo $(X'_t, t \in I)$, che abbia le proprietà richieste, e che sia ***stocasticamente equivalente a*** $(X_t, t \in I)$, ovvero una ***versione del processo*** $(X_t, t \in I)$[59] ed ha quindi le stesse distribuzioni finito-dimensionali di $(X_t, t \in I)$

Per proprietà tipo la continuità si può al massimo sperare di procedere nel seguente modo: si dimostra che, se ristretto ad un opportuno insieme numerabile di tempi S, il processo X_t risulta a traiettorie continue, tranne a parte un eventuale insieme di probabilità nulla. Si definisce poi un nuovo processo X'_t definito come limite delle traiettorie ristrette ad S. Si spera poi di dimostrare che il processo così ottenuto abbia le stesse distribuzioni finito-dimensionali del processo X_t.

[59] Si ricordi che , dato un processo $(X_t, t \in I)$. Un processo $(X'_t, t \in I)$ si dice ***stocasticamente equivalente a*** $(X_t, t \in I)$ se

$$\mathbb{P}(X'_t \neq X_t) = 0 \text{ per ogni } t \geq 0.$$

In tale caso si dice anche X'_t è una ***versione*** di X_t.

Problemi di questo tipo sono connessi al problema della separabilità. Gli interessati possono consultare il libro di Billingsley, Probability and measures [3].

A titolo di esempio vediamo come si può procedere con il processo di Poisson, con $I = [0,1]$. Il Teorema di Kolmogorov assicura che esiste una misura di probabilità $\mathbb{P}$ su $(\mathbb{R}^I, \mathcal{R}^I)$, in modo che il processo $X_t(x(\cdot)) = x(t)$ abbia le distribuzioni finito-dimensionali uguali a quelle del processo di Poisson. Ovviamente non possiamo dire nulla sulle traiettorie, ad esempio non possiamo affermare che siano non decrescenti, costanti a tratti e con salti unitari. Proviamo a mostrare come procedere per ottenere una versione X'_t con traiettorie non decrescenti. Consideriamo l'insieme D dei diadici. La condizione che $X_t|_D$ sia a traiettorie non decrescenti e a valori in $\mathbb{N}$ diviene $X_t|_D \in \mathcal{I}$, dove

$$\mathcal{I} = \bigcap_n \bigcap_{0 \le i < 2^n} \{x(\cdot) \text{ tali che } x(\frac{i}{2^n}) \in \mathbb{N} \text{ e } x(\frac{i}{2^n}) \le x(\frac{i+1}{2^n})\}.$$

Inoltre

$$\mathbb{P}(\bigcap_{0 \le i < 2^n} \{x(\cdot) \text{ tali che } x(\frac{i}{2^n}) \in \mathbb{N} \text{ e } x(\frac{i}{2^n}) \le x(\frac{i+1}{2^n})\}) = 1 \quad \text{per ogni } n,$$

e quindi vale anche che $(X_t, t \in D)$ è a traiettorie non decrescenti con probabilità 1.
Si noti inoltre che

$$\lim_{s \downarrow t, s \in D} x(s)$$

esiste per ogni $x(\cdot)$ appartenente a $\mathcal{I}$.

Si definisca ora

$$\begin{cases} X'_t(x(\cdot)) := X_t(x(\cdot)) & \text{se } x(\cdot) \in \mathcal{I} \text{ e } t \in D \\ X'_t(x(\cdot)) := \lim_{s \downarrow t, s \in D} X_s(x(\cdot)) = \lim_{s \downarrow t, s \in D} x(s), & \text{se } x(\cdot) \in \mathcal{I} \text{ e } t \notin D \\ X'_t(x(\cdot)) := 0 & \text{se } x(\cdot) \notin \mathcal{I} \text{ e per ogni } t \in (0,1) \end{cases}$$

Inoltre ovviamente $\{X'_t = X_t\}$ è tutto $\mathbb{R}^I$, per $t \in D$, ed è, se $t \notin D$,

$$\bigcap_{\epsilon \in \mathbb{Q}^+} \bigcup_{\delta \in \mathbb{Q}^+} \bigcap_{\substack{t \le s \le t+\delta \\ s \in D}} \{x(\cdot) \text{ tali che } \mid x(s) - x(t) \mid < \epsilon\}$$

e, se intersecato con $\mathcal{I}$, coincide con

$$\bigcup_{\delta \in \mathbb{Q}^+} \bigcap_{\substack{t \le s \le t+\delta \\ s \in D}} \{x(\cdot) \text{ tali che } \mid x(s) - x(t) \mid = 0\}.$$

Infine [60]

$$\mathbb{P}(\bigcap_{\substack{t \le s \le t+\delta \\ s \in D}} \{x(\cdot) \text{ tali che } \mid x(s) - x(t) \mid = 0\}) \ge \exp\{-\lambda\delta\} \to 1$$

Di conseguenza $\mathbb{P}(X'_t \neq X_t) = 0$ per ogni $t \ge 0$, ovvero il processo X'_t è una versione del processo X_t, ed ha quindi le stesse distribuzioni finito-dimensionali di X_t e gode della proprietà di monotonia, continuità a destra delle traiettorie e di essere a valori in $\mathbb{N}$.

[60] Si osservi che

$$\mathbb{P}(\bigcap_{\substack{t \le s \le t+\delta \\ s \in D}} \{x(\cdot) \text{ tali che } \mid x(s) - x(t) \mid = 0\}) = \lim_{n \to \infty} \mathbb{P}(\bigcap_{\substack{t \le s \le t+\delta \\ s \in D_n}} \{x(\cdot) \text{ tali che } \mid x(s) - x(t) \mid = 0\})$$

e che

$$* * * \mathbb{P}(\bigcap_{\substack{t \le s \le t+\delta \\ s \in D_n}} \{x(\cdot) \text{ tali che } \mid x(s) - x(t) \mid = 0\}) \ge * * * \mathbb{P}(x(t) = x(i/2^n) = x(t+\delta), \text{ per ogni } i \text{ tale che } t \le i/2^n \le t+\delta) = \exp\{-\lambda\delta\}.$$

4.10 Appendice: dimostrazione del criterio di Chensov-Kolmogorov

Per comodità del lettore riportiamo in appendice sia la definizione di funzione hölderiana che l'enunciato del criterio (Proposizione 4.3).

Una funzione $f(x)$ è **hölderiana** di esponente γ se per ogni x esistono un $\delta(x) > 0$ e un $L_\gamma(x)$ tali che, per ogni y per il quale $|y - x| \leq \delta(x)$, si abbia

$$|f(x) - f(y)| \leq L_\gamma(x)|x - y|^\gamma.$$

Nel caso in cui $\delta(x)$ e $L_\gamma(x)$ possano essere presi in modo indipendente da x, per $x \in I$, si dice che f è **uniformemente hölderiana** nell'insieme I.

Proposizione 4.3 (Criterio di Chensov-Kolmogorov) *Sia X_t un processo per cui esistono α, β e C strettamente positivi, per cui*

$$\mathbb{E}[|X_t - X_s|^\beta] \leq C|t - s|^{1+\alpha}.$$

Allora esiste una versione $\widetilde{X}_t$ di X_t, a traiettorie continue. Inoltre le traiettorie sono uniformemente hölderiane di esponente γ, per ogni $\gamma < \frac{\alpha}{\beta}$, in ogni intervallo limitato.

Dimostrazione. Per semplicità consideriamo il caso dell'intervallo $[0, 1]$. Sia come al solito D l'insieme dei diadici e $D_n = \{s = \frac{k}{2^n}, \text{ per } 0 \leq k \leq 2^n\}$. Si definisca Γ_n

$$\Gamma_n = \{\omega \text{ t.c. } \max_{k \leq 2^n - 1} \left|X_{k/2^n} - X_{(k+1)/2^n}\right| > 2^{-n\gamma}\}$$

Se mostriamo che

$$\sum_{n=1}^{\infty} \mathbb{P}(\Gamma_n) < +\infty,$$

allora, per il Lemma di Borel-Cantelli,

$$\mathbb{P}(\cap_{m \geq 1} \cup_{n \geq m} \Gamma_n) = 0$$

ovvero

$$\mathbb{P}(\cup_{m \geq 1} \cap_{n \geq m} \cap \Gamma_n^c) = 1.$$

Ciò significa che esiste un evento $\mathcal{N}$, con $\mathbb{P}(\mathcal{N}) = 0$, e che per ogni $\omega \in \mathcal{N}^c$ esiste un $m = m(\omega)$ tale che per ogni $n \geq m = m(\omega)$

$$\left|X_{k/2^n} - X_{(k+1)/2^n}\right| \leq 2^{-n\gamma}.$$

Mostreremo ora che per $\omega \in \mathcal{N}^c$ le traiettorie sono uniformemente hölderiane su D con

$$\delta = \delta(\omega) = 1/2^{m(\omega)} \text{ ed } L_\gamma = \Big(\frac{2}{2^\gamma - 1} + 1\Big)2^\gamma,$$

e quindi sono hölderiane su D per quasi ogni ω.

Sia quindi $\omega \in \mathcal{N}^c$.

Cominciamo con l'osservare che se

$$r \in [j/2^m, (j+1)/2^m), \text{ con } m \geq m(\omega), \text{ ed} r \in D,$$

allora

$$r = j/2^m + \sum_{h=m+1}^{m+i} \alpha_h \frac{1}{2^h},$$

per un $i \geq 0$, con $\alpha_h \in \{0, 1\}$. Posto

$$r_0 = j/2^m \text{ ed } r_s = j/2^m + \sum_{h=m+1}^{m+s} \alpha_h \frac{1}{2^h},$$

si ha che $r = r_i$ e che

$$\left|X_{j/2^m} - X_r\right| \le \sum_{s=1}^{i} \left|X_{r_s} - X_{r_{s-1}}\right| \le \sum_{h=m+1}^{m+i} 2^{-h\gamma} \le 2^{-m\gamma} \sum_{l=1}^{i} 2^{-l\gamma} \le 2^{-m\gamma} \frac{1}{2^\gamma - 1}.$$

Se invece $|r-u| \le 1/2^m$ con $r, u \in D$, allora sono possibili due casi:

(i) **esiste** j **per cui** $r, u \in [j/2^m, (j+1)/2^m)$, e allora

$$|X_u - X_r| \le \left|X_{j/2^m} - X_u\right| + \left|X_{j/2^m} - X_r\right| \le \frac{2}{2^\gamma - 1} 2^{-m\gamma}$$

(ii) **esiste** j **per cui** $r \in [j/2^m, (j+1)/2^m)$ **ed** $u \in [(j+1)/2^m, (j+2)/2^m)$, e allora

$$|X_u - X_r| \le \left|X_{(j+1)/2^m} - X_u\right| + \left|X_{j/2^m} - X_{(j+1)/2^m}\right| + \left|X_{j/2^m} - X_r\right| \le \left(\frac{2}{2^\gamma - 1} + 1\right) 2^{-m\gamma}$$

Quindi, per quasi ogni ω, si ha definitivamente, per ogni $r, u \in D$, con $|r-u| \le 1/2^m$,

$$|X_u - X_r| \le M_\gamma 2^{-m\gamma},$$

con $M_\gamma = \frac{2}{2^\gamma - 1} + 1$. In particolare se $|r-u| \le 1/2^m(\omega)$,ed $m \ge m(\omega)$ è tale che $1/2^{m+1} \le |r-u| \le 1/2^m$, allora

$$|X_u - X_r| \le L_\gamma 2^{-(m+1)\gamma} \le L_\gamma |r-u|^\gamma.$$

Avremo quindi, come preannunciato, che le traiettorie sono uniformemente hölderiane su D con $\delta = \delta(\omega) = 1/2^{m(\omega)}$ ed $L_\gamma = \left(\frac{2}{2^\gamma - 1} + 1\right) 2^\gamma$, per quasi ogni ω. [1]

Si definisce, sempre per $\omega \in \mathcal{N}^c$, $\widetilde{X}_t(\omega)$ come il limite di $X_{r_n}(\omega)$ per una successione r_n di diadici convergente a t, e per $\omega \in \mathcal{N}$ si può definire, ad esempio, $\widetilde{X}_t \equiv 0$. È facile vedere che la proprietà di hölderianità si conserva. Il fatto che si tratta di una versione di X_t, deriva dal fatto che, per l'ipotesi fatta, X_t è continuo in L^β e quindi è continuo in probabilità e quindi contemporaneamente si ha

$$X_{r_n} \stackrel{Pr}{\to} X_t \text{ e } X_{r_n} \stackrel{q.c.}{\to} \widetilde{X}_t,$$

da cui subito $\mathbb{P}(X_t = \widetilde{X}_t) = 1$.

A questo punto rimane solo da controllare che $\sum_{n=1}^{\infty} \mathbb{P}(\Gamma_n) < +\infty$, e infatti:

$$\sum_{n=1}^{\infty} \mathbb{P}(\Gamma_n) \le \sum_{n=1}^{\infty} \sum_{k=0}^{2^n-1} \mathbb{P}(\left|X_{k/2^n} - X_{(k+1)/2^n}\right| > 2^{-n\gamma}) \le \sum_{n=1}^{\infty} \sum_{k=0}^{2^n-1} \mathbb{E}[\left|X_{k/2^n} - X_{(k+1)/2^n}\right|^\beta]/2^{-n\gamma\beta} \le$$

$$\le \sum_{n=1}^{\infty} \sum_{k=0}^{2^n-1} (1/2^n)^{1+\alpha} 2^{n\gamma\beta}) = \sum_{n=1}^{\infty} 2^n (1/2^n)^{1+\alpha} 2^{n\gamma\beta}) = \sum_{n=1}^{\infty} 2^{-n(\alpha-\gamma\beta)} < +\infty$$

purché $\alpha - \gamma\beta > 0$, e ciò è vero per come abbiamo preso γ. (Si noti che è fondamentale nella prova che $1 + \alpha > 1$, in quanto altrimenti non ci sarebbe speranza di ottenere nessun tipo di convergenza ovvero la maggiorazione di $\mathbb{P}(\Gamma_n)$ non avrebbe nessun significato in quanto risulterebbe un numero maggiore di 1)

□

[1] Si noti che questa parte della dimostrazione è puramente deterministica: nel senso che se una funzione $x(\cdot)$ definita sui diadici di $[0,1]$, gode della proprietà che esiste un m tale che per ogni $n \ge m$ e per $k < 2^n$ si ha $|x(k/2^n) - x((k+1)/2^n)| \le 2^{-n\gamma}$, allora $x(\cdot)$ è hölderiana sui diadici di $[0,1]$

Capitolo 5

Proprietà del moto browniano

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

5.1 Trasformazioni del moto browniano

Nei seguenti casi, dato un moto browniano W_t, si costruisce un altro processo che risulta ancora un moto browniano:

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

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

3) Il processo definito da $\widehat{W}_t := tW_{1/t}$, per $t > 0$, e $\widehat{W}_0 := 0$, è ancora un moto browniano, rispetto alla sua filtrazione naturale.

Si supponga ora di prendere una versione continua di W_t.

4) Per ogni ****τ, tempo d'arresto finito con probabilità 1, il processo $\widetilde{W}_t^{(\tau)} := W_{\tau+t} - W_\tau$ è ancora un moto Browniano rispetto alla filtrazione $\mathcal{F}_{\tau+t}^W$.

Le proprietà 1),2) e 3) sono di immediata verifica[1].

La proprietà 4), invece, non è immediata, e corrisponde a quella proprietà che, nei processi di Markov, viene detta Proprietà di Markov Forte.

5.2 Proprietà di Markov forte per il processo di Wiener

In questa sezione si suppone di prendere una versione continua di W_t.

Proposizione 5.1. *Sia τ un tempo d'arresto finito con probabilità 1. Allora il processo $Y_t := W_{t+\tau} - W_\tau$ è un processo di Wiener standard ed è indipendente da $\mathcal{F}_\tau$.*

[1] ****Le proprietà 1) e 2) sono di immediata verifica anche se si parte da una versione continua del processo di Wiener e si aggiunge la proprietà che $\widetilde{W}_t$ e $W_t^{(c)}$ anche sono a traiettorie continue. Anche la proprietà 3) è semplice da verificare: se si vuole dimostrare che il processo $\widehat{W}_t$ è un processo di Wiener a traiettorie continue, tutto è semplice da verificare, tranne la continuità in 0. Per quest'ultima proprietà ci limitiamo qui ad osservare che $\widehat{W}_t$ converge in probabilità a 0 per t che tende a 0. Per la convergenza quasi certa, si veda invece la Sezione 5.5

Come conseguenza, per ogni funzionale Φ, misurabile nello spazio delle traiettorie continue,

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

In questo senso la proprietà 4), ovvero la Proposizione 5.1, è una generalizzazione della proprietà di Markov, che invece di valere solo per tempi deterministici t vale anche per tempi d'arresto τ.

Più precisamente

Definizione 5.1 (Proprietà di Markov)**.** *La* **proprietà di Markov** *(***rispetto ad una filtrazione** $\mathcal{F}_t \supseteq \mathcal{F}_t^X$*) per un generico processo* X_t *è la seguente:*

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

Si osservi che, nel caso dei processi di Markov regolari

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

dove $P(u, v, x, \cdot)$ sono le probabilità di transizione regolari.

Definizione 5.2 (Proprietà di Markov forte)**.** *La* **proprietà di Markov forte**[2] **rispetto ad una filtrazione** $\mathcal{F}_t \supseteq \mathcal{F}_t^X$*) è invece*

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

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

Questa proprietà di misurabilità è sempre verificata per processi X_t *cadlag*, in realtà potrebbe bastare un poco meno e precisamente la progressiva misurabilità (si veda a questo proposito la Sezione 3.11).

Dimostrazione. (della Proposizione 5.1, ovvero della proprietà di Markov forte per il processo di Wiener standard)

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

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

infatti, prendendo $C = \Omega$, si ottiene che il vettore $(Y_{t_1}, \cdots, Y_{t_k})$ ha la stessa distribuzione del processo di Wiener, e di conseguenza si ottiene immediatamente l'indipendenza tra $\mathcal{F}_\tau$ e $\sigma\{Y_{t_1}, \cdots, Y_{t_k}\}$. Per l'arbitrarietà di $0 \leq t_1 \leq \cdots \leq t_k$, ciò equivale all'indipendenza del processo Y_t da $\mathcal{F}_\tau$.

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

$$\begin{aligned}
\mathbb{P}(Y_{t_1} \in A_1, \cdots, Y_{t_k} \in A_k, C) &= \sum_{h=1}^{\infty} \mathbb{P}(Y_{t_1} \in A_1, \cdots, Y_{t_k} \in A_k, C, \tau = s_h) \\
&= \sum_{h=1}^{\infty} \mathbb{P}(W_{t_1+s_h} - W_{s_h} \in A_1, \cdots, W_{t_k+s_h} - W_{s_h} \in A_k, C, \tau = s_h) \\
&\qquad (C \cap \{\tau = s_h\} \in \mathcal{F}_{s_h}, \text{ poiché } C \in \mathcal{F}_\tau \text{ e } W_t \text{ ha incrementi indipendenti}) \\
&= \sum_{h=1}^{\infty} \mathbb{P}(W_{t_1+s_h} - W_{s_h} \in A_1, \cdots, W_{t_k+s_h} - W_{s_h} \in A_k)\mathbb{P}(C, \tau = s_h) \\
&\qquad (W \text{ ha incrementi omogenei}) \\
&= \sum_{h=1}^{\infty} \mathbb{P}(W_{t_1} \in A_1, \cdots, W_{t_k} \in A_k)\mathbb{P}(C, \tau = s_h) = \mathbb{P}(W_{t_1} \in A_1, \cdots, W_{t_k} \in A_k)\mathbb{P}(C)
\end{aligned}$$

[2] Si faccia attenzione che la proprietà di Markov forte non è sempre verificata per un generico processo di Markov.

II CASO: τ generico.

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

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

in quanto le funzioni continue determinano univocamente la legge di una variabile aleatoria. Per il Caso I, presa $\{\tau_m\}$ una successione[3] di tempi d'arresto, tale che $\tau_m \geq \tau$ e $\tau_m \to \tau$

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

Si osservi che la precedente relazione vale addirittura per ogni $C \in \mathcal{F}_{\tau_m}$, e quindi per ogni $C \in \mathcal{F}_\tau$: infatti $\tau_m \geq \tau$ implica $\mathcal{F}_{\tau_m} \supseteq \mathcal{F}_\tau$.

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

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

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

□

Osservazione 5.1. *È evidente che questa dimostrazione (e quindi la proprietà di Markov forte) rimane valida se al posto del moto browniano si mette un qualsiasi processo ad incrementi indipendenti ed omogenei, con traiettorie* cadlag.

Terminiamo mostrando come la proprietà di Markov forte implica immediatamente l'affermazione 4) all'inizio del capitolo, ossia che $Y_t = W_{\tau+t} - W_\tau$ è un processo di Wiener rispetta alla filtrazione $\mathcal{F}_{\tau+t}$. Infatti manca solo di dimostrare l'indipendenza di $Y_t - Y_s$ da $\mathcal{F}_{\tau+s}$, per $s \leq t$: ciò segue immediatamente osservando che $Y_t - Y_s = W_{\tau+t} - W_\tau - (W_{\tau+s} - W_\tau) = W_{\tau+t} - W_{\tau+s}$ e applicando la precedente Proposizione 5.1 al tempo d'arresto $\tau + s$.

5.3 Principio di riflessione

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

$$\tau_a = \inf\{t > 0 \text{ t.c. } W_t \geq a\},$$

cioè il tempo di prima uscita da $(-\infty, a)$, che è un tempo d'arresto se, come al solito prendiamo la versione continua di W_t. Allora

$$\mathbb{P}(\tau_a \leq t) = 2\mathbb{P}(W_t \geq a). \tag{5.1}$$

Dalla precedente eguaglianza si può ricavare che allora

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

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

[3] Per la proprietà **2)** dei tempi di arresto (Sezione 3.4), per ogni tempo d'arresto esiste sempre una successione con tali proprietà.

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

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

e quindi la sua densità è

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

e il suo valore atteso è infinito:

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

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

Inoltre questa proprietà permette di calcolare ***la distribuzione del massimo di un moto browniano*** sull'intervallo di tempo $[0, t]$, (per $a \geq 0$):

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

Dimostrazione intuitiva del principio di riflessione (5.1)

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

in quanto ovviamente

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

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

$$\mathbb{P}(W_t \geq a \mid \tau_a \leq t) = \mathbb{P}(W_t \geq W_{\tau_a} \mid \tau_a \leq t) = \mathbb{P}(W_t - W_{\tau_a} \geq 0 \mid \tau_a \leq t) =$$

$$= \mathbb{P}(Y_{t-\tau_a} \geq 0 \mid \tau_a \leq t) = \mathbb{P}(Y_{t-\tau_a} \geq 0)$$

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

Prima di dare la dimostrazione formale occorre dimostrare che si può usare la proprietà di Markov forte, ossia occorre dimostrare che τ_a è un tempo d'arresto finito. Ciò può essere dimostrato utilizzando il fatto che[4]

$$\mathbb{P}\big(\sup_{t \geq 0} W(t) = +\infty, \quad \inf_{t \geq 0} W(t) = -\infty\big) = 1.$$

Inoltre va ricordato che la trasformata di Laplace individua univocamente una funzione nel seguente senso:

[4]Questo risultato può essere espresso dicendo che il processo di Wiener è ricorrente (cioè ogni punto viene raggiunto con probabilità 1). Il fatto che il valore atteso di τ_a è infinito, come abbiamo già notato, si può essere espresso dicendo che il processo di Wiener è ricorrente nullo. Per una dimostrazione, ****si veda la Proposizione 6.11, della Sezione 5.5, o equivalentemente gli appunti del prof. Bertini, Proposizione 1.7.

Lemma 5.2 (Trasformata di Laplace, Feller [6], Billingsley [3]). *Se f_1 ed f_2 sono due funzioni misurabili per cui*

$$\int_0^\infty |f_i(t)|\, e^{-s_0\, t}\, dt < \infty$$

e per cui

$$\int_0^\infty f_1(t)\, e^{-s\, t}\, dt = \int_0^\infty f_2(t)\, e^{-s\, t}\, dt, \qquad \textit{per ogni } s \geq s_0, \tag{5.2}$$

allora $f_1 = f_2$ quasi ovunque in $[0, \infty)$.

(Per la dimostrazione del Lemma 5.2, si veda la sottosezione 5.3.1.)

Dimostrazione formale del principio di riflessione (5.1) Faremo vedere che le trasformate di Laplace rispetto a t, di $\mathbb{P}(W_t \geq a)$ e di $\frac{1}{2}\mathbb{P}(\tau_a \leq t)$ sono uguali, utilizzando il fatto che $W_{\tau_a} = a$, per la continuità delle traiettorie di W_t, e che $W_{\tau+s} - W_\tau$ è indipendente da $\mathcal{F}_\tau$, e quindi da τ: qualunque sia $\alpha \geq 0$,

$$\begin{aligned}
&\int_0^\infty e^{-\alpha t}\mathbb{P}(W_t \geq a)dt \qquad \text{(è necessario usare la congiunta misurabilità in } (t,\omega) \text{ di } W_t\text{, per scambiare gli integrali)}\\
&= {**}\mathbb{E}[\int_0^\infty e^{-\alpha t}\mathbb{I}_{\{W_t \geq a\}}dt]{**} = \mathbb{E}[\int_0^\infty e^{-\alpha t}\mathbf{1}_{[a,+\infty)}(W_t)dt] \qquad \text{(tenendo conto che } W_t < a \text{ per } t < \tau_a)\\
&= \mathbb{E}[\int_{\tau_a}^\infty e^{-\alpha t}\mathbf{1}_{[a,+\infty)}(W_t)dt] = \mathbb{E}[\int_{\tau_a}^\infty e^{-\alpha\tau_a}e^{-\alpha(t-\tau_a)}\mathbf{1}_{[0,+\infty)}(W_t - W_{\tau_a})dt]\\
&\qquad (W_t - W_{\tau_a} = Y_{t-\tau_a} \text{ e cambio di variabile } s = t - \tau_a\ {**})\\
&= \mathbb{E}[e^{-\alpha\tau_a}\int_0^\infty e^{-\alpha s}\mathbf{1}_{[0,+\infty)}(Y_s)ds]\\
&\qquad (\{Y_t, t \geq 0\} \text{ e } \tau_a \text{ sono indipendenti})\\
&= \mathbb{E}[e^{-\alpha\tau_a}]\mathbb{E}[\int_0^\infty e^{-\alpha s}\mathbf{1}_{[0,+\infty)}(Y_s)ds] = \mathbb{E}[e^{-\alpha\tau_a}]\int_0^\infty e^{-\alpha s}\mathbb{P}(Y_s \geq 0)ds]\\
&= \frac{1}{2\alpha}\mathbb{E}[e^{-\alpha\tau_a}].
\end{aligned}$$

In modo analogo

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

□

Può essere interessante, alla luce del precedente risultato, trovare il valore della trasformata di Laplace della funzione $t \mapsto \mathbb{P}(\tau_a \leq t)$ o, equivalentemente, il valore di $\mathbb{E}[e^{-\alpha\tau_a}]$, la trasformata di Laplace della distribuzione di τ_a.

Esempio 5.1 (trasformata di Laplace di τ_a). *Sia $a > 0$ e $\tau_a = \inf\{t > 0 : W_t \geq a\}$. Allora, per ogni $\alpha \geq 0$*

$$\mathbb{E}[\exp\{-\alpha\, \tau_a\}] = \exp\{-a\sqrt{2\,\alpha}\}.$$

Come già visto in Esempio 4.7, il processo $Z_t^\theta = \exp\{\theta\, W_t - \frac{1}{2}\,\theta^2\, t\}$ è una martingala, e quindi $Z_{t\wedge\tau_a}^\theta$ è una martingala limitata con

$$\mathbb{E}[Z_{t\wedge\tau_a}^\theta] = \mathbb{E}[Z_0^\theta] = 1$$

Inoltre $Z_{t\wedge\tau_a}^\theta$ converge, per t che tende all'infinito, a $Z_{\tau_a}^\theta$. Inoltre $0 \leq Z_{t\wedge\tau_a}^\theta \leq \exp\{\theta\, a\}$, e dal teorema della convergenza dominata si deduce che,

$$\mathbb{E}[Z_{\tau_a}^\theta] = \mathbb{E}[\exp\{\theta\, W_{\tau_a} - \tfrac{1}{2}\,\theta^2\,\tau_a\}] = 1$$

Poiché $W_{\tau_a} = a$ si ottiene che

$$\mathbb{E}[\exp\{\theta\, a - \tfrac{1}{2}\,\theta^2\,\tau_a\}] = 1 \quad \Leftrightarrow \quad \mathbb{E}[\exp\{-\tfrac{1}{2}\,\theta^2\,\tau_a\}] = \exp\{-\theta\, a\}.$$

basta poi porre $\frac{1}{2}\,\theta^2 = \alpha$.

Osservazione 5.2. *Si noti che il calcolo effettuato nel precedente Esercizio 5.1, poteva anche essere effettuato utilizzando l'espressione della densità di τ_a, ossia calcolando direttamente*

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

Invece noi abbiamo ottenuto che

$$\int_0^\infty e^{-\alpha\, t} \sqrt{\frac{1}{2\pi}}\, \frac{|a|}{t^{3/2}} \exp\{-a^2/2t\}\, dt = \exp\{-|a|\,\sqrt{2\,\alpha}\}, \tag{5.3}$$

utilizzando il fatto che il processo $Z^\theta_{t\wedge\tau_a}$ è una martingala[5].

Per terminare osserviamo che in alcuni testi viene dato come principio di riflessione la seguente proprietà (la cui dimostrazione è simile alla dimostrazione della proprietà di Markov forte):

Principio di riflessione: seconda formulazione
Per ogni tempo d'arresto τ finito, il processo "riflesso"

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

è ancora un moto browniano.

5.3.1 Trasformata di Laplace

In questa sezione diamo uno schema di dimostrazione del Lemma 5.2

Risultato preliminare 1: Se Y_λ è una variabile aleatoria di Poisson di parametro λ allora la variabile aleatoria Y_λ/λ converge in probabilità a 1 per λ che tende ad infinito (ciò è una semplice conseguenza della disuguaglianza di Chebyshev). Di conseguenza, per la funzione di ripartizione G_λ di Y_λ/λ vale

$$G_\lambda(t) = \sum_{k=0}^{\lfloor \lambda t \rfloor} \frac{\lambda^k}{k!}\, e^{-\lambda} \underset{\lambda\to\infty}{\longrightarrow} \begin{cases} 0, & t < 1 \\ 1, & t > 1 \end{cases}$$

(in realtà, con l'aiuto del teorema centrale del limite, si potrebbe dimostrare anche che $G_\lambda(1)$ tende a $1/2$ per $\lambda \to \infty$)

Risultato preliminare 2: Sia μ una misura di probabilità su $[0, \infty)$ e sia nota

$$M(s) = \int_0^\infty e^{-s\, t}\, \mu(dt) \qquad s \ge s_0$$

Allora, M ammette tutte le derivate $M^{(k)}(s)$ di ordine k, per $s > 0$, e vale la seguente formula di inversione

$$F_\mu(x) = \mu[0, x] = \lim_{s\to\infty} \sum_{k=0}^{\lfloor s\, x \rfloor} (-1)^k\, \frac{s^k}{k!}\, M^{(k)}(s),$$

[5] Si osservi infine che, derivando entrambi i membri della precedente uguaglianza (5.3) rispetto ad α si ottiene **

$$-\int_0^\infty e^{-\alpha\, t} \sqrt{\frac{1}{2\pi}}\, \frac{|a|}{t^{1/2}} \exp\{-a^2/2t\}\, dt = -\frac{1}{\sqrt{2\alpha}}\, |a| \exp\{-|a|\,\sqrt{2\,\alpha}\}.$$

Questo fatto verrà utilizzato per il calcolo del risolvente del nucleo del calore (si veda la sezione 5.6

per ogni x di continuità per μ (ossia per ogni x tale che $\mu(\{x\}) = 0$.
La dimostrazione dipende dal fatto che

$$M^{(k)}(s) = (-1)^k \int_0^\infty t^k\, e^{-s\,t}\, \mu(dt)$$

da cui

$$\begin{aligned}
\sum_{k=0}^{\lfloor s\,x\rfloor} (-1)^k\, \frac{s^k}{k!}\, M^{(k)}(s) &= \sum_{k=0}^{\lfloor s\,x\rfloor} (-1)^k\, \frac{s^k}{k!}\, (-1)^k \int_0^\infty t^k\, e^{-s\,t}\, \mu(dt) \\
&= \int_0^\infty \sum_{k=0}^{\lfloor s\,x\rfloor} \frac{(st)^k}{k!}\, e^{-s\,t}\, \mu(dt) \\
&= \int_0^\infty \sum_{k=0}^{\lfloor st\,\frac{x}{t}\rfloor} \frac{(st)^k}{k!}\, e^{-s\,t}\, \mu(dt) \\
&= \int_0^\infty G_{st}(\tfrac{x}{t})\, \mu(dt) \longrightarrow \mu\big((0,x)\big) + \frac{1}{2}\, \mu(\{x\}).
\end{aligned}$$

L'ultima uguaglianza dipende dal teorema della convergenza dominata ($0 \le G_\lambda(x) \le 1$) e dal fatto che

$$G_{st}(\tfrac{x}{t}) \underset{s\to\infty}{\longrightarrow} \begin{cases} 0, & x < t \\ \frac{1}{2}, & x = t \\ 1, & x > t \end{cases}$$

A questo punto la dimostrazione del Lemma 5.2 è basata, nel caso in cui f_i siano non negative, sul fatto che, posto

$$C_i := \int_0^\infty f_i(t)\, e^{-s_0\,t}\, dt,$$

per la (5.2) $C_1 = C_2 = C$ e le due misure $\mu_i(dt) := \frac{1}{C_i}\, f_i(t)\, e^{-s_0\,t}\, dt$ hanno la stessa trasformata di Laplace e quindi coincidono. Quindi, per ogni $x \ge 0$, si ha $\int_0^x f_1(t)\, e^{-s_0\,t}\, dt = \int_0^x f_2(t)\, e^{-s_0\,t}\, dt$, e necessariamente le due densità $f_i(t)\, e^{-s_0\,t}$ coincidono quasi ovunque su $[0,\infty)$.

