

Note di probabilità: processi di Wiener ed applicazioni

PAOLO CARESSA

Roma, agosto 2008

Negli ultimi anni, professionalmente, ho avuto necessità di approfondire la teoria dei processi stocastici, in particolare dei processi di Lévy: fra le varie fonti che ho usato ci sono anche queste note, che ora trascrivo in formato elettronico: riproducono il mio lavoro per un mio seminario presso l'Università di Firenze, tenuto nel 1996 per sostenere l'esame finale di un corso di dottorato in teoria delle probabilità. Ho fatto loro premettere qualche breve cenno introduttivo sui fondamenti della teoria probabilistica e sulle variabili aleatorie: metto il tutto a disposizione con l'avvertenza che si tratta di materiale *as it is*. Si presuppone la conoscenza della teoria della misura di Lebesgue, quale per esempio è impartita nei corsi istituzionali di analisi superiore o di metodi matematici e, per la seconda parte, le nozioni di base sugli spazi di Hilbert: per queste e altre nozioni di base si potrà consultare il mio e-book [1].

Indice

1	Probabilità su insiemi	1
2	Probabilità sui boreliani	12
3	Variabili aleatorie	23
4	Indipendenza, attese condizionate e martingale	30
5	Funzioni caratteristiche e variabili gaussiane	37
6	Teorema di Kolmogorov e processi stocastici	45
7	Processo di Wiener	49
8	Misura di Wiener	59
9	Integrale di Wiener	65
10	Path Integrals	68
11	Formula di Trotter	72
12	Formula di Feynmann–Kac	75
13	La formula di Itô	77

1 Probabilità su insiemi

Il sesto problema della famosa lista di Hilbert¹ chiedeva di dare fondamento assiomatico alla fisica, intendendo con questo la meccanica e la teoria probabilistica (considerata come base della meccanica statistica): una risposta soddisfacente l'ha data Kolmogorov trent'anni dopo², sistemando definitivamente la teoria delle probabilità in seno alla teoria della misura di Lebesgue, maturata in quegli anni: cominciamo col rammentare qualche tratto di questa sistemazione assiomatica.

Definizione 1.1 Una σ -algebra di sottoinsiemi di un insieme Ω è una famiglia $\mathcal{A} \subset \mathcal{P}(\Omega)$ di sottoinsiemi di Ω tale che³

- (1) $\emptyset \in \mathcal{A}$
- (2) $A \in \mathcal{A} \implies \complement A = \Omega \setminus A \in \mathcal{A}$
- (3) $\{A_n\}_{n \in \mathbb{N}} \subset \mathcal{A} \implies \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$

Per esempio $\Omega \in \mathcal{A}$ (per (1) e (2)) e $A \cap B \in \mathcal{A}$ se $A, B \in \mathcal{A}$ (per (2) e (3)). In particolare, l'intersezione di σ -algebre di sottoinsiemi di uno stesso insieme è ancora una σ -algebra, il che vuol dire che, dato un sottoinsieme qualsiasi (o più in generale una qualsiasi famiglia di sottoinsiemi) di Ω ha senso parlare della σ -algebra da esso generata, che è l'intersezione delle σ -algebre che contengono il sottoinsieme dato.

Definizione 1.2 Se $\Omega \neq \emptyset$ e \mathcal{A} è una σ -algebra di sottoinsiemi di Ω , una funzione $P : \mathcal{A} \rightarrow [0, 1]$ tale che

- (1) $P(\emptyset) = 0$
- (2) $P(\Omega) = 1$
- (3) Se $\{A_n\}_{n \in \mathbb{N}} \subset \mathcal{A}$ è una successione di elementi a due a due disgiunti ($i \neq j \implies A_i \cap A_j = \emptyset$) allora

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n=0}^{\infty} P(A_n)$$

¹D. Hilbert, *Mathematische Probleme. Vortrag, gehalten auf dem internationalen Mathematiker Congress zu Paris 1900*, Gött. Nachr. 1900, 253-297, Vandenhoeck & Ruprecht, Göttingen.

²A.N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse Der Mathematik, Springer Verlag, 1933.

³Le notazioni insiemistiche meno comuni qui adottate sono le seguenti: \emptyset denota l'insieme vuoto, $\complement A = \Omega \setminus A$ il complemento di A in un insieme Ω fissato e $\mathcal{P}(\Omega)$ denota l'insieme delle parti di Ω , cioè la famiglia dei suoi sottoinsiemi.

si dice una probabilità su Ω (relativa alla σ -algebra \mathcal{A}). In questo caso, la tripla (Ω, \mathcal{A}, P) si dice uno spazio di probabilità e gli elementi di \mathcal{A} si dicono eventi.

Notiamo che $P(B \setminus A) = P(B) - P(A)$ se $A \subset B$. Una funzione $P : \mathcal{A} \rightarrow \mathbb{R}$ che soddisfi le (1) e (3) ma non necessariamente la (2) si dice *misura* su Ω relativa alla σ -algebra \mathcal{A} .

Esempio 1.3 Se Ω è uno spazio metrico separabile, allora la σ -algebra generata dalla topologia di Ω (cioè dalla famiglia dei suoi insiemi aperti) si dice σ -algebra dei boreliani di Ω , e si denota con $\mathfrak{B}(\Omega)$. Una misura su $\mathfrak{B}(\Omega)$ si dice misura boreliana.

Esempio 1.4 Se $\Omega = \{\omega_1, \dots, \omega_n\}$ e $\mathcal{A} = \mathcal{P}(\Omega)$ allora, se P è una qualsiasi probabilità su Ω , per ogni $A \subset \Omega$ si ha che

$$P(A) = P\left(\bigcup_{\omega \in A} \{\omega\}\right) = \sum_{\omega \in A} P(\{\omega\})$$

e quindi $P(A)$ è determinata dai valori $\{P(\omega_1), \dots, P(\omega_n)\}$, cioè da una funzione $p : \Omega \rightarrow [0, 1]$ tale che

$$p(\omega_1) + \dots + p(\omega_n) = 1$$

Per esempio $p(\omega_n) = 1/n$ soddisfa questa condizione.

La proprietà di σ -additività (3) non trova, a priori, nessuna giustificazione: va considerata una ipotesi verificata nelle applicazioni e che rende la teoria trattabile. Se la si sostituisce con l'ipotesi di finita additività

$$P\left(\bigcup_{n=0}^N A_n\right) = \sum_{n=0}^N P(A_n) \quad (\text{F})$$

si ottiene il concetto di *probabilità (o misura) finitamente additiva*: evidentemente, poggiando sull'induzione, la (F) equivale alla

$$A \cap B = \emptyset \implies P(A \cup B) = P(A) + P(B)$$

Teorema 1.5 Sia (Ω, \mathcal{A}, P) uno spazio di probabilità e sia P finitamente additiva: allora

$$(1) \quad A \subset B \implies P(A) \leq P(B) \quad (\text{monotonia})$$

(2) Se $\mathcal{B} \subset \mathcal{A}$ è tale che per ogni $B \in \mathcal{B}$ si abbia $P(B) > 0$ ed inoltre per ogni $B_1, B_2 \in \mathcal{B}$ si abbia che $B_1 \neq B_2 \implies B_1 \cap B_2 = \emptyset$ allora \mathcal{B} è al più una famiglia numerabile.

DIMOSTRAZIONE: La (1) segue facilmente dal fatto che $B = A \cup (B \setminus A)$ e quindi, per la (D), $P(B) = P(A) + P(B \setminus A) \geq P(A)$. Per avere la (2) notiamo che, ponendo

$$\mathcal{B}_n = \left\{ B \in \mathcal{B} \mid \frac{1}{n+1} < P(B) \leq \frac{1}{n} \right\}$$

abbiamo che

$$\mathcal{B} = \bigcup_{n=1}^{\infty} \mathcal{B}_n$$

(come segue dal fatto che P assume valori in $[0, 1]$.) Ora mostriamo che⁴

$$|\mathcal{B}_n| \leq n$$

Da questo seguirà ovviamente che \mathcal{B} è al più numerabile. In effetti, se per assurdo in \mathcal{B}_n ci fossero più di n elementi distinti, siano essi B_1, \dots, B_{n+1} , allora

$$P(B_1 \cup \dots \cup B_{n+1}) = P(B_1) + \dots + P(B_{n+1}) > \underbrace{\frac{1}{n+1} + \dots + \frac{1}{n+1}}_{n+1 \text{ volte}} = 1$$

il che è assurdo, dato che $B_1 \cup \dots \cup B_{n+1} \subset \Omega$ e quindi

$$1 = P(\Omega) \geq P(B_1 \cup \dots \cup B_{n+1})$$

QED

Esempio 1.6 Consideriamo $\Omega = \mathbb{N}$ e definiamo

$$P(\{n\}) = 0$$

(ed ovviamente $P(\mathbb{N}) = 1$.) Se consideriamo l'algebra di insiemi \mathcal{A} (che non è una σ -algebra) generata⁵ dai sottoinsiemi $\{n\}$ al variare di $n \in \mathbb{N}$ e dal singolo insieme \mathbb{N} , abbiamo che

$$\mathcal{A} = \{A \subset \Omega \mid |A| < \infty \text{ oppure } |\Omega \setminus A| < \infty\}$$

⁴Con $|X|$ o $\#X$ denotiamo la cardinalità di un insieme, cioè il numero dei suoi elementi.

⁵Un'algebra \mathcal{A} generata da una famiglia di sottoinsiemi è la più piccola famiglia contenente tali sottoinsiemi e tale che per ogni $A, B \in \mathcal{A}$ si abbia $A \cup B, A \cap B, \complement A \in \mathcal{A}$.

cioè che \mathcal{A} è l'algebra degli insiemi finiti e cofiniti in \mathbb{N} : possiamo estendere P ad una funzione $P : \mathcal{A} \rightarrow [0, 1]$ come

$$P(A) = \begin{cases} 0 & |A| < \infty \\ 1 & |\Omega - \setminus A| < \infty \end{cases}$$

ottenendo una probabilità finitamente additiva ma non σ -additiva su \mathcal{A} , dato che per esempio

$$0 = \sum_{n=0}^{\infty} P(\{n\}) \neq P\left(\bigcup_{n \in \mathbb{N}} \{n\}\right) = P(\Omega) = 1$$

Esempio 1.7 (DISTRIBUZIONE DI POISSON) Se $\Omega = \{\omega_n\}_{n \in \mathbb{N}}$ è numerabile e se definiamo

$$p(\{\omega_n\}) = p_n$$

dove la successione $\{p_n\}$ è tale che $p_n \geq 0$ e $\sum_{n=0}^{\infty} p_n = 1$, allora la funzione

$$P(A) = \sum_{\omega_n \in A} p_n$$

è una probabilità su $\mathcal{P}(\Omega)$. Per esempio, se $\Omega = \mathbb{N}$, una tale successione è data dalla

$$p_n = \frac{e^{-\lambda} \lambda^n}{n!}$$

(con $\lambda > 0$), che si dice distribuzione di Poisson di parametro λ .

Teorema 1.8 Se (Ω, \mathcal{A}, P) è uno spazio di probabilità finitamente additiva allora per ogni successione $\{A_n\}$ di elementi di \mathcal{A} a due a due disgiunti abbiamo che

$$\sum_{n=0}^{\infty} P(A_n) \leq P\left(\bigcup_{n \in \mathbb{N}} A_n\right)$$

DIMOSTRAZIONE: Infatti possiamo scrivere l'unione della famiglia $\{A_n\}$ come una unione disgiunta di due insiemi nel modo seguente:

$$\bigcup_{n \in \mathbb{N}} A_n = \left(\bigcup_{n=0}^N A_n\right) \cup \left(\bigcup_{n=N+1}^{\infty} A_n\right)$$

ed ottenere, per ogni $N \in \mathbb{N}$,

$$\sum_{n=0}^N P(A_n) = P\left(\bigcup_{n=0}^N A_n\right) \leq P\left(\bigcup_{n=0}^N A_n\right) + P\left(\bigcup_{n=N+1}^{\infty} A_n\right) = P\left(\bigcup_{n=0}^{\infty} A_n\right)$$

QED

Teorema 1.9 *Se (Ω, \mathcal{A}, P) è uno spazio di probabilità e $\{A_n\}_{n \in \mathbb{N}} \subset \mathcal{A}$, allora valgono le seguenti proprietà:*

(1) Disuguaglianza di Boole:

$$P\left(\bigcup_{n=0}^{\infty} A_n\right) \leq \sum_{n=0}^{\infty} P(A_n)$$

(2) Continuità inferiore: se $i < j \implies A_i \subset A_j$ allora

$$P\left(\bigcup_{n=0}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n)$$

(3) Continuità superiore: se $i > j \implies A_i \subset A_j$ allora

$$P\left(\bigcap_{n=0}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n)$$

DIMOSTRAZIONE: (1) Consideriamo gli insiemi

$$B_n = A_n \setminus \left(\bigcup_{i=0}^{n-1} A_i\right)$$

Evidentemente $B_n \subset A_n$ e gli insiemi $\{B_n\}$ sono a due a due disgiunti: allora dalla monotonia ($P(B_n) \leq P(A_n)$) segue che

$$P\left(\bigcup_{n=0}^{\infty} A_n\right) = P\left(\bigcup_{n=0}^{\infty} B_n\right) = \sum_{n=0}^{\infty} P(B_n) \leq \sum_{n=0}^{\infty} P(A_n)$$

(2) Poniamo $A_{-1} = \emptyset$, e notiamo che

$$\begin{aligned} \bigcup_{n=0}^{\infty} (A_n \setminus A_{n-1}) &= \bigcup_{n=0}^{\infty} (A_n \cap (\Omega \setminus A_{n-1})) = \left(\bigcup_{n=0}^{\infty} A_n\right) \cap \left(\bigcup_{n=0}^{\infty} (\Omega \setminus A_{n-1})\right) \\ &= \left(\bigcup_{n=0}^{\infty} A_n\right) \cap \left(\Omega \setminus \bigcup_{n=0}^{\infty} A_{n-1}\right) \\ &= \left(\bigcup_{n=0}^{\infty} A_n\right) \cap (\Omega \setminus \emptyset) = \bigcup_{n=0}^{\infty} A_n \end{aligned}$$

da cui, dato che $\bigcup_{n \in \mathbb{N}} A_n = \bigcup_{n \in \mathbb{N}} (A_n \setminus A_{n-1})$ e $\{A_n \setminus A_{n-1}\}$ sono a due a due disgiunti, segue

$$\begin{aligned} P\left(\bigcup_{n=0}^{\infty} A_n\right) &= P\left(\bigcup_{n=0}^{\infty} (A_n \setminus A_{n-1})\right) = \sum_{n=0}^{\infty} P(A_n \setminus A_{n-1}) \\ &= \sum_{n=0}^{\infty} (P(A_n) - P(A_{n-1})) = \lim_{N \rightarrow \infty} \sum_{n=0}^N (P(A_n) - P(A_{n-1})) \\ &= \lim_{N \rightarrow \infty} (P(A_N) - P(A_{-1})) = \lim_{n \rightarrow \infty} P(A_n) \end{aligned}$$

(3) Poniamo $B_n = A_0 \setminus A_n$, ottenendo una successione $B_0 \subset B_1 \subset B_2 \subset \dots$ che, per la (2), soddisfa alla

$$P\left(\bigcup_{n=0}^{\infty} B_n\right) = \lim_{n \rightarrow \infty} P(B_n)$$

Ma

$$\bigcup_{n=0}^{\infty} B_n = \bigcup_{n=0}^{\infty} (A_0 \setminus A_n) = \bigcup_{n=0}^{\infty} (A_0 \cap (\Omega \setminus A_n)) = A_0 \cap \bigcup_{n=0}^{\infty} (\Omega \setminus A_n)$$

pertanto

$$P(A_0) - P\left(\bigcap_{n=0}^{\infty} A_n\right) = P\left(\bigcup_{n=0}^{\infty} B_n\right) = \lim_{n \rightarrow \infty} P(B_n) = \lim_{n \rightarrow \infty} (P(A_0) - P(A_n))$$

e, dato che $P(A_0) \leq P(\Omega) = 1$, possiamo elidere $P(A_0)$ da ambo i membri (perché non è infinito!) ed avere finalmente

$$P\left(\bigcap_{n=0}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n)$$

QED

Combinando questi due teoremi abbiamo il

Corollario 1.10 *Una probabilità finitamente additiva che soddisfa la disuguaglianza di Boole è σ -additiva e viceversa.*

Data una successione di insiemi $\{A_n\}_{n \in \mathbb{N}}$, definiamo i suoi limiti inferiore e superiore come

$$\begin{aligned} \liminf_{n \rightarrow \infty} A_n &= \bigcup_{n=0}^{\infty} \bigcap_{k=n}^{\infty} A_k \\ \limsup_{n \rightarrow \infty} A_n &= \bigcap_{n=0}^{\infty} \bigcup_{k=n}^{\infty} A_k \end{aligned}$$

Quando questi limiti coincidono scriviamo semplicemente $\lim_{n \rightarrow \infty} A_n$ per denotarli entrambi.

Corollario 1.11 *Se $\{A_n\}$ è una famiglia di eventi e P è una probabilità allora*

$$\lim_{n \rightarrow \infty} A_n = A \implies \lim_{n \rightarrow \infty} P(A_n) = P(A)$$

DIMOSTRAZIONE: Infatti, per il teorema e le definizioni di limite appena date,

$$\begin{aligned} P(A) &= P\left(\liminf_{n \rightarrow \infty} A_n\right) = P\left(\bigcup_{n=0}^{\infty} \bigcap_{k=n}^{\infty} A_k\right) = \lim_{n \rightarrow \infty} P\left(\bigcap_{k=n}^{\infty} A_k\right) \\ &\leq \liminf_{n \rightarrow \infty} P(A_n) \leq \limsup_{n \rightarrow \infty} P(A_n) \leq \lim_{n \rightarrow \infty} P\left(\bigcup_{k=n}^{\infty} A_k\right) \\ &= P\left(\bigcap_{n=0}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) = P\left(\limsup_{n \rightarrow \infty} A_n\right) = P(A) \end{aligned}$$

quindi il limite numerico $\lim_{n \rightarrow \infty} P(A_n)$ esiste ed è uguale a $P(A)$.

QED

Teorema di Borel–Cantelli 1.12 *Se $\{A_n\}$ è una successione di eventi in uno spazio di probabilità (Ω, \mathcal{A}, P) allora*

$$\sum_{n=1}^{\infty} P(A_n) < \infty \implies P\left(\limsup_{n \rightarrow \infty} A_n\right) = 0$$

DIMOSTRAZIONE: Ricordiamo che $\limsup_n A_n = \bigcap_n \bigcup_{k \geq n} A_k$. Pertanto

$$P\left(\limsup_{n \rightarrow \infty} A_n\right) = \lim_{n \rightarrow \infty} P\left(\bigcup_{k=n}^{\infty} A_k\right) \leq \lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} P(A_k) = 0$$

se la serie numerica $\sum_n P(A_n)$ è convergente.

QED

Teorema 1.13 *Le seguenti condizioni sono equivalenti su una funzione $P : \mathcal{A} \rightarrow [0, 1]$:*

(1) P è una probabilità σ -additiva, cioè se $i \neq j \implies A_i \cap A_j = \emptyset$ allora

$$\sum_{n=0}^{\infty} P(A_n) = P\left(\bigcup_{n \in \mathbb{N}} A_n\right)$$

(2) P è una probabilità finitamente additiva e se $\{A_n\}$ è una successione di elementi di \mathcal{A} tale che $i < j \implies A_i \subset A_j$ allora

$$\lim_{n \rightarrow \infty} P(A_n) = P\left(\bigcup_{n \in \mathbb{N}} A_n\right)$$

(3) P è una probabilità finitamente additiva e se $\{A_n\}$ è una successione di elementi di \mathcal{A} tale che $i > j \implies A_i \subset A_j$ allora

$$\bigcap_{n \in \mathbb{N}} A_n = \emptyset \implies \lim_{n \rightarrow \infty} P(A_n) = 0$$

DIMOSTRAZIONE: Le uniche implicazioni non banali da dimostrare sono (2) \implies (1) e (3) \implies (1). Mostriamo la (1) \implies (2), cioè che se $\{A_n\}$ sono a due a due disgiunti allora

$$\sum_{n=0}^{\infty} P(A_n) = P\left(\bigcup_{n \in \mathbb{N}} A_n\right)$$

Consideriamo la successione di insiemi

$$B_n = \bigcup_{i=0}^n A_i$$

tale che $i < j \implies B_i \subset B_j$, e che l'unione delle famiglie $\{A_n\}$ e $\{B_n\}$ sia la stessa: allora, per la (2) del teorema precedente,

$$\begin{aligned} P\left(\bigcup_{n \in \mathbb{N}} A_n\right) &= P\left(\bigcup_{n \in \mathbb{N}} B_n\right) = \lim_{n \rightarrow \infty} P(B_n) = \lim_{n \rightarrow \infty} P\left(\bigcup_{i=0}^n A_i\right) \\ &= \lim_{n \rightarrow \infty} \sum_{i=0}^n P(A_i) = \sum_{n=0}^{\infty} P(A_n) \end{aligned}$$

(3) \implies (1) si dimostra in modo analogo, notando che in questo caso l'unione della famiglia $\{B_n\}$ è A_0 e che

$$\lim_{n \rightarrow \infty} B_n = \emptyset$$

(come segue dal fatto che gli $\{A_k\}$ sono a due a due disgiunti.) Dunque, per la (3) del teorema precedente,

$$\lim_{n \rightarrow \infty} P(B_n) = 0$$

pertanto

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = P(A_0) = P\left(\bigcup_{i=0}^{n-1} A_i\right) + P(B_n) = \sum_{i=0}^{n-1} P(A_i) + P(B_n)$$

Ma, per $n \rightarrow \infty$ si ha $P(B_n) \rightarrow 0$, dunque

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n=0}^{\infty} P(A_n)$$

QED

D'ora in avanti ci limiteremo a considerare probabilità σ -additive, se non altrimenti specificato⁶.

Spesso per costruire una probabilità su una σ -algebra, si parte da una probabilità definita su una sottoalgebra che generi la σ -algebra cui siamo interessati: nel caso di probabilità finitamente additive c'è il seguente teorema del tutto soddisfacente.

Teorema 1.14 *Se P è una probabilità finitamente additiva su un'algebra $\mathcal{A} \subset \mathcal{P}(\Omega)$ allora si può estendere ad una probabilità finitamente additiva su $\mathcal{P}(\Omega)$.*

DIMOSTRAZIONE: Se f è un funzionale lineare positivo⁷ sullo spazio vettoriale

$$W = \{X : \Omega \rightarrow \mathbb{R} \mid \exists \alpha \in \mathbb{R} \forall \omega \in \Omega \mid X(\omega) < \alpha\}$$

delle funzioni limitate su Ω , tale che $f(1) = 1$ (dove a sinistra 1 denota la funzione costante 1), allora possiamo associare a f una probabilità P_f come

$$P_f(A) = f(\chi_A)$$

dove χ_A è la *funzione indicatore*⁸ di A (altrimenti detta funzione caratteristica, locuzione che in teoria delle probabilità ha un altro significato). Evidentemente $0 \leq P_f(\omega) \leq 1$: infatti, dato che f è positivo e $\chi_A \geq 0$ si ha $0 \leq P_f$, mentre dato $f(1) = 1$ e che $A \subset B \implies f(\chi_A) \leq f(\chi_B)$ abbiamo che $P_f \leq 1$.

⁶Notiamo che nella definizione originale di Kolmogorov la (3) del teorema precedente è assunta assiomaticamente al posto della σ -additività.

⁷Cioè tale che $f(X) \geq 0$.

⁸Cioè $\chi_A(\omega) = 1$ se $\omega \in A$ e $\chi_A(\omega) = 0$ se $\omega \notin A$.

Inoltre, dato che f è lineare e che, se $A \cap B = \emptyset$ allora $\chi_{A \cup B} = \chi_A + \chi_B$, troviamo che

$$A \cap B = \emptyset \implies P_f(A \cup B) = P_f(A) + P_f(B)$$

Ora consideriamo una probabilità finitamente additiva su un'algebra $\mathcal{A} \subset \mathcal{P}(\Omega)$ e poniamo, per $A \in \mathcal{A}$,

$$f(\chi_A) = P(A)$$

definendo così un funzionale lineare positivo sullo spazio vettoriale V generato⁹ dalle funzioni $\{\chi_A\}_{A \in \mathcal{A}}$, cioè

$$f\left(\sum_{i=1}^n a_i \chi_{A_i}\right) = \sum_{i=1}^n a_i P(A_i)$$

Verifichiamo che questa definizione è ben posta: dato che una funzione $\sum_{i=1}^n a_i \chi_{A_i}$ in V assume solo un numero finito di valori $\{x_1, \dots, x_m\}$ non nulli, allora

$$\sum_{i=1}^n a_i \chi_{A_i} = \sum_{i=1}^m x_i \chi_{\Omega_i}$$

dove

$$\Omega_i = \left\{ x \in \mathbb{R} \mid \sum_{i=1}^n a_i \chi_{A_i}(x) = \omega_i \right\}$$

In questa rappresentazione della nostra funzione di V , gli insiemi Ω_i sono disgiunti e quindi la definizione

$$f\left(\sum_{i=1}^m x_i \chi_{\Omega_i}\right) = \sum_{i=1}^m x_i P(\Omega_i)$$

è ben posta. Ora possiamo definire un funzionale T su W come

$$T(X) = \sup_{X \in W} X$$

Questo funzionale è "sublineare" ($T(X + Y) \leq T(X) + T(Y)$) ed inoltre

$$f(X) \leq T(X)$$

⁹Cioè il più piccolo sottospazio dello spazio vettoriale delle funzioni $f : \Omega \rightarrow \mathbb{R}$ che contenga le χ_A : si tratta dello spazio delle funzioni che sono combinazioni lineari di funzioni del tipo χ_A .

Infatti, se $X = \sum_i a_i \chi_{A_i}$, definiamo i *costituenti* degli $\{A_i\}$ come

$$C_k = A_1^* \cap \cdots \cap A_n^*$$

dove A_i^* può essere sia A_i che $\complement A_i$ (quindi ci sono 2^n costituenti possibili): ovviamente $C_i \cap C_j = \emptyset$ se $i \neq j$, quindi

$$f(X) = f\left(\sum_{i=1}^n a_i \chi_{A_i}\right) = f\left(\sum_{k=1}^{2^n} c_k \chi_{C_k}\right) = \sum_{k=1}^{2^n} c_k P(C_k)$$

Ma $\sum_k c_k P(C_k)$ è combinazione convessa degli c_k e quindi sta fra il loro estremo inferiore e il loro estremo superiore, col che $f(X) \leq T(X)$.

