Lecture Notes
Introduction to Stochastic Analysis

Olivier Menoukeu Pamen

Contents

Chapter 1

Preliminaries

1.1 Mathematical Preliminaries

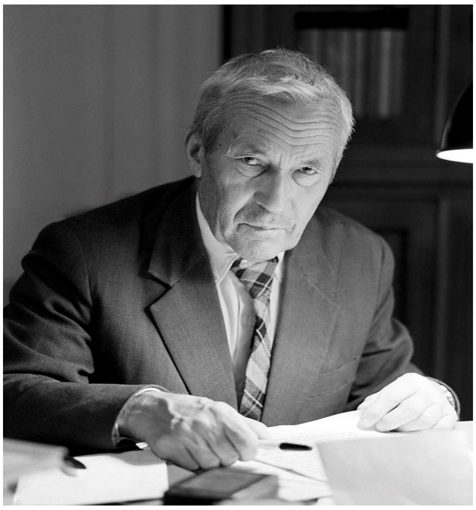


Figure 1.1: Andrey N. Kolmogorov (25-04-1903 to 20-10-1987): Soviet mathematician

1.1.1 Measure theory

Set Theory

Let Ω be a set (or sample space) e.g a set of all possible outcomes of an experiment.

Let $\omega \in \Omega$ be an element of Ω (a sample point). Denote by $\mathscr{F}$ the collection of all subsets of Ω including Ω and $\emptyset$.

Definition 1.1.1. *A collection $\Sigma \subset \mathscr{F}$ of subsets of Ω is called an algebra iff*

- $\Omega, \emptyset \in \Sigma$.
- *If $F \in \Sigma$ then $F^{\complement} = \Omega \setminus F \in \Sigma$, where $F^{\complement}$ is the complement of F in Ω.*
- *For any finite collection of elements $F_1, F_2, \cdots F_n$ in Σ we have $\cap_{i=1}^{n} F_i \in \Sigma$ (Closed under finite intersection).*

Example 1.1.2. $\Omega = \{1,2,3\}$
$\Sigma = \{\emptyset, \Omega, \{1\}, \{2,3\}\}$. *Then Σ is an algebra.*

Definition 1.1.3 (σ-algebra)**.** *A collection $\Sigma \subset \mathscr{F}$ of subsets of Ω is called a σ-algebra iff Σ is an algebra and for any countable collection of elements in Σ their union is also in Σ. That is $F_1, F_2 \cdots\cdots \in \Sigma \Rightarrow F = \cup_{n=1}^{\infty} F_n \in \Sigma$.*

Theorem 1.1.4. *If Σ_i, $i \in I$ is an arbitrary collection of σ-algebras then their intersection $\cap_{i \in I} \Sigma_i$ is a σ-algebra.*
■

Remark 1.1.5. *The union of two σ-algebra is not always a σ-algebra.*

Example 1.1.6.

$$\begin{aligned} \Omega &= \{1,2,3\} \\ \Sigma_1 &= \{\emptyset; \Omega; \{1\}; \{2,3\}\} \\ \Sigma_2 &= \{\emptyset; \Omega; \{2\}; \{1,3\}\} \end{aligned}$$

$$\Sigma_1 \cup \Sigma_2 = \{\emptyset; \Omega; \{1\}; \{2\}; \{1,3\}; \{2,3\}\}$$

$$\begin{aligned} &\{1\} \in \Sigma_1 \cup \Sigma_2 \qquad\qquad \textit{but} \quad \{1,2\} \notin \Sigma_1 \cup \Sigma_2 \\ &\{2\} \in \Sigma_1 \cup \Sigma_2 \end{aligned}$$

Example of algebra which is not a σ-algebra
Let Ω be an infinite set
Σ = {collection of subsets of Ω which are finite or have finite complement}.

Definition 1.1.7. *Let $\mathscr{C}$ be a class of subsets of Ω, then $\sigma(\mathscr{C})$, the σ-algebra generated by $\mathscr{C}$ is the smallest σ-algebra Σ on Ω such that $\mathscr{C} \subset \Sigma$. It is the intersection of all σ-algebras of Ω containing $\mathscr{C}$. In particular, we have the following theorem.*

Theorem 1.1.8. *If $\mathscr{C}$ is any collection of subsets of Ω, there exists a smallest σ-algebra $\sigma(\mathscr{C})$ containing $\mathscr{C}$ i.e $\mathscr{C} \subset \sigma(\mathscr{C})$, namely*

$$\sigma(\mathscr{C}) = \cap\{\mathscr{H}, \quad \mathscr{H} \;\; \sigma\text{-algebra of } \Omega \text{ and } \mathscr{C} \subset \mathscr{H}\}.$$

■

1.1.2 Measure Spaces

Let Ω be a set and $\mathscr{F}$ be a σ-algebra on Ω.

Definition 1.1.9. *The pair $(\Omega, \mathscr{F})$ is called a measurable space. An element of $\mathscr{F}$ is called $\mathscr{F}$-measurable subset of Ω.*

Definition 1.1.10. *Let Ω be a set, let $\sum_0$ be an algebra on Ω and let μ_0 be a non-negative set function $\mu_0 : \sum_0 \mapsto [0, \infty]$. Then μ_0 is called additive if*

1. $\mu_0(\emptyset) = 0$ *and*
2. *For $F, G \in \sum_0$, $F \cap G = \emptyset \Rightarrow \mu_0(F \cup G) = \mu_0(F) + \mu_0(G)$.*

The function μ_0 is called σ-additive if

1. $\mu_0(\emptyset) = 0$ *and*
2. *For a sequence $(F_n)_{n \in N}$ of disjoint sets in $\sum_0$, with $F = \bigcup F_n \in \sum_0$ (this is true if $\sum_0$ is a σ-algebra) Then*

$$\mu_0(F) = \sum_n \mu_0(F_n).$$

Definition 1.1.11. *The triple $(\Omega, \mathscr{F}, \mu)$ consisting of a set Ω , a σ- algebra on Ω and a measure μ (countably additive map from $\mathscr{F} \mapsto [0; \infty]$) is called a measure space. The measure space $(\Omega, \mathscr{F}, \mu)$ is called*

- *finite if $\mu(\Omega) < \infty$*
- *σ-finite if there is a sequence $(F_n)_{n\in\mathbb{N}}$ of elements of $\mathscr{F}$ such that $\mu(F_n) < \infty \; \forall \; n \in \mathbb{N}$ and $\bigcup F_n = \Omega$*

Definition 1.1.12. *Let Ω be a set, $\mathscr{F}$ be a σ-algebra on Ω and $\mathbb{P}$ be a measure on $\mathscr{F}$ such that $\mathbb{P}(\Omega) = 1$. Then $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space and $\mathbb{P}$ is a probability measure.*

Definition 1.1.13.

- *An element of F of $\mathscr{F}$ is called μ-null if $\mu(F) = 0$.*
- *A statement S about $w \in \Omega$ is said to hold almost everywhere (a.e) if*

$$F := \{w, S(w) \text{ is false } \} \in \mathscr{F} \text{ and } \mu(F) = 0.$$

Remark 1.1.14. *Let $(\Omega, \mathscr{F}, \mu)$ be a measure space, then*

(i) *For any $A, B \in \mathscr{F}$, then $\mu(A \cup B) \leq \mu(A) + \mu(B)$.*

(ii) *For every $A_1, A_2, \ldots, A_n \in \mathscr{F}$, $\mu(\cup_{i=1}^n A_i) \leq \sum_i \mu(A_i)$.*

(iii) *For any $A, B \in \mathscr{F}$, $A \subset B \Rightarrow \mu(A) \leq \mu(B)$.*

(iv) *For any $A, B \in \mathscr{F}$, if $B \subset A$ and $\mu(B) < \infty$ then $\mu(A \backslash B) = \mu(A) - \mu(B)$.*

(v) *For any $A, B \in \mathscr{F}$, if $\mu(\Omega) < \infty$ then $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$.*

1.1.3 Events and Random Variable

Definition 1.1.15. *A statement S about outcomes is said to be true almost surely (a.s) or with probability 1 if*

$$F := \{\omega, S(\omega) \text{ is true } \} \in \mathscr{F} \text{ and } \mathbb{P}(F) = 1.$$

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. Any event $A \in \mathscr{F}$ for which $\mathbb{P}(A) = 0$ is called $\mathbb{P}$-**null**. It seems reasonable to assume that any subset B of a null set A is itself null. However this may have no meaning since B may not be an event. i.e $\mathbb{P}(B)$ may not be defined.

Definition 1.1.16. *A probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is called complete if $\mathscr{F}$ contains all subsets G of Ω with $\mathbb{P}$-outer measure zero i.e., with*

$$\mathbb{P}^*(G) = \inf\{\mathbb{P}(F), F \in \mathscr{F}, \; G \subset F\} = 0$$

that is all subset of null sets are events.

Definition 1.1.17. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $(E, \mathscr{E})$ a measurable space. A random variable is a measurable function $X : \Omega \mapsto E$; that is $X^{-1} : \mathscr{E} \mapsto \mathscr{F}$*

$$X^{-1}(B) = \{\omega \in \Omega; X(\omega) \in B\} \in \mathscr{F}$$

The random variable is real random variable if $E = \mathbb{R}$ and $\mathscr{E} = \mathscr{B}(\mathbb{R})$ and in this case

$$\{\omega \in \Omega; X(\omega) \leq x\} \in \mathscr{F}, \quad \forall \, x \in \mathbb{R}. \tag{1.1}$$

Definition 1.1.18. *The σ-algebra generated by a random variable X, denoted by $\sigma(X)$, is the smallest σ-algebra which X is measurable with respect to.*

Example 1.1.19. *Let*

$$\Omega = \{1,2,3,4,5,6\}, \quad \mathscr{F} = \{\emptyset, \Omega, \{1,3,5\}, \{2,4,6\}\}$$

$$X(\omega) = \begin{cases} 1 & \text{if } \omega \text{ even} \\ 0 & \text{else} \end{cases}$$

and

$$Y(\omega) = \begin{cases} 1 & \text{if } \omega < 4 \\ 0 & \text{else.} \end{cases}$$

1. *Are X, Y random Variables on* $(\Omega, \mathscr{F})$*?*

2. *What is the smallest* σ*-algebra that makes Y a random variable?*

Solution 1.1.20.

1. $X(\omega)$ *is* $\mathscr{F}$*-measurable , in fact*

$$\{\omega \in \Omega;\ X(\omega) = x\} = \begin{cases} \{2,4,6\} \in \mathscr{F} & \text{if } x = 1 \\ \{1,3,5\} \in \mathscr{F} & \text{if } x = 0 \\ \emptyset \in \mathscr{F} & \text{if } x \in \Re \backslash \{0,1\} \end{cases}$$

Therefore $X(\omega)$ *is a random variable on* $(\Omega, \mathscr{F})$ *thus* $\mathscr{F}$*-measurable.*

$$\{\omega \in \Omega;\ Y(\omega) = y\} = \begin{cases} \{1,2,3\} \notin \mathscr{F} & \text{if } y = 1 \\ \{4,5,6\} \notin \mathscr{F} & \text{if } y = 0 \\ \emptyset \in \mathscr{F} & \text{if } y \in \Re \backslash \{0,1\} \end{cases}$$

Therefore $Y(\omega)$ *is not* $\mathscr{F}$*-measurable.*

2.
$$Y(\omega) = \begin{cases} 1 & \text{if } \omega \in \{1,2,3\} \\ 0 & \text{if } \omega \in \{4,5,6\} \end{cases}$$

$$\mathscr{P} = \{\{1,2,3\}, \{4,5,6\}\}$$

$$\sigma(\mathscr{P}) = \{\emptyset, \Omega, \{1,2,3\}, \{4,5,6\}\}$$

Example 1.1.21. *Let* $(\Omega, \mathscr{F})$ *be a measurable space. Define the indicator function on an event* $A \subseteq \Omega$ *by*

$$I_A : \Omega \to \{0,1\}$$

$$\omega \mapsto I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \ \text{ i.e } \ \omega \in A^c. \end{cases}$$

1. *Show that if* $A \in \mathscr{F}$ *the* I_A *is* $\mathscr{F}$*-measurable.*

2. $I_{A \cap B} = I_A I_B$

3. $I_{A \cup B} + I_{A \cap B} = I_A + I_B$

4. *Let* $A_1, \ldots, A_n$ *be a partition of* Ω*.* $\sum_{i=1}^{n} I_{A_i}(\omega) = 1$ *for* $\omega \in \Omega$*.*

Solution 1.1.22.

1. $I_A^{-1}(\{1\}) = \{\omega \in \Omega;\ I_A(\omega) = 1\} = A \in \mathscr{F}$
   $I_A^{-1}(\{0\}) = \{\omega \in \Omega;\ I_A(\omega) = 0\} = A^{\complement} \in \mathscr{F}$
   Let $x \in \mathbb{R} \setminus \{0,1\}$*. Then* $I_A^{-1}(\{x\}) = \{\omega \in \Omega;\ I_A(\omega) = x\} = \emptyset \in \mathscr{F}$
   Hence I_A *is* $\mathscr{F}$*-measurable.*

2. *If*
$$I_{A \cap B}(\omega) = \begin{cases} 1 & \text{if } \omega \in A \cap B \\ 0 & \text{if } \omega \notin A \cap B. \end{cases}$$

.

$$\begin{aligned} I_A(\omega) I_B(\omega) = 1 \quad &\Leftrightarrow \quad I_A(\omega) = 1 \quad \text{and } I_B(\omega) = 1. \\ &\Leftrightarrow \quad \omega \in A \quad \text{and } \omega \in B \\ &\Leftrightarrow \quad \omega \in A \cap B \\ &\Leftrightarrow \quad I_{A \cap B}(\omega) = 1. \end{aligned}$$

We also have the following lemma that characterises a real $\sigma(X)$-measurable r.v.

Lemma 1.1.23. *Let $X : (\Omega, \mathscr{F}, \mathbb{P}) \mapsto (E, \mathscr{E})$ be a r.v. Then every real $\sigma(X)$-measurable r.v. is of the form $f(X)$, where $f : (E, \mathscr{E}) \mapsto (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ is a measurable map.*

The r.v.'s $X_1; \ldots; X_m$, taking values respectively in $(E_1, \mathscr{E}_1), \ldots, (E_m, \mathscr{E}_m)$, are said to be independent if, for every $A_1 \in \mathscr{E}_1, \ldots, A_m \in \mathscr{E}_m$, we have

$$\mathbb{P}(X_1 \in A_1, \ldots, X_m \in A_m) = \mathbb{P}(X_1 \in A_1) \ldots \mathbb{P}(X_m \in A_m).$$

The events $A_1, \ldots, A_m \in \mathscr{F}$ are said to be independent if and only if

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_p}) = \mathbb{P}(A_{i_1}) \ldots \mathbb{P}(A_{i_p})$$

for every choice of $p \in \{1, \ldots, m\}$ and every choice of $1 \leq i_1 < i_2 < \ldots < i_p \leq m$.

If $\mathscr{F}_1; \ldots; \mathscr{F}_m$ are sub-σ-algebras of $\mathscr{F}$, then there are said to be independent if for every $A_1 \in \mathscr{F}_1, \ldots, A_m \in \mathscr{F}_m$, it holds:

$$\mathbb{P}(A_1 \cap \ldots \cap A_p) = \mathbb{P}(A_1) \ldots \mathbb{P}(A_p)$$

Therefore the r.v.'s $X_1, \ldots, X_m$ are independent if and only if the σ-algebras $\sigma(X_1), \ldots, \sigma(X_m)$ they generate are independent.

Let X be a r.v. with values in the measurable space $(E, \mathscr{E})$. Let us now define the set function μ_X on $\mathscr{E}$ by

$$\mu_X(A) := \mathbb{P}(X^{-1}(A)). \tag{1.2}$$

Then μ_X is a probability measure on $\mathscr{E}$ and called the law of X also known as the image or pullback measure of $\mathbb{P}$ via X. We also have the following proposition

Proposition 1.1.24. *Let $X : (\Omega, \mathscr{F}, \mathbb{P}) \mapsto (E, \mathscr{E})$ be a r.v., μ_X be the law of X. Then a measurable function $f : (E, \mathscr{E}) \mapsto (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ is μ_X-integrable if and only if f is $\mathbb{P}$-integrable, and then we have*

$$\int_\Omega f(X(\omega)) \mathrm{d}\mathbb{P}(\omega) = \int_E f(x) \mathrm{d}\mu_X(x).$$

The above proposition provides a formula for the computation of integrals with respect to an image law.

If X be a real random variable on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$. If $\int_\Omega |X(\omega)| \mathrm{d}\mathbb{P}(\omega) < \infty$. Then the number defined by

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

is called the expectation of X(with respect to $\mathbb{P}$). More generally if $g : \mathbb{R}^n \mapsto \mathbb{R}$ is a Borel measurable function and $\int_\Omega |g(X(\omega))| \mathrm{d}\mathbb{P}(\omega) < \infty$ then

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

Definition 1.1.25 (Marginal law)**.** *Let μ be a probability measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ and let $\mu_i : \mathbb{R}^n \mapsto \mathbb{R}$ be its projection on the i-th coordinate. The i-th marginal law of μ denoted by μ_i is the image of μ by π_i. It is given by:*

1. *(Discrete r.v.):*
$$\mu_i(X_i = x) = \sum_{(x_1, \ldots, x_n) \in S_X : x_i = x} \mu(X_1 = 1, \ldots, X_i = x, \ldots, X_n = x_n), \tag{1.3}$$
where the sum is over the set $\{(x_1, \ldots, x_n) \in S_X : x_i = x\}$ and S_X is the support of X.

2. *(Continuous r.v.)*
$$\mu_i(A) = \int_{\mathbb{R}^n} 1_A(x_i) \mu(\mathrm{d}x_1, \cdots, \mathrm{d}x_n),\ A \in \mathscr{B}(\mathbb{R}) \tag{1.4}$$

Remark 1.1.26. *Let $X = (X_1, \ldots, X_n)$ be an n-dimensional r.v. with law μ. Then its i-th marginal μ_i coincides with the law of X_i.*

Example 1.1.27. *Let X be a 2×1 random vector with support*

$$S_X = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right\}$$

and joint probability mass function

$$\mu(x_1,x_2) = \begin{cases} 1/2 \text{ if } x_1 = 1 \text{ and } x_2 = 2 \\ 1/3 \text{ if } x_1 = 1 \text{ and } x_2 = 0 \\ 1/6 \text{ if } x_1 = 2 \text{ and } x_2 = 2 \\ 0 \text{ otherwise} \end{cases}$$

Find the marginal law of X_1 at x_1.

The marginal probability mass function of X_1 evaluated at the point $x_1 = 2$ is given by

$$\mu_1(X_i = 2) = \sum_{(x_1,x_2)\in S_X : x_1=2} \mu(X_1 = 1, X_2 = x) = \mu(2,1) = 1/6$$

The marginal probability mass function of X_1 evaluated at the point $x_1 = 1$ is given by

$$\mu_1(X_i = 1) = \sum_{(x_1,x_2)\in S_X : x_1=1} \mu(X_1 = 1, X_2 = x) = \mu(1,2) + \mu(1,0) = 1/2 + 1/3 = 5/6$$

For all the other points, it is equal to zero. Thus it holds

$$\mu_1(x_1) = \begin{cases} 5/6 \text{ if } x_1 = 1 \\ 1/6 \text{ if } x_1 = 2 \\ 0 \text{ otherwise} \end{cases}$$

Definition 1.1.28. *Let μ be a probability measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$. We say that μ has a density (with respect to Lebesgue measure) if there exists a Borel measurable non negative function f such that for every*

$$\mu(A) = \int_A f(x)\mathrm{d}x \text{ for every } A \in \mathscr{B}(\mathbb{R}^n).$$

Remark 1.1.29.

- *Suppose μ has a density f. Then its i-th marginal μ_i also has a density f_i given by*

$$f_i(x) = \int f(y_1,\dots,y_{i-1},x,y_{i+1},\dots,y_n)\mathrm{d}y_1\cdots\mathrm{d}y_{i-1}\mathrm{d}y_{i+1}\cdots\mathrm{d}y_n.$$

- *Let X be an $\mathbb{R}^n$-valued r.v. with density f. Let $b \in \mathbb{R}^n$ and A an $m\times n$ invertible matrix; One can show that the r.v. Y defined by $Y = AX + b$ also has a density function g given by*

$$g(y) = |det\, A|^{-1} f(A^{-1}(y-b)). \tag{1.5}$$

Let $X = (X_1,\dots,X_n)$ be an n-dimensional r.v. Then its covariance matrix is the matrix $C = (c_{ij})_{i,j}$ defined by:

$$c_{i,j} = Cov(X_i,X_j) = \mathbb{E}[X_iX_j] - \mathbb{E}[X_i]\mathbb{E}[X_j]. \tag{1.6}$$

Definition 1.1.30 (Characteristic function)**.** *If X is an n-dimensional r.v. and μ_X be its law. The characteristic function $\hat{\mu}_X$ of μ_X is defined by*

$$\hat{\mu}_X(\theta) = \int e^{i\langle\theta,x\rangle}\mathrm{d}\mu_X(x) = \mathbb{E}[e^{i\langle\theta,X\rangle}],\ \ \theta \in \mathbb{R}^n \text{ (similar to the Fourier transform).}$$

Note that

(C1) If $\hat{\mu}(\theta) = \hat{\nu}(\theta)$ for every $\theta \in \mathbb{R}^n$, then $\mu = \nu$.

(C2) Let X and Y be two independent r.v.'s with laws μ and ν, respectively. Then $\hat{\mu}_{X+Y}(\theta) = \hat{\mu}_X(\theta)\hat{\nu}_Y(\theta)$

(C3) If $X_1,\dots,X_n$ are r.v.'s with distribution respectively given by $\mu_1,\dots,\mu_n$ then they are independent if and only if, for μ be the law of $X=(X_1,\dots,X_m)$, it holds

$$\hat{\mu}(\theta_1,\dots,\theta_n)=\hat{\mu}_1(\theta_1)\dots\hat{\mu}_n(\theta_n).$$

(C4) Suppose that μ has a finite mathematical expectation (respectively a finite moment of order 2) then $\hat{\mu}$ is differentiable (twice differentiable) and it holds

$$\frac{\partial\hat{\mu}}{\partial\theta_j}(\theta)=i\int x_j e^{i\langle\theta,x\rangle}\mathrm{d}\mu(x),\quad \frac{\partial^2\hat{\mu}}{\partial\theta_j\theta_k}(\theta)=-\int x_k x_j e^{i\langle\theta,x\rangle}\mathrm{d}\mu(x)$$

Particularly

$$\frac{\partial\hat{\mu}}{\partial\theta_j}(0)=i\int x_j\mathrm{d}\mu(x),\quad \frac{\partial^2\hat{\mu}}{\partial\theta_j\theta_k}(0)=i\int x_j x_k\mathrm{d}\mu(x)$$

that is $\hat{\mu}'(0)=i\mathbb{E}[X]$ and $\hat{\mu}'(0)=-\mathbb{E}[X^2]$.

(C5) Let $b\in\mathbb{R}^d$ and $A\in\mathbb{R}^{d\times n}$ matrix. Let $Y=AX+b$ is a $\mathbb{R}^d$-valued r.v. and let ν be its law. Then for $\theta\in\mathbb{R}^d$

$$\hat{\nu}(\theta)=\mathbb{E}[e^{i\langle\theta,AX+b\rangle}]=e^{i\langle\theta,b\rangle}\mathbb{E}[e^{i\langle A^*\theta,X\rangle}]=e^{i\langle\theta,b\rangle}\hat{\mu}(A^*\theta), \tag{1.7}$$

where A^* is the transpose of A.

1.2 Gaussian laws

Definition 1.2.1. *A probability μ on $\mathbb{R}$ is said to be $N(a,\sigma^2)$ (normal, or Gaussian, with mean a and variance σ^2), with $a\in\mathbb{R}$ and $\sigma>0$, if it has density with respect to the Lebesgue measure given by*

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

In addition, its characteristic is given by

$$\hat{\mu}(\theta)=e^{i\theta a-\frac{1}{2}\sigma^2\theta^2}.$$

G1 Let $X_1,\dots,X_d$ be independent and normally distributed r.v with mean 0 and variance 1 and let $X=(X_1,\dots,X_d)$ then the vector X has density

$$f(x)=\frac{1}{(2\pi)^{d/2}}e^{-\frac{1}{2}|x|^2}=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x_1^2}\dots\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x_d^2} \tag{1.8}$$

and its characteristic function is given by

$$e^{-\frac{1}{2}|\theta|^2}=e^{-\frac{1}{2}\theta_1^2}\dots e^{-\frac{1}{2}\theta_d^2}.$$

G2 Let $b\in\mathbb{R}^d$ and A be a $d\times d$ matrix, then the random variable $Y=AX+b$ has the following characteristic function

$$e^{i\langle\theta,b\rangle}e^{-\frac{1}{2}\langle A^*\theta,A^*\theta\rangle}=e^{i\langle\theta,b\rangle}e^{-\frac{1}{2}\langle\Gamma\theta,\theta\rangle}, \tag{1.9}$$

where $\Gamma=AA^*$.

Definition 1.2.2. *Let $b\in\mathbb{R}^d$ and let Γ be a $d\times d$ positive semi-definite matrix. A law μ on $\mathbb{R}^d$ is said to be $N(b,\Gamma)$ (normal with mean b and covariance matrix Γ) if its characteristic function is given by* (1.9).

The following result also holds:

Theorem 1.2.3. *Affine transformations map Gaussian laws into Gaussian laws. More specifically, if X and Y are real r.v.'s with a normal joint law, then $X+Y$ is also normal.*

■

Remark 1.2.4. *Suppose the r.v. $X = (X_1, \dots, X_n)$ is normal and has a diagonal covariance matrix. Then its components $X_1, \dots, X$ are independent r.v.'s.*

Definition 1.2.5. *A family $\mathfrak{F}$ of d-dimensional r.v.'s defined on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is said to be a Gaussian family if, for every choice of $X_1, \dots, X_n \in \mathfrak{F}$, the dn-dimensional r.v. $X = (X_1, \dots, X_n)$ is Gaussian.*

Proposition 1.2.6. *Let $\mathfrak{F}$ be a family of d-dimensional r.v. Then $\mathfrak{F}$ is a Gaussian family iff for every $X_1, \dots, X_n \in \mathfrak{F}$ and $\alpha_1, \dots, \alpha_n \in \mathbb{R}^n$, the r.v. $\langle \alpha_1, X_1 \rangle + \dots + \langle \alpha_n, X_n \rangle$ is Gaussian.*

Proof. Without loss of generality we suppose $d = 1$.

Suppose first that $X = (X_1, \dots, X_n)$ is Gaussian and let $\alpha \in \mathbb{R}^n$, the using Theorem 1.2.3, it holds that $\langle \alpha, X \rangle$ is Gaussian as a linear function of a Gaussian vector.

Next suppose that that every finite linear combination of the r.v.'s of $\mathfrak{F}$ is Gaussian. More precisely, every r.v. of $\mathfrak{F}$ is Gaussian. Thus for $X, Y \in \mathfrak{F}$, it holds

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|X|^2]^{1/2} \mathbb{E}[|Y|^2]^{1/2} < \infty.$$

Therefore $Cov[X,Y]$ is well defined. Let $X_1, \dots, X_n \in \mathfrak{F}$ and let $X = (X_1, \dots, X_n)$ and let μ the law of X. Next we compute the characteristic function $\hat{\mu}$ of X. Let $\alpha_1, \dots, \alpha_n \in \mathbb{R}^n$. We know by hypothesis that $\langle \alpha, X \rangle = \langle \alpha_1, X_1 \rangle + \dots + \langle \alpha_n, X_n \rangle$ is normally distributed with mean $a \in \mathbb{R}$ and variance $\sigma^2 \geq 0$ (i.e., $\langle \alpha, X \rangle \sim N(a, \sigma^2)$). Let μ and Γ denote the expectation and covariance matrix of X, respectively.

$$\begin{aligned} a =& \mathbb{E}[\langle \alpha, X \rangle] = \sum_{i=1}^{n} \alpha_i \mathbb{E}[|X_i|] = \langle \alpha, \mu \rangle \\ \sigma^2 =& \mathbb{E}[\langle \alpha, X \rangle^2] - a^2 = \sum_{i,j=1}^{n} \alpha_i \alpha_j \Big(\mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j] \Big) = \sum_{i,j=1}^{n} \alpha_i \alpha_j \Gamma_{i,j}. \end{aligned}$$

Hence

$$\hat{\mu}(\alpha) = \mathbb{E}[e^{i\langle \alpha, X \rangle}] = e^{ia} e^{-\sigma^2/2} = e^{i\langle \alpha, \mu \rangle} e^{-\frac{1}{2}\langle \Gamma \alpha, \alpha \rangle}.$$

The result follows. □

1.3 Conditional Expectation

1.3.1 Definition and examples

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space on which athe random variable X is defined. Also let $\mathbb{E}[|X|] < \infty$. Recall the following definition.

Definition 1.3.1 (With respect to an event)**.** *Let X be a random variable on $(\Omega, \mathscr{F}, \mathbb{P})$ where Ω is finite, $\mathbb{P}(\{\omega\}) > 0$ for all $\omega \in \Omega$ and $0 < \mathbb{P}(A) \leq 1$ for $A \in \mathscr{F}$. Then*

$$\begin{aligned} \mathbb{E}[X|A] &= \sum_{w \in \Omega} X(\omega) \mathbb{P}(\{\omega\}|A) \qquad (1.10) \\ \textit{where} \quad \mathbb{P}(\{\omega\}|A) &= \frac{P(\{\omega\} \cap A)}{\mathbb{P}(A)} = \begin{cases} \frac{\mathbb{P}(\{\omega\})}{\mathbb{P}(A)} & \textit{if } \omega \in A, \\ 0 & \textit{if } \omega \notin A. \end{cases} \end{aligned}$$

Definition 1.3.2. *Let $\mathscr{G}$ be a sub-σ-algebra of $\mathscr{F}$. The conditional expectation of X given $\mathscr{G}$ is any random variable Y that satisfies:*

(i) Y is $\mathscr{G}$-measurable,

ii) $\int_A Y \mathrm{d}\mathbb{P} = \int_A X \mathrm{d}\mathbb{P}$ for all $A \in \mathscr{G}$.

We use the notation $Y = \mathbb{E}[X|\mathscr{G}]$.

If $\mathscr{G}$ is the σ-algebra generated by some random variable Z, i.e. $\mathscr{G} = \sigma_Z := \sigma(Z)$, then we use the notation $\mathbb{E}[X|Z]$. Property (ii) is called "partial averaging".

Example 1.3.3. *If $\mathscr{G} = \{\emptyset, \Omega\}$, then $\mathbb{E}[X|\mathscr{G}] = \mathbb{E}[X]$. Indeed, since $\mathbb{E}[X]$ is a constant, then it is $\mathscr{G}$-measurable (the σ-algebra of $\mathbb{E}[X]$ is $\{\emptyset, \Omega\}$). For it to be a conditional expectation it must also satisfy the partial averaging property, i.e.*

$$\int_A \mathbb{E}[X]\mathrm{d}\mathbb{P} = \int_A X\mathrm{d}\mathbb{P} \textit{ for all } A \in \mathscr{G}.$$

This trivially holds for $A = \emptyset$, whereas for $A = \Omega$ it follows from the definition of expected value, i.e.

$$\int_\Omega E[X]\mathrm{d}\mathbb{P} = E[X]\mathbb{P}(\Omega) = \mathbb{E}[X] := \int_\Omega X\mathrm{d}\mathbb{P}.$$

Example 1.3.4. *Let $\mathscr{G} = \{\Omega, A, A^c, \Omega\}$ where $A \in \mathscr{F}$. The conditional expectation of X given $\mathscr{G}$ can take at most two values (since only such random variables can generate a σ-algebra of the form $\{\emptyset, A, A^c, \Omega\}$),*

$$\mathbb{E}[X|\mathscr{G}] = \begin{cases} c_1 & \textit{if } \omega \in A, \\ c_2 & \textit{if } \omega \in A^c. \end{cases}$$

The two values c_1 and c_2 are determined from the requirement that $\mathbb{E}[X|\mathscr{G}]$ must satisfy the partial averaging property, i.e. we must have

$$\int_B \mathbb{E}[X|\mathscr{G}]\mathrm{d}\mathbb{P} = \int_B X\mathrm{d}\mathbb{P} \ \textit{ for all } \ B \in \mathscr{G}.$$

This trivially holds for $B = \emptyset$, whereas for $B = A$ and $B = A^c$, we have

$$\int_A \mathbb{E}[X|\mathscr{G}]\mathrm{d}\mathbb{P} = c_1\mathbb{P}(A) = \int_A X\mathrm{d}\mathbb{P}.$$

If $0 < \mathbb{P}(A) < 1$, then

$$c_1 = \frac{1}{\mathbb{P}(A)}\int_A X\mathrm{d}\mathbb{P}.$$

Similarly,

$$\int_{A^c} \mathbb{E}[X|\mathscr{G}]\mathrm{d}\mathbb{P} = c_2\mathbb{P}(A^c) = \int_{A^c} X\mathrm{d}\mathbb{P}.$$

That is

$$c_2 = \frac{1}{\mathbb{P}(A^c)}\int_{A^c} X\mathrm{d}\mathbb{P}.$$

Example 1.3.5. *If $X(\omega) = c$, then $\mathbb{E}[X|\mathscr{G}] = c$.*
To show this, note that the σ-algebra generated by c is $\{\emptyset, \Omega\}$, and thus $\mathbb{E}[X|\mathscr{G}]$ is always $\mathscr{G}$-measurable. The partial averaging also holds since

$$\int_A c\mathrm{d}\mathbb{P} = \int_A \mathbb{E}[X|\mathscr{G}]\mathrm{d}\mathbb{P} = \int_A X\mathrm{d}\mathbb{P} \textit{ for all } A \in \mathscr{G}.$$

Example 1.3.6. *If X is $\mathscr{G}$-measurable, then $\mathbb{E}[X|\mathscr{G}] = X$.*
To show this, note that by assumption, X is $\mathscr{G}$-measurable. Also, the partial averaging trivially holds.

