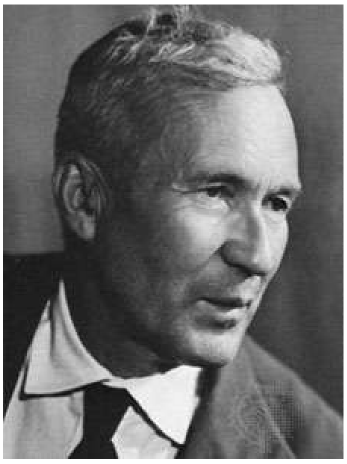
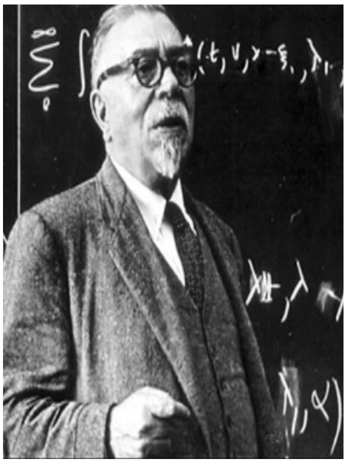
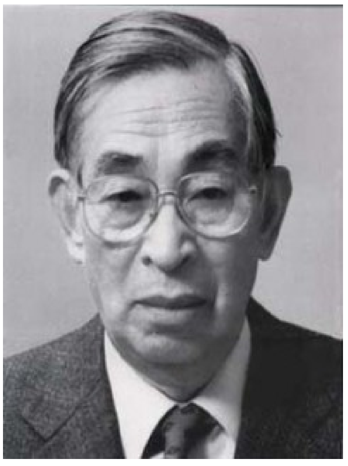
Stochastic Processes

Discrete and Continuous Time
Markov, Martingales, Brownian motion
Itô Calculus







by
YURI F. SAPORITO

Lecture notes mostly based on

“Introduction to Stochastic Processes” by Hoel, Port and Stone,

“Markov Chains” by J. R. Norris,

“Stochastic Calculus and Financial Applications” by J. M. Steele

and

“Backward Stochastic Differential Equations - From Linear to Fully Nonlinear Theory” by J. Zhang.

Contents

Part I

Discrete Time

Chapter 1

Discrete-Time Markov Chains

1.1 Introduction

We start by fixing a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a countable set I, which will be the state space of the stochastic processes considered in this chapter.

Definition 1.1.1 (Markov chain)**.** Let $(X_n)_{n\in\mathbb{N}}$ be a stochastic process taking values in I. We say that $(X_n)_{n\in\mathbb{N}}$ is a *Markov chain* if it satisfies the *Markov property*:

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j \mid X_n = i), \tag{1.1}$$

for any $n \geq 0$ and any $i, i_0, \ldots, i_{n-1}, j \in I$.

A Markov chain is a process that the past realizations do not give more knowledge to forecast the next step than just knowing the present state. In other words, given the present, past and future are independent.

Proposition 1.1.2. *If a stochastic process $(X_n)_{n\in\mathbb{N}}$ satisfies*

(ii)' for any $n \geq 0$ and any $i, i_0, \ldots, i_{n-1}, j \in I$,

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \text{ depends only on } (i,j), \tag{1.2}$$

then $(X_n)_{n\in\mathbb{N}}$ is Markov chain.

Proof. Define a_{ij} as

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = a_{ij}.$$

Then, we have to prove that, for any $n \geq 0$,

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = a_{ij}.$$

We will use induction. Notice that for $n = 0$ the result is clear from (1.2). Assume the result is true for any $k \leq n-1$. This means that

$$\mathbb{P}(X_{k+1} = j \mid X_k = i, X_{k-1} = i_{k-1}, \ldots, X_0 = i_0) = \mathbb{P}(X_{k+1} = j \mid X_k = i),$$

for any $k \leq n-1$ and $i_0, \ldots, i_{k-1}, i, j \in I$. Let's prove it is true for n:

$$\mathbb{P}(X_{n+1} = j, X_n = i) = \sum_{i_0,\ldots,i_{n-1}\in I} \mathbb{P}(X_{n+1} = j, X_n = i, X_{n-1} = i_{n-1} \ldots, X_0 = i_0)$$
$$= \sum_{i_0,\ldots,i_{n-1}\in I} \mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0)$$
$$\mathbb{P}(X_n = i \mid X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \cdots \mathbb{P}(X_1 = i_1 \mid X_0 = i_0)\mathbb{P}(X_0 = i_0)$$
$$= \sum_{i_0,\ldots,i_{n-1}\in I} p_{ij}\mathbb{P}(X_n = i \mid X_{n-1} = i_{n-1})\mathbb{P}(X_{n-1} = i_{n-1} \mid X_{n-2} = i_{n-2})$$

$$
\begin{aligned}
&\cdots \mathbb{P}(X_1 = i_1 \mid X_0 = i_0)\mathbb{P}(X_0 = i_0) \\
&= a_{ij} \sum_{i_1,\dots,i_{n-1}\in I} \mathbb{P}(X_n = i \mid X_{n-1} = i_{n-1})\mathbb{P}(X_{n-1} = i_{n-1} \mid X_{n-2} = i_{n-2}) \\
&\cdots \mathbb{P}(X_2 = i_2 \mid X_1 = i_1) \sum_{i_0\in I} \mathbb{P}(X_1 = i_1 \mid X_0 = i_0)\mathbb{P}(X_0 = i_0) \\
&= a_{ij} \sum_{i_2,\dots,i_{n-1}\in I} \mathbb{P}(X_n = i \mid X_{n-1} = i_{n-1})\mathbb{P}(X_{n-1} = i_{n-1} \mid X_{n-2} = i_{n-2}) \\
&\cdots \sum_{i_1\in I} \mathbb{P}(X_2 = i_2 \mid X_1 = i_1)\mathbb{P}(X_1 = i_1) \\
&= \cdots = a_{ij} \sum_{i_{n-1}\in I} \mathbb{P}(X_n = i \mid X_{n-1} = i_{n-1})\mathbb{P}(X_{n-1} = i_{n-1}) = a_{ij}\mathbb{P}(X_n = i)
\end{aligned}
$$

and this finishes the proof. □

A distribution on I is any sequence $(\lambda_i)_{i\in I}$ such that $\lambda_i \geq 0$ and $\sum_{i\in I} \lambda_i = 1$. A (possibly infinite) matrix $P = (p_{ij})_{i,j\in I}$ is said to be *stochastic* if each row $(p_{ij})_{j\in I}$, $i \in I$, is a distribution:

$$\sum_{j\in I} p_{ij} = 1.$$

Definition 1.1.3 (Markov Chain)**.** Let $(X_n)_{n\in\mathbb{N}}$ be a stochastic process taking values in I. We say that $(X_n)_{n\in\mathbb{N}}$ is a (time-homogeneous) *Markov chain* with *initial distribution* λ and *transition matrix* $P = (p_{ij})_{i,j\in I}$, and write, Markov(λ, P), if

(i) $\mathbb{P}(X_0 = i) = \lambda_i$, for any $i \in I$, and

(ii) for any $n \geq 0$ and any $i, i_0, \dots, i_{n-1}, j \in I$,

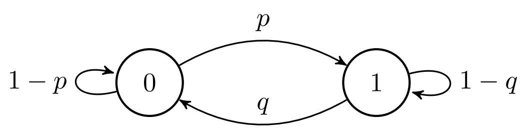
$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = p_{ij}$$

1.1.1 Examples

Example 1.1.4 (2 State Markov Chain)**.** Let us consider a Markov chain with $I = \{0, 1\}$, P given as

$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix},$$

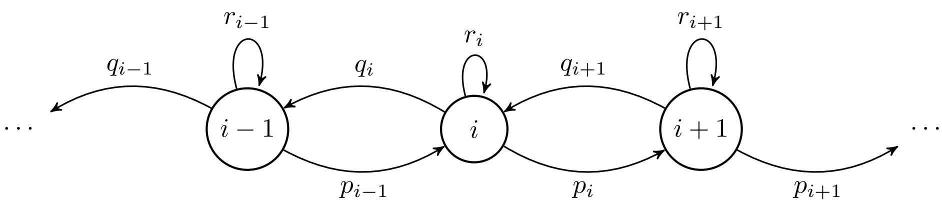
and a general initial distribution λ.



Example 1.1.5 (Birth and Death Chains)**.** Let us consider a Markov chain with $I = \{0, \dots, d\}$, where d could be $+\infty$, and

$$p_{ij} = \begin{cases} q_i, & \text{if } j = i-1, \\ r_i, & \text{if } j = i, \\ p_i, & \text{if } j = i+1, \end{cases}$$

where $q_0 = 0$ and $p_d = 0$, when $d < +\infty$.



An interesting question that we will answer later is to compute the probability of hitting i before hitting j, assuming the chain starts at i_0, where $i < i_0 < j$.

Example 1.1.6 (Random Walk)**.** Consider a sequence of iid r.v's $(\xi_n)_{n\in\mathbb{N}}$ taking integer values and with common probability mass function f. Assume also that X_0 is a r.v. independent of $(\xi_n)_{n\in\mathbb{N}}$ and also taking integer values. Define then, for any $n \in \mathbb{N}$,

$$X_n = X_0 + \sum_{k=1}^{n} \xi_k.$$

It is easy to see that $(X_n)_{n\in\mathbb{N}}$ is a Markov chain with $I = \mathbb{Z}$. Indeed,

$$\begin{aligned}&\mathbb{P}(X_0 = i_0, \dots, X_n = i_n)\\ &= \mathbb{P}(X_0 = i_0, \xi_1 = i_1 - i_0, \xi_2 = i_2 - i_1, \dots, \xi_n = i_n - i_{n-1})\\ &= \lambda_{i_0} \prod_{k=0}^{n-1} f(i_{k+1} - i_k),\end{aligned}$$

where λ is the initial distribution. Moreover, we can conclude $p_{ij} = f(j - i)$, for any $i, j \in \mathbb{Z}$. The simple random walk is the special case where $f(1) + f(-1) = 1$.

Before continuing with the examples of Markov chains, we state an important definition:

Definition 1.1.7. An state $i \in I$ is said to be *absorbing* if $p_{ii} = 1$.

Example 1.1.8 (Branching Process)**.** Let us consider particles that randomly generate new particles of the same type. Let X_n be the number of particles in the n-th generation. Assume that these new particles are generated independently and accordingly with a r.v. ξ taking integers and non-negative values. Let f be the probability mass function of ξ, i.e. $f(j) = \mathbb{P}(\xi = j)$.

Some facts about this process:

- The state 0 is absorbent.
- For $i \geq 1$, $p_{ij} = \mathbb{P}(\sum_{k=1}^{i} \xi_k = j)$, where $(\xi_k)_{k\in\{1,\dots,i\}}$ is an iid sequence of r.v's with distribution f.
- $p_{1j} = f(j)$.
- $\xi = 0$ means that particle has disappeared.

Assume this chain starts with one particle. An interesting question that we will answer later is to compute the probability of extinction, i.e. the probability of hitting the state 0 in finite time. Notice that, if ρ is this probability with $X_0 = 1$, but the chain has actually started with i particles, the probability of extinction is ρ^i.

Example 1.1.9 (Queuing Chain)**.** Let us divide the time in consistent periods, say in hours. Suppose there are clients arriving and waiting for a certain service at the beginning of a period. In this case, exactly one client is served during this period. Otherwise, if there are no clients waiting, no clients are served.

Let ξ_n be the number of new clients that arrive at the n-th step. Let us assume $(\xi_n)_{n\in\mathbb{N}}$ are iid and take integer values. Denote their common probability mass function by f. We will then denote the number of clients waiting for service at time n by X_n. Therefore, from the explanation of the process above, we have

$$\begin{aligned}X_n = 0 &\Rightarrow X_{n+1} = \xi_{n+1},\\ X_n \geq 1 &\Rightarrow X_{n+1} = X_n + \xi_{n+1} - 1.\end{aligned}$$

Obviously, $(X_n)_{n\in\mathbb{N}}$ is a Markov chain with state space $I = \mathbb{N}$. Moreover,

$$p_{ij} = \begin{cases} f(j), & \text{if } i = 0,\\ \\ f(j - i + 1), & \text{if } i \geq 1.\end{cases}$$

The next examples are particular cases of the birth and death chains.

Example 1.1.10 (Ehrenfest Chain)**.** Consider two boxes labeled 1 and 2 and d marbles labeled from 1 to d. Initially, some marbles are placed in box 1 and the others in box 2. We then randomly pick a number between 1 and d and swap the ball with that number from its box to the other. Repeat this procedure infinitely many times, each of them independently of the other.

Define X_n as the number of marbles in box 1 after repeating the experiment above n times. Obviously, X_n is a Markov chain (can you prove this?). Note that the state space is $\{0, \dots, d\}$.

Let us compute the transition probability. Assume that box 1 has i marbles at time n. The probability of randomly picking a marble from box 1 is i/d. In this case, at time $n+1$, there will be $i-1$ marbles in box 1. Similarly, with probability $(d-i)/d$, there will be $i+1$ marbles in box 1. Therefore,

$$p_{ij} = \begin{cases} \dfrac{i}{d}, & \text{if } j = i-1, \\[2ex] \dfrac{d-i}{d}, & \text{if } j = i+1. \end{cases}$$

Example 1.1.11 (Modified Ehrenfest Chain)**.** Let us consider the experiment described to build the Ehrenfest chain. However, now we randomly chose a box to put the marble we have just picked. In this case, it is easy to see that the state space is $\{0, \dots, d\}$ and

$$p_{ij} = \begin{cases} \dfrac{i}{2d}, & \text{if } j = i-1, \\[2ex] \dfrac{1}{2}, & \text{if } j = i, \\[2ex] \dfrac{d-i}{2d}, & \text{if } j = i+1. \end{cases}$$

Example 1.1.12 (Gambler's Ruin Chain)**.** Assume that a gambler enters the casino with d dollars and that he makes a series of bets of 1 dollar against the house. Assume that he wins with probability p and loses with probability $q = 1-p$. We further assume that when his capital becomes 0, it will stay 0 forever. Additionally, we can modify this problem so that the gambler stops gambling when he hits k dollars of wealth.

Denote by X_n the gambler's wealth at time n. Clearly, $(X_n)_{n\in\mathbb{N}}$ is a Markov chain. Moreover,

$$p_{ij} = \begin{cases} p, & \text{if } j = i+1, \\[2ex] q, & \text{if } j = i-1, \end{cases}$$

for $i \geq 1$, and $p_{00} = 1$.

1.1.2 Equivalent Definitions and Some Formulas

In this section we will see some other characterization of Markov chains and the relation of multiple step probabilities with the one-step transition probability matrix.

Theorem 1.1.13. *$(X_n)_{n\in\mathbb{N}}$ is Markov(λ, P) if, and only if, for any $n \geq 0$ and any $i_0, \dots, i_n \in I$,*

$$\mathbb{P}(X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \lambda_{i_0} \prod_{k=0}^{n-1} p_{i_k, i_{k+1}} \tag{1.3}$$

Proof. Assume $(X_n)_{n\in\mathbb{N}}$ is Markov(λ, P). Then,

$$\begin{aligned} &\mathbb{P}(X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \\ &= \mathbb{P}(X_0 = i_0)\mathbb{P}(X_1 = i_1 \mid X_0 = i_0)\mathbb{P}(X_2 = i_2 \mid X_1 = i_1, X_0 = i_0) \end{aligned}$$

$$\begin{aligned}&\cdots \mathbb{P}(X_n = i_n \mid X_{n-1} = i_{n-1}, \ldots, X_{n-1} = i_{n-1})\\&= \lambda_{i_0} \prod_{k=0}^{n-1} p_{i_k, i_{k+1}}.\end{aligned}$$

Suppose now that (1.3) is true. Then $\mathbb{P}(X_0 = i_0) = \lambda_{i_0}$ (taking $n = 0$) and

$$\begin{aligned}&\mathbb{P}(X_n = i_n \mid X_{n-1} = i_{n-1}, \ldots, X_0 = i_0)\\&= \frac{\mathbb{P}(X_n = i_n, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0)}{\mathbb{P}(X_{n-1} = i_{n-1}, , \ldots, X_0 = i_0)} = p_{i_{n-1}, i_n},\end{aligned}$$

which implies that $(X_n)_{n\in\mathbb{N}}$ is Markov(λ, P) by Proposition 1.1.2.

□

Remark 1.1.14. It is important in what follows to characterize the information contained in the r.v's $X_0, \ldots, X_m$. Since these r.v's take values in a countable set I, we have $A \in \sigma(X_0, \ldots, X_m)$ if and only if $A = \cup_{k=1}^{+\infty} A_k$, where A_k is of the form $A_k = \{X_0 = i_0, \ldots, X_m = i_m\}$.

Theorem 1.1.15. *Let $(X_n)_{n\in\mathbb{N}}$ be Markov(λ, P). Then, given $X_m = i$, $(X_{m+n})_{n\in\mathbb{N}}$ is Markov(δ_i, P) and it is independent of $X_0, \ldots, X_m$.*

Proof. Notice that if we prove that for any $A \in \sigma(X_0, \ldots, X_m)$, we have

$$\begin{aligned}&\mathbb{P}(\{X_m = i_m, \ldots, X_{m+n} = i_{m+n}\} \cap A \mid X_m = i)\\&= \delta_i(i_m) p_{i_m, i_{m+1}} \cdots p_{i_{m+n-1}, i_{n+m}} \mathbb{P}(A \mid X_m = i),\end{aligned}$$

the theorem is proved. Now, the equation above is easily proved by noticing that A can be written as $A = \cup_{k=1}^{+\infty} A_k$, where A_k is of the form $A_k = \{X_0 = i_0, \ldots, X_m = i_m\}$. □

Definition 1.1.16 (Alternative Definition)**.** $(X_n)_{n\in\mathbb{N}}$ is a Markov chain if there exist a sequence of iid r.v's $(\xi_n)_{n\in\mathbb{N}}$ and $F : I \times \mathbb{R} \longrightarrow I$ measurable such that

$$X_n = F(X_{n-1}, \xi_n), \ \forall\, n \in \mathbb{N}.$$

Remark 1.1.17. The definition above is more practical, although restrictive. One can show that for any stochastic matrix P, there exists a Markov chain (in the sense of the alternative definition) such that

$$\mathbb{P}(X_n = j \mid X_{n-1} = i) = p_{ij}.$$

In particular, for any Markov chain $(X_n)_{n\in\mathbb{N}}$ (in the sense of the original definition), there exists another process $(\bar{X}_n)_{n\in\mathbb{N}}$ built using the algorithm of the alternative definition that has the same law as $(X_n)_{n\in\mathbb{N}}$.

Before we proceed, let us fix some notation

$$p_{ij}^{(n)} = (P^n)_{ij} = \sum_{i_1, \ldots, i_{n-1} \in I} p_{i, i_1} \cdots p_{i_{n-1}, j}.$$

The proof of the next proposition is elementary and left to the reader.

Proposition 1.1.18. *Let $(X_n)_{n\in\mathbb{N}}$ be Markov(λ, P). Then, for any $n, m \geq 0$,*

(i) $\mathbb{P}(X_{n+m} = j \mid X_m = i) = p_{ij}^{(n)}$

(ii) $\mathbb{P}(X_n = j) = \sum_{k\in I} \lambda_k p_{kj}^{(n)}$

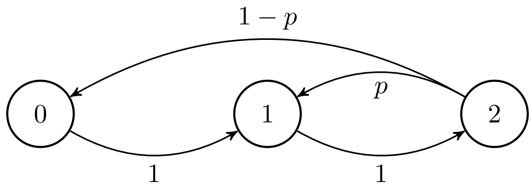
(iii) (Chapman-Kolmogorov) $p_{ij}^{(n+m)} = \sum_{k\in I} p_{ik}^{(n)} p_{kj}^{(m)}$

Proposition 1.1.19. *Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain. If $g : I \longrightarrow J$ is a one-to-one measurable function, then $(g(X_n))_{n\in\mathbb{N}}$ is also a Markov chain.*

Proof. If follows easily from $g(X_n) = j \Leftrightarrow X_n = g^{-1}(j)$.

$\square$

Example 1.1.20. The proposition above breaks if g is not one-to-one. Consider the following example:



Consider now $g : \{0, 1, 2\} \longrightarrow \{a, b\}$ given as $g(0) = a$ and $g(1) = g(2) = b$. We know that $a \to b$ in this new chain, with probability 1. However, just knowing we are at b does not give us the knowledge to what will be a probability to go to a or stay at b. Indeed, denote $Y_n = g(X_n)$ and notice

$$\mathbb{P}(Y_3 = a \mid Y_2 = b, Y_1 = b, Y_0 = a) = \mathbb{P}(X_3 = 0 \mid X_2 = 2, X_1 = 1, X_0 = 0) = 1 - p,$$
$$\mathbb{P}(Y_3 = a \mid Y_2 = b, Y_1 = a, Y_0 = b) = \mathbb{P}(X_3 = 0 \mid X_2 = 1, X_1 = 0, X_0 = 2) = 0.$$

1.2 Strong Markov Property

Definition 1.2.1. A r.v. $\tau : \Omega \longrightarrow \mathbb{N} \cup \{+\infty\}$ is a *stopping time* for the sequence $(X_n)_{n\in\mathbb{N}}$ if

$$\{\tau = n\} \in \sigma(X_0, X_1, \dots, X_n).$$

This means that to decide if $\tau = n$ one needs only to observe $(X_n)_{n\in\mathbb{N}}$ up to time n.

Definition 1.2.2. The *hitting time* of a Markov chain $(X_n)_{n\in\mathbb{N}}$ of a subset $A \subset I$ is the r.v.

$$T_A(w) = \inf\{n \geq 1 \ ; \ X_n(w) \in A\}.$$

When $A = \{i\}$, we use the notation $T_i = T_{\{i\}}$.

Example 1.2.3.

1. *Hitting time:* T_i is a stopping time:

$$\{T_i = n\} = \{X_0 \neq i, X_1 \neq i, \dots, X_{n-1} \neq i, X_n = i\}.$$

   In general, T_A is a stopping time.

2. *Last exit time:* in general, $L_A = \sup\{n \geq 0 \ ; \ X_n \in A\}$ is not a stopping time since $\{L_A = n\}$ depends on the values of $(X_{n+m})_{m\in\mathbb{N}}$.

Remark 1.2.4. It is easy to prove that

$$p_{ij}^{(n)} = \sum_{m=1}^{n} \mathbb{P}(T_j = m \mid X_0 = i) p_{jj}^{(n-m)} \tag{1.4}$$

If the state j is absorbing, then $p_{ij}^{(n)} = \mathbb{P}(T_j \leq n \mid X_0 = i)$, which follows from the Equation (1.4) and from the fact $p_{jj}^{(n-m)} = 1$, since j is absorbing.

The following definition states the meaning of the information contained in the sequence $X_0, \dots, X_\tau$. Since τ is random, it is not straightforward to generalize the definition discussed in Remark 1.1.14.

Definition 1.2.5. If τ is a stopping time for the sequence $(X_n)_{n\in\mathbb{N}}$, then we define

$$\sigma(X_0, \dots, X_\tau) = \{B \in \mathcal{F} \ ; \ B \cap \{\tau = n\} \in \sigma(X_0, \dots, X_n), \ \forall \ n \in \mathbb{N}\}.$$

That is, if $\tau = n$, the set B must be in $\sigma(X_0, \dots, X_n)$. It is very important the fact τ is a stopping time, since this implies that $\{\tau = n\}$ is in $\sigma(X_0, \dots, X_n)$.

Theorem 1.2.6 (Strong Markov Property)**.** *Let $(X_n)_{n\in\mathbb{N}}$ be Markov(λ, P) and τ a stopping time for $(X_n)_{n\in\mathbb{N}}$. Then, given $\tau < +\infty$ and $X_\tau = i$, $(X_{\tau+n})_{n\in\mathbb{N}}$ is Markov(δ_i, P) and it is independent of $X_0, \dots, X_\tau$.*

Proof. Notice that if we prove that for any $A \in \sigma(X_0, \dots, X_\tau)$, we have

$$\begin{aligned}&\mathbb{P}(\{X_\tau = i_0, \dots, X_{\tau+n} = i_n\} \cap A \mid \tau < +\infty, X_\tau = i)\\ &= \delta_i(i_0) p_{i_0,i_1} \cdots p_{i_{n-1},i_n} \mathbb{P}(A \mid \tau < +\infty, X_\tau = i),\end{aligned}$$

the theorem is proved. But the equation above follows from:

$$\begin{aligned}&\mathbb{P}(\{X_\tau = i_0, \dots, X_{\tau+n} = i_n\} \cap A \cap \{\tau = m\} \cap \{X_\tau = i\})\\ &= \mathbb{P}(\{X_m = i_0, \dots, X_{m+n} = i_n\} \cap \underbrace{A \cap \{\tau = m\} \cap \{X_m = i\}}_{\in \sigma(X_0, \dots, X_m)})\\ &= \mathbb{P}(X_0 = i_0, \dots, X_n = i_n \mid X_0 = i) \mathbb{P}(A \cap \{\tau = m\} \cap \{X_\tau = i\}),\end{aligned}$$

summing m and dividing by $\mathbb{P}(\tau < +\infty, X_\tau = i)$. □

1.3 Class Structure; Transient and Recurrent States

Definition 1.3.1. We say $j \in I$ is accessible from $i \in I$ (and write $i \to j$) if

$$\mathbb{P}\left(\bigcup_{n=0}^{+\infty} \{X_n = j\} \;\middle|\; X_0 = i \right) = \mathbb{P}(X_n = j \text{ for some } n \geq 0 \mid X_0 = i) > 0.$$

It is easy to see that if $i \to j$ and $j \to k$, then $i \to k$.

We say that i and j communicate with each other (and write $i \leftrightarrow j$) if $i \to j$ and $j \to i$. Notice that $\leftrightarrow$ is a equivalence relation in I.

We call a subset C of I a *class*. A class C is said to be *communicating* if

$$\forall\ i, j \in C, i \leftrightarrow j.$$

A class C is said to be *closed* if

$$i \in C, i \to j \Rightarrow j \in C.$$

A class is said to be *irreducible* if the class is closed and communicating. Moreover, notice that if $i \in I$ is an absorbing state, then the set $\{i\}$ is closed.

Proposition 1.3.2. *The following are equivalent:*

(i) $i \to j$.

(ii) There exist $i_1, \dots, i_{n-1}$ *in* I *such that*

$$p_{i,i_1} \cdots p_{i_{n-1},j} > 0.$$

(iii) There exists $n \in \mathbb{N}$ *such that* $p_{ij}^{(n)} > 0$.

Proof. Notice that

$$p_{ij}^{(n)} \leq \mathbb{P}\left(\bigcup_{m=0}^{+\infty} \{X_m = j\} \;\middle|\; X_0 = i \right) \leq \sum_{m=0}^{+\infty} p_{ij}^{(m)},$$

which proves the equivalence between (i) and (iii). To see the equivalence between (ii) and (iii), notice that

$$p_{ij}^{(n)} = \sum_{i_1, \dots, i_{n-1} \in I} p_{i,i_1} \cdots p_{i_{n-1},j}.$$

□

Let $(X_n)_{n\in\mathbb{N}}$ be Markov(λ, P) and define

$$\rho_{ij} = \mathbb{P}(T_j < +\infty \mid X_0 = i) = 1 - \mathbb{P}(T_j = +\infty \mid X_0 = i), \tag{1.5}$$

i.e. ρ_{ij} is the probability the chain, starting at i, hit the state j at some point in the future. In particular, ρ_{ii} is the probability of the chain to visit its initial state i again in the future.

Definition 1.3.3. A state i is

- *recurrent* if $\rho_{ii} = 1$;
- and *transient* if $\rho_{ii} < 1$.

Moreover, we say a chain is recurrent (or transient) if all its states are recurrent (or transient).

Remark 1.3.4. Notice that if i is a transient state, with probability $1 - \rho_{ii}$, the chain will never visit i again. Additionally, if the state i is absorbing, then it will clearly be recurrent. Moreover, $i \to j$ is equivalent to $\rho_{ij} > 0$.

Remark 1.3.5. Define

$$f_{ij}^{(n)} = \mathbb{P}(T_j = n \mid X_0 = i) = \mathbb{P}(X_1 \neq j, \dots, X_{n-1} \neq j, X_n = j \mid X_0 = i).$$

Then

$$\rho_{ij} = \sum_{n=1}^{+\infty} f_{ij}^{(n)}.$$

This could be used to compute ρ_{ij}. Another way would be

$$\rho_{ij} = 1 - \lim_{m\to+\infty} \mathbb{P}(T_j \geq m \mid X_0 = i).$$

It is very intuitive that once a chain return to its initial state it behaves as the chain had started again. We will make this idea precise in what follows. Fix a state $i \in I$ and define the rth passage of the chain through the state i as

$$T_i^{(r)} = \inf\{n \geq T_i^{(r-1)} + 1 \ ; \ X_n = i\},$$

with $T_i^{(0)} = 0$ and $T_i^{(1)} = T_i$. These are clearly stopping times. Moreover, we define the rth excursion to the state i as

$$S_i^{(r)} = T_i^{(r)} - T_i^{(r-1)},$$

if $T_i^{(r-1)} < +\infty$, and 0, otherwise. The next lemma proves the following intuitive idea: "when the chain returns to j it is as if it has just restarted at j".

Lemma 1.3.6. *For $r \geq 1$, given $T_i^{(r-1)} < +\infty$, $S_i^{(r)}$ is independent of $X_0, \dots, X_{T_i^{(r-1)}}$ and*

$$\mathbb{P}(S_i^{(r)} = n \mid T_i^{(r-1)} < +\infty) = \mathbb{P}(T_i = n \mid X_0 = i).$$

This implies that the sequence of r.v's $(S_i^{(r)})_{r\geq 1}$ is iid with the same distribution as $T_i \mid X_0 = i$.

Proof. Define $\tau = T_i^{(r-1)}$, which is a stopping time. We will assume from now on that $\tau < +\infty$. Notice that $X_\tau = i$ and, by the strong Markov property, $(X_{\tau+n})_{n\in\mathbb{N}}$ is Markov(δ_i, P) and independent of $X_0, \dots, X_\tau$. Furthermore,

$$S_i^{(r)} = \inf\{n \geq 1 \ ; \ X_{\tau+n} = i\},$$

and hence $S_i^{(r)}$ is the hitting time of i of the Markov chain $(X_{\tau+n})_{n\in\mathbb{N}}$. Hence,

$$\mathbb{P}(S_i^{(r)} = n \mid T_i^{(r-1)} < +\infty) = \mathbb{P}(T_i = n \mid X_0 = i).$$

$\square$

1.3.1 Number of Visits and its Relation to Recurrence

Denote N_i the number of times the chain visits the state i. Precisely, we have

$$N_i = \sum_{n=1}^{+\infty} 1_{\{i\}}(X_n). \tag{1.6}$$

Notice that $N_i \geq 1 \Leftrightarrow T_i < +\infty$ and then, $\rho_{ij} = \mathbb{P}(N_j \geq 1 \mid X_0 = i)$.

By the Markov property, the probability of a Markov chain, starting at i, hit j at m (and not before) and then, after n steps, hit j for the second time is $\mathbb{P}(T_j = m \mid X_0 = i)\mathbb{P}(T_j = n \mid X_0 = j)$. Thus,

$$\begin{aligned}
\mathbb{P}(N_j \geq 2 \mid X_0 = i) &= \sum_{m=1}^{+\infty}\sum_{n=1}^{+\infty} \mathbb{P}(T_j = m \mid X_0 = i)\mathbb{P}(T_j = n \mid X_0 = j) \\
&= \left(\sum_{m=1}^{+\infty} \mathbb{P}(T_j = m \mid X_0 = i)\right)\left(\sum_{n=1}^{+\infty} \mathbb{P}(T_j = n \mid X_0 = j)\right) \\
&= \rho_{ij}\rho_{jj}
\end{aligned}$$

Moreover, one can easily see that

$$\mathbb{P}(N_j \geq m \mid X_0 = i) = \rho_{ij}\rho_{jj}^{m-1},$$

which implies

$$\mathbb{P}(N_j = m \mid X_0 = i) = \begin{cases} \rho_{ij}(1-\rho_{jj})\rho_{jj}^{m-1}, & m > 0, \\ 1 - \rho_{ij}, & m = 0. \end{cases}$$

Definition 1.3.7 (Average Number of Visits)**.** The *average number of visits* is defined as

$$g_{ij} = \mathbb{E}[N_j \mid X_0 = i].$$

Notice

$$g_{ij} = \sum_{n=1}^{+\infty} \mathbb{E}[1_{\{j\}}(X_n) \mid X_0 = i] = \sum_{n=1}^{+\infty} p_{ij}^{(n)}.$$

We hence have the following result:

Theorem 1.3.8.

(i) Let j be a transient state. Then, for any $i \in I$,

- $\mathbb{P}(N_j < +\infty \mid X_0 = i) = 1$
- $g_{ij} = \dfrac{\rho_{ij}}{1-\rho_{jj}} < +\infty$

(ii) However, if j is a recurrent state, we find

- $\mathbb{P}(N_j = +\infty \mid X_0 = j) = 1$
- $g_{jj} = +\infty$
- $\mathbb{P}(N_j = +\infty \mid X_0 = i) = \mathbb{P}(T_j < +\infty \mid X_0 = i) = \rho_{ij}$*, for any* $i \in I$
- $\rho_{ij} = 0 \Rightarrow g_{ij} = 0$
- $\rho_{ij} > 0 \Rightarrow g_{ij} = +\infty$

Remark 1.3.9. If j is transient, then the chain will visit it only a finite number of times, regardless where the chain starts. Additionally, the expected number of visits of j is finite.

In the case where j is recurrent, when the chain starts at j, it will visit it infinitely many times.

Proof.

