Lecture notes in stochastic analysis

Contents

1 Overview of the measure theory for Probability **3**
- 1.1 The definition of measure . . . 3
- 1.2 Measurable functions and random variables . . . 4
- 1.3 Integration, expectations, conditional expectations . . . 6
- 1.4 Examples . . . 14

2 Stochastic processes **17**

3 Brownian Motion **19**
- 3.1 Existence of a Brownian motion . . . 19
- 3.2 Some properties of a Brownian motion . . . 25
- 3.3 Local properties of a Brownian path . . . 28

4 Markov Processes **29**
- 4.1 Continuous-time Markov processes . . . 29
- 4.2 Markov property . . . 32

5 Martingales **34**
- 5.1 Filtrations and Stopping times . . . 34
- 5.2 Definition and examples of martingales . . . 36
- 5.3 Discrete-time martingales . . . 37
- 5.4 Optional stopping theorem . . . 40
- 5.5 Local martingales . . . 44
- 5.6 The variance process . . . 47
- 5.7 Semimartingales . . . 50

6 Stochastic Integral **50**
- 6.1 Construction of the stochastic integral and Itô's isometry . . . 50
- 6.2 Properties of the stochastic integral . . . 56
- 6.3 An extension of the stochastic intergral and multi-dimensional stochastic intergrals . . . 60

7 Itô's formula **60**
7.1 The one-dimensional Itô formula 60
7.2 The multi-dimensional Itô formula 65
7.3 Itô representation theorem 68
7.4 The martingale representation theorem 70

8 Stochastic Differential Equations **71**
8.1 Formal definitions 71
8.2 An existence and uniqueness theorem 72
8.3 The Tanaka equation 74

9 Stochastic partial differential equations **76**

1. Overview of the measure theory for Probability

This chapter reviews some basic topics and concepts of the measure theory and probability which are important for understanding the rest of the course.

1.1 The definition of measure

Definition 1.1. *Let Ω be a set. A* ***σ-algebra*** *$\mathcal{F}$ is a family of subsets of Ω possessing the following properties:*

(1) The empty set $\varnothing$ belongs to $\mathcal{F}$.

(2) If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$, where $A^c = \Omega \setminus A$ is the complement of A in Ω.

(3) If $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Definition 1.2. *A* ***measure*** *μ on $\mathcal{F}$ is a function $\mathcal{F} \to [0, \infty)$ such that for any sequence of subsets $\{A_i\}_{i=1}^{\infty}$, $A_i \in \mathcal{F}$, with $A_i \cap A_j = \varnothing$, $i \neq j$,*

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

Definition 1.3. *A* ***signed measure*** *on $\mathcal{F}$ is a function $\mathcal{F} \to (-\infty, \infty)$ where the rest of the properties is the same as for the measure.*

Definition 1.4. *We say that some property A holds* ***almost everywhere*** *(a.e.) if the set of those point $x \in E$ where the property A does not hold is zero.*

Definition 1.5. *A* ***probability measure*** *P on $\mathcal{F}$ is a function $\mathcal{F} \to [0, 1]$ satisfying the following conditions:*

(1) $P(\varnothing) = 0$, $P(\Omega) = 1$.

(2) If a sequence of subsets $\{A_i\}_{i=1}^{\infty}$, $A_i \in \mathcal{F}$, is such that $A_i \cap A_j = \varnothing$, $i \neq j$, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Sometimes, if a σ-algebra $\mathcal{F}$ of subsets of Ω is specified, we may use the expression "a probability measure on Ω" keeping in mind that it is actually a probability measure on the σ-algebra $\mathcal{F}$ of subsets of Ω. The $\mathcal{F}$-measurable subsets of Ω are often thought as *events*, while $P(A)$ is thought as the probability that the event A occurs. If $P(A) = 1$, we say that the event A occurs *almost surely (a.s.)*.

The triple $(\Omega, \mathcal{F}, P)$ is called a *probability space*.

1.2 Measurable functions and random variables

Let (E, μ) be a measurable space, i.e. a set E with the measure μ on a σ-algebra $\mathcal{G}$ of subsets of E. A function $Y : E \to \mathbb{R}^n$ is called *$\mathcal{F}$-measurable* if for any Borel subset $U \subset \mathbb{R}^n$,

$$Y^{-1}(U) = \{x \in E : Y(x) \in U\} \in \mathcal{G}.$$

Let $(\Omega, \mathcal{F}, P)$ be a probability space.

Definition 1.6. *An $\mathcal{F}$-measurable function $X : \Omega \to \mathbb{R}^n$ is called a* ***random variable****.*

Let $\mathcal{B}$ be the Borel σ-algebra of $\mathbb{R}^n$. Every random variable X generates a σ-algebra $\mathcal{H}_X$:

$$\mathcal{H}_X = \{X^{-1}(B), B \in \mathcal{B}\}.$$

Clearly, X is $\mathcal{H}_X$-measurable. Moreover, $\mathcal{H}_X$ is the smallest σ-algebra among σ-algebras with the property that X is measurable with respect to this σ-algebra.

Let I be a finite or infinite index set.

Definition 1.7.

(1) Two subsets $A \in \mathcal{F}$ and $B \in \mathcal{F}$ are called ***independent*** *if*

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

(2) σ-algebras $\{\mathcal{H}_i\}$, $i \in I$, are called ***independent*** *if for any choice of sets $H_{i_1} \in \mathcal{H}_{i_1}$, ..., $H_{i_k} \in \mathcal{H}_{i_k}$*

$$P(H_{i_1} \cap \cdots \cap H_{i_k}) = P(H_{i_1}) \cdots P(H_{i_k}).$$

(3) Random variables $\{X_i\}$, $i \in I$, are called ***independent*** *if the σ-algebras $\mathcal{H}_{X_i}$ generated by X_i are independent.*

Example (Independent sets). Let

$$\Omega = \{HH, HT, TH, TT\}$$

be the sample space corresponding to two tosses of a fair coin, equipped with the probability measure

$$P(\{\omega\}) = \tfrac{1}{4} \quad \text{for every } \omega \in \Omega.$$

Define the events

$$A = \{\text{the first toss is Heads}\} = \{HH, HT\},$$

$$B = \{\text{the second toss is Heads}\} = \{HH, TH\}.$$

Then

$$P(A) = \tfrac{1}{2}, \qquad P(B) = \tfrac{1}{2},$$

and

$$A \cap B = \{HH\}, \qquad P(A \cap B) = \tfrac{1}{4}.$$

Hence,

$$P(A \cap B) = \tfrac{1}{4} = \tfrac{1}{2} \cdot \tfrac{1}{2} = P(A)\,P(B),$$

and therefore the events A and B are independent.

Theorem 1.8. *If $f : E \to \mathbb{R}^n$ is a measurable function, and $g : \mathbb{R}^n \to \mathbb{R}^m$ is another measurable function, then the composition $g \circ f$ is a measurable function.*

Remark. We assume that on $\mathbb{R}^n$ we are give the Borel σ-algebra.
Assumption. Below we assume that the σ-algebra $\mathcal{G}$ of subsets of the measurable space E is augmented with all subsets of zero μ-measure sets.

Definition 1.9. *A measure μ defined on a σ-algebra augmented with all subsets of null sets is said to be* ***complete****.*

Theorem 1.10. *A funtion $f : E \to \mathbb{R}$ is measurable if and only if the set $\{x \in E : f(x) < c\}$ is measurable for every $c \in \mathbb{R}$.*

Theorem 1.11. *The sum, the difference, and the product of two measurable functions f and g are measurable. If $g(x) \neq 0$ for all $x \in E$, then $\frac{f}{g}$ is also measurable.*

Definition 1.12. *Let $f_n : E \to \mathbb{R}$, and $f : E \to \mathbb{R}$. We say that $f_n(x)$ converges to $f(x)$* ***almost everywhere*** *(a.e.) on E if the measure of those $x \in E$ where the convergence does not hold is zero.*

Theorem 1.13. *If a sequence of measurable functions f_n converges to f almost everywhere, then f is also measurable.*

Definition 1.14. *A sequence of functions $f_n : E \to \mathbb{R}$ is said to converge* ***in measure*** *to a function $f : E \to \mathbb{R}$ if for every $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mu\{x : |f_n(x) - f(x)| \geqslant \varepsilon\} = 0.$$

Theorem 1.15. *If a sequence of measurable functions $\{f_n\}$ converges to a function f almost everywhere, then it converges to the same functions in measure.*

Theorem 1.16. *If a sequence of measurable functions $\{f_n\}$ converges to a function f in measure, then there exists a subsequence $\{f_{n_k}\}$ of $\{f_n\}$ that converges to f almost everywhere.*

1.3 Integration, expectations, conditional expectations

Below we assume that a measure μ is given on a complete σ-algebra of subsets of E.

Definition 1.17. *A function f on (E, μ) is called **simple** if it takes no more than countable number of values.*

Theorem 1.18. *A function f taking no more than countable number of different values*

$$y_1, \dots, y_n, \dots,$$

is measurable if and only if the sets

$$A_n = \{x : f(x) = y_n\}$$

are measurable.

Theorem 1.19. *A function f is measurable if and only if one can represent it as a limit of a uniformly convergent sequence of simple measurable functions.*

Let $f = \sum_{i=1}^{\infty} y_i \mathbb{1}_{A_i}$ where $y_i \neq y_j$ if $i \neq j$ and the sets A_i are μ-measurable.

Definition 1.20. *The integral of f over a μ-measurable set A with respect to the measure μ is defined as follows:*

$$\int_A f(x)\mu(dx) = \sum_{i=1}^{\infty} y_i \mu(A_i \cap A).$$

*If the series on the right-hand side converges, then the function f is called **integrable** on the set A.*

Below we list some properties of the integral for simple functions.

1. Additivity:

$$\int_A [f(x) + g(x)]\mu(dx) = \int_A f(x)\mu(dx) + \int_A g(x)\mu(dx),$$

   moreover, the existence of the integrals on the right-hand side implies the existence of the integral on the left-hand side.

2. Multiplicativity: If $k \in \mathbb{R}$, then

$$\int_A kf(x)\mu(dx) = k \int_A f(x)\mu(dx),$$

   moreover, the existence of the integrals on the right-hand side implies the existence of the integral on the left-hand side.

3. If f is bounded on A by a constant M, then

$$\Big|\int_A f(x)\mu(dx)\Big| \leqslant M\mu(A).$$

Definition 1.21. *A function f is called integrable on A if there exists a sequence of simple functions $\{f_n\}$ which converges to f uniformly. The limit*

$$\lim_{n\to\infty}\int_A f_n(x)\mu(dx) \tag{1}$$

*is denoted $\int_A f(x)\mu(dx)$ and called the **integral** of f over A with respect to μ.*

The integral of f over A with respect to μ is well defined. Indeed, limit (1) exists for any uniformly convergent system of simple functions since

$$\Big|\int_A f_n(x)\mu(dx) - \int_A f_m(x)\mu(dx)\Big| \leqslant \mu(A)\sup_{x\in A}|f_n(x) - f_m(x)|.$$

The above inequality also implies that limit (1) does not depend on the choice of a sequence of simple functions that uniformly converges to f. Properties 1, 2, and 3 easily follow for the integral defined for an arbitrary integrable function. Below we will list some additional properties.

4. If $f(x) \geqslant g(x)$, then

$$\int_A f(x)\mu(dx) \geqslant \int_A g(x)\mu(dx).$$

5. If $\mu(A) = 0$, then $\int_A f(x)\mu(dx) = 0$.

6. If $f(x) = g(x)$ almost everywhere on A, then

$$\int_A f(x)\mu(dx) = \int_A g(x)\mu(dx).$$

7. σ-additivity: If $A = \cup_n A_n$ where $A_i \cap A_j = \emptyset$ $(i \neq j)$, then

$$\int_A f(x)\mu(dx) = \sum_n \int_{A_n} f(x)\mu(dx).$$

   Moreover, from the existence of the integral on the left-hand side it follows that each integral over A_n on the right-hand side exists, and the series converges.

8. If $\mu(A) \neq 0$, and

$$\int_A |f(x)|\mu(dx) = 0,$$

   then $f(x) = 0$ on A a.e..

9. Chebyshev's inequality: if $f(x) \geqslant 0$ and $c > 0$, then

$$\mu\{x \in A : \ f(x) > c\} \leqslant \frac{1}{c} \int_A f(x)\mu(dx).$$

10. Absolute continuity of the integral: If f is integrable on A, then for every $\varepsilon > 0$, there exists a $\delta > 0$ such that for every μ-measurable subset $C \subset A$ with $\mu(C) < \delta$,

$$\Big| \int_C f(x)\mu(dx) \Big| < \varepsilon.$$

Passing to the limit under the integral sign

Theorem 1.22 (Lebesgue's [or dominated convergence] theorem)**.** *Let f_n converge to f on A a.e. and for all n*

$$|f_n(x)| \leqslant \phi(x) \text{ a.e.}$$

where ϕ is integrable on A. Then, the function f is integrable on A, and

$$\lim_{n\to\infty} \int_A f_n(x)\mu(dx) = \int_A f(x)\mu(dx).$$

Theorem 1.23 (Beppo Levy's [or monotone convergence] theorem)**.** *Let $f_1(x) \leqslant f_2(x) \leqslant \cdots \leqslant f_n(x) \leqslant \cdots$ on A, moreover, all f_n are integrable and their integrals are bounded:*

$$\int_A f_n(x)\mu(dx) \leqslant K.$$

Then, almost everywhere on A there exists a limit

$$\lim_{n\to\infty} f_n(x) = f(x).$$

Moreover, the function $f(x)$ is integrable on A, and

$$\lim_{n\to\infty} \int_A f_n(x)\mu(dx) = \int_A f(x)\mu(dx).$$

Theorem 1.24 (Fatou Lemma)**.** *Let a sequence of integrable non-negative functions f_n converge to f a.e. on A, and*

$$\int_A f_n(x)\mu(dx) \leqslant K.$$

Then f is integrable on A, and

$$\int_A f(x)\mu(dx) \leqslant K.$$

Fubini's theorem

Let $A \subset U \times V$. Define

$$A_x = \{y \in Y : (x, y) \in A\} \quad (x \text{ is fixed})$$
$$A_y = \{x \in X : (x, y) \in A\} \quad (y \text{ is fixed}).$$

Theorem 1.25. *Let the measures μ_U and μ_V be defined on U and V, respectively, σ-additive, and complete. Further let $\mu = \mu_U \otimes \mu_V$ and the function $f(x, y)$ be integrable on the set $A \subset U \times V$. Then*

$$\int_A f(x)\mu(dx) = \int_U \mu_U(dx) \int_{A_x} f(x, y)\mu_V(dy) = \int_V \mu_V(dy) \int_{A_y} f(x, y)\mu_U(dx). \quad (2)$$

Corollary 1.26. *Let one the integrals*

$$\int_U \mu_U(dx) \int_{A_x} |f(x, y)| \, \mu_V(dy) \quad \text{or}$$
$$\int_V \mu_V(dy) \int_{A_y} |f(x, y)| \, \mu_U(dx).$$

exists. Then f is integrable on A and (2) holds.

Distributions, expectations, characteristic functions

Every random variable X induces a probability distribution μ_X on $\mathbb{R}^n$ defined on the σ-algebra of Borel subsets of $\mathbb{R}^n$ by the formula:

$$\mu_X(B) = P\big(X^{-1}(B)\big). \quad (3)$$

Definition 1.27. *The measure μ_X defined by (3) is called the **distribution** of X.*

Definition 1.28. *The number*

$$E[X] = \int_\Omega X(\omega)P(d\omega) = \int_{\mathbb{R}^n} x \, \mu_X(dx)$$

*is called the **expectation** of X.*

Definition 1.29. *The **characteristic function** of a random variable $X : \Omega \to \mathbb{R}^n$ is the function $\varphi_X : \mathbb{R}^n \to \mathbb{C}$ (where $\mathbb{C}$ denotes the complex numbers) defined by*

$$\varphi_X(y_1, \ldots, y_n) = \mathbb{E}[\exp\{i\,(y_1X_1 + \cdots + y_nX_n)\}] = \int_{\mathbb{R}^n} e^{i(y,x)_n} \, (P \circ X^{-1})(dx)$$

where $X_1, \ldots, X_n$ are the coordinates of X in $\mathbb{R}^n$, $(y, x)_n$ is the scalar product in $\mathbb{R}^n$.

Theorem 1.30. *The characteristic function of X determines the distribution $P \circ X^{-1}$ of X uniquely.*

Proof. The characteristic function is actually the Fourier transform of the measure $P \circ X^{-1}$, and therefore defines this measure uniquely. □

Conditional expectations

Definition 1.31. *Let a measure μ and a signed measure ν be defined on a σ-algebra $\mathcal{G}$ of the space E. The signed measure ν is said to be* ***absolutely continuous*** *with respect to the measure μ if $\nu(A) = 0$ for any set $A \in \mathcal{G}$ with $\mu(A) = 0$.*

Theorem 1.32 (Radon-Nikodym's theorem)**.** *Let μ be a measure on a σ-algebra $\mathcal{G}$ of the space E, and let the measure ν, defined on the same σ-algebra $\mathcal{G}$, be absolutely continuous with respect to μ. Then, there exists an integrable function f on E such that for any $A \in \mathcal{G}$*

$$\nu(A) = \int_A f(x)\mu(dx).$$

The function f is called the derivative of ν with respect to μ and denoted $f = \frac{d\nu}{d\mu}$.

Now let $(\Omega, \mathcal{F}, P)$ be a probability space, X be a random variable and $\mathcal{B} \subset \mathcal{F}$ be a sub-σ-algebra.

Definition 1.33 (Conditional expectation)**.** *A $\mathcal{B}$-measurable random variable Y is called the* ***conditional expectation*** *of X with respect to $\mathcal{B}$, and denoted $Y = \mathbb{E}[X|\mathcal{B}]$, if for any $\mathcal{B}$-measurable set A, it holds that*

$$\int_A Y(\omega)P(d\omega) = \int_A X(\omega)P(d\omega).$$

Conditional expectation always exists by Radon-Nikodym's theorem. Indeed, consider the restriction of the measure P to $\mathcal{H}$. We denote this restriction by the same symbol P. Define the signed measure $\nu(A) = \int_A X(\omega)P(d\omega)$ on the σ-algebra $\mathcal{H}$. Clearly, $\nu(A)$ is absolutely continuous with respect to P. By Radon-Nikodym's theorem, there exists a $\mathcal{H}$-measurable function Y such that $\nu(A) = \int_A Y(\omega)P(d\omega)$. The latter implies that $Y = \mathbb{E}[X|\mathcal{H}]$.

Theorem 1.34 (Properties of conditional expectations)**.** *Let, as before, $\mathcal{H}$ be a sub-σ-algebra of $\mathcal{F}$.*

1. *If a random variable X is $\mathcal{H}$-measurable, then $\mathbb{E}[X|\mathcal{H}] = X$ a.s.*
2. *For any square integrable $\mathcal{H}$-measurable Z (that is, $Z \in L_2(\Omega)$), $\mathbb{E}[ZX] = \mathbb{E}[Z\,\mathbb{E}[X|\mathcal{H}]]$.*
3. $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|\mathcal{H}]]$.
4. - *Additivity:* $\mathbb{E}[X + Y|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}] + \mathbb{E}[Y|\mathcal{H}]$ *a.s.*
   - *Linearity:* $\mathbb{E}[cX + d|\mathcal{H}] = c\,\mathbb{E}[X|\mathcal{H}] + d$ *a.s., where* $c, d \in \mathbb{R}$.
5. *If $X, Z \in L^2(\Omega)$ and Z is $\mathcal{H}$-measurable, where $\mathcal{H} \subset \mathcal{F}$, then $\mathbb{E}[ZX|\mathcal{H}] = Z\mathbb{E}[X|\mathcal{H}]$ a.s..*
6. *If $\mathcal{G} \subset \mathcal{H}$ is a sub-σ-algebra, then $\mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}]$ a.s.*

7. *If $X \leqslant Y$ where Y is another random variable, then $\mathbb{E}[X|\mathcal{H}] \leqslant \mathbb{E}[Y|\mathcal{H}]$ a.s..*

8. *If $X \in L^1(\Omega)$ is independent of $\mathcal{H}$, then $\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[X]$.*

Proof of Property 5. Set $Y := \mathbb{E}[X \mid \mathcal{H}]$. First assume that Z takes at most countably many values and can be written as

$$Z = \sum_{m=1}^{N} c_m \, \mathbb{1}_{B_m},$$

where $B_m \in \mathcal{H}$ are pairwise disjoint. For any $A \in \mathcal{H}$,

$$\int_A ZX \, \mathbb{P}(d\omega) = \sum_{m=1}^{N} c_m \int_{A \cap B_m} X \, \mathbb{P}(d\omega) = \sum_{m=1}^{N} c_m \int_{A \cap B_m} Y \, \mathbb{P}(d\omega) = \int_A ZY \, \mathbb{P}(d\omega).$$

Now if $Z \geqslant 0$ is an $L_2(\Omega)$ random variable, we approximate it pointwisee by step functions Z_n taking a finite number of values and with the property that $Z_n \leqslant Z_{n+1} \leqslant Z$ a.s. Indeed, take $N = 2^n$ and define

$$\begin{aligned} A_{mN} &= \big\{\omega \in \Omega : \frac{m}{N} \leqslant Z(\omega) < \frac{m+1}{N}\big\}, \qquad -N^2 \leqslant m \leqslant N^2 - 1. \\ Z_n(\omega) &:= \sum_{m=0}^{N^2-1} \frac{m}{N} \, \mathbb{1}_{A_{mN}}(\omega). \end{aligned} \tag{4}$$

Suppose first $X \geqslant 0$. Then, for every $n \in \mathbb{N}$,

$$Z_n Y \leqslant Z_{n+1} Y \quad \text{and} \quad \int_A Z_n Y \mathbb{P}(d\omega) = \int_A Z_n X \mathbb{P}(d\omega) \leqslant \int_A ZX \mathbb{P}(d\omega).$$

Futhermore,

$$\lim_{n\to\infty} Z_n Y = ZY \qquad \text{a.s.}$$

By the monotone converge theorem, $\int_A ZY \mathbb{P}(d\omega) < \infty$ and we can pass to the limit under the integral sign. Therefore,

$$\int_A ZX \, \mathbb{P}(d\omega) = \lim_{n\to\infty} \int_A Z_n X \, \mathbb{P}(d\omega) = \lim_{n\to\infty} \int_A Z_n Y \, \mathbb{P}(d\omega) = \int_A ZY \, \mathbb{P}(d\omega).$$

Since this holds for all $A \in \mathcal{H}$, it follows that

$$\mathbb{E}[ZX \mid \mathcal{H}] = Z\mathbb{E}[X \mid \mathcal{H}] \quad \text{a.s.}$$

if $Z \geqslant 0$ and $X \geqslant 0$. To show the general case, we note that $Z = Z_+ - Z_-$ and $X = X_+ - X_-$, where $Z_+ := \max\{Z, 0\}$ and $Z_- := \max\{-Z, 0\}$. Note that Z_+, Z_- are $\mathcal{H}$-measurable.

Applying the result proved above to each product $Z_\pm X_\pm$, we obtain

$$\mathbb{E}[Z_+X_+ \mid \mathcal{H}] = Z_+\mathbb{E}[X_+ \mid \mathcal{H}], \qquad \mathbb{E}[Z_+X_- \mid \mathcal{H}] = Z_+\mathbb{E}[X_- \mid \mathcal{H}],$$

$$\mathbb{E}[Z_- X_+ \mid \mathcal{H}] = Z_- \mathbb{E}[X_+ \mid \mathcal{H}], \qquad \mathbb{E}[Z_- X_- \mid \mathcal{H}] = Z_- \mathbb{E}[X_- \mid \mathcal{H}].$$

By Property 4,

$$\mathbb{E}[ZX \mid \mathcal{H}] = Z\,\mathbb{E}[X \mid \mathcal{H}] \quad \text{a.s.}$$

This completes the proof. □

Proof of Property 4. Let $\mathcal{H} \subset \mathcal{F}$ and assume $X, Y \in L^1(\Omega)$. Set

$$U := \mathbb{E}[X \mid \mathcal{H}], \qquad V := \mathbb{E}[Y \mid \mathcal{H}].$$

We first prove additivity. The random variable $U + V$ is $\mathcal{H}$-measurable and integrable. For every $A \in \mathcal{H}$,

$$\int_A (X+Y)\,\mathbb{P}(d\omega) = \int_A X\,\mathbb{P}(d\omega) + \int_A Y\,\mathbb{P}(d\omega) = \int_A U\,\mathbb{P}(d\omega) + \int_A V\,\mathbb{P}(d\omega) = \int_A (U+V)\,\mathbb{P}(d\omega).$$

Therefore

$$\mathbb{E}[X + Y \mid \mathcal{H}] = U + V = \mathbb{E}[X \mid \mathcal{H}] + \mathbb{E}[Y \mid \mathcal{H}] \quad \text{a.s.}$$

Next, let $c, d \in \mathbb{R}$. The random variable $cU + d$ is $\mathcal{H}$-measurable and integrable. For every $A \in \mathcal{H}$,

$$\int_A (cX + d)\,\mathbb{P}(d\omega) = c \int_A U\,\mathbb{P}(d\omega) + d\,\mathbb{P}(A) = \int_A (cU + d)\,\mathbb{P}(d\omega).$$

Thus $cU + d$ satisfies the defining property of $\mathbb{E}[cX + d \mid \mathcal{H}]$, and we conclude that

$$\mathbb{E}[cX + d \mid \mathcal{H}] = c\,\mathbb{E}[X \mid \mathcal{H}] + d \quad \text{a.s.}$$

□

Proof of Property 6. Let $X \in L^1(\Omega)$ and let $\mathcal{G} \subset \mathcal{H} \subset \mathcal{F}$ be σ-algebras. Set $Y := \mathbb{E}[X \mid \mathcal{H}]$. We show that

$$\mathbb{E}[Y \mid \mathcal{G}] = \mathbb{E}[X \mid \mathcal{G}] \quad \text{a.s.}$$

Both $\mathbb{E}[Y \mid \mathcal{G}]$ and $\mathbb{E}[X \mid \mathcal{G}]$ are $\mathcal{G}$-measurable and integrable. By uniqueness of conditional expectation, it suffices to verify that

$$\int_A \mathbb{E}[Y \mid \mathcal{G}]\,\mathbb{P}(d\omega) = \int_A \mathbb{E}[X \mid \mathcal{G}]\,\mathbb{P}(d\omega) \qquad \forall\ A \in \mathcal{G}. \tag{5}$$

Fix $A \in \mathcal{G} \subset \mathcal{H}$. Using the defining property of conditional expectation, we obtain

$$\int_A \mathbb{E}[Y \mid \mathcal{G}]\,\mathbb{P}(d\omega) = \int_A Y\,\mathbb{P}(d\omega).$$

On the other hand, by the defining property of $\mathbb{E}[X \mid \mathcal{G}]$,

$$\int_A \mathbb{E}[X \mid \mathcal{G}]\,\mathbb{P}(d\omega) = \int_A X\,\mathbb{P}(d\omega).$$

But $\int_A X\,\mathbb{P}(d\omega) = \int_A Y\,\mathbb{P}(d\omega)$. This implies (5). □

Proof of Property 7. Let $X, Y \in L^1(\Omega)$ with $X \leqslant Y$ almost surely, and let $\mathcal{H} \subset \mathcal{F}$ be a σ-algebra. Set

$$U := \mathbb{E}[X \mid \mathcal{H}], \qquad V := \mathbb{E}[Y \mid \mathcal{H}].$$

Assume, by contradiction, that the set

$$A := \{U > V\}$$

has positive probability. Since U and V are $\mathcal{H}$-measurable, we have $A \in \mathcal{H}$. On A one has $U - V > 0$, hence

$$\int_A (U - V)\, \mathbb{P}(d\omega) > 0.$$

Using the defining property of conditional expectation and the fact that $A \in \mathcal{H}$, we obtain

$$\int_A U\, \mathbb{P}(d\omega) = \int_A X\, \mathbb{P}(d\omega), \qquad \int_A V\, \mathbb{P}(d\omega) = \int_A Y\, \mathbb{P}(d\omega).$$

Therefore,

$$\int_A (U - V)\, \mathbb{P}(d\omega) = \int_A (X - Y)\, \mathbb{P}(d\omega) > 0$$

which is false since $X \leqslant Y$ almost surely. □

Proof of Property 8. Set $c := \mathbb{E}[X]$. Since c is $\mathcal{H}$-measurable, it suffices to show that for every $A \in \mathcal{H}$,

$$\int_A X\, d\mathbb{P} = \int_A c\, d\mathbb{P} = c\, \mathbb{P}(A). \tag{6}$$

We proceed in steps.

Step 1: Indicators. Let $B \in \mathcal{F}$ be independent of $\mathcal{H}$ and set $X = \mathbb{1}_B$. Then for any $A \in \mathcal{H}$,

$$\int_A \mathbb{1}_B\, d\mathbb{P} = \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)\, \mathbb{E}[\mathbb{1}_B],$$

which is (6) for this X.

Step 2: Uniform approximation by simple functions. Define, for $n \in \mathbb{N}$ and $m \in \mathbb{Z}$,

$$B_{mn} := \left\{ \frac{m}{n} \leqslant X < \frac{m+1}{n} \right\},$$

and set

$$X_n := \sum_{m=-\infty}^{\infty} \frac{m}{n}\, \mathbb{1}_{B_{mn}}.$$

Then each X_n is measurable with respect to $\sigma(X)$ and satisfies

$$0 \leqslant X(\omega) - X_n(\omega) < \frac{1}{n} \quad \text{for all } \omega \in \Omega.$$

Consequently, $X_n \to X$ uniformly on Ω.

$$\int_A X_n \mathbb{P}(d\omega) = \sum_{m=-\infty}^{\infty} \frac{m}{n} \int_{A \cap B_{mn}} \mathbb{P}(d\omega) = \sum_{m=-\infty}^{\infty} \frac{m}{n}\, \mathbb{P}(A)\mathbb{P}(B_{mn}) = \mathbb{P}(A)\, \mathbb{E}[X_n].$$

Passing to the limit as $n \to \infty$ finishes the proof. □

1.4 Examples

EXAMPLE 1.(Wiener's probability space). Let $\Omega = [0,1]$. Define a σ-algebra $\mathcal{F}$ as the σ-algebra of all Lebesgue-measurable subsets of $[0,1]$, and let P be the Lebesgue measure on $[0,1]$. The triple $(\Omega, \mathcal{F}, \mathbb{P})$ builds a probability space. This probability space is called *Wiener's probability space*.