Example 1.3.7. *If X is independent of $\mathscr{G}$, then $\mathbb{E}[X|\mathscr{G}] = \mathbb{E}[X]$.*
Indeed, let $I_A(\omega)$ denote the indicator function for some $A \in \mathscr{G}$, i.e.

$$I_A(\omega) := \begin{cases} 1 & \textit{if } \omega \in A \\ 0 & \textit{if } \omega \notin A. \end{cases}$$

The σ-algebra $I_A(\omega)$ generates is $\{\emptyset, A, A^c, \Omega\}$. Thus, since $I_A(\omega)$ is $\mathscr{G}$-measurable, and X is independent of $\mathscr{G}$, we have that X is independent of I_A.
Since $\mathbb{E}[X]$ is a constant, it is clear that it is $\mathscr{G}$-measurable. We must also show that it satisfies the partial averaging property, i.e.

$$\int_A \mathbb{E}[X]\mathrm{d}\mathbb{P} = \int_A X\mathrm{d}\mathbb{P} \textit{ for all } A \in \mathscr{G}.$$

However, this certainty holds:

$$\int_A \mathbb{E}[X]\mathrm{d}\mathbb{P} = \mathbb{E}[X]\mathbb{P}(A),$$

and

$$\int_A X\mathrm{d}\mathbb{P} = \int_\Omega XI_A\mathrm{d}\mathbb{P} = \mathbb{E}[X]\mathbb{E}[I_A] = \mathbb{E}[X]\mathbb{P}(A).$$

Note that the random variable $X(\omega) = c$ is both (and always) $\mathscr{G}$-measurable and independent of $\mathscr{G}$.

1.3.2 Properties of conditional expectation

Here we state and prove several basic properties of conditional expectation.

(P1) (Linearity) Let X and Y be two integrable random variables and $c_1, c_2 \in \mathbb{R}$. Then

$$\mathbb{E}[c_1X + c_2Y|\mathscr{G}] = c_1\mathbb{E}[X|\mathscr{G}] + c_2\mathbb{E}[Y|\mathscr{G}].$$

(P2) (Monotonicity) Let X and Y be two integrable random variables such that $X \leq Y$. Then

$$\mathbb{E}[X|\mathscr{G}] \leq \mathbb{E}[Y|\mathscr{G}] \text{ a.s..}$$

The conditional probability of an event A given $\mathscr{G}$, is defined as:

$$\mathbb{P}(A|\mathscr{G}) := \mathbb{E}[I_A|\mathscr{G}].$$

(P3) (Chebyshev's inequality) If X is integrable and $\alpha > 0$, then

$$\mathbb{P}(|X| \geq \alpha|\mathscr{G}) \leq \frac{1}{\alpha^2}\mathbb{E}[X^2|\mathscr{G}].$$

(P4) (Integrability) If $Y := \mathbb{E}[X|\mathscr{G}]$, then $\mathbb{E}[|Y|] < \infty$.

(P5) (Tower property) Let $\mathscr{F}_1 \subset \mathscr{F}_2 \subset \mathscr{F}$. Then:

(i) $\mathbb{E}[\mathbb{E}[X|\mathscr{F}_1]|\mathscr{F}_2] = \mathbb{E}[X|\mathscr{F}_1]$,

(ii) $\mathbb{E}[\mathbb{E}[X|\mathscr{F}_2]|\mathscr{F}_1] = \mathbb{E}[X|\mathscr{F}_1]$.

(P6) (Take out what is known) If Y and XY are integrable and X is $\mathscr{G}$-measurable, then

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

(P7) If X is integrable, then $|\mathbb{E}[X|Y]| \leq \mathbb{E}[|X||Y]$.

This is a special case of the **Jensen's inequality** which states that if ϕ is a convex function and $\mathbb{E}[|X|] < \infty$ and $\mathbb{E}[\phi(X)] < \infty$, then

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

Chapter 2

Martingales

Originally, martingale pertained to a class of betting strategies that was popular in 18th-century France.

Definition 2.0.1. *A stochastic process $X := \{X_t : t \in I\}$ is a family of random variables defined on a common probability space, indexed by the index set I.*

Usually I represents time. If $I = \{0, 1, 2, \cdots\}$, then the process is called a **discrete-time** (or discrete-parameter) stochastic process. If $I = \mathbb{R}^+$ (the set of non-negative real numbers), then the process is called a **continuous-time** stochastic process.

Martingales are examples of stochastic processes. We consider both the discrete-time and continuous-time martingales in this chapter.

2.1 Discrete-time martingales

Definition 2.1.1. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. The family of σ-algebras $\{\mathscr{F}_n\}_{n=0}^{\infty}$ is called a filtration if $\mathscr{F}_n \subset \mathscr{F}_{n+1} \subset \mathscr{F}$ for all n.*

Thus, an **increasing** family of σ-algebras is called a filtration, and is used to model the increase of information with time.

Definition 2.1.2. *A probability space with a filtration, i.e. $(\Omega, \mathscr{F}, \{\mathscr{F}_n\}_{n=0}^{\infty}, \mathbb{P})$, is called a filtered probability space.*

Definition 2.1.3. *A stochastic process $\{X_n\}_{n=0}^{\infty}$ is said to be adapted to the filtration $\{\mathscr{F}_n\}_{n=0}^{\infty}$ if X_n is $\mathscr{F}_n$-measurable for all $n \geq 0$.*

Definition 2.1.4. *If $\mathscr{F}_n := \sigma\{X_0, X_1, \ldots, X_n\}$ i.e. the σ-algebra generated by the random variables $X_0, X_1, \ldots, X_n$, then $\{\mathscr{F}_n\}_{n=0}^{\infty}$ is called the natural filtration of the process $\{X_n\}_{n=0}^{\infty}$.*

It is clear that the process is always adapted to its natural filtration. We can now give the definition of a martingale.

Definition 2.1.5. *A process $\{X_n\}_{n=0}^{\infty}$ is called a martingale relative to $(\{\mathscr{F}_n\}_{n=0}^{\infty}, \mathbb{P})$ if:*

(i) $\{X_n\}_{n=0}^{\infty}$ *is adapted to* $\{\mathscr{F}_n\}_{n=0}^{\infty}$,

(ii) $\mathbb{E}[|X_n|] < \infty$ *for all* $n \geq 0$,

(iii) $\mathbb{E}[X_n|\mathscr{F}_{n-1}] = X_{n-1}$, *a.s. for all* $n \geq 1$.

The process is called a supermartingale if instead of (iii) above we have

$$\mathbb{E}[X_n|\mathscr{F}_{n-1}] \leq X_{n-1} \text{ a.s. } \forall n \geq 1.$$

The process is called a submartingale if instead of *iii*.) above we have

$$\mathbb{E}[X_n|\mathscr{F}_{n-1}] \geq X_{n-1} \text{ a.s. } \forall n \geq 1.$$

Note that a process $\{X_n\}_{n=0}^{\infty}$ is a submartingale if and only if $\{-X_n\}_{n=0}^{\infty}$ is a supermartingale. Also note that a process is a martingale if and only if it is a submartingale and a supermartingale.
Property (iii) in the above definition is called the martingale property and is equivalent to

$$\mathbb{E}[X_n|\mathscr{F}_m] = X_m \text{ for all } n > m.$$

Indeed, let $\mathbb{E}[X_n|\mathscr{F}_{n-1}] = X_{n-1}$ for all $n \geq 1$. Then:

$$\begin{aligned}\mathbb{E}[X_n|\mathscr{F}_m] =& \mathbb{E}[\mathbb{E}[X_n|\mathscr{F}_{n-1}]|\mathscr{F}_m] = \mathbb{E}[X_{n-1}|\mathscr{F}_m] \\ =& \ldots = \mathbb{E}[X_m|\mathscr{F}_m] = X_m.\end{aligned}$$

Example 2.1.6. *Let X be an integrable random variable and $\{\mathscr{F}_n\}_{n=0}^{\infty}$ a filtration. Show that he process $\{X_n\}_{n=0}^{\infty}$ defined as*

$$X_n := \mathbb{E}[X|\mathscr{F}_n]$$

is a martingale.

Proof. Indeed, since $\mathbb{E}[X|\mathscr{F}]$ is $\mathscr{F}_n$-measurable, it is clear that $\{X_n\}_{n=0}^{\infty}$ is $\{\mathscr{F}_n\}_{n=0}^{\infty}$-adapted. Also, since the conditional expectation is integrable, then so is (X_n). The martingale property also holds:

$$\mathbb{E}[X_{n+1}|\mathscr{F}_n] = \mathbb{E}[\mathbb{E}[X|\mathscr{F}_{n+1}]|\mathscr{F}_n] = \mathbb{E}[X|\mathscr{F}_n] = X_n.$$

□

Example 2.1.7. *Let $Y_n, n \geq 1$, be a sequence of independent and integrable random variables with mean zero. We define the filtration as: $\mathscr{F}_0 := \{\emptyset, \Omega\}$ and*

$$\mathscr{F}_n := \sigma\{Y_1, \cdots, Y_n\} \textit{ for } \; n \geq 1.$$

The process $\{X_n\}_{n=0}^{\infty}$ defined as: $X_0 = 0$, and

$$X_n := \sum_{i=1}^{n} Y_i \textit{ for } \; n \geq 1,$$

*called the **additive random walk**, is a martingale.*

Proof. Indeed, the adaptivity and integrability of (X_n) are clear. The martingale property also holds:

$$\begin{aligned}\mathbb{E}[X_n|\mathscr{F}_{n-1}] =& \mathbb{E}[Y_1 + \cdots + Y_n|\mathscr{F}_{n-1}] \\ =& Y_1 + \cdots + Y_{n-1} + \mathbb{E}[Y_n|\mathscr{F}_{n-1}] \\ =& X_{n-1} + \mathbb{E}[Y_n] = X_{n-1},\end{aligned}$$

where independence and the fact that the mean is zero were used □

Example 2.1.8. *Let $Y_n, n \geq 1$, be a sequence of independent and integrable random variables with mean one. The filtration is defined as in the previous example, whereas the process $\{X_n\}_{n=0}^{\infty}$ is defined as: $X_0 = 1$, and*

$$X_n := \prod_{i=1}^{n} Y_i \textit{ for all } n \geq 1.$$

*Then $\{X_n\}_{n=0}^{\infty}$, called the **multiplicative random walk**, is a martingale.*

Proof. Indeed, the adaptivity and integrability of X_n are clear. The martingale property is:

$$\begin{aligned}\mathbb{E}[X_n|\mathscr{F}_{n-1}] =& \mathbb{E}[\prod_{i=1}^{n} Y_i|\mathscr{F}_{n-1}] \\ =& \prod_{i=1}^{n-1} Y_i \mathbb{E}[Y_n|\mathscr{F}_{n-1}] \\ =& X_{n-1}\mathbb{E}[Y_n] = X_{n-1}.\end{aligned}$$

□

Example 2.1.9. *Let f and g be two probability density functions with $g(x) \neq 0$ for all $x \in \mathbb{R}$. Let $Y_n, \neq 1$, be a sequence of independent and identically distributed random variables with a common pdf g. Let the filtration be defined as in the previous example, and the process $\{X_n\}_{n=0}^{\infty}$ be defined as $X_0 = 1$,*

$$X_n := \frac{f(Y_1)}{g(Y_1)} \times \frac{f(Y_2)}{g(Y_2)} \times \cdots \frac{f(Y_n)}{g(Y_n)} \text{ for } n \geq 1.$$

Show that (X_n) is a martingale.

Proof. The adaptivity of $\{X_n\}_{n=0}^{\infty}$ is clear. For the integrability we have

$$\begin{aligned}\mathbb{E}[|X_n|] = \mathbb{E}[X_n] &= \mathbb{E}\Big[\prod_{i=1}^{n} \frac{f(Y_i)}{g(Y_i)}\Big] \\ &= \prod_{i=1}^{n} \mathbb{E}\Big[\frac{f(Y_i)}{g(Y_i)}\Big] = \prod_{i=1}^{n} \int_{-\infty}^{\infty} \frac{f(x)}{g(x)} g(x)\mathrm{d}x = \prod_{i=1}^{n} \int_{-\infty}^{\infty} f(x)\mathrm{d}x = \prod_{i=1}^{n} 1 = 1 \text{ for all } n \geq 1.\end{aligned}$$

The martingale property also holds:

$$\begin{aligned}\mathbb{E}[X_n|\mathscr{F}_{n-1}] &= \mathbb{E}\Big[\prod_{i=1}^{n} \frac{f(Y_i)}{g(Y_i)}\Big|\mathscr{F}_{n-1}\Big] \\ &= \prod_{i=1}^{n-1} \frac{f(Y_i)}{g(Y_i)} \mathbb{E}\Big[\frac{f(Y_n)}{g(Y_n)}|\mathscr{F}_{n-1}\Big] \\ &= X_{n-1}\mathbb{E}\Big[\frac{f(Y_n)}{g(Y_n)}\Big] = X_{n-1}.\end{aligned}$$

□

Example 2.1.10. *The stochastic process $X = \{X_t : t \in \{0,1,2,3\}\}$ represents the evolution of a stock price using a binomial tree. The interval of time considered is of length four days. At first, we assume the stock price can go up with probability with probability 1/2 and down with probability 1/2 (we call it $\mathbb{P}$). Secondly, we assume the stock price can go up with probability with probability 1/3 and down with probability 2/3 (we call it $\mathbb{Q}$).*

	$X_0(\omega)$	$X_1(\omega)$	$X_2(\omega)$	$X_3(\omega)$
ω_1	1	2	4	8
ω_2	1	2	4	2
ω_3	1	2	1	2
ω_4	1	2	1	1/2
ω_5	1	1/2	1	2
ω_6	1	1/2	1	1/2
ω_7	1	1/2	1/4	1/2
ω_8	1	1/2	1/4	1/8

(i) What is the filtration generated by the process X?

(ii) Find $\mathbb{E}^{\mathbb{P}}[X_3/X_2]$, the conditional expectation of X_3 given X_2,

(iii) Find $\mathbb{E}^{\mathbb{P}}[X_3/\mathscr{F}_2]$, the conditional expectation of X_3 given $\mathscr{F}_2$.

(iv) Is X a martingale with respect to $(\mathscr{F}, \mathbb{P})$?

(v) Is X a martingale with respect to $(\mathscr{F}, \mathbb{Q})$?

2.1.1 Martingale transform and Doob decomposition

Definition 2.1.11. *A process $\{C_n\}_{n=1}^{\infty}$ is said to be predictable (or previsible) with respect to filtration $\{\mathscr{F}_n\}_{n=0}^{\infty}$. If C_n is $\mathscr{F}_{n-1}$-measurable for all $n \geq 1$.*

From the definition it is clear that if a process is predictable then we know its value at time n based on the information up to time $n-1$.
The process $\{Y_n\}_{n=0}^{\infty}$ defined as: $Y_0 = 0$ and

$$Y_n := \sum_{i=1}^{n} C_i(X_i - X_{i-1}), n \geq 1,$$

for some process $\{X_n\}_{n=0}^{\infty}$, is called the martingale transform of X by C, and is denoted by $Y = C \cdot X$. Note that

$$Y_n - Y_{n-1} = C_n(X_n - X_{n-1}).$$

The martingale transform is a discrete-time analogue of a stochastic integral. It appears in discrete-time mathematical finance where X can represent the asset price, C the trading strategy, and Y investor's wealth.

Theorem 2.1.12. *Let $\{X_n\}_{n=0}^{\infty}$ and $\{C_n\}_{n=1}^{\infty}$ be adapted and predictable processes, respectively.*

(i) If $\{X_n\}_{n=0}^{\infty}$ is a martingale and $\mathbb{E}[|C_n(X_n - X_{n-1})|] < \infty$ for all $n \geq 1$, then $\{Y_n\}_{n=0}^{\infty}$ is a martingale.

(ii) If $\{X_n\}_{n=0}^{\infty}$ is a submartingale (supermartingale) and $0 \leq C_n(\omega) \leq K$, for all $n \geq 0$ and all $\omega \in \Omega$, where $K > 0$, then $\{Y_n\}_{n=0}^{\infty}$ is a submartingale (supermartingale).

■

Proof. We only prove (i) as the proof of (ii) is similar.

$$\begin{aligned}\mathbb{E}[Y_n - Y_{n-1}|\mathscr{F}_{n-1}] =& \mathbb{E}[C_n(X_n - X_{n-1})|\mathscr{F}_{n-1}] \\ =& C_n \mathbb{E}[X_n - X_{n-1}|\mathscr{F}_{n-1}] = 0,\end{aligned}$$

and thus $\mathbb{E}[Y_n|\mathscr{F}_n] = Y_{n-1}$ for all $n \geq 1$ (as Y_{n-1} is $\mathscr{F}_{n-1}$-measurable).

□

Theorem 2.1.13 (Doob decomposition)**.** *$\{X_n\}_{n=0}^{\infty}$ is a submartingale, then there exists a martingale $\{M_n\}_{n=0}^{\infty}$ and a process $\{A_n\}_{n=0}^{\infty}$ satisfying $A_n \geq A_{n-1}$ a.s. and A_n is $\mathscr{F}_{n-1}$-measurable, such that*

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

with $A_0 = M_0 = 0$. This decomposition is unique.

■

Proof. Let $A_0 = 0$ and define

$$A_n := \sum_{i=1}^{n} \mathbb{E}[X_i - X_{i-1}|\mathscr{F}_{i-1}], n \geq 1.$$

It is clear that A_n is $\mathscr{F}_{n-1}$-measurable. Moreover, since X is assumed to be a submartingale, we have $\mathbb{E}[X_i - X_{i-1}|\mathscr{F}_{i-1}] \geq 0$, and thus $A_n \geq A_{n-1}$, a.s..

Further note that

$$A_n - A_{n-1} = \mathbb{E}[X_n - X_{n-1}|\mathscr{F}_{n-1}],$$

and thus

$$\mathbb{E}[X_n - A_n|\mathscr{F}_{n-1}] = X_{n-1} - A_{n-1}.$$

If we define $M_n := X_n - A_n - X_0$, then $\mathbb{E}[M_n|\mathscr{F}_{n-1}] = M_{n-1}$.
This proves the **existence**. For **uniqueness**, let there exist processes $\{M_n\}_{n=0}^{\infty}, \{A_n\}_{n=0}^{\infty}, \{N_n\}_{n=0}^{\infty}, \{B_n\}_{n=0}^{\infty}$, such that

$$\begin{aligned}X_n &= X_0 + M_n + A_n, \\ X_n &= X_0 + N_n + B_n.\end{aligned}$$

Then,

$$M_n - N_n = B_n - A_n.$$

Since A_n and B_n are predictable, then so is $B_n - A_n$ and so is $M_n - N_n$. Thus,

$$\begin{aligned} M_n - N_n =& \mathbb{E}[M_n - N_n | \mathscr{F}_{n-1}] \\ =& M_{n-1} - N_{n-1} = \cdots = M_0 - N_0 = 0. \end{aligned}$$

This implies that $M_n = N_n$-a.s. for all n, and this in turn implies that $A_n = B_n$-a.s for all n, i.e. this proves that the Doob decomposition is unique $a.s$. □

2.1.2 Stopping times

Definition 2.1.14. *Let $\{\mathscr{F}_n\}_{n=0}^{\infty}$ be a given filtration. A map*

$$\tau : \Omega \to \{0, 1, 2, \cdots, \infty\}$$

*is called a **random time**. If*

$$\{\tau \leq n\} := \{\omega \in \Omega : \tau(\omega) \leq n\} \in \mathscr{F}_n, \forall n \leq \infty,$$

then τ is called a stopping time.

Here $\mathscr{F}_\infty := \lim_{n\to\infty} \mathscr{F}_n$. Note that τ can take the value ∞. The statements

$$\{\tau \leq n\} \in \mathscr{F}_n, \forall n \leq \infty \tag{2.1}$$

and

$$\{\tau = n\} \in \mathscr{F}_n, \forall n \leq \infty. \tag{2.2}$$

are equivalent. Indeed, let (2.1) holds. Then,

$$\{\tau = n\} = \{\tau \leq n\} \setminus \{\tau \leq n-1\} = \{\tau \leq n\} \cap \{\tau \leq n-1\}^c \in \mathscr{F}_n.$$

Let (2.2) holds. Then,

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

since $\{\tau = k\} \in \mathscr{F}_k \subset \mathscr{F}_n$ for all $a \leq k \leq n$.

Example 2.1.15. *$\tau := m$, where $n \in \mathbb{N}_0 := \{0, 1, 2, \cdots\}$ is a stopping time, Indeed, it is clearly a random time, and*

$$\{\tau = n\} = \begin{cases} \emptyset & \text{if } n \neq m, \\ \Omega & \text{if } n = m. \end{cases}$$

Thus a stooping time is a generalisation of the single instant of time.

Example 2.1.16 (Hitting time)**.** *Let $X = \{X_n\}_{n=0}^{\infty}$ be an adapted process and $B \in \mathscr{B}$, i.e. B is a Borel set. Also let*

$$\tau := \inf\{n \geq 0 : X_n \in B\}.$$

This is called the hitting time as the time of the first entry of X into B. It is clear that τ is a random time, i.e. a map $\tau : \Omega \to \{0, 1, \cdots, \infty\}$, where $\tau = \infty$ corresponds to the case of X never entering B. Moreover,

$$\{\tau \leq n\} = \bigcup_{k \leq 0}^{n} \{X_k \in B\} \in \mathscr{F}_n,$$

since X is an adapted process.

Example 2.1.17. *Let S and U be two stopping time with respect to a given filtration. Then $\tau = S + U$ is also a stopping time. Indded it is clear that τ is a random time. Moreover*

$$\{\tau = n\} = \bigcup_{k=0}^{n} [\{S = k\} \cap \{U = n-k\}] \in \mathscr{F}_n,$$

since for each $0 \leq k \leq n$, $\{S = k\} \in \mathscr{F}_k \subset \mathscr{F}_n$, and $\{U = n-k\} \in \mathscr{F}_{n-k} \in \mathscr{F}_n$.

Let $\{X_n\}_{n=0}^{\infty}$ be a stochastic process and τ a bounded stopping time, i.e. there exists a non-negative integer c such that $\mathbb{P}\{\tau \leq c\} = 1$.
The stopped random variable $X_\tau(\omega)$ is defined as:

$$X_\tau(\omega) := X_{\tau(\omega)}(\omega),$$

i.e. if for some $\omega \in \Omega$ we have $\tau(\omega) = m$, then $X_\tau(\omega) = X_m(\omega)$.
The stopped random process $X_{\tau\wedge n}$ is defined as:

$$X_{\tau\wedge n}(\omega) := X_{\tau(\omega)\wedge n}(\omega),$$

i.e. if for some $\omega \in \Omega$ we have $\tau(\omega) = m$, then $X_{\tau\wedge n}(\omega) = X_{m\wedge n}(\omega)$.

Theorem 2.1.18. *Let $(\Omega, \mathscr{F}, \{\mathscr{F}_n\}_{n=0}^{\infty}, \mathbb{P})$ be a filtered probability space and $\{X_n\}_{n=0}^{\infty}$ be a $(\{\mathscr{F}_n\}_{n=0}^{\infty}, \mathbb{P})$-martingale (super/sub.). Then $X_{\tau\wedge n}$ is a martingale. Moreover,*

$$\mathbb{E}[X_{\tau\wedge n}] = \mathbb{E}[X_0]$$

(and with appropriate inequalities in the case of super./sub.).
■

Proof. Let the process (C_n^τ) be defined as:

$$C_n^\tau := I_{\tau\geq n}, \text{ for } n \geq 1.$$

Then, C_n^τ is predictable. Indeed, the two possible values of C_n^τ are:

$$C_n^\tau = \begin{cases} 1 & \text{if } \omega \in \{\tau \geq n\}, \\ 0 & \text{if } \omega \in \{\tau \leq n-1\}. \end{cases}$$

The σ-algebra generated by C_n^τ is $\{\emptyset, \{\tau \leq n-1\}, \{\tau \leq n-1\}^c, \Omega\}$ which is in $\mathscr{F}_{n-1}$, and thus C_n^τ is predictable.
The martingale transform of (X_n) by C_n^τ is:

$$(C^\tau \cdot X)_n = \sum_{i=1}^{n} C_i^\tau (X_i - X_{i-1}).$$

If $\tau \geq n$, then

$$(C^\tau \cdot X)_n = \sum_{i=1}^{n} (X_i - X_{i-1}) = X_n - X_0 = X_{\tau\wedge n} - X_0.$$

If $\tau < n$, say $\tau = m < n$ then

$$(C^\tau \cdot X)_n = \sum_{i=1}^{m} (X_i - X_{i-1}) = X_m - X_0 = X_{\tau\wedge n} - X_0.$$

Therefore, $(C^\tau \cdot X)_n = X_{\tau\wedge n} - X_0$.
If (X_n) is a martingale, then so is $((C^\tau \cdot X)_n)$ and so is $X_{\tau\wedge n}$. We also know that in this case it holds:

$$\mathbb{E}[X_{\tau\wedge n}] = \mathbb{E}[X_0].$$

□

Example 2.1.19. *Consider the stochastic process described in Example 2.1.10. Let τ be the first time the process takes the value 2.*

- *Show that τ is a $\{\mathscr{F}\}_{t\in\{0,1,2,3\}}$-stopping time.*
- *Describe the process $Y = \{X_{t\wedge\tau}\}$.*
- *Is $Y = \{X_{t\wedge\tau}\}$ an $(\{\mathscr{F}\}_{t\in\{0,1,2,3\}}, \mathbb{P})$-martingale? an $(\{\mathscr{F}\}_{t\in\{0,1,2,3\}}, \mathbb{Q})$-martingale?*

Theorem 2.1.20. *If (X_n) is a martingale and τ a bounded stopping time, then $\mathbb{E}[X_\tau] = \mathbb{E}[X_0]$.*
■

Proof. We can write X_τ as:

$$X_\tau = \sum_{n=0}^{\infty} X_n(\omega) I_{\{\tau(\omega)=n\}}.$$

Then,

$$\mathbb{E}[X_\tau(\omega)] = \mathbb{E}\Big[\sum_{n=0}^{\infty} X_n(\omega) I_{\{\tau(\omega)=n\}}\Big] = \mathbb{E}\Big[\sum_{n=0}^{c} X_n(\omega) I_{\{\tau(\omega)=n\}}\Big] = \sum_{n=0}^{c} \mathbb{E}\Big[X_n(\omega) I_{\{\tau(\omega)=n\}}\Big].$$

Since $\{\tau(\omega) = n\} \in \mathscr{F}_n$, the indicator function $I_{\{\tau(\omega)=n\}}$ is $\mathscr{F}_n$-measurable. Thus,

$$\begin{aligned}
\mathbb{E}[X_\tau(\omega)] &= \sum_{n=0}^{c} \mathbb{E}\Big[\mathbb{E}\Big[X_n(\omega) I_{\{\tau(\omega)=n\}}|\mathscr{F}_n\Big]\Big] = \sum_{n=0}^{c} \mathbb{E}\Big[\mathbb{E}\Big[X_n(\omega)|\mathscr{F}_n\Big] I_{\{\tau(\omega)=n\}}\Big] \\
&= \sum_{n=0}^{c} \mathbb{E}\Big[\mathbb{E}\Big[X_c(\omega)|\mathscr{F}_n\Big] I_{\{\tau(\omega)=n\}}\Big] = \sum_{n=0}^{c} \mathbb{E}\Big[X_c(\omega) I_{\{\tau(\omega)=n\}}\Big] \\
&= \mathbb{E}\Big[X_c(\omega) \sum_{n=0}^{c} I_{\{\tau(\omega)=n\}}\Big] = \mathbb{E}[X_c(\omega)] = \mathbb{E}[X_0].
\end{aligned}$$

□

The above theorem is a special case of the Doob's optional sampling theorem.

Theorem 2.1.21 (Doob's weak maximal inequality)**.** . *Let $M = \{M_j : j \geq 0\}$ be a submartingale and let $M^* = \{M_j^* : J \geq 0\}$ be the process defined by $M_n^* := \max_{1 \leq j \leq n} M_j$. Then for all $\lambda > 0$, we have*

$$\lambda \mathbb{P}(M_n^* \geq \lambda) \leq \mathbb{E}[M_n 1_{\{M_n^* \geq \lambda\}}].$$

■

Proof. Let τ be the stopping time defined by

$$\tau = \begin{cases} \min\{k : M_k \geq \lambda\} & \text{if } M_n^* \geq \lambda \\ n & \text{if } M_n^* < \lambda \end{cases}$$

Observe that $\{M_n^* \geq \lambda\} = \{M_\tau \geq \lambda\}$. Thus

$$\lambda \mathbb{P}(M_n^* \geq \lambda) = \lambda \mathbb{P}(M_\tau \geq \lambda) = \mathbb{E}[\lambda 1_{\{M_\tau \geq \lambda\}}] \leq \mathbb{E}[M_\tau 1_{\{M_\tau \geq \lambda\}}] = \mathbb{E}[M_\tau 1_{\{M_n^* \geq \lambda\}}].$$

Therefore it is enough to show that $\mathbb{E}[M_\tau 1_{\{M_n^* \geq \lambda\}}] \leq \mathbb{E}[M_n 1_{\{M_n^* \geq \lambda\}}]$. Since τ is a bounded stopping time ($\tau \leq n$) and M_τ is a submartingale, we have $\mathbb{E}[M_\tau] \leq \mathbb{E}[M_n]$. Therefore

$$\mathbb{E}[M_\tau 1_{\{M_n^* < \lambda\}}] + \mathbb{E}[M_\tau 1_{\{M_n^* \geq \lambda\}}] \leq \mathbb{E}[M_n 1_{\{M_n^* < \lambda\}}] + \mathbb{E}[M_n 1_{\{M_n^* \geq \lambda\}}].$$

Note that by definition of τ we have $M_\tau 1_{\{M_n^* < \lambda\}} = M_n 1_{\{M_n^* < \lambda\}}$. Hence the above inequality becomes:

$$\mathbb{E}[M_\tau 1_{\{M_n^* \geq \lambda\}}] \leq \mathbb{E}[M_n 1_{\{M_n^* \geq \lambda\}}].$$

The proof is completed. □

Lemma 2.1.22. *Let X, Y be two nonnegative random variables satisfying*

$$\mathbb{P}\{Y \geq \lambda\} \leq \mathbb{E}[X 1_{\{Y \geq \lambda\}}] \textit{ for all } \lambda > 0.$$

Then for all $p > 1$, it holds

$$\mathbb{E}[Y^p] \leq \Big(\frac{p}{p-1}\Big)^p \mathbb{E}[X^p].$$

Proof. Notice that we have $x^p = \int_0^x p\lambda^{p-1}\mathrm{d}\lambda$. For $X \geq 0$, $\mathbb{E}[X^p]$ can be expressed using Fubini's theorem as:

$$\mathbb{E}[X^p] = \mathbb{E}\Big[\int_0^X p\lambda^{p-1}\mathrm{d}\lambda\Big] = \mathbb{E}\Big[\int_0^\infty 1_{\{\lambda\leq X\}} p\lambda^{p-1}\mathrm{d}\lambda\Big] = \int_0^\infty \mathbb{P}\{\lambda \leq X\} p\lambda^{p-1}\mathrm{d}\lambda$$

Similarly, using the theorem hypothesis,

$$\mathbb{E}[Y^p] = \int_0^\infty \mathbb{P}\{\lambda \leq Y\} p\lambda^{p-1}\mathrm{d}\lambda \leq \int_0^\infty p\lambda^{p-2}\mathbb{E}[X1_{\{Y\geq\lambda\}}]\mathrm{d}\lambda.$$

Set $q = \frac{p}{p-1}$. Using once more Fubini's theorem and Hölder inequality, the right side of the above equation can be bounded as

$$\int_0^\infty p\lambda^{p-2}\mathbb{E}[X1_{\{Y\geq\lambda\}}]\mathrm{d}\lambda = \mathbb{E}\Big[X\int_0^Y p\lambda^{p-2}\mathrm{d}\lambda\Big] = q\mathbb{E}[XY^{p-1}] \leq \|X\|_p\|Y^{p-1}\|_q.$$

Combining this with the above gives

$$\mathbb{E}[Y^p] \leq q(\mathbb{E}[X^p])^{\frac{1}{p}}(\mathbb{E}[Y^p])^{\frac{1}{q}}.$$

Thus if we assume $\mathbb{E}[Y^p] < \infty$, the above can be rewritten as

$$(\mathbb{E}[Y^p])^{\frac{1}{p}} \leq q(\mathbb{E}[X^p])^{\frac{1}{p}}.$$

The result follows from taking the p^{th} power on both sides.

Notice that if we assume that $\mathbb{E}[Y^p] = \infty$, then for any $n \in \mathbb{N}$, the random variable $Y_n = Y \wedge n$ satisfies the hypothesis of the lemma, and the result follows by letting $n \uparrow 1$ and applying the monotone convergence theorem. □

Theorem 2.1.23 (Doob's L^p maximal inequality)**.** *Let $M = \{M_n : n \geq 0\}$ be a submartingale and let $M^* = \{M_j^* : J \geq 0\}$ be the process defined by $M_n^* := \max_{1\leq j\leq n} M_j$ and let $p > 1$. Then*

$$\mathbb{E}[|M_n^*|^p] \leq \Big(\frac{p}{p-1}\Big)^p \mathbb{E}[|M_n|^p].$$

■

Proof. Since $M = \{M_n : n \geq 0\}$ is a submartingale, so is $|M|t = \{|M|_n : n \geq 0\}$. Thus we may assume w.l.o.g that $M_n \geq 0$. Using Theorem 2.1.21, we have

$$\lambda\mathbb{P}(M_n^* \geq \lambda) \leq \mathbb{E}[M_n 1_{\{M_n^*\geq\lambda\}}].$$

Finally applying Lemma 2.1.22 with $X = M_n$ and $Y = M_n^*$, we obtain the desired result. □

2.2 Continuous-time martingales and Brownian motion

In this section we consider the continuous-time martingales, i.e. martingales whose index is the set of non-negative real numbers. We begin with analogous definitions to those in discrete-time.

Definition 2.2.1. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. The family of σ-algebras $\{\mathscr{F}_t\}_{t\geq 0}$ is called a* ***filtration*** *if for any $0 \leq s \leq t$ we have*

$$\mathscr{F}_s \subset \mathscr{F}_t \subset \mathscr{F}.$$

Definition 2.2.2. *A stochastic process $X = \{X(t), t \geq 0\}$ is said to be càdlàg if it almost surely has sample paths which are right continuous with left limit.*

Definition 2.2.3. *Two stochastic processes are called versions (modifications) of another if :*

$$\mathbb{P}(\{X(t) = Y(t)\}) = \mathbb{P}(\{\omega \in \Omega, X(t,\omega) = Y(t,\omega)\}) = 1 \quad \textit{for all} \quad t \geq 0.$$

Example 2.2.4. *Let τ be a uniformly distributed random variable on $[0,1]$ and let $X = \{X(t), t \in [0,1]\}$ and $Y = \{Y(t), t \in [0,1]\}$ defined by;*
$X(t) = 0$ for all $t \in [0,1]$ and

$$Y(t) = \begin{cases} 0 \text{ if } t \in [0,1] \backslash \{\tau\}, \\ 1 \text{ if } t = \tau. \end{cases}$$

Then $\mathbb{P}(Y(t) \neq 0) = \mathbb{P}(\tau = t) = 0$. Hence $\mathbb{P}(Y(t) = 0) = 1 - \mathbb{P}(Y(t) \neq 0) = 1$. Therefore

$$\mathbb{P}(X(t) = Y(t)) = 1 - \mathbb{P}(X(t) \neq Y(t)) = 1 - P(t = \tau).$$

Hence $X(t)$, and $Y(t)$ are versions of one another.