(i) Notice that, $0 \le \rho_{jj} < 1$ and then

$$\mathbb{P}(N_j = +\infty \mid X_0 = i) = \lim_{m\to+\infty} \mathbb{P}(N_j \ge m \mid X_0 = i) = \lim_{m\to+\infty} \rho_{ij}\rho_{jj}^{m-1} = 0.$$

Moreover,

$$g_{ij} = \mathbb{E}[N_j \mid X_0 = i] = \sum_{m=1}^{+\infty} \mathbb{P}(N_j \ge m \mid X_0 = i) = \sum_{m=1}^{+\infty} \rho_{ij}\rho_{jj}^{m-1} = \frac{\rho_{ij}}{1-\rho_{jj}}$$

(ii) In this case,

$$\mathbb{P}(N_j = +\infty \mid X_0 = i) = \lim_{m\to+\infty} \mathbb{P}(N_j \ge m \mid X_0 = i) = \lim_{m\to+\infty} \rho_{ij}\cancelto{1}{\rho_{jj}^{m-1}} = \rho_{ij}.$$

This clearly implies $\rho_{ij} > 0 \Rightarrow g_{ij} = +\infty$ (a non-negative r.v. with positive probability to attain infinity has infinite expectation). Moreover, if $\rho_{ij} = 0$, then $\mathbb{P}(T_j = m \mid X_0 = i) = 0$, for any $m \in \mathbb{N}$. This implies, by Equation (1.4), that $p_{ij}^{(n)} = 0$, for any $n \in \mathbb{N}$. Then, $g_{ij} = 0$.

$\square$

Remark 1.3.10. Notice that, if j is transient, then $g_{ij} = \sum_{n=1}^{+\infty} p_{ij}^{(n)} < +\infty$, which implies

$$\lim_{n\to+\infty} p_{ij}^{(n)} = 0, \ \forall\, i \in I.$$

Recurrence and transient could be reformulated as follows:

Theorem 1.3.11.

- *The state j is recurrent if $\sum_{n=1}^{+\infty} p_{jj}^{(n)} = +\infty$. Moreover, if $\rho_{ij} > 0$, $\sum_{n=1}^{+\infty} p_{ij}^{(n)} = +\infty$.*
- *The state j is transient if $\sum_{n=1}^{+\infty} p_{jj}^{(n)} < +\infty$. Moreover, $\sum_{n=1}^{+\infty} p_{ij}^{(n)} < +\infty$.*

Theorem 1.3.12. *Let i be a recurrent state of a given chain and assume that $i \to j$. Then j is also recurrent and $\rho_{ij} = \rho_{ji} = 1$, which implies $j \to i$.*

Proof. Assume that $i \ne j$, otherwise there is nothing to prove. Notice that, since $i \to j$, we have

$$\rho_{ij} = \mathbb{P}(T_j < +\infty \mid X_0 = i) > 0.$$

Define then

$$n_0 = \min\{n \in \mathbb{N} \ ; \ \mathbb{P}(T_j = n \mid X_0 = i) > 0\}$$

and note that $n_0 > 0$. Moreover, by Equation (1.4), $p_{ij}^{(m)} = 0$, for any $1 \le m < n_0$. Since $p_{ij}^{(n_0)} > 0$, there exist $j_1, \dots, j_{n_0-1} \in I$ such that

$$\mathbb{P}(X_1 = j_1, \dots, X_{n_0-1} = j_{n_0-1}, X_{n_0} = j \mid X_0 = i) = p_{i,j_1} \cdots p_{j_{n_0-1},j} > 0.$$

Notice that none of $j_1, \dots, j_{n_0-1} \in I$ are equal i or j, otherwise there would be a faster way to go from i to j, contradicting the minimality of n_0. Let us show now that $\rho_{ji} = 1$. By contradiction, assume $\rho_{ji} < 1$. Then, the chain has probability $1 - \rho_{ji}$ of start at j and never pass through i. Moreover, the chain has positive probability of start at i, visit $j_1, \dots, j_{n_0-1}, j$ and then never pass through i after n_0:

$$p_{i,j_1} \cdots p_{j_{n_0-1},j}(1 - \rho_{ji}) > 0.$$

Hence, the chain would never return to i with positive probability, which contradicts the fact i is recurrent. Thus, $\rho_{ji} = 1$.

Let us now verify that j is also recurrent. Note there exists n_1 such that $p_{ji}^{(n_1)} > 0$. Then

$$p_{jj}^{(n_0+n_1+n)} \ge p_{ji}^{(n_1)} p_{ii}^{(n)} p_{ij}^{(n_0)}.$$

Therefore,

$$\begin{aligned}
g_{jj} &= \sum_{k=1}^{+\infty} p_{jj}^{(k)} \geq \sum_{n=1}^{+\infty} p_{jj}^{(n_0+n_1+n)} \\
&\geq \sum_{n=1}^{+\infty} p_{ji}^{(n_1)} p_{ii}^{(n)} p_{ij}^{(n_0)} = p_{ji}^{(n_1)} p_{ij}^{(n_0)} \overbrace{\sum_{n=1}^{+\infty} p_{ii}^{(n)}}^{g_{ii}} = +\infty,
\end{aligned}$$

and then j is recurrent.

We have shown that if i is recurrent and $i \to j$, then j is recurrent and $\rho_{ji} = 1$. But, this implies that $j \to i$. Hence, by the lengthy argument above, $\rho_{ij} = 1$. ☐

Remark 1.3.13. An equivalent definition of closed class is: $C \subset I$ is a closed class if

$$\rho_{ij} = 0, \ \forall i \in C \text{ and } j \notin C.$$

Theorem 1.3.14. *If C is an irreducible class, then all its elements are either recurrent or transient. If C is finite, then it must be recurrent. In particular, an irreducible Markov chain with finite state space is necessarily recurrent.*

Proof. We only need to prove that a finite irreducible class is recurrent. By contradiction, assume that some state in C is transient. Then, by irreducibility, all the states are transient, which implies, $p_{ij}^{(n)} \to 0$, for any $j \in C$, when $n \to +\infty$. Hence,

$$1 = \lim_{n\to+\infty} \sum_{j\in C} \mathbb{P}(X_n = j \mid X_0 = i) = \lim_{n\to+\infty} \sum_{j\in C} p_{ij}^{(n)} = \sum_{j\in C} \lim_{n\to+\infty} p_{ij}^{(n)} = 0,$$

which is clearly a contradiction. Therefore, all its states are recurrent. ☐

Theorem 1.3.15. *Denote by $I_T \subset I$ and $I_R \subset I$ the sets of transient and recurrent states, respectively. Notice $I = I_T \cup I_R$ and this union is disjoint. If the set I_R is non-empty, then I_R is the (finite or countable) disjoint union of irreducible classes.*

Proof. Consider $i \in I_R$ and define $C_i = \{j \in I_R \ ; \ \rho_{ij} > 0\}$. Since i is recurrent, $i \in C_i$. Let us verify that C_i is closed and communicative.

- C_i is closed: take $j \in C_i$ and assume that $j \to k$. We have to show that $k \in C_i$. Since j is recurrent, so is k. Moreover, it is easy to see that $i \to k$, which implies $k \in C_i$.
- C_i is communicative: take $j, k \in C_i$. We will show that $j \to k$. Since i is recurrent and $i \to j$, we have that $j \to i$. Hence, $j \to i \to k \Rightarrow j \to k$.

We will prove now that if C and D are irreducible classes of I_R, then either C and D are disjoints or they are the same set. By contradiction, assume $C \cap D \neq \emptyset$ and let $k \in C \cap D$. So, for any $j \in C$, $k \to j$. Now, since D is closed, $k \in D$ and $k \to j$, we must have $j \in D$. Hence, if $C \cap D \neq \emptyset$, we have $C = D$.

Therefore, one may write I_R is the disjoint union of some of the C_i's. ☐

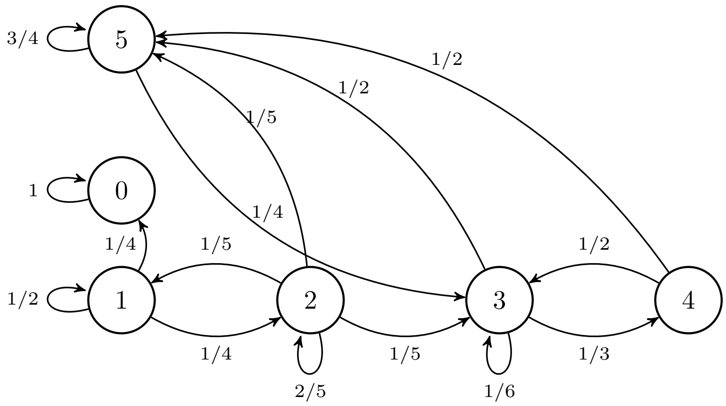
Remark 1.3.16. The theorem above helps us to understand the behavior of a Markov chain. If the chain starts at some of the C_i's, it will stay at C_i forever and it will visit all its states infinitely many times. Now, if it starts at I_T, it will leave I_T to enter some of C_i's.

1.3.2 Examples

Example 1.3.17. Let us consider a Markov chain with $I = \{0, 1, 2, 3, 4, 5\}$ and transition probability

$$P = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
1/4 & 1/2 & 1/4 & 0 & 0 & 0 \\
0 & 1/5 & 2/5 & 1/5 & 0 & 1/5 \\
0 & 0 & 0 & 1/6 & 1/3 & 1/2 \\
0 & 0 & 0 & 1/2 & 0 & 1/2 \\
0 & 0 & 0 & 1/4 & 0 & 3/4
\end{bmatrix}.$$

This chain can be represented by



Let us analyze the states of this chain:

- 0 is an absorbing state, and hence, recurrent;
- $\{3, 4, 5\}$ is a finite, irreducible class, and hence, recurrent;
- 1 and 2 are transient states.

Example 1.3.18 (2 State Markov Chain)**.** If $0 < p, q < 1$, it is easy to see that the chain is recurrent (finite number of states and irreducible).

Example 1.3.19 (Birth and Death Chains)**.** Let us assume $p_i > 0$ and $q_i > 0$. In this case, the chain is clearly irreducible. We want to see if it is recurrent or transient. If $d < +\infty$, it is obvious the chain will be recurrent. Let us then assume $d = +\infty$ and compute $\rho_{00} = \mathbb{P}(T_0 < +\infty \mid X_0 = 0)$.

Notice this chain can only jump one state (to the left or to the right) at each time step. Then

$$\rho_{00} = r_0 + p_0 \mathbb{P}(T_0 < +\infty \mid X_1 = 1) = r_0 + p_0 \rho_{10}. \tag{1.7}$$

Let us then compute ρ_{10}. Assuming we start at 1 implies $1 \leq T_2 < T_3 < \cdots < T_n < \cdots$. Moreover, $T_n \geq n - 1$, and then $T_n \to +\infty$ a.s. when $n \to +\infty$. Hence, the sets $A_n = \{T_0 < T_n\}$ are increasing and

$$\mathbb{P}(T_0 < +\infty \mid X_0 = 1) = \lim_{n \to +\infty} \mathbb{P}(T_0 < T_n \mid X_0 = 1)$$

Therefore, we should compute $u_i = \mathbb{P}(T_j < T_k \mid X_0 = i)$, where $j < i < k$. This is the probability of, starting at $i \in (j, k)$, hitting j before k. It is easy to see that $u_{j-1} = 1$, $u_{k+1} = 0$ and, for $j \leq m \leq k$,

$$u_m = q_m u_{m-1} + r_m u_m + p_m u_{m+1}.$$

Since $r_m = 1 - p_m - q_m$, we find

$$u_{m+1} - u_m = \frac{q_m}{p_m}(u_m - u_{m-1}).$$

Using recurrence:

$$\begin{aligned} u_{m+1} - u_m &= \frac{q_m}{p_m}(u_m - u_{m-1}) = \frac{q_m}{p_m}\frac{q_{m-1}}{p_{m-1}}(u_{m-1} - u_{m-2}) \\ &= \cdots = \frac{q_m}{p_m}\frac{q_{m-1}}{p_{m-1}} \cdots \frac{q_j}{p_j}(u_j - u_{j-1}). \end{aligned}$$

Define $\gamma_0 = 1$ and

$$\gamma_m = \frac{q_1 \cdots q_m}{p_1 \cdots p_m}$$

Then

$$u_{m+1} - u_m = \frac{\gamma_m}{\gamma_{j-1}}(u_j - u_{j-1}).$$

Summing from $m = j-1$ to k, we get

$$\underbrace{u_{k+1} - u_{j-1}}_{=-1} = \sum_{m=j-1}^{k} (u_{m+1} - u_m) = \frac{u_j - u_{j-1}}{\gamma_{j-1}} \sum_{m=j-1}^{k} \gamma_m \Rightarrow u_j - u_{j-1} = -\frac{\gamma_{j-1}}{\sum_{m=j-1}^{k} \gamma_m}.$$

Hence,

$$u_{l+1} - u_l = -\frac{\gamma_l}{\sum_{m=j-1}^{k} \gamma_m}.$$

Now, summing from $l = i$ to k, we finally find (using $u_k = 0$)

$$\mathbb{P}(T_j < T_k \mid X_0 = i) = u_i = \frac{\sum_{m=i}^{k} \gamma_m}{\sum_{m=j-1}^{k} \gamma_m}$$

or

$$\mathbb{P}(T_j > T_k \mid X_0 = i) = \frac{\sum_{m=j-1}^{i-1} \gamma_m}{\sum_{m=j-1}^{k} \gamma_m}.$$

In particular,

$$\begin{aligned}\rho_{10} = \mathbb{P}(T_0 < +\infty \mid X_0 = 1) &= \lim_{n\to+\infty} \mathbb{P}(T_0 < T_n \mid X_0 = 1) \\ &= \lim_{n\to+\infty} \frac{\sum_{m=1}^{n} \gamma_m}{\sum_{m=0}^{n} \gamma_m} = \frac{\sum_{m=1}^{+\infty} \gamma_m}{\sum_{m=0}^{+\infty} \gamma_m} \\ &= 1 - \frac{1}{\sum_{m=0}^{+\infty} \gamma_m}.\end{aligned}$$

We are then able to prove the following result:

Theorem 1.3.20. *An irreducible birth and death chain is recurrent if and only if*

$$\sum_{m=1}^{+\infty} \frac{q_1 \cdots q_m}{p_1 \cdots p_m} = \sum_{m=1}^{+\infty} \gamma_m = +\infty \tag{1.8}$$

Proof. Firstly, notice

$$\rho_{10} = 1 - \frac{1}{\sum_{m=0}^{+\infty} \gamma_m} = 1 \Leftrightarrow \sum_{m=1}^{+\infty} \gamma_m = +\infty.$$

In words, $\rho_{10} = 1$ is equivalent to (1.8). So, the implication ($\Rightarrow$) follows easily from above (since recurrence of the chain implies $\rho_{10} = 1$). To verify the opposite implication assume (1.8) is true. Then, $\rho_{10} = 1$. By Equation (1.7), we conclude

$$\rho_{00} = r_0 + p_0\rho_{10} = r_0 + p_0 = 1 \Rightarrow 0 \text{ is a recurrent state.}$$

Since the chain is irreducible, the chain itself is recurrent. □

Example 1.3.21 (Branching Process)**.** Denote

$$\rho = \rho_{10} = \mathbb{P}(T_0 < +\infty \mid X_0 = 1).$$

We have argued that $\rho_{i0} = \rho^i$ and that 0 is a absorbing state. Our intuition tells us that this chain either hits 0 or goes to $+\infty$. Thus,

$$\mathbb{P}\left(\lim_{n\to+\infty} X_n = +\infty \; | X_0 = i\right) = 1 - \rho^i.$$

Hence, every state different than 0 is transient. We will now study ρ. Notice that

$$\rho = \rho_{10} = p_{10} + \sum_{i=1}^{+\infty} p_{1i}\rho_{i0} = f(0) + \sum_{i=1}^{+\infty} f(i)\rho^i,$$

where $f(i) = \mathbb{P}(\xi = i)$. Define then, for $0 \leq t \leq 1$,

$$\Phi(t) = \sum_{i=0}^{+\infty} f(i)t^i$$

Hence, ρ is a solution of $t = \Phi(t)$ in the interval $[0,1]$. Moreover, it is easy to see that $\Phi(1) = 1$, $\Phi(0) = f(0)$ and

$$\Phi'(t) = \sum_{i=1}^{+\infty} if(i)t^{i-1}.$$

This gives $\Phi'(1) = \mathbb{E}[\xi] = \mu$ and that Φ' is non-decreasing in $[0,1)$. Additionally,

$$\Phi''(t) = \sum_{i=2}^{+\infty} i(i-1)f(i)t^{i-2},$$

which is non-negative in $[0,1)$.

We assume $f(i) < 1$, for any $i \in \mathbb{N}$, since the case $f(i) = 1$ for some i is trivial. Moreover, we assume $f(0) > 0$, otherwise it is straightforward that $\rho = 0$. Therefore, one can notice the following

- $\mu < 1$: Since $\Phi'(1) < 1$ and Φ' is non-decreasing, we have $\Phi'(t) < 1$ in $[0,1]$.
- $\mu = 1$: In this case, $\Phi'(1) = 1$ and we must have $f(i) > 0$, for some $i \geq 2$. Otherwise, $\mu < 1$. Then, Φ' is strictly increasing, which implies, as in the case $\mu < 1$, that $\Phi'(t) < 1$ in $[0,1)$.
- $\mu > 1$: By the continuity of Φ' and from $\Phi'(1) > 1$ and $\Phi'(0) = f(1) < 1$, there must exist $\rho_0 \in (0,1)$ such that $\Phi'(\rho_0) = 1$. Since $\Phi''(t) > 0$ in $(0,1)$, this ρ_0 is unique and $\Phi'(t) < 1$, for $0 < t < \rho_0$, and $\Phi'(t) > 1$, for $\rho_0 < t < 1$.

Now, we are able to conclude that:

- $\mu \leq 1$: We have seen that in this case $\Phi'(t) < 1$ in $[0,1)$. Then

  $$\frac{d}{dt}(\Phi(t) - t) < 0 \text{ in } [0,1)$$

  Since $\Phi(1) - 1 = 0$, we must have $\Phi(t) - t > 0$ in $[0,1)$. Hence, $\Phi(t) = t$ has no solution in $[0,1)$ and then $\rho = 1$. Therefore, when $\mu \leq 1$, the extinction is inevitable.
- $\mu > 1$: In this case, we have discovered that there exists a unique $\rho_0 \in (0,1)$ such that $\Phi'(\rho_0) = 1$. Then,

  $$\frac{d}{dt}(\Phi(t) - t) \begin{cases} < 0, & \text{for } t \in (0, \rho_0), \\ = 0, & \text{for } t = \rho_0, \\ > 0, & \text{for } (\rho_0, 1). \end{cases}$$

  This implies there exists $0 < \rho < \rho_0$ such that $\Phi(\rho) = \rho$ and that $\Phi(t) = t$ cannot have a solution for $t \neq \rho$. Hence, when $\mu > 1$, the probability of extinction is strictly less than 1.

Example 1.3.22 (Simple Random Walk in $\mathbb{Z}$)**.** Let us assume $p = f(1) > 0, f(0) = 0$ and $f(-1) = 1 - p = q$, which implies the chain is irreducible. Let us then analyze the state 0 and see if it is either transient or recurrent (it will depend on p). For this, we will use the test

$$\sum_{n=0}^{+\infty} p_{00}^{(n)} \lesseqgtr +\infty.$$

Firstly, notice that $p_{00}^{(2n+1)} = 0$. Now, to compute $p_{00}^{(2n)}$, notice that the chain must have given n steps to the right and n to the left. Hence,

$$p_{00}^{(2n)} = \binom{2n}{n} p^n q^n.$$

We will use now Stirling's formula (that describes the asymptotic behavior of the factorial):

$$\lim_{n\to+\infty} \frac{n!}{\sqrt{2\pi n}(n/e)^n} = 1$$

Then, the asymptotic behavior of $p_{00}^{(2n)}$ is given by

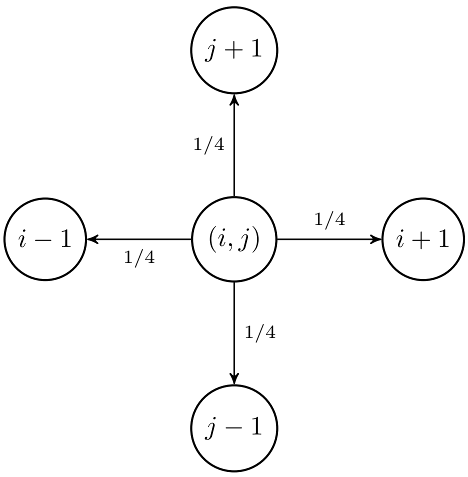
$$\begin{aligned}
\lim_{n\to+\infty} p_{00}^{(2n)} &= \lim_{n\to+\infty} \binom{2n}{n} p^n q^n = \lim_{n\to+\infty} \frac{(2n)!}{(n!)^2} p^n q^n \\
&= \lim_{n\to+\infty} \frac{(2n)!}{\sqrt{2\pi(2n)}(2n/e)^{2n}} \frac{(\sqrt{2\pi n}(n/e)^n)^2}{(n!)^2} \frac{\sqrt{2\pi(2n)}(2n/e)^{2n}}{(\sqrt{2\pi n}(n/e)^n)^2} p^n q^n \\
&= \cancelto{1}{\lim_{n\to+\infty} \frac{(2n)!}{\sqrt{2\pi(2n)}(2n/e)^{2n}}} \cancelto{1}{\left(\lim_{n\to+\infty} \frac{\sqrt{2\pi n}(n/e)^n}{n!} \right)^2} \\
& \lim_{n\to+\infty} \frac{\sqrt{2\pi(2n)}(2n/e)^{2n}}{2\pi n \ (n/e)^{2n}} p^n q^n = \lim_{n\to+\infty} \frac{(4pq)^n}{\sqrt{\pi n}}
\end{aligned}$$

Hence, we find

$$\sum_{n=0}^{+\infty} p_{00}^{(n)} \begin{cases} = +\infty, & \text{if } p = 1/2, \\ < +\infty, & \text{if } p \neq 1/2. \end{cases}$$

Therefore, the simple random walk is transient if $p \neq 1/2$ and recurrent if $p = 1/2$.

Example 1.3.23 (Simple Symmetric Random Walk in $\mathbb{Z}^2$)**.**



One can show that the projections of X_n on the lines $y = \pm x$ are independent simple symmetric random walks in $\frac{1}{\sqrt{2}}\mathbb{Z}$. Notice that $X_n = 0$ happens if and only if each of these projections hit 0 as well. Then,

$$p_{00}^{(2n)} = \left(\binom{2n}{n} \left(\frac{1}{2}\right)^{2n} \right)^2.$$

By Stirling's formula, we can see that

$$\lim_{n\to+\infty} \frac{p_{00}^{(2n)}}{\frac{1}{\pi n}} = 1$$

which clearly implies that

$$\sum_{n=0}^{+\infty} p_{00}^{(n)} = +\infty.$$

So, the simple symmetric random walk in two dimensions is recurrent. The same argument used in the one-dimensional case (and that we will use in the three-dimensional case) could be used here as well.

Example 1.3.24 (Simple Symmetric Random Walk in $\mathbb{Z}^3$)**.** A simple random walk can step up and down, north and south, and east and west, each with probability 1/6. As before, it is possible to hit 0 only in an even number of steps. In three dimensions, we must have the same numbers of steps for up and down, north and south, and east and west. Let this numbers be i, j and k. We must have $i + j + k = n$ in other to come back to 0 in $2n$ steps. Hence

$$\begin{aligned} p_{00}^{(2n)} &= \sum_{\substack{i,j,k\geq 0\\ i+j+k=n}} \frac{(2n)!}{(i!j!k!)^2}\left(\frac{1}{6}\right)^{2n} \\ &= \binom{2n}{n}\left(\frac{1}{2}\right)^{2n} \sum_{\substack{i,j,k\geq 0\\ i+j+k=n}} \binom{n}{i\ j\ k}^2\left(\frac{1}{3}\right)^{2n}. \end{aligned}$$

By the definition of the multinomial coefficient, we have

$$\sum_{\substack{i,j,k\geq 0\\ i+j+k=n}} \binom{n}{i\ j\ k} = 3^n.$$

Moreover, when $n = 3m$, we get

$$\binom{n}{i\ j\ k} \leq \binom{n}{m\ m\ m}.$$

Hence, with $n = 3m$, we find

$$\begin{aligned} \sum_{\substack{i,j,k\geq 0\\ i+j+k=n}} \binom{n}{i\ j\ k}^2\left(\frac{1}{3}\right)^{2n} &\leq \binom{n}{m\ m\ m}\left(\frac{1}{3}\right)^{n} \sum_{\substack{i,j,k\geq 0\\ i+j+k=n}} \binom{n}{i\ j\ k}\left(\frac{1}{3}\right)^{n} \\ &= \binom{n}{m\ m\ m}\left(\frac{1}{3}\right)^{n}. \end{aligned}$$

This implies

$$p_{00}^{(6m)} \leq \binom{2n}{n}\left(\frac{1}{2}\right)^{2n}\binom{n}{m\ m\ m}\left(\frac{1}{3}\right)^{n}.$$

Now, by Stirling's formula,

$$\lim_{n\to+\infty} \frac{\binom{2n}{n}\left(\frac{1}{2}\right)^{2n}\binom{n}{m\ m\ m}\left(\frac{1}{3}\right)^{n}}{\frac{1}{2(2\pi)^{3/2}}\left(\frac{6}{n}\right)^{3/2}} = 1.$$

Hence,

$$\sum_{m=0}^{+\infty} p_{00}^{(6m)} < +\infty.$$

Additionally, $p_{00}^{(6m)} \geq (1/6)^2 p_{00}^{(2(3m-1))}$ and $p_{00}^{(6m)} \geq (1/6)^4 p_{00}^{(2(3m-2))}$, which combined gives

$$\sum_{n=0}^{+\infty} p_{00}^{(2n)} < +\infty.$$

Therefore, differently from the one and two-dimensional cases, the simple symmetric random walk in three dimensions is transient. One can prove that this is true for any dimension bigger or equal than 3. Furthermore, the same is observed in the continuous-time version of the simple symmetric random walk, the Brownian motion.

Example 1.3.25 (Martingale)**.** We say a Markov(λ, P) chain with state space $I = \{0, \dots, d\}$ is a *martingale* if

$$\sum_{j=0}^{d} j\ p_{ij} = i,\ \forall\ i = 0, \dots, d. \tag{1.9}$$

This means

$$\begin{aligned}
&\mathbb{E}[X_{n+1} \,|X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0] \\
&= \sum_{j=0}^{d} j\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \\
&= \sum_{j=0}^{d} j \; p_{ij} = i = X_n.
\end{aligned}$$

In words, the expected value of X_{n+1} given $X_0, \ldots, X_n$ is equal X_n. Later we will study martingale with a lot more details. We would like to point out that obviously being a Markov chain is not necessary in any way for a process to be a martingale.

Let us now study recurrence and transience of states of a martingale Markov chain. Notice that, by Equation (1.9), we have

$$\sum_{j=0}^{d} j \; p_{0j} = 0 \Rightarrow p_{0j} = 0, \;\forall\, j = 1, \ldots, d \text{ and } p_{00} = 1.$$

Then, 0 is absorbing for any martingale Markov chain with state space $I = \{0, \ldots, d\}$. Additionally,

$$\sum_{j=0}^{d} j \; p_{dj} = d \Rightarrow p_{dj} = 0, \;\forall\, j = 1, \ldots, d-1 \text{ and } p_{dd} = 1,$$

then d must also be absorbing. To understand why $p_{dd} = 1$ notice that when a positive random variable has average equals its maximum value, it means this random variable is constant and equal its maximum value. Assume now that 0 and d are the only two absorbing states of this chain. Therefore, $\{1, \ldots, d-1\}$ must be transient, if any of these states communicates with 0 or d.

Let us now compute ρ_{id}. Firstly, notice that since we are assuming 0 and d are the only two absorbing states and that some of the other states communicate with either 0 or d, we conclude $\rho_{i0} + \rho_{id} = 1$. To compute ρ_{id}, we have

$$\begin{aligned}
\mathbb{E}[X_n \mid X_0 = i] &= \sum_{j=0}^{d} j \; \mathbb{P}(X_n = j \mid X_0 = i) = \sum_{j=0}^{d} j p_{ij}^{(n)} \\
&= \sum_{j=1}^{d-1} j p_{ij}^{(n)} + d \; \mathbb{P}(T_d \leq n, X_n = d \mid X_0 = i) \\
&= \sum_{j=1}^{d-1} j p_{ij}^{(n)} + d \; \mathbb{P}(T_d \leq n \mid X_0 = i),
\end{aligned}$$

where the last equality follows from the fact d is absorbing. Letting $n \to +\infty$ and using the fact that $\{1, \ldots, d\}$ are transient, we find

$$\lim_{n\to+\infty} \mathbb{E}[X_n \mid X_0 = i] = d \; \mathbb{P}(T_d < +\infty \mid X_0 = i) = d\rho_{id}.$$

On the other hand, since $(X_n)_{n\in\mathbb{N}}$ is a martingale, $\mathbb{E}[X_n \mid X_0 = i] = i$ (prove this). Hence,

$$\rho_{id} = \frac{i}{d} \text{ and } \rho_{i0} = 1 - \frac{i}{d}.$$

1.4 Stationary Distributions

Definition 1.4.1 (Stationary Distribution)**.** A probability distribution $(\pi_i)_{i\in I}$ is said to be a *stationary distribution* of a Markov chain $(X_n)_{n\in\mathbb{N}}$ with state space I and with transition probability P if

$$\sum_{i\in I} \pi_i p_{ij} = \pi_j, \;\forall\, j \in I.$$

Remark 1.4.2. In matrix notation, $\pi P = \pi$, i.e. π is a stationary distribution if π is a left eigenvector for the eigenvalue 1 of the matrix P and π is a probability distribution.

Definition 1.4.3 (Equilibrium Distribution)**.** A probability distribution $(\pi_i)_{i\in I}$ is said to be a *equilibrium distribution* of a Markov chain $(X_n)_{n\in\mathbb{N}}$ with state space I and with transition probability P if

$$\lim_{n\to+\infty} p_{ij}^{(n)} = \pi_j, \ \forall\, i,j \in I.$$

Stationary distributions can be understood as a initial probability distribution of the chain such that the distribution of X_n does not depend on n, for any $n \in \mathbb{N}$. Indeed, we have the following result:

Proposition 1.4.4. *The distribution of X_n is independent of n if and only if the initial distribution is stationary.*

Proof.

($\Leftarrow$) Assume the distribution of X_0 is π (which is stationary). Then

$$\begin{aligned}\mathbb{P}(X_n = j) &= \sum_{i\in I} \pi_i p_{ij}^{(n)} = \sum_{i\in I} \pi_i \sum_{\ell\in I} p_{i\ell} p_{\ell j}^{(n-1)} \\ &= \sum_{\ell\in I} p_{\ell j}^{(n-1)} \left(\sum_{i\in I} \pi_i p_{i\ell} \right) = \sum_{\ell\in I} \pi_\ell p_{\ell j}^{(n-1)} \\ &= \cdots = \sum_{\ell\in I} \pi_\ell p_{\ell j} = \pi_j,\end{aligned}$$

which does not depend on n. In matrix notation, this argument looks simpler

$$\pi P^n = \pi P(P^{n-1}) = \pi P^{n-1} = \cdots = \pi P = \pi.$$

($\Rightarrow$) Now let λ be the initial distribution and assume that the distribution of X_n does not depend on n. Then

$$\lambda_j = \mathbb{P}(X_0 = j) = \mathbb{P}(X_1 = j) = \sum_{i\in I} \lambda_i p_{ij}.$$

Hence, λ is a stationary distribution of $(X_n)_{n\in\mathbb{N}}$.

□

Remark 1.4.5. In the proof above, we have proved that if π is a stationary distribution, then

$$\sum_{i\in I} \pi_i p_{ij}^{(n)} = \pi_j, \ \forall\, j \in I,$$

for any $n \in \mathbb{N}$.

One can easily prove that a equilibrium distribution is invariant when the state space I is finite. But before a necessary theoretical result:

Theorem 1.4.6 (Dominated Convergence Theorem for Sequences)**.** *Consider $(a_i)_{i\in I}$ a sequence of non-negative real numbers such that $\sum_{i\in I} a_i < +\infty$ and $(b_i^{(n)})_{i\in I}$, for each $n \in \mathbb{N}$, a sequence of real numbers such that $|b_i^{(n)}| \le 1$, for all $i \in I$ and $n \in \mathbb{N}$, and*

$$\lim_{n\to+\infty} b_i^{(n)} = b_i, \ \forall\, i \in I.$$

Then

$$\lim_{n\to+\infty} \sum_{i\in I} a_i b_i^{(n)} = \sum_{i\in I} a_i b_i.$$

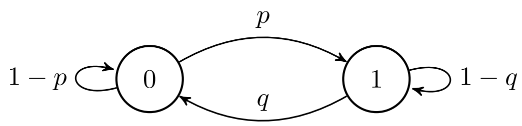
Remark 1.4.7. One can also analyze the unconditional probability $\mathbb{P}(X_n = j)$ using the DCT above. Consider $(X_n)_{n\in\mathbb{N}}$ Markov(λ, P) and let π be the equilibrium distribution. Then:

$$\lim_{n\to+\infty} \mathbb{P}(X_n = j) = \lim_{n\to+\infty} \sum_{i\in I} \lambda_i p_{ij}^{(n)} = \sum_{i\in I} \lambda_i \lim_{n\to+\infty} p_{ij}^{(n)} = \sum_{i\in I} \lambda_i \pi_j = \pi_j.$$

1.4.1 Examples

Before continuing, let us analyze the concept of stationary distributions in some examples.