EXAMPLE 2.(Gaussian measure, Gaussian random variable). A probability measure γ on the real line is called *Gaussian* if its density with respect to Lebesgue measure is given by the formula:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\Big(-\frac{(x-a)^2}{2\sigma^2}\Big).$$

It holds that

$$\int_{\mathbb{R}} p(x)dx = 1. \tag{7}$$

This means that the measure γ of a Borel subset $A \subset \mathbb{R}$ is defined by

$$\gamma(A) = \int_A p(x)\,dx.$$

The numbers a and σ^2 are called the *mean* and resp. the *variance*. A real-valued random X variable is called *Gaussian* if its distribution $\mu_X = \mathbb{P} \circ X^{-1}$ is a Gaussian measure. If $a = 0$ and $\sigma^2 = 1$ the random variable X is called *standard Gaussian* random variable. A Gaussian random variable is also called a normally distributed random variable.

Proof of (7). Make the change of variables $y = \frac{x-a}{\sigma}$, so that $x = a + \sigma y$ and $dx = \sigma\, dy$. Then

$$\int_{\mathbb{R}} p(x)\,dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right) dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{y^2}{2}}\, dy.$$

Set

$$I := \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\, dy.$$

Then

$$I^2 = \left(\int_{\mathbb{R}} e^{-\frac{x^2}{2}}\, dx\right)\left(\int_{\mathbb{R}} e^{-\frac{y^2}{2}}\, dy\right) = \int_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2}}\, dx\, dy,$$

where we used Fubini's theorem. Passing to polar coordinates $x = r\cos\theta$, $y = r\sin\theta$, with Jacobian r, we obtain

$$I^2 = \int_0^{2\pi} \int_0^{\infty} e^{-\frac{r^2}{2}}\, r\, dr\, d\theta.$$

Compute the radial integral by the substitution $u = \frac{r^2}{2}$:

$$\int_0^{\infty} e^{-\frac{r^2}{2}}\, r\, dr = \int_0^{\infty} e^{-u}\, du = 1.$$

Hence

$$I^2 = 2\pi, \qquad \text{and therefore} \qquad I = \sqrt{2\pi}.$$

□

EXAMPLE 3.(Multi-normal distribution). An $\mathbb{R}^n$-valued random variable X is called **multi-normal** of its distribution has the density:

$$p_X(x_1, \dots, x_n) = \frac{\sqrt{|A|}}{(2\pi)^{\frac{n}{2}}} \exp\Big(-\frac{1}{2} \sum_{j,k=1}^{n} (x_j - m_j) a_{jk} (x_k - m_k)\Big) \tag{8}$$

where $m = (m_1, \dots, m_n) \in \mathbb{R}^n$ and $A = \{a_{jk}\} \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix. The vector m is called the *mean*, and the matrix $C = \{c_{jk}\} = A^{-1}$ is called the *covariance matrix* of X. One can explicitly compute the Fourier transform of the distribution of the random variable X. Indeed,

$$\begin{aligned} \varphi_X(y_1, \dots, y_n) = \int_{\mathbb{R}^n} e^{i(y,x)} \, (P \circ X^{-1})(dx) = \int_{\mathbb{R}^n} e^{i(y,x)} p_X(x) \, dx \\ = \exp\Big(-\frac{1}{2} \sum_{j,k} y_j c_{jk} y_k + i \sum_j y_j m_j\Big). \end{aligned} \tag{9}$$

EXAMPLE 4.(Discrete random variable, Poisson distribution, Poisson random variable). A random variable X is called a *discrete random variable* if it can be written in the from:

$$X = \sum_i u_i \, \mathbb{1}_{\Omega_i},$$

where $\mathbb{1}_{\Omega_i}$ is the indicator function of Ω_i, Ω_i are disjoint subsets of Ω such that $\cup_i \Omega_i = \Omega$, and u_i are the values that the random variable X takes with a non-zero probability. Namely,

$$P(\Omega_i) = P(\omega : X(\omega) = u_i) > 0.$$

A *Poisson random variable* X is a discrete random variable taking non-negative integer values $k = 0, 1, 2, \dots$ and having a *Poisson probability distribution* (with some parameter λ), i.e.

$$P(X = k) = \frac{e^{\lambda} \, \lambda^k}{k!}.$$

EXAMPLE 5.(Conditional probability) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $D \in \mathcal{F}$ with $\mathbb{P}(D) \neq 0$ and $\mathbb{P}(D^c) \neq 0$, where $D^c := \Omega \setminus D$. Consider the σ-algebra

$$\mathcal{H} = \{\varnothing, \ D, \ D^c, \ \Omega\} = \sigma(D).$$

Let $X \in L^1(\Omega)$. Compute the conditional expectation $\mathbb{E}[X \mid \mathcal{H}]$. Define

$$y_1 := \int_D X \, \mathbb{P}(d\omega), \qquad y_2 := \int_{D^c} X \, \mathbb{P}(d\omega).$$

Then

$$\int_\Omega X\,\mathbb{P}(d\omega) = \int_D X\,\mathbb{P}(d\omega) + \int_{D^c} X\,\mathbb{P}(d\omega) = y_1 + y_2.$$

Define the random variable

$$Y(\omega) := \frac{y_1}{\mathbb{P}(D)}\ \mathbb{1}_D(\omega) + \frac{y_2}{\mathbb{P}(D^c)}\ \mathbb{1}_{D^c}(\omega).$$

Clearly, Y is $\mathcal{H}$-measurable. We verify the defining property of conditional expectation. Indeed,

$$\begin{aligned}
&\int_D Y\,\mathbb{P}(d\omega) = \frac{y_1}{\mathbb{P}(D)}\,\mathbb{P}(D) = y_1 = \int_D X\,\mathbb{P}(d\omega),\\
&\int_{D^c} Y\,\mathbb{P}(d\omega) = \frac{y_2}{\mathbb{P}(D^c)}\,\mathbb{P}(D^c) = y_2 = \int_{D^c} X\,\mathbb{P}(d\omega),\\
&\int_\Omega Y\,\mathbb{P}(d\omega) = \int_D Y\,\mathbb{P}(d\omega) + \int_{D^c} Y\,\mathbb{P}(d\omega) = y_1 + y_2 = \int_\Omega X\,\mathbb{P}(d\omega),\\
&\int_\varnothing Y\,\mathbb{P}(d\omega) = \int_\varnothing X\,\mathbb{P}(d\omega) = 0.
\end{aligned}$$

Hence,

$$\mathbb{E}[X \mid \mathcal{H}] = \frac{y_1}{\mathbb{P}(D)}\ \mathbb{1}_D(\omega) + \frac{y_2}{\mathbb{P}(D^c)}\ \mathbb{1}_{D^c}(\omega).$$

Now let $A \in \mathcal{F}$ and $X := \mathbb{1}_A$. Then,

$$y_1 = \mathbb{P}(A \cap D), \qquad y_2 = \mathbb{P}(A \cap D^c)$$

Defining

$$\mathbb{P}(A \mid D) := \frac{\mathbb{P}(A \cap D)}{\mathbb{P}(D)}. \tag{10}$$

we obtain

$$\mathbb{E}[\mathbb{1}_A \mid \sigma(D)] = \mathbb{P}(A \mid D)\,\mathbb{1}_D + \mathbb{P}(A \mid D^c)\,\mathbb{1}_{D^c}\,.$$

This implies that

$$\mathbb{P}(A \mid D) = \mathbb{E}[\mathbb{1}_A \mid \sigma(D)]\big|_D.$$

Definition 1.35. *The number $\mathbb{P}(A \mid D)$, defined by (10), is called the conditional probability of the event A given D.*

Remark 1. *Interpretation.* The conditional probability $\mathbb{P}(A \mid D)$ represents the probability that the event A occurs under the information that the event D has occurred. Let

$$D = \{\text{the person is sick}\}, \qquad A = \{\text{the person is coughing}\}.$$

Assume that $\mathbb{P}(A) = 0.05$ and $\mathbb{P}(A \mid D) = 0.75$. This means that coughing is relatively rare in the general population, but it is very likely among sick individuals.

2. Stochastic processes

Definition 2.1. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. A* ***stochastic process*** *is a parametrized collection of $\mathbb{R}^n$-valued random variables:*

$$\{X_t\}_{t\in T}$$

defined on the probability space $(\Omega, \mathcal{F}, P)$.

T here is a parameter set. Usually it is the halfline $[0, \infty)$, or an interval $[a, b] \subset [0, \infty)$. It can also be the set of non-negative integers $\mathbb{Z}_+$. Let us emphasize that in the definition of a stochastic process, for each fixed $t \in T$,

$$\omega \mapsto X_t(\omega)$$

is a random variable. Let us fix $\omega_0 \in \Omega$ and consider the function

$$t \mapsto X_t(\omega_0).$$

This function is called a *path* of the process X_t.

Definition 2.2. *A non-decreasing family $\{\mathcal{F}_t\}$ of sub-σ-algebras of $\mathcal{F}$ is called a* ***filtration****, i.e. for all $0 \leqslant s < t < \infty$,*

$$\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}.$$

Definition 2.3. *A* ***filtered probability space*** *$(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ is a probability space with a filtration $\{\mathcal{F}_t\}$ of the σ-algebra $\mathcal{F}$.*

Definition 2.4. *A stochastic process X_t is called* ***adapted*** *with respect to the filtration $\mathcal{F}_t$ if for every $t \geqslant 0$ the random variable X_t is $\mathcal{F}_t$-measurable.*

Definition 2.5. *The* ***finite-dimensional distributions*** *of a process X_t are the measures $\mu_{t_1,\dots,t_k}$ on $\mathbb{R}^{nk}$, $k = 1, 2, \dots$, defined by:*

$$\mu_{t_1,\dots,t_k}(A_1 \times \cdots \times A_k) = \mathbb{P}(X_{t_1} \in A_1, \cdots, X_{t_k} \in A_k),$$

where $A_1, \dots, A_k$ are Borel subsets in $\mathbb{R}^n$.

Let stochastic processes X_t and Y_t be defined on probability spaces $(\Omega, \mathcal{F}, P)$ and resp. on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$.

Definition 2.6. *The processes X_t and Y_t are said to have the same finite-dimensional distributions if for any $k \in \mathbb{N}$, for any choice of $t_1, \dots, t_k \in T$, the finite-dimensional distributions $\mu^X_{t_1,\dots,t_k}$ and $\mu^Y_{t_1,\dots,t_k}$ of X_t and resp. Y_t coinside.*

Let us consider now a converse problem. Suppose we are given a family $\{\nu_{t_1,\dots,t_k}, k \in \mathbb{N}, t_i \in T\}$ of probability measures on $\mathbb{R}^{nk}$. We would like to construct a stochastic process X_t whose finite-dimenstional distributions coincide with $\nu_{t_1,\dots t_k}$.

Theorem 2.7 (Kolmogorov's extension theorem)**.** *For all $t_1, \dots t_k$, $k \in \mathbb{N}$, let $\nu_{t_1,\dots,t_k}$ be probability measures on $\mathbb{R}^{nk}$ satisfying the two following properties (consistency conditions):*

1. $\nu_{t_{\sigma(1)},\dots,t_{\sigma(k)}}(F_1 \times \cdots \times F_k) = \nu_{t_1,\dots t_k}(F_{\sigma^{-1}(1)} \times \cdots \times F_{\sigma^{-1}(k)})$;
   for all permutations σ of $\{1, 2, \dots, k\}$ and for all Borel subsets $F_i \subset \mathbb{R}^n$.
2. $\nu_{t_1,\dots t_k}(F_1 \times \cdots \times F_k) = \nu_{t_1,\dots,t_k,t_{k+1},\dots,t_{k+m}}(F_1 \times \cdots \times F_k \times \underbrace{\mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{m})$
   for all $m \in \mathbb{N}$ and for all Borel subsets $F_i \subset \mathbb{R}^n$.

Then there exists a probability space $(\Omega, \mathcal{F}, P)$ and an $\mathbb{R}^n$-valued stochastic process X_t on it, such that

$$\nu_{t_1,\dots,t_k}(F_1 \times \cdots \times F_k) = \mathbb{P}(X_{t_1} \in F_1, \cdots, X_{t_k} \in F_k)$$

for all $t_i \in T$, $k \in \mathbb{N}$, and all Borell subsets $F_i \subset \mathbb{R}^n$.

Proof. Without proof. A proof can be found, for example, in the book by Koralov, L., Sinai, Y. "Theory of probability and random processes", 2007, p. 167. □

Let X_t and Y_t be stochastic proesses given on the same probability space $(\Omega, \mathcal{F}, P)$.

Definition 2.8. *We say that the process Y_t is a **version** or **modification** of X_t if for all $t \in T$*

$$P\{\omega : X_t(\omega) = Y_t(\omega)\} = 1.$$

Note that the finite-dimensional distributions of X_t and Y_t coincide. However, their path properties can be different. If the process Y_t in Definition 2.8 has continuous paths, then it is called a *continuous path modification* of X_t.

Theorem 2.9 (Kolmogorov's continuity theorem)**.** *Suppose that the process X_t, $t \geqslant 0$, satisfies the following assumption: for all $\Lambda > 0$ there exist positive constants M, α, and β such that*

$$\mathbb{E}[|X_s - X_t|^\alpha] \leqslant M|s - t|^{1+\beta},$$

for all $0 \leqslant s, t \leqslant \Lambda$. Then, there exists a continuous path modification of X_t.

Proof. Without proof. A proof can be found in the book by Kunita, H. "Stochastic flows and stochastic differential equations", 1997, p. 31. □

Examples of basic stochastic processes

Brownian motion. Let $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ be a filtered probability space.

Definition 2.10. *A **(standard, one-dimesional) Brownian motion** is a continuous $\mathcal{F}_t$-adapted stochastic process B_t, $t \geqslant 0$, defined on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ and possessing the following properties:*

1. *$B_0 = 0$ a.s.*

2. *For all $0 \leqslant s < t$, the increment $B_t - B_s$ is independent of $\mathcal{F}_s$.*

3. *For all $0 \leqslant s < t$, the increment $B_t - B_s$ is a Gaussian random variable with mean zero and variance $t - s$*

Poisson point process. A Poisson process N_t on the interval $[0, \infty)$ counts a number of times some premitive event has occured during the time interval $[0, t]$. More precisely, let $(\Omega, \mathcal{F}, P)$ be a probability space. A *Poisson process* with the parameter λ is a process with the following properties:

1. $N_0 = 0$ a.s.

2. N_t is a process ith independent increments, i.e. for all $0 \leqslant t_1 \leqslant \cdots \leqslant t_k$, the random variables N_{t_1}, $N_{t_2} - N_{t_1}$, ..., $N_{t_k} - N_{t_{k-1}}$ are independent.

3. For any $0 \leqslant s < t < \infty$, the random variable $N_t - N_s$ has the Poisson distribution with the parameter $\lambda(t - s)$, i.e.

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

3. Brownian Motion

3.1 Existence of a Brownian motion

A Brownian motion is a stochastic process named after the Scottish botanist Robert Brown who observed that pollen grains suspended in liquid permormed an irregular motion. To describe this motion mathematically it is natural to introduce a stochastic process $B_t(\omega)$ and interpret it as the position of the pollen grain ω at time t.

We intend to specify a family $\{\nu_{t_1,\dots,t_k}\}$ of probability measures satisfying the consitency conditions (1) and (2) of the Kolmogorov entension theorem. The latter theorem will imply the existence of a stochstic processes with the finite-dimensional distributions $\{\nu_{t_1,\dots,t_k}\}$. Fix an $x \in \mathbb{R}^n$, and for all $y \in \mathbb{R}^n$ and $t > 0$ define a function:

$$p(t, x, y) = \frac{1}{(2\pi t)^{\frac{n}{2}}} \exp\Big(-\frac{|x-y|^2}{2t}\Big). \tag{11}$$

If $t = 0$ we set $p(0, x, y) = \delta_x(y)$. The $\delta_x(y)$ is the Dirac measure, i.e. $\delta_x(A) = 1$ if $x \in A$ and $\delta_x(A) = 0$ if $x \notin A$, where $A \subset \mathbb{R}$ is a Borel set.

Lemma 3.1 (Chapman–Kolmogorov identity for the heat kernel)**.** *It holds that*

$$\int_{\mathbb{R}^n} p(s, x, z)p(t, z, y)\, dz = p(s + t, x, y).$$

Proof. A direct computation yields

$$\int_{\mathbb{R}^n} p(s,x,z)p(t,z,y)\,dz = \frac{1}{(2\pi)^n (st)^{n/2}} \int_{\mathbb{R}^n} \exp\left(-\frac{|x-z|^2}{2s} - \frac{|z-y|^2}{2t}\right) dz.$$

Completing the square in the exponent, we obtain

$$-\frac{|x-z|^2}{2s} - \frac{|z-y|^2}{2t} = -\frac{|x-y|^2}{2(s+t)} - \frac{1}{2}\left(\frac{1}{s}+\frac{1}{t}\right)\left|z - \frac{tx+sy}{s+t}\right|^2.$$

Let r be defined by

$$\frac{1}{r} = \frac{1}{s} + \frac{1}{t}, \qquad \text{so that} \qquad r = \frac{st}{s+t}.$$

Since

$$\frac{1}{(2\pi r)^{\frac{n}{2}}} \int_{\mathbb{R}^n} \exp\left(-\frac{\left|z - \frac{tx+sy}{s+t}\right|^2}{2r}\right) du = 1,$$

we conclude that

$$\int_{\mathbb{R}^n} p(s,x,z)p(t,z,y)\,dz = \frac{1}{(2\pi)^{n/2}(s+t)^{n/2}} \exp\left(-\frac{|x-y|^2}{2(s+t)}\right) = p(s+t,x,y).$$

□

Let $0 < t_1 < t_2 < \cdots < t_k$, $F_i \in \mathcal{B}(\mathbb{R}^n)$, $i = 1,\dots,k$. We define a measure $\nu_{t_1,\dots,t_k}$ on $\mathbb{R}^{nk}$ by

$$\begin{aligned}&\nu_{t_1,\dots,t_k}(F_1 \times \cdots \times F_k)\\ &= \int_{F_1} p(t_1,x,x_1)\,dx_1 \int_{F_2} p(t_2-t_1,x_1,x_2)\,dx_2 \dots \int_{F_k} p(t_k - t_{k-1}, x_{k-1}, x_k)\,dx_k. \quad (12)\end{aligned}$$

If the sequence of $\{t_1, t_2, \dots, t_k\}$ is not arranged in the ascending order, then we find a permutation σ that puts $\{t_1, t_2, \dots, t_k\}$ in the the ascending order, and define $\nu_{t_1,\dots,t_k}$ by formula (1) of the Kolmogorov extension theorem.

If $\{0, t_1, t_2, \dots, t_k\}$ are not necessary different numbers, then we define $\nu_{t_1,\dots,t_k}$ as follows. If $t_{i-1} = t_i$, $i = 1,\dots,k$, $t_0 = 0$, we set

$$\nu_{t_1,\dots,t_k}(F_1 \times \cdots \times F_k) := \nu_{t_1,\dots,t_{i-1},t_i,\dots,t_k}(F_1 \times \cdots \times F_i \cap F_{i+1} \times F_{i+2} \times \cdots \times F_k).$$

This procedure leads to the same outcome as if we replace $p(t_i - t_{i-1}, x_{i-1}, x_i)dx_i$ with $\delta_{x_{i-1}}(dx_i)$. In a similar manner we exclude all repetated numbers from the sequence $\{0, t_1, t_2, \dots, t_k\}$.

Thus, by construction, the system of measures $\nu_{t_1,\dots,t_k}$ satisfies consistency condition (1) of the latter theorem. By Lemma 3.1 and since $\int_{\mathbb{R}^n} p(t,x,y)\,dy = 1$, condition (2) of the Kolmogorov extension theorem holds as well (a verification is immediate).

Now the Kolmogorov extension theorem implies the existence of a probability space $(\Omega, \mathcal{F}, \mathbb{P}^x)$ and a stochastic process B_t, $t \geqslant 0$, on it whose finite-dimensional distributions are the measures $\nu_{t_1,\dots,t_k}$. Specifically, if $0 \leqslant t_1 < t_2 \leqslant \dots < t_k$, then

$$\begin{aligned}&\mathbb{P}^x(B_{t_1} \in F_1, \dots, B_{t_k} \in F_k)\\ &= \int_{F_1} p(t_1, x, x_1)\, dx_1 \int_{F_2} p(t_2 - t_1, x_1, x_2)\, dx_2 \dots \int_{F_k} p(t_k - t_{k-1}, x_{k-1}, x_k)\, dx_k. \quad (13)\end{aligned}$$

The process B_t constructed above is not unique. There exists several quadruples $(B_t, \Omega, \mathcal{F}, \mathbb{P}^x)$ such that relation (13) holds. However, it is possible to choose a version of the process B_t with a.s. continuous paths. For this purpose we have to verify Kolmogorov's continuity theorem. We are going to use the formula

$$\mathbb{E}^x[|B_t - B_s|^4] = n(n+2)|t-s|^2 \quad (14)$$

which follows from some properties of the process B_t that we will prove below. The symbol $\mathbb{E}^x$ denotes the expectation with respect to $\mathbb{P}^x$. Kolmogorov's continuity theorem and (14) imply that there is a version of B_t with a.s. continuous paths.

Definition 3.2. *A version of B_t, constructed above, with a.s. continuous paths is called a* ***Brownian motion*** *starting at x.*

Definition 3.3. *The filtration $\mathcal{G}_t := \sigma\{B_s, s \in [0,t]\}$ is called the* ***natural filtration*** *of B_t.*

Theorem 3.4. *Let $n = 1$. Then, the Brownian motion in the sense of Definition 3.2 is also a Brownian motion in the sense of Defininition 2.10 plus x.*

Theorem 3.5 (Approximation of a σ-algebra by an algebra)**.** *Let (X, Σ, μ) be a measure space with $\mu(X) < \infty$, and suppose that*

$$\Sigma = \sigma(\mathcal{A}),$$

where $\mathcal{A}$ is an algebra of subsets of X. Then for every $E \in \Sigma$ and every $\varepsilon > 0$, there exists $A \in \mathcal{A}$ such that

$$\mu(E \triangle A) < \varepsilon,$$

where $E \triangle A = (E \setminus A) \cup (A \setminus E)$ denotes the symmetric difference.

Lemma 3.6. *Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}^n$ be a random variable and let*

$$\mu_X(B) := \mathbb{P}(X \in B) = \mathbb{P} \circ X^{-1}(B), \qquad B \in \mathcal{B}(\mathbb{R}^n),$$

be its distribution. Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a Borel measurable function such that $\varphi(X)$ is integrable. Then,

$$\mathbb{E}[\varphi(X)] = \int_{\mathbb{R}^n} \varphi(x)\, \mu_X(dx).$$

Proof. First assume that $\varphi = \mathbb{1}_B$ for some $B \in \mathcal{B}(\mathbb{R})$. Then

$$\mathbb{E}[\mathbb{1}_B(X)] = \mathbb{P}(X \in B) = \mu_X(B) = \int_{\mathbb{R}} \mathbb{1}_B(x)\, \mu_X(dx).$$

By linearity, the equality holds for every simple function $\varphi(x) = \sum_{k=1}^m a_k\, \mathbb{1}_{B_k}(x)$

Now let $\varphi \geqslant 0$ be Borel measurable. There exists a sequence of simple functions $\varphi_n \uparrow \varphi$ pointwise, see, e.g., (4). By the Monotone Convergence Theorem,

$$\mathbb{E}[\varphi(X)] = \lim_{n\to\infty} \mathbb{E}[\varphi_n(X)] = \lim_{n\to\infty} \int_{\mathbb{R}} \varphi_n(x)\, \mu_X(dx) = \int_{\mathbb{R}} \varphi(x)\, \mu_X(dx).$$

Finally, for general integrable φ, write $\varphi = \varphi^+ - \varphi^-$ and apply the previous step to φ^+ and φ^- separately. □

Remark 2. In particular, the conclusion holds when $\mu = \mathbb{P}$ is any probability measure.

Lemma 3.7. *For any sequence $0 < t_1 < \cdots < t_k$, $\nu_{t_1,\ldots,t_k}$ is a Borel measure on $\mathbb{R}^{nk}$, absolutely continuous with respect to Lebesgue measure whose density is*

$$f_{t_1,\ldots,t_k}(x, x_1, \ldots, x_k) = \prod_{i=1}^{k} p(t_i - t_{i-1}, x_{i-1}, x_i), \qquad x_0 = x. \tag{15}$$

where $p(t,\cdot,\cdot)$ is given by (11). Moreover, $\nu_{t_1,\ldots,t_k}$ is a distribution of the $\mathbb{R}^{nk}$-valued random variable

$$Z := (B_{t_1}, \ldots, B_{t_k}), \qquad \text{i.e.} \quad \nu_{t_1,\ldots,t_k} = \mathbb{P}^x \circ Z^{-1}.$$

Proof. We define two probability measures on $\mathcal{B}(\mathbb{R}^{nk})$.

$$\mu(E) := \mathbb{P}^x\big(Z \in E\big) = \mathbb{P}^x \circ Z^{-1}(E), \qquad E \in \mathcal{B}(\mathbb{R}^{nk}).$$

Also define

$$\nu(E) := \int_E f_{t_1,\ldots,t_k}(x, x_1, \ldots, x_k) dx_1 \ldots dx_k, \qquad E \in \mathcal{B}(\mathbb{R}^{nk}),$$

where $f_{t_1,\ldots,t_k}(x, x_1, \ldots, x_k)$ is given by (15), and note that ν is also a probability measure on $\mathcal{B}(\mathbb{R}^{nk})$.

By (12) and the construction of B_t,

$$\begin{aligned}\nu(F_1 \times \ldots \times F_k) = \int_{F_1\times\ldots\times F_k} f_{t_1,\ldots,t_k}(x, x_1, \ldots, x_k) dx_1 \ldots dx_k &= \mathbb{P}^x(B_{t_1} \in F_1, \ldots, B_{t_k} \in F_k) \\ &= \mathbb{P}^x \circ Z^{-1}(F_1 \times \ldots \times F_k) = \mu(F_1 \times \ldots \times F_k),\end{aligned}$$

for all Borel sets $F_1, F_2, \ldots, F_k \subset \mathbb{R}^{nk}$. Sets of the form $F_1 \times \ldots \times F_k$ generate the Borel σ-algebra $\mathcal{B}(\mathbb{R}^{nk})$. Therefore,

$$\mu(E) = \nu(E) \quad \text{for all } E \in \mathcal{B}(\mathbb{R}^{nk}).$$

□

Proof of Theorem 3.4. Let B_t be a Brownian motion in the sense of Definition 3.2, i.e., constructed above. Note that $\mathbb{P}^x(B_0 = x) = 1$. Indeed, for any $A \subset \mathbb{R}$ Borel set, $\mathbb{P}^x(B_0 \in A) = 1$ if $x \in A$ and $\mathbb{P}^x(B_0 \in A) = 0$ if $x \neq A$. Hence, B_t verifies item 1 of Defininition 2.10.

Let us show that B_t verifies item 3 of Defininition 2.10. By Lemma 3.7, the law of (B_s, B_t) is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^2$ and has density

$$f_{s,t}(y, z) = p(s, x, y)\, p(t - s, y, z),$$

that is,

$$\mathbb{P}^x\big((B_s, B_t) \in E\big) = \int_E p(s, x, y)\, p(t - s, y, z)\, dy\, dz, \qquad E \in \mathcal{B}(\mathbb{R}^2).$$

Define $\phi : \mathbb{R}^2 \to \mathbb{R}$ by $\phi(y, z) = z - y$. By Lemma 3.6, For every Borel set $A \subset \mathbb{R}$,

$$\begin{aligned}\mathbb{P}^x(B_t - B_s \in A) = \mathbb{P}^x\big(\phi(B_s, B_t) \in A\big) &= \mathbb{E}^x[\mathbb{1}_A(\phi(B_s, B_t))] \\ &= \int_{\mathbb{R}^2} \mathbb{1}_A(y - z) p(s, x, y)\, p(t - s, y, z) dy dz.\end{aligned}$$

Perform the change of variables $w = z - y$. Then

$$\int_{\mathbb{R}} \mathbb{1}_A(z - y)\, p(t - s, y, z)\, dz = \int_A p(t - s, 0, w)\, dw.$$

Therefore,

$$\begin{aligned}\int_{\mathbb{R}^2} \mathbb{P}^x(B_t - B_s \in A) &= \int_{\mathbb{R}^2} \mathbb{1}_A(z - y)\, p(s, x, y)\, p(t - s, y, z)\, dz\, dy \\ &= \int_{\mathbb{R}} p(s, x, y)\, dy \int_A p(t - s, 0, w)\, dw = \int_A p(t - s, 0, w)\, dw \qquad (16)\end{aligned}$$

since $\int_{\mathbb{R}} p(s, x, y)\, dy = 1$.

It remains to show that B_t verifies item 2 of Defininition 2.10. Let $0 \leqslant r_1 < \cdots < r_k \leqslant s < t$ and let $D \subset \mathcal{B}(\mathbb{R}^k)$ $A \in \mathcal{B}(\mathbb{R})$. Define

$$\begin{aligned}&X := (B_{r_1}, \ldots, B_{r_k}), \\ &E := X^{-1}(D) = \{\omega \in \Omega : X(\omega) \in D\} \in \sigma(B_{r_1}, \ldots, B_{r_k}), \\ &Z := B_t - B_s, \\ &C := Z^{-1}(A) = \{\omega \in \Omega : Z(\omega) \in A\} \in \sigma(Z).\end{aligned} \qquad (17)$$

We prove that

$$\mathbb{P}^x(E \cap C) = \mathbb{P}^x(E)\, \mathbb{P}^x(C). \qquad (18)$$

By Lemma 3.7, the distribution of the $\mathbb{R}^{n+2}$-valued random variable

$$(B_{r_1}, \ldots, B_{r_k}, B_s, B_t) = (X, Z)$$

is

$$p(r_1, x, x_1)\, p(r_2 - r_1, x_1, x_2) \cdots p(s - r_k, x_k, y)\, p(t - s, y, z) dx_1 \dots dx_k dy dz$$

If $r_k = s$, instead of $p(s - r_k, x_k, y)dy$ we write $\delta_{x_k}(dy)$. By (17) and Lemma 3.6,

$$\mathbb{P}^x(E \cap C) = \mathbb{E}^x[\mathbb{1}_D(X)\, \mathbb{1}_A(B_t - B_s)] = \int_{\mathbb{R}^{k+2}} \mathbb{1}_D(x_1, \dots, x_k)\, \mathbb{1}_A(y - z)$$
$$\times \Big(\prod_{i=1}^{k} p(r_i - r_{i-1}, x_{i-1}, x_i) \Big) p(s - r_k, x_n, y)\, p(t - s, y, z)\, dx_1 \cdots dx_k\, dy dz$$
$$= \int_D \prod_{i=1}^{k} p(r_i - r_{i-1}, x_{i-1}, x_i)\, dx_1 \dots dx_k \int_{\mathbb{R}^2} \mathbb{1}_A(z-y) p(s - r_k, x_k, y)\, p(t-s, y, z) dy dz$$
$$= \Big(\int_D \prod_{i=1}^{k} p(r_i - r_{i-1}, x_{i-1}, x_i)\, dx_1 \dots dx_k \Big) \int_A p(t - s, 0, w) dw$$
$$\mathbb{P}^x(X \in D)\, \mathbb{P}^x(B_t - B_s \in A) = \mathbb{P}^x(E)\, \mathbb{P}^x(C),$$

where $x_0 := x$ and and we have used (16). This proves that $\sigma(B_t - B_s)$ is independent of $\sigma\{B_{r_1}, \dots, B_{r_k}\}$.