Possiamo allora applicare il teorema di Hahn–Banach (cfr. [8, §III.2], [1, §6.5]) per estendere f un funzionale su tutto W e quindi P ad una misura su tutto $\mathcal{P}(\Omega)$.

QED

Nel caso di probabilità qualsiasi non abbiamo un risultato così completo: tuttavia esiste il seguente fondamentale risultato di estensione, per la cui dimostrazione si rimanda a [1] o [10, §12.2].

Teorema di Carathéodory 1.15 *Se m è una misura su un'algebra \mathcal{A} di sottoinsiemi di Ω allora esiste una misura \tilde{m} sulla σ -algebra $\tilde{\mathcal{A}}$ generata da \mathcal{A} che estenda m (cioè tale che per ogni $A \in \mathcal{A}$ si abbia $\tilde{m}(A) = m(A)$). Se m è σ -finita allora \tilde{m} è unica.*

L'idea è che la misura \tilde{m} si ottiene restringendo a $\tilde{\mathcal{A}}$ la *misura esterna*

$$m^* A = \inf \sum_{n=0}^{\infty} m(A_n)$$

al variare di $\{A_n\}$ in tutte le successioni di \mathcal{A} tali che $A \subset \bigcup_n A_n$.

Per esempio prendendo gli intervalli aperti della retta reale \mathbb{R} abbiamo un'algebra sulla quale esiste una misura naturale, la lunghezza dell'intervallo, dunque, per il teorema di Carathéodory, possiamo dedurre l'esistenza della misura di Lebesgue sui boreliani della retta reale. Il completamento di questa misura fornisce la classica misura di Lebesgue sulla retta, nel senso della

Definizione 1.16 *Uno spazio di misura (Ω, \mathcal{A}, m) è completo se \mathcal{A} contiene tutti gli insiemi che sono sottoinsiemi di insiemi di misura nulla, cioè se $A \in \mathcal{A}$, $m(A) = 0$ e $B \subset A$ allora $B \in \mathcal{A}$.*

Come è noto, la misura di Lebesgue è completa, mentre la sua restrizione alla σ -algebra dei boreliani non lo è, dato che esiste un sottoinsieme dell'insieme di Cantor (che è boreliano in quanto chiuso, e ha misura nulla) che non è boreliano. In generale si può aggirare il problema dell'incompletezza di una misura per mezzo del seguente teorema, per una cui dimostrazione rinviamo pure a [1] o [10].

Teorema 1.17 *Se (Ω, \mathcal{A}, m) è uno spazio di misura allora esiste uno spazio di misura completo $(\Omega, \mathcal{A}', m')$ tale che $\mathcal{A} \subset \mathcal{A}'$ e m' estende m , cioè $A \in \mathcal{A} \implies m'(A) = m(A)$.*

Questo conclude i nostri richiami di teoria della misura: nel paragrafo 3 daremo degli ulteriori richiami sul concetto di funzione misurabile e integrabile in uno spazio di misura, e sul concetto di integrale di una funzione misurabile esteso ad un insieme misurabile di uno spazio di misura (per le dimostrazioni dei risultati qui riportati senza dimostrazione e per maggiori dettagli cfr. [8], [10], [1]).

2 Probabilità sui boreliani

Definizione 2.1 *Se $\Omega \neq \emptyset$, una semialgebra di sottoinsiemi di Ω è una famiglia $\mathcal{S} \subset \mathcal{P}(\Omega)$ tale che*

- (1) $\emptyset, \Omega \in \mathcal{S}$.
- (2) $A, B \in \mathcal{S} \implies A \cap B \in \mathcal{S}$.
- (3) $\forall A \in \mathcal{S} \exists C_1, \dots, C_n \in \mathcal{S}$ disgiunti e tali che $A = C_1 \cup \dots \cup C_n$.

Per esempio gli intervalli (anche non limitati) della retta reale formano una semialgebra di insiemi.

Definizione 2.2 *Se \mathcal{S} è una semialgebra di insiemi di Ω , allora l'algebra generata da \mathcal{S} è la famiglia $\mathcal{A}(\mathcal{S})$ i cui elementi sono tutte le possibili unioni finite e disgiunte di elementi di \mathcal{S} .*

Se \mathcal{S} è una semialgebra di sottoinsiemi di Ω e $P : \mathcal{S} \rightarrow [0, 1]$ è una funzione additiva tale che $P(\Omega) = 1$, possiamo estendere P all'algebra $\mathcal{A}(\mathcal{S})$ generata da \mathcal{S} come

$$P \left(\bigcup_{i=1}^n S_i \right) = \sum_{i=1}^n P(S_i)$$

(dove $S_1 \cap \dots \cap S_n = \emptyset$): si tratta ovviamente di una estensione ben definita, ed ancora additiva. In realtà vale il seguente risultato (per la dimostrazione si veda [8, §5.2, teorema 3]).

Teorema 2.3 *Se m è una misura σ -additiva su una semialgebra \mathcal{S} di sottinsiemi di Ω , allora l'estensione di m all'algebra $\mathcal{A}(\mathcal{S})$ è ancora σ -additiva.*

Esempio 2.4 *Se $\Omega = [a, b]$ allora la famiglia dei suoi intervalli limitati e connessi è una semialgebra: se definiamo*

$$P(A) = \frac{l(A)}{b-a}$$

dove $l(A)$ è la lunghezza dell'intervallo $A \in \mathcal{S}$, allora il teorema precedente ci fornisce la misura di Lebesgue sui boreliani di $[a, b]$. Se $\Omega = \mathbb{R}$ allora la famiglia

$$\mathcal{S} = \{(x, y] \mid -\infty \leq x \leq y < \infty\} \cup \{(x, \infty) \mid -\infty \leq x\}$$

è una semialgebra. In entrambi i casi la σ -algebra da essi generata è quella dei boreliani.

Più in generale, se Ω è un boreliano di \mathbb{R}^n di misura di Lebesgue $m(\Omega) < \infty$, e se

$$f(x) = \frac{1}{m(\Omega)} \chi_{\Omega}(x)$$

per ogni boreliano $A \subset \Omega$ possiamo definire una probabilità

$$P(A) = \int_A f(x) dx = \frac{1}{m(\Omega)} \int_A dx = \frac{m(A)}{m(\Omega)}$$

sui boreliani di Ω . La funzione f si dice *densità di probabilità* associata alla probabilità P .

Definizione 2.5 *Se $\Omega \neq \emptyset$, una famiglia $\mathcal{K} \subset \mathcal{P}(\Omega)$ si dice classe compatta se per ogni successione $\{A_n\} \subset \mathcal{K}$ che ha la proprietà dell'intersezione finita¹⁰ si ha $\bigcap_{n \in \mathbb{N}} A_n \neq \emptyset$.*

Questa definizione risulta motivata dal seguente lemma (che non dimostreremo (cfr. [10, §12.3])

Lemma 2.6 *Se \mathcal{S} è una semialgebra e $P : \mathcal{S} \rightarrow [0, 1]$ una probabilità finitamente additiva tale che, per ogni $A \in \mathcal{S}$:*

$$P(A) = \sup\{P(K) \mid K \in \mathcal{K} \text{ classe compatta, } K \subset A\}$$

allora P è σ -additiva su \mathcal{S} .

¹⁰Cioè tale che ogni sua sottofamiglia finita abbia intersezione non vuota.

utilizzando il quale possiamo caratterizzare le probabilità sui boreliani:

Definizione 2.7 Una funzione di ripartizione è una funzione $F : \mathbb{R} \rightarrow [0, 1]$ tale che

- (1) $\lim_{x \rightarrow -\infty} F(x) = 0$.
- (2) $\lim_{x \rightarrow +\infty} F(x) = 1$.
- (3) F è non decrescente e continua superiormente.

Esempio 2.8

- (1) L'esempio più elementare è la funzione di Heaviside (normalizzata)

$$H(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1 & \text{se } x \geq 0 \end{cases}$$

- (2) Un altro esempio è la funzione

$$F(x) = \frac{\pi + 2 \arctan(x)}{2\pi}$$

che è una funzione di ripartizione continua.

Teorema 2.9 Esiste una corrispondenza biunivoca fra probabilità sui boreliani della retta reale e funzioni di ripartizione, espressa dalla relazione

$$F(x) = P((-\infty, x])$$

DIMOSTRAZIONE: Supponiamo che P sia una probabilità sui boreliani di \mathbb{R} , e definiamo F come nell'enunciato: dobbiamo dimostrare che soddisfa le (1)–(3) della definizione precedente. Intanto è ovviamente una funzione monotona non decrescente (dato che $x < y \implies (-\infty, x] \subset (-\infty, y]$), inoltre

$$\lim_{x \rightarrow -\infty} F(x) = \lim_{x \rightarrow -\infty} P((-\infty, x]) = P\left(\bigcap_{n \in \mathbb{Z}} (-\infty, n]\right) = P(\emptyset) = 0$$

$$\lim_{x \rightarrow +\infty} F(x) = \lim_{x \rightarrow +\infty} P((-\infty, x]) = P\left(\bigcup_{n \in \mathbb{Z}} (-\infty, n]\right) = P(\mathbb{R}) = 1$$

Ora notiamo che

$$\begin{aligned} P((a, b]) &= P((-\infty, b] \setminus (-\infty, a]) \\ &= P((-\infty, b]) - P((-\infty, a]) = F(b) - F(a) \end{aligned}$$

pertanto, avendosi $(a, b] = \bigcap_{n \in \mathbb{N}} (a, b + 1/(1+n)]$, troviamo che

$$\begin{aligned} F(b) - F(a) &= P((a, b]) = P\left(\bigcap_{n \in \mathbb{N}} (a, b + \frac{1}{1+n}]\right) \\ &= \lim_{n \rightarrow \infty} P\left((a, b + \frac{1}{1+n}]\right) = \lim_{n \rightarrow \infty} F\left(b + \frac{1}{1+n}\right) - F(a) \end{aligned}$$

da cui

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

cioè la continuità superiore della F .

Mostriamo ora come data una funzione di ripartizione F sia possibile definire una misura che verifichi l'equazione dell'enunciato del teorema. Consideriamo la semialgebra \mathcal{S} degli intervalli (qualsiasi, inclusi anche \emptyset e \mathbb{R}) della retta reale e definiamo su di essa la seguente probabilità

$$\begin{cases} P((-\infty, b]) = F(b) \\ P((a, b]) = F(b) - F(a) \\ P([a, b)) = F(b) - F(a^-) \\ P([a, b]) = F(b) - F(a^-) \\ P(\{x\}) = F(x) - F(x^-) \end{cases}$$

Si tratta ovviamente di una probabilità finitamente additiva su \mathcal{S} : vogliamo mostrare che soddisfa alle condizioni del lemma, in modo da poter dire che si estende ad una probabilità sulla σ -algebra generata da \mathcal{S} (che poi è esattamente la σ -algebra dei boreliani della retta reale).

Si consideri la famiglia

$$\mathcal{K} = \{[a, b] \mid a \leq b\}$$

(formata cioè dagli intervalli chiusi e limitati e dai singoli punti $[a, a] = \{a\}$.) Si tratta di una classe compatta: se $\{A_n = [a_n, b_n]\} \subset \mathcal{K}$ gode della proprietà dell'intersezione finita, supponiamo per assurdo che $\bigcap_n A_n = \emptyset$; allora, prendendo il complemento di questa equazione ed usando le leggi di de Morgan,

$$\begin{aligned} \mathbb{R} &= \bigcup_{n \in \mathbb{N}} \mathcal{C}[a_n, b_n] = \bigcup_{n \in \mathbb{N}} ((-\infty, a_n) \cup (b_n, \infty)) \\ &= (-\infty, \sup_{n \in \mathbb{N}} a_n) \cup (\inf_{n \in \mathbb{N}} b_n, \infty) \end{aligned}$$

da cui deduciamo che $\inf_{n \in \mathbb{N}} b_n < \sup_{n \in \mathbb{N}} a_n$ (perché \mathbb{R} è connesso), e quindi che esistono $h, k \in \mathbb{N}$ tali che $b_h < a_k$, da cui

$$A_h \cap A_k = [a_h, b_h] \cap [a_k, b_k] = \emptyset$$

il che contraddice la proprietà dell'intersezione finita per gli $\{A_n\}$.

Ora che sappiamo che \mathcal{K} è una classe compatta, non resta che verificare che, per ogni $A \in \mathcal{S}$,

$$P(A) = \sup\{P(K) \mid K \subset A, K \in \mathcal{K}\}$$

per applicare il lemma e dedurre la σ -additività dell'estensione di P ai boreliani. Questo può farsi considerando ciascun tipo di insieme in \mathcal{S} : per esempio se $A = (a, b)$ allora abbiamo

$$P((a, b]) = F(b) - F(a) = \sup_{n \in \mathbb{N}} (F(b_n) - F(a_n)) = \sup_{n \in \mathbb{N}} P([a_n, b_n])$$

dove $\{a_n\}$ è una successione decrescente e convergente ad a e $\{b_n\}$ è una successione crescente e convergente a b tali che $a < a_n < b_n < b$.

In modo analogo si verificano gli altri casi.

QED

Dato che $P(x) = F(x) - F(x^-)$ è chiaro che i punti di continuità avranno misura nulla, quindi

Corollario 2.10 *F è continua in x se e solo se $P(\{x\}) = 0$.*

In particolare, dato che una funzione monotona possiede al più una quantità numerabile di punti di discontinuità,

Corollario 2.11 *Una probabilità boreliana P su \mathbb{R} è tale che $P(\{x\}) > 0$ al più per una quantità numerabile di $x \in \mathbb{R}$.*

Questi risultati ci spingono a dare la seguente generalizzazione di questi concetti:

Definizione 2.12 *Una misura di Lebesgue–Stieltjes è una qualsiasi misura sui boreliani della retta reale, e una funzione di distribuzione è una funzione non decrescente e continua superiormente sulla retta reale.*

Ovviamente sussisterà una generalizzazione del teorema precedente al caso di questa nuova definizione, ma con una precisazione che ora andiamo a svolgere: in effetti, se F è una funzione di distribuzione, sarà naturale definire la misura di Lebesgue–Stieltjes corrispondente come

$$m((a, b)) = F(b) - F(a^-)$$

Notiamo tuttavia che $G(x) = F(x) + c$ (dove c è una costante) induce la medesima misura: questo non è un problema, dato che la relazione $G(x) = F(x) + c$ fra F e G è di equivalenza sull'insieme delle funzioni di distribuzione, col che, sostanzialmente ripetendo la dimostrazione del teorema precedente, abbiamo il

Teorema 2.13 *Esiste una corrispondenza biunivoca fra le classi di equivalenza di funzioni di distribuzione e le misure di Lebesgue–Stieltjes sulla retta reale.*

Si noti che nel caso di funzioni di ripartizione, la relazione di equivalenza si riduce all'uguaglianza.

Esempio 2.14 *La funzione di distribuzione $F(x) = x$ induce la misura di Lebesgue in \mathbb{R} sui boreliani.*

Esempio 2.15 *La funzione di Heaviside induce la misura di Dirac δ_0 concentrata nell'origine: infatti*

$$\delta_0((-\infty, x]) = \begin{cases} 0 & \text{se } x \leq 0 \\ 1 & \text{se } x > 0 \end{cases} = \begin{cases} 0 & \text{se } 0 \notin (-\infty, 0) \\ 1 & \text{se } 0 \in (-\infty, 0) \end{cases}$$

pertanto, dato un qualsiasi aperto A

$$\delta_0(A) = \begin{cases} 0 & \text{se } 0 \notin A \\ 1 & \text{se } 0 \in A \end{cases}$$

Evidentemente, se f è una funzione boreliana non negativa e A è un boreliano allora

$$\int_A f \delta_0 = \begin{cases} 0 & \text{se } 0 \notin A \\ f(0) & \text{se } 0 \in A \end{cases}$$

La misura di Dirac concentrata nell'origine è mutuamente singolare con la misura di Lebesgue nel senso della seguente

Definizione 2.16 *Se m e m' sono misure su una σ -algebra \mathcal{A} di sottoinsiemi di un insieme non vuoto Ω , si dice che m e m' sono mutuamente singolari (e si scrive $m \perp m'$) se esistono due insiemi disgiunti $A, B \in \mathcal{A}$ tali che $\Omega = A \cup B$ ma $m(A) = 0$ e $m'(B) = 0$.*

All'opposto abbiamo la nozione di assoluta continuità:

Definizione 2.17 *Se m e m' sono misure su una σ -algebra \mathcal{A} di sottoinsiemi di un insieme non vuoto Ω , si dice che m' è assolutamente continua rispetto a m (e si scrive $m' \ll m$) se per ogni $A \in \mathcal{A}$ se $m(A) = 0$ allora anche $m'(A) = 0$.*

Esempio 2.18 *Se m è una misura su $\mathcal{A} \subset \mathcal{P}(\Omega)$ e $f : \Omega \rightarrow \mathbb{R}$ è una funzione non negativa e misurabile, la misura definita come*

$$m'(A) = \int_A f \, dm$$

è assolutamente continua rispetto a m , come segue ovviamente dal fatto che l'integrale esteso ad un insieme di misura nulla è nullo.

Se (Ω, \mathcal{A}, m) è uno spazio di misura σ -finito (cioè è unione al più numerabile di insiemi di misura finita: per esempio uno spazio di probabilità) allora l'esempio precedente esaurisce la classe delle misure assolutamente continue: vale infatti il seguente e fondamentale

Teorema di Radon–Nikodym. *Se (Ω, \mathcal{A}, m) è uno spazio di misura σ -finito e se m' è una misura assolutamente continua rispetto a m allora esiste una funzione misurabile non negativa $f : \Omega \rightarrow \mathbb{R}$ tale che, per ogni $A \in \mathcal{A}$:*

$$m'(A) = \int_A f \, dm$$

unica a meno di equivalenza quasi ovunque¹¹ (cioè se esiste una g che soddisfa la stessa equazione allora $f = g$ q.o.).

Per una dimostrazione rimandiamo a [8], [10] e, per la classica dimostrazione di von Neumann che usa gli spazi di Hilbert, a [1].

La funzione f del teorema di Radon–Nikodym si dice *derivata di Radon–Nikodym* di m' rispetto a m e si denota di solito

$$f = \frac{dm'}{dm}$$

Il seguente teorema è un altro pilastro della teoria della misura, per il quale si rimanda sempre alle fonti già citate:

¹¹Rammentiamo che in teoria della misura la locuzione “*quasi ovunque*” applicata a una proprietà significa che quella proprietà può non essere verificata al più su un insieme di misura nulla.

Teorema di Lebesgue. *Se (Ω, \mathcal{A}, m) è uno spazio di misura σ -finito e se m' è una misura σ -finita su \mathcal{A} allora esistono due uniche misure m_0 e m_1 tali che $m_0 \perp m$, $m_1 \ll m$ e*

$$m' = m_0 + m_1$$

Applichiamo questo teorema alla situazione seguente: P sia una probabilità sui boreliani, e definiamo la sua *parte discreta* come

$$P_d(A) = \sum_{x \in A} P(\{x\})$$

(questa definizione ha senso perché c'è al più un insieme numerabile di punti che abbiamo misura positiva rispetto a P .)

Ora, per il teorema di Lebesgue, possiamo decomporre $P - P_d$ come

$$P - P_d = P_0 + P_1$$

con $P_0 \perp P$ e $P_1 \ll P$, da cui abbiamo una decomposizione

$$P = P_0 + P_1 + P_d$$

cui, a livello di funzioni di ripartizione, corrisponde la

$$F = \alpha_0 F_0 + \alpha_1 F_1 + \alpha_d F_d$$

(le costanti $\alpha_0, \alpha_1, \alpha_d$ servono a normalizzare la F in modo che $F < 1$) dove:

- F_0 è continua ed ha derivata nulla q.o.
- $F_1(x) = \int_{-\infty}^x f'(t) dt$ dove f è la funzione di ripartizione associata a P (teorema di Radon–Nikodym).
- $\alpha_0 F_0 + \alpha_1 F_1$ è continua (dato che $(P_0 + P_1)(\{x\}) = 0$).
- F_d è una funzione che possiede al più un'infinità numerabile di discontinuità (salti) la somma delle cui ampiezze sia 1.

Teorema 2.19 *Se P è una probabilità boreliana sui reali e f è la sua funzione di ripartizione allora, se λ denota la misura di Lebesgue,*

$$P \ll \lambda \iff f \in AC(\mathbb{R})$$

(dove $AC(\mathbb{R})$ è lo spazio delle funzioni reali assolutamente continue) e

$$\frac{dP}{d\lambda} = f'$$

q.o. nel senso della misura P .

Passiamo ora al caso di probabilità boreliane su \mathbb{R}^n : l'unico concetto che non si generalizza banalmente, nel considerare le funzioni di ripartizione, è la monotonia. Consideriamo l'intervallo

$$I = (a_1, b_1] \times \cdots \times (a_n, b_n]$$

in \mathbb{R}^n e l'insieme dei suoi vertici

$$V(I) = \{x \in \mathbb{R}^n \mid \forall i \ x_i = a_i \text{ oppure } x_i = b_i\}$$

(dove $x = (x_1, \dots, x_n)$): se $x \in V(I)$ poniamo

$$S(x) = \begin{cases} 1 & \text{se } x_i = a_i \text{ per un numero pari di indici} \\ -1 & \text{se } x_i = a_i \text{ per un numero dispari di indici} \end{cases}$$

Possiamo allora associare, ad ogni funzione $F : \mathbb{R}^n \rightarrow \mathbb{R}$ e ad ogni intervallo I della forma precedente, l'*incremento*

$$\Delta_I F = \sum_{x \in V(I)} F(x) S(x)$$

Evidentemente, se $n = 1$, abbiamo $V(I) = \{a_1, b_1\}$ e quindi l'incremento $\Delta_I F$ è esattamente $F(b_1) - F(a_1)$.

Definizione 2.20 Una funzione di ripartizione $F : \mathbb{R}^n \rightarrow [0, 1]$ è una *funzione tale che*

- (1) $\lim_{x_i \rightarrow \infty} F(x) = 0$ per qualche indice i ;
- (2) $\lim_{x_i \rightarrow \infty} F(x) = 1$ per ogni indice i ;
- (3) F è *semicontinua superiormente* in ciascun punto x : per ogni $\varepsilon > 0$ esiste $\delta_\varepsilon > 0$ in modo che per ogni $y \in \mathbb{R}^n$ tale che $x_i \leq y_i < x_i + \delta_\varepsilon$ (per ogni indice i) si abbia

$$|F(x) - F(y)| < \varepsilon$$

- (4) Per ogni intervallo I si abbia che $\Delta_I F \geq 0$.

Si noti che l'ultima condizione è l'analogo della monotonia nel caso $n = 1$: in effetti implica (ma non è equivalente) alla monotonia in ogni singola variabile. Sempre in modo analogo al caso $n = 1$ si dimostra allora il seguente

Teorema 2.21 *Esiste una corrispondenza biunivoca fra le probabilità sui boreliani di \mathbb{R}^n e le funzioni di ripartizione, data dalla*

$$F(x_1, \dots, x_n) = P((-\infty, x_1] \times \cdots \times (-\infty, x_n])$$

L'unico passo della dimostrazione che si discosta dal caso $n = 1$ è quello relativo all'ultima proprietà di F , cioè la non negatività del suo incremento rispetto ad un qualsiasi intervallo: limitiamoci al caso $n = 2$. Necessitiamo del seguente lemma

Principio di inclusione–esclusione 2.22 *Se P è una probabilità e A_1, \dots, A_n sono nella σ -algebra dove P è definita, allora*

$$\begin{aligned} P(A_1 \cup \cdots \cup A_n) &= \sum_i P(A_i) - \sum_{i < j} P(A_i \cap A_j) + \\ &+ \sum_{i < j < k} P(A_i \cap A_j \cap A_k) + \cdots + (-1)^{n+1} P(A_1 \cap \cdots \cap A_n) \end{aligned}$$

(dove gli indici variano tutti fra 1 e n .)

DIMOSTRAZIONE: Per induzione su n : se $n = 1$ l'enunciato è banale ($P(A_1) = P(A_1)$). Il caso $n = 2$ è immediato, dato che

$$\begin{aligned} P(A_1 \cup A_2) &= P(A_1 \cup (A_2 \setminus A_1)) = P(A_1) + P(A_2 \setminus A_1) \\ &= P(A_1) + P(A_2) - P(A_1 \cap A_2) \end{aligned}$$

Supponiamo ora l'enunciato vero nel caso n e dimostriamolo nel caso $n + 1$. Abbiamo cioè $n + 1$ insiemi A_1, \dots, A_{n+1} : poniamo $A = A_1 \cup \cdots \cup A_n$; allora, per il caso $n = 2$ abbiamo che

$$P(A_1 \cup \cdots \cup A_{n+1}) = P(A \cup A_{n+1}) = P(A) + P(A_{n+1}) - P(A \cap A_{n+1})$$

Ma, per ipotesi induttiva, (e dato che $(A_i \cup A_{n+1}) \cap (A_j \cup A_{n+1}) = A_i \cap A_j \cup A_{n+1}$)

$$\begin{aligned} P(A \cap A_{n+1}) &= P((A_1 \cap A_{n+1}) \cup \cdots \cup (A_n \cap A_{n+1})) \\ &= \sum_i P(A_i \cap A_{n+1}) - \sum_{i < j} P(A_i \cap A_j \cap A_{n+1}) + \\ &+ \sum_{i < j < k} P(A_i \cap A_j \cap A_k \cap A_{n+1}) \\ &+ \cdots + (-1)^{n+1} P(A_1 \cap \cdots \cap A_n \cap A_{n+1}) \end{aligned}$$

(gli indici variano da 1 a n). Usando questa equazione e di nuovo l'ipotesi induttiva troviamo infine che

$$\begin{aligned}
P(A_1 \cup \dots \cup A_{n+1}) &= P(A \cup A_{n+1}) = P(A) + P(A_{n+1}) - P(A \cap A_{n+1}) \\
&= \sum_i^{1\dots n} P(A_i) - \sum_{i<j}^{1\dots n} P(A_i \cap A_j) + \sum_{i<j<k}^{1\dots n} P(A_i \cap A_j \cap A_k) + \\
&\quad + \dots + (-1)^{n+1} P(A_1 \cap \dots \cap A_n) + P(A_{n+1}) - P(A \cap A_{n+1}) \\
&= \sum_i^{1\dots n+1} P(A_i) - \sum_{i<j}^{1\dots n+1} P(A_i \cap A_j) + \sum_{i<j<k}^{1\dots n+1} P(A_i \cap A_j \cap A_k) + \\
&\quad + \dots + (-1)^{n+2} P(A_1 \cap \dots \cap A_{n+1})
\end{aligned}$$

QED

Ora consideriamo una probabilità boreliana P su \mathbb{R}^2 e la sua funzione di ripartizione

$$F(x_1, x_2) = P((-\infty, x_1] \times (-\infty, x_2])$$

Denotiamo con $A_{x,y} = (-\infty, x] \times (-\infty, y]$ ed osserviamo che

$$\begin{aligned}
\Delta_I F &= P(A_{b_1, b_2}) - P(A_{a_1, b_2}) - P(A_{a_2, b_1}) + P(A_{a_1, a_2}) \\
&= P(A_{b_1, b_2}) - (P(A_{a_1, b_2}) + P(A_{a_2, b_1}) - P(A_{a_1, a_2})) \\
&= P(A_{b_1, b_2}) - P(A_{a_1, b_2} \cup A_{a_2, b_1})
\end{aligned}$$

per il principio di inclusione-esclusione e dato che $A_{a_1, a_2} = A_{a_1, b_2} \cap A_{a_2, b_1}$. Inoltre $A_{a_1, b_2} \cap A_{a_2, b_1} \subset A_{b_1, b_2}$, pertanto troviamo che

$$\Delta_I F \geq 0$$

Viceversa, se F è una funzione di ripartizione basta porre

$$P(I) = \Delta_I F$$

ed estendere per additività alla semialgebra \mathcal{S} dei rettangoli di \mathbb{R}^n , utilizzando il solito criterio di estensione di una probabilità da una semialgebra ad una σ -algebra per ottenere la probabilità boreliana desiderata.