Example 1.4.8 (2 State Markov Chain)**.** Let us consider the 2 State Markov chain studied before:



What is the stationary distribution of this Markov chain? The equation $\pi P = \pi$ can be written as

$$\begin{cases} (1-p)\pi_0 + q\pi_1 = \pi_0, \\ p\pi_0 + (1-q)\pi_1 = \pi_1. \end{cases}$$

This implies $-p\pi_0 + q\pi_1 = 0$ and, obviously, $\pi_0 + \pi_1 = 1$. So, we find

$$\pi_0 = \frac{q}{p+q} \text{ and } \pi_1 = \frac{p}{p+q}.$$

Let us now compute $\mathbb{P}(X_n = 0)$. Notice that

$$\begin{aligned} \mathbb{P}(X_n = 0) &= \mathbb{P}(X_n = 0, X_{n-1} = 0) + \mathbb{P}(X_n = 0, X_{n-1} = 1) \\ &= \mathbb{P}(X_n = 0 \mid X_{n-1} = 0)\mathbb{P}(X_{n-1} = 0) \\ &+ \mathbb{P}(X_n = 0 \mid X_{n-1} = 1)\mathbb{P}(X_{n-1} = 1) \\ &= (1-p)\mathbb{P}(X_{n-1} = 0) + q\mathbb{P}(X_{n-1} = 1) \\ &= (1-p-q)\mathbb{P}(X_{n-1} = 0) + q \end{aligned}$$

Then,

$$\begin{aligned} &\mathbb{P}(X_1 = 0) = (1-p-q)\lambda_0 + q, \\ &\mathbb{P}(X_2 = 0) = (1-p-q)\mathbb{P}(X_1 = 0) + q \\ &= (1-p-q)^2\lambda_0 + q(1 + (1-p-q)) \\ &\cdots\cdots \\ &\mathbb{P}(X_n = 0) = (1-p-q)^n\lambda_0 + q\sum_{j=0}^{n-1}(1-p-q)^j \end{aligned}$$

Notice that, for $0 < p, q < 1$, we have

$$\sum_{j=0}^{n-1}(1-p-q)^j = \frac{1-(1-p-q)^n}{p+q}.$$

Hence,

$$\begin{aligned} \mathbb{P}(X_n = 0) &= \frac{q}{p+q} + (1-p-q)^n\left(\lambda_0 - \frac{q}{p+q}\right), \\ \mathbb{P}(X_n = 1) &= \frac{p}{p+q} + (1-p-q)^n\left(\lambda_1 - \frac{p}{p+q}\right). \end{aligned}$$

Letting $n \to +\infty$, we find

$$\lim_{n\to+\infty}\mathbb{P}(X_n = 0) = \frac{q}{p+q} \text{ and } \lim_{n\to+\infty}\mathbb{P}(X_n = 1) = \frac{p}{p+q},$$

which is the stationary distribution!

Example 1.4.9 (Birth and Death Chains)**.** To find the stationary distribution (if it exists) of a birth and death chain, we need to find π such that

$$\sum_{i=0}^{+\infty}\pi_i p_{ij} = \pi_j.$$

We are assuming $d = +\infty$ here. Using the particular form P takes in this example:

$$\begin{aligned}\pi_0 r_0 + \pi_1 q_1 &= \pi_0,\\ \pi_{j-1}p_{j-1} + \pi_j r_j + \pi_{j+1}q_{j+1} &= \pi_j.\end{aligned}$$

Using $r_j = 1 - p_j - q_j$, we find

$$\begin{aligned}-\pi_0 p_0 + \pi_1 q_1 &= 0,\\ \pi_{j+1}q_{j+1} - \pi_j p_j &= q_j \pi_j - \pi_{j-1}p_{j-1}.\end{aligned}$$

Then, by induction,

$$\pi_{j+1}q_{j+1} - \pi_j p_j = 0 \Rightarrow \pi_{j+1} = \frac{p_j}{q_{j+1}}\pi_j \Rightarrow \pi_j = \frac{p_0\cdots p_{j-1}}{q_1\cdots q_j}\pi_0.$$

It is obvious that the π above is positive. We have also to require that $\sum \pi_j = 1$, and then it is necessary to require:

$$\sum_{j=1}^{+\infty}\frac{p_0\cdots p_{j-1}}{q_1\cdots q_j} < +\infty \tag{1.10}$$

Define $\lambda_j = \frac{p_0\cdots p_{j-1}}{q_1\cdots q_j}$, $j \geq 1$, and $\lambda_0 = 1$. In this case,

$$\pi_j = \frac{\lambda_j}{\sum_{i=0}^{+\infty}\lambda_i}.$$

Therefore, if (1.10) is true, the chain has a unique stationary distribution. On the other hand, if (1.10) is not verified, the chain has no stationary distribution. When $d < +\infty$, it is easy to see there always exists the stationary distribution.

Example 1.4.10. Consider the following Markov chain: $(\xi_n)_{n\in\mathbb{N}}$ denotes the number of particles added to a box at time n. We assume this sequence is iid with distribution Poisson(λ). Assume that, independently of everything else, every particle at the box at n has probability p to remain in the box at time $n+1$ and $1-p$ to be removed. The state space of this chain is $I = \mathbb{N}$. Mathematically, we may write:

$$X_{n+1} = \xi_{n+1} + R_n,$$

where

$$\mathbb{P}(R_n = k \mid X_n = i) = \binom{i}{k}p^k(1-p)^{i-k}, \text{ for any } k \in \{0,\dots,i\}.$$

Let us compute the transition probability of this chain:

$$\begin{aligned}p_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i) &= \sum_{k=0}^{i\vee j}\mathbb{P}(R_n = k, \xi_{n+1} = j-k \mid X_n = i)\\ &= \sum_{k=0}^{i\vee j}\mathbb{P}(\xi_{n+1} = j-k)\mathbb{P}(R_n = k \mid X_n = i)\\ &= \sum_{k=0}^{i\vee j} e^{-\lambda}\frac{\lambda^{j-k}}{(j-k)!}\binom{i}{k}p^k(1-p)^{i-k}\end{aligned}$$

Notice that $p_{ij} > 0$, for any $i, j \in I$. One can easily show that, if $X_n \sim$ Poisson(ϕ), then $R_n \sim$ Poisson($p\phi$) (can you prove this?). Let us use this to prove that the stationary distribution of $(X_n)_{n\in\mathbb{N}}$ is Poisson(ϕ), where we want to determine ϕ as well. Hence, assume $X_0 \sim$ Poisson(ϕ). We have seen that $R_0 \sim$ Poisson($p\phi$). Since $X_1 = \xi_1 + R_0$ and ξ_1 is independent of R_0, we conclude that $X_1 \sim$ Poisson($\lambda + p\phi$). So the distribution of X_0 is stationary, we must have $X_0 \sim X_1$, which implies $\phi = \lambda + p\phi$. Thus,

$$\phi = \frac{\lambda}{1-p}.$$

Therefore, the stationary distribution of this chain is Poisson($\frac{\lambda}{1-p}$)

$$\pi_i = e^{-\frac{\lambda}{1-p}}\frac{\lambda^i}{(1-p)^i i!}.$$

1.4.2 Reversible Markov Chains

What happens when we look at a Markov chain running backwards? It makes sense that for this to work, we need to start in equilibrium (with the stationary distribution).

Theorem 1.4.11. *Consider $(X_n)_{n\in\mathbb{N}}$ Markov(π, P), assumed to be irreducible, where π is a stationary distribution for it (which we assume to exist). Fix $N \in \mathbb{N}$ and define $Y_n = X_{N-n}$. Then $(Y_n)_{n\in\{0,\dots,N\}}$ is Markov$(\pi, \hat{P})$ and irreducible, where*

$$\pi_j \hat{p}_{ji} = \pi_i p_{ij}, \ \forall\ i, j \in I. \tag{1.11}$$

Proof. It is easy to see that $\hat{P}$ is a stochastic matrix and π is stationary under $\hat{P}$. Moreover,

$$\begin{aligned}\mathbb{P}(Y_0 = i_0, \dots, Y_N = i_N) &= \mathbb{P}(X_0 = i_N, \dots, X_N = i_0)\\ &= \pi_{i_N} p_{i_N, i_{N-1}} \cdots p_{i_1, i_0}\\ &= \pi_{i_0} \hat{p}_{i_1, i_0} \cdots \hat{p}_{i_N, i_{N-1}},\end{aligned}$$

which implies that $(Y_n)_{n\in\{0,\dots,N\}}$ is Markov$(\pi, \hat{P})$. It is also clear that $(Y_n)_{n\in\{0,\dots,N\}}$ is irreducible. $\square$

Definition 1.4.12 (Detailed Balance)**.** We say that a stochastic matrix P and a distribution λ are in *detailed balance* if

$$\lambda_i p_{ij} = \lambda_j p_{ji}, \ \forall\ i, j \in I$$

The next proposition is trivially verified. The main practical use of this proposition is to find the stationary distribution: you could find the distribution that satisfies the detailed balance equation above.

Proposition 1.4.13. *If P and λ are in detailed balance, then λ is invariant for P.*

Of course, not all chains have their stationary distribution in detailed balance.

Definition 1.4.14. $(X_n)_{n\in\mathbb{N}}$ Markov(λ, P) and irreducible. We say $(X_n)_{n\in\mathbb{N}}$ is *reversible* if λ and P are in detailed balance. In particular, λ is invariant under P.

1.4.3 Average Number of Visits to a Recurrent State

It is well-known that a sequence of real number $(a_n)_{n\in\mathbb{N}}$ could be divergent, but its running average $\frac{1}{n}\sum_{k=1}^n a_k$ converges. In this section, we will see that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n p_{ij}^{(k)}$$

always exists for any $i, j \in I$. We then define

$$\begin{aligned}N_j^{(n)} &= \sum_{k=1}^n 1_{\{j\}}(X_k),\\ g_{ij}^{(n)} &= \mathbb{E}[N_j^{(n)} \mid X_0 = i] = \sum_{k=1}^n p_{ij}^{(k)}.\end{aligned}$$

In words, $N_j^{(n)}$ is the number of visits the chain $(X_n)_{n\in\mathbb{N}}$ does to the state j up to time n. If j is transient, we easily conclude

$$\begin{aligned}\lim_{n\to+\infty} \frac{N_j^{(n)}}{n} &= 0 \text{ a.s},\\ \lim_{n\to+\infty} \frac{g_{ij}^{(n)}}{n} &= 0,\end{aligned}$$

for any $i \in I$, since both numerators converge to a finite number, when $n \to +\infty$.

Let us now consider a recurrent state $j \in I$ and define

$$m_j = \mathbb{E}[T_j \mid X_0 = j].$$

In words, m_j is the average return time of the state j.

Theorem 1.4.15. *Let j be a recurrent state. Then*

(i) $\lim_{n\to+\infty} \frac{N_j^{(n)}}{n} = \frac{1_{\{T_j<+\infty\}}}{m_j}$ *a.s.*

(ii) $\lim_{n\to+\infty} \frac{g_{ij}^{(n)}}{n} = \frac{\rho_{ij}}{m_j}$

Proof. Remember the following definitions (see Lemma 1.3.6):

$$\begin{aligned} T_j^{(k)} &= \min\{n \in \mathbb{N} \; ; \; N_j^{(n)} = k\}, \; k \geq 1, \\ S_j^{(1)} &= T_j^{(1)} = T_j, \\ S_j^{(k)} &= T_j^{(k)} - T_j^{(k-1)}, \; k \geq 2. \end{aligned}$$

The r.v. $S_j^{(k)}$ is the waiting time between the $(k-1)$-th visit to j and k-th visit, or what we called the rth excursion of the state j. Notice $T_j^{(k)} = S_j^{(1)} + \cdots + S_j^{(k)}$. Moreover, $(S_j^{(k)})_{k\in\mathbb{N}}$ is an iid sequence with $\mathbb{E}[S_j^{(1)} \mid X_0 = j] = \mathbb{E}[T_j \mid X_0 = j] = m_j$, by Lemma 1.3.6. For simplicity, we assume $m_j < +\infty$; the case $m_j = +\infty$ can be easily handled. By the Law of the Large Number:

$$\lim_{k\to+\infty} \frac{T_j^{(k)}}{k} = \lim_{m\to+\infty} \frac{S_j^{(1)} + \cdots + S_j^{(k)}}{k} = m_j \text{ a.s.}$$

Additionally, we have

$$T_j^{(N_j^{(n)})} \leq n < T_j^{(N_j^{(n)}+1)} \Rightarrow \frac{T_j^{(N_j^{(n)})}}{N_j^{(n)}} \leq \frac{n}{N_j^{(n)}} < \frac{T_j^{(N_j^{(n)}+1)}}{N_j^{(n)}}$$

Since j is recurrent, $N_j^{(n)} \to +\infty$ a.s. when $n \to +\infty$ on $\{T_j < +\infty\}$. Therefore, taking care of the case $T_j = +\infty$, we find

$$\lim_{n\to+\infty} \frac{N_j^{(n)}}{n} = \frac{1_{\{T_j<+\infty\}}}{m_j} \text{ a.s.}$$

Furthermore, notice $0 \leq N_j^{(n)} \leq n$ and hence (ii) can be proved by a simple application of the Dominated Convergence Theorem. □

The theorem above shows that the behavior of the states should be further divide by the fact $m_j < +\infty$. We then have the following definition:

Definition 1.4.16. Consider a recurrent state j. We say j is

- *null recurrent* if $m_j = +\infty$.
- *positive recurrent* if $m_j < +\infty$.

Remark 1.4.17. If j is null recurrent or transient, then

$$\lim_{n\to+\infty} \frac{N_j^{(n)}}{n} = 0 \text{ and } \lim_{n\to+\infty} \frac{g_{ij}^{(n)}}{n} = 0.$$

In the the case j is positive recurrent, we have

$$\lim_{n\to+\infty} \frac{N_j^{(n)}}{n} = \frac{1}{m_j} \text{ and } \lim_{n\to+\infty} \frac{g_{jj}^{(n)}}{n} = \frac{1}{m_j}.$$

Remark 1.4.18. Notice that we can understand the quantity

$$\lim_{n\to+\infty} \frac{N_j^{(n)}}{n}$$

as the long-run proportion of the time the chain spent in state j. As we concluded, this quantity is zero if j is transient or null recurrent. However, if j is positive recurrent, this quantity is $\frac{1_{\{T_j<+\infty\}}}{m_j}$. As we will see, if the chain is positive recurrent and irreducible, this is equivalent to π_j.

Theorem 1.4.19. *If i is positive recurrent and $i \to j$, then j is positive recurrent.*

Proof. We have previously proved that j is recurrent and that $j \to i$. Hence, there exist $n_1, n_2 \in \mathbb{N}$ such that

$$p_{ji}^{(n_1)} > 0 \text{ and } p_{ij}^{(n_2)} > 0.$$

Moreover,

$$p_{jj}^{(n_1+k+n_2)} \geq p_{ji}^{(n_1)} p_{ii}^{(k)} p_{ij}^{(n_2)} \Rightarrow \frac{g_{jj}^{(n_1+n+n_2)}}{n} \geq p_{ji}^{(n_1)} p_{ij}^{(n_2)} \frac{g_{ii}^{(n)}}{n}$$

Letting $n \to +\infty$ and using the fact i is positive recurrent, we find

$$\lim_{n\to+\infty} \frac{g_{jj}^{(n)}}{n} \geq p_{ji}^{(n_1)} p_{ij}^{(n_2)} \frac{1}{m_i} > 0 \Rightarrow m_j < +\infty \Rightarrow j \text{ is positive recurrent.}$$

□

Remark 1.4.20. If C is a irreducible class, then all its states are either transient, null recurrent or positive recurrent.

Theorem 1.4.21. *If C is a closed and finite class, then at least one of its states is positive recurrent. Hence, if C is also communicative, all its states are positive recurrent.*

Proof. Since C is closed and finite,

$$\sum_{j\in C} p_{ij}^{(k)} = 1,\ \forall\ i \in C \Rightarrow \sum_{k=1}^{n} \sum_{j\in C} \frac{p_{ij}^{(k)}}{n} = 1 \Rightarrow \sum_{j\in C} \frac{g_{ij}^{(n)}}{n} = 1.$$

Letting $n \to +\infty$, we cannot have

$$\lim_{n\to+\infty} \frac{g_{ij}^{(n)}}{n} = 0,$$

otherwise we would find $1 = 0$. Therefore, j cannot be transient or null recurrent. □

Corollary 1.4.22. *A Markov chain with finite state space does not have null recurrent states.*

1.4.4 Existence and Uniqueness of Stationary Distributions

Let π be a stationary distribution and fix $k \in \mathbb{N}$. Then

$$\sum_{i\in I} \pi_i p_{ij}^{(k)} = \pi_j,\ \forall\ j \in I$$

It is easy to see that

$$\sum_{i\in I} \pi_i \frac{g_{ij}^{(n)}}{n} = \sum_{i\in I} \pi_i \frac{1}{n} \sum_{k=1}^{n} p_{ij}^{(k)} = \frac{1}{n} \sum_{k=1}^{n} \sum_{i\in I} \pi_i p_{ij}^{(k)} = \frac{1}{n} \sum_{k=1}^{n} \pi_j = \pi_j$$

Theorem 1.4.23. *Let π be a stationary distribution. If j is transient or null recurrent, then $\pi_j = 0$.*

Proof. We have seen that

$$\pi_j = \sum_{i\in I} \pi_i \frac{g_{ij}^{(n)}}{n},$$

for any $n \in \mathbb{N}$. Then

$$\pi_j = \lim_{n\to+\infty} \sum_{i\in I} \pi_i \frac{g_{ij}^{(n)}}{n}.$$

We want to apply the DCT. For this, notice that

$$0 \leq \frac{g_{ij}^{(n)}}{n} \leq 1$$

Since π is a probability distribution, we find

$$\pi_j = \sum_{i\in I} \pi_i \lim_{n\to+\infty} \frac{g_{ij}^{(n)}}{n} = 0,$$

because j is transient or null recurrent. $\square$

Corollary 1.4.24. *A Markov chain without positive recurrent states cannot have a stationary distribution.*

Theorem 1.4.25. *A positive recurrent and irreducible Markov chain has a unique stationary distribution given by*

$$\pi_j = \frac{1}{m_j}, \ \forall\, j \in I. \tag{1.12}$$

Proof. We know that

$$\lim_{n\to+\infty} \frac{g_{ij}^{(n)}}{n} = \frac{1}{m_j}, \ \forall\, i, j \in I.$$

Let us show that if this chain has a stationary distribution, it must be given by (1.12), i.e. uniqueness. Remember

$$\pi_j = \lim_{n\to+\infty} \sum_{i\in I} \pi_i \frac{g_{ij}^{(n)}}{n} = \sum_{i\in I} \pi_i \lim_{n\to+\infty} \frac{g_{ij}^{(n)}}{n} = \sum_{i\in I} \pi_i \frac{1}{m_j} = \frac{1}{m_j}.$$

Let us now verify that (1.12) is a stationary distribution. It is clearly positive. Moreover, we have to verify $\sum_{i\in I} \frac{1}{m_i} = 1$ and $\sum_{i\in I} \frac{1}{m_i} p_{ij} = \frac{1}{m_j}$. Indeed,

$$\begin{aligned}
\sum_{j\in I} \frac{g_{ij}^{(n)}}{n} &= \sum_{j\in I} \sum_{m=1}^{n} \frac{p_{ij}^{(m)}}{n} = \frac{1}{n} \sum_{m=1}^{n} \cancelto{1}{\sum_{j\in I} p_{ij}^{(m)}} = 1,\\
\sum_{i\in I} \frac{g_{ki}^{(n)}}{n} p_{ij} &= \sum_{m=1}^{n} \frac{p_{kj}^{(m+1)}}{n} = \frac{g_{kj}^{(n+1)}}{n} - \frac{p_{kj}}{n}.
\end{aligned}$$

Now, if I is finite, we can let $n \to +\infty$ and conclude

$$\begin{aligned}
1 = \lim_{n\to+\infty} \sum_{j\in I} \frac{g_{ij}^{(n)}}{n} &= \sum_{j\in I} \lim_{n\to+\infty} \frac{g_{ij}^{(n)}}{n} = \sum_{j\in I} \frac{1}{m_j},\\
\sum_{i\in I} \frac{1}{m_i} p_{ij} &= \sum_{i\in I} \lim_{n\to+\infty} \frac{g_{ki}^{(n)}}{n} p_{ij} = \lim_{n\to+\infty} \sum_{i\in I} \frac{g_{ki}^{(n)}}{n} p_{ij}\\
&= \lim_{n\to+\infty} \left(\frac{g_{kj}^{(n+1)}}{n} - \frac{p_{kj}}{n}. \right) = \frac{1}{m_j}.
\end{aligned}$$

This shows that (1.12) is a stationary distribution when I is finite. In the countable case, notice that, for any $I_1 \subset I$ finite, we find

$$\sum_{j\in I_1} \frac{g_{ij}^{(n)}}{n} \le \sum_{j\in I} \frac{g_{ij}^{(n)}}{n} = 1 \Rightarrow \sum_{j\in I_1} \frac{1}{m_j} \le 1.$$

Since this argument works for any $I_1 \subset I$ finite, we conclude

$$\sum_{j\in I} \frac{1}{m_j} \le 1.$$

Similarly, we conclude

$$\sum_{i\in I} \frac{1}{m_i} p_{ij} \le \frac{1}{m_j}.$$

By contradiction, assume that

$$\sum_{i\in I}\frac{1}{m_i}p_{ij_0}<\frac{1}{m_{j_0}},$$

for some j_0. Then

$$\sum_{j\in I}\frac{1}{m_j}>\sum_{j\in I}\sum_{i\in I}\frac{1}{m_i}p_{ij}=\sum_{i\in I}\frac{1}{m_i}\sum_{j\in I}p_{ij}=\sum_{i\in I}\frac{1}{m_i}$$

which is blatantly false. Hence,

$$\sum_{i\in I}\frac{1}{m_i}p_{ij}=\frac{1}{m_j}.$$

Define now $c=1/(\sum_{i\in I}\frac{1}{m_i})$ (which is a positive number since $\sum_{i\in I}\frac{1}{m_i}\le 1$) and $\pi_j=c/m_j$. It is easy to verify that π is also a stationary distribution, and by uniqueness, we conclude that $c=1$. □

Corollary 1.4.26. *An irreducible Markov chain has a unique stationary distribution if and only if it is positive recurrent.*

Notice that the proof of Theorem 1.4.25 could be adapted to prove the next theorem.

Theorem 1.4.27. *Suppose $(X_n)_{n\in\mathbb{N}}$ has an equilibrium distribution π (at lease one $\pi_j>0$). Then π is the unique stationary distribution.*

Proof. Indeed, repeat the argument of the aforementioned proof changing $1/m_j$ to π_j and $g_{ij}^{(n)}/n$ to $p_{ij}^{(n)}$. □

Remark 1.4.28. In the case when $\sum_{\ell\in I}p_{\ell j}<+\infty$, the proof of $\pi=\pi P$ for equilibrium distribution π is a direct application of the DCT. Notice that,

$$\pi_j=\lim_{n\to+\infty}p_{ij}^{(n)}=\lim_{n\to+\infty}p_{ij}^{(n+1)}=\lim_{n\to+\infty}\sum_{\ell\in I}p_{i\ell}^{(n)}p_{\ell j}.$$

We could then simply apply the DCT with $a_\ell=p_{\ell j}$ and $b_\ell^{(n)}=p_{\ell j}^{(n)}$. This would imply that

$$\pi_j=\sum_{\ell\in I}\pi_\ell p_{\ell j},$$

i.e. $\pi=\pi P$.

The same argument of the proof of Theorem 1.4.25 can be used to prove that

Theorem 1.4.29. *Consider a Markov chain and let C be a irreducible and positive recurrent class. Then, this chain has a unique stationary distribution such that $\pi_j=0$, for $j\notin C$ (concentrated in C), and it is given by*

$$\pi_i=\begin{cases}\dfrac{1}{m_i}, & \text{if } i\in C\\[2ex] 0, & \text{if } i\notin C\end{cases},$$

Remark 1.4.30 (Non-Uniqueness of Stationary Distributions)**.** Assume a Markov chain has two distinct, irreducible and positive recurrent classes, C_0 and C_1. Then, there are two stationary distributions, π_0 and π_1, concentrated in C_0 and C_1, respectively. Define then $\pi_\alpha=(1-\alpha)\pi_0+\alpha\pi_1$, for $\alpha\in[0,1]$. It is easy to see that π_α is a stationary distribution for every α.

Example 1.4.31 (Birth and Death Chains)**.** We have seen that an irreducible birth and death chain has stationary distribution if and only if

$$\sum_{i=1}^{+\infty}\frac{p_0\cdots p_{i-1}}{q_1\cdots q_i}<+\infty.$$

In particular, the condition above is equivalent to the chain being positive recurrent. Additionally, we proved that the chain is transient if and only if

$$\sum_{i=1}^{+\infty}\frac{q_1\cdots q_i}{p_1\cdots p_i}<+\infty.$$

Therefore, the chain will be null recurrent if and only if both conditions above fail, i.e.

$$\sum_{i=1}^{+\infty} \frac{p_0 \cdots p_{i-1}}{q_1 \cdots q_i} = +\infty \text{ and } \sum_{i=1}^{+\infty} \frac{q_1 \cdots q_i}{p_1 \cdots p_i} = +\infty.$$

Theorem 1.4.32 (Ergodic Theorem). *Let $(X_n)_{n\in\mathbb{N}}$ a positive recurrent, irreducible Markov chain. Then, if $f : I \longrightarrow \mathbb{R}$ is bounded,*

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{k=1}^{n} f(X_k) = \sum_{i\in I} f(i)\pi_i = \bar{f}.$$

Proof. Notice that

$$\frac{1}{n} \sum_{k=1}^{n} f(X_k) = \sum_{i\in I} \frac{1}{n} N_i^{(n)} f(i).$$

Let J a finite subset of I and assume, without loss of generality, that $|f| \le 1$. Then

$$\begin{aligned} \left| \frac{1}{n} \sum_{k=1}^{n} f(X_k) - \bar{f} \right| &= \left| \sum_{i\in I} \left(\frac{1}{n} N_i^{(n)} - \pi_i \right) f(i) \right| \\ &\le \sum_{i\in J} \left| \frac{1}{n} N_i^{(n)} - \pi_i \right| + \sum_{i\notin J} \left| \frac{1}{n} N_i^{(n)} - \pi_i \right| \\ &\le \sum_{i\in J} \left| \frac{1}{n} N_i^{(n)} - \pi_i \right| + \sum_{i\notin J} \left(\frac{1}{n} N_i^{(n)} + \pi_i \right) \end{aligned}$$

Remember $\sum_{i\in I} N_i^{(n)} = n$. Hence,

$$\sum_{i\notin J} \frac{1}{n} N_i^{(n)} = 1 - \sum_{i\in J} \frac{1}{n} N_i^{(n)} = \sum_{i\in I} \pi_i - \sum_{i\in J} \frac{1}{n} N_i^{(n)} = \sum_{i\notin J} \pi_i - \sum_{i\in J} \left(\frac{1}{n} N_i^{(n)} - \pi_i \right).$$

This implies

$$\left| \frac{1}{n} \sum_{k=1}^{n} f(X_k) - \bar{f} \right| \le 2 \sum_{i\in J} \left| \frac{1}{n} N_i^{(n)} - \pi_i \right| + 2 \sum_{i\notin J} \pi_i.$$

Fix $\varepsilon > 0$. Since π is a proper distribution, there exists J finite such that

$$\sum_{i\notin J} \pi_i < \frac{\varepsilon}{4}.$$

Moreover, since $\frac{1}{n} N_i^{(n)} \to \pi_i$ a.s. and J is finite, we have

$$\sum_{i\in J} \left| \frac{1}{n} N_i^{(n)} - \pi_i \right| < \frac{\varepsilon}{4}.$$

These inequalities yield the result. □

1.4.5 Convergence to the Stationary Distribution

In this section we will study the cases when the equilibrium distribution is equal to the stationary distribution. An example where there is no convergence:

Example 1.4.33. Consider the Markov chain with two states $I = \{0, 1\}$ and transition probability

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

It is easy to see that $P^{2n} = I$ and $P^{2n+1} = P$. Hence, $p_{ij}^{(n)}$ does not converge, for any $i, j \in I$.

The main issue with the example above is that $p_{ij}^{(n)}$ can always take the value 0 when n is growing to infinity. It turns out that this is equivalent to the chain having a periodic behavior. We will precisely define the notion below.
Before proceeding, let us fix some notation:

Definition 1.4.34. Let n and d in $\mathbb{N}$. We say d is a *divisor* of n if $n/d \in \mathbb{N}$. Now, if $A \subset \mathbb{N}$ is non-empty, we define the *greatest common divisor* as

$$\gcd(A) = \max\{d \in \mathbb{N}\ ;\ n/d \in \mathbb{N},\ \forall\ n \in A\}.$$

Obviously,

$$1 \leq \gcd(A) \leq \min A$$
$$1 \in A \Rightarrow \gcd(A) = 1.$$

Definition 1.4.35 (Period)**.** A *period* of a state i such $\rho_{ii} > 0$ is defined as

$$d_i = \gcd\{n \in \mathbb{N}\ ;\ p_{ii}^{(n)} > 0\}.$$

Notice that $\rho_{ii} > 0$ implies $p_{ii}^{(n)} > 0$, for some $n \in \mathbb{N}$.

Proposition 1.4.36. *If $i \leftrightarrow j$, then $d_i = d_j$.*

Proof. Let n_i and n_j such that $p_{ij}^{(n_i)} > 0$ e $p_{ji}^{(n_j)} > 0$. Then

$$p_{ii}^{(n_i+n_j)} \geq p_{ij}^{(n_i)} p_{ji}^{(n_j)} > 0 \Rightarrow d_i \text{ divides } n_i + n_j$$

Now, take $m \in \{n \in \mathbb{N}\ ;\ p_{jj}^{(n)} > 0\}$ and note

$$p_{ii}^{(n_i+m+n_j)} \geq p_{ij}^{(n_i)} p_{jj}^{(m)} p_{ji}^{(n_j)} > 0 \Rightarrow d_i \text{ divides } n_i + m + n_j$$

Then, d_i must divide m, for any $m \in \{n \in \mathbb{N}\ ;\ p_{jj}^{(n)} > 0\}$. Since d_j is the greatest common divisor of the set $\{n \in \mathbb{N}\ ;\ p_{jj}^{(n)} > 0\}$, we conclude $d_i \leq d_j$. Similarly, $d_j \leq d_i$ and therefore, $d_i = d_j$. ☐

Definition 1.4.37. A Markov chain is said to be periodic with period d, if $d_i = d$, for all $i \in I$. Additionally, we say the chain is aperiodic if it periodic with period 1. Moreover, if the chain is irreducible, by the proposition above, we have to verify the period of only one element. In particular, a sufficient condition for an irreducible Markov chain to be aperiodic is that $p_{ii} > 0$, for some $i \in I$.

Example 1.4.38 (Birth and Death Chains)**.** A birth and death chain with $r_i > 0$, for some i, is aperiodic. Now, assume $r_i = 0$, for all $i \in I$. It is easy to see that $p_{ii}^{(2n+1)} = 0$, for any n. Moreover, $p_{00}^{(2)} = p_0 q_1 > 0$. Hence, the period of this chain is 2. In particular, the Ehrenfest chain is periodic with period 2.

The connection between aperidic chains and $p_{ij}^{(n)} > 0$ for large n is shown in the next lemma from Number Theory. Notice that property 2 below is always satisfied for Markov chains.

Lemma 1.4.39. *Let $A \subset \mathbb{N}$ be non-empty such that*

1. $\gcd\{A\} = 1$, *and*
2. $m, n \in A \Rightarrow m + n \in A$.

Then, there exists $n_0 \in \mathbb{N}$ such that $\{n_0, n_0 + 1, \ldots\} \subset A$.

Theorem 1.4.40. *Let $(X_n)_{n\in\mathbb{N}}$ be Markov(λ, P). Assume it is aperiodic, irreducible and positive recurrent with stationary distribution π. Then*

$$\lim_{n\to+\infty} \mathbb{P}(X_n = j) = \pi_j,\ \forall\ j \in I.$$

Proof. The technique used in this proof is called coupling. Consider $(Y_n)_{n\in\mathbb{N}}$ Markov(π, P) independent of $(X_n)_{n\in\mathbb{N}}$ (the initial distribution is the stationary distribution of $(X_n)_{n\in\mathbb{N}}$). Note that $\mathbb{P}(Y_n = j) = \pi_j$. It is easy to notice that $W_n = (X_n, Y_n)$ is Markov$(\mu, \tilde{P})$, where

$$\begin{aligned}\tilde{p}_{(i,l)(j,k)} &= p_{ij}p_{lk},\\ \mu_{(i,l)} &= \lambda_i\pi_l.\end{aligned}$$

Fix a state $a \in I$ and define the stopping time

$$\tau = \inf\{n \in \mathbb{N} \; ; \; X_n = Y_n = a\}$$

Firstly, we will show $\mathbb{P}(\tau < +\infty) = 1$. Since P is aperiodic, by Lemma 1.4.39, we have

$$\tilde{p}^{(n)}_{(i,l)(j,k)} = p^{(n)}_{ij}p^{(n)}_{lk} > 0,$$

for n large enough. Indeed, let n_0 be given by Lemma 1.4.39 for the state j. Since the original chain is irreducible, let $n_1 \geq 0$, such that $p^{(n_1)}_{ij} > 0$, and $n_2 \geq n_0 + n_1$ such that $p^{(n_2)}_{lk} > 0$. This implies that, taking $n = n_2$,

$$\tilde{p}^{(n_2)}_{(i,l)(j,k)} = p^{(n_2)}_{ij}p^{(n_2)}_{lk} \geq p^{(n_1)}_{ij}p^{(n_2-n_1)}_{jj}p^{(n_2)}_{lk} > 0,$$

This implies that $(W_n)_{n\in\mathbb{N}}$ is irreducible. It is easy to see that $\tilde{\pi}_{(i,l)} = \pi_i\pi_l$ is a stationary distribution of $(W_n)_{n\in\mathbb{N}}$, which implies that this chain is positive recurrent. Since $\tau = T_{(a,a)}$, the hitting time of (a, a) for the chain $(W_n)_{n\in\mathbb{N}}$, we conclude $\mathbb{P}(\tau < +\infty) = 1$.

Define now

$$Z_n = \begin{cases} X_n, & \text{if } n < \tau, \\ Y_n, & \text{if } n \geq \tau. \end{cases}$$

Let us verify that $(Z_n)_{n\in\mathbb{N}}$ is Markov(λ, P). By the strong Markov property, $(X_{\tau+n}, Y_{\tau+n})_{n\in\mathbb{N}}$ is Markov$(\delta_{(a,a)}, \tilde{P})$ and independent of $(X_0, Y_0), \ldots, (X_\tau, Y_\tau)$. Symmetrically, $(Y_{\tau+n}, X_{\tau+n})_{n\in\mathbb{N}}$ is Markov$(\delta_{(a,a)}, \tilde{P})$ and also independent of $(X_0, Y_0), \ldots, (X_\tau, Y_\tau)$. Hence, if we define

$$V_n = \begin{cases} Y_n, & \text{if } n < \tau, \\ X_n, & \text{if } n \geq \tau, \end{cases}$$

we conclude that $(Z_n, V_n)_{n\in\mathbb{N}}$ is Markov$(\mu, \tilde{P})$. Therefore, integrating V out, we conclude $(Z_n)_{n\in\mathbb{N}}$ is Markov(λ, P). This implies $\mathbb{P}(Z_n = j) = \mathbb{P}(X_n = j)$. Moreover,

$$\mathbb{P}(Z_n = j) = \mathbb{P}(X_n = j, \tau > n) + \mathbb{P}(Y_n = j, \tau \leq n).$$

Therefore

$$\begin{aligned}|\mathbb{P}(X_n = j) - \pi_j| &= |\mathbb{P}(Z_n = j) - \mathbb{P}(Y_n = j)|\\ &= |\mathbb{P}(X_n = j, \tau > n) - \mathbb{P}(Y_n = j, \tau > n)| \leq \mathbb{P}(\tau > n),\end{aligned}$$

proving the result. □

Remark 1.4.41. Since the initial distribution, λ, did not play any roll in the previous theorem, we can easily conclude that

$$\lim_{n\to+\infty} p^{(n)}_{ij} = \pi_j, \; \forall \; i, j \in I.$$

Additionally, if the chain has an irreducible, positive recurrent and aperiodic class $C \subset I$, then there exists a unique stationary distribution concentrated on C such that

$$\lim_{n\to+\infty} p^{(n)}_{ij} = \pi_j = \frac{1}{m_j}, \; \forall \; i, j \in C.$$

Let us try to understand why the argument above does not work when the chain is periodic.