Definition 2.2.5. *Let $X = \{X(t), t \in [0,1]\}$ and $Y = \{Y(t), t \geq 0\}$ be two stochastic processes on $(\Omega, \mathscr{F}, \mathbb{P})$. Then the processes are said to be indistinguishable if:*

$$\mathbb{P}(\{X(t) = Y(t) \quad for \quad t \geq 0\}) = 1.$$

That is

$$\mathbb{P}(\{\omega, \exists t, X(t,\omega) \neq Y(t,\omega)\}) = 0.$$

Definition 2.2.6. *A stochastic process $\{X(t), t \geq 0\}$ is called a martingale with respect to $(\{\mathscr{F}_t\}_{t \geq 0}, \mathbb{P})$ if:*

(i) $\{X(t), t \geq 0\}$ *is* $\{\mathscr{F}_t\}_{t\geq 0}$*-adapted,*

(ii) $\mathbb{E}[|X(t)|] < \infty$ *for all* $t \geq 0$,

(iii) $\mathbb{E}[X(t)|\mathscr{F}_s] = X(s)$ *for all* $0 \leq s \leq t$.

Figure 2.1: Robert Brown (21-12-1773 to 10-6-1858): Scottish botanist and paleobotanist

A very important example of a continuous-time martingale is the following stochastic process.

Definition 2.2.7 (Standard Wiener or Brownian (motion) process)**.** *A real-valued stochastic process $W = \{W(t), t \geq 0\}$ defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is called a standard Wiener (or Brownian motion) process if*

(i) $W(0) = 0$ *a.s.,*

(ii) $W(t) - W(s) \sim \mathscr{N}(0, t-s)$ *for all* $0 \leq s < t < \infty$,

(iii) the increments $W(t)-W(u)$ and $W(u)-W(s)$ are ***independent*** *for all $0 \le s < u < t$,*

Theorem 2.2.8. *Let W be a Brownian motion. Then there exists a modification of W which has continuous paths a.s.*

■

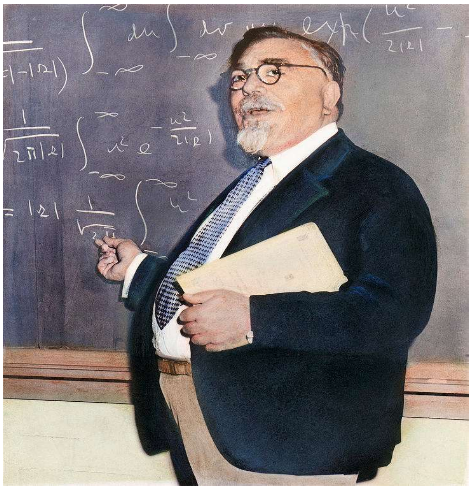


Figure 2.2: Norbert Wiener (26-11-1894 to 18-03-1964): American mathematician and philosopher

Proposition 2.2.9. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. The process W is a standard Brownian motion if and only if*

1. *$W(0) = 0$ a.s.,*
2. *$\alpha_1, \dots, \alpha_n \in \mathbb{R}$, $0 \le t_1 < \dots < t_n$, it holds that $W(t_1), \dots, W(t_n)$ is an n-dimensional standard normal r.v.,*
3. *$\mathbb{E}[W(t)W(s)] = \min(s,t)$.*

Proof. Suppose that W is a Brownian motion.

1. Follows immediatly from the definition. For 2., using Proposition 1.2.6, we need to prove that for $\alpha_1, \dots, \alpha_n \in \mathbb{R}$, $0 \le t_1 < \dots < t_n$

$$\alpha_1 W(t_1) + \ldots + \alpha_n W(t_n)$$

is a normal r.v. We prove this by induction. Suppose $n = 1$ then from (ii) in Definition 2.2.7 with $s = 0$, we have the required result. Suppose that the result is true for $n-1$ and let us show that it holds for n. We have

$$\alpha_1 W(t_1) + \ldots + \alpha_n W(t_n) = \alpha_1 W(t_1) + (\ldots \alpha_{n-1} - \alpha_n) W(t_{n-1}) + \alpha_n (W(t_n) - W(t_{n-1}))$$

We know from (iii) in Definition 2.2.7 that the second term on the right side of the above is independent of the first term and that it is normally distributed. Thus the above sum is normal as a sum of independent normally distributed r.v.

For 3., without loss of generality, let $s \le t$ (and the proof is similar for $s > t$).

$$\begin{aligned}\mathbb{E}[W(t)W(s)] &= \mathbb{E}[(W(t) - W(s) + W(s))W(s)] = \mathbb{E}[(W(t) - W(s))W(s)] + \mathbb{E}[W^2(s)] \\ &= \mathbb{E}[W(t) - W(s)]\mathbb{E}[W(s)] + s = s = \min(s,t).\end{aligned}$$

Conversely, suppose that W satisfies 1., 2., and 3. then (i) in Definition 2.2.7 is immediate. For (ii). Let us first observe that for $0 < s < t$, since (B_s, B_t) is standard normal from 2., it follows that $B_t - B_s$ is also a normal r.v. as a linear combination of (B_s, B_t) (see Proposition 1.2.6). B_t and B_s are centred and so is $B_t - B_s$. In addition

$$\mathbb{E}[(W(t) - W(s))^2] = \mathbb{E}[W^2(t)] + \mathbb{E}[W^2(t)] - 2\mathbb{E}[W(s)W(t)] = t + s - 2s = t - s$$

and (ii) follows.

Finally for (iii) since the Brownian motion is a Gaussian process, we need to show that $W(t)-W(s)$ is uncorrelated to $W(s)$. It follows from the fact that:

$$\mathbb{E}[(W(t)-W(s))W(s)] = \min(s,t)-s = 0.$$

The proof is completed. □

Theorem 2.2.10 (Wiener)**.** *The Brownian motion exists.*

■

From the definition it follows that any $u>0$, we have

$$W(s+u)-W(s) \sim \mathcal{N}(0,u),\ s\geq 0,$$

i.e. $\{W(t), t\geq 0\}$ has **stationary** increments. Also note from the definition that

$$W(t) \sim \mathcal{N}(0,t), t>0.$$

Theorem 2.2.11. *$(W(t), t\geq 0)$ is a martingale with respect to its natural filtration ($\{\mathscr{F}_t, t\geq 0\}$ defined by $\mathscr{F}_t = \sigma(W(s): 0\leq s\leq t)$.*

■

Proof. It is clear that $W(t)$ is adapted to its natural filtration. $(W(t), t>0)$ is integrable:

$$\begin{aligned}\mathbb{E}[|W(t)|] &= \int_{-\infty}^{\infty} |x| \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}\,\mathrm{d}x = 2\int_0^{\infty} x\frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}\,\mathrm{d}x\\ &= \sqrt{\frac{2t}{\pi}}\int_0^{\infty} -e^{-\frac{x^2}{2t}}\,\mathrm{d}\Big(\frac{x^2}{2t}\Big) = \sqrt{\frac{2t}{\pi}}\Big[-e^{-\frac{x^2}{2t}}\Big]_0^{\infty} = \sqrt{\frac{2t}{\pi}} < \infty.\end{aligned}$$

The martingale property holds: for any $0<s<t$, we have

$$\begin{aligned}\mathbb{E}[W(t)|\mathscr{F}_s] =& \mathbb{E}[W(t)-W(s)+W(s)|\mathscr{F}_s]\\ =& \mathbb{E}[W(t)-W(s)|\mathscr{F}_s] + \mathbb{E}[W(s)|\mathscr{F}_s]\\ =& \mathbb{E}[W(t)-W(s)] + W(s) = W(s).\end{aligned}$$

□

Example 2.2.12. *The process $\{W^2(t)-t, t\geq 0\}$ is a martingale with respect to the natural filtration of $W(t), t\geq 0$. Indeed, adaptiveness is immediate, and for integrability we have:*

$$\mathbb{E}[|W^2(t)-t|] \leq \mathbb{E}[|W^2(t)|] + \mathbb{E}[|t|] = \mathbb{E}[W^2(t)] + \mathbb{E}[t] = t+t = 2t < \infty.$$

The martingale property also holds:

$$\begin{aligned}\mathbb{E}[W^2(t)-t|\mathscr{F}_s] =& \mathbb{E}[(W(t)-W(s)+W(s))^2 - t|\mathscr{F}_s]\\ =& \mathbb{E}[(W(t)-W(s))^2 + 2(W(t)-W(s))W(s) + W^2(s) - t|\mathscr{F}_s]\\ =& \mathbb{E}[(W(t)-W(s))^2] + W^2(s) - t = t-s+W^2(s)-t = W^s - s.\end{aligned}$$

Example 2.2.13. *The process $e^{uW(t)-\frac{1}{2}u^2t}$ is a martingale for any $u\in\mathbb{R}$ with respect to the natural filtration of $W(t)$. Again, the adaptiveness is immediate, and for integrability we have:*

$$\begin{aligned}\mathbb{E}\Big[e^{uW(t)-\frac{1}{2}u^2t}\Big] &= \int_{-\infty}^{\infty} e^{ux-\frac{1}{2}u^2t}\frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}\,\mathrm{d}x\\ &= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-ut)^2}{2t}}\,\mathrm{d}x = 1 < \infty.\end{aligned}$$

The martingale property also holds:

$$\begin{aligned}\mathbb{E}\Big[e^{uW(t)-\frac{1}{2}u^2t}|\mathscr{F}_s\Big] =& \mathbb{E}\Big[e^{u(W(t)-W(s))+uW(s)-\frac{1}{2}u^2t}|\mathscr{F}_s\Big]\\ =& \mathbb{E}\Big[e^{u(W(t)-W(s))}\Big]e^{uW(s)-\frac{1}{2}u^2t} = e^{\frac{1}{2}u^2(t-s)}e^{uW(s)-\frac{1}{2}u^2t} = e^{uW(s)-\frac{1}{2}u^2s},\end{aligned}$$

where we have used the fact that $W(t)$ is $\mathscr{F}_t$-measurable and the increment of the Brownian motion are independent.

Example 2.2.14. *Let x_0, μ, σ be given constants. The process*

$$X(t) := x_0 + \mu t + \sigma W(t), t \geq 0,$$

is called ***general*** *Brownian motion. It is clear that*

$$X_t \sim \mathscr{N}(x_0 + \mu t, \sigma^2 t),$$

whereas the distribution of the increment is (for $s < t$):

$$X(t) - X(s) = \mu(t-s) + \sigma[W(t) - W(s)] \sim \mathscr{N}(\mu(t-s), \sigma^2(t-s)).$$

Note that $X(t), t \geq 0$ is ***not*** *a martingale. Indeed,*

$$\begin{aligned} \mathbb{E}\Big[X(t)|\mathscr{F}_s\Big] &= \mathbb{E}\Big[X(t) - X(s) + X(s)|\mathscr{F}_s\Big] \\ &= \sigma\mathbb{E}[W(t) - W(s)] + \mu(t-s) + X(s) = \mu(t-s) + X(s) \neq X(s). \end{aligned}$$

2.2.1 Further properties of Brownian motion

Theorem 2.2.15. *Let $\{W(t), t \geq 0\}$ be a Brownian motion on $(\Omega, \mathscr{F}, \mathbb{P})$. The following processes are Brownian motions:*

(i) (reflection) $X_1(t) := -W(t)$, $t \geq 0$,

(ii) (shift) $X_2(t) := W(t+c) - W(c)$ for any $c > 0$,

(iii) (self-similarity) $X_3(t) := W(c^2 t)/c$ for $c \neq 0$.

■

Proof. In each case we need to show that the defining properties of Brownian motion hold. It is clear that all three processes have continuous paths and have value zero at time zero. Therefore, we only need to show that the increments are independent and normally distributed.

(i) $X_1(t) - X_1(s) = -(W(t) - W(s)) \sim \mathscr{N}(0, t-s)$.
The increments of $X_1(t)$ are independent since so are those of $W(t)$.

(ii)

$$\begin{aligned} X_2(t) - X_2(s) &= W(t+c) - W(c) - W(s+c) + W(c) \\ &= W(t+c) - W(s+c) \sim \mathscr{N}(0, t-s). \end{aligned}$$

Again, the increments of $X_2(t)$ are independent since are those of $W(t)$.

iii.) $X_3(t) - X_3(s) = \frac{1}{c}W(c^2 t) - \frac{1}{c}W(c^2 s) \sim \mathscr{N}(0, t-s)$.
Again, the increments of $X_3(t)$ are independent since are those of $W(t)$.

□

Example 2.2.16. *Let $X(t) := W(t) - tW(1)$ for $0 \leq t \leq 1$. Then:*

$$\begin{aligned} \mathbb{E}[X(t)] &= \mathbb{E}[W(t) - tW(1)] = 0, \\ Var[X(t)] &= \mathbb{E}[X^2(t)] = \mathbb{E}[W^2(t) - 2tW(t)W(1) + t^2 W^2(1)] \\ &= t - 2t^2 + t^2 = t - t^2. \end{aligned}$$

Thus, since the distribution of $X(t)$ is ***not*** *$\mathscr{N}(0, t)$, it is not a standard Brownian motion. This process is called a Brownian bridge; note that $X(0) = X(1) = 0$.*

Example 2.2.17. *If X us a positive random variable and*

$$\log X \sim \mathcal{N}(\mu, \sigma^2),$$

then X is said to have a ***log-normal*** *distribution. If $f(x)$ is the pdf of X, then:*

$$\begin{aligned} f(x) = \frac{\mathrm{d}}{\mathrm{d}x}\mathbb{P}\{X \le x\} &= \frac{\mathrm{d}}{\mathrm{d}x}\mathbb{P}\{\log X \le \log x\} \\ &= \frac{\mathrm{d}}{\mathrm{d}x}\int_{-\infty}^{\log x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\log x-\mu)^2}{2\sigma^2}} \frac{1}{x} \text{ for } x > 0. \end{aligned}$$

The process

$$X(t) := e^{\mu t+\sigma W(t)}, t \ge 0,$$

is called the ***geometric Brownian motion****. It is clear that*

$$\log X(t) = \mu t + \sigma W(t) \sim \mathcal{N}(\mu t, \sigma^2 t),$$

i.e. $X(t)$ is a log-normal random variable. The geometric Brownian motion is used as a model for stock prices.

Theorem 2.2.18 (Kolmogorov)**.** *Let $\{X(t)\ ,\ t \in I\}$ be a stochastic process such that*

$$\mathbb{E}[|X(t) - X(s)|^{\beta}] \le C|t-s|^{1+\alpha}$$

for some positive α, β, C. Then there exists a continuous versions of $X(t)$ that is a continuous process.

■

Theorem 2.2.19.

1. *For any $\alpha < \frac{1}{2}$ almost all sample paths of a Brownian motion are Hölder continuous with exponent α on every bounded interval.*
2. *For any $\alpha > \frac{1}{2}$ almost all sample paths of a Brownian motion are nowhere Hölder continuous with exponent α.*

■

Theorem 2.2.20. *Almost all sample paths of a Brownian motion are nowhere differentiable.*

■

Theorem 2.2.21 (Lévy characterisation)**.** *Let $X = \{X(t), t \ge 0\}$ be a continuous process and let $\{\mathcal{F}_t\}_{t\ge 0}$ be an increasing family of σ-algebras such that $X(t)$ is $\mathcal{F}_t$-measurable and $X(t)$ and $X^2(t) - t$ is a martingale w.r.t $\{\mathcal{F}_t\}_{t\ge 0}$ then $X(t)$ is a Brownian motion.*

■

The variation in the interval $[a,b]$ of a function $f : \mathbb{R} \mapsto \mathbb{R}$ is the quantity defined by

$$V_b^a(f) = \sup_{\Pi} \sum_{i=1}^{n} |f(t_{i+1}) - f(t_i)|,$$

the supremum is taken among all finite partitions $(t_0 = a < t_1 < \cdots < t_{n+1} = b)$ of the interval $[a,b]$. f is said to be of finite variation if $V_b^a(f) < \infty$ for every $a, b \in \mathbb{R}$.

Theorem 2.2.22. *Let $W = \{W(t), t \ge 0\}$ be a Brownian motion and $\Pi_n = \{t_{n,0}, \ldots, t_{n,m_n}\}$ be a partition of a finite closed interval $[a,b]$ with mesh $\Delta_n = \max_{0\le j\le m_n}(t_{n_j} - t_{n_{j-1}})$ converging to zero as $n \to \infty$ $(t_{n,0} = a < t_{n,1} < \cdots < t_{n,m_n} = b)$.*

Let $S_n = \sum_{j=1}^{M_n}(W(t_{n_j}) - W(t_{n_{j-1}}))^2$. Then $S_n \to (b-a)$ in $L^2(\Omega)$ that is $\lim_{n\to\infty}\mathbb{E}[(S_n - (b-a))^2] = 0$.

■

Proof. We wish to show that $\lim_{n\to\infty}\mathbb{E}[(S_n-(b-a))^2]=0$. For simplicity $M_n = n$, $t_{n_j}=t_j$

$$\begin{aligned}\mathbb{E}[(S_n-(b-a))^2] =&\mathbb{E}\Big[\Big(\sum_{j=1}^{M_n}(W(t_j)-W(t_{j-1}))^2-(b-a)\Big)^2\Big]\\ =&\mathbb{E}\Big[\Big(\sum_{j=1}^{M_n}(W(t_j)-W(t_{j-1}))^2-(\sum_{j=1}^{n}t_j-t_{j-1})\Big)^2\Big]\\ =&\mathbb{E}\Big[\Big(\sum_{j=1}^{M_n}(W(t_j)-W(t_{j-1}))^2-(t_j-t_{j-1})\Big)^2\Big]\\ =&\sum_{j=1}^{n}\Big[\mathbb{E}\Big((W(t_j)-W(t_{j-1}))^2-(t_j-t_{j-1})\Big)^2\Big]\end{aligned}$$

(Since the summands are independent and with mean 0).

$$\begin{aligned}=&\sum_{j=1}^{n}\Big\{\mathbb{E}\Big(\Big\{\frac{(W(t_j)-W(t_{j-1}))^2}{(t_j-t_{j-1})}-1\Big\}(t_j-t_{j-1})\Big)^2\Big\}\\ =&\sum_{j=1}^{n}\{\mathbb{E}[(Y_j^2-1)(t_j-t_{j-1})]^2\}\end{aligned}$$

With $Y_j=\frac{W(t_j)-W(t_{j-1})}{(t_j-t_{j-1})^{1/2}}$, (the Y_j normally distributed with mean 0 and variance 1 and independent)

$$\begin{aligned}=&\sum_{j=1}^{n}(t_j-t_{j-1})^2\mathbb{E}[(Y_j^2-1)^2]\\ =&\mathbb{E}(Y_1^2-1)^2\sum_{j=1}^{n}(t_j-t_{j-1})(t_j-t_{j-1})\\ \le&\mathbb{E}[(Y_1^2-1)^2]\Delta_n\sum_{j=1}^{n}(t_j-t_{j-1})\\ =&\mathbb{E}[(Y_1^2-1)^2]\Delta_n(b-a)\to 0 \text{ as } n\to\infty\end{aligned}$$

□

Proposition 2.2.23. *$\{W(t),t\ge 0\}$ be a Brownian motion on a probability space $(\Omega,\mathscr{F},\mathbb{P})$. For every $a\ge 0$ we have*

$$\mathbb{P}(M(t)\ge a)=2\mathbb{P}(W(t)\ge a)=\frac{2}{\sqrt{2\pi t}}\int_a^\infty e^{-\frac{x^2}{t}}\,\mathrm{d}x. \tag{2.3}$$

Proof. We have

$$\begin{aligned}\mathbb{P}(W(t)\ge a) =&\mathbb{P}(W(t)\ge a, M(t)\ge a)+\mathbb{P}(W(t)\ge a, M(t)<a)\\ =&\mathbb{P}(W(t)\ge a, M(t)\ge a)+0,\end{aligned}$$

by definition of $M(t)$. In addition, we have

$$\begin{aligned}\mathbb{P}(W(t)\ge a, M(t)\ge a) =&\mathbb{P}(W(t)\ge a|M(t)\ge a)\mathbb{P}(M(t)\ge a)\\ =&\mathbb{P}(W(t)\ge a|T_a\le t)\mathbb{P}(M(t)\ge a).\end{aligned}$$

We also know that $W(T_a+t)-a=W(T_a+t)-W(T_a)$ is a Brownian motion and by symmetry propertie of a Brownian motion $W(t)$ we have $\mathbb{P}(W(t)\ge 0)=\frac{1}{2}$ for every t. Thus

$$\mathbb{P}(W(t)\ge a|T_a\le t) =\mathbb{P}(W(T_a+(t-T_a))-a\ge 0|T_a\le t)=\frac{1}{2}.$$

Substituting this above yields

$$\mathbb{P}(W(t) \geq a) = \frac{1}{2}\mathbb{P}(M(t) \geq a)$$

and the result follows.

$\square$

Proposition 2.2.24. *$\{W(t), t \geq 0\}$ be a Brownian motion on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. For every $a \geq 0, y \geq 0$ we have*

$$\mathbb{P}(M(t) \geq a, W(t) \leq a-y) = \mathbb{P}(W(t) > a+y). \tag{2.4}$$

Proof. As before we use the definition of $M(t), T_a$ and the symmetry property of the Brownian motion to have:

$$\begin{aligned}
\mathbb{P}(W(t) > a+y) =& \mathbb{P}(W(t) > a+y, M(t) \geq a) + \mathbb{P}(W(t) > a+y, M(t) < a) \\
=& \mathbb{P}(W(t) \geq a, M(t) \geq a) + 0 \\
=& \mathbb{P}(W(t) > a+y | M(t) \geq a)\mathbb{P}(M(t) \geq a) \\
=& \mathbb{P}(W(t) - a > y | T_a \leq t)\mathbb{P}(M(t) \geq a) \\
=& \mathbb{P}(W(T_a + (t-T_a)) - a > y | T_a \leq t)\mathbb{P}(M(t) \geq a) = \mathbb{P}(W(T_a + (t-T_a)) - a \leq -y | T_a \leq t)\mathbb{P}(M(t) \geq a) \\
=& \mathbb{P}(W(t) \leq -y | T_a \leq t)\mathbb{P}(M(t) \geq a) = \mathbb{P}(W(t) \leq a-y, M(t) \geq a),
\end{aligned}$$

where in the 6th equality, we have used the fact that since $W(T_a + (t-T_a)) - a$ is a Brownian motion and its symmetry property.

$\square$

Corollary 2.2.25. *The joint density of $(W(t), M(t))$; is given by*

$$f(b,a) = \begin{cases} \frac{2}{\sqrt{2\pi t^3}}(2a-b)e^{-\frac{1}{2t}(2a-b)^2} \textit{ for } a > 0, b < a \\ 0 \textit{ otherwise} \end{cases} \tag{2.5}$$

The next result shows that the local maxima of the Brownian motion are distinct

Proposition 2.2.26. *$\{W(t), t \geq 0\}$ be a Brownian motion on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let $0 \geq p < q < r < s$. Then*

$$\sup_{p \leq t \leq q} W(t) \neq \sup_{r \leq t \leq s} W(t). \tag{2.6}$$

Proof. Fix any rational number $0 \geq p < q < r < s$, we show that

$$\mathbb{P}(\sup_{p \leq t \leq q} W(t) = \sup_{r \leq t \leq s} W(t)) = 0.$$

Set

$$\begin{aligned}
X &= \sup_{p \leq t \leq q} W(t) - W(r), \\
Y &= \sup_{r \leq t \leq s} W(t) - W(r).
\end{aligned}$$

Since $\{W(r) - W(t), p \leq t \leq q\}$ and $\{W(t) - W(r), r \leq t \leq s\}$ are independent, we have that the random variable X, Y are independent. By using the Markov property, we see that $\{W(t) - W(r), r \leq t\}$ is a Brownian motion. Now let $S(t)$ be the process defined by

$$S(t) = \sup_{t \geq r} W(t) - W(r).$$

We claim that $S(t)$ is a continuous random variable for $t \geq r$. Indeed, using the fact that the Brwonian motion has stationary increments,

$$\begin{aligned}
\mathbb{P}(S(t) \geq a) =& \mathbb{P}(\sup_{t \geq r} W(t) - W(r) \geq a) \\
=& \mathbb{P}(\sup_{t \geq r}(W(t) - W(r)) \geq a) \\
=& \mathbb{P}(\sup_{t \geq r} W(t-r) \geq a) \\
=& 2\mathbb{P}(W(t-r) \geq a) \text{ reflection principle} \\
=& \mathbb{P}(|W(t-r)| \geq a) \text{ symmetry}.
\end{aligned}$$

Therefore

$$\begin{aligned}
\mathbb{P}(\sup_{p\le t\le q} W(t)=\sup_{r\le t\le s} W(t)) =&\mathbb{P}(\sup_{p\le t\le q} W(t)-W(r)=\sup_{r\le t\le s} W(t)-W(r))\\
=&\mathbb{P}(X-Y=0)\\
=&\int_{\mathbb{R}^2} 1_{\{0\}}(x+y)\mathbb{P}_{(X,-Y)}(\mathrm{d}x,\mathrm{d}y)\\
=&\int_{\mathbb{R}}\int_{\mathbb{R}} 1_{\{0\}}(x+y)\mathbb{P}_{(-Y)}(\mathrm{d}y)\mathbb{P}_{(X)}\mathrm{d}x \text{ by independence}\\
=&\int_{\mathbb{R}}\int_{\mathbb{R}} 1_{\{-x\}}(y)\mathbb{P}_{(-Y)}(\mathrm{d}y)\mathbb{P}_{(X)}\mathrm{d}x\\
=&\int_{\mathbb{R}} \mathbb{P}(-Y=-x)\mathbb{P}_{(X)}\mathrm{d}x=0 \text{ since } Y \text{ is continuous,}
\end{aligned}$$

From which we get

$$\mathbb{P}(\bigcup_{0\ge p<q<r<s,\text{rational numbers}} \sup_{p\le t\le q} W(t)=\sup_{r\le t\le s} W(t))=0.$$

The result follows. □

Chapter 3

Stochastic Integrals

In this Chapter we wish to define the notion of stochastic integration. We start with a motivating example.

A motivating example

A machine is producing one type of product. The raw materials are processed by the machine, and the finished products are stored in a buffer.

Suppose at time t the production rate is $u(t)$ and the inventory level in the buffer is $X(t)$. Let $X(0) = X_0$ be the inventory at time $t = 0$, if the demand rate for this product is a known function $Z(t)$ then the relationship these quantities can be described by

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t} = u(t) - Z(t),\ t \geq 0 \text{ and } X(0) = X_0. \tag{3.1}$$

Assume that the demand rate satisfies

$$\frac{\mathrm{d}Z(t)}{\mathrm{d}t} = a(t)Z(t),\ Z(0) = Z_0 \text{ (constant)}, \tag{3.2}$$

where $a(t)$ is the relative demand rate at time t.

Suppose the cost of having the inventory X and production rate u per unit time is $f(x,u)$. A typical example of h is

$$f(x,u) = c_1 x^+ + c_2 x^- + pu,$$

where $c_1, c_2 > 0$ are the marginal cost/penalty for surplus and backlog, respectively, and p is the unit cost of production.

The production management wants to choose a $u(\cdot)$ so as to minimize the total discounted cost over the planning horizon $[0,T]$.

More specifically their objective is to minimize:

$$J(u(\cdot)) = \int_0^T e^{-\gamma t} f(X(t), u(t))\mathrm{d}t, \tag{3.3}$$

where $\gamma > 0$ is the discount rate.

If the machine has a maximum production rate k, then any production plan must be subject to

$$0 < u(t) < k,\ \forall t \in [0,T].$$

If the buffer size is $b > 0$, then the inventory level $X(t)$ must satisfy the constraint $X(t) \leq b$. Any production plan that satisfies the above constraint is called an admissible plan.

What happens if $a(t)$ is not deterministic that is if $a(t)$ is subject to some random environmental effects so that we have

$$a(t) = r(t) + \text{“}\mathit{noise}\text{”?}.$$

Here, we do not know the exact behaviour of the noise term, only its probability distribution. The function $r(t)$ is assumed to be nonrandom. A reasonable mathematical interpretation of the noise term is the so called “white noise” which is formally the derivative of the Brownian motion.

Substituting $a(t)$ to (3.2) gives

$$\frac{\mathrm{d}Z(t)}{\mathrm{d}t} = r(t)Z(t) + \text{“noise”}Z(t),\ Z(0) = Z_0(\text{constant}), \tag{3.4}$$

(3.4) can also be written in an integral form as follows:

$$Z(t) = Z_0 + \int_0^t r(s)Z(s)\mathrm{d}s + \int_0^t Z(s)\text{“noise”}\mathrm{d}s \tag{3.5}$$

If we let $\dot{W}(t) = \text{“noise”}$ then $\dot{W}(t) = \frac{\mathrm{d}W(t)}{\mathrm{d}t}$ i.e., “noise”$\mathrm{d}s = \mathrm{d}W(s)$ and hence (3.5) becomes

$$Z(t) = Z_0 + \int_0^t r(s)Z(s)\mathrm{d}s + \int_0^t Z(s)\mathrm{d}W(s). \tag{3.6}$$

What meaning do we give to the integral $\int_0^t Z(s)\mathrm{d}W(s)$?

Let $\{W(t), t \geq 0\}$ be a one-dimensional standard Brownian motion and $\{\mathscr{F}_t\}_{t\geq 0}$ its natural filtration. If $(f(t), t \geq 0\ )$ is an adapted process and $T \geq 0$, we are interested in giving meaning to the integral

$$I_T(f) = \int_0^T f(t)\mathrm{d}W(t).$$

Here $W(t)$ is called the integrator, whereas $f(t)$ is the integrand. If the paths of $W(t)$ were differentiable, that is if

$$\frac{\mathrm{d}W(t)}{\mathrm{d}t} \text{ existed,}$$

then $I_T(f)$ could have been written as

$$I_T(f) = \int_0^T f(t)\frac{\mathrm{d}W(t)}{\mathrm{d}t}\mathrm{d}t.$$

However, it can be shown that $\frac{\mathrm{d}W(t)}{\mathrm{d}t}$ does not exist. Thus the above approach will not work.

In this chapter, we give the definition of $I_T(f)$ and consider its properties. We begin with a simpler case of integrands $f(t)$, and then show how to proceed for the general class of integrands (the square-integrable processes). We start we the notion of Riemann-Stieltjes integral.

3.1 Riemann-Stieltjes integral

1. A bounded function f defined an a finite closed interval $[a,b]$ is called Riemann integrable if the following limit exists:

$$\int_a^b f(t)\mathrm{d}t = \lim_{\Delta_n\to 0}\sum_{i=1}^n f(\tau_i)(t_i - t_{i-1}), \tag{3.7}$$

   where $\Pi_n = \{t_0, t_1, \ldots, t_n\}$ is a partition of $[a,b]$; $\Delta_n = \max_{1\leq i\leq n}|t_i - t_{i-1}|$; $a = t_0 < t_1 < \cdots < t_n = b$. τ_i is an evaluation in the interval $[t_i, t_{i-1}]$. If f is continuous then f is Riemann integrable.

2. Let g be a monotonically increasing function on $[a,b]$. A bounded f defined on $[a,b]$ is said to be Riemann-Stieltjes integrable with respect to g, if the following limit exits:

$$\int_a^b f(t)\mathrm{d}g(t) = \lim_{\Delta_n\to 0}\sum_{i=1}^n f(\tau_i)(g(t_i) - g(t_{i-1})). \tag{3.8}$$

   What happens if $f = g$ is only continuous? that is $\int_a^b f(t)\mathrm{d}f(t)$?

   Can $\int_a^b f(t)\mathrm{d}f(t)$ be defined as Riemann-Stieltjes integral?

Defined the left and right Riemann integral. Let R_n and L_n denote the right and left sum of the Riemann-Stieltjes integral, respectively. Then

$$\begin{aligned} R_n &= \sum_{i=1}^{n} f(t_i)(f(t_i) - f(t_{i-1})), \\ L_n &= \sum_{i=1}^{n} f(t_{i-1})(f(t_i) - f(t_{i-1})). \end{aligned}$$

Do we always have $\lim_{\Delta_n \to 0} R_n = \lim_{\Delta \to 0} L_n$? We have

$$R_n + L_n = \sum_{i=1}^{n} \left\{ f^2(t_i) - f^2(t_{i-1}) \right\}$$

and

$$\begin{aligned} R_n - L_n &= \sum_{i=1}^{n} \left\{ f^2(t_i) - f(t_i)f(t_{i-1}) + f^2(t_{i-1}) - f(t_i)f(t_{i-1}) \right\} \\ &= \sum_{i=1}^{n} \left\{ f^2(t_i) - 2f(t_i)f(t_{i-1}) + f^2(t_{i-1}) \right\} \\ &= \sum_{i=1}^{n} \left(f(t_i) - f(t_{i-1}) \right)^2. \end{aligned}$$

We then get

$$\begin{aligned} R_n &= \frac{1}{2} \sum_{i=1}^{n} \left\{ f^2(t_i) - f^2(t_{i-1}) \right\} + \left(f(t_i) - f(t_{i-1}) \right)^2, \\ L_n &= \frac{1}{2} \sum_{i=1}^{n} \left\{ f^2(t_i) - f^2(t_{i-1}) \right\} - \left(f(t_i) - f(t_{i-1}) \right)^2. \end{aligned}$$

Hence $\lim_{\Delta_n \to 0} R_n \neq \lim_{\Delta_n \to 0} L_n$. if $\sum_{i=1}^{n} \left(f(t_i) - f(t_{i-1}) \right)^2 \neq 0$ and thus the left and right integral do not coincedes. Since the Brownian motion is and example of continuous function with quadratic variation which is not zero, wer conclude that the integral $\int_a^b W(t) \mathrm{d}W(t)$ cannot be described as a Riemann-Stieltjes integral.

3.2 Wiener Integral

We consider the following integral $\displaystyle\int_a^b f(t) \mathrm{d}W(t, \omega)$, where f is a deterministic function and W is the Brownian motion. Let $L^2([a,b])$ be the space of squared integrable functions that is $\displaystyle\int_a^b |f(t)|^2 \mathrm{d}t < \infty$. We define the stochastic integral in two steps:

Step 1: Suppose that f is a step (simple) function defined by

$$f(t) = \sum_{i=1}^{n} a_i 1_{[t_{i-1}, t_i)},$$

where $t_0 = a$, $t_n = b$. We define the stochastic integral $I(f)$ by

$$I(f) = \sum_{i=1}^{n} a_i (W(t_i) - W(t_{i-1})). \tag{3.9}$$

We have $I(\alpha f + \beta g) = \alpha I(f) + \beta I(g)$ for any $\alpha, \beta \in \mathbb{R}$ and f, g are two step functions.