Nel caso di misure qualsiasi abbiamo la seguente

Definizione 2.23

- (1) Una funzione di distribuzione $F : \mathbb{R}^n \rightarrow \mathbb{R}$ è una funzione
- (a) continua superiormente;

(b) tale che $\Delta_I F \geq 0$.

(2) Una misura di Lebesgue–Stieltjes su \mathbb{R}^n è una qualsiasi misura sui boreliani che abbia valori finiti sugli insiemi compatti.

Come ci si può attendere, queste due nozioni sono equivalenti in virtù del teorema seguente, analogo al precedente.

Teorema 2.24 *Esiste una corrispondenza biunivoca fra (classi di equivalenza di) funzioni di distribuzione e misure di Lebesgue–Stieltjes sui boreliani di \mathbb{R}^n .*

Esempio 2.25 Se $F(x_1, \dots, x_n) = \prod_{i=1}^n x_i$ troviamo la misura di Lebesgue sui boreliani di \mathbb{R}^n .

3 Variabili aleatorie

Definizione 3.1 *Se (Ω, \mathcal{A}, P) è uno spazio di probabilità e (Y, \mathcal{B}) è uno spazio misurabile allora*

- (1) $X : \Omega \rightarrow Y$ misurabile si dice ente aleatorio.
- (2) $X : \Omega \rightarrow \mathbb{R}$ misurabile si dice variabile aleatoria (VA).
- (3) $X : \Omega \rightarrow \mathbb{R}^n$ misurabile si dice vettore aleatorio.

Ovviamente

$$X_1, \dots, X_n \text{ è una VA} \iff (X_1, \dots, X_n) \text{ è un vettore aleatorio}$$

Definizione 3.2 *Se $X : \Omega \rightarrow Y$ è un ente aleatorio, la misura*

$$m(B) = P(X^{-1}(B))$$

su \mathcal{B} è una probabilità che si dice misura immagine di X o legge di X o ancora distribuzione di X . Si denota anche come

$$m = X_*(P) \quad \text{oppure} \quad m(B) = P(X \in B)$$

Se $Y = \mathbb{R}^n$ allora m è una probabilità sui boreliani, corrispondente alla funzione di ripartizione

$$\begin{aligned} F(x_1, \dots, x_n) &= m((-\infty, x_1] \times \dots \times (-\infty, x_n]) \\ &= P \circ X^{-1}((-\infty, x_1] \times \dots \times (-\infty, x_n]) \\ &= P(X_1 \leq x_1, \dots, X_n \leq x_n) \end{aligned}$$

dove $X = (X_1, \dots, X_n)$. Viceversa, se $X : \Omega \rightarrow Y$ è una funzione a valori in uno spazio di probabilità (Y, \mathcal{B}, m) vogliamo stabilire quando esiste una probabilità su Ω in modo che la legge di X sia m : la σ -algebra che consideriamo su Ω è l'immagine inversa di \mathcal{B} :

$$\mathcal{A} = \{X^{-1}(B)\}_{B \in \mathcal{B}}$$

Teorema 3.3 *Su Ω esiste una probabilità P la cui legge di X sia m se e solo se per ogni $B \in \mathcal{B}$ se $m(B) > 0$ allora $X^{-1}(B) \neq \emptyset$.*

DIMOSTRAZIONE: Se tale P esiste allora si ha

$$m(B) = P(X \in B)$$

sicché $m(B) > 0$ implica $P(X^{-1}(B)) > 0$ cioè $X^{-1}(B) \neq \emptyset$.

Se invece per ogni $B \in \mathcal{B}$ si ha che $m(B) > 0$ implica $X^{-1}(B) \neq \emptyset$, dato che un generico elemento di una σ -algebra su Ω è della forma $X^{-1}(B)$, basterà definire

$$P(X^{-1}(B)) = m(B)$$

per avere la corrispondente probabilità su Ω .

QED

Dato che le VA sono funzioni misurabili, godono delle stesse proprietà di queste ultime: per esempio sono VA gli indicatori degli insiemi misurabili, e le funzioni semplici, cioè le loro combinazioni lineari

$$s = \sum_{i=1}^n a_i \chi_{A_i}$$

Una funzione semplice è dunque una VA che assume un numero finito di valori $\{a_1, \dots, a_n\}$: ogni tale funzione ammette una rappresentazione standard

$$s = \sum_{i=1}^n x_i \chi_{E_i}$$

dove $E_i = s^{-1}(\{a_i\})$ in modo che $E_i \cap E_j = \emptyset$ se $i \neq j$.

Rinfreschiamo qualche nozione di teoria dell'integrazione secondo Lebesgue, che si applica alle funzioni misurabili, dunque in particolare alle VA: supporremo che (Ω, \mathcal{A}, P) sia uno spazio di probabilità. Cominciamo dal seguente risultato di teoria della misura (cfr. [8], [10] o [1]):

Teorema 3.4 Se $f : \Omega \rightarrow \mathbb{R}$ è misurabile allora esiste una successione $\{f_i\}$ di funzioni semplici su Ω in modo che, per ogni $\omega \in \Omega$:

$$\lim_{n \rightarrow \infty} f_n(\omega) = f(\omega)$$

Inoltre, se $\sup_A |f| < \infty$ allora la convergenza è uniforme, e se $f \geq 0$ possiamo scegliere la successione $\{f_n\}$ in modo che $0 \leq f_1 \leq f_2 \leq \dots \leq f_n \leq \dots \leq f$.

L'integrale di una funzione semplice non negativa si definisce ovviamente come

$$\int f dP = \sum_{i=1}^n a_i P(E_i)$$

Se $f \geq 0$ è una qualsiasi VA allora poniamo

$$\int f dP = \sup_{\substack{0 \leq \varphi \leq f \\ \varphi \text{ semplice}}} \int \varphi dP$$

In generale, per funzioni che possono assumere valori negativi, poniamo $f = f_+ - f_-$ dove f_+ è uguale a f dove f è non negativa e nulla altrove, e f_- è uguale a f dove f è negativa e nulla altrove. Allora si definisce

$$\int f dP = \int f_+ dP - \int f_- dP$$

Evidentemente questa definizione ha senso solo se almeno uno dei due integrali è finito: in generale, se $\int |f| dP = \int f_+ dP + \int f_- dP < \infty$ la funzione f si dice *integrabile*.

Possiamo definire, ovviamente, l'integrale di una VA esteso ad un evento che non sia tutto Ω :

$$\int_A f dP = \int_{\Omega} f \chi_A dP$$

Se $f \geq 0$ è misurabile, la funzione $\alpha : \mathcal{A} \rightarrow \mathbb{R}$ definita come

$$\alpha(A) = \int_A f dP$$

è una misura: dato che, per la formula di cambiamento di variabile nell'integrale di Lebesgue,

$$\int g d\alpha = \int gf dP$$

si dice che f è la *densità di α rispetto a P* .

Definizione 3.5 Una successione $\{X_n\}$ di VA converge in probabilità alla VA X se

$$\forall \varepsilon > 0 \exists n_\varepsilon \in \mathbb{N} \forall n > n_\varepsilon \quad P\{|X_n - X| > \varepsilon\} \rightarrow 0$$

$\{X_n\}$ è una successione di Cauchy in probabilità se

$$\forall \varepsilon > 0 \exists n_\varepsilon \in \mathbb{N} \forall n, m > n_\varepsilon \quad P\{|X_n - X_m| > \varepsilon\} \rightarrow 0$$

Il seguente teorema esprime il cambiamento di variabile in teoria delle probabilità:

Teorema 3.6 Se (Ω, \mathcal{A}, P) è uno spazio di probabilità, (Y, \mathcal{F}) uno spazio misurabile, $X : \Omega \rightarrow Y$ un ente aleatorio e $g : Y \rightarrow \overline{\mathbb{R}}$ misurabile allora

$$\int g(X(\omega)) dP(\omega) = \int g(x) d(P \circ X^{-1})(x)$$

DIMOSTRAZIONE: Se $g = \chi_A$ con A boreliano in \mathbb{R} allora

$$g(X(\omega)) = \begin{cases} 0 & \text{se } X(\omega) \notin A \\ 1 & \text{se } X(\omega) \in A \end{cases}$$

sicché $g \circ X = \chi_{X^{-1}(A)}$, pertanto

$$\begin{aligned} \int g(X(\omega)) dP(\omega) &= \int \chi_{X^{-1}(A)} dP(\omega) = P(X^{-1}(A)) \\ &= \int \chi_A(x) d(P \circ X^{-1})(x) = \int g(x) d(P \circ X^{-1})(x) \end{aligned}$$

Ora, se g è una qualsiasi funzione misurabile ci riconduciamo a questo caso per linearità e monotonia dell'integrale. Precisamente, se

$$g = \sum_i a_i \chi_{A_i}$$

allora

$$\begin{aligned} \int g(X(\omega)) dP(\omega) &= \sum_i a_i \int \chi_{A_i}(X(\omega)) dP(\omega) \\ &= \sum_i a_i \int \chi_{A_i}(x) d(P \circ X^{-1})(x) \\ &= \int g(x) d(P \circ X^{-1})(x) \end{aligned}$$

Se $g \geq 0$ è una qualsiasi funzione misurabile, fissata una successione $\{\varphi_n\}$ di funzioni semplici ad essa convergente, come nel teorema precedente, abbiamo che, per il caso appena discusso,

$$\begin{aligned} \int g(X(\omega)) dP(\omega) &= \lim_n \int \varphi_n(X(\omega)) dP(\omega) \\ &= \lim_n \int \varphi_n(x) d(P \circ X^{-1})(x) = \int g(x) d(P \circ X^{-1})(x) \end{aligned}$$

Se g è qualsiasi, usando la decomposizione $g = g_+ - g_-$ e il caso precedente si dimostra il teorema in generale.

QED

La seguente definizione è di importanza cruciale:

Definizione 3.7 *Se (Ω, \mathcal{A}, P) è uno spazio di probabilità e X è una VA allora l'attesa o speranza matematica di X è*

$$\mathbb{E}[X] = \int X(\omega) dP(\omega)$$

Ovviamente l'attesa di una funzione semplice è

$$\mathbb{E}[S] = \sum_i a_i P(E_i)$$

Inoltre, se $g : \mathbb{R} \rightarrow \mathbb{R}$ è misurabile e $X : \Omega \rightarrow \mathbb{R}$ è una VA allora $g \circ X$ è ancora una VA la cui attesa è

$$\mathbb{E}[g \circ X] = \int g(X(\omega)) dP(\omega) = \int_{\mathbb{R}} g(x) d(P \circ X^{-1})(x)$$

Esempio 3.8 *Se g è l'identità $g(x) = x$ allora*

$$\mathbb{E}[X] = \int_{\mathbb{R}} x d(P \circ X^{-1})(x)$$

Se $m = P \circ X^{-1}$ e se F è la sua funzione di ripartizione, supponendo che m sia della forma

$$m(A) = \int_A f(x) dx$$

con $f \geq 0$, allora

$$\int x d(P \circ X^{-1}) = \int x dm = \int x f(x) dx$$

Consideriamo ora uno spazio di misura (Ω, \mathcal{A}, m) con m concentrata nel punto $\omega \in \Omega$:

$$m(\{\omega\}) = c > 0 \quad \text{e} \quad m(\Omega \setminus \{\omega\}) = 0$$

Ne viene che

$$\begin{aligned} \int f \, dm &= \int_{\Omega \setminus \{\omega\}} f \, dm + \int_{\{\omega\}} f \, dm \\ &= \int \chi_{\{\omega\}} f \, dm = f(\omega)m(\{\omega\}) = cf(\omega) \end{aligned}$$

Se m è concentrata in un insieme numerabile e se $f \geq 0$ allora

$$\begin{aligned} \int f \, dm &= \lim_{n \rightarrow \infty} \int f \chi_{\{\omega_1, \dots, \omega_n\}} \, dm = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(\omega_i)m(\{\omega_i\}) \\ &= \sum_{i=1}^{\infty} f(\omega_i)m(\{\omega_i\}) \end{aligned}$$

Infine, se f non è positiva, dove ha senso vale la seguente identità

$$\int f \, dm = \sum_{i=1}^{\infty} f_+(\omega_i)m(\{\omega_i\}) - \sum_{i=1}^{\infty} f_-(\omega_i)m(\{\omega_i\})$$

nel qual caso abbiamo che

$$\int f \, dm = \sum_{i=1}^{\infty} f(\omega_i)m(\{\omega_i\})$$

se e solo se la serie a secondo membro converge assolutamente (questo è necessario perché l'ordine delle ω_i è ininfluenza, e la somma di una serie convergente non dipende dall'ordine dei suoi termini solo se la serie converge assolutamente, per il teorema di Riemann–Dini).

Nel caso di una VA X che assuma un numero numerabile di valori $\{\omega_i\}$ su uno spazio di probabilità (Ω, \mathcal{A}, P) abbiamo dunque che

$$\begin{aligned} \mathbb{E}[X] &= \int X(\omega) \, dP(\omega) = \int \omega \, d(P \circ X^{-1})(\omega) \\ &= \sum_{i=1}^{\infty} \omega_i (P \circ X^{-1})(\{\omega_i\}) = \sum_{i=1}^{\infty} \omega_i P(X = \omega_i) \end{aligned}$$

Se X non è positiva,

$$\mathbb{E}[X] \text{ esiste finita} \iff \sum_{i=1}^{\infty} \omega_i P(X = \omega_i) \text{ converge assolutamente}$$

ed in tal caso

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \omega_i P(X = \omega_i)$$

Per finire questi richiami di teoria dell'integrazione applicati al nostro caso probabilistico, rammentiamo, nella forma a noi consona, alcuni risultati ben noti ($A \in \mathcal{A}$ e X, Y sono VA)

$$(1) \int_A X dP \geq \int_A Y dP \iff X \geq Y \text{ q.o.}$$

$$(2) \int_A X dP = \int_A Y dP \iff X = Y \text{ q.o.}$$

(3) Se $\mathbb{E}[X] < \infty$ e $\mathbb{E}[Y] < \infty$ allora $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$ per ogni $a, b \in \mathbb{R}$.

(4) *Disuguaglianza di Čebišef*: se $f > 0$ è monotona non decrescente allora

$$\forall x > 0 \quad P(|X| > x) \leq \frac{\mathbb{E}[f(|X|)]}{f(x)}$$

(5) *Disuguaglianza di Jensen*: se g è continua e convessa su (a, b) e se $a < X < b$ q.o. allora

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

(6) *Disuguaglianza di Hölder*: se $1/p + 1/q = 1$ e $p > 1$ allora

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|X|^p]^{1/p} \mathbb{E}[|Y|^q]^{1/q}$$

(7) *Disuguaglianza di Cauchy*:

$$\mathbb{E}[XY]^2 \leq \mathbb{E}[X^2] \mathbb{E}[Y^2]$$

(8) *Disuguaglianza di Minkowski*: se $p > 1$:

$$\mathbb{E}[|X + Y|^p]^{1/p} \leq \mathbb{E}[|X|^p]^{1/p} + \mathbb{E}[|Y|^p]^{1/p}$$

4 Indipendenza, attese condizionate e martingale

Definizione 4.1 Due eventi $A, B \in \mathcal{A}$ in uno spazio di probabilità (Ω, \mathcal{A}, P) sono indipendenti se

$$P(A \cap B) = P(A)P(B)$$

Evidentemente Ω e \emptyset sono indipendenti da qualsiasi altro evento; più in generale, un evento di probabilità zero è indipendente da qualsiasi altro.

Definizione 4.2 Se $P(B) > 0$ diciamo che la quantità

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

è la probabilità di A condizionata a B , ovvero la probabilità di A dato B .

Se A e B sono indipendenti, si ha ovviamente che $P(A|B) = P(A)$ cioè che la probabilità di A dato B non dipende in effetti da quella di B (che però deve essere positiva affinché la definizione abbia senso). Se invece A e B sono incompatibili, cioè se $P(A \cap B) = 0$, allora $P(A|B) = 0$: in altri termini se due eventi si escludono a vicenda, dal verificarsi dell'uno sappiamo che l'altro non può essersi verificato.

Attenzione: il lettore che conosce un po' di logica, interpretando l'algebra degli eventi come l'algebra di Boole del calcolo proposizionale, potrebbe presumere che $P(A|B)$ sia la probabilità che $B \Rightarrow A$; *questo è falso*. Infatti $B \Rightarrow A$ è logicamente equivalente a $\neg B \vee A$ cioè, insiemisticamente, $\complement B \cup A$, mentre evidentemente

$$P(\complement B \cup A) \neq P(A|B)$$

(basti pensare al caso $B = \complement A$, in cui $P(\complement B \cup A) = P(A)$, mentre $P(A|\complement A) = 0$).

Inoltre, se $C = \complement B$ allora

$$P(A) = P(A|B)P(B) + P(A|C)P(C)$$

Da questo possiamo derivare l'interessante

Teorema di Bayes 4.3 Se $C = \complement B$ allora

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

DIMOSTRAZIONE: Poiché $P(B|A)P(A) = P(A \cap B) = P(A|B)P(B)$ ne viene $P(B|A)P(A) = P(A|B)P(B)$ e, usando la precedente formula per $P(A)$, troviamo la formula di Bayes. QED

L'interesse del teorema di Bayes sta nel fatto che consente di esprimere $P(B|A)$ in termini di $P(A|B)$: di solito si dice che $P(B|A)$ è la *probabilità a posteriori* interpretando B come una ipotesi la cui probabilità è condizionata dall'osservazione del dato A . Questa interpretazione è utilizzata per esempio nella costruzione dei sistemi esperti, cioè di programmi per calcolatore che cercano di dedurre la "causa" di una osservazione¹².

Ovviamente una famiglia $\{A_i\}$ di eventi è indipendente se

$$P\left(\bigcap_i A_i\right) = \prod_i P(A_i)$$

Ricordiamo che il lemma di Borel-Cantelli 1.12 afferma che la probabilità del limite superiore di una successione di eventi è nulla se la serie delle loro probabilità è convergente: vogliamo dimostrare che invece, se questa serie diverge, la probabilità del limite superiore è uno, a patto che gli eventi siano indipendenti:

Teorema 4.4 *Se $\{A_n\}$ è una successione di eventi indipendenti, allora*

$$P\left(\limsup_{n \rightarrow \infty} A_n\right) = \begin{cases} 0 \\ 1 \end{cases} \quad \text{se} \quad \sum_{n=1}^{\infty} P(A_n) \quad \begin{cases} \text{converge} \\ \text{diverge} \end{cases}$$

DIMOSTRAZIONE: Per il lemma di Borel-Cantelli, sappiamo che se la serie converge allora la probabilità è zero (anche senza l'ipotesi di indipendenza). Mostriamo quindi che, se la serie diverge, la probabilità del limite è uno, mostrando che la probabilità del complementare è zero:

$$\begin{aligned} P\left(\mathbb{C} \limsup_{n \rightarrow \infty} A_n\right) &= P\left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \mathbb{C}A_k\right) = \lim_{n \rightarrow \infty} P\left(\bigcap_{k=n}^{\infty} \mathbb{C}A_k\right) \\ &= \lim_{n \rightarrow \infty} \prod_{k=n}^{\infty} P(\mathbb{C}A_k) = \lim_{n \rightarrow \infty} \prod_{k=n}^{\infty} (1 - P(A_k)) = 0 \end{aligned}$$

¹²Per esempio da quale malattia B sia afflitto un paziente in base ai sintomi A ; è più facile determinare $P(A|B)$ che $P(B|A)$ in base all'osservazione: per conoscere $P(A|B)$ basta consultare una enciclopedia medica (che enumera le malattie e per ciascuna di esse fornisce i sintomi), mentre per conoscere $P(B|A)$ non si può leggere l'enciclopedia medica scorrendo tutte le malattie fino a trovare i sintomi che il paziente possiede!

Abbiamo usato le leggi di de Morgan al primo passaggio, l'indipendenza degli $\{A_n\}$ al terzo passaggio e l'ipotesi di divergenza della serie all'ultimo: infatti dato che la serie diverge, la successione $\{P(A_n)\}$ non può tendere a zero, e quindi i prodotti $(1 - P(A_n))$ sono definitivamente minori di uno.

QED

Vogliamo ora dimostrare la cosiddetta “legge del tutto-o-niente” di Kolmogorov: per farlo abbiamo bisogno di una

Definizione 4.5 *Due insiemi \mathcal{I}, \mathcal{J} di eventi si dicono indipendenti se per ogni $A \in \mathcal{I}$ e per ogni $B \in \mathcal{J}$ A e B sono indipendenti. Una famiglia $\{C_i\}_{i \in I}$ di insiemi di eventi si dice che i suoi elementi sono indipendenti se per ogni sottoinsieme finito $\{i_1, \dots, i_n\} \subset I$ e per ogni $A_{i_1} \in \mathcal{I}_{i_1}, \dots, A_{i_n} \in \mathcal{I}_{i_n}$ questi eventi sono indipendenti.*

Possiamo quindi parlare di σ -algebre indipendenti: se $\{\mathcal{B}_n\}$ è una successione di σ -algebre indipendenti e se scegliamo un evento da ciascuna di queste σ -algebre, a formare una successione $\{A_n\}$, il teorema precedente ci dice che il limite superiore di $\{A_n\}$ può solo avere probabilità zero oppure uno.

Consideriamo ora una famiglia $\{\mathcal{B}_n\}$ di σ -algebre: possiamo da questa produrre una successione monotona decrescente (rispetto all'inclusione) di σ -algebre, prendendo le σ -algebre generate dagli elementi della successione $\{\mathcal{B}_n\}$ presi da un certo punto in poi:

$$\mathcal{B}'_n = \sigma(\mathcal{B}_k)_{k \geq n}$$

L'intersezione \mathcal{B} delle \mathcal{B}'_n ancora una σ -algebra che chiamiamo *limite superiore della successione di σ -algebre $\{\mathcal{B}_n\}$* . Per costruzione, questo limite rimane lo stesso se si cambia un numero finito di elementi della successione $\{\mathcal{B}_n\}$.

Teorema del tutto-o-niente di Kolmogorov 4.6 *Se $\{\mathcal{B}_n\}$ sono σ -algebre indipendenti, allora ogni evento che sta in \mathcal{B} ha probabilità zero oppure uno.*

DIMOSTRAZIONE: Se $A \in \mathcal{B}$ allora per ogni $k \in \mathbb{N}$ si ha che $A \in \mathcal{B}'_k$, e quindi A è indipendente dalla σ -algebra generata da $\{\mathcal{B}_1, \dots, \mathcal{B}_{k-1}\}$. Dato che k è qualsiasi, questo vuol dire che A è indipendente dalla σ -algebra generata dall'intera successione $\{\mathcal{B}_n\}$, e, dato che A appartiene a questa σ -algebra, questo significa che A è indipendente da A , cioè $P(A) = P(A \cap A) = P(A)^2$, il che è possibile solo se $P(A)$ è zero oppure uno. QED

La rilevanza del concetto di probabilità condizionata consiste nel fatto che consente di formalizzare il seguente problema: supponiamo di osservare dei fenomeni e di voler capire come queste osservazioni ci possano aiutare nel prevederne la probabilità di altri. Dal punto di vista matematico le osservazioni

corrispondono agli eventi di una σ -algebra \mathcal{B} , e quindi quello che vogliamo definire è $P(A|\mathcal{B})$.

Consideriamo prima il caso di una σ -algebra semplice \mathcal{B} su un insieme Ω : questo vuol dire che Ω è unione di una successione $\{E_n\}$ di insiemi disgiunti (chiamati *atomi*) e che ogni elemento di \mathcal{B} è unione finita di qualche E_n (dobbiamo che supporre che $\emptyset \in \mathcal{B}$). Notiamo che in questo caso una funzione misurabile è per forza semplice, dato che deve essere costante sugli atomi.