Il caso generale si tratta considerando che l'ipotesi equivale a considerare che le funzioni $f_1^+ + f_2^-$ e $f_2^+ + f_1^-$, sono funzioni non negative con la stessa trasformata di Laplace.

5.4 Tempi di uscita da una striscia

Anche nel caso del moto Browniano si possono ottenere delle applicazioni simili a quelle della rovina del giocatore. Più in generale si può considerare il processo di Wiener con coefficiente di drift μ e coefficiente di diffusione σ^2, cioè

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

e il tempo τ di prima uscita da una striscia

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

e cercare di calcolare la probabilità[6] $p \equiv p(x,a,b) := \mathbb{P}_x(X_\tau = -a)$, e la trasformata di Laplace $\mathbb{E}_x[\exp\{-\alpha\tau\}]$, per $x \in (-a,b)$.

[6]Stiamo considerando W_t come il processo canonico $W_t(x(\cdot)) = x(t)$, nello spazio delle traiettorie continue, e come misura su tale spazio la misura ottenuta dalla misura di Wiener (che è concentrata sulle traiettorie con $x(0) = 0$), attraverso la trasformazione che porta una funzione $x(\cdot)$ in $x(\cdot) + x$, di modo che la misura $\mathbb{P}_x$ è concentrata sulle traiettorie per le quali $x(0) = x$. Alternativamente si può pensare che $\mathbb{P}_x$ è la misura indotta sullo spazio delle traiettorie continue dal processo $B_t + x$, dove questa volta però B_t è un qualunque processo di Wiener con $B_0 = 0$.

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

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

è una martingala con

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

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

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

e quindi anche

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

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

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

da cui subito, tenendo conto del valore di θ,

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

e quindi

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

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

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

Come prima si ottiene che

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

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

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

si ottiene immediatamente il seguente sistema lineare nelle incognite

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

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

di cui si ricava la soluzione, ovvero

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

avendo posto

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

Basta poi considerare la somma

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

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

Esercizio 5.1. *Nel caso $x = 0$, $\mu > 0$ mandare a ad infinito in modo da ottenere la trasformata di Laplace di $\tau := \tau(-\infty, b)$*

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

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

Si osservi che ponendo $\sigma^2 = 1$ e mandando μ a zero si ottiene che

$$\lim_{\mu\to 0} \mathbb{E}_0[\exp\{-\alpha\tau(-\infty,b)\}] = \exp\left\{-\sqrt{2\,\alpha}\, b\right\}.$$

Si confronti il risultato con l'Esempio 5.1.

Esercizio 5.2. *Si consideri $\tau(-a,a)$ nel caso del moto browniano o Wiener standard, cioè con μ=0 e σ=1, e con $x = 0$ ed $a = b$.*

soluzione:

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

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

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

Esercizio 5.3 (Revuz-Yor [11], Prop-II(3.7))**.** *Ritrovare il risultato del precedente Esercizio 5.2, utilizzando la stessa tecnica dell'Esempio 5.1 e la martingala*

$$M_t^\theta := \left(Z_t^\theta + Z_t^{-\theta}\right)/2 = \cosh(\theta\, W_t)\, \exp\{-\tfrac{\theta^2}{2}\, t\}.$$

*(**suggerimento:** la martingala $M^\theta_{t\wedge\tau(-a,a)}$ è limitata da $\cosh(\theta\, a)$)*

5.5 Proprietà delle traiettorie di un processo di Wiener ****

Vediamo ora un risultato che permette di affermare che il processo di Wiener (a traiettorie continue) è ricorrente, ovvero che $\mathbb{P}(\tau_a < \infty) = 1$, ogni punto $a \in \mathbb{R}$. Per la proprietà di Markov forte questo fatto implica poi che il processo di Wiener passa infinite volte[7] in ogni punto $a \in \mathbb{R}$.

Proposizione 5.3. *Il processo di Wiener visita ogni punto $a \in \mathbb{R}$ almeno una volta, ossia*

$$\mathbb{P}(\sup_{t\geq 0} W_t = +\infty,\ \inf_{t\geq 0} W_t = -\infty) = 1.$$

Dimostrazione. Basta ovviamente dimostrare separatamente che $\mathbb{P}(\sup_{t\geq 0} W_t = +\infty) = 1$ e che $\mathbb{P}(\inf_{t\geq 0} W_t = -\infty) = 1$. Inoltre, grazie al fatto che $(-W_t)_{t\geq 0}$ è ancora un processo di Wiener (per la proprietà di invarianza di scala della sua legge), basta dimostrare solamente la prima uguaglianza. La variabile aleatoria[8] generalizzata, a valori in $[0, +\infty)\cup\{\infty\}$ definita come

$$W^* := \sup_{t\geq 0} W_t,$$

può assumere (con probabilità 1) solo i valori 0 e $+\infty$, ossia $\mathbb{P}(W^* \in \{0, +\infty\}) = 1$. Ciò discende di nuovo dalla proprietà di invarianza della sua legge per cambiamenti di scala, in quanto, per ogni $c > 0$, W^* e cW^* hanno la stessa distribuzione[9]. Non rimane quindi che dimostrare che $\mathbb{P}(W^* = 0) = 0$. A questo scopo notiamo che

$$\begin{aligned}\mathbb{P}(W^* = 0) &\leq \mathbb{P}(W_1 \leq 0,\ W_s \leq 0,\ \forall\, s \geq 1)\\ &= \mathbb{P}(W_1 \leq 0,\ W_s \leq 0,\ \forall\, s \geq 1,\ \overline{W}^* = 0) + \mathbb{P}(W_1 \leq 0,\ W_s \leq 0,\ \forall\, s \geq 1\ \overline{W}^* = +\infty)\end{aligned}$$

dove si è posto $\overline{W}^* = \sup_{t\geq 0} W_{t+1} - W_1$, e si è tenuto conto che $\overline{W}^*$ ha la stessa legge di W^*, e quindi assume solo i valori 0 e $+\infty$. Inoltre $\overline{W}^*$ è indipendente da W_1, di conseguenza

$$\begin{aligned}\mathbb{P}(W^* = 0) &\leq \mathbb{P}(W_1 \leq 0,\ W_s \leq 0,\ \forall\, s \geq 1,\ \sup_{t\geq 0} W_{t+1} - W_1 = 0) + \mathbb{P}(\emptyset)\\ &\leq \mathbb{P}(W_1 \leq 0,\ \sup_{t\geq 0} W_{t+1} - W_1 = 0) = \mathbb{P}(W_1 \leq 0)\,\mathbb{P}(\sup_{t\geq 0} W_{t+1} - W_1 = 0) = \frac{1}{2}\,\mathbb{P}(\overline{W}^* = 0)\end{aligned}$$

Poiché $\mathbb{P}(\overline{W}^* = 0) = \mathbb{P}(W^* = 0)$, la precedente relazione diviene

$$\mathbb{P}(W^* = 0) = \frac{1}{2}\,\mathbb{P}(W^* = 0)$$

che implica appunto che $\mathbb{P}(W^* = 0) = 0$. □

In questa sezione vogliamo mostrare anche una semplice proprietà che implica la proprietà di continuità in $t = 0$ del processo $\widehat{W}_t$ definito nella proprietà 3), all'inizio di questo capitolo, come $\widehat{W}_t := t\,W_{1/t}$, per $t > 0$ e $\widehat{W}_0 := 0$. Per dimostrare che

$$\mathbb{P}(\lim_{t\to 0^+} \widehat{W}_t = \widehat{W}_0 = 0) = 1,$$

basta ovviamente dimostrare la seguente proprietà.

Proposizione 5.4. *Sia $(W_s)_{s\geq 0}$ un processo di Wiener, allora*

$$\mathbb{P}(\lim_{s\to+\infty} \frac{W_s}{s} = 0) = 1.$$

[7] Per ottenere che ogni stato è visitato un numero infinito di volte, si può, ad esempio, sfruttare la proprietà di Markov forte: ogni volta che il processo di Wiener passa per un punto a, poi il processo $W_{t+\tau_a} - W_\tau$ visiterà di nuovo ogni punto almeno una volta, e quindi, ad esempio, anche lo zero. A questo punto si ricomincia daccapo e quindi il processo visiterà a almeno un'altra volta con probabilità 1, e così via.

[8] Si può affermare che $W^* := \sup_{t\geq 0} W_t$ è una variabile aleatoria, in quanto, essendo a traiettorie continue,

$$W^* := \sup_{t\geq 0} W_t = \sup_{t\geq 0,\, t\in\mathbb{Q}} W_t.$$

[9] Infatti per ogni $c > 0$ si ha che

$$cW^* := c\sup_{t\geq 0} W_t = \sup_{s\geq 0} c\,W_{s/c^2} = \sup_{s\geq 0} c\,W_{s/c^2} = \big(W^{(c)}\big)^*.$$

Dimostrazione. L'idea della dimostrazione è la seguente:

$$\frac{W_s}{s} = \frac{\lfloor s \rfloor}{s}\,\frac{W_s}{\lfloor s \rfloor} = \frac{\lfloor s \rfloor}{s}\Big(\frac{W_s - W_{\lfloor s \rfloor}}{\lfloor s \rfloor} + \frac{W_{\lfloor s \rfloor}}{\lfloor s \rfloor}\Big).$$

Ovviamente $\lim_{s\to\infty} \frac{\lfloor s \rfloor}{s} = 1$ e inoltre

$$\frac{W_{\lfloor s \rfloor}}{\lfloor s \rfloor} = \frac{1}{\lfloor s \rfloor}\sum_{k=1}^{\lfloor s \rfloor}\big[W_k - W_{k-1}\big]$$

quindi per ottenere la convergenza quasi certa a 0 basta applicare la legge forte dei grandi numeri.

Per quanto riguarda il termine

$$\frac{W_s - W_{\lfloor s \rfloor}}{\lfloor s \rfloor}$$

si può osservare che

$$|W_s - W_{\lfloor s \rfloor}| \le \sup_{u\in[\lfloor s \rfloor,\lfloor s \rfloor+1]}\big|W_u - W_{\lfloor s \rfloor}\big| \le \sup_{u\in[\lfloor s \rfloor,\lfloor s \rfloor+1]}\big(W_u - W_{\lfloor s \rfloor}\big) + \sup_{u\in[\lfloor s \rfloor,\lfloor s \rfloor+1]}\big((-W_u) - (-W_{\lfloor s \rfloor})\big)$$

Di conseguenza, posto $X_n^* = \sup_{u\in[n,n+1]}\big|W_u - W_n\big|$, si ha che $\{X_n^*\}_{n\ge 0}$ forma una successione di variabili aleatorie indipendenti ed identicamente distribuite, che ammettono tutti i momenti finiti. Ciò implica[10] che

$$\mathbb{P}(\lim_{n\to\infty}\frac{X_n^*}{n} = 0) = 1,$$

e quindi la tesi. □

****Qui si potrebbe mettere la dimostrazione di quest'ultima affermazione, a scopo dimostrativo. Ma in realtà manca anche la legge forte dei grandi numeri e il lemma di Borel Cantelli !!!!

[10] A scopo dimostrativo diamo la dimostrazione che

$$\mathbb{P}(\lim_{n\to\infty}\frac{X_n^*}{n} = 0) = 1.$$

Si tratta di dimostrare che

$$\mathbb{P}\big(\omega:\ \forall\varepsilon > 0,\ \exists m = m(\omega)\ \text{ tale che } \forall n \ge m\ \big|\frac{X_n^*(\omega)}{n}\big| \le \varepsilon\big) = 1$$

ossia, tenuto conto che basta considerare ε razionali,

$$\mathbb{P}\Big(\bigcap_{\varepsilon>0,\varepsilon\in\mathbb{Q}}\bigcup_{m\ge 1}\bigcap_{n\ge m}\{\big|\frac{X_n^*}{n}\big| \le \varepsilon\}\Big) = 1 \ \Leftrightarrow\ \forall\varepsilon > 0, \varepsilon\in\mathbb{Q}\quad \mathbb{P}\Big(\bigcup_{m\ge 1}\bigcap_{n\ge m}\{\big|\frac{X_n^*}{n}\big| \le \varepsilon\}\Big) = 1.$$

Basta quindi dimostrare che per ogni $\varepsilon > 0$ si ha

$$\mathbb{P}\Big(\bigcap_{m\ge 1}\bigcup_{n\ge m}\{\big|\frac{X_n^*}{n}\big| > \varepsilon\}\Big) = 0.$$

Ciò discende dal Lemma di Borel-Cantelli, che assicura che una condizione sufficiente per la precedente uguaglianza è che

$$\sum_{n=1}^{\infty}\mathbb{P}\Big(\big|\frac{X_n^*}{n}\big| > \varepsilon\Big) < \infty.$$

Infatti, per la disuguaglianza di Markov (nota anche come disuguaglianza di Chebyshev generalizzata), qualunque sia $\beta > 1$

$$\sum_{n=1}^{\infty}\mathbb{P}\Big(\big|\frac{X_n^*}{n}\big| > \varepsilon\Big) \le \sum_{n=1}^{\infty}\frac{1}{(\varepsilon\, n)^\beta}\mathbb{E}\Big(\big|X_n^*\big|^\beta\Big) = \frac{\mathbb{E}\Big(\big|X_1^*\big|^\beta\Big)}{\varepsilon^\beta}\sum_{n=1}^{\infty}\frac{1}{n^\beta} < \infty.$$

Si noti che non abbiamo usato l'indipendenza delle variabili aleatorie (quella serve per la legge dei grandi numeri) mentre abbiamo usato la loro identica distribuzione, ma la dimostrazione funzionerebbe lo stesso sotto l'ipotesi che $\sup_{n\ge 1}\mathbb{E}\big(|X_1^*|^\beta\big) \le M < \infty$.

5.6 Calcolo del risolvente del nucleo del calore ****

Lo scopo di questa sezione è calcolare, utilizzando i risultati precedenti, il risolvente del semigruppo dato dal nucleo del calore $g_t(y-x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{(y-x)^2}{2t}}$:

$$\begin{aligned} R_\lambda f(x) &:= \int_0^\infty e^{-\lambda\, t} \int_{\mathbb{R}} f(y)\, g_t(y-x)\, dy\, dt \\ &= \int_{\mathbb{R}} f(y) \int_0^\infty e^{-\lambda\, t} g_t(y-x)\, dt\, dy \end{aligned} \tag{5.4}$$

che risulta uguale, ed è quello che vogliamo dimostrare a

$$= \int_{\mathbb{R}} f(y)\, \frac{e^{-\sqrt{2\lambda}\,|y-x|}}{\sqrt{2\lambda}}\, dy. \tag{5.5}$$

L'interesse sta nel fatto che, posto $A = \frac{1}{2}\frac{d^2}{dx^2}$, su C_b^2

$$R_\lambda = (\lambda - A)^{-1} = (\lambda - \frac{1}{2}\frac{d^2}{dx^2})^{-1},$$

ossia, per ogni $f \in C_b$, si ha $R_\lambda f \in C_b^2$

$$(\lambda - \frac{1}{2}\frac{d^2}{dx^2}) R_\lambda f(x) = f(x),$$

o in altre parole, $R_\lambda f(x)$ è la soluzione del problema

$$\lambda\, g(x) - \frac{1}{2}\frac{d^2}{dx^2}\, g(x) = f(x), \qquad g \in C_b^2.$$

ATTENZIONE, meglio mettere C_b^{un}, lo spazio delle funzioni uniformemente contiue e limitate, e $C_b^{2,un}$...

Tale uguaglianza nasce dal fatto (qui ne diamo solo una dimostrazione formale) che, posto

$$T_t f(x) = \int_{\mathbb{R}} f(y)\, g_t(y-x)\, dy$$

si ha che

$$R_\lambda f(x) := \int_0^\infty e^{-\lambda\, t}\, T_t f(x)\, dt.$$

Tenendo conto che

$$\frac{d}{dt} T_t f(x) = A\, T_t f(x) = \frac{1}{2}\frac{d^2}{dx^2}\, T_t f(x),$$

si ottiene quindi che

$$\begin{aligned} (\lambda - A)\, R_\lambda f = \lambda\, R_\lambda f - A\, R_\lambda f &= \lambda \int_0^\infty e^{-\lambda\, t}\, T_t f\, dt - A \int_0^\infty e^{-\lambda\, t}\, T_t f\, dt \\ &= \int_0^\infty \lambda\, e^{-\lambda\, t}\, T_t f\, dt - \int_0^\infty e^{-\lambda\, t}\, A\, T_t f\, dt \\ &= -\left(\int_0^\infty T_t f\, de^{-\lambda\, t} + \int_0^\infty e^{-\lambda\, t}\, dT_t f \right) \\ &= -\int_0^\infty d\Big(e^{-\lambda\, t}\, T_t f\Big) = -(e^{-\lambda\, t}\, T_t f\Big|_0^\infty \\ &= -0 + e^{-\lambda\, 0}\, T_0 f = f \end{aligned}$$

Tornando al calcolo esplicito, per dimostrare che vale la (5.5), basta dimostrare che

$$\int_0^\infty e^{-\lambda\, t} g_t(y-x)\, dt = \frac{e^{-\sqrt{2\lambda}\,|y-x|}}{\sqrt{2\lambda}}. \tag{5.6}$$

L'idea è la seguente. Noi abbiamo già calcolato, per $\alpha > 0$ e $a(= y - x) > 0$,

$$\int_0^\infty e^{-\alpha t} \mathbb{P}(W_t \geq a) dt = \frac{1}{2\alpha} \mathbb{E}[e^{-\alpha \tau_a}]$$

e abbiamo anche calcolato $\mathbb{E}[\exp\{-\alpha\, \tau_a\}] = \exp\{-a\, \sqrt{2\,\alpha}\}$. Di conseguenza

$$\int_0^\infty e^{-\alpha t} \mathbb{P}(W_t \geq a) dt = \frac{1}{2\alpha}\, \exp\{-a\, \sqrt{2\,\alpha}\},$$

e allora, derivando ambo i membri della precedente uguaglianza si ottiene

$$-\int_0^\infty e^{-\alpha t}\, g_t(a)\, dt = -\sqrt{2\,\alpha}\, \frac{1}{2\alpha}\, \exp\{-a\, \sqrt{2\,\alpha}\} = -\frac{1}{\sqrt{2\alpha}}\, \exp\{-a\, \sqrt{2\,\alpha}\},$$

cioè la tesi, per $\lambda = \alpha$ ed $a > 0$. Per il caso $a = -|a| < 0$, basta poi osservare che $\tau_{-|a|}$ ha la stessa legge di $\tau_{|a|}$, e che $\mathbb{P}(W_t \leq -|a|) = \mathbb{P}(W_t \geq |a|)$.

Capitolo 6

Integrale stocastico: cenni

Sia $W = (W_t)_t$ un moto browniano continuo standard. L'obiettivo è di dare un significato ad espressioni del tipo

$$\int_0^T X_s(\omega) dW_s(\omega) \tag{6.1}$$

dove l'integrando $(X_s)_{0 \leq s \leq T}$ è un processo che gode di proprietà da precisare. Ricordiamo che se una funzione g è a variazione finita, allora è possibile definire l'integrale

$$\int_0^T h(t) dg(t)$$

per ogni funzione h misurabile e limitata[1].

Come abbiamo visto precedentemente nella sezione 4.4, le traiettorie del moto browniano non sono a variazione finita, quindi l'operazione (6.1) non si può definire traiettoria per traiettoria.
La (6.1) si chiamerà *integrale stocastico*. Questo tipo di calcolo è fondamentale per la costruzione e lo studio di nuovi processi.

In questa sezione $(\mathcal{F}_t)$ denota una filtrazione rispetto alla quale il processo di Wiener è una martingala. In particolare si può considerare $\mathcal{F}_t = \mathcal{F}_t^W$.

6.1 Integrale stocastico per processi integrabili, a partire dai processi elementari

Consideriamo le due seguenti classi di processi

Definizione 6.1 (Processi elementari predicibili)**.** *I processi del tipo*

$$X_t(\omega) = \sum_{i=0}^{n-1} C_i(\omega) I_{(t_i, t_{i+1}]}(t), \tag{6.2}$$

[1]† Se g è a variazione finita continua a destra, allora si può scrivere $g = g^+ - g^-$ con g^+ ed g^- funzioni crescenti in senso lato, non negative continue a destra e limitate. Allora l'integrale di Lebesgue-Stieltjes è definito da

$$\int_0^T h(t) dg(t) := \int_0^T h(t) \mu_g(dt),$$

dove

$$\mu_g = \mu_{g^+} - \mu_{g^-},$$

e $\mu_{g^\pm}$ sono le misure su $[0, T]$ univocamente definite da

$$\mu_{g^\pm}((a, b]) = g^\pm(b) - g^\pm(a).$$

Nel caso in cui g sia continua, e anche h lo sia allora, l'integrale di Lebesgue-Stieltjes coincide con l'integrale di Riemann-Stieltjes, ovvero con

$$\lim_{n \to \infty} \sum_k h(t_k^*)[g(t_k) - g(t_{k-1})],$$

dove $\{t_k\}_k$ è una partizione di $[0, T]$ e $t_k^* \in [t_{k-1}, t_k]$.

dove $\alpha = t_0 < t_1 < ... < t_n = \beta$, e le C_i sono variabili aleatorie reali $\mathcal{F}_{t_i}$-misurabili vengono detti **processi elementari $(\mathcal{F}_t)$-predicibili**, *o, più semplicemente,* **processi elementari predicibili** *(se non ci sono dubbi sulla filtrazione).*

Nel seguito la ***famiglia dei processi elementari predicibili*** viene denotata con $\mathcal{E}([\alpha, \beta])$.

Definizione 6.2. *I processi del tipo*

$$X_t(\omega) = \sum_{i=0}^{n-1} C_i(\omega) I_{[t_i, t_{i+1})}(t), \tag{6.3}$$

dove $\alpha = t_0 < t_1 < ... < t_n = \beta$, e le C_i sono variabili aleatorie reali $\mathcal{F}_{t_i}$-misurabili vengono chiamati **processi elementari $(\mathcal{F}_t)$-opzionali** *o, più semplicemente,* **processi elementari opzionali** *(se non ci sono dubbi sulla filtrazione).*

Nel seguito la ***famiglia dei processi elementari opzionali*** viene denotata con $\mathcal{E}_{op}([\alpha, \beta])$.

Si noti che i processi considerati negli Esempi 3.11 e 3.12, per $\alpha = 0$ e $\beta = T$, nel definire l'integrale stocastico, sono esempi di processi elementari predicibili. Ricordiamo che l'integrale stocastico rispetto al processo di Wiener per tali processi viene definito come

$$\mathcal{I}_t^W(X) = \sum_{i=1}^{n} C_{i-1}(\omega)\big(W_{t_i} - W_{t_{i-1}}\big) \left(= \sum_{i=0}^{n-1} C_i(\omega)\big(W_{t_{i+1}} - W_{t_i}\big) \right).$$

Inoltre i processi elementari predicibili (e anche quelli opzionali) sono esempi di processi (congiuntamente) misurabili:

Definizione 6.3 (Processi (congiuntamente) misurabili)**.** *Sia $X : [\alpha, \beta] \times \Omega :\mapsto \mathbb{R}$, $(t, \omega) \mapsto X(t, \omega)\big(= X_t(\omega)\big)$. Sia $\mathcal{B}([\alpha, \beta]) \times \mathcal{F}$ la σ-algebra prodotto su $[\alpha, \beta] \times \Omega$, ovvero la σ-algebra generata da $\{J \times A \subseteq [\alpha, \beta] \times \Omega;\ J \in \mathcal{B}([\alpha, \beta]), A \in \mathcal{F}\}$. Il processo $(X_t)_t$ si dice* **(congiuntamente) misurabile** *se la funzione X è $\mathcal{B}([\alpha, \beta]) \times \mathcal{F}$-misurabile.*

In realtà si tratta di processi progressivamente misurabili e continui a sinistra (continui a destra nel caso dei processi opzionali).

Negli Esempi 3.11 e 3.12 veniva anche richiesto che le variabili aleatorie coinvolte nella definizione del processo predicibile elementare fossero di quadrato sommabile (o addirittura limitati). Con questa richiesta (ovvero se C_i sono anche di quadrato integrabile) il processo elementare predicibile (o opzionale) X è un elemento del seguente spazio di processi, che è anche uno spazio vettoriale, metrico e completo.[2]

Definizione 6.4. *Lo spazio $\mathbb{L}^2\big([\alpha, \beta] \times \Omega\big) = \mathbb{L}^2\big([\alpha, \beta] \times \Omega, \mathcal{B}([\alpha, \beta]) \times \mathcal{F}, \lambda \times \mathbb{P}\big)$, dove λ indica la misura di Lebesgue, è lo spazio delle classi di equivalenza dei processi (congiuntamente) misurabili per i quali*

$$\mathbb{E}\left[\int_\alpha^\beta |X_s|^2 ds\right] < \infty. \tag{6.4}$$

Parlando di classi di equivalenza si intende che identifichiamo due processi X e X' se

$$\mathbb{E}\left[\int_\alpha^\beta |X_s - X'_s|^2 ds\right] = 0.$$

Si può definire quindi la chiusura di

$$\mathcal{E}^2([\alpha, \beta]) = \{X \in \mathcal{E},\ \text{con } C_i \text{ di quadrato sommabile}\}$$

[2]† Il fatto che $\mathbb{L}^2\big([\alpha, \beta] \times \Omega\big)$ sia uno spazio vettoriale è ovvio, in quanto se (6.4) vale per due processi X_1 e X_2, allora vale anche per la sua combinazione lineare $Y = a_1 X_1 + a_2 X_2$.
Inoltre $\mathbb{L}^2\big([\alpha, \beta] \times \Omega\big)$ è uno spazio metrico completo con la metrica

$$d(X, X') := \left(\mathbb{E}\left[\int_\alpha^\beta |X_s - X'_s|^2 ds\right]\right)^{\frac{1}{2}}.$$

In realtà si tratta di uno spazio di Hilbert.

in tale spazio metrico, rispetto alla metrica

$$d(X,X') := \left(\mathbb{E}\left[\int_\alpha^\beta |X_s - X'_s|^2 ds \right] \right)^{\frac{1}{2}}.$$

Denoteremo tale chiusura come

$$\overline{\mathcal{E}^2}([\alpha,\beta]).$$

È importante identificare lo spazio $\overline{\mathcal{E}^2}([\alpha,\beta])$, perché è quello su cui (come vedremo) sappiamo definire l'integrale stocastico[3]. È altrettanto importante individuare almeno una classe ampia di processi che vi appartengono: è possibile mostrare che se X è un processo adattato e X ha traiettorie continue, ed è un processo in $\mathbb{L}^2\big([\alpha,\beta]\times\Omega\big)$, ovvero se vale (6.4), allora X appartiene a $\overline{\mathcal{E}^2}([\alpha,\beta])$: infatti se $\{t_i^n\}_{i=0,\cdots,m_n}$ è una partizione di $[\alpha,\beta]$ con $\max_i |t_i^n - t_{i-1}^n| \to 0$, allora la successione di processi $\{X^n\}$ definiti da

$$X_t^n(\omega) = \sum_{i=1}^{m_n} X_{t_{i-1}^n}(\omega) I_{(t_{i-1}^n, t_i^n]}(t),$$

è una successione di processi elementari e predicibili in $\mathbb{L}^2\big([\alpha,\beta]\times\Omega\big)$ che converge ad X nella metrica d, ovvero che

$$\lim_{n\to\infty} \mathbb{E}\left[\int_\alpha^\beta |X_t - X_t^n|^2 dt \right] = 0.$$

Se $X \in \overline{\mathcal{E}^2}([\alpha,\beta])$ allora, per definizione, esiste una successione di processi elementari e predicibili che soddisfa la precedente relazione. Di conseguenza la successione X_n è una successione di Cauchy in $\mathbb{L}^2\big([\alpha,\beta]\times\Omega\big)$, ovvero

$$\lim_{n,m\to\infty} \mathbb{E}\left[\int_\alpha^\beta |X_t^m - X_t^n|^2 dt \right] = 0.$$

Sappiamo inoltre che, per $\alpha = 0$ e $\beta = T$, per ogni processo elementare $\{X_n\}$, il processo $\mathcal{I}^W(X^n) = (\mathcal{I}_t^W(X^n))_{t\in[0,T]}$ definisce una martingala a traiettorie continue . Ovviamente anche la differenza $\mathcal{I}_t^W(X^n) - \mathcal{I}_t^W(X^m)$ è una martingala. Inoltre, chiaramente, per l'integrale stocastico dei processi elementari valgono le proprietà di linearità, e quindi

$$\mathcal{I}_t^W(X^n) - \mathcal{I}_t^W(X^m) = \mathcal{I}_t^W(X^n - X^m).$$

Dall'Esempio 3.12 sappiamo che, per ogni $t \in [0,T]$,

$$\mathbb{E}\left[\left|\mathcal{I}_t^W(X^n - X^m)\right|^2 \right] = \mathbb{E}\left[\int_0^t |X_t^m - X_t^n|^2 dt \right]. \tag{6.5}$$

Inoltre applicando la disuguaglianza di Doob per $p = 2$, estesa alle martingale continue, alla martingala continua $\mathcal{I}^W(X^n - X^m)$, sappiamo che

$$\mathbb{E}\left[\max_{t\in[0,T]} \left|\mathcal{I}_t^W(X^n - X^m)\right|^2 \right] \le 4\mathbb{E}\left[\left|\mathcal{I}_T^W(X^n - X^m)\right|^2 \right]. \tag{6.6}$$

D'altra parte, ovviamente, si ha anche che

$$\left|\mathcal{I}_T^W(X^n - X^m)\right|^2 \le \max_{t\in[0,T]} \left|\mathcal{I}_t^W(X^n - X^m)\right|^2.$$

[3]La classe **$\overline{\mathcal{E}^2}([\alpha,\beta])$ che abbiamo definito comprende tutti i processi $(\mathcal{F}_t)$-predicibili di $\mathbb{L}^2\big([\alpha,\beta]\times\Omega\big)$. Ricordiamo qui la definizione di processi $(\mathcal{F}_t)$-predicibili: i processi predicibili sono i processi da $[\alpha,\beta]\times\Omega \to \mathbb{R}$ e misurabili rispetto alla σ-algebra su $[\alpha,\beta]\times\Omega$, generata dai processi elementari e predicibili, o equivalentemente dai processi $(\mathcal{F}_t)$-adattati e continui a sinistra. Più precisamente si definisce $\mathcal{P} = \mathcal{P}\big((\mathcal{F}_t)_t\big)$ la σ-algebra su $[\alpha,\beta]\times\Omega$ generata dai processi elementari predicibili. Lo spazio $\overline{\mathcal{E}^2}([\alpha,\beta])$ è allora lo spazio $\mathbb{L}^2\big([\alpha,\beta]\times\Omega,\mathcal{P},\lambda\times\mathbb{P}\|_{\mathcal{P}}\big)$. Tale spazio è un sottospazio chiuso di $\mathbb{L}^2\big([\alpha,\beta]\times\Omega,\big)$ in quanto i processi elementari predicibili sono congiuntamente misurabili e quindi $\mathcal{P}$ è contenuta nella σ- algebra prodotto. Va anche detto che, nel caso del processo di Wiener, si potrebbe estendere l'integrale anche ai processi opzionali, la cui definizione è simile ai processi predicibili, ma a partire dai processi elementari opzionali.