Let $\mathcal{F}_s := \sigma\{B_r, r \in [0, s]\}$. Define the class

$$\mathfrak{C}_s := \{E \in \mathcal{F}_s : \mathbb{P}^x(E \cap C) = \mathbb{P}^x(E)\, \mathbb{P}^x(C) \text{ for all sets } C \in \sigma(Z)\}$$

and introduce the algebra

$$\mathcal{A}_s := \bigcup_{n \geqslant 1} \bigcup_{0 \leqslant r_1 < \dots < r_n \leqslant s} \sigma\big(B_{r_1}, \dots, B_{r_n}\big).$$

To see that $\mathcal{A}_s$ is an algebra, we note that $\Omega, \varnothing \in \mathcal{A}_s$ and $\mathcal{A}_s$ is clearly closed under taking complements. Take $A \in \sigma(B_{r_1}, \dots, B_{r_n})$ and $B \in \sigma(B_{u_1}, \dots, B_{u_m})$ with

$$0 \leqslant r_1 < \dots < r_n \leqslant s \quad \text{and} \quad 0 \leqslant u_1 < \dots < u_m \leqslant s.$$

Let $t_1 < \dots < t_k$ be the increasing rearrangement of the finite set

$$\{r_1, \dots, r_n, u_1, \dots, u_m\} \subset [0, s].$$

Then

$$\sigma(B_{r_1}, \dots, B_{r_n}) \subset \sigma(B_{t_1}, \dots, B_{t_k}), \qquad \sigma(B_{u_1}, \dots, B_{u_m}) \subset \sigma(B_{t_1}, \dots, B_{t_k}).$$

Therefore, $A, B \in \sigma(B_{t_1}, \dots, B_{t_k})$ and $A \cup B \in \sigma(B_{t_1}, \dots, B_{t_k})$ because the latter is a σ-albebra. By (18), for any set $E \in \mathcal{A}_s$ it holds that

$$\mathbb{P}^x(E \cap C) = \mathbb{P}^x(E)\, \mathbb{P}^x(C) \text{ for all sets } C \in \sigma(Z)$$

Now let $E \in \mathcal{F}_s$. Note that

$$\mathcal{F}_s = \sigma(\mathcal{A}_s).$$

Therefore, there exist a sequence $\{E_n\}_{n=1}^{\infty}$ of subsets of $\mathcal{A}_s$ with the property

$$\mathbb{P}^x(E_n \triangle E) < \frac{1}{n}$$

This implies that

$$|\mathbb{P}^x(E) - \mathbb{P}^x(E_n)| \leqslant \mathbb{P}^x(E_n \triangle E) \to 0 \qquad \text{as } n \to \infty.$$

Furthermore, for any $C \in \sigma(Z)$ (as well as any other $C \in \mathcal{F}$),

$$(E_n \cap C) \triangle (E \cap C) = (E_n \triangle E) \cap C$$

and hence

$$|\mathbb{P}^x(E_n \cap C) - \mathbb{P}^x(E \cap C)| \leqslant \mathbb{P}^x(E_n \triangle E) \to 0 \qquad \text{as } n \to \infty.$$

We obtain

$$\mathbb{P}^x(E \cap C) = \lim_{n\to\infty} \mathbb{P}^x(E_n \cap C) = \lim_{n\to\infty} \mathbb{P}^x(E_n)\, \mathbb{P}^x(C) = \mathbb{P}^x(E)\, \mathbb{P}^x(C),$$

for any set $C \in \sigma(Z)$. Therefore,

$$\mathcal{F}_s = \mathfrak{C}_s.$$

This immediately implies that $\mathcal{F}_s$ is independent of $\sigma(Z)$, which is the same as $B_t - B_s$ is independent of $\mathcal{F}_s$. □

Corollary 3.8. *$\sigma\{B_t - B_s\}$ is independent of $\sigma\{B_{r_1}, \ldots, B_{r_k}\}$ for all $k \in \mathbb{N}$ and all $0 \leqslant r_1 < \cdots < r_k \leqslant s$.*

Our arguments can be summarized in the theorem below:

Theorem 3.9. *A Brownian motion exists.*

3.2 Some properties of a Brownian motion

Proposition 3.10. *B_t possesses the following properties:*

$$\begin{aligned}
&\mathbb{E}^x[B_t] = x,\\
&\mathbb{E}^x[\,|B_t - x|^2] = nt,\\
&\mathbb{E}^x[\,(B_t - x, B_s - x)_{\mathbb{R}^n}] = n\min\{s,t\},\\
&\mathbb{E}^x[\,|B_t - B_s|^2] = n|t-s|.
\end{aligned}$$

Proof. It suffices to prove the above properties for the one-dimensional case. Therefore, without loss of generality we assume that the process B_t is one-dimensional. In the following computation we assume that $s < t$. We obtain:

$$\mathbb{E}^x[B_t] = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} y\, e^{-\frac{(y-x)^2}{2t}}\, dy = x,$$

$$\mathbb{E}^x[\,(B_t - x)^2] = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} (y-x)^2\, e^{-\frac{(y-x)^2}{2t}}\, dy = t,$$

$$\begin{aligned}
&\mathbb{E}^x[\,(B_t - x)(B_s - x)] \\
&= \frac{1}{\sqrt{2\pi s}} \int_{-\infty}^{\infty} (y_1 - x)\, e^{-\frac{(y_1-x)^2}{2s}}\, dy_1 \cdot \frac{1}{\sqrt{2\pi(t-s)}} \int_{-\infty}^{\infty} (y_2 - x)\, e^{-\frac{(y_2-y_1)^2}{2(t-s)}}\, dy_2 \\
&= \frac{1}{\sqrt{2\pi s}} \int_{-\infty}^{\infty} (y_1 - x)^2\, e^{-\frac{(y_1-x)^2}{2s}}\, dy_1 = s,
\end{aligned}$$

$$\mathbb{E}^x[\,(B_t - B_s)^2] = \mathbb{E}^x[(B_t - x)^2] + \mathbb{E}^x[\,(B_s - x)^2] - 2\,\mathbb{E}^x[\,(B_t - x)(B_s - x)] = t - s.$$

□

Definition 3.11. A stochastic process X_t is called **stationary**, if for every $n \in \mathbb{N}$, every choice of times $t_1, \ldots, t_n \in \mathbb{R}$, and every $h \in \mathbb{R}$

$$(X_{t_1}, \ldots, X_{t_n}) \quad \text{and} \quad (X_{t_1+h}, \ldots, X_{t_n+h}),$$

have the same distribution.

Definition 3.12. A stochastic process $(X_t)_{t \geqslant 0}$ is said to have *stationary increments* if for every $n \in \mathbb{N}$, every $0 \leqslant t_1 < \cdots < t_n$, and every $h \geqslant 0$,

$$(X_{t_1+h} - X_h, \ldots, X_{t_n+h} - X_h) \quad \text{and} \quad (X_{t_1} - X_0, \ldots, X_{t_n} - X_0).$$

have the same distribution.

Proposition 3.13. *B_t has **stationary increments**. In particular, for any fixed t, the processes $\{B_{t+h} - B_h\}_{h>0}$, have the same distibution as $B_t - x$.*

Proof. Left as an exercise. □

For the subsequent property of a Brownian motion we will need the theorem below.

Theorem 3.14. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let X and Y be $L^1(\Omega)$-random variables. If X and Y are independent (that is, the σ-algebras $\sigma(X)$ and $\sigma(Y)$ are independent), then*

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

Proof. We proceed in several steps. *Step 1: Indicator functions.* Let $A \in \sigma(X)$ and $B \in \sigma(Y)$. By independence of the σ-algebras,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Therefore,

$$\mathbb{E}[\mathbb{1}_A\,\mathbb{1}_B] = \mathbb{E}[\mathbb{1}_{A\cap B}] = \mathbb{P}(A\cap B) = \mathbb{P}(A)\mathbb{P}(B) = \mathbb{E}[\mathbb{1}_A]\mathbb{E}[\mathbb{1}_B].$$

Step 2: Simple functions. Let

$$X = \sum_{i=1}^{n} a_i\,\mathbb{1}_{A_i}, \qquad Y = \sum_{j=1}^{m} b_j\,\mathbb{1}_{B_j},$$

where $A_i \in \sigma(X)$ and $B_j \in \sigma(Y)$. Then

$$XY = \sum_{i,j} a_i b_j\,\mathbb{1}_{A_i\cap B_j}\,.$$

Taking expectations and using Step 1,

$$\mathbb{E}[XY] = \sum_{i,j} a_i b_j \mathbb{P}(A_i\cap B_j) = \sum_{i,j} a_i b_j \mathbb{P}(A_i)\mathbb{P}(B_j) = \mathbb{E}[X]\mathbb{E}[Y].$$

Step 3: Nonnegative measurable functions. Assume $X \geqslant 0$ and $Y \geqslant 0$. There exist increasing sequences of simple functions (X_n) and (Y_n) such that

$$X_n \uparrow X, \qquad Y_n \uparrow Y.$$

See, e.g, (4). By Step 2,

$$\mathbb{E}[X_nY_n] = \mathbb{E}[X_n]\mathbb{E}[Y_n].$$

By the Monotone Convergence Theorem,

$$\mathbb{E}[XY] = \lim_{n\to\infty} \mathbb{E}[X_nY_n] = \lim_{n\to\infty} \mathbb{E}[X_n]\mathbb{E}[Y_n] = \mathbb{E}[X]\mathbb{E}[Y].$$

Step 4: General integrable random variables. For general integrable X and Y, write

$$X = X^+ - X^-, \qquad Y = Y^+ - Y^-.$$

Each of the random variables $X^\pm$ and $Y^\pm$ is nonnegative and independent. Applying Step 3 and expanding bilinearly yields

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

□

Proposition 3.15. *B_t has **independent increments**, i.e. for all $0 \leqslant t_1 \leqslant \cdots \leqslant t_k$ the random variables*

$$B_{t_1} - x, B_{t_2} - B_{t_1}, \ldots, B_{t_k} - B_{t_{k-1}} \tag{19}$$

are independent.

Proof. Let $H_i \in \sigma\{B_{t_i} - B_{t_{i-1}}\} \subset \sigma\{B_{t_i}, B_{t_{i-1}}\}$, $i = 1, \dots, k$. Note that $B_0 = x$ a.s. By Corollary 3.8, $\sigma\{B_{t_k} - B_{t_{k-1}}\}$ is independent of $\sigma\{B_{t_{k-1}}, \dots, B_{t_1}\}$. Therefore,

$$\mathbb{P}(H_1 \cap \dots H_{k-1} \cap H_k) = \mathbb{P}(H_1 \cap \dots H_{k-1})\mathbb{P}(H_k).$$

The statement can be proved by induction or continuing the same argument. □

Note that Proposition 3.15 were proved without any assumption on the continuity of paths of B_t. In Section 4.6 we left formula (14) without proof. Now we prove this formula.

Proposition 3.16. *Equality* (14) *holds.*

Proof. Let $t > s \geqslant 0$, and let $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. We have

$$\begin{aligned}
\mathbb{E}^x[|B_t - B_s|^4] == \frac{1}{(2\pi(t-s))^{\frac{n}{2}}} \int_{\mathbb{R}^n} |x|^4 \, e^{-\frac{|x|^2}{2(t-s)}} \, dx \\
= \sum_{i=1}^{n} \frac{1}{(2\pi(t-s))^{\frac{1}{2}}} \int_{-\infty}^{\infty} x_i^4 \, e^{-\frac{x_i^2}{2(t-s)}} \, dx_i \\
+ \sum_{i \neq j} \frac{1}{(2\pi(t-s))^{\frac{1}{2}}} \int_{-\infty}^{\infty} x_i^2 \, e^{-\frac{x_i^2}{2(t-s)}} \, dx_i \, \frac{1}{(2\pi(t-s))^{\frac{1}{2}}} \int_{-\infty}^{\infty} x_j^2 \, e^{-\frac{x_j^2}{2(t-s)}} \, dx_j \\
= 3n(t-s)^2 + n(n-1)(t-s)^2 = n(n+2)(t-s)^2.
\end{aligned}$$

We just used the following facts: $|x|^4 = (x_1^2 + \dots + x_n^2)^2 = \sum_{i=1}^n x_i^4 + \sum_{i \neq j} x_i^2 x_j^2$, $\frac{1}{(2\pi(t-s))^{\frac{1}{2}}} \int_{-\infty}^{\infty} x_i^2 \, e^{-\frac{x_i^2}{2(t-s)}} \, dx_i = (t-s)$, as well as the computation of the integral below:

$$\begin{aligned}
\frac{1}{(2\pi(t-s))^{\frac{1}{2}}} \int_{-\infty}^{\infty} x_i^4 \, e^{-\frac{x_i^2}{2(t-s)}} \, dx_i = -\frac{1}{(2\pi(t-s))^{\frac{1}{2}}} (t-s) \int_{-\infty}^{\infty} x_i^3 d\Big(e^{-\frac{x_i^2}{2(t-s)}}\Big) \\
= -\frac{1}{(2\pi(t-s))^{\frac{1}{2}}} (t-s) \, x_i^3 \, e^{-\frac{x_i^2}{2(t-s)}} \Big|_{-\infty}^{\infty} + 3(t-s) \frac{1}{(2\pi(t-s))^{\frac{1}{2}}} \int_{-\infty}^{\infty} x_i^2 \, e^{-\frac{x_i^2}{2(t-s)}} \, dx_i \\
= 3(t-s)^2.
\end{aligned}$$

□

3.3 Local properties of a Brownian path

Here we list some properties of a Brownian path without proof. The proofs can be found in the books of Karatzas and Shreve (1991), Revuz and Yor (2001).

Hölder continuity. For almost all $\omega \in \Omega$, the function $t \mapsto B_t$ is locally Hölder continuous of order α for every $\alpha < \frac{1}{2}$. In other words, for all $T > 0$, $0 < \alpha < \frac{1}{2}$, and for almost all $\omega \in \Omega$ there exists a constant $C_{T,\alpha}(\omega)$ such that for all $s, t \in [0, T)$

$$|B_t(\omega) - B_s(\omega)| \leqslant C_{T,\alpha}(\omega) \, |t - s|^{\alpha}.$$

Modulus of continuity. For Brownian paths

$$\overline{\lim}_{\delta\to 0} \sup_{\substack{s,t<T:\\ |t-s|<\delta}} \frac{|B_t - B_s|}{\sqrt{2\delta \ln(1/\delta)}} = 1 \quad \text{a.s.}$$

In general the *modulus of continuity* of a continuous function f on $[0,T]$ is a function $\varepsilon_f(\delta)$ defined by the formula

$$\varepsilon_f(\delta) = \sup\{|f(t) - f(s)| : t, s \in [0,T),\ |t-s| < \delta\}$$

for sufficiently small δ. Thus, for almost every Brownian path $B_\bullet(\omega)$

$$\varepsilon_{B_\bullet(\omega)}(\delta) \leqslant \sqrt{2\delta \ln(1/\delta)}$$

for all $\delta < \delta_0$ where δ_0 is sufficiently small.

Nowhere differentiability. Brownian paths are a.s. nowhere locally Hölder continuous of order $\alpha \geqslant \frac{1}{2}$. In particular, Brownian paths are nowhere differentiable.

Infinite variation. Brownian paths are of infinite variation on any interval $[s,t]$ a.s., i.e. a.s.

$$\sup \sum_{i=1}^{n} |B_{t_i} - B_{t_{i-1}}| = \infty$$

where the supremum is taken over all subdivisions $s \leqslant t_1 \leqslant \cdots \leqslant t_n \leqslant t$ of the interval $[s,t]$.

4. Markov Processes

Intuitevely speaking, the process X_t is Markov if, to make a prediction at time t on what is going to happen in the future, it is useless to know the whole past up to time t, but only the present state X_t at time t.

4.1 Continuous-time Markov processes

Definition 4.1. *Let $\mathcal{B}$ be the Borel σ-algebra of subsets of $\mathbb{R}^n$. A function $\pi : \mathbb{R}^n \times \mathcal{B} \to [0,1]$ is called a **transition probability** if*

1. *for every $x \in \mathbb{R}^n$, the map $A \mapsto \pi(x,A)$ is a probability measure on $\mathbb{R}^n$;*
2. *for every $A \in \mathcal{B}$, the map $x \mapsto \pi(x,A)$ is $\mathcal{B}$-measurable.*

Definition 4.2. *A **transition function** on $(\mathbb{R}^n, \mathcal{B})$ is a family $P_{s,t}$, $0 \leqslant s < t$, of transition probabilities on $(\mathbb{R}^n, \mathcal{B})$ such that for every three numbers $s < r < t$,*

$$\int_{\mathbb{R}^n} P_{r,t}(y,A) P_{s,r}(x,dy) = P_{s,t}(x,A) \tag{20}$$

for every $x \in \mathbb{R}^n$ and for every $A \in \mathcal{B}$.

Relation (20) is called the *Chapman-Kolmogorov* equation. The transition function is said to be *homogeneous* if $P_{s,t}$ depends on s and t only through the difference $t-s$, i.e. $P_{s,t} = P_{s-t}$, where P_t is the notation for $P_{0,t}$. Equation (20) takes the form:

$$P_{s+t}(x, A) = \int P_t(y, A) P_s(x, dy) \tag{21}$$

for every $s, t \geqslant 0$.

Definition 4.3. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A process X_t is a Markov process with transition function $P_{s,t}$ if for any Borel-measurable function $f : \mathbb{R}^n \to \mathbb{R}$,*

$$\mathbb{E}[f(X_t) \,|\, \sigma(X_u, u \leqslant s)] = (P_{s,t} f)(X_s) \quad \mathbb{P} - a.s.$$

In the above definition, $P_{s,t} f$ is a function $\mathbb{R}^n \to \mathbb{R}$, such that for any $x \in \mathbb{R}^n$ it is defined as follows:

$$(P_{s,t} f)(x) := \int_{\mathbb{R}^n} f(y)\, P_{s,t}(x, dy). \tag{22}$$

Remark 3. Intuitevely speaking, if the process is at position x at time s, then $P_{s,t}(x, A)$ is the probability that it will be in the set A at time t.

Once the present position X_s is known, the entire past of the process up to time s is irrelevant for predicting the state of the process at time t.

Definition 4.4. *The process X_t is said to be homogeneous if its transition function is homogeneous.*

For a homogeneous Markov process

$$\mathbb{E}[f(X_t) \,|\, \sigma(X_u, u \leqslant s)] = (P_{t-s} f)(X_s) \quad P - \text{a.s.}$$

In the following, let $\nu = P \circ X_0^{-1}$ denote the initial distribution of the process X_t.

Theorem 4.5. *A process X_t is a Markov process with transition function $P_{s,t}$ and initial measure ν if and only if for any sequence $0 = t_0 < t_1 < \cdots t_k$ and and for any Borel-measurable functions $f_i : \mathbb{R}^n \to \mathbb{R}$, $0 \leqslant i \leqslant k$,*

$$\begin{aligned} &\mathbb{E}\Big[\prod_{i=1}^{k} f_i(X_{t_i})\Big] \\ &= \int_{\mathbb{R}^n} \nu(dx_0) f_0(x_0) \int_{\mathbb{R}^n} P_{0,t_1}(x_0, dx_1) f_1(x_1) \ldots \int_{\mathbb{R}^n} P_{t_{k-1},t_k}(x_{k-1}, dx_k) f_k(x_k). \end{aligned} \tag{23}$$

Proof. We will only prove the "if" statement, i.e. we prove that if the process X_t is Markov, then formula (23) holds. The converse statement, i.e. if formula (23) holds

then the process X_t is Markov, will be left without proof. We have:

$$\mathbb{E}\Big[\prod_{i=0}^{k} f_i(X_{t_i})\Big] = \mathbb{E}\Big[\prod_{i=0}^{k-1} f_i(X_{t_i})\,\mathbb{E}\big[f_k(X_{t_k})\,|\,\sigma(X_s, s\leqslant t_{k-1})\big]\Big]$$
$$= \mathbb{E}\Big[\prod_{i=0}^{k-1} f_i(X_{t_i})\,(P_{t_{k-1},t_k}f_k)(X_{t_{k-1}})\Big] = \mathbb{E}\Big[\prod_{i=0}^{k-2} f_i(X_{t_i})\,f_{k-1}(X_{t_{k-1}})(P_{t_{k-1},t_k}f_k)(X_{t_{k-1}})\Big]. \tag{24}$$

When passing from the first to the second expectation in (24) we used the fact that $f(X_{t_i})$, $0\leqslant i\leqslant k-1$, are $\sigma(X_s, s\leqslant t_{k-1})$-measurable. The latter expectation in (24) is of the product of $k-1$ functions, but has the same form as the first expectation of the product of k functions. The $(k-1)$th function in the latter product is $f_{k-1}\,P_{t_{k-1},t_k}f_k$. We can proceed the same way to come to a product of $k-2$ functions, etc. After the last step we obtain:

$$\mathbb{E}\Big[\prod_{i=0}^{k} f_i(X_{t_i})\Big] = \mathbb{E}\Big[\,f(X_0)P_{0,t_1}(f_1P_{t_1,t_2}(f_2P_{t_2,t_3}\dots f_{k-1}(P_{t_{k-1}t_k}f_k)\dots))(X_0)\Big]. \tag{25}$$

Using equality (22) as well as the fact that for any Borel-measurable function $g:\mathbb{R}^n\to\mathbb{R}$

$$\mathbb{E}[g(X_0)] = \int_\Omega g(X_0(\omega))\mathbb{P}(d\omega) = \int_{\mathbb{R}^n} g(x)\,\nu(dx),$$

we conclude that the right-hand side of (25) equals to the right-hand side of (23). The proof of the converse statement, i.e. if formula (23) holds then the process X_t is Markov, will be left without proof. A proof can be found in the book by D. Revuz and M. Yor "Continuous martingales and Brownian motion", Chapter III. □

Note that identity (23) shows that if we know the transition function of a Markov process, we can construct its finite-dimensional distributions.

Theorem 4.6 (Existence of a Markov process)**.** *Given a transition function $P_{s,t}$ on $(\mathbb{R}^n,\mathcal{B})$ and a probability measure μ on $(\mathbb{R}^n,\mathcal{B})$, there exists a probability space $(\Omega,\mathcal{F},\mathbb{P}^\mu)$ and a stochastic process X_t on it which is Markov with transition function $P_{s,t}$ and initial measure μ.*

Proof. Let us define the finite-dimensional distributions. Fix a sequence $0<t_1<t_2<t_k$, and define:

$$\nu_{t_1,\dots,t_k}(F_1\times\dots\times F_k) = \int_{F_1} P_{0,t_1}(x,dx_1)\int_{F_2} P_{t_1,t_2}(x_1,dx_2)\dots\int_{F_k} P_{t_{k-1},t_k}(x_{k-1},dx_k).$$
$$\nu_{0,t_1,\dots,t_k}(F_0\times F_1\times\dots\times F_k) = \int_{F_0}\mu(dx_0)\int_{F_1} P_{0,t_1}(x_0,dx_1)\int_{F_2} P_{t_1,t_2}(x_1,dx_2)$$
$$\dots\int_{F_k} P_{t_{k-1},t_k}(x_{k-1},dx_k).$$

Let $\{t_1, t_2, \dots, t_k\}$ be an arbitrary sequence (not necessary ascending). Let σ be the permutation of $\{1, \dots, k\}$ such that $t_{\sigma(1)} < t_{\sigma(2)} < \dots < t_{\sigma(k)}$. Define

$$\nu_{t_1,\dots,t_k}(F_1 \times \dots \times F_k) = \nu_{t_{\sigma(1)},\dots,t_{\sigma(k)}}(F_{\sigma(1)} \times \dots \times F_{\sigma(k)}).$$

The latter definition implies consistency condition 1 of the Kolmogorov extension theorem. Further let $\tilde{\sigma}$ be the permutation of $\{1, \dots, k, k+1, \dots, k+m\}$ such that $t_{\tilde{\sigma}(1)} < \dots t_{\tilde{\sigma}(k)} < t_{\tilde{\sigma}(k+1)} < \dots < t_{\tilde{\sigma}(k+m)}$. Define $F_i = \mathbb{R}^n$ for $k+1 \leqslant i \leqslant k+m$. We obtain:

$$\begin{aligned}
&\nu_{t_1,\dots,t_k,t_{k+1},\dots,t_{k+m}}(F_1 \times \dots \times F_k \times \underbrace{\mathbb{R}^n \times \dots \times \mathbb{R}^n}_{m}) = \nu_{t_1,\dots,t_k,t_{k+1},\dots,t_{k+m}}(F_1 \times \dots \times F_{k+m}) \\
&= \nu_{t_{\tilde{\sigma}(1)},\dots,t_{\tilde{\sigma}(k)},t_{\tilde{\sigma}(k+1)},\dots,t_{\tilde{\sigma}(k+m)}}(F_{\tilde{\sigma}(1)} \times \dots \times F_{\tilde{\sigma}(k+m)}) = \nu_{t_{\sigma(1)},\dots,t_{\sigma(k)}}(F_{\sigma(1)} \times \dots \times F_{\sigma(k)}) \qquad (26) \\
&= \nu_{t_1,\dots,t_k}(F_1 \times \dots \times F_k).
\end{aligned}$$

The third equality in (26) holds by the Chapman-Kolmogorov equations. Indeed, some of the sets $F_{\tilde{\sigma}(i)}$, $1 \leqslant i \leqslant k+m$, equal to $\mathbb{R}^n$ (namely, those that do not coincide with one of the sets $F_{\sigma(i)}$, $1 \leqslant i \leqslant k$), and therefore integration over these sets can be excluded. Specifically, we exclude integration over those $F_{\tilde{\sigma}(i)}$ which are equal to $\mathbb{R}^n$ as follows: for any three numbers $t_{i-1} < t_i < t_{i+1}$, for any function $f : \mathbb{R}^n \to \mathbb{R}$, for any Borel subset $F \subset \mathbb{R}^n$,

$$\begin{aligned}
&\int_{\mathbb{R}^n} P_{t_{i-1},t_i}(x_{i-1}, dx_i) \int_F P_{t_i,t_{i+1}}(x_i, dx_{i+1}) f(x_{i+1}) \\
&= \int_F f(x_{i+1}) \int_{\mathbb{R}^n} P_{t_{i-1},t_i}(x_{i-1}, dx_i) P_{t_i,t_{i+1}}(x_i, dx_{i+1}) = \int_F f(x_{i+1}) P_{t_{i-1},t_{i+1}}(x_{i-1}, dx_{i+1})
\end{aligned}$$

and therefore, time t_i and the integration over $\mathbb{R}^n$ are excluded. This proves (26) and consistency condition 2 of the Kolmogorov extension theorem. The latter theorem implies the existence of a probability space and a stochastic processes X_t on it whose finite-dimensional distributions are the measures $\nu_{t_1,\dots,t_k}$. Let us prove that X_t is a Markov process. We will apply Theorem 4.5. Indeed, let $f_i : \mathbb{R}^n \to \mathbb{R}$ be Borel-measurable functions. We obtain:

$$\begin{aligned}
\mathbb{E}\Big[\prod_{i=0}^{k} f_i(X_{t_i})\Big] &= \int_{\mathbb{R}^{nk}} \prod_{i=0}^{k} f_i(x_i)\, \nu_{0,t_1,\dots,t_k}(dx_0 dx_1 \cdots dx_k) \\
&= \int_{\mathbb{R}^n} f_0(x_0)\mu(dx_0) \int_{\mathbb{R}^n} f_1(x_1) P_{0,t_1}(x_0, dx_1) \dots \int_{\mathbb{R}^n} f_k(x_k)\, P_{t_{k-1},t_k}(x_{k-1}, dx_k).
\end{aligned}$$

By Theorem 4.5 the process X_t is Markov. □

4.2 Markov property

Let $(\mathbb{R}^n, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra of subsets of $\mathbb{R}^n$, be a measurable space. Further let $\Omega = (\mathbb{R}^n)^{[0,\infty)} = \{\varphi : [0,\infty) \to \mathbb{R}^n\}$, let $\pi_t : \Omega \to \mathbb{R}^n$, $\pi_t(\varphi) = \varphi(t)$

be the evaluation map, and let the σ-algebra of subsets of Ω be defined as follows: $\mathcal{F} = \big(\mathcal{B}\big)^{[0,\infty)} = \sigma\{F \subset \Omega : \pi_t(F) \in \mathcal{B}\ \forall t \in [0,\infty)\}$. Finally, let X_t be the *coordinate process*, i.e. $X_t(\omega) = \omega(t)$. The theorem below is similar to Theorem 4.6 but states the existence of a probability measure $\mathbb{P}$ and a Markov process on a certain probability space which is $(\Omega, \mathcal{F}, \mathbb{P})$.