Se una σ -algebra \mathcal{A} qualsiasi ed una sua sotto- σ -algebra semplice \mathcal{B} , ed una probabilità P su \mathcal{B} tale che $P(E_n) > 0$ per ogni n . Se $A \in \mathcal{A}$ allora definiamo la *probabilità condizionata rispetto alla σ -algebra \mathcal{B}* come

$$P(A|\mathcal{B}) = \sum_n P(A|E_n)\chi(E_n) = \sum_n \frac{P(A \cap E_n)}{P(E_n)}\chi(E_n)$$

In questo modo la probabilità condizionata è una VA \mathcal{B} -misurabile: possiamo usarla per definire l'*attesa condizionata* come

$$\mathbb{E}[\xi|\mathcal{B}] = \int_{\Omega} \xi P(d\omega|\mathcal{B}) = \sum_n \frac{\chi(E_n)}{P(E_n)} \int_{E_n} \xi dP$$

Notiamo che, per $\xi = \chi(A)$ ritroviamo la probabilità condizionata:

$$\mathbb{E}[\chi(A)|\mathcal{B}] = \sum_n \frac{\chi(E_n)}{P(E_n)} P(A \cap E_n) = P(A|\mathcal{B})$$

Se η è una VA \mathcal{B} -misurabile e limitata, possiamo scrivere

$$\eta = \sum_n a_n \chi(E_n)$$

da cui

$$(*) \quad \mathbb{E}[\eta \mathbb{E}[\xi|\mathcal{B}]] = \sum_n a_n \int_{E_n} \xi dP = \sum_n \int_{E_n} \eta \xi dP = \mathbb{E}[\eta \xi]$$

In particolare, se $\eta = \chi(B)$ (con $B \in \mathcal{B}$) allora

$$\int_A \mathbb{E}[\xi|\mathcal{B}] dP = \int_B \xi dP$$

Usiamo questa ultima formula per definire l'*attesa condizionata* nel caso generale (cioè dove \mathcal{B} non è necessariamente semplice).

Definizione 4.7 L'attesa (o speranza) matematica condizionata $\mathbb{E}[\xi|\mathcal{B}]$ di una VA ξ integrabile rispetto a una σ -algebra \mathcal{B} è una VA \mathcal{B} -misurabile e tale che

$$\forall B \in \mathcal{B} \quad \int_B \mathbb{E}[\xi|\mathcal{B}] dP = \int_B \xi dP$$

Notiamo che $\mathbb{E}[\xi|\mathcal{B}]$ è una VA laddove $\mathbb{E}[\xi]$ è una costante.

Teorema 4.8 Data una VA ξ integrabile e una σ -algebra \mathcal{B} rispetto a cui ξ sia misurabile, l'attesa condizionata $\mathbb{E}[\xi|\mathcal{B}]$ esiste ed è unica q.o. (rispetto a P).

DIMOSTRAZIONE: Date ξ e \mathcal{B} come nell'enunciato, la misura

$$\mu(B) = \int_B \xi dP$$

è σ -finita su \mathcal{B} assolutamente continua rispetto a P , e quindi il teorema di Radon–Nikodym esiste una funzione $g : \Omega \rightarrow \mathbb{R}$ non negativa e \mathcal{B} -misurabile tale che

$$\mu(B) = \int_B g(\omega) dP(\omega)$$

Ponendo $\mathbb{E}[\xi|\mathcal{B}] = g$ troviamo la VA richiesta, che è unica q.o. sempre per il teorema di Radon–Nikodym.

QED

Di nuovo ispirandoci al caso delle σ -algebre semplici, diamo la seguente

Definizione 4.9 La probabilità condizionata di un evento A rispetto a una σ -algebra \mathcal{B} è la VA

$$P(A|\mathcal{B}) = \mathbb{E}[\chi(A)|\mathcal{B}]$$

In virtù della definizione di attesa condizionata, possiamo caratterizzare la probabilità condizionata per mezzo della relazione

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

In effetti l'esistenza della probabilità condizionata potrebbe dedursi dal teorema di Radon–Nikodym, definendo $P(A|\mathcal{B})$ come una VA \mathcal{B} -misurabile e tale che valga la relazione precedente.

Proposizione 4.10 *Sia ξ una VA e \mathcal{B} una σ -algebra: allora*

(1) *Se ξ è \mathcal{B} -misurabile allora $\mathbb{E}[\xi|\mathcal{B}] = \xi$.*

(2) *Se ξ_1 e ξ_2 sono VA integrabili allora*

$$\mathbb{E}[\xi_1 + \xi_2|\mathcal{B}] = \mathbb{E}[\xi_1|\mathcal{B}] + \mathbb{E}[\xi_2|\mathcal{B}]$$

(3) *Se $\xi \leq \eta$ allora $\mathbb{E}[\xi|\mathcal{B}] \leq \mathbb{E}[\eta|\mathcal{B}]$*

(4) *Se $\{\xi_n\}$ è una successione monotona non decrescente di VA non negative allora*

$$\lim_{n \rightarrow \infty} \mathbb{E}[\xi_n|\mathcal{B}] = \mathbb{E}\left[\lim_{n \rightarrow \infty} \xi_n \mid \mathcal{B}\right]$$

(5) *Se α è una VA integrabile e \mathcal{B} -misurabile, allora $\mathbb{E}[\alpha|\mathcal{B}] = \alpha$.*

DIMOSTRAZIONE: La (1) segue immediatamente dal fatto che, se $\mathbb{E}[\xi|\mathcal{B}]$ è una \mathcal{B} -misurabile che soddisfa l'equazione della definizione 4.7 ed ovviamente anche ξ se è \mathcal{B} -misurabile soddisfa questa definizione, quindi per l'unicità abbiamo q.o. che $\mathbb{E}[\xi|\mathcal{B}] = \xi$.

La (2) segue in modo analogo dalla linearità dell'integrale: infatti la VA $\mathbb{E}[\xi_1|\mathcal{B}] + \mathbb{E}[\xi_2|\mathcal{B}]$ soddisfa la definizione di attesa condizionata per la VA $\xi_1 + \xi_2$ dato che

$$\begin{aligned} \int_B (\mathbb{E}[\xi_1|\mathcal{B}] + \mathbb{E}[\xi_2|\mathcal{B}]) dP &= \int_B \mathbb{E}[\xi_1|\mathcal{B}] dP + \int_B \mathbb{E}[\xi_2|\mathcal{B}] dP \\ &= \int_B \xi_1 dP + \int_B \xi_2 dP = \int_B (\xi_1 + \xi_2) dP \end{aligned}$$

(per definizione di $\mathbb{E}[\xi_1|\mathcal{B}]$ e $\mathbb{E}[\xi_2|\mathcal{B}]$.)

La (3) segue dall'osservare che se $\eta \geq 0$ allora $\int_B \mathbb{E}[\eta|\mathcal{B}] dP \geq 0$ per ogni $B \in \mathcal{B}$, pertanto $\mathbb{E}[\eta|\mathcal{B}] \geq 0$. Ora usiamo la (2) notando che, se η è qualsiasi, $\eta = \xi + (\eta - \xi)$ da cui

$$\mathbb{E}[\eta|\mathcal{B}] = \mathbb{E}[\xi + (\eta - \xi)|\mathcal{B}] = \mathbb{E}[\xi|\mathcal{B}] + \mathbb{E}[\eta - \xi|\mathcal{B}] \geq \mathbb{E}[\xi|\mathcal{B}]$$

dato che $\eta - \xi \geq 0$.

La (4) segue pure dall'unicità di $\mathbb{E}[\xi|\mathcal{B}]$ e dal teorema di convergenza di Beppo Levi: per ogni $B \in \mathcal{B}$

$$\int_B \lim_{n \rightarrow \infty} \mathbb{E}[\xi_n|\mathcal{B}] dP = \lim_{n \rightarrow \infty} \int_B \mathbb{E}[\xi_n|\mathcal{B}] dP = \lim_{n \rightarrow \infty} \int_B \xi_n dP = \int_B \xi dP$$

Infine la (5): notiamo che l'abbiamo già dimostrata nel caso in cui \mathcal{B} sia una σ -algebra semplice (cfr. equazione (*) a pagina 33): allora basta usare

la (4), esprimendo sia α che ξ come limiti di successioni non decrescenti non negative (stiamo quindi supponendo $\alpha \geq 0$ e $\xi \geq 0$) per passare al limite la relazione (*). La limitazione di non negatività su α e ξ non lede la generalità: lo si vede usando la (2) come nella dimostrazione della (3).

QED

Notiamo che la (2) e la (5) implicano una sorta di linearità: se ξ_1 e ξ_2 sono VA integrabili e se α_1, α_2 sono VA integrabili e \mathcal{B} -misurabili allora

$$\mathbb{E}[\alpha_1 \xi_1 + \alpha_2 \xi_2 | \mathcal{B}] = \alpha_1 \mathbb{E}[\xi_1 | \mathcal{B}] + \alpha_2 \mathbb{E}[\xi_2 | \mathcal{B}]$$

Notiamo anche una interessante interpretazione geometrica della relazione

$$\mathbb{E}[\alpha \mathbb{E}(\xi | \mathcal{B})] = \mathbb{E}[\alpha \xi]$$

che segue immediatamente dalla (5). Supponiamo che $\mathbb{E}[\xi^2] < \infty$: le VA siffatte si possono interpretare come elementi dello spazio di Hilbert $L^2(\Omega, \mathcal{A}, P)$. Quest'ultimo contiene il sottospazio di Hilbert $\mathcal{H} = L^2(\Omega, \mathcal{A}, P)$ delle VA \mathcal{B} -misurabili e tali che $\mathbb{E}[\xi^2] < \infty$: notiamo che $\mathbb{E}[\xi | \mathcal{B}] \in \mathcal{H}$ e che per ogni $\alpha \in \mathcal{H}$, in virtù della relazione precedente, abbiamo che

$$\mathbb{E}[\alpha(\mathbb{E}[\xi | \mathcal{B}] - \xi)] = \mathbb{E}[\alpha \mathbb{E}[\xi | \mathcal{B}] - \alpha \xi] = 0$$

Questo vuol dire che il vettore $\mathbb{E}[\xi | \mathcal{B}] - \xi$ è ortogonale al sottospazio \mathcal{H} e quindi che $\mathbb{E}[\xi | \mathcal{B}]$ è la *proiezione ortogonale di ξ sul sottospazio \mathcal{H}* .

La proposizione precedente ammette un corollario immediato per le probabilità condizionate, ponendo $\chi(A)$ al posto di ξ :

Corollario 4.11 *Se $A, B \in \mathcal{A}$ e \mathcal{B} è una sotto- σ -algebra di \mathcal{A} allora*

- (1) *Se $A \in \mathcal{B}$ allora $P(A|\mathcal{B}) = \chi(A)$.*
- (2) *Se $A \cap B = \emptyset$ allora $P(A \cup B | \mathcal{B}) = P(A|\mathcal{B}) + P(B|\mathcal{B})$ q.o.*
- (3) *Se $A \subset B$ allora $P(A|\mathcal{B}) \leq P(B|\mathcal{B})$.*
- (4) *Se $\{A_n\}$ è una successione di eventi a due a due disgiunti allora*

$$P\left(\bigcup_{n=1}^{\infty} A_n | \mathcal{B}\right) = \sum_{n=1}^{\infty} P(A_n | \mathcal{B})$$

- (5) *Se $A \in \mathcal{A}$ e $B \in \mathcal{B}$ allora $P(A \cap B | \mathcal{B}) = \chi(B)P(A|\mathcal{B})$.*

Usando la nozione di attesa condizionata possiamo dare la definizione di martingala, un esempio elementare di processo stocastico, cioè di famiglia di VA dipendenti da un parametro reale: in più, una martingala ha la proprietà di “non avere memoria”.

Consideriamo uno spazio di probabilità (Ω, \mathcal{A}, P) e un insieme non vuoto di numeri reali $T \subset \mathbb{R}$, i cui elementi vanno interpretati come differenti istanti di osservazione dei fenomeni che vogliamo descrivere: questi fenomeni sono rappresentati da eventi di σ -algebre. Supponiamo quindi che esista, per ogni $t \in T$, una σ -algebra \mathcal{B}_t che contiene i possibili eventi all'istante t : dato che immaginiamo un processo evolutivo nel tempo, supponiamo che la famiglia $\{\mathcal{B}_t\}$ sia monotona, cioè che

$$s < t \implies \mathcal{B}_s \subset \mathcal{B}_t$$

Definizione 4.12 *Una famiglia $\{\xi_t\}_{t \in T}$ di VA si dice adattata ad una famiglia monotona di σ -algebre $\{\mathcal{B}_t\}$ se per ogni $t \in T$, la VA $\xi(t)$ è \mathcal{B}_t -misurabile.*

Una famiglia monotona di σ -algebre si dice anche una *famiglia filtrata*.

Definizione 4.13 *Una famiglia $\{\xi_t\}_{t \in T}$ di VA adattata ad una famiglia monotona di σ -algebre $\{\mathcal{B}_t\}$ si dice:*

- (1) martingala se $s < t$ allora $\mathbb{E}[\xi(t)|\mathcal{B}_s] = \xi(s)$.
- (2) submartingala se $s < t$ allora $\mathbb{E}[\xi(t)|\mathcal{B}_s] \geq \xi(s)$.
- (3) supermartingala se $s < t$ allora $\mathbb{E}[\xi(t)|\mathcal{B}_s] \leq \xi(s)$.

Il processo di Wiener avrà la proprietà di essere una martingala: in generale le martingale sono strumenti teorici e pratici di importanza capitale¹³ sui quali non posso tuttavia soffermarmi ulteriormente.

5 Funzioni caratteristiche e variabili gaussiane

Se X è una VA reale allora e^{-itX} è una VA complessa integrabile (dato che $|e^{i\vartheta}| = 1$ se $\vartheta \in \mathbb{R}$!) sicché è definita la funzione di variabile reale

$$\widehat{X}(\xi) = \mathbb{E} [e^{-i\xi x}]$$

¹³Nel vero senso del termine: in finanza permettono di calcolare i prezzi degli strumenti derivati, sotto opportune ipotesi di modellizzazione dei mercati finanziari.

che si dice *funzione caratteristica* (FC) della VA X : non è altro (a meno di un fattore $1/\sqrt{2\pi}$) che la trasformata di Fourier.

Se F_X è la funzione di distribuzione di X

$$F_X(x) = P(X < x)$$

allora

$$\widehat{X}(\xi) = \int e^{-i\xi x} dF_X(x) = \int e^{-i\xi x} d\mu_X(x)$$

dove μ_X è la distribuzione di X : $\mu_X(B) = P \circ X^{-1}(B)$. Ritroviamo la classica trasformata di Fourier se μ_X è della forma $f(x)dx$ ove dx denota la misura di Lebesgue.

Proposizione 5.1

- (1) $\widehat{X_1 + X_2}(\xi) = \widehat{X_1}(\xi)\widehat{X_2}(\xi)$ se X_1 e X_2 sono indipendenti.
- (2) $\widehat{aX}(\xi) = \widehat{X}(a\xi)$ se a è costante.
- (3) \widehat{X} determina univocamente la distribuzione di μ_X .

Questo segue dalle proprietà elementari della trasformata di Fourier¹⁴.

Utilizziamo ora un calcolo standard di teoria di Fourier per calcolare la FC della distribuzione normale in dimensione uno. Rammentiamo per prima cosa che (cfr. per esempio [1, §8.8.5.1])

$$\int_{\mathbb{R}} e^{-t^2} dt = \sqrt{\pi}$$

Pertanto la funzione

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

è una funzione di distribuzione: troviamone la FC:

$$\begin{aligned} \widehat{\Phi}(\xi) &= \int e^{-i\xi x} d\Phi(x) = \frac{1}{\sqrt{2\pi}} \int e^{-i\xi x - \frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int e^{-\frac{\xi^2}{2} - \frac{(x+i\xi)^2}{2}} dx = \frac{e^{-\frac{\xi^2}{2}}}{\sqrt{2\pi}} \int_{\mathbb{R}'} e^{-\frac{t^2}{2}} dt \end{aligned}$$

dove

$$\mathbb{R}' = \{x + i\xi \in \mathbb{C} \mid x \in \mathbb{R}\}$$

¹⁴cfr. per esempio: [7], [8], [1].

è la retta del piano complesso parallela all'asse reale $\mathbb{R} = \{x + iy \in \mathbb{C} \mid y = 0\}$, sicché l'integrale su \mathbb{R} è lo stesso che su \mathbb{R}' (invarianza per traslazioni della misura di Lebesgue), da cui

$$\widehat{\Phi}(\xi) = \frac{e^{-\frac{\xi^2}{2}}}{\sqrt{2\pi}} \sqrt{2\pi} = e^{-\frac{\xi^2}{2}}$$

Ora, se X è una VA con distribuzione $\Phi(x) dx$, dato che

$$\left(\frac{d^k}{d\xi^k} \widehat{X} \right)_{\xi=0} = (-i)^k \mathbb{E} [X^k]$$

allora

$$\mathbb{E} [X] = \left(\frac{d}{d\xi} \widehat{X}(\xi) \right)_{\xi=0} = \left(-\xi e^{-\frac{\xi^2}{2}} \right)_{\xi=0} = 0$$

ed inoltre

$$\begin{aligned} \text{Var}(X) &= \mathbb{E} [X - \mathbb{E} [X^2]] = \mathbb{E} [X^2] = \frac{1}{(-i)^2} \left(\frac{d^2}{d\xi^2} \widehat{X}(\xi) \right)_{\xi=0} \\ &= - \left(-e^{-\frac{\xi^2}{2}} + \xi^2 e^{-\frac{\xi^2}{2}} \right)_{\xi=0} = 1 \end{aligned}$$

Pertanto, per la VA $Y = \sigma X + \mu$ (con $\sigma > 0$) troviamo

$$\mathbb{E} [Y] = \mathbb{E} [\mu] = \mu \quad \text{e} \quad \text{Var}(Y) = \mathbb{E} [(Y - \mu)^2] = \mathbb{E} [(\sigma X)^2] = \sigma^2$$

La distribuzione di F_Y è data da:

$$\begin{aligned} F_Y(x) &= P(Y < x) = P \left(X < \frac{x - \mu}{\sigma} \right) = \Phi \left(\frac{x - \mu}{\sigma} \right) \\ &= \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \end{aligned}$$

e la FC di Y è data da:

$$\begin{aligned} \widehat{Y}(\xi) &= \mathbb{E} [e^{-i\xi Y}] = e^{-i\xi\mu} \mathbb{E} [e^{-i\xi\sigma Y}] \\ &= e^{-i\xi\mu} \widehat{X}(\sigma\xi) = e^{-i\xi\mu - \frac{\sigma^2\xi^2}{2}} \end{aligned}$$

Definizione 5.2 *La distribuzione*

$$N(\mu, \sigma^2)(x) = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

si dice distribuzione normale e, nel caso $\mu = 0$ e $\sigma = 1$, distribuzione standard.

Notiamo che la FC di una VA X soddisfa alle seguenti condizioni:

(1) \widehat{X} è definita positiva, cioè per ogni $\xi_1, \dots, \xi_n \in \mathbb{R}$ e $z_1, \dots, z_n \in \mathbb{C}$, si ha

$$\sum_{i,j} z_i \overline{z_j} \widehat{X}(\xi_i - \xi_j) \geq 0$$

(2) \widehat{X} è uniformemente continua.

(3) $\widehat{X}(0) = 1$.

Queste proprietà caratterizzano le FC in virtù del

Teorema di Bochner 5.3 *Se Φ è una funzione positiva, uniformemente continua e tale che $\Phi(0) = 1$ allora esiste un'unica misura di probabilità μ sulla retta reale rispetto alla σ -algebra dei boreliani tale che*

$$\Phi(\xi) = \int_{\mathbb{R}} e^{i\xi x} d\mu(x)$$

per il quale si rimanda a [7, §IV.2] o, in una veste più astratta e generale, a [1, Teorema 14.2.17].

Definizione 5.4 *Una VA $X : \Omega \rightarrow \mathbb{R}^n$ si dice variabile gaussiana se la FC della sua distribuzione ha la forma*

$$\widehat{X}(\xi) = \mathbb{E} [e^{-i\langle \xi, x \rangle}] = e^{-i\langle m, \xi \rangle - \frac{\langle A\xi, \xi \rangle}{2}}$$

ove $m, \xi \in \mathbb{R}^n$ e $A = ((a_{ij}))$ è una matrice simmetrica semidefinita positiva e $\langle v, w \rangle = v_1 w_1 + \dots + v_n w_n$ denota il prodotto scalare standard in \mathbb{R}^n .

Cominciamo col caratterizzare le VA gaussiane.

Teorema 5.5 *La funzione*

$$\Psi(\xi) = e^{-i\langle m, \xi \rangle - \frac{\langle A\xi, \xi \rangle}{2}}$$

è la FC di una VA $X : \Omega \rightarrow \mathbb{R}^n$ se e solo se A è una matrice simmetrica semidefinita positiva il cui nucleo è ortogonale al supporto della distribuzione di X .

DIMOSTRAZIONE: Sia X una VA gaussiana: a conti fatti abbiamo che

$$\begin{aligned} \left(\frac{\partial}{\partial \xi_j} \widehat{X}(\xi) \right)_{\xi=0} &= -i \mathbb{E}[X_j] = -im_j \\ \left(\frac{\partial^2}{\partial \xi_j \partial \xi_k} \widehat{X}(\xi) \right)_{\xi=0} &= \mathbb{E}[X_j X_k] = m_j m_k + a_{jk} \end{aligned}$$

ne viene che A è reale e simmetrica. Inoltre

$$\langle A\xi, \xi \rangle = \mathbb{E}[\langle X - m, \xi \rangle^2] \geq 0$$

Sia ora $r \leq n$ il rango di A : a meno di cambiare coordinate, possiamo supporre (teorema di Sylvester sulla forma canonica di una matrice simmetrica) che sia

$$\langle A\xi, \xi \rangle = \sum_{j=1}^r \lambda_j \zeta_j^2$$

dove $(\zeta_1, \dots, \zeta_n)$ sono le coordinate in \mathbb{R}^n e quindi

$$\langle A\xi, \xi \rangle = \mathbb{E}[\langle X - m, C\zeta \rangle^2]$$

dove C è la matrice invertibile del cambiamento di base che mette A nella forma canonica: $\xi = C\zeta$. Dunque, se $\xi \in \ker A$, abbiamo che

$$\mathbb{E}[\langle X - m, C\zeta \rangle^2] = 0 \implies \langle X - m, C\zeta \rangle = 0 \quad \text{q.o.}$$

e quest'ultima è l'equazione dell'iperpiano che contiene il supporto della distribuzione di X , che ha dimensione r .

Viceversa, se A è una matrice simmetrica semidefinita positiva, la funzione

$$\Psi(\xi) = e^{-i\langle m, \xi \rangle - \frac{\langle A\xi, \xi \rangle}{2}}$$

è assolutamente integrabile ed è anche derivabile: possiamo quindi prenderne la trasformata di Fourier

$$f(x) = \frac{1}{(2\pi)^{n/2}} \int \Psi(\xi) e^{-i\langle \xi, x \rangle} d\xi$$

e, per la formula di inversione, esprimerla come

$$\Psi(x) = \frac{1}{(2\pi)^{n/2}} \int f(x) e^{i\langle \xi, x \rangle} dx$$

Ora supponiamo che sia $\det A \neq 0$ e che C sia una matrice che diagonalizza A (la cui esistenza è data dal teorema spettrale, essendo A simmetrica). Allora (indichiamo con C^* l'aggiunta, cioè la trasposta coniugata)

$$C^*AC = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

(con $\lambda_i > 0$.) Poiché C è una matrice ortogonale (reale), abbiamo che $C^* = C^{-1}$ e quindi

$$\begin{aligned} f(x) &= \frac{1}{(2\pi)^{n/2}} \int e^{i\langle x-m, \xi \rangle - \frac{\langle A\xi, \xi \rangle}{2}} d\xi \\ &= \frac{1}{(2\pi)^{n/2}} \int e^{i\langle x-m, C\zeta \rangle - \frac{\langle AC\zeta, C\zeta \rangle}{2}} d\zeta \end{aligned}$$

dato che la forma di volume $d\xi$ è invariante per trasformazioni ortogonali. Ora: dato che

$$\begin{cases} \langle AC\zeta, C\zeta \rangle = \langle C^*AC\zeta, \zeta \rangle = \sum_k \lambda_k \zeta_k^2 \\ \langle x-m, C\zeta \rangle = \langle C^*(x-m), \zeta \rangle = \sum_k x_k^* \zeta_k \end{cases}$$

ove $x^* = C(x-m)$, abbiamo che

$$\begin{aligned} f(x) &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{-i\sum_k x_k^* \zeta_k - \frac{1}{2}\sum_k \lambda_k \zeta_k^2} d\zeta \\ &= \prod_{k=1}^n \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}} e^{-ix_k^* \zeta_k - \frac{1}{2}\lambda_k \zeta_k^2} d\zeta_k \\ &= \prod_{k=1}^n \frac{1}{(2\pi)^{n/2}} (2\pi)^{n/2} e^{-\frac{1}{2}\lambda_k x_k^{*2}} \\ &= \prod_{k=1}^n \frac{1}{\sqrt{\lambda_k}} e^{-\frac{x_k^{*2}}{2\lambda_k}} = \frac{1}{\sqrt{\det A}} e^{-\frac{1}{2}\langle C^*A^{-1}Cx^*, x^* \rangle} \end{aligned}$$

Con ciò abbiamo dimostrato che

$$f(x) = \frac{1}{\sqrt{\det A}} e^{-\frac{1}{2}\langle C^*A^{-1}Cx^*, x^* \rangle} = \frac{1}{\sqrt{\det A}} e^{-\frac{1}{2}\sum_{j,k} \frac{A_{kj}(x_j-m_j)(x_k-m_k)}{\det A}}$$

Definizione 5.6 Una famiglia $\{X_i\}_{i \in I}$ di VA si dice un sistema gaussiano se ogni combinazione lineare finita di suoi elementi è una VA gaussiana. Se $I \subset \mathbb{R}$ il sistema gaussiano si dice processo gaussiano.

Se (X_1, \dots, X_n) è un sistema gaussiano poniamo, se $\lambda = (\lambda_1, \dots, \lambda_n)$ sono parametri qualsiasi:

$$(\lambda, x) = \sum \lambda_i x_i$$

Allora si ha

$$\mathbb{E} [e^{i\xi(\lambda, x)}] = e^{m(\lambda) - \frac{1}{2}\xi^2 v(\lambda)}$$

dove abbiamo posto

$$m(\lambda) = \mathbb{E} \left[\sum_i \lambda_i x_i \right] \quad \text{e} \quad v(\lambda) = \mathbb{E} \left[\left(\sum_i \lambda_i x_i - m(\lambda) \right)^2 \right]$$

Ovviamente $m(\lambda) = \sum m_i \lambda_i$ ($m_i = \mathbb{E}[X_i]$) è lineare, e $v(\lambda)$ è una forma quadratica semidefinita positiva sullo spazio \mathbb{R}^n dei parametri λ .