Passando ai valori attesi, nella precedente disuguaglianza, e tenendo conto della (6.6), si ha

$$\mathbb{E}\left[\left|\mathcal{I}_T^W(X^n - X^m)\right|^2\right] \le \mathbb{E}\left[\max_{t\in[0,T]}\left|\mathcal{I}_t^W(X^n - X^m)\right|^2\right] \le 4\mathbb{E}\left[\left|\mathcal{I}_T^W(X^n - X^m)\right|^2\right].$$

Utilizzato la relazione (6.5) per $t = T$, si ottiene che

$$\mathbb{E}\left[\int_0^T |X_t^m - X_t^n|^2 dt\right] \le \mathbb{E}\left[\max_{t\in[0,T]}\left|\mathcal{I}_t^W(X^n - X^m)\right|^2\right] \le 4\mathbb{E}\left[\int_0^T |X_t^m - X_t^n|^2 dt\right] \tag{6.7}$$

In altre parole, se i processi elementari X^n sono una successione di Cauchy in $\mathbb{L}^2\big([0,T]\times\Omega\big)$, allora la successione di martingale a traiettorie continue $\mathcal{I}^W(X^n) = (\mathcal{I}_t^W(X^n))_{t\in[0,T]}$, risulta una successione di Cauchy[4] rispetto alla metrica

$$\tilde{d}(M, M') := \left(\mathbb{E}\left[\max_{t\in[0,T]}|M_t - M_t'|^2\right]\right)^{\frac{1}{2}};$$

in quanto la (6.7) si può riscrivere come

$$d^2(X^n, X^m) \le \tilde{d}^2(\mathcal{I}^W(X^n), \mathcal{I}^W(X^m)) \le 4d^2(X^n, X^m)$$

Viceversa se i processi elementari X^n in $\mathbb{L}^2\big([0,T]\times\Omega\big)$ sono tali che la successione $\mathcal{I}^W(X^n) = (\mathcal{I}_t^W(X^n))_{t\in[0,T]}$ è una successione di Cauchy rispetto alla metrica $\tilde{d}$ allora la successione di processi X^n risulta di Cauchy in $\mathbb{L}^2\big([0,T]\times\Omega\big)$ rispetto alla metrica d.

Da questa relazione si può dedurre l'esistenza di un processo limite, in questa metrica $\tilde{d}$, della successione di processi $\mathcal{I}^W(X_n)$. Il processo limite[5] viene chiamato integrale stocastico di X e denotato anche esso con

$$\mathcal{I}_t^W(X) = \int_0^t X_s dW_s.$$

Si osservi che la convergenza considerata implica la convergenza in probabilità di

$$\sup_{t\imath[0,T]}\left|\mathcal{I}_t^W(X_n) - \mathcal{I}_t^W(X)\right|$$

Inoltre il processo limite $\mathcal{I}_t^W(X)$ è a sua volta una martingala di quadrato integrabile[6], e che la sua variazione quadratica è il processo $\int_0^t X_s ds$**[7].

Ribadiamo che la definizione viene (in queste note) data quindi solo per i processi $X \in \overline{\mathcal{E}^2}([\alpha,\beta])$.

È chiaro che l'integrale stocastico $\int_0^t X_s dW_s$ è lineare in X. Inoltre $\int_0^t X_s dW_s$ è una martingala a traiettorie continue. Nel caso dei processi elementari ciò discende dal risultato esaminato nell'Esempio 3.12:

[4] In altre parola, stiamo considerando $\mathcal{I}^W(X^n) = (\mathcal{I}_t^W(X^n))_{t\in[0,T]}$ come un elemento dello spazio metrico completo $L^2(\Omega; C[0,T], \mathbb{P})$. Ovviamente identifichiamo processi che sono indistinguibili, ossia che differiscono su un insieme di probabilità nulla.

[5] Si può mostrare che il limite non dipende dalla particolare successione $\{X^n\}$ di processi elementari, scelta per approssimare il processo X.

[6] La convergenza di $\mathcal{I}_t^W(X_n)$ a $\mathcal{I}_t^W(X)$ nella metrica $\tilde{d}$ implica la convergenza **a zero di

$$\mathbb{E}\left[\left|\mathcal{I}_t^W(X_n) - \mathcal{I}_t^W(X)\right|\right], \qquad \forall t\in[0,T], **$$

ed implica l'esistenza di una sottosuccessione $\mathcal{I}_t^W(X_{n_k})$ convergente quasi certamente ed in norma uniforme. Inoltre, se una successione di $\mathcal{G}_t$-martingale $M_t^{(k)}$ converge, per k che tende ad infinito, a un processo M_t, allora il processo limite M_t è una $\mathcal{G}_t$-martingala: infatti, per ipotesi, qualunque siano $k \ge 1$, $s \le t$ e $A \in \mathcal{G}_s$, si ha

$$\mathbb{E}[M_t^{(k)}\,\mathbb{I}_A] = \mathbb{E}[M_s^{(k)}\,\mathbb{I}_A].$$

Basta allora passare al limite per k che tende ad infinito, per ottenere che

$$\mathbb{E}[M_t\,\mathbb{I}_A] = \mathbb{E}[M_s\,\mathbb{I}_A],$$

in quanto, ad esempio

$$\left|\mathbb{E}[M_t^{(k)}\,\mathbb{I}_A] - \mathbb{E}\left[M_t\,\mathbb{I}_A\right]\right| \le \mathbb{E}\left[\left|M_t^{(k)} - M_t\right|\,\mathbb{I}_A\right] \le \mathbb{E}\left[\left|M_t^{(k)} - M_t\right|\right].$$

[7] **La convergenza di di $\mathcal{I}_t^W(X_n)$ a $\mathcal{I}_t^W(X)$ nella metrica $\tilde{d}$ implica la convergenza a zero di

$$\mathbb{E}\left[\left|\mathcal{I}_t^W(X_n) - \mathcal{I}_t^W(X)\right|^2\right], \qquad \forall t\in[0,T],$$

Lemma 6.1. *Se X è un processo elementare predicibile, come nella (6.3), con C_i di quadrato integrabile, allora*

$$\mathbb{E}\left(\int_\alpha^\beta X_t dW_t \Big| \mathcal{F}_\alpha\right) = 0$$
$$\mathbb{E}\left[\left(\int_\alpha^\beta X_t dW_t\right)^2 \Big| \mathcal{F}_\alpha\right] = \mathbb{E}\left[\int_\alpha^\beta {X_t}^2 \Big| \mathcal{F}_\alpha\right].$$

In particolare, l'integrale stocastico di un processo elementare di $\mathcal{E}^2$ è una variabile aleatoria centrata di quadrato integrabile e

$$\mathbb{E}\left[\left(\int_\alpha^\beta X_t dW_t\right)^2\right] = \mathbb{E}\left[\int_\alpha^\beta {X_t}^2 dt\right].$$

Le proprietà di linearità, di martingala, di continuità e le proprietà precedenti sui valori medi si ottengono per i processi più generali, con un passaggio al limite (**vedere le note 6 e 7).

6.1.1 Esempi di calcolo di integrali stocastici

Procediamo con qualche esempio di calcolo di integrale stocastico.

Esempio 6.1 (integrale stocastico di $X_t = W_t$)**.** *In questo esempio proviamo a calcolare*

$$\int_\alpha^\beta 2W_s dW_s.$$

Essendo $X_t = 2W_t$ un processo a traiettorie continue, con

$$\mathbb{E}\left[\int_\alpha^\beta |2W_s|^2 ds\right] = \int_\alpha^\beta \mathbb{E}\left[|2W_s|^2\right] ds = \int_\alpha^\beta 4s ds < \infty$$

possiamo utilizzare come processo approssimante[8]

$$X_t^n = \sum_{i=1}^{m_n} W_{t_{i-1}^n}(\omega) I_{(t_{i-1}^n, t_i^n]}(t),$$

dove si considera la partizione $t_0^n = \alpha < t_1^n < \cdots < t_{m_n}^n = \beta$. Così

$$\int_\alpha^\beta 2W_s dW_s = \lim_n \sum_{i=1}^{m_n} W_{t_{i-1}^n}(W_{t_i^n} - W_{t_{i-1}^n}).$$

---

oltre ovviamente all'ipotesi di convergenza a zero di

$$\mathbb{E}\left[\int_0^t |X_s - X_s^n|^2 ds\right].$$

La prima convergenza a zero implica la convergenza a zero di

$$\mathbb{E}\left[\left|\left(\mathcal{I}_t^W(X_n)\right)^2 - \left(\mathcal{I}_t^W(X)\right)^2\right|\right], \qquad \forall t \in [0,T],$$

mentre la seconda implica la convergenza a zero di

$$\mathbb{E}\left[\left|\int_0^t |X_s|^2 ds - \int_0^t |X_s^n|^2 ds\right|\right].$$

Quindi la successione di martingale $\left(\mathcal{I}_t^W(X_n)\right)^2 - \int_0^t |X_s^n|^2 ds$ converge in L^1 al processo $\left(\mathcal{I}_t^W(X)\right)^2 - \int_0^t |X_s|^2 ds$, che risulta quindi una martingala (si procede come nella precedente nota 6).

[8]**Questo tipo di approssimazione si ottiene per ogni processo continuo a sinistra e adattato.

Considerando che

$$\begin{aligned}
2\sum_k b_k(b_{k+1}-b_k) &= \sum_k [2b_{k+1}b_k - 2b_k^2] = \sum_k [2b_{k+1}b_k - 2b_k^2 - b_{k+1}^2 + b_{k+1}^2] \\
&= \sum_k [(2b_{k+1}b_k - b_k^2 - b_{k+1}^2) + b_{k+1}^2 - b_k^2] = -\sum_k [b_{k+1}-b_k]^2 + \sum_k [b_{k+1}^2 - b_k^2] \\
&= -\sum_k [b_{k+1}-b_k]^2 + [b_{k_{max}}^2 - b_{k_{min}}^2]
\end{aligned}$$

ponendo $b_k = W_{t^n_{i-1}}$ *si ottiene*

$$\int_\alpha^\beta 2W_s dW_s = \lim_n -\sum_{i=1}^{m_n} (W_{t^n_i} - W_{t^n_{i-1}})^2 + W_\beta^2 - W_\alpha^2.$$

da cui, ricordando il Lemma 4.4, si ottiene che

$$\int_\alpha^\beta 2W_s dW_s = -(\beta-\alpha) + W_\beta^2 - W_\alpha^2. \tag{6.8}$$

C'è un altro caso in cui si può calcolare l'integrale stocastico: Se $f \in \mathbb{L}^2([0,T],\lambda)$ è una funzione deterministica, allora $f \in \overline{\mathcal{E}^2}([0,T])$ e quindi ha senso l'integrale stocastico di Itô, che in questo caso coincide con l'integrale stocastico di Wiener, e gode allora di una importante proprietà.

Proposizione 6.2. *Se f è una funzione deterministica, con $f \in \mathbb{L}^2([0,T])$, allora il processo*

$$\int_0^t f(s)dW_s$$

è gaussiano, di media zero e varianza $\int_0^t f^2(s)ds$.

Dimostrazione. L'idea della dimostrazione è la seguente, ed è sostanzialmente l'idea seguita d Wiener, nel definire l'integrale stocastico per processi deterministici.

Sia f è una funzione deterministica e C^1, allora si osserva che vale la seguente formula di integrazione per parti (questa era la definizione originaria di Wiener)

$$\int_0^T f(s)dW_s := f(T)\,W_T - f(0)\,W_0 - \int_0^T W_s\, f'(s)\,ds. \tag{6.9}$$

Per dimostrare questa formula di integrazione per parti basta osservare che, se si considera la partizione $\pi:\ t_0^n = 0 < t_1^n < \cdots < t_{2^n}^n = T$, con $t_i^n = \frac{i}{2^n}T$, allora, il processo deterministico $f(s)$ si può approssimare con

$$f(s) = \lim_{n\to\infty} \sum_{i=1}^{2^n} f(t_{i-1}^n)\,\mathbf{1}_{(t_{i-1}^n, t_i^n]}(s)$$

$$\begin{aligned}
\int_0^T f(s)dW_s &= \lim_{n\to\infty} \sum_{i=1}^{2^n} f(t_{i-1}^n)\,[W_{t_i^n} - W_{t_{i-1}^n}] \qquad \text{(aggiungendo e togliendo i termini } f(t_i^n)\,W_{t_i^n}) \\
&= \lim_{n\to\infty} \sum_{i=1}^{2^n} \Big([f(t_i^n)\,W_{t_i^n} - f(t_{i-1}^n)\,W_{t_{i-1}^n}] + [f(t_{i-1}^n) - f(t_i^n)]\,W_{t_i^n} \Big) \\
&= [f(T)\,W_T - f(0)\,W_0] + \\
&= \frac{1}{2}\,[f(T)\,W_T - f(0)\,W_0] + \sum_{i=1}^{2^n} \big(f'(t_i^n)[t_{i-1}^n - t_i^n] + |t_{i-1}^n - t_i^n|\,O_i(1) \big)\ W_{t_i^n}
\end{aligned}$$

dove $O_i(1)$ è infinitesimo uniformemente in i, e quindi, essendo $f'(t)\ W_t(\omega)$ continue e quindi limitate in $[0,T]$ si ottiene che

$$\int_0^T f(s)dW_s = f(T)\,W_T - f(0)\,W_0 - \lim_{n\to\infty}\sum_{i=1}^{2^n} f'(t_i^n)\,W_{t_i^n}\,(t_i^n - t_{i-1}^n),$$

e quindi la (6.9).

Si osservi che, se invece di considerare l'intervallo $[0,T]$ si fosse considerato l'intervallo $[\alpha,\beta]$, avremmo ottenuto

$$\int_\alpha^\beta f(s)dW_s = f(\beta)\,W_\beta - f(\alpha)\,W_\alpha - \lim_{|\pi|\to 0}\sum_{i=1}^{2^n} W_{t_i^n} f'(t_i^n)\,(t_i^n - t_{i-1}^n) \tag{6.10}$$

Dalla (6.10) si ottiene immediatamente che l'integrale stocastico è una variabile aleatoria gaussiana, in quanto limite di variabili aleatorie gaussiane[9]. Ovviamente si tratta di variabili gaussiane a media nulla ed è facile vedere che la varianza di $f(\beta)\,W_\beta - f(\alpha)\,W_\alpha - \sum_{i=1}^{2^n} W_{t_i^n} f'(t_i^n)\,(t_{i-1}^n - t_{i-1}^n)$ converge al valore $\int_\alpha^\beta f^2(s)ds$. Per il caso generale di funzioni f, basta considerare che lo spazio $C^1([0,T])$ è denso in $\mathbb{L}^2([0,T])$. □

Osservazione 6.1. *La proposizione precedente implica in particolare che*

$$\mathbb{E}\left[\exp\left(\int_0^t f(s)dW_s\right)\right] = \exp\left(\frac{1}{2}\int_0^t f^2(s)ds\right).$$

Infatti si riconosce a sinistra la media dell'esponenziale di una variabile aleatoria gaussiana centrata Z, con varianza σ_Z^2, cioè la sua trasformata di Laplace $\mathcal{L}(\theta) := \mathbb{E}[\exp(\theta Z)]$ calcolata in $\theta = 1$, che è uguale all'esponenziale della sua varianza divisa per 2 (ovvero $\mathcal{L}(\theta) = \exp\{\theta^2\sigma_Z^2/2\}$.

Definizione 6.5. *Sia $M^2([\alpha,\beta])$ il sottospazio di $\mathbb{L}^2\big([\alpha,\beta]\times\Omega\big)$ definito come lo spazio delle classi d'equivalenza dei* **processi $\mathcal{F}_t^W$-progressivamente misurabili**[10] *tali che*

$$\mathbb{E}\left[\int_\alpha^\beta |X_s|^2 ds\right] < \infty.$$

Osservazione 6.2. *Si può definire l'integrale stocastico anche per il processi elementari opzionali e si può dimostrare che l'integrale stocastico è una isometria tra $M^2([0,T])$ (il sottospazio di $\mathbb{L}^2\big([0,T]\times\Omega\big)$ dei processi progressivamente misurabili) e lo spazio delle v.a. $\mathbb{L}^2(\Omega,\mathcal{F}_T,\mathbf{P})$.*

Alla luce della precedente osservazione ci si può chiedere se tutte le variabili aleatorie di $\mathbb{L}^2(\Omega,\mathcal{F},\mathbf{P})$ si possono ottenere come integrale stocastico di un processo di $M^2([0,T])$. Ciò non può essere dal momento che ogni integrale stocastico in $M^2([0,T])$ ha media nulla. Tuttavia si può vedere che, se la filtrazione $(\mathcal{F}_t)_{t\geq 0} = (\mathcal{F}_t^W)_{t\geq 0}$ è quella naturale completata del moto browniano W, allora ogni variabile aleatoria $Z \in \mathbb{L}^2(\Omega,\mathcal{F},\mathbf{P})$, con $\mathcal{F} = \mathcal{F}_T$, si può rappresentare nella forma

$$Z = c + \int_0^T X_s dW_s,$$

con $X \in M^2([0,T])$ e $c \in \mathbb{R}$ e, naturalmente, con $\mathbb{E}(Z) = c$.

Prendendo $Z = M_T$, questa proprietà è equivalente alla proprietà che ogni martingala[11] $(M_t)_{t\in[0,T]}$ di quadrato integrabile si possa rappresentare tramite un'integrale stocastico, come

$$M_t = \mathbb{E}[M_T|\mathcal{F}_t] = c + \int_0^t X_s dW_s$$

[9] Ciò si può dimostrare, ad esempio, usando il fatto che il limite di funzioni caratteristiche di tipo gaussiano converge se e solo se convergono valore atteso e varianza. Ovviamente si considera gaussiano ache il caso degenere di variabili con varianza nulla.

[10] Più in generale, data una filtrazione $(\mathcal{F}_t)$, i processi $\mathcal{F}_t$-progressivamente misurabili sono i processi $X : \Omega\times[0,\infty) \mapsto \mathbb{R}$ tali che qualunque sia t la restrizione di X a funzione $\Omega\times[0,t]$ è misurabile rispetto alla σ-algebra prodotto $\mathcal{F}_t \times \mathcal{B}([0,t])$.

[11] Si tratta sempre del caso in cui la filtrazione 'e la filtrazione generata da W, completata e resa continua a destra.

6.1.2 Integrale stocastico per processi localmente di quadrato integrabile

In questa sezione affrontiamo il problema della definizione dell'integrale stocastico anche per una classe più ampia di processi[12]. La seguente proposizione afferma che, se l'integrando è continuo, allora l'integrale stocastico è il limite di particolari[13] somme di Riemann, in analogia con l'integrale di Lebesgue.

Proposizione 6.3. *Se X è un processo dello spazio $\Lambda^2[\alpha,\beta]$, ovvero è progressivamente misurabile e*

$$\mathbb{P}\left(\int_\alpha^\beta X_t^2 dt < \infty\right) = 1,$$

allora si può dare un senso all'integrale stocastico $\int_\alpha^\beta X_t dW_t$, come limite in probabilità di integrali di processi elementari opzionali. Se inoltre X è un processo continuo, allora per ogni successione $(\pi_n)_n$ di partizioni $\alpha = t_0^n < t_1^n < ... < t_{m_n}^n = \beta$ con $|\pi_n| \to 0$ si ha

$$\sum_{k=1}^{m_n} X(t_{k-1}^n)(W_{t_k^n} - W_{t_{k-1}^n}) \underset{n\to+\infty}{\longrightarrow} \int_\alpha^\beta X(t)dW_t \qquad \textit{in probabilità.}$$

La dimostrazione di tale risultato è basata su due risultati:

Lemma 6.4 (Baldi [1], Lemma???)**.** *Se $X \in \Lambda^2([\alpha,\beta])$, allora*

1 esiste una successione $(Y^{(n)})_n$ di processi continui appartenenti a $\Lambda^2([\alpha,\beta])$ e tali che

$$\lim_{n\to\infty} \int_\alpha^\beta |X_t - Y_t^{(n)}|^2 dt = 0 \quad q.c.$$

2 esiste una successione $(X^{(n)})_n$ di processi elementari appartenenti a $\Lambda^2([\alpha,\beta])$ e tali che

$$\lim_{n\to\infty} \int_\alpha^\beta |X_t - X_t^{(n)}|^2 dt = 0 \quad q.c.$$

Lemma 6.5 (Baldi [1], lemma ???)**.** *Per ogni processo elementare $X \in \Lambda^2([\alpha,\beta])$ e per ogni $\varepsilon > 0$, $\rho > 0$*

$$\mathbf{P}\left(\left|\int_\alpha^\beta X_t dW_t\right| > \varepsilon\right) \le \mathbf{P}\left(\int_\alpha^\beta X_t^2 dt > \rho\right) + \frac{\rho}{\varepsilon^2}.$$

Dimostrazione. In realtà si può dimostrare un risultato più forte:

$$\mathbf{P}\left(\sup_{b\in[\alpha,\beta]}\left|\int_\alpha^b X_t dW_t\right| > \varepsilon\right) \le \mathbf{P}\left(\int_\alpha^\beta X_t^2 dt > \rho\right) + \frac{\rho}{\varepsilon^2}. \tag{6.11}$$

La dimostrazione è basata sul fatto che, posto $X_t^\rho = X_t\,\mathbf{1}_{[0,\rho]}\big(\int_\alpha^t X_s^2 ds\big)$,

$$\begin{aligned}\mathbf{P}\left(\sup_{b\in[\alpha,\beta]}\left|\int_\alpha^b X_t dW_t\right| > \varepsilon\right) &= \mathbf{P}\left(\sup_{b\in[\alpha,\beta]}\left|\int_\alpha^b X_t dW_t\right| > \varepsilon, \int_\alpha^\beta X_t^2 dt > \rho\right) + \mathbf{P}\left(\sup_{b\in[\alpha,\beta]}\left|\int_\alpha^b X_t dW_t\right| > \varepsilon, \int_\alpha^\beta X_t^2 dt \le \rho\right)\\ &\le \mathbf{P}\left(\int_\alpha^\beta X_t^2 dt > \rho\right) + \mathbf{P}\left(\sup_{b\in[\alpha,\beta]}\left|\int_\alpha^b X_t^\rho dW_t\right| > \varepsilon, \int_\alpha^\beta X_t^2 dt \le \rho\right)\end{aligned}$$

[12] Questa parte non è stata ancora rivista e usa un approccio diverso da quello visto a lezione: si consiglia di vedere invece gli appunti del Prof. Bertini[2].

[13] Una somma di Riemann è una somma del tipo

$$\sum_{k=1}^{m_n} X(t_{k-1}^{n*})(W_{t_k^n} - W_{t_{k-1}^n})$$

con $t_{k-1}^{n*} \in [t_{k-1}^n, t_k^n]$ La differenza nel caso dell'integrale stocastico di Itô, sta nel fatto che va sempre scelto $t_{k-1}^{n*} = t_{k-1}^n$.

dove si è usato il fatto che, se $\int_\alpha^\beta X_t^2 dt \le \rho$ allora $X_t = X_t^\rho$,

$$\le \mathbf{P}\left(\int_\alpha^\beta X_t^2 dt > \rho\right) + \mathbf{P}\left(\sup_{b\in[\alpha,\beta]}\left|\int_\alpha^b X_t^\rho dW_t\right| > \varepsilon\right)$$
$$\le \mathbf{P}\left(\int_\alpha^\beta X_t^2 dt > \rho\right) + \frac{\mathbf{E}\left(\left|\int_\alpha^\beta X_t^\rho dW_t\right|^2\right)}{\varepsilon^2} \le \mathbf{P}\left(\int_\alpha^\beta X_t^2 dt > \rho\right) + \frac{\rho}{\varepsilon^2},$$

dove si è usata la disuguaglianza di Kolmogorov e il fatto che X_t^ρ è di quadrato integrabile, e quindi

$$\mathbf{E}\left(\left|\int_\alpha^\beta X_t^\rho dW_t\right|^2\right) = \mathbf{E}\left(\int_\alpha^\beta (X_t^\rho)^2\, d_t\right) \le \rho$$

□

Definiamo l'integrale stocastico per ogni processo $X \in \Lambda^2([\alpha,\beta])$. Per il Lemma 6.4 esiste una successione di processi elementari $(X_n)_n \subset \Lambda^2([\alpha,\beta])$ tale che

$$\int_\alpha^\beta |X_n(t) - X(t)|^2 dt \longrightarrow_{n\to\infty} 0 \qquad \text{in probabilità.}$$

Quindi, per ogni $\eta > 0$, $\delta > 0$ esiste n_0 tale che per $n, m > n_0$

$$\mathbf{P}\left(\int_\alpha^\beta |X_n(t) - X(t)|^2 dt > \delta\right) < \eta$$

e, per il Lemma 6.5, per ogni $\rho > 0$

$$\mathbf{P}\left(\left|\int_\alpha^\beta X_n(t)dW_t - \int_\alpha^\beta X_m(t)dW_t\right| > \varepsilon\right) \le \mathbf{P}\left(\int_\alpha^\beta |X_n(t) - X_m(t)|^2 dt > \rho\right) + \frac{\rho}{\varepsilon^2}.$$

Scegliamo, ora, prima ρ abbastanza piccolo perché sia $\frac{\rho}{\varepsilon^2} < \frac{\varepsilon}{2}$ e poi n_0 abbastanza grande perché per $n, m > n_0$ il primo termine a destra sia strettamente minore di $\frac{\varepsilon}{2}$. La successione

$$\left(\int_\alpha^\beta X_n(t)dW_t\right)_n$$

è di Cauchy, e dunque convergente, in probabilità. Indichiamo con

$$\int_\alpha^\beta X_t dW_t$$

il suo limite. Esso si chiama *integrale stocastico* di X rispetto a W.

6.1.3 Osservazioni su alcune proprietà degli integrali stocastici

Osservazione 6.3. *Sappiamo che, per ogni a e $b \in \mathbb{R}$*

$$\int_\alpha^\beta [aX_t + bY_t]dW_t = a\int_\alpha^\beta X_t dW_t + b\int_\alpha^\beta Y_t dW_t,$$

se X e Y sono processi per cui ha senso l'integrale stocastico, ad esempio se $X, Y \in \overline{\mathcal{E}^2}([\alpha,\beta])$, o se $X, Y \in \Lambda^2([\alpha,\beta])$. Ma se $A = A(\omega)$ è una variabile aleatoria (per semplicità prendiamo $b = 0$), non possiamo dire altrettanto, o meglio dipende dalle proprietà di misurabilità di A. Infatti, mentre $A\int_\alpha^\beta X_t dW_t$ si può scrivere sempre, sappiamo dare un significato a $\int_\alpha^\beta A(\omega)X_t dW_t$ solo se $AX \in \overline{\mathcal{E}}^2([\alpha,\beta])$ (o se $AX \in \Lambda^2([\alpha,\beta])$). In particolare AX_t deve essere progressivamente misurabile, di conseguenza è necessario che AX_t sia $\mathcal{F}_\alpha$-misurabile. Affinché ciò si verifichi è sufficiente che A sia $\mathcal{F}_\alpha$-misurabile. Per le condizioni di integrabilità è sufficiente che A sia una v.a. limitata.

Dunque possiamo enunciare il seguente teorema.

Teorema 6.6. *Sia A una variabile aleatoria $\mathcal{F}_\alpha$-misurabile limitata; allora per $X \in \overline{\mathcal{E}}^2([\alpha,\beta])$ (o per $X \in \Lambda^2([\alpha,\beta])$)*

$$\int_\alpha^\beta AX(t)dW_t = A\int_\alpha^\beta X(t)dW_t.$$

6.2 Calcolo stocastico e formula di Itô

Sia X un processo tale che per ogni $0 \le t_1 < t_2 \le T$

$$X_{t_2} - X_{t_1} = \int_{t_1}^{t_2} F_t dt + \int_{t_1}^{t_2} G_t dW_t,$$

dove F_t e G_t soddisfano opportune condizioni.[14]

Diremo allora che X ammette il ***differenziale stocastico***

$$dX_t = F_t dt + G_t dW_t.$$

Un processo che ammette differenziale stocastico si chiama anche un ***processo di Itô***.

Come primo esempio banale, prendiamo $X_t = W_t$ allora ovviamente $dW_t = dW_t$ e quindi $F_t = F_t^W = 0$ e $G_t = G_t^W = 1$. Lo stesso vale anche per $X_t = x + W_t$, con x costante reale, visto che nella definizione del differenziale, la costante x scompare.

Come secondo esempio consideriamo $X_t = W_t^2$ (o anche $X_t = x + W_t^2$). Ricordiamo che per la (6.8) si ha che qualunque siano $0 \le t_1 < t_2 \le T$, si ha

$$X_{t_2} - X_{t_1} = W_{t_2}^2 - W_{t_1}^2 = \int_{t_1}^{t_2} 2W_s dW_s + (t_2 - t_1),$$

ovvero in forma differenziale

$$dX_t = dW_t^2 = 2W_t dW_t + dt,$$

che è come dire che $F_t = F_t^{W^2} = 1$ e $G_t = G_t^{W^2} = 2W_t$. Si noti la differenza con il calcolo del differenziale usuale.[15]

Osservazione 6.4. *Il differenziale stocastico, se esiste, è unico, nel senso che se X ammette una rappresentazione del tipo precedente, allora F e G sono determinati univocamente*[16].

Definiamo

$$< X >_t = \int_0^t {G_s}^2 ds.$$

Il processo $< X >$ non è altro che il processo crescente associato alla martingala $\int_0^t G_s dW_s$ che compare nella sua definizione. In maniera simile, se Y è un altro processo di Itô, avente differenziale stocastico

$$dY_t = H_t dt + K_t dW_t,$$

porremo

$$< X, Y >_t = \int_0^t G_s K_s ds.$$

[14] Per quanto riguarda le proprietà di misurabilità, si richiede che sia F_T che G_t siano $\mathcal{F}_t^W$-progressivamente misurabili. Inoltre si richiede che abbiano le opportune proprietà di integrabilità per cui abbia senso fare gli integrali.

[15] Il motivo sta nel fatto che il processo di Wiener non ha traiettorie a variazione limitata o meglio possiamo interpretare il Lemma 4.4 come $(W_{t+\delta} - W_t)^2 = O(\delta)$.

[16] A meno di identificazione di processi. (DA CHIARIRE MEGLIO)

Proposizione 6.7 (Differenziale del prodotto). *Se X_i, $i = 1, 2$, sono due processi aventi differenziale stocastico*

$$dX_i(t) = F_i(t)dt + G_i(t)dW_t,$$

allora

$$\begin{aligned} d(X_1(t)X_2(t)) &= X_1(t)dX_2(t) + X_2(t)dX_1(t) + G_1(t)G_2(t)dt \\ &= X_1(t)dX_2(t) + X_2(t)dX_1(t) + d < X_1, X_2 >_t . \end{aligned}$$

Ovvero, se $t_1 < t_2$,

$$\begin{aligned} X_1(t_2)X_2(t_2) - X_1(t_1)X_2(t_1) &= \int_{t_1}^{t_2} [X_1(t)F_2(t) + X_2(t)F_1(t)]dt \\ &+ \int_{t_1}^{t_2} [X_1(t)G_2(t) + X_2(t)G_1(t)]dW_t \\ &+ \int_{t_1}^{t_2} G_1(t)G_2(t)dt. \end{aligned}$$

Dimostrazione. Daremo la dimostrazione solo nel caso in cui $F_i(t)$ e $G_i(t)$ sono processi aleatori costanti nell'intervallo di estremi t_1 e t_2, ossia $F_i(t) = F_i$ e $G_i(t) = G_i$, dove F_i e G_i sono variabili aleatorie $\mathcal{F}_{t_1}$-misurabili. Infatti, una volta ottenuta la validità della formula per il differenziale stocastico del prodotto in questo caso, si ottiene immediatamente la validità per il caso in cui $F_i(t)$ e $G_i(t)$ sono processi aleatori elementari (in quanto l'integrale stocastico è lineare). Ed infine il caso generale si ottiene passando al limite, considerando successioni di processi elementari $F_i^{(n)}(t)$ e $G_i^{(n)}(t)$ che convergono a $F_i(t)$ e $G_i(t)$, nel senso che

$$\mathbb{P}\left(\int_0^T |F_i^{(n)}(t) - F_i(t)|\, dt \to 0\right) = 1$$

$$\mathbb{P}\left(\int_0^T |G_i^{(n)}(t) - G_i(t)|^2\, dt \to 0\right) = 1.$$

Basta infine considerare che i processi $X_i^{(n)}(t)$, definiti con differenziale stocastico $dX_i^{(n)}(t) = F_i^{(n)}(t)dt + G_i^{(n)}(t)dW_t$, convergono ai processi $X_i(t)$ nel senso che converge in probabilità a zero $\sup_{t\in[0,T]} |X_i^{(n)}(t) - X(t)|$.