Theorem 4.7. *Given a transition function $P_{s,t}$ on $(\mathbb{R}^n, \mathcal{B})$, for any probability measure μ on $(\mathbb{R}^n, \mathcal{B})$, there exists a unique probability measure $\mathbb{P}^\mu$ on $(\Omega, \mathcal{F})$ such that the coordinate process X_t is Markov with transition function $P_{s,t}$ and initial measure μ.*

Proof. Without proof. A proof can be found in the book by D. Revuz and M. Yor "Continuous martingales and Brownian motion", Chapter III. □

Let $(\Omega, \mathcal{F})$ be a measurable space. Note that in Theorem 4.6 to an initial measure μ we associated a probability measure $\mathbb{P}^\mu$. For the probability measure $\mathbb{P}^{\delta_x}$ which is associated to initial measure δ_x we will use the notation $\mathbb{P}^x$. Furthermore, the symbols $\mathbb{E}^\mu$ and $\mathbb{E}^x$ will be used for the expectations relative to the measures $\mathbb{P}^\mu$ and resp. $\mathbb{P}^x$.

Proposition 4.8. *Let Z be an $\mathcal{F}$-measurable random variable which is either positive or bounded. Then,*

$$\mathbb{E}^\mu[Z] = \int_{\mathbb{R}^n} \mu(dx)\, \mathbb{E}^x[Z].$$

Proof. Without proof. A proof can be found in the book by D. Revuz and M. Yor "Continuous martingales and Brownian motion", Chapter III. □

For every $t > 0$, we define the *shift operator* $\theta_t : \Omega \to \Omega$ as follows:

$$\theta_t(\omega) = \omega(t + \,\cdot\,), \quad \text{i.e. } \big(\theta_t(\omega)\big)(s) = \omega(t+s) \quad \forall s > 0.$$

Theorem 4.9 (Markov property)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P}^\mu)$ be a probability space and X_t be a homogeneous Markov process. Further let Z be an $\mathcal{F}$-measurable random variable such that it is either positive or bounded. Then*

$$\mathbb{E}^\mu[Z \circ \theta_t \,|\, \sigma(X_s, s \leqslant t)] = \mathbb{E}^{X_t}[Z].$$

Remark 4. The expectation of the right-hand side of this formula is understood as $\mathbb{E}^x[Z]$ with X_t substituted for x. In other words one can say that it is the composition of two maps: $\omega \mapsto X_t(\omega)$ and $x \mapsto \mathbb{E}^x[Z]$.

Remark 5. The homogenuity of X_t means

$$\begin{aligned}&\mathbb{P}^\mu(X_0 \in A_0, X_{t_1} \in A_1, \dots, X_{t_n} \in A_n)\\ &\qquad = \int_{A_0} \mu(dx) \int_{A_1} P_{t_1}(x, dx_1) \int_{A_2} P_{t_2 - t_1}(x_1, dx_2) \dots \int_{A_n} P_{t_n - t_{n-1}}(x_{n-1}, dx_n).\end{aligned}$$

Idea of proof. By the definition of conditional probability, we have to show that for any subset $A \in \sigma(X_s, s \leqslant t)$,

$$\mathbb{E}^\mu[(Z \circ \theta_t)\, \mathbb{1}_A] = \int_A (Z \circ \theta_t)(\omega)\, \mathbb{P}^\mu(d\omega) = \int_A \mathbb{E}^{X_t}[Z](\omega)\, \mathbb{P}^\mu(d\omega).$$

A more general statement would be to prove

$$\mathbb{E}^\mu[(Z \circ \theta_t)\, Y] = \mathbb{E}^\mu[\mathbb{E}^{X_t}[Z]\, Y]$$

for any $\sigma(X_s, s \leqslant t)$-measurable and positive random variable Y. Consider the case when $Y = \prod_{i=0}^k f_i(X_{t_i})$ and $Z = \prod_{j=1}^m g_j(X_{\tau_j})$, where f_i and g_j are positive Borel-measurable functions and $0 = t_0 < t_1 < \cdots < t_k = t$, $0 < \tau_1 < \cdots < \tau_m$. Without loss of generality we can assume that $t_0 = 0$ and $t_k = t$ multiplying Y by functions identically equal to 1 if necessary. We obtain:

$$\begin{aligned}
\mathbb{E}^\mu\Big[Y\,(Z \circ \theta_t)\Big] &= \mathbb{E}^\mu\Big[\prod_{i=0}^k f_i(X_{t_i})\, \prod_{j=0}^m g_j(X_{t+\tau_j})\Big] \\
&= \int_{\mathbb{R}^n} \mu(dx_0) f_0(x_0) \int_{\mathbb{R}^n} P_{t_1}(x_0, dx_1) f_1(x_1) \ldots \int_{\mathbb{R}^n} P_{t_k - t_{k-1}}(x_{k-1}, dx_k) f_k(x_k) \times \\
&\quad \times \int_{\mathbb{R}^n} P_{t_{k+1} - t}(x_k, dx_{k+1}) g_0(x_{k+1}) \ldots \int_{\mathbb{R}^n} P_{\tau_m - \tau_{m-1}}(x_{m-1}, dx_m) g_m(x_m) \\
&= \int_{\mathbb{R}^n} \mu(dx_0) f_0(x_0) \int_{\mathbb{R}^n} P_{t_1}(x_0, dx_1) f_1(x_1) \ldots \\
&\ldots \int_{\mathbb{R}^n} P_{t - t_{k-1}}(x_{k-1}, dx_k) f_k(x_k)\, \mathbb{E}^{x_k}\Big[\prod_{j=1}^m g(X_{\tau_j})\Big] = \mathbb{E}^\mu\Big[\prod_{i=0}^k f_i(X_{t_i})\, \mathbb{E}^{X_t} \prod_{j=0}^m g_j(X_{\tau_j})\Big] \\
&= \mathbb{E}^\mu\Big[Y\, \mathbb{E}^{X_t}[Z]\Big].
\end{aligned}$$

The general case that deals an arbitrary $\mathcal{F}$-measurable random variable Z and a $\sigma(X_s, s \leqslant t)$-measurable random variable Y we leave without proof. A proof can be found in the book by D. Revuz and M. Yor "Continuous martingales and Brownian motion", Chapter III. □

5. Martingales

5.1 Filtrations and Stopping times

Filtrations play a fundamental role in the theory of stochastic processes, namely, in the definition of the basic objects of our consideration - martingales. In Section 2 we introduced the concept of a *filtration* and defined processed *adapted* with respect to filtration. Let us introduce further definitions. With every filtration $\mathcal{F}_t$ we associate two other filtrations:

$$\mathcal{F}_{t^-} = \sigma\Big(\bigcup_{s<t} \mathcal{F}_s\Big), \qquad \mathcal{F}_{t^+} = \bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon}.$$

The filtration $\mathcal{F}_\infty$ denotes

$$\mathcal{F}_\infty = \sigma\Big(\bigcup_{t<\infty} \mathcal{F}_t\Big)$$

Note that

$$\mathcal{F}_\infty = \sigma\Big(\bigcup_{t<\infty} \mathcal{F}_{t^-}\Big) = \sigma\Big(\bigcup_{t<\infty} \mathcal{F}_{t^+}\Big).$$

By convention, $\mathcal{F}_{0^-} = \mathcal{F}_0$. We always have the inclusion $\mathcal{F}_{t^-} \subseteq \mathcal{F}_t \subseteq \mathcal{F}_{t^+}$.

Definition 5.1. *If $\mathcal{F}_t = \mathcal{F}_{t^+}$, the filtration $\mathcal{F}_t$ is called **right-continuous**. If $\mathcal{F}_t = \mathcal{F}_{t^-}$, the filtration $\mathcal{F}_t$ is called **left-continuous**.*

Let $\mathcal{F}_t$ be a filtration.

Definition 5.2. *A **stopping time** relative to the filtration $\mathcal{F}_t$ is a random variable with values in $[0,\infty)$ such that for every $t \geqslant 0$*

$$\{\omega : T(\omega) \leqslant t\} \in \mathcal{F}_t.$$

Definition 5.3. *Let T be a stopping time.*

$$\mathcal{F}_T := \{A \in \mathcal{F}_\infty : \ A \cap \{T \leqslant t\} \in \mathcal{F}_t \quad \textit{for all } t\}$$

Lemma 5.4. *$\mathcal{F}_T$ is a σ-algebra.*

Proof. We verify the defining properties of a σ-algebra.

(i) $\Omega \in \mathcal{F}_T$. Clearly $\Omega \in \mathcal{F}_\infty$. Moreover, since T is a stopping time,

$$\Omega \cap \{T \leqslant t\} = \{T \leqslant t\} \in \mathcal{F}_t \quad \text{for all } t \geqslant 0.$$

Hence $\Omega \in \mathcal{F}_T$.

(ii) Closed under complements. Let $A \in \mathcal{F}_T$. Then $A^c \in \mathcal{F}_\infty$, and for each $t \geqslant 0$,

$$A^c \cap \{T \leqslant t\} = \{T \leqslant t\} \setminus \big(A \cap \{T \leqslant t\}\big).$$

Since both $\{T \leqslant t\}$ and $A \cap \{T \leqslant t\}$ belong to $\mathcal{F}_t$, and $\mathcal{F}_t$ is a σ-algebra, it follows that $A^c \cap \{T \leqslant t\} \in \mathcal{F}_t$. Thus $A^c \in \mathcal{F}_T$.

(iii) Closed under countable unions. Let $(A_n)_{n\geqslant 1} \subset \mathcal{F}_T$. Then $\bigcup_{n=1}^\infty A_n \in \mathcal{F}_\infty$, and for each $t \geqslant 0$,

$$\left(\bigcup_{n=1}^\infty A_n\right) \cap \{T \leqslant t\} = \bigcup_{n=1}^\infty \big(A_n \cap \{T \leqslant t\}\big).$$

Each set $A_n \cap \{T \leqslant t\}$ belongs to $\mathcal{F}_t$, and since $\mathcal{F}_t$ is closed under countable unions, the right-hand side lies in $\mathcal{F}_t$. Hence $\bigcup_{n=1}^\infty A_n \in \mathcal{F}_T$.

Having verified these properties, we conclude that $\mathcal{F}_T$ is a σ-algebra. □

Let us give examples of stopping times.

Proposition 5.5. *Let $A \subset \mathbb{R}^n$ be an closed set, and let the process X_t have continuous paths. Define*

$$T_A(\omega) = \inf\{t \geqslant 0,\ X_t(\omega) \notin A\}.$$

Then T_A is a stopping time relative to the natural filtration $\mathcal{F}_t^0 = \sigma(X_s, s \leqslant t)$.

Remark. The stopping time T_A is called the *exit time* of the process X_t from the set A.

Proof. We have to prove that for every $t \geqslant 0$, $\{T_A \leqslant t\} \in \mathcal{F}_t^0$, or, which is the same, that $\{T_A > t\} \in \mathcal{F}_t^0$. But

$$\{\omega :\ T_A(\omega) > t\} = \bigcup_{s\in[0,t]} \{\omega :\ X_s(\omega) \in A\} = \bigcup_{s\in\mathbb{Q}\cap[0,t]} \{\omega :\ X_s(\omega) \in A\} \in \mathcal{F}_t^0.$$

□

Proposition 5.6. *Let $A \subset \mathbb{R}^n$ be an open set, and let the process X_t have continuous paths. Define*

$$\tilde{T}_A(\omega) = \inf\{t \geqslant 0,\ X_t(\omega) \in A\}.$$

Then $\tilde{T}_A$ is a stopping time relative to the natural filtration $\mathcal{F}_t^0 = \sigma(X_s, s \leqslant t)$.

Remark. The stopping time $\tilde{T}_A$ is called the *hitting time* of A.

Definition 5.7. *Let X_t be a process and T be a stopping time. The process $X_t^T = X_{t\wedge T}$ is called the **stopped** process.*

Remark. In the above definition $t \wedge T$ denotes $\min\{t, T\}$.

5.2 Definition and examples of martingales

Let $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ be a filtered probability space.

Definition 5.8. *An $\mathcal{F}_t$-adapted process X_t is called a **martingale** if $\mathbb{E}|X_t| < \infty$ and*

$$\mathbb{E}\big[X_t \,|\, \mathcal{F}_s\big] = X_s \quad \text{a.s. for all } \ s \leqslant t.$$

Theorem 5.9. *Let B_t be a one-dimensional $\mathcal{F}_t$-Brownian motion. Then, the following processes are $\mathcal{F}_t$-martingales*

1. *B_t itself;*
2. *$B_t^2 - t$;*
3. *$M_t^\alpha = \exp\Big(\alpha B_t - \frac{\alpha^2}{2}\, t\Big)$ for all $\alpha \in \mathbb{R}$.*

Proof. 1. $\mathbb{E}[B_t|\mathcal{F}_s] = \mathbb{E}[B_s|\mathcal{F}_s]+\mathbb{E}[B_t-B_s|\mathcal{F}_s]$. Since B_s is $\mathcal{F}_s$ measurable and B_t-B_s is independent of $\mathcal{F}_s$, $\mathbb{E}[B_t - B_s|\mathcal{F}_s] = \mathbb{E}[B_t - B_s] = 0$.

$$\mathbb{E}[B_t|\mathcal{F}_s] = B_s.$$

2. Note that $\mathbb{E}\big|B_t^2 - t\big| \leqslant 2t$. Next,

$$\mathbb{E}\big[B_t^2 \,|\, \mathcal{F}_s^0\big] = \mathbb{E}\big[\big(B_s + (B_t - B_s)\big)^2 \,|\, \mathcal{F}_s\big] = B_s^2 + 2B_s\,\mathbb{E}\big[B_t - B_s \,|\, \mathcal{F}_s\big] + \mathbb{E}\big[(B_t - B_s)^2 \,|\, \mathcal{F}_s\big] = B_s^2 + (t-s).$$

The latter equality implies that a.s.

$$\mathbb{E}\big[B_t^2 - t \,|\, \mathcal{F}_s\big] = B_s^2 - s.$$

3. Note that

$$\mathbb{E}\big[\exp(\alpha B_t)\big] = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} e^{\alpha x}\, e^{-\frac{x^2}{2t}}\, dx = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} e^{-\frac{(x-\alpha t)^2}{2t}}\, e^{\frac{\alpha^2 t}{2}}\, dx = e^{\frac{\alpha^2 t}{2}}. \quad (27)$$

We have:

$$\mathbb{E}\big[\exp(\alpha B_t) \,|\, \mathcal{F}_s\big] = \exp(\alpha B_s)\,\mathbb{E}\big[\exp(\alpha(B_t - B_s)) \,|\, \mathcal{F}_s\big] = \exp(\alpha B_s)\,\mathbb{E}\big[\exp(\alpha(B_t - B_s))\big] = \exp(\alpha B_s)\exp\Big(\frac{\alpha^2(t-s)}{2}\Big).$$

We used relation (27) and the fact that $B_t - B_s$ is independent of the σ-algebra $\mathcal{F}_s$. The latter relation implies:

$$\mathbb{E}\Big[\exp\big(\alpha B_t - \frac{\alpha^2 t}{2}\big) \,|\, \mathcal{F}_s\Big] = \exp\big(\alpha B_s - \frac{\alpha^2 s}{2}\big).$$

□

5.3 Discrete-time martingales

In the definition of a martingale one can replace continuous time processes and filtrations with discrete time processes and filtrations. Namely, a discrete time process X_n, $n \in \mathbb{Z}_+$ ($\mathbb{Z}_+ = \mathbb{N} \cup \{0\}$), is called *adapted* with respect to a discrete-time filtration $\mathcal{F}_n$, if for any $n \in \mathbb{Z}_+$, the random variable X_n is $\mathcal{F}_n$-measurable. An adapted discrete-time process X_n is called a *martingale* if $\mathbb{E}|X_n| < \infty$ for any $n \in \mathbb{Z}_+$ and $\mathbb{E}[X_n \,|\, \mathcal{F}_{n-1}] = X_{n-1}$.

Proposition 5.10. *Let X_n, $n \in \mathbb{Z}_+$, be a martingale with respect to a discrete filtration $\mathcal{F}_n$, and let H_n, $n \in \mathbb{Z}_+$, be a positive bounded process such that H_n is $\mathcal{F}_{n-1}$-measurable for $n > 1$. Define the process Y_n:*

$$Y_0 = X_0, \qquad Y_n = Y_{n-1} + H_n(X_n - X_{n-1}).$$

Then, the process Y_n is a martingale. In particular, if T is an integer-valued stopping time, then the stopped process X_n^T is a martingale.

Proof. The first sentence (stating that Y_n is a martingale) can be verified immediately. Let us prove that the stopped process $X_n^T = X_{n\wedge T}$ is a martingale. Take $H_n = \mathbb{1}_{\{n\leqslant T\}}$ and note that $H_n = 1 - \mathbb{1}_{\{T\leqslant n-1\}}$, and therefore it is $\mathcal{F}_{n-1}$-measurable. One can easily verify that if Y_n is constructed with the help of $H_n = \mathbb{1}_{\{n\leqslant T\}}$ is exactly the stopped process X_n^T. Indeed,

$$Y_n = X_0 + \sum_{k=1}^{n} \mathbb{1}_{\{k\leqslant T\}}(\omega)(X_k - X_{k-1})$$

If $\omega \in \{n \leqslant T\}$, then all indicators are equal to 1, and the sum telescopes:

$$Y_n = X_0 + \sum_{k=1}^{n}(X_k - X_{k-1}) = X_n = X_{n\wedge T}.$$

If $\omega \in \{n > T\}$, then $\mathbb{1}_{\{k\leqslant T\}} = 1$ for $k \leqslant T$ and $\mathbb{1}_{\{k\leqslant T\}} = 0$ for $k > T$, hence

$$Y_n = X_0 + \sum_{k=1}^{T}(X_k - X_{k-1}) = X_T = X_{n\wedge T}.$$

This proves that X_n^T is a martingale. □

Definition 5.11. *The process Y_n defined in Proposition 5.10 will be denoted by $(H \cdot X)_n$. More specifically,*

$$(H \cdot X)_n = X_0 + \sum_{k=1}^{n} H_n(X_k - X_{k-1}).$$

Theorem 5.12 (A discrete-time version of the optional stopping theorem)**.** *Let X_n be a martingale and let S and T be two integer-valued stopping times such that for every ω*

$$S(\omega) \leqslant T(\omega) \leqslant M < \infty,$$

where $M > 0$ is a constant. Then

$$X_S = \mathbb{E}[X_T \,|\, \mathcal{F}_S] \qquad a.s. \tag{28}$$

Lemma 5.13. *If $(X_n)_{n\geqslant 0}$ is adapted to $\mathcal{F}_n$, and S is an integer-valued stopping time, then X_S is $\mathcal{F}_S$-measurable.*

Proof. Let $B \subset \mathbb{R}$ be a Borel set. Since S is integer-valued,

$$\Omega = \bigcup_{k=0}^{\infty}\{\omega \in \Omega : S(\omega) = k\}.$$

Consequently,

$$\{X_S \in B\} = \bigcup_{k=0}^{\infty}\big(\{S = k\} \cap \{X_k \in B\}\big).$$

For any $n \in \mathbb{Z}_+$,

$$\{X_S \in B\} \cap \{S \leqslant n\} = \bigcup_{k=0}^{n} \big(\{S = k\} \cap \{X_k \in B\}\big) \in \mathcal{F}_n.$$

Indeed, $\{S = k\} \cap \{X_k \in B\} \in \mathcal{F}_k \subset \mathcal{F}_n$ for $k \leqslant n$. Hence, X_S is $\mathcal{F}_S$-measurable. □

Proof of Theorem 5.12. Take $H_n = \mathbb{1}_{\{n \leqslant T\}} - \mathbb{1}_{\{n \leqslant S\}}$. Note that $(H \cdot X)_n - X_0$ is linear in H. Hence

$$(H \cdot X)_n - X_0 = (\mathbb{1}_{\{n \leqslant T\}} \cdot X)_n - X_0 - \big((\mathbb{1}_{\{n \leqslant S\}} \cdot X)_n - X_0\big) = X_{n \wedge T} - X_{n \wedge S}.$$

If $n > M$, then

$$(H \cdot X)_n - X_0 = X_T - X_S. \tag{29}$$

Next, since $(H \cdot X)_n$ is a martingale, $\mathbb{E}[(H \cdot X)_n] = \mathbb{E}[X_0]$. Hence,

$$\mathbb{E}[X_T] = \mathbb{E}[X_S]. \tag{30}$$

Let us apply (30) to the pair of stopping times $S^B := S\,\mathbb{1}_B + M\,\mathbb{1}_{B^c}$ and $T^B := T\,\mathbb{1}_B + M\,\mathbb{1}_{B^c}$, where $B \in \mathcal{F}_S$. Note that S^B and T^B are indeed stopping times since

$$\{S^B < n\} \in \mathcal{F}_n \quad \text{and} \quad \{T^B < n\} \in \mathcal{F}_n.$$

Indeed,

$$\{S^B < n\} \cap B = \{S < n\} \cap B \in \mathcal{F}_n.$$

Further, if $M \geqslant n$,

$$\{S^B < n\} \cap B^c = \{M < n\} \cap B^c = \varnothing \cap B^c \in \mathcal{F}_n.$$

If $M < n$, then, $S \leqslant M < n$ a.s. Hence,

$$\{S^B < n\} \cap B^c = B^c = \{S < n\} \cap B^c \in \mathcal{F}_n.$$

Since $X_{T^B} = X_T\,\mathbb{1}_B + X_M\,\mathbb{1}_{B^c}$ and $X_{S^B} = X_S\,\mathbb{1}_B + X_M\,\mathbb{1}_{B^c}$, equality (30) implies:

$$\mathbb{E}[X_T\,\mathbb{1}_B + X_M\,\mathbb{1}_{B^c}] = \mathbb{E}[X_S\,\mathbb{1}_B + X_M\,\mathbb{1}_{B^c}].$$

Therefore, for any $\mathcal{F}_S$-measurable set B,

$$\int_B X_T(\omega)\,P(d\omega) = \int_B X_S(\omega)\,P(d\omega).$$

By Lemma 5.13, this is equivalent to (28). □

5.4 Optional stopping theorem

Lemma 5.14. *Let T be a stopping time. Define $T_k = +\infty$ if $T \geqslant k$, and $T_k = q2^{-k}$, if $(q-1)2^{-k} \leqslant T < q2^{-k}$, where $q2^{-k} < k$, $q \in \mathbb{Z}$. Then, T_k is a sequence of stopping times such that the stopping time T is the decreasing limit of T_k as $k \to \infty$.*

Proof. As k increases by one, each interval $[(q-1)2^{-k},\, q2^{-k}]$ gets divided into two intervals of equal length. This shows that the limit $\{T_k\}$ is decreasing to T as $k \to \infty$. Now note that the set $\{T < \tau\} \in \mathcal{F}_\tau$. Indeed,

$$\{T < \tau\} = \bigcup_{n=1}^{\infty} \Big\{ T \leqslant \tau - \frac{1}{n} \Big\}$$

But $\{T \leqslant \tau - \frac{1}{n}\} \in \mathcal{F}_{\tau - \frac{1}{n}} \subset \mathcal{F}_\tau$, and therefore $\{T < \tau\} \in \mathcal{F}_\tau$.

Note that

$$\omega \in T^{-1}[(q-1)2^{-k},\, q2^{-k}) \iff T(\omega) \in [(q-1)2^{-k},\, q2^{-k}) \iff T2^k \in [q-1, q)$$
$$q = [T\,2^k] + 1,$$

where $[\,\cdot\,]$ denotes the integer part. These arguments imply that

$$T_k = ([T\,2^k] + 1)\,2^{-k}, \qquad \omega \in \{T(\omega) < k\},$$
$$T_k = +\infty, \qquad \omega \in \{T(\omega) \geqslant k\}.$$

Let $t < k$. Analogously, we define $t_k = ([t\,2^k] + 1)\,2^{-k}$, so that $[t\,2^k] = t_k 2^k - 1$. Clearly,

$$t_k 2^k - 1 = [t2^k] \leqslant t2^k < [t2^k] + 1 = t_k 2^k \implies t_k - 2^{-k} \leqslant t < t_k.$$

Since $2^k T_k$ takes integer values and $(q-1)2^{-k} \leqslant T < q2^{-k}$ with $q = [T\,2^k] + 1$,

$$\begin{aligned} \{T_k \leqslant t\} &= \{2^k T_k \leqslant 2^k t\} = \{2^k T_k \leqslant [2^k t]\} = \{2^k T_k \leqslant 2^k t_k - 1\} \\ &= \{[T\,2^k] + 1 \leqslant 2^k t_k - 1\} = \{T\,2^k < 2^k t_k - 1\} = \{T < t_k - 2^{-k}\} \\ &\qquad\qquad \subset \mathcal{F}_{t_k - 2^{-k}} \subset \mathcal{F}_t. \quad (31) \end{aligned}$$

If $t \geqslant k$, then by definition of T_k we have

$$T_k(\omega) = +\infty \quad \iff \quad T(\omega) \geqslant k.$$

Hence,

$$\{T < k\} = \{T2^k < k2^k\} = \{[T2^k] \leqslant k2^k - 1\} = \{T_k 2^k \leqslant k2^k\} = \{T_k \leqslant k\} \subset \{T_k \leqslant t\},$$
$$\{T_k \leqslant t\} \subset \{T_k < \infty\} = \{T < k\}.$$

Hence,

$$\{T_k \leqslant t\} = \{T < k\} \subset \mathcal{F}_k \subset \mathcal{F}_t.$$

This shows that T_k is a stopping time. □

Theorem 5.15 (Doob's martingale convergence theorem)**.** *Let X_t be a right-continuous $\mathcal{F}_t$-martingale with the property that*

$$\sup_{t>0} \mathbb{E}[X_t^-] < \infty,$$

where $X_t^- = \max(-X_t, 0)$. Then, the pointwise limit

$$X_\infty(\omega) = \lim_{t\to\infty} X_t(\omega)$$

exists a.s., and

$$X_t = \mathbb{E}[X_\infty \,|\, \mathcal{F}_t].$$

Proof. Without proof. A proof can be found in the book by Z. Brzezniak, T. Zastawniak "Basic stochastic processes: a course through exercises", Chapter 4. □

Definition 5.16. *A family of random variables X_τ, $\tau \in \mathcal{A}$, is called* ***uniformly integrable*** *if for any $\varepsilon > 0$ there exists a constant $K > 0$ such that*

$$\sup_{\tau\in\mathcal{A}} \int_{\{\omega:|X_\tau(\omega)|>K\}} |X_\tau(\omega)|\, P(d\omega) < \varepsilon.$$

Theorem 5.17. *Let $\mathcal{F}$ be a σ-algebra. Suppose $X_n \to X$ where X_n and X are random variables such that one of the following conditions is true*

(i) $|X_n| \leqslant Z$ a.s. for all n and $\mathbb{E}|Z| < \infty$.

(ii) The family X_n, $n \in \mathbb{N}$ is uniformly integrable.

Then,

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

Proof. Without proof. A proof can be found in the book by D. Revuz and M. Yor "Continuous martingales and Brownian motion", Chapter II. □

Lemma 5.18. *Let X_t be a continuous martingale and $M > 0$. Then, $X_t^M = X_{t\wedge M}$ is a uniformly integrable martingale.*

Proof. If $s < t \leqslant M$, then $X_s^M = X_s$ and $X_t^M = X_t$. If $M \leqslant s < t$, then $X_t^M = X_s^M = X_M$. If $s < M < t$, then $X_s^M = X_s$ and $X_t^M = X_M$. In all cases, the martingale property is obvious. Thus, X_t^M is a martingale. Further, since $X_t^M = \mathbb{E}[X_M|\mathcal{F}_t]$ for all $t \geqslant 0$, it holds that

$$|X_t^M| = \big|\mathbb{E}[X_M|\mathcal{F}_t]\big| \leqslant \mathbb{E}[|X_M||\mathcal{F}_t] \tag{32}$$

Hence,

$$\begin{aligned}\int_{\{\omega:|X_t^M(\omega)|>K\}} |X_t^M(\omega)|\, P(d\omega) &= \mathbb{E}[\mathbb{1}_{\{|X_t^M|>K\}}\, |X_t^M|\,] \\ &\leqslant \mathbb{E}\big[\,\mathbb{1}_{\{|X_t^M|>K\}}\, \mathbb{E}[|X_M||\mathcal{F}_t]\,\big] = \mathbb{E}\big[\mathbb{E}[\mathbb{1}_{\{|X_t^M|>K\}}\, |X_M||\mathcal{F}_t]\,\big] \\ &= \mathbb{E}[\mathbb{1}_{\{|X_t^M|>K\}}\, |X_M|] \leqslant \mathbb{E}[\mathbb{1}_{\{\sup_{t\in[0,M]}|X_t^M|>K\}}\, |X_M|] \to 0 \quad \text{as } K\to\infty. \end{aligned} \tag{33}$$

Indeed, since X_t is continuous, $\sup_{t\in[0,M]}|X_t|<\infty$ a.s. Hence, $\mathbb{1}_{\{\sup_{t\in[0,M]}|X^M_t|>K\}}\to 0$ as $K\to\infty$ a.s. One can pass to the limit under the expectation sign by the dominated convergence theorem. The lemma is proved. □

Lemma 5.19. *Let $T\leqslant M$ be a bounded stopping time and let $T_n=([T\,2^n]+1)\,2^{-n}$. Further let X_t be a continuous martingale. Then the family of random variables $\{X_{T_n}\}$, $n\in\mathbb{N}$, is uniformly integrable.*

Proof. By definition,

$$T_n=\sum_{q=1}^{\infty}q2^{-n}\,\mathbb{1}_{A_{qn}}\qquad\text{where } A_{qn}:=(q-1)2^{-n}\leqslant T<q2^{-n}.$$

Hence, setting $N:=M+1$ and noticing that $X_{T_n}=X^N_{T_n}$, we have

$$X_{T_n}=\sum_{q=1}^{\infty}X^N_{q2^{-n}}\,\mathbb{1}_{A_{qn}}\,.$$

By (32) and similar to (33),

$$\mathbb{E}[\mathbb{1}_{\{|X_{T_n}|>K\}}\,|X_{T_n}|\,]=\sum_{q=1}^{\infty}\mathbb{E}[\mathbb{1}_{\{|X^N_{q2^{-n}}|>K\}}\,\mathbb{1}_{A_{qn}}\,|X^N_{q2^{-n}}|]$$

$$\sum_{q=1}^{\infty}=\mathbb{E}\big[\mathbb{E}[\mathbb{1}_{\{|X^N_{q2^{-n}}|>K\}}\,\mathbb{1}_{A_{qn}}\,|X_N||\mathcal{F}_{q2^{-n}}]\big]\leqslant\sum_{q=1}^{\infty}\mathbb{E}[\mathbb{1}_{\{\sup_{t\in[0,N]}|X_t|>K\}}\,\mathbb{1}_{A_{qn}}\,|X_N|]$$

$$=\mathbb{E}[\mathbb{1}_{\{\sup_{t\in[0,N]}|X_t|>K\}}\,|X_N|]\to 0,\qquad K\to+\infty.$$

□

Theorem 5.20 (Optional stopping theorem)**.** *Let X_t, $t\geqslant 0$, be a continuous martingale, and let $S\leqslant T\leqslant M$ be bounded stopping times. Then,*

$$\mathbb{E}[X_T\,|\,\mathcal{F}_S]=X_S.\tag{34}$$

In particular, the stopped process $X^T_t=X_{t\wedge T}$ is a martingale with respect to $\mathcal{F}_{t\wedge T}$.