Proof. Exercise □

Lemma 3.2.1. *For a step function f, $I(f)$ is Gaussian with mean 0 and variance $\mathbb{E}[|I(f)|^2] = \int_a^b |f(t)|^2 \mathrm{d}t$. That is $\int_a^b f(t) \mathrm{d}W(t) \sim \mathcal{N}(0, \int_a^b |f(t)|^2 \mathrm{d}t)$*

Proof. $W(t_i)-W(t_{i-1})$ are Gaussian for $1\le i\le n$ and $W(t_0), W(t_1)-W(t_0),\dots,W(t_n)-W(t_{n-1})$ are independent, it follows that their linear combination is also a Gaussian process. $I(f)$ given by (3.9) is a Gaussian with mean zero.

$$\mathbb{E}[|I(f)|^2]=\mathbb{E}\Big[\sum_{i,j=1}^{n} a_i a_j (W(t_i)-W(t_{i-1}))(W(t_j)-W(t_{j-1}))\Big].$$

Suppose that $i=j$, then:

$$\begin{aligned}\mathbb{E}[|I(f)|^2]=&\mathbb{E}\Big[\sum_{i=1}^{n} a_i^2 (W(t_i)-W(t_{i-1}))^2\Big]\\ =&\sum_{i=1}^{n} a_i^2\mathbb{E}[W(t_i)-W(t_{i-1})]^2\\ =&\sum_{i=1}^{n} a_i^2(t_i-t_{i-1}).\end{aligned}$$

Suppose that $i\neq j$. Without lost of generality, set $i<j$

$$\begin{aligned}&\mathbb{E}\Big[\sum_{i,j=1}^{n} a_i a_j (W(t_i)-W(t_{i-1}))(W(t_j)-W(t_{j-1}))\Big]\\ =&\sum_{i,j=1}^{n} a_i a_j \mathbb{E}(W(t_i)-W(t_{i-1}))\mathbb{E}(W(t_j)-W(t_{j-1})) \text{ by independence}\\ =&0 \quad \text{since } \mathbb{E}(W(t_i)-W(t_{i-1}))=0.\end{aligned}$$

Hence

$$\begin{aligned}\mathbb{E}[|I(f)|^2]=&\sum_{i=1}^{n} a_i^2(t_i-t_{i-1})\\ =&\int_a^b |f(t)|^2\mathrm{d}t\end{aligned}$$

or

$$E\Big(\int_a^b f(t)\mathrm{d}W(t,\omega)\Big)^2=\int_a^b |f(t)|^2\mathrm{d}t.$$

□

Step 2: We wish to define the integral for square integrable functions, that is, for functions $f\in L^2([a,b])$. Denote $L^2(\Omega)$ be the space of square integrable random variables. For $f\in L^2([a,b])$, there exists a sequence $\{f_n\}_{n\ge1}$ of step functions such that $\{f_n\}_{n\ge1}$ converges to f in $L^2([a,b])$.

Lemma 3.2.2. *$\{I(f)_n\}_{n=1}^{\infty}$ is a Cauchy sequence in $L^2(\Omega)$.*

Proof. For $n,m\ge1$

$$\begin{aligned}\|I(f_n)-I(f_m)\|^2_{L^2(\Omega)}=&\|\int_a^b f_n(t)\mathrm{d}W(t,\omega)-\int_a^b f_m(t)\mathrm{d}W(t,\omega)\|^2_{L^2(\Omega)}\\ =&\|\int_a^b (f_n(t)-f_m(t))\mathrm{d}W(t,\omega)\|^2_{L^2(\Omega)} \text{ (By linearity)}\\ =&\int_a^b |f_n(t)-f_m(t)|^2\mathrm{d}t \text{ (By Lemma 3.2.1)}\\ =&\|f_n-f_m\|^2_{L^2([a,b])}.\end{aligned}$$

Since $\{f_n\}_{n\ge1}$ converges to f in $L^2(a,b)$, then $\{f_n\}_{n\ge1}$ is a Cauchy sequence in $L^2([a,b])$. Therefore $\{I(f_n)\}_{n=1}^{\infty}$ is a Cauchy sequence in $L^2(\Omega)$. □

Since $L^2(\Omega)$ is complete, then $\{I(f_n)\}_{n=1}^{\infty}$ converges in $L^2(\Omega)$. We will call its limit the Wiener integral of f and denote by $I(f)$ i.e.

$$I(f) = \lim_{n\to\infty} I(f_n) \quad \text{in } L^2(\Omega). \tag{3.10}$$

Next, we will show that $I(f)$ is well defined that is, it does not depend on the choice of the sequence $\{f_n\}_{n\geq 1}$.

Lemma 3.2.3. *Let $\{f_n\}_{n\geq 1}$ and $\{g_m\}_{m\geq 1}$ be two sequence of step functions that converge to f in $L^2([a,b])$. Then*

$$\lim_{n\to\infty} I(f_n) = \lim_{m\to\infty} I(g_m) \quad \text{in } L^2(\Omega).$$

Proof. As before we have

$$\begin{aligned}
\|I(f_n)-I(g_m)\|^2_{L^2(\Omega)} =& \|I(f_n-g_m)\|^2_{L^2(\Omega)}\\
=& \|f_n-g_m\|^2_{L^2([a,b])}\\
=& \|f_n-f+f-g_m\|^2_{L^2(\Omega)}\\
=& \int_a^b |f_n-f+f-g_m|^2(t)\mathrm{d}t\\
\leq& 2\Big(\int_a^b |f_n-f|^2(t)\mathrm{d}t + \int_a^b |g_m-f|^2(t)\mathrm{d}t\Big) \text{triangle inequality}\\
=& 2\left(\|f_n-f\|^2_{L^2([a,b])} + \|g_m-f\|^2_{L^2([a,b])}\right)\\
\to& 0 \text{ as } n,m\to\infty.
\end{aligned}$$

□

Definition 3.2.4. *Let $f\in L^2([a,b])$. The limit $I(f)$ defined in (3.10) is called the Wiener Integral of f and denoted by $I(f)(w) = \int_a^b f(t)\mathrm{d}W(t,\omega)$, $\omega\in\Omega$ a.s. For simplicity we write $\displaystyle\int_a^b f(t)\mathrm{d}W(t)$.*

Theorem 3.2.5. *For each $f\in L^2([a,b])$ the Wiener Integral $\int_a^b f(t)\mathrm{d}W(t)$ is a Gaussian process with mean 0 and variance $\|f\|^2_{L^2([a,b])} = \int_a^b |f(t)|^2\mathrm{d}t$.*

■

Proof. It follows from the $L^2(\Omega)$ convergence of Gaussian process. □

Corollary 3.2.6. *If $f,g\in L^2([a,b])$, then $\mathbb{E}[I(f)g(f)] = \int_a^b f(t)g(t)\mathrm{d}t$.*

Theorem 3.2.7. *Let $f\in L^2([a,b])$. Then the stochastic process $\{M(t), t\geq 0\}$ defined by $M(t) = \int_a^t f(s)\mathrm{d}W(s)$, $a\leq t\leq b$ is a martingale w.r.t. $\mathscr{F}_t = \sigma(W(s), s\leq t)$*

■

Proof.

1. $M(t)$ is $\mathscr{F}_t$-measurable (by definition).
2. $M(t)$ is integrable for all $t\geq 0$. In fact

$$\begin{aligned}
\mathbb{E}[|M(t)|] \leq& \mathbb{E}[|M(t)|^2]^{1/2} \text{ Cauchy-Schwartz inequality}\\
=& \Big(\int_a^t |f(s)|^2\mathrm{d}s\Big)^{1/2} \qquad \text{by Theorem 3.2.5}\\
\leq& \Big(\int_a^b |f(s)|^2\mathrm{d}s\Big)^{1/2} = \|f\|_{L^2([a,b])} < \infty.
\end{aligned}$$

3. Let us now show the martingale property that is $\mathbb{E}[M(t)|\mathscr{F}_u] = M(u)$ a.s. $(a\leq u\leq t\leq b)$ We have

$$M(t) = \int_a^t f(s)\mathrm{d}W(s) = \int_a^u f(s)\mathrm{d}W(s) + \int_u^t f(s)\mathrm{d}W(s) \tag{3.11}$$

$$= M(u) + \int_u^t f(s)\mathrm{d}W(s). \tag{3.12}$$

Observe that $M(u)$ is $\mathscr{F}_u$-measurable by construction. Thus taking the conditional expectation on both integral gives

$$\mathbb{E}[M(t)|\mathscr{F}_u] = M(u) + \mathbb{E}\Big[\int_u^t f(s)\mathrm{d}W(s)|\mathscr{F}_u\Big].$$

The result follows if we can show that

$$\mathbb{E}\Big[\int_s^t f(r)\mathrm{d}W(r)|\mathscr{F}_s\Big] = 0.$$

We will prove this in two steps. We first start wit step function.

Step 1: Assume that f is a step function $f = \sum_{i=1}^n a_i 1_{[t_{i-1},t_i)}$, $t_0 = s$ and $t_n = t$. Then

$$\int_s^t f(u)\mathrm{d}W(u) = \sum_{i=1}^n a_i(W(t_i) - W(t_{i-1})).$$

Taking the conditional expectation on both side of the above equality and use the fact thaty $W(t_i) - W(t_{i-1})$ are independent of $\mathscr{F}_s$, we get

$$\begin{aligned}\mathbb{E}\Big[\int_s^t f(u)\mathrm{d}W(u)|\mathscr{F}_s\Big] &= \sum_{i=1}^n a_i \mathbb{E}[W(t_i) - W(t_{i-1})|\mathscr{F}_s] \\ &= \sum_{i=1}^n a_i \mathbb{E}[W(t_i) - W(t_{i-1})] = 0.\end{aligned}$$

Step 2: Let $f \in L^2([a,b])$ and $\{f_n\}_{n=1}^\infty$ be a sequence of step functions such that f_n converges to f in $L^2([a,b])$. We wish to show that

$$\mathbb{E}\Big[\int_s^t f(r)\mathrm{d}W(r)|\mathscr{F}_s\Big] = 0.$$

We first show that

$$\mathbb{E}\Big[\int_s^t f_n(u)\mathrm{d}W(u)|\mathscr{F}_s\Big] \text{ converges to } \mathbb{E}\Big[\int_s^t f(u)\mathrm{d}W(u)|\mathscr{F}_s\Big] \text{ in } L^2(\Omega).$$

Let us start by recalling the Jensen inequality

$$\mathbb{E}[X|\mathscr{F}_s]^2 \leq \mathbb{E}[X^2|\mathscr{F}_s].$$

Using the above inequality, we obtain

$$\mathbb{E}\Big[\int_s^t (f_n - f)(u)\mathrm{d}W(u)|\mathscr{F}_s\Big]^2 \leq \mathbb{E}[\Big(\int_s^t (f_n - f)(u)\mathrm{d}W(u)\Big)^2|\mathscr{F}_s\Big].$$

The required convergence limit is computed as

$$\begin{aligned}\mathbb{E}\Big[\mathbb{E}\Big[\int_s^t (f_n - f)(u)\mathrm{d}W(u)\Big|\mathscr{F}_s\Big]^2\Big] \leq &\mathbb{E}\Big[\mathbb{E}\Big[\Big(\int_s^t (f_n - f)(u)\mathrm{d}W(u)\Big)^2|\mathscr{F}_s\Big]\Big] \\ =&\mathbb{E}\Big[\Big(\int_s^t (f_n - f)(u)\mathrm{d}W(u)\Big)^2\Big] \text{ by tower property} \\ =&\int_s^t (f_n - f)^2(u)\mathrm{d}u (\text{by isometry}) \\ \to& 0 \text{ as } n \to \infty.\end{aligned}$$

Hence $\mathbb{E}\Big[\int_s^t f_n(u)\mathrm{d}W(u)|\mathscr{F}_s\Big]$ converges to $\mathbb{E}\Big[\int_s^t f_n(u)\mathrm{d}W(u)|\mathscr{F}_s\Big]$ in $L^2(\Omega)$.

Convergence in $L^2(\Omega)$ implies convergence in probability which also means that there exists a subsequence $(f_{n_k})_{k\geq 1}$ such that

$$\lim_{k\to 0} \mathbb{E}\Big[\int_s^t f_{n_k}(u)\mathrm{d}W(u)|\mathscr{F}_s\Big] = \mathbb{E}\Big[\int_s^t f(u)\mathrm{d}W(u)|\mathscr{F}_s\Big] \text{ a.s.}$$

□

3.3 Stochastic (Itô's) integral

Assume that the deterministic function f is replaced by $f(t,\omega)$ with $f(t,\omega)$ a stochastic process. How do we define $\int_a^b f(t,\omega)\mathrm{d}W(t,\omega)$ such that $M(t)=\int_a^t f(s,\omega)\mathrm{d}W(t,\omega)$ is a martingale?

An Idea

Consider the following example with

$$f(t)=W(t) \text{ i.e } \int_a^b W(t)\mathrm{d}W(t).$$

From the right and left limit of the Riemann-Stieltjes integral, we have

$$\begin{aligned} L_n &= \sum_{i=1}^n W(t_{i-1})(W(t_i)-W(t_{i-1})), \\ R_n &= \sum_{i=1}^n W(t_i)(W(t_i)-W(t_{i-1})), \\ R_n-L_n &= \sum_{i=1}^n (W(t_i)-W(t_{i-1}))^2. \end{aligned}$$

We know from Theorem 2.2.22 that

$$R_n-L_n \to b-a \text{ in } L^2(\omega).$$

On the other hand, we have

$$R_n+L_n=\sum_{i=1}^n W^2(t_i)-W^2(t_{i-1})=W^2(b)-W^2(a).$$

Then we obtain the following limit in $L^2(\Omega)$.

$$\lim_{n\to\infty} R_n = \frac{1}{2}(W^2(b)-W^2(a)+(b-a)), \tag{3.13}$$

$$\lim_{n\to\infty} L_n = \frac{1}{2}(W^2(b)-W^2(a)-(b-a)). \tag{3.14}$$

Assume that $b=t$ and $a=0$, then

$$\begin{aligned} \lim_{n\to\infty} R_n &= \frac{1}{2}(W^2(t)+t)=R(t), \\ \lim_{n\to\infty} L_n &= \frac{1}{2}(W^2(t)-t)=L(t). \end{aligned}$$

$\mathbb{E}[R(t)]=t$ is not constant, that is $R(t)$ is not a martingale. This suggest that we take the left integral.

Definition 3.3.1. *Let $L^p([a,b]\times\Omega)=L^p_\omega([a,b]), 0\le p<\infty$ be the set of stochastic process $\{f(t,\omega), t\in[a,b]\}$ such that*

(i) for every $t\in[a,b]$ the map $(u,\omega)\longmapsto f(u,\omega)$ is $\mathscr{B}([0,t])\otimes\mathscr{F}-$measurable (progressively measurable).

(ii) $\int_a^b \mathbb{E}[|f(t)|^p]\mathrm{d}t<\infty$.

Definition 3.3.2. *Let $L^p_{Loc}([a,b]\times\Omega)=L^p_{\omega,Loc}([a,b]), 0\le p<\infty$ be the set of stochastic process $\{f(t,\omega), t\in[a,b]\}$ such that*

(i) f is progressively measurable.

(ii) $\int_a^b |f(t)|^p\mathrm{d}t<\infty$, a.s.

Remark 3.3.3.

(a) (i) implies that $\{f(t,\omega)\}_{t\ge0}$ is adapted to the filtration $\{\mathscr{F}_t\}_{t\ge0}$.

(b) Every right continuous process that is $\{\mathscr{F}_t\}_{t\geq 0}$-adapted is $\{\mathscr{F}_t\}_{t\geq 0}$-progressively measurable.

(c) $L^p_{Loc}([0,\infty[\times\Omega)$ (resp. $L^p([0,\infty[\times\Omega)$) stands for the space of the processes $\{f(t,\omega), t\geq 0\}$ such that $\{f(t,\omega), t\in[0,T]\}$ belongs to $L^p_{Loc}([0,T]\times\Omega)$ (resp. $L^p([0,T]\times\Omega)$)

(d) Every continuous and adapted process $\{f(t,\omega), t\geq 0\}$ belongs to $L^p_{Loc}([a,b]\times\Omega)$ for every $p\geq 0$. (this follows from (b) and the fact that the continuous map $s\mapsto |X_s(\omega)|^p$ is integrable on every bounded interval.)

(e) Let $\{g(t,\omega), t\geq 0\}$ be bounded and progressively measurable process. If $\{f(t,\omega), t\geq 0\}$ belongs to $L^p_{Loc}([a,b]\times\Omega)$ for every $p\geq 0$, then so is $\{(fg)(t,\omega), t\geq 0\}$.

Again we start with by taking a step (simple) stochastic function **Step 1** Assume that f is a step stochastic process in $L^p_{\omega,loc}([a,b])$

$$f(t,\omega)=\sum_{i=1}^{n}a_{i-1}(\omega)1_{[t_{i-1},t_i)}(t),\ a=t_0<t_1<\ldots<t_n=b, \tag{3.15}$$

where a_{i-1} is $\mathscr{F}_{t_{i-1}}$ -measurable. Note that the expression (3.15) is right-continuous and thus taking $a=\{a_i, 0\leq i\leq n\}$ adapted to $\{\mathscr{F}_{t_i}\}_{0\leq i\leq n}$ ensures that f is progressively measurable.

Note that $\{g(t,\omega), t\geq 0\}\in L^2_\omega([a,b]$ and only if $\mathbb{E}[|a_{i-1}|^2]<\infty$.

We define

$$I(f)=\sum_{i=1}^{n}a_{i-1}(\omega)(W(t_i)-W(t_{i-1})). \tag{3.16}$$

It is clear that the linearity is satisfied.

Lemma 3.3.4. *Let $I(f)$ be defined by equation* (3.16)*, then*

$$\mathbb{E}[I(f)]=0 \text{ and } \mathbb{E}[|I(f)|^2]=\int_a^b\mathbb{E}[|f(t)|^2]\mathrm{d}t.$$

Proof. In order to show that $\mathbb{E}[I(f)]=0$ we show that the expectation of each summand in (3.16) is zero.

$$\begin{aligned}
&\mathbb{E}[a_{i-1}(\omega)(W(t_i)-W(t_{i-1}))]\\
=&\mathbb{E}[\mathbb{E}[a_{i-1}(\omega)(W(t_i)-W(t_{i-1}))|\mathscr{F}_{t_{i-1}}]] \text{ by Tower property}\\
=&\mathbb{E}[a_{i-1}(\omega)\mathbb{E}[(W(t_i)-W(t_{i-1}))|\mathscr{F}_{t_{i-1}}]] \text{ since } a_{i-1} \text{ is } \mathscr{F}_{t_{i-1}}\text{-measurable}\\
=&\mathbb{E}[a_{i-1}(\omega)\mathbb{E}[B(t_i)-B(t_{i-1})]] \text{ by independance property}\\
=&\mathbb{E}[a_{i-1}(\omega)\times 0]=0.
\end{aligned}$$

To prove that $\mathbb{E}[|I(f)|^2]=\int_a^b\mathbb{E}[|f(t)|^2]$ we first write

$$|I(f)|^2=\sum_{i,j=1}^{n}a_{i-1}a_{j-1}[(W(t_i)-W(t_{i-1}))(W(t_j)-W(t_{j-1}))].$$

Using the conditional expectational once more, we get the result. □

Lemma 3.3.5. *Let $f\in L^2_\omega([a,b])$ (resp. $L^2_{\omega,loc}([a,b])$). Then there exists a sequence $\{f_n\}_{n=1}^\infty$ of step stochastic processes in $L^2_\omega([a,b])$ (resp. $L^2_{\omega,loc}([a,b])$) such that*

$$\lim_{n\to\infty}\int_a^b\mathbb{E}[|f_n(t)-f(t)|^2]\mathrm{d}t=0. \tag{3.17}$$

resp

$$\lim_{n\to\infty}\int_a^b|f_n(t)-f(t)|^2\mathrm{d}t=0,\quad a.s. \tag{3.18}$$

Let $f\in L^2_\omega([a,b])$ then choose a sequence $\{f_n\}_{n=1}^\infty$ of step stochastic processes such that equation (3.17) holds. $I(f_n)$ is a Cauchy sequence in $L^2(\Omega)$. We define

$$I(f)=\lim_{n\to\infty}I(f_n) \text{ in } L^2(\Omega). \tag{3.19}$$

Definition 3.3.6. *The limit $I(f)$ defined by equation* (3.19) *is called Itô's integral of f and is denoted*

$$I(f)=\int_a^b f(t)\mathrm{d}W(t).$$

Proposition 3.3.7. *The above integral satisfies*

$$\mathbb{E}[I(f)]=0$$

and

$$\mathbb{E}\Big[\Big(\int_a^b f(t)\mathrm{d}W(t)\Big)^2\Big]=\int_a^b \mathbb{E}[|f(t)|^2]\mathrm{d}t \text{ Itô isometry}.$$

Theorem 3.3.8. *Let $\{X(t),t\geq 0\}\in L^2([0,T]\times\Omega)$ and $I_t(X)=\int_0^t X(s)\mathrm{d}W(s)$. Then $(I_t(X),0\leq t\leq T)$ is a square integrable $(\{\mathscr{F}_t\}_{t\geq 0},\mathbb{P})$-martingale whose increasing process is*

$$A(t)=\int_0^t X^2(s)\mathrm{d}s.$$

A is the process such that $M=I^2(X)-A$ is a martingale

■

Theorem 3.3.9. *Let $\tau<T$ be an $\{\mathscr{F}_t\}_{t\geq 0}$-stopping time. Let $\{X(t),t\geq 0\}$ be a process in $L^2([0,T]\times\Omega)$. Then $\{X(t)I_{t<\tau},t\geq 0\}\in L^2([0,T]\times\Omega)$ and we have*

$$\int_0^\tau X(s)\mathrm{d}W(s)=\int_0^T X(s)I_{s<\tau}\mathrm{d}W(s),\ a.s \tag{3.20}$$

■

Next we wish to define the stochastic integral for processes $\{X(t),t\geq 0\}$ in $L^2_{loc}([0,T]\times\Omega)$. The idea is to approximate $\{X(t),t\geq 0\}\in L^2_{loc}([0,T]\times\Omega)$ by processes in $L^2([0,T]\times\Omega)$.

Remark 3.3.10. *In fact for every $n\geq 1$, let τ_n be the random time define by $\tau_n=\{t\leq T;\int_0^t X^2(s)\mathrm{d}s>n\}$ with the understanding that $\tau_n=T$ if $\int_0^t X^2(s)\mathrm{d}s\leq n$. Then $\{\tau_n\}_{n\geq 1}$ is a stopping time and the process $\{X_n(t),t\geq 0\}$ defined by $X_n(t)=X(t)I_{t<\tau_n}$ belongs to $L^2([0,T]\times\Omega)$. Indeed, we have from Theorem 3.3.9*

$$\int_0^T X_n^2(t)\mathrm{d}t=\int_0^T X^2(t)I_{t<\tau_n}\mathrm{d}t=\int_0^{\tau_n\wedge T} X^2(t)\mathrm{d}t\leq n$$

Thus $\{X_n(t),t\geq 0\}\in L^2([0,T]\times\Omega)$. Thus, for every $n\geq 1$ the stochastic integral

$$I_n(t)=\int_0^T X_n(t)\mathrm{d}W(t)$$

is well defined.

Definition 3.3.11. *Let $\{X(t),t\geq 0\}\in L^2_{loc}([0,T]\times\Omega)$. Its stochastic integral is defined as*

$$I(t)=\int_0^T X(t)\mathrm{d}W(t)=\lim_{n\to\infty}\int_0^T X(t)I_{t<\tau_n}\mathrm{d}W(t),\ a.s.$$

Lemma 3.3.12. *Let $\{X(t),t\geq 0\}\in L^2_{loc}([0,T]\times\Omega)$. Then for every $\varepsilon>0,\rho>0$*

$$\mathbb{P}\big(\big|\int_a^b X(t)\mathrm{d}W(t)\big|>\varepsilon\big)\leq\mathbb{P}\big(\big|\int_a^b X^2(t)\mathrm{d}t\big|>\rho\big)+\frac{\rho}{\varepsilon^2}.$$

Proposition 3.3.13. *Let $\{X(t),t\geq 0\}$ and $\{X_n(t),t\geq 0\}_{n\geq 1}\in L^2_{loc}([0,T]\times\Omega)$. Suppose*

$$\int_a^b \big|X(t)-X^n(t)\big|^2\mathrm{d}t\to 0 \text{ as } n\to\infty \text{ (in probability)}.$$

Then

$$\int_a^b X_n(t)\mathrm{d}W(t)\to\int_a^b X(t)\mathrm{d}W(t) \text{ as } n\to\infty \text{ (in probability)}$$

Proof. Using Lemma 3.3.12,

$$\mathbb{P}\big(\big|\int_a^b X_n(t)\mathrm{d}W(t) - \int_a^b X(t)\mathrm{d}W(t)\big| > \varepsilon\big) = \mathbb{P}\big(\big|\int_a^b \{X_n(t) - X(t)\}\mathrm{d}W(t)\big| > \varepsilon\big)$$
$$\leq \mathbb{P}\big(\int_a^b \big|X_n(t) - X(t)\big|^2 \mathrm{d}t > \rho\big) + \frac{\rho}{\varepsilon^2}$$

Let $\alpha > 0$ and choose ρ so that $\frac{\rho}{\varepsilon^2} \leq \frac{\alpha}{2}$. Since $\int_a^b \big|X(t) - X^n(t)\big|^2 \mathrm{d}t \to 0$ as $n \to \infty$ (in probability), choose n_0 such that for $n > n_0$

$$\mathbb{P}\big(\int_a^b \big|X_n(t) - X(t)\big|^2 \mathrm{d}t > \rho\big) < \frac{\alpha}{2}.$$

Thus for $n > n_0$, it holds

$$\mathbb{P}\big(\big|\int_a^b X_n(t)\mathrm{d}W(t) - \int_a^b X(t)\mathrm{d}W(t)\big| > \varepsilon\big) < \alpha.$$

□

Proposition 3.3.14. *Let $\{X(t), t \geq 0\} \in L^2_{loc}([0,T] \times \Omega)$ be a continuous process. Then for every sequence $(\Delta_n)_n$ of partitions $a = t_{n,0} < t_{n,1} < \ldots < t_{n,m_n} = b$ with $\Delta_n = \max|t_{n,k+1} - t_{n,k}| \to 0$ we have*

$$\sum_{k=0}^{m_n-1} X(t_{n,k})\{W(t_{n,k+1}) - W(t_{n,k})\} \to \int_a^b X(t)\mathrm{d}W(t) \text{ as } n \to \infty \text{ (in probability).}$$

Proof. Define the process $\{X_n(t), t \geq 0\}_{n \geq 1}$ by:

$$X_n(t) = \sum_{k=0}^{m_n-1} X(t_{n,k}) I_{[t_{n,k}, t_{n,k+1}[}(t).$$

Then $\{X_n(t), t \geq 0\}_{n \geq 1}$ is a step process and we have

$$\sum_{k=0}^{m_n-1} X(t_{n,k})\{W(t_{n,k+1}) - W(t_{n,k})\} = \int_a^b X_n(t)\mathrm{d}W(t).$$

Since the paths of the processes are continuous, we have

$$\int_a^b \big|X(t) - X^n(t)\big|^2 \mathrm{d}t \to 0 \text{ as } n \to \infty \text{ (a.s.)}$$

Thanks to Proposition 3.3.13,

$$\sum_{k=0}^{m_n-1} X(t_{n,k})\{W(t_{n,k+1}) - W(t_{n,k})\} = \int_a^b X_n(t)\mathrm{d}W(t) \to \int_a^b X(t)\mathrm{d}W(t) \text{ as } n \to \infty \text{ (in probability)}$$

□

3.3.1 Local Martingales

Note that if $f \in L^2_\omega([a,b])$, $I(t)$ is a martingale. However, this is not true in general if $f \in L^2_{\omega,loc}([a,b])$, since in this case $I(f)$ might not be integrable. In this section, we see how $(I_t(f), t \geq 0)$ can be approximated by martingale.

Definition 3.3.15. *Let $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P})$ be a filtered probability space. Let $M = (M(t), t \geq 0)$ be a process. Then M is called an $(\{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P})$-local martingale if there exists an increasing sequence of stopping time $\{\tau_n\}_n$ such that*

1. *$\tau \nearrow +\infty$ as $n \to \infty$ a.s.*
2. *$(M(t \wedge \tau_n), t \geq 0)$ is an $(\{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P})$-martingale for every n.*

Remark 3.3.16. *In fact if $\{X(t), t \geq 0\} \in L^2_{loc}([0,T] \times \Omega)$, then its stochastic integral with respect to the Brownian motion is a local martingale. Indeed, let τ_n be the stopping time defined by $\tau_n := \{t \leq T; \int_0^t X^2(s)\mathrm{d}s > n\}$ and let $(I_t(X), t \geq 0)$ be the process defined by*

$$I_t(X) = \int_0^t X(s)\mathrm{d}W(s).$$

Then

$$I_{t\wedge\tau_n}(X) = \int_0^{t\wedge\tau_n} X(s)\mathrm{d}W(s) = \int_0^t X(s)I_{s<\tau_n}\mathrm{d}W(s).$$

Since $s \mapsto X(s)I_{s<\tau_n}$ belongs to $L^2([0,T] \times \Omega)$ (see Remark 3.3.10), it follows that $(I_{t\wedge\tau_n}(X), t \geq 0)$ is a square integrable martingale.

Proposition 3.3.17.

1. *A positive local martingale is a supermartingale.*
2. *A bounded local martingale is a martingale.*

Proposition 3.3.18. *Let M be a continuous local martingale. Then there exists a unique continuous increasing process $A = (A(t), t \geq 0)$ such that $X_t = M^2(t) - A(t)$ is a continuous local martingale. A is called the associated increasing process to M.*

Proposition 3.3.19. *Let $\{X(t), t \geq 0\} \in L^2_{loc}([0,T] \times \Omega)$ and $I_t(X) = \int_0^t X(s)\mathrm{d}W(s)$. Then $(I_t(X), 0 \leq t \leq T)$ is a local martingale whose increasing process is*

$$A(t) = \int_0^t X^2(s)\mathrm{d}s.$$

Proof. Thanks to Remark 3.3.16, $(I_t(X), 0 \leq t \leq T)$ is a local martingale. The proof is completed if we prove that $(I_t^2(X) - A(t), 0 \leq t \leq T)$ is a local martingale. As before, let τ_n be the stopping time defined by $\tau_n := \{t \leq T; \int_0^t X^2(s)\mathrm{d}s > n\}$. Then $(I_{t\wedge\tau_n}(X), t \geq 0)$ defined by

$$I_{t\wedge\tau_n}(X) = \int_0^{t\wedge\tau_n} X(s)\mathrm{d}W(s) = \int_0^t X(s)I_{s<\tau_n}\mathrm{d}W(s).$$

is a square integrable martingale. We also know by the definition of an increasing associated to a martingale $M \in L^2([0,T] \times \Omega)$, the process

$$I^2_{t\wedge\tau_n}(X) - A(t\wedge\tau_n) = \big(\int_0^t X(s)I_{s<\tau_n}\mathrm{d}W(s)\big)^2 - \int_0^t X^2(s)I_{s<\tau_n}\mathrm{d}s.$$

is a martingale. Thus the result follows. □

We denote by $(\langle M\rangle_t, t \geq 0)$ the associated increasing process of the local martingale M.

Corollary 3.3.20. *If M and N are continuous local martingales, then there exists a unique process A with finite variation such that the process $\{Z(t), t \geq 0\}$ defined by $Z(t) = M(t)N(t) - A(t)$ is a continuous local martingale.*

3.3.2 Itô's formula

Motivation

Consider the Leibniz-Newton chain rule for differentiation. Let f, g be two differentiable functions, then $f(g(t))$ is differentiable and

$$\frac{\mathrm{d}}{\mathrm{d}t}(f(g(t))) = f'(g(t))g'(t).$$

This implies

$$f(g(t)) - f(g(a)) = \int_a^t f'(g(s))g'(s)\mathrm{d}s = \int_a^t f'(g(s))\mathrm{d}g(s). \tag{3.21}$$

What happens if we replace $g(t)$ by $W(t)$?
The equality

$$\frac{\mathrm{d}}{\mathrm{d}t}(f(W(t))) = f'(W(t))W'(t)$$

has no meaning since the path of Brownian motion are no where differentiable in time.
Question 1: Is it true that

$$f(W(t)) - f(W(a)) = \int_a^t f'(W(s))\mathrm{d}W(s)?$$

Let choose $f(x) = x^2$. Then we have $f'(x) = 2x$. Thus the above question become: Is is true that

$$W^2(t) - W^2(a) = 2\int_a^t W(s)\mathrm{d}W(s)?$$

or

$$\int_a^t W(s)\mathrm{d}W(s) = \frac{1}{2}(W^2(t) - W^2(a))?$$

We know from our previous computations that

$$\int_a^t W(s)\mathrm{d}W(s) = \frac{1}{2}[W^2(t) - W^2(a) - (t-a)].$$

Hence the answer to **Question 1** is NO.
What is then the decomposition ?

$$f(W(t)) - f(W(a)) = \sum_{i=1}^n f(W(t_i)) - f(W(t_{i-1}))\ t_0 = a\ \ t_n = t$$

Suppose that f is twice continuously differentiable then

$$f(x) - f(x_0) = f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0 + \lambda(x - x_0))(x - x_0)^2;\ \text{where } 0 < \lambda < 1$$

So,

$$\begin{aligned} f(W(t)) - f(W(a)) = &\sum_{i=1}^n f'(W(t_{i-1}))(W(t_i) - W(t_{i-1})) \\ &+ \frac{1}{2}\sum_{i=1}^n f''(W(t_{i-1}) + \lambda(W(t_i) - W(t_{i-1})))(W(t_i) - W(t_{i-1}))^2 \end{aligned}$$

The first term on the right converges to $\int_a^t f'(W(s))\mathrm{d}W(s)$ in $L^2(\Omega)$ and the second term converges in certain sense (we will not show that here) to $\frac{1}{2}\int_a^t f''(W(s))\mathrm{d}s$.