Torniamo al caso dei processi con $F_i(t) = F_i$ e $G_i(t) = G_i$. Per semplicità di notazione prenderemo $t_1 = 0$ e $t_2 = t$, di modo che

$$X_i(t) = X_i(0) + F_i\, t + G_i\, W_t,$$

con $X_i(0)$, F_i e G_i variabili aleatorie $\mathcal{F}_0$-misurabili. Si tratta allora di controllare

$$X_1(t)\, X_2(t) - X_1(0)\, X_2(0) := \big(X_1(0) + F_1\, t + G_1\, W_t\big)\, \big(X_2(0) + F_2\, t + G_2\, W_t\big) - X_1(0)\, X_2(0),$$

coincide con

$$\begin{aligned} &\int_0^t X_1(u)[F_2\, du + G_2\, dW_u] + \int_0^t X_2(u)[F_1\, du + G_1\, dW_u] + G_1\, G_2\, t \\ &= \int_0^t [X_1(0) + F_1\, u + G_1\, W_u]\, [F_2\, du + G_2\, dW_u] + \int_0^t [X_2(0) + F_2\, u + G_2\, W_u]\, [F_1\, du + G_1\, dW_u] + G_1\, G_2\, t \end{aligned}$$

Considerando che le variabili aleatorie $X_i(0)$, F_i e G_i sono $\mathcal{F}_0$-misurabili, e quindi possono essere portate fuori dall'integrale stocastico, è facile convincersi (con un poco di pazienza) che la dimostrazione si basa sul fatto che

$$t^2 = 2\int_0^t s\, ds, \qquad t\, W_t = \int_0^t s\, dW_s + \int_0^t W_s\, ds, \qquad \text{e infine che } W_t^2 = 2\int_0^t W_s\, dW_s + t.$$

È interessante notare che nella seconda uguaglianza si è usata la formula di integrazione per parti (6.9) (nella dimostrazione della Proposizione 6.2) nel caso in cui l'integrando è una funzione deterministica e C^1. Va infine notato che per ottenere la formula di integrazione per parti si usa solo il fatto che il processo di Wiener ammette traiettorie continue. Invece l'ultima uguaglianza (si veda l'Esempio 6.1) vale in quanto la variazione quadratica del processo di Wiener tende a t (Lemma 4.4).

□

Siano f una funzione regolare di (t, x) e X un processo di Itô, con differenziale stocastico (6.2). Ci chiediamo quale sia il differenziale stocastico del processo $(Y_t = f(t, X_t))_t$. La risposta è fornita dalla formula di Itô.

Teorema 6.8 (Formula di Itô)**.** *Sia*

$$dX_t = F_t dt + G_t dW_t,$$

e sia $f(t, x)$ una funzione continua, con derivate parziali prime f_t ed f_x continue, e con la derivata parziale seconda f_{xx} continua, allora

$$\begin{aligned} dY_t = df(t, X_t) &= f_t(t, X_t)dt + f_x(t, X_t)dX_t + \frac{1}{2} f_{xx}(t, X_t)G_t^2 dt \\ &= f_t(t, X_t)dt + f_x(t, X_t)F_t dt + f_x(t, X_t)G_t dW_t + \frac{1}{2} f_{xx}(t, X_t)G_t^2 dt. \end{aligned}$$

In particolare per $X_t = W_t$ (ovvero per $F_t = 0$ e $G_t = 1$) si ha

$$df(t, W_t) = f_t(t, W_t)dt + f_x(t, W_t)dW_t + \frac{1}{2} f_{xx}(t, W_t)dt.$$

In realtà il Teorema 6.8 sopra enunciato non è completo in quanto mancano le ipotesi che permettono di assicurare che abbia senso, ad esempio, l'integrale stocastico di $f_x(t, X_t)G_t$ rispetto a dW_t. Un'ipotesi sufficiente è che le derivate siano limitate, tuttavia non è un'ipotesi necessaria.

Uno schema della dimostrazione si ottiene scrivendo per ogni partizione di $[t', t'']$

$$f(t'', X_{t''}) - f(t', X_{t'}) = \sum_k \left[f(t_k, X_{t_k}) - f(t_{k-1}, X_{t_{k-1}}) \right]$$

e utilizzando la formula di Taylor[17]

$$f(t, x) = f(t_0, x_0) + f_t(t_0, x_0)(t - t_0) + f_x(t_0, x_0)(x - x_0) + \frac{1}{2} f_{xx}(t_0, x_0)(x - x_0)^2 + o(t - t_0) + o\big((x - x_0)^2\big)$$

e tenendo conto del fatto che

$$\begin{aligned} (X_{t_k} - X_{t_{k-1}})^2 &\simeq \big(F_{t_{k-1}}(t_k - t_{k-1}) + G_{t_{k-1}}(W_{t_k} - W_{t_{k-1}})\big)^2 \\ &= F_{t_{k-1}}^2 (t_k - t_{k-1})^2 + 2F_{t_{k-1}}G_{t_{k-1}}(t_k - t_{k-1})(W_{t_k} - W_{t_{k-1}}) + G_{t_{k-1}}^2 (W_{t_k} - W_{t_{k-1}})^2 \\ &= O\big((t_k - t_{k-1})^2\big) + o\big((t_k - t_{k-1})\big) + G_{t_{k-1}}^2 (W_{t_k} - W_{t_{k-1}})^2 \end{aligned}$$

con $(W_{t_k} - W_{t_{k-1}})^2$ che si comporta come $(t_k - t_{k-1})$ (si riveda la sezione in cui si dimostra che il processo di Wiener non ha traiettorie a variazione limitata)

Esempio 6.2. *Sia $f(x)$ una funzione di classe C^1 e sia $F(x) = \int_0^x f(y)\, dy$ allora la formula di Itô, applicata alla funzione $F(x)$ nel caso in cui $X_t = W_t$ ci assicura che*

$$F(W_t) - F(W_s) = \int_s^t f(W_u)\, dW_u + \frac{1}{2} \int_s^t f'(W_u)\, du.$$

[17] Veramente mancano i termini di ordine $O\big((t - t_0)(x - x_0)\big)$. Tuttavia successivamente, se si prende $t = t_k$, $t_0 = t_{k-1}$, $x = W_{t_k}$ e $x_0 = W_{t_{k-1}}$, si ha che $|(t - t_0)(x - x_0)| = |t_k - t_{k-1}| \cdot |W_{t_k} - W_{t_{k-1}}|$, che è un infinitesimo di ordine superiore a $|t_k - t_{k-1}|$ in quanto il processo di Wiener è a traiettorie continue. Questo è un punto fondamentale, perché mostra come per ottenere la formula di Itô serva la continuità delle traiettorie, oltre al fatto che il processo di Wiener sia una martingala con variazione quadratica uguale a t.

L'interesse di questa formula sta nel fatto che possiamo calcolare l'integrale stocastico $\int_s^t f(W_u)\,dW_u$ *usando solo l'integrale ordinario, in quanto possiamo riscrive l'uguaglianza precedente come segue*

$$\int_s^t f(W_u)\,dW_u = F(W_t) - F(W_s) - \frac{1}{2}\int_s^t f'(W_u)\,du.$$

**Un altro schema di dimostrazione della formula di Itô si basa sull'idea di dimostrare, attraverso la formula del differenziale del prodotto, che la formula di Itô vale per ogni $f(t,x) = x^n$, e quindi, per linearità, per i polinomi. Poi per ogni funzione del tipo $f(t,x) = a(t)\,p(x)$ con $a(\cdot)$ una funzione C^1 e $p(\cdot)$ un polinomio, e quindi per ogni funzione $f(t,x)$ che si possa ottenere come limite di funzioni del tipo $f_n(t,x) = \sum_{k=0}^{m_n} a_k(t)\,p_k(x)$. **Tali funzioni si possono ottenere, grazie al Teorema di Stone Weierstrass e i limiti sono uniformi sui compatti.

Il precedente Teorema 6.8 ammette diverse generalizzazioni, come ad esempio la seguente.

Teorema 6.9 (Formula di Itô multidimensionale)**.** *Siano X_i, $i = 1,...,m$, processi che ammettono differenziale stocastico*

$$dX_i(t) = F_i(t)dt + G_i(t)dW_t \qquad i = 1,...,m$$

e sia $f : \mathbb{R}^+ \times \mathbb{R}^m \to \mathbb{R}$ una funzione continua in (t,x), derivabile con continuità una volta in t e due volte in x. Allora posto $X_t = (X_1(t),...,X_m(t))$, il processo $(f(t,X_t))_t$ ammette differenziale stocastico

$$\begin{aligned} df(t,X_t) &= \left(f_t(t,X_t) + \sum_{i=1}^m f_{x_i}(t,X_t)F_i(t) + \frac{1}{2}\sum_{i,j=1}^m f_{x_ix_j}(t,X_t)G_i(t)G_j(t) \right) dt \\ &+ \sum_{i=1}^m f_{x_i}(t,X_t)G_i(t)dW_t \\ &= \left(f_t(t,X_t) + \sum_{i=1}^m f_{x_i}(t,X_t)dX_i(t) + \frac{1}{2}\sum_{i,j=1}^m f_{x_ix_j}(t,X_t)G_i(t)G_j(t) \right) dt. \end{aligned}$$

In presenza di un processo di Wiener a dimensione d, ossia in presenza di d processi di Wiener indipendenti $W_t^{(1)}$, ... ,$W_t^{(d)}$ si può considerare anche il caso in cui i processi X_i, $i = 1,...,m$ ammettono differenziale stocastico definito come segue

$$dX_i(t) = F_i(t)dt + \sum_{k=1}^d G_i^{(k)}(t)dW_t^{(k)} \qquad i = 1,...,m$$

Allora il processo $(f(t,X_t))_t$ ammette differenziale stocastico

$$df(t,X_t) = \left(f_t(t,X_t) + \sum_{i=1}^m f_{x_i}(t,X_t)F_i(t) + \frac{1}{2}\sum_{i,j=1}^m f_{x_ix_j}(t,X_t)\sum_{k=1}^d G_i^{(k)}(t)G_j^{(k)}(t) \right) dt \tag{6.12}$$

$$+ \sum_{i=1}^m f_{x_i}(t,X_t)G_i^{(k)}(t)dW_t^{(k)} \tag{6.13}$$

$$= \left(f_t(t,X_t) + \sum_{i=1}^m f_{x_i}(t,X_t)dX_i(t) + \frac{1}{2}\sum_{i,j=1}^m f_{x_ix_j}(t,X_t)\sum_{k=1}^d G_i^{(k)}(t)G_j^{(k)}(t) \right) dt \tag{6.14}$$

$$= \left(f_t(t,X_t) + \sum_{i=1}^m f_{x_i}(t,X_t)dX_i(t) + \frac{1}{2}\sum_{i,j=1}^m f_{x_ix_j}(t,X_t) ** (G(t)'G(t))_{i,j} \right) dt. \tag{6.15}$$

dove abbiamo indicato con $G(t)'$ la matrice trasposta della matrice che al posto di indice (k,i) ha il valore $G_i^{(k)}(t)$.

Dimostrazione. Per la prima formula, nel caso di un unico processo di Wiener, una dimostrazione formale si basa sul considerare prima il caso dei polinomi, dimostrando la formula per induzione e utilizzando la formula del differenziale del prodotto. Per ottenere la formula nel caso generale si tratta di considerare ***(i)*** il fatto che i polinomi approssimano

uniformemente sui compatti le funzioni regolari, insieme alle loro derivate, ***(ii)*** il fatto che ci si può ridurre al caso in cui X_t rimane in un compatto, per ogni t, introducendo opportuni tempi d'arresto $\tau_M = \inf\{s\, t.c. |X_s| \geq M\}$.

Per la seconda formula il procedimento è lo stesso, ma bisogna utilizzare una formula per il differenziale del prodotto nel caso in cui ci siano diversi processi di Wiener indipendenti: **se

$$dX_i(t) = F_i(t)dt + \sum_{k=1}^{d} G_i^{(k)}(t)dW_t^{(k)}, \qquad dX_j(t) = F_j(t)dt + \sum_{k=1}^{d} G_j^{(k)}(t)dW_t^{(k)},$$

allora

$$d(X_i\, X_j)(t) = X_i(t)dX_j(t) + X_j(t)dX_i(t) + \sum_{k=1}^{d} G_i^{(k)}(t)G_j^{(k)}(t)\, dt. \tag{6.16}$$

Si tratta di una generalizzazione della Proposizione 6.7. Anche in questo caso la dimostrazione è simile a quella nel caso unidimensionale, ma bisogna utilizzare il fatto che, **poiché per $h \neq k$ i processi di Wiener $(W_t^{(h)})$ e $(W_t^{(k)})$ sono indipendenti, allora

$$2\, W_t^{(h)}\, W_t^{(k)} = \int_0^t W_s^{(h)}\, dW_s^{(k)} + \int_0^t W_s^{(k)}\, dW_s^{(h)}.$$

** □

Esercizio 6.1. *Dati due processi di Wiener $W_t^{(1)}$ e $W_t^{(2)}$ indipendenti, dimostrare che*

$$\tilde{W}_t^{(1)} = \frac{1}{\sqrt{2}}\,[W_t^{(1)} + W_t^{(2)}], \qquad \tilde{W}_t^{(2)} = \frac{1}{\sqrt{2}}\,[W_t^{(1)} - W_t^{(2)}],$$

sono processi di Wiener indipendenti. Dimostrare che

$$2\, W_t^{(1)}\, W_t^{(2)} = \int_0^t W_s^{(1)}\, dW_s^{(2)} + \int_0^t W_s^{(2)}\, dW_s^{(1)}, **$$

(suggerimento: si consideri $\int_0^t \tilde{W}_s^{(1)}\, d\tilde{W}_s^{(1)}$ e lo si calcoli utilizzando sia il fatto che è un processo di Wiener, sia riscrivendolo come somma di integrali stocastici)

Osservazione 6.5. *Se indichiamo con f' il gradiente di f rispetto a x. La formula di Itô si può scrivere in maniera più compatta:*

$$df(t, X_t) = f_t(t, X_t)dt + f'(t, X_t)dX(t) + \frac{1}{2}\sum_{i,j=1}^{m} f_{x_i x_j}(t, X_t)d < X_i, X_j >_t,$$

***dove $d < X_i, X_j >_t = \sum_{k=1}^{d} G_i^{(k)}(t)G_j^{(k)}(t)\, dt$ (che compare nella formula (6.16) del differenziale del prodotto di $X_i(t)$ con $X_j(t)$.*

Il differenziale stocastico si comporta, dunque, in maniera diversa da quello ordinario per la presenza dell'ultimo termine a destra.

6.2.1 Moto browniano geometrico e il suo differenziale stocastico

Ricordiamo che un ***moto browniano geometrico*** è un processo $(S_t)_t$ definito da

$$S_t = s_0 \exp\left\{\sigma W_t + \left(\mu - \frac{1}{2}\sigma^2\right) t\right\}$$

dove μ e σ sono costanti, con $\sigma \neq 0$.
Definiamo

$$f(t, x) = s_0 \exp\left\{\sigma x + \left(\mu - \frac{1}{2}\sigma^2\right) t\right\}$$

così

$$S_t = f(t, W_t).$$

Allora

$$f_t(t,x) = \left(\mu - \frac{1}{2}\sigma^2\right) f(t,x), \qquad f_x(t,x) = \sigma f(t,x), \qquad f_{xx}(t,x) = \sigma^2 f(t,x).$$

Per la formula di Itô, applicata a $X_t = W_t$ (così $F_t = F_t^W = 0$ e $G_t = G_t^W = 1$**)

$$\begin{aligned} dS_t = df(t,W_t) &= f_t(t,W_t)dt + f_x(t,W_t)dW_t + \frac{1}{2}f_{xx}(t,W_t)dt \\ &= \left(\mu - \frac{1}{2}\sigma^2\right) f(t,W_t)dt + \sigma f(t,W_t)dW_t + \frac{1}{2}\sigma^2 f(t,W_t)dt \\ &= \mu S_t dt + \sigma S_t dW_t. \end{aligned}$$

Così il moto browniano geometrico in forma differenziale è dato da

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

e nella sua forma integrale

$$S_t = s_0 + \int_0^t \mu S_u du + \int_0^t \sigma S_u dW_u.$$

In particolare quindi

$$\begin{aligned} \mathbb{E}[S_t] &= s_0 + \mathbb{E}[\int_0^t \mu S_u du] + \mathbb{E}[\int_0^t \sigma S_u dW_u] \\ &= s_0 + \int_0^t \mu \mathbb{E}[S_u] du, \end{aligned}$$

da cui

$$\mathbb{E}[S_t] = s_0 \exp\{\mu t\}.$$

Riutilizzando ancora la formula di Itô per $X_t = S_t$ (così $F_t = F_t^S = \mu S_t$ e $G_t = G_t^S = \sigma S_t$) per $f(t,x) = x^2$ (per cui $f_t = 0$, $f_x = 2x$ ed $f_{xx} = 2$) si ottiene [18]

$$\begin{aligned} dS_t^2 &= 2S_t dS_t + \frac{1}{2}2\sigma^2 S_t^2 dt \\ &= 2\mu S_t^2 dt + 2\sigma S_t^2 dW_t + \sigma^2 S_t^2 dt \\ &= (2\mu + \sigma^2) S_t^2 dt + 2\sigma S_t^2 dW_t. \end{aligned}$$

Integrando quest'ultima espressione, e prendendo i valori attesi (tralasciando il problema di mostrare che l'integrale stocastico è effettivamente una martingala a media nulla) si ottiene che

$$\begin{aligned} \mathbb{E}[S_t^2] &= s_0^2 + \mathbb{E}[\int_0^t (2\mu + \sigma^2) S_u^2 du] + 2\sigma \mathbb{E}[\int_0^t S_u^2 dW_u] \\ &= s_0^2 + \int_0^t (2\mu + \sigma^2) \mathbb{E}[S_u^2] du, \end{aligned}$$

da cui

$$\mathbb{E}[S_t^2] = s_0^2 \exp\{(2\mu + \sigma^2)t\}$$

[18] Per ottenere il differenziale di S_t^2 si può considerare che

$$\begin{aligned} df(t,S_t) &= f_t(t,S_t)dt + f_x(t,S_t)dS_t + \frac{1}{2}f_{xx}(t,S_t)(G^S)_t^2 dt \\ &= 2S_t\left(\mu S_t dt + \sigma S_t dW_t\right) + \sigma^2 S_t^2 dt \end{aligned}$$

oppure si può considerare che $S_t^2 = \tilde{f}(t,W_t)$ con

$$\tilde{f}(t,x) = s_0^2 \exp\left\{2\sigma x + \left(2\mu - \sigma^2\right) t\right\}$$

e applicare la formula di Itô al processo $X_t = W_t$. Si consiglia il lettore di controllare che si ottiene lo stesso risultato con questo metodo.

e quindi si può ricavare facilmente la varianza.[19] Il moto browniano geometrico è un processo da prendere in considerazione per modellizzare l'evoluzione di quantità che devono sempre rimanere positive. Anche per questo motivo, è un modello usato per descrivere l'evoluzione dei prezzi nei mercati finanziari. Nel modello di Black-Scholes, come già visto, il titolo rischioso segue un moto browniano geometrico.

6.2.2 Funzioni armoniche e tempi di uscita per il processo di Wiener

Consideriamo qui un processo di Wiener d-dimensionale, ossia un vettore $W_t = (W_t^{(1)}, \cdots, W_t^{(d)})$, con $W_t^{(k)}$ processi di Wiener indipendenti. La formula di Itô (6.12), applicata a una funzione $f(t,x)$, che appartiene a $C^{1,2}$, e al processo W_t, per il quale $m = d$, $F_i(t) = 0$, $G_i^{(k)}(t) = 0$ per $k \neq i$ e $G_i^{(i)}(t) = 1$ (ossia la matrice $G(t)$ è la matrice identità I), diviene, tenendo conto che $(G(t)'G(t)) = I$

$$df(t, W_t) = f_t(t, W_t)\, dt + \sum_{i=1}^{m} f_{x_i}(t, W_t) dW_t^{(i)} + \frac{1}{2} \sum_{i=1}^{m} f_{x_i x_i}(t, W_t)\, dt \tag{6.17}$$

che possiamo riscrivere, in modo rapido come, utilizzando la simbologia del gradiente $\nabla f(x) = {}^{**}\big(f_{x_1}(x), \cdots, f_{x_d}(x)\big)$ e del Laplaciano $\triangle f(x)\big(= \nabla \cdot \nabla f(x)\big) = \sum_{i=1}^{m} f_{x_i x_i}(x)$, e scrivendo $dW_t = (dW_t^{(1)}, \cdots, dW_t^{(d)})$

$$= f_t(t, X_t)\, dt + \nabla f(t, W_t) \cdot dW_t + \frac{1}{2} \triangle f(t, X_t) dt. \tag{6.18}$$

Osservazione 6.6. *Si osservi che il precedente risultato vale anche se, invece di considerare un processo di Wiener che parte da* 0 *si considera un processo di Wiener che parte da* x*, ossia se* $W_t = x + B_t$*, dove* B_t *è un processo di Wiener con* $B_0 = 0$*.*

Questa formula permette di ottenere il seguente risultato:

Proposizione 6.10. *Sia* $f(t,x)$ *una funzione che appartiene a* $C^{1,2}$ *e per la quale valga*

$$f_t(t,x) = -\frac{1}{2} \triangle f(t,x),$$

allora il processo $f(t, W_t)$ *è una martingala locale. Se inoltre*

$$\mathbb{E}\left[\int_0^T |f_{x_i}(t, W_t)|^2\, dt\right] < \infty, \qquad i = 1, \cdots, d$$

allora il processo $f(t, W_t)$ *è una martingala per* $t \in [0, T]$*.*

6.2.3 Applicazione ai tempi di uscita da una striscia per il processo di Wiener con drift

Sia ${}^{**}X_t (= X_t^x) = x + \mu\, t + \sigma\, W_t$ e posto, come nella Sezione 5.4, τ il tempo di prima uscita da una striscia

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

vogliamo calcolare di nuovo la probabilità $p \equiv p(x, a, b) := \mathbb{P}(X_\tau^x = -a)$ **che il processo esca dalla striscia passando per $-a$.

A volte si preferisce usare la seguente notazione: x_t è il processo canonico sullo spazio delle funzioni continue, $\tau(x(\cdot)) = \inf\{t > 0 : x(t) \notin (-a, b)\}$ e $\mathbb{P}_x$ è la legge del processo X_t^x e allora la probabilità di uscire dalla striscia $(-a, b)$ viene denotata come $\mathbb{P}_x(x(\tau) = -a)$. Situazioni simili sono state incontrate in precedenza[20].

[19] Ovviamente la $Var(S_t)$ si può ricavare anche a direttamente, si consiglia il lettore di eseguire questo calcolo e controllare che i due metodi portano allo stesso risultato, come deve essere.

[20] **Si vedano
(i) l'osservazione 6.6 precedente, in cui si considera $W_t = x + B_t$, dove B_t è un processo di Wiener con $B_0 = 0$;
(ii) la nota 6 **della sezione 5.4.

L'idea per ottenere tale probabilità è la seguente: se troviamo una funzione h continua, per la quale il processo $M_t := h(X_t)$ è una martingala, allora anche $M_{t\wedge\tau}$ è una martingala, limitata (dal $\max\{|h(x)|;\ x \in [-a,b]\}$) e quindi, passando al limite nella uguaglianza $\mathbb{E}_x[M_{t\wedge\tau}] = \mathbb{E}_x[M_0] = h(x)$, si ottiene che

$$\mathbb{E}_x[M_\tau] = \mathbb{E}_x[M_0] = h(x).$$

Allora, tenendo conto che

$$\mathbb{E}_x[M_\tau] = \mathbb{E}_x[h(X_\tau)] = h(-a)\,\mathbb{P}_x(X_\tau = -a) + h(b)\,\mathbb{P}_x(X_\tau = b),$$

se aggiungiamo la condizione che $h(-a) = 1$ e $h(b) = 0$, si ottiene che

$$\mathbb{E}_x[M_\tau] = \mathbb{E}_x[h(X_\tau)] = \mathbb{P}_x(X_\tau = -a) = h(x).$$

Poniamo ora $f(t,y) = h(\mu\, t + \sigma\, y)$. Allora la condizione che $h(X_t)$ sia una martingala diviene, per la Proposizione 6.10 precedente

$$f_t(t,y) + \frac{1}{2}\, f_{yy}(t,y) = 0 \qquad \Leftrightarrow \qquad \mu\, h'(\mu\, t + \sigma\, y) + \frac{1}{2}\,\sigma^2\, h''(\mu\, t + \sigma\, y) = 0$$

ossia, otteniamo l'equazione per h

$$\mu\, h'(x) + \frac{1}{2}\,\sigma^2\, h''(x) = 0 \qquad \Leftrightarrow \qquad h''(x) = -\frac{2\,\mu}{\sigma^2}\, h'(x),$$

da cui

$$h'(x) = C_1\, e^{-\frac{2\,\mu}{\sigma^2}\, x}, \qquad h(x) = C_0 + C_1 \int_0^x e^{-\frac{2\,\mu}{\sigma^2}\, z}\, dz = \widetilde{C}_0 + \widetilde{C}_1\, e^{-\frac{2\,\mu}{\sigma^2}\, x},$$

**per opportune costanti, da determinare in modo che valgano le condizioni al bordo

$$h(-a) = 1, \quad h(b) = 0. \qquad **$$

Tutte queste funzioni godono della proprietà di integrabilità richiesta dalla Proposizione 6.10, per ogni $T > 0$. Le condizioni ai bordi sono chiaramente soddisfatte se si prende

$$h(x) = \frac{e^{-\frac{2\,\mu}{\sigma^2}\, b} - e^{-\frac{2\,\mu}{\sigma^2}\, x}}{e^{-\frac{2\,\mu}{\sigma^2}\, b} - e^{-\frac{2\,\mu}{\sigma^2}\, (-a)}},$$

che quindi è la probabilità cercata $\mathbb{P}_x(X_{\tau(-a,b)} = -a)$ (si confronti con quanto ottenuto nella Sezione 5.4). Ovviamente si ottiene anche che

$$\mathbb{P}_x(X_{\tau(-a,b)} = b) = \frac{e^{-\frac{2\,\mu}{\sigma^2}\, (-a)} - e^{-\frac{2\,\mu}{\sigma^2}\, x}}{e^{-\frac{2\,\mu}{\sigma^2}\, (-a)} - e^{-\frac{2\,\mu}{\sigma^2}\, b}} = \frac{e^{\frac{2\,\mu}{\sigma^2}\, a} - e^{-\frac{2\,\mu}{\sigma^2}\, x}}{e^{\frac{2\,\mu}{\sigma^2}\, a} - e^{-\frac{2\,\mu}{\sigma^2}\, b}} \tag{6.19}$$

Esempio 6.3 (Distribuzione dell'estremo superiore di un processo di Wiener con drift)**.** *Possiamo applicare il risultato precedente* (6.19) *mandando a ad infinito si ottiene che, nel caso $x = 0$*

$$\lim_{a\to\infty} \mathbb{P}(X_{\tau(-a,b)} = b) = \mathbb{P}(X_{\tau(-\infty,b)} = b) = \mathbb{P}(\sup_{t\geq 0} X_t \geq b) = \begin{cases} 1 & \text{se } \mu > 0 \\ e^{\frac{2\,\mu}{\sigma^2}\, b} = e^{-\frac{2\,|\mu|}{\sigma^2}\, b} & \text{se } \mu < 0 \end{cases}$$

*Nel caso di un processo di Wiener con drift negativo, il calcolo precedente assicura che la variabile aleatoria **$X^* := \sup_{t\geq 0} X_t$ è esponenziale di parametro $\frac{2\,|\mu|}{\sigma^2}$ **(cioè $\mathbb{P}(X^* > y) = e^{-\lambda y}$, per $y > 0$ e $\lambda = \frac{2\,|\mu|}{\sigma^2}$), mentre nel caso in cui il drift sia positivo si ottiene che $X^* := \sup_{t\geq 0} X_t$ è una variabile aleatoria (generalizzata e degenere) che vale $+\infty$.*

Osservazione 6.7. *In realtà, nella derivazione delle probabilità $\mathbb{P}_x(X_{\tau(-a,b)} = -a)$ e $\mathbb{P}_x(X_{\tau(-a,b)} = b)$, avremmo dovuto controllare che il tempo $\tau = \tau(-a,b)$ di uscita dalla striscia fosse finito per essere sicuri che $M_\tau = h(X_\tau)$ avesse senso e per poter si calcolare il valore atteso come $h(-a)\,\mathbb{P}_x(X_\tau = -a) + h(b)\,\mathbb{P}_x(X_\tau = b)$.*

A questo scopo possiamo osservare che addirittura si può calcolare il valore atteso di τ, come si può dimostrare utilizzando le seguenti Proposizioni 6.11 e 6.16, ma lo faremo dopo aver enunciato e dimostrato queste due proposizioni.

Proposizione 6.11. *Se esiste una funzione $\phi \in C^2$, che sia non negativa nell'intervallo $[-a, b]$ per la quale*

$$\phi'(x)\,\mu + \frac{1}{2}\,\phi''(x)\,\sigma^2 \le -1,$$

*allora $\tau = \tau(-a, b)$, **il tempo di prima uscita da $(-a, b)$, ha valore atteso **finito, e quindi τ è finito con probabilità 1.*

Dimostrazione. L'idea è quella di dimostrare che esiste un numero M tale che, per ogni t si ha $\mathbb{E}_x[t \wedge \tau] \le M$.

Allora per il Teorema della convergenza monotona

$$\mathbb{E}_x[\tau] = \mathbb{E}_x\left[\lim_{t\to\infty} t \wedge \tau\right] = \lim_{t\to\infty} \mathbb{E}_x[t \wedge \tau] \le M$$

Si consideri che, applicando la formula di Itô al processo di Wiener con drift X_t, per una funzione $f(x) \in C^2$, si ottiene che

$$df(X_t) = f'(X_t)\,\mu\,dt + f'(X_t)\,\sigma\,dW_t + \frac{1}{2}\,f''(X_t)\,\sigma^2\,dt.$$

Prendendo $f = \phi$ e calcolando l'integrale tra 0 e $t \wedge \tau$ si ottiene

$$\phi(X_{t\wedge\tau}) = \phi(X_0) + \int_0^{t\wedge\tau} \left[\mu\,\phi'(X_s) + \frac{1}{2}\,\phi''(X_s)\,\sigma^2\right] ds + \int_0^{t\wedge\tau} \sigma\,\phi'(X_s)\,dW_s$$

da cui

$$\phi(X_{t\wedge\tau}) - \phi(X_0) = \int_0^{t\wedge\tau} \left[\mu\,\phi'(X_s) + \frac{1}{2}\,\phi''(X_s)\,\sigma^2\right] ds + \int_0^{t\wedge\tau} \sigma\,\phi'(X_s)\,dW_s \le -t \wedge \tau + \int_0^{t\wedge\tau} \sigma\,\phi'(X_s)\,dW_s$$

Il valore atteso di

$$\int_0^{t\wedge\tau} \sigma\,\phi'(X_s)\,dW_s = \int_0^t \mathbb{I}_{\{s\le\tau\}}\sigma\,\phi'(X_s)\,dW_s$$

è nullo, infatti

$$\mathbb{E}_x\left[\int_0^t \mathbb{I}_{\{s\le\tau\}}\sigma^2\,|\phi'(X_s)|^2\,ds\right] \le t\,{**}\,\sigma^2 \max_{x\in[-a,b]} |\phi'(x)|^2 < \infty$$

Quindi

$$\mathbb{E}_x[\phi(X_{t\wedge\tau}) - \phi(X_0)] \le -\mathbb{E}_x[t \wedge \tau] \qquad \Leftrightarrow \qquad \mathbb{E}_x[\phi(X_{t\wedge\tau})] - \phi(x) \le -\mathbb{E}_x[t \wedge \tau]$$

e quindi

$$\mathbb{E}_x[t \wedge \tau] \le -\mathbb{E}_x[\phi(X_{t\wedge\tau})] + \phi(x) \le 2 \max_{y\in[-a,b]} \phi(y) =: M$$

□

In realtà si può fare di meglio.