Example 1.4.42. It is easy to see that the period of the chain in Example 1.4.33 is 2. Moreover, the (unique) stationary distribution is given by $\pi = [1/2, 1/2]$. Hence, in the argument above we start $(Y_n)_{n\in\mathbb{N}}$ from either 0 or 1 with equal probability. Assuming $(X_n)_{n\in\mathbb{N}}$ starts from 0 and $(Y_n)_{n\in\mathbb{N}}$ starts from 1, then by the periodicity of the chain, these two processes will never meet and the proof breaks.

Theorem 1.4.43. *Let $(X_n)_{n\in\mathbb{N}}$ be Markov(λ, P) irreducible and positive recurrent with stationary distribution π. If the chain is periodic with period d, then for each $i, j \in I$, there exists $r \in \mathbb{N}$ such that $p_{ij}^{(n)} = 0$, unless $n = md + r$, and in this case, we find*

$$\lim_{n\to+\infty} p_{ij}^{(md+r)} = d\pi_j.$$

Proof. Define the chain $Y_n = X_{nd}$. The initial distribution does not change, but the new transition matrix is now $Q = P^d$. Moreover, it is easy to see that this chain is now aperiodic and that the expected return time to a state j for the chain $(Y_n)_{n\in\mathbb{N}}$ is m_j/d, where m_j is the expected return time to j for $(X_n)_{n\in\mathbb{N}}$. Hence, we are able to apply the previous theorem and conclude, for any $j \in I$,

$$\lim_{n\to+\infty} p_{ij}^{(nd)} = \lim_{n\to+\infty} q_{ij}^{(n)} = \frac{d}{m_j} = d\pi_j,$$

for any $i \in I$ such that $i \leftrightarrow j$ under $(Y_n)_{n\in\mathbb{N}}$. In particular,

$$\lim_{n\to+\infty} p_{jj}^{(nd)} = d\pi_j, \ \forall\, j \in I.$$

Consider now $i, j \in I$ and define $l = \min\{n \ ; \ p_{ij}^{(n)} > 0\}$. We will verify that $p_{ij}^{(n)} > 0$ only when $n - l$ is a multiple of d. Let t be such $p_{ij}^{(t)} > 0$. Then, by the definition of l,

$$p_{jj}^{(t+l)} \geq p_{ji}^{(t)} p_{ij}^{(l)} > 0,$$

which implies that $t + l$ is a multiple of d, by the definition of period. The same argument shows $n + t$ is multiple of d if $p_{ij}^{(n)} > 0$. Hence, so is $n - l$, i.e. $n = kd + l$, for some k. We have then proved that if $p_{ij}^{(n)} > 0$, then $n = md + r$, for some m and $0 \leq r < d$ (just right l as a multiple of d plus a rest r). In other words, $p_{ij}^{(n)} = 0$ unless $n = md + r$, for some m. Furthermore,

$$p_{ij}^{(md+r)} = \sum_{k=0}^{m} \mathbb{P}(T_j = kd + r \mid X_0 = i) p_{jj}^{(d(m-k))} = \sum_{k=0}^{+\infty} \mathbb{P}(T_j = kd + r \mid X_0 = i)\alpha_m^{(k)},$$

where $\alpha_m^{(k)} = p_{jj}^{(d(m-k))}$, if $k \leq m$, and $\alpha_m^{(k)} = 0$, otherwise. Notice that $\alpha_m^{(k)}$ converges to $d\pi_j$, when $m \to +\infty$, for any k. Then, by the Dominated Convergence Theorem, we conclude

$$\lim_{m\to+\infty} p_{ij}^{(md+r)} = d\pi_j \sum_{k=0}^{+\infty} \mathbb{P}(T_j = kd + r \mid X_0 = i) = d\pi_j \mathbb{P}(T_j < +\infty \mid X_0 = i) = d\pi_j,$$

which proves the theorem. □

1.4.6 Numerical Implementation: Computing π

We will consider the case of $d = |I| < +\infty$. Let us first make some linear algebra considerations. Notice that P will be irreducible (i.e. communicative, since the whole chain is obviously closed) if, for each $i, j \in I$, there exists n such that $(P^n)_{ij} > 0$. Additionally, the chain will be aperiodic if there exists n such that $P^n > 0$ (meaning all elements of P^n are strictly positive); see the proof of Theorem 1.4.40. In Linear Algebra, we say that P is regular.

Since the I is finite, this implies that all states are positive recurrent. Hence, there exists a unique $\pi > 0$ (all entries strictly positive) such that $\pi = \pi P$, i.e. 1 is eigenvalue of P, π its eigenvector and the sequence generated as $\pi^{(n)} = \pi^{(n-1)} P$ converges to π for any $\pi^{(0)}$ (by Theorem 1.4.40).

Let us now analyze the convergence of $\pi^{(n-1)}$ (called the power method) more closely. By the Perron-Frobenius Theorem (from Linear Algebra), we have that any other eigenvalue λ of P satisfies $|\lambda| < 1$. Assume for simplicity that P is diagonalizable. This means there exists n orthonormal eigenvectors $\pi, v_2, \ldots, v_d$ for each of the eigenvalues $1, \lambda_2, \ldots, \lambda_d$. Then

$$\pi^{(0)} = c_1\pi + c_2 v_2 + \cdots + c_d v_d,$$

which implies that

$$\pi^{(n-1)}P = \pi^{(0)}P^n = (c_1\pi + c_2 v_2 + \cdots + c_d v_d)P^n = c_1\pi + c_2\lambda_2^n v_2 + \cdots + c_d\lambda_d^n v_d.$$

Therefore, since $|\lambda_i| < 1$, taking n to infinity, the sequence converges to a multiple of π (which we known it will be π itself). The speed of convergence of this method is determined by the second largest eigenvalue, λ_2. It is also important to notice that this is a direct method, meaning that only matrix multiplications are involved.

1.5 Application in Network Science: PageRank

We want to assign each webpage p a number that measures its importance, denoted here by $I(p)$ and called p's PageRank (from Google). Assume that page p_j has ℓ_j links. If one of them forward us to page p_i, them p_j will give $1/\ell_j$ of its importance to p_i. Hence, if we denote by B_i all the pages that links to p_i, we set

$$I(p_i) = \sum_{p_j \in B_i} \frac{1}{\ell_j} I(p_j).$$

Define then the following "almost" stochastic matrix H

$$H_{ij} = \begin{cases} \dfrac{1}{\ell_i}, & \text{if } P_i \in B_j, \\ 0, & \text{otherwise.} \end{cases}$$

This matrix is "almost" stochastic because all of its entries are non-negative, its rows sum to one unless the page for that row has no links. We can then rewrite the equation for $I(p_i)$ as

$$I = H\, I,$$

where $I = (I(p_1), \ldots, I(p_N))$, where N is the total number of webpages on the internet. Therefore, if H were stochastic, then I would be its stationary distribution. Additionally, if H were aperiodic, we could find I following the power method described in Section 1.4.6.

So, in order to make use of this method we will modify H so we have an aperiodic Markov chain. We start by changing the zero rows of H by the vector $\frac{1}{N}\mathbf{1}$. Let us call this matrix S, which is obviously stochastic. In order to make it irreducible and aperidic, we consider

$$G = \alpha S + (1-\alpha)\frac{1}{N}E,$$

where $\alpha \in (0,1)$ and E is a $N \times N$ matrix of ones. It can be show that the second largest eigenvalue of G satisfies $|\lambda_2| = \alpha$. A sensible choice for α is 0.85. Applying then the power method to G we find an approximation of its stationary distribution, denoted here by π. The PageRank of a webpage p_i is then π_i.

1.6 Application in Statistics: Markov Chain Monte Carlo

Assuming we know how to simulate iid random variables with $U[0,1]$ distribution. For any cdf F with known F^{-1} (it could depend on some numerical method), we could simulate iid samples of random variable with cdf F by using $F^{-1}(U)$:

$$\mathbb{P}(F^{-1}(U) \le x) = \mathbb{P}(U \le F(x)) = F(x).$$

For instance, the Exp(λ) distribution has cdf $F(x) = 1 - e^{-\lambda x}$ and hence $F^{-1}(u) = -\frac{1}{\lambda}\log(1-u)$. Since $1 - U \sim U[0,1]$, we might simulate from Exp(λ) using $-\frac{1}{\lambda}\log(1-U)$. More complicated distributions could be simulated using ingenious methods, see, for example, the Box-Muller method to simulate $N(0,1)$.

Here we are interested on simulating general distributions with unknown normalizing constant. This arises naturally in Bayesian statistics. The method is based on the theory of Markov Chain and it is called Markov Chain Monte Carlo (MCMC). We will present the method in the setting of this chapter, namely the state space (where the random variable takes value) will be assumed discrete. We would like to make it clear that discrete distributions could be easily simulated using the method described in the paragraph above (F^{-1} is known). The MCMC method is used to simulate distributions on complicated spaces.

Nonetheless, the nature of the method can be understood by looking at the discrete case and that is what we will do here.

Let I be the state space (the values we would like to simulate from). We will continue using i, j, k to denote the elements of this space. This, as before, does note mean that the values are integers.

The name of the particular MCMC algorithm we will present here is Metropolis-Hastings. The idea of the algorithm is to create a Markov chain with a given distribution as the stationary distribution. For this, consider any transition probability $Q = (q_{ij})_{i,j\in I}$ and a distribution $(\pi_i)_{i\in I}$ with $\pi_i = cf_i$. We know $(f_i)_{i\in I}$, but we do not know c.

Let us define then the Markov chain $(X_n)_{n\in\mathbb{N}}$ as follows: choose any $X_0 \in I$ and for $n = 1, 2, \ldots$, if $X_{n-1} = i$,

1. draw j from $(q_{ij})_{j\in I}$;
2. define the probability of acceptance:

$$\alpha_{ij} = \min\left\{1, \frac{f_j q_{ji}}{f_i q_{ij}}\right\}.$$

3. Set $X_n = j$ with probability α_{ij} and keep $X_n = i$ with probability $1 - \alpha_{ij}$.

The main assumption is that the first step of the method above can be easily performed. In our case, it is as hard as simulating π itself (up to computing c, but it should not be hard in the countable case). However, as mentioned before, the setting we are using is just to keep the computations (and definitions) simple. Always think in the uncountable case.
We will now find the transition probability of $(X_n)_{n\in\mathbb{N}}$.

$$\begin{aligned}
p_{ij} &= \mathbb{P}(X_n = j \mid X_{n-1} = i) \\
&= \mathbb{P}(X_n = j \mid X_{n-1} = i,\ \text{change accepted})\mathbb{P}(\text{change accepted} \mid X_{n-1} = i) \\
&+ \mathbb{P}(X_n = j \mid X_{n-1} = i,\ \text{change rejected})\mathbb{P}(\text{change rejected} \mid X_{n-1} = i) \\
&= q_{ij}\alpha_{ij} + 1_{\{i=j\}}(1 - a(i)),
\end{aligned}$$

where

$$a(i) = \mathbb{P}(\text{change accepted} \mid X_{n-1} = i) = \sum_{k\in I} q_{ik}\alpha_{ik}.$$

We will now verify that $(\pi_i)_{i\in I}$ is the stationary distribution of P by checking the detailed balance equation:

$$\pi_i p_{ij} = \pi_j p_{ji} \Leftrightarrow f_i p_{ij} = f_j p_{ji}.$$

Indeed,

$$\begin{aligned}
f_i p_{ij} &= f_i q_{ij}\alpha_{ij} + f_i 1_{\{i=j\}}(1 - a(i)) \\
&= \min\{f_i q_{ij}, f_j q_{ji}\} + f_j 1_{\{i=j\}}(1 - a(j)) \\
&= f_j q_{ji}\alpha_{ji} + f_j 1_{\{i=j\}}(1 - a(j)) = f_j p_{ji}.
\end{aligned}$$

The chain is clearly aperiodic (if the probability of rejection is strictly positive). The irreducibility of the chain will depend on the support of f. Assuming here $f_i > 0$ for all i, the chain will be irreducible. These facts are much more complicated in the uncountable case.

Therefore, to simulate from π one needs to simulate the chain $(X_n)_{n\in\mathbb{N}}$ up to very large n. By the theorems presented in the section before, the law of X_n is converging to π and hence we will be simulating (approximately) from π. There are diagnosis techniques to check if the chain has converged (for instance the autocorrelation function of X for large n).

1.7 Application in Machine Learning: Markov Decision Processes (Reinforcement Learning)

A Markov Decision Process (MDP) is a tuple (S, R, A), where $(S_n)_{n\in\mathbb{N}}$ models the state of the environment taking value in I, $(R_n)_{n\in\mathbb{N}}$ is the reward and $(A_n)_{n\in\mathbb{N}}$ is the action taken by the agent. Moreover, we assume that $A_n \in \mathcal{A}(S_n)$, where $\mathcal{A}(i)$ is the set of the available actions to the agent when the state is equal i. The goal of the agent at time n is to choose action A_n to maximize (in some sense) the future gain:

$$G_n = R_{n+1} + \gamma R_{n+2} + \cdots = \sum_{k=0}^{+\infty} \gamma^k R_{n+k+1},$$

where $\gamma \in (0, 1)$ is a discount rate. We are writing this problem in its infinite horizon formulation. In order to consider the finite horizon case, one could assume that $R_n = 0$, for $n \geq N$, for some given N. The main assumption is the following version of the Markov property:

$$\begin{aligned}&\mathbb{P}(R_{n+1} = r, S_{n+1} = j \mid S_0, A_0, R_1, \ldots, S_{n-1}, A_{n-1}, R_n, S_n = i, A_n = a)\\ &= \mathbb{P}(R_{n+1} = r, S_{n+1} = j \mid S_n = i, A_n = a) = p(j, r|i, a),\end{aligned}$$

where we are also assuming that the process is time homogeneous. Let us also define the following quantities:

$$\begin{aligned}p(r \mid i, a) &= \mathbb{P}(R_{n+1} = r \mid S_n = i, A_n = a) = \sum_j p(j, r|i, a),\\ r(i, a) &= \mathbb{E}[R_{n+1} \mid S_n = i, A_n = a] = \sum_r r p(r|i, a),\\ p(j \mid i, a) &= \mathbb{P}(S_{n+1} = j \mid S_n = i, A_n = a) = \sum_r p(j, r|i, a),\\ r(i, a, j) &= \mathbb{E}[R_{n+1} \mid S_n = i, A_n = a, S_{n+1} = j] = \frac{\sum_r r p(j, r|i, a)}{p(j|i, a)}.\end{aligned}$$

Policies and Value Function

A policy is a mapping from each state $i \in I$ to an action $a \in \mathcal{A}(i)$.[1] The value of a state i under policy π is defined as

$$V_\pi(i) = \mathbb{E}_\pi[G_n \mid S_n = i],$$

where $\mathbb{E}_\pi$ is the expectation assuming the agent follows policy π. Additionally, we define the action-value function under policy π as

$$Q_\pi(i, a) = \mathbb{E}_\pi[G_n \mid S_n = i, A_n = a].$$

It is straightforward that

$$V_\pi(i) = Q_\pi(i, \pi(i)).$$

Notice that

$$\begin{aligned}Q_\pi(i, a) &= \mathbb{E}_\pi[R_{n+1} \mid S_n = i, A_n = a] + \gamma \mathbb{E}_\pi\left[\sum_{k=0}^{+\infty} \gamma^k R_{n+k+2} \mid S_n = i, A_n = a\right]\\ &= r(i, a) + \gamma \sum_{j\in I} p(j|i, a) \mathbb{E}_\pi\left[\sum_{k=0}^{+\infty} \gamma^k R_{n+k+2} \mid S_{n+1} = j, S_n = i, A_n = a\right]\\ &= r(i, a) + \gamma \sum_{j\in I} p(j|i, a) \mathbb{E}_\pi\left[\sum_{k=0}^{+\infty} \gamma^k R_{n+k+2} \mid S_{n+1} = j\right]\\ &= r(i, a) + \gamma \sum_{j\in I} p(j|i, a) V_\pi(j).\end{aligned}$$

[1] We could also consider a policy being a mapping from each state $i \in I$ and action $a \in \mathcal{A}(i)$ to the probability of taking action a when the state is i.

Hence

$$Q_\pi(i,a) = r(i,a) + \gamma \sum_{j\in I} p(j|i,a) V_\pi(j).$$

Moreover, the value function V_π satisfies the following recursive property

$$V_\pi(i) = Q_\pi(i,\pi(i)) = r(i,\pi(i)) + \gamma \sum_{j\in I} p(j|i,\pi(i)) V_\pi(j).$$

This is called the Bellman equation for the value function. Notice that, if we denote the transition matrix $p(j|i,\pi(i))$ by P_π and the rewards $r(i,\pi(i))$ by r_π, we find the matrix version of the equation above

$$V_\pi = r_\pi + \gamma P_\pi V_\pi.$$

Hence, when feasible, we may use

$$V_\pi = (I - \gamma P_\pi)^{-1} r_\pi.$$

Optimal Policy

The main idea is to choose the policy that yields the best value function. We then define the optimal value function as

$$V_*(i) = \sup_\pi V_\pi(i).$$

An optimal policy, π_*, is a policy that achieves the supremum above for each $i \in I$:

$$\pi_*(i) \in \arg\max_\pi V_\pi(i).$$

Notice that π_* does not need to be unique. Moreover, $V_* = V_{\pi_*}$.

Similarly, we consider the optimal action-value function:

$$Q_*(i,a) = \sup_\pi Q_\pi(i,a).$$

First, we will relate Q_* and V_*. Notice

$$\begin{aligned} Q_*(i,a) = \sup_\pi Q_\pi(i,a) &= \sup_\pi \left(r(i,a) + \gamma \sum_{j\in I} p(j|i,a) V_\pi(j) \right) \\ &= r(i,a) + \gamma \sum_{j\in I} p(j|i,a) V_*(j). \end{aligned}$$

Additionally,

$$V_*(i) = \sup_{a\in\mathcal{A}(i)} Q_*(i,a).$$

Therefore

$$\begin{aligned} V_*(i) = \sup_{a\in\mathcal{A}(i)} Q_*(i,a) &= \sup_{a\in\mathcal{A}(i)} \mathbb{E}[R_{n+1} + \gamma V_*(S_{n+1}) \mid S_n = i, A_n = a] \\ &= \sup_{a\in\mathcal{A}(i)} \left(r(i,a) + \gamma \sum_{j\in I} p(j|i,a) V_*(j) \right). \end{aligned}$$

Using the formula that relates Q_* and V_*, we find the following Bellman equation for Q_*:

$$Q_*(i,a) = r(i,a) + \gamma \sum_{j\in I} p(j|i,a) \sup_{a'\in\mathcal{A}(j)} Q_*(j,a').$$

Example 1.7.1. Let us assume that we are controlling a robot that "searches and destroys" enemy's robots. We assume that once it finds an enemy's robot, the destruction is automatic. The state of the environment is the level of the robot's battery and it could be H (high) or L (low). The actions available are

$$\mathcal{A}(H) = \{S, W\} \text{ and } \mathcal{A}(L) = \{S, W, Re\},$$

where S is search, W is wait and Re is recharge. The dynamics work accordingly the following table:

i	j	a	$p(j\|i,a)$	$r(i,a,j)$
H	H	S	α	r_s
H	L	S	$1-\alpha$	r_s
L	H	S	$1-\beta$	-3
L	L	S	β	r_s
H	H	W	1	r_w
H	L	W	0	r_w
L	H	W	0	r_w
L	L	W	1	r_w
L	H	Re	1	0
L	L	Re	0	0

Some remarks: $r_s > r_w > 0$ and, when $i = \text{L}$ and we choose $a = \text{S}$ there is a chance that the battery of the robot is depleted (that happens with probability $1-\beta$. In this case, the robot must be rescued and then its battery is recharged to H. We set the reward to -3 (a very low number) in order to model the rescue costs.
Using Bellman equation, we find

$$\begin{aligned} V_*(H) &= \max\{r_s + \gamma(\alpha V_*(H) + (1-\alpha)V_*(L)), \\ &\quad r_w + \gamma V_*(H)\}, \\ V_*(L) &= \max\{\beta r_s - 3(1-\beta) + \gamma((1-\beta)V_*(H) + \beta V_*(L)), \\ &\quad r_w + \gamma V_*(H), \\ &\quad \gamma V_*(H)\}. \end{aligned}$$

Existence of V_* and Dynamic Programming

For each policy π and function $W : I \longrightarrow \mathbb{R}$, define the linear operators

$$\begin{cases} (T_\pi W)(i) = r(i,\pi(i)) + \gamma \sum_{j\in I} p(j|i,\pi(i))W(j), \\ (TW)(i) = \max_{a\in\mathcal{A}} \left(r(i,a) + \gamma \sum_{j\in I} p(j|i,a)W(j) \right), \end{cases}$$

where we are assuming, for simplicity, that $\mathcal{A}(i) = \mathcal{A}$, for all $i \in I$. Notice that $T_\pi \leq T$ and it is straigthforward to verify that both T^π and T are monotonic and contractions (with constant γ) in $L^\infty(I)$. By Banach fixed point theorem, V_π and V_* are fixed points of T_π and T, respectively. Moreover

$$V_* = \lim_{k\to+\infty} T^k W \text{ and } V_\pi = \lim_{k\to+\infty} T_\pi^k W,$$

for any $W : I \longrightarrow \mathbb{R}$.

This result gives us an numerical method to compute V_*. Let us assume I is finite. We can then identify W by a vector in $\mathbb{R}^N$, where $N = |I|$. Then, given any vector $V_0 \in \mathbb{R}^N$, the sequence $V_k = TV_{k-1}$ converges to V_* and, for a large K, we can approximate the optimal policy as

$$\pi_K(i) = \arg\max_{a\in A} \left(r(i,a) + \gamma \sum_{j\in I} p(j|i,a)V_K(j) \right).$$

Similarly, we could consider the Q function instead of the value function. This is called the value iteration method.

Another approach is the so-called policy iteration. Consider any initial policy π_0. For any k, we do a policy evaluation V_{π_k} and then a policy improvement:

$$\pi_{k+1}(i) = \arg\max_{a\in\mathcal{A}} \left(r(i,a) + \gamma \sum_{j\in I} p(j|i,a)V_{\pi_k}(j) \right).$$

We end the iteration when $V_{\pi_k} = V_{\pi_{k+1}}$. In the finite case there are a finite number of policies and then this condition is eventually met. Moreover, by the monotonicity of T and T_π, we find that $V_{\pi_{k+1}} \geq V_{\pi_k}$. Indeed, by the definitions of T, T_π and π_{k+1},

$$V_{\pi_k} = T_{\pi_k} V_{\pi_k} \leq T V_{\pi_k} = T_{\pi_{k+1}} V_{\pi_k}.$$

Therefore, since $T_{\pi_{k+1}}$ is monotonic,

$$V_{\pi_k} \leq T_{\pi_{k+1}} V_{\pi_k} \Rightarrow V_{\pi_k} \leq T_{\pi_{k+1}} V_{\pi_k} \leq T^2_{\pi_{k+1}} V_{\pi_k} \leq \cdots \leq T^n_{\pi_{k+1}} V_{\pi_k},$$

which implies

$$V_{\pi_k} \leq \lim_{n\to+\infty} T^n_{\pi_{k+1}} V_{\pi_k} = V_{\pi_{k+1}}.$$

The same could be achieved using the Q function.

1.8 Application in Statistics: Hidden Markov Models (HMM)

Let $(X_n)_{n\in\mathbb{N}}$ be Markov(λ, P) defined on the state space I. Moreover, let $(S_n)_{n\in\mathbb{N}}$ be a stochastic process defined on a set of signals $\mathcal{S}$. We assume that

$$\mathbb{P}(S_n = s \mid X_n = i, S_{n-1} = s_{n-1}, X_{n-1} = i_{n-1}, \ldots, S_0 = s_0, X_0 = i_0) = \mathbb{P}(S_n = s \mid X_n = i) = \alpha_{is},$$

where, for each $i \in I$, the sequence $(\alpha_{is})_{s\in\mathcal{S}}$ is a probability distribution on $\mathcal{S}$ and called output/emission probabilities. The parameters of this model are the matrices P, $A = (\alpha_{is})_{I,\mathcal{S}}$ and the distribution λ. We denote them by θ and when necessary we will make the conditioning on it explicit.

We observe the signal process S, but the underlying Markov chain X is unobservable. This type of processes is called hidden Markov chain. Notice that S alone is clearly not a Markov chain, but, conditional on X_n, the future $S_n, X_{n+1}, S_{n+1}, \ldots$ is independent of the past $S_{n-1}, X_{n-1}, \ldots, S_0, X_0$.

Define now $\mathbf{S}_n = (S_0, \ldots, S_n)$, $\mathbf{s}_n = (s_0, \ldots, s_n)$, $\mathbf{X}_n = (X_0, \ldots, X_n)$, $\mathbf{i}_n = (i_0, \ldots, i_n)$. By the assumption above, we find that

$$\mathbb{P}(\mathbf{S}_n = \mathbf{s}_n, \mathbf{X}_n = \mathbf{i}_n \mid \theta) = \lambda_{i_0} \prod_{k=1}^{n} p_{i_{k-1},i_k} \alpha_{i_k,s_k}.$$

We are going to analyze three problems under this model: calculating the probability of observing a sequence $\mathbf{s}_n$ (likelihood); finding the most likely hidden path $\mathbf{x}_n$ given an observation $\mathbf{s}_n$ (decoding); and estimating the parameters θ given an observation $\mathbf{s}_n$ (inference).

Calculating Likelihood of a sequence $\mathbf{s}_n$

Notice that

$$\mathbb{P}(\mathbf{S}_n = \mathbf{s}_n) = \sum_{\mathbf{i}_n} \mathbb{P}(\mathbf{S}_n = \mathbf{s}_n, \mathbf{X}_n = \mathbf{i}_n) = \sum_{\mathbf{i}_n} \lambda_{i_0} \prod_{k=1}^{n} p_{i_{k-1},i_k} \alpha_{i_k,s_k}.$$

However, there are too many terms in this sum to be computationally feasible. Define then

$$F_n(\mathbf{s}_n, j) = \mathbb{P}(\mathbf{S}_n = \mathbf{s}_n, X_n = j).$$

Then

$$\mathbb{P}(X_n = j \mid \mathbf{S}_n = \mathbf{s}_n) = \frac{\mathbb{P}(\mathbf{S}_n = \mathbf{s}_n, X_n = j)}{\mathbb{P}(\mathbf{S}_n = \mathbf{s}_n)} = \frac{F_n(\mathbf{s}_n, j)}{\sum_{i\in I} F_n(\mathbf{s}_n, i)}.$$

Moreover

$$\begin{aligned}
F_n(\mathbf{s}_n, j) &= \mathbb{P}(\mathbf{S}_{n-1} = \mathbf{s}_{n-1}, S_n = s_n, X_n = j) \\
&= \sum_{i\in I} \mathbb{P}(\mathbf{S}_{n-1} = \mathbf{s}_{n-1}, X_{n-1} = i, X_n = j, S_n = s_n) \\
&= \sum_{i\in I} \mathbb{P}(\mathbf{S}_{n-1} = \mathbf{s}_{n-1}, X_{n-1} = i)\mathbb{P}(X_n = j, S_n = s_n \mid \mathbf{S}_{n-1} = \mathbf{s}_{n-1}, X_{n-1} = i)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i \in I} F_{n-1}(\mathbf{s}_{n-1}, i)\mathbb{P}(X_n = j, S_n = s_n \mid X_{n-1} = i) \\
&= \sum_{i \in I} F_{n-1}(\mathbf{s}_{n-1}, i)\mathbb{P}(S_n = s_n \mid X_n = j)\mathbb{P}(X_n = j \mid X_{n-1} = i) \\
&= \sum_{i \in I} F_{n-1}(\mathbf{s}_{n-1}, i)\alpha_{j,s_n} p_{ij} = \alpha_{j,s_n} \sum_{i \in I} F_{n-1}(\mathbf{s}_{n-1}, i)p_{ij}.
\end{aligned}
$$

So, we have found a recursive formula for F_n. Notice that the first term is given by

$$F_0(s_0, j) = \mathbb{P}(X_0 = j, S_0 = s_0) = \lambda_j \alpha_{j,s_0}.$$

One can use the probabilities F_n to compute the distribution of a given sequence of signals (i.e. the likelihood of $\mathbf{S}_n$ given the parameters θ):

$$\mathbb{P}(\mathbf{S}_n = \mathbf{s}_n | \theta) = \sum_{j \in I} F_n(\mathbf{s}_n, j | \theta).$$

This is called the forward approach. A different way, called backward approach, uses the probabilities $\mathbb{P}(S_{k+1} = s_{k+1}, \dots, S_n = s_n \mid X_k = i)$.

Most Likely Hidden Path

We will now consider the problem of estimating the hidden states once we observed the $\mathbf{s}_n$, also called decoding. This will be accomplished by maximizing $\mathbb{P}(\mathbf{X}_n = \mathbf{i}_n \mid \mathbf{S}_n = \mathbf{s}_n)$ in $\mathbf{i}_n$. By the conditional probability formula, this is equivalent to maximizing $\mathbb{P}(\mathbf{X}_n = \mathbf{i}_n, \mathbf{S}_n = \mathbf{s}_n)$. To solve this problem, define

$$V_k(j) = \max_{\mathbf{i}_{k=1}} \mathbb{P}(\mathbf{X}_{k-1} = \mathbf{i}_{k-1}, X_k = j, \mathbf{S}_k = \mathbf{s}_k).$$

Similarly to what we have done above:

$$
\begin{aligned}
&\mathbb{P}(\mathbf{X}_{k-1} = \mathbf{i}_{k-1}, X_k = j, \mathbf{S}_k = \mathbf{s}_k) \\
&= \mathbb{P}(S_k = s_k, X_k = j \mid \mathbf{X}_{k-1} = \mathbf{i}_{k-1}, \mathbf{S}_{k-1} = \mathbf{s}_{k-1})\mathbb{P}(\mathbf{X}_{k-1} = \mathbf{i}_{k-1}, \mathbf{S}_{k-1} = \mathbf{s}_{k-1}) \\
&= \mathbb{P}(S_k = s_k, X_k = j \mid X_{k-1} = i_{k-1})\mathbb{P}(\mathbf{X}_{k-1} = \mathbf{i}_{k-1}, \mathbf{S}_{k-1} = \mathbf{s}_{k-1}) \\
&= \mathbb{P}(S_k = s_k \mid X_k = j, X_{k-1} = i_{k-1})\mathbb{P}(X_k = j \mid X_{k-1} = i_{k-1})\mathbb{P}(\mathbf{X}_{k-1} = \mathbf{i}_{k-1}, \mathbf{S}_{k-1} = \mathbf{s}_{k-1}) \\
&= \alpha_{j,s_k} p_{i_{k-1},j}\mathbb{P}(\mathbf{X}_{k-1} = \mathbf{i}_{k-1}, \mathbf{S}_{k-1} = \mathbf{s}_{k-1}).
\end{aligned}
$$

Hence

$$V_k(j) = \alpha_{j,s_k} \max_{i_{k-1}} V_{k-1}(i_{k-1}) p_{i_{k-1},j}.$$

Moreover,

$$V_0(j) = \mathbb{P}(X_0 = j, S_0 = s_0) = \lambda_j \alpha_{j,s_0}.$$

This recursion allows us to compute $V_k(j)$ for each $j \in I$. To find the sequence that maximizes $\mathbb{P}(\mathbf{X}_n = \mathbf{i}_n, \mathbf{S}_n = \mathbf{s}_n)$ proceed as follows: let i_n be the maximizing state for $V_n(j)$, and write

$$V_n(i_n) = \max_{\mathbf{i}_{n-1}} \mathbb{P}(\mathbf{X}_{n-1} = \mathbf{i}_{n-1}, X_n = i_n, \mathbf{S}_n = \mathbf{s}_n) = \alpha_{i_n,s_n} \max_i p_{i,i_n} V_{n-1}(i).$$

Then, define i_{n-1} the value of i that attains the maximum above and perform a recursion of the equation above up to time 0. This is called Viterbi Algorithm.