Proof. Let $S_n=([S\,2^n]+1)\,2^{-n}$, $T_n=([T\,2^n]+1)\,2^{-n}$ (as in Lemma 5.14). However since S and T are bounded, we define S_n and T_n by the above formulas everywhere on Ω. Fix $n\in\mathbb{N}$ and define a martingale

$$Y_m=X_{T_n\wedge m2^{-n}}$$

with respect the filtration $\mathcal{G}_m=\mathcal{F}_{m2^{-n}}$. Since $X_{t2^{-n}}$ is an $\mathcal{F}_{t2^{-n}}$-martingale, $X_{m2^{-n}}$ is a $\mathcal{G}_m$-martingale. Hence, $Y_m=X_{(\tilde T\wedge m)2^{-n}}=X^{\tilde T}_{m2^{-n}}$, where $\tilde T:=[T\,2^n]+1$, is a martingale by Proposition 5.10. Note that

$$\{\tilde T\leqslant m\}=\{T_n\leqslant m2^{-n}\}\in\mathcal{F}_{m2^{-n}}=\mathcal{G}_m,$$

so $\tilde{T}$ is a $\mathcal{G}_m$ stopping time. Likewise, $\tilde{S} = [S\,2^n] + 1$ is also a $\mathcal{G}_m$-stopping time.

We apply Theorem 5.12 to the pair of stopping times $\tilde{S}$ and $\tilde{T}$. We obtain:

$$\mathbb{E}[X_{T_n} \,|\, \mathcal{F}_{S_n}] = \mathbb{E}[Y_{\tilde{T}} \,|\, \mathcal{G}_{\tilde{S}}] = Y_{\tilde{S}} = X_{S_n}. \tag{35}$$

Indeed, $T_n = \tilde{T}2^{-n}$, and hence $Y_{\tilde{T}} = X_{T_n \wedge \tilde{T}2^{-n}} = X_{T_n}$. Further, $\mathcal{G}_{\tilde{S}} = \mathcal{F}_{\tilde{S}2^{-n}} = \mathcal{F}_{S_n}$. The second equality in (35) holds by Theorem 5.12. Finally, the third equality in (35) follows again from the definition of the process Y_m.

Let $A \in \mathcal{F}_S$. Further let t_k be defined as in Lemma 5.14, so that $t_k - 2^{-k} \leqslant t$. By (31),

$$A \cap \{S_n \leqslant t\} = A \cap \{S < t_k - 2^{-k}\} \subset \mathcal{F}_{t_k - 2^{-k}} \subset \mathcal{F}_t.$$

Hence $A \in \mathcal{F}_{S_n}$ for all $n \in \mathbb{N}$. By (35),

$$\int_A X_{S_n} \mathbb{P}(d\omega) = \int_A X_{T_n} \mathbb{P}(d\omega). \tag{36}$$

By Lemma 5.19, the families $\{X_{S_n}\}_{n\in\mathbb{N}}$ and $\{X_{T_n}\}_{n\in\mathbb{N}}$ are uniformly integrable. Let $\varepsilon > 0$ ne arbitrary and $K > 0$ be as it is dictated by the definition of the uniform integrability. Then, (36) implies

$$\begin{aligned}
&\int_A \mathbb{1}_{\{|X_{S_n}|>K\}} X_{S_n} \mathbb{P}(d\omega) + \int_A \mathbb{1}_{\{|X_{S_n}|\leqslant K\}} X_{S_n} \mathbb{P}(d\omega) \\
&= \int_A \mathbb{1}_{\{|X_{T_n}|>K\}} X_{T_n} \mathbb{P}(d\omega) + \int_A \mathbb{1}_{\{|X_{T_n}|\leqslant K\}} X_{T_n} \mathbb{P}(d\omega).
\end{aligned}$$

We obtain

$$\int_A \mathbb{1}_{\{|X_{S_n}|\leqslant K\}} X_{S_n} \mathbb{P}(d\omega) = \int_A \mathbb{1}_{\{|X_{T_n}|\leqslant K\}} X_{T_n} \mathbb{P}(d\omega) + r_{n,K,\varepsilon}, \quad \text{with } \sup_n |r_{K,n,\varepsilon}| < \varepsilon.$$

Passing to the limit as $n \to \infty$, by the bounded convergence theorem

$$\int_A \mathbb{1}_{\{|X_S|\leqslant K\}} X_S \mathbb{P}(d\omega) = \int_A \mathbb{1}_{\{|X_T|\leqslant K\}} X_T \mathbb{P}(d\omega) + \lim_{n\to\infty} r_{n,K,\varepsilon}.$$

Passing to the limit as $K \to \infty$, we obtain that

$$\int_A X_S \, \mathbb{P}(d\omega) = \int_A X_T \, \mathbb{P}(d\omega) + + \lim_{K\to\infty} \lim_{n\to\infty} r_{n,K,\varepsilon},$$

where $|\lim_{K\to\infty} \lim_{n\to\infty} r_{n,K,\varepsilon}| \leqslant \varepsilon$. Since $\varepsilon > 0$ is arbitrary, we have

$$\int_A X_S \, \mathbb{P}(d\omega) = \int_A X_T \, \mathbb{P}(d\omega) \qquad \forall A \in \mathcal{F}_S.$$

Let us show that X_S is $\mathcal{F}_S$ measurable. Note that by Lemma 5.13, $X_{S_n} = Y_{\tilde{S}}$ is $\mathcal{G}_{\tilde{S}} = \mathcal{F}_{S_n}$-measurable. Let $B \in \mathcal{B}(\mathbb{R}^n)$ be an open set. By (31),

$$\{X_{S_n} \in B\} \cap \{S < t_n - 2^{-n}\} = \{X_{S_n} \in B\} \cap \{S_n \leqslant t\} \in \mathcal{F}_t.$$

Hence, the function

$$\mathbb{1}_B(S_n)\,\mathbb{1}_{(S,+\infty)}(t_n - 2^{-n})$$

is $\mathcal{F}_t$-measurable. Since B is an open set, $|S_n - S| \leqslant \frac{1}{n}$, and $t_n - 2^{-n} \to t$ as $n \to \infty$, we obtain that the limit function

$$\mathbb{1}_B(S)\,\mathbb{1}_{(S,+\infty)}(t)$$

is $\mathcal{F}_t$-measurable. This implies

$$\{X_S \in B\} \cap \{S \leqslant t\} \in \mathcal{F}_t \quad \Longleftrightarrow \quad X_S^{-1}(B) \in \mathcal{F}_S \quad \forall\, B \text{ (open set)}$$

Hence,

$$X_S^{-1}(\mathcal{B}(\mathbb{R}^n)) = \sigma\{X_S^{-1}(B),\ B \text{ is an open set}\} \in \mathcal{F}_S$$
$$\Longleftrightarrow \quad X_S \quad \text{is } \mathcal{F}_S\text{-measurable.}$$

Finally we conclude that the stopped process X_t^T is a martingale by applying identity (34) to the stopping times $T \wedge s$ and $T \wedge t$ where $s < t$. □

5.5 Local martingales

Definition 5.21. *Let X_t be an adapted process with respect to the filtration $\mathcal{F}_t$. The process X_t is called a* ***local martingale*** *with respect to the filtration $\mathcal{F}_t$ if there exists a sequence of stopping times*

$$0 = T_0 \leqslant T_1 \leqslant \cdots T_n \leqslant \cdots$$

such that $T_n \to \infty$ as $n \to \infty$ and the process $X_t^{T_n}$ is an $\mathcal{F}_t$-martingale.

The sequence T_n is called a *reducing system* of stopping times. Every martingale is a local martingale since every sequence of stopping time increasing to infinity is reducing by Theorem 5.20.

Theorem 5.22. *Let X_t be $\mathcal{F}_t$-adapted and S be a stopping time. Then, X_t^S is an $\mathcal{F}_t$-martingale if and only if it is an $\mathcal{F}_{S\wedge t}$-martingale.*

Proof. Let X_t^S be an $\mathcal{F}_t$-martingale, then, by the Optional stopping theorem, X_t^S is also an $\mathcal{F}_{S\wedge t}$-martingale. Indeed, let $s < t$. We have

$$\mathbb{E}[X_t^S | \mathcal{F}_{s\wedge S}] = X_{s\wedge S}^S = X_s^S.$$

Indeed, $s \wedge S < t$ can be regarded as bounded stopping times.

Let us prove now the opposite. Suppose X_t^S is an $\mathcal{F}_{S\wedge t}$-martingale. We show that it is an $\mathcal{F}_t$-martingale for every n. We have to show that

$$\mathbb{E}[X_t^S | \mathcal{F}_r] = X_r^S \qquad \text{for all } r < t \ \text{ a.s.}$$

Which is equivalent to showing that for every $A \in \mathcal{F}_r$,

$$\int_A X_r^S \mathbb{P}(d\omega) = \int_A X_t^S \mathbb{P}(d\omega).$$

We claim that $A \cap \{S > r\} \in \mathcal{F}_{S\wedge r}$. Let us prove this claim. Take $\tau \in \mathbb{R}$. We have to prove that

$$A \cap \{S > r\} \cap \{S \wedge r < \tau\} \in \mathcal{F}_\tau.$$

If $\tau \leqslant r$, then $\{S > r\} \cap \{S \wedge r < \tau\} = \varnothing$, so this case is trivial. If $\tau > r$, then $\{S \wedge r < \tau\} = \Omega$. Now since $A \in \mathcal{F}_r$ and $\{S > r\} \in \mathcal{F}_r$, we have

$$A \cap \{S > r\} \cap \{S \wedge r < \tau\} = A \cap \{S > r\} \in \mathcal{F}_r \subset \mathcal{F}_\tau \qquad \text{for } \tau > r, \tag{37}$$

which proves the claim.

Since $A \cap \{S > r\} \in \mathcal{F}_{S\wedge r}$ and X_t^S is an $\mathcal{F}_{S\wedge t}$-martingale, we obtain:

$$\int_{A\cap\{S>r\}} X_t^S P(d\omega) = \int_{A\cap\{S>r\}} X_r^S P(d\omega).$$

On the other hand, taking into account that on the set $\{S \leqslant r\}$, one has $S \leqslant r < t$, we have:

$$\begin{aligned}\int_{A\cap\{S\leqslant r\}} X_t^S P(d\omega) = \int_{A\cap\{S\leqslant r\}} X_{S\wedge t} P(d\omega) = \int_{A\cap\{S\leqslant r\}} X_S P(d\omega) \\ = \int_{A\cap\{S\leqslant r\}} X_{S\wedge r} P(d\omega) = \int_{A\cap\{S\leqslant r\}} X_r^S P(d\omega)\end{aligned} \tag{38}$$

Adding (37) to (38), we obtain that

$$\int_A X_t^S P(d\omega) = \int_A X_r^S P(d\omega).$$

We also note that if $B \in \mathcal{B}(\mathbb{R}^n)$, then

$$\begin{aligned}&X_r^S = X_r \mathbb{1}_{\{S>r\}} + X_S \mathbb{1}_{\{S\leqslant r\}}, \\ &(X_S \mathbb{1}_{\{S\leqslant r\}})^{-1}(B) = X_S^{-1}(B) \cap \{S \leqslant r\} \in \mathcal{F}_r, \quad \text{if } 0 \notin B, \\ &(X_S \mathbb{1}_{\{S\leqslant r\}})^{-1}(B) = \big(X_S^{-1}(B) \cap \{S \leqslant r\}\big) \cup \{S > r\} \in \mathcal{F}_r, \quad \text{if } 0 \in B.\end{aligned} \tag{39}$$

Hence, X_r^S is $\mathcal{F}_r$-measurable. This proves that $\mathbb{E}[X_t^S|\mathcal{F}_r] = X_r^S$. □

Theorem 5.23. *Let X_t, $t \geqslant 0$, be a continuous local martingale. Then the sequence of stopping times*

$$T_n = \inf\{t \geqslant 0 : |X_t| > n\} \tag{40}$$

is always reducing. In particular, one can always find a sequence of stopping times that reduces X_t to a bounded martingale.

Proof. Note that by Proposition 5.5 every T_n is a stopping time with the property $T_n \leqslant T_{n+1}$ a.s. Let us show that $\lim_{n\to\infty} T_n = +\infty$ a.s. Let $T := \lim_{n\to\infty} T_n$. We have

$$\{T = +\infty\} = \cap_{N=1}^{\infty}\{T > N\} = \cap_{N=1}^{\infty} \cup_{n=1}^{\infty} \{T_n > N\} = \cap_{N=1}^{\infty} \cup_{n=1}^{\infty} \{ \sup_{t\in[0,N]} |X_t| \leqslant n\}$$

Note that

$$\begin{aligned}
&\lim_{m\to\infty} \mathbb{P}\{ \sup_{t\in[0,N]} |X_t| \leqslant m\} = 1,\\
&\mathbb{P}\big\{ \cup_{n=1}^{\infty} \{ \sup_{t\in[0,N]} |X_t| \leqslant n\}\big\} = \lim_{m\to\infty} \mathbb{P}\big\{ \cup_{n=1}^{m} \{ \sup_{t\in[0,N]} |X_t| \leqslant n\}\big\}\\
&\qquad = \lim_{m\to\infty} \mathbb{P}\big\{ \sup_{t\in[0,N]} |X_t| \leqslant m\big\} = 1.
\end{aligned}$$

Hence $\mathbb{P}\{T = +\infty\} = 1$.

Let $0 < s < t$, and let S_n be a reducing sequence of stopping times. We apply the optional stopping theorem (Theorem 5.20) to each martingale $X_t^{S_n}$ and conclude that for every $m > 0$ the process $X^{S_n}_{T_m\wedge t}$ is a martingale. This implies:

$$\mathbb{E}[X^{S_n}_{t\wedge T_m} \,|\, \mathcal{F}_{s\wedge T_m}] = X^{S_n}_{s\wedge T_m}.$$

By Theorem 5.17 we can pass to the limit as $n \to \infty$ in the both parts of the above equality. We obtain that a.s.

$$\mathbb{E}[X_t^{T_m} \,|\, \mathcal{F}_{s\wedge T_m}] = \mathbb{E}[X_{t\wedge T_m} \,|\, \mathcal{F}_{s\wedge T_m}] = X_{s\wedge T_m} = X_s^{T_m}.$$

Hence, $X_t^{T_m}$ is an $\mathcal{F}_{t\wedge T_m}$-martingale, and consequently, an $\mathcal{F}_t$-martingale. Note that $|X_t^{T_n}| \leqslant n$, and therefore, $X_t^{T_n}$ is a bounded martingale. □

Remark. Instead of the sequence $\{T_n\}$ we could consider any sequence of stopping times $T'_n \leqslant T_n$ such that $\lim_{n\to\infty} T'_n = \infty$. Clearly, $X_t^{T'_n}$ would be a bounded maringale.

Lemma 5.24. *Let X_t be a martingale and T be a stopping time. Then, X_t^T is also a martingale. In particular, if T_n is given by (40) for some $n \in \mathbb{Z}_+$, $X_t^{T_n}$ is a bounded martingale.*

Proof. By (39), X_t^T is $\mathcal{F}_t$-adapted. The statement holds by the optional stopping theorem. □

Definition 5.25. *Let $\tau > 0$ be a random time. A process X_t, $t \geqslant 0$, is called a local martingale on $[0, \tau)$ if there is a sequence of stopping time $T_n \uparrow \tau$ such that $X_t^{T_n}$ is a martingale.*

Theorem 5.26. *Suppose X_t, $t \geqslant 0$, is a local martingale, and for every $t > 0$,*

$$\mathbb{E}\Big[\sup_{0\leqslant s\leqslant t} |X_s|\Big] < \infty. \tag{41}$$

Then, X_t is a martingale.

Proof. Clearly, $\mathbb{E}|X_t| < \infty$. Now let $\{T_n\}$ be a reducing sequence of stopping times so that

$$\mathbb{E}[X_t^{T_n} \,|\, \mathcal{F}_s] = X_s^{T_n} \qquad \text{a.s.} \tag{42}$$

Assumption (41) allows us to apply Theorem 5.17 since $|X_t^{T_n}| \leqslant \sup_{0\leqslant s\leqslant t} |X_s|$ and the latter function is integrable. Taking the limit in (42) as $n \to \infty$, we obtain:

$$\mathbb{E}[X_t \,|\, \mathcal{F}_s] = X_s \qquad \text{a.s.}$$

This proves that X_t is a martingale. □

Corollary 5.27. *A bounded local martingale is a martingale.*

5.6 The variance process

Let $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ be a filtered probability space.

Definition 5.28. *An $\mathcal{F}_t$-adapted process X_t is called a submartingale if $\mathbb{E}|X_t| < \infty$ and*

$$\mathbb{E}(X_t|\mathcal{F}_s) \geqslant X_s \quad \textit{a.s. for all} \quad s \leqslant t.$$

Definition 5.29. *A family of random variables X_τ, $\tau \in \mathcal{A}$, is called* ***uniformly integrable*** *if for any $\varepsilon > 0$ there exists a constant $K > 0$ such that*

$$\sup_{\tau\in\mathcal{A}} \int_{\{|X_\tau|>K\}} |X_t(\omega)|\, P(d\omega) < \varepsilon.$$

Theorem 5.30 (Doob–Meyer's theorem)**.** *Let X_t be a continuous submartingale such that for every constant $a > 0$, the family of rundom variables*

$$\{X_\tau,\ \tau \textit{ is a stopping time with } \tau \leqslant a\}$$

is uniformly integrable. Then there exists a unique decomposition

$$X_t = M_t + A_t$$

where M_t is a continuous martingale and A_t is a continuous non-decreasing $\mathcal{F}_t$-adapted process starting at zero and such that $\mathbb{E}(A_t) < \infty$.

Proof. Without proof. □

Lemma 5.31. *Let X_t be a square integrable continuous martingale. Then X_t^2 is a continuous submartingale satisfying the assumptions of Theorem 5.30.*

Proof. Let τ be a stopping time with $\tau \leqslant a$ where a is a constant. By the Optional stopping theorem, we obtain:

$$0 \leqslant \mathbb{E}[(X_a - X_\tau)^2|\mathcal{F}_\tau] = \mathbb{E}[X_a^2|\mathcal{F}_\tau] - 2X_\tau\mathbb{E}[X_a|\mathcal{F}_\tau] + X_\tau^2 = \mathbb{E}[X_a^2|\mathcal{F}_\tau] - X_\tau^2.$$

Since the left-hand side is positive, we obtain that

$$X_\tau^2 \leqslant \mathbb{E}[X_a^2|\mathcal{F}_\tau].$$

Note that $\{X_\tau^2 > K\} \subset \{\sup_{\tau\leqslant a} X_\tau^2 > K\}$, where the supremum is over all stopping times $\tau \leqslant a$. We have

$$\int_\Omega \mathbb{1}_{\{X_\tau^2>K\}} X_\tau^2(\omega)\mathbb{P}(d\omega) \leqslant \mathbb{E}\big[\,\mathbb{1}_{\{X_\tau^2>K\}}\,\mathbb{E}[X_a^2|\mathcal{F}_\tau]\big] = \mathbb{E}\big[\mathbb{E}[\mathbb{1}_{\{X_\tau^2>K\}}\,X_a^2|\mathcal{F}_\tau]\big]$$
$$= \mathbb{E}\big[\,\mathbb{1}_{\{X_\tau^2>K\}}\,X_a^2\big] \leqslant \mathbb{E}\big[\,\mathbb{1}_{\{\sup_{\tau\leqslant a} X_\tau^2>K\}}\,X_a^2\big] = \mathbb{E}\big[\,\mathbb{1}_{\{\sup_{t\in[0,a]} X_t^2>K\}}\,X_a^2\big] \to 0, \quad K\to\infty$$

where τ in the above formula always means a stopping time. Let $\Omega' \subset \Omega$, $\mathbb{P}(\Omega') = 1$, be such that tralectories $X_t(\omega)$ are continuous for $\omega \in \Omega'$. Note that for every $\omega \in \Omega'$, $X^2_{\tau(\omega)}(\omega) \leqslant \sup_{t\in[0,a]} X_t^2(\omega)$. Hence, $\sup_{\tau\leqslant a} X_\tau^2 \leqslant \sup_{t\in[0,a]} X_t^2$ and therefore, $\sup_{\tau\leqslant a} X_\tau^2 = \sup_{t\in[0,a]} X_t^2$. Moreover, by the continuity of X_t,

$$\mathbb{1}_{\{\sup_{t\in[0,a]} X_t^2>K\}} = 1 - \mathbb{1}_{\{\sup_{t\in[0,a]} X_t^2\leqslant K\}} \to 0, \quad K\to\infty \quad \text{a.s.}$$

□

In general, if X_t is a martingale, X_t^2 fails to be a martingale.

Theorem 5.32. *Let X_t be a square integrable continuous $\mathcal{F}_t$-martingale. Then, there exists a process $\langle X\rangle_t$, $t \geqslant 0$, which is continuous, non-decreasing, starting at zero, and such that the process $X_t^2 - \langle X\rangle_t$ is a continuous martingale. The proces $\langle X\rangle_t$ is a.s. unique.*

Proof. By Lemma 5.31, X_t^2 is a submartingale sarisfying the assumptions of Doob-Meyer's theorem. Therefore, there exists a continuous martingale M_t and a non-decreasing $\mathcal{F}_t$ continuous apadted process, denoted by $\langle X\rangle_t$, starting at zero and such that

$$X_t^2 = M_t + \langle X\rangle_t.$$

This also implies that $X_t^2 - \langle X\rangle_t$ is a continuous martingale. By the uniqueness of the Doob-Meyer's decomposition, the process $\langle X\rangle_t$ is unique. □

Definition 5.33. *The process $\langle X\rangle_t$ defined in Theorem 5.32 is called the* ***variance process****.*

Theorem 5.34. *Every continuous $\mathcal{F}_t$-local martingale starting at zero which is of bounded variation on every compact interval is constantly equal to zero.*

Proof. Let

$$V_t = \sup\Big\{\sum_{k=1}^n |X_{t_k} - X_{t_{k-1}}| : 0 = t_0 < t_1 < \cdots < t_n = t\Big\}$$

be the variation of X_t on $[0,t]$. Since X_t is continuous, and V_t is bounded on every compact interval, V_t is continuous. Let $S_k = \inf\{s : V_s > k\}$. Note that if $t \leqslant S_k$, then $V_t \leqslant k$, and hence, $|X_t| \leqslant k$. By Theorem 5.5, S_k is a reducing sequence of stopping times, and the process $Y_t = X_t^{S_k} := X_{S_k \wedge t}$ is a bounded martingale with respect to $\mathcal{F}_t$ with $Y_0 = 0$. Note that:

$$\mathbb{E}[(Y_t - Y_s)^2 \,|\, \mathcal{F}_s] = \mathbb{E}[Y_t^2 \,|\, \mathcal{F}_s] - 2Y_s\,\mathbb{E}[Y_t \,|\, \mathcal{F}_s] + Y_s^2 = \mathbb{E}[Y_t^2 - Y_s^2 \,|\, \mathcal{F}_s]. \tag{43}$$

Take a partition of $\{0 = t_0 < t_1 < \cdots < t_n = t\}$ of the interval $[0,t]$. We have:

$$\begin{aligned}\mathbb{E}[Y_t^2] = \mathbb{E}\Big[\sum_{i=1}^n (Y_{t_i}^2 - Y_{t_{i-1}}^2)\Big] = \mathbb{E}\Big[\sum_{i=1}^n (Y_{t_i} - Y_{t_{i-1}})^2\Big] &\leqslant \mathbb{E}\Big[\sup_{i\geqslant 1} |Y_{t_i} - Y_{t_{i-1}}| \sum_{i=1}^n |Y_{t_i} - Y_{t_{i-1}}|\Big] \\ &\leqslant \mathbb{E}\Big[\sup_{i\geqslant 1} |Y_{t_i} - Y_{t_{i-1}}|\, V(t \wedge S_k)\Big] \leqslant k\,\mathbb{E}\Big[\sup_{i\geqslant 1} |Y_{t_i} - Y_{t_{i-1}}|\Big].\end{aligned}$$

Let the mesh $\delta_n := \max_{i\geqslant 1} |t_i - t_{i-1}|$ tend to zero as $n \to \infty$, and note that since $|Y_{t_i}| = |X_{S_k \wedge t_i}| \leqslant k$ for all $1 \leqslant i \leqslant n$, then $\sup_{i\geqslant 1}\Big[|Y_{t_i} - Y_{t_{i-1}}|\Big] \leqslant 2k$. By Lebesgue's theorem we can pass to the limit under the expectation sign. This implies that

$$\lim_{n\to\infty} \mathbb{E} \sup_{i\geqslant 1}\Big[|Y_{t_i} - Y_{t_{i-1}}|\Big] = 0.$$

Therefore $\mathbb{E}[Y_t^2] = 0$, and hence, $Y_t = X_{S_k \wedge t} = 0$ a.s. This proves that there exists a set $\Omega' \subset \Omega$ of full P-measure such that $X_t^{S_k} = 0$ for all rational $t \geqslant 0$ and for all $\omega \in \Omega'$. By continuity of paths, $X_t^{S_k} = 0$ for all $t \geqslant 0$ and for all $\omega \in \Omega'$. Finally we can find another set $\Omega'' \subset \Omega'$ of full P-measure such that $X_t = 0$ for all $t \geqslant 0$ and for all $\omega \in \Omega''$.

We further nore that $\lim_{k\to\infty} S_k = \infty$ by the same argument as in Lemma 5.23. □

Lemma 5.35. *Let X_t be a martingale and T be a stopping time. Then $\langle X^T\rangle_t = \langle X\rangle_t^T$.*

Proof. By what was proved $X_t^2 - \langle X\rangle_t$ is a martingale. By the optional stopping theorem $(X_t^T)^2 - \langle X\rangle_t^T$ is also a martingale. By uniqueness of the variance process for a martingale, $\langle X^T\rangle_t = \langle X\rangle_t^T$. □

Suppose X_t and Y_t, $t \geqslant 0$, are martingales. Then, in general, X_tY_t fails to be a martingale.

Theorem 5.36. *Let X_t and Y_t be martingales which are continuous and start at zero. Then there exists a unique, continuous process $\langle X,Y\rangle_t$ that has bounded variation on every compact interval and starts at zero, such that $X_tY_t - \langle X,Y\rangle_t$ is a martingale.*

Proof. Define

$$\langle X,Y\rangle_t = \frac{1}{4}\big(\langle X+Y\rangle_t - \langle X-Y\rangle_t\big).$$

Clearly, $\langle X, Y\rangle_t$ is continuous, predictable, and of bounded variation. The latter property holds since $\langle X, Y\rangle_t$ is defined as a difference of two increasing processes. Finally,

$$X_tY_t - \langle X, Y\rangle_t = \frac{1}{4}\Big((X_t + Y_t)^2 - \langle X + Y\rangle_t - \big((X_t - Y_t)^2 - \langle X - Y\rangle_t\big)\Big).$$

As a difference of two martingales, $X_tY_t - \langle X, Y\rangle_t$ is a martingale. Let us prove the uniqueness of $\langle X, Y\rangle_t$. Suppose there are two processes A_t and B_t satisfying requirements of the theorem. Then

$$A_t - B_t = (X_tY_t - A_t) - (X_tY_t - B_t)$$

is a continuous martingale which has bounded variation on every compact interval. By Theorem 5.34, $A_t - B_t$ is constantly equal to zero. This implies the uniqueness of the process $\langle X, Y\rangle_t$. □

5.7 Semimartingales

Let $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ be a probability space.

Definition 5.37. *A stochastic process X_t, $t \geqslant 0$, is called a **semimartingale** if there is an $\mathcal{F}_t$-local martingale M_t, $t \geqslant 0$, and a continuous $\mathcal{F}_t$-adapted process A_t, $t \geqslant 0$, of locally bounded variation and starting at 0 such that $X_t = M_t + A_t$ for all $t \geqslant 0$.*

Theorem 5.38. *If X_t is a continuous semimartingale, then the decomposition $X_t = M_t + A_t$ is unique.*

Proof. Suppose $X = M + A = M' + A'$. Then $A - A' = M' - M$ is a continuous local martingale locally of bounded variation and starting at zero. By Theorem 5.34, it is constantly equal to zero. □

6. Stochastic Integral

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let B_t be an n-dimensional Brownian motion, and let $\mathcal{F}_t = \sigma(B_s, s \leqslant t)$ be the natural filtration.

6.1 Construction of the stochastic integral and Itô's isometry

The class of integrands

Definition 6.1. *Let $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ be a filtered probability space. Let $S, T \in \mathbb{R}$, $0 \leqslant S \leqslant T$ and let B_t be an $\mathcal{F}_t$-Brownian motion. Define the class of functions $\Phi = \Phi(S, T)$*

$$f(t, \omega) : [0, \infty) \times \Omega \to \mathbb{R}$$

as follows:

1. $(t,\omega) \mapsto f(t,\omega)$ *is* $\mathcal{B} \times \mathcal{F}$*-measurable, where* $\mathcal{B}$ *denotes the* σ*-algebra of Borel subsets of* $[0,\infty)$.

2. $f(t,\omega)$ *is* $\mathcal{F}_t$*-adapted.*

3. $\mathbb{E}\Big[\int\limits_S^T f^2(t,\omega)\,dt\Big] < \infty$.

Now we are going to define the stochastic integral

$$I[f](t) = \int_S^t f(s,\omega)\,dB_s(\omega), \qquad t \in [S,T],$$

for functions $f \in \Phi$, where B_t is a one-dimensional Brownian motion. The idea is the following: first we define the stochastic integral for simple functions from Φ. Then we show that every function $f \in \Phi$ can be represented as a limit of simple functions ϕ_n in some sense. Then we define the stochastic integral $\int f dB_t$ as a limit $\int \phi_n dB_t$.