Theorem 3.3.21. *Let f be a $\mathscr{C}^2$ function then*

$$f(W(t)) - f(W(a)) = \int_a^t f'(W(s))\mathrm{d}W(s) + \frac{1}{2}\int_a^t f''(W(s))\mathrm{d}s$$

■

Theorem 3.3.22. *Let f be a $\mathscr{C}^{1,2}([a,b]\times\mathbb{R})$ function. Then*

$$\begin{aligned} f(t,W(t)) =& f(a,W(a))+\int_a^t \frac{\partial f}{\partial s}(s,W(s))\mathrm{d}s+\int_a^t \frac{\partial f}{\partial x}(s,W(s))\mathrm{d}W(s) \\ &+\frac{1}{2}\int_a^t \frac{\partial^2 f}{\partial x^2}(s,W(s))\mathrm{d}s \end{aligned} \tag{3.22}$$

■

Example 3.3.23. *Use Itô's formula to find*

$$\int_0^t sW(s)\mathrm{d}W(s).$$

Choose f such that

$$\frac{\partial f}{\partial x}(t,x)=tx.$$

Then

$$f(t,x)=\frac{1}{2}tx^2+C(t)+K \quad ; \quad \frac{\partial f}{\partial t}=\frac{1}{2}x^2+C'(t) \quad ; \quad \frac{\partial f}{\partial x}=xt \quad ; \quad \frac{\partial^2 f}{\partial^2 x}=t$$

Using Itô's formula, we get

$$\begin{aligned} f(t,W(t)) =& f(0,W(0))+\int_0^t \frac{\partial f}{\partial s}(s,W(s))\mathrm{d}s+\int_0^t \frac{\partial f}{\partial x}(s,W(s))\mathrm{d}W(s) \\ &+\frac{1}{2}\int_0^t \frac{\partial^2 f}{\partial x^2}(s,W(s))\mathrm{d}s \end{aligned}$$

Substituting f yields:

$$\frac{1}{2}tW^2(t)+C(t)+K=\frac{1}{2}\times 0+C(0)+K+\int_0^t \frac{1}{2}W^2(s)+C'(s)\mathrm{d}s+\int_0^t sW(s)\mathrm{d}W(s)+\frac{1}{2}\int_0^t s\mathrm{d}s$$

From which we get

$$\begin{aligned} \int_0^t sW(s)\mathrm{d}W(s) =& \frac{1}{2}tW^2(t)-\frac{1}{2}\int_0^t s\mathrm{d}s-\frac{1}{2}\int_0^t W^2(s)\mathrm{d}s \\ =& \frac{1}{2}tW^2(t)-\frac{1}{2}\left[\frac{t^2}{2}\right]-\frac{1}{2}\int_0^t W^2(s)\mathrm{d}s \\ =& \frac{1}{2}tW^2(t)-\frac{t^2}{4}-\frac{1}{2}\int_0^t W^2(s)\mathrm{d}s \end{aligned}$$

Definition 3.3.24. *An Itô process is a stochastic process of the form:*

$$X(t)=X(a)+\int_a^t \mu(s)\mathrm{d}s+\int_a^t B(s)\mathrm{d}W(s) \quad ; \quad a\geq t\geq b, \tag{3.23}$$

where $X(a)$ is $\mathscr{F}_a-$measurable, $B\in L^2_\omega([a,b])$ and $\mu\in L^1_\omega([a,b])$

For an Itô process, the Itô's formula gives the differential of $f(t,X(t))$.

Theorem 3.3.25 (Itô's formula)**.** *Let f be a given function in $C^{1,2}$. Let X be an Itô process as in the previous definition. Then we have*

$$\mathrm{d}f(t,X(t))=\Big[\frac{\partial f}{\partial t}+\frac{\partial f}{\partial x}\mu(t)+\frac{1}{2}\frac{\partial^2 f}{\partial x^2}B(t)^2\Big]\mathrm{d}t+\frac{\partial f}{\partial x}B(t)\mathrm{d}W(t),\ \ t\geq 0,$$

or in integral form

$$f(t,X(t))=f(0,X(0))+\int_0^t \Big[\frac{\partial f}{\partial s}+\frac{\partial f}{\partial x}\mu(s)+\frac{1}{2}\frac{\partial^2 f}{\partial x^2}B(s)^2\Big]\mathrm{d}s+\int_0^t \frac{\partial f}{\partial x}B(s)\mathrm{d}W(s),\ \ t\geq 0,$$

provided that the drift is integrable and $\frac{\partial f}{\partial x}B$ is square-integrable.

■

We omit the proof of this important result due to its length. The main ideas of the proof are the Taylor expansion of $f(t,X(t))$ and the inclusion of the quadratic term $(\mathrm{d}X)^2$ due to the non-zero quadratic variation of the stochastic integral.

Example 3.3.26. *Compute the integral* $\int_0^t W(s)\mathrm{d}W(s)$.
We apply the Itô's formula to $f(x)=x^2$ *as follows:*

$$
\begin{aligned}
W^2(t) =& W^2(0)+\int_0^t \mathrm{d}s+2\int_0^t W(s)\mathrm{d}W(s)\\
=&.t+2\int_0^t W(s)\mathrm{d}W(s)
\end{aligned}
$$

Thus

$$
\int_0^t W(s)dW(s)=\frac{1}{2}W^2(t)-\frac{t}{2}.
$$

Example 3.3.27. *Similarly compute* $\int_0^t W^2(s)\mathrm{d}W(s)$.
Once more, we apply Itô's formula to the function $f(x)=x^3$ *to get*

$$
W^3(t) =3\int_0^t W(s)ds+3\int_0^t W^2(s)\mathrm{d}W(s).
$$

From which we get

$$
\int_0^t W^2(s)\mathrm{d}W(s) =\frac{1}{3}W^3(t)-\int_0^t W(s)\mathrm{d}s.
$$

Example 3.3.28. *Let* $\mathscr{B}_k(t):=\mathbb{E}\big[W^k(t)\big]$, $k=0,1,\cdots$. *Use Itô's formula to show that :*

(i) $\mathscr{B}_k(t)=\frac{1}{2}k(k-1)\int_0^t \mathscr{B}_{k-2}(s)\mathrm{d}s,\ \ k\geq 2,$

(ii) $\mathbb{E}\big[W^{2k}(t)\big]=\frac{(2k)!t^k}{2^k k!},\ \ k=1,2,\cdots.$

By Itô's formula to $f(x)=x^3$, *we have*

$$
W^k(t)=\int_0^t \frac{1}{2}k(k-1)W^{k-2}(s)\mathrm{d}s+\int_0^t kW^{k-1}(s)\mathrm{d}W(s),
$$

Thus (since the stochastic integral is a martingale)

$$
\begin{aligned}
\mathscr{B}_k(t) =&\mathbb{E}[W^k(t)]=\frac{1}{2}k(k-1)\int_0^t \mathbb{E}[W^{k-2}(t)]\mathrm{d}s+0\\
=&\frac{1}{2}k(k-1)\int_0^t \mathscr{B}_{k-2}(s)\mathrm{d}s,\ \ k\geq 2
\end{aligned}
$$

.

We prove (ii)by induction. Thus, if $k=1$, *we have* $\mathbb{E}[W^2(t)]=t$. *We assume that (ii) holds for* $k=m$, *that is,*

$$
\mathbb{E}\big[W^{2m}(t)\big]=\frac{(2m)!t^m}{2^m m!}.
$$

We now prove that (ii) holds for $k=m+1$:

$$
\begin{aligned}
\mathbb{E}\big[W^{2(m+1)}(t)\big] =&\frac{1}{2}2(m+1)[2(m+1)-1]\int_0^t \mathbb{E}\big[W^{2(m+1)-2}(s)\big]\mathrm{d}s\\
=&\frac{1}{2}2(m+1)2(m+1)\int_0^t \mathbb{E}\big[W^{2m}(s)\big]\mathrm{d}s\\
=&\frac{1}{2}\frac{2(m+1)(2m+1)(2m)!}{2^m m!}\ \frac{t^{m+1}}{m+1}\\
=&\frac{[2(m+1)]!}{2^{m+1}(m+1)!}t^{m+1}.
\end{aligned}
$$

The conclusion now follows by induction.

3.3.3 Multidimensional Itô's formula

In this section we derive the $d$dimensional Itô's formula

Definition 3.3.29. *A d-dimensional Brownian motion is a process $W(t) := (W_1(t), \dots, W_d(t))$ such that $W_i(t)$, $i = 1, \cdots, d$ are one-dimensional standard Brownian motions, and $W_i(t)$ and $W_j(t)$ are independent if $i \neq j$.*

Consider the following Itô processes:

$$\begin{cases} \mathrm{d}X_1(t) = \mu_1(t)\mathrm{d}t + B_{11}(t)\mathrm{d}W_1(t) + \cdots + B_{1d}(t)\mathrm{d}W_d(t), \\ \vdots \\ \mathrm{d}X_n(t) = \mu_n(t)dt + B_{n1}(t)\mathrm{d}W_1(t) + \cdots + B_{nd}(t)\mathrm{d}W_d(t), \\ X_1(0), \dots, X_n(0), \text{ are given,} \end{cases}$$

where the processes μ_i, $i = 1, \cdots, n$ are integrable and $B_{ij}(t)$, $i = 1, \dots, n$, $j = 1, \dots, d$, are square-integrable (of course, all processes are assumed to be adapted). These can be written in a vector-matrix form as:

$$\begin{cases} \mathrm{d}X(t) = \mu(t)dt + B(t)\mathrm{d}W(t), \\ X(0) \text{ is given,} \end{cases}$$

where

$$X(t) := \begin{bmatrix} X_1(t) \\ \vdots \\ X_n(t) \end{bmatrix}, \mu(t) := \begin{bmatrix} \mu_1(t) \\ \vdots \\ \mu_n(t) \end{bmatrix}, B(t) := \begin{bmatrix} B_{11}(t) & \cdots & B_{1d}(t) \\ \vdots & & \vdots \\ B_{n1}(t) & \cdots & B_{nd}(t) \end{bmatrix}.$$

The process $X(t)$ is called an n-dimensional Itô process. Let $f : [0, \infty) \times \mathbb{R}^n \to \mathbb{R}$ be a given $C^{1,2}$-function such that

$$\frac{\partial f}{\partial t}, \frac{\partial f}{\partial x} := \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}, \frac{\partial^2 f}{\partial x^2} := \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix},$$

have continuous elements. The following is the multi-dimensional Itô's formula, where $^\intercal$ denotes the transpose of a matrix and Tr the trace of a matrix (that is the sum of its diagonal elements).

Theorem 3.3.30. *Let f be a function as above. The differential of the process $f(t, X(t))$ is given by:*

$$\mathrm{d}f = \left\{ \frac{\partial f}{\partial t} + \left(\frac{\partial f}{\partial x} \right)^\intercal \mu + \frac{1}{2} \mathrm{Tr} \left[\frac{\partial^2 f}{\partial x^2} B B^\intercal \right] \right\} \mathrm{d}t + \left(\frac{\partial f}{\partial x} \right)^\intercal B \mathrm{d}W(t),$$

whenever the drift is integrable and $\left(\frac{\partial f}{\partial x}\right)^\intercal B$ is square-integrable.

■

Theorem 3.3.31 (Itô's product rule)**.** *Consider the following two Itô processes*

$$\begin{cases} \mathrm{d}X_1(t) = \mu_1(t)\mathrm{d}t + B_1(t)\mathrm{d}W(t), \\ \mathrm{d}X_2(t) = \mu_2(t)dt + B_2(t)\mathrm{d}W(t), \\ X_1(0), X_2(0), \text{ are given,} \end{cases}$$

where $W(t)$ is one dimensional. Then $X_1(t)X_2(t)$ is also an Itô's process and

$$\mathrm{d}\big[X_1(t)X_2(t)\big] = (\mathrm{d}X_1)X_2 + X_1(\mathrm{d}X_2) + B_1 B_2 \mathrm{d}t.$$

■

Proof. Note that in this case we have $n = 2$ and $d = 1$. We apply Itô's formula to the function

$$f(t, x_1, x_2) = x_1 x_2.$$

Its partial derivatives are

$$\frac{\partial f}{\partial t} = 0, \frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \frac{\partial^2 f}{\partial x^2} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

By the multi-dimensional Itô's formula we have:

$$\begin{aligned}
\mathrm{d}f = \mathrm{d}(X_1 X_2) =& \begin{bmatrix} X_2 & X_1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \mathrm{d}t + \frac{1}{2} \mathrm{Tr}\Big(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \begin{bmatrix} B_1 & B_2 \end{bmatrix} \Big) \mathrm{d}t \\
&+ \begin{bmatrix} X_2 & X_1 \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \mathrm{d}t \\
=& \big(X_2 \mu_1 + X_1 \mu_2\big) \mathrm{d}t + \frac{1}{2} \mathrm{Ttr}\Big(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} B_1^2 & B_1 B_2 \\ B_1 B_2 & B_2^2 \end{bmatrix} \Big) \mathrm{d}t \\
&+ \big(X_2 B_1 + X_1 B_2\big) \mathrm{d}W \\
=& \big(X_2 \mu_1 + X_1 G_2\big) \mathrm{d}t + \frac{1}{2} \big(B_1 B_2 + B_1 B_2\big) \mathrm{d}t + \big(X_2 B_1 + X_1 B_2\big) \mathrm{d}W \\
=& X_2 \mathrm{d}X_1 + X_1 \mathrm{d}X_2 + B_1 B_2 \mathrm{d}t.
\end{aligned}$$

In applying Itô's product rule, we adapt the formaliser:

$$(\mathrm{d}t)^2 = 0, \mathrm{d}t \mathrm{d}W(t) = 0, (\mathrm{d}W(t))^2 = \mathrm{d}t.$$

We can now write Itô's product rule as:

$$\mathrm{d}(X_1 X_2) = (\mathrm{d}X_1) X_2 + X_1 \mathrm{d}(X_2) + (\mathrm{d}X_1)(\mathrm{d}X_2).$$

□

Chapter 4

Stochastic Differential Equations

In this chapter, we discuss the concept of stochastic differential equation driven by Brownian motion and its solution in the strong and weak sense. We will also discuss some related topics.

4.1 Existence and Uniqueness of Solutions

Let $b_i : [0,\infty[\times\mathbb{R}^d \mapsto \mathbb{R}$, $\sigma_{ij} : [0,\infty[\times\mathbb{R}^d \mapsto \mathbb{R}$, $1 \leq i \leq d$; $1 \leq j \leq m$, be Borel measurable functions representing the components of d-dimensional drift vector $b(t,x) = (b_i(t,x))_{1\leq i\leq d}$ and the $d\times m$-dispersion matrix $\sigma(t,x) = (\sigma_{ij}(t,x))_{1\leq i\leq d, 1\leq j\leq m}$.

Let $W = (W(t), 0 \leq t < \infty)$ be an m-dimensional Brownian motion. We are given the stochastic differential equation (SDE)

$$\begin{cases} \mathrm{d}X(t) = b(t,X(t))\mathrm{d}t + \sigma(t,X(t))\mathrm{d}W(t),\ 0 \leq t < \infty, \\ X(0) = \xi \end{cases} \tag{4.1}$$

or componentwise as

$$\begin{cases} \mathrm{d}X_i(t) = b_i(t,X(t))\mathrm{d}t + \sum_{j=1}^{m} \sigma_{ij}(t,X(t))\mathrm{d}W_j(t),\ 0 \leq t < \infty \\ X(0) = \xi_i \end{cases}$$

Example 4.1.1.

$$\begin{pmatrix} \mathrm{d}X_1(t) \\ \mathrm{d}X_2(t) \end{pmatrix} = \begin{pmatrix} b_1(t,X_1(t),X_2(t)) \\ b_2(t,X_1(t),X_2(t)) \end{pmatrix} \mathrm{d}t + \begin{pmatrix} \sigma_{11}(t,X_1(t),X_2(t)) & \sigma_{12}(t,X_1(t),X_2(t)) \\ \sigma_{21}(t,X_1(t),X_2(t)) & \sigma_{22}(t,X_1(t),X_2(t)) \end{pmatrix} \begin{pmatrix} \mathrm{d}W_1(t) \\ \mathrm{d}W_2(t) \end{pmatrix}$$

$$\mathrm{d}X_1(t) = b_1(t,X_1(t),X_2(t))\mathrm{d}t + \sigma_{11}(t,X_1(t),X_2(t))\mathrm{d}W_1(t) + \sigma_{12}(t,X_1(t),X_2(t))\mathrm{d}W_2(t).$$

Here $X = (X(t), 0 \leq t < \infty)$ is a suitable d-dimensional process with continuous sample path representing the solution of equation (4.1).

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $W = (W(t), 0 \leq t < \infty)$ be a Brownian motion on it. Let $\mathscr{F}_t^W = \sigma(W(s), 0 \leq s \leq t)$. We assume that the space is rich enough to accommodate the random vector ξ taking values in $\mathbb{R}^d$, independent of $\mathscr{F}_\infty^W = \sigma(\bigcup_{t\geq 0} \mathscr{F}_t^W)$ and with a given distribution $N(\mathscr{O}) = \mathbb{P}(\xi \in \mathscr{O})$; $\mathscr{O} \in \mathscr{B}(\mathbb{R}^d)$.

Consider the following filtration $\mathscr{H}_t = \sigma(\xi) \vee \mathscr{F}_t^W = \sigma(\xi, W(s), 0 \leq s \leq t)$, as well as the collection of null sets $\mathscr{N} = \{M \subset \Omega; \exists G \in \mathscr{H}_\infty \text{ with } M \subseteq G \text{ and } \mathbb{P}(G) = 0\}$ and create an augmented filtration

$$\mathscr{F}_t = \sigma(\mathscr{H}_t \cup \mathscr{N}),\ 0 \leq t < \infty;\ \mathscr{F}_\infty = \sigma(\bigcup_{s\geq 0} \mathscr{F}_s). \tag{4.2}$$

Remark 4.1.2. *Since $(W(t), 0 \leq t < \infty)$ is a d-dimensional Brownian motion with respect to $\{\mathscr{H}_t\}_{t\geq 0}$, it is also a Brownian motion with respect to $\{\mathscr{F}_t\}_{t\geq 0}$.*

Definition 4.1.3. *A* strong solution *of the SDE* (4.1) *on the given probability space* $(\Omega, \mathscr{F}, \mathbb{P})$ *and with respect to the fixed Brownian motion* $W = (W(t), 0 \le t < \infty)$ *and initial condition* ξ *is a stochastic process* $X = (X(t), 0 \le t < \infty)$ *with continuous sample paths and with the following properties:*

(i) $X = (X(t), 0 \le t < \infty)$ *is adapted to the filtration* $\{\mathscr{F}_t\}_{t \ge 0}$;

(ii) $\mathbb{P}(X(0) = \xi) = 1$;

(iii) It holds for every $1 \le i \le d$, $1 \le j \le m$ *and* $0 \le t < \infty$ *that*

$$\mathbb{P}\left(\int_0^t |b_i(s, X(s))|^2 + |\sigma_{ij}(s, X(s))|^2 \mathrm{d}s < \infty\right) = 1;$$

(iv) The integral version of (4.1)

$$X(t) = X(0) + \int_0^t b(s, X(s))\mathrm{d}s + \int_0^t \sigma(s, X(s))\mathrm{d}W(s), \tag{4.3}$$

holds almost surely.

Below we plot path of the Brownian and of geometric Brownian motion representing the price of an asset

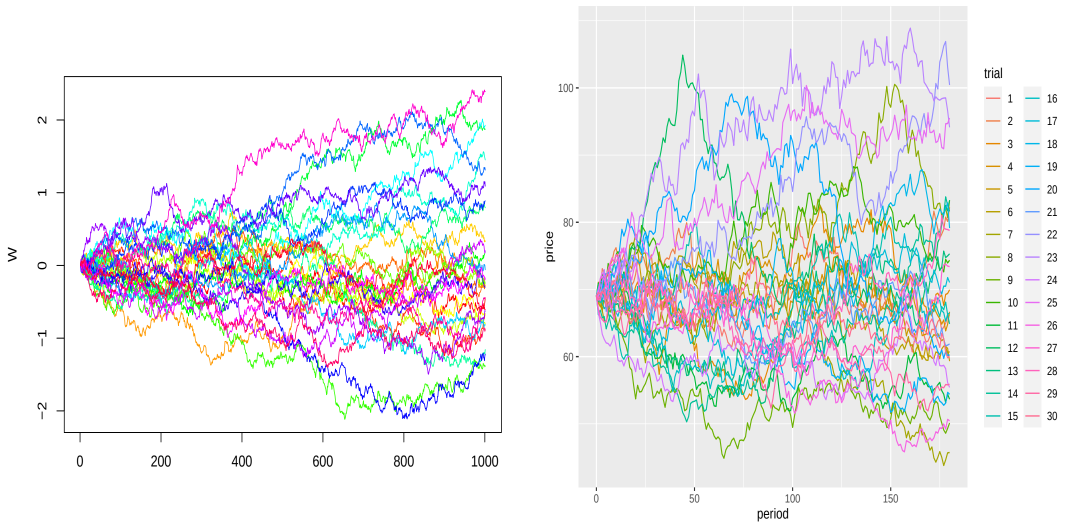


Figure 4.1: Brownian motion and Geometric Brownian motion

Definition 4.1.4. *Let* b *and* σ *be given. Suppose that whenever* $W = (W(t), 0 \le t < \infty)$ *is an* m*-dimensional Brownian motion on some* $(\Omega, \mathscr{F}, \mathbb{P})$, ξ *is an independent* d*-dimensional vector,* $\{\mathscr{F}_t\}_{t \ge 0}$ *is given by* (4.2) *and* $X(t)$ *and* $\bar{X}(t)$ *are two solutions of* (4.1) *relative to* W *with initial condition* ξ *then* $\mathbb{P}(X(t) = \bar{X}(t), 0 \le t < \infty) = 1$. *Under these condition. we say that* strong uniqueness *holds for* (4.1).

In fact, the version of the Brownian motion is given in advance in the formulation of strong solution and the solution constructed from it is $\mathscr{F}_t$-adapted.

Definition 4.1.5. *Assume now that we are only given* b *and* σ *and asked for a triplet* $(\tilde{X}, \tilde{W}, \{\tilde{\mathscr{F}}_t\}_{t \ge 0})$ *on a probability space* $(\Omega, \mathscr{F}, \mathbb{P})$ *such that conditions (iii) and (iv) in Definition 4.1.3 hold, then* $(\tilde{X}, \tilde{W}, \{\tilde{\mathscr{F}}_t\}_{t \ge 0})$ *is called* weak solution.

Definition 4.1.6. *Suppose whenever* $(X, W, \{\mathscr{F}_t\}_{t \ge 0})$ *and* $(\tilde{X}, W, \{\tilde{\mathscr{F}}_t\}_{t \ge 0})$ *are weak solutions to* (4.1) *with common Brownian motion (relative to different filtration) on a common probability space and with common initial value, that is* $\mathbb{P}(X(0) = \tilde{X}(0)) = 1$. *If the two processes are indistinguishable, that is* $\mathbb{P}(X(t) = \tilde{X}(t), 0 \le t < \infty) = 1$, *we say that* pathwise uniqueness *exists for equation*(4.1).

Now, we are ready to give the main theorem of this section

Theorem 4.1.7. *Let $T > 0$ be fixed. Suppose that the coefficients $b(t,x)$ and $\sigma(t,x)$ satisfy global Lipschtix and linear growth conditions, that is*

$$\|b(t,x) - b(t,y)\| + \|\sigma(t,x) - \sigma(t,y)\| \le k_1\|x - y\| \tag{4.4}$$

$$\|b(t,x)\| + \|\sigma(t,x)\| \le k_2(1 + \|x\|), \tag{4.5}$$

for every $0 \le t \le T$, $x, y \in \mathbb{R}^d$ and k_1, k_2 positive constants. Let ξ be an $\mathbb{R}^d$-valued random vector defined on some probability space and which is independent of the Brownian filtration $\{\mathscr{F}_t^W\}_{0 \le t \le T}$ and such that $\mathbb{E}[|\xi|^2] < \infty$. Let $\{\mathscr{F}_t\}_{t \ge 0}$ defined as in (4.2)*. Then, there exists a unique strong t-continuous solution $X = (X(t), 0 \le t < \infty)$ of equation* (4.1)*. Moreover,*

$$\mathbb{E}\left[\int_0^t |X(s)|^2 \mathrm{d}s\right] < \infty \text{ for } 0 \le s \le t < T.$$

■

Proof of Theorem 4.1.7. To be given during online lecture. □

Remark 4.1.8.

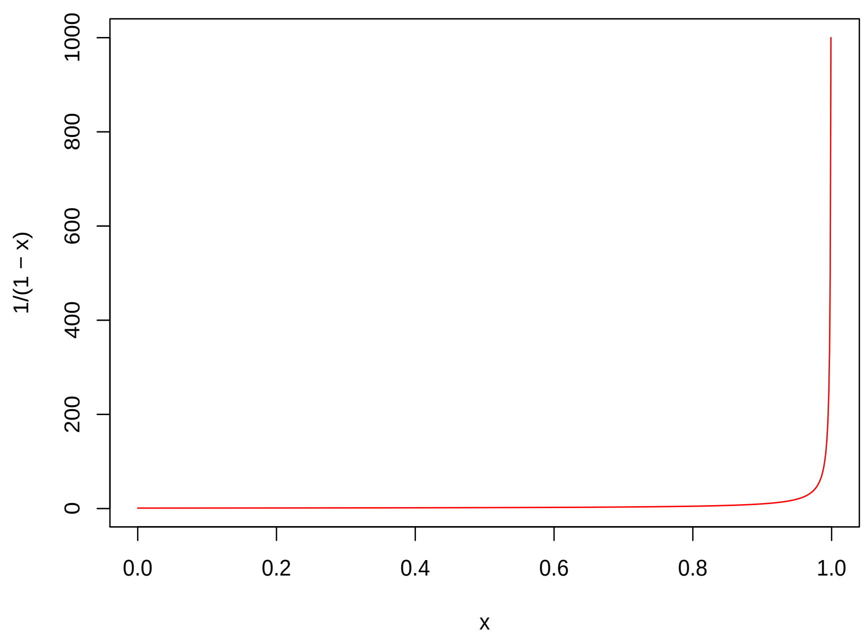
1) *Let us see the importance of* (4.5) *for SDE with $\sigma = 0$ that is an ODE.*

   Consider the following ODE

$$\begin{cases} \dot{X}(t) = \dfrac{\mathrm{d}X(t)}{\mathrm{d}t} = X^2(t) \\ X(0) = 1. \end{cases}$$

   This correspond to $b(x) = x^2$, which does not satisfy (4.5)*. The solution to this ODE is $X(t) = \frac{1}{1-t}$, $t \in \mathbb{R}_+ \backslash \{1\}$. It is hence difficult to find a global solution (define for all t) in this case.*

   More generally, (4.5) *ensures that the solution $X(t)$ does not explode in a finite time as shown in the graph below.*



2) *Let us see the importance of* (4.4) *in an ODE.*

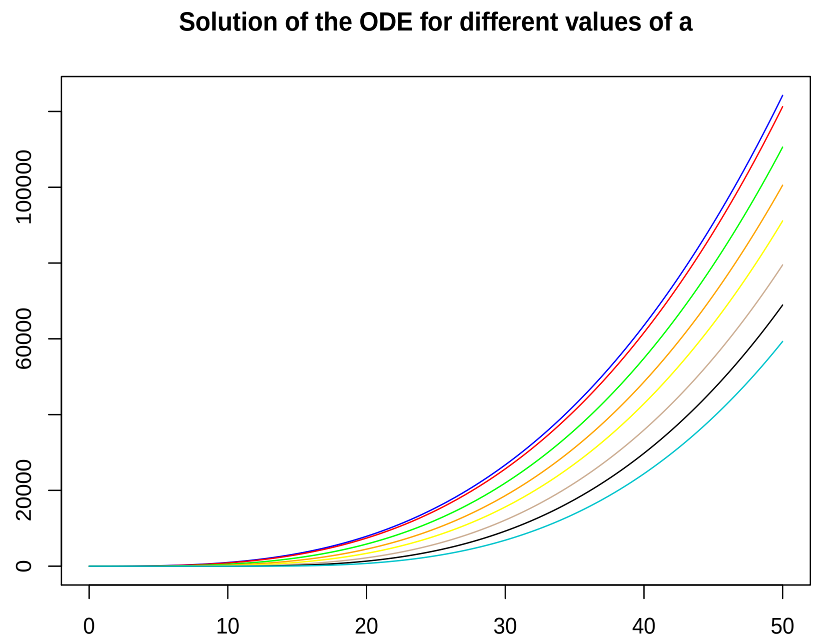
$$\begin{cases} \dfrac{\mathrm{d}X(t)}{\mathrm{d}t} = 3X^{2/3}(t) \\ X(0) = 0. \end{cases}$$

This corresponds to $b(x) = 3x^{2/3}$ *which does not satisfy* (4.4)*. This equation has more than one solution. In fact, for* $a > 0$*, the function*

$$X(t) = \begin{cases} 0 & \text{if } t \leq a, \\ (t-a)^3 & \text{if } t > a, \end{cases}$$

solves this ODE. Hence, (4.4) *guaranties that the SDE has a unique solution.*

Note that is it possible to show existence and uniqueness of strong solution of SDE under weaker conditions on the coefficients. However we do not discuss that in this lecture.



Before proving the theorem, we will recall some results

Lemma 4.1.9 (Gronwall inequality)**.** *Suppose that* $g(t)$ *is a continuous function satisfying*

$$0 \leq g(t) \leq \alpha(t) + \beta \int_0^t g(s)\mathrm{d}s,$$

with $\beta \geq 0$ *and* $\alpha : [0,T] \mapsto \mathbb{R}$ *integrable. Then*

$$g(t) \leq \alpha(t) + \beta \int_0^t \alpha(s)e^{\beta(t-s)}\mathrm{d}s, \, 0 \leq t \leq T.$$

Lemma 4.1.10 (Borel-Cantelli)**.** *Suppose that* $\{A_n, n \geq 1\}$ *is a sequence of events in a probability space. If*

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty, \text{ then } \mathbb{P}(A_{i_0}) = 0.$$

The event

$$A_{i_0} = \{A_n \text{occurs for infinitely many } n\} = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n.$$

Lemma 4.1.11 (Fatou's Lemma). *If $X_n \geq 0$ for all n, then*

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n\right] \leq \liminf_{n\to\infty} \mathbb{E}[X_n].$$

Lemma 4.1.12 (Chebyshev inequality). *Let $X \in L^p(\Omega)$ then*

$$\mathbb{P}(|X| \geq k) \leq \frac{\mathbb{E}[|X|^p]}{k^p}$$

Lemma 4.1.13. *Let $f = (f(t,\omega), 0 \leq t \leq T)$ be a process such that*

(1) $(t,\omega) \mapsto f(t,\omega)$ is $\mathscr{B}([0,T]) \otimes \mathscr{F}$-measurable;

(2) f is $\{\mathscr{F}_t^W\}_{0\leq t\leq T}$-adapted;

(3) $\mathbb{E}\int_0^T |f(s)|^2 \mathrm{d}s < \infty$.

Then

$$\mathbb{E}\left[\sup_{0\leq t\leq T}\left(\int_0^t f(s)\mathrm{d}W(s)\right)^2\right] \leq 4\mathbb{E}\int_0^T |f(s)|^2 \mathrm{d}s.$$

Example 4.1.14. *The mean-reverting Orstein-Uhlenbeck process is the solution $(X(t), 0 \leq t \leq T)$ to the SDE*

$$\mathrm{d}X(t) = (m - X(t))\mathrm{d}t + \sigma \mathrm{d}W(t),$$

where m, σ are real constants, and $(W(t), 0 \leq t \leq T)$ is a Brownian motion

a) Solve this equation

b) Find $\mathbb{E}[X(t)]$ and $Var(X(t))$

Proof of Theorem 4.1.7. **Uniqueness:** Let $X(t)$ and $\bar{X}(t)$ be two strong solution to the SDE (4.1) with the same initial condition, that is

$$\begin{cases} \mathrm{d}X(t) = b(t,X(t))\mathrm{d}t + \sigma(t,X(t))\mathrm{d}W(t) \\ X(0) = \xi \end{cases} \tag{4.6}$$

$$\begin{cases} \mathrm{d}\bar{X}(t) = b(t,\bar{X}(t))\mathrm{d}t + \sigma(t,\bar{X}(t))\mathrm{d}W(t) \\ \bar{X}(0) = \xi \end{cases} \tag{4.7}$$

Writting (4.6) and (4.7) in the integral form and taking the difference, we have

$$X(t) - \bar{X}(t) = \int_0^t (b(s,X(s)) - b(s,\bar{X}(s)))\,\mathrm{d}s + \int_0^t (\sigma(s,X(s)) - \sigma(s,\bar{X}(s)))\,\mathrm{d}W(s).$$

Now use $(a+b)^2 \leq 2(a^2+b^2)$ to get

$$|X(t) - \bar{X}(t)|^2 \leq 2\left[\left|\int_0^t (b(s,X(s)) - b(s,\bar{X}(s)))\,\mathrm{d}s\right|^2 + \left|\int_0^t (\sigma(s,X(s)) - \sigma(s,\bar{X}(s)))\,\mathrm{d}W(s)\right|^2\right].$$

Taking expectation on both sides gives,

$$\mathbb{E}|X(t) - \bar{X}(t)|^2 \leq 2\left[\mathbb{E}\left|\int_0^t (b(s,X(s)) - b(s,\bar{X}(s)))\,\mathrm{d}s\right|^2 + \mathbb{E}\left|\int_0^t (\sigma(s,X(s)) - \sigma(s,\bar{X}(s)))\,\mathrm{d}W(s)\right|^2\right].$$

Using Cauchy-Schwartz inequality for the first term and Itô isometry for the second term (in the right side), we get

$$
\begin{aligned}
&\mathbb{E}|X(t)-\bar{X}(t)|^2\\
&\leq 2\left[t\mathbb{E}\int_0^t\left|b(s,X(s))-b(s,\bar{X}(s))\right|^2\mathrm{d}s+\mathbb{E}\int_0^t\left|\sigma(s,X(s))-\sigma(s,\bar{X}(s))\right|^2\mathrm{d}s\right]\\
&\leq 2\left(tk_1^2\int_0^t\mathbb{E}\left|X(s)-\bar{X}(s)\right|^2\mathrm{d}s+k_1^2\int_0^t\mathbb{E}\left|X(s)-\bar{X}(s)\right|^2\mathrm{d}s\right)\\
&\leq 2k_1^2(1+t)\int_0^t\mathbb{E}\left|X(s)-\bar{X}(s)\right|^2\mathrm{d}s
\end{aligned}
$$

where the second inequality is coming from the Lipschitz condition. Using Gronwall's Lemma (with $\alpha(t)=0$, $g(t)=\mathbb{E}|X(t)-\bar{X}(t)|^2$ and $\beta=2k_1(1+t)$), we have

$$\mathbb{E}\left|X(t)-\bar{X}(t)\right|^2\leq 0$$

that is

$$\mathbb{E}\left|X(t)-\bar{X}(t)\right|^2=0 \text{ for each } t\in[0,T]. \text{ Hence } X(t)=\bar{X}(t) \text{ almost surely for each } t.$$

Let $\{q_1,q_2,\cdots\}$ be a counting of rational numbers in $[0,T]$; for each q_n, there exists Ω_n of full mass (that is $\mathbb{P}(\Omega_n)=1$) such that $X(q_n,\omega)=\bar{X}(q_n,\omega)$ for all $\omega\in\Omega_n$. Choose $\Omega'=\bigcap_{n=1}^{\infty}\Omega_n$ then $\mathbb{P}(\Omega')=1$. For each $\omega\in\Omega'$, $X(q_n,\omega)=\bar{X}(q_n,\omega)$ for all n.