Proposizione 6.12. *Se esiste una funzione $v \in C^2$ e tale che*

$$v'(x)\,\mu + \frac{1}{2}\,v''(x)\,\sigma^2 = -1, \qquad v(-a) = v(b) = 0,$$

*allora **$\tau = \tau(-a, b)$, il tempo di prima uscita da $(-a, b)$, ha valore atteso $v(x)$, ossia ***

$$\mathbb{E}_x[\tau] = v(x).$$

Dimostrazione. Procedendo in modo simile alla dimostrazione della precedente Proposizione 6.11 si ottiene

$$v(X_{t\wedge\tau}) - v(X_0) = \int_0^{t\wedge\tau} \left[\mu\,v'(X_s) + \frac{1}{2}\,v''(X_s)\,\sigma^2\right] ds + \int_0^{t\wedge\tau} \sigma\,v'(X_s)\,dW_s = -t \wedge \tau + \int_0^{t\wedge\tau} \sigma\,v'(X_s)\,dW_s$$

e quindi

$$\mathbb{E}_x[v(X_{t\wedge\tau})] - \phi(x) = -\mathbb{E}_x[t \wedge \tau].$$

Poiché dalla Proposizione 6.11 sappiamo che τ è finito, allora $X_{t\wedge\tau}$ tende a X_τ e, per la continuità di v si ha $v(X_{t\wedge\tau})$ converge a $v(X_\tau) = 0$. Inoltre $|v(X_{t\wedge\tau})| \le \max_{y\in[-a,b]} |v(y)|$ e quindi

$$\mathbb{E}_x[\tau] = \lim_{t\to\infty} \mathbb{E}_x[t \wedge \tau] = \phi(x) - \lim_{t\to\infty} \mathbb{E}_x[v(X_{t\wedge\tau})] = \phi(x).$$

□

A questo punto non rimane che trovare la soluzione dell'equazione $v'(x)\,\mu + \frac{1}{2}\,v''(x)\,\sigma^2 = -1$, ossia

$$v''(x) = -\frac{2\,\mu}{\sigma^2}\,v'(x) - \frac{2}{\sigma^2}$$

con le condizioni al bordo **

$$v(-a) = v(b) = 0.$$

**Ricordiamo che la soluzione generale dell'equazione lineare

$$v''(x) = -\alpha(x)\,v'(x) - \beta(x) \tag{6.20}$$

(nel nostro caso $\alpha(x) = \frac{2\mu}{\sigma^2}$ e $\beta(x) = \frac{2}{\sigma^2}$) è data da

$$\begin{aligned} v'(x) &= e^{-\int_{-a}^{x}\alpha(y)\,dy}\left(C_1 - \int_{-a}^{x}\beta(z)\,e^{\int_{-a}^{z}\alpha(y)\,dy}\,dz\right) \\ &= C_1\,e^{-\int_{-a}^{x}\alpha(y)\,dy} - \int_{-a}^{x}\beta(z)\,e^{-\int_{z}^{x}\alpha(y)\,dy}\,dz \qquad (6.21) \\ v(x) &= C_0 + C_1\int_{-a}^{x} e^{-\int_{-a}^{z}\alpha(y)\,dy}\,dz - \int_{-a}^{x}\int_{-a}^{w}\beta(z)\,e^{-\int_{z}^{w}\alpha(y)\,dy}\,dz\,dw \qquad (6.22) \end{aligned}$$

prendendo $C_0 = 0$ e $C_1 = \frac{\int_{-a}^{b}\int_{-a}^{w}\beta(z)\,e^{-\int_{z}^{w}\alpha(y)\,dy}\,dz\,dw}{\int_{-a}^{b}e^{-\int_{-a}^{z}\alpha(y)\,dy}\,dz}$ si ottiene la soluzione cercata, cioè che soddisfa le condizioni al bordo, ossia **

$$v(x) = \frac{\int_{-a}^{b}\int_{-a}^{w}\beta(z)\,e^{-\int_{z}^{w}\alpha(y)\,dy}\,dz\,dw}{\int_{-a}^{b}e^{-\int_{-a}^{z}\alpha(y)\,dy}\,dz}\int_{-a}^{x}e^{-\int_{-a}^{z}\alpha(y)\,dy}\,dz - \int_{-a}^{x}\int_{-a}^{w}\beta(z)\,e^{-\int_{z}^{w}\alpha(y)\,dy}\,dz\,dw$$

quindi, fissati a piacere x_0 e x_1 in $(-a,b)$, A_0, A_1, B_0, B_1 in $\mathbb{R}$, e posto

$$\varphi(x) = A_0 + A_1\int_{x_1}^{x}\exp\left\{-\int_{x_0}^{z}\alpha(y)\,dy\right\}dz, \quad \text{e} \quad \psi(w) = B_0 + B_1\int_{x_1}^{x}dw\int_{x_1}^{w}\beta(z)\exp\left\{-\int_{x_0}^{z}\alpha(y)\,dy\right\}dz,$$

la soluzione $v(x)$ si scrive

$$v(x) = \big(\psi(b) - \psi(-a)\big)\,\frac{\varphi(x) - \varphi(-a)}{\varphi(b) - \varphi(-a)} - \big(\psi(x) - \psi(-a)\big)$$

Nel nostro caso, prendendo $x_0 = x_1 = 0$ e $A_0 = A_1 = B_0 = B_1 = 0$, si ha

$$\varphi(x) = \int_0^x \exp\left\{-\int_0^z \frac{2\,\mu}{\sigma^2}\,dy\right\}dz = \frac{1 - e^{-\frac{2\,\mu}{\sigma^2}x}}{\frac{2\,\mu}{\sigma^2}},$$

e

$$\psi(x) = \frac{2}{\sigma^2}\int_0^x \varphi(w)dw = \frac{2}{\sigma^2}\,\frac{1}{\frac{2\,\mu}{\sigma^2}}\left[x - \frac{1 - e^{-\frac{2\,\mu}{\sigma^2}x}}{\frac{2\,\mu}{\sigma^2}}\right] = \frac{x}{\mu} - \frac{1}{\mu}\varphi(x).$$

e quindi

$$\begin{aligned} v(x) = &\left(\frac{b}{\mu} - \frac{(-a)}{\mu}\right)\frac{\varphi(x) - \varphi(-a)}{\varphi(b) - \varphi(-a)} - \left(\frac{x}{\mu} - \frac{(-a)}{\mu}\right) \\ &+ \frac{1}{\mu}\left[\big(\varphi(b) - \varphi(-a)\big)\,\frac{\varphi(x) - \varphi(-a)}{\varphi(b) - \varphi(-a)} - \big(\varphi(x) - \varphi(-a)\big)\right] \end{aligned}$$

da cui, essendo identicamente nulla l'espressione tra parentesi quadre nella seconda linea,

$$v(x) = \left(\frac{b}{\mu} - \frac{(-a)}{\mu}\right)\frac{e^{-\frac{2\,\mu}{\sigma^2}x} - e^{-\frac{2\,\mu}{\sigma^2}(-a)}}{e^{-\frac{2\,\mu}{\sigma^2}b} - e^{-\frac{2\,\mu}{\sigma^2}(-a)}} - \left(\frac{x}{\mu} - \frac{(-a)}{\mu}\right).$$

6.2.4 Applicazione: ricorrenza e transienza per il processo di Wiener

Esempio 6.4 (Il processo di Wiener in $\mathbb{R}^2$ è ricorrente)**.** *Applicheremo il risultato della Proposizione 6.10 al caso in cui*

$$f(x) = f(x_1, x_2) = \log(\sqrt{x_1^2 + x_2^2}) = \log(|x|),$$

che è una funzione armonica, cioè vale $^{**}\triangle f(x) = 0$*, o ad una sua trasformazione lineare, che è sempre armonica.*

Strettamente parlando questa funzione non è C^2*, e verifica* $^{**}\triangle f(x) = 0$*, solo per* $|x| > 0$ *e quindi non si può usare la formula di Itô. Tuttavia noi vogliamo usare questa funzione solo nell'insieme* $r \le |x| \le R$ *e possiamo tranquillamente pensare che sia regolare, in tutto* $\mathbb{R}^2$*. Si tratta allora di ripetere il ragionamento della Proposizione 6.10*

Sia x *tale che* $r < |x| < R$ *e supponiamo che* W_t *sia un processo di Wiener che parte da* x*. Siano* $\tau_r = \inf\{t > 0$ *tali che* $|W_t| \le r\}$ *e* $\tau_R = \inf\{t > 0$ *tali che* $|W_t| \ge R\}$*. Posto* $\tau = \tau_r \wedge \tau_R$ *possiamo affermare che*

$$\log(|W_{t\wedge\tau}|) = \log(|x|) + \int_0^{t\wedge\tau} \frac{1}{|W_s|^2 **} \big(W_t^{(1)}\, dW_t^{(1)} + W_t^{(2)}\, dW_t^{(2)}\big)$$

è una martingala. Posto

$$h(y) = \frac{\log(R) - \log(|y|)}{\log(R) - \log(r)}$$

anche

$$h(W_{t\wedge\tau}) = \frac{\log(R) - \log(|W_{t\wedge\tau}|)}{\log(R) - \log(r)}$$

è una martingala. Mandando t *all'infinito si ottiene che*

$$\mathbb{P}_x(\tau = \tau_r) = \frac{\log(R) - \log(|x|)}{\log(R) - \log(r)}. \tag{6.23}$$

Infatti

$$h(x) = \mathbb{E}_x[h(W_\tau)] = 1\,\mathbb{P}_x(\tau = \tau_r) + 0\,\mathbb{P}_x(\tau = \tau_R) = \mathbb{P}_x(\tau = \tau_r),$$

***in quanto* $h(y) = 1$ *se* $|y| = r$ *mentre* $h(y) = 0$ *se* $|y| = R$*.* **

Mandando R *ad infinito nella* (6.23)*, si ottiene che, partendo da* x*, con* $|x| > r$ *la probabilità di toccare, prima o poi, la sfera di raggio* r *vale* 1*. Questo fatto si esprime dicendo che il processo di Wiener è ricorrente.*

Esempio 6.5 (Il processo di Wiener in $\mathbb{R}^d$, per $d \ge 3$ è transiente)**.** *Nel caso* $d \ge 3$ *si può ripetere il ragionamento del precedente Esempio 6.4, a partire però della funzione armonica*

$$f(x) = \frac{1}{|x|^{d-2}} = |x|^{2-d}$$

Si arriva allora all'identità

$$\mathbb{P}_x(\tau = \tau_r) = \frac{R^{2-d} - |x|^{2-d}}{R^{2-d} - r^{2-d}}. \tag{6.24}$$

Mandando R *ad infinito nella* (6.24)*,* ***essendo* $2 - d < 0$*, si ottiene che, partendo da* x*, con* $|x| > r$ *la probabilità di toccare, prima o poi, la sfera di raggio* r *vale* $\frac{|x|^{2-d}}{r^{2-d}} = \frac{r^{d-2}}{|x|^{d-2}} < 1$*. Questo fatto si esprime dicendo che il processo di Wiener è transiente.*

6.3 Equazioni differenziali stocastiche

Introduciamo in questa sezione la nozione di equazione differenziale stocastica.
Siano $b(x,t) = (b_i(x,t))_{1\le i\le m}$ e $\sigma(x,t) = (\sigma_{ij}(x,t))_{1\le i\le m, 1\le j\le d}$ funzioni (congiuntamente) misurabili definite su $\mathbb{R}^m \times [0,T]$ a valori in $\mathbb{R}^m$ e in $M(m,d)$ rispettivamente, dove $M(m,d)$ indica l'insieme delle matrici $m \times d$.

Definizione 6.6. *Sia $(W_t)_t$ un moto browniano d-dimensionale standard. Sia $x \in \mathbb{R}^m$ ed $s \ge 0$. Diremo che il processo $(\xi_t)_{t\in[s,T]}$ è soluzione dell'equazione differenziale stocastica*

$$\begin{cases} d\xi_t = b(\xi_t,t)dt + \sigma(\xi_t,t)dW_t \\ \xi_s = x, \end{cases} \tag{6.25}$$

con dato iniziale (s,x) $x \in \mathbb{R}^m$, se, per ogni $t \in [s,T]$ si ha

$$\xi_t = x + \int_s^t b(\xi_u,u)du + \int_s^t \sigma(\xi_u,u)dW_u. \tag{6.26}$$

La definizione è analoga nel caso in cui invece di dato iniziale deterministico, si considera un dato iniziale aleatorio η, che sia $\mathcal{F}_s$ misurabile.

Osservazione 6.8. *Nella definizione precedente, si richiede implicitamente che $s \to b(\xi_s,s)$ e $s \to \sigma(\xi_s,s)$ siano processi per cui abbia senso fare i rispettivi integrali, deterministico e stocastico, insieme alle condizioni di progressiva misurabilità.*

Infine, ricordiamo che σ^2 viene chiamato ***coefficiente di diffusione*** e b viene chiamato ***drift*** (o ***deriva***).

Definizione 6.7. *Si dice che il sistema (6.25) ha* **soluzione forte** *se per ogni $(\Omega,\mathcal{F},\mathcal{F}_t,\mathbf{P})$ e per ogni moto browniano standard $W = (W_t)_t$, esiste un processo ξ tale che $(\xi_t)_t$ sia soluzione di (6.25).*

La nozione di soluzione forte va confrontata con la nozione di soluzione debole, che però daremo in seguito, nella sottosezione 6.3.6 (Definizione 6.8), per la quale, invece, né il moto browniano (rispetto al quale si fanno le integrazioni), né lo spazio di probabilità su cui esso è definito sono assegnati a priori.

6.3.1 Unicità per EDS: il caso Lipschitz globale

In questa sezione diamo l'enunciato del risultato che garantisce l'esistenza e l'unicità della soluzione per una equazione differenziale stocastica nelle ipotesi di lipschitzianità globale, ma diamo la dimostrazione solo dell'unicità. La dimostrazione dell'esistenza verrà data solo dopo aver esaminato alcuni semplici esempi di equazioni con soluzione esplicitamente calcolabile (Sezione 6.3.3).

Teorema 6.13. *Supponiamo che i coefficienti $b(t,x)$ e $\sigma(t,x)$, oltre alla condizione di congiunta misurabilità, soddisfino la seguente condizione di Lipschitz globale:*

$$|b(x,t) - b(\bar{x},t)|^2 + |\sigma(x,t) - \sigma(\bar{x},t)|^2 \le K|x-\bar{x}|^2, \qquad ** \tag{6.27}$$

e la seguente condizione di crescita

$$**|b(x,t)|^2 + |\sigma(x,t)|^2 \le K\left(1+|x|^2\right), \tag{6.28}$$

per una costante $K > 0$.

Allora l'equazione differenziale stocastica

$$dX_t = b(X_t,t)dt + \sigma(X_t,t)dW_t, \quad t \ge s, \qquad X_s = x \tag{6.29}$$

ammette una soluzione che è un processo progressivamente misurabile, con $\mathbb{E}\big[\int_s^T |X_u|^2 du\big] < \infty$, per ogni $T \ge s$. Inoltre esiste una sola soluzione con queste caratteristiche, nel senso che, se $\bar{X}_t$ è un'altra soluzione, allora $\mathbb{P}\big(X_t = \bar{X}_t,$ per ogni $t\big) = 1$. Nel seguito denoteremo tale soluzione con $X_t^{s,x}$.

La tesi del teorema rimane valida se invece di considerare il dato iniziale deterministico si considera il dato iniziale stocastico, di quadrato integrabile, ossia si considera

$$dX_t = b(X_t, t)dt + \sigma(X_t, t)dW_t, \quad t \geq s, \qquad\qquad X_s = \eta, \tag{6.30}$$

con η una variabile aleatoria di quadrato integrabile, $\mathcal{F}_s$-misurabile.

Dimostrazione. (del Teorema 6.13)
Unicità

Siano X_t e $\widetilde{X}_t$ due soluzioni di (6.29) (o di (6.29)), con la proprietà che $\mathbb{E}\big[\int_s^T |X_u|^2 du\big] < \infty$ e $\mathbb{E}\big[\int_s^T |\widetilde{X}_u|^2 du\big] < \infty$. Si consideri che

$$X_t - \widetilde{X}_t = \int_s^t [b(X_u, u) - b(\widetilde{X}_u, u)]\, du + \int_s^t [\sigma(X_u, u) - \sigma(\widetilde{X}_u, u)]\, dW_u$$

per cui, usando la disuguaglianza $(a+b)^2 \leq 2\,a^2 + 2\,b^2$, e la disuguaglianza di Cauchy

$$\begin{aligned} |X_t - \widetilde{X}_t|^2 &\leq 2\left(\int_s^t [b(X_u, u) - b(\widetilde{X}_u, u)]\, du\right)^2 + 2\left(\int_s^t [\sigma(X_u, u) - \sigma(\widetilde{X}_u, u)] dW_u\right)^2 \\ &\leq \int_s^t 1^2 du \int_s^t [b(X_u, u) - b(\widetilde{X}_u, u)]^2 du + 2\left(\int_s^t [\sigma(X_u, u) - \sigma(\widetilde{X}_u, u)] dW_u\right)^2 \end{aligned}$$

utilizzando la condizione di Lipschitz (6.27), si ha

$$\leq 2\,T \int_s^t K\, |X_u - \widetilde{X}_u|^2 du + 2\left(\int_s^t [\sigma(X_u, u) - \sigma(\widetilde{X}_u, u)] dW_u\right)^2$$

Passando ai valori attesi si ottiene che, per $t \leq T$

$$\mathbb{E}\big[|X_t - \widetilde{X}_t|^2\big] \leq 2\,T\,K \int_s^t \mathbb{E}\big[|X_u - \widetilde{X}_u|^2\big] du + 2\mathbb{E}\left[\left(\int_s^t [\sigma(X_u, u) - \sigma(\widetilde{X}_u, u)] dW_u\right)^2\right]$$

e quindi,per la isometria dell'integrale stocastico,

$$\mathbb{E}\big[|X_t - \widetilde{X}_t|^2\big] \leq T\,K \int_s^t \mathbb{E}\big[|X_u - \widetilde{X}_u|^2\big] du + 2 \int_s^t \mathbb{E}\left[[\sigma(X_u, u) - \sigma(\widetilde{X}_u, u)]^2\right] du$$

e usando di nuovo la condizione di Lipschitz (6.27), si ha

$$\mathbb{E}\big[|X_t - \widetilde{X}_t|^2\big] \leq 2\,T\,K \int_s^t \mathbb{E}\big[|X_u - \widetilde{X}_u|^2\big]\, du + 2\,K \int_s^t \mathbb{E}\big[|X_u - \widetilde{X}_u|^2\big]\, du = C \int_s^t \mathbb{E}\big[|X_u - \widetilde{X}_u|^2\big]\, du,$$

con $C := 2(T+1)\,K$.

A questo punto non rimane che applicare il Lemma di Gronwall[21] alla funzione $v(t) := \mathbb{E}\big[|X_t - \widetilde{X}_t|^2\big]$, che per

[21] Ricordiamo che il Lemma di Gronwall garantisce che se una funzione $v(t)$, misurabile, non negativa e integrabile ha la proprietà che

$$v(t) \leq C_0 + C_1 \int_0^t v(u)\, du,$$

allora si ha $v(t) \leq C_0\, e^{C_1\, t}$.
Una dimostrazione potrebbe essere la seguente: definire $w(t) = \int_0^t v(u)\, du$, e considerare la funzione $f(t) := e^{-C_1 t}\, w(t)$. Allora

$$f'(t) := -C_1\, e^{-C_1 t}\, w(t) + e^{-C_1 t}\, w'(t) \leq -C_1\, e^{-C_1 t}\, w(t) + e^{-C_1 t}\, \big(C_0 + C_1 w(t)\big) = C_0\, e^{-C_1 t}.$$

Di conseguenza

$$e^{-C_1 t}\, w(t) = f(t) = f(0) + \int_0^t f'(u)\, du \leq 0 + \int_0^t C_0\, e^{-C_1\, u}\, du = C_0\, \frac{1 - e^{-C_1 t}}{C_1} \qquad \Leftrightarrow \qquad w(t) \leq C_0\, \frac{e^{C_1 t} - 1}{C_1}$$

da cui

$$v(t) \leq C_0 + C_1 \int_0^t v(u)\, du, = C_0 + C_1\, w(t) \leq C_0 + C_1\, C_0\, \frac{e^{C_1 t} - 1}{C_1} = C_0\, e^{C_1 t}.$$

ipotesi è una funzione finita (ed ovviamente non negativa), ed ottenere che $\mathbb{E}\big[|X_t - \widetilde{X}_t|^2\big] = 0$.

Ciò garantisce che $\mathbb{P}[X_t = \widetilde{X}_t] = 1$, per ogni t. Da ciò discende che $\mathbb{P}[X_t = \widetilde{X}_t,$ per ogni t razionale$] = 1$ ed infine, tenendo conto che essendo X_t e $\widetilde{X}_t$ soluzioni di un'equazione differenziale stocastica devono avere traiettorie continue, si ottiene che $\mathbb{P}[X_t = \widetilde{X}_t,$ per ogni $t] = 1$.

Si noti che la condizione che si tratti di soluzioni di $M_2[s,T]$ è stata usata, oltre che per poter utilizzare il Lemma di Gronwall, anche per poter utilizzare l'isometria: poiché vale la condizione di crescita (6.28), si ha che

$$[\sigma(X_u,u) - \sigma(\widetilde{X}_u,u)]^2 \le 2\,|\sigma(X_u,u)|^2 + 2\,|\sigma(\widetilde{X}_u,u)|^2 \le 2\,K\left(1+|X_u|^2\right) + 2\,K\left(1+|\widetilde{X}_u|^2\right)$$

e ciò garantisce che $\mathbb{E}\Big[\int_s^T |X_u - \widetilde{X}_u|^2\,du\Big] < \infty$. □

6.3.2 Esempi di soluzioni di equazioni differenziali stocastiche: il processo di Orstein-Ulhenbeck

Tenendo conto della definizione di soluzione forte di equazione differenziale stocastica, grazie a quanto visto nella sottosezione 6.2.1, possiamo dire che il moto browniano geometrico è soluzione forte dell'equazione differenziale

$$d\xi_t = \mu\xi_t dt + \sigma\xi_t dW_t, \quad \xi_0 = s_0. \tag{6.31}$$

A titolo di esempio vediamo come si può dimostrare che il processo di Orstein-Ulhenbeck è soluzione dell'equazione differenziale

$$d\xi_t = -\lambda\xi_t dt + \sigma dW_t, \quad \xi_0 = x. \tag{6.32}$$

L'idea è quella di ripetere il metodo della variazione delle costanti per le equazioni differenziali ordinarie[22]

Si cerca quindi la soluzione del tipo $X_t = \eta_t \exp\{-\lambda t\}$, dove η_t è un processo che ammetta differenziale stocastico

$$d\eta_t = F_t dt + G_t dW_t, \quad ovvero \quad \eta_t = x + \int_0^t F_s ds + \int_0^t G_s dW_s,$$

con F_t e G_t da determinare, appunto in modo che il differenziale di $\eta_t \exp\{-\lambda t\}$ soddisfi l'equazione differenziale stocastica (6.32).Il processo $\exp\{-\lambda t\}$ è deterministico, di conseguenza,

$$d(\exp\{-\lambda t\}) = -\lambda \exp\{-\lambda t\} dt$$

per la formula del differenziale stocastico del prodotto si ottiene che

$$\begin{aligned} dX_t &= d(\exp\{-\lambda t\}\eta_t) = \exp\{-\lambda t\} d\eta_t + \eta_t d(\exp\{-\lambda t\}) \\ &= \exp\{-\lambda t\}(F_t dt + G_t dW_t) - \lambda \exp\{-\lambda t\}\eta_t dt \\ &= \exp\{-\lambda t\} F_t dt + \exp\{-\lambda t\} G_t dW_t - \lambda X_t dt. \end{aligned}$$

[22] Ricordiamo che per risolvere l'equazione differenziale ordinaria

$$\dot{y}_t = -\lambda y_t + v_t$$

si può considerare prima l'equazione

$$\dot{y}_t = -\lambda y_t,$$

la cui soluzione è data da

$$y_t^0 = C \exp\{-\lambda t\}.$$

Poi si cerca la soluzione y_t del tipo $y_t = C_t \exp\{-\lambda t\}$, da cui, essendo

$$\frac{d}{dt} y_t = \frac{d}{dt}\Big(C_t \exp\{-\lambda t\}\Big) = -\lambda \exp\{-\lambda t\} C_t + \dot{C}_t \exp\{-\lambda t\} = -\lambda y_t + \dot{C}_t \exp\{-\lambda t\},$$

si ottiene che deve essere necessariamente

$$\dot{C}_t \exp\{-\lambda t\} = v_t.$$

ovvero

$$C_t = C_0 + \int_0^t \exp\{\lambda s\} v_s\, ds.$$

In conclusione la soluzione è data da

$$y_t = \exp\{-\lambda t\}\left(C_0 + \int_0^t \exp\{\lambda s\} v_s\, ds\right)$$

Da ciò si ricava immediatamente che deve essere $F_t = 0$ ed $\exp\{-\lambda t\}G_t = \sigma$, ovvero

$$G_t = \sigma \exp\{\lambda t\}.$$

Da quest'ultima espressione si ricava immediatamente che

$$\eta_t = x + \int_0^t \sigma \exp\{\lambda s\} dW_s,$$

e quindi

$$X_t = \exp\{-\lambda t\}\eta_t = \exp\{-\lambda t\}\left(x + \int_0^t \sigma \exp\{\lambda s\} dW_s\right),$$

ovvero il processo di Orstein-Ulhenbeck.

Una volta ottenuta la soluzione è facile verificare che è effettivamente una soluzione, riapplicando la formula del differenziale stocastico del prodotto.[23]

Come applicazione delle proprietà dell'integrale stocastico si vede quindi subito come si possono calcolare valore medio e varianza del processo X_t di Orstein-Ulhenbeck.

Infatti

$$\mathbb{E}[X_t] = \exp\{-\lambda t\}x + \exp\{-\lambda t\}\mathbb{E}\left[\int_0^t \sigma \exp\{\lambda s\} dW_s\right] = \exp\{-\lambda t\}x,$$

dove l'ultima uguaglianza vale in quanto l'integrale stocastico ha valore atteso nullo.

Da ciò si può ricavare anche la varianza, in quanto

$$(X_t - \mathbb{E}[X_t])^2 = (X_t - \exp\{-\lambda t\}x)^2 = \exp\{-2\lambda t\}\left(\int_0^t \sigma \exp\{\lambda s\} dW_s\right)^2,$$

e quindi, passando ai valori attesi,

$$\begin{aligned} Var(X_t) &= \exp\{-2\lambda t\}\mathbb{E}\left[\left(\int_0^t \sigma \exp\{\lambda s\} dW_s\right)^2\right] \\ &= \mathbb{E}\left[\exp\{-2\lambda t\}\int_0^t \sigma^2 \exp\{2\lambda s\} ds\right] = \frac{\sigma^2}{2\lambda}\left(1 - \exp\{-2\lambda t\}\right). \end{aligned}$$

Anche la covarianza si può ricavare facilmente osservando che

$$\begin{aligned} &(X_t - \mathbb{E}[X_t])\,(X_{t'} - \mathbb{E}[X_{t'}]) \\ &= \exp\{-\lambda t\}\left(\int_0^t \sigma \exp\{\lambda s\} dW_s\right)\exp\{-\lambda t'\}\left(\int_0^{t'} \sigma \exp\{\lambda s\} dW_s\right) \\ &= \exp\{-\lambda(t+t')\}M_t M_{t'}, \end{aligned}$$

dove $M_t = \int_0^t \sigma \exp\{\lambda s\} dW_s$. Poiché, per ogni $\mathcal{G}_t$-martingala di quadrato integrabile, se $t \le t'$, si ha

$$\mathbb{E}[M_t M_{t'}] = \mathbb{E}[M_t \mathbb{E}[M_{t'}|\mathcal{G}_t]] = \mathbb{E}[M_t M_t] = \mathbb{E}[M_t^2].$$

[23] Ovvero si tratta di ripercorrere i passi precedenti al contrario:

$$d(\exp\{-\lambda t\}) = -\lambda \exp\{-\lambda t\} dt,$$

$$d\left(x + \int_0^t \sigma \exp\{\lambda s\} dW_s\right) = 0 \cdot dt + \sigma \exp\{\lambda t\} dW_t.$$

quindi se

$$X_t = \exp\{-\lambda t\}\left(x + \int_0^t \sigma \exp\{\lambda s\} dW_s\right),$$

allora

$$dX_t = \left(x + \int_0^t \sigma \exp\{\lambda s\} dW_s\right)(-\lambda \exp\{-\lambda t\} dt) + \exp\{-\lambda t\}\sigma \exp\{\lambda t\} dW_t = -\lambda X_t + \sigma dW_t.$$

si ottiene che, per $t \le t'$,

$$Cov(X_t, X_{t'}) = \exp\{-\lambda(t+t')\}\mathbb{E}\left[\int_0^t \sigma^2 \exp\{2\lambda s\} ds\right].$$

6.3.3 Esistenza per EDS: il caso Lipschitz globale

Dimostrazione. (del Teorema 6.13)
Esistenza
La dimostrazione dell'esistenza si basa sul classico procedimento di costruzione delle soluzioni con le approssimazioni di Picard. Si ponga $X_t^{(0)} = \eta$ (oppure $X_t^{(0)} = x$) e si definiscano per induzione

$$X_t^{(k+1)} := \eta + \int_s^t b(X_u^{(k)}, u)\, du + \int_s^t \sigma(X_u^{(k)}, u)\, dW_u$$

• Come primo punto dimostriamo che $X_t^{(k)}$ hanno la proprietà che $\sup_{t\le T} \mathbb{E}\big[|X_t^{(k)}|^2 du\big] < \infty$.