Parameter Estimation

We want to derive a procedure to find the maximum likelihood estimator:

$$\hat{\theta} = \arg\max_\theta \mathbb{P}(\mathbf{S}_n = \mathbf{s}_n | \theta).$$

Remember that

$$F_k(\mathbf{s}_k, i) = \mathbb{P}(\mathbf{S}_k = \mathbf{s}_k, X_k = i | \theta)$$

and

$$\begin{aligned}\mathbb{P}(\mathbf{S}_n = \mathbf{s}_n, X_k = i|\theta) &= \mathbb{P}(S_{k+1} = s_{k+1}, \dots, S_n = s_n | X_k = i, \mathbf{S}_k = \mathbf{s}_k, \theta)\mathbb{P}(\mathbf{S}_k = \mathbf{s}_k, X_k = i, \theta) \\ &= \mathbb{P}(S_{k+1} = s_{k+1}, \dots, S_n = s_n | X_k = i, \theta) F_k(\mathbf{s}_k, i|\theta).\end{aligned}$$

We then define

$$B_k(\mathbf{s}_{k+1:n}, i|\theta) = \mathbb{P}(S_{k+1} = s_{k+1}, \dots, S_n = s_n | X_k = i, \theta),$$

where $\mathbf{s}_{k+1:n} = (s_{k+1}, \dots, s_n)$. We also define $B_n(\cdot, i) = 1$. Similarly to what we did for F, we can find

$$B_k(\mathbf{s}_{k+1:n}, i|\theta) = \sum_{j\in I} p_{ij}\alpha_{j,s_{k+1}} B_{k+1}(\mathbf{s}_{k+2:n}, j|\theta).$$

Hence

$$\mathbb{P}(X_k = i|\mathbf{S}_n = \mathbf{s}_n, \theta) = \frac{\mathbb{P}(\mathbf{S}_n = \mathbf{s}_n, X_k = i|\theta)}{\mathbb{P}(\mathbf{S}_n = \mathbf{s}_n|\theta)} = \frac{B_k(\mathbf{s}_{k+1:n}, i|\theta)F_k(\mathbf{s}_k, i|\theta)}{\mathbb{P}(\mathbf{S}_n = \mathbf{s}_n|\theta)} =: \gamma_k(\mathbf{s}_n, i|\theta),$$

where

$$\mathbb{P}(\mathbf{S}_n = \mathbf{s}_n|\theta) = \sum_{i\in I}\mathbb{P}(\mathbf{S}_n = \mathbf{s}_n, X_k = i|\theta) = \sum_{i\in I} B_k(\mathbf{s}_{k+1:n}, i|\theta)F_k(\mathbf{s}_k, i|\theta).$$

Moreover, consider

$$\xi_k(\mathbf{s}_n, i, j|\theta) = \mathbb{P}(X_k = i, X_{k+1} = j \mid \mathbf{S}_n = \mathbf{s}_n, \theta)$$

and notice that

$$\xi_k(\mathbf{s}_n, i, j|\theta) = \frac{\mathbb{P}(X_k = i, X_{k+1} = j, \mathbf{S}_n = \mathbf{s}_n|\theta)}{\mathbb{P}(\mathbf{S}_n = \mathbf{s}_n|\theta)} = \frac{F_k(\mathbf{s}_k, i|\theta)p_{ij}\alpha_{j,s_{k+1}}B_{k+1}(\mathbf{s}_{k+2:n}, j|\theta)}{\mathbb{P}(\mathbf{S}_n = \mathbf{s}_n|\theta)}.$$

Moreover

$$\gamma_k(\mathbf{s}_n, i|\theta) = \mathbb{E}[1_{\{X_k = i\}} \mid \mathbf{S}_n = \mathbf{s}_n, \theta] \text{ and } \xi_k(\mathbf{s}_n, i, j|\theta) = \mathbb{E}[1_{\{X_k = i, X_{k+1} = j\}} \mid \mathbf{S}_n = \mathbf{s}_n, \theta],$$

which implies that

$$\sum_{k=1}^{n}\gamma_k(\mathbf{s}_n, i|\theta) = \text{ expected number of visits to state } i$$

and

$$\sum_{k=1}^{n}\xi_k(\mathbf{s}_n, i, j|\theta) = \text{ expected number of transitions from } i \text{ to } j.$$

Hence, we can use γ_k and ξ_k to estimate P, A and λ:

$$\begin{aligned}\bar{p}_{ij} &= \frac{\sum_{k=1}^{n}\xi_k(\mathbf{s}_n, i, j|\theta)}{\sum_{k=1}^{n}\gamma_k(\mathbf{s}_n, i|\theta)}, \\ \bar{\alpha}_{i,s} &= \frac{\sum_{k=1}^{n}\gamma_k(\mathbf{s}_n, i|\theta)1_{\{S_k = s\}}}{\sum_{k=1}^{n}\gamma_k(\mathbf{s}_n, i|\theta)}, \\ \bar{\lambda}_i &= \gamma_0(\mathbf{s}_n, i|\theta).\end{aligned}$$

The Baum–Welch algorithm to estimate θ is given by: given an initial parameter θ_0, iterate the equations above to create sequence θ_k that should converge to $\hat{\theta}$. This is a special case of the Expectation–Maximization (EM) algorithm.

Chapter 2

Discrete-Time Martingales

2.1 Preamble: Conditional Expectation

For a throughout introduction to measure theoretic probability and the statement of results that will be intensively used in what follows, see Chapter 1 of my lecture notes "Probability and Finance".

2.1.1 For L^2 random variables

First of all, it is important to realize the following equivalent definition of the expectation of a r.v. X: $\mathbb{E}[X]$ minimizes the mean-square error, i.e.

$$\mathbb{E}[(X - \mathbb{E}[X])^2] = \min_{x \in \mathbb{R}} \mathbb{E}[(X - x)^2].$$

In words: with no additional information, the best predictor of X, in the sense of minimizing the mean-square error, is its expectation. How does this change if we have access to additional information? For this, let us consider the throwing of a fair dice and X is the number that comes up. The additional information is the parity of that number. If we know the number is even, it is clear that $\mathbb{E}[X \mid \mathit{even}] = (2+4+6)/3 = 4$. Now, it is also clear that $\mathbb{E}[X \mid \mathit{odd}] = (1+3+5)/3 = 3$. Therefore, it is intuitive that $\mathbb{E}[X \mid \mathit{parity}]$ should be a random variable and given by

$$\mathbb{E}[X \mid \mathit{parity}] = \mathbb{E}[X \mid \mathit{odd}] 1_{odd} + \mathbb{E}[X \mid \mathit{even}] 1_{even}$$

Given a random variable X in $(\Omega, \mathcal{F}, \mathbb{P})$, we would like to give a precise mathematical sense to the "closest to" (or the best predictor of) X and measurable with respect to a sub-σ-algebra of $\mathcal{G} \subset \mathcal{F}$. If $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, we have the following definition:

Definition 2.1.1. The *conditional expectation of X with respect to $\mathcal{G}$*, denoted by $\mathbb{E}[X \mid \mathcal{G}]$, is given by

$$\mathbb{E}\left[(X - \mathbb{E}[X \mid \mathcal{G}])^2\right] = \inf_{Y \in L^2(\Omega, \mathcal{G}, \mathbb{P})} \mathbb{E}[(X - Y)^2].$$

In other words, $\mathbb{E}[X \mid \mathcal{G}]$ is the orthogonal projection of X onto the (closed) subspace $L^2(\Omega, \mathcal{G}, \mathbb{P})$. By the orthogonality property, we have

$$\mathbb{E}[(X - \mathbb{E}[X \mid \mathcal{G}])Y] = 0, \tag{2.1}$$

for any $Y \in L^2(\Omega, \mathcal{G}, \mathbb{P})$;

Remark 2.1.2. Notice that by the closeness of $L^2(\Omega, \mathcal{G}, \mathbb{P})$, we have $\mathbb{E}[X \mid \mathcal{G}] \in L^2(\Omega, \mathcal{G}, \mathbb{P})$. In particular, $\mathbb{E}[X \mid \mathcal{G}]$ is measurable with respect to $\mathcal{G}$.

Example 2.1.3. Assume $\mathcal{G}$ is finite, i.e. generated by a partition of Ω, $B_1, \ldots, B_n$. Let's show that

$$\mathbb{E}[X \mid \mathcal{G}] = \sum_{i=1}^{n} \frac{\mathbb{E}[X 1_{B_i}]}{\mathbb{P}(B_i)} 1_{B_i}.$$

It is clear that $\mathcal{G}$-measurable r.v. are of the form: $Y = \sum_{i=1}^n y_i 1_{B_i}$. Hence, we are interested in the following minimization problem:

$$\inf_{y_1,\dots,y_n} \mathbb{E}\left[\left(X - \sum_{i=1}^n y_i 1_{B_i}\right)^2\right] = \inf_{y_1,\dots,y_n} \mathbb{E}\left[\sum_{i=1}^n (X - y_i)^2 1_{B_i}\right].$$

Therefore, taking derivative with respect to y_i, we clearly see that

$$2\mathbb{E}[(X - y_i)1_{B_i}] = 0 \Leftrightarrow y_i = \frac{\mathbb{E}[X1_{B_i}]}{\mathbb{P}(B_i)}.$$

For example, if $\mathcal{G} = \sigma(Y)$, where Y is a discrete random variable taking values $y_1, \dots, y_n$, then

$$\mathbb{E}[X \mid Y] = \mathbb{E}[X \mid \sigma(Y)] = \sum_{i=1}^n \frac{\mathbb{E}[X1_{Y=y_i}]}{\mathbb{P}(Y = y_i)} 1_{Y=y_i}.$$

Corollary 2.1.4. *The conditional expectation* $\mathbb{E}[X \mid \mathcal{G}]$ *is unique a.s.*

Proof. The proof follows easily from the orthogonality property, Equation (2.1). If there exists $Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})$ that is also the orthogonal projection of X onto $L^2(\Omega, \mathcal{G}, \mathbb{P})$, then

$$\mathbb{E}[(X - \mathbb{E}[X \mid \mathcal{G}])1_B] = 0 = \mathbb{E}[(X - Z)1_B], \ \forall\, B \in \mathcal{G}.$$

Hence, $\mathbb{E}[(Z - \mathbb{E}[X \mid \mathcal{G}])1_B] = 0$, for any $B \in \mathcal{G}$, and this implies that $Z = \mathbb{E}[X \mid \mathcal{G}]$ a.s.

□

Proposition 2.1.5. *The conditional expectation satisfies the following properties:*

1. *Normalization:* $\mathbb{E}[1 \mid \mathcal{G}] = 1$;
2. *Linearity:* $\mathbb{E}[\alpha X + Y \mid \mathcal{G}] = \alpha\mathbb{E}[X \mid \mathcal{G}] + \mathbb{E}[Y \mid \mathcal{G}]$;
3. *Positivity:* $X \geq 0 \Rightarrow \mathbb{E}[X \mid \mathcal{G}] \geq 0$;
4. *Monotonicity:* $X \geq Y \Rightarrow \mathbb{E}[X \mid \mathcal{G}] \geq \mathbb{E}[Y \mid \mathcal{G}]$;

The proof follows from the orthogonality property, see Equation (2.1).

2.1.2 For L^1 random variables

Since the expectation $\mathbb{E}$ is defined for any L^1 random variable, we would like to define $\mathbb{E}[\cdot \mid \mathcal{G}]$ for the same class of r.v's. This is done using the orthogonality property. More precisely, that property implies the *covariance matching* property:

$$\mathbb{E}[XY] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]\, Y], \ \forall\, Y \in L^2(\Omega, \mathcal{G}, \mathbb{P}).$$

Hence, we may consider a smaller class of random variable for Y to be able to consider a wider class of random variable for X.

Definition 2.1.6. Consider a r.v. $X \in L^1$. We say $Z = \mathbb{E}[X \mid \mathcal{G}]$ is the *conditional expectation of* X *with respect to* $\mathcal{G} \subset \mathcal{F}$ if

(i) Z is $\mathcal{G}$-measurable;

(ii) $\mathbb{E}[X\ 1_B] = \mathbb{E}[Z\ 1_B]$, for any $B \in \mathcal{G}$.

Theorem 2.1.7. *For any* $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$*, the conditional expectation* $\mathbb{E}[X \mid \mathcal{G}]$ *exists and it is unique.*

Proof. The uniqueness follows as in the L^2 case. To prove existence, consider the case when $X \geq 0$. The general case is proved by taking the positive and negative part of a general random variable. Define now $X_n = X \wedge n$. Notice that X_n is bounded and hence in $L^2(\Omega, \mathcal{F}, \mathbb{P})$. We are then able to consider $Z_n = \mathbb{E}[X_n \mid \mathcal{G}]$. Notice that, by Properties 3 and 4 of the conditional expectation, $Z_n \geq 0$ and $Z_n \leq Z_{n+1}$. Hence, we may define the limit $Z = \lim_{n\to+\infty} Z_n$ and the limit Z is $\mathcal{G}$-measurable. Now, by the MCT:

$$\mathbb{E}[X 1_B] = \lim_{n\to+\infty} \mathbb{E}[X_n 1_B] = \lim_{n\to+\infty} \mathbb{E}[Z_n 1_B] = \mathbb{E}[Z 1_B],$$

for any $B \in \mathcal{G}$.

□

Example 2.1.8. Take two r.v's X and Y that have a (continuous) joint probability density function $f_{X,Y}$ in $\mathbb{R}^2$. Consider $g : \mathbb{R} \longrightarrow \mathbb{R}$ Borel-measurable and such that $g(X)$ is integrable. We would like to compute (a version of) $\mathbb{E}[g(X) \mid Y] = \mathbb{E}[g(X) \mid \sigma(Y)]$. First, notice that $\sigma(Y) = \{\{Y \in B\}\ ;\ B \in \mathcal{B}(\mathbb{R})\}$. Moreover, this conditional expectation needs to satisfy $\mathbb{E}[\mathbb{E}[g(X) \mid Y] 1_{\{Y\in B\}}] = \mathbb{E}[g(X) 1_{\{Y\in B\}}]$, for any $B \in \mathcal{B}(\mathbb{R})$. The r.v. inside the expectation in the left-hand side depends only on Y and hence (formally)

$$\mathbb{E}[\mathbb{E}[g(X) \mid Y] 1_{\{Y\in B\}}] = \int_B \mathbb{E}[g(X) \mid Y = y] f_Y(y) dy,$$

where f_Y is the marginal density of Y:

$$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y) dx.$$

Therefore, we want to find a similar expression for $\mathbb{E}[g(X) 1_B]$:

$$\mathbb{E}[g(X) 1_{\{Y\in B\}}] = \int_B \int_{\mathbb{R}} g(x) f_{X,Y}(x, y) dx dy = \int_B \left(\int_{\mathbb{R}} g(x) \frac{f_{X,Y}(x, y)}{f_Y(y)} dx \right) f_Y(y) dy,$$

where we understand $0/0 = 0$. Define then

$$\phi(y) = \int_{\mathbb{R}} g(x) \frac{f_{X,Y}(x, y)}{f_Y(y)} dx,$$

for $f_Y(y) > 0$, and 0 otherwise. Thus $\mathbb{E}[g(X) 1_{\{Y\in B\}}] = \mathbb{E}[\phi(Y) 1_{\{Y\in B\}}]$, which implies $\mathbb{E}[g(X) \mid Y] = \phi(Y)$ a.s. We may understand this result as

$$\mathbb{E}[g(X) \mid Y] = \int_{\mathbb{R}} g(x) \mu(Y, dx),$$

where $\mu(y, dx) = \frac{f_{X,Y}(x,y)}{f_Y(y)} 1_{\{f_Y(y)>0\}} dx = f_{X|Y}(x|y) dx$. The function $f_{X|Y}(\cdot|y)$ is then called the *conditional probability density of X given $Y = y$.*

Remark 2.1.9. All the properties outlined in Proposition 2.1.5 are satisfied for $X \in L^1$.

Definition 2.1.10. Two sub-σ-algebras $\mathcal{G}$ and $\mathcal{H}$ are said to be independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B), \forall\ A \in \mathcal{G} \text{ and } B \in \mathcal{H}.$$

Moreover, a r.v. X is independent of $\mathcal{G}$ if $\sigma(X)$ is independent of $\mathcal{G}$.

Proposition 2.1.11 (More Properties of the Conditional Expectation)**.** *The conditional expectation satisfies the following additional properties:*

6. *Tower property: $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F} \Rightarrow \mathbb{E}[\mathbb{E}[X \mid \mathcal{H}] \mid \mathcal{G}] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}] = \mathbb{E}[X \mid \mathcal{H}]$;*

7. *Pulling out what's known: if Y is $\mathcal{G}$-measurable and $XY \in L^1$, then $\mathbb{E}[XY \mid \mathcal{G}] = Y\mathbb{E}[X \mid \mathcal{G}]$;*

8. *Independence: if X is independent of $\mathcal{G}$, then $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X]$;*

9. *Jensen Inequality: if $\phi : \mathbb{R} \longrightarrow \mathbb{R}$ is convex, then $\mathbb{E}[\phi(X) \mid \mathcal{G}] \geq \phi(\mathbb{E}[X \mid \mathcal{G}])$;*

10. *Monotone Convergence Theorem (MCT): if $(X_n)_{n\in\mathbb{N}}$ is a sequence of non-negative r.v's converging to X and such that $X_n \leq X_{n+1}$, then $\mathbb{E}[X_n \mid \mathcal{G}] \nearrow \mathbb{E}[X \mid \mathcal{G}]$ a.s.*

11. *Dominated Convergence Theorem (DCT): if $(X_n)_{n\in\mathbb{N}}$ is a sequence of r.v's converging to X and such that $|X_n| \leq Y \in L^1$, then $\mathbb{E}[X_n \mid \mathcal{G}] \to \mathbb{E}[X \mid \mathcal{G}]$ a.s. Moreover, $\mathbb{E}[|X_n - X| \mid \mathcal{G}] \to 0$ a.s.*

12. *Fatou's Lemma: if $(X_n)_{n\in\mathbb{N}}$ is a sequence of non-negative r.v's, then $\mathbb{E}[\liminf_{n\to+\infty} X_n \mid \mathcal{G}] \leq \liminf_{n\to+\infty} \mathbb{E}[X_n \mid \mathcal{G}]$ a.s.*

Proposition 2.1.12. *Consider a $\mathcal{G}$-measurable r.v. Y and a bounded and measurable function $\phi : \mathbb{R} \times \Omega \longrightarrow \mathbb{R}$. Assume that, for every $y \in \mathbb{R}$, the r.v. $\phi(y, \cdot)$ is independent of $\mathcal{G}$. Then*

$$\mathbb{E}[\phi(Y, \omega) \mid \mathcal{G}] = \psi(Y),$$

where $\psi(y) = \mathbb{E}[\phi(y, \cdot)]$.

2.1.3 Conditional Probability

When learning conditioning in Probability Theory for the first time, we see the notion of *conditional probability*:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \tag{2.2}$$

This notion makes a lot of sense: we are re-weighting the chances of A by the fact that B happened. Using this notion, a lot of results regarding conditional probability and expectation are derived. However, this is not enough for the generality we require for the study of martingales. That is why we developed the notions of conditional expectation above. We have not discussed conditional probability yet and the notion of conditional expectation did not depend on some notion of conditional probability. In this more general setting, we consider the opposite direction: the definition of conditional probability will be given using the notion of conditional expectation. Remember that, we can write: $\mathbb{P}(A) = \mathbb{E}[1_A]$. Considering this, we define:

Definition 2.1.13 (Conditional Probability)**.** For any $A \in \mathcal{F}$, the *conditional probability of A given $\mathcal{G}$* is given by

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

Notice that $\mathbb{P}(\cdot \mid \mathcal{G})$ is a (integrable) random variable, not a number as in (2.2). It is straightforward to prove it is a probability measure taking values in L^1. Indeed, consider $(A_n)_{n\in\mathbb{N}}$ disjoints. Then, by the MCT

$$\mathbb{P}\left(\bigcup_{n=1}^{+\infty} A_n \;\middle|\; \mathcal{G}\right) = \mathbb{E}\left[\sum_{n=1}^{+\infty} 1_{A_n} \;\middle|\; \mathcal{G}\right] = \sum_{n=1}^{+\infty} \mathbb{E}[1_{A_n} \mid \mathcal{G}] = \sum_{n=1}^{+\infty} \mathbb{P}(A_n \mid \mathcal{G}) \text{ a.s.}$$

We would like to point out that the set of probability 1 where the equation above holds depends on the sequence $(A_n)_{n\in\mathbb{N}}$. Therefore, there may not exist a version of the conditional probability $\mathbb{P}(\cdot \mid \mathcal{G})$ such that the countable additivity property above holds for any sequence $(A_n)_{n\in\mathbb{N}}$. This may happen because there are uncountable many sequences $(A_n)_{n\in\mathbb{N}}$. This justifies the need for the following definition:

Definition 2.1.14 (Regular Conditional Probability)**.** A family of probability distributions on $(\Omega, \mathcal{F})$ denoted by $(\mu(\omega, \cdot))_{\omega\in\Omega}$ is called a *regular conditional probability* given $\mathcal{G}$ if, for each $A \in \mathcal{F}$, $\mu(\cdot, A) = \mathbb{P}(A \mid \mathcal{G})$ a.s.

Regarding the existence of regular conditional probabilities, we have the following theorem, which we state without proving.

Theorem 2.1.15. *For any real r.v. X defined on $(\Omega, \mathcal{F}, \mathbb{P})$, there exists a regular conditional probability for $\mathbb{P} \circ X^{-1}$, where $(\mathbb{P} \circ X^{-1})(F) = \mathbb{P}(X \in F)$.*

2.2 Definition and Basic Properties

A filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ is a sequence of σ-algebras satisfying $\mathcal{F}_n \subset \mathcal{F}_{n+1}$. Usually, we will assume $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Filtrations model the flow of information. For example, given a sequence of r.v's $(X_n)_{n\in\mathbb{N}}$, the σ-algebras defined as $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ is a filtration. Additionally, we use the notation $Y \in \mathcal{F}_n$ meaning that Y is $\mathcal{F}_n$-measurable. In the case where $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$, it means there exists f measurable such that $Y = f(X_1, \dots, X_n)$. A set A belongs to $\mathcal{F}_n$ if $1_A \in \mathcal{F}_n$.

A martingale is a mathematical model that describes a “fair” game.

Definition 2.2.1 (Martingale)**.** A sequence of r.v's $(M_n)_{n\in\mathbb{N}}$ is said a *martingale* with respect to filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ if, for each $n \in \mathbb{N}$,

(i) Adaptedness: $M_n \in \mathcal{F}_n$;

(ii) Integrability: $\mathbb{E}[|M_n|] < +\infty$;

(iii) Martingale property: $\mathbb{E}[M_n \mid \mathcal{F}_{n-1}] = M_{n-1}$.

We could further assume that $\mathbb{P}(M_0 = c) = 1$, for some $c \in \mathbb{R}$. For simplicity, we may considered the case where the information flow is given through a sequence of r.v's $(X_n)_{n\in\mathbb{N}}$.

Remark 2.2.2. Notice that the martingale property is equivalent to $\mathbb{E}[(M_n - M_{n-1})1_A] = 0$, for all $A \in \mathcal{F}_{n-1}$. Moreover, by the tower property, $\mathbb{E}[M_N \mid \mathcal{F}_n] = M_n$, for any $n \leq N$.

Remark 2.2.3. The martingale property implies that the martingale has constant expectation. Indeed, by the tower property, we find:

$$\mathbb{E}[M_n] = \mathbb{E}[\mathbb{E}[M_n \mid \mathcal{F}_{n-1}]] = \mathbb{E}[M_{n-1}].$$

2.2.1 Examples

Let us now see some examples of martingales:

(1) Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent r.v's with $\mathbb{E}[X_n] = 0$ and define

$$S_n = \sum_{i=1}^{n} X_i.$$

We assume $S_0 = 0$. Let us verify that $(S_n)_{n\in\mathbb{N}}$ is a martingale with respect to $(X_n)_{n\in\mathbb{N}}$. For (i), consider $f_n(x_1, \dots, x_n) = \sum_{i=1}^{n} x_i$, which is clearly measurable. Moreover, it is easy to see that $\mathbb{E}[|S_n|] \leq \sum_{i=1}^{n} \mathbb{E}[|X_i|] = n\mathbb{E}[|X_1|] < +\infty$. Finally, notice that

$$\begin{aligned}
\mathbb{E}[S_n \mid \mathcal{F}_{n-1}] &= \sum_{i=1}^{n} \mathbb{E}[X_i \mid \mathcal{F}_{n-1}] \\
&= \sum_{i=1}^{n-1} X_i + \underbrace{\mathbb{E}[X_n \mid \mathcal{F}_{n-1}]}_{=\mathbb{E}[X_n],\text{ by indep.}} \\
&= S_{n-1} + \cancelto{0}{\mathbb{E}[X_n]} = S_{n-1}
\end{aligned}$$

(2) Take X as above but satisfying $\mathrm{Var}(X_n) = \sigma^2 < +\infty$. Notice that

$$\begin{aligned}
\mathbb{E}[S_n^2] &= \mathbb{E}[(S_{n-1} + X_n)^2] = \mathbb{E}[S_{n-1}^2] + 2\mathbb{E}[S_{n-1}X_n] + \mathbb{E}[X_n^2] \\
&= \mathbb{E}[S_{n-1}^2] + 2\mathbb{E}[S_{n-1}]\cancelto{0}{\mathbb{E}[X_n]} + \sigma^2 = \mathbb{E}[S_{n-1}^2] + \sigma^2,
\end{aligned}$$

which implies that $\mathbb{E}[S_n^2] = n\sigma^2$. So, S_n^2 is not a martingale. However, define $M_n = S_n^2 - n\sigma^2$. Notice that (i) and (ii) are easily verified. To verify the martingale property:

$$\begin{aligned}
\mathbb{E}[M_n|\mathcal{F}_{n-1}] &= \mathbb{E}[S_n^2 - n\sigma^2|\mathcal{F}_{n-1}] \\
&= \mathbb{E}[(S_{n-1} + X_n)^2 - n\sigma^2|\mathcal{F}_{n-1}] \\
&= \mathbb{E}[S_{n-1}^2 + 2S_{n-1}X_n + X_n^2 - n\sigma^2|\mathcal{F}_{n-1}] \\
&= \mathbb{E}[S_{n-1}^2|\mathcal{F}_{n-1}] \\
&+ 2\mathbb{E}[S_{n-1}X_n|\mathcal{F}_{n-1}] + \mathbb{E}[X_n^2|\mathcal{F}_{n-1}] - n\sigma^2 \\
&= S_{n-1}^2 + 2S_{n-1}\cancelto{0}{\mathbb{E}[X_n]} + \cancelto{\sigma^2}{\mathbb{E}[X_n^2]} - n\sigma^2 \\
&= S_{n-1}^2 - (n-1)\sigma^2 = M_{n-1}
\end{aligned}$$

(3) Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent r.v's with $X_n \geq 0$ and $\mathbb{E}[X_n] = 1$. Define $M_n = \prod_{i=1}^{n} X_i$ with $M_0 = 1$. Let us verify the martingale property:

$$\begin{aligned}\mathbb{E}[M_n|\mathcal{F}_{n-1}] &= \mathbb{E}[M_{n-1}X_n|\mathcal{F}_{n-1}] \\ &= M_{n-1} \underbrace{\mathbb{E}[X_n \mid \mathcal{F}_{n-1}]}_{=\mathbb{E}[X_n],\text{ by indep.}} = M_{n-1}\cancelto{1}{\mathbb{E}[X_n]} = M_{n-1}\end{aligned}$$

(4) Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of iid r.v's with moment generation function $\phi(\lambda) = \mathbb{E}[e^{\lambda X_n}] < +\infty$. Notice that $e^{\lambda X_n}/\phi(\lambda)$ has mean 1. Hence, using the example above, we may conclude that

$$M_n = \frac{e^{\lambda \sum_{i=1}^{n} X_i}}{\phi(\lambda)^n}$$

is a martingale. This is a parametric family of martingales indexed by λ. If $\lambda_0 \neq 0$ satisfies $\phi(\lambda_0) = 1$, then $e^{\lambda_0 S_n}$ is a martingale.

2.2.2 Additional Properties

Proposition 2.2.4. *Let $(M_n^{(k)})_{n\in\mathbb{N}}$ be a martingale such that $M_n^{(k)} \to M_n$ a.s. and in L^1, when $k \to +\infty$. Then $(M_n)_{n\in\mathbb{N}}$ is a martingale.*

Proof. Notice that, for any $A \in \mathcal{F}_{n-1}$,

$$\begin{aligned}\mathbb{E}\left[(M_n - M_{n-1})\ 1_A\right] &= \mathbb{E}\left[\lim_{k\to+\infty}(M_n^{(k)} - M_{n-1}^{(k)})\ 1_A\right] \\ &= \lim_{k\to+\infty} \mathbb{E}\left[(M_n^{(k)} - M_{n-1}^{(k)})\ 1_A\right] = 0.\end{aligned}$$

□

Definition 2.2.5 (Sub-Martingale)**.** $(M_n)_{n\in\mathbb{N}}$ is a *sub-martingale* with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$ if satisfies (i) and (ii) of the martingale definition and the *sub-martingale property*:

$$\mathbb{E}[M_n \mid \mathcal{F}_{n-1}] \geq M_{n-1},\ \forall\ n \in \mathbb{N}.$$

Proposition 2.2.6. *Let $(M_n)_{n\in\mathbb{N}}$ be a sub-martingale and consider ϕ convex and increasing. Then, $(\phi(M_n))_{n\in\mathbb{N}}$ is a sub-martingale. If $(M_n)_{n\in\mathbb{N}}$ is a martingale, we do not need to assume that ϕ is increasing.*

Proof. It follows directly from Jensen inequality:

$$\mathbb{E}[\phi(M_n) \mid \mathcal{F}_{n-1}] \geq \phi(\mathbb{E}[M_n \mid \mathcal{F}_{n-1}]) \geq \phi(M_{n-1}),\ \forall\ n \in \mathbb{N}.$$

□

Definition 2.2.7 (Non-Anticipative)**.** A sequence of r.v's $(A_n)_{n\in\mathbb{N}}$ is *non-anticipative process* with respect to the filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ if $A_n \in \mathcal{F}_{n-1}$, for all $n \in \mathbb{N}$.

Definition 2.2.8 (Martingale Transformation)**.** The *martingale transformation* of the process $(M_n)_{n\in\mathbb{N}}$ by the process $(A_n)_{n\in\mathbb{N}}$ is defined as

$$(A \cdot M)_n = M_0 + \sum_{i=1}^{n} A_i(M_i - M_{i-1}).$$

Notice that this is a discretized version of what we would like to call the stochastic integral of A with respect to M; in symbols $\int AdM$.

Theorem 2.2.9. *Assume $(M_n)_{n\in\mathbb{N}}$ is a martingale with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$ and $(A_n)_{n\in\mathbb{N}}$ is a sequence of bounded r.v's and non-anticipative with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$. Then, the martingale transformation of $(M_n)_{n\in\mathbb{N}}$ by $(A_n)_{n\in\mathbb{N}}$ is also a martingale with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$.*

Proof. Note that property (i) of the martingale definition is clearly satisfied. To verify property (ii), notice that

$$\mathbb{E}[|(A \cdot M)_n|] \leq \mathbb{E}[|M_0|] + 2C \sum_{i=1}^{n} \mathbb{E}[|M_i|],$$

where C is the bound of the process $(A_n)_{n\in\mathbb{N}}$. Let us now verify the martingale property:

$$\begin{aligned}\mathbb{E}[(A \cdot M)_n - (A \cdot M)_{n-1}|\mathcal{F}_{n-1}] &= \mathbb{E}[A_n(M_n - M_{n-1})|\mathcal{F}_{n-1}] \\ &= A_n\mathbb{E}[M_n - M_{n-1}|\mathcal{F}_{n-1}] = 0\end{aligned}$$

$\square$

For the next two propositions, we assume that $M_0 = 0$. With small modifications, the results hold true in the general case.

Proposition 2.2.10. *Let $(M_n)_{n\in\mathbb{N}}$ be a stochastic process adapted to $(\mathcal{F}_n)_{n\in\mathbb{N}}$. Then, $(M_n)_{n\in\mathbb{N}}$ is a martingale if and only if*

$$\mathbb{E}\left[(A \cdot M)_n\right] = 0,$$

for all $n \in \mathbb{N}$ and all non-anticipative, bounded processes $(A_n)_{n\in\mathbb{N}}$.

Proof. The direction $(\Rightarrow)$ is trivial. For the other direction, for any $j \in \mathbb{N}$ and $B \in \mathcal{F}_j$, consider the non-anticipative processes $(A_n)_{n\in\mathbb{N}}$ defined as $A_{j+1} = 1_B$, and $A_i = 0$, if $i \neq j+1$. Then, $(A \cdot M)_n = 1_B(M_{j+1} - M_j)$, if $n \geq j+1$ and 0, otherwise. Hence,

$$\mathbb{E}\left[1_B(M_{j+1} - M_j)\right] = 0, \ \forall\ B \in \mathcal{F}_j,$$

which implies $\mathbb{E}\left[M_{j+1} \mid \mathcal{F}_j\right] = M_j$, as desired. $\square$

Proposition 2.2.11. *If $(M_n)_{n\in\mathbb{N}}$ is a submartingale and $(A_n)_{n\in\mathbb{N}}$ be a non-anticipative process satisfying $0 \leq A \leq C$, then*

$$\mathbb{E}\left[(A \cdot M)_n\right] \leq C\mathbb{E}[M_n],$$

for all $n \in \mathbb{N}$.

Proof. Notice that

$$\begin{aligned}\mathbb{E}[(A \cdot M)_n] &= \mathbb{E}\left[\sum_{k=1}^{n} A_k(M_k - M_{k-1})\right] \\ &= \sum_{k=1}^{n} \mathbb{E}\left[A_k(M_k - M_{k-1})\right] = \sum_{k=1}^{n} \mathbb{E}\left[\mathbb{E}\left[A_k(M_k - M_{k-1}) \mid \mathcal{F}_{k-1}\right]\right] \\ &= \sum_{k=1}^{n} \mathbb{E}\Big[A_k \underbrace{\mathbb{E}\left[(M_k - M_{k-1}) \mid \mathcal{F}_{k-1}\right]}_{\geq 0}\Big] \leq C \sum_{k=1}^{n} \mathbb{E}\Big[\mathbb{E}\left[(M_k - M_{k-1}) \mid \mathcal{F}_{k-1}\right]\Big] \\ &= C \sum_{k=1}^{n} \mathbb{E}\Big[M_k - M_{k-1}\Big] = C\mathbb{E}[M_n].\end{aligned}$$

$\square$

Theorem 2.2.12 (Doob's Decomposition Theorem)**.** *Let $(Y_n)_{n\in\mathbb{N}}$ be a submartingale. There exists a martingale $(M_n)_{n\in\mathbb{N}}$ and an increasing non-antecipative process $(A_n)_{n\in\mathbb{N}}$ starting at 0 such that*

$$Y_n = M_n + A_n.$$

This decomposition is unique a.s.

Proof. Define M and A as follows $M_0 = Y_0$, $A_0 = 0$ and

$$\begin{aligned}M_{n+1} &= M_n + Y_{n+1} - \mathbb{E}\left[Y_{n+1} \mid \mathcal{F}_n\right], \\ A_{n+1} &= A_n + \mathbb{E}\left[Y_{n+1} \mid \mathcal{F}_n\right] - Y_n.\end{aligned}$$

It is easy to prove that M and A satisfy the desired properties. To prove uniqueness, let M' and A' be another Doob's decomposition. Then

$$Y_{n+1} - Y_n = (M'_{n+1} - M'_n) + (A'_{n+1} - A'_n) = (M_{n+1} - M_n) + (A_{n+1} - A_n).$$

Taking the conditional expectation with respect to $\mathcal{F}_n$, we find $A'_{n+1} - A'_n = A_{n+1} - A_n$. Since A and A' both start at 0, we find $A_n = A'_n$, which implies that $M_n = M'_n$.