Since $X(t,\omega)-\bar{X}(t,\omega)$ is a continuous process, there exists Ω'' such that $\mathbb{P}(\Omega'')=1$ and $X(t,\omega)-\bar{X}(t,\omega)$ is continuous in $t\in[0,T]$, that is for all rational numbers in $[0,T]$.

Put $\Omega_0=\Omega'\cap\Omega''$ then $\mathbb{P}(\Omega_0)=1$ and for all $\omega\in\Omega_0$ the function $X(t,\omega)-\bar{X}(t,\omega)$ is continuous in t. Hence, for each $\omega\in\Omega_0$, $X(t,\omega)=\bar{X}(t,\omega)\forall t\in[0,T]$ i.e.,

$$\mathbb{P}\left(\left\{\omega,X(t,\omega)=\bar{X}(t,\omega),\,\forall t\in[0,T]\right\}\right)=\mathbb{P}(\Omega_0)=1.$$

Existence: It is similar to the proof of existence of ODE. We use Picard's iteration.

Define the sequence $\{X^{(n)}\}$ of stochastic processes inductively by

$$\begin{cases} X^{(0)}(t)=\xi\\ X^{(n+1)}(t)=\xi+\int_0^t b(s,X^{(n)}(s))\mathrm{d}s+\int_0^t\sigma(s,X^{(n)}(s))\mathrm{d}W(s)\end{cases}\tag{4.8}$$

1) We will first show that the sequence $\{X^{(n)}\}$ is adapted to $\{\mathfrak{F}_t\}_{0\leq t\leq T}$ and $\mathbb{E}\int_0^T|X^{(n)}(s)|^2\mathrm{d}s<\infty$ for all n.

By hypothesis, ξ is $\mathfrak{F}_t$-measurable for all $t>0$ and $\mathbb{E}[|\xi|^2]<\infty$. Thus

$$X^{(1)}(t)=\xi+\int_0^t b(s,\xi)\mathrm{d}s+\int_0^t\sigma(s,\xi)\mathrm{d}W(s)$$

$\mathfrak{F}_t$-measurable for all $t>0$. Thus the process $\{X^{(1)}(t),0\leq t\leq T\}$ is adapted to $\{\mathfrak{F}_t\}_{0\leq t\leq T}$, and

$$|X^{(1)}(t)|^2\leq 3\left(|\xi|^2+\left|\int_0^t b(s,\xi)\mathrm{d}s\right|^2+\left|\int_0^t\sigma(s,\xi)\mathrm{d}W(s)\right|^2\right)$$

Taking expectation and using Cauchy-Schwartz inequality, Itô's isometry and the linear growth condition, we get

$$
\begin{aligned}
\mathbb{E}|X^{(1)}(t)|^2\leq&3\left(\mathbb{E}|\xi|^2+t\mathbb{E}\int_0^t|b(s,\xi)|^2\mathrm{d}s+\mathbb{E}\int_0^t|\sigma(s,\xi)|^2\mathrm{d}s\right)\\
\leq&3\left(\mathbb{E}|\xi|^2+2tk_2^2\int_0^t\mathbb{E}(1+|\xi|^2)\mathrm{d}s+2k_2^2\int_0^t\mathbb{E}(1+|\xi|^2)\mathrm{d}s\right)\\
\leq&3\left(\mathbb{E}|\xi|^2+2k_2^2(1+t)t(1+\mathbb{E}|\xi|^2)\right)\leq C(1+\mathbb{E}|\xi|^2)<\infty
\end{aligned}\tag{4.9}
$$

i.e.

$$\mathbb{E}\int_0^T|X^{(1)}(t)|^2\mathrm{d}t=\int_0^T\mathbb{E}|X^{(1)}(t)|^2\mathrm{d}t<\infty.$$

The constant C in (4.9) depends on T and k_2.

Assume that the process $\{X^{(n)}, 0 \le t \le T\}$ is $\{\mathfrak{F}_t\}_{0\le t\le T}$-adapted, then since b and σ are continuous, then $\{X^{(n+1)}, 0 \le t \le T\}$ is $\{\mathfrak{F}_t\}_{0\le t\le T}$-adapted.

We will show that $\mathbb{E}\int_0^T |X^{(n+1)}(t)|^2 \mathrm{d}t < \infty$. Using (4.8), and Hölder inequality, we have

$$|X^{(n+1)}(t)|^2 \le 3\Big(|\xi|^2 + t\int_0^t |b(s, X^{(n)}(s))|^2 \mathrm{d}s + \Big|\int_0^t \sigma(s, X^{(n)}(s))\mathrm{d}W(s)\Big|^2\Big).$$

Taking the expectation on both sides, using Itô's isometry and the linear growth condition give

$$\begin{aligned}
\mathbb{E}|X^{(n+1)}(t)|^2 &\le 3\Big(\mathbb{E}|\xi|^2 + 2tk_2^2\int_0^t (1+\mathbb{E}|X^{(n)}(s)|^2)\mathrm{d}s + 2k_2^2\int_0^t (1+\mathbb{E}|X^{(n)}(s)|^2)\mathrm{d}s\Big)\\
&\le 3\Big(\mathbb{E}|\xi|^2 + 2k_2^2 T(1+T) + 2k_2^2(T+1)\int_0^t \mathbb{E}|X^{(n)}(s)|^2\mathrm{d}s\Big)\\
&\le C(1+\mathbb{E}|\xi|^2) + C\int_0^t \mathbb{E}|X^{(n)}(s)|^2\mathrm{d}s \qquad (4.10)
\end{aligned}$$

where C is a constant that depends on k_2 and T. Using similar estimate for $\mathbb{E}|X^n(t)|^2$ and substituting into (4.10) gives

$$\begin{aligned}
\mathbb{E}|X^{(n+1)}(t)|^2 &\le C(1+\mathbb{E}|\xi|^2) + C\int_0^t \Big\{C(1+\mathbb{E}|\xi|^2) + C\int_0^s \mathbb{E}|X^{(n-1)}(s_1)|^2\mathrm{d}s_1\Big\}\mathrm{d}s\\
&\le C(1+\mathbb{E}|\xi|^2) + C^2 t(1+\mathbb{E}|\xi|^2) + C^2\int_0^t\int_0^s \mathbb{E}|X^{(n-1)}(s_1)|^2\mathrm{d}s_1\mathrm{d}s
\end{aligned}$$

Hence using this iterative argument we have

$$\mathbb{E}|X^{(n+1)}(t)|^2 \le \Big(c + c^2 t + \frac{c^3t^2}{2!} + \cdots + \frac{c^{(n+2)}t^{n+1}}{(n+1)!}\Big)(1+\mathbb{E}|\xi|^2) \le C(1+\mathbb{E}|\xi|^2)e^{Ct} \qquad (4.11)$$

integrating both side from 0 to T gives

$$\int_0^T \mathbb{E}|X^{(n+1)}(t)|^2\mathrm{d}t < \infty.$$

Hence the sequence $\{X^{(n)}\}_{n=0}^{\infty}$ is a t-continuous adapted sequence such that $\int_0^T \mathbb{E}|X^{(n)}(t)|^2\mathrm{d}t < \infty$.

2) Now we show that the sequence converges uniformly

$$\begin{aligned}
\mathbb{E}|X^{(n+1)}(t) - X^{(n)}(t)|^2 \le 2\Big(&\mathbb{E}\Big|\int_0^t \Big(b(s, X^{(n)}(s)) - b(s, X^{(n-1)}(s))\Big)\mathrm{d}s\Big|^2\\
&+\mathbb{E}\Big|\int_0^t \Big(\sigma(s, X^{(n)}(s)) - \sigma(s, X^{(n-1)}(s))\Big)\mathrm{d}W(s)\Big|^2\Big)
\end{aligned}$$

Using once more Hölder inequality, Itô's isometry and Lipschitz condition, we get

$$\begin{aligned}
&\mathbb{E}|X^{(n+1)}(t) - X^{(n)}(t)|^2\\
\le& 2\Big(k_1^2 t\int_0^t \mathbb{E}|X^{(n)}(s) - X^{(n-1)}(s)|^2\mathrm{d}s + k_1^2\int_0^t \mathbb{E}|X^{(n)}(s) - X^{(n-1)}(s)|^2\mathrm{d}s\Big)\\
\le& 2k_1^2(1+T)\int_0^t \mathbb{E}|X^{(n)}(s) - X^{(n-1)}(s)|^2\mathrm{d}s\\
\le& 2k_1^2(1+T)\int_0^t \Big\{2k_1^2(1+T)\int_0^s \mathbb{E}|X^{(n-1)}(s_1) - X^{(n-2)}(s_1)|^2\mathrm{d}s_1\Big\}\mathrm{d}s\\
&\vdots\\
\le& \underbrace{2^{n+1}k_1^{2(n+1)}(1+T)^{n+1}}_{C^{n+1}}\int_0^t\int_0^s \cdots\Big\{\int_0^{s_{n-1}} \mathbb{E}|X^{(1)}(s_n) - X^{(0)}(s_n)|^2\mathrm{d}s_n\Big\}\cdots\mathrm{d}s_1\mathrm{d}s
\end{aligned}$$

But,

$$\begin{aligned}
\mathbb{E}|X^{(1)}(s_n)-X^{(0)}(s_n)|^2 &= \mathbb{E}|X^{(1)}(s_n)-\xi|^2 \\
&\leq 2\Big(\mathbb{E}\Big|\int_0^{s_n} b(s,\xi)\mathrm{d}s\Big| + \mathbb{E}\Big|\int_0^{s_n} \sigma(s,\xi)\mathrm{d}W(s)\Big|^2\Big) \\
&\leq 2\Big(s_n\mathbb{E}\int_0^{s_n}|b(s,\xi)|^2\mathrm{d}s + \mathbb{E}\int_0^{s_n}|\sigma(s,\xi)|^2\mathrm{d}s\Big) \\
&\leq 2\Big(k_2^2 s_n\int_0^{s_n}(1+\mathbb{E}|\xi|^2)\mathrm{d}s + \int_0^{s_n}(1+\mathbb{E}|\xi|^2)\mathrm{d}s\Big) \\
&\leq 2(T+1)k_2^2\int_0^{s_n}(1+\mathbb{E}|\xi|^2)\mathrm{d}s \leq \underbrace{2(T+1)k_2^2(1+\mathbb{E}|\xi|^2)}_{A}s_n = As_n
\end{aligned}$$

Hence

$$\mathbb{E}|X^{(n+1)}(t)-X^{(n)}(t)|^2 \leq C^{(n+1)}\int_0^t\int_0^s\cdots\int_0^{s_{n-1}} As_n\mathrm{d}s_n\cdots\mathrm{d}s_1\mathrm{d}s \leq K^{n+1}\frac{t^{n+1}}{(n+1)!}$$

On the other hand

$$\begin{aligned}
|X^{(n+1)}(t)-X^{(n)}(t)|^2 \leq 2\Big(t\int_0^t&|b(s,X^{(n)}(s))-b(s,X^{(n-1)}(s))|^2\mathrm{d}s \\
&+\Big|\int_0^t\Big(\sigma(s,X^{(n)}(s))-\sigma(s,X^{(n-1)}(s))\Big)\mathrm{d}W(s)\Big|^2\Big)
\end{aligned}$$

taking the supremum on both sides gives

$$\begin{aligned}
\sup_{0\leq t\leq T}|X^{(n+1)}(t)-X^{(n)}(t)|^2 \leq 2\Big(T\sup_{0\leq t\leq T}\int_0^t&|b(s,X^{(n)}(s))-b(s,X^{(n-1)}(s))|^2\mathrm{d}s \\
&+\sup_{0\leq t\leq T}\Big|\int_0^t\Big(\sigma(s,X^{(n)}(s))-\sigma(s,X^{(n-1)}(s))\Big)\mathrm{d}W(s)\Big|^2\Big)
\end{aligned}$$

Taking the expectation on both sides and using Lemma 4.1.13, we have

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{0\leq t\leq T}|X^{(n+1)}(t)-X^{(n)}(t)|^2\Big] \\
\leq&2\Big(T\mathbb{E}\int_0^T|b(s,X^{(n)}(s))-b(s,X^{(n-1)}(s))|^2\mathrm{d}s + 4\mathbb{E}\int_0^T|\sigma(s,X^{(n)}(s))-\sigma(s,X^{(n-1)}(s))|^2\mathrm{d}s\Big) \\
\leq&2\Big(Tk_1^2\mathbb{E}\int_0^T\mathbb{E}|X^{(n)}(s)-X^{(n-1)}(s)|^2\mathrm{d}s + 4k_1^2\int_0^T\mathbb{E}|X^{(n)}(s)-X^{(n-1)}(s)|^2\mathrm{d}s\Big) \\
\leq&4(T+1)k_1^2\int_0^T\mathbb{E}|X^{(n)}(s)-X^{(n-1)}(s)|^2\mathrm{d}s \\
\leq&4(T+1)k_1^2\int_0^T K^n\frac{s^n}{n!}\mathrm{d}s \leq A_1^{n+1}\frac{T^{n+1}}{(n+1)!}
\end{aligned}$$

Using the Chebyshev inequality (Lemma 4.1.12), we have

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{0\leq t\leq T}|X^{(n+1)}(t)-X^{(n)}(t)| > \frac{1}{2^n}\Big) \\
\leq&\frac{\mathbb{E}\big[\sup_{0\leq t\leq T}|X^{(n+1)}(t)-X^{(n)}(t)|^2\big]}{(2^{-n})^2} \leq \frac{4^nA_1^{n+1}T^{n+1}}{(n+1)!} \leq \frac{A_2^{n+1}T^{n+1}}{(n+1)!}
\end{aligned}$$

It follows from the Borel-cantelli Lemma (Lemma 4.1.10) that

$$\mathbb{P}\left(\sup_{0\leq t\leq T}|X^{(n+1)}(t)-X^{(n)}(t)| > \frac{1}{2^n};\text{i.o.}\right) = 0$$

That is there exists an event $\Omega^+ \in \mathfrak{F}$ such that $\mathbb{P}(\Omega^+) = 1$ and an integer value random variable $N(\omega)$ such that for every $\omega \in \Omega^+$

$$\sup_{0\le t\le T} |X^{(n+1)}(t) - X^{(n)}(t)| \le \frac{1}{2^n} \text{ if } n > N(\omega).$$

It follows that, the partial sum

$$X^{(k)}(t) = \xi + \sum_{n=0}^{k-1}\left(X^{(n+1)}(t) - X^{(n)}(t)\right),$$

converges uniformly in $t \in [0,T]$ with the limit denoted by $X(t)$. Then $X(t)$ is continuous process, $\mathfrak{F}_t$-measurable for each $t \in [0,T]$.

3) Let show that the limit satisfy the integral relation. For almost all ω, we have

$$b(s,X^{(n)}(s)) \to b(s,X(s)) \text{ uniformly in } t \in [0,T]$$
$$\sigma(s,X^{(n)}(s)) \to \sigma(s,X(s)) \text{ uniformly in } t \in [0,T]$$

and hence also

$$\int_0^t b(s,X^{(n)}(s))\mathrm{d}s \to \int_0^t b(s,X(s))\mathrm{d}s \text{ in } t \in L^2(\Omega)(\text{ using Hölder (or Cauchy) inequality})$$
$$\int_0^t \sigma(s,X^{(n)}(s))\mathrm{d}W(s) \to \int_0^t \sigma(s,X(s))\mathrm{d}W(s) \text{ in } L^2(\Omega)(\text{ using Itô isometry})$$

Therefore, taking the limit in the integral representation of $X^{(n+1)}(t)$, we get

$$X(t) = \xi + \int_0^t b(s,X(s))\mathrm{d}s + \int_0^t \sigma(s,X(s))\mathrm{d}W(s),$$

that is $X(t) = (X(t), 0 \le t \le T)$ is a solution to (4.1).

4) Let show that $\int_0^T \mathbb{E}|X(t)|^2\mathrm{d}t < 0$.

We know from (4.11) that

$$\mathbb{E}|X^{(n+1)}(t)|^2 \le C(1+\mathbb{E}|\xi|^2)e^{ct}$$

Taking the limit when $n\uparrow\infty$ and using Fatou's Lemma, we conclude that

$$\mathbb{E}|X(t)|^2 \le C(1+\mathbb{E}|\xi|^2)e^{ct},$$

i.e., $\int_0^T \mathbb{E}|X(t)|^2\mathrm{d}t < 0$. □

4.2 The Martingale Representation Theorem

In this section $W(t) = (W_1(t),\dots,W_m(t))$ is an m-dimensional Brownian motion. We denote by $L^2_\omega([0,T])$ the space of stochastic processes $f = (f(t,\omega), 0 \le t \le T)$ satisfying conditions (1),(2) and (3) of Lemma 4.1.13.

The objective of this section is to prove that any $\mathscr{F}_T^W$-martingale (w.r.t. $\mathbb{P}$) can be represented as an Itô-integral. This result is known as the martingale representation theorem. We will first prove the following

Theorem 4.2.1 (Itô representation theorem)**.** *Let $F \in L^2(\mathscr{F}_T^W, \mathbb{P})$. Then, there exists a unique stochastic process $f(t,\omega) \in L^2_\omega([0,T])$ such that*

$$F = \mathbb{E}[F] + \int_0^T f(t,\omega)\mathrm{d}W(t). \tag{4.12}$$

■

In order to prove Theorem4.2.1, we need the following result

Lemma 4.2.2. *The linear span of random variables of the type*

$$\exp\left\{\int_0^T h(t)\mathrm{d}W(t) - \frac{1}{2}\int_0^T h^2(t)\mathrm{d}t\right\},\ h \in L^2[0,T]\,(\textit{deterministic}) \tag{4.13}$$

is dense in $L^2(\mathscr{F}_T^W, \mathbb{P})$.

Proof of Theorem 4.2.1. We will prove the result for $m = 1$(The prove in the general case is similar). We first show the result for F of the form (4.13) and the result follows by Lemma 4.2.2.

Assume that F has the form (4.13), that is

$$\exp\left\{\int_0^T h(t)\mathrm{d}W(t) - \frac{1}{2}\int_0^T h^2(t)\mathrm{d}t\right\},$$

for some $h \in L^2[0,T]$. Define

$$Y(t,w) := \exp\left\{\int_0^t h(s)\mathrm{d}W(s) - \frac{1}{2}\int_0^t h^2(s)\mathrm{d}s\right\}, 0 \le t \le T.$$

Then, by the Itô's formula (applied to $f(x) = e^x$ and $\mathrm{d}X(t) = -1/2h^2(t)\mathrm{d}t + h(t)\mathrm{d}W(t)$, $X(0) = 0$). We have

$$\begin{aligned}
\mathrm{d}Y(t) = \mathrm{d}f(X(t)) &= \frac{\partial f}{\partial x}(X(t))\mathrm{d}X(t) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(X(t))(\mathrm{d}X(t))^2 \\
&= e^{X(t)}\left(-\frac{1}{2}h^2(t)\mathrm{d}t + h(t)\mathrm{d}W(t)\right) + \frac{1}{2}e^{X(t)}h^2(t)\mathrm{d}t \\
&= e^{X(t)}h(t)\mathrm{d}W(t) = Y(t)h(t)\mathrm{d}W(t).
\end{aligned}$$

Hence,

$$\begin{aligned}
Y(t) &= Y(0) + \int_0^t Y(s)h(s)\mathrm{d}W(s) \\
&= 1 + \int_0^t Y(s)h(s)\mathrm{d}W(s);\ \ Y(0) = e^{X(0)} = 1.
\end{aligned}$$

Therefore, $F = Y(T) = 1 + \int_0^T Y(s)h(s)\mathrm{d}W(s)$ and $\mathbb{E}[F] = 1$.

It follows that (4.12) holds in this case. By linearity, (4.12) also holds for linear combinations of functions of the form (4.13).

So if $F \in L^2(\mathscr{F}_T^W, \mathbb{P})$ is arbitrary, we approximate $F \in L^2(\mathscr{F}_T^W, \mathbb{P})$ by linear combination F_n of the form (4.13). Then for each n, we have

$$F_n = \mathbb{E}[F_n] + \int_0^T f_n(s)\mathrm{d}W(s), \text{ where } f_n \in L^2_\omega([0,T])$$

i.e.

$$\int_0^T f_n(s)\mathrm{d}W(s) = F_n - \mathbb{E}[F_n],$$

We claim that $\{f_n\}_{n\ge1}$ is a Cauchy sequence in $L^2([0,T]\times\Omega)$. Indeed, for $n,m \ge 1$

$$\int_0^T (f_n(s) - f_m(s))\mathrm{d}W(s) = F_n - F_m - \mathbb{E}[F_m - F_n].$$

By Itô isometry

$$\begin{aligned}
\mathbb{E}\Big(\int_0^T (f_n - f_m)^2(s)\mathrm{d}s\Big) =& \mathbb{E}\Big(\int_0^T (f_n(s) - f_m(s))\mathrm{d}W(s)\Big)^2 \\
=& \mathbb{E}\Big[\Big(F_n - F_m - \mathbb{E}[F_m - F_n]\Big)^2\Big] \\
\le& 2\mathbb{E}(F_n - F_m)^2 + 2\mathbb{E}\Big[\Big(\mathbb{E}[F_m - F_n]\Big)^2\Big].
\end{aligned}$$

Thus by Jensen inequality

$$\int_0^T \mathbb{E}(f_n - f_m)^2(s)\mathrm{d}s \le 2\Big(\mathbb{E}(F_n - F_m)^2 + \mathbb{E}(F_m - F_n)^2\Big),$$

that is $\int_0^T \mathbb{E}(f_n - f_m)^2(s)\mathrm{d}s \to 0$ as $n,m \to \infty$. So $\{f_n\}$ is a Cauchy sequence in $L^2([0,T]\times\Omega)$ and hence converges to some $f \in L^2([0,T]\times\Omega)$. Since $f_n \in L^2_\omega([0,T])$, we have $f \in L^2_\omega([0,T])$.

In fact, a subsequence of $\{f_n(t,\omega)\}$ converges to $f(t,\omega)$ for a.a. $(t,\omega) \in [0,T] \times \Omega$. Therefore $f(t,\cdot)$ is $\mathscr{F}_t^W$-measurable. So by modifying $f(t,\omega)$ on a t-set of measure 0 we can obtain that $f(t,\omega)$ is $\mathscr{F}_t^W$-adapted.

Again using the Itô isometry we see that

$$\begin{aligned} F = \lim_{n\to\infty} F_n = \lim_{n\to\infty} \Big(\mathbb{E}[F_n] + \int_0^T f_n(s)\mathrm{d}W(s)\Big) \\ =\mathbb{E}[F] + \int_0^T f(s)\mathrm{d}W(s), \end{aligned}$$

the limit being taken in $L^2(\mathscr{F}_T^W,\mathbb{P})$. Hence the representation (4.12) holds for all $F \in L^2(\mathscr{F}_T^W,\mathbb{P})$.

The uniqueness follows from the Itô isometry. Suppose that there exists $f_1, f_2 \in L^2_\omega([0,T])$ such that (4.12) holds, that is

$$F = \mathbb{E}[F] + \int_0^T f_1(t)\mathrm{d}W(t) = \mathbb{E}[F] + \int_0^T f_2(t)\mathrm{d}W(t),$$

by Itô isometry

$$\begin{aligned} 0 =&\mathbb{E}\Big(\int_0^T (f_1(t) - f_2(t))\mathrm{d}W(t)\Big)^2 \\ =&\int_0^T \mathbb{E}(f_1 - f_2)^2(t)\mathrm{d}t \end{aligned}$$

and therefore $f_1(t) = f_2(t)$ for a.a. $(t,\omega) \in [0,T] \times \Omega$. □

Theorem 4.2.3 (Martingale representation Theorem)**.** *Let $W(t) = (W_1(t),\dots,W_m(t))$ be an m-dimensional Brownian motion. Suppose that $M = \{M(t), t \geq 0\}$ is a $\{\mathscr{F}_t^W\}_{t\geq 0}$-martingale (w.r.t. $\mathbb{P}$) and that $M(t) \in L^2(\mathbb{P})$ for all $t \geq 0$. Then there exists a unique stochastic process $g(s,\omega)$ such that $g \in L^2_\omega([0,T])$ for all $t \geq 0$ and*

$$M(t) = \mathbb{E}[M(0)] + \int_0^t g(s)\mathrm{d}W(s),\ \textit{a.s. for all } t \geq 0 \tag{4.14}$$

■

Proof. For $m = 1$. By Theorem 4.2.1 applied to $F = M(t)$ with $T = t$, we have that for all t there exists a unique $f^{(t)} \in L^2_\omega([0,t])$ such that

$$\begin{aligned} M(t) = \mathbb{E}[M(t)] + \int_0^t f^{(t)}(s,\omega)\mathrm{d}W(s) \\ = \mathbb{E}[M(0)] + \int_0^t f^{(t)}(s,\omega)\mathrm{d}W(s). \end{aligned}$$

Now, assume $0 \leq t_1 < t_2$. Then

$$\begin{aligned} M(t_1) = \mathbb{E}[M(t_2)|\mathscr{F}_{t_1}] =&\mathbb{E}\Big[\mathbb{E}[M(0)] + \int_0^{t_2} f^{(t_2)}(s)\mathrm{d}W(s)|\mathscr{F}_{t_1}\Big] \\ =&\mathbb{E}[M(0)] + \mathbb{E}\Big[\int_0^{t_2} f^{(t_2)}(s)\mathrm{d}W(s)|\mathscr{F}_{t_1}\Big] \\ =&\mathbb{E}[M(0)] + \int_0^{t_1} f^{(t_2)}(s)\mathrm{d}W(s). \end{aligned} \tag{4.15}$$

where the martingale property for stochastic integral was used to obtain the last equality. Note that by the representation of $M(t_2)$, $f^{(t_2)}$ is in $L^2_\omega([0,t_2])$.

On the other hand, it follows from Theorem 4.2.1 that there exists $f^{(t_1)} \in L^2_\omega([0,t_1])$ such that

$$M(t_1) = \mathbb{E}[M(0)] + \int_0^{t_1} f_1(s)\mathrm{d}W_s. \tag{4.16}$$

Comparing (4.15) and (4.16) and using Ttô isometry

$$0 = \mathbb{E}\Big(\int_0^{t_1} (f^{(t_2)} - f^{(t_1)})(s)\mathrm{d}W(s)\Big)^2 = \int_0^{t_1} \mathbb{E}(f^{(t_2)} - f^{(t_1)})^2(s)\mathrm{d}s,$$

and therefore, $f^{(t_2)}(s) = f^{(t_1)}(s)$ for a.a. $(s,\omega) \in [0,t_1] \times \Omega$. So we can define $f(s,\omega)$ for a.a. $[0,\infty[\times\Omega$ by setting $f(s,\omega) = f^{(N}(s,\omega)$ if $s \in [0,N]$ and then we get

$$\begin{aligned}M(t) &= \mathbb{E}[M(0)] + \int_0^t f^{(t)}(s,\omega)\mathrm{d}W(s)\\ &= \mathbb{E}[M(0)] + \int_0^t f(s)\mathrm{d}W(s), \text{ for all } t \geq 0.\end{aligned}$$

□

Example 4.2.4. *In each of the cases below, find the process $f(t) \in L^2_\omega([0,T])$ such that $F = \mathbb{E}[F] + \int_0^T f(t)\mathrm{d}W(t)$.*

a) $F = W(T)$, $\mathbb{E}[F] = \mathbb{E}[W(T)] = 0$, $W(T) = \int_0^T 1\mathrm{d}W(s)$ *i.e.* $f = 1$.

b) $F = \int_0^T W(t)\mathrm{d}t$. *We have* $\mathbb{E}[F] = 0$. *Put* $X(t) = t$ *and* $Y(t) = W(t)$ *then by Itô product rule*

$$tW(t) = \int_0^t W(s)\mathrm{d}s + \int_0^t s\mathrm{d}W(s).$$

That is

$$TW(T) = \int_0^T W(t)\mathrm{d}t + \int_0^T t\mathrm{d}W(t),$$

i.e.

$$\begin{aligned}\int_0^T W(t)\mathrm{d}t &= TW(T) - \int_0^T t\mathrm{d}W(t)\\ &= T\int_0^T \mathrm{d}W(t) - \int_0^T t\mathrm{d}W(t)\\ &= \int_0^T (T-t)\mathrm{d}W(t),\end{aligned}$$

i.e. $F = \int_0^T (T-t)\mathrm{d}W(t)$; $f(t) = T - t$.

c) $F = W^2(T)$, $\mathbb{E}[F] = \mathbb{E}[W^2(T)] = T$. *By Itô's formula*

$$\begin{aligned}F = W^2(T) &= W^2(0) + \int_0^T 2W(t)\mathrm{d}W(t) + \frac{1}{2}\int_0^T 2\mathrm{d}t\\ &= T + \int_0^T 2W(t)\mathrm{d}W(t)\\ &= \mathbb{E}[F] + \int_0^T 2W(t)\mathrm{d}W(t),\end{aligned}$$

$f(t) = 2W(t)$.

4.3 The Girsanov theorem

This section examines the Girsanov theorem which is a very important result in stochastic analysis.

We first start with the following Lévy characterisation of Brownian motion.

Theorem 4.3.1 (Lévy characterisation of Brownian motion)**.** *Let $X(t) = (X_1(t),\dots,X_n(t))$ be a continuous stochastic process on a probability space $(\Omega,\mathscr{F},\mathbb{P})$ in $\mathbb{R}^n$. Then the following assertions are equivalent:*

(a) $X(t)$ is a Brownian motion w.r.t $\mathbb{P}$. That is, the law of $X(t)$ w.r.t $\mathbb{P}$ is the same as the law of an n-dimensional Brownian motion.

(b) (i) $X(t) = (X_1(t),\dots,X_n(t))$ is a martingale w.r.t $\mathbb{P}$ (and w.r.t its own filtration) and

(ii) $X_i(t)X_j(t) - \delta_{ij}t$ is a martingale w.r.t $\mathbb{P}$ (and w.r.t its own filtration) for all $i,j \in \{1,2,\dots,n\}$.

■

We will need the following result.

Lemma 4.3.2. *Let μ and ν be two probability measures on a measurable space $(\Omega,\mathscr{F})$ such that $\mathrm{d}\nu(\omega) = f(\omega)\mathrm{d}\mu(\omega)$ for some $f \in L^1(\mu)$. Let X be a random variable on $(\Omega,\mathscr{F})$ such that*

$$\mathbb{E}_\nu[|X|] = \int_\Omega |X(\omega)|f(\omega)\mathrm{d}\mu(\omega) < \infty.$$

Let $\mathscr{H}$ be a σ-algebra such that $\mathscr{H} \subset \mathscr{F}$. Then

$$\mathbb{E}_\nu[X|\mathscr{H}]\mathbb{E}_\mu[f|\mathscr{H}] = \mathbb{E}_\mu[Xf|\mathscr{H}]. \tag{4.17}$$

Proof. By the definition of the conditional expectation, we get that for $H \in \mathscr{H}$,

$$\int_H \mathbb{E}_\nu[X|\mathscr{H}]f\mathrm{d}\mu = \int_H \mathbb{E}_\nu[X|\mathscr{H}]\mathrm{d}\nu = \int_H X\mathrm{d}\nu = \int_H Xf\mathrm{d}\mu.$$

Thus

$$\int_H \mathbb{E}_\nu[X|\mathscr{H}]f\mathrm{d}\mu = \int_H \mathbb{E}_\mu[Xf|\mathscr{H}]\mathrm{d}\mu \tag{4.18}$$

Now, using the Tower property, we have

$$\begin{aligned}\int_H \mathbb{E}_\nu[X|\mathscr{H}]f\mathrm{d}\mu =& \mathbb{E}_\mu[\mathbb{E}_\nu[X|\mathscr{H}]fI_H] = \mathbb{E}_\mu[\mathbb{E}_\mu[\mathbb{E}_\nu[X|\mathscr{H}]fI_H|\mathscr{H}]]\\ =& \mathbb{E}_\mu[\mathbb{E}_\nu[X|\mathscr{H}]\mathbb{E}_\mu[fI_H|\mathscr{H}]] = \mathbb{E}_\mu[I_H\mathbb{E}_\nu[X|\mathscr{H}]\mathbb{E}_\mu[f|\mathscr{H}]]\end{aligned}$$

Thus

$$\int_H \mathbb{E}_\nu[X|\mathscr{H}]f\mathrm{d}\mu = \int_H \mathbb{E}_\nu[X|\mathscr{H}]\mathbb{E}_\mu[f|\mathscr{H}]\mathrm{d}\mu. \tag{4.19}$$

Putting (4.18) and (4.19) together yield

$$\int_H \mathbb{E}_\nu[X|\mathscr{H}]\mathbb{E}_\mu[f|\mathscr{H}]\mathrm{d}\mu = \int_H \mathbb{E}_\mu[Xf|\mathscr{H}]\mathrm{d}\mu.$$

Since this is true for all $H \in \mathscr{H}$, the result follows. □

Theorem 4.3.3 (The Girsanov theorem I)**.** *Let $Y = \{Y(t), t \geq 0\} \in \mathbb{R}^n$ be an Itô process of the form*

$$\mathrm{d}Y(t) = a(t,\omega)\mathrm{d}t + \mathrm{d}W(t);\ 0 \leq t \leq T, Y(0) = 0, \tag{4.20}$$

where $T < \infty$ is a given constant and $W = \{W(t), t \geq 0\}$ is an n-dimensional Brownian motion. Set

$$M(t) = \exp\Big(-\int_0^t a(s,\omega)\mathrm{d}W(s) - \frac{1}{2}\int_0^t |a(s,\omega)|^2\mathrm{d}s\Big);\ t \leq T. \tag{4.21}$$

Assume that $a(s,\omega)$ satisfies the Novikov's condition

$$\mathbb{E}_{\mathbb{P}}\big[\exp\big\{\frac{1}{2}\int_0^T |a(s,\omega)|^2\mathrm{d}s\big\}\big] < \infty, \tag{4.22}$$

where $\mathbb{E}_{\mathbb{P}}$ is the expectation w.r.t $\mathbb{P}$. Define the measure $\mathbb{Q}$ on $(\Omega,\mathscr{F}_T^W)$ by

$$\mathrm{d}\mathbb{Q}(\omega) = M(T,\omega)\mathrm{d}\mathbb{P}(\omega). \tag{4.23}$$

Then $Y = \{Y(t), t \geq 0\}$ is an n-dimensional Brownian motion w.r.t the probability law $\mathbb{Q}$, for $t \leq T$.
■

Remark 4.3.4.