A questo scopo osserviamo che, ovviamente $\mathbb{E}[|X_t^{(0)}|^2] = \mathbb{E}[|\eta|^2] < \infty$, e che la precedente proprietà si ottiene per induzione utilizzando il fatto che se Y_t è un processo progressivamente misurabile e che gode della proprietà che $\sup_{t\le T} \mathbb{E}\big[|Y_t|^2 du\big] \le L < \infty$, allora anche il processo

$$Z_t := \eta + \int_s^t b(Y_u, u)\, du + \int_s^t \sigma(Y_u, u)\, dW_u$$

gode di una proprietà simile, ovviamente con un'altra costante: utilizzando il fatto che $(a+b+c)^2 \le 3a^2 + 3b^2 + 3c^2$, la disuguaglianza di Cauchy e la condizione di crescita (6.28), si ottiene

$$\begin{aligned}
|Z_t|^2 &\le 3\,|\eta|^2 + 3\left(\int_s^t b(Y_u, u)\, du\right)^2 + 3\left(\int_s^t \sigma(Y_u, u)\, dW_u\right)^2 \\
&\le 3\,|\eta|^2 + 3\,T\int_s^t |b(Y_u, u)|^2\, du + 3\left(\int_s^t \sigma(Y_u, u)\, dW_u\right)^2 \\
&\le 3\,|\eta|^2 + 3\,T\int_s^t K(1+|Y_u|^2)\, du + 3\left(\int_s^t \sigma(Y_u, u)\, dW_u\right)^2.
\end{aligned}$$

Passando al valore atteso, utilizzando l'isometria dell'integrale stocastico di Itô, e di nuovo la condizione di crescita (6.28), si ottiene

$$\begin{aligned}
\mathbb{E}\big[|Z_t|^2\big] &\le 3\,\mathbb{E}\big[|\eta|^2\big] + 3\,T\int_s^t K(1+\mathbb{E}\big[|Y_u|^2)\big]\, du + 3\,\mathbb{E}\left[\left(\int_s^t \sigma(Y_u, u)\, dW_u\right)^2\right] \\
&\le 3\,\mathbb{E}\big[|\eta|^2\big] + 3\,T\int_s^t K(1+\mathbb{E}\big[|Y_u|^2)\big]\, du + 3\int_s^t K(1+\mathbb{E}\big[|Y_u|^2)\big]\, du \\
&\le 3\,\mathbb{E}\big[|\eta|^2\big] + 3\,T\int_s^t K\big(1+\sup_{u\in[s,t]}\mathbb{E}\big[|Y_u|^2\big]\big)\, du + 3\int_s^t K\big(1+\sup_{u\in[s,t]}\mathbb{E}\big[|Y_u|^2\big]\big)\, du \\
&\le 3\,\mathbb{E}\big[|\eta|^2\big] + 3\,T^2K(1+L) + 3\,TK(1+L)
\end{aligned}$$

• Torniamo ora alla successione $X_t^{(k)}$. Lo scopo è dimostrare che è una successione per la quale vale la seguente disuguaglianza:

$$\mathbb{E}\big[\sup_{u\le t}|X_u^{(k+1)} - X_u^{(k)}|^2\big] \le 2\,(T+4)\,K\int_s^t \mathbb{E}\big[|X_u^{(k)} - X_u^{(k-1)}|^2\big]\, du \tag{6.33}$$

che a sua volta[24] implica che

$$\mathbb{E}\big[|X_t^{(k+1)} - X_t^{(k)}|^2\big] \le 2\,(T+4)\,K \int_s^t \mathbb{E}\big[|X_u^{(k)} - X_u^{(k-1)}|^2\big]\,du,$$

ossia, posto $C := 2\,(T+4)\,K$, e $g_k(t) := \mathbb{E}\big[|X_t^{(k)} - X_t^{(k-1)}|^2\big]$, che

$$g_{k+1}(t) \le C \int_s^t g_k(u)\,du.$$

Da questa disuguaglianza, insieme al fatto che

$$g_1(t) \le C\,(1 + \mathbb{E}[|\eta|^2])\,(t-s),$$

segue facilmente, per induzione, che[25]

$$g_k(t) = \mathbb{E}\big[|X_t^{(k)} - X_t^{(k-1)}|^2\big] \le \frac{C^k\,(t-s)^k}{k!}\,(1 + \mathbb{E}[|\eta|^2]). \tag{6.34}$$

**Infine, combinando (6.34) e la (6.33), si ottiene che

$$\begin{aligned}\mathbb{E}\big[\sup_{u\le t}|X_u^{(k+1)} - X_u^{(k)}|^2\big] &\le 2\,(T+4)\,K \int_s^t g_k(u)\,du \\ &\le C \int_s^t \frac{C^k\,(u-s)^k}{k!}\,(1 + \mathbb{E}[|\eta|^2])\,du \\ &= \frac{C^{k+1}\,(t-s)^{k+1}}{(k+1)!}\,(1 + \mathbb{E}[|\eta|^2]).**\end{aligned} \tag{6.35}$$

La dimostrazione della (6.33), è sostanzialmente simile a quella effettuata per l'unicità:

$$X_t^{(k+1)} - X_t^{(k)} = \int_s^t [b(X_u^{(k)}, u) - b(X_u^{(k-1)}, u)]\,du + \int_s^t [\sigma(X_u^{(k)}, u) - \sigma(X_u^{(k-1)}, u)]\,dW_u$$

per cui, usando la disuguaglianza $(a+b)^2 \le 2\,a^2 + 2\,b^2$, e la disuguaglianza di Cauchy

$$\begin{aligned}|X_t^{(k+1)} - X_t^{(k)}|^2 &\le 2\left(\int_s^t [b(X_u^{(k)}, u) - b(X_u^{(k-1)}, u)]\,du\right)^2 + 2\left(\int_s^t [\sigma(X_u^{(k)}, u) - \sigma(X_u^{(k-1)}, u)]dW_u\right)^2 \\ &\le \int_s^t 1^2 du \int_s^t [b(X_u^{(k)}, u) - b(X_u^{(k-1)}, u)]^2 du + 2\left(\int_s^t [\sigma(X_u^{(k)}, u) - \sigma(X_u^{(k-1)}, u)]dW_u\right)^2\end{aligned}$$

[24] Ovviamente $\mathbb{E}\big[|X_t^{(k+1)} - X_t^{(k)}|^2\big] \le \mathbb{E}\big[\sup_{u\le t}|X_u^{(k+1)} - X_u^{(k)}|^2\big]$
Tuttavia si potrebbe dimostrare, usando la stessa tecnica usata per la dimostrazione dell'unicità della soluzione che

$$\mathbb{E}\big[|X_t^{(k+1)} - X_t^{(k)}|^2\big] \le 2\,(T+1)\,K \int_s^t \mathbb{E}\big[|X_u^{(k)} - X_u^{(k-1)}|^2\big]\,du.$$

[25] Infatti

$$g_1(t) = \mathbb{E}\big[|X_t^{(1)} - X_t^{(0)}|^2\big] = \mathbb{E}\big[|X_t^{(1)} - \eta|^2\big] = \mathbb{E}\left[\left(\int_s^t b(\eta, u)\,du + \int_s^t \sigma(\eta, u)\,dW_u\right)^2\right]$$

e utilizzando la condizione di crescita (6.28) si ha che

$$\begin{aligned}g_1(t) &\le 2\,T \int_s^t \mathbb{E}\left[|b(\eta, u)|^2\right]\,du + 2\int_s^t \mathbb{E}\left[|\sigma(\eta, u)|^2\right]\,du \\ &\le 2\,T\,(t-s)K\,\left(1 + \mathbb{E}\left[|\eta|^2\right]\right) + 2\,(t-s)K\,\left(1 + \mathbb{E}\left[|\eta|^2\right]\right) \le C\,(1 + \mathbb{E}[|\eta|^2])(t-s)\end{aligned}$$

per ogni $C \ge 2\,(T+1)\,K$.
Basta poi verificare che l'ipotesi induttiva $g_k(t) \le \frac{C^k\,(t-s)^k}{k!}\,(1 + \mathbb{E}[|\eta|^2])$ insieme alla relazione $g_{k+1}(t) \le C \int_s^t g_k(u)\,du$, per ottenere che

$$g_{k+1}(t) \le C \int_s^t g_k(u)\,du \le C \int_s^t \frac{C^k\,(u-s)^k}{k!}\,(1 + \mathbb{E}[|\eta|^2])\,du = \frac{C^{k+1}\,(t-s)^{k+1}}{(k+1)!}\,(1 + \mathbb{E}[|\eta|^2]).$$

utilizzando la condizione di Lipschitz (6.27), si ha

$$\leq 2\,T \int_s^t K\,|X_u^{(k)} - X_u^{(k-1)}|^2 du + 2\left(\int_s^t [\sigma(X_u^{(k)}, u) - \sigma(X_u^{(k-1)}, u)] dW_u\right)^2$$

Considerando l'estremo superiore, si ottiene che

$$\begin{aligned}\sup_{t'\leq t} |X_{t'}^{(k+1)} - X_{t'}^{(k)}|^2 &\leq 2\,T \sup_{t'\leq t} \int_s^{t'} K\,|X_u^{(k)} - X_u^{(k-1)}|^2 du + 2 \sup_{t'\leq t}\left(\int_s^{t'} [\sigma(X_u^{(k)}, u) - \sigma(X_u^{(k-1)}, u)] dW_u\right)^2 \\ &\leq 2\,T \int_s^t K\,|X_u^{(k)} - X_u^{(k-1)}|^2 du + 2 \sup_{t'\leq t}\left(\int_s^{t'} [\sigma(X_u^{(k)}, u) - \sigma(X_u^{(k-1)}, u)] dW_u\right)^2\end{aligned}$$

Passando ai valori attesi, ed utilizzando la disuguaglianza di Doob si ottiene che, per $t \leq T$

$$\mathbb{E}\big[\sup_{t'\leq t} |X_{t'}^{(k+1)} - X_{t'}^{(k)}|^2\big] \leq 2\,T\,K \int_s^t \mathbb{E}\big[|X_u^{(k)} - X_u^{(k-1)}|^2\big] du + 2\cdot 4\,\mathbb{E}\left[\left(\int_s^t [\sigma(X_u^{(k)}, u) - \sigma(X_u^{(k-1)}, u)] dW_u\right)^2\right]$$

e quindi, per l'isometria dell'integrale stocastico,

$$\mathbb{E}\big[|X_t^{(k)} - X_t^{(k-1)}|^2\big] \leq T\,K \int_s^t \mathbb{E}\big[|X_u^{(k)} - X_u^{(k-1)}|^2\big] du + 8 \int_s^t \mathbb{E}\left[[\sigma(X_u^{(k)}, u) - \sigma(X_u^{(k-1)}, u)]^2\right] du$$

e usando di nuovo la condizione di Lipschitz (6.27), si ha

$$\mathbb{E}\big[|X_t^{(k)} - X_t^{(k-1)}|^2\big] \leq 2\,T\,K \int_s^t \mathbb{E}\big[|X_u^{(k)} - X_u^{(k-1)}|^2\big]\, du + 8\,K \int_s^t \mathbb{E}\big[|X_u^{(k)} - X_u^{(k-1)}|^2\big]\, du$$

• A questo punto si può facilmente provare che la successione $X_t^{(k)}$ converge quasi certamente in norma uniforme su ciascun intervallo limitato verso un processo X_t:

Dalla (6.33) e dalla (6.34), si ottiene immediatamente che

$$\mathbb{P}\big(A_k\big) = \mathbb{P}\big(\sup_{u\leq t} |X_u^{(k)} - X_u^{(k-1)}|^2 \geq 2^{-k}\big) \leq \mathbb{E}\big[\sup_{u\leq t} |X_u^{(k)} - X_u^{(k-1)}|^2\big]\, 2^k \leq \frac{**2^k C^k\, T^k}{k!}\,(1 + \mathbb{E}[|\eta|^2]),$$

dove gli eventi A_k sono definiti implicitamente. Il fatto che

$$\sum_k \mathbb{P}\big(A_k\big) < \infty$$

implica, per il Lemma di Borel Cantelli[26], che la probabilità che A_k si verifichino infinitamente spesso è nulla. Ciò equivale ad affermare che la probabilità che i complementari A_k^c si verifichino definitivamente vale 1, ossia:

[26] Ricordiamo che il Lemma di Borel Cantelli assicura che la condizione di convergenza della serie

$$\sum_k \mathbb{P}\big(A_k\big) < \infty$$

è sufficiente affinché

$$\mathbb{P}\big(A_k\ i.s.\ \big) = \mathbb{P}\big(\forall\, n\, \exists\, m \geq n \text{ tale che} A_m \text{ si verifica }\big) = \mathbb{P}\big(\cap_n \cup_{m\geq n} A_n\big) = 0.$$

Ciò è ovviamente equivalente a

$$\mathbb{P}\big(A_k^c\ def.\ \big) = \mathbb{P}\big(\exists\, n\, \forall\, m \geq n\ A_m^c \text{ si verifica }\big) = \mathbb{P}\big(\cup_n \cap_{m\geq n} A_n^c\big) = 1,$$

cioè, esiste un evento E di probabilità 1, per il quale se $\omega \in E$, allora esiste un $n = n(\omega)$, tale che per ogni $m \geq n$ si verifica A_m^c, cioè $\omega \in A_m^c$.
Inoltre, il Lemma di Borel Cantelli assicura che la condizione di convergenza è necessaria, nel caso in cui gli eventi A_k siano indipendenti.

esiste un evento E di probabilità 1, per il quale se $\omega \in E$, allora esiste un $n = n(\omega)$, tale che per ogni $m \geq n$ si ha

$$\sup_{u \leq t} |X_u^{(k)}(\omega) - X_u^{(k-1)}(\omega)|^2 < 2^{-k}.$$

Quindi, per ogni $p \geq 0$

$$\sup_{u \leq t} |X_u^{(n+p)}(\omega) - X_u^{(n-1)}(\omega)|^2 < \sum_{k \geq n} 2^{-k} = 2^{n-1}.$$

Ciò mostra che la successione è di Cauchy e che quindi esiste un processo continuo X_t, che è il limite (con probabilità 1), in norma uniforme di $X_t^{(k)}$.

**• Va osservato che la maggiorazione in (6.35) implica che la successione dei processi $X^{(k)}$ ha la proprietà che

$$\sup_{p \geq 0} \mathbb{E}\big[\big(\sup_{u \in [s,t]} |X_u^{(n)} - X_u^{(n+p)}|\big)^2\big]$$

tende a zero per n che tende all'infinito. Infatti[27]

$$\sup_{p \geq 0} \mathbb{E}^{1/2}\big[\big(\sup_{u \in [s,t]} |X_u^{(n)} - X_u^{(n+p)}|\big)^2\big] \leq \sup_{p \geq 0} \sum_{k=n}^{n+p-1} \mathbb{E}^{1/2}\big[\big(\sup_{u \in [s,t]} |X_u^{(k+1)} - X_u^{(k)}|\big)^2\big]$$
$$\leq \sum_{k=n}^{\infty} \left(\frac{C^{k+1}\,(t-s)^{k+1}}{(k+1)!} \right)^{1/2} (1 + \mathbb{E}[|\eta|^2])^{1/2}$$

tende a zero in quanto $\sum_{k=1}^{\infty} \left(\frac{\alpha^k}{k!}\right)^{1/2}$ è una serie convergente.

In altre parole la successione dei processi $X^{(k)}$ è di Cauchy nello spazio $L^2(\Omega, C[s,t])$ ed ammette limite in tale spazio. D'altra parte sappiamo, dal punto precedente che la successione converge quasi certamente, in norma uniforme al processo X, quindi questo limite deve essere necessariamente tale processo, ossia

$$\mathbb{E}\big[\big(\sup_{u \in [s,t]} |X_u^{(n)} - X_u|\big)^2\big]$$

tende a zero per n che tende all'infinito.

In questa dimostrazione entrano in gioco lo spazio $L^2(\Omega, C[s,t])$, ma è possibile evitare di utilizzare questo strumento, procedendo come suggerito ad esempio nel libro di Öksendal [9]:

Da (6.34) segue immediatamente che, per ogni $t \in [s,T]$

$$\sup_{p \geq 0} \mathbb{E}^{1/2}\big[\big(|X_t^{(n)} - X_t^{(n+p)}|\big)^2\big] \leq \sup_{p \geq 0} \sum_{k=n}^{n+p-1} \mathbb{E}^{1/2}\big[\big(|X_t^{(k+1)} - X_t^{(k)}|\big)^2\big]$$
$$\leq \sum_{k=n}^{\infty} \left(\frac{C^{k+1}\,(t-s)^{k+1}}{(k+1)!} \right)^{1/2} (1 + \mathbb{E}[|\eta|^2])^{1/2}$$

tende a zero in quanto $\sum_{k=1}^{\infty} \left(\frac{\alpha^k}{k!}\right)^{1/2}$ è una serie convergente. Quindi, qualunque sia $t \in [s,T]$ la successione di variabili aleatorie $X_t^{(k)}$ è di Cauchy in $L^2(\Omega, \mathbb{P})$ e perciò ammette limite in tale spazio. Tale limite deve coincidere con

[27] Si tenga conto del fatto che se (Y_u) e (Z_u) sono due processi per i quali vale

$$\mathbb{E}\big[\big(\sup_{u \in [s,t]} |Y_u|\big)^2\big] < \infty, \qquad \mathbb{E}\big[\big(\sup_{u \in [s,t]} |Z_u|\big)^2\big] < \infty$$

allora

$$\mathbb{E}^{1/2}\big[\big(\sup_{u \in [s,t]} |Y_u + Z_u|\big)^2\big] \leq \mathbb{E}^{1/2}\big[\big(\sup_{u \in [s,t]} |Y_u| + \sup_{u \in [s,t]} |Z_u|\big)^2\big] \leq \mathbb{E}^{1/2}\big[\big(\sup_{u \in [s,t]} |Y_u|\big)^2\big] + \mathbb{E}^{1/2}\big[\big(\sup_{u \in [s,t]} |Z_u|\big)^2\big].$$

X_t perché sappiamo già che la successione dei processi $(X_t^{(k)})_{t\in[s,T]}$ converge quasi certamente al processo $(X_t)_{t\in[s,T]}$ in norma uniforme. Di conseguenza, poiché per quasi ogni ω

$$\lim_{n\to\infty}\int_s^T |X_u^{(k)} - X_u^{(n)}|^2 du = \int_s^T |X_u^{(k)} - X_u|^2,$$

per il lemma di Fatou,

$$\mathbb{E}\Big[\int_s^T |X_u^{(k)} - X_u|^2 du\Big] \le \limsup_{n\to\infty} \mathbb{E}\Big[\int_s^T |X_u^{(k)} - X_u^{(n)}|^2 du\Big]$$

e prendendo quindi il limite per k che tende ad infinito si ottiene

$$0 \le \limsup_{k\to\infty} \mathbb{E}\Big[\int_s^T |X_u^{(k)} - X_u|^2 du\Big] \le \limsup_{k\to\infty}\limsup_{n\to\infty} \mathbb{E}\Big[\int_s^T |X_u^{(k)} - X_u^{(n)}|^2 du\Big] = 0$$

da cui si ottiene

$$\lim_{k\to\infty} \mathbb{E}\Big[\int_s^T |X_u^{(k)} - X_u|^2 du\Big] = 0 \tag{6.36}$$

• A questo punto non rimane che
(i) osservare che la convergenza (quasi certa) in norma uniforme implica la convergenza sia degli integrali ordinari che degli integrali stocastici

$$\int_s^t b(X_u^{(k)}, u)\, du \to \int_s^t b(X_u, u)\, du, \qquad \int_s^t \sigma(X_u^{(k)}, u)\, dW_u \to \int_s^t \sigma(X_u, u)\, dW_u,$$

dove il secondo limite deriva dal fatto che

$$\mathbb{E}\left[\int_s^t \big(\sigma(X_u^{(k)}, u) - \sigma(X_u, u)\big)^2\, du\right] \to 0. \tag{6.37}$$

**Quest'ultima proprietà deriva[28] dalla proprietà di Lipschitz per σ e dal limite (6.36).
(ii) ottenere che il processo limite è soluzione dell'equazione differenziale stocastica passando al limite nella relazione che definisce $X_t^{(k+1)}$:

$$\begin{array}{ccccccc} X_t^{(k+1)} & = & \eta & + & \displaystyle\int_s^t b(X_u^{(k)}, u)\, du & + & \displaystyle\int_s^t \sigma(X_u^{(k)}, u)\, dW_u \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ X_t & = & \eta & + & \displaystyle\int_s^t b(X_u, u)\, du & + & \displaystyle\int_s^t \sigma(X_u, u)\, dW_u. \end{array}$$

□

[28]**La dimostrazione delle proprietà in (6.37) e in (6.36), non sono necessarie, infatti (si veda ad esempio la prima edizione del libro di Baldi [1]) si potrebbe anche dimostrare che il limite degli integrali stocastici

$$\int_s^t \sigma(X_u^{(k)}, u)\, dW_u \to \int_s^t \sigma(X_u, u)\, dW_u$$

si ottiene lo stesso: infatti basta la convergenza quasi certa (e quindi la convergenza in probabilità), di $\int_s^t \big(\sigma(X_u^{(k)}, u) - \sigma(X_u, u)\big)^2\, du$ a zero. Tuttavia poi bisognerebbe dimostrare che il processo limite gode della proprietà che

$$\mathbb{E}\Big[\int_s^T |X_u|^2 du\Big] < \infty,$$

che invece si ottiene subito dalla (6.36).

6.3.4 Dipendenza continua dal dato iniziale e proprietà di Markov per soluzioni di EDS

In questa sezione consideriamo i seguenti problemi: sotto le ipotesi di Lipschitz globale (6.27) e di crescita (6.28), che assicurano l'esistenza e l'unicità della soluzione forte $X_t^{s,x}$ dell'equazione differenziale stocastica

$$\xi_t = x + \int_s^t b(\xi_u, u)\, du + \int_s^t \sigma(\xi_u, u)\, dW_u$$

vogliamo studiare

- la dipendenza della soluzione dal dato iniziale di $X_t^{s,x}$;
- la proprietà di Markov della famiglia di processi $X_t^{s,x}$.

Per quanto riguarda la dipendenza dal dato iniziale, vale il seguente risultato **di continuità in media quadratica rispetto alle condizioni iniziali.

Proposizione 6.14. *Nelle condizioni del Teorema 6.13 di esistenza e unicità, **esiste una costante C tale che, per ogni $t \le T$,*

$$\mathbb{E}\left[\sup_{u\le t} |X_u^{s,x} - X_u^{s,y}|^2\right] \le 3\,|x-y|^2\, e^{C\,(t-s)}$$

Dimostrazione. La dimostrazione si ottiene a partire dalla disuguaglianza

$$\begin{aligned}\mathbb{E}\big[\sup_{u\le t} |X_u^{s,x)} - X_u^{s,y}|^2\big] &\le 3\,|x-y|^2 + 3\,(T+4)\,K \int_s^t \mathbb{E}\big[|X_u^{s,x} - X_u^{s,y)}|^2\big]\, du \\ &\le 3\,|x-y|^2 + 3\,(T+4)\,K \int_s^t \mathbb{E}\big[\sup_{u'\le u} |X_{u'}^{s,x} - X_{u'}^{s,y)}|^2\big]\, du\end{aligned}$$

e dal Lemma di Gronwall, con $C = 3(T+4)K$. La precedente disuguaglianza si ottiene in modo simile a quanto fatto nella dimostrazione dell'esistenza della soluzione, ma a partire dall'eguaglianza

$$X_t^{s,x} - X_t^{s,y} = x - y + \int_s^t [b(X_u^{s,x}, u) - b(X_u^{s,y}, u)]\, du + \int_s^t [\sigma(X_u^{s,x}, u) - \sigma(X_u^{s,y}, u)]\, dW_u,$$

e utilizzando il fatto che $(a+b+c)^2 \le 3\,a^2 + 3\,b^2 + 3\,c^2$. □

In realtà si può dimostrare che valgono altre disuguaglianze come, ad esempio

$$\mathbb{E}\left[|X_t^{s,x} - X_{t'}^{s',y}|^p\right] \le c\,\left(|x-y|^p + |t-t'|^{p/2} + |s-s'|^{p/2}\right).$$

A sua volta tale disuguaglianza permette di affermare che esiste una versione di $X_t^{s,x}$ che è continua (congiuntamente) in (x, s, t), $s \le t$, grazie ad una generalizzazione del criterio di Chensov-Kolmogorov. Per approfondimenti rimandiamo ad esempio al libro di Baldi [1].

Queste proprietà di continuità permettono di affermare che, per ogni funzione uniformemente continua e limitata la funzione

$$x \mapsto T_{s,t} f(x) := \mathbb{E}\,[f(X_t^{s,x})]$$

sia una funzione uniformemente continua. In realtà a questo scopo è sufficiente[29] la continuità in media quadratica rispetto ai dati iniziali, ossia il risultato della Proposizione 6.14.

[29] Per la continuità della funzione $T_{s,t} f$ basta controllare che

$$\begin{aligned}|T_{s,t} f(x) - T_{s,t} f(y)| &= \left|\mathbb{E}\left[f(X_t^{s,x})\right] - \mathbb{E}\left[f(X_t^{s,y})\right]\right| \le \mathbb{E}\left[\left|f(X_t^{s,x}) - f(X_t^{s,y})\right|\right] \\ &\le \mathbb{E}\left[\mathbb{I}_{|X_t^{s,x} - X_t^{s,y}|\le\theta}\left|f(X_t^{s,x}) - f(X_t^{s,y})\right|\right] + \mathbb{E}\left[\mathbb{I}_{|X_t^{s,x} - X_t^{s,y}|>\theta}\left|f(X_t^{s,x}) - f(X_t^{s,y})\right|\right].\end{aligned}$$

Per quanto riguarda la proprietà di Markov, cominciamo con l'osservare che, nel caso in cui $f = \mathbf{1}_I$, **con I boreliano, la funzione indicatrice di I, possiamo considerare

$$p(s,x,t,I) := T_{s,t}\mathbf{1}_I(x)\big) = \mathbb{E}\left[\mathbf{1}_I(X_t^{s,x})\right] = \mathbb{P}\left(X_t^{s,x} \in I\right),$$

come possibile famiglia delle probabilità di transizione.

Ovviamente vale che

$$I \mapsto p(s,x,t,I) := \mathbb{P}\left(X_t^{s,x} \in I\right)$$

definisce una misura di probabilità per ogni scelta di x e di $s \le t$. Tuttavia, in generale, $p(s,x,t,I)$, come funzione di x, non è una funzione continua. Tuttavia la proprietà di continuità di $T_{s,t}f$ per le funzioni continue, a sua volta permette di affermare che, per ogni boreliano I, e comunque scelti $s \le t$, la funzione

$$x \mapsto p(s,x,t,I) := \mathbb{P}\left(X_t^{s,x} \in I\right)$$

è misurabile[30].

Per ottenere che $p(s,x,t,I)$ sia una famiglia di probabilità di transizione, rimane da verificare che soddisfi la condizione di Chapman-Kolmogorov. A questo scopo facciamo le seguenti osservazioni.

Il Teorema 6.13 di esistenza e unicità garantisce che per ogni $s < r < t$ vale la seguente identità:

$$X_t^{s,x} = X_t^{r,X_r^{s,x}},$$

cioè che, in ogni istante $t \ge r$, la soluzione ottenuta a partire dal'istante r, con condizione iniziale $X_r^{s,x}$ deve coincidere con la soluzione $X_t^{s,x}$, ottenuta a partire dall'istante s, con condizione iniziale x. Questa identità è facile da dimostrare[31].

Posto $w(f,\theta) := \sup\{|f(z) - f(z')|,\ |z - z'| \le \theta\}$, il modulo di continuità di f e $\|f\| = \sup_z |f(z)|$ si ha

$$\begin{aligned}|T_{s,t}f(x) - T_{s,t}f(y)| &\le w(f,\theta)\,\mathbb{P}\left(|X_t^{s,x} - X_t^{s,y}| \le \theta\right) + 2\,\|f\|\,\mathbb{P}\left(|X_t^{s,x} - X_t^{s,y}| > \theta\right)\\ &\le w(f,\theta)\ + 2\,\|f\|\,\mathbb{E}\left(|X_t^{s,x} - X_t^{s,y}|^2\right)/\theta^2\end{aligned}$$

La continuità segue nel seguente modo: per ogni $\varepsilon > 0$ sia $\theta = \theta_\varepsilon$ tale che $w(f,\theta_\epsilon) \le \varepsilon/2$ e sia $\delta = \delta_\varepsilon\big(= \delta(\varepsilon,\theta_\epsilon)\big)$ tale che,**se $|x - y| < \delta_\varepsilon$, allora**$2\,\|f\|\,\mathbb{E}\left(|X_t^{s,x} - X_t^{s,y}|^2\right)/\theta_\varepsilon^2 \le \varepsilon/2$, **in modo che**possiamo affermare che

$$\forall\,\varepsilon > 0 \quad \exists\,\delta_\epsilon, \quad \text{tale che } \forall\,x\ y \ \text{ con } |x - y| \le \delta_\varepsilon \quad |T_{s,t}f(x) - T_{s,t}f(y)| \le \varepsilon.$$

Dalla dimostrazione è chiaro che non serve la continuità delle traiettorie, e neppure la continuità in media quadratica, ma basta solo la continuità in probabilità: basta infatti sostituire la richiesta su δ_ε con la richiesta che $\delta = \delta_\varepsilon\big(\ = \delta(\varepsilon,\theta_\epsilon)\big)$ sia tale che $2\,\|f\|\,\mathbb{P}\left(|X_t^{s,x} - X_t^{s,y}| > \theta_\varepsilon\right) \le \varepsilon/2$.

[30] L'idea è quella di utilizzare la continuità per ottenere che per ogni compatto Γ la probabilità $p(s,x,t,\Gamma)$ si può approssimare con $T_{s,t}f_n(x)$, per una opportuna successione di funzioni uniformemente continue, e che il limite puntuale di funzioni continue è misurabile. Si tratta poi di utilizzare la proprietà di regolarità interna delle misure in $\mathbb{R}^d$.

[31] Il punto fondamentale consiste nell' osservare che

$$\begin{aligned}X_t^{s,x} &= x + \int_s^t b(X_u^{s,x},u)\,du + \int_s^t \sigma(X_u^{s,x},u)\,dW_u\\ &= x + \int_s^r b(X_u^{s,x},u)\,du + \int_s^r \sigma(X_u^{s,x},u)\,dW_u + \int_r^t b(X_u^{s,x},u)\,du + \int_r^t \sigma(X_u^{s,x},u)\,dW_u\\ &= X_r^{s,x} + \int_r^t b(X_u^{s,x},u)\,du + \int_r^t \sigma(X_u^{s,x},u)\,dW_u\end{aligned}$$

e che $X_r^{s,x}$ è una variabile aleatoria $\mathcal{F}_r$-misurabile e di quadrato integrabile: per il problema di esistenza e unicità necessariamente $X_t^{s,x}$ è la soluzione dell'equazione differenziale stocastica con dato iniziale $(r, X_r^{s,x})$.

Per completare la dimostrazione si dovrebbe anche osservare che se ξ_t è soluzione di

$$\xi_t = \eta + \int_r^t b(\xi_u,u)\,du + \int_r^t \sigma(\xi_u,u)\,dW_u, \qquad \text{con } \eta \text{ una variabile aleatoria } \mathcal{F}_r\text{-misurabile, } \mathbb{E}[|\eta|^2] < \infty,$$

allora

$$\xi_t = X_t^{r,\eta}.$$

Per ottenere la condizione di Chapman-Kolmogorov osserviamo innanzi tutto che l'equazione differenziale stocastica

$$\xi_t = x + \int_r^t b(\xi_u, u)\, du + \int_r^t \sigma(\xi_u, u)\, dW_u,$$

equivale alla equazione differenziale stocastica

$$\xi_t = x + \int_r^t b(\xi_u, u)\, du + \int_r^t \sigma(\xi_u, u)\, d\widetilde{W}_u^{(r)},$$

dove $\widetilde{W}_u^{(r)} := W_{r+u} - W_r$ è un processo di Wiener, rispetto alla filtrazione $\widetilde{\mathcal{F}}_t = \mathcal{F}_{t+r}$: continuità, distribuzione e misurabilità sono ovvie, la proprietà degli incrementi indipendenti rispetto a $\widetilde{\mathcal{F}}_t = \mathcal{F}_{t+r}$, deriva dall'osservazione che se $t \le t'$, allora $\widetilde{W}_{t'}^{(r)} - \widetilde{W}_t^{(r)} := W_{t'+r} - W_r - (W_{t+r} - W_r) = W_{t'+r} - W_{t+r}$. Si osservi inoltre che quindi il processo $\widetilde{W}^{(r)}(\cdot, \omega) := \left(\widetilde{W}_u^{(r)}\right)_{u \ge 0}$ è indipendente da $\widetilde{\mathcal{F}}_0 = \mathcal{F}_r$.

Possiamo quindi affermare che esiste un funzionale

$$\varphi : \mathbb{R}^m \times \{0 \le s \le t < \infty\} \times C([0,\infty), \mathbb{R}^d) \to C([0,\infty), \mathbb{R}^m) \qquad (x, s, t, w(\cdot)) \mapsto \varphi(x, r, t, w(\cdot))$$

per il quale

$$X_t^{r,x}(\omega) = \varphi(r, x, t, \widetilde{W}^{(r)}(\cdot, \omega)).$$

Di conseguenza possiamo riscrivere la relazione $X_t^{s,x} = X_t^{r, X_r^{s,x}}$ come

$$X_t^{s,x}(\omega) = \varphi(r, X_r^{s,x}(\omega), t, \widetilde{W}^{(r)}(\cdot, \omega)),$$

dove è fondamentale osservare che $X_r^{s,x}(\omega)$ è $\mathcal{F}_r$-misurabile e $\widetilde{W}^{(r)}(\cdot, \omega)$ è indipendente da $\mathcal{F}_r$.

Questo fatto è fondamentale nella dimostrazione dell'equazione di Chapman-Kolmogorov. Infatti

$$\begin{aligned} p(s, x, t, I) &= \mathbb{P}\left(X_t^{s,x} \in I\right) = \mathbb{E}\left[\mathbf{1}_I(X_t^{s,x})\right] = \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_I(X_t^{s,x}) | \mathcal{F}_r\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_I(\varphi(r, X_r^{s,x}(\omega), t, \widetilde{W}^{(r)}(\cdot, \omega))) | \mathcal{F}_r\right]\right] \end{aligned}$$

A questo punto basta applicare una proprietà dei valori attesi condizionali[32] per la quale possiamo affermare che

$$\begin{aligned} \mathbb{E}\left[\mathbf{1}_I(\varphi(r, X_r^{s,x}(\omega), t, \widetilde{W}^{(r)}(\cdot, \omega))) | \mathcal{F}_r\right] &= \mathbb{E}\left[\mathbf{1}_I(\varphi(r, y, t, \widetilde{W}^{(r)}(\cdot, \omega)))\right]\Big|_{y = X_r^{s,x}(\omega)} \\ &= \mathbb{E}\left[\mathbf{1}_I(X_t^{r,y})\right]\Big|_{y = X_r^{s,x}(\omega)} = p(r, y, t, I)\Big|_{y = X_r^{s,x}(\omega)} = p(r, X_r^{s,x}(\omega), t, I). \end{aligned}$$

Questo fatto è banale se η assume un numero finito di valori. Ogni variabile aleatoria di quadrato integrabile si può approssimare, in media quadratica, con variabili aleatorie η_n, che assumono un numero finito di valori. Dalla relazione $\mathbb{E}[|\eta_n - \eta|^2] \to 0$ si ottiene facilmente che le soluzioni $\xi_n(t) = X_t^{r,\eta_n}$ convergono a ξ_t nel senso che $\mathbb{E}[\sup_{t \le T} |\xi_n(t) - \xi_t|^2] \to 0$. Passando eventualmente a sottosuccessioni, possiamo affermare che $\eta_n(\omega)$ converge ad $\eta(\omega)$ e $\xi_n(t, \omega) = X_t^{r,\eta_n(\omega)}(\omega)$ converge a $\xi_t(\omega)$, per ω che appartiene ad un insieme di probabilità 1. La continuità rispetto ad x di $X_t^{r,x}(\omega)$ permette di ottenere la tesi:

$$\xi_t(\omega) = \lim_{n \to \infty} \xi_n(t, \omega) = \lim_{n \to \infty} X_t^{r,\eta_n(\omega)}(\omega) = X_t^{r,\eta(\omega)}(\omega).$$

[32] Ci riferiamo alla proprietà seguente (per la dimostrazione si veda Lemma 3.9 (pag. 55 di Baldi [1]):
In uno spazio $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathcal{G}$ e $\mathcal{H}$ sono due sotto-σ-algebre di $\mathcal{F}$, indipendenti, X è una variabile aleatoria a valori in uno spazio misurabile $(S, \mathcal{S})$, con X misurabile rispetto a **$\mathcal{H}$**, infine $\psi(x, \omega)$ è una funzione a valori reali, congiuntamente misurabile rispetto a $\mathcal{S} \times \mathcal{G}$, con la variabile aleatoria $\Psi(\omega) = \psi(X(\omega), \omega)$ integrabile. Allora, posto

$$\Phi(x) := \mathbb{E}[\psi(x, \omega)],$$

si ha che

$$\Phi\big(X(\omega)\big) = \mathbb{E}\left[\Psi | \mathcal{H}\right] \Big(= \mathbb{E}\left[\psi(X, \cdot) | \mathcal{H}\right] \Big)$$

L'idea della dimostrazione è basata sull'osservazione che ***(i)*** tale proprietà è banale nel caso in cui $\psi(x, \omega) = f(x)\, Z(\omega)$ con f misurabile e deterministica e Z variabile aleatoria $\mathcal{G}$-misurabile, ***(ii)*** che basta dimostrarla per funzioni $\psi(x, \omega)$ non negative, ***(iii)*** ogni funzione $\psi(x, \omega)$ misurabile rispetto a $\mathcal{S} \times \mathcal{G}$, si può approssimare monotonamente con combinazioni lineari del tipo $f(x)\, Z(\omega)$.