□

2.3 Optional Stopping Theorems

Definition 2.3.1 (Stopping Time)**.** A r.v. τ taking value is $\mathbb{N} \cup \{+\infty\}$ is a *stopping time* with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$ if

$$\{\tau \leq n\} \in \mathcal{F}_n, \ \forall \ n \in \mathbb{N}.$$

Intuitively, if we use the r.v. τ to decide when we are going to stop a game, this decision could not depend on the events yet to come. For example, when giving directions, saying that you should turn on the first right is a stopping time, but saying you should turn on the last right is NOT a stopping time.

Definition 2.3.2 (Stopped Process)**.** When τ is a finite stopping time ($\mathbb{P}(\tau < +\infty) = 1$), Y_τ is defined as

$$Y_\tau = \sum_{k=0}^{+\infty} 1_{\{\tau=k\}} Y_k.$$

Additionally, notice that for any stopping τ, $n \wedge \tau = \min\{n, \tau\}$ is a finite stopping time and, in this case,

$$Y_{n\wedge\tau} = \sum_{k=0}^{+\infty} 1_{\{n\wedge\tau=k\}} Y_k = \sum_{k=0}^{n-1} 1_{\{\tau=k\}} Y_k + Y_n 1_{\{\tau\geq n\}}.$$

We call $(Y_{n\wedge\tau})_{n\in\mathbb{N}}$ the *stopped process.*

Theorem 2.3.3. *If $(M_n)_{n\in\mathbb{N}}$ is a martingale with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$, then the stopped process $(M_{n\wedge\tau})_{n\in\mathbb{N}}$ is also a martingale with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$.*

Proof. Without loss of generality, we may assume $M_0 = 0$. Note that $A_k = 1_{\{\tau\geq k\}}$, for $k \in \mathbb{N}$, are bounded r.v's and non-anticipative, since τ is a stopping time. This follows the fact $A_k = 1 - 1_{\{\tau\leq k-1\}}$ and $1_{\{\tau\leq k-1\}} \in \mathcal{F}_{k-1}$. Finally,

$$\begin{aligned}
\sum_{k=1}^{n} A_k(M_k - M_{k-1}) &= \sum_{k=1}^{n} 1_{\{\tau\geq k\}}(M_k - M_{k-1}) \\
&= \sum_{k=1}^{n} 1_{\{\tau=k\}} M_k + \sum_{k=1}^{n} 1_{\{\tau\geq k+1\}} M_k - \sum_{k=1}^{n} 1_{\{\tau\geq k\}} M_{k-1} \\
&= \sum_{k=1}^{n} 1_{\{\tau=k\}} M_k + \sum_{k=2}^{n+1} 1_{\{\tau\geq k\}} M_{k-1} - \sum_{k=1}^{n} 1_{\{\tau\geq k\}} M_{k-1} \\
&= \sum_{k=1}^{n} 1_{\{\tau=k\}} M_k + 1_{\{\tau\geq n+1\}} M_n \\
&= \sum_{k=1}^{n-1} 1_{\{\tau=k\}} M_k + 1_{\{\tau\geq n\}} M_n = M_{n\wedge\tau}.
\end{aligned}$$

This means that $(M_{n\wedge\tau})_{n\in\mathbb{N}}$ is a martingale transformation of the martingale $(M_n)_{n\in\mathbb{N}}$ by the bounded and non-anticipative process $(A_n)_{n\in\mathbb{N}}$ defined above. Therefore, $(M_{n\wedge\tau})_{n\in\mathbb{N}}$ is a martingale with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$.

□

Remark 2.3.4. This does not imply that $\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$. Additional assumptions must be made.

Theorem 2.3.5 (First Optional Stopping Theorem). *Let $(M_n)_{n\in\mathbb{N}}$ a martingale and τ a finite stopping time ($\tau < +\infty$ a.s.) such that $(M_{\tau\wedge n})_{n\in\mathbb{N}}$ is bounded. Then*

$$\mathbb{E}[M_\tau] = \mathbb{E}[M_0].$$

Proof. Since $\tau < +\infty$ a.s, $M_{\tau\wedge n} \longrightarrow M_\tau$ a.s, as $n \to +\infty$. Moreover, by the boundedness of $(M_{\tau\wedge n})_{n\in\mathbb{N}}$, one can apply the DCT to conclude

$$\mathbb{E}[M_\tau] = \lim_{n\to+\infty} \mathbb{E}[M_{\tau\wedge n}] = \mathbb{E}[M_0],$$

by the theorem above. □

Theorem 2.3.6 (Doob's Optional Stopping Theorem). *Let $(M_n)_{n\in\mathbb{N}}$ be a martingale and τ a stopping time. Then, $\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$, if one of the following assumptions holds:*

(i) τ *is bounded;*

(ii) $(M_{\tau\wedge n})_{n\in\mathbb{N}}$ *is bounded and* $\tau < +\infty$ *a.s;*

(iii) $\mathbb{E}[\tau] < +\infty$ *and* $|M_k - M_{k-1}| \leq C$*, for some* $C > 0$*;*

(iv) $\mathbb{E}[\tau] < +\infty$ *and* $\mathbb{E}\left[|M_k - M_{k-1}| \mid \mathcal{F}_{k-1}\right] \leq C$*, for some* $C > 0$*;*

Proof. We have already proved this result under assumptions (i) and (ii). Let us prove (iii). Notice now that

$$M_{\tau\wedge n} = M_0 + \sum_{k=0}^{n-1} (M_{k+1} - M_k) 1_{\{\tau > k\}}.$$

Because $\tau < +\infty$ a.s,

$$M_{\tau\wedge n} \overset{n\to+\infty}{\longrightarrow} M_\tau.$$

Additionally,

$$|M_{\tau\wedge n}| \leq |M_0| + \sum_{k=0}^{n} |M_{k+1} - M_k| 1_{\{\tau>k\}} \leq |M_0| + C\sum_{k=0}^{n} 1_{\{\tau>k\}} \leq |M_0| + C\sum_{k=0}^{+\infty} 1_{\{\tau>k\}}.$$

Finally, notice

$$\mathbb{E}\left[\sum_{k=0}^{+\infty} 1_{\{\tau>k\}}\right] = \sum_{k=0}^{+\infty} \mathbb{P}(\tau > k) = \mathbb{E}[\tau] < +\infty,$$

which implies we can apply the DCT to conclude the result. Part (iv) follows similarly. □

It is interesting to investigate what would be the information available at a stopping time τ:

Definition 2.3.7.

$$\mathcal{F}_\tau = \{A \in \mathcal{F}\ ; A \cap \{\tau \leq n\} \in \mathcal{F}_n,\ \forall\ n \in \mathbb{N}\}.$$

Proposition 2.3.8. *Let τ and σ be stopping times such that $\tau \leq \sigma$. Then $\mathcal{F}_\tau \subset \mathcal{F}_\sigma$.*

Proof. Consider $A \in \mathcal{F}_\tau$. Notice that $\{\sigma \leq n\} \subset \{\tau \leq n\}$. Then

$$A \cap \{\sigma \leq n\} = \underbrace{A \cap \{\tau \leq n\}}_{\in\mathcal{F}_n} \cap \underbrace{\{\sigma \leq n\}}_{\in\mathcal{F}_n} \in \mathcal{F}_n$$

finishing the proof. □

Theorem 2.3.9 (Optional Sampling Theorem). *Let $(M_n)_{n\in\mathbb{N}}$ be a martingale and consider τ and σ bounded stopping times with $\tau \leq \sigma$. Then*

$$\mathbb{E}\left[M_\sigma \mid \mathcal{F}_\tau\right] = M_\tau.$$

Proof. Boundedness here means that $\tau \leq \sigma \leq N$, for some $N > 0$. Let ρ be one of the bounded stopping times (it does not matter which one now). We clearly have $|M_\rho| \leq \sum_{i=1}^N |M_i|$, implying that this r.v. is integrable. Let us prove that $M_\rho = \mathbb{E}\left[M_N \mid \mathcal{F}_\rho \right]$. For any $A \in \mathcal{F}_\rho$, we get

$$\begin{aligned}\mathbb{E}[M_\rho \ 1_A] &= \sum_{k=1}^N \mathbb{E}[M_\rho \ 1_A \ 1_{\{\rho=k\}}] = \sum_{k=1}^N \mathbb{E}[M_k \ \underbrace{1_{A\cap\{\rho=k\}}}_{\in\mathcal{F}_k}] \\ &= \sum_{k=1}^N \mathbb{E}[M_N \ 1_A \ 1_{\{\rho=k\}}] = \mathbb{E}[M_N \ 1_A].\end{aligned}$$

Hence, since $\mathcal{F}_\tau \subset \mathcal{F}_\sigma$, we find

$$\mathbb{E}\left[M_\sigma \mid \mathcal{F}_\tau \right] = \mathbb{E}\left[\mathbb{E}\left[M_N \mid \mathcal{F}_\sigma \right] \mid \mathcal{F}_\tau \right] = \mathbb{E}\left[M_N \mid \mathcal{F}_\tau \right] = M_\tau.$$

□

Remark 2.3.10. In the proof above, we have proved the following result: if τ is a bounded stopping time (by N), then $\mathbb{E}\left[M_N \mid \mathcal{F}_\tau \right] = M_\tau$.

2.4 Doob's Inequalities

Definition 2.4.1 (Maximal Sequence)**.** The *maximal sequence* of $(X_n)_{n\in\mathbb{N}}$ is defined as

$$X_n^* = \sup_{k\leq n} X_k.$$

Lemma 2.4.2 (Index Shifting Inequality)**.** *Let $(M_n)_{n\in\mathbb{N}}$ be any sub-martingale. Then, for $n \geq k$,*

$$\mathbb{E}[M_k 1_A] \leq E[M_n 1_A], \ \forall \ A \in \mathcal{F}_k.$$

Proof. It is easy to see that the sub-martingale property (and, of course, the martingale property as well) is equivalent to

$$M_k \leq \mathbb{E}[M_n \mid \mathcal{F}_k], \ \forall \ n \geq k.$$

Hence, for any $A \in \mathcal{F}_k$,

$$M_k 1_A \leq 1_A \mathbb{E}[M_n \mid \mathcal{F}_k] = \mathbb{E}[M_n 1_A \mid \mathcal{F}_k] \Rightarrow \mathbb{E}[M_k 1_A] \leq \mathbb{E}[M_n 1_A].$$

□

Theorem 2.4.3 (Doob's Maximal Inequality)**.** *Let $(M_n)_{n\in\mathbb{N}}$ be a non-negative sub-martingale and take $\lambda > 0$. Then*

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

Proof. Let us first define the stopping time

$$\tau = \min\{n \in \mathbb{N} \ ; \ M_n \geq \lambda\}$$

To verify that τ is a stopping time, notice that

$$\{\tau > n\} = \{M_m < \lambda, \ \forall \ m \leq n\} = \bigcap_{m=1}^n \{M_m < \lambda\} \in \mathcal{F}_n \Rightarrow \{\tau \leq n\} = \{\tau > n\}^c \in \mathcal{F}_n.$$

Moreover

$$\lambda 1_{\{\tau\leq n\}} \leq M_\tau 1_{\{\tau\leq n\}} = \sum_{k=0}^n M_k 1_{\{\tau=k\}}$$

Let us now compute the expectation on both sides of the above inequality. Firstly, notice that $\{\tau = k\} \in \mathcal{F}_k$ and thus, for any $k \leq n$,

$$\mathbb{E}[M_k 1_{\{\tau=k\}}] \leq \mathbb{E}[M_n 1_{\{\tau=k\}}].$$

Hence, since

$$\{M_n^* \geq \lambda\} = \left\{\sup_{m \leq n} M_m \geq \lambda\right\} = \{\tau \leq n\},$$

we find

$$\begin{aligned}
\lambda \mathbb{P}(M_n^* \geq \lambda) &= \lambda \mathbb{P}(\tau \leq n) = \mathbb{E}[\lambda 1_{\{\tau \leq n\}}] \\
&= \mathbb{E}\left[\sum_{k=0}^{n} M_k 1_{\{\tau = k\}}\right] \leq \sum_{k=0}^{n} \mathbb{E}[M_k 1_{\{\tau = k\}}] \\
&\leq \sum_{k=0}^{n} \mathbb{E}[M_n 1_{\{\tau = k\}}] = \mathbb{E}\left[M_n \sum_{k=0}^{n} 1_{\{\tau = k\}}\right] \\
&= \mathbb{E}[M_n 1_{\{\tau \leq n\}}] = \mathbb{E}[M_n 1_{\{M_n^* \geq \lambda\}}].
\end{aligned}$$

□

Remark 2.4.4. Let $p \geq 1$. Then

$$\mathbb{P}(M_n^* \geq \lambda) \leq \frac{\mathbb{E}[|M_n|^p]}{\lambda^p}.$$

To prove this simply notice that $(|M_n|^p)_{n \in \mathbb{N}}$ is a non-negative sub-martingale, if M is a martingale in L^p.

Lemma 2.4.5. *Let X and Y be non-negative r.v's such that $Y \in L^p$ for some $p > 1$ and*

$$\lambda \mathbb{P}(X \geq \lambda) \leq \mathbb{E}[Y 1_{\{X \geq \lambda\}}], \ \forall \ \lambda \geq 0. \tag{2.3}$$

Then

$$\|X\|_p \leq \frac{p}{p-1} \|Y\|_p. \tag{2.4}$$

In particular, $X \in L^p$.

Proof. Since we do not know if $X \in L^p$, define $X_n = \min\{X, n\}$. Notice that $X_n \to X$ a.s. and that, by Fatou's Lemma,

$$\|X\|_p \leq \liminf_{n \to +\infty} \|X_n\|_p.$$

Hence, if we prove that X_n satisfies Equation (2.4), the lemma is proved. However, we need to verify that condition (2.3) is true for the pair X_n and Y. Notice that

$$1_{\{X_n \geq \lambda\}} = 1_{\{X \geq \lambda\}} 1_{\{n \geq \lambda\}}.$$

Thus,

$$\begin{aligned}
\mathbb{E}[Y\ 1_{\{X_n \geq \lambda\}}] &= \mathbb{E}[Y\ 1_{\{X \geq \lambda\}}] 1_{\{n \geq \lambda\}} \geq \lambda\ \mathbb{P}(X \geq \lambda) 1_{\{n \geq \lambda\}} \\
&= \lambda\ \mathbb{E}[1_{\{X \geq \lambda\}}] 1_{\{n \geq \lambda\}} = \lambda\ \mathbb{E}[1_{\{X \geq \lambda\}} 1_{\{n \geq \lambda\}}] = \lambda \mathbb{P}(X_n \geq \lambda)
\end{aligned}$$

Therefore, condition (2.3) hold for X_n and Y and we can continue the proof of the lemma. Notice that

$$z^p = p \int_0^z t^{p-1} dt = p \int_0^{+\infty} t^{p-1} 1_{\{t \leq z\}} dt.$$

Then, exchanging z by X_n and taking expectation, we find

$$\mathbb{E}[X_n^p] = p \int_0^{+\infty} t^{p-1} \mathbb{P}(X_n \geq t) dt.$$

To prove the equation above, one just need to apply Tonelli's Theorem since the integrand is positive. Therefore,

$$\begin{aligned}
\mathbb{E}[X_n^p] &\leq p \int_0^{+\infty} t^{p-2} \mathbb{E}[Y 1_{\{X_n \geq t\}}] dt = p \mathbb{E}\left[Y \int_0^{+\infty} t^{p-2} 1_{\{X_n \geq t\}} dt\right] \\
&= p \mathbb{E}\left[Y \int_0^{X_n} t^{p-2} dt\right] = p \mathbb{E}\left[Y \frac{X_n^{p-1}}{p-1}\right].
\end{aligned}$$

By Hölder's inequality, we conclude (notice that $\frac{1}{p}+\frac{1}{q}=1 \Rightarrow q=\frac{p}{p-1}$)

$$\begin{aligned}\|X_n\|_p^p = \mathbb{E}[X_n^p] &\leq \frac{p}{p-1}\mathbb{E}\left[YX_n^{p-1}\right] \leq \frac{p}{p-1}\|Y\|_p\|X_n^{p-1}\|_q \\ &= \frac{p}{p-1}\|Y\|_p\|X_n\|_p^{p-1} \Rightarrow \|X_n\|_p \leq \frac{p}{p-1}\|Y\|_p.\end{aligned}$$

□

Theorem 2.4.6 (Doob's L^p Inequality). *Let $(M_n)_{n\in\mathbb{N}}$ be a non-negative sub-martingale. Then, for any $p>1$ and $n \geq 0$,*

$$\|M_n^*\|_p \leq \frac{p}{p-1}\|M_n\|_p.$$

Proof. Notice that if $M_n \notin L^p$ for some p, then M_n^* will not belong to L^p, since $M_n^* \geq M_n$. In this case, the inequality above holds (the equality, actually). Hence, we may assume that $M_n \in L^p$. The theorem then follows from the lemma above.

□

2.5 Martingale Convergence

Theorem 2.5.1 (Martingale Convergence Theorem in L^2). *Let $(M_n)_{n\in\mathbb{N}}$ be a martingale such that*

$$\mathbb{E}[M_n^2] \leq C, \ \forall \ n \in \mathbb{N}.$$

Then, there exists a r.v. M_∞ with $\mathbb{E}[M_\infty^2] \leq C$ such that

$$M_n \longrightarrow M_\infty, \ \textit{a.s. and in } L^2$$

Proof. Fix $m \in \mathbb{N}$ and define the non-negative sub-martingale $X_k = (M_{k+m}-M_m)^2$. By Doob's Maximal Inequality, for any $\lambda > 0$,

$$\mathbb{P}\left(\sup_{k\leq n} X_k \geq \lambda\right) \leq \frac{\mathbb{E}[X_n]}{\lambda}.$$

Additionally,

$$\mathbb{E}[X_n] = \mathbb{E}[(M_{n+m}-M_m)^2] = \mathbb{E}\left[\left(\sum_{k=m+1}^{n+m} d_k\right)^2\right],$$

where $d_k = M_k - M_{k-1}$. Notice that $\mathbb{E}[d_k d_j]=0$, for $k \neq j$. In fact, with out loss of generality, $j>k$, and

$$\mathbb{E}[d_k d_j] = \mathbb{E}[\mathbb{E}[d_k d_j \mid \mathcal{F}_k]] = \mathbb{E}[d_k \underbrace{\mathbb{E}[d_j \mid \mathcal{F}_k]}_{=0}] = 0.$$

Then,

$$\mathbb{E}[X_n] = \mathbb{E}\left[\left(\sum_{k=m+1}^{n+m} d_k\right)^2\right] = \sum_{k=m+1}^{n+m}\mathbb{E}[d_k^2] \leq \sum_{k=m+1}^{+\infty}\mathbb{E}[d_k^2]$$

Moreover, notice that

$$\begin{aligned} C \geq \mathbb{E}[M_n^2] = \mathbb{E}\left[\left(\sum_{k=1}^{n} d_k\right)^2\right] &= \sum_{k=1}^{n}\mathbb{E}[d_k^2] \Rightarrow \sum_{k=1}^{n}\mathbb{E}[d_k^2] \leq C, \ \forall \ n \in \mathbb{N} \\ &\Rightarrow \sum_{k=1}^{+\infty}\mathbb{E}[d_k^2] \leq C.\end{aligned}$$

Therefore,

$$\mathbb{P}\left(\sup_{k\geq 0}|M_{k+m}(w)-M_m(w)| \geq \varepsilon\right) \leq \lim_{n\to+\infty}\mathbb{P}\left(\sup_{k\leq n}|M_{k+m}-M_m| \geq \varepsilon\right) = \lim_{n\to+\infty}\mathbb{P}\left(\sup_{k\leq n} X_k \geq \varepsilon^2\right)$$

$$\leq \lim_{n\to+\infty} \frac{\mathbb{E}[X_n]}{\varepsilon^2} \leq \frac{1}{\varepsilon^2} \sum_{k=m+1}^{+\infty} \mathbb{E}[d_k^2].$$

Now, for every $i \in \mathbb{N}$, choose $n_i \in \mathbb{N}$ such that

$$\sum_{k=n_i+1}^{+\infty} \mathbb{E}[d_k^2] \leq 8^{-i}.$$

Thus, for every $i \in \mathbb{N}$,

$$\mathbb{P}\left(\sup_{k\geq 0} |M_{k+n_i}(w) - M_{n_i}(w)| \geq 2^{-i}\right) \leq 2^{-i}.$$

Define now

$$A_i = \left\{\sup_{k\geq 0} |M_{k+n_i}(w) - M_{n_i}(w)| \geq 2^{-i}\right\}$$

and notice that

$$\sum_{i=1}^{+\infty} \mathbb{P}(A_i) < +\infty.$$

By Borel-Cantelli lemma,

$$\mathbb{P}\left(\bigcup_{i=1}^{+\infty} \bigcap_{j=i}^{+\infty} A_j^c\right) = 1.$$

This means that, almost surely, there exists $i \in \mathbb{N}$, such that, for all $j \geq i$,

$$\sup_{k\geq 0} |M_{k+n_j} - M_{n_j}| < 2^{-j} \Rightarrow |M_{k+n_j} - M_{n_j}| < 2^{-j},\ \forall\, k \geq 0.$$

Hence, for every $\varepsilon > 0$, we can choose $j \geq i$, such that $2^{-j} < \varepsilon/2$ and $n_0 = n_j$. This implies that, for any $n, m \geq n_0$, we find

$$|M_n - M_m| = |M_{k_n+n_j} - M_{k_m+n_j}| \leq |M_{k_n+n_j} - M_{n_j}| + |M_{k_n+n_j} - M_{n_j}| < \varepsilon.$$

Then, $(M_n)_{n\in\mathbb{N}}$ is a.s. Cauchy, which implies it converges a.s. Let M_∞ be the limit of this sequence and notice:

$$M_\infty = \sum_{k=1}^{+\infty} d_k.$$

Hence,

$$\mathbb{E}\left[(M_\infty - M_n)^2\right] = \sum_{k=n+1}^{+\infty} \mathbb{E}[d_k^2] \overset{n\to+\infty}{\longrightarrow} 0 \Rightarrow M_n \to M_\infty \text{ in } L^2.$$

□

The next theorem shows how localization argument can be used to improve some results.

Theorem 2.5.2 (Another Martingale Convergence Theorem)**.** *Let $(M_n)_{n\in\mathbb{N}}$ be a martingale such that*

$$\mathbb{E}[|M_n|] \leq C \text{ and } |M_{n+1} - M_n| \leq C,\ \forall\, n \in \mathbb{N}.$$

Then, there exists a r.v. M_∞ with $\mathbb{E}[|M_\infty|] \leq C$ such that

$$M_n \longrightarrow M_\infty,\ a.s.$$

Proof. Define the stopping time $\tau = \inf\{n \in \mathbb{N}\ ;\ |M_n| \geq \lambda\}$. We know that $(M_{n\wedge\tau})_{n\in\mathbb{N}}$ is a martingale. Moreover,

$$|M_{n\wedge\tau}| - |M_{n\wedge\tau-1}| \leq |M_{n\wedge\tau} - M_{n\wedge\tau-1}| \leq C \Rightarrow |M_{n\wedge\tau}| \leq C + |M_{n\wedge\tau-1}| \leq C + \lambda,$$

by the definition of τ. Hence,

$$\mathbb{E}[(M_{n\wedge\tau})^2] \leq (C+\lambda)^2,$$

which implies, by the Martingale Convergence Theorem in L^2, that $(M_{n\wedge\tau})_{n\in\mathbb{N}}$ converges almost surely. Notice that, in the set $\{\tau = +\infty\}$, we have $M_{n\wedge\tau} = M_n$, for all $n \in \mathbb{N}$. Therefore, $(M_n)_{n\in\mathbb{N}}$ converges almost surely on $\{\tau = +\infty\}$. Define now

$$S = \{w \in \Omega\ ;\ (M_n(w))_{n\in\mathbb{N}} \text{ does not converge}\} \subset \{\tau < +\infty\} = \left\{\sup_{n\in\mathbb{N}} |M_n| \geq \lambda\right\}.$$

By the Doob's Maximal Inequality, for any $n \in \mathbb{N}$,

$$\mathbb{P}\left(\sup_{n\leq N} |M_n| \geq \lambda\right) \leq \frac{\mathbb{E}[|M_N|]}{\lambda} \leq \frac{C}{\lambda}.$$

Thus

$$\mathbb{P}(S) \leq \mathbb{P}\left(\sup_{n\in\mathbb{N}} |M_n| \geq \lambda\right) = \lim_{N\to+\infty} \mathbb{P}\left(\sup_{n\leq N} |M_n| \geq \lambda\right) \leq \frac{C}{\lambda}.$$

Since λ is arbitrary, we conclude $\mathbb{P}(S) = 0$. Therefore, there exists a r.v. M_∞ such that $M_n \to M_\infty$ a.s. and, by Fatou's Lemma,

$$\mathbb{E}[|M_\infty|] \leq \liminf_{n\to+\infty} \mathbb{E}[|M_n|] \leq C$$

□

This localization argument has its limitations. For example, in the theorem above, the boundedness of the increments is unnecessary as we will show.

Theorem 2.5.3 (Martingale Convergence Theorem in L^1)**.** *Let $(M_n)_{n\in\mathbb{N}}$ be a martingale such that*

$$\mathbb{E}[|M_n|] \leq C,\ \forall\ n \in \mathbb{N}.$$

Then, there exists a r.v. M_∞ with $\mathbb{E}[|M_\infty|] \leq C$ such that

$$M_n \longrightarrow M_\infty \ \text{a.s.}$$

The convergence might not be in L^1. It will be in L^1 if the martingale is uniform integrable, see Section 2.5.1

To prove this theorem we will need some preliminary results.

Definition 2.5.4 (Up-Crossings)**.** Let $(c_i)_{i\in[0,n]}$ be a sequence of real numbers. We define the *number of up-crossings of* $[a, b]$ *by* $(c_i)_{i\in[0,n]}$ as the biggest integer k such that there exist $(i_m)_{m\in[0,k]}$ and $(j_m)_{m\in[0,k]}$ satisfying $c_{i_m} \leq a$ and $c_{j_m} \geq b$, for any $0 \leq m \leq k$.

Proposition 2.5.5 (Up-Crossing Inequality)**.** *Given a sub-martingale $(M_n)_{n\in\mathbb{N}}$, denote the number of up-crossings of the interval $[a, b]$ by the sequence $(M_k)_{k\in[0,n]}$ by $N_n(a, b)$. Then*

$$\mathbb{E}[N_n(a, b)] \leq \frac{\mathbb{E}[(M_n - a)^+]}{b - a},$$

where $(x - a)^+ = \max\{x - a, 0\}$.

Proof. Firstly, notice that $((M_n - a)^+)_{n\in\mathbb{N}}$ is a non-negative sub-martingale (since $(x - a)^+$ is increasing and convex). Moreover, the number of up-crossing of the interval $[a, b]$ by $(M_k)_{k\in[0,n]}$ is equal the number of up-crossing of the interval $[0, b - a]$ by the sequence $((M_k - a)^+)_{n\in[0,n]}$.

Let us now define an "investment" strategy based on the up-crossings of $((M_n - a)^+)_{n\in\mathbb{N}}$: we buy one unit of $(M_n - a)^+$ when it hits zeros and then sell this one unit when gets to or higher than $b - a$. This strategy define a sequence $(A_n)_{n\in\mathbb{N}}$ that is non-anticipative and can assume only the values $\{0, 1\}$. Indeed, when $(M_n - a)^+$ hits zeros, we set $A_{n+1} = 1$ and keep this until the process gets over $b - a$, when we set $A_{n+1} = 0$. This means that

$$\mathbb{E}[(A \cdot (M - a)^+)_n] \leq \mathbb{E}[(M_n - a)^+].$$

Indeed, notice that, since $A_k \in \{0, 1\}$, it follows by Proposition 2.2.10. On the other hand, this strategy makes at least $(b - a)N_n(a, b)$, which implies $(b - a)N_n(a, b) \leq (A \cdot (M - a)^+)_n$. Therefore, the inequality holds.

□

Proof. (of Martingale Convergence Theorem in L^1)
Let $a < b$ be rational numbers and define

$$A_{ab} = \left\{ w \in \Omega \ ; \ \liminf_{n\to+\infty} M_n(w) \leq a < b \leq \limsup_{n\to+\infty} M_n(w) \right\}.$$

Notice that $(M_n)_{n\in\mathbb{N}}$ converges almost surely if and only if $\mathbb{P}(A_{ab}) = 0$ for any rational numbers $a < b$. Moreover, $N_n(a,b)$ is a monotonic sequence of positive r.v, which means it converges a.s. to a r.v. $N_\infty(a,b)$, that could be infinite. One can prove that

$$A_{ab} \subset \{N_\infty(a,b) = +\infty\}$$

In fact, for $w \in \{N_\infty(a,b) < +\infty\}$, one must have one of the following situations $\liminf M_n(w) > b$, $\limsup M_n(w) < a$ or $a < \liminf M_n(w) \leq \limsup M_n(w) < b$. Hence, $w \notin A_{ab}$. Then, if we prove $\mathbb{E}[N_\infty(a,b)] < +\infty$, this will imply that $\mathbb{P}(N_\infty(a,b) = +\infty) = 0$ and thus, by the argument above, $\mathbb{P}(A_{ab}) = 0$. Thus, since $|x-a| \geq (x-a)^+$,

$$\begin{aligned}\mathbb{E}[N_n(a,b)] &\leq \frac{1}{b-a}\mathbb{E}[(M_n-a)^+] \leq \frac{1}{b-a}\mathbb{E}[|M_n-a|] \\ &\leq \frac{1}{b-a}(|a| + \mathbb{E}[|M_n|]) \leq \frac{|a|+C}{b-a}.\end{aligned}$$

By the MCT, we have

$$\mathbb{E}[N_\infty(a,b)] = \lim_{n\to+\infty} \mathbb{E}[N_n(a,b)] \leq \frac{|a|+C}{b-a} < +\infty$$

Therefore, M_n converges a.s. to a r.v. M_∞. To verify that $M_\infty \in L^1$, we apply Fatou's lemma:

$$\mathbb{E}[|M_\infty|] \leq \liminf_{n\to+\infty} \mathbb{E}[|M_n|] \leq C < +\infty.$$

□

Corollary 2.5.6. *Let $(S_n)_{n\in\mathbb{N}}$ be the simple random walk and define $\tau = \inf\{n \ ; \ S_n = A$ or $S_n = -B\}$. Then $\mathbb{P}(\tau < +\infty) = 1$*

Proof. In fact, $M_n = S_{n\wedge\tau}$ is a martingale and it cannot converge for any $w \in \{\tau = +\infty\}$. In fact,

$$|M_{n+1}(w) - M_n(w)| = |S_{n+1}(w) - S_n(w)| = 1.$$

However, $(M_n)_{n\in\mathbb{N}}$ is a bounded martingale. Then, by Martingale Convergence Theorem in L^1, $(M_n)_{n\in\mathbb{N}}$ converges almost surely. Therefore, we must have $\mathbb{P}(\tau < +\infty) = 1$.

□

Example 2.5.7. Let $(X_n)_{n\in\mathbb{N}}$ a sequence of iid r.v's with $\mathbb{P}(X_n = 0) = \mathbb{P}(X_n = 2) = 1/2$. Define $M_n = \prod_{k=1}^n X_k$. It is easy to see that M is a martingale that satisfies the Martingale Convergence Theorem and that $M_\infty = 0$. However, $\mathbb{E}[M_\infty] = 0 \neq 1 = \mathbb{E}[M_n]$. Hence we cannot guarantee the L^1 convergence. For this, one would need to study the class of *uniform integrable martingales.*

2.5.1 Uniform Integrability

We want to answer the following question: *what is the additional, optimal condition that $(X_n)_{n\in\mathbb{N}}$ needs to satisfy in order to have the equivalence?*

$$X_n \xrightarrow{\mathbb{P}} X \Leftrightarrow X_n \xrightarrow{L^1} X \tag{2.5}$$

The extra condition is obviously needed since $X_n = n1_{[0,1/n]}$ converges to 0 in probability (actually a.s.) and it does not converge to 0 in L^1. Example 2.5.7 is another case where we have a.s. convergence but not in L^1.

Firstly, consider the following equivalent characterization of integrability:

Lemma 2.5.8. *$X \in L^1$ if and only if the following holds:*

$$\forall\ \varepsilon > 0,\ \exists\ \delta > 0 \text{ s.t. } \mathbb{E}[|X|1_A] < \varepsilon,\ \forall\ A \text{ satisfying } \mathbb{P}(A) < \delta.$$

Proof. Let's prove the ($\Leftarrow$)-direction. Fix $\varepsilon > 0$ and find $\delta > 0$ as in the property above. Find now x such that $\mathbb{P}(|X| > x) < \delta$. Then, for $y > x$, we have

$$\mathbb{E}[|X|1_{\{|X|\le y\}}] \le \mathbb{E}[|X|1_{\{|X|\le x\}}] + \mathbb{E}[|X|1_{\{|X|>x\}}] \le x + \varepsilon.$$

Hence, the left-hand side of the inequality is bounded for large y. Letting $y \to +\infty$ and using MCT, we find $\mathbb{E}[|X|] < +\infty$. Let us now prove the converse. By the DCT, we have $\mathbb{E}[|X|1_{\{|X|>y\}}] \longrightarrow 0$ as $y \to +\infty$. Then, for any fixed $\varepsilon > 0$, choose $y > 0$ such that $\mathbb{E}[|X|1_{\{|X|>y\}}] < \varepsilon/2$. Hence, for any set A, we find

$$\mathbb{E}[|X|1_A] \le \mathbb{E}[|X|1_{A\cap\{|X|\le y\}}] + \mathbb{E}[|X|1_{\{|X|>y\}}] \le y\mathbb{P}(A) + \frac{\varepsilon}{2}.$$

Hence, choosing $\delta = \frac{\varepsilon}{2y}$ gives us the desired result. ☐

Let us now verify the version of the Dominated Convergence Theorem in the case of convergence in probability.