Ora: se $\det(v) > 0$ allora la distribuzione congiunta di (X_1, \dots, X_n) è una misura gaussiana (basterà guardare alla FC). Se $\det(v) = 0$ il nucleo di v è lo spazio dei $\lambda \in \mathbb{R}^n$ per i quali la FC è costante, cioè (cfr. la dimostrazione del teorema precedente) (λ, X) è costante q.o. e $m(\lambda) = (\lambda, X)$. Quindi

Corollario 5.7 Se $X : \Omega \rightarrow \mathbb{R}^n$ è un sistema gaussiano allora esistono un vettore $m \in \mathbb{R}^n$ ed un sottospazio $V \subset \mathbb{R}^n$ tali che

- (1) l'immagine di X è uguale a $m + V$ q.o.
- (2) La distribuzione di X è a supporto in $m + V$ ed è ivi gaussiana.

È possibile dare una descrizione più astratta dei processi gaussiani: consideriamo una VA gaussiana di media nulla; ciò non lede la generalità, dato che l'immagine di una VA gaussiana in \mathbb{R}^n è q.o. del tipo $m + V$, ove m è una costante (la media) e $V \subset \mathbb{R}^n$ è un sottospazio, e la distribuzione delle VA ha supporto in $m + V$, ed è ivi gaussiana, il che vuol dire che, a meno di traslazioni, può assumersi che $m = 0$.

Ora, in generale, una VA gaussiana X di media zero ha CF

$$\widehat{X}(\xi) = e^{-\frac{1}{2}a\xi^2}$$

col che

$$d\mu_X(x) = \begin{cases} \frac{1}{2\pi a} e^{-\frac{1}{2a}x^2} dx & \text{se } a \neq 0 \\ \delta_x & \text{se } a = 0 \end{cases}$$

In \mathbb{R}^n abbiamo

$$d\mu_X(x) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{\det A}} e^{-\frac{1}{2}\langle A^{-1}x, x \rangle} dx$$

dove $A = ((a_{ij}))$ è la matrice di covarianza.

Viceversa, una matrice $n \times n$ simmetrica A è la matrice di covarianza di una VA gaussiana se e solo se è semidefinita positiva. Infatti, in tal caso, la

$$\varphi(\xi_1, \dots, \xi_n) = e^{-\frac{1}{2}\langle A\xi, \xi \rangle}$$

definisce una funzione positiva e quindi, per il teorema di Bochner, una misura di probabilità.

6 Teorema di Kolmogorov e processi stocastici

L'esistenza della misura prodotto (che non abbiamo discusso, ma che si trova in ogni testo di teoria della misura, cfr. [8], [10], [1]) si può generalizzare al caso del prodotto di una qualsiasi famiglia di spazi di misura e consente di dedurre, a partire da una famiglia di distribuzioni, l'esistenza di una famiglia di VA *indipendenti* le cui distribuzioni siano quelle date. Tuttavia quello che spesso si richiede, specie nella teoria dei processi stocastici, è un risultato analogo nel caso di distribuzioni che non siano indipendenti fra loro, ed il teorema di Kolmogorov soddisfa a questo requisito.

Se $\{(\Omega_t, \mathcal{A}_t)\}_{t \in T}$ sono spazi misurabili e se $\mathcal{P} = \{P_{t_1, \dots, t_n} \mid n \in \mathbb{N}^+, t_i \in T\}$ è una famiglia di probabilità, dove P_{t_1, \dots, t_n} è una probabilità sullo spazio misurabile prodotto $(\Omega_{t_1} \times \dots \times \Omega_{t_n}, \mathcal{A}_{t_1} \times \dots \times \mathcal{A}_{t_n})$, vogliamo costruire uno spazio di probabilità (Ω, \mathcal{A}, P) ed una famiglia $\{X_t\}$ di enti aleatori su $(\Omega_t, \mathcal{A}_t)$ tali che, per ogni $n \in \mathbb{N}^+$, la n -pla $(X_{t_1}, \dots, X_{t_n})$ abbia come distribuzione la misura P_{t_1, \dots, t_n} su $(\Omega_{t_1} \times \dots \times \Omega_{t_n}, \mathcal{A}_{t_1} \times \dots \times \mathcal{A}_{t_n})$.

La famiglia \mathcal{P} non può essere arbitraria: infatti, supponendo di aver risolto il problema, cioè di avere

$$P_{t_1, \dots, t_n} \left(\prod_{j=1}^n A_{t_j} \right) = P \left(\bigcap_{j=1}^n \{X_{t_j} \in A_{t_j}\} \right)$$

allora deve aversi

$$P_{t_1, \dots, t_{n+m}} \left(\prod_{j=1}^{n+m} A_{t_j} \right) = P_{t_1, \dots, t_n} \left(\prod_{j=1}^n A_{t_j} \right) \quad (C_1)$$

(se per $n < j \leq n+m$ si ha $A_{t_j} = \Omega_{t_j}$). Inoltre, per ogni permutazione $\sigma \in S_n$ degli indici $(1, \dots, n)$ si deve pure avere

$$P_{t_1, \dots, t_n} \left(\prod_{j=1}^n A_{t_j} \right) = P_{t_{\sigma(1)}, \dots, t_{\sigma(n)}} \left(\prod_{j=1}^n A_{t_{\sigma(j)}} \right) \quad (C_2)$$

Le relazioni (C_1) e (C_2) si dicono *condizioni di compatibilità* della famiglia \mathcal{P} : il teorema di Kolmogorov afferma in sostanza che queste condizioni sono anche sufficienti per l'esistenza dello spazio di probabilità (Ω, \mathcal{A}, P) e della famiglia $\{X_t\}$, a patto che gli spazi misurabili $(\Omega_t, \mathcal{A}_t)$ soddisfino a certe condizioni di regolarità, che, per esempio nel caso degli spazi metrici completi con le σ -algebre dei boreliani, sono spesso verificate.

Teorema di Kolmogorov 6.1 *Se $\{(\Omega_t, \mathcal{A}_t)\}_{t \in T}$ è una famiglia di spazi misurabili e se $\{P_{t_1, \dots, t_n} \mid n \in \mathbb{N}^+, t_i \in T\}$ è una famiglia di probabilità sugli spazi misurabili $\{(\prod_{j=1}^n \Omega_{t_j}, \prod_{j=1}^n \mathcal{A}_{t_j})\}_{j \in \mathbb{N}^+}$ e se sono soddisfatte le seguenti condizioni:*

$$(1) \quad \forall n, m \in \mathbb{N}^+ \quad \forall \{t_j\} \subset T \quad \forall A_{t_1} \in \mathcal{A}_{t_1}, \dots, A_{t_n} \in \mathcal{A}_{t_n}$$

$$P_{t_1, \dots, t_{n+m}} \left(\prod_{j=1}^n A_{t_j} \times \prod_{j=n+1}^{n+m} \Omega_{t_j} \right) = P_{t_1, \dots, t_n} \left(\prod_{j=1}^n A_{t_j} \right)$$

$$(2) \quad \forall n \in \mathbb{N}^+ \quad \forall \{t_j\} \subset T \quad \forall A_{t_1} \in \mathcal{A}_{t_1}, \dots, A_{t_n} \in \mathcal{A}_{t_n} \quad \forall \sigma \in S_n$$

$$P_{t_1, \dots, t_n} \left(\prod_{j=1}^n A_{t_j} \right) = P_{t_{\sigma(1)}, \dots, t_{\sigma(n)}} \left(\prod_{j=1}^n A_{t_{\sigma(j)}} \right)$$

(S_n è il gruppo delle permutazioni su n oggetti.)

$$(3) \quad \forall t \in T \quad \exists \mathcal{C}_t \subset \mathcal{A}_t \text{ classe compatta}^{16} \quad \forall n \in \mathbb{N}^+ \quad \forall \{t_j\} \subset T \quad \forall A_{t_1, \dots, t_n} \in \prod_{j=1}^n \Omega_{t_j}$$

$$P_{t_1, \dots, t_n}(A_{t_1, \dots, t_n}) = \sup_{K \in \prod_{j=1}^n \mathcal{C}_{t_j}} P_{t_1, \dots, t_n}(K)$$

Allora esiste uno spazio di probabilità (Ω, \mathcal{A}, P) ed una famiglia di enti aleatori $\{X_t\}_{t \in T}$ tali che P_{t_1, \dots, t_n} sia la distribuzione di $(X_{t_1}, \dots, X_{t_n})$.

¹⁶Cfr. definizione 2.5.

Per una dimostrazione elementare si veda [3, §II.2], mentre una dimostrazione elegante che usa un po' di analisi funzionale è in [12, §1.2]: qui ci limitiamo a qualche commento.

Per cominciare, si noti che le ipotesi (1) e (2) sono condizioni di compatibilità della famiglia $\{P_{t_1, \dots, t_n}\}$, mentre (3) è una condizione di regolarità degli spazi $\{(\Omega_t, \mathcal{A}_t)\}$: per esempio se questi spazi sono spazi metrici completi, le classi \mathcal{C}_t sono esattamente quelle formate dai compatti, e quindi la (3) si riduce alla regolarità degli spazi di probabilità $\left(\prod_{j=1}^n \Omega_{t_j}, \prod_{j=1}^n \mathcal{A}_{t_j}, P_{t_1, \dots, t_n}\right)$ ove \mathcal{A}_t è la σ -algebra dei boreliani di Ω_t .

Osserviamo poi che il teorema di Kolmogorov fornisce un risultato di *unicità* nel senso seguente: se $(\Omega', \mathcal{A}', P')$ e $\{X'_t\}$ sono tali che P_{t_1, \dots, t_n} sia la distribuzione di $(X'_{t_1}, \dots, X'_{t_n})$ allora, se \mathcal{A} (risp. \mathcal{A}') è la più piccola σ -algebra rispetto alla quale $\{X_t\}$ (risp. $\{X'_t\}$) sono misurabili, esiste un isomorfismo di spazi di probabilità fra (Ω, \mathcal{A}, P) e $(\Omega', \mathcal{A}', P')$ che porta le X_t nelle X'_t .

Ecco una tipica applicazione del teorema di Kolmogorov:

Teorema 6.2 *Se \mathcal{H} è uno spazio di Hilbert reale separabile allora esiste uno spazio di probabilità (Ω, \mathcal{A}, P) e, per ogni $v \in \mathcal{H}$, una VA X_v in modo che la funzione $v \mapsto X_v$ sia lineare e, per ogni $v_1, \dots, v_n \in \mathcal{H}$, $(X_{v_1}, \dots, X_{v_n})$ sia una VA gaussiana di covarianza $(\langle v_i, v_j \rangle)$. Questo spazio di probabilità è unico nel senso del teorema di Kolmogorov.*

DIMOSTRAZIONE: Sia $\{e_n\}$ una sistema ortonormale¹⁷ su \mathcal{H} e poniamo, per ogni successione $\{\xi_n\}$ definitivamente nulla,

$$\varphi(\xi) = e^{-\frac{1}{2} \sum_i \xi_i^2}$$

Dato che la forma quadratica

$$e^{-\frac{1}{2} \sum_{i,j} \langle e_i, e_j \rangle \xi_i \xi_j}$$

è definita positiva (per l'ortonormalità del sistema prescelto), per il teorema di Kolmogorov, esistono (Ω, \mathcal{A}, P) e $\{X_n\}$ sistema di variabili gaussiane¹⁸ di covarianza $\delta_{ij} = \langle e_i, e_j \rangle$.

¹⁷Per le nozioni relative agli spazi di Hilbert si può vedere: [8], [10], [1].

¹⁸Basta in realtà considerare $\Omega = \mathbb{R}^\infty$, $X_i = x_i$ e $P = \bigotimes_{n=1}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_n^2} dx_n$.

Ora, data una combinazione lineare $v = a_1 e_1 + \cdots + a_n e_n$, poniamo

$$X_v = \sum a_i X_i$$

in modo che

$$\int |X_v|^2 dP = \|v\|^2$$

Per continuità, X si estende ad una mappa $\mathcal{H} \rightarrow L^2(\Omega, dP)$ in modo che

$$\int e^{iX_v} dP = e^{-\frac{1}{2}\|v\|^2}$$

QED

La potenza di questo teorema sta nel fatto che ci fornisce una “unica” famiglia di VA gaussiane in un dato spazio di probabilità parametrizzata dagli elementi di uno spazio di Hilbert.

Il caso più elementare è $\mathcal{H} = \mathbb{R}$: in questo caso la famiglia si dice un processo:

Definizione 6.3 *Un processo stocastico è una famiglia $\{X_t\}_{t \geq 0}$ di VA $X_t : (\Omega, \mathcal{A}, P) \rightarrow (\Psi, \mathcal{P})$ a valori in uno spazio misurabile (Ψ, \mathcal{P}) e parametrizzata dai numeri reali non negativi.*

Spesso ci si limita a processi parametrizzati da un intervallo $[0, T]$ di numeri reali. In altre parole, un processo è una funzione $X : \mathbb{R}^+ \times \Omega \rightarrow \Psi$ tale che, per ogni $t \in \mathbb{R}^+$, le funzioni $\omega \mapsto X_t(\omega)$ sono misurabili. Fissato $\omega \in \Omega$, la funzione $t \mapsto X_t(\omega)$ si dice *traiettoria* per ω del processo.

Esattamente come nel caso delle martingale, possiamo associare una filtrazione ad un processo stocastico: precisamente definiamo, per ogni $t \geq 0$, \mathbb{B}_t come (il completamento de) la σ -algebra generata dagli insiemi $X_s^{-1}(A)$ dove $A \in \Psi$ e $s \leq t$. In questo modo $s \leq t$ implica $\mathbb{B}_s \subset \mathbb{B}_t$ e ovviamente le X_t sono misurabili rispetto a \mathbb{B}_t .

Come si vede bene, il teorema di Kolmogorov consente di definire in modo unico un processo stocastico la cui distribuzione (delle sue VA) sia assegnata. Tuttavia, la forma “debole” di unicità è responsabile della possibile esistenza di più “versioni” di uno stesso processo, cioè di processi equivalenti ma con traiettorie dalle caratteristiche distinte: per esempio, se (Ω, \mathcal{A}, P) è $[0, 1]$ con la misura di Lebesgue sui boreliani, il processo $X_t = 1$ e il processo

$$Y_t(x) = \begin{cases} 1 & \text{se } t - x \notin \mathbb{Q} \\ 2 & \text{se } t - x \in \mathbb{Q} \end{cases}$$

$(t, x \in [0, 1])$ sono identici dal punto di vista del teorema di Kolmogorov: infatti, per ogni $t \in [0, 1]$, X e Y sono uguali q.o. Eppure, fissato x , la mappa $t \mapsto X_t$ è continua, mentre $t \mapsto Y_t(x)$ non lo è.

Ciò mostra che se $\{X_t\}_{t \in [0,1]}$ è una famiglia di VA su (Ω, \mathcal{A}, P) e se \mathcal{A} è la σ -algebra minimale allora l'insieme $\{x \mid t \mapsto X_t(x)\}$ può non essere misurabile!

Questo problema emergerà nel caso del processo di Wiener, e lo aggireremo mostrando che

$$(*) \quad |X_t - X_s| \leq C_x |t - s|^\alpha$$

per qualche fissato α e per ogni $t, s \in [0, 1] \cap \mathbb{Q}$: allora, dato che solo infinità numerabili entrano in gioco, la (*) vale indipendentemente dalla versione del processo.

Infatti, scelto un esplicito Ω per il processo, si consideri X_t per $t \in \mathbb{Q}$ e sia Ω_0 l'insieme degli $\omega \in \Omega$ per i quali la (*) è verificata: se $\omega \in \Omega_0$, $X_t(\omega)$ definita per $t \in \mathbb{Q}$ si estende ad un'unica funzione continua $\tilde{X}_t(\omega)$ per ogni $t \in \mathbb{R}$ e si possono definire le \tilde{X}_t su Ω_0 ed estenderle a zero su Ω/Ω_0 . Ovviamente la mappa $t \mapsto \tilde{X}_t(\omega)$ è ora continua per ogni $\omega \in \Omega$, e la distribuzione congiunta di $(X_{t_1}, \dots, X_{t_n})$ è la stessa di $(\tilde{X}_{t_1}, \dots, \tilde{X}_{t_n})$. Ciò è ovvio per $t \in \mathbb{Q}$ e si estende ad ogni t per continuità delle \tilde{X} e L^2 -continuità della X .

Abbiamo dunque, data la (*), una versione continua del nostro processo.

7 Processo di Wiener

Intuitivamente, il processo di Wiener W_t è il limite di cammini aleatori elementari, nel senso seguente¹⁹: consideriamo la misura ν su $\{-1, 1\}$

$$\nu(1) = \nu(-1) = \frac{1}{2}$$

e la misura $\mu = \otimes_{n \in \mathbb{N}} \nu_n$ su $\Omega = \{-1, 1\}^{\mathbb{N}}$ dove ciascuna ν_n è ν . Se $p_n : \Omega \rightarrow \{-1, 1\}$ è la funzione di proiezione sull' n -esima componente e

$$X_n = \sum_{i=1}^n p_i$$

allora la famiglia di VA $\{X_n\}$ si dice *cammino aleatorio*. Dato che i punti di Ω sono successioni di -1 e 1 , abbiamo che

$$\mathbb{E}[p_n p_m] = \int p_n p_m d\mu = \int \delta_{nk} \delta_{km} d\nu_k = \delta_{nk} \delta_{km} = \delta_{nm}$$

¹⁹Il lettore può immaginare di spostarsi a caso su una retta: ci sono quindi due valori possibili per lo spostamento, destra e sinistra, ad ogni istante, supponendo che ogni volta ci si sposti lungo una stessa distanza.

e pertanto

$$\mathbb{E} [X_n^2] = \mathbb{E} \left[\sum_{i,j}^{1,\dots,n} p_i p_j \right] = \sum_{i,j}^{1,\dots,n} \delta_{ij} = n$$

Possiamo ora dimostrare le che VA

$$Y_n = \frac{1}{\sqrt{n}} X_n$$

tendono ad una VA gaussiana per $n \rightarrow \infty$.

Teorema di de Moivre–Gauss–Laplace 7.1 *Se μ_n è la distribuzione di Y_n allora*

$$\text{w-lim}_{n \rightarrow \infty} \mu_n = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$

dove il simbolo w-lim denota il limite debole.

DIMOSTRAZIONE: Dalla teoria di Fourier (in particolare dalla formula di Parseval) segue che una successione $\{P_n\}$ di probabilità in \mathbb{R}^n converge debolmente a P se e solo se la successione $\{\widehat{P}_n\}$ converge puntualmente a \widehat{P} (cfr. [7, §VI.2]).

La distribuzione del movimento aleatorio p_n è

$$\frac{1}{2} (\delta_{-1} + \delta_1)$$

(dove δ_x denota la misura di Dirac concentrata in $\{x\}$) e quindi

$$\widehat{Y}_n(\xi) = \widehat{\sum_{j=1}^n \frac{1}{\sqrt{n}} p_j(\xi)} = \prod_{j=1}^n \widehat{p}_j \left(\frac{\xi}{\sqrt{n}} \right)$$

Ma

$$\begin{aligned} \widehat{p}_j(\xi) &= \int e^{-i\xi x} P \circ p_j^{-1}(dx) = \frac{1}{2} \int e^{-i\xi x} (\delta_{-1} + \delta_1)(dx) \\ &= \frac{1}{2} (e^{i\xi} + e^{-i\xi}) = \cos \xi \end{aligned}$$

e dunque la distribuzione di Y_n è $\cos(\xi/\sqrt{n})^n$:

$$\widehat{Y}_n(\xi) = \cos \left(\frac{\xi}{\sqrt{n}} \right)^n$$

Infine, osservando che

$$\begin{aligned} \lim_{n \rightarrow \infty} \cos \left(\frac{\xi}{\sqrt{n}} \right)^n &= \lim_{n \rightarrow \infty} \left(1 - 2 \frac{\xi^2}{n} + o(n^{-2}) \right)^n \\ &= \lim_{n \rightarrow \infty} \left(1 - \frac{\xi^2}{2n} \right)^n = e^{-\frac{\xi^2}{2}} \end{aligned}$$

e ricordando che $\widehat{e^{-\frac{x^2}{2}}} = e^{-\frac{\xi^2}{2}}$, troviamo la tesi del teorema.

QED

Ora, per definire il processo di Wiener, quello che vorremmo scrivere è:

$$“ W_t = \lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} X_{[nt]} ”$$

(dove $[nt]$ è la parte intera del numero $[nt]$) a significare che W_t descrive un moto aleatorio sulla retta.

Il nostro W_t deve quindi essere un processo gaussiano di varianza t perché intuitivamente

$$“ \mathbb{E} \left[\left(\frac{1}{\sqrt{n}} X_{[ny]} \right)^2 \right] = t ”$$

Inoltre, dato che $X_n - X_m$ e X_m sono indipendenti, per $m < n$:

$$\begin{aligned} \mathbb{E} [(X_n - X_m) X_m] &= \mathbb{E} [X_n X_m] - m = \mathbb{E} [Y_1 X_m] + \dots + \mathbb{E} [Y_n X_m] - m \\ &= \mathbb{E} \left[Y_1 \sum_{i=1}^m Y_i \right] + \dots + \mathbb{E} \left[Y_n \sum_{i=1}^m Y_i \right] - m \\ &= \underbrace{1 + \dots + 1}_{m \text{ volte}} + \underbrace{0 + \dots + 0}_{n-m \text{ volte}} - m = 0 \end{aligned}$$

pertanto ci aspettiamo che $W_t - W_s$ sia indipendente da W_s se $s < t$, cioè che $\mathbb{E} [(W_t - W_s) W_s] = 0$ da cui, se $s < t$ allora $\mathbb{E} [W_t W_s] = s$.

Possiamo allora usare queste proprietà per definire il processo di Wiener.

Definizione 7.2 *Un processo di Wiener (o moto browniano) è una famiglia $\{W_t\}_{t \geq 0}$ di VA gaussiane con media zero e covarianza*

$$\mathbb{E} [W_t W_s] = \min(s, t)$$

La teoria delle VA gaussiane ci assicura che questa definizione è ben posta: naturalmente la covarianza deve essere una funzione positiva, dato che se $0 \leq s_1 \leq s_2 \leq \dots \leq s_n$ e $z_1, \dots, z_n \in \mathbb{C}$ allora

$$\sum_{i,j}^{1,\dots,n} \bar{z}_i z_j \min(s_i, s_j) = \sum_{i=1}^n (s_i - s_{i-1}) \left| \sum_{j=1}^n z_j \right|^2 \geq 0$$

Osserviamo inoltre che

$$\mathbb{E} [W_t^2] = t$$

e che per ogni $t_1 \leq t_2$ e $s_1 \leq s_2$ abbiamo che

$$\mathbb{E} [(W_{t_2} - W_{t_1})(W_{s_2} - W_{s_1})] = \lambda([t_1, t_2] \cap [s_1, s_2])$$

dove λ denota la misura di Lebesgue in \mathbb{R} . In particolare

$$\mathbb{E} [(W_t - W_s)W_s] = \lambda([s, t] \cap [0, s]) = 0$$

se $s < t$.

Proposizione 7.3

- (1) *Il processo di Wiener è ad accrescimenti indipendenti.*
- (2) *W_t e $W_t^2 - t$ sono martingale.*
- (3) *Se $a \in \mathbb{R}^+$ e $\lambda \neq 0$ allora i processi $W_{t+a} - W_a$ e $\frac{1}{\lambda}W_{\lambda^2 t}$ sono processi di Wiener.*

DIMOSTRAZIONE: La (1) è stata appena dimostrata. Per vedere la (2) basta notare che $W_t = W_s + (W_t - W_s)$ e, dato che $W_t - W_s$ è indipendente da $\{W_u\}_{0 \leq u < s}$, si ha, per ogni $t > s$:

$$\mathbb{E} [W_t | \{W_u\}_{0 \leq u < s}] = W_s$$

Infine la (3): il processo $W_{t+a} - W_a$ è gaussiano di media

$$\mathbb{E} [W_{t+a} - W_a] = \mathbb{E} [W_{t+a}] - \mathbb{E} [W_a] = 0$$

e di covarianza

$$\begin{aligned} & \mathbb{E} [(W_{t+a} - W_a)(W_{s+a} - W_a)] \\ &= \mathbb{E} [W_{t+a}W_{s+a}] - \mathbb{E} [W_{t+a}W_a] - \mathbb{E} [W_aW_{s+a}] + \mathbb{E} [W_aW_a] \\ &= \min(t+a, s+a) - \min(t+a, a) - \min(s+a, a) + a \\ &= \min(t+a, s+a) - 2a - a = \min(t, s) \end{aligned}$$

Analogamente

$$\mathbb{E} \left[\frac{1}{\lambda} W_{\lambda^2 t} \right] = \frac{1}{\lambda} \mathbb{E} [W_{\lambda^2 t}]$$

e

$$\mathbb{E} \left[\frac{1}{\lambda^2} W_{\lambda^2 t} W_{\lambda^2 s} \right] = \frac{\min(\lambda^2 t, \lambda^2 s)}{\lambda^2} = \min(s, t)$$

QED

Notiamo in particolare che la (3) implica essere W_{-t} un processo di Wiener.

Vogliamo ora studiare la natura delle traiettorie del processo di Wiener, in particolare se queste possono essere assunte continue, derivabili, etc.

Due processi stocastici X_t e Y_t si dicono *equivalenti* se ovviamente si ha $X_t = Y_t$ q.o: evidentemente in questo caso hanno la stessa distribuzione. Dal punto di vista delle traiettorie possono invece differire in modo essenziale, come mostra il seguente teorema.

Teorema di regolarità di Kolmogorov 7.4 *Se (Ω, \mathcal{A}, P) è uno spazio di probabilità e $\{X_t\}_{t \in [0, T]}$ è un processo reale per il quale esistano $C, \alpha, \beta > 0$ in modo che, per ogni s, t si abbia*

$$\mathbb{E} [|X_s - X_t|^\beta] \leq C|s - t|^{1+\alpha}$$

Allora su Ω esiste un processo $\{Y_t\}_{t \in [0, T]}$ equivalente a X_t ma tale che:

- (1) $P(Y_t \in \{\text{funzioni continue in } t\}) = 1.$
- (2) $\forall \gamma \in (0, \alpha/\beta)$ Y_t è γ -hölderiana q.o.