(1) The Novikov's condition (4.22) *guarantee that the process M defined by* (4.21) *is a martingale. The theorem holds if we assume only that $M = \{M(t), t \geq 0\}$ is a martingale w.r.t $(P, \{\mathscr{F}_t^W\}_{t\geq 0})$.*

(2) The transformation $\mathbb{P} \mapsto \mathbb{Q}$ given by (4.23) *is called the Girsanov transformation of measures.*

Proof. We will assume without loss of generality that $a(s,\omega)$ is bounded. Using the Lévy characterisation Theorem 4.3.1, it is enough to show that

(i) $Y(t) = (Y_1(t), \ldots, Y_n(t))$ is a martingale w.r.t $\mathbb{Q}$ and

(ii) $Y_i(t)Y_j(t) - \delta_{ij}t$ is a martingale w.r.t $\mathbb{Q}$ for all $i, j \in \{1, 2, \ldots, n\}$.

Let first show (i). By its definition Y is adapted and integrable. Next we wish to compute $\mathbb{E}_{\mathbb{Q}}[Y_i(t)|\mathscr{F}_s]$. Put $K(t) = M(t)Y(t)$, then from Itô's formula, we have

$$\begin{aligned}
\mathrm{d}K_i(t) =& M(t)\mathrm{d}Y_i(t) + Y_i(t)\mathrm{d}M(t) + \mathrm{d}Y_i(t)\mathrm{d}M(t)\\
=& M(t)(a_i(t)\mathrm{d}t + \mathrm{d}W_i(t)) + Y_i(t)M(t)\sum_{k=1}^{n}(-a_k)\mathrm{d}W_k(t) + (\mathrm{d}W_i(t))\big(-M(t)\sum_{k=1}^{n}a_k\mathrm{d}W_k(t)\big)\\
=& M(t)\Big(\mathrm{d}W_i(t) - Y_i(t)\sum_{k=1}^{n}a_k\mathrm{d}W_k(t)\Big)\\
=& M(t)\gamma^{(i)}(t)\mathrm{d}W(t),
\end{aligned}$$

where $\gamma^{(i)}(t) = (\gamma_1^{(i)}(t), \ldots, \gamma_n^{(i)}(t))$, with $\gamma_j^{(i)}(t) = \begin{cases} -Y_i(t)a_j(t), & \text{for } j \neq i \\ 1 - Y_i(t)a_i(t), & \text{for } j = i. \end{cases}$

Hence $K_i(t)$ is a martingale w.r.t P, that is $K_i(t)$ is integrable.
In addition

$$\mathbb{E}_Q[|Y_i(t)|] = \int_\Omega |Y_i(t)|M(t)\mathrm{d}\mathbb{P}(\omega) = \int_\Omega |K_i(t)|\mathrm{d}\mathbb{P}(\omega) = \mathbb{E}_{\mathbb{P}}[|K_i(t)|] < \infty.$$

Using Lemma 4.3.2, we have

$$\mathbb{E}_{\mathbb{Q}}[Y_i(t)|\mathscr{F}_s]\mathbb{E}_{\mathbb{P}}[M(t)|\mathscr{F}_s] = \mathbb{E}_P[Y_i(t)M(t)|\mathscr{F}_s].$$

that is

$$\mathbb{E}_{\mathbb{Q}}[Y_i(t)|\mathscr{F}_s] = \frac{\mathbb{E}_{\mathbb{P}}[K_i(t)|\mathscr{F}_s]}{\mathbb{E}_{\mathbb{P}}[M(t)|\mathscr{F}_s]} = \frac{K(s)}{M(s)} = Y_i(s)$$

which shows that $Y_i(s)$ is a martingale w.r.t $\mathbb{Q}$.
Hence (i) is proved.

The proof of (ii) follows in a similar way.
Put $Z_{ij}(t) = Y_i(t)Y_j(t) - \delta_{ij}t$. Then by Itô's formula

$$\begin{aligned}
\mathrm{d}Z_{ij}(t) =& \mathrm{d}Y_i(t)Y_j(t) - \delta_{ij}\mathrm{d}t\\
=& Y_i(t)(a_j(t)\mathrm{d}t + \mathrm{d}W_j(t)) + Y_j(t)(a_i(t)\mathrm{d}t + \mathrm{d}W_i(t))\\
&+ (a_j(t)\mathrm{d}t + \mathrm{d}W_j(t))(a_i(t)\mathrm{d}t + \mathrm{d}W_i(t)) - \delta_{ij}\mathrm{d}t\\
=& (Y_i(t)a_j(t) + a_i(t)Y_j(t))\mathrm{d}t + Y_i(t)\mathrm{d}W_j(t) + Y_j(t)\mathrm{d}W_i(t).
\end{aligned}$$

One can show that Z_{ij} is a adapted and integrable.

Now put $K_{ij}(t) = M(t)Z_{ij}(t)$. Then Itô's formula yields

$$\begin{aligned}
\mathrm{d}K_{ij}(t) =& M(t)\mathrm{d}Z_{ij}(t) + Z_{ij}(t)\mathrm{d}M(t) + \mathrm{d}Z_{ij}(t)\mathrm{d}M(t)\\
=& M(t)\{(a_i(t)Y_j(t) + a_j(t)Y_i(t))\mathrm{d}t + Y_i(t)\mathrm{d}W_j(t) + Y_j(t)\mathrm{d}W_i(t)\}\\
&+ Z_{ij}(t)M(t)\sum_{k=1}^{n}(-a_k)\mathrm{d}W_k(t)\\
&+ (Y_i(t)\mathrm{d}W_j(t) + Y_j(t)\mathrm{d}W_i(t))\big(M(t)\sum_{k=1}^{n}(-a_k)\mathrm{d}W_k(t)\big)\\
=& M(t)\big(Y_i(t)\mathrm{d}W_j(t) + Y_j(t)\mathrm{d}W_i(t) - Z_{ij}(t)\sum_{k=1}^{n}a_k\mathrm{d}W_k(t)\big)\\
=& M(t)\gamma^{(i,j)}(t)\mathrm{d}W_t,
\end{aligned}$$

where $\gamma^{(i,j)}(t) = (\gamma_1^{(i,j)}(t), \dots, \gamma_n^{(i,j)}(t))$ with $\gamma_m^{(i,j)}(t) = \begin{cases} Y_i(t) - Z_{ij}(t)a_j, \; m = j \\ Y_j(t) - Z_{ij}(t)a_i, \; m = i \\ -Z_{ij}(t)a_m, \; m \neq \{i,j\}. \end{cases}$
Hence $K_{ij}(t)$ is a martingale w.r.t $\mathbb{P}$ and thus the result follows as in the proof of (i). □

Remark 4.3.5. *The Girsanov Theorem I 4.3.3 means that for all Borel sets $F_1, \dots, F_k \subset \mathbb{R}^n$ and $t_1, \dots, t_k \leq T$, $k = 1, 2, \dots$ We have*

$$\mathbb{Q}(Y(t_1) \in \mathscr{F}_1, \dots, Y(t_k) \in \mathscr{F}_k) = \mathbb{P}(W(t_1) \in \mathscr{F}_1, \dots, W(t_k) \in \mathscr{F}_k).$$

Theorem 4.3.6 (The Girsanov Theorem II)**.** *Let $Y = \{Y(t), t \geq 0\} \in \mathbb{R}^n$ be an Itô process of the form*

$$\mathrm{d}Y(t) = \beta(t,\omega)\mathrm{d}t + \theta(t,\omega)\mathrm{d}W(t); \; t \leq T \tag{4.24}$$

where $W(t) \in \mathbb{R}^n$, $\beta(t,\omega) \in \mathbb{R}^n$ and $\theta(t,\omega) \in \mathbb{R}^{n\times m}$.
Suppose that there exist processes $u(t,\omega) \in \mathscr{W}_{\mathscr{H}}^m$ and $\alpha(t,\omega) \in \mathscr{W}_{\mathscr{H}}^n$ such that

$$\theta(t,\omega)u(t,\omega) = \beta(t,\omega) - \alpha(t,\omega) \tag{4.25}$$

and assume that $u(t,\omega)$ satisfies the Novikov's condition

$$\mathbb{E}\Big[\exp\Big(\frac{1}{2}\int_0^T |u(t,\omega)|^2 \mathrm{d}t\Big)\Big] < \infty. \tag{4.26}$$

Put

$$M(t) = \exp\big(-\int_0^t u(s,\omega)\mathrm{d}W(s) - \frac{1}{2}\int_0^t |u(s,\omega|^2 \mathrm{d}s\big); \; t \leq T \tag{4.27}$$

and

$$\mathrm{d}Q(\omega) = M(T,\omega)\mathrm{d}\mathbb{P}(\omega) \;\; on \; \mathscr{F}_T^W.$$

Then

$$\widehat{W}(t) = \int_0^t u(s,\omega)\mathrm{d}s + W(t); \; t \leq T \tag{4.28}$$

is a Brownian motion w.r.t Q and in terms of $\hat{W}(t)$ the process $Y(t)$ has the following stochastic representation

$$\mathrm{d}Y(t) = \alpha(t,\omega)\mathrm{d}t + \theta(t,\omega)\mathrm{d}\widehat{W}(t,\omega). \tag{4.29}$$

■

We will first recall the definition of the space $\mathscr{W}_{\mathscr{H}}$ in the theorem.

Definition 4.3.7. *$\mathscr{W}_{\mathscr{H}} = \bigcap_{t\geq 0} \mathscr{W}_{\mathscr{H}}(0,T)$, where $\mathscr{W}_{\mathscr{H}}(0,T)$ is the space of stochastic processes f satisfying*

(i) $(t,\omega) \mapsto f(t,\omega)$ *is $\mathscr{B} \times \mathscr{F}$-measurable, where $\mathscr{B}$ is the Borel σ-algebra on $[0,\infty)$.*

(ii) There exists an increasing family of σ-algebras $\mathscr{H}_t$; $t \geq 0$ such that

(a) $W(t)$ is a martingale w.r.t $\mathscr{H}_t$ and

(b) $f(t,\omega)$ is $\mathscr{H}_t$-adapted.

(iii) $\mathbb{P}\Big(\int_0^T f^2(t,\omega)\mathrm{d}t < \infty\Big) = 1.$

Proof. We know from Theorem 4.3.3 that $\widehat{W}(t)$ is a Brownian motion w.r.t $\mathbb{Q}$. Hence substituting (4.28) in (4.24), we get by (4.25)

$$\begin{aligned} \mathrm{d}Y(t) &= \beta(t,\omega)\mathrm{d}t + \theta(t,\omega)(\mathrm{d}\widehat{W}(t) - u(t,\omega)\mathrm{d}t) \\ &= (\beta(t,\omega) - \theta(t,\omega)u(t,\omega))\mathrm{d}t + \theta(t,\omega)\mathrm{d}\widehat{W}(t) \\ &= \alpha(t,\omega)\mathrm{d}t + \theta(t,\omega)\mathrm{d}\widehat{W}(t). \end{aligned}$$

□

Theorem 4.3.8 (The Girsanov Theorem III)**.** *Let $X(t) = X^x(t) \in \mathbb{R}^n$ and $Y(t) = Y^x(t) \in \mathbb{R}^n$ be an Itô diffusion and and Itô process, respectively, of the forms*

$$\begin{cases} \mathrm{d}X(t) & = b(X(t))\mathrm{d}t + \sigma(X(t))\mathrm{d}W(t);\ t \leq T, X(0) = x \\ \mathrm{d}Y(t) & = (\gamma(t,\omega) + b(Y(t)))\mathrm{d}t + \sigma(Y(t))\mathrm{d}W(t);\ t \leq T, Y(0) = x \end{cases} \tag{4.30}$$

where the functions $b : \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ satisfy the condition of existence and uniqueness of solution of the SDEs and $\gamma(t,\omega) \in \mathscr{W}^n_{\mathscr{H}}$, $x \in \mathbb{R}^n$. Suppose that there exists a process $u(t,\omega) \in \mathscr{W}^n_{\mathscr{H}}$ such that

$$\sigma(t,\omega)u(t,\omega) = \gamma(t,\omega) \tag{4.31}$$

and assume that $u(t,\omega)$ satisfies the Novikov's condition

$$\mathbb{E}\big[\exp\big(\frac{1}{2}\int_0^T |u(t,\omega)|^2\mathrm{d}t\big)\big] < \infty. \tag{4.32}$$

Put

$$M(t) = \exp\Big(-\int_0^t u(s,\omega)\mathrm{d}W(s) - \frac{1}{2}\int_0^t |u(s,\omega|^2\mathrm{d}s\Big);\ t \leq T \tag{4.33}$$

and

$$\mathrm{d}\mathbb{Q}(\omega) = M(T,\omega)\mathrm{d}P(\omega) \ \ on\ \mathscr{F}^B_T. \tag{4.34}$$

Then

$$\widehat{W}(t) = \int_0^t u(s,\omega)\mathrm{d}s + W(t);\ t \leq T \tag{4.35}$$

is a Brownian motion w.r.t Q. Moreover,

$$\mathrm{d}Y(t) = b(Y(t))\mathrm{d}t + \sigma(Y(t))\mathrm{d}\widehat{W}(t,\omega). \tag{4.36}$$

Therefore the $\mathbb{Q}$-law of $Y^x(t)$ is the same as the $\mathbb{P}$-law of $X^x(t)$, $t \leq T$.
■

Proof. Set $\theta(t,\omega) = \sigma(t,\omega)$, $\beta(t,\omega) = \gamma(t,\omega) + b(Y(t))$ and $\alpha(t,\omega) = b(Y(t))$ in Theorem 4.3.6. Then the representation holds.
In addition, since strong (or weak) solutions of SDEs are weakly unique. It follows that the $\mathbb{Q}$-law of $Y^x(t)$ is the same as the $\mathbb{P}$-law of $X^x(t)$; $t \leq T$. □

Example 4.3.9. *Let $Y(t)$ be a weak (strong) solution to the SDE*

$$\mathrm{d}Y(t) = b(Y(t))\mathrm{d}t + \sigma(Y(t))\mathrm{d}W(t) \tag{4.37}$$

where $b : \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^n \to \mathbb{R}^{n\times m}$, $W(t) \in \mathbb{R}^n$.
Suppose that σ is invertible and find conditions under which the SDE given by

$$\mathrm{d}X(t) = a(X(t))\mathrm{d}t + \sigma(X(t))\mathrm{d}W(t) \tag{4.38}$$

with $a : \mathbb{R}^n \to \mathbb{R}^n$, has a weak solution.

Write $\mathrm{d}Y(t) = (b(Y(t)) - a(Y(t)) + a(Y(t)))\mathrm{d}t + \sigma(Y(t))\mathrm{d}W(t)$.
Suppose that we can find $u_0 : \mathbb{R}^n \to \mathbb{R}^n$ such that $\sigma(y)u_0(y) = b(y) - a(y);\ y \in \mathbb{R}^n$, that is, $u_0 = \sigma^{-1}(b-a)$.
If $u(t,\omega) = U_0(Y(t,\omega))$ satisfies the Novikov's condition, one defines

$$M(t) = \exp\big(-\int_0^t u(s,\omega)\mathrm{d}W(s) - \frac{1}{2}\int_0^t u^2(s,\omega)\mathrm{d}s\big);\ t \leq T,$$

$$\mathrm{d}\mathbb{Q}(\omega) = M(T,\omega)\mathrm{d}\mathbb{P}(\omega) \ \ \text{on}\ \mathscr{F}^B_T$$

and

$$\widehat{W}(t) = \int_0^t u(s,\omega)\mathrm{d}s + W(t);\ t \leq T\ (\mathbb{Q}\text{-B.M}).$$

Then

$$\mathrm{d}Y(t) = a(Y(t))\mathrm{d}t + \sigma(Y(t))\mathrm{d}\widehat{W}(t). \tag{4.39}$$

Thus we found a $\mathbb{Q}$-Brownian motion $\hat{W}(t)$ such that (4.39) is valid. Hence $(Y(t), \widehat{W}(t))$ is a weak solution to (4.38).

Example 4.3.10. *Construct a weak solution to the following SDE*

$$\mathrm{d}X(t) = sign(X(t))\mathrm{d}t + \mathrm{d}W(t);\ X(0) = x \in \mathbb{R}. \tag{4.40}$$

Comparing this with the preceding theorem, we set $\sigma = 1,\ b = \mathrm{sign}(x)$ and $\mathrm{d}Y(t) = \mathrm{d}W(t)$.
That is, $\mathrm{d}Y(t) = (-\mathrm{sign}(Y(t)) + \mathrm{sign}(Y(t))\mathrm{d}t + \mathrm{d}W(t)$ and $\gamma(t,\omega) = -\mathrm{sign}(Y(t))$ Set $u_0 = \sigma^{-1}\gamma = -\mathrm{sign}(Y(t))$.

$$\begin{aligned}
\mathbb{E}\big[\exp\big(\frac{1}{2}\int_0^T u_0^2(t,\omega)\mathrm{d}t\big)\big] &= \mathbb{E}\big[\exp\big(\frac{1}{2}\int_0^T (\mathrm{sign}(X(t)))^2\mathrm{d}t\big)\big] \\
&= \mathbb{E}\big[\exp\big(\frac{1}{2}T\big)\big] < \infty \quad ((\mathrm{sign}(X(t)))^2 = 1),
\end{aligned}$$

that is the Novikov's condition is satisfied. Define

$$\begin{aligned}
M(t) &= \exp\big(-\int_0^T u_0(s,\omega)\mathrm{d}W(s) - \frac{1}{2}\int_0^T u_0^2(s,\omega)\mathrm{d}s\big) \\
&= \exp\big(\int_0^T \mathrm{sign}(Y(s))\mathrm{d}W(s) - \frac{1}{2}\int_0^T 1\mathrm{d}s\big) \\
&= \exp\big(\int_0^T \mathrm{sign}(W(s))\mathrm{d}W(s) - \frac{1}{2}t\big).
\end{aligned}$$

Let $T < \infty$ and put $\mathrm{d}\mathbb{Q}(\omega) = M(T,\omega)\mathrm{d}P(\omega)$ on $\mathscr{F}_T^W$.
Then

$$\hat{W}(t) := -\int_0^t \mathrm{sign}(W(s))\mathrm{d}s + W(t)$$

is a $\mathbb{Q}$-Brownian motion for $t \leq T$ and

$$\begin{aligned}
\mathrm{d}W(t) = \mathrm{d}Y(t) &= \mathrm{sign}(W(s))\mathrm{d}s + \mathrm{d}\widehat{W}(t) \\
&= \mathrm{sign}(Y(s))\mathrm{d}s + \mathrm{d}\widehat{W}(t).
\end{aligned}$$

Setting $Y(0) = x$ yields to the conclusion that $(Y(t), \widehat{W}(t))$ is a weak solution to (4.40).

4.4 Diffusions: Basic Properties

One may think of a solution to a stochastic differential equation as the mathematical description of the motion of a small particle in a moving fluid: such processes are called (Itô) diffusion.

Throughout this section, we assume that the process is a time-homogeneous diffusion.

Definition 4.4.1. *A (time-homogeneous) Itô diffusion is a stochastic process $X_t(\omega) = X(t,\omega) : [0,T] \times \Omega \mapsto \mathbb{R}^d$ satisfying a stochastic differential equation of the form*

$$\mathrm{d}X(t) = b(X(t))\mathrm{d}t + \sigma(X(t))\mathrm{d}W(t),\ t \geq s; X(s) = x. \tag{4.41}$$

Where $W = \{W(t); t \geq 0\}$ is a Brownian motion and $b : \mathbb{R}^d \mapsto \mathbb{R}^d$ $\sigma : \mathbb{R}^d \mapsto \mathbb{R}^{d\times m}$ satisfy the condition

$$|b(x) - b(y)| + |\sigma(x) - \sigma(y)| \leq D|x-y|,\ x,y \in \mathbb{R}^d.$$

Remark 4.4.2. *Note that, when the coefficients are time-homogeneous then the Lipschitz continuity is enough to guarantee existence and uniqueness of strong solution.*

We denote the unique solution of (4.12) by $X(t) = X_t = X_t^{s,x}; t \geq s$. If $s = 0$, we write X_t^x for $X_t^{0,x}$.

The notion of time-homogeneous is understood in the following sense: Note that

$$X_{s+h}^{s,x} = x + \int_s^{s+h} b(X_u^{s,x})\mathrm{d}u + \int_s^{s+h} \sigma(X_u^{s,x})\mathrm{d}W_u.$$

Put $u = s + v$ and $\tilde{W}_v = W_{s+v} - W_s$, $v \geq 0$. Then

$$X_{s+h}^{s,x} = x + \int_0^h b(X_{s+v}^{s,x})\mathrm{d}v + \int_0^h \sigma(X_{s+v}^{s,x})\mathrm{d}\tilde{W}_v.$$

On the other hand

$$X_h^{0,x} = x + \int_0^h b(X_v^{0,x})\mathrm{d}v + \int_0^h \sigma(X_v^{0,x})\mathrm{d}W_v,$$

since $\{\tilde{W}_v\}_{v\geq 0}$ and $\{W_v\}_{v\geq 0}$ have the same $\mathbb{P}^0$-distribution. It follows by weak uniqueness of the solution of SDE

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t;\ X_0 = 0,$$

that $\{X_{s+h}^{s,x}\}_{h\geq 0}$ and $\{X_h^{0,x}\}_{h\geq 0}$ have the same $\mathbb{P}^0$-distribution i.e. is $\{X_t\}_{t\geq 0}$ time-homogeneous.

The probability Law $\mathbb{Q}^x$ of $\{X_t\}_{t\geq 0}$, for $x \in \mathbb{R}^d$ gives the distribution of $\{X_t\}_{t\geq 0}$, assuming that $X_0 = x$.

As in the previous section, $\{\mathscr{F}_t^W, t \geq 0\}$ is the σ-algebra generated by $\{W_r; r \leq t\}$. We denote by $\mathscr{M}_t$ the σ-algebra generated by $\{X_r; r \leq t\}$. We know that X_t is measurable w.r.t. $\mathscr{F}_t^W$, so $\mathscr{M}_t \subseteq \mathscr{F}_t^W$.

The following results and properties give the future behaviour of the process X_t based on its present behaviour.

4.4.1 The Markov and Strong Markov Properties

The Markov property says that the future behaviour of the process given what happened up to time t is the same as the behaviour obtained when starting the process at X_t.

Theorem 4.4.3 (Markov property)**.** *Let f be a bounded Borel function from $\mathbb{R}^d$ to $\mathbb{R}$. Then for $t, h \geq 0$*

$$\mathbb{E}^x[f(X_{t+h})|\mathscr{F}_t^W]_\omega = \mathbb{E}^{X_t(\omega)}[f(X_h)] \tag{4.42}$$

■

Here and in the following $\mathbb{E}^x$ denotes the expectation w.r.t. the probability measure $\mathbb{Q}^x$. Hence $\mathbb{E}^y[f(X_h)]$ means $\mathbb{E}[f(X_h^y)]$, where $\mathbb{E}$ denotes the expectation w.r.t. the measure $\mathbb{P}^0$. The right hand side of (4.42) means the function $\mathbb{E}^y[f(X_h)]$ evaluated at $y = X_t(\omega)$.

Remark 4.4.4. *Theorem4.4.3 states that X_t is a Markov process w.r.t. the family of σ-algebras $\{\mathscr{F}_t^W\}_{t\geq 0}$. Since $\mathscr{M}_t \subseteq \mathscr{F}_t^W$, this implies that X_t is also a Markov process w.r.t. the σ-algebras $\{\mathscr{M}_t\}_{t\geq 0}$. In fact,*

$$\begin{aligned}\mathbb{E}^x[f(X_{t+h})|\mathscr{M}_t] =& \mathbb{E}^x\left[\mathbb{E}^x[f(X_{t+h})|\mathscr{F}_t^W]|\mathscr{M}_t\right]\\ =& \mathbb{E}^x\left[\mathbb{E}^{X_t}[f(X_h)]|\mathscr{M}_t\right]\\ =& \mathbb{E}^{X_t}[f(X_h)],\end{aligned}$$

where the last inequality follows since $\mathbb{E}^{X_t}[f(X_h)]$ is $\mathscr{M}_t$-measurable.

The following property known as the strong Markov property states that a relation of the form (4.42) continues to hold if the time is replaced by a random time $\tau)\omega$ of a more general type called stopping time.

Definition 4.4.5. *Let $\{\mathscr{N}_t\}_{t\geq 0}$ be an increasing family of σ-algebras (of subset of Ω). A function $\tau : \Omega \to [0, \infty]$ is called a (strict) stopping time w.r.t. $\{\mathscr{N}_t\}_{t\geq 0}$ if $\{\omega, \tau(\omega) \leq t\} \in \mathscr{N}_t$ for all $t \geq 0$.*

Definition 4.4.6. *Let τ be a stopping time w.r.t. $\{\mathscr{N}_t\}$ and let $\mathscr{N}_\infty$ be the smallest σ-algebracontaining $\mathscr{N}_t$ for all $t \geq 0$. Then the σ-algebra $\mathscr{N}_\tau$ consists of all sets $N \in \mathscr{N}_\infty$ such that $N \cap \{\tau \leq t\} \in \mathscr{N}_t$, for all $t \geq 0$.*

In this case, when $\mathscr{N}_t = \mathscr{M}_t$, an alternative description is $\mathscr{M}_t$ is the σ-algebra generated by $\{X_{\min(s,\tau)}, s \geq 0\}$. Similarly, if $\mathscr{N}_t = \mathscr{F}_t^W$, we get $\mathscr{F}_t^W$ is the σ-algebra generated by $\{W_{\min(s,\tau)}, s \geq 0\}$.

Theorem 4.4.7 (The strong Markov property)**.** *Let f be a bounded Borel function on $\mathbb{R}^d$, τ a stopping time w.r.t. $\mathscr{F}_t^W$, $\tau < \infty$ a.s.. Then*

$$\mathbb{E}^x[f(X_{\tau+h})|\mathscr{F}_\tau^W]_\omega = \mathbb{E}^{X_{\tau(\omega)}}[f(X_h)], \text{ for all } h \geq 0. \tag{4.43}$$

■

4.4.2 Generator of an Itô Diffusion

In many applications, it is fundamental to associate a second order partial differential operator $\mathscr{A}$ to an Itô diffusion $X(t)$.

Definition 4.4.8. *Let $\{X(t), t \geq 0\}$ be a (time-homogeneous) Itô diffusion in $\mathbb{R}^d$. The (infinitesimal) generator $\mathscr{A}$ of X_t is defined by*

$$\mathscr{A}f(x) = \lim_{t \downarrow 0} \frac{\mathbb{E}^x[f(X(t))] - f(x)}{t}, x \in \mathbb{R}^d. \tag{4.44}$$

The set of functions $f : \mathbb{R}^d \mapsto \mathbb{R}$ such that the limit exists at x is denoted $\mathscr{D}_{\mathscr{A}}(x)$, while $\mathscr{D}_{\mathscr{A}}$ denotes the set of functions for which the limit exists for all $x \in \mathbb{R}^d$.

The following result is needed to find the relation between $\mathscr{A}$ and the coefficients b, σ in the stochastic differential equation(4.41)

Lemma 4.4.9. *Let $Y(t) = Y_t = Y_t^x$ be an Itô process in $\mathbb{R}^d$ of the form*

$$Y_t^x(\omega) = x + \int_0^t u(s,\omega)\mathrm{d}s + \int_0^t \sigma(s,\omega)\mathrm{d}W_s(\omega),$$

where W is a d-dimensional Brownian motion. Let $f \in C_0^2(\mathbb{R}^d)$ i.e., $f \in C^2(\mathbb{R}^d)$ and f has compact support, and let τ be a stopping time with respect to $\{\mathscr{F}_t^W\}$ and assume that $\mathbb{E}^x[\tau] < \infty$. Assume that $u(t,\omega)$ and $\sigma(t,\omega)$ are bounded on the set (t,ω) such that $Y(t,\omega)$ belongs to the support of f. Then

$$\mathbb{E}^x[f(Y(\tau))] = f(x) + \mathbb{E}^x\Big[\int_0^\tau \Big\{\sum_i u_i(s,\omega)\frac{\partial f}{\partial x_i}(Y(s)) + \frac{1}{2}\sum_{i,j}(\sigma\sigma^T)_{i,j}(s,\omega)\frac{\partial^2 f}{\partial x_i \partial x_j}(Y(s))\Big\}\mathrm{d}s\Big], \tag{4.45}$$

where $\mathbb{E}^x$ is the expectation w.r.t. the natural probability law R^x for Y_t starting at x.

$$R^x(Y_{t_1} \in F_1, \cdots, Y_{t_k} \in F_k) = \mathbb{P}^0(Y_{t_1}^x \in F_1, \cdots, Y_{t_k}^x \in F_k), \ F_i \text{ Borel sets.}$$

Proof. Let $Z = f(Y)$ and apply Itô's formula to get

$$\begin{aligned}
\mathrm{d}Z(t) &= \sum_{i=1}^d \frac{\partial f}{\partial x_i}\mathrm{d}Y_i(t) + \frac{1}{2}\sum_{i,j}\frac{\partial^2 f}{\partial x_i \partial x_j}\mathrm{d}Y_i(t)\mathrm{d}Y_j(t) \\
&= \sum_{i=1}^d u_i \frac{\partial f}{\partial x_i}\mathrm{d}t + \frac{1}{2}\sum_{i,j}\frac{\partial^2 f}{\partial x_i \partial x_j}(\sigma \mathrm{d}W(t))_i(\sigma \mathrm{d}W(t))_j + \sum_i \frac{\partial f}{\partial x_i}(\sigma \mathrm{d}W(t))_i
\end{aligned}$$

But

$$\begin{aligned}
(\sigma \mathrm{d}W(t))_i(\sigma \mathrm{d}W)_j &= \big(\sum_k \sigma_{ik}\mathrm{d}W_k(t)\big)\big(\sum_n \sigma_{jn}\mathrm{d}W_n(t)\big) \\
&= \big(\sum_k \sigma_{ik}\sigma_{jk}\big)\mathrm{d}t = (\sigma\sigma^T)_{ij}\mathrm{d}t
\end{aligned}$$

leading to

$$f(Y(t)) = f(Y(0)) + \int_0^t \Big\{\sum_i u_i(s,\omega)\frac{\partial f}{\partial x_i}(Y(s)) + \frac{1}{2}\sum_{i,j}(\sigma\sigma^T)_{i,j}(s,\omega)\frac{\partial^2 f}{\partial x_i \partial x_j}(Y(s))\Big\}\mathrm{d}s + \sum_{i,k}\int_0^t \sigma_{ik}(s,\omega)\frac{\partial f}{\partial x_i}(Y(s))\mathrm{d}W_k(s).$$

Hence,

$$\mathbb{E}^x[f(Y(\tau))] = f(x) + \mathbb{E}^x\Big[\int_0^\tau \Big\{\sum_i u_i(s,\omega)\frac{\partial f}{\partial x_i}(Y(s)) + \frac{1}{2}\sum_{i,j}(\sigma\sigma^T)_{i,j}(s,\omega)\frac{\partial^2 f}{\partial x_i \partial x_j}(Y(s))\Big\}\mathrm{d}s\Big]$$
$$+\sum_{i,k}\mathbb{E}^x\Big[\int_0^\tau \sigma_{ik}\frac{\partial f}{\partial x_i}(Y(s))\mathrm{d}W_k(s)\Big].$$

The result is proved if $\mathbb{E}^x\Big[\int_0^\tau \sigma_{ik}\frac{\partial f}{\partial x_i}(Y(s))\mathrm{d}W_k(s)\Big] = 0$. In fact, if g is bounded Borel function, $|g| \le M$. Then for all integer k we have

$$\mathbb{E}^x\Big[\int_0^{\tau\wedge k} g(Y(s))\mathrm{d}W(s)\Big] = \mathbb{E}^x\Big[\int_0^k \mathscr{X}_{\{s<\tau\}} g(Y(s))\mathrm{d}W(s)\Big] = 0,$$

since $g(Y_s)$ and $\mathscr{X}_{\{s<\tau\}}$ are both $\mathscr{F}_s^W$-measurable. In addition

$$\mathbb{E}^x\Big[\int_0^\tau g(Y(s))\mathrm{d}W(s) - \int_0^{\tau\wedge k} g(Y(s))\mathrm{d}W(s)\Big]^2 = \mathbb{E}^x\Big[\int_{\tau\wedge k}^\tau g(Y(s))\mathrm{d}W(s)\Big]^2$$
$$= \mathbb{E}^x\Big[\int_{\tau\wedge k}^\tau g^2(Y_s)\mathrm{d}s\Big] \le M^2(\mathbb{E}^x(\tau) - \mathbb{E}^x(\tau\wedge k)) \to 0 \text{ as } k \to \infty.$$

Therefore, $0 = \lim_{k\to\infty}\mathbb{E}^x\Big[\int_0^{\tau\wedge k} g(Y_s)\mathrm{d}W(s)\Big] = \mathbb{E}^x\Big[\int_0^\tau g(Y_s)\mathrm{d}W(s)\Big]$. The result follows. □

Theorem 4.4.10. *[Generator of a diffusion] Let X_t be an Itô diffusion*

$$\mathrm{d}X(t) = b(X(t))\mathrm{d}t + \sigma(X(t))\mathrm{d}W(t).$$

If $f \in C_0^2(\mathbb{R}^d)$ then $f \in \mathscr{D}_{\mathscr{A}}$ and

$$\mathscr{A} f(x) = \sum_i b_i(x)\frac{\partial f}{\partial x_i} + \frac{1}{2}\sum_{i,j}(\sigma\sigma^T)_{i,j}(x)\frac{\partial^2 f}{\partial x_i \partial x_j}. \tag{4.46}$$

■

Proof. It follows from Lemma4.4.9 that

$$\begin{aligned}
\mathscr{A} f(x) &= \lim_{t\downarrow 0}\frac{\mathbb{E}^x[f(X(t))] - f(x)}{t} \\
&= \lim_{t\downarrow 0}\frac{\mathbb{E}^x\Big[\int_0^t \Big(\sum_i b_i(X(s))\frac{\partial f}{\partial x_i}(X(s)) + \frac{1}{2}\sum_{i,j}(\sigma\sigma^T)_{i,j}(X(s))\frac{\partial^2 f}{\partial x_i \partial x_j}(X(s))\Big)\mathrm{d}s\Big]}{t} \\
&= \sum_i b_i(x)\frac{\partial f}{\partial x_i} + \frac{1}{2}\sum_{i,j}(\sigma\sigma^T)_{i,j}(x)\frac{\partial^2 f}{\partial x_i \partial x_j}
\end{aligned}$$

□

Example 4.4.11. *Find the generator of the following process*

1) $\mathrm{d}X(t) = rX(t)\mathrm{d}t + \alpha X(t)\mathrm{d}W(t)$

2)

$$\begin{bmatrix}\mathrm{d}X^1(t)\\ \mathrm{d}X^2(t)\end{bmatrix} = \begin{bmatrix}0\\1\end{bmatrix}\mathrm{d}t + \begin{bmatrix}1 & 0\\ 0 & X^1(t)\end{bmatrix}\begin{bmatrix}\mathrm{d}W^1(t)\\ \mathrm{d}W^2(t)\end{bmatrix}$$

3)

$$\begin{cases} \mathrm{d}X^1(t) = \mathrm{d}t;\ X^1(0) = t_0 \\ \mathrm{d}X^2(t) = \mathrm{d}W(t);\ X^2(0) = x_0 \end{cases}$$

that is $\mathrm{d}X(t) = b\mathrm{d}t + \sigma \mathrm{d}W(t)$; *with*

$$b = \begin{pmatrix} 1 \\ 0 \end{pmatrix},\ \sigma = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \text{ and } X(0) = \begin{pmatrix} t_0 \\ x_0 \end{pmatrix}.$$

Solution

1) $b(x) = rx$; $\sigma(s) = \alpha x$, $Af = rx\frac{\partial f}{\partial x} + \frac{1}{2}\alpha^2 x^2 \frac{\partial^2 f}{\partial x^2}$.