Theorem 2.5.9. *If $X_n \xrightarrow{\mathbb{P}} X$ and $|X_n| \le Y$ for some $Y \in L^1$. Then $X_n \xrightarrow{L^1} X$.*

Proof. Fix $\varepsilon > 0$ and notice that

$$\lim_{n\to+\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

Moreover, there exists a subsequence $(n_k)_{k\in\mathbb{N}}$ converging to X a.s. Then, since $|X_n| \le Y$, we must have $|X| \le Y$. Hence, $|X_n - X| \le 2Y$. Thus,

$$\begin{aligned}\mathbb{E}[|X_n - X|] &= \mathbb{E}[|X_n - X|1_{\{|X_n-X|>\varepsilon\}}] + \mathbb{E}[|X_n - X|1_{\{|X_n-X|\le\varepsilon\}}] \\ &\le \mathbb{E}[|X_n - X|1_{\{|X_n-X|>\varepsilon\}}] + \varepsilon \le \varepsilon + 2\mathbb{E}[Y1_{\{|X_n-X|>\varepsilon\}}].\end{aligned}$$

By the lemma above, there exists $\delta > 0$ such that $\mathbb{E}[Y1_{\{|X_n-X|>\varepsilon\}}] < \varepsilon$, if $\mathbb{P}(|X_n - X| > \varepsilon) < \delta$. Hence, choosing n large enough such that this condition is satisfied concludes the proof. ☐

Remark 2.5.10. Notice that requiring that

$$\lim_{\varepsilon\to+\infty} \sup_{n\in\mathbb{N}} \mathbb{E}[|X_n|1_{\{|X_n|>\varepsilon\}}] = 0$$

would be enough to prove the previous theorem.

Definition 2.5.11 (Uniform Integrability)**.** A set of r.v's $\{X_\lambda\ ;\ \lambda \in \Lambda\}$ is said to be *uniform integrable (u.i)* if

$$\lim_{a\to+\infty} \sup_{\lambda\in\Lambda} \mathbb{E}[|X_\lambda|1_{\{|X_\lambda|>a\}}] = 0$$

We have seen that integrability means

$$\lim_{a\to+\infty} \mathbb{E}[|X|1_{\{|X|>a\}}] = 0.$$

This is a condition on the tail of X, i.e. the probability of $|X| > a$ cannot be too large when compared to a. Hence, the u.i. property tells us that every element of the set of r.v's $\{X_\lambda\ ;\ \lambda \in \Lambda\}$ is integrable, but additionally that the tail probability of $|X_\lambda|$ converge to 0 fast enough when compared to a and uniformly in λ.

A equivalent characterization of u.i. is given in the next lemma:

Lemma 2.5.12. *$\{X_\lambda\ ;\ \lambda \in \Lambda\}$ is u.i. if and only if*

(i) $\sup_{\lambda\in\Lambda} \mathbb{E}[|X_\lambda|] < +\infty$;

(ii) for all $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\mathbb{P}(A) < \delta \Rightarrow \sup_{\lambda\in\Lambda} \mathbb{E}[|X_\lambda|1_A] < \varepsilon.$$

Proof. ($\Rightarrow$): For any $a > 0$, we have

$$\sup_{\lambda\in\Lambda} \mathbb{E}[|X_\lambda|] \le a + \sup_{\lambda\in\Lambda} \mathbb{E}[|X_\lambda| 1_{\{|X_\lambda|>a\}}.$$

Since $\{X_\lambda \ ; \ \lambda \in \Lambda\}$ is u.i., then (i) is true. Now, for any set A,

$$\mathbb{E}[|X_\lambda| 1_A] \le \mathbb{E}[|X_\lambda| 1_{\{|X_\lambda|>a\}}] + a\mathbb{P}(A).$$

If we fix $\varepsilon > 0$, by the u.i property, we can choose $a > 0$ such that

$$\mathbb{E}[|X_\lambda| 1_{\{|X_\lambda|>a\}}] < \frac{\varepsilon}{2},$$

for all $\lambda \in \Lambda$. Hence, choosing A such that $\mathbb{P}(A) < \varepsilon/2a$ yields the result.

($\Leftarrow$): Fix $\varepsilon > 0$ and choose δ as in (ii) above. Take $a > 0$ such that

$$\sup_{\lambda\in\Lambda} \mathbb{E}[|X_\lambda|] < \delta a.$$

Then

$$\mathbb{E}[|X_\lambda|] \ge \mathbb{E}[|X_\lambda| 1_{\{|X_\lambda|>a\}}] \ge a\mathbb{P}(|X_\lambda| > a) \Rightarrow \sup_{\lambda\in\Lambda} \mathbb{P}(|X_\lambda| > a) \le \frac{1}{a} \sup_{\lambda\in\Lambda} \mathbb{E}[|X_\lambda|] < \delta.$$

Now, by property (ii), we conclude

$$\mathbb{E}[|X_\lambda| 1_{\{|X_\lambda|>a\}}] < \varepsilon,$$

for any $\lambda \in \Lambda$, which means that $\{X_\lambda \ ; \ \lambda \in \Lambda\}$ is u.i. □

We are then ready to prove the main theorem regarding uniform integrability:

Theorem 2.5.13. *Let $X_n \xrightarrow{\mathbb{P}} X$. Then, the following statements are equivalents:*

(i) $(X_n)_{n\in\mathbb{N}}$ *is u.i.;*

(ii) $X_n, X \in L^1$ *and* $X_n \xrightarrow{L^1} X$*;*

(iii) $X_n \in L^1$ *and* $\mathbb{E}[|X_n|] \to \mathbb{E}[|X|]$.

Proof. The proof will follow the strategy: $(i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (i)$. The implication $(ii) \Rightarrow (iii)$ is trivial. We will focus on the other two.

$(i) \Rightarrow (ii)$: Clearly, $X_n \in L^1$. Moreover, there exists a sub-sequence $(n_k)_{k\in\mathbb{N}}$ of $(X_n)_{n\in\mathbb{N}}$ converging to X a.s. This implies, by Fatou's lemma, that $X \in L^1$. Indeed,

$$\mathbb{E}[|X|] = \mathbb{E}\left[\lim_{k\to+\infty} |X_{n_k}|\right] \le \liminf_{k\to+\infty} \mathbb{E}[|X_{n_k}|] \le \sup_{n\in\mathbb{N}} \mathbb{E}[|X_n|] < +\infty.$$

Fix $\varepsilon > 0$ and notice that

$$\mathbb{E}[|X_n - X|] \le \varepsilon + \mathbb{E}[|X_n| 1_{\{|X_n - X|>\varepsilon\}}] + \mathbb{E}[|X| 1_{\{|X_n - X|>\varepsilon\}}].$$

By the convergence in probability, $\mathbb{P}(|X_n - X| > \varepsilon) \longrightarrow 0$ as $n \to +\infty$. By uniform integrability, the second term goes to zero. Moreover, since $X \in L^1$, the third term goes to 0 as well. Hence, for any $\varepsilon > 0$,

$$\limsup_{n\to+\infty} \mathbb{E}[|X_n - X|] \le \varepsilon.$$

Letting $\varepsilon \to 0^+$, we conclude $X_n \xrightarrow{L^1} X$.

$(iii) \Rightarrow (i)$: Define $u_a(x) = |x| 1_{[-a,a]}(x)$ and notice

$$\mathbb{E}[|X_n| 1_{\{|X_n|>a\}}] = \mathbb{E}[|X_n|] - \mathbb{E}[u_a(X_n)].$$

Notice that u_a is bounded and continuous on $(-a,a)$. Moreover, $X_n \xrightarrow{\mathcal{L}} X$ (since the convergence in probability holds). If we denote Δ the set of discontinuities of the cdf of X (which would be at most countable), we have

$$\mathbb{E}[u_a(X_n)] \overset{n\to+\infty}{\longrightarrow} \mathbb{E}[u_a(X)],$$

for $-a, a \notin \Delta$. Hence, in this case (and noticing the assumptions in (iii)), we have

$$\mathbb{E}[|X_n|1_{\{|X_n|>a\}}] = \mathbb{E}[|X_n|] - \mathbb{E}[u_a(X_n)] \overset{n\to+\infty}{\longrightarrow} \mathbb{E}[|X|] - \mathbb{E}[u_a(X)] = \mathbb{E}[|X|1_{\{|X|>a\}}].$$

Since $X \in L^1$ and Δ is at most countable, for any $\varepsilon > 0$, there exists $b > 0$, $b \notin \Delta$, such that

$$\mathbb{E}[|X|1_{\{|X|>b\}}] < \varepsilon.$$

Hence, there exists $n_0 \in \mathbb{N}$ such that, for any $n \geq n_0$,

$$\mathbb{E}[|X_n|1_{\{|X_n|>a\}}] < 2\varepsilon.$$

Moreover, since any finite set of integrable r.v's is u.i., there exists $c > 0$ such that, for any $n < n_0$,

$$\mathbb{E}[|X_n|1_{\{|X_n|>c\}}] < 2\varepsilon.$$

Take $a > \max\{b, c\}$ and notice

$$\mathbb{E}[|X_n|1_{\{|X_n|>a\}}] < 2\varepsilon,$$

for any $n \in \mathbb{N}$. Therefore, $(X_n)_{n\in\mathbb{N}}$ is u.i. □

Example 2.5.14. We will now list examples of classes of uniform integrable sequences:

(i) $|X_\lambda| \leq Y$, with $Y \in L^1$. Indeed, it follows from the fact $|X_\lambda|1_{\{|X_\lambda|>a\}} \leq |Y|1_{\{|Y|>a\}}$.

(ii) there exists $\gamma > 0$ and $K > 0$ such that $\mathbb{E}[|X_\lambda|^{1+\gamma}] \leq K$, for all λ. In fact, we must have

$$\mathbb{E}[|X_\lambda|1_{\{|X_\lambda|>a\}}] \leq \mathbb{E}\left[|X_\lambda|\frac{|X_\lambda|^\gamma}{a^\gamma}1_{\{|X_\lambda|>a\}}\right] \leq \frac{\mathbb{E}[|X_\lambda|^{1+\gamma}]}{a^\gamma} \leq \frac{K}{a^\gamma} \overset{a\to+\infty}{\longrightarrow} 0$$

(iii) if φ is such that $\lim\limits_{x\to+\infty}\dfrac{\varphi(x)}{x} = +\infty$ and $\mathbb{E}[\varphi(|X_\lambda|)] \leq K$, for all λ. Notice that for any $M > 0$, there exists $b > 0$ such that $\varphi(x)/x > M$, for $x > b$. Similarly to the argument above, we find, for $a \geq b$,

$$|X_\lambda|1_{\{|X_\lambda|>a\}} \leq \frac{\varphi(|X_\lambda|)}{M}1_{\{|X_\lambda|>a\}}.$$

Proposition 2.5.15. *Let $X \in L^1$. Then, the set*

$$\{\mathbb{E}[X \mid \mathcal{G}] \ ; \ \mathcal{G} \textit{ sub-}\sigma\textit{algebra of } \mathcal{F}\}$$

is u.i.

To prove this proposition, we will use the following lemma that says that L^1 is somehow similar to L^p, for $p > 1$.

Lemma 2.5.16. *If $X \in L^1$, then there exists a convex, increasing function φ such that $\lim\limits_{x\to+\infty}\dfrac{\varphi(x)}{x} = +\infty$ and*

$$\mathbb{E}[\varphi(|X|)] < +\infty.$$

Proof of Lemma. Notice that

$$\mathbb{E}[|X|] = \int_0^{+\infty} \mathbb{P}(|X| \geq x)dx < +\infty.$$

One can prove there exists an increasing, positive function f with $f(x) \to +\infty$ as $x \to +\infty$ and[1]

$$\int_0^{+\infty} f(x)\mathbb{P}(|X| \geq x)dx < +\infty.$$

[1] $f(x) = \left(\int_x^{+\infty} \mathbb{P}(|X| \geq y)dy\right)^{-\alpha}$ with $0 < \alpha < 1$.

Define then

$$\varphi(x) = \int_0^x f(y)dy,$$

which is clearly convex, increasing and satisfies $\lim_{x\to+\infty} \frac{\varphi(x)}{x} = +\infty$ (by L'Hôpital). Moreover

$$\mathbb{E}[\varphi(|X|)] = \mathbb{E}\left[\int_0^{|X|} f(x)dx\right] = \mathbb{E}\left[\int_0^{+\infty} f(x)1_{\{|X|\geq x\}}dx\right] = \int_0^{+\infty} f(x)\mathbb{P}(|X| \geq x)dx < +\infty.$$

□

Proof of Proposition. By the lemma above, there exists φ increasing, convex and satisfying $\lim_{x\to+\infty} \frac{\varphi(x)}{x} = +\infty$ such that

$$\mathbb{E}[\varphi(|X|)] < +\infty.$$

Moreover, since φ is increasing,

$$\varphi(|\mathbb{E}[X \mid \mathcal{G}]|) \leq \varphi(\mathbb{E}[|X| \mid \mathcal{G}]).$$

Hence, by Jensen's inequality,

$$\mathbb{E}[|\varphi(\mathbb{E}[X \mid \mathcal{G}])|] \leq \mathbb{E}[\varphi(\mathbb{E}[|X| \mid \mathcal{G}])] \leq \mathbb{E}[\mathbb{E}[\varphi(|X|) \mid \mathcal{G}]] = \mathbb{E}[\varphi(|X|)] < +\infty.$$

Therefore, by the third item in Example 2.5.14, we conclude the proof. □

Part II

Continuous Time

Chapter 3

Continuous-Time Markov Chain

3.1 Review of Exponential Distribution

We will need the following results about the Exponential distribution, $\tau \sim \text{Exp}(\lambda)$:

- Density: $f(t) = \lambda e^{-\lambda t} 1_{[0,+\inf)}(t)$;
- $\mathbb{E}[\tau] = \frac{1}{\lambda}$, $\text{Var}(\tau) = \frac{1}{\lambda^2}$, $\mathbb{P}(\tau > t) = e^{-\lambda t}$;
- τ is exponential if and only is memoryless: $\mathbb{P}(\tau > t + s \mid \tau > s) = \mathbb{P}(\tau > t)$;
- if τ_i is iid $\text{Exp}(\lambda)$, then $\tau_1 + \cdots + \tau_n \sim \text{Gamma}(n, \lambda)$;
- $\mathbb{P}(\tau_1 < \tau_2) = \mathbb{P}(\tau_1 = \min\{\tau_1, \tau_2\}) = \dfrac{\lambda_1}{\lambda_1 + \lambda_2}$;
- if $\tau_i \sim \text{Exp}(\lambda_i)$ are independent, $\min\{\tau_1, \dots, \tau_n\} \sim Exp(\lambda_1 + \cdots + \lambda_n)$;
- if $\tau_i \sim \text{Exp}(\lambda_i)$ are independent, $\mathbb{P}(\tau_k = \min\{\tau_1, \dots, \tau_n\}) = \dfrac{\lambda_k}{\lambda_1 + \cdots + \lambda_n}$;
- if $\tau_i \overset{iid}{\sim} \text{Exp}(\lambda)$ and $N \sim \text{Geom}(p)$ are independent, then $\sum_{i=1}^{N} \tau_i \sim \text{Exp}(\lambda p)$.

3.2 Introduction

Instead of considering that a process changes at pre-determined times, that we have indexed by $\mathbb{N}$, it may be interesting, by a modeling standing point, to consider processes that can change at any instant. We start considering an increasing sequence of continuous and positive random variables, $(\tau_n)_{n\in\mathbb{N}}$, i.e. $\tau_n \geq \tau_{n-1} \geq \tau_0 = 0$, and we assume

$$\lim_{n\to+\infty} \tau_n = +\infty.$$

We will not considering the finite case in these notes. We define then the process X_t as

$$X_t = \begin{cases} i_0, \; 0 \leq t < \tau_1, \\ i_1, \; \tau_1 \leq t < \tau_2, \\ \vdots \end{cases} = \sum_{n=0}^{+\infty} i_n 1_{[\tau_n, \tau_{n+1})}(t),$$

where $i_n \in I$ is random. We must assume $i_n \neq i_{n-1}$. We first separate the states I in two classes: absorbing and non-absorbing. We say $i \in I$ is absorbing if

$$\mathbb{P}(X_s = i \mid X_t = i) = 1,$$

for all $s \geq t$. For $i \in I$ that is non-absorbing, fix a cdf in $[0, +\infty)$, F_i, and a distribution over I, $(a_{ij})_{j\in I}$. We then set

$$\mathbb{P}(\tau_1 \leq t, X_{\tau_1} = j \mid X_0 = i) = F_i(t) a_{ij}.$$

Define

$$P_{ij}(t) = \mathbb{P}(X_t = j \mid X_0 = i).$$

Notice that

$$P_{ij}(0) = 1_{\{i=j\}} \text{ and } \mathbb{P}(X_t = j) = \sum_{i \in I} \mathbb{P}(X_0 = i) P_{ij}(t).$$

We cannot use, in general, $P_{ij}(t)$ to compute the probability of paths of X. In what follows, we will assume that $(X_t)_{t \geq 0}$ is Markovian:

$$\mathbb{P}(X_t = j \mid X_s = i, X_{s_n} = i_n, \ldots, X_{s_1} = i_1) = P_{ij}(t-s) \tag{3.1}$$

for any $i, j, i_1, \ldots, i_n \in I$ and $s_1 \leq s_2 \leq \cdots s_n \leq s \leq t$. In this case, we then have

$$\mathbb{P}(X_{t_1} = i_1, \ldots, X_{t_n} = i_n) = \mathbb{P}(X_{t_1} = i_1) P_{i_1, i_2}(t_2 - t_1) \cdots P_{i_{n-1}, i_n}(t_n - t_{n-1}).$$

It also follows from the Markov property, the Chapman-Kolmogorov equation:

$$P_{ij}(t+s) = \sum_{k \in I} \mathbb{P}(X_t = k, X_{t+s} = j \mid X_0 = i) = \sum_{k \in I} P_{ik}(t) P_{kj}(s).$$

It is outside the scope of these notes, but one can show that (3.1) is equivalent to:

$$\mathbb{P}(\tau_1 > t + s \mid \tau_1 > s, X_0 = i) = \mathbb{P}(\tau_1 > t \mid X_0 = i).$$

This is the memoryless property and therefore, since τ_1 is assume to be continuous and positive, we must have

$$\tau_1 \mid X_0 = i \sim Exp(q_i)$$

for some rate $q_i > 0$.

Remark 3.2.1. A non-rigorous argument that the memoryless property above should hold is (where we suppress the condition on $X_0 = i$):

$$\begin{aligned}
\mathbb{P}(\tau_1 > s + t \mid \tau_1 > s) &= \mathbb{P}(X_u = i,\ u \in [0, s+t] \mid X_u = i,\ u \in [0, s]) \\
&= \mathbb{P}(X_u = i,\ u \in [s, s+t] \mid X_u = i,\ u \in [0, s]) \\
&= \mathbb{P}(X_u = i,\ u \in [s, s+t] \mid X_s = i) \\
&= \mathbb{P}(X_u = i,\ u \in [0, t] \mid X_0 = i) = \mathbb{P}(\tau_1 > t).
\end{aligned}$$

Proposition 3.2.2.

$$P_{ij}(t) = 1_{\{i=j\}} e^{-q_i t} + \int_0^t q_i e^{-q_i s} \left(\sum_{k \neq i} a_{ik} P_{kj}(t-s) \right) ds$$

Proof. The proof is trivial if i is absorbing. So, let's assume i in non-absorbing. Then

$$P_{ij}(t) = \mathbb{P}(\tau_1 > t, X_t = j \mid X_0 = i) + \mathbb{P}(\tau_1 \leq t, X_t = j \mid X_0 = i).$$

Note that

$$\mathbb{P}(\tau_1 \leq t, X_{\tau_1} = k, X_t = j \mid X_0 = i) = \int_0^t q_i e^{-q_i s} a_{ik} P_{kj}(t-s) ds.$$

Moreover

$$\mathbb{P}(\tau_1 > t, X_t = j \mid X_0 = i) = 1_{\{i=j\}} \mathbb{P}(\tau_1 > t \mid X_0 = i) = 1_{\{i=j\}} e^{-q_i t}.$$

□

We may take the derivative of the formula above with respect to t and find that

$$P'_{ij}(t) = -q_i P_{ij}(t) + q_i \sum_{k \neq i} a_{ik} P_{kj}(t)$$

If you don't know how to verify this computation, search for Leibniz integral rule. Setting $t = 0$, we find

$$P'_{ij}(0) = -q_i 1_{\{i=j\}} + q_i a_{ij} 1_{\{i \neq j\}}.$$

Therefore, we define the *infinitesimal generator matrix* $\mathbf{Q}$ as

$$q_{ij} = P'_{ij}(0) = \begin{cases} -q_i, & i = j, \\ q_i a_{ij}, & i \neq j. \end{cases}$$

Notice that $\sum_{j\in I} q_{ij} = -q_i + q_i \sum_{j\neq i} a_{ij} = 0$. Moreover,

$$P'_{ij}(t) = \sum_{k\in I} q_{ik} P_{kj}(t).$$

This is called **Backward Kolmogorov equation**. In matrix notation, it may be written as $\mathbf{P}'(t) = \mathbf{QP}(t)$. One could also differentiate with respect to s the Chapman-Kolmogorov (under some smoothness condition) and conclude that

$$P'_{ij}(t+s) = \sum_{k\in I} P_{ik}(t) P'_{kj}(s).$$

Setting $s = 0$, we find

$$P'_{ij}(t) = \sum_{k\in I} P_{ik}(t) P'_{kj}(0) = \sum_{k\in I} P_{ik}(t) q_{kj}.$$

This is called **Forward Kolmogorov equation**. In matrix notation: $\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{Q}$.

Stationary Distribution

We define the stationary distribution similarly:

$$\pi = \pi\mathbf{P}(t) \text{ or } \pi_j = \sum_{i\in I} \pi_i P_{ij}(t),$$

for all $i \in I$ and $t \geq 0$. The most important fact about π is that $\pi\mathbf{Q} = 0$. Let us prove this in the case where we can interchange the order of sum and differentiation (for instance, we might have assumed that the state space was finite). Notice that, by the Forward Equation and if π is a stationary distribution

$$\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{Q} \Rightarrow \pi\mathbf{P}(t)' = \pi\mathbf{P}(t)\mathbf{Q} = \pi\mathbf{Q} \Rightarrow (\pi\mathbf{P}(t))' = \pi\mathbf{Q}$$

Then, if $\pi = \pi\mathbf{P}(t)$, we find $\pi\mathbf{Q} = 0$. On the other hand, by the Backward Equation and if $\pi\mathbf{Q} = 0$,

$$\mathbf{P}'(t) = \mathbf{QP}(t) \Rightarrow \pi\mathbf{P}'(t) = \pi\mathbf{QP}(t) = 0 \Rightarrow (\pi\mathbf{P}(t))' = 0$$

Hence, it follows that $\pi\mathbf{P}(t)$ is constant and then, $\pi\mathbf{P}(t) = \pi\mathbf{P}_0 = \pi$.

3.2.1 The Embedded Process

Firstly, define the jump times of a continuous-time Markov Chain as

$$\tau_n = \inf\{t \geq \tau_{n-1} \ ; \ X_t \neq X_{\tau_{n-1}}\},$$

with $\tau_0 = 0$. We have seen that

$$\mathbb{P}(\tau_{n+1} - \tau_n > t \mid X_{\tau_n} = i) = e^{-tq_i}.$$

The embedded process is then defined as $Y_n = X_{\tau_n}$. One can prove that $(Y_n)_{n\in\mathbb{N}}$ is a discrete-time Markov chain with transition probability equal

$$p_{ij} = \frac{q_{ij}}{q_i}.$$

Moreover, the definitions of transient, recurrent (null and positive) are the same as in the discrete case, and under some mild assumptions, the embedded process has the same characteristics as the continuous-time process. Additionally, if $(\lambda_i)_{i\in I}$ is a stationary distribution of Y, then

$$\sum_{i\in I} \lambda_i p_{ij} = \lambda_j.$$

By the definition of p_{ij}, we have

$$\sum_{i\neq j} \lambda_i \frac{q_{ij}}{q_i} = \lambda_j \Leftrightarrow \sum_{i\neq j} \frac{\lambda_i}{q_i} q_{ij} = \lambda_j \Leftrightarrow \sum_{i\neq j} \frac{\lambda_i}{q_i} q_{ij} = -\frac{\lambda_j}{q_j} q_{jj} \Leftrightarrow \sum_{i\in I} \frac{\lambda_i}{q_i} q_{ij} = 0,$$

which implies that $\pi = \lambda/q$ is the stationary distribution for X. Notice that if $q_i = q_j$, then $\pi = \lambda$.

3.2.2 Reversibility

We say that a continuous-time Markov chain $(X_t)_{t\geq 0}$ with generator $\mathbf{Q}$ is reversible if there exists a probability measure $(\pi_i)_{i\in I}$ such that

$$\pi_i q_{ij} = \pi_j q_{ji}.$$

Moreover, we say that π is reversible for X. That means that, initializing the process $(X_t)_{t\geq 0}$ with π, we have the same probability of observing a jump from i to j and a jump from j to i.
If π is reversible for X, then π is invariant for X. Indeed

$$\sum_{i\in I} \pi_i q_{ij} = \sum_{i\in I} \pi_j q_{ji} = \pi_j \sum_{i\in I} q_{ji} = 0.$$

If the generator of X is symmetric, it is easy to see that

$$\pi_i = \frac{\sum_{j\neq i} \pi_j}{\sum_{k\in I}\sum_{j\neq k} \pi_j} = \frac{q_i}{\sum_{k\in I} g_k}$$

is reversible for X.

Proposition 3.2.3. *π is reversible for X if and only if, for all $t \geq 0$ and $i, j \in I$,*

$$\pi_i p_{ij}(t) = \pi_j p_{ji}(t).$$

Proof. ($\Leftarrow$): Taking derivative of the equation above, we find

$$\pi_i p'_{ij}(t) = \pi_i \sum_{k\in I} p_{ik}(t) q_{kj} = \sum_{k\in I} \pi_i p_{ik}(t) q_{kj} = \sum_{k\in I} \pi_k p_{ki}(t) q_{kj}$$

$$\pi_j p'_{ji}(t) = \pi_j \sum_{k\in I} q_{jk} p_{ki}(t) = \sum_{k\in I} \pi_j p_{ki}(t) q_{jk}$$

Then, for all $t \geq 0$,

$$\sum_{k\in I} p_{ki}(t)(\pi_k q_{kj} - \pi_j q_{jk}) = 0,$$

which implies π is reversible for X.

($\Rightarrow$): Assume $\sup_{i\in I} q_i = \gamma < +\infty$. Define then the following transition probability matrix Q:

$$Q_{ij} = \frac{q_{ij}}{\gamma} \text{ and } Q_{ii} = 1 - \frac{q_i}{\gamma}.$$

If Y is a discrete-time Markov chain with transition probability Q and N is a Poisson process with parameter γ, both independent, we can write

$$X_t = Y_{N_t},$$

following similar construction of X we have done before. Notice that, for $i \neq j$,

$$\pi_i q_{ij} = \pi_j q_{ji} \Leftrightarrow \pi_i Q_{ij} = \pi_j Q_{ji}$$

and that, for any $i_1, \dots, i_n$,

$$\pi_i Q_{i,i_1} \cdots Q_{i_n,j} = \pi_j Q_{j,i_n} \cdots Q_{i_1,i}.$$

Summing over all $i_1, \dots, i_n$, we find

$$\pi_i Q^n_{ij} = \pi_j Q^n_{ji}.$$

Moreover, by our characterization of X, we have

$$p_{ij}(t) = \sum_{n\geq 0} \mathbb{P}(N_t = n) Q^n_{ij},$$

which implies

$$\pi_i p_{ij}(t) = \sum_{n\geq 0} \mathbb{P}(N_t = n)\pi_i Q^n_{ij} = \sum_{n\geq 0} \mathbb{P}(N_t = n)\pi_j Q^n_{ji} = \pi_j p_{ji}(t).$$

□

Proposition 3.2.4. *Consider* $0 = t_1 < \ldots < t_n < t$ *and* $i_1, \ldots, i_n$ *in* I. *Define* $s_i = t - t_{n-i}$. *Then*

$$\mathbb{P}_\pi(X_{t_1} = i_1, \ldots, X_{t_n} = i_n) = \mathbb{P}_\pi(X_{s_1} = i_n, \ldots, X_{s_n} = i_1)$$

Proof. By the previous proposition, we have

$$\pi_i p_{ij}(t) = \pi_j p_{ji}(t).$$

Moreover,

$$\begin{aligned}\mathbb{P}_\pi(X_{t_1} = i_1, \ldots, X_{t_n} = i_n) &= \sum_{i\in I} \pi_i \sum_{j\in I} p_{i,i_1}(t_1) \cdots p_{i_{n-1},i_n}(t_n - t_{n-1}) p_{i_n,j}(t - t_n) \\ &= \sum_{i\in I}\sum_{j\in I} \pi_i p_{i_1,i}(s_n - s_{n-1}) \cdots p_{i_n,i_{n-1}}(s_1 - s_0) p_{j,i_n}(s_0) \\ &= \mathbb{P}_\pi(X_{s_1} = i_n, \ldots, X_{s_n} = i_1).\end{aligned}$$

□

3.3 Examples

3.3.1 Two-States Markov Chain

Let $(X_t)_{t\geq 0}$ be a continuous-time Markov chain with state space $I = \{0, 1\}$ and generator

$$\mathbf{Q} = \begin{bmatrix} -\mu & \mu \\ \lambda & -\lambda \end{bmatrix}$$

where $\lambda, \mu > 0$. The interpretation of this chain goes as follows: when the chain arrives the state 0, an exponential random variable with parameter μ is sampled and its value will determine the holding time until proceeding to state 1. An analogous situation happens when the chain arrives at state 1 but with parameter λ.

The forward equations are:

$$p'_{ij}(t) = \sum_{k=1}^{2} p_{ik}(t) q_{kj},$$

which can be read as

$$\begin{cases} p'_{00}(t) = \lambda p_{01}(t) - \mu p_{00}(t) \\ p'_{01}(t) = -\lambda p_{01}(t) + \mu p_{00}(t) \\ p'_{10}(t) = -\mu p_{10}(t) - \lambda p_{11}(t) \\ p'_{11}(t) = \mu p_{10}(t) - \lambda p_{11}(t). \end{cases}$$

Note that last two equations are the same as the first two but with μ and λ interchanged, so it's enough to solve the first two. Note also that

$$p_{00}(t) = 1 - p_{01}(t).$$

Using this formula, the first ODE turns to

$$p'_{00}(t) = \lambda p_{01}(t) - \mu p_{00}(t) = \lambda(-p_{00}(t) + 1) - \mu p_{00}(t) = -(\lambda + \mu) p_{00}(t) + \lambda.$$

Trying a solution of the form $p_{00}(t) = C_1 e^{-(\mu+\lambda)t} + C_2$, we have

$$p'_{00}(t) = -(\mu + \lambda) C_1 e^{-(\mu+\lambda)t} = -(\mu + \lambda) p_{00}(t) + C_2(\lambda + \mu).$$

So, $C_2 = \lambda/(\lambda + \mu)$. Furthermore, $p_{00}(0) = 1$, which means

$$C_1 + \frac{\lambda}{\lambda + \mu} = 1 \Rightarrow C_1 = \frac{\mu}{\lambda + \mu}.$$

Finally,

$$p_{00}(t) = \frac{\mu}{\lambda + \mu} e^{-(\mu+\lambda)t} + \frac{\lambda}{\lambda + \mu}.$$

Moreover,

$$p_{01}(t) = -\frac{\mu}{\lambda+\mu}e^{-(\mu+\lambda)t} + \frac{\mu}{\lambda+\mu}.$$

Interchanging μ and λ, we get

$$p_{10}(t) = -\frac{\lambda}{\lambda+\mu}e^{-(\mu+\lambda)t} + \frac{\lambda}{\lambda+\mu} \text{ and } p_{11}(t) = \frac{\lambda}{\lambda+\mu}e^{-(\mu+\lambda)t} + \frac{\mu}{\lambda+\mu}.$$

Additionally, to find the stationary distribution, we have to solve $\pi G = 0$, where $\pi = [\pi_0, \pi_1]$. So,

$$[\pi_0, \pi_1]\begin{pmatrix} -\mu & \mu \\ \lambda & -\lambda \end{pmatrix} = [-\mu\pi_0 + \lambda\pi_1, \quad \mu\pi_0 - \lambda\pi_1] = [0,\ 0].$$

Therefore, using the fact that $\pi_0 + \pi_1 = 1$ and $\pi_1 = \frac{\mu}{\lambda}\pi_0$, we have

$$\pi_0 = \frac{\lambda}{\mu+\lambda} \text{ and } \pi_1 = \frac{\mu}{\mu+\lambda}.$$

So, since $\lambda, \mu > 0$, we have

$$\lim_{t\to+\infty} e^{-(\mu+\lambda)t} = 0,$$

which implies that $p_{00}(t) \to \pi_0$, $p_{01}(t) \to \pi_1$, $p_{10}(t) \to \pi_0$ and $p_{11}(t) \to \pi_1$, when $t \to +\infty$.

3.3.2 Birth and Death Chains

This process is the continuous-time generalization of the example we studied in discrete time. The state space is $I = \mathbb{N}$ and the infinitesimal generator is given by

$$q_{ij} = \begin{cases} \lambda_i, \text{ if } j = i+1, \\ \mu_i, \text{ if } j = i-1, \\ -(\mu_i + \lambda_i), \text{ if } j = i, \\ 0, \text{ otherwise.} \end{cases}$$

λ_i is called the birth rate and μ_i, the death rate. Notice that $q_i = \lambda_i + \mu_i$ and

$$a_{i,i+1} = \frac{\lambda_i}{\mu_i + \lambda_i} \text{ and } a_{i,i-1} = \frac{\mu_i}{\mu_i + \lambda_i}.$$

We say the process is pure birth if $\mu \equiv 0$. If $\lambda \equiv 0$, we call it pure death. Certain assumptions on λ must be considered to guarantee that the chain does not explode. For instance, it is enough to assume $\lambda_i \leq A + Bi$.