DIMOSTRAZIONE: L'idea della dimostrazione è la seguente: una funzione continua è completamente determinata dai valori che assume su un sottoinsieme denso, che nel nostro caso sarà

$$\mathbb{D} = \left\{ \frac{m}{2^n} \right\}_{n, m \in \mathbb{N}}$$

Sia dunque $\mathbb{D}_T = \mathbb{D} \cap [0, T]$: dare una funzione hölderiana in $C^\gamma[0, T]$ equivale a darne una su $C^\gamma(\mathbb{D}_T)$ e quindi basterà dimostrare che X_t è hölderiano su \mathbb{D}_T per avere un $H \in \mathcal{A}$ tale che $P(H) = 1$ e $X_t(\omega) \in C^\gamma(\mathbb{D}_T)$ per ogni $\omega \in H$. A questo punto, per ogni $\omega \in H$ potremo (per ogni $t \in [0, T]$) estendere X_t univocamente su H ottenendo in questo modo Y_t : al di fuori di H , Y_t si può definire in modo arbitrario (per esempio ponendolo a zero), dato che $P(\Omega \setminus H) = 0$.

Vediamo in dettaglio questo procedimento. Siano

$$\mathbb{D}_n = \left\{ \frac{k}{2^n} \mid k \in \mathbb{N}, \frac{k}{2^n} \in [0, T] \right\}$$

in modo che $\bigcup_{n \in \mathbb{N}} \mathbb{D}_n = \mathbb{D}_T$. Dato che per il momento siamo interessati al processo in quanto funzione di T piuttosto che di ω , scriviamolo come $X_\omega(t)$ o anche solo $X(t)$.

Se $q > 0$, un modo per controllare l'oscillazione di $X(t)$ sui punti di \mathbb{D}_T è di porre

$$A_q = \left\{ \omega \in \Omega \mid \exists N \forall n \geq N \forall k \quad \left| X_\omega \left(\frac{k+1}{2^n} \right) - X_\omega \left(\frac{k}{2^n} \right) \right| < q^n \right\}$$

Assumiamo per il momento il seguente lemma, la cui dimostrazione sarà data dopo quella del teorema per non interrompere il flusso di pensiero:

Lemma.

- (1) $A_q \in \mathcal{A}$ e, se $q > 2^{-\alpha/\beta}$, allora $P(A_q) = 1$.
 (2) $\forall \omega \in A_q \exists N \forall m \geq n \geq N \forall s, t \in \mathbb{D}_m$:

$$|t - s| < \frac{1}{2^n} \implies |X(t) - X(s)| < 2 \sum_{i=n+1}^m q^i$$

La seconda parte del lemma ci consente di controllare l'oscillazione di $X(t)$: in effetti, se $s, t \in \mathbb{D}_T$ sono tali che $|s - t| < 1/2^N$ e se $n \geq N$ è tale che

$$\frac{1}{2^{n+1}} \leq |t - s| < \frac{1}{2^n}$$

allora esiste m in modo che $s, t \in \mathbb{D}_m$ e possiamo applicare la (2) del lemma.

Se poi $q < 1$ allora

$$|X(t) - X(s)| < 2 \sum_{i=n+1}^m q^i < 2 \sum_{i=n+1}^{\infty} q^i = 2 \frac{q^{n+1}}{1 - q}$$

A questo punto ci siamo quasi: vogliamo solo determinare un $q < 1$ in modo che la stima precedente divenga del tipo $|X(t) - X(s)| < C|t - s|^\gamma$.

Per definizione di n abbiamo che $|t - s| \geq 1/2^{n+1}$ sicché

$$|t - s|^\gamma \geq \left(\frac{1}{2^{n+1}} \right)^\gamma$$

e quindi basta fissare γ tale che $1/(2^{n+1})^\gamma = q^{n+1}$, cioè $\gamma = -\log_2 q$, da cui

$$|X(t) - X(s)| < 2 \frac{2^{n+1}}{1 - q} = \frac{2}{q - 1} \left(\frac{1}{2^{n+1}} \right)^\gamma < \frac{2}{q - 1} |t - s|^\gamma$$

Detto in altri termini, fissiamo un q tale che $2^{-\alpha/\beta} < q < 1$, poniamo $\gamma = -\log_2 q$ e consideriamo l'insieme

$$B = \left\{ \omega \in \Omega \mid \exists N \forall s, t \in \mathbb{D}_T \quad |s - t| < \frac{1}{2^N} \right\}$$

Per ogni $\omega \in B$ abbiamo allora che

$$|X(t) - X(s)| \leq C|s - t|^\gamma$$

il che vuol dire $A_q \subset B$ e quindi, a meno di completare la σ -algebra \mathcal{A} , che $P(B) = 1$ e $B \in \mathcal{A}$.

Questo ci dà l'hölderianità su B e per i punti in \mathbb{D}_T tali che $|t-s| < 1/2^N$: quest'ultima restrizione si fa cadere facilmente, dato che per ogni $s, t \in [0, T]$ esiste sempre un $n \in \mathbb{N}$ tale che

$$n \leq T2^N$$

in modo che ci sono n punti $t_1 = s < t_2 < \dots < t_n = t$ tali che

$$|t_{i+1} - t_i| < \frac{1}{2^N}$$

sicché

$$|X(t) - X(s)| \leq \sum_{i=1}^{n-1} |X(t_{i+1}) - X(t_i)| \leq Cn|t_{i+1} - t_i|^\gamma \leq C'|t-s|^\gamma$$

Dunque: abbiamo dimostrato che $X(t)$ è γ hölderiana (per $t \in \mathbb{D}_T$) sull'insieme $B \in \mathcal{A}$ che ha probabilità 1. Dato che \mathbb{D}_T è denso in $[0, T]$, possiamo estendere $X(t)$ a $[0, T]$ in modo unico per $\omega \in B$ ed estenderla a zero al di fuori di B .

Con ciò troviamo un processo Y_t che è equivalente a X_t , dato che

$$\forall \omega \in B \quad Y_t(\omega) = \lim_{\substack{s \rightarrow t \\ s \in \mathbb{D}_T}} X_s(\omega)$$

Ma per ipotesi $\mathbb{E}[|X_t - X_s|^\beta] \leq C|t-s|^{1+\alpha}$ pertanto, qualunque sia $\epsilon > 0$:

$$P(|X_t - X_s| > \epsilon) = P(|X_t - X_s|^\beta > \epsilon^\beta) \leq \frac{C}{\epsilon^\beta} |t-s|^{1+\alpha}$$

da cui se ne deduce

$$\lim_{\substack{s \rightarrow t \\ s \in \mathbb{D}_T}} P(|X_t - X_s| > \epsilon) = 0$$

il che significa, per definizione di Y_t , $P(|X_t - Y_t| > \epsilon) = 0$ e dunque abbiamo un processo equivalente a X_t ma con traiettorie continue e hölderiane q.o.

QED

La dimostrazione del teorema precedente non può dirsi completa finché non avremo dimostrato il

Lemma.

- (1) $A_q \in \mathcal{A}$ e, se $q > 2^{-\alpha/\beta}$, allora $P(A_q) = 1$.
- (2) $\forall \omega \in A_q \exists N \forall m \geq n \geq N \forall s, t \in \mathbb{D}_m$:

$$|t-s| < \frac{1}{2^n} \implies |X(t) - X(s)| < 2 \sum_{i=n+1}^m q^i$$

DIMOSTRAZIONE: (1) Ricordiamo che

$$A_q = \left\{ \omega \in \Omega \mid \exists N \forall n \geq N \forall k \quad \left| X_\omega \left(\frac{k+1}{2^n} \right) - X_\omega \left(\frac{k}{2^n} \right) \right| < q^n \right\}$$

Dunque, in particolare:

$$A_1 = \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \bigcup_{k=0}^{2^n T-1} \left\{ \omega \in \Omega \mid \left| X_\omega \left(\frac{k+1}{2^n} \right) - X_\omega \left(\frac{k}{2^n} \right) \right| < q^n \right\}$$

il che mostra come $A_q \in \mathbb{A}$ (per \mathcal{A} -misurabilità delle X). Ora vogliamo mostrare che, se $q > 2^{-\alpha/\beta}$ allora $P(\Omega \setminus A_q) = 0$. Si ha che

$$\Omega \setminus A = \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} A_q^{(n)} = \limsup_{n \rightarrow \infty} A_q^{(n)}$$

dove ovviamente si è posto

$$A_q^{(n)} = \bigcup_{k=0}^{2^n T-1} \left\{ \omega \in \Omega \mid \left| X_\omega \left(\frac{k+1}{2^n} \right) - X_\omega \left(\frac{k}{2^n} \right) \right| \geq q^n \right\}$$

In questo modo, il fatto che $P(\Omega \setminus A) = P(\limsup_n A_q^{(n)}) = 0$ segue dal lemma 1.12 di Borel–Cantelli se facciamo vedere che

$$\sum_{n=0}^{\infty} P(A_q^{(n)}) < \infty$$

Per vederlo mostriamo che la serie è maggiorata da una serie geometrica: intanto abbiamo che

$$P(A_q^{(n)}) \leq \sum_{k=0}^{2^n T-1} P \left(\omega \mid \left| X_\omega \left(\frac{k+1}{2^n} \right) - X_\omega \left(\frac{k}{2^n} \right) \right| \geq q^n \right)$$

Inoltre, l'ipotesi del teorema di regolarità di Kolmogorov (che usiamo dato che il lemma lo applichiamo nella dimostrazione di quel teorema!) implica

$$P(|X_t - X_s| \geq q^n) = P(|X_t - X_s|^\beta \geq q^{n\beta}) \leq \frac{1}{q^{n\beta}} \mathbb{E}(|X_t - X_s|^\beta) \leq \frac{C}{q^{n\beta}} |t - s|^{1+\alpha}$$

da cui troviamo la maggiorazione richiesta:

$$P(A_q^{(n)}) \leq \sum_{k=0}^{2^n T-1} \frac{C}{q^{n\beta}} \left(\frac{1}{2^n} \right)^{1+\alpha} \leq TC \left(\frac{1}{2^\alpha q^\beta} \right)^n$$

col che, non appena $2^\alpha q^\beta > 1$, abbiamo maggiorato la serie degli $P(A_q^{(n)})$ con una serie geometrica convergente.

Passiamo ora al punto (2) del lemma: per definizione di A_q , se $\omega \in A_q$ allora esiste N in modo che

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

per ogni $n \geq N$ e con $(k+1)/2^n \in \mathbb{D}_T$. Ora procediamo per induzione su $m \geq n$.

Se $m = n$ allora $s = t$ e non c'è nulla da dimostrare. Se $s', t' \in \mathbb{D}_{m-1}$ sono gli elementi di \mathbb{D}_{m-1} più vicini rispettivamente a s e a t ed interni all'intervallo $[s, t]$, allora, per ipotesi induttiva (il teorema vale per $m-1$):

$$|X(t') - X(s')| < 2 \sum_{i=n+1}^{m-1} q^i$$

Ma, per definizione di A_q , $|X(t) - X(t')| < q^m$ e $|X(s) - X(s')| < q^m$, sicché il solito argomento dei "3 epsilon" fornisce

$$\begin{aligned} |X(t') - X(s')| &\leq |X(t) - X(t')| + |X(t') - X(s')| + |X(s') - X(s)| \\ &< q^m + 2 \sum_{i=n+1}^{m-1} q^i + q^m = 2 \sum_{i=n+1}^m q^i \end{aligned}$$

e quindi il lemma è dimostrato.

QED

Naturalmente a noi interessa applicare il teorema di Kolmogorov al processo di Wiener: in questo caso troviamo che possiamo rendere il processo $\frac{1}{2}$ -hölderiano: se $s < t$ allora

$$\mathbb{E} [|W_t - W_s|^{2n}] = (2n-1)!! |t-s|^n$$

dato che W_t è un processo gaussiano di media zero e varianza $t-s$, per cui abbiamo che

$$\frac{\alpha}{\beta} = \frac{n-1}{2n}$$

cioè $\gamma \in (0, \frac{1}{2})$.

Osserviamo che questo risultato non può migliorarsi, in particolare non è possibile trovare un processo (equivalente a un processo di Wiener) le cui traiettorie siano C^1 e quindi derivabili.

Teorema di Paley–Wiener–Zygmund 7.5 *Se $\alpha > \frac{1}{2}$ allora per quasi ogni traiettoria $\omega(t)$ di W_t abbiamo che*

$$\inf_{t \in [0,1]} \limsup_{h \rightarrow 0} \frac{|\omega_{t+h} - \omega_t|}{|h|^\alpha} = \infty \quad q.o.$$

In particolare, W_t non può avere q.o. traiettorie derivabili.

DIMOSTRAZIONE: Sia $k \in \mathbb{Z}$ tale che

$$k \left(\alpha - \frac{1}{2} \right) > 1$$

e sia $\omega(t)$ una traiettoria di W_t tale che

$$\inf_{t \in [0,1]} \limsup_{h \rightarrow 0} \frac{|\omega_{t+h} - \omega_t|}{|h|^\alpha} < \infty$$

Ciò vuol dire che esistono t , C e h_0 tali che per ogni $|h| \leq h_0$ si abbia

$$|\omega(t+h) - \omega(t)| \leq C|h|^\alpha$$

Siano ora $m, i \in \mathbb{Z}$ tali che, dato $n \in \mathbb{Z}$,

$$\frac{k+1}{m} < h_0 \quad \text{e} \quad i = [tn] + 1$$

Cioè, se $n > m$, $i/n, (i+1)/n, \dots, (i+k)/n \in [t, t+h_0]$. Ma allora

$$\left| \omega \left(\frac{i+j}{n} \right) - \omega \left(\frac{i+j-1}{n} \right) \right| \leq \frac{C}{n^\alpha} (|k|^\alpha + |k+1|^\alpha)$$

per $j = 1, \dots, k$ (dato che $|\omega(s) - \omega(u)| \leq C(|s-t|^\alpha + |u-t|^\alpha)$ per la disuguaglianza triangolare) e quindi

$$\begin{aligned} & \{ \omega \mid \omega \text{ traiettoria } \alpha\text{-h\"olderiane} \} \subset \\ & \subset \bigcup_{\substack{D>1 \\ m>1}} \bigcap_{n \geq m} \bigcup_{i=0}^{n-k+1} \left\{ \omega \mid \left| \omega \left(\frac{j}{n} \right) - \omega \left(\frac{j-1}{n} \right) \right| \leq \frac{D}{n^\alpha}, j = i+1, \dots, i+k \right\} \end{aligned}$$

Dimostriamo allora che l'insieme

$$\bigcap_{n \geq m} \bigcup_{i=0}^{n-k+1} \left\{ \omega \mid \left| \omega \left(\frac{j}{n} \right) - \omega \left(\frac{j-1}{n} \right) \right| \leq \frac{D}{n^\alpha}, j = i+1, \dots, i+k \right\}$$

ha misura nulla: ciò segue immediatamente dalla relazione

$$\lim_{n \rightarrow \infty} n \left(\mathbb{E} \left[\omega \left| \omega \left(\frac{1}{n} \right) \right| \leq \frac{d}{n^\alpha} \right] \right)^k = 0$$

Infatti $\omega(j/n) - \omega((j-1)/n)$ sono indipendenti da $\omega(1/n)$ ma hanno la stessa distribuzione.

Vediamo dunque perché questo limite è zero: dato che la distribuzione di $\omega(1)$ è continua, a meno di fattori di scala abbiamo che

$$\mathbb{E} \left[\omega \left| \omega \left(\frac{1}{n} \right) \right| \leq \frac{D}{n^\alpha} \right] = \mathbb{E} \left[\omega \mid |\omega(1)| \leq \frac{D}{n^{\alpha - \frac{1}{2}}} \right] = o(n^{-\alpha + \frac{1}{2}})$$

A questo punto il limite risulta essere nullo in quanto $k(\alpha - \frac{1}{2}) > 1$.

QED

8 Misura di Wiener

Il processo di Wiener è una descrizione astratta del moto browniano, che governa molti processi fisici ed economici, a partire dai movimenti delle particelle in un gas: quest'ultimo caso è stato formulato matematicamente da Einstein nel 1905²⁰.

Consideriamo il caso di dimensione uno, cioè la proiezione del moto browniano su una retta. La densità delle particelle per unità di lunghezza all'istante t sarà denotata $u(t, x)$ e supporremo il movimento uniforme, cioè che, se $\varphi(\tau, y)$ denota la proporzione delle particelle mosse da x a $x+Y$ nell'intervallo di tempo τ , allora

$$u(t + \tau, x) dx = \left(\int_{\mathbb{R}} u(t, x - y) \varphi(\tau, y) dy \right) dx$$

Una ipotesi naturale è che φ sia simmetrica rispetto all'origine spaziale O e che

$$\int_{\mathbb{R}} y^\varphi(\tau, y) dy = D\tau$$

²⁰Una descrizione matematica del moto browniano è stata anche formulata da Bachelier nella sua tesi del 1901 per fornire un modello per l'andamento delle azioni in borsa: settant'anni dopo gli economisti si accorsero della cosa, e nel 1997 Black e Scholes presero in Nobel per l'economia per aver derivato una formula sottesa dalla teoria di Bachelier–Einstein.

dove D è una costante (coefficiente di diffusione). Se espandiamo $u(t + \tau, x)$ in serie di Taylor, troviamo

$$\begin{aligned} u(t, x) + \tau \frac{\partial}{\partial t} u(x, t) + o(\tau) \\ = \int_{\mathbb{R}} \left(u(t, x) - y \frac{\partial}{\partial x} u(t, x) + \frac{1}{2} y^2 \frac{\partial^2}{\partial x^2} u(x, t) + \dots \right) \varphi(\tau, y) dy \end{aligned}$$

che, per l'ipotesi di simmetricità di φ , ci conduce all'equazione di diffusione del calore

$$\frac{\partial u}{\partial t} = \frac{1}{2} D \frac{\partial^2 u}{\partial x^2}$$

Consideriamo la condizione iniziale per cui la posizione originaria di una particella sia y , vale a dire

$$u(0, t) = \delta(x - y)$$

ed otteniamo la soluzione

$$u(t, 0) = \frac{1}{\sqrt{2\pi Dt}} e^{-\frac{(x-y)^2}{2Dt}}$$

Per i valori $y = 0$ e $D = 1/2$, la probabilità che $x(t_1) \in (a_1, b_1), \dots, x(t_n) \in (a_n, b_n)$ è, per $0 < t_1 < \dots < t_n$:

$$\frac{1}{\sqrt{\pi^n t_1 (t_2 - t_1) \dots (t_n - t_{n-1})}} \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} e^{-\frac{x_1^2}{t_1} - \frac{(x_2 - x_1)^2}{t_2 - t_1} - \dots - \frac{(x_n - x_{n-1})^2}{t_n - t_{n-1}}} dx_1 \dots dx_n$$

che è uguale a

$$\int_{\prod_{i=1}^n (a_i, b_i)} u(t_1, x_1) u(t_2 - t_1, x_2 - x_1) \dots u(t_n - t_{n-1}, x_n - x_{n-1}) dx$$

In termini probabilistici, questa espressione definisce una misura sui cilindri nello spazio di tutte le funzioni $x(t)$ tali che

$$a_i < x(t) < b_i$$

Ma allora, per il teorema di Kolmogorov, esiste un'unica misura che la estende a tutto lo spazio $C_0[0, \infty)$ delle funzioni $f : [0, \infty) \rightarrow \mathbb{R}$ continue e tali che $f(0) = 0$. Questa misura è la *misura di Wiener*.

Per capire qual'è il legame fra questa misura ed il processo di Wiener W_t , dobbiamo considerare il legame fra la funzione

$$u(t, x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}}$$

ed il processo di Wiener, e questo legame è dato dal semigruppò di operatori e^{-tH_0} dove

$$H_0 = -\frac{1}{2}\Delta = -\frac{1}{2}\sum_{i=1}^n \frac{\partial^2}{\partial x_i^2}$$

è l'hamiltoniana di una particella libera in uno spazio a n dimensioni. Facciamo quindi una digressione su questo operatore differenziale.

Se $\alpha \in \mathbb{C}$ e $\Re\alpha > 0$ allora la funzione $e^{-\lambda^2\alpha} \in L^2(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$ è essenzialmente limitata e a quadrato integrabile, ed è l'antitrasformata di Fourier di una funzione gaussiana:

$$e^{-\lambda^2\alpha} = \frac{1}{(2\alpha)^{n/2}} \widehat{e^{-\frac{x^2}{4\alpha}}}$$

per cui si ha che

$$(e^{-\alpha H_0})(x) = \left(\frac{1}{4\pi\alpha}\right)^{n/2} \int e^{-\frac{|x-y|^2}{4\alpha}} \varphi(y) dy$$

Ma se $\varphi \in L^1(\mathbb{R}^2) \cap L^2(\mathbb{R}^2)$ allora, dato che

$$\lim_{\epsilon \rightarrow 0} e^{-i(t-i\epsilon)H_0} \varphi = e^{-itH_0} \varphi$$

(il limite è inteso nella norma L^2) deve esistere, per il teorema della convergenza dominata, una sottosuccessione che converge q.o.

$$\begin{aligned} (e^{-itH_0} \varphi)(x) &= \lim_{\epsilon \rightarrow 0} (e^{-i(t-i\epsilon)H_0} \varphi)(x) \\ &= \lim_{\epsilon \rightarrow 0} \left(\frac{1}{4\pi i(t-i\epsilon)}\right)^{n/2} \int e^{-\frac{|x-y|^2}{4i(t-i\epsilon)}} \varphi(y) dy \\ &= \left(\frac{1}{4\pi it}\right)^{n/2} \int e^{i\frac{|x-y|^2}{4t}} \varphi(y) dy \end{aligned}$$

Se invece φ sta solo in $L^2(\mathbb{R}^n)$ ma non in $L^1(\mathbb{R}^n)$, allora prendiamo

$$\chi_R(x) = \begin{cases} 1 & \text{se } |x| \leq R \\ 0 & \text{se } |x| > R \end{cases}$$

in modo che $\chi_R(x)\varphi \in L^1(\mathbb{R}^n)$ e $\chi_R\varphi$ converge a f in norma L^2 al tendere di $R \rightarrow \infty$: ora invochiamo il teorema di Plancherel, che afferma la convergenza delle trasformate di Fourier $\widehat{\chi_R\varphi}$ (sempre in norma L^2 per $R \rightarrow \infty$) a \widehat{f} , dove precisamente

$$\widehat{f}(x) = L^2\text{-}\lim_{R \rightarrow \infty} \left(\frac{1}{2\pi}\right)^{n/2} \int_{|x| \leq R} e^{-i\langle \xi, x \rangle} f(x) dx$$

Nel nostro caso abbiamo

$$(e^{-itH_0}\varphi)(x) = L^2\text{-}\lim_{R \rightarrow \infty} \left(\frac{1}{4\pi it}\right)^{n/2} \int e^{i\frac{|x-y|^2}{4t}} \varphi(y) dy$$

il che dimostra che e^{-itH_0} è un operatore integrale il cui nucleo è il *propagatore della particella libera*:

$$K(x, y|t) = \left(\frac{1}{4\pi it}\right)^{n/2} e^{i\frac{|x-y|^2}{4t}}$$

In altri termini, abbiamo mostrato che il nucleo dell'operatore e^{-tH_0} è proprio $u(x, t)$:

$$(e^{-tH_0}f)(x) = \int \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2t}|x-y|^2} f(y) dy$$

Ma allora, dato che $u(x-y, t)$ varia con continuità in L^2 al variare di x , la funzione

$$\int u(x-y, t) f(y) dy$$

è continua per ogni $f \in L^2$, sicché $(e^{-tH_0}f)(x)$ è definito per quasi ogni x come operatore integrale di nucleo u .

Il legame col processo di Wiener W_t è dato dal

Teorema 8.1 *Se $f_1, \dots, f_n \in L^\infty(\mathbb{R})$ ed anche $f_n \in L^2(\mathbb{R})$, e se $0 < s_1 < \dots < s_n$, allora*

$$\mathbb{E}[f_1(W(s_1)) \cdots f_n(W(s_n))] = (e^{-t_1 H_0} f_1 \cdots e^{-t_n H_0} f_n)(0)$$

con $t_1 = s_1, t_2 = s_2 - s_1, \dots, t_n = s_n - s_{n-1}$.

DIMOSTRAZIONE: Basta dimostrare che la distribuzione congiunta di $(W_{s_1}, \dots, W_{s_n})$ è (scriviamo $P_t(x, y) = u(x-y, t)$)

$$P_{t_1}(0, x_1) \cdots P_{t_n}(0, x_n) dx_1 \otimes \cdots \otimes dx_n$$

Ma, dato che $W_{s_1}, W_{s_2} - W_{s_1}, \dots, W_{s_n} - W_{s_{n-1}}$ sono indipendenti e di varianze t_1, t_2, \dots, t_n , la loro distribuzione congiunta è, in quanto variabili gaussiane,

$$P_{t_1}(0, y_1) \cdots P_{t_n}(0, y_n) dy_1 \otimes \cdots \otimes dy_n$$

da cui, per mezzo delle sostituzioni

$$\begin{cases} y_1 = x_1 \\ y_2 = x_2 - x_1 \\ \vdots \\ y_n = x_n - x_{n-1} \end{cases}$$

(che corrispondono ad una trasformazione di Jacobiano 1) otteniamo il risultato voluto.

QED

Questo teorema vale in dimensione qualsiasi: per formularlo abbiamo bisogno della

Definizione 8.2 *Il processo di Wiener N -dimensionale $W(t)$ è la famiglia di VA a valori in \mathbb{R}^N le cui componenti sono processi di Wiener indipendenti: cioè*

$$\begin{aligned} W(t) &= (W_1(t), \dots, W_n(t)) \\ \mathbb{E}[W_i(t)W_j(s)] &= \delta_{ij} \min(t, s) \end{aligned}$$

Detto ciò, il teorema precedente si estende come segue

Teorema 8.3 *Se $H_0 = -\frac{1}{2}\Delta$ su $L^2(\mathbb{R}^N)$ e se $f_1, \dots, f_n \in L^\infty(\mathbb{R}^N)$ con inoltre $f_n \in L^2(\mathbb{R}^N)$, e se $0 < s_1 < \dots < s_n$, allora*

$$\mathbb{E}[f_1 W(s_1) \cdots f_n W(s_n)] = (e^{-t_1 H_0} f_1 \cdots e^{-t_n H_0} f_n)(0)$$

con $t_1 = s_1, t_2 = s_2 - s_1, \dots, t_n = s_n - s_{n-1}$.