Ricapitolando abbiamo che

$$p(s,x,t,I) = \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_I(\varphi(r, X_r^{s,x}(\omega), t, \widetilde{W}^{(r)}(\cdot,\omega)))|\mathcal{F}_r\right]\right] = \mathbb{E}\left[p(r, X_r^{s,x}(\omega), t, I)\right],$$

ossia, considerando che $p(s,x,r,dy)$ è la legge di $X_r^{s,x}$

$$p(s,x,t,I) = \int_{\mathbb{R}^m} p(r,y,t,I)\, p(s,x,r,dy),$$

che è appunto l'uguaglianza di Chapman-Kolmogorov.

6.3.5 Tempi di uscita da una striscia per una soluzione di una EDS

Quanto fatto nella Sottosezione 6.2.3 si può ripetere nel caso di un processo $X_t = X_t^{x,0}$, **soluzione di una equazione differenziale stocastica in $\mathbb{R}^1$, nel caso omogeneo, ossia con $b(y,t) = b(y)$ e $\sigma(y,t) = \sigma(y) > 0$, con b e σ che soddisfano le condizioni di esistenza e unicità, almeno nell'intervallo di interesse, che in questo paragrafo poniamo uguale a $[a,b]$, invece che a $[-a,b]$. In particolare possiamo generalizzare la Proposizione 6.11.

Proposizione 6.15. *Se esiste una funzione $\phi \in C^2$, che sia non negativa nell'intervallo $[a,b]$, e per la quale*

$$\phi'(y)\, b(y) + \frac{1}{2}\, \phi''(y)\, \sigma^2(y) \le -1,$$

*allora il valore atteso di $\tau = \tau(a,b) = \inf\{t > 0 :\ X_t \notin [a,b]\}$ **è finito, e quindi τ è finito con probabilità 1.*

Dimostrazione. La dimostrazione è simile a quella della Proposizione 6.11, va solo osservato che $\max_{y\in[a,b]} |\phi'(y)\, \sigma(y)| < \infty$. □

**

Anche il risultato della Proposizione 6.16, relativo al valore atteso di τ, si estende in modo simile.

Proposizione 6.16. *Se esiste una funzione $v \in C^2$ e tale che*

$$b(y)v'(y) + \frac{1}{2}\, \sigma^2(y)v''(y) = -1, \quad y \in (a,b) \qquad v(a) = v(b) = 0,$$

allora $\tau = \tau(a,b)$, il tempo di prima uscita da (a,b), ha valore atteso $v(x)$, ossia

$$\mathbb{E}_x\left[\tau\right] = v(x).$$

Per la soluzione generale di tale equazione si rimanda a (6.20) e a (6.22), ponendo $\alpha(x) = \frac{2b(x)}{\sigma^2(x)}$ e $\beta(x) = \frac{2}{\sigma^2(x)}$.

**Se τ è finito con probabilità 1, possiamo porci il problema del calcolo della probabilità di uscire dalla striscia passando per a. In altri termini posti

$$\tau_a := \inf\{t > 0 :\ X_t \notin [a, +\infty)\} \quad \text{e} \quad \tau_b := \inf\{t > 0 :\ X_t \notin (-\infty, b]\},$$

di modo che

$$\tau(a,b) = \tau_a \wedge \tau_b,$$

ci chiediamo quanto vale la probabilità

$$\mathbb{P}(\tau_a < \tau_b) = \mathbb{P}(X_\tau = a).$$

Per effettuare questo calcolo ci serviamo sempre della tecnica di trovare una funzione h, che sia A-armonica, di modo che il processo $h(X_t)$ sia una martingala.**

La condizione per una funzione h di essere A-armonica, ossia che

$$Ah(x) = h'(y)\, b(y) + \frac{1}{2}\, h''(y)\, \sigma^2(y) = 0, \qquad y \in [a,b].$$

**Aggiungendo la condizione che $h(a) = 1$ e $h(b) = 0$, la funzione $h(x)$ permette di calcolare la probabilità di uscire da una striscia passando per a, ossia

$$h(x) = \mathbb{P}_x(X_\tau = a) = \mathbb{P}_x\big(h(X_\tau)\big).$$

La soluzione $h(x)$ si può scrivere come

$$h(x) = \frac{\varphi(b) - \varphi(x)}{\varphi(b) - \varphi(a)}$$

dove

$$\varphi'(x) = C_1 \exp\left\{-\int_{x_0}^{x} \frac{2\,b(y)}{\sigma^2(y)}\,dy\right\},$$

da cui

$$\varphi(x) = C_0 + C_1 \int_{x_1}^{x} \exp\left\{-\int_{x_0}^{z} \frac{2\,b(y)}{\sigma^2(y)}\,dy\right\}\,dz.$$

Si osservi che si può sempre prendere qualsiasi C_0 e C_1 senza modificare la definizione di h, in particolare si possono prendere $C_0 = 0$ e $C_1 = 1$.

Si osservi che, tranne nel caso in cui $C_1 = 0$, e quindi φ è costante, la funzione φ è monotona, in quanto φ' ha segno costante.

Consideriamo ora il caso in cui $\sigma(x) > 0$ in tutto l'intervallo aperto (r_1, r_2), **ma non necessariamente agli estremi. Allora, per ogni $\varepsilon > 0$ e per ogni $b < r_2$, la probabilità di uscire da $[r_1 + \varepsilon, b]$ passando da $r_1 + \varepsilon$ vale

$$\mathbb{P}_x\left(X_{\tau(r_1+\varepsilon,b)} = r_1 + \varepsilon\right) = \frac{\varphi(b) - \varphi(x)}{\varphi(b) - \varphi(r_1 + \varepsilon)},$$

dove φ è definita come sopra con x_0 ed x_1 appartenenti all'intervallo (r_1, r_2).

Assumiamo inoltre che

$$\lim_{r \to r_1^+} \varphi(r) = -\infty.$$

Allora per ogni x e b, si può rendere la probabilità $\mathbb{P}_x\left(X_{\tau(r_1+\varepsilon,b)} = r_1 + \varepsilon\right)$ arbitrariamente piccola, pur di scegliere ε sufficientemente piccolo.

In questo caso l'estremo r_1 non può essere raggiunto in un tempo finito prima di raggiungere b, in quanto per raggiungere r_1 deve comunque passare nell'intervallo $(r_1, r_1 + \varepsilon)$ **(e quindi, di nuovo, raggiunge $r_1 + \varepsilon$ prima di raggiungere b). Ossia, posto $\tau_{r_1} = \inf\{t > 0 :\ X_t \le r_1\}$, $\tau_{r_1+\varepsilon} = \inf\{t > 0 :\ X_t < r_1 + \varepsilon\}$ e $\tau_b = \inf\{t > 0 :\ X_t > b\}$, se $\mathbb{P}_x(\tau_{r_1} < \tau_b) > 0$ allora $0 < \mathbb{P}_x(\tau_{r_1} < \tau_b) \le \mathbb{P}_x(\tau_{r_1+\varepsilon} < \tau_b)$, mentre il membro a destra si può rendere arbitrariamente vicino a zero per ε sufficientemente piccolo (anzi non può essere raggiunto affatto, in quanto, una volta raggiunto b, il processo dovrebbe comunque ritornare nell'intervallo aperto (r_1, b) prima di raggiungere r_1). Analogamente non può nemmeno verificarsi, con probabilità positiva, il caso in cui il processo $X_t = X_t^x$ tenda ad r_1 per t che tende ad infinito: infatti allora con la stessa probabilità raggiungerebbe $r_1 + \varepsilon$ prima di b.

Inoltre non può accadere che il limite di **X_t, per t che tende ad infinito, valga proprio r_1 con probabilità positiva $\alpha > 0$, perché allora, qualunque sia $\varepsilon > 0$, la probabilità di raggiungere $r_1 + \varepsilon$ prima di b dovrebbe essere almeno α, mentre sappiamo che tale probabilità tende a zero per ε che tende a zero.

In conclusione: se

$$\lim_{x \to r_1^+} \varphi(x) = \lim_{x \to r_1^+} \int_{x_1}^{x} \exp\left\{-\int_{x_0}^{z} \frac{2\,b(y)}{\sigma^2(y)}\,dy\right\}\,dz = -\infty,$$

allora il bordo r_1 non può essere mai raggiunto, neanche al limite per t che tende ad infinito. In tale caso si dice che r_1 è un ***bordo naturale***.

Tale conclusione può anche essere riscritta prendendo $x_1 = x_0 = \overline{x}$ e dicendo che r_1 è un bordo naturale se

$$\int_{r_1}^{\overline{x}} \exp\left\{-\int_{\overline{x}}^{z} \frac{2\,b(y)}{\sigma^2(y)}\,dy\right\}\,dz = +\infty.$$

6.3.6 Soluzioni deboli di equazioni differenziali stocastiche

Definizione 6.8 (soluzioni deboli di EDS)**.** *Si dice che l'equazione differenziale stocastica* (6.25) *ammette una* soluzione debole *se esistono uno spazio* $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathcal{F}}_t, \widetilde{\mathbf{P}})$*, un moto browniano* $\widetilde{W}$ *e un processo* $\widetilde{\xi}_t$*, che soddisfano il sistema* (6.25) *con* $\widetilde{W}$ *al posto di* W*, ossia*

$$\begin{cases} d\widetilde{\xi}_t = b(\widetilde{\xi}_t, t)dt + \sigma(\widetilde{\xi}_t, t)d\widetilde{W}_t \\ \widetilde{\xi}_s = x, \qquad\qquad\qquad\qquad\qquad\qquad x \in \mathbb{R}^m \end{cases} \tag{6.38}$$

Per dare un esempio di equazione differenziale debole consideriamo il seguente esempio:

Esempio 6.6 (Equazione di Tanaka)**.** *Si consideri l'equazione differenziale stocastica, **detta* **equazione di Tanaka***,*

$$dX_t = \operatorname{sign}(X_t)\, dW_t, \qquad X_0 = 0,$$

dove

$$\operatorname{sign}(x) = \begin{cases} +1 & \text{se } x > 0, \\ -1 & \text{se } x \leq 0. \end{cases}$$

Questa equazione ammette sempre una soluzione debole: Sia X_t *un processo di Wiener, e si definisca*

$$B_t := \int_0^t \operatorname{sign}(X_s)\, dX_s$$

Allora, poiché chiaramente $\operatorname{sign}^2(x) = 1$

$$X_t = \int_0^t \operatorname{sign}^2(X_s)\, dX_s = \int_0^t \operatorname{sign}(X_s)\, dB_s.$$

Tuttavia si può dimostrare che tale equazione non ammette soluzioni forti.

La dimostrazione è basata sulla formula di Tanaka, che, per ogni processo di Wiener W_t*, assicura che per ogni funzione convessa* h[33] *esiste un processo continuo e crescente e adattato (alla filtrazione generata dal processo di Wiener stesso) tale che*

$$h(W_t) = h(W_s) + \int_s^t h'_-(W_u)\, dW_u + A_t^h - A_s^h. \tag{6.39}$$

Nel caso in cui consideriamo la funzione $h(x) = |x|$*, tale processo si indica con* $L_t = L_t^{(W)}$ *ed è detto tempo locale (in* 0*) di* $(W_t)_{t\geq 0}$*, e si ottiene che*

$$|W_t| = \int_0^t \operatorname{sign}(W_s)\, dW_s + L_t.$$

Inoltre si può dimostrare che $L_t = L_t^{(W)}$ *è adattato anche rispetto alla filtrazione* $\mathcal{F}_t^{|W|}$ *generata da* $|W|$ *(strettamente contenuta in* $\mathcal{F}_t^W$*).*

Sia ora W_t *un processo di Wiener fissato e sia, per assurdo,* X_t *una soluzione forte **dell'equazione di Tanaka. Allora* X_t *è una martingala continua con* $X_t^2 - \int_0^t \operatorname{sign}^2{**}(X_s)\, ds = X_t^2 - t$ *una martingala. Quindi per il teorema di caratterizzazione di Levy, ***X_t *è un processo di Wiener. Applicando la formula di Tanaka per* $h(x) = |x|$ *al processo di Wiener* X_t *si ottiene che il processo*

$$\int_0^t \operatorname{sign}(X_s)\, dX_s = |X_t| - L_t^{(X)}$$

è un processo adattato alla filtrazione $\mathcal{F}_t^{|X|}$*, che è strettamente contenuta nella filtrazione* $\mathcal{F}_t^X$*. Quest'ultima è ovviamente contenuta nella filtrazione* $\mathcal{F}_t^W$*, essendo una soluzione forte dell'equazione di Tanaka. L'assurdo esce fuori nel momento in cui si osserva che*

$$\operatorname{sign}(X_s)\, dX_s = \operatorname{sign}(X_s)\operatorname{sign}(X_s)\, dW_s = dW_s,$$

si arriverebbe quindi all'assurdo che il processo W_t *è adattato ad una filtrazione **(ossia a* $\mathcal{F}_t^{|X|}$*) che è strettamente contenuta nella filtrazione generata da esso stesso **(ossia* $\mathcal{F}_t^{|X|} \subset \mathcal{F}_t^X \subseteq \mathcal{F}_t^W$*).*

[33] Attenzione è fondamentale che si tratti del caso a dimensione 1.

6.3.7 Trasformazione di Girsanov e soluzioni deboli

Il teorema seguente è conosciuto sotto il nome di teorema di Girsanov ed è la generalizzazione del risultato dell'Esercizio 4.1.

Teorema 6.17 (Teorema di Girsanov)**.** *Sia $(\theta_t)_{0\le t\le T}$ un processo progressivamente misurabile e limitato da una costante K, ossia $\sup_{s\in[0,T]}|\theta_s(\omega)|\le K$. Allora il processo $(L_t)_{0\le t\le T}$, definito come*

$$L_t = \exp\left\{\int_0^t \theta_s dW_s - \frac{1}{2}\int_0^t \theta_s^2 ds\right\}, \tag{6.40}$$

è una martingala non negativa con $\mathbb{E}^{\mathbb{P}}[L_t]=1$.

Inoltre, posta $\mathbb{Q}$, la misura di probabilità assolutamente continua rispetto a $\mathbb{P}$ su $\mathcal{F}_T$, con $d\mathbb{Q}=L_T d\mathbb{P}$, il processo $(B_t)_{0\le t\le T}$, definito da $B_t = W_t - \int_0^t \theta_s ds$, è un moto browniano standard.

Osservazione 6.9. *Il risultato rimane vero anche senza l'ipotesi di limitatezza per θ, ma è necessario che il processo L_t definito da (6.40) sia una martingala (in generale L_t è una supermartingala non negativa, con $\mathbb{E}^{\mathbb{P}}[L_t]\le 1$)*

Una condizione sufficiente affinché $(L_t)_{0\le t\le T}$ sia una martingala è la seguente

$$\mathbb{E}\left[\exp\left\{\frac{1}{2}\int_0^T \theta_t^2 dt\right\}\right] < \infty,$$

nota come condizione di Novikov.

Dimostrazione. La dimostrazione del fatto che L_t sia una $\mathbb{P}$ martingala si basa sul fatto che

$$dL_t = \theta_t\, L_t\, dW_t, \qquad L_t = 1 + \int_0^t \theta_s\, L_s\, dW_s \tag{6.41}$$

che si dimostra usando la formula di Itô alla funzione $f(x)=e^x$ al processo $X_t := \int_0^t \theta_s dW_s - \frac{1}{2}\int_0^t \theta_s^2 ds$.

Non rimane che dimostrare che $L_t \in M^2[0,T]$, cioè che $\mathbb{E}\big[\int_0^T L_t^2\, dt\big] < \infty$. Infatti allora, grazie alla limitatezza di θ_s anche il processo $\theta_t\, L_t \in M^2[0,T]$, e quindi dalla forma integrale di (6.41), si ottiene che L_t è una martingala[34].

[34] La dimostrazione della proprietà di integrabilità di L_t può essere fatta in diversi modi: procedendo come nella dimostrazione dell'esistenza della soluzione di una equazione differenziale stocastica con coefficienti Lipschitziani (in fondo si può pensare che L_t sia la soluzione di una equazione differenziale stocatica, ma con coefficienti aleatori e progressivamente misurabili); oppure introducendo il tempo d'arresto $\tau_M = \inf\{t>0\, t.c.\ L_t \ge M\}$, di modo che allora

$$L_{t\wedge\tau_M} = 1 + \int_0^{t\wedge\tau_M} \theta_s\, L_s\, dW_s = 1 + \int_0^{t\wedge\tau_M} \theta_s\, L_{s\wedge\tau_M}\, dW_s = 1 + \int_0^t \mathbb{I}_{\{s\le\tau_M\}}\theta_s\, L_{s\wedge\tau_M}\, dW_s$$

Allora

$$L^2_{t\wedge\tau_M} \le 2 + 2\left(\int_0^t \mathbb{I}_{\{s\le\tau_M\}}\theta_s\, L_{s\wedge\tau_M}\, dW_s\right)^2$$

e

$$\mathbb{E}\left[L^2_{t\wedge\tau_M}\right] \le 2 + 2\mathbb{E}\left[\int_0^t \mathbb{I}_{\{s\le\tau_M\}}|\theta|_s^2\, L^2_{s\wedge\tau_M}\, ds\right] \le 2 + 2\mathbb{E}\left[\int_0^t K^2\, L^2_{s\wedge\tau_M}\, ds\right]$$

Il lemma di Gronwall implica allora che $\mathbb{E}\left[L^2_{t\wedge\tau_M}\right] \le 2\, e^{2K^2 t}$, e passando al limite per M che tende ad infinito, con il lemma di Fatou, si ottiene che $\mathbb{E}\left[L_t^2\right] \le 2\, e^{2K^2 t}$, che conclude la dimostrazione,**previo la verifica che la successione τ_M sia una successione crescente (questo è ovvio) che converge all'infinito per M che tende all'infinito. **La dimostrazione è standard, ma vale la pena di essere riportata: essendo τ_M una successione crescente in senso lato, il limite di $\tau_M(\omega) =: \tau_\infty(\omega)$ esiste, si tratta di dimostrare che τ_∞vale infinito con probabilità 1. Ciò equivale a dimostrare che $\mathbb{P}(\tau_\infty \le T) = 0$ per ogni $T>0$. Poiché

$$\mathbb{P}(\tau_\infty \le T) = \lim_{M\to\infty} \mathbb{P}(\tau_M \le T),$$

basta dimostrare che tale limite vale 0. A questo scopo si osservi che, per ogni M si ha

$$\tau_M = \inf\{t>0\, t.c.\ L_t \ge M\} = \inf\{t>0\, t.c.\ X_t \ge \log(M)\},$$

e quindi

$$\{\tau_M \le T\} = \{\sup_{t\in[0,T]} X_t > \log(M)\}, \quad \text{dove} X_t := \int_0^t \theta_s dW_s - \frac{1}{2}\int_0^t \theta_s^2 ds,$$

da cui

$$\mathbb{P}(\tau_M \le T) = \mathbb{P}(\sup_{t\in[0,T]} X_t > \log(M)) \le \frac{\mathbb{E}(\sup_{t\in[0,T]}|X_t|^2}{\log^2(M)} \to 0 \quad \text{per } M\to\infty.$$

Si osservi che, anche **nel caso in cui θ_s è un processo a valori nei complessi si ha ancora che L_t è una martingala (ovviamente a valori nei complessi, però).

Per quanto riguarda invece il fatto che $B_t = W_t - \int_0^t \theta_s\, ds$ sia un moto browniano rispetto alla misura $\mathbb{Q}$ si basa sul fatto che è sufficiente mostrare che, per ogni u il processo $M_t^u := e^{i\,u\,B_t + \frac{1}{2}\,u^2\,t}$ è una martingala rispetto alla misura di probabilità $\mathbb{Q}$. Sappiamo che ciò equivale ad affermare che il processo $L_t\, M_t^u$ è una martingala rispetto a $\mathbb{P}$. Di nuovo la dimostrazione si basa sul fatto che

$$\begin{aligned} L_t\, M_t^u &= \exp\left\{\int_0^t \theta_s dW_s - \frac{1}{2}\int_0^t \theta_s^2 ds + i\,u\,B_t + \frac{1}{2}\,u^2\,t\right\} \\ &= \exp\left\{\int_0^t \theta_s dW_s - \frac{1}{2}\int_0^t \theta_s^2 ds + i\,u\,W_t - i\,u + \int_0^t \theta_s\, ds \frac{1}{2}\,u^2\,t\right\} \\ &= \exp\left\{\int_0^t (\theta_s + i\,u) dW_s - \frac{1}{2}\int_0^t (\theta_s + i\,u)^2 ds\right\} \end{aligned}$$

Si tratta allora di applicare a $\widetilde{\theta}_s = \theta_s + i\,u$ la generalizzazione del risultato ottenuto nella prima parte del teorema al caso di processi a valori nei complessi.

□

L'interesse del precedente teorema sta in una delle sue applicazioni principali, il risultato di esistenza per le soluzioni di equazioni differenziale stocastiche.

Sia $b(x,t)$ limitato e misurabile e si supponga che $\sigma(x,t)$, oltre a soddisfare le condizioni di Lipschitz globale (6.27) e di crescita (6.28), soddisfi anche la condizione che $\sigma(x,t) \geq a > 0$. Allora esiste una soluzione forte X_t del problema $dX_t = \sigma(X_t,t)\, dW_t$, $X_0 = x$, inoltre, **definito il processo L_t come

$$L_t = \exp\left\{\int_0^t \theta_s dW_s - \frac{1}{2}\int_0^t \theta_s^2 ds\right\}, \qquad \text{**con } \theta_t \text{ definito da} \quad \theta_t := \frac{b(X_t,t)}{\sigma(X_t,t)},$$

**si ha che L_t è una martingala, in quanto θ_t è limitato. Di conseguenza il processo $B_t = W_t - \int_0^t \frac{b(X_s,s)}{\sigma(X_s,s)}\, ds$ è un moto browniano rispetto alla misura $\mathbb{Q}$ del teorema di Girsanov.

Quindi

$$\begin{aligned} X_t &= x + \int_0^t \sigma(X_s,s)\, dW_s = x + \int_0^t \sigma(X_s,s)\, d\left(B_s + \frac{b(X_s,s)}{\sigma(X_s,s)}\, ds\right) \\ &= x + \int_0^t \sigma(X_s,s)\, dB_s + \int_0^t \sigma(X_s,s)\frac{b(X_s,s)}{\sigma(X_s,s)}\, ds \\ &= x + \int_0^t \sigma(X_s,s)\, dB_s + \int_0^t b(X_s,s)\, ds \end{aligned}$$

che dimostra che rispetto alla misura $\mathbb{Q}$ il processo X_t è soluzione dell'equazione differenziale stocastica $d\xi_t = b(\xi_t,t)\, dt + \sigma(\xi_t,t)\, dW_t$, $\xi_0 = x$. Si tratta quindi di una soluzione debole, ma attenzione non si richiede alcuna condizione di continuità per il coefficiente $b(x,t)$.

Altra applicazione, simile, del teorema di Girsanov sta nel fatto che permette di cambiare il drift in una equazione differenziale stocastica. Infatti, se sotto $\mathbb{P}$ il processo ξ_t è soluzione dell'equazione

$$d\xi_t = a(t,\xi_t)dt + \sigma(t,\xi_t)dW_t, \qquad \xi_0 = x,$$

allora sotto $\mathbb{Q}$, vale la seguente relazione

$$d\xi_t = a(t,\xi_t)dt + \sigma(t,\xi_t)(dB_t - \theta_t dt) = \big(a(t,\xi_t) - \theta_t\big)\, dt + \sigma(t,\xi_t)\, dB_t, \qquad \xi_0 = x.$$

Il valore atteso di $\sup_{t\in[0,T]} |X_t|^2$ è finito, grazie all'isometria di Ito e alla disuguaglianza di Doob, in quanto,

$$\sup_{t\in[0,T]} |X_t|^2 \leq 2 \sup_{t\in[0,T]} \left(\int_0^t \theta_s dW_s\right)^2 + 2\left(\frac{1}{2}\int_0^T \theta_s^2 ds\right)^2,$$

e il processo θ_s è limitato uniformemente in s ed ω da K. **

Può quindi essere possibile, cambiando misura di probabilità, rendere un processo soluzione di una equazione differenziale stocastica una martingala locale, prendendo $\theta_t = a(t, \xi_t)$ (ovviamente purché il processo L_t sia una martingala).

Capitolo 7

Rappresentazione di soluzioni di equazioni alle derivate parziali

7.1 Rappresentazione di Feynman-Kac per il processo di Wiener

Teorema 7.1. *Siano $f(x)$ e $c(x)$ due funzioni limitate. Si consideri il problema*

$$\begin{cases} u_t(t,x) = \frac{1}{2}\, u_{xx}(t,x) + c(x)\, u(x) \quad t > 0, \\ u(0,x) = f(x). \end{cases} \tag{7.1}$$

Se $u(t,x)$ è una soluzione limitata[1] *e di classe $C^{1,2}$ del problema* (7.1)*, allora $u(t,x)$ può essere rappresentata come*[2]

$$u(t,x) = \mathbf{E}_x \left[f(x(t))\, e^{\int_0^t c(x(s))\, ds} \right] \qquad \Leftrightarrow \qquad u(t,x) = \mathbb{E} \left[f(x+W_t)\, e^{\int_0^t c(x+W_s)\, ds} \right] \tag{7.2}$$

Ne discende quindi che esiste una sola soluzione limitata del problema (7.1)*.*

L'idea della dimostrazione sta nel trovare una martingala $(M_s)_{s\in[0,t]}$ che in un certo senso interpola $f(x)$ (che coincide con $f(x+W_0)$ sotto $\mathbb{P}$) e $f(x+W_t)\, e^{\int_0^t c(x+W_s)\, ds}$ La martingala cercata è definita come

$$M_s := u(t-s, x+W_s)\, e^{\int_0^s c(x+W_{s'})\, ds'}.$$

Infatti, se, come dimostreremo fra poco, $(M_s)_{s\in[0,t]}$ è una martingala, allora **$\mathbb{E}[M_0] = \mathbb{E}[M_t]$. Inoltre

$$M_0 = u(t, x+W_0)\, (= u(t,x)\mathbb{P} - q.c.)\,, \qquad M_t = u(0, x+W_t)\, e^{\int_0^t c(x+W_{s'})\, ds'} = f(x+W_t)\, e^{\int_0^t c(x+W_{s'})\, ds'}$$

e quindi

$$u(t,x) = \mathbb{E}[M_0] = \mathbb{E}[M_t] = \mathbb{E}\left[f(x+W_t)\, e^{\int_0^t c(x+W_{s'})\, ds'} \right].$$

Non rimane che dimostrare che $(M_s)_{s\in[0,t]}$ è una martingala. Possiamo utilizzare la formula di Itô alla funzione $g(s,y) = u(t-s,y)$ e al processo **$(W_s^x := x + W_s)_{s\in[0,t]}$ ****(tenendo conto che $dW_s^x = dW_s$)

$$dg(s, W_s^x) = g_s(s, W_s^x)\, ds + g_y(s, W_s^x)\, dW_s + \frac{1}{2}\, g_{yy}(s, W_s^x)\, ds$$

[1] **La condizione di limitatezza per la funzione u è necessaria se deve valere la rappresentazione (7.2). Infatti la funzione definita dal membro a destra di (7.2), cioè

$$(t,x) \to \mathbb{E}\left[f(x+W_t)\, e^{\int_0^t c(x+W_s)\, ds} \right]$$

è limitata uniformemente in $(x,t) \in \mathbb{R} \times [0,T]$, da $\|f\|\, e^{\|}c\|T$, (con $\sup_x |f(x)| \le \|f\|$ e similmente $\sup_x |c(x)| \le \|c\|$).

[2] ****Se con $\mathbf{E}_x$ si denota il valore atteso rispetto alla misura indotta $\mathbf{P}_x$ sullo spazio delle funzioni continue, dalla soluzione di $dX_t = dW_t,\ X_0 = x$, ossia il processo $X_t^x (= X_t^{0,x}) = x + W_t$, allora la seguente uguaglianza

$$\mathbf{E}_x \left[f(x(t))\, e^{\int_0^t c(x(s))} \right] = \mathbb{E}\left[f(x+W_t)\, e^{\int_0^t c(x+W_s)\, ds} \right]$$

discende dalla formula del cambio di misura.

ossia, tenuto conto che $g_s(s,y) = -u_t(t-s,y)$, $g_y(s,y) = u_x(t-s,y)$ e $g_{yy}(s,y) = u_{xx}(t-s,y)$,

$$du(t-s,W_s^x) = -u_t(t-s,W_s^x)\,ds + u_x(t-s,W_s^x)\,dW_s + \frac{1}{2}\,u_{xx}(t-s,W_s^x)\,ds$$

e quindi tenuto conto che, essendo u soluzione del problema (7.1), si ha $-u_t(t-s,y)+\frac{1}{2}\,u_{xx}(t-s,y) = -c(y)\,u(t-s,y)$

$$du(t-s,W_s^x) = -c(W_{t-s}^x)\,u(t-s,W_s^x)\,ds + u_x(t-s,W_s^x)\,dW_s.$$

D'altro canto, posto $Y_s := e^{\int_0^s c(x+W_{s'})\,ds'}\Big(= e^{\int_0^s c(W_{s'}^x)\,ds'}\Big)$, ovviamente $dY_s = c(W_s^x)\,Y_s\,ds$. Quindi, usando la formula del differenziale del prodotto di ottiene

$$\begin{aligned}
dM_s &= d\left(u(t-s,W_s^x)\,Y_s\right) = Y_s\,du(t-s,W_s^x) + u(t-s,W_s^x)\,dY_s \\
&= -Y_s\,c(W_{t-s}^x)\,u(t-s,W_s^x)\,ds + Y_s\,u_x(t-s,W_s^x)\,dW_s + u(t-s,W_s^x)\,c(W_s^x)\,Y_s\,ds \\
&= Y_s\,u_x(t-s,W_s^x)\,dW_s = e^{\int_0^t c(W_s^x)\,ds}\,u_x(t-s,W_s^x)\,dW_s
\end{aligned}$$

Da questa ultima espressione si vede subito che $(M_s)_{s\in[0,t]}$ è una martingala locale. Per mostrare che è una martingala abbiamo due possibilità: la prima più semplice aggiungere, ad esempio, un'ipotesi del tipo $\sup_{r\in[0,T]}\,|u_x(r,y)| \le C\,e^{a|y|}$ per un $a\in\mathbb{R}$, e poi dimostrare che

$$\mathbb{E}\left[\int_0^t \left|e^{\int_0^t c(W_s^x)\,ds}\,u_x(t-s,W_s^x)\right|^2 ds\right] \le e^{2\,\|c\|_\infty\,t}\,\mathbb{E}\left[\int_0^t C^2\,e^{2\,a\,|W_s^x|}\,ds\right] < \infty.$$

Tuttavia questa ipotesi non serve in quanto dalla definizione sappiamo che

$$\sup_{s\in[0,t]}|M_s| \le \|u\|_\infty\,e^{\|c\|_\infty\,t},$$

e questo fatto garantisce che la martingala locale M_s è una martingala[3].

7.1.1 Applicazione: Formula dell'arcoseno (di Lévy)

Sia τ_t il tempo trascorso dal processo di Wiener nella semiretta $(0,+\infty)$, ossia

$$\tau_t = \int_0^t \mathbf{1}_{(0,\infty)}(W_s)\,ds.$$

vale il seguente risultato: ****per ogni $x\in(0,1)$,

$$\mathbb{P}(\tau_t \le xt) = \frac{2}{\pi}\,\arcsin\sqrt{x} = \int_0^x \frac{du}{\sqrt{u(1-u)}}.$$

Prendendo $f(x)=1$ e $c(x) = -\lambda\,\mathbf{1}_{(0,\infty)}(x)$ possiamo tradurre il problema di trovare la trasformata di Laplace di τ_t in un problema di equazioni alla derivate parziali

$$\mathbb{E}_x\left[e^{-\lambda\,\tau_t}\right] = \mathbb{E}_x\left[e^{\int_0^t c(W_s)\,ds}\right].$$

Per i dettagli rimandiamo ad esempio al libro di Steele [15].