2)

$$b = \begin{pmatrix} 1 \\ 0 \end{pmatrix};\ \sigma = \begin{pmatrix} 1 & 0 \\ 0 & x_1 \end{pmatrix};\ \sigma^T = \begin{pmatrix} 1 & 0 \\ 0 & x_1 \end{pmatrix};\ \sigma\sigma^T = \begin{pmatrix} 1 & 0 \\ 0 & x_1^2 \end{pmatrix} = \begin{pmatrix} (\sigma\sigma^T)_{11} & (\sigma\sigma^T)_{12} \\ (\sigma\sigma^T)_{21} & (\sigma\sigma^T)_{22} \end{pmatrix}$$

$\mathscr{A} f = \frac{\partial f}{\partial x} + \frac{1}{2}\frac{\partial^2 f}{\partial x_1^2} + \frac{1}{2}x_1^2 \frac{\partial^2 f}{\partial x_2^2}$.

3) $\mathscr{A} f = \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}$, $f = f(t,x) \in C_0^2(\mathbb{R}^d)$.

Example 4.4.12. *Find an Itô diffusion whose generator is given by*

$$\mathscr{A} f = 2\frac{\partial f}{\partial x} + 4x^2 \frac{\partial^2 f}{\partial x^2}.$$

Theorem 4.4.13 (Dynkin's formula)**.** *Let $f \in C_0^2(\mathbb{R}^n)$. Suppose τ is a stopping time such that $\mathbb{E}^x[\tau] < \infty$. Then*

$$\mathbb{E}^x[f(X(\tau))] = f(x) + \mathbb{E}^x\Big[\int_0^\tau \mathscr{A} f(X(s))\mathrm{d}s\Big].$$

■

Example 4.4.14. *Let W be an m-dimensional Brownian motion starting in some point a and $\tau_R = \inf\{t \geq 0 / \|W(t)\| \geq R\}$. Show that*

$$\mathbb{E}^a[\tau_R] = \frac{R^2 - \|a\|^2}{m}$$

holds for $\|a\| < R$.

Solution (Online session) We use Dynkin's formula. Consider the open ball of radius R in $\mathbb{R}^m$

$$K_R = \{x \in \mathbb{R}^m, \|x\| < R\}.$$

Then, τ can also be seen as the first exit time from this ball.

Let $f \in C_0^2(\mathbb{R})$ such that $f(x) = \|x\|^2$ for $\|x\| < R$. The generator $\mathscr{A} f$ is given by

$$\mathscr{A} f(x) = m = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1} \frac{\partial^2 f}{\partial x_i \partial x_j}(x), \text{ for } \|x\| < R.$$

Then the Dynkin's formula yields

$$\begin{aligned}\mathbb{E}^a[f(W(\tau_R \wedge n))] &= f(a) + \mathbb{E}^a\Big[\int_0^{\tau_R \wedge n} m\mathrm{d}s\Big] \\ &= f(a) + m\mathbb{E}^a[\tau_R \wedge n].\end{aligned}$$

By monotone convergence, we have

$$\mathbb{E}^a[\tau_R] = \lim_{n\to\infty} \mathbb{E}^a[\tau_R \wedge n] = \lim_{n\to\infty} \frac{\mathbb{E}^a[\|W(\tau_R \wedge n)\|^2] - \|a\|^2}{m}.$$

By the dominated convergence theorem we have $\lim_{n\to\infty} \mathbb{E}^a[\|W(\tau_R \wedge n)\|^2] = R^2$, that is

$$\mathbb{E}^a[\tau_R] = \frac{R^2 - \|a\|^2}{m}.$$

Chapter 5

Applications

In this chapter, we discuss some applications and link to PDEs.

5.1 Kolmogorov's equation

Let X_t be an Itô diffusion in $\mathbb{R}^n$ with generator A. Let $f \in C_0^2(\mathbb{R}^n)$ and $\tau = t$ in Dynkin's formula. Set

$$u(t,x) = \mathbb{E}^x[f(X(t))].$$

Then u is differentiable with respect to t and

$$\frac{\partial u}{\partial t}(t,x) = \mathbb{E}^x[\mathscr{A} f(X(t))], \tag{5.1}$$

where $\mathscr{A}$ represents the generator of $X(t)$.

Is it possible to express the right hand side of (5.1) in terms of u?

The following result gives an answer to the question.

Theorem 5.1.1 (Kolmogorov's equation)**.** *Let $f \in C_0^2(\mathbb{R}^n)$ and $X(t)$ a diffusion process in $\mathbb{R}^n$ and with generator $\mathscr{A}$.*

(a) Define

$$u(t,x) = \mathbb{E}^x[f(X(t))]. \tag{5.2}$$

Then $u(t,\cdot) \in \mathscr{D}_{\mathscr{A}}$ for each t and

$$\begin{cases} \frac{\partial u}{\partial t}(t,x) & = \mathscr{A} u(t,x),\ t > 0, x \in \mathbb{R}^n \\ u(0,x) & = f(x),\ x \in \mathbb{R}^n. \end{cases} \tag{5.3}$$

In (5.3)*, the right hand side is to be interpret as $\mathscr{A}$ applied to the function $x \mapsto u(t,x)$.*

(b) Moreover, if $v(t,x) \in C^{1,2}(\mathbb{R} \times \mathbb{R}^n)$ is a bounded function satisfying (5.3) *then $v(t,x) = u(t,x)$, given by* (5.2)*.*

■

Proof. (a) Let $g(x) = u(t,x) = \mathbb{E}^x[f(X(t))]$. Since $t \mapsto u(t,x)$ is differentiable, we have

$$\begin{aligned}
\frac{\partial u}{\partial t} &= \lim_{r \downarrow 0} \frac{u(t+r,x) - u(t,x)}{r} \\
&= \lim_{r \downarrow 0} \frac{1}{r}\big(\mathbb{E}^x[f(X(t+r))] - \mathbb{E}^x[f(X(t))]\big) \\
&= \lim_{r \downarrow 0} \frac{1}{r}\big(\mathbb{E}^x[\mathbb{E}^x[f(X(t+r))]|\mathscr{F}_r] - \mathbb{E}^x[f(X(t))]\big) \quad \text{(Tower property)} \\
&= \lim_{r \downarrow 0} \frac{1}{r}\big(\mathbb{E}^x[\mathbb{E}^{X_r}[f(X(t))]] - \mathbb{E}^x[f(X(t))]\big) \quad \text{(Markov property)} \\
&= \lim_{r \downarrow 0} \frac{1}{r}\big(\mathbb{E}^x[g(X(r))] - g(x)\big),
\end{aligned}$$

that is $\mathscr{A}u = \lim_{r\downarrow 0} \frac{1}{r}\big(\mathbb{E}^x[g(X(r))] - g(x)\big)$ exists. Hence $\frac{\partial u}{\partial t} = \mathscr{A}u$.

(b) Let $v(t,x) \in C^{1,2}(\mathbb{R}\times\mathbb{R}^n)$ and satisfying (5.3) then

$$\begin{cases} \tilde{\mathscr{A}}v(t,x) & := -\frac{\partial v}{\partial t}(t,x) + \mathscr{A}v(t,x),\ t>0, x\in\mathbb{R}^n \\ v(0,x) & = f(x),\ x\in\mathbb{R}^n. \end{cases} \tag{5.4}$$

Fix $(s,x)\in\mathbb{R}\times\mathbb{R}^n$. Define the process $Y(t)\in\mathbb{R}^{n+1}$ by $Y(t) = (s-t, X^{0,x}(t)), t\geq 0$. Then $Y(t)$ has generator $\tilde{\mathscr{A}}$ and using (5.4) and Dynkin's formula, we have, for all $t>0$

$$\mathbb{E}^{s,x}[v(Y(t\wedge\tau_R))] = v(s,x) + \mathbb{E}^{s,x}\Big[\int_0^{t\wedge\tau_R} \tilde{A}v(Y(r))\mathrm{d}r\Big] = v(s,x),$$

where $\tau_R = \inf\{t>0; |X(t)|\geq R\}$.
Letting $R\to\infty$ we get

$$v(s,x) = \mathbb{E}^{s,x}[v(Y(t))],\ \forall t\geq 0.$$

Choose $t=s$,

$$v(s,x) = \mathbb{E}^{s,x}[v(Y(s))] = \mathbb{E}^{s,x}[v(0,X^{0,x}(s))] = \mathbb{E}[f(X^{0,x}(t))] = \mathbb{E}^x[f(X(s))]$$

□

5.2 The Feynman-Kac formula

The following formula gives the probabilistic representation of solutions to some partial differential equations.

Theorem 5.2.1 (The Feynman-Kăc formula)**.** *Let $f\in C_0^2(\mathbb{R}^n), q\in C(\mathbb{R}^n)$. Let X_t be a diffusion process with generator $\mathscr{A}$. Assume that q is lower bounded.*

(a) Put

$$v(t,x) = \mathbb{E}^x\Big[\exp\Big\{-\int_0^t q(X(s))\mathrm{d}s\Big\} f(X(t))\Big] \tag{5.5}$$

then

$$\begin{cases} \frac{\partial v}{\partial t}(t,x) & = \mathscr{A}v(t,x) - qv(t,x),\ t>0, x\in\mathbb{R}^n \\ v(0,x) & = f(x),\ x\in\mathbb{R}^n, \end{cases} \tag{5.6}$$

(b) Conversely, if $w(t,x)\in C^{1,2}(\mathbb{R}\times\mathbb{R}^n)$ is bounded on $K\times\mathbb{R}^n$ for each compact $K\subset\mathbb{R}$ and w solves (5.6)*, then $w(t,x) = v(t,x)$, given by* (5.5)*.*

■

Example 5.2.2.

1. Let Δ be the Laplace operator on $\mathbb{R}^n$.
Write down a bounded solution g of the Cauchy problem

$$\begin{cases} \frac{\partial g}{\partial t}(t,x) - \frac{1}{2}\Delta_x g(t,x) & = 0,\ t>0, x\in\mathbb{R}^n \\ g(0,x) & = \phi(x),\ x\in\mathbb{R}^n, \end{cases}$$

where $\phi\in C_0^2(\mathbb{R}^n)$ is given.

2. *Show that the solution u(t,x) of the initial value problem*

$$\begin{cases} \frac{\partial u}{\partial t} & = \frac{1}{2}\beta^2 x^2 \frac{\partial^2 u}{\partial x^2} + \alpha x \frac{\partial u}{\partial x};\ t > 0, x \in \mathbb{R} \\ u(0,x) & = f(x),\ x \in \mathbb{R} \quad (f \in C_0^2(\mathbb{R}) \textit{ is given}) \end{cases}$$

can be expressed as follows:

$$\begin{aligned} u(t,x) &= \mathbb{E}\big[f(x\exp\big\{\big(\alpha - \frac{1}{2}\beta^2\big)t + \beta B_t\big\})\big] \\ &= \frac{1}{\sqrt{2\pi t}} \int_{\mathbb{R}} f(x\exp\big\{\big(\alpha - \frac{1}{2}\beta^2\big)t + \beta y\big\}) \exp\big(-\frac{y^2}{2t}\big) \mathrm{d}y. \end{aligned}$$

3. *Show that the solution of the initial value problem*

$$\begin{cases} \frac{\partial u}{\partial t}(t,x) & = \frac{1}{2}\Delta u(t,x) + \rho u(t,x);\ t > 0, x \in \mathbb{R} \\ u(0,x) & = f(x),\ x \in \mathbb{R} \quad (f \in C_0^2(\mathbb{R}) \textit{ is given}) \end{cases}$$

(where $\rho \in \mathbb{R}$ is a constant) can be expressed by

$$u(t,x) = \frac{1}{\sqrt{2\pi t}} \exp(\rho t) \int_{\mathbb{R}^n} f(y) \exp\big(-\frac{(x-y)^2}{2t}\big) \mathrm{d}y.$$

Solution

1. We use Theorem 5.1.1 to find that the operator $\mathscr{A} f = \frac{1}{2}\Delta_x f$ is the generator of the process $\mathrm{d}X(t) = \mathrm{d}W(t)$ and using Theorem 5.2.1, we have $g(t,x) = \mathbb{E}^x[\phi(X(t))]$.

   Assume that $\phi(t,x) = e^{2x} + 4x^2$. Find $g(t,x)$. We have that $\mathrm{d}X(t) = \mathrm{d}W(t) \Rightarrow X(t) = X_0 + W(t)$. Then,

$$\begin{aligned} g(t,x) &= \mathbb{E}^x[\phi(X(t))] \\ &= \mathbb{E}[e^{2X(t)} + 4X^2(t)|X_0 = x] \\ &= \mathbb{E}[e^{2(x+W(t))} + 4(x+W(t))^2] \\ &= e^{2x}\mathbb{E}[e^{2W(t)}] + 4\mathbb{E}[x^2 + 2xW(t) + W^2(t)] = e^{2x}e^{2t} + 4x^2 + 4t. \end{aligned}$$

2. We use Theorem 5.1.1.The diffusion process associated to the generator $\mathscr{A} f = \frac{1}{2}\beta^2 x^2 \frac{\partial^2 f}{\partial x^2} + \alpha x \frac{\partial f}{\partial x}$ is given by $\mathrm{d}X(t) = \alpha X(t)\mathrm{d}t + \beta X(t)\mathrm{d}W(t)$.
   Using Theorem 5.2.1, $u(t,x) = \mathbb{E}^x[f(X(t))]$. From Itô's formula, we have

$$X(t) = X_0 \exp\Big\{\big(\alpha - \frac{1}{2}\beta^2\big)t + \beta W(t)\Big\}.$$

   Then

$$\begin{aligned} u(t,x) &= \mathbb{E}^x[f(X(t))] \\ &= \mathbb{E}[f(X(t))|X_0] \\ &= \mathbb{E}\big[x\exp\big\{\big(\alpha - \frac{1}{2}\beta^2\big)t + \beta W(t)\big\}\big] \\ &= \frac{1}{\sqrt{2\pi t}} \int_{\mathbb{R}} f(x\exp\big\{\big(\alpha - \frac{1}{2}\beta^2\big)t + \beta y\big\}) \exp\big(-\frac{y^2}{2t}\big) \mathrm{d}y \end{aligned}$$

3. We use Theorem 5.2.1. The process associated to the generator $\mathscr{A} f = \frac{1}{2}\Delta f$ is given by $\mathrm{d}X(t) = \mathrm{d}W(t)$, that

is, $X(t) = X_0 + W(t)$. It follows from the Feyman-Kac formula, with $q(x) = -\rho$ that:

$$
\begin{aligned}
u(t,x) &= \mathbb{E}^x\big[\exp\big\{-\int_0^t q(X(s))\mathrm{d}s\big\} f(X(t))\big] \\
&= \mathbb{E}\big[\exp\big\{-\int_0^t q(X(s))\mathrm{d}s\big\} f(X(t))|X_0 = x\big] \\
&= \mathbb{E}\big[\exp\big\{\int_0^t \rho \mathrm{d}s\big\} f(x+W(t))\big] \\
&= \exp(\rho t)\frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}} f(x+y)\exp\big\{-\frac{y^2}{2t}\big\}\mathrm{d}y \\
&= \exp(\rho t)\frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}} f(y)\exp\big\{-\frac{(x-y)^2}{2t}\big\}\mathrm{d}y.
\end{aligned}
$$

5.3 Parabolic Equations and Feynman-Kac formula

In this section we wish to discuss the Kolmogorov backward equation which is a parabolic equation and its probabilistic representation. Let $D \subset \mathbb{R}^m$ be a bounded connected open set having a C^2 boundary and let $Q = [0,T[\times D$ and

$$
\mathscr{A} = \sum_i b_i(t,x)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j}(\sigma\sigma^T)_{i,j}(t,x)\frac{\partial^2}{\partial x_i \partial x_j} \tag{5.7}
$$

be a differential operator on $\bar{Q}$ such that

$$
\langle a(t,x)z,z\rangle \geq \lambda|z|^2, \text{ where } \lambda > 0 \text{ for every } (t,x) \in Q, z \in \mathbb{R}^m; \text{ (uniformly ellipticity)} \tag{5.8}
$$

$$
a \text{ and } b \text{ are Lipschitz continuous in } \bar{Q}, \tag{5.9}
$$

where $a(t,x) = (\sigma\sigma^T)(t,x)$.

Before stating the Feyman-Kac formula, we recall the following result

Theorem 5.3.1. *Let ϕ be a continuous function on $\bar{D}$ and g a continuous functionon $\partial D \times [0,T]$ such that $g(T,x) = \phi(x)$ if $x \in \partial D$; let c and f be Hölder continuous functions $\bar{Q} \to \mathbb{R}$. Suppose that* (5.8) *and* (5.9) *hold.*

(a) Then, there exists a unique function $u \in C^{2,1}(Q) \cap C(\bar{Q})$ such that

$$
\begin{cases}
\mathscr{A}u - cu + \frac{\partial u}{\partial t} = f \text{ on } Q, \\
u(T,x) = \phi(x) \text{ on } D, \\
u(t,x) = g(t,x) \text{ on } [0,T] \times \partial D.
\end{cases} \tag{5.10}
$$

(b) Let $X = (X(t), t \geq o)$ be the diffusion associated to the SDE with coefficients b and σ. Then we have the following representation formula

$$
\begin{aligned}
u(t,x) =& \mathbb{E}^{x,t}\Big[g(\tau, X(\tau))e^{-\int_t^\tau c(s,X(s))\mathrm{d}s}1_{\{\tau<T\}}\Big] + \mathbb{E}^{x,t}\Big[\phi(X(T))e^{-\int_t^T c(s,X(s))\mathrm{d}s}1_{\{\tau\geq T\}}\Big] \\
&- \mathbb{E}^{x,t}\Big[\int_t^{\tau\wedge T} f(s,X(s))e^{-\int_t^s c(u,X(u))\mathrm{d}u}\mathrm{d}s\Big],
\end{aligned} \tag{5.11}
$$

where τ is the exit time from D.

■

We have the following Feynman-Kac formula

Theorem 5.3.2. *Let $c : [0,T] \times \mathbb{R}^m \to \mathbb{R}$ be a continuous function that is bounded below (i.e. $c(t,x) \geq -K > -\infty$) and w a function in $C^{1,2}([0,T] \times \mathbb{R}^m)$ and a solution of the problem*

$$
\begin{cases}
\mathscr{A}w - cw + \frac{\partial w}{\partial t} = f \text{ on } [0,T[\times\mathbb{R}^m, \\
w(T,x) = \phi(x).
\end{cases} \tag{5.12}
$$

Assume in addition that the coefficients a and b are Lipschitz continuous on $[0,T]\times\mathbb{R}^m$ and that a is uniformly elliptic. Furthermore, assume that ϕ and f are continuous and are either both positives or of polynomial growth. Assume, finally, that w is of polynomial growth. Then we have the following representation formula

$$w(t,x) = \mathbb{E}^{x,t}\Big[\phi(X(T))e^{-\int_t^T c(s,X(s))\mathrm{d}s}\Big] - \mathbb{E}^{x,t}\Big[\int_t^T f(s,X(s))e^{-\int_t^s c(u,X(u))\mathrm{d}u}\mathrm{d}s\Big], \tag{5.13}$$

■

We also have the following result

Theorem 5.3.3. *Let us assume hypotheses of Theorem 5.3.2 are satisfied. Suppose in addition that for every $R>0$ there exists a $\lambda_R>0$ such that $\langle a(t,x)z,z\rangle \geq \lambda_R|z|^2$ for (t,x), $|x|\leq R$, $0\leq t\leq T$, $z\in\mathbb{R}^m$. Furthermore, suppose that c and f are locally Hölder continuous.*

Then there exists a function u, continuous on $[0,T]\times\mathbb{R}^m$ and $C^{1,2}([0,T[\times\mathbb{R}^m)$, which is a solution of

$$\begin{cases} \mathscr{A}u - cu + \frac{\partial u}{\partial t} = f \text{ on } [0,T[\times\mathbb{R}^m, \\ u(T,x) = \phi(x). \end{cases} \tag{5.14}$$

This solution u is given by

$$u(t,x) = \mathbb{E}^{x,t}\Big[\phi(X(T))e^{-\int_t^T c(s,X(s))\mathrm{d}s}\Big] - \mathbb{E}^{x,t}\Big[\int_t^T f(s,X(s))e^{-\int_t^s c(u,X(u))\mathrm{d}u}\mathrm{d}s\Big], \tag{5.15}$$

where X is the diffusion associated to $\mathscr{A}$. In addition, u is the unique solution to (5.14) *with polinomial growth.*

If we addume in addition that σ is bounded, then u is the unique solution with exponential growth ;provided that ϕ and f have exponential growth.

■

Example 5.3.4. *Let us consider the following problem*

$$\begin{cases} \frac{1}{2}\frac{\partial^2 u}{\partial x^2} + \frac{\partial u}{\partial t} = 0 \text{ on } [0,T[\times\mathbb{R} \\ u(T,x) = x^2 \end{cases}$$

Find a solution of the above equation.

The diffusion associated to the Laplacian operator is the Brownian motion so that

$$\mathrm{d}X(t) = \mathrm{d}W(t)$$

Thus integrating from t to T we get $X(T) = X(t) + W(T) - W(t)$ Using the previous Theorem , we get

$$\begin{aligned} \mathbb{E}^{t,x}[X^2(t)] =& \mathbb{E}^{t,x}[(X(t)+W(T)-W(t))^2] \\ =& \mathbb{E}[(x+W(T)-W(t))^2] \\ =& \mathbb{E}[(x+W(T-t))^2] \\ =& \mathbb{E}[x^2+2xW(T-t)+W^2(T-t)] \\ =& x^2+2x\mathbb{E}[W(T-t)]+\mathbb{E}[W^2(T-t)] = x^2+T-t \end{aligned}$$

Example 5.3.5. *Let us consider the following problem*

$$\begin{cases} \frac{\sigma^2}{2}x^2\frac{\partial^2 u}{\partial x^2} + bx\frac{\partial u}{\partial x} + \frac{\partial u}{\partial t} = 0 \text{ on } [0,T[\times\mathbb{R} \\ u(T,x) = x \end{cases}$$

Find a solution of the above problem. Find the solution when terminal condition is replaced by x^2.

The diffusion associated to the above operator is the Brownian motion so that

$$\mathrm{d}X(t) = bX(t)\mathrm{d}t + \sigma X(t)\mathrm{d}W(t)$$

Thus using Itô's formula, we get

$$\begin{aligned} X(T) =& X(0)e^{(b-\frac{\sigma^2}{2})T+\sigma W(T)} \\ X(t) =& X(0)e^{(b-\frac{\sigma^2}{2})t+\sigma W(t)}. \end{aligned}$$

So that

$$X(T) = X(t)e^{(b-\frac{\sigma^2}{2})(T-t)+\sigma(W(T)-W(t))}.$$

Using once more, the previous Theorem , we get

$$\begin{aligned} \mathbb{E}^{t,x}[X(T)] =& \mathbb{E}^{t,x}\left[X(t)e^{(b-\frac{\sigma^2}{2})(T-t)+\sigma(W(T)-W(t))}\right] \\ =& \mathbb{E}\left[xe^{(b-\frac{\sigma^2}{2})(T-t)+\sigma(W(T)-W(t))}\right] \\ =& xe^{(b-\frac{\sigma^2}{2})(T-t)}\mathbb{E}\left[e^{\sigma W(T-t)}\right] \\ =& xe^{(b-\frac{\sigma^2}{2})(T-t)}e^{\frac{\sigma^2}{2}(T-t)}. \end{aligned}$$

In the second case, we have

$$\begin{aligned} \mathbb{E}^{t,x}[X^2(t)] =& \mathbb{E}^{t,x}\left[X^2(t)e^{2(b-\frac{\sigma^2}{2})(T-t)+2\sigma(W(T)-W(t))}\right] \\ =& \mathbb{E}\left[x^2e^{2(b-\frac{\sigma^2}{2})(T-t)+2\sigma(W(T)-W(t))}\right] \\ =& x^2e^{2(b-\frac{\sigma^2}{2})(T-t)}\mathbb{E}\left[e^{2\sigma W(T-t)}\right] \\ =& x^2e^{2(b-\frac{\sigma^2}{2})(T-t)}e^{2\sigma^2(T-t)} \\ =& x^2e^{(2b+\sigma^2)(T-t)}. \end{aligned}$$

Example 5.3.6. *Find the solution $u(t,x)$ of the initial value problem*

$$\begin{cases} \frac{\partial u}{\partial t} & = \frac{1}{2}\sigma^2\frac{\partial^2 u}{\partial x^2} + (a-\beta x)\frac{\partial u}{\partial x};\ t>0, x\in\mathbb{R} \\ u(0,x) & = f(x),\ x\in\mathbb{R} \end{cases}$$

for $f(x) = x$ and $e^{\theta x}$ respectively. Check the solution satisties the equation.

First, we find the diffusion associated to the above genertator.

$$\mathrm{d}X(t) = (a-\beta X(t))\mathrm{d}t + \sigma\mathrm{d}W(t)$$

$Y(t) = e^{\beta t}X(t)$ and $f(t,x) = e^{\beta t}x$. We have then: $\frac{\partial f}{\partial x} = e^{\beta t}$; $\frac{\partial^2 f}{\partial x^2} = 0$; $\frac{\partial f}{\partial x} = \beta e^{\beta t}x$

$$\begin{aligned} Y(t) =& Y(0) + \int_0^t \beta e^{\beta s}X(s)\mathrm{d}s + \int_0^t e^{\beta s}\mathrm{d}X(s) + 0 \\ =& Y(0) + \int_0^t \beta e^{\beta s}X(s)\mathrm{d}s + \int_0^t e^{\beta s}(a-\beta X(t))\mathrm{d}s + \int_0^t e^{\beta s}\sigma\mathrm{d}W(t) \\ =& X(0) + \int_0^t e^{\beta s}a\mathrm{d}s + \int_0^t e^{\beta s}\sigma\mathrm{d}W(s) \\ =& X(0) + \frac{a}{\beta}[e^{\beta t}-1] + \int_0^t e^{\beta s}\sigma\mathrm{d}W(s) \end{aligned}$$

Thus, we have

$$
\begin{aligned}
e^{\beta t}X(t) =& X(0)+\frac{a}{\beta}[e^{\beta t}-1]+\int_0^t e^{\beta s}\sigma \mathrm{d}W(s)\\
X(t) =& e^{-\beta t}X(0)+\frac{a}{\beta}[1-e^{-\beta t}]+e^{-\beta t}\int_0^t e^{\beta s}\sigma \mathrm{d}W(s)
\end{aligned}
$$

Taking expectation on both sides gives

$$
\begin{aligned}
\mathbb{E}^x[X(t)] =& e^{-\beta t}x+\frac{a}{\beta}[1-e^{-\beta t}]+e^{-\beta t}\mathbb{E}^x\left[\int_0^t e^{\beta s}\sigma \mathrm{d}W(s)\right]\\
=& xe^{-\beta t}+\frac{a}{\beta}[1-e^{-\beta t}] \qquad (5.16)
\end{aligned}
$$

Thus

$$
u(t,x)=\mathbb{E}^x[X(t)]=xe^{-\beta t}+\frac{a}{\beta}[1-e^{-\beta t}] \qquad (5.17)
$$

In addition, we have that

$$
\begin{aligned}
\frac{\partial u}{\partial t} =& -\beta xe^{-\beta t}+ae^{-\beta t}\\
(a-\beta x)\frac{\partial u}{\partial x} =& e^{-\beta t}(a-\beta x)\\
\frac{1}{2}\sigma^2\frac{\partial^2 u}{\partial x^2} =& 0
\end{aligned}
$$

Thus u clearly satisfies the equation.

For $f(x)=e^{\theta x}$, we have

$$
\begin{aligned}
\mathbb{E}^x[e^{\theta X(t)}] =& \mathbb{E}^x\left[\exp\left\{\theta\left(e^{-\beta t}X(0)+\frac{a}{\beta}[1-e^{-\beta t}]+e^{-\beta t}\int_0^t e^{\beta s}\sigma \mathrm{d}W(s)\right)\right\}\right]\\
=& \exp\left\{\theta\left(e^{-\beta t}x+\frac{a}{\beta}[1-e^{-\beta t}]\right)\right\}\mathbb{E}\left[\exp\left\{\theta\left(e^{-\beta t}\int_0^t e^{\beta s}\sigma \mathrm{d}W(s)\right)\right\}\right]\\
=& \exp\left\{\theta\left(e^{-\beta t}x+\frac{a}{\beta}[1-e^{-\beta t}]\right)\right\}\mathbb{E}\left[\exp\left\{\int_0^t \theta\sigma e^{-\beta(t-s)}\mathrm{d}W(s)\right\}\right]
\end{aligned}
$$

Using Exercise 2 tutorial 7, we have

$$
\begin{aligned}
u(t,x)=\mathbb{E}^x[e^{\theta X(t)}] =& \exp\left\{\theta\left(e^{-\beta t}x+\frac{a}{\beta}[1-e^{-\beta t}]\right)\right\}\mathbb{E}\left[\exp\left\{\int_0^t \theta\sigma e^{-\beta(t-s)}\mathrm{d}W(s)\right\}\right]\\
=& \exp\left\{\theta\left(e^{-\beta t}x+\frac{a}{\beta}[1-e^{-\beta t}]\right)\right\}\exp\left\{\frac{\theta^2\sigma^2}{2}\int_0^t e^{-2\beta(t-s)}\mathrm{d}s\right\}\\
=& \exp\left\{\theta\left(e^{-\beta t}x+\frac{a}{\beta}[1-e^{-\beta t}]\right)\right\}\exp\left\{\frac{\theta^2\sigma^2}{4\beta}(1-e^{-2\beta t})\right\}\\
=& \exp\left\{\theta\left(e^{-\beta t}x+\frac{a}{\beta}[1-e^{-\beta t}]\right)+\frac{\theta^2\sigma^2}{4\beta}(1-e^{-2\beta t})\right\}
\end{aligned}
$$

In addition, we have that

$$
\begin{aligned}
\frac{\partial u}{\partial t} =& (-\theta x\beta e^{-\beta t}+a\theta e^{-\beta t}+\frac{\theta^2\sigma^2}{2}e^{-2\beta t})\exp\left\{\theta\left(e^{-\beta t}x+\frac{a}{\beta}[1-e^{-\beta t}]\right)+\frac{\theta^2\sigma^2}{4\beta}(1-e^{-2\beta t})\right\}\\
=& (-\theta x\beta e^{-\beta t}+a\theta e^{-\beta t}+\frac{\theta^2\sigma^2}{2}e^{-2\beta t})u(t,x)\\
\frac{\partial u}{\partial x} =& \theta e^{-\beta t}\exp\left\{\theta\left(e^{-\beta t}x+\frac{a}{\beta}[1-e^{-\beta t}]\right)+\frac{\theta^2\sigma^2}{4\beta}(1-e^{-2\beta t})\right\}=\theta e^{-\beta t}u(t,x)\\
\frac{\partial^2 u}{\partial x^2} =& \theta^2 e^{-2\beta t}\exp\left\{\theta\left(e^{-\beta t}x+\frac{a}{\beta}[1-e^{-\beta t}]\right)+\frac{\theta^2\sigma^2}{4\beta}(1-e^{-2\beta t})\right\}=\theta^2 e^{-2\beta t}u(t,x)
\end{aligned}
$$

So that

$$
\begin{aligned}
\frac{1}{2}\sigma^2\frac{\partial^2 u}{\partial x^2} =& \frac{\theta^2\sigma^2}{2}e^{-2\beta t}u(t,x) \\
(a-\beta x)\frac{\partial u}{\partial x} =& (a-\beta x)\theta e^{-\beta t}u(t,x)
\end{aligned}
$$

Putting all this together yield the result.

In case we were for example looking at the backward equation, we can write $X(T)$ in terms of $X(t)$ as follows using Itô's formula

$$
\begin{aligned}
Y(T) =& Y(t) + \int_t^T \beta e^{\beta s} X(s)\mathrm{d}s + \int_t^T e^{\beta s}\mathrm{d}X(s) + 0 \\
=& X(t)e^{\beta t} + \int_t^T e^{\beta s} a\mathrm{d}s + \int_t^T e^{\beta s}\sigma \mathrm{d}W(s) \\
=& X(t)e^{\beta t} + \frac{a}{\beta}[e^{\beta T} - e^{\beta t}] + \int_t^T e^{\beta s}\sigma \mathrm{d}W(s)
\end{aligned}
$$

Finally, we have

$$
X(T) = X(t)e^{-\beta(T-t)} + \frac{a}{\beta}[1 - e^{-\beta(T-t)}] + \int_t^T e^{-\beta(T-s)}\sigma \mathrm{d}W(s)
$$

From here one can also compute the expectation.