Questi teoremi ovviamente privilegiano lo zero: per avere l'invarianza per traslazioni (come la possiede H_0), bisogna introdurre una misura non normalizzata.

Definizione 8.4 *Se dx denota la misura di Lebesgue su \mathbb{R}^N e se (B, \mathcal{B}, dW) è lo spazio di misura per il processo di Wiener (dato dal teorema di Kolmogorov), la misura di Wiener è la misura prodotto $d\mu_0 = dx \otimes dW$ su $\mathbb{R}^N \times B$. Gli elementi sui quali si integrerà rispetto a questa misura sono i cammini $\omega(t) = x + W_t$.*

Teorema 8.5 *Se $f_0, \dots, f_n \in L^\infty(\mathbb{R}^N)$ con $f_0, f_n \in L^2(\mathbb{R}^N)$, e se $0 \leq s_0 < s_1 < \dots < s_n$, allora*

$$\int f_0(\omega(s_0)) \cdots f_n(\omega(s_n)) d\mu_0(\omega) = \langle \overline{f_0}, e^{-t_1 H_0} f_1 \cdots e^{-t_n H_0} f_n \rangle$$

con $t_1 = s_1, t_2 = s_2 - s_1, \dots, t_n = s_n - s_{n-1}$.

DIMOSTRAZIONE: Se $s_0 = 0$, posto $\omega = x + W$ il teorema precedente ci dice che

$$\begin{aligned} \int f_0(\omega(s_0)) \cdots f_n(\omega(s_n)) d\mu_0(\omega)(x + W(t)) \\ &= \int f_0(W(s_0) + x) \cdots f_n(W(s_n) + x) dW(W(t)) dx \\ &= \int \mathbb{E} [f_0(x) f_1(W(s_1) + x) \cdots f_n(W(s_n) + x)] dx \\ &= \int f_0(x) (e^{-t_1 H_0} f_1 \cdots e^{-t_n H_0} f_n)(x) dx \\ &= \langle \overline{f_0}, e^{-t_1 H_0} f_1 \cdots e^{-t_n H_0} f_n \rangle \end{aligned}$$

Se $s_0 > 0$, sia β_R la funzione che vale 1 sulla palla di raggio R in \mathbb{R}^N e zero altrove: il teorema della convergenza dominata implica allora che

$$\begin{aligned} \int f_0(\omega(s_0)) \cdots f_n(\omega(s_n)) d\mu_0(\omega) \\ &= \lim_{R \rightarrow \infty} \int \beta_R(\omega(0)) f_0(\omega(s_0)) \cdots f_n(\omega(s_n)) d\mu_0(\omega) \\ &= \lim_{R \rightarrow \infty} \langle \overline{f_0} e^{-s_0 H_0} \beta_R, e^{-t_1 H_0} f_1 \cdots e^{-t_n H_0} f_n \rangle \end{aligned}$$

e, dato che $\overline{f_0} e^{-s_0 H_0} \beta_R \rightarrow \overline{f_0}$ in L^2 per il teorema di Beppo Levi, si ha la tesi.

QED

Integrare rispetto a μ_0 vuol dire integrare sullo “spazio dei cammini” $\omega(t)$, il che ha rilevanza fondamentale nel dare senso all’espressione $d\text{path}$ che si trova nei lavori di Feynmann sui fondamenti della teoria quantistica dei campi: il processo di Wiener è quindi centrale in questa teoria gigantesca e bellissima.

Nei prossimi paragrafi ci dilungheremo sull’integrazione nello spazio dei cammini: ne daremo prima una definizione astratta, matematicamente pulita e formalmente assai elegante, per poi dare qualche idea dei calcoli che intervengono nelle applicazioni quantistiche.

9 Integrale di Wiener

Vogliamo ora provare a dare un senso ad espressioni della forma

$$\int_0^T f(t) dW_t$$

La prima idea che può venirci in mente, cioè quella di definirlo come un integrale di Stieltjes²¹: ma questo non è possibile, dato che la funzione W_t non è mai a variazione limitata²² non essendo quasi mai differenziabile. Il modo col quale Wiener aggirò questa difficoltà fu di assumere che la derivata f' stesse in $L^2[0, T]$ e di utilizzare la formula di integrazione per parti

$$\int_0^T f(t) dW_t = f(t)W_t - \int_0^T f'W_t dt$$

In questo modo ci si può ricondurre al caso delle funzioni differenziabili.

L'approccio che vogliamo qui seguire è quello di Nelson, un po' più astratto. Nello spazio $L^2[0, T]$ (che è lo spazio delle funzioni che vogliamo integrare) consideriamo le funzioni $\chi_t = \chi_{[0, T]}$, calcoliamo il prodotto scalare in L^2 di

$$\langle W_t, W_s \rangle = \mathbb{E}[W_t W_s] = \min(s, t) = \langle \chi_t, \chi_s \rangle$$

ed usiamo la seguente proprietà universale delle mappe bilineari:

Lemma 9.1 *Se I è un insieme e $F : I \times I \rightarrow \mathbb{C}$ è una funzione hermitiana positiva, cioè*

- (1) $F(i, j) = \overline{F(j, i)}$
- (2) $\sum_{i, j \in I} z_i \overline{z_j} F(i, j) \geq 0$.

per ogni scelta di $\{z_i\}_{i \in I} \subset \mathbb{C}$ quasi tutti nulli (cioè tali che solo per un numero finito di indici i si abbia $z_i \neq 0$) allora:

²¹Cioè un integrale associato a una misura di Stieltjes: si veda per esempio [8, §VI.6] o [1, §4.5].

²²Rammentiamo che una *funzione a variazione limitata* $f : [a, b] \rightarrow \mathbb{R}$ è una funzione tale che esista una costante C in modo che per ogni $a = x_0 < x_1 < \dots < x_n = b$ si abbia

$$\sum_{i=1}^n |f(x_i) - f(x_{i-1})| \leq C$$

La proprietà saliente delle funzioni a variazione limitata è di avere derivata finita q.o.

- (1) *Esiste uno spazio di Hilbert \mathcal{H}_I ed una funzione $X : I \rightarrow \mathcal{H}_I$ tali che*
- (a) \mathcal{H}_I è generato (topologicamente) da $\{X(i)\}_{i \in I}$.
 - (b) $F(i, j) = \langle X(i), X(j) \rangle$.
- (2) *Se $Y : I \rightarrow \mathcal{H}$ è una funzione a valori in uno spazio di Hilbert \mathcal{H} e tale che $F(i, j) = \langle Y(i), Y(j) \rangle$ allora esiste un'unica isometria $U : \mathcal{H}_I \rightarrow \mathcal{H}$ tale che $UX = Y$.*

DIMOSTRAZIONE: Questo teorema è un classico risultato universale di esistenza e unicità a meno di isometrie, simile per esempio al teorema di esistenza del prodotto tensoriale: la dimostrazione è del tutto naturale. Precisamente, consideriamo lo spazio vettoriale $V = \mathbb{C}^{(I)}$ delle funzioni $\varphi : I \rightarrow \mathbb{C}$ che sono non nulle al più che su un numero finito di elementi di I , e dotiamolo della forma bilineare

$$\langle \varphi, \psi \rangle_V = \sum_{i, j \in I} F(i, j) \varphi(i) \overline{\psi(j)}$$

(la somma è finita per definizione di V). Allora la funzione $X : I \rightarrow V$ definita come l'indicatore dell'insieme $\{i\}$

$$X(i) = \chi_{\{i\}}$$

è tale che

$$F(i, j) = \langle X(i), X(j) \rangle$$

La forma bilineare $\langle \cdot, \cdot \rangle$ su V è hermitiana ma potrebbe essere degenere: consideriamo quindi lo spazio quoziente $\overline{V} = V/K$ dove K è il nucleo di $\langle \cdot, \cdot \rangle$, che viene quindi ad essere uno spazio pre-hilbertiano rispetto al prodotto scalare indottovi da $\langle \cdot, \cdot \rangle$: non resta che definire \mathcal{H}_I come il completamento di \overline{V} rispetto a questo prodotto e notare che X passa al quoziente in $\overline{V} \subset \mathcal{H}_I$ (a meno di isometrie) per avere la (1).

Dimostriamo ora la (2): per ipotesi abbiamo $Y : I \rightarrow \mathcal{H}$, quindi usiamola per definire una funzione $S : V \rightarrow \mathcal{H}$ come

$$S(\varphi) = \sum_{i \in I} \varphi(i) Y(i)$$

Allora

$$\langle S(\varphi), S(\psi) \rangle = \sum_{i, j \in I} \varphi(i) \overline{\psi(j)} \langle Y(i), Y(j) \rangle = \langle \varphi, \psi \rangle$$

Questo vuol dire che se $\varphi \in K$ (il nucleo di V) allora $\|S(\varphi)\| = 0$ e quindi S passa al quoziente inducendo una funzione $\overline{S} : \overline{V} \rightarrow \mathcal{H}$ che preserva il prodotto hilbertiano. Questa mappa si estende allora in modo unico ad una isometria $\tilde{S} : \mathcal{H}_I \rightarrow \mathcal{H}$ (dato che \overline{V} è denso in \mathcal{H}_I).

QED

Notiamo che, per $I = \mathcal{H}_1 \times \mathcal{H}_2$ e $F((v_1, w_1), (v_2, w_2)) = \langle v_1, w_1 \rangle \langle w_1, w_2 \rangle$, il teorema precedente ci fornisce lo spazio di Hilbert $\mathcal{H}_1 \otimes \mathcal{H}_2$ prodotto tensoriale.

Torniamo ora a considerare il processo di Wiener W_t : le funzioni $\{\chi_t\}$ generano (topologicamente) $L^2[0, T]$ e quindi, per il lemma, esiste un'unica isometria fra $L^2[0, T]$ ed il sottospazio \mathcal{H}_T generato da $\{W_t\}_{0 \leq t \leq T}$.

Definizione 9.2 *L'isometria $L^2[0, T] \rightarrow \mathcal{H}_T$ si dice integrale di Wiener e si denota con*

$$f \mapsto \int_0^T f(t) dW_t$$

Questa terminologia è motivata per esempio dal fatto che sulle funzioni χ_t l'isometria si comporta come un integrale:

$$\sum_i a_i \chi_{(0, t_i)} \mapsto \sum_i a_i W_{t_i}$$

da cui

$$\sum_i a_i \chi_{(t_i, t_{i+1})} \mapsto \sum_i a_i (W_{t_{i+1}} - W_{t_i})$$

In questo modo abbiamo determinato l'integrale di Wiener sulle funzioni semplici e, per densità di queste in $L^2[0, T]$, su una qualsiasi funzione a quadrato integrabile, che è limite di funzioni semplici.

Teorema 9.3 *Se $f \in BV[0, T]$ (funzione a variazione limitata) allora*

$$\int_0^t f(t) dW_t = - \int_0^T d(f(f)W_t)$$

DIMOSTRAZIONE: Se f è una funzione semplice allora

$$\int_0^T \sum_i a_i (\chi_i - \chi_{i-1}) dW_t = \sum_i a_i \int_{t_{i-1}}^{t_i} dW_t = \sum_i a_i (W_{t_i} - W_{t_{i-1}})$$

In generale, una funzione in BV si può approssimare in L^2 con funzioni semplici (cfr. [8, §VI.2]) e la misura df lo è nella topologia debole. QED

Poiché quasi ogni traiettoria di W_t è continua, l'integrale di Wiener è definito sullo spazio dei cammini continui, sui quali è un funzionale: se $\omega(t)$ è una funzione continua di t , il teorema precedente si scrive per esempio come segue:

$$\left(\int_0^T f(t) dW_t \right) (\omega(t)) = \int_0^T f(t) d\omega(t) = - \int_0^T d(\omega(t)f(t))$$

Se $f \in AC$ ritroviamo la formula di Wiener:

$$\begin{aligned} \int_0^T f(t) d\omega(t) &= - \int_0^T d(\omega(t)f(t)) \\ &= f(T)\omega(T) - f(0)\omega(0) - \int_0^T \frac{df(t)}{dt} \omega(t) dt \end{aligned}$$

L'applicazione che vogliamo dare dell'integrale di Wiener è, come si è detto, alla teoria quantistica, precisamente per dare senso ai "path integral" introdotti da Feynmann, che sono alla base del suo approccio alla teoria dei campi, la cosiddetta QED (Quantum ElectroDynamics), la più precisa teoria fisica, nel senso dei riscontri sperimentali, mai prodotta...

10 Path Integrals

Nella teoria quantistica la probabilità di trovare un sistema fisico nello stato ψ al tempo t sapendo che si trovava nello stato φ al tempo 0, si calcola tramite un *propagatore* $K(x_0, x_1|t)$, mercé la formula

$$\left| \int_{\mathbb{R}} \int_{\mathbb{R}} \varphi(x_0) \overline{\psi(x)} K(x_0, x|t) dx_0 dx \right|$$

La funzione K si ottiene come soluzione dell'*equazione di Schrödinger*

$$\begin{cases} i\hbar \frac{\partial K}{\partial t} = -\frac{\hbar^2}{2} \frac{\partial^2 K}{\partial x^2} + V(x)K \\ K(x_0, x_1|0) = \delta(x - x_0) \end{cases}$$

dove $V = V(x)$ è il potenziale (stiamo scrivendo l'equazione nel caso di dimensione uno, per semplicità). Feynmann espresse la soluzione di questa equazione parabolica (che è in fin dei conti una variante dell'equazione del calore) come

$$K(x_0, x|t) = \int e^{i\hbar^{-1}S[x]} d\text{path}$$

ove l'integrale è esteso a tutti i cammini $x = x(t)$ tali che $x(0) = x_0$ e $S[x]$ è l'azione classica del sistema:

$$S = \int_0^t \left(\frac{1}{2} \left(\frac{dx}{ds} \right)^2 - V(x(s)) \right) ds$$

fornita dal principio variazionale.

L'imbarazzante notazione $d\text{path}$ viene spiegata da Feynmann nel modo seguente:

$$K(x_0, x|t) = \lim_{N \rightarrow \infty, \Delta \rightarrow 0, N\Delta \rightarrow 1} \frac{1}{(2\pi i \Delta \hbar)^{N/2}} \int_{\mathbb{R}} \cdots \\ \cdots \int_{\mathbb{R}} \exp \left[\frac{i}{\hbar} \left(\frac{(x_1 - x_0)^2}{2\Delta} + \cdots + \frac{(x - x_{N-1})^2}{2\Delta} - \Delta \sum_{k=1}^{N-1} V(x_k) \right) \right] dx_1 \cdots dx_N$$

La teoria di Wiener consente di dare un senso a questa formula.

Per cominciare c'è una analogia formale tra il limite precedente e la misura i Wiener di un cilindro:

$$\frac{1}{\sqrt{\pi^n t_1(t_2 - t_1) \cdots (t_n - t_{n-1})}} \int_{a_1}^{b_1} \cdots \\ \cdots \int_{a_n}^{b_n} \exp \left[-\frac{x_1^2}{t_1} - \frac{(x_2 - x_1)^2}{t_2 - t_1} - \cdots - \frac{(x_n - x_{n-1})^2}{t_n - t_{n-1}} \right] dx_1 \cdots dx_n$$

Questo suggerisce di esprimere $K(x_0, x|t)$ come un integrale di Wiener. Supponendo che $x_0 = 0$, consideriamo intanto l'integrale di Wiener di un funzionale $F[x(t)]$ definito sullo spazio $C_0[0, T]$ dei cammini continui fra 0 e T tali che $x(0) = 0$.

Sostituiamo $x(t)$ con la curva spezzata $x_n(t)$ che coincide con $x(t)$ nei punti

$$x(0) = 0, \quad x(t_1) = x_1, \quad \dots, \quad x(t_n) = x_n$$

ove t_1, \dots, t_{n-1} dividono $[0, T]$ in n parti uguali di lunghezza $\Delta t = t/n$. Allora

$$\int f[x(t)] dW = \lim_{n \rightarrow \infty} \frac{1}{(\pi \Delta t)^{n/2}} \int_{\mathbb{R}} \cdots \\ \cdots \int_{\mathbb{R}} F(x_1, \dots, x_n) \exp \left[-\frac{x_1^2}{\Delta t} - \sum_{k=1}^{n-1} \frac{(x_{k+1} - x_k)^2}{\Delta t} \right] dx_1 \cdots dx_n$$

dove $F(x_1, \dots, x_n) = F[x_n(t)]$. Simbolicamente scriviamo

$$\int_{C_0[0, T]} F[x(t)] dW = \frac{1}{N} \int_{\mathbb{R}^n} F[x(t)] \exp \left[-\int_0^T \left(\frac{\partial x}{\partial s} \right)^2 ds \right] \prod_{k=1}^n dx_1(t)$$

ove N è una costante di normalizzazione tale che

$$\int_{C_0[0, T]} dW = 1$$

Questa scrittura è sensata, perché se $\prod_{k=1}^n dx_1(t)$ è sostituito dai prodotti dei differenziali delle coordinate $x(s)$ in un numero finito di punti t_1, \dots, t_n , l'integrale

$$\int_0^T \left(\frac{\partial x}{\partial s} \right)^2 ds$$

diviene

$$\sum_{j=0}^n \frac{(x(t_{j+1}) - x(t_j))^2}{t_{j+1} - t_j}$$

In particolare, se $F[x(t)]$ dipende solo dal valore di $x(t)$ in un numero finito di punti, l'integrale di Wiener si riduce ad un normale integrale multiplo in \mathbb{R}^n : per esempio, se

$$F[x(t)] = x(t_1) \cdots x(t_k)$$

allora $\int x(t_1) dW = \mathbb{E}[W(t_1)] = 0$ e, in generale,

$$\int x(t_1) \cdots x(t_{2k+1}) dW = 0$$

mentre $\int x(t_1)x(t_2) dW = \mathbb{E}[W(t_1)W(t_2)] = \min(t_1, t_2)$, e, in generale,

$$\int x(t_1) \cdots x(t_{2k}) dW = \sum \min(t_{i_1}, t_{i_2}) \cdots \min(t_{i_{2k-1}}, t_{i_{2k}})$$

dove la sommatoria è estesa a tutte le partizioni dell'insieme $\{1, 2, \dots, 2k\}$ in coppie.

Un altro esempio di calcolo dell'integrale di Wiener con le formule precedenti si ha per il funzionale

$$F[x(t)] = \exp \left[\lambda \int_0^t p(s) s^2(s) ds \right]$$

dove $\lambda \in \mathbb{R}$ e $p(s) \geq 0$. In effetti,

$$F(x_1, \dots, x_n) = \exp \left[\lambda \sum_{j=1}^n p(j\Delta t) x^2(j\Delta t) \Delta t \right] = \exp \left[\lambda \Delta t \sum_{j=1}^n p_j x_j^2 \right]$$

con le posizioni:

$$\Delta t = \frac{t}{n}, \quad p_j = p(j\Delta t), \quad x_j = x(j\Delta t)$$

Ma se ricordiamo la formula

$$\int_{\mathbb{R}^n} e^{-\langle Ax, x \rangle} dx = \frac{\pi^{n/2}}{\det A}$$

Esempio 10.1 Se $p(s) = 1$ allora $D(s) = \sqrt{\cos \lambda}(t - s)$, e troviamo quindi

$$\int_C \exp \left[\lambda \int_0^t x^2(s) ds \right] dW(x) = \sqrt{\sec t \sqrt{\lambda}}$$

per $\sqrt{\lambda} < \pi/2t$.

11 Formula di Trotter

Per chiudere il cerchio che collega i path integral e la misura di Wiener occorre in qualche modo rendere conto della spiegazione di Feynmann di questo legame, cioè la “formula”

$$K(x_0, x|t) = \int e^{\frac{iS[x]}{\hbar}} d\text{path}$$

Fin qui abbiamo espresso il propagatore K come un integrale di Wiener: il tassello mancante è riempito con la *formula di Trotter*, che permette di dimostrare, cosa che fece M. Kac, la formula di Feynmann nel caso del semigruppato e^{-tH_0} . Naturalmente nella teoria quantistica l’oggetto centrale è e^{-itH_0} , per il quale non è possibile tuttavia sviluppare la teoria che stiamo dando nel caso di e^{-tH_0} : detta brutalmente, la teoria di Feynmann non ha senso a meno che, con Kac, non si sposti l’attenzione su e^{-tH_0} , facendo perdere senso alla variabile t come “tempo reale”; se si è disposti ad accettare l’idea di un tempo immaginario, la teoria diviene formalmente corretta...

L’idea è di esprimere e^{-itH} come un limite

$$(*) \quad e^{-itH} = \lim_{n \rightarrow \infty} \left(e^{-it\frac{H_0}{n}} e^{-it\frac{V}{n}} \right)^n$$

e quindi di trattare gli esponenziali come nel caso dei gruppi di Lie (cfr. [1, §15.3]). Allora, dato che il nucleo di e^{-itH_0} è

$$K(x, y|t) = \left(\frac{1}{\sqrt{2\pi it}} \right)^n e^{i\frac{|x-y|^2}{2t}}$$

il nucleo di $\left(e^{-it\frac{H_0}{n}} e^{-it\frac{V}{n}} \right)^n$ è

$$K^{(n)}(x, y|t) = \left(\frac{\sqrt{n}}{\sqrt{2\pi it}} \right)^{n^2} \int e^{iS(x_0, \dots, x_n|t)} dx_1 \cdots dx_n$$

dove abbiamo posto

$$S(x_0, \dots, x_n|t) = \sum_{k=1}^n \frac{1}{2} |x_{k-1} - x_k|^2 \left(\frac{n}{t} \right) - \sum_{k=1}^n V(x_k) \left(\frac{t}{n} \right)$$

che è l'approssimazione dell'azione

$$S(\omega) = \frac{1}{2} \int_0^T \dot{\omega}^2(s) ds - \int_0^T V(\omega(s)) ds$$

per un cammino poligonale che attraversi gli x_k al tempo kt/n . Al limite, l'espressione formale

$$K(x, y|t) = \int e^{iS(\omega)} d\omega$$

è quindi valida, per il nucleo integrale di e^{-itH} , ove con $d\omega$ intendiamo l'integrale esteso a tutti i cammini fra x e y al tempo t .

Quello che ci manca per rendere conto della formula (*), e quindi chiudere il cerchio, è il seguente teorema.

Formula di Trotter 11.1 *Se A e B sono operatori autoaggiunti²³ in uno spazio di Hilbert separabile \mathcal{H} tali che $A+B$, definito sul dominio $\mathcal{D}(A) \cap \mathcal{D}(B)$, sia autoaggiunto, allora*

$$e^{it(A+B)} = \lim_{n \rightarrow \infty} \left(e^{i\frac{t}{n}A} e^{i\frac{t}{n}B} \right)^n$$

(dove il limite è nella topologia forte). Se inoltre A e B sono limitati inferiormente, allora

$$e^{-it(A+B)} = \lim_{n \rightarrow \infty} \left(e^{-i\frac{t}{n}A} e^{-i\frac{t}{n}B} \right)^n$$

DIMOSTRAZIONE: (Nelson) Siano

$$S_t = e^{it(A+B)}, \quad V_t = e^{itA}, \quad W_t = e^{itB}, \quad U_t = V_t W_t$$

e sia, per qualche $\psi \in \mathcal{H}$, $\psi_t = S_t \psi$. Allora

$$\begin{aligned} \|(S_t - U_{t/n}^n) \psi\| &= \|(S_{t/n}^n - U_{t/n} S_{t/n}^{n-1} + U_{t/n} S_{t/n}^{n-1} - U_{t/n} S_{t/n}^{n-2} + \dots \\ &\quad \dots + U_{t/n}^{n-1} S_{t/n} - U_{t/n}^n S_{t/n}^0) \psi\| \\ &= \left\| \sum_{k=0}^{n-1} U_{t/n}^k (S_{t/n} - U_{t/n}) S_{t/n}^{k-j-1} \right\| \psi \Big\| \\ &\leq n \sup_{s \in [0, t]} \|(S_{t/n} - U_{t/n}) \psi_s\| \end{aligned}$$

²³Non necessariamente limitati, quindi densamente definiti.

Inoltre, se $\varphi \in \mathcal{D}(A) \cap \mathcal{D}(B)$,

$$\begin{aligned} \lim_{s \rightarrow 0^+} \frac{1}{s}(S_s - 1)\varphi &= i(A + B)\varphi, \\ \lim_{s \rightarrow 0^+} \frac{1}{s}(U_s - 1)\varphi &= \lim_{s \rightarrow 0^+} \left(V_s \frac{1}{s} (e^{isB} - 1) + \frac{1}{s} (V_s - 1) \right) \varphi \\ &= iV_0 B \varphi + \lim_{s \rightarrow 0^+} \frac{1}{s} (e^{isA} - 1) \varphi \\ &= iB\varphi + iA\varphi \end{aligned}$$

Ne viene quindi

$$\lim_{s \rightarrow 0^+} (n \|(S_{t/n} - U_{t/n})\varphi\|) = 0$$

che, unitamente alla disequazione scritta poco sopra, ci dice che $\|(S_t - U_{t/n}^n)\psi\| \rightarrow 0$. Se i nostri operatori fossero ovunque definiti, quindi continui, avremmo concluso: invece abbiamo che le relazioni appena dedotte valgono nello spazio di Banach $\mathcal{D} = \mathcal{D}(A) \cap \mathcal{D}(B)$, normato da

$$\|\varphi\|_{\mathcal{D}} = \|(A + B)\varphi\| + \|\varphi\|$$

In esso, la successione $\{n(S_{t/n} - U_{t/n})\}$ è infinitesima in norma, quindi è una successione di operatori continui da \mathcal{D} in \mathcal{H} , tali che

$$\forall \varphi \quad \sum_{n \in \mathbb{N}} \|n(S_{t/n} - U_{t/n})\varphi\| < \infty$$

Ma allora, per il teorema di Banach–Steinhaus (cfr. [1, §6.5]) esiste una costante $C > 0$ tale che

$$\|n(S_{t/n} - U_{t/n})\varphi\| \leq C\|\varphi\|_{\mathcal{D}}$$

Questo ci dice che il limite $\|(S_t - U_{t/n}^n)\psi\| \rightarrow 0$ è uniforme sui compatti di \mathcal{D} .