[3] Vale infatti il seguente risultato:

Proposizione. *Sia M_t una martingala locale, ****a traiettorie continue (oppure caldlag e con la proprietà che, per ogni $t\in[0,T]$, si abbia $\mathbb{P}(M_t = M_{t^-}) = 1$)**e con la proprietà che $\sup_{t\in[0,T]}|M_t| \le B$, con B una costante deterministica. Allora M_t è una martingala.*

Dimostrazione. Per ipotesi sappiamo che esiste una successione di tempi d'arresto τ_n per i quali $M_{t\wedge\tau_n}$ è una martingala, ossia

$$\mathbb{E}\left[M_{t\wedge\tau_n}\middle|\mathcal{F}_s\right] = M_{s\wedge\tau_n}.$$

Per la continuità delle traiettorie $\lim_{n\to\infty} M_{t\wedge\tau_n} = M_t$ (se fosse una martingala *cadlag* si avrebbe $\lim_{n\to\infty} M_{t\wedge\tau_n} = M_{t^-}$, ma coinciderebbe quasi certamente con M_t per l'ipotesi fatta). Ovviamente lo stesso vale con s al posto di t. Inoltre, per ogni $t\le T$, grazie all'ipotesi di limitatezza, si ha che $|M_{t\wedge\tau_n}| \le B$. Quindi, per il teorema della convergenza dominata sotto il segno di valore atteso condizionale,

$$M_s = \lim_{n\to\infty} M_{s\wedge\tau_n} = \lim_{n\to\infty}\mathbb{E}\left[M_{t\wedge\tau_n}\middle|\mathcal{F}_s\right] = \mathbb{E}\left[\lim_{n\to\infty} M_{t\wedge\tau_n}\middle|\mathcal{F}_s\right] = \mathbb{E}\left[M_t\middle|\mathcal{F}_s\right],$$

ossia che M_t è una martingala.

□

7.1.2 Generalizzazioni

Questo risultato ammette diverse generalizzazioni, sia a processi più generali, sia sotto condizioni più deboli. Tuttavia è necessario dare alcune condizioni sulla classe delle soluzioni in cui vale un risultato di unicità.

Va anche tenuto presente che spesso, quando ci si riferisce alla formula di Feynman-Kac, si considera il problema con condizione finale, invece che iniziale, ossia al problema

$$\begin{cases} -u_t(t,x) = \frac{1}{2}\, u_{xx}(t,x) + c(x)\, u(x), \quad t \in [0,T), \\ u(T,x) = f(x). \end{cases} \tag{7.3}$$

In tale caso la rappresentazione diviene

$$u(t,x) = \mathbb{E}\left[f(x+W_{T-t})\, e^{\int_0^{T-t} c(x+W_s)\, ds}\right] = \mathbf{E}_x\left[f(x(T-t))\, e^{\int_0^{T-t} c(x(s))\, ds}\right] \text{****} \tag{7.4}$$

o come si scrive usualmente, prendendo come istante iniziale t

$$u(t,x) = \mathbb{E}\left[f(W_T)\, e^{\int_t^T c(W_s)\, ds} | W_t = x\right] \tag{7.5}$$

Va inoltre osservato che non serve alcuna ipotesi di regolarità sul dato iniziale (o finale f) e neppure sulla funzione c.

Infine notiamo il fatto che ci sia una rappresentazione permette, a volte di fare il seguente procedimento:

• Definire

$$\overline{u}(t,x) = \mathbb{E}_x\left[f(W_t)\, e^{\int_0^t c(W_s)\, ds}\right]$$

• Provare a verificare se tale funzione $\overline{u}$ è soluzione del problema cercato.

Altre generalizzazioni sono possibili considerando un processo ****X_t^x al posto di $x + W_t$, con X_t^x soluzione dell'equazione differenziale stocastica,

$$dX_t = b(X_t)\, dt + \sigma(X_t)\, dW_t, \qquad X_0 = x,$$

sostituendo l'operatore $\frac{1}{2}\frac{\partial^2}{\partial x^2}$ con l'operatore

$$L = b(x)\, \frac{\partial}{\partial x} + \frac{1}{2}\, \sigma(x)\, \frac{\partial^2}{\partial x^2},$$

ossia sostituendo $\frac{1}{2}u_{xx}(t,x)$ con $b(x)\, u_x(t,x) + \frac{1}{2}\, \sigma(x)\, u_{xx}(t,x)$, nell'equazione (7.1).

Teorema 7.2. *Siano $f(x)$ e $c(x)$ due funzioni limitate. Si consideri il problema*

$$\begin{cases} u_t(t,x) = b(x)\, u_x(t,x) + \frac{1}{2}\, \sigma^2(x) u_{xx}(t,x) + c(x)\, u(x) \quad t > 0, \\ u(0,x) = f(x). \end{cases} \tag{7.6}$$

Si assuma che i coefficienti b e σ soddisfino condizioni che assicurano l'esistenza e l'unicità (in legge) della soluzione X_t^x dell'equazione differenziale stocastica (ad esempio condizioni di lipschitzianità globale)

$$dX_t = b(X_t)dt + \sigma(X_t)\, dW_t,\ X_0 = x.$$

Se $u(t,x)$ è una soluzione limitata e di classe $C^{1,2}$ del problema (7.6), allora $u(t,x)$ può essere rappresentata come[4]

$$u(t,x) = \mathbf{E}_x\left[f(x(t))\, e^{\int_0^t c(x(s))\, ds}\right] \qquad \Leftrightarrow \qquad u(t,x) = \mathbb{E}\left[f(X_t^x)\, e^{\int_0^t c(X_s^x)\, ds}\right]. \tag{7.7}$$

Ne discende quindi che esiste una sola soluzione limitata del problema (7.6).

[4] ****Se con $\mathbf{E}_x$ si denota il valore atteso rispetto alla (unica per ipotesi) misura indotta $\mathbf{P}_x$ sullo spazio delle funzioni continue, dalla soluzione di $dX_t = b(X_t)dt + \sigma(X_t)\, dW_t,\ X_0 = x$. Posto X_t^x tale soluzione, allora la seguente uguaglianza

$$\mathbf{E}_x\left[f(x(t))\, e^{\int_0^t c(x(s))}\right] = \mathbb{E}\left[f(X_t^x)\, e^{\int_0^t c(X_s^x)\, ds}\right]$$

discende dalla formula del cambio di misura.

L'idea della dimostrazione, come nel caso precedente, sta nel trovare una martingala $(M_s)_{s\in[0,t]}$ che in un certo senso interpola $f(x)$ $\big($che coincide con $f(X_0^x)$ sotto $\mathbb{P}$, o con $f(x(0))$ sotto $\mathbf{P}_x\big)$ e $f(X_t^x)\,e^{\int_0^t c(X_s^x)\,ds}$. La martingala cercata è definita[5] come

$$M_s := u(t-s, X_s^x)\,e^{\int_0^t c(X_s^x)\,ds}.$$

Infatti, se, come dimostreremo fra poco, $(M_s)_{s\in[0,t]}$ è una martingala, allora

$$M_0 = u(t, X_0^x)\,(= u(t,x)\ \mathbb{P}-q.c.)\,, \qquad M_t = u(0, X_t^x)\,e^{\int_0^t c(X_s^x)\,ds} = f(X_t^x)\,e^{\int_0^t c(X_s^x)\,ds}$$

e quindi

$$u(t,x) = \mathbb{E}\,[M_0] = \mathbb{E}\,[M_t] = \mathbb{E}\left[f(X_t^x)\,e^{\int_0^t c(X_s^x)\,ds}\right].$$

Non rimane che dimostrare che $(M_s)_{s\in[0,t]}$ è una martingala. Possiamo utilizzare la formula di Itô alla funzione $g(s,y) = u(t-s,y)$ e al processo $(X_s^x)_{s\in[0,t]}$

$$dg(s, X_s^x) = g_s(s, X_s^x)\,ds + g_y(s, X_s^x)\,dX_s^x + \frac{1}{2}\,g_{yy}(s, X_s^x)\,\sigma^2(X_s^x)\,ds$$

ossia, tenuto conto che $g_s(s,y) = -u_t(t-s,y)$, $g_y(s,y) = u_x(t-s,y)$ e $g_{yy}(s,y) = u_{xx}(t-s,y)$,

$$du(t-s, X_s^x) = -u_t(t-s, X_s^x)\,ds + u_x(t-s, X_s^x)\,\big(b(X_s^x)\,ds + \sigma(X_s^x)\,dW_s\big) + \frac{1}{2}\,\sigma^2(X_s^x)\,u_{xx}(t-s, X_s^x)\,ds$$

e quindi tenuto conto che, essendo u soluzione del problema (7.6), si ha

$$-\,u_t(t-s,y) + b(y)\,u_x(t-s,y) + \frac{1}{2}\,\sigma^2(y)u_{xx}(t-s,y) = -c(y)\,u(t-s,y)$$

e quindi

$$du(t-s, X_s^x) = -c(X_{t-s}^x)\,u(t-s, X_s^x)\,ds + u_x(t-s, X_s^x)\,\sigma(X_s^x)\,dW_s.$$

D'altro canto, posto $Y_s := e^{\int_0^s c(X_{s'}^x)\,ds'}$, ovviamente $dY_s = c(X_s^x)\,Y_s\,ds$. Quindi, usando la formula del differenziale del prodotto di ottiene

$$\begin{aligned} dM_s &= d\,(u(t-s, X_s^x)\,Y_s) = Y_s\,du(t-s, X_s^x) + u(t-s, X_s^x)\,dY_s \\ &= -Y_s\,c(X_{t-s}^x)\,u(t-s, X_s^x)\,ds + Y_s\,u_x(t-s, X_s^x)\,dW_s + u(t-s, X_s^x)\,c(X_s^x)\,Y_s\,ds \\ &= Y_s\,u_x(t-s, X_s^x)\,dW_s = e^{\int_0^t c(X_s^x)\,ds}\,u_x(t-s, X_s^x)\,dW_s \end{aligned}$$

Da questa ultima espressione si vede subito che M_s è una martingala locale. Per mostrare che è una martingala abbiamo due possibilità: la prima più semplice aggiungere un'ipotesi del tipo $\sup_{r\in[0,T]}|u_x(r,y)| \le C\,e^{\alpha y^2}$, per α sufficientemente piccolo e dimostrare che

$$\mathbb{E}\left[\int_0^t \left|e^{\int_0^t c(X_s^x)\,ds}\,u_x(t-s, X_s^x)\right|^2 ds\right] \le e^{2\,\|c\|_\infty\,t}\,\mathbb{E}\left[\int_0^t C^2\,e^{2\,a\,(X_s^x)^2}\,ds\right] < \infty.$$

Tuttavia questa ipotesi non serve in quanto dalla definizione sappiamo che

$$\sup_{s\in[0,t]}|M_s| \le \|u\|_\infty\,e^{\|c\|_\infty\,t},$$

e questo fatto garantisce che la martingala locale M_s è una martingala (si veda la Proposizione nella Nota 3).

[5] Si potrebbe equivalentemente, anzi meglio, nel caso si avesse solo esistenza e unicità della soluzione in senso debole, definire una martingala, sotto la misura $\mathbf{P}_x$, nello spazio delle funzioni continue, $\mathcal{M}_t$ che interpola $f(x)$ e $f(x(t))\,e^{\int_0^t c(x(s))\,ds}$.

7.1.3 Connessioni con il problema di Dirichlet

(DA Finire di SCRIVERE e RIVEDERE!!!)

**Per problema di Dirichlet si intende un problema del tipo

$$\begin{cases} Lu(x) = 0 & x \in \overset{\circ}{D}, \\ u(x) = f(x) & x \in \partial D \end{cases}$$

dove D è un dominio regolare di $\mathbb{R}^n$, connesso, ed L è un operatore differenziale del secondo ordine.

Sotto opportune ipotesi si ottiene una rappresentazione della soluzione $u(x)$,

$$u(x) = \mathbb{E}\big[f(X_\tau^x)\big],$$

dove $(X_t^x)_{t\geq 0}$ è la soluzione di una equazione differenziale stocastica multidimensionale, con la proprietà che

$$g(X_t) = g(X_0) + \int_0^t Lg(X_s)\,ds + \int_0^t \nabla g(X_s)\,dW_s,$$

e

$$\tau = \tau_D = \inf\{t > 0 \,:\, X_t^x \notin D\}.$$

Nel capitolo precedente, in particolare nelle Sezioni 6.3.5 e 6.2.3, abbiamo visto il problema di Dirichlet in dimensione 1, in un caso particolare. In $\mathbb{R}$ si tratta di prendere $D = [ab]$, $Lu = b(x)u'(u) + \frac{1}{2}\,u''(x)$

$$\begin{cases} Lu(x) = 0 & x \in (a,b), \\ u(a) = f(a) & u(b) = f(b). \end{cases}$$

Abbiamo considerato solo il caso in cui $f(a) = 1$ e $f(b) = 0$. Tuttavia la dimostrazione che la soluzione sia data da

$$u(x) = \mathbb{E}\big[f(X_{\tau(a,b)}^x)\big]$$

si effettua nello stesso modo.

7.1.4 Il processo di Bessel

(NON RIGUARDATO!!)

Sia dato un processo di Wiener $W_t = (W_t^{(1)}, \cdots, W_t^{(n)})$, con $n \geq 2$ e con $W_0 = x$, per un punto $x \neq 0$. Allora si ha che[6]

$$\begin{aligned} |W_t| &= x + \int_0^t \sum_{i=1}^n \frac{W_s^{(i)}}{|W_s|}\, dW_s^{(i)} + \frac{n-1}{2} \int_0^t \frac{1}{|W_s|}\, ds \\ &= B_t + \frac{n-1}{2} \int_0^t \frac{1}{|W_s|}\, ds, \end{aligned}$$

dove B_t è un processo di Wiener: infatti è a traiettorie continue, è una martingala, in quanto somma di martingale, e inoltre è una martingala anche il processo

$$B_t^2 - \int_0^t \sum_{i=1}^n \left(\frac{W_s^{(i)}}{|W_s|} \right)^2 ds = B_t^2 - \int_0^t \frac{\sum_{i=1}^n \left(W_s^{(i)} \right)^2}{|W_s|^2} ds = B_t^2 - t,$$

e quindi B_t è un processo di Wiener, grazie al teorema di caratterizzazione di Levy.

Abbiamo quindi che $|W_t|$ è, rispetto alla misura $\mathbb{P}_x$, soluzione debole dell'equazione differenziale stocastica

$$d\xi_t = \frac{n-1}{2} \frac{1}{\xi_t}\, dt + dW_t, \qquad \xi_0 = x.$$

Si può inoltre dimostrare che il processo di Bessel è definito anche sostituendo n con $\delta \geq 1$ e che, nel caso $\delta \geq 2$ un semigruppo (Markoviano e omogeneo) di operatori $T_t^{(\delta)}$ con probabilità di transizione

$$p^{(\delta)}(t,x,y)dy = \mathbf{1}_{(0,\infty)}(y)\, \frac{x}{t} \left(\frac{y}{x} \right)^{\delta/2} I_{\frac{\delta}{2}-1}(xy/t)\, \exp\{-(x^2+y^2)/2t\}, \qquad x > 0,$$

dove I_ν è la funzione di Bessel modificata[7] di indice ν e

$$p^{(\delta)}(t,0,y)dy = \mathbf{1}_{(0,\infty)}(y)\, \Gamma(\delta/2)\, 2^{\frac{\delta}{2}-1}\, t^{-\frac{\delta}{2}}\, y^{\delta-1}\, \exp\{-y^2/2t\}.$$

---

[6] Dalla formula di Itô, applicata al processo di Wiener W_t e alla funzione

$$f(x) = f(x_1, \cdots, x_n) = |(x_1, \cdots, x_n)| = \sqrt{x_1^2 + \cdots + x_n^2}$$

per la quale

$$f_{x_i}(x) = \frac{1}{2} \frac{2\, x_i}{\sqrt{x_1^2 + \cdots + x_n^2}} = \frac{x_i}{|x|},$$

$$f_{x_i x_i}(x) = \frac{1}{|x|} - \frac{1}{2} \frac{2\, x_i^2}{(x_1^2 + \cdots + x_n^2)^{\frac{3}{2}}} = \frac{1}{|x|} - \frac{x_i^2}{|x|^3}$$

si ottiene

$$\begin{aligned} d|W_t| &= \sum_{i=1}^n \frac{W_t^{(i)}}{|W_t|}\, dW_t^{(i)} + \frac{1}{2} \sum_{i=1}^n \left(\frac{1}{|W_t|} - \frac{(W_t^{(i)})^2}{|W_t|^3} \right) dt \\ &= \sum_{i=1}^n \frac{W_t^{(i)}}{|W_t|}\, dW_t^{(i)} + \frac{1}{2} \left(n \frac{1}{|W_t|} - \frac{|W_t|^2}{|W_t|^3} \right) dt \\ &= \sum_{i=1}^n \frac{W_t^{(i)}}{|W_t|}\, dW_t^{(i)} + \frac{n-1}{2} \frac{1}{|W_t|}\, dt. \end{aligned}$$

Questo differenziale ha senso fino al primo istante in cui $|W_t| = 0$. Già abbiamo visto che, il processo di Wiener è transiente per $n \geq 3$ (ma va detto che vale anche per $n = 2$, caso in cui è ricorrente) tale tempo vale infinito con probabilità 1.

[7] Ricordiamo che funzione di Bessel modificata $I_\nu(x)$ è la soluzione dell'equazione differenziale del secondo ordine

$$z^2\, w'' + z\, w' - (z^2 + \nu^2) = 0.$$

$$I_\nu(x) = \left(\frac{x}{2} \right)^\nu \sum_{i=0}^\infty \frac{1}{\Gamma(\nu+i+1)\, i!} \left(\frac{x}{4} \right)^{2i},$$

Per il caso $n = 1$ le cose non sono così semplici: ci vuole una formula a parte: la formula di Tanaka.

Le cose sono più semplici per il caso in cui si consideri invece il modulo al quadrato di un processo di Wiener n-dimensionale[8], con $n \geq 1$ e qualunque sia x:

$$|W_t|^2 = |x|^2 + 2\int_0^t \sum_{i=1}^n W_s^{(i)}\, dW_s^{(i)} + n\, t$$

Tenendo conto del fatto che

$$\int_0^t \sum_{i=1}^n W_s^{(i)}\, dW_s^{(i)} = \int_0^t \sum_{i=1}^n |W_s|\, \frac{W_s^{(i)}}{|W_s|}\, dW_s^{(i)}$$

e che, come sappiamo $B_t = \int_0^t \sum_{i=1}^n, \frac{W_s^{(i)}}{|W_s|}\, dW_s^{(i)}$ è un processo di Wiener, possiamo affermare che il processo

$$Y_t := |W_t|^2, \qquad Y_0 = y = |x|^2$$

è soluzione dell'equazione differenziale stocastica

$$dY_t = y + 2\int_0^t \sqrt{Y_s}\, dB_s + n\, t$$

Questo processo è un caso particolare di una classe più generale di processi: i processi di Bessel al quadrato di dimensione δ, con $\delta \geq 0$ (square of a δ dimensional Bessel process $BESQ^\delta$), che sono definiti come le soluzioni dell'equazione differenziale

$$d\xi_t = y + 2\int_0^t \sqrt{\xi_s}\, d\beta_s + \delta\, t, \qquad y \geq 0$$

dove β_t è un processo di Wiener.

Se invece definissi

$$\mathcal{I}_\nu(x) = \sum_{i=0}^\infty \frac{1}{\Gamma(\nu+i+1)\, i!} \left(\frac{x}{2}\right)^{\nu+2i},$$

quale equazione differenziale soddisferebbe?

$$\frac{d}{dx}\, \mathcal{I}_\nu(x) = \sum_{i=0}^\infty \frac{1}{\Gamma(\nu+i+1)\, i!} (\nu+2i) \left(\frac{x}{2}\right)^{\nu+2i-1} \frac{1}{2},$$

$$\frac{d^2}{dx^2}\, \mathcal{I}_\nu(x) = \sum_{i=0}^\infty \frac{1}{\Gamma(\nu+i+1)\, i!} (\nu+2i)(\nu+2i-1) \left(\frac{x}{2}\right)^{\nu+2i-2} \frac{1}{4},$$

quindi

$$x^2 \frac{d^2}{dx^2}\, \mathcal{I}_\nu(x) = \sum_{i=0}^\infty \frac{1}{\Gamma(\nu+i+1)\, i!} (\nu+2i)(\nu+2i-1) \left(\frac{x}{2}\right)^{\nu+2i} \frac{1}{4}$$

$$x\, \frac{d}{dx}\, \mathcal{I}_\nu(x) = \sum_{i=0}^\infty \frac{1}{\Gamma(\nu+i+1)\, i!} (\nu+2i) \left(\frac{x}{2}\right)^{\nu+2i} \frac{1}{2},$$

Inoltre ricordiamo che, ovviamente per $\nu \leq 0$

$$I_\nu(x) = I_{-\nu}(x).$$

[8]Dalla formula di Itô, applicata al processo di Wiener W_t e alla funzione

$$f(x) = f(x_1, \cdots, x_n) = |(x_1, \cdots, x_n)|^2 = x_1^2 + \cdots + x_n^2$$

per la quale

$$f_{x_i}(x) = 2\, x_i, \qquad f_{x_i x_i}(x) = 2$$

si ottiene

$$d|W_t|^2 = \sum_{i=1}^n 2\, W_t^{(i)}\, dW_t^{(i)} + \frac{1}{2} \sum_{i=1}^n 2\, dt = 2 \sum_{i=1}^n W_t^{(i)}\, dW_t^{(i)} + n\, dt$$

Si può dimostrare che tale equazione differenziale ammette soluzione forte unica $Y_t = Y_t^{y,\delta}$.

Sia Q_y^δ la legge (sullo spazio delle funzioni continue) del processo soluzione della precedente equazione differenziale. Allora vale la seguente proprietà

Proposizione 7.3 (Revuz e Yor [11]). *Per ogni δ, $\delta' \geq 0$ e per ogni y, $y' \geq 0$, si ha*

$$Q_{y+y'}^{\delta+\delta'} = Q_y^\delta * Q_{y'}^{\delta'}$$

ossia $Q_{y+y'}^{\delta+\delta'}$ è la convoluzione di Q_y^δ e $Q_{y'}^{\delta'}$.

Dimostrazione. L'idea della dimostrazione sta nel fatto che, se definiamo il processo

$$X_t := Y_t + Y_t' = y + y' + 2\int_0^t \left(\sqrt{Y_s}\, dB_s + \sqrt{Y_s'}\, dB_s'\right) + \left(\delta + \delta'\right) t$$

dove B_t e B_t' sono due processi di Wiener indipendenti, e $Y_t = Y_t^{y,\delta}$ e $Y_t' = Y_t^{y',\delta'}$ sono $BESQ^\delta$ e $BESQ^{\delta'}$, allora X_t ha la legge $Q_y^\delta * Q_{y'}^{\delta'}$. Inoltre si può dimostrare che X_t è anche soluzione dell'equazione

$$\begin{aligned}
X_t &= y + y' + 2\int_0^t \mathbb{I}_{\{X_s>0\}} \sqrt{X_s}\, \frac{\sqrt{Y_s}\, dB_s + \sqrt{Y_s'}\, dB_s'}{\sqrt{X_s}} + 2\int_0^t \mathbb{I}_{\{X_s=0\}} \left(\sqrt{Y_s}\, dB_s + \sqrt{Y_s'}\, dB_s'\right) + \left(\delta + \delta'\right) t \\
&= y + y' + 2\int_0^t \mathbb{I}_{\{X_s>0\}} \sqrt{X_s}\, \frac{\sqrt{Y_s}\, dB_s + \sqrt{Y_s'}\, dB_s'}{\sqrt{X_s}} + \left(\delta + \delta'\right) t \\
&= y + y' + 2\int_0^t \mathbb{I}_{\{X_s>0\}} \sqrt{X_s}\, \frac{\sqrt{Y_s}\, dB_s + \sqrt{Y_s'}\, dB_s'}{\sqrt{X_s}} + 2\int_0^t \mathbb{I}_{\{X_s=0\}} \sqrt{X_s}\, dB_s'' + \left(\delta + \delta'\right) t
\end{aligned}$$

dove B_t'' è un processo di Wiener indipendente da B_t e B_t'. Ovviamente le precedenti identità sono da intendersi quasi certamente e sono valide perché $\int_0^t \mathbb{I}_{\{X_s=0\}} X_s\, ds = 0$, $\int_0^t \mathbb{I}_{\{X_s=0\}} Y_s\, ds = 0$ e $\int_0^t \mathbb{I}_{\{X_s=0\}} Y_s'\, ds = 0$, in quanto dove $X_s = 0$, allora ovviamente anche $Y_s = Y_s' = 0$.

Inoltre basta verificare che il processo definito come

$$\gamma_t = \int_0^t \mathbb{I}_{\{X_s>0\}} \frac{\sqrt{Y_s}\, dB_s + \sqrt{Y_s'}\, dB_s'}{\sqrt{X_s}} + 2\int_0^t \mathbb{I}_{\{X_s=0\}}\, dB_s''$$

è ancora un processo di Wiener, per cui

$$X_t = y + y' + \int_0^t \sqrt{X_s}\, d\gamma_s + \left(\delta + \delta'\right) t$$

□

Proposizione 7.4. *Sia μ una misura (non negativa) su $\mathbb{R}^+$ tale che $\int_{\mathbb{R}^+} (1+s)\, \mu(ds) < \infty$. Allora esistono due numeri $A = A_\mu$ e $B = B_\mu$, che dipendono dalla misura μ, tali che, per ogni y e δ in $[0,\infty)$, si ha*

$$\mathbb{E}\left[\exp\left\{-\int_{\mathbb{R}^+} Y_s^{y,\delta}\, \mu(ds)\right\}\right] = (A_\mu)^\delta\, (B_\mu)^y\,.$$

In particolare, prendendo $\mu(ds) = \lambda\, \delta_t(ds)$, esistono due numeri $A = A_{\lambda,t}$ e $B = B_{\lambda,t}$, che dipendono da λ e t e tali che, per ogni y e δ in $[0,\infty)$, si ha

$$\mathbb{E}\left[e^{-\lambda\, Y_t^{y,\delta}}\right] = (A_{\lambda,t})^\delta\, (B_{\lambda,t})^y\,.$$

Si può trovare facilmente la legge del processo, nel caso $\delta = 1$ in quanto sappiamo che $Y_t^{y,\delta} = |\sqrt{y} + W_t|^2$, per un processo di Wiener unidimensionale, allora la trasformata di Laplace[9]

$$\mathbb{E}\left[e^{-\lambda\, Y_t^{y,1}}\right] = \mathbb{E}_{\sqrt{y}}\left[e^{-\lambda\,|W_t|^2}\right] = (1 + 2\,\lambda\, t)^{-\frac{1}{2}}\ \exp\{-\lambda y/(1 + 2\lambda\, t)\}$$

dove W_t è qui un processo di Wiener unidimensionale.

Poi, grazie alla proposizione precedente, si può affermare che

$$\mathbb{E}\left[e^{-\lambda\, Y_t^{y,\delta}}\right] = (1 + 2\,\lambda\, t)^{-\frac{\delta}{2}}\ \exp\{-\lambda y/(1 + 2\lambda\, t)\}$$

da cui, invertendo la trasformata di Laplace, si ottengono le probabilità di transizione, per $\delta > 0$

$$q^{(\delta)}(t, x, y)dy = \mathbf{1}_{(0,\infty)}(y)\,\frac{1}{2}\,\left(\frac{y}{x}\right)^{\nu(\delta)/2}\ I_{\nu(\delta)}(\sqrt{xy}/t)\ \exp\{-(x + y)/2t\}, \qquad x > 0,$$

dove $\nu(\delta) = \frac{\delta}{2} - 1$, e I_ν è la funzione di Bessel modificata, mentre

$$q^{(\delta)}(t, 0, y)dy = \mathbf{1}_{(0,\infty)}(y)\,\Gamma^{-1}(\delta/2)\,(2t)^{-\frac{\delta}{2}}\,y^{\delta-1}\ \exp\{-y/2t\}.$$

Il generatore del processo $BESQ^\delta$ (almeno formalmente) è dato da

$$Lf(x) = \delta\, f'(x) + x\, f''(x)$$

9

$$\begin{aligned}
\mathbb{E}_{\sqrt{y}}\left[e^{-\lambda\,|W_t|^2}\right] &= \mathbb{E}_0\left[e^{-\lambda\,|W_t+\sqrt{y}|^2}\right] = \int_{\mathbb{R}} \frac{1}{\sqrt{2\,\pi\, t}}\, e^{-\lambda\, z^2}\, e^{-\frac{(z-\sqrt{y})^2}{2\,t}}\, dz \\
&= \int_{\mathbb{R}} \frac{1}{\sqrt{2\,\pi\, t}}\, e^{-\frac{2\,\lambda\, z^2\, t+(z-\sqrt{y})^2}{2\,t}}\, dz = \int_{\mathbb{R}} \frac{1}{\sqrt{2\,\pi\, t}}\, e^{-\frac{2\,\lambda\, z^2\, t+z^2-2\,z\sqrt{y}+y}{2\,t}}\, dz \\
&= \int_{\mathbb{R}} \frac{1}{\sqrt{2\,\pi\, t}}\, e^{-\frac{(2\,\lambda\, t+1)\, z^2-2\,z\sqrt{y}+y}{2\,t}}\, dz = \int_{\mathbb{R}} \frac{1}{\sqrt{2\,\pi\, t}}\, e^{-\frac{(2\,\lambda\, t+1)\,(z^2-2\,z\sqrt{y}(2\,\lambda\, t+1)^{-}1+y(2\,\lambda\, t+1)^{-}1}{2\,t}}\, dz \\
&= \int_{\mathbb{R}} \frac{1}{\sqrt{2\,\pi\, t}}\, e^{-\frac{(2\,\lambda\, t+1)\,\left(z^2-2\,z\sqrt{y}(2\,\lambda\, t+1)^{-}1+y(2\,\lambda\, t+1)^{-}2\right)}{2\,t}}\, dz\, e^{-\frac{y[1-(2\,\lambda\, t+1)^{-}1]}{2\,t}} \\
&= \frac{1}{(2\,\lambda\, t+1)^{\frac{1}{2}}}\, e^{-\frac{y[(2\,\lambda\, t+1)-1]}{2\,t\,(2\,\lambda\, t+1)}} = (2\,\lambda\, t+1)^{-\frac{1}{2}}\, e^{-y\,\frac{\lambda}{2\,\lambda\, t+1}}
\end{aligned}$$

Bibliografia

[1] Baldi, P. *Equazione differenziali stocastiche e applicazioni.* Quaderni dell'Unione Matematica Italiana, 28. Bologna: Pitagora Editrice. VIII, 309 p. (seconda edizione), 2000.

[2] Bertini, L. Note del corso di Dottorato Metodi probabilistici nelle equazioni alle derivate parziali. Dip. Mat. Guido Castelnuovo- Univ. di Roma La Sapienza - Disponibili sul Web all'indirizzo `http://carpenter.mat.uniroma1.it/lorenzo/didattica/dottorato`, 2007.

[3] Billingsley, P. *Probability and measure*, third ed. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1995. A Wiley-Interscience Publication.

[4] Billingsley, P. *Convergence of probability measures*, second ed. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, 1999. A Wiley-Interscience Publication.

[5] Breiman, L. *Probability*, vol. 7 of *Classics in Applied Mathematics.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992. Corrected reprint of the 1968 original.

[6] Feller, W. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

[7] Karatzas, I., and Shreve, S. E. *Brownian motion and stochastic calculus*, vol. 113 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 1988.

[8] Lamberton, D., and Lapeyre, B. *Introduction to stochastic calculus applied to finance.* Chapman & Hall, London, 1996. Translated from the 1991 French original by Nicolas Rabeau and François Mantion.

[9] Øksendal, B. *Stochastic differential equations*, sixth ed. Universitext. Springer-Verlag, Berlin, 2003. An introduction with applications.

[10] Pinsky, M. A. Brownian continuity modulus via series expansions. *J. Theoret. Probab. 14*, 1 (2001), 261–266.

[11] Revuz, D., and Yor, M. *Continuous martingales and Brownian motion*, third ed., vol. 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, 1999.

[12] Ross, S. M. *Stochastic processes*, second ed. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, 1996.

[13] Ross, S. M. *An introduction to mathematical finance.* Cambridge University Press, Cambridge, 1999. Options and other topics.

[14] Shiryaev, A. N. *Essentials of stochastic finance. Facts, models, theory,*, vol. 3 of *Advanced Series on Statistical Science & Applied Probability.* World Scientific Publishing Co. Inc., River Edge, NJ, 1999. Translated from the Russian manuscript by N. Kruzhilin.

[15] Steele, J. M. *Stochastic calculus and financial applications*, vol. 45 of *Applications of Mathematics (New York).* Springer-Verlag, New York, 2001.

[16] Stroock, D. W., and Varadhan, S. R. S. *Multidimensional diffusion processes.* Classics in Mathematics. Springer-Verlag, Berlin, 2006. Reprint of the 1997 edition.

[17] Williams, D. *Probability with martingales.* Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.