Definition of the Stochastic integral for simple functions

A function $\phi \in \Phi$ is called **simple** if it has the form:

$$\phi(t,\omega) = \sum_{j=0}^{N-1} e_j(\omega) \cdot \mathbb{1}_{[t_j,t_{j+1})} \tag{44}$$

where $S = t_0 < t_1 < \cdots t_N = T$ is a partition and e_j is $\mathcal{F}_{t_j}$-measurable. Note that $\phi \in \Phi$, it is $\mathcal{F}_t$-adapted. For functions of form (44) we define the stochastic integral by the formula:

$$I[\phi] = \int_S^T \phi(t,\omega)\,dB_t(\omega) = \sum_{j=0}^{N-1} e_j(\omega)(B_{t_{j+1}} - B_{t_j})(\omega) \tag{45}$$

Itô's isometry

Theorem 6.2 (Itô's isometry for simple functions)**.** *If* $\phi \in \Phi$ *is bounded and simple then*

$$\mathbb{E}\Big[\Big(\int_S^T \phi(t,\omega)dB_t(\omega)\Big)^2\Big] = \mathbb{E}\Big[\int_S^T \phi^2(t,\omega)\,dt\Big]. \tag{46}$$

Proof. Define $\Delta B_j = B_{t_{j+1}} - B_{t_j}$. Let us prove that

$$\mathbb{E}[e_i e_j \Delta B_i \Delta B_j] = \begin{cases} 0, & \text{if } i \neq j, \\ \mathbb{E}[e_j^2]\,(t_{j+1} - t_j), & \text{if } i = j. \end{cases}$$

Indeed, if $i < j$ then the random variables $e_i\, e_j\, \Delta B_i$ and ΔB_j are independent. Indeed, $e_i(B_{t_{i+1}} - B_{t_i})$ is $\mathcal{F}_{t_{i+1}}$-measurable and therefore is $\mathcal{F}_{t_j}$-measurable $(i{+}1 \leqslant j)$,

e_j is $\mathcal{F}_{t_j}$-measurable. Therefore, the product $e_i\, e_j\, \Delta B_i$ is $\mathcal{F}_{t_j}$-measurable. On the other hand, ΔB_{t_j} is independent of $\mathcal{F}_{t_j}$. Hence, $e_i\, e_j\, \Delta B_i$ and ΔB_j are independent for $i < j$. This implies that

$$\begin{aligned}
&\mathbb{E}[e_i e_j \Delta B_i\, \Delta B_j | \mathcal{F}_{t_j}] = e_i e_j \Delta B_i\, \mathbb{E}[B_{t_{j+1}} - B_{t_j} | \mathcal{F}_{t_j}] = 0, \quad \text{and} \\
&\mathbb{E}[e_i^2 \Delta B_i^2 | \mathcal{F}_{t_i}] = e_i^2\, \mathbb{E}[(B_{t_{i+1}} - B_{t_i})^2 | \mathcal{F}_{t_i}] = e_i^2\, \mathbb{E}[(B_{t_{i+1}} - B_{t_i})^2] = e_i^2 (t_{i+1} - t_i). \\
&\mathbb{E}[e_i^2 \Delta B_i^2] = \mathbb{E}[\mathbb{E}[e_i^2 \Delta B_i^2 | \mathcal{F}_{t_i}]] = \mathbb{E}[e_i^2](t_{i+1} - t_i).
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathbb{E}\Big[\Big(\int_S^T \phi(t,\omega) dB_t(\omega)\Big)^2\Big] = \sum_{i,j=0}^{N-1} \mathbb{E}[e_i e_j\, \Delta B_i \Delta B_j] = \sum_{j=1}^{N-1} \mathbb{E}[e_j^2]\, (t_{j+1} - t_j) \\
= \mathbb{E}\Big[\int_S^T \phi^2(t,\omega)\, dt\Big].
\end{aligned}$$

□

Approximation of the integrands from Φ by simple functions

Lemma 6.3. *Let $g \in \Phi$ be bounded and $g(\cdot, \omega)$ be continuous for each ω. Then, there exist simple functions $\phi_n \in \Phi$ such that*

$$\lim_{n\to\infty} \mathbb{E}\Big[\int_S^T (g - \phi_n)^2 dt\Big] = 0.$$

Proof. Define $\phi_n = \sum_{j=0}^{N-1} g(t_j, \omega)\ \mathbb{1}_{[t_j, t_{j+1})}(t)$. Since $g \in \Phi$, $g(t_j, \cdot)$ is $\mathcal{F}_{t_j}$-measurable, and therefore ϕ_n is $\mathcal{F}_t$-adapted. Clearly, since every indicator function $\mathbb{1}_{[t_j, t_{j+1})}(t)$ is $\mathcal{B}$-measurable, and every $g(t_j, \omega)$ is $\mathcal{F}_{t_j}$-measurable (and therefore $\mathcal{F}$-measurable), ϕ_n is $\mathcal{B} \times \mathcal{F}$-measurable. Hence $\phi_n \in \Phi$. Since $g(\cdot, \omega)$ is continuous for every ω, then

$$\lim_{n\to\infty} (g - \phi_n)^2 = 0 \qquad \text{for each } \omega.$$

Since g is bounded, ϕ_n is bounded as well, and therefore, by Lebesgue's theorem,

$$\lim_{n\to\infty} \mathbb{E}\Big[\int_S^T (g - \phi_n)^2 dt\Big] = 0.$$

□

Mollifiers

$$\rho(x) = \begin{cases} I^{-1} \exp\Big\{ -\frac{1}{1-x^2} \Big\}, & x \in (-1, 1), \\ 0, & x \notin (-1, 1), \end{cases}$$

where

$$I = \int_{-1}^{1} \exp\Big\{ -\frac{1}{1-x^2} \Big\} dx.$$

$$\rho_n(x) = n\rho(xn).$$

Definition 6.4. *Let ρ_n, $n \in \mathbb{N}$, be a sequence of Lebesgue-integrable functions on $\mathbb{R}$ possessing the properties:*

1. $\rho_n \geqslant 0$;
2. $\int_{-\infty}^{\infty} \rho_n(x)dx = 1$;
3. $\lim_{n\to\infty} \sup_{x\notin I} |\rho_n(x)| = 0$ *for any interval $I \subset \mathbb{R}$ containing* 0.

The sequence of functions ρ_n is called an ***approximate identity***.

Proposition 6.5. *Let $\{\rho_n\}$ be an approximate identity, and let $f \in L_p$. Then $\{f * \rho_n\}$, where $(f * \rho_n)(t) := \int_{-\infty}^{\infty} \rho_n(s-t)\, f(s)\, ds$, converges to f in the L_p-norm, $1 \leqslant p < \infty$.*

Proof. Without proof. A proof can be found in the book by Evans "Partial differential equations", 2010. □

Lemma 6.6. *Let $h \in \Phi$ be bounded. Then there exist bounded functions $g_n \in \Phi$ such that $g_n(\cdot, \omega)$ is continuous for every ω and n, and*

$$\lim_{n\to\infty} \mathbb{E}\Big[\int_S^T (h - g_n)^2\, dt\Big] = 0.$$

Proof. Let $M > 0$ be such that $|h(t,\omega)| \leqslant M$ for all (t,ω). For each $n \in \mathbb{N}$, let ψ_n be a non-negative, continuous function on $\mathbb{R}$ such that

1. $\psi_n(x) = 0 \quad \text{for} \quad x \leqslant -\frac{1}{n}$ and $x \geqslant 0$,
2. and $\int_{-\infty}^{\infty} \psi_n(x)dx = 1$.

Example of ψ_n: $\psi_n(x) = \rho_{2n}\big(x + \frac{1}{2n}\big)$. The sequence ψ_n is an approximate identity. Without loss of generality, we let $h = 0$ on $[0, S]$. Define

$$g_n(t,\omega) = \int_0^t \psi_n(s-t)\, h(s,\omega)\, ds \tag{47}$$

and note that the following properties of these functions. If $s > t$, then $\psi_n(s-t) = 0$, and, therefore, the integration in (47) can be extended to $[0,\infty)$. Furthermore, if $t \geqslant \frac{1}{n}$ and $s < 0$, then $s - t \leqslant -\frac{1}{n}$, and hence, $\psi_n(s-t) = 0$. This implies that for $t \geqslant \frac{1}{n}$, we can extend the integration in (47) to $(-\infty,\infty)$. Thus, if $t \geqslant \frac{1}{n}$, then $g_n(t,\omega) = (h(\cdot,\omega) * \psi_n)(t)$. Note that g_n is t-continuous for each ω and n, and $|g_n(t,\omega)| \leqslant M$. Since $h(s,\omega)$ is $\mathcal{B}\times\mathcal{F}$-measurable, so is the product $\psi_n(s-t)\, h(s,\omega)$,

and therefore the integral on the right-hand side of (47) is a limit of $\mathcal{B}\times\mathcal{F}$-measurable functions. Next, $h(s,\omega)$ is $\mathcal{F}_s$-measurable, and therefore $\mathcal{F}_t$-measurable since $s \leqslant t$. Moreover, the integral on the right-hand side of (47) can be represented as a limit of $\mathcal{F}_t$-measurable functions. Indeed, it sufficies to represent $h(s,\omega)$ as a limit of $\mathcal{F}_s$-measurable step functions. Therefore $g_n(t,\omega)$ is $\mathcal{F}_t$-measurable for all $t \geqslant 0$. Since g_n is bounded by M,

$$\mathbb{E}\Big[\int_S^T g_n(s,\omega)^2\,ds\Big] < \infty$$

Hence, $g_n \in \Phi$. By Proposition 6.5

$$\int_S^T (g_n(s,\omega) - h(s,\omega))^2\,ds \leqslant 2\int_S^T \big(g_n(s,\omega) - (h(\cdot,\omega) * \psi_n)(s)\big)^2\,ds + 2\int_S^T ((h(\cdot,\omega) * \psi_n)(t) - h(s,\omega))^2\,ds \to 0 \quad \text{as } n \to \infty.$$

Since both functions g_n and h are bounded, by Lebesgue's theorem,

$$\lim_{n\to\infty} \mathbb{E}\Big[\int_S^T (g_n(s,\omega) - h(s,\omega))^2\Big] = 0.$$

□

Lemma 6.7. *Let $f \in \Phi$. Then there exists a sequence $\{h_n\}$, $h_n \in \Phi$, such that it is bounded and*

$$\lim_{n\to\infty} \mathbb{E}\Big[\int_S^T (f - h_n)^2\,ds\Big] = 0. \tag{48}$$

Proof. Define

$$h_n(t,\omega) = \begin{cases} -n, & \text{if } f(t,\omega) < -n, \\ f(t,\omega) & \text{if } -n \leqslant f(t,\omega) \leqslant n, \\ n, & \text{if } f(t,\omega) > n. \end{cases}$$

Note that $|h_n(t,\omega)| \leqslant |f(t,\omega)|$. Hence $\mathbb{E}\big[\int_S^T h_n(s,\omega)^2 ds\big] < \mathbb{E}\big[\int_S^T f(s,\omega)^2 ds\big] < \infty$. Clearly h_n possesses also other properties of the class Φ by its construction. Therefore, $h_n \in \Phi$. Next, $h_n \to f$ for each $\omega \in \Omega$ and each $t \in [S,T]$. By Lebesgue's theorem, in (48) we can pass to the limit under the expectation and the integral signs which will imply (48). □

Summarizing Lemmas 6.3, 6.6, and 6.7, we obtain the following:

Lemma 6.8. *Let $f \in \Phi$, then there exists a system ψ_n of simple functions such that*

$$\lim_{n\to\infty} \mathbb{E}\Big[\int_S^T (f - \psi_n)^2\,ds\Big] = 0.$$

Proof. Proof follows from Lemmas 6.3, 6.6, and 6.7. □

Definition of the stochastic integral and Itô's isometry

Theorem 6.9. *Let $f \in \Phi$, and let $\psi_n \in \Phi$ be a sequence of simple functions such that*

$$\lim_{n\to\infty} \mathbb{E}\Big[\int_S^T (f-\psi_n)^2\,dt\Big] = 0. \tag{49}$$

Then, for every $t \in [S,T]$, there exist a random variable $I[f](t)$ such that

$$\lim_{n\to\infty} \mathbb{E}\,\Big(\,I[f](t) - I[\psi_n](t)\,\Big)^2 = 0$$

where $I[\psi_n](t) = \int_S^t \psi_n(s,\omega)ds$ is the stochastic integral of the simple function ψ_n defined by (45). Moreover $I[f](t)$ does not depend on the choice of a sequence ψ_n converging to f in the $L_2(\Omega\times[S,T])$-norm (as in (49)).

Definition 6.10 (Stochastic integral)**.** *The random variable $I[f](t)$ defined in the theorem above is called the* ***stochastic integral*** *of f with respect to a Brownian motion. It is denoted by the symbol*

$$I[f](t) = \int_S^t f(s,\omega)\,dB_s(\omega).$$

Proof of Theorem 6.9. By Theorem 6.2 (Itô's isometry for simple functions) we obtain:

$$\mathbb{E}\Big[\int_S^t (\psi_m-\psi_n)^2\,ds\Big] = \mathbb{E}\Big(\,I[\psi_n](t) - I[\psi_m](t)\,\Big)^2. \tag{50}$$

Since ψ_n converges to f in the $L_2(\Omega\times[S,T])$-norm, it is a Cauchy sequence with respect to this norm. Therefore, the sequence $I[\psi_n](t)$ on the right-hand side of (50) is a Cauchy sequence in the $L_2(\Omega)$-norm. But the $L_2(\Omega)$-space is complete. Hence, there exists a limit of $I[\psi_n](t)$ in this space, i.e. there exists a random variable $I[f](t)$ such that

$$\lim_{n\to\infty} \mathbb{E}\Big(\,I[f](t) - I[\psi_n](t)\,\Big)^2 = 0.$$

Suppose ψ_n' is another sequence of simple functions that converges to f with respect to the $L_2(\Omega\times[S,T])$-norm. Then, Itô's isometry implies:

$$\mathbb{E}\Big[\int_S^t (\psi_n'-\psi_n)^2\,ds\Big] = \mathbb{E}\Big(\,I[\psi_n](t) - I[\psi_n'](t)\,\Big)^2.$$

The left-hand side of this equality converges to zero since both sequences ψ_n and ψ_n' converge to f. Therefore the right-hand side converges to zero too. This proves that

$$\lim_{n\to\infty} I[\psi_n](t) = \lim_{n\to\infty} I[\psi_n'](t)$$

in the $L_2(\Omega)$-norm. Hence the stochastic integral $I[f](t) = \int_S^t f(s,\omega)dB_s(\omega)$ does not depend on the choice of the approximating system ψ_n. □

Corollary 6.11 (Itô's isometry)**.** *For any* $f \in \Phi$,

$$\mathbb{E}\Big[\Big(\int_S^t f(t,\omega)dB_s\Big)^2\Big] = \mathbb{E}\Big[\int_S^t f(s,\omega)^2\,ds\Big] \tag{51}$$

Proof. For any simple function ψ_n by Theorem 6.2 it holds that

$$\mathbb{E}\Big[\Big(\int_S^t \psi_n(s,\omega)dB_s\Big)^2\Big] = \mathbb{E}\Big[\int_S^t \psi_n(s,\omega)^2\,ds\Big]$$

Passing to the limit as $n \to \infty$ in the both parts of the above idenity we obtain (51). Namely, the convergence of expectations holds by the simple norm inequality: $\big|\|f\| - \|g\|\big| \leqslant \|f-g\|$ which was applied to the L_2-norms above. □

6.2 Properties of the stochastic integral

Theorem 6.12. *Let* $f, g \in \Phi(S,T)$*, where* $0 \leqslant S < U < T$*. Then*

(i) $\int\limits_S^t f\,dB_s = \int\limits_S^u f\,dB_s + \int\limits_u^t f\,dB_s$ *a.s.*

(ii) $\int\limits_S^t (cf+g)dB_t = c\cdot\int\limits_S^t f\,dB_s + \int\limits_S^t g\,dB_s$ *a.s., where* c *is a constant*

(iii) $\mathbb{E}\big[\int\limits_S^t f\,dB_s\big] = 0$ *for all* $S \leqslant t \leqslant T$*.*

(iv) $\int_S^t f\,dB_s$ *is* $\mathcal{F}_t$*-measurable for all* $S \leqslant t \leqslant T$*.*

Proof. Properties (i) and (ii) clearly hold for simple functions. If $f \in \Phi$ is arbitrary we take a sequence of simple functions ψ_n converging to f as $n \to \infty$. Taking limits from the both sides in (i) and (ii) in the $L_2(\Omega)$-norm we obtain (i) and (ii) for f. Let us prove (iii). Again, (iii) clearly hold when f is a simple function of form (44), i.e. when the stochastic integral $\int\limits_S^t f\,dB_s$ is given by (45). Indeed, for each summand in (45) we obtain:

$$\mathbb{E}\big[e_j(\omega)(B_{t_{j+1}} - B_{t_j})\big] = \mathbb{E}[e_j]\,\mathbb{E}[B_{t_{j+1}} - B_{t_j}] = 0.$$

The equality holds since e_j is $\mathcal{F}_{t_j}$-measurable and $B_{t_{j+1}} - B_{t_j}$ is independent of $\mathcal{F}_{t_j}$. Next if $\psi_n \to f$ as $n \to \infty$ in the $L_2(\Omega \times [S,t])$-norm, then

$$\begin{aligned}\Big|\,\mathbb{E}\Big[\int_S^t f\,dB_s\Big] - \mathbb{E}\Big[\int_S^t \psi_n\,dB_s\Big]\,\Big| &\leqslant \mathbb{E}\Big|\int_S^t f\,dB_s - \int_S^t \psi_n\,dB_s\Big| \\ &\leqslant \Big(\mathbb{E}\Big(\int_S^t f\,dB_s - \int_S^t \psi_n\,dB_s\Big)^2\Big)^{\frac{1}{2}} \to 0 \quad \text{as } n \to \infty.\end{aligned}$$

To show (iv), note that if $\psi \in \Phi$ is simple, then

$$\int_S^t \psi(s,\omega)\, dB_s = \sum_{j=0}^{N-1} e_j(\omega)\,(B_{t_{j+1}} - B_{t_j}),$$

where $t_{j_N} = t$, is an $\mathcal{F}_t$-measurable function. If $\psi_n \to f$ as $n \to \infty$ with respect to the $L_2(\Omega \times [S,s])$-norm, then $\int_S^t \psi_n\, dB_s \to \int_S^t f\, dB_s$ with respect to the $L_2(\Omega)$-norm, and therefore the limit $\int_S^t f\, dB_s$ is $\mathcal{F}_t$-measurable as a limit of $\mathcal{F}_t$-measurable functions. □

Note that the stochastic integral $\int\limits_0^t f\, dB_s$ can be also regarded as a stochastic process where t is the time. We are going to show its continuity in t and that it is a martingale. We will need the following lemma.

Lemma 6.13. *Suppose for each n, $Z_t^{(n)}$, $t \geqslant 0$, is a martingale with respect to the filtration $\mathcal{F}_t$, and for each $t \geqslant 0$, $Z_t^{(n)} \to Z_t$ as $n \to \infty$ with respect to the L_p-norm, $p \geqslant 1$. Then Z_t, $t \geqslant 0$, is a martingale.*

Proof. It suffices to prove the statement for the case $p = 1$ since the L_1-convergence is the weakest. We obtain:

$$\begin{aligned}\mathbb{E}\Big|\mathbb{E}[Z_t^{(n)} \,|\, \mathcal{F}_s] - \mathbb{E}[Z_t \,|\, \mathcal{F}_s]\Big| = \mathbb{E}\Big|\mathbb{E}[Z_t^{(n)} - Z_t \,|\, \mathcal{F}_s]\Big| &\leqslant \mathbb{E}\Big[\mathbb{E}[|Z_t^{(n)} - Z_t| \,|\, \mathcal{F}_s]\Big] \\ &= \mathbb{E}|Z_t^{(n)} - Z_t| \to 0 \quad \text{as } n \to \infty.\end{aligned}$$

Since $\mathbb{E}[Z_t^{(n)} \,|\, \mathcal{F}_s] = Z_s^{(n)}$, we obtain from the above inequalities that

$$\lim_{n\to\infty} Z_s^{(n)} = \mathbb{E}[Z_t \,|\, \mathcal{F}_s]$$

On the other hand

$$\lim_{n\to\infty} Z_s^{(n)} = Z_s$$

by assumption. Therefore $\mathbb{E}[Z_t \,|\, \mathcal{F}_s] = Z_s$. □

Theorem 6.14. *The stochastic integral integral $\int\limits_0^t f\, dB_s$ is a martingale.*

Proof. First we prove the statement when f is a simple function. Let us prove that if $s \leqslant t$, then

$$\mathbb{E}\Big[\int_s^t \psi\, dB_r \,|\, \mathcal{F}_s\Big] = 0.$$

Indeed, if $\psi = \sum_{j=0}^{N-1} e_j\ \mathbb{1}_{[t_{j+1},t_j)}$, then $\int_s^t \psi\, dB_r = \sum_{j=p}^{N-1} e_j\,(B_{t_{j+1}} - B_{t_j})$ where $t_p = s$. Compute the conditional expectation with respect to $\mathcal{F}_s$ of each summand of the stochastic integral. We obtain:

$$\begin{aligned}\mathbb{E}[e_j\,(B_{t_{j+1}} - B_{t_j}) \,|\, \mathcal{F}_s] = \mathbb{E}\big[\,\mathbb{E}\,[e_j\,(B_{t_{j+1}} - B_{t_j}) \,|\, \mathcal{F}_{t_j} \,|\, \mathcal{F}_s]\big] \\ = \mathbb{E}\big[e_j\,\mathbb{E}[(B_{t_{j+1}} - B_{t_j}) \,|\, \mathcal{F}_{t_j}] \,|\, \mathcal{F}_s\big] = 0.\end{aligned}$$

Clearly, $\mathbb{E}[(B_{t_{j+1}} - B_{t_j}) \,|\, \mathcal{F}_{t_j}] = 0$. Moreover, e_j is $\mathcal{F}_{t_j}$-measurable, and therefore was written outside the conditional expectation sign. By additivity, $\mathbb{E}\big[\int_s^t \psi \, dB_r \,|\, \mathcal{F}_s\big] = 0$. Now again applying the additivity property of the stochastic integral (property (i)), we obtain:

$$\mathbb{E}[\int_0^t \psi \, dB_r \,|\, \mathcal{F}_s\Big] = \mathbb{E}\Big[\int_0^s \psi \, dB_r \,|\, \mathcal{F}_s\Big] + \mathbb{E}\Big[\int_s^t \psi \, dB_r \,|\, \mathcal{F}_s\Big]$$
$$= \mathbb{E}\Big[\int_0^s \psi \, dB_r \,|\, \mathcal{F}_s\Big] = \int_0^s \psi \, dB_r.$$

By what was proved $\mathbb{E}\Big[\int_s^t \psi \, dB_r \,|\, \mathcal{F}_s\Big] = 0$. Moreover, Property (iv) of Theorem 6.12 implies that $\int_0^s \psi \, dB_r$ is an $\mathcal{F}_s$-measurable random variable. Note that I_n converge to I with respect to the L_2-norm. Let us apply Lemma 6.13 which says that in this case $I(t, \cdot)$ is a martingale. □

Lemma 6.15. *Let M_t be a continuous martingale, and let $0 < T < \infty$. Then for any $\varepsilon > 0$,*

$$P\Big(\sup_{[0,T]} |M_t| \geqslant \varepsilon\Big) \leqslant \frac{1}{\varepsilon^p} \mathbb{E}\big[|M_T|^p\big].$$

Proof. Without proof. □

Lemma 6.16 (Borel-Cantelli lemma)**.** *Let $E_k \subset \Omega$ be events such that*

$$\sum_{k=0}^{\infty} \mathbb{P}(E_k) < \infty,$$

then

$$\mathbb{P}\Big(\bigcap_{n=0}^{\infty} \bigcup_{k=n}^{\infty} E_k\Big) = 0,$$

Proof. Without proof. □

Theorem 6.17. *Let $f \in \Phi(0, T)$, then there exist a t-continuous version of the stochastic integral $\int_0^t f \, dB_s$.*

Remark. In other words, we have to prove that there exists a continuous process J_t such that for every $0 \leqslant t \leqslant T$

$$J_t = \int_0^t f \, dB_s \quad \text{a.s.}$$

Proof. Let ψ_n be a sequence of simple functions converging to f, and let $I_n(t, \omega) = \int_0^t \psi_n(s, \omega) \, dB_s$, $I(t, \omega) = \int_0^t f(s, \omega) \, dB_s$. Note that

$$I_n(t, \omega) = \sum_{j=0}^{N-2} e_j(\omega)(B_{t_{j+1}} - B_{t_j}) + e_{N-1}(\omega)(B_t - B_{t_{N-1}}),$$

and therefore it is a continuous martingale. Then $I_n(t,\cdot)-I_m(t,\cdot)$ is also a continuous martingale. By Lemma 6.15,

$$P\Big[\sup_{[0,T]}|I_n(t,\omega)-I_m(t,\omega)|>\varepsilon\Big]\leqslant\frac{1}{\varepsilon^2}\,\mathbb{E}[(I_n(T,\omega)-I_m(T,\omega))^2]$$
$$=\frac{1}{\varepsilon^2}\,\mathbb{E}\Big[\int_0^T(\psi_n-\psi_m)^2ds\Big]\to 0$$

as $n,m\to\infty$. Consequently,

$$\sup_{[0,T]}|I_n(t,\omega)-I_m(t,\omega)|>2^{-k}\Big]\to 0.$$

Hence, we can choose a subsequence $n_k\uparrow\infty$ such that

$$\mathbb{P}\Big[\sup_{[0,T]}|I_{n_k}(t,\omega)-I_{n_{k+1}}(t,\omega)|>2^{-k}\Big]<2^{-k}.$$

Consider the event

$$E_k=\big\{\omega:\sup_{[0,T]}|I_{n_k}(t,\omega)-I_{n_{k+1}}(t,\omega)|>2^{-k}\big\}.$$

$$\sum_{k=1}^{\infty}\mathbb{P}(E_k)<\sum_{k=1}^{\infty}2^{-k}<\infty.$$

By the Borel-Cantelli lemma,

$$\mathbb{P}\Big(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}E_k\Big)=0.$$

That is,

$$\mathbb{P}\Big[\forall\,n\geqslant 0,\,\exists\,k\geqslant n:\ \sup_{t\in[0,T]}|I_{n_{k+1}}-I_{n_k}|>2^{-k}\Big]=0.$$

Hence,

$$\mathbb{P}\Big[\exists\,n\geqslant 0,\,\forall\,k\geqslant n:\ \sup_{t\in[0,T]}|I_{n_{k+1}}-I_{n_k}|\leqslant 2^{-k}\Big]=1.$$

For almost all $\omega\in\Omega$,

$$\exists\,n\geqslant 0,\,\forall\,k\geqslant n:\ \sup_{t\in[0,T]}|I_{n_{k+1}}-I_{n_k}|\leqslant 2^{-k}.$$

This implies that for almost all $\omega\in\Omega$, I_{n_k} converges uniformly in $t\in[0,T]$. Let J_t be the limit of I_{n_k}. Clearly J_t is continuous in t a.s. Next, since $I_{n_k}(t,\omega)\to I(t,\omega)$ as $k\to\infty$ for all t in the $L_2(\Omega)$-norm, then $I(t,\cdot)=J_t$ a.s. for all $t\in[0,T]$. The theorem is proved. □

Everywhere below the stochastic integral $\int_0^t f\,dB_s$ means its continuous version.

6.3 An extension of the stochastic intergral and multi-dimensional stochastic intergrals

Define a multi-dimensional stochastic integral. Let B_t^k be the k-th coordinate of an n-dimensional Brownian motion B_t. Assume that the filtration $\mathcal{F}_t$ in Definition 6.1 is such that each B_t^k is an $\mathcal{F}_t$-Brownian motion. Example of $\mathcal{F}_t$ is the following:

$$\mathcal{F}_t^n = \sigma(B_{s_1}^1, \ldots, B_{s_n}^n; s_1, \ldots, s_n \leqslant t). \tag{52}$$

Each Brownian motion B_t^k is a martingale with respect to $\mathcal{F}_t^n$.

Definition 6.18. *Let $B_t = (B_t^1, \ldots B_t^n)$ be an n-dimensional Brownian motion, and let $v = v_{ij}(t, \omega)$ be an $n \times m$ matrix such that $v_{ij} \in \Phi$ for all $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant 1$. Let us define the stochastic integral as follows:*

$$\int_S^t v\, dB_s = \int_S^t \begin{pmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & & \vdots \\ v_{n1} & \cdots & v_{mn} \end{pmatrix} \begin{pmatrix} dB_s^1 \\ \vdots \\ dB_s^n \end{pmatrix}$$

where the right-hand side of this formula is an m-dimensional vector whose ith component equals to

$$\sum_{j=1}^{n} \int_S^t v_{ij}\, dB_s^j.$$

7. Itô's formula

7.1 The one-dimensional Itô formula

Definition 7.1 (1-dimensional Itô process.)**.** *Let B_t be a 1-dimensional $\mathcal{F}_t$-Brownian motion on $(\Omega, \mathcal{F}, P)$. A 1-dimensional Itô process is a stochastic process X_t on $(\Omega, \mathcal{F}, P)$ of the form:*

$$X_t = X_0 + \int_0^t u(s, \omega) ds + \int_0^t v(s, \omega) dB_s \tag{53}$$

where $u, v \in \Phi(0, T, \mathcal{F}_t)$ for all $T > 0$.

Note that, an Itô process X_t is a semimartingale. For the theorem below, we will need the concept of the *covariance process* of two semimartingales.

Definition 7.2. *Let $X_t = A_t + M_t$ and $Y_t = C_t + N_t$ be two semimartingales where M_t and N_t are their martingale parts. The covariance process $\langle X, Y \rangle_t$ is defined as follows:*

$$\langle X, Y \rangle_t = \langle M, N \rangle_t.$$

Theorem 7.3. *Let X_t be an Itô process given by (53), and let $g(t,x) \in \mathrm{C}^2([0,\infty)\times\mathbb{R})$, i.e. g is twice continuously differentiable on $[0,\infty)\times\mathbb{R}$). Then*

$$Y_t = g(t, X_t)$$

is again an Itô process, and

$$Y_t = g(0,X_0) + \int_0^t \frac{\partial g}{\partial t}(s,X_s)\,ds + \int_0^t \frac{\partial g}{\partial x}(s,X_s)\,dX_s + \frac{1}{2}\int_0^t \frac{\partial^2 g}{\partial x^2}(s,X_s) d\langle X\rangle_s. \tag{54}$$

Remark. In (54), $dX_t = u(t,\omega)\,dt + v(t,\omega)\,dB_t$ according to (53). Also, the last integral is understood as a Riemann-Stieltjes integral with respect to the increasing funcion $\langle X\rangle_s$.