Per concludere la dimostrazione non resta che notare come, dato $\psi \in \mathcal{D}$, la funzione $F : [0, t] \rightarrow \mathcal{D}$ definita come $F(s) = \psi_s = S\psi_s$ è continua, dunque l'insieme $\{\psi_s\}_{s \in [0, t]}$ è compatto in \mathcal{D} , da cui

$$\lim_{n \rightarrow \infty} n \sup_{s \in [0, t]} \|(S_{t/n} - U_{t/n})\psi_s\| = 0$$

il che implica, per la disuguaglianza più sopra dimostrata, che $\|(S_t - U_{t/n}^n)\psi\| \rightarrow 0$, cioè che $S_t = e^{it(A+B)}$ è limite forte degli $U_{t/n}^n$.

QED

12 Formula di Feynmann–Kac

Possiamo finalmente dimostrare la celebre formula di Feynmann–Kac, che lega il semigruppò e^{-tH} all'integrale di Wiener.

Teorema 12.1 *Se $B = C_\infty(\mathbb{R}^N)$ è lo spazio delle funzioni continue nulle all'infinito, $V \in B$, $d\mu = dx \otimes dW$ è la misura di Wiener su $\mathbb{R}^n \times B$ e $H = H_0 + V$ è autoaggiunto su $\mathcal{D}(H_0)$ allora, per ogni $f, g \in B$:*

$$\langle f, e^{-tH} g \rangle = \int \overline{f(\omega(0))} g(\omega(t)) \exp \left[- \int_0^t V(\omega(s)) ds \right] d\mu(\omega)$$

dove, dato che $s \mapsto V(\omega(s))$ è continua per quasi ogni ω , $\int_0^t V(\omega(s)) ds$ può intendersi come integrale di Riemann.

DIMOSTRAZIONE: Usiamo immediatamente la formula di Trotter:

$$\langle f, e^{-tH} g \rangle = \lim_{n \rightarrow \infty} \langle f, \left(e^{-\frac{t}{n}H_0} e^{-\frac{t}{n}V} \right)^n g \rangle$$

Ma, per il teorema 8.5 si ha allora che

$$\int f_0(\omega(s_0)) \cdots f_n(\omega(s_n)) d\mu(\omega) = \langle \overline{f_0}, e^{-t_1 H_0} f_1 \cdots e^{-t_n H_0} f_n \rangle$$

ove, al solito, $t_k = s_k - s_{k-1}$, da cui

$$(*) \quad \langle f, e^{-tH} g \rangle = \lim_{n \rightarrow \infty} \int \overline{f(\omega(0))} g(\omega(t)) \exp \left[- \frac{t}{n} \sum_{k=0}^{n-1} V(\omega(t_k/n)) \right] d\mu(\omega)$$

dove

$$\lim_{n \rightarrow \infty} \frac{t}{n} \sum_{k=0}^{n-1} V(\omega(t_k/n)) = \int_0^t V(\omega(s)) ds$$

per quasi ogni ω . Inoltre si ha

$$\begin{aligned} \left| \overline{f(\omega(0))} g(\omega(t)) \exp \left[- \frac{t}{n} \sum_{k=0}^{n-1} V(\omega(t_k/n)) \right] \right| &\leq \\ &\leq |\overline{f(\omega(0))}| |g(\omega(t))| \exp[t\|V\|_\infty] \end{aligned}$$

Sempre per il teorema 8.5, si ha pure che

$$\int |\overline{f(\omega(0))}| |g(\omega(t))| d\mu = \langle |f|, e^{-tH_0} |g| \rangle < \infty$$

cioè che $|\overline{f(\omega(0))}| |g(\omega(t))| \exp[t\|V\|_\infty] \in L^1$, il che consente di applicare il teorema della convergenza dominata alla (*), trovando quindi la tesi.

QED

La formula del teorema precedente coinvolge la misura $d\mu = dx \otimes dW$ su $\mathbb{R}^n \times B$, cioè integra lungo cammini del tipo $x + \omega(t)$. Il risultato originale di Feynmann, e lo spirito della nostra trattazione, che guardano a W_t come all'oggetto centrale, chiedono una riformulazione di questo risultato in termini della sola misura di Wiener. Precisamente si ha

Teorema 12.2 *Se $V \in B$ e $H = H_0 + V \in C(\mathbb{R}^N)$ allora, per ogni $f \in C(\mathbb{R}^N)$:*

$$(e^{-tH} f)(0) = \int \exp \left[- \int_0^t V(W_s) ds \right] f(W_t) dW_t$$

DIMOSTRAZIONE: Notiamo innanzi tutto che ponendo

$$(e^{-tH} f)(x) = \frac{1}{\sqrt{2\pi t}} \int e^{-\frac{|x-y|^2}{2t}} f(y) dy$$

definiamo in $C(\mathbb{R}^N)$ un semigruppino continuo di operatori e, per funzioni in $C(\mathbb{R}^N) \cap L^\infty(\mathbb{R}^N)$, si ha (cfr. teorema 8.1)

$$\mathbb{E} [f_1(W(s_1)) \cdots f_n(W(s_n))] = (e^{-s_1 H_0} f_1 \cdots e^{-(s_n - s_{n-1}) H_0} f_n)(0)$$

Allora, la formula di Trotter in $C(\mathbb{R}^N) \cap L^\infty(\mathbb{R}^N)$ ci permette di ottenere la tesi come nel caso del teorema precedente.

QED

La validità di questi risultati dipende sensibilmente dal tipo di regolarità di V , che nei problemi applicativi può avere comportamenti anche capricciosi: per esempio, se V è limitata si può espandere l'integrale $\int V(\omega(s)) ds$ in serie di potenze, mentre se $V(x) \rightarrow \infty$ per $|x| \rightarrow \infty$ è necessario considerare particolari condizioni di crescita.

Esempio 12.3 *Consideriamo l'equazione di Schrödinger che regola la dinamica quantistica non relativistica in finiti gradi di libertà: se $\mathcal{H} = L^2(\mathbb{R}^N)$ è lo spazio di Hilbert delle funzioni d'onda, vogliamo studiare una particella libera di massa 1 (la cui hamiltoniana è $H_0 = -\Delta/2$) soggetta al potenziale $V(x, t)$; se $H = H_0 + V$, l'equazione di Schrödinger è*

$$\begin{cases} \frac{\partial}{\partial t} \psi(x, t) = iH\psi(x, t) \\ \psi(x, 0) = \psi(x) \end{cases}$$

Le soluzioni si ricercano in $L^2(\mathbb{R}^N)$. Ora, se V non dipende da t e soddisfa alle condizioni che implicano che H è autoaggiunto, allora la

$$U(t) = e^{itH}$$

definisce un gruppo ad un parametro tale che

$$\frac{d}{dt}U(t)\varphi = iH(U(t)\varphi)$$

Pertanto $\psi = U(t)\varphi$ risolve l'equazione di Schrödinger (questo è l'approccio di Weyl–von Neumann alla teoria).

In generale, se consideriamo

$$F(x, y|t) = \int \exp \left[- \int_0^t V(\omega(s)) ds \right] dW$$

e se V è tale che l'operatore H ammetta un sistema di autofunzioni $\{f_n\}$ normalizzate di autovalori $\{\lambda_n\}$, cioè

$$-\Delta f_n + V f_n = \lambda_n f_n$$

allora è noto che F può esprimersi come

$$g(x - y|t)F(x, y|t) = \sum_n e^{-\lambda_n t} f_n(x)f_n(y)$$

ove g è un nucleo di Gauss. Ne segue che, ponendo

$$\psi(x, y|0, t) = g(x - y|t)F(x, y|t)$$

è possibile dimostrare la

$$\psi(x, y|t_1, t_2) = \int_{-\infty}^{\infty} \psi(x, y|t_0, t_1)\psi(x, y|t_1, t_2) dy$$

e quindi che $\psi(x_0, x|0, t)$ è soluzione fondamentale dell'equazione di Schrödinger.

13 La formula di Itô

Poiché si tratta della costruzione che sta alla base del calcolo stocastico, che trova innumerevoli applicazioni in fisica e nella finanza, voglio concludere queste note dando la definizione “costruttiva” dell'integrale di Wiener che si può far risalire a Itô: come notevole applicazione ne trarremo la celeberrima formula di Itô, sovente citata e dimostrata in modo trasandato e impreciso in molti testi di finanza.

Lemma 13.1 *Se, per $\alpha > 0$ e $n \in \mathbb{N}$, definiamo*

$$f(W, n, \alpha) = \sum_{k=1}^{2^n} \left| W\left(\frac{k}{2^n}\right) - W\left(\frac{k-1}{2^n}\right) \right|^\alpha$$

allora:

$$(1) \quad \alpha < 2 \implies f(W, n, \alpha) \longrightarrow \infty.$$

$$(2) \quad \alpha = 2 \implies f(W, n, \alpha) \longrightarrow 1.$$

$$(3) \quad \alpha > 2 \implies f(W, n, \alpha) \longrightarrow 0.$$

dove i limiti sono intesi in probabilità e, nei casi (2) e (3) anche in L^p .

DIMOSTRAZIONE: Poniamo

$$c_\alpha = \frac{1}{\sqrt{2\pi}} \int |x|^\alpha e^{-\frac{1}{2}x^2} dx$$

(di modo che $c_2 = 1$), e

$$d(n, \alpha) = \mathbb{E} \left[\left| W\left(\frac{k}{2^n}\right) - W\left(\frac{k-1}{2^n}\right) \right|^\alpha \right] = \frac{c_\alpha}{\sqrt{2^{n\alpha}}}$$

Dato che, al variare di k , le variabili $W\left(\frac{k}{2^n}\right) - W\left(\frac{k-1}{2^n}\right)$ sono indipendenti, si ha che

$$\begin{aligned} & \mathbb{E} \left[\left(\sum_{k=1}^{2^n} \left| W\left(\frac{k}{2^n}\right) - W\left(\frac{k-1}{2^n}\right) \right|^\alpha - d(n, \alpha) \right)^2 \right] \\ &= \sum_{k=1}^{2^n} \mathbb{E} \left[\left(\left| W\left(\frac{k}{2^n}\right) - W\left(\frac{k-1}{2^n}\right) \right|^\alpha - d(n, \alpha) \right)^2 \right] \\ &\leq \sum_{k=1}^{2^n} \mathbb{E} \left[\left| W\left(\frac{k}{2^n}\right) - W\left(\frac{k-1}{2^n}\right) \right|^{2\alpha} \right] = 2^n d(n, 2\alpha) \end{aligned}$$

(si rammenti che $\mathbb{E}[(X - \mathbb{E}[X])^2] \leq \mathbb{E}[X^2]$). Dunque

$$\mathbb{E} \left[\left| f(W, n, \alpha) - 2^{n(1-\frac{1}{2}\alpha)} c_\alpha \right| \right] \leq 2^{n(1-\alpha)} c_{2\alpha}$$

Se $\alpha < 2$ allora, per ogni k fissato e n abbastanza grande si ha

$$\frac{1}{2} c_\alpha 2^{n(1-\frac{\alpha}{2})} > k$$

Per questi n si ha pure che

$$\mathbb{E} [f(W, n, \alpha) < k] \leq \mathbb{E} \left[f - 2^{n(1-\frac{\alpha}{2})} c_\alpha < -\frac{1}{2} c_\alpha 2^{n(1-\frac{\alpha}{2})} \right] \leq \frac{4c_{2\alpha}}{2^n c_\alpha^2}$$

Ma allora, per il lemma di Borel–Cantelli, $f(W, n, \alpha) \geq k$ per questi n abbastanza grandi e, dato che k è qualsiasi, abbiamo la (1).

La (2) segue in modo analogo, solo che in questo caso si stima

$$\mathbb{E} [|f - 1| > \epsilon]$$

invece di $\mathbb{E} [f < k]$ e nella (3) si stima $\mathbb{E} [|f| > \epsilon]$.

QED

La nostra intenzione è di definire l'integrale di Wiener cominciando dal caso di funzioni C^1 come

$$\int_0^1 f(W) dW = \lim_{n \rightarrow \infty} \sum_{m=1}^{2^n} f \left(W \left(\frac{m-1}{2^n} \right) \right) \left(W \left(\frac{m}{2^n} \right) - W \left(\frac{m-1}{2^n} \right) \right)$$

Itô riesce a dare significato a questa espressione nel senso della convergenza puntuale: qui ci contentiamo di quella L^2 . Prima di dimostrare il teorema di convergenza, svolgiamo una fondamentale osservazione.

Notiamo per prima cosa il fatto importantissimo che gli incrementi usati nella somma il cui limite dà l'integrale (secondo la formula precedente) sono “futuri”, vale a dire calcoliamo f nell'estremo inferiore di un intervallo e la moltiplichiamo per la differenza fra gli estremi; il motivo di questa scelta sarà chiaro una volta che si sia calcolato $\int_0^1 W dW$, come segue: se poniamo

$$I_+ = \sum_{m=1}^{2^n} f \left(W \left(\frac{m}{2^n} \right) \right) \left(W \left(\frac{m}{2^n} \right) - W \left(\frac{m-1}{2^n} \right) \right)$$

$$I_- = \sum_{m=1}^{2^n} f \left(W \left(\frac{m}{2^n} \right) \right) \left(W \left(\frac{m}{2^n} \right) - W \left(\frac{m-1}{2^n} \right) \right)$$

allora, per ogni n si ha $W(1)^2 = I_+(n) + I_-(n)$, dato che

$$\begin{aligned} \left(W \left(\frac{m}{2^n} \right) + W \left(\frac{m-1}{2^n} \right) \right) \left(W \left(\frac{m}{2^n} \right) - W \left(\frac{m-1}{2^n} \right) \right) \\ = W^2 \left(\frac{m}{2^n} \right) - W^2 \left(\frac{m-1}{2^n} \right) \end{aligned}$$

Ma per il lemma precedente abbiamo che $I_+(n) - I_-(n)$ tende a 1 se $n \rightarrow \infty$, cioè che

$$\lim_{n \rightarrow \infty} I_\pm(n) = \frac{1}{2} (W(1)^2 \pm 1)$$

Ne viene, tenendo conto che $W(0) = 0$,

$$\int_0^1 W dW = \frac{1}{2} (W(1)^2 - 1)$$

il che mostra che le ordinarie regole del calcolo integrale non valgono per espressioni di questo tipo: infatti $\int x dx = \frac{1}{2}x^2$.

Attenzione: si potrebbe pensare di ovviare questa stranezza definendo $\int_0^1 W dW$ come il limite di $\frac{1}{2}(I_-(n) + I_+(n))$, col che si avrebbe la formula usuale; ma un momento di riflessione mostra che $\int_0^1 W^k dW$ non sarebbe comunque pari a $\frac{1}{k+1}W^{k+1}$ per $k > 1$.

Teorema 13.2 *Se $f \in C^1(\mathbb{R})$ e sia f che f' sono limitate allora il limite*

$$\int_0^1 f(W) dW = \lim_{n \rightarrow \infty} \sum_{m=1}^{2^n} f\left(W\left(\frac{m-1}{2^n}\right)\right) \left(W\left(\frac{m}{2^n}\right) - W\left(\frac{m-1}{2^n}\right)\right)$$

esiste in $L^2[0, 1]$ ed inoltre

$$\left\| \int_0^1 f(W) dW \right\|^2 = \mathbb{E} \left[\int_0^1 f(W(s))^2 ds \right]$$

DIMOSTRAZIONE: Sia

$$J_n(f) = \sum_{m=1}^{2^n} q_{m,n} = \sum_{m=1}^{2^n} f\left(W\left(\frac{m-1}{2^n}\right)\right) \left(W\left(\frac{m}{2^n}\right) - W\left(\frac{m-1}{2^n}\right)\right)$$

Per $k \neq j$ abbiamo che $\mathbb{E}[q_{k,n}q_{j,n}] = 0$, essendo per $k < j$ un fattore di $W\left(\frac{j}{2^n}\right) - W\left(\frac{j-1}{2^n}\right)$ indipendente dagli altri addendi che formano $q_{k,n}q_{j,n}$, ed essendo $\mathbb{E}\left[W\left(\frac{k}{2^n}\right)\right] = 0$.

Inoltre, $f\left(W\left(\frac{m-1}{2^n}\right)\right)$ è indipendente da $W\left(\frac{m}{2^n}\right) - W\left(\frac{m-1}{2^n}\right)$ sicché

$$\mathbb{E}[q_{m,n}^2] = \mathbb{E} \left[f\left(W\left(\frac{m-1}{2^n}\right)\right)^2 \right] \frac{1}{2^n}$$

vale a dire

$$\|J_n(f)\|^2 = \sum_{m=1}^{2^n} \mathbb{E} \left[f\left(W\left(\frac{m-1}{2^n}\right)\right)^2 \right] \frac{1}{2^n}$$

Ma dato che sia f che W sono continue, abbiamo che

$$\lim_{n \rightarrow \infty} \sum_{m=1}^{2^n} f\left(W\left(\frac{m-1}{2^n}\right)\right)^2 \frac{1}{2^n} = \int_0^1 f(W(s))^2 ds$$

e dato che la sommatoria è dominata da $\|f\|_\infty^2$, basterà mostrare la convergenza di $J_n(f)$ per avere la tesi del teorema.

Ed in effetti un calcolo analogo ai precedenti mostra che

$$\begin{aligned} \|J_{n+1}(f) - J_n(f)\|^2 &= \sum_{m=0}^{2^n-1} \mathbb{E} \left[\left(f \left(W \left(\frac{2m+1}{2^{n+1}} \right) \right) - f \left(W \left(\frac{2m}{2^{n+1}} \right) \right) \right)^2 \right] \\ &\leq \frac{1}{2^{n+1}} C \sum_{m=0}^{2^n-1} \mathbb{E} \left[\left(W \left(\frac{2m+1}{2^{n+1}} \right) - W \left(\frac{2m}{2^{n+1}} \right) \right)^2 \right] \\ &= \frac{C}{2^{n+2}} \end{aligned}$$

dove $C = \|f'\|_\infty^2$ (di modo che $|f(x) - f(y)|^2 \leq C|x - y|^2$). Pertanto $J_n(f)$ è una successione di Cauchy in L^2 , dunque convergente.

QED

È possibile estendere questo integrale al caso di funzioni qualsiasi purché soddisfino alla

$$\mathbb{E} \left[\int_0^1 \chi_{(-a,a)}(W(s)) f(W(s))^2 ds \right] < \infty$$

(per esempio se $f \in L^\infty \cap L^2$ intorno allo zero), potendosi in questo caso definire, per qualsiasi a ,

$$\int_0^1 f(W(s)) \chi_a(W(s)) ds$$

Di nuovo sfruttiamo la continuità di W in $[0, 1]$, che ne implica la limitatezza e quindi, fissato W , il fatto che il limite

$$\lim_{a \rightarrow 0} \int_0^1 \chi_{(-a,a)}(W(s)) f(W(s)) ds$$

esiste finito (quasi ovunque rispetto a W) e, se a è abbastanza grande, non dipende da a .

In altri termini stiamo estendendo l'integrale da C^1 a L^2 per L^2 -continuità, cioè stiamo definendo una mappa lineare (che poi è una isometria) $L^2(0, 1) \rightarrow L^2(C(0, 1), dW)$.

Inoltre, questo integrale può chiaramente definirsi rispetto alla misura di Wiener: basta integrare lungo la parte N -dimensionale usando la misura di Lebesgue

$$\int_0^1 f(\omega(s)) ds$$

La costruzione è la stessa, con la differenza che

$$\mathbb{E} \left[\int_0^1 |f(\omega(s))|^2 ds \right] = \int |f(x)|^2 dx^N$$

In questo modo abbiamo un integrale $\int_0^t f(\omega(s)) d\omega$ per quasi ogni cammino ω .

Formula di Itô 13.3 *Se $f : \mathbb{R}^N \times [0, T] \rightarrow \mathbb{R}$ che sia L^2 assieme ai suoi differenziali primi e secondi lungo la componente \mathbb{R}^N e f' lungo la componente $[0, T]$ allora per quasi ogni ω :*

$$\begin{aligned} f(\omega(t), t) &= f(\omega(0), 0) + \int_0^t \nabla f(\omega(s), s) d\omega \\ &\quad + \frac{1}{2} \int_0^t \Delta f(\omega(s), s) ds + \int_0^t f'(\omega(s), s) ds \end{aligned}$$

DIMOSTRAZIONE: Per argomenti di densità ci possiamo limitare al caso $f \in C_0^\infty$: cominciamo con lo scrivere

$$f(\omega(t), t) - f(\omega(0), 0) = A_n + B_n + C_n$$

dove, ponendo $\delta\omega_{n,m} = \omega(m2^{-n}) - \omega((m-1)2^{-n})$, abbiamo definito

$$\begin{aligned} A_n &= \sum_{m=1}^{2^n} \nabla f \left(\omega \left(t \frac{m-1}{2^n} \right), t \frac{m-1}{2^n} \right) \delta\omega_{n,m} \\ B_n &= \sum_{m=1}^{2^n} \left[f \left(\omega \left(t \frac{m}{2^n} \right), t \frac{m-1}{2^n} \right) - f \left(\omega \left(t \frac{m-1}{2^n} \right), t \frac{m-1}{2^n} \right) \right] - A_n \\ C_n &= \sum_{m=1}^{2^n} \left[f \left(\omega \left(t \frac{m}{2^n} \right), t \frac{m}{2^n} \right) - f \left(\omega \left(t \frac{m}{2^n} \right), t \frac{m-1}{2^n} \right) \right] \end{aligned}$$

Ovviamente queste definizioni un po' artificiali si ispirano al lemma 13.1 e sono fatte in modo che

$$\lim_n A_n = \int_0^t \nabla f(\omega(s), s) ds$$

(in L^2). A questo punto notiamo le seguenti disuguaglianze, che seguono dalla definizione di gradiente e laplaciano (in senso debole):

$$(1) \left| f(x) - f(y) - \nabla f(y) \cdot (x - y) - \frac{1}{2} \sum_{h,k} \frac{\partial^2 f}{\partial y_h \partial y_k} (x - y)_h (x - y)_k \right| \leq C |x - y|^3$$

$$(2) |f(x, t) - f(x, s) - f'(x, s)(t - s)| \leq C |t - s|^2$$

Utilizzando la (1) possiamo dunque affermare che B_n è pari, a meno di un termine additivo infinitesimo (nel senso L^2) a

$$B'_n = \frac{1}{2} \sum_{m=1}^{2^n} \sum_{h,k} \frac{\partial^2 f}{\partial y_h \partial y_k} \left(\omega \left(\frac{m-1}{2^n} \right), \frac{m-1}{2^n} \right) (\delta \omega_{n,m})_h (\delta \omega_{n,m})_k$$

che, come nella dimostrazione del lemma 13.1, è pari, sempre a meno di un termine additivo infinitesimo L^2 , a

$$B'_n = \frac{1}{2} \sum_{m=1}^{2^n} \sum_{h,k} \frac{\partial^2 f}{\partial y_h \partial y_k} \left(\omega \left(\frac{m-1}{2^n} \right), \frac{m-1}{2^n} \right) \frac{\delta_{hk}}{2^n}$$

Ne viene pertanto che, nella convergenza L^2 ,

$$\lim_n B_n = \frac{1}{2} \int_0^t \Delta f(\omega(s), s) ds$$

Infine dalla disuguaglianza (2) segue immediatamente che

$$\lim_n C_n = \int_0^t f'(\omega(s), s) ds$$

QED

La formula di Itô è una formula integrale che un ruolo analogo alla formula di Taylor, o meglio al teorema di Lagrange, nel caso del calcolo stocastico: spesso la si scrive nella notazione simbolica

$$df = \nabla f d\omega + f' dt + \frac{1}{2} \Delta f dt$$

Come si vede il fatto “bizzarro” è che nell’espressione differenziale (del primo ordine) compaiono le derivate seconde!

Esempio 13.4 Consideriamo ancora $H_0 = -\frac{1}{2}\Delta$ e sia

$$f(x, s) = (e^{-(t-s)H_0}g)(x)$$

di modo che $f' = H_0 f$ cioè $f' + \frac{1}{2}\Delta f = 0$: se applichiamo il lemma di Itô abbiamo allora che

$$f(W_t, t) = f(W_0, 0) + \int_0^t \nabla f(W_s, s) dW$$

Se invece $H = H_0 + V$ e $f(x, s) = (e^{-(t-s)H}g)(x)$, ne viene

$$(*) \quad f(W_t, t) = f(W_s, s) + \int_s^t \nabla f(W_u, u) dW + \int_s^t V(W_u) f(W_u, u) du$$

In particolare

$$f(W_0, 0) = f(W_t, t) - \int_0^t \nabla f(W_s, s) dW - \int_0^t V(W_u) f(W_u, u) du$$

Usando la formula di Itô è possibile dare una dimostrazione “diretta” della formula di Feynmann–Kac: cfr. [12, §V.14] per una trattazione elegante orientata alla teoria quantistica, o [6, §4.4] per un approccio più analitico più orientato alla teoria delle equazioni differenziali. Per le applicazioni finanziarie si può consultare [11] che fornisce una introduzione formalmente corretta ai metodi stocastici in finanza.

Bibliografia

- [1] P. Caressa, *Metodi matematici della meccanica quantistica*, free e-book alla URL <http://www.caressa.it>
- [2] I.M. Gel'fand, A.M. Yaglom, *J. Math. Phys.* 1 (1960), 48–69.
- [3] I. Guikhman, A.Skorokhod, *Introduction à la théorie des processus aléatoires*, MIR, 1990 (tr. dal russo).
- [4] T. Hida, *Brownian Motion*, Springer, 1980.
- [5] M. Kac, *Functional Integration and some of its Applications*, Lezioni Fermiane, Pisa, 1980.
- [6] I. Karatzas, S. Shreve, *Brownian Motion and Stochastic Calculus*, Springer, 1998.
- [7] Y. Katznelson, *An Introduction to Harmonic Analysis*, Dover, 1976.
- [8] A.N. Kolmogorov, S.V. Fomin, *Elementi di teoria delle funzioni e analisi funzionale*, Editori Riuniti, 1980 (tr. dal russo).
- [9] E. Nelson, *Dynamical Theories of Brownian Motion*, Princeton, 1967, ora on line alla URL .
- [10] H. Royden, *Real Analysis*, MacMillan, 1988.
- [11] S. Shreve, *Stochastic Calculus and Finance, II: Continuous-Time Models*, Springer, 2004.
- [12] B. Simon, *Functional Integration and Quantum Physics*, Academic Press, 1979.