Lemma 7.4. *Let X_t be an Itô process given by (53). Then*

$$\langle X\rangle_t = \left\langle \int_0^\bullet v(s,\cdot)\,dB_s \right\rangle_t = \int_0^t v^2(s,\cdot)\,ds.$$

Proof. Left as an exercise. □

Remark. Equivalently (54) can be written as follows:

$$g(t,X_t) = g(0,X_0) + \int_0^t \Big(\frac{\partial g}{\partial s}(s,X_s) + u(s,\cdot)\,\frac{\partial g}{\partial x}(s,X_s) + \frac{1}{2}\,v(s,\cdot)^2\,\frac{\partial^2 g}{\partial x^2}(s,X_s)\Big)ds + \int_0^t v(s,\cdot)\,\frac{\partial g}{\partial x}(s,X_2)\,dB_s. \tag{55}$$

Note that (55) is an Itô process in the sense of Definition 7.1.

Proof of Theorem 7.3. First we assume that $g \in \mathrm{C}_b^3([0,T]\times\mathbb{R})$, i.e. three times continuously differentiable with bounded derivatives and that $u, v \in \Phi(0,T)$ are bounded simple functions associated to a partition $\mathcal{P}_0 = \{0 = \tau_0 < \cdots < \tau_n = t\}$. Further, we consider a partition $\mathcal{P} = \{0 = t_0 < \cdots < t_n = t\}$ which is always a refinement of $\mathcal{P}_0$. Using Taylor's expansion we obtain:

$$\begin{aligned} g(t,X_t) = g(0,X_0) + \sum_j \Delta g(t_j,X_{t_j}) &= g(0,X_0) + \sum_j \frac{\partial g}{\partial t}\,\Delta t_j + \sum_j \frac{\partial g}{\partial x}\,\Delta X_j \\ &+ \frac{1}{2}\sum_j \frac{\partial^2 g}{\partial t^2}\,(\Delta t_j)^2 + \sum_j \frac{\partial^2 g}{\partial t\partial x}\,(\Delta t_j)(\Delta X_j) + \frac{1}{2}\sum_j \frac{\partial^2 g}{\partial x^2}\,(\Delta X_j)^2 + \sum_j R_j \end{aligned}$$

where $\frac{\partial g}{\partial t}$, $\frac{\partial g}{\partial x}$, $\frac{\partial^2 g}{\partial t\partial x}$ etc. are evaluated at the points (t_j, X_{t_j}), $\Delta t_j = t_{j+1} - t_j$, $\Delta X_j = X_{t_{j+1}} - X_{t_j}$, $\Delta g(t_j,X_{t_j}) = g(t_{j+1},X_{t_{j+1}}) - g(t_j,X_{t_j})$, and $R_j = O(|\Delta t_j|^3 + |\Delta X_j|^3)$ for all j. As $\Delta t_j \to 0$,

$$\sum_j \frac{\partial g}{\partial t}\Delta t_j = \sum_j \frac{\partial g}{\partial t}(t_j,X_{t_j})\Delta t_j \to \int_0^t \frac{\partial g}{\partial s}(s,X_s)\,ds$$

$$\sum_j \frac{\partial g}{\partial x}\Delta X_j = \sum_j \frac{\partial g}{\partial x}(t_j,X_{t_j})\Delta X_j \to \int_0^t \frac{\partial g}{\partial x}(s,X_s)\,dX_s.$$

Since u and v are simple functions, we have $\Delta X_j = u_j \Delta t_j + v_j \Delta B_j$, where $\Delta B_j = B_{t_{j+1}} - B_{t_j}$, $u_j = u(t_j, \omega)$ and $v_j = v(t_j, \omega)$. Therefore,

$$\sum_j \frac{\partial^2 g}{\partial x^2} (\Delta X_j)^2 = \sum_j \frac{\partial^2 g}{\partial x^2} u_j^2 (\Delta t_j)^2 + 2 \sum_j \frac{\partial^2 g}{\partial x^2} u_j v_j (\Delta t_j)(\Delta B_j) + \sum_j \frac{\partial^2 g}{\partial x^2} v_j^2 (\Delta B_j)^2. \quad (56)$$

The first term in (56) is bounded, and tends to zero pointwise as the mesh of $\mathcal{P}$ tends to zero. For the second term we have:

$$\mathbb{E}\Big(\sum_j \frac{\partial^2 g}{\partial x^2} u_j v_j (\Delta t_j)(B_{t_{j+1}} - B_{t_j}) \Big)^2$$

$$= \sum_{i,j} \mathbb{E}\Big(\frac{\partial^2 g}{\partial x^2} u_j v_j \Big) \Big(\frac{\partial^2 g}{\partial x^2} u_i v_i \Big) (\Delta t_i)(\Delta t_j)(B_{t_{i+1}} - B_{t_i})(B_{t_{j+1}} - B_{t_j})$$

$$= \sum_j \mathbb{E}\Big(\frac{\partial^2 g}{\partial x^2} u_j v_j \Big)^2 (\Delta t_j)^2 (B_{t_{j+1}} - B_{t_j})^2 = \sum_j \mathbb{E}\Big(\frac{\partial^2 g}{\partial x^2} u_j v_j \Big)^2 (\Delta t_j)^3 \to 0 \quad \text{as } \Delta t_j \to 0$$

by boundedness of u, v, and $\frac{\partial^2 g}{\partial x^2}$, and by the fact that $\mathbb{E}[(B_{t_{j+1}} - B_{t_j})^2] = \Delta t_j$. Let us prove that the last term in (56) tends to $\frac{\partial^2 g}{\partial x^2} v^2$ with respect to the $L^2(\Omega)$-norm as the mesh $\max_j |\Delta t_j|$ tends to zero. Define $a(t) = \frac{\partial^2 g}{\partial x^2}(t, X_t)\, v^2(t, \omega)$, and $a_j = a(t_j)$. We have:

$$\mathbb{E}\Big(\sum_j a_j (\Delta B_j)^2 - \sum_j a_j \Delta t_j \Big)^2 = \sum_{i,j} \mathbb{E}\big[a_i a_j \big((\Delta B_i)^2 - \Delta t_i \big) \big((\Delta B_j)^2 - \Delta t_j \big) \big]. \quad (57)$$

If $i < j$ then $a_i\, a_j \big((\Delta B_i)^2 - \Delta t_i \big)$ is $\mathcal{F}_{t_j}$-measurable, and therefore

$$\mathbb{E}\big[a_i a_j \big((\Delta B_i)^2 - \Delta t_i \big) \big((\Delta B_j)^2 - \Delta t_j \big) \big] = \mathbb{E}\big[a_i a_j ((\Delta B_i)^2 - \Delta t_i)\, \mathbb{E}[\big((\Delta B_j)^2 - \Delta t_j \big) | \mathcal{F}_{t_j}] \big] = 0.$$

Similarly for the case $i > j$. For the case $i = j$, we obtain:

$$\sum_j \mathbb{E}\big[a_j^2 \big((\Delta B_j)^2 - \Delta t_j \big)^2 \big] = \mathbb{E}\big[a_j^2\, \mathbb{E}\big[\big((\Delta B_j)^2 - \Delta t_j \big)^2 | \mathcal{F}_{t_j} \big] \big] = \mathbb{E}[a_j^2]\, \mathbb{E}\big((\Delta B_j)^2 - \Delta t_j \big)^2$$

since a_j^2 is $\mathcal{F}_{t_j}$-measurable and $(\Delta B_j)^2$ is independent of $\mathcal{F}_{t_j}$. Further, we have:

$$\mathbb{E}\big((\Delta B_j)^2 - \Delta t_j \big)^2 = \mathbb{E}\big[(\Delta B_j)^4 - 2 (\Delta B_j)^2 \Delta t_j + (\Delta t_j)^2 \big] = 3(\Delta t_j)^2 - 2(\Delta t_j)^2 + (\Delta t_j)^2 = 2(\Delta t_j)^2.$$

Therefore,

$$\sum_j \mathbb{E}\big[a_j^2 \big((\Delta B_j)^2 - \Delta t_j \big)^2 \big] = 2 \sum_j \mathbb{E}[a_j^2]\, (\Delta t_j)^2$$

which tends to zero as the mesh $\max|\Delta t_j|$ goes to zero. This implies, in turn, that the right-hand side of (57) converges to zero as $\max|\Delta t_j| \to 0$. But

$$\sum_j a_j \Delta t_j \to \int_0^t a(s)\,ds, \quad \text{as } \max_j |\Delta t_j| \to 0,$$

and therefore,

$$\sum_j a_j (\Delta B_j)^2 \to \int_0^t a(s)\,ds \quad \text{as } \max_j |\Delta t_j| \to 0 \quad \text{in } L_2(\Omega),$$

$$\mathbb{E}\Big(\sum_j a_j (\Delta B_j)^2 - \int_0^t a(s)\,ds\Big)^2 \to 0 \quad \text{as } \max_j |\Delta t_j| \to 0.$$

Clearly, $\sum_j R_j \to 0$ as $\max_j |\Delta t_j| \to \infty$. Indeed, there exists a bounded function $\varphi : \mathbb{R} \to \mathbb{R}$ (it depends on the bounded third-order derivatives of g by Taylor's formula) such that

$$R_j = \big((\Delta t_j)^3 + |\Delta B_j|^3\big)\varphi\big(t_j, t_{j+1}, B_{t_j}, B_{t_{j+1}}\big).$$

Note that

$$\sum_j (\Delta t_j)^3 \varphi\big(t_j, t_{j+1}, B_{t_j}, B_{t_{j+1}}\big) \to 0 \quad \text{as } |\mathcal{P}| \to 0 \text{ a.s.}$$

Further

$$\mathbb{E}\Big(\sum_j |\Delta B_j|^3 \varphi\big(t_j, t_{j+1}, B_{t_j}, B_{t_{j+1}}\big)\Big)^2 \leqslant C\mathbb{E}\sum_{i,j} |\Delta B_j|^3 |\Delta B_i|^3$$

$$\leqslant C\sum_{i\neq j} \mathbb{E}|\Delta B_j|^3 \mathbb{E}|\Delta B_i|^3 + C\sum_i \mathbb{E}|\Delta B_i|^6 \leqslant C_1 \sum_{i\neq j} (\Delta t_j)^{\frac{3}{2}} (\Delta t_i)^{\frac{3}{2}} + C_1 \sum_i (\Delta t_i)^3$$

$$= C_1\Big(\sum_i (\Delta t_i)^{\frac{3}{2}}\Big)^2 \leqslant C_1 |\mathcal{P}| t^2 \to 0 \quad \text{as } |\mathcal{P}| \to 0$$

This proves Itô's formula for the case when $g \in \mathrm{C}_b^3([0,\infty)\times\mathbb{R})$ and u and v are bounded simple functions. To prove the general case let us note that we can always approximate $u, v \in \Phi(0,T)$ by bounded simple functions with respect to the $L_2(\Omega)$-norm. Let $u, v \in \Phi(0,T)$ and let u_n, v_n be bounded simple functions converging to u, v, respectively, with respect to the $L_2([0,T]\times\Omega)$-norm. Let

$$X_t^{(n)} = X_0 + \int_0^t u_n(s)\,ds + \int_0^t v_n(s)\,dB_s$$

Then,

$$X_t^{(n)} - X_t = \int_0^t (u_n(s) - u(s))\,ds + \int_0^t (v_n(s) - v(s))\,dB_s,$$

$$\mathbb{E}|X_t^{(n)} - X_t|^2 \leq 2\,\mathbb{E}\Big|\int_0^t (u_n(s) - u(s))\,ds\Big|^2 + 2\,\mathbb{E}\Big|\int_0^t (v_n(s) - v(s))\,dB_s\Big|^2,$$

$$\mathbb{E}|X_t^{(n)} - X_t|^2 \leq 2t\,\mathbb{E}\int_0^t (u_n(s) - u(s))^2\,ds + 2\,\mathbb{E}\int_0^t (v_n(s) - v(s))^2\,ds.$$

Let us find subsequences $u_{n_k}, v_{n_k}, X^{(n_k)}$ such that for every $t > 0$,

$$\lim_{k\to\infty} |X_t^{(n_k)} - X_t| = \lim_{k\to\infty} \int_0^t (u_{n_k}(s) - u(s))^2 ds = \lim_{k\to\infty} \int_0^t (v_{n_k}^2(s) - v^2(s))^2 ds = 0.$$

Passing to the limit in (55), we conclude that (55) holds for X_t with arbitrary $u, v \in \Phi(0,T)$. We show the convergence for two terms in (55). For the rest of the terms it can be done likewise. For simplicity of notation, we again write v_n and $X_t^{(n)}$ instead of v_{n_k} and $X_t^{(n_k)}$. We have

$$\begin{aligned}
&\int_0^t v_n^2 g_{xx}(s, X_s^{(n)}) ds - \int_0^t v^2 g_{xx}(s, X_s) ds \\
&= \int_0^t (v_n \pm v)^2 g_{xx}(s, X_s^{(n)}) ds - \int_0^t v^2 g_{xx}(s, X_s) ds = \int_0^t (v_n - v)^2 g_{xx}(s, X_s^{(n)}) ds \\
&\qquad + 2\int_0^t v(v_n - v) g_{xx}(s, X_s^{(n)}) ds + \int_0^t v^2 (g_{xx}(s, X_s) - g_{xx}(s, X_s^{(n)}) ds.
\end{aligned}$$

In the last term on the right-hand side, we can pass to the limit under the integral sign by the dominated convergence theorem. The first and the second term can be bounded by

$$\sup|g_{xx}| \int_0^t (v_n - v)^2 ds + 2\sup|g_{xx}| \sqrt{\int_0^t (v_n - v)^2 ds} \sqrt{\int_0^t v^2 ds} \to 0, \quad n \to \infty.$$

Furher, subtracting and adding $\int_0^t v(s)\, g_x(s, X_s^{(n)})\, dB_s$, by Itô's isometry,

$$\begin{aligned}
&\Big\| \int_0^t v(s)\, g_x(s, X_s)\, dB_s - \int_0^t v_n(s)\, g_x(s, X_s^{(n)}) dB_s \Big\|_{L_2(\Omega)} \\
&\leqslant \Big\| \int_0^t (v(s) - v_n(s))\, g_x(s, X_s^{(n)}) dB_s \Big\|_{L_2(\Omega)} + \Big\| \int_0^t v(s) \big(g_x(s, X_s^{(n)}) - g_x(s, X_s)\big) dB_s \Big\|_{L_2(\Omega)} \\
&\qquad = \|(v - v_n)\, g_x(\,\cdot\,, X_\cdot^{(n)})\|_{L_2([0,t]\times\Omega)} + \|v\,\big(g_x(\,\cdot\,, X_\cdot^{(n)}) - g_x(\,\cdot\,, X_\cdot)\|_{L_2([0,t]\times\Omega)} \\
&\qquad\qquad \to 0 \quad (n \to \infty).
\end{aligned}$$

Thus, u and v in (55) can be arbitrary functions from $\Phi(0,T)$ for all $T > 0$.

Now take $g \in \mathrm{C}^2([0,\infty) \times \mathbb{R})$ such that $\operatorname{supp} g(t,\,\cdot\,) \subset K$ for all $t > 0$, where K is a compact. Let $g_n := \rho_{\frac{1}{n}} * g$ (the convolution is considered in (t,x)). Then, $\|g - g_n\|_{\mathrm{C}^2([0,t]\times K_1)} \to 0$, as $n \to \infty$, where $K_1 \supset K$ is a 1-neighborhood of K. Hence $\|g_n\|_{\mathrm{C}^2([0,t]\times K_1)}$ is uniformly bounded. By the dominated convergence theorem, (55) holds for $g \in \mathrm{C}^2([0,\infty \times \mathbb{R})$ such that $\operatorname{supp} g(t,\,\cdot\,) \subset K$ for all $t > 0$, where K is a compact.

Now let $g \in \mathrm{C}^2([0,\infty) \times \mathbb{R})$ and $\chi_R := \rho * \mathbb{1}_{B_{R+1}(0)}$, where $\rho \in \mathrm{C}_c^\infty(\mathbb{R})$ is a standard mollifier. Thus $\chi_R \in \mathrm{C}^\infty(\mathbb{R})$ and the derivatives are bounded for all $k \geqslant 0$ uniformly in $R > 0$. Indeed,

$$D^k \chi_R(x) = D^k(\rho * \mathbb{1}_{B_{R+1}(0)})(x) = (D^k \rho) * \mathbb{1}_{B_{R+1}(0)}(x) = \int_{\mathbb{R}} D^k \rho(x-y)\, \mathbb{1}_{B_{R+1}(0)}(y)\, dy.$$

Hence

$$|D^k \chi_R(x)| \leqslant \|D^k \rho\|_{L^1(\mathbb{R})}.$$

Let

$$g_R(t,x) := \chi_R(x) g(t,x).$$

Note that $g_R(t,x) \in \mathrm{C}_b^2([0,t]\times\mathbb{R})$ and $\operatorname{supp} g_R(t,\,\cdot\,) \subset B_{R+2}(0)$. Moreover $g_R(t,x) = g(t,x)$ for $|x| \leqslant R$. Then, as $R \to \infty$, $g_R(t,x) \to g(t,x)$ for each $x \in \mathbb{R}$. Further we have that there exist a constant C_k, depending only on k, such that

$$|D^k g_R(t,x)| \leqslant C_k \sum_{\ell=1}^{k} |D^\ell g(t,x)|.$$

In particular, $|g_R(t,x)| \leqslant |g(t,x)|$. Therefore, in (55), we can pass to the limit as $R \to \infty$ by the dominated convergence theorem. This proves Itô's formula.

□

7.2 The multi-dimensional Itô formula

Let $B_t = (B_t^1, \dots, B_t^m)$ be an m-dimensional $\mathcal{F}_t$-Brownian motion.

Definition 7.5 (n-dimensional Itô process)**.** *A process X_t is called an n-dimensional **Itô process** if it can be written in the form:*

$$\begin{cases} X_t^1 = X_0^1 + \int_0^t u_1(s,\omega)\,ds + \sum_{j=1}^m \int_0^t v_{1j}(s,\omega)\,dB_s^j \\ \vdots \\ X_t^n = X_0^n + \int_0^t u_n(s,\omega)\,ds + \sum_{j=1}^m \int_0^t v_{nj}(s,\omega)\,dB_s^j \end{cases} \tag{58}$$

where $u_i \in L_2([0,\infty)\times\Omega)$, $1 \leqslant i \leqslant n$, are $\mathcal{F}_t$-adapted, and $v_{ij} \in \Phi(0,T,\mathcal{F}_t)$ for all $T > 0$, $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant m$.

In matrix notation we can rewrite (58) as follows:

$$dX_t = u_t dt + v_t dB_t, \tag{59}$$

where

$$X_t = \begin{pmatrix} X_t^1 \\ \vdots \\ X_t^2 \end{pmatrix}, \quad u_t = \begin{pmatrix} u_1(t,\cdot) \\ \vdots \\ u_n(t,\cdot) \end{pmatrix}, \quad v_t = \begin{pmatrix} v_{11} & \cdots & v_{1m} \\ \vdots & & \vdots \\ v_{n1} & \cdots & v_{nm} \end{pmatrix}, \quad dB_t = \begin{pmatrix} dB_t^1 \\ \vdots \\ dB_t^m \end{pmatrix}.$$

Lemma 7.6. *Let $f, g \in \Phi(0,T)$ and B^i, B^j be independent Brownian motions. Then*

$$\mathbb{E}\left[\int_0^t f_r\,dB_r^i \int_0^t g_r\,dB_r^j\right] = 0.$$

Proof.

$$\varphi_n \to f_s \text{ in } L^2([0,T]\times\Omega), \qquad \varphi_n \text{ simple},$$
$$\psi_n \to g_s \text{ in } L^2([0,T]\times\Omega), \qquad \psi_n \text{ simple},$$

Let $\mathcal{P} := \{0 = t_0 < \cdots < t_n = t\}$ be a partition. Then,

$$\varphi_n = \sum_{k=0}^{n-1} e_k \, \mathbb{1}_{[t_k, t_{k+1})}, \qquad \psi_n = \sum_{k=0}^{n-1} \hat{e}_k \, \mathbb{1}_{[t_k, t_{k+1})},$$

where

$$e_k \text{ is } \mathcal{F}_{t_k}\text{-measurable}, \qquad \hat{e}_\ell \text{ is } \mathcal{F}_{t_\ell}\text{-measurable}.$$

We have,

$$\mathbb{E}\left[\left(\sum_{k=0}^{n-1} e_k (B^i_{t_{k+1}} - B^i_{t_k})\right)\left(\sum_{\ell=0}^{n-1} \hat{e}_\ell (B^i_{t_{\ell+1}} - B^i_{t_\ell})\right)\right] = \sum_{k,\ell=0}^{n-1} \mathbb{E}\left[e_k \hat{e}_\ell (B^i_{t_{k+1}} - B^i_{t_k})(B^j_{t_{\ell+1}} - B^j_{t_\ell})\right]$$

Assume $t_\ell < t_k$ and let $\Delta_k B^i := B^i_{t_{k+1}} - B^i_{t_k}$, $\Delta_\ell B^j := B^j_{t_{\ell+1}} - B^j_{t_\ell}$. We have

$$\mathbb{E}\big[e_k \hat{e}_\ell \Delta_k B^i \Delta_\ell B^j\big] = \mathbb{E}\big[e_k \hat{e}_\ell \Delta_k B^i \, \mathbb{E}[\Delta_\ell B^j \,\big|\, \mathcal{F}_{t_\ell} = 0]\big]$$

Indeed, $\mathcal{F}_{t_{k+1}} \subset \mathcal{F}_{t_\ell}$ and $\Delta_k B^i$ is $\mathcal{F}_{t_{k+1}}$-measurable.

Now let $t_k = t_\ell$ (i.e. $k = \ell$). We have

$$\mathbb{E}\big[\mathbb{E}\big[e_k \hat{e}_k \, \Delta_k B^i \, \Delta_k B^j \;\big|\; \mathcal{F}_{t_k}\big]\big] = \mathbb{E}\big[e_k \hat{e}_k \, \mathbb{E}\big[\Delta_k B^i \, \Delta_k B^j \;\big|\; \mathcal{F}_{t_k}\big]\big] = \mathbb{E}\big[e_k \hat{e}_k \, \mathbb{E}\big[\Delta_k B^i \, \Delta_k B^j\big]\big] = 0$$

since $\mathbb{E}[\Delta_k B^i \, \Delta_k B^j] = 0$ by the independence of B^i_t and B^j_t. □

Lemma 7.7 (Itô's isometry)**.** *Let B_t be an m-dimensional Brownian motion and the integran v_t be as above. Then,*

$$\mathbb{E}\left|\int_S^t v_r \, dB_r\right|^2 = \mathbb{E}\int_S^t |v_r|^2 \, dr$$

Proof.

$$\mathbb{E}\left|\sum_{\ell=1}^n \int_S^t v_r^{k\ell} \, dB_r^\ell\right|^2 = \sum_{\ell=1}^n \mathbb{E}\left|\int_S^t v_r \, dB_r\right|^2 = \mathbb{E}\int_S^t |v_r|^2 \, dr \int_S^t (v_r^{k\ell})^2 \, dr$$
$$+ 2\sum_{i\neq j} \mathbb{E}\left[\int_S^t v_r^{ki} \, dB_r^i \int_S^t v_r^{kj} \, dB_r^j\right] = \sum_{\ell=1}^n \mathbb{E}\int_S^t (v_r^{k\ell})^2 \, dr = \mathbb{E}\int_S^t \sum_{\ell=1}^n (v_r^{k\ell})^2 \, dr.$$

By Lemma 7.6,

$$\mathbb{E}\left[\int_S^t v_r^{ki}\,dB_r^i \int_S^t v_r^{kj}\,dB_r^j\right] = 0.$$

Let $I(t) := \int_S^t v_r\,dB_r$, and $I^k(t)$ be its k-th component. We have

$$\mathbb{E}|I(t)|^2 = \sum_{k=1}^m \mathbb{E}|I^k(t)|^2 = \mathbb{E}\int_S^t \sum_{k=1}^m \sum_{\ell=1}^n (v_r^{k\ell})^2\,dr = \mathbb{E}\int_S^t |v_r|^2\,dr.$$

□

Theorem 7.8 (Itô's formula)**.** *Let X_t be an n-dimensional Itô's process given by (59), and let $g(t,x) = (g_1(t,x), \cdots, g_p(t,x)) \in \mathrm{C}^2([0,\infty)\times\mathbb{R}^n, \mathbb{R}^p)$. Then the process*

$$Y_t(\omega) = g(t, X_t)$$

is again an Itô process, whose k'th component Y_t^k is given by

$$Y_t^k = g^k(0,X_0) + \int_0^t \frac{\partial g^k}{\partial s}(s,X_s)\,ds + \sum_{i=1}^n \int_0^t \frac{\partial g^k}{\partial x_i}(s,X_s)\,dX_s^i + \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \int_0^t \frac{\partial^2 g^k}{\partial x_i \partial x_j}(s,X_s)\,d\langle X^i, X^j\rangle_s$$

or equivalently,

$$Y_t^k = g^k(0,X_0) + \int_0^t \frac{\partial g^k}{\partial s}(s,X_s)\,ds + \sum_{i=1}^n \int_0^t \frac{\partial g^k}{\partial x_i}(s,X_s)\,u_i(s,\cdot)\,ds$$
$$+\sum_{i=1}^n\sum_{j=1}^m \int_0^t \frac{\partial g^k}{\partial x_i}(s,X_s)\,v_{ij}(s,\cdot)\,dB_s^j + \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n\sum_{p=1}^m \int_0^t \frac{\partial^2 g^k}{\partial x_i\partial x_j}(s,X_s)\,(v_{ip}v_{jp})(s,\cdot)\,ds.$$

Lemma 7.9. *Let B_t^i and B_t^j be independent Brownian motions. Then*

$$\langle B^i, B^j\rangle_t = 0.$$

Proof. It suffices to prove that $B_t^i B_t^j$ is an $\mathcal{F}_t$-martingale. Indeed, taking into account that both B_t^i and B_t^j are $\mathcal{F}_t$-martingales, for $s < t$ we obtain:

$$0 = \mathbb{E}[(B_t^i - B_s^i)]\,\mathbb{E}[(B_t^j - B_s^j)] = \mathbb{E}[(B_t^i - B_s^i)(B_t^j - B_s^j)] = \mathbb{E}[(B_t^i - B_s^i)(B_t^j - B_s^j)\,|\,\mathcal{F}_s]$$
$$= \mathbb{E}[B_t^i B_t^j\,|\,\mathcal{F}_s] - B_s^i B_s^j.$$

□

Lemma 7.10. *Let X_t^i and X_t^j be two 1-dimensional Itô's processes with the martingale parts $\int_0^t v_i(s,\cdot)\,dB_s^i$ and resp. $\int_0^t v_j(s,\cdot)\,dB_s^j$. Then*

$$\begin{aligned}\langle X^i, X^j\rangle_t &= \Big\langle \int_0^\bullet v_i(s,\cdot)\,dB_s^i\,,\int_0^\bullet v_j(s,\cdot)\,dB_s^j\Big\rangle_t = \int_0^t (v^iv^j)(s,\cdot)\,d\langle B^i,B^j\rangle_s\\ &= \int_0^t (v^iv^j)(s,\cdot)\,\delta_{ij}\,ds\end{aligned}$$

where $\delta_{ij}=0$ if $i\neq j$ and $\delta_{ii}=1$.

Proof. Left as an exercise. □

The proof of Theorem 7.8 is similar to the proof of Theorem 7.3, and therefore is omitted.

7.3 Itô representation theorem

Let $B_t=(B_t^1,\dots,B_t^n)$ be an n-dimensional Brownian motion and let $\mathcal{F}_t=\mathcal{F}_t^n$ be the filtration defined by (52).

Lemma 7.11. *The linear span of random variables of the type*

$$\exp\Big\{\int_0^T h(t)dB_t-\frac{1}{2}\int_0^T h^2(t)\,dt\Big\} \tag{60}$$

where $h\in L_2([0,T])$ is deterministic, is dense in $L_2(\mathcal{F}_T,\Omega)$.

Proof. Without proof. □

Theorem 7.12 (Itô's representation theorem)**.** *Let $B_t=(B_t^1,\dots,B_t^n)$ be an n-dimensional Brownian motion, and let the filtration $\mathcal{F}_t^n$ be defined by (52). Let $F\in L_2(\mathcal{F}_T^n,\Omega)$. Then there exists a unique stochastic process $f(t,\omega)\in\Phi(0,T,\mathcal{F}_t^n)$ such that*

$$F(\omega)=\mathbb{E}[F]+\int_0^T f(t,\omega)\,dB_t. \tag{61}$$

Proof. For simplicity we consider the case $n=1$. The proof in the general case is similar. First we assume that F has the form (60), i.e.

$$F(\omega)=\exp\Big(\int_0^T h(t)dB_t-\frac{1}{2}\int_0^T h^2(t)dt\Big) \tag{62}$$

for some $h(t)\in L_2[0,T]$. Define

$$Y_t=\exp\Big(\int_0^t h(s)dB_s-\frac{1}{2}\int_0^t h^2(s)ds\Big),\quad 0\leqslant t\leqslant T. \tag{63}$$

By Itô's formula applied to $Y_t = \exp(X_t)$ where $X_t = \int_0^T h(t)dB_t - \frac{1}{2}\int_0^T h^2(t)dt$. Note that $\frac{d}{dx}\exp(X_t) = \frac{d^2}{dx^2}\exp(X_t) = \exp(X_t) = Y_t$. Also, note that by Lemma 7.10,

$$\langle X\rangle_t = \left\langle \int_0^\bullet h(s)dB_s \right\rangle_t = \int_0^t h^2(s)\,ds.$$

We obtain:

$$Y_t = Y_0 + \int_0^t Y_s\left(h(s)\,dB_s - \frac{1}{2}h^2(s)\,ds\right) + \frac{1}{2}\int_0^t Y_s\,h^2(s)\,ds = 1 + \int_0^t Y_s\,h(s)\,dB_s.$$

This implies that $\mathbb{E}[F] = 1$, and that formula (61) holds for the dase when F is given by (62). By linearity, (61) holds for linear combinations of funcions of form (60).

Now let $F \in L_2(\mathcal{F}_T, \Omega)$ be arbitrary. We approximate F by linear combinations F_n of functions of form (60). For every n we have:

$$F_n(\omega) = \mathbb{E}[F_n] + \int_0^T f_n(s,\omega)\,dB_s(\omega)$$

where $f_n \in \Phi(0,T)$. By Itô's isometry and by the fact that $\mathbb{E}\Big[\int_0^T (f_n - f_m)\,dB_s\Big] = 0$, we obtain:

$$\begin{aligned}\mathbb{E}[(F_n - F_m)^2] &= \mathbb{E}\Big[\big(\mathbb{E}[F_n - F_m] + \int_0^T (f_n - f_m)\,dB_s\big)^2\Big] \\ &= \big(\mathbb{E}[F_n - F_m]\big)^2 + 2\,\mathbb{E}[F_n - F_m]\mathbb{E}\Big[\int_0^T (f_n - f_m)\,dB_s\Big] + \mathbb{E}\Big(\int_0^T (f_n - f_m)\,dB_s\Big)^2 \\ &= \big(\mathbb{E}[F_n - F_m]\big)^2 + \int_0^T \mathbb{E}[(f_n - f_m)^2]\,dt.\end{aligned}$$

Therefore

$$\int_0^T \mathbb{E}[(f_n - f_m)^2]\,dt = \mathbb{E}[(F_n - F_m)^2] - \big(\mathbb{E}[F_n - F_m]\big)^2 \to 0, \quad \text{as } n, m \to \infty.$$

Hence f_n is a Cauchy sequence in $L_2(\Omega \times [0,T])$, and hence converges to some $f \in L_2(\Omega \times [0,T])$. This implies that $f \in \Phi(0,T)$. Indeed, there is a subsequence of $\{f_n\}$ that converges to f for almost all $(\omega, t) \in \Omega \times [0,T]$. Therefore, f is $\mathcal{F}_t$-measurable for almost all $t \in [0,T]$. By modifying f on a t-set of measure zero we obtain that $f(t,\omega)$ is $\mathcal{F}_t$- adapted. Similar, the map $(t,\omega) \mapsto f(t,\omega)$ is $\mathcal{B} \times \mathcal{F}$-measurable. Clearly $\mathbb{E}\int_0^T f^2(t,\cdot)dt < \infty$. We obtain:

$$F = \lim_{n\to\infty} F_n = \lim_{n\to\infty}\Big(\mathbb{E}[F_n] + \int_0^T f_n\,dB_t\Big) = \mathbb{E}[F] + \int_0^T f\,dB_t$$

where the limit is taken in $L_2(\mathcal{F}_T, \Omega)$. Also, we used Itô's isometry to show the convergence of stochastic integrals.

The uniqueness follows from Itô's isometry. Suppose there exists two different functions $f_1, f_2 \in \Phi(0,T)$ such that

$$F(\omega) = \mathbb{E}[F] + \int_0^T f_1(t,\omega)\, dB_t = \mathbb{E}[F] + \int_0^T f_2(t,\omega)\, dB_t.$$

Then

$$\int_0^T (f_1 - f_2)\, dB_t = 0,$$

and therefore

$$\|f_1 - f_2\|^2_{L_2(\Omega\times[0,T])} = \mathbb{E}\int_0^T (f_1 - f_2)^2 dt = \mathbb{E}\Big(\int_0^T (f_1 - f_2)\, dB_t\Big)^2 = 0.$$

This implies that $f_1(t,\omega) = f_2(t,\omega)$ for almost all $(t,\omega) \in [0,T]\times\Omega$. □

7.4 The martingale representation theorem

Theorem 7.13. *Let $B_t = (B_t^1, \dots, B_t^n)$ be an n-dimensional Brownian motion, and let M_t be an $\mathcal{F}_t^n$-martingale, where the filtration $\mathcal{F}_t^n$ is defined by (52), such that $M_t \in L_2(\Omega)$ for all $t \geqslant 0$. Then there exists a unique stochastic process $g(s,\omega)$ such that $g(s,\cdot) \in \Phi(0,t)$ for all $t \geqslant 0$ with respect to $\mathcal{F}_s^n$, and*

$$M_t = \mathbb{E}[M_0] + \int_0^t g(s,\omega)\, dB_s \quad \text{a.s., for all } t \geqslant 0.$$

Proof. Again, we prove the theorem for the case when $n = 1$ for simplicity. By Itô's representation theorem (Theorem 7.13), which we apply to $T = t$ and $F = M_t$, we obtain that for every $t \geqslant 0$, there exists a unique $f^t(s,\omega) \in L_2(\Omega, \mathcal{F}_t)$ such that

$$M_t = \mathbb{E}[M_t] + \int_0^t f^t(s,\omega)\, dB_s = \mathbb{E}[M_0] + \int_0^t f^t(s,\omega)\, dB_s, \tag{64}$$

since for the martingale M_t it holds that $\mathbb{E}[M_t] = \mathbb{E}[M_0]$. Note that in the above formula, the integrand of the stochastic integral depends on t. Let us prove that $f^t(s,\omega)$ is actually does not depend on t. Assume $0 \leqslant t_1 < t_2$. Then,

$$M_{t_1} = \mathbb{E}[M_{t_2} \,|\, \mathcal{F}_{t_1}] = \mathbb{E}[M_0] + \mathbb{E}\Big[\int_0^{t_2} f^{t_2}(s,\omega)\, dB_s \,|\, \mathcal{F}_{t_1}\Big]$$
$$= \mathbb{E}[M_0] + \int_0^{t_1} f^{t_2}(s,\omega)\, dB_s.$$

On the other hand, by (64),

$$M_{t_1} = \mathbb{E}[M_0] + \int_0^{t_1} f^{t_1}(s,\omega)\, dB_s.$$

By the uniqueness result of Itô's representation theorem (Theorem 7.12), for almost all $s \in [0, t_1]$,

$$f^{t_1}(s, \omega) = f^{t_2}(s, \omega) \quad \text{a.s..}$$

Let us define $f(s, \omega)$ for almost all $s \in [0, \infty)$ by setting

$$f(s, \omega) = f^t(s, \omega) \quad \text{a.s. for almost all } s \in [0, t].$$

$f(s, \omega)$ is well-defined on $[0, t]$. Indeed, if $t_1 < t$, then for $s \in [0, t_1]$, a.s.,

$$f(s, \omega) = f^{t_1}(s, \omega) = f^t(s, \omega).$$

We obtain that for almost all $t \geqslant 0$,

$$M_t = \mathbb{E}[M_0] + \int_0^t f^t(s, \omega)\, dB_s = \mathbb{E}[M_0] + \int_0^t f(s, \omega)\, dB_s.$$

By t-continuity of the stochastic integral, the latter equality holds for all $t \geqslant 0$ a.s. □

8. Stochastic Differential Equations

Stochastic differential equations can be defined in several contexts that vary their generality. Informally, a stochastic differential equation (SDE) can be written in the form:

$$dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t.$$

Below we define weak and strong solutions of the above SDE and prove the existence and uniqueness theorem.

8.1 Formal definitions

Let $T > 0$, and let the functions $b : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : [0, T] \times \mathbb{R}^n \to \mathbb{R}^{n \times m}$ be measurable. Further let Z be a random variable, and let $\mathcal{F}_t^Z$ be the filtration generated by Z and B_s, $s \leqslant t$.

Definition 8.1. *Given a probability space $(\Omega, \mathcal{F}_t^Z, \mathcal{F}_T^Z, \mathbb{P})$, a process X_t is called a **strong solution** to the SDE*

$$dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t \tag{65}$$

on $[0, T]$ with the initial condition $X_0 = Z$ if X_t is $\mathcal{F}_t^Z$-adapted and

$$X_t = Z + \int_0^t b(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s. \tag{66}$$

Definition 8.2. *A pair of processes $(\tilde{X}_t, \tilde{B}_t)$ is called a **weak solution** to SDE (65) on $[0, T]$ with the initial condition $X_0 = Z$ if there exists a filtered probability space $(\Omega, \mathcal{H}_t, \mathcal{H}, P)$ such that $\tilde{B}_t$ is an $\mathcal{H}_t$-Brownian motion and $\tilde{X}_t$ is $\mathcal{H}_t$-adapted and*

$$\tilde{X}_t = Z + \int_0^t b(s, \tilde{X}_s)\, ds + \int_0^t \sigma(s, \tilde{X}_s)\, d\tilde{B}_s.$$

8.2 An existence and uniqueness theorem

Theorem 8.3 (Existence and uniqueness)**.** *Let, as before, $T > 0$ and let the functions $b : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$ be measurable and satisfying conditions (i) and (ii) below:*

(i) $|b(t,x)| + |\sigma(t,x)| \leqslant C\,(1+|x|)$, $x \in \mathbb{R}^n$, $t \in [0,T]$, *for some constant* $C > 0$;

(ii) $|b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leqslant D|x-y|$, $x, y \in \mathbb{R}^n$, $t \in [0,T]$, *for some constant* $D > 0$.

Let Z be a random variable which is independent of the σ-algebra $\mathcal{F}_\infty^m$ and such that

$$\mathbb{E}|Z|^2 < \infty.$$

Then SDE (65) has a unique strong t-continuous, $\mathcal{F}_t^Z$-adapted solution X_t on $[0,T]$ with the initial condition $X_0 = Z$.

Proof. Introduce the space $\mathcal{E}$ of $\mathcal{F}_t^Z$-adapted stochastic proceses with the finite squared norm

$$\|X\|^2 = \sup_{t\in[0,T]} \mathbb{E}|X_s|^2.$$

and define the map

$$\Gamma : \mathcal{E} \to \mathcal{E}, \quad \Gamma(X)(t) = Z + \int_0^t b(s, X_s)\,ds + \int_0^t \sigma(s, X_s)\,dB_s.$$

First we note that Γ maps $\mathcal{E}$ to $\mathcal{E}$. Take $X \in \mathcal{E}$. There exists a constant $C_1 > 0$ such that on $[0,T]$,

$$\begin{aligned}
\mathbb{E}|\Gamma(X)(t)|^2 &\leqslant 3\Big(\mathbb{E}|Z|^2 + \mathbb{E}\Big|\int_0^t b(s,X_s)\,ds\Big|^2 + \mathbb{E}\Big|\int_0^t \sigma(s,X_s)\,dB_s\Big|^2\Big) \\
&\leqslant 3\Big(\mathbb{E}|Z|^2 + t\int_0^t \mathbb{E}|b(s,X_s)|^2\,ds + \int_0^t \mathbb{E}|\sigma(s,X_s)|^2\,ds\Big) \\
&\leqslant C_1\Big(1 + \mathbb{E}|Z|^2 + \int_0^t \mathbb{E}|X_s|^2\,ds\Big) \\
&\leqslant C_1\Big(1 + \mathbb{E}|Z|^2 + T\sup_{s\in[0,T]} \mathbb{E}|X_s|^2\Big) < \infty.
\end{aligned}$$

Therefore,

$$\sup_{t\in[0,T]} \mathbb{E}|\Gamma(X)(t)|^2 < \infty$$

and hence, $\Gamma(X) \in \mathcal{E}$. Let $X^{(1)}, X^{(2)} \in E$. We have

$$\mathbb{E}|\Gamma(X^{(1)})(t) - \Gamma(X^{(2)})(t)|^2 \leqslant 2Dt\int_0^t \mathbb{E}|X_s^{(1)} - X_s^{(2)}|^2 ds + 2D\int_0^t \mathbb{E}|X_s^{(1)} - X_s^{(2)}|^2 ds.$$

Set $K := 2D(T+1)$. Then the estimate above implies that for every $t \in [0,T]$,

$$\mathbb{E}\big|\Gamma(X^{(1)})(t) - \Gamma(X^{(2)})(t)\big|^2 \leqslant K \int_0^t \mathbb{E}\big|X_s^{(1)} - X_s^{(2)}\big|^2 \, ds. \tag{67}$$

We claim that for every $n \geqslant 1$,

$$\mathbb{E}\big|\Gamma^n(X^{(1)})(t) - \Gamma^n(X^{(2)})(t)\big|^2 \leqslant K^n \int_0^t \frac{(t-s)^{n-1}}{(n-1)!} \, \mathbb{E}\big|X_s^{(1)} - X_s^{(2)}\big|^2 \, ds. \tag{68}$$

Indeed, for $n = 1$ this is exactly (67). Assume (68) holds for all $k \leqslant n$. Then, applying (67) with

$$X^{(1)} \text{ replaced by } \Gamma^n(X^{(1)}), \qquad X^{(2)} \text{ replaced by } \Gamma^n(X^{(2)}),$$

we obtain

$$\begin{aligned}
\mathbb{E}\big|\Gamma^{n+1}(X^{(1)})(t) - \Gamma^{n+1}(X^{(2)})(t)\big|^2 &\leqslant K \int_0^t \mathbb{E}\big|\Gamma^n(X^{(1)})(s) - \Gamma^n(X^{(2)})(s)\big|^2 \, ds \\
&\leqslant K^{n+1} \int_0^t \int_0^s \frac{(s-r)^{n-1}}{(n-1)!} \mathbb{E}\big|X_r^{(1)} - X_r^{(2)}\big|^2 \, dr \, ds.
\end{aligned}$$

By Fubini's theorem,

$$\begin{aligned}
\mathbb{E}\big|\Gamma^{n+1}(X^{(1)})(t) - \Gamma^{n+1}(X^{(2)})(t)\big|^2 &\leqslant K^{n+1} \int_0^t \left(\int_r^t \frac{(s-r)^{n-1}}{(n-1)!} \, ds \right) \mathbb{E}\big|X_r^{(1)} - X_r^{(2)}\big|^2 \, dr \\
&= K^{n+1} \int_0^t \frac{(t-r)^n}{n!} \, \mathbb{E}\big|X_r^{(1)} - X_r^{(2)}\big|^2 \, dr
\end{aligned}$$

which proves (68) for $n+1$. From (68), we have

$$\mathbb{E}\big|\Gamma^n(X^{(1)})(t) - \Gamma^n(X^{(2)})(t)\big|^2 \leqslant \frac{K^n t^n}{n!} \sup_{r \in [0,t]} \mathbb{E}\big|X_r^{(1)} - X_r^{(2)}\big|^2 \, dr$$

Taking the supremum over $t \in [0,T]$, we obtain

$$\|\Gamma^n(X^{(1)}) - \Gamma^n(X^{(2)})\|^2 \leqslant \frac{(KT)^n}{n!} \, \|X^{(1)} - X^{(2)}\|^2.$$

Since

$$\frac{(cT)^n}{n!} \longrightarrow 0 \qquad \text{as } n \to \infty,$$

there exists $N \in \mathbb{N}$ such that

$$\frac{(KT)^N}{N!} < 1.$$

Therefore Γ^N is a contraction on E. By Banach's fixed point theorem, Γ^N has a unique fixed point $X \in E$. Then

$$\Gamma^N(\Gamma X) = \Gamma(\Gamma^N X) = \Gamma X,$$

so ΓX is also a fixed point of Γ^N. By uniqueness of the fixed point of Γ^N, it follows that

$$\Gamma X = X.$$

Thus X is a fixed point of Γ in $\mathcal{E}$. If Y is another fixed point of Γ, then it is also a fixed point of Γ^N, hence $Y = X$. Therefore the fixed point of Γ is unique in $\mathcal{E}$.

Let us prove that there is t-continuous version of the solution X_t to (66). The uniqueness in $\mathcal{E}$ means

$$\mathbb{E}|X_t - \Gamma(X)(t)|^2 = 0 \qquad \forall t \in [0,T]. \tag{69}$$

Define $\hat{X}_t := \Gamma(X)(t)$ for all $t \in [0,T]$ and note that $\hat{X}_t$ is t-continuous. (69) implies that

$$\mathbb{P}(X_t = \hat{X}_t) = 1.$$

Therefore, $\hat{X}_t$ is a continuous modification of X_t. We also note that by the Lipschitz condition (ii), $\Gamma(X)(t) = \Gamma(\hat{X})(t)$ for all $t \in [0,T]$. Hence,

$$\hat{X}_t = \Gamma(\hat{X})(t) \qquad \text{a.s.},$$

so $\hat{X}_t$ is a t-continuous solution to (66). □

Definition 8.4. *We say that a solution (X_t, B_t) to SDE (65) is **weakly unique** if for any other weak solution $(\tilde{X}_t, \tilde{B}_t)$ to (65) it holds that (X_t, B_t) and $(\tilde{X}_t, \tilde{B}_t)$ are identical in law.*

Theorem 8.5 (Weak uniqueness)**.** *Let σ and b satisfy assumptions of Theorem 8.3, and, moreover, are continuous with respect to the first argument. Then, a weak or strong solution to SDE (65) is weakly unique.*

Remark. Clearly, every strong solution to (65) is also a weak solution. When we say a strong solution X_t is "weakly unique" we mean that the solution (X_t, B_t) is weakly unique.

8.3 The Tanaka equation

There are stochastic differential equations which do not have *strong solutions* but have a *unique weak solution*. The example below discusses such an SDE. Consider the one-dimensional stochastic differential equation:

$$dX_t = \operatorname{sign}(X_t)\, dB_t, \quad X_0 = 0 \tag{70}$$

where

$$\operatorname{sign}(x) = \begin{cases} +1 & x \geqslant 0, \\ -1, & x < 0. \end{cases}$$

Note that here $\sigma(x,t) = \operatorname{sign}(x)$ does not satisfy the Lipschitz condition (Condition (ii) of Theorem 8.3), so Theorem 8.3 cannot be applied here.

Proposition 8.6. *SDE (70) has no strong solution.*

Lemma 8.7 (Tanaka's formula)**.** *Let B_t be a 1-dimensional Brownian motion. Then,*

$$|B_t| = |B_0| + \int_0^t \operatorname{sign}(B_s)\, dB_s + L_t(\omega), \tag{71}$$

where

$$L_t = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \lambda\{s \in [0,t] :\ B_s \in (-\varepsilon, \varepsilon)\}$$

and λ is the Lebesgue measure.

Remark. L_t is called the *local time* for a Brownian motion at 0. Formula (71) is known as *Tanaka's formula* for a Brownian motion.

Proof of Lemma 8.7. Without proof. □

Lemma 8.8 (How to recognize a Brownian motion)**.** *Let B_t be an m-dimensional Bronian motion, and $v : [0,\infty) \times \Omega \to \mathbb{R}^{n\times m}$ satisfy the assumptions (see Section 6.3) under which the stochastic integral is well defined. Further let*

$$Y_t = \int_0^t v(s,\omega)\, dB_s.$$

Then Y_t be a Brownian motion if and only if

$$v(t,\omega)v^T(t,\omega) = I_n$$

for almost all (t,ω). I_n in the above formula is the $n \times n$ identity matrix.

Proof. Without proof. □

Proof of Proposition 8.6. Let $\hat{B}_t$ be a Brownian motion and let $\hat{\mathcal{F}}_t$ be its natural filtration (i.e. $\sigma(\hat{B}_s, s \leqslant t)$). Define the process

$$Y_t = \int_0^t \operatorname{sign}(\hat{B}_s)\, d\hat{B}_s.$$

By Tanaka's formula (Lemma 8.7),

$$Y_t = |\hat{B}_t| - |\hat{B}_0| - \hat{L}_t,$$

where $\hat{L}_t$ is the local time for $\hat{B}_t$ at 0. Let $\mathcal{F}_t = \sigma(|\hat{B}_s|, s \leqslant t)$. Clearly $\mathcal{F}_t \subset \hat{\mathcal{F}}_t$, where the inclusion is strict. Note that

$$L_t = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_{[0,t]} \mathbb{1}_{[0,\varepsilon)}(|B_s|)\lambda(ds).$$

Hence, the σ-algebra $\mathcal{N}_t = \sigma(Y_s, s \leqslant t) = \sigma(|\hat{B}_s|, s \leqslant t) \subset \hat{\mathcal{F}}_t$.

Now suppose X_t is a strong solution to (70). Then, since $\big(\operatorname{sign}(X_t)\big)^2 = 1$, by Lemma 8.8, X_t is a Brownian motion starting at 0. Multiplying both parts of (70) by $\operatorname{sign}(X_t)$ we obtain:

$$dB_t = \operatorname{sign}(X_t)\, dX_t.$$

Let us apply the above argument to $\hat{B}_t = X_t$ (since we proved that X_t is a Brownian motion) and $Y_t = \int_0^t \operatorname{sign}(X_r)\, dX_r = B_t$. We obtain that the natural filtration $\mathcal{F}_t$ of B_t is strictly contained in $\mathcal{M}_t = \sigma(X_s, s \leqslant t)$. This contradicts to the fact that X_t is a strong solution since a strong solution is $\mathcal{F}_t$-adapted. Thus, a stong solution to (70) does not exist. □

Let us find a weak solution to (70).

Lemma 8.9. *SDE (70) has a weak solution.*

Proof. Choose X_t to be any Brownian motion $\hat{B}_t$. Let us define

$$\tilde{B}_t = \int_0^t \operatorname{sign}(\hat{B}_s)\, d\hat{B}_s = \int_0^t \operatorname{sign}(X_s)\, dX_s.$$

Again, by Lemma 8.8, $\tilde{B}_t$ is a Brownian motion. Finally the SDE

$$d\tilde{B}_t = \operatorname{sign}(X_t) dX_t$$

is equivalent to

$$dX_t = \operatorname{sign}(X_t)\, d\tilde{B}_t$$

which can be obtained by means of multiplication of the both parts by $\operatorname{sign}(X_t)$. Therefore, $(X_t, \tilde{B}_t)$ is a weak solution.

Let us prove the weak uniqueness. Let $(\bar{X}_t, \bar{B}_t)$ be another weak solution. Another application of Lemma 8.8 shows that $\bar{X}_t$ is a Brownian motion. Hence, all weak solutions are identical in law. □

9. Stochastic partial differential equations

Semigroup and stochastic PDE interpretation. Let H be a separable Hilbert space and $A : H \supset D(A) \to H$ be a (possibly unbounded) linear operator. The operator A is called the generator of a strongly continuous semigroup $(S(t))_{t \geqslant 0}$ on H if

$$S(0) = I, \quad S(t+s) = S(t)S(s), \quad \forall t, s \geqslant 0,$$

and

$$\lim_{t \downarrow 0} \|S(t)x - x\|_H = 0, \quad \forall x \in H.$$

The generator is defined by

$$Ax := \lim_{t \downarrow 0} \frac{S(t)x - x}{t}, \quad x \in D(A),$$

where

$$D(A) = \left\{ x \in H : \lim_{t \downarrow 0} \frac{S(t)x - x}{t} \text{ exists in } H \right\}.$$

In applications, A is typically a partial differential operator:

$$Af(x) = \sum_{i,j=1}^{d} a_{ij}(x)\, \partial_{ij}^2 f(x) + \sum_{i=1}^{d} b_i(x)\, \partial_i f(x).$$

Above, $a_{ij} \in C^1,\ b_i \in L^\infty$. The matrix $a(x) = (a_{ij}(x))$ is symmetric,

$$a_{ij}(x) = a_{ji}(x),$$

and uniformly elliptic, i.e. there exists $\lambda > 0$ such that

$$\sum_{i,j=1}^{d} a_{ij}(x)\xi_i\xi_j \geqslant \lambda|\xi|^2 \qquad \text{for all } x \in D,\ \xi \in \mathbb{R}^d.$$

Two important cases of H are: (i) the whole space $H = L^2(\mathbb{R}^d)$, where A is defined on a Sobolev space such as $H^2(\mathbb{R}^d)$, (ii) a bounded domain $D \subset \mathbb{R}^d$, where $H = L^2(D)$ and $D(A)$ incorporates boundary conditions, e.g. $H^2(D) \cap H_0^1(D)$.

In this setting, stochastic evolution equations of the form

$$dX_t = (AX_t + F(t, X_t))\, dt + G(t, X_t)\, dB_t.$$

Assume that for a fixed horizon $T > 0$, the semigroup is bounded:

$$\|S(t)\|_{\mathcal{L}(H)} \leqslant M_T, \qquad t \in [0, T].$$

Let $(B_t)_{t \geqslant 0}$ be an m-dimensional Brownian motion defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geqslant 0}, \mathbb{P})$. Let

$$F : [0, T] \times H \to H, \qquad G : [0, T] \times H \to \mathcal{L}(\mathbb{R}^m, H)$$

be measurable functions. Consider the stochastic evolution equation

$$\begin{cases} dX_t = (AX_t + F(t, X_t))\, dt + G(t, X_t)\, dB_t, \\ X(0) = Z, \end{cases} \tag{72}$$

where $Z \in L^2(\Omega; H)$ is $\mathcal{F}_0$-measurable.

Assumptions

(H1) Lipschitz condition: there exists $L > 0$ such that

$$\|F(t,x) - F(t,y)\| + \|G(t,x) - G(t,y)\|_{\mathcal{L}(\mathbb{R}^m,H)} \leqslant L\|x-y\|,$$

for all $t \in [0,T]$, $x, y \in H$.

(H2) Linear growth: there exists $C > 0$ such that

$$\|F(t,x)\|^2 + \|G(t,x)\|^2_{\mathcal{L}(\mathbb{R}^m,H)} \leqslant C(1+\|x\|^2).$$

Definition of a mild solution

An adapted process $X : [0,T] \to H$ with continuous trajectories is called a *mild solution* of (72) if for all $t \in [0,T]$,

$$X_t = S(t)Z + \int_0^t S(t-s)F(s,X_s)\,ds + \int_0^t S(t-s)G(s,X_s)\,dB_s, \quad \text{a.s.}$$

Note that $\int_0^t S(t-s)G(t,X_s)\,dB_s$ is H-valued. Let $(e_k)_{k\geqslant 1}$ be an orthonormal basis in H. The H-valued stochastic integral can be interpreted componentwise by considering the real-valued processes

$$\left\langle \int_0^t S(t-s)G(t,X_s)\,dB_s, e_k \right\rangle = \int_0^t \langle S(t-s)G(t,X_s), e_k\rangle\,dB_s,$$

which are well-defined one-dimensional Itô integrals for each $k \geqslant 1$. For each $N \in \mathbb{N}$, consider the finite-dimensional projection

$$\sum_{k=1}^N \left\langle \int_0^t S(t-s)G(t,X_s)\,dB_s, e_k \right\rangle e_k.$$

By Itô's isometry in $\mathbb{R}^N$, we have

$$\mathbb{E}\sum_{k=1}^N \left| \int_0^t \langle S(t-s)G(t,X_s), e_k\rangle\,dB_s \right|^2 = \mathbb{E}\int_0^t \sum_{k=1}^N |\langle S(t-s)G(t,X_s), e_k\rangle|^2 ds.$$

Passing to the limit as $N \to \infty$, we obtain

$$\mathbb{E}\left\| \int_0^t S(t-s)G(t,X_s)\,dB_s \right\|_H^2 = \mathbb{E}\int_0^t \|S(t-s)G(t,X_s)\|_H^2 ds.$$

Define

$$\mathcal{E} := \Big\{ X \text{ adapted} : \|X\|_{\mathcal{E}}^2 := \sup_{t\in[0,T]} \mathbb{E}\|X_t\|^2 < \infty \Big\}.$$

Define $\Gamma : \mathcal{E} \to \mathcal{E}$ by

$$(\Gamma X)(t) := S(t)Z + \int_0^t S(t-s)F(s,X_s)\,ds + \int_0^t S(t-s)G(s,X_s)\,dB_s.$$

Since the semigroup $S(t)$ is bounded, there exists $M_T > 0$ such that $\|S(t)\| \leqslant M_T$ for all $t \in [0,T]$. Let $X \in \mathcal{E}$. Then

$$\mathbb{E}\|(\Gamma X)(t)\|^2 \leqslant C\Big(\mathbb{E}\|S(t)Z\|^2 + \mathbb{E}\Big\|\int_0^t S(t-s)F(s,X_s)\,ds\Big\|^2 + \mathbb{E}\Big\|\int_0^t S(t-s)G(s,X_s)\,dB_s\Big\|^2\Big).$$

Using $\|S(t)\| \leqslant M_T$, we estimate in one step

$$\mathbb{E}\|(\Gamma X)(t)\|^2 \leqslant C\Big(M_T^2\mathbb{E}\|Z\|^2 + M_T^2\int_0^t \mathbb{E}\|F(s,X_s)\|^2 ds + M_T^2\int_0^t \mathbb{E}\|G(s,X_s)\|^2 ds\Big).$$

By the linear growth assumptions on F, G,

$$\mathbb{E}\|(\Gamma X)(t)\|^2 \leqslant C\Big(1 + \mathbb{E}\|Z\|^2 + \int_0^t \mathbb{E}\|X_s\|^2 ds\Big).$$

Taking the supremum over $t \in [0,T]$ and using that $X \in \mathcal{E}$, we obtain

$$\sup_{t\in[0,T]} \mathbb{E}\|(\Gamma X)(t)\|^2 < \infty.$$

Hence $\Gamma X \in \mathcal{E}$. Let $X^{(1)}, X^{(2)} \in \mathcal{E}$. Since $\|S(t)\| \leqslant M_T$,

$$\begin{aligned}\mathbb{E}\big|\Gamma(X^{(1)})(t) - \Gamma(X^{(2)})(t)\big|^2 &\leqslant 2M_T^2 t\int_0^t \mathbb{E}\big|F(s,X_s^{(1)}) - F(s,X_s^{(2)})\big|^2 ds \\ &\quad + 2M_T^2\int_0^t \mathbb{E}\big\|G(s,X_s^{(1)}) - G(s,X_s^{(2)})\big\|^2 ds \leqslant C\int_0^t \mathbb{E}\big\|X_s^{(1)} - X_s^{(2)}\big\|^2 ds,\end{aligned}$$

Thus the estimate reduces to the same form as in the SDE case, i.e. by proving that

$$\|\Gamma^n X - \Gamma^n Y\|_{\mathcal{E}}^2 \leqslant \frac{(CT)^n}{n!}\|X - Y\|_{\mathcal{E}}^2.$$

so that Γ^N is a contraction on $\mathcal{E}$ for some $N \in \mathbb{N}$. From here, we proceed exactly as in the case of an SDE.