Notice that the Backward and Forward Kolmogorov equations become

$$\begin{cases} P'_{ij}(t) = \mu_i P_{i-1,j}(t) - (\lambda_i + \mu_i)P_{ij}(t) + \lambda_i P_{i+1,j}(t), \\ \\ P'_{ij}(t) = \lambda_{j-1}P_{i,j-1}(t) - (\lambda_j + \mu_j)P_{ij}(t) + \mu_{j+1}P_{i,j+1}(t), \end{cases}$$

Using the integral-form of the solution of a first-order linear ODE (as the one above), we find that

$$P_{ij}(t) = P_{ij}(0)e^{-(\lambda_i+\mu_i)t} + \int_0^t e^{-(\lambda_i+\mu_i)(t-s)}(\mu_i P_{i-1,j}(s) + \lambda_i P_{i+1,j}(s))ds.$$

Similar things could be said if we considered the Forward Kolmogorov equation.

The examples we will consider below are all examples of Birth-Death chains.

3.3.3 Branching

Let X_t denote the number of particles at time t and assume $X_0 = i$. Let ξ_k be the waiting time of particle k among the i particles at the beginning. Assume these random variables are iid Exp(λ). After these waiting time, each particle splits in 2 with probability p and dies with probability $1-p$. Everything happens independently. Hence,

$$\tau_1 = \min\{\xi_1, \ldots, \xi_i\} \sim Exp(i\lambda), \quad a_{i,i+1} = p \text{ and } a_{i,i-1} = 1-p.$$

This implies that $\lambda_i = q_i a_{i,i+1} = i\lambda p$ and $\mu_i = q_i a_{i,i-1} = i\lambda(1-p)$.

3.3.4 Poisson Process

The Poisson process is a pure birth process with $\lambda_i = \lambda$. This implies that

$$\begin{aligned} &P_{ij}(t) = 0, \text{ if } j < i, \\ &P_{ii}(t) = \mathbb{P}(\tau_1 > t \mid X_0 = i) = e^{-\lambda t}, \\ &P'_{ij}(t) = \lambda P_{i,j-1}(t) - \lambda P_{ij}(t), \text{ if } j > i. \end{aligned}$$

This implies that

$$P_{ij}(t) = P_{ij}(0)e^{-\lambda t} + \lambda \int_0^t e^{-\lambda(t-s)} P_{i,j-1}(s) ds = \lambda \int_0^t e^{-\lambda(t-s)} P_{i,j-1}(s) ds.$$

Hence, by induction

$$\begin{aligned} P_{i,i+1}(t) &= \lambda \int_0^t e^{-\lambda(t-s)} e^{-\lambda s} ds = \lambda t e^{-\lambda t}, \\ P_{i,i+2}(t) &= \lambda \int_0^t e^{-\lambda(t-s)} \lambda s e^{-\lambda s} ds = \frac{(\lambda t)^2}{2} e^{-\lambda t}, \\ P_{i,i+k}(t) &= \frac{(\lambda t)^k}{k!} e^{-\lambda t}. \end{aligned}$$

Notice that $P_{ij}(t) = P_{0,j-i}(t)$, which gives us $X_t - i \mid X_0 = i \sim Poisson(\lambda t)$. In general, $X_t - X_s \sim Poisson(\lambda(t-s))$. Indeed,

$$\begin{aligned} \mathbb{P}(X_t - X_s = j) &= \sum_{i=0}^{+\infty} \mathbb{P}(X_s = i, X_t = j + i) = \sum_{i=0}^{+\infty} \mathbb{P}(X_s = i) P_{i,j+i}(t-s) \\ &= \sum_{i=0}^{+\infty} \mathbb{P}(X_s = i) P_{0,j}(t-s) = P_{0,j}(t-s). \end{aligned}$$

Moreover, similar computations show that the increments $X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent.

Hence, if $(X_t)_{t \geq 0}$ is a Poisson process with rate λ, we concluded that

(i) $X_0 = 0$;

(ii) increments are independent;

(iii) $X_t - X_s \sim Poisson(\lambda(t-s))$.

Furthermore, if we assume that a process X satisfies (i), (ii) and (iii), one might show that the waiting time of this process are independent $\text{Exp}(\lambda)$.

3.3.5 Branching with Immigration

Additionally to the branching mechanism, assume that particles could immigrate accordingly to an independent Poisson process with rate ν. In this case, we find

$$\tau_1 = \min\{\xi_1, \ldots, \xi_i, \eta\} \sim Exp(i\lambda + \nu),$$

where $\eta \sim Exp(\nu)$ is the first arrival of the Poisson process. Moreover,

$$\mathbb{P}(\tau_1 = \eta) = \frac{\nu}{i\lambda + \nu},$$

which implies

$$a_{i,i+1} = \frac{\nu}{i\lambda + \nu} + \frac{i\lambda}{i\lambda + \nu} p \text{ and } a_{i,i-1} = \frac{i\lambda}{i\lambda + \nu}(1-p).$$

Therefore, $\lambda_i = \nu + i\lambda p$ and $\mu_i = i\lambda(1-p)$.

3.3.6 Pure Birth

In this case, we have

$$
\begin{aligned}
&P_{ij}(t) = 0, \text{ if } j < i,\\
&P_{ii}(t) = e^{-\lambda_i t},\\
&P'_{ij}(t) = \lambda_{j-1} P_{i,j-1}(t) - \lambda_j P_{ij}(t), \text{ if } j > i.
\end{aligned}
$$

Hence,

$$P_{ij}(t) = \lambda_{j-1} \int_0^t e^{-\lambda_j (t-s)} P_{i,j-1}(s) ds.$$

When $j = i+1$ we may compute the integral above and find

$$P_{i,i+1}(t) = \begin{cases} \frac{\lambda_i}{\lambda_{i+1}-\lambda_i}(e^{-\lambda_i t} - e^{-\lambda_{i+1} t}), \text{ if } \lambda_{i+1} \neq \lambda_i,\\ \lambda_i + e^{-\lambda_i t}, \text{ if } \lambda_{i+1} = \lambda_i. \end{cases}$$

In the case when $\lambda_i = i\lambda$, we may continue computing P_{ij} and find, for $j \geq i$,

$$P_{ij}(t) = \binom{j-1}{j-i} e^{-i\lambda t}(1 - e^{-\lambda t})^{j-i}.$$

3.3.7 Laplace Transform and Birth-Death Processes

Assume the r.v. X has density f and define $\hat{f}_X(u) = \mathbb{E}[e^{-uX}]$. The following properties are true:

- X and Y are independent $\Rightarrow \hat{f}_{X+Y} = \hat{f}_X \hat{f}_Y$;
- If $\displaystyle\int_{\mathbb{R}} |\hat{f}(x+iy)| dy < +\infty$, then $f_X(t) = \dfrac{1}{2\pi i} \displaystyle\int_{x-i\infty}^{x+i\infty} e^{tu} \hat{f}_X(u) du$;
- If $F_X(t) = \displaystyle\int_{-\infty}^t f_X(x)dx$, then $\hat{F}_X(u) = \dfrac{1}{u}\hat{f}_X(u)$.

Moreover, if $(X_t)_{t\geq 0}$ is a birth-death process, define

$$\tau_{ij} = \inf\{t \geq 0 \ ; \ X_0 = i \text{ and } X_t = j\}.$$

If $i > j$, we may write $\tau_{ij} = \tau_{i,i-1} + \cdots + \tau_{j+1,j}$, where all these times are independents. Let $f_{i,j}$ be the density of τ_{ij}. Then

$$\hat{f}_{ij} = \prod_{k=i}^{j+1} \hat{f}_{k,k-1}.$$

Let us compute $\hat{f}_{k,k-1}$. Denote $\tau_k = \inf\{t \geq 0 \ ; \ X_t = k\}$ and τ is the time the first event of X happens (i.e. $\tau \mid X_0 = k \sim \mathit{Exp}(\lambda_k + \mu_k)$). Hence

$$
\begin{aligned}
\hat{f}_{k,k-1}(u) = \mathbb{E}[e^{-u\tau_{k-1}} \mid X_0 = k] &= \frac{\mu_k}{\mu_k + \lambda_k}\mathbb{E}[e^{-u\tau} \mid X_0 = k]\\
&+ \frac{\lambda_k}{\mu_k+\lambda_k}\mathbb{E}[e^{-u\tau} \mid X_0 = k]\mathbb{E}[e^{-u\tau_k} \mid X_0 = k+1]\mathbb{E}[e^{-u\tau_{k-1}} \mid X_0 = k]\\
&= \frac{\mu_k}{\mu_k+\lambda_k}\frac{\lambda_k + mu_k}{\lambda_k+\mu_k+u} + \frac{\lambda_k}{\mu_k+\lambda_k}\frac{\lambda_k+mu_k}{\lambda_k+\mu_k+u}\hat{f}_{k+1,k}(u)\hat{f}_{k,k-1}(u).
\end{aligned}
$$

Therefore

$$\hat{f}_{k,k-1}(u) = -\frac{1}{\lambda_{k-1}}\frac{-\lambda_{k-1}\mu_k}{\lambda_k + \mu_k + u - \lambda_k \hat{f}_{k+1,k}(u)}$$

One might show that

$$\hat{f}_{k,k-1}(u) = -\frac{1}{\lambda_{k-1}}\Phi_{m=k}^{+\infty}\frac{-\lambda_{m-1}\mu_m}{\lambda_m+\mu_m+u},$$

where

$$\Phi_{n=1}^{+\infty}\frac{a_n}{b_n} = \lim_{n\to+\infty} t_1 \circ \cdots \circ t_n(0),$$

with $t_i(v) = \dfrac{a_i}{b_i + v}$.

3.4 Poisson Process

In this section we further study the Poisson process. Before, some notation:

Definition 3.4.1. A function f is said to be of $o(h)$ if

$$\lim_{h\to 0}\frac{f(h)}{h}=0.$$

We will use the following characterization of the Poisson process:

Definition 3.4.2 (Poisson Process)**.** A process $(N_t)_{t\geq 0}$ taking values in $\mathbb{N}$ is a *Poisson process with rate* λ if

(i) $N_0 = 0$ q.c;

(ii) $N_t - N_s$ is independent of N_s, if $s < t$;

(iii) $\mathbb{P}(N_{t+h} - N_t = 0) = 1 - \lambda h + o(h)$;

(iv) $\mathbb{P}(N_{t+h} - N_t = 1) = \lambda h + o(h)$;

(v) $\mathbb{P}(N_{t+h} - N_t \geq 2) = o(h)$.

N_t is understood as the number of arrivals by time t.

Remark 3.4.3. Notice that, since the increments, $N_{t+h} - N_t$, are always positive, the process $(N_t)_{t\geq 0}$ is increasing. Moreover, the increments are stationary, meaning that the distribution of the increments depends only on h.

Remark 3.4.4. In general, a the increments probabilities of birth and death process can be written as

$$P_{i,i+k}(h)=\mathbb{P}(X_{t+h}-X_t=k\mid X_t=i)=\begin{cases}\lambda_i h+o(h), & \text{if } k=1,\\ \mu_i h+o(h), & \text{if } k=-1,\\ 1-(\lambda_i+\mu_i)h+o(h), & \text{if } k=0,\\ o(h), & \text{otherwise.}\end{cases}$$

Theorem 3.4.5. *N_t is distributed as a Poisson(λt):*

$$\mathbb{P}(N_t=j)=e^{-\lambda t}\frac{(\lambda t)^j}{j!}.$$

Proof. Notice that

$$\begin{aligned}\mathbb{P}(N_{t+h}=j)&=\sum_{i=0}^{+\infty}\mathbb{P}(N_{t+h}=j\mid N_t=i)\mathbb{P}(N_t=i)=\sum_{i=0}^{+\infty}\mathbb{P}(N_{t+h}-N_t=j-i\mid N_t=i)\mathbb{P}(N_t=i)\\&=\sum_{i=0}^{+\infty}\mathbb{P}(N_{t+h}-N_t=j-i)\mathbb{P}(N_t=i)=\sum_{i=0}^{j}\mathbb{P}(N_{t+h}-N_t=j-i)\mathbb{P}(N_t=i)\\&=\lambda h\mathbb{P}(N_t=j-1)+(1-\lambda h)\mathbb{P}(N_t=j)+o(h)\end{aligned}$$

Hence, denoting $p_j(t)=\mathbb{P}(N_t=j)$, we conclude:

$$\begin{aligned}p_j(t+h)&=\lambda h p_{j-1}(t)+(1-\lambda h)p_j(t)+o(h),\\p_0(t+h)&=(1-\lambda h)p_0(t)+o(h),\end{aligned}$$

and $p_j(0)=\delta_0(j)$. This implies

$$\begin{aligned}p_j'(t)&=\lambda p_{j-1}(t)-\lambda p_j(t),\\p_0'(t)&=-\lambda p_0(t).\end{aligned}$$

Hence, $p_0(t)=e^{-\lambda t}$ and

$$p_1'(t)=\lambda e^{-\lambda t}-\lambda p_1(t)\Rightarrow (p_1(t)e^{\lambda t})'=e^{\lambda t}(p_1'(t)+\lambda p_1(t))=\lambda\Rightarrow p_1(t)=\lambda t e^{-\lambda t}.$$

By induction, we find the desired formula. ∎

There are many ways to describe the process $(N_t)_{t\geq 0}$. Define the *arrival times*

$$T_n = \inf\{t \ ; \ N_t = n\},$$

with $T_0 = 0$, and the *interarrival times*

$$X_n = T_n - T_{n-1}.$$

It is interesting to notice that we can reconstruct $(N_t)_{t\geq 0}$ using $(X_n)_{n\in\mathbb{N}}$. Indeed,

$$T_n = \sum_{i=1}^{n} X_i \text{ and } N_t = \max\{n \ ; \ T_n \leq t\}.$$

What is the distribution of $(X_n)_{n\in\mathbb{N}}$?

Proposition 3.4.6. *$(X_n)_{n\in\mathbb{N}}$ are independent and identically distributed as $exp(\lambda)$.*

Proof. Notice $\mathbb{P}(X_1 > t) = \mathbb{P}(N_t = 0) = e^{-\lambda t}$, which means $X_1 \sim \exp(\lambda)$. Notice that

$$\mathbb{P}(X_2 > h \mid X_1 = t_1) = \mathbb{P}(\text{ no arrivals in } (t_1, t_1 + h]) = e^{-\lambda h}.$$

Hence, $X_2 \sim \exp(\lambda)$ and it is independent of X_1. By induction, the result follows.

□

This implies that $T_n \sim \Gamma(n, \lambda)$.

Example 3.4.7. Assume that people immigrate to Brazil at a Poisson rate of $\lambda = 1$ per day.

(i) What is the expected time until the fifth immigrant arrives?

$$\mathbb{E}[T_5] = \frac{5}{\lambda} = 5 \text{ days}.$$

(ii) What is the probability that the elapsed time between the fifth and the sixth arrival exceeds two days

$$\mathbb{P}(X_6 > 2) = e^{-2\lambda} = e^{-2}.$$

Theorem 3.4.8.

$$(T_1, \ldots, T_n) \mid N_t = n \sim (\overline{U}_1, \ldots, \overline{U}_n),$$

where $(\overline{U}_1, \ldots, \overline{U}_n)$ is the sorted version of $(U_1, \ldots, U_n)$, where $U_i \overset{iid}{\sim} U[0,t]$.

Proof. If f is the probability density of $(\overline{U}_1, \ldots, \overline{U}_n)$, one can show that

$$f(t_1, \ldots, t_n) = \frac{n!}{t^n}, \text{ if } 0 \leq t_1 \leq \cdots \leq t_n \leq t.$$

Hence, taking $0 \leq t_1 \leq \cdots \leq t_n \leq t$, we have

$$\begin{aligned}
&\mathbb{P}(T_1 \in (t_1, t_1 + dt_1), \ldots, T_n \in (t_n, t_n + dt_n) \mid N_t = n) \\
&= \frac{\mathbb{P}(T_1 \in (t_1, t_1 + dt_1), \ldots, T_n \in (t_n, t_n + dt_n), T_{n+1} > t)}{\mathbb{P}(N_t = n)} \\
&= \frac{\mathbb{P}(X_1 \in (t_1, t_1 + dt_1), \ldots, X_n \in (t_n - t_{n-1}, t_n - t_{n-1} + dt_n), X_{n+1} > t - t_n)}{\mathbb{P}(N_t = n)} \\
&= \frac{\left(\displaystyle\prod_{i=1}^{n} \lambda e^{-\lambda(t_i - t_{i-1})} dt_i\right) e^{-\lambda(t - t_n)}}{e^{-\lambda t} \dfrac{(\lambda t)^n}{n!}} = \frac{n!}{t^n} dt_1 \cdots dt_n.
\end{aligned}$$

□

Theorem 3.4.9 (Superposition). *Let $(N_i)_{i=1,\dots,k}$ be k independent Poisson processes with rate λ_k, respectively. Hence, $N = N_1 + \cdots + N_k$ is a Poisson process with rate $\lambda = \lambda_1 + \cdots + \lambda_k$. Moreover, one can show that the probability that an arrival is of type i is λ_i/λ.*

Proof. It is clear that N also have stationary and independent increments (since it will the sum of increments of N_i's). Now, since $N_i(t) \sim Poisson(\lambda_i t)$, it is also clear that $N(t) \sim Poisson(\lambda t)$.

□

Theorem 3.4.10 (Splitting). *Consider k classes of events, each with probability p_i of happening. Assume that each arrival can be of type i with probability p_i, independent of everything else. Hence, if N_i denotes the counting process of arrivals of type i, then N_i is a Poisson process with rate λp_i.*

Proof. Let $R(i) \sim Geom(p_i)$ and notice that

$$X_1(i) = \sum_{n=1}^{R(i)} X_n,$$

where X_n is the interarrival time of n^{th} event and $X_1(i)$ is the time of arrival of first event of type i. Notice that $R(i)$ is independent of $(X_n)_{n\in\mathbb{N}}$. Hence, $X_1(i) \sim \exp(\lambda p_i)$ (prove this). Similarly, $X_r(i) \sim \exp(\lambda p_i)$ and they form an iid sequence. Hence, N_i is a Poisson process with rate λp_i.

□

3.5 Application in Finance: Order Book Dynamics

This sections is based on the paper "Stochastic Model for Order Book Dynamics" by Cont, Stoikov and Talreja.

Consider a financial asset that is traded based on an order book. There are two types of order a trader can post:

- *Limited:* it is an order to buy (bid) or sell (ask) a determined quantity of the asset at a given price. Its parameters are type (bid or ask), quantity and price. It guarantees the traded price but it might take too long to be executed. Moreover, limited order could be canceled at any time.
- *Market:* it is an order to buy or sell a determined quantity of the asset at the best price. Its parameters are type (bid or ask) and quantity. It guarantees immediate transaction but the it final traded price is unknown.

The *order book* is the summary of all limited orders. Usually, the market assumes the smallest *tick*, i.e. the asset price must be a multiple of ΔS (the tick). We will assume the asset price cannot be larger than $n\Delta S$. This is not a strong assumption if we consider our modeling for a time period that is not too long.

The model starts by considering $X(t) = (X_1(t), \dots, X_n(t))$, where $|X_p(t)|$ is the quantity available of the asset at time t and price $p\Delta S$. If $X_p(t) < 0$, we have $-X_p(t)$ bid orders at $p\Delta S$. If $X_p(t) > 0$, we have $X_p(t)$ ask orders at $p\Delta S$.

The main assumption of the model is that there is a common unit size of the order. That means that $X_p(t)$ tells us how many orders of that unit size is available at that price. This assumption could be weakened, but it considerably complicates the model.

The best prices (bid and ask) are defined as

$$p_A(t) = \inf\{p \ ; \ X_p(t) > 0\} \vee (n+1),$$
$$p_B(t) = \sup\{p \ ; \ X_p(t) < 0\} \wedge 0.$$

Moreover, we define the mid price as $p_M(t) = (p_A(t) + p_B(t))/2$ and the spread is defined as $s(t) = p_A(t) - p_B(t)$.

In order to define the possible moves in the order book, define $x_{p\pm 1} = x \pm e_p$, where $x \in \mathbb{Z}^n$ and e_p is a vector of zeros and a one at the position p.

- limited bid order at $p < p_A(t)$: $x \mapsto x_{p-1}$;
- limited ask order at $p > p_B(t)$: $x \mapsto x_{p+1}$;
- market bid order at: $x \mapsto x_{p_A(t)-1}$;
- market ask order at: $x \mapsto x_{p_B(t)+1}$;
- canceling of limited bid order at $p < p_A(t)$: $x \mapsto x_{p+1}$;
- canceling of limited ask order at $p > p_B(t)$: $x \mapsto x_{p-1}$;

We will assume that each of events above happens accordingly a birth-death process. More precisely:

- limited orders arrive at a distance i from the opposite best price at a rate λ_i;
- market orders arrive at rate μ;
- canceling order arrive at a distance i from from the opposite best price at rate $\theta_i x$, where x is the quantity of orders at that price.

All the events above happen independently from one another. Therefore,

- $x \mapsto x_{p-1}$ with rate $\lambda_{p_A(t)-p}$, if $p < p_A(t)$;
- $x \mapsto x_{p+1}$ with rate $\lambda_{p-p_B(t)}$, if $p > p_B(t)$;
- $x \mapsto x_{p_B(t)+1}$ with rate μ;
- $x \mapsto x_{p_A(t)-1}$ with rate μ;
- $x \mapsto x_{p+1}$ with rate $\theta_{p_A(t)-p}|x_p|$, if $p < p_A(t)$;
- $x \mapsto x_{p-1}$ with rate $\theta_{p-p_B(t)}|x_p|$, if $p > p_B(t)$;

We hence define

$$X_A(t) = X_{p_A(t)}(t) \text{ and } X_B(t) = X_{p_B(t)}(t).$$

Assume the spread $s(0) = s_0$. Then one might prove that X_V and X_B are independent birth-death process with birth rate λ_{s_0} and death rate $\mu + i\theta_{s_0}$. This is the precise result; for a more rigorous discussion of this see Cont *et al.*'s paper.

Using the model above, we would like to answer probabilistic questions regarding the market. For instance, we would like to know the direction of the price movement. More precisely, if we define

$$\tau = \inf\{t \geq 0 \ ; \ p_M(t) \neq p_M(0)\},$$

we would like to compute

$$\mathbb{P}(p_M(\tau) > p_M(0) \mid X_A(0) = a, X_B(0) = b, s(0) = s_0).$$

Define then

$$\tau_A = \inf\{t \geq 0 \ ; \ X_A(t) = 0\},$$
$$\tau_B = \inf\{t \geq 0 \ ; \ X_B(t) = 0\}.$$

Hence, **assuming** $s_0 = 1$, we find

$$p_M(\tau) > p_M(0) \Leftrightarrow \tau_A < \tau_B,$$

which means we want to compute $\mathbb{P}(\tau_A - \tau_B < 0)$. We will do this using the Laplace transform technique, see the Section 3.3.7 for a discussion of the results in the general case. Notice first that the Laplace transform of $\tau_A - \tau_B$, by independence, is given by

$$\hat{f}(u) = \hat{g}_A(u)\hat{g}_B(-u),$$

where

$$\hat{g}_A(u) = \left(-\frac{1}{\lambda_{s_0}}\right)^a \prod_{i=1}^{a} \Phi_{k=i}^{+\infty} \frac{-\lambda_{s_0}(\mu + k\theta_{s_0})}{\lambda_{s_0} + \mu + k\theta_{s_0} + u},$$

and similar formula for $\hat{g}_B$. In order to compute $\mathbb{P}(\tau_A - \tau_B < 0)$, we need to invert (in the Laplace sense) the function $\frac{1}{u}\hat{g}_A(u)\hat{g}_B(-u)$ and compute this inverse at 0.

3.6 Queueing Modeling

The modeling tool we have study so far is useful to study queueing phenomena. For instance, assume that customers arrive according a Poisson process with rate λ on a website and stays on it for an exponential time with rate μ. Everything happens independently. So, there are three ingredients for this model: arrival, departure and number of servers. Kendall's notation to describe this model is $M/M/k$. The first letter tells how customers arrive; M is Markovian, which is equivalent to being a Poisson process in this case. The second letter tells us how customers departs (or how the service time is modeled). The third and final letter is the number of servers, $k \in \mathbb{N} \cup \{+\infty\}$.

3.6.1 Infinite server queue - $M/M/\infty$

This is a birth-death process with $\lambda_i = \lambda$ and $\mu_i = i\mu$.

Assume $X_0 = i$ and define

- $X_t^{(1)}$: the number of customers arriving in $(0, t]$ and still on the website by time t;
- $X_t^{(2)}$: the number of initial customers on the website by time t.

Notice these processes are independent and $X_t = X_t^{(1)} + X_t^{(2)}$. Hence, in order to compute $P_{ij}(t)$, we will use

$$P_{ij}(t) = \sum_{k=0}^{i \wedge j} \mathbb{P}(X_t^{(2)} = k \mid X_0 = i)\mathbb{P}(X_t^{(1)} = j - k).$$

Notice that one of the initial customers is still on the website by time t with probability $e^{-\mu t}$. Hence, $X_t^{(2)} \mid X_0 = i \sim Bin(i, e^{-\mu t})$. To find the distribution of $X_t^{(1)}$, notice that if a customer arrives at time $s \in (0, t]$, he is still on the website at time t with probability $e^{-\mu(t-s)}$. Assuming customers arrive uniformly in $[0, t]$, the probability that the customer will be on the website by time t is

$$p = \int_0^t e^{-\mu(t-s)} \frac{1}{t} ds = \frac{1 - e^{-\mu t}}{\mu t}.$$

Therefore, if Y_t denote the number of customers that arrived in $(0, t]$, then $X_t^{(1)} \mid Y_t = n \sim Bin(n, p)$. Then

$$\mathbb{P}(X_t^{(1)} = j) = \sum_{n=j}^{+\infty} \mathbb{P}(X_t^{(1)} = j \mid Y_t = n)\mathbb{P}(Y_t = n) = \frac{(\lambda t p)^j}{j!} e^{-\lambda t p},$$

which shows that $X_t^{(1)} \sim Poisson(\lambda t p)$ and $\lambda t p = \frac{\lambda}{\mu}(1 - e^{-\mu t})$. All these probabilities can be put together to find $P_{ij}(t)$.
From detailed balance, one may conclude that the invariant distribution is $\pi \sim Poisson(\lambda/\mu)$ and exists for any λ and μ.

3.6.2 One server queue - $M/M/1$

This is a birth-death process with $\lambda_i = \lambda$ and $\mu_i = \mu$.

Define $\rho = \frac{\lambda}{\mu}$. From detailed balance,

$$\pi_i \lambda = \pi_{i+1}\mu \Rightarrow \pi_i = \rho^i \pi_0.$$

Hence, the process has an stationary distribution if and only if $\rho < 1$. In this case, it is clear that

$$\pi_i = (1-\rho)\rho^i.$$

This also implies that the chain is positive recurrent if $\rho < 1$. Moreover, notice that the embedded process is a random walk on the positive integers with reflection at 0 and probability to jump to the right being $p = \frac{\lambda}{\lambda+\mu}$. This discrete-time Markov chain is transient if and only if $p > 1/2$, which give us $\lambda > \mu$. The case $\lambda = \mu$ yields $p = 1/2$ and the chain is null recurrent. The same classification holds for the continuous-time process.

Additionally, in equilibrium, the waiting time of a customer that arrives at time t is $Exp(\mu - \lambda)$. Indeed, in equilibrium means that $X_0 \sim \pi$. The waiting time of a customer that arrives at time t is given by

$$W = \sum_{i=1}^{X_t+1} T_i,$$

where X_t is the length of the queue at time t and T_i is the service time of the customer in the queue. Notice that $T_i \sim Exp(\mu)$ and $X_t + 1 \sim Geom(\rho)$ starting from 1. Hence, $W \sim Exp(\mu(1-\rho))$.

Another nice result for $M/M/1$ queues if the so-called Burke's Theorem.

Theorem 3.6.1. *In equilibrium, the number of customers that departed up to time t, denoted by D_t, is a Poisson process with rate λ. Moreover, X_t is independent of D_s, $s \le t$.*

Proof. Notice that X in equilibrium ($X_0 \sim \pi$) is reversible. This means that for any $T > 0$ fixed, if we denote $\hat{X}_t = X_{T-t}$, then $(\hat{X}_t)_{t\in[0,T]}$ and $(X_t)_{t\in[0,T]}$ have the same distribution. Moreover, when $\hat{X}$ has a jump of size $+1$ at time t, then X has a jump of size -1 at time $T-t$, i.e. departures in X mean arrivals for $\hat{X}$. □

3.7 Renewal Processes

Renewal processes is a mathematical tool for modeling arrival of certain events.

Definition 3.7.1. Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of iid positive r.v's with $\mathbb{P}(X_1 > 0) > 0$. These will model the inter-arrival times. The arrival time of the n-th event is then

$$T_n = \sum_{i=1}^{n} X_i.$$

We then the define the number of arrivals up to time t as

$$N_t = \max\{n \in \mathbb{N} \ ; \ T_n \le t\}.$$

The process $(N_t)_{t\ge 0}$ is called the renewal process for the sequence $(X_i)_{i\in\mathbb{N}}$.

Remark 3.7.2. For any fixed t, $N_t + 1$ is a stopping time for the inter-arrival sequence $(X_n)_{n\in\mathbb{N}}$, i.e. $\{N_t + 1 = n\}$ depends only on $X_1, \dots, X_n$. Indeed, $N_t = n-1$ if and only if $T_{n-1} \le t < T_n$. Hence, the event $\{N_t + 1 = n\}$ depends only on $X_1, \dots, X_n$.

Lemma 3.7.3 (Wald's Equation)**.** *If τ is a stopping time for $(X_i)_{i\in\mathbb{N}}$, then*

$$\mathbb{E}\left[\sum_{i=1}^{\tau} X_i\right] = \mathbb{E}[\tau]\mathbb{E}[X_1]$$

Proof. Notice that

$$\sum_{i=1}^{\tau} X_i = \sum_{i=1}^{+\infty} X_i 1_{\{i\le\tau\}}.$$

Moreover, $\{i \le \tau\} = \{i > \tau\}^c$ depends only on $X_1, \dots, X_{i-1}$. Hence, it is independent of X_i. Since, there are only positive terms in the infinite sum above, we have

$$\mathbb{E}\left[\sum_{i=1}^{\tau} X_i\right] = \sum_{i=1}^{+\infty} \mathbb{E}[X_i 1_{\{i\le\tau\}}] = \sum_{i=1}^{+\infty} \mathbb{E}[X_i]\mathbb{E}[1_{\{i\le\tau\}}]$$

$$= \mathbb{E}[X_1] \sum_{i=1}^{+\infty} \mathbb{P}(i \leq \tau) = \mathbb{E}[X_1]\mathbb{E}[\tau].$$

□

The next theorem tells us the asymptotic behavior of N.

Theorem 3.7.4 (Elementary Renewal Theorem)**.** *If $\mathbb{E}[X_1] = 1/\lambda < +\infty$, then*

$$\lim_{t\to+\infty} \frac{N_t}{t} = \lambda \text{ a.s. and } \lim_{t\to+\infty} \frac{\mathbb{E}[N_t]}{t} = \lambda.$$

Proof. Trivially

$$T_{N_t} \leq t < T_{N_t+1} \Rightarrow \frac{T_{N_t}}{N_t} \leq \frac{t}{N_t} < \frac{T_{N_t+1}}{N_t}.$$

Then, the first result follows from $T_n/n \to 1/\lambda$ a.s. (LLN) and $N_t \to +\infty$. Let us now prove the second limit. Notice that $N_t + 1$ is a stopping time for the sequence $(X_i)_{i\in\mathbb{N}}$. By Wald's equation

$$t < \mathbb{E}[T_{N_t+1}] = \mathbb{E}[N_t + 1]\mathbb{E}[X_1] = (\mathbb{E}[N_t] + 1)\frac{1}{\lambda}.$$

Then

$$\liminf_{t\to+\infty} \frac{\mathbb{E}[N_t]}{t} \geq \lambda.$$

Consider now $X_i^a = \min\{X_i, a\}$ and note $N_t \leq N_t^a$ and

$$T^a_{N_t^a+1} = T^a_{N_t^a} + X^a_{N_t^a+1} \leq t + a.$$

Hence,

$$t + a \geq \mathbb{E}[T^a_{N_t^a+1}] = (\mathbb{E}[N_t^a] + 1)\mathbb{E}[X_1^a] \geq (\mathbb{E}[N_t] + 1)\mathbb{E}[X_1^a].$$

This implies

$$\frac{\mathbb{E}[N_t]}{t} \leq \left(\frac{1}{\mathbb{E}[X_1^a]} + \frac{a}{t\mathbb{E}[X_1^a]}\right) - \frac{1}{t}.$$

Therefore,

$$\limsup_{t\to+\infty} \frac{\mathbb{E}[N_t]}{t} \leq \frac{1}{\mathbb{E}[X_1^a]}.$$

Letting $a \to +\infty$ (and using MCT), we find the desired result. □

Theorem 3.7.5 (CLT for Renewal Processes)**.** *If $\mathbb{E}[X_1] = 1/\lambda < +\infty$ and $Var(X_1) = \sigma^2 < +\infty$, then, we $t \to +\infty$,*

$$Z_t = \frac{N_t - \lambda t}{\sigma\sqrt{\lambda^3 t}} \xrightarrow{\mathcal{L}} N(0,1).$$

Proof. Notice

$$Y_n = \frac{T_n - n/\lambda}{\sigma\sqrt{n}} \xrightarrow{\mathcal{L}} N(0,1).$$

and that $\mathbb{P}(N_t < n) = \mathbb{P}(T_n > t)$. Define, for each x, $n_t = \lambda t + x\sigma\sqrt{\lambda^3 t}$ (better yet, the integer part of n_t). Hence

$$\begin{aligned}\mathbb{P}(Z_t < x) &= \mathbb{P}(N_t < n_t) = \mathbb{P}(T_{n_t} > t)\\ &= \mathbb{P}\left(\frac{T_{n_t} - n_t/\lambda}{\sigma\sqrt{n_t}} > \frac{t - n_t/\lambda}{\sigma\sqrt{n_t}}\right) = \mathbb{P}\left(Y_{n_t} > -\frac{x}{\sqrt{1 + (x\sigma)/\sqrt{t/\lambda}}}\right).\end{aligned}$$

Letting $t \to +\infty$, we find the result. □

Usually, arrivals are connected to rewards of a certain kind. This could be modeled as follows. Let $(Y_i)_{i\in\mathbb{N}}$ be a iid sequence of r.v's with common mean $\nu = \mathbb{E}[Y]$. The sequences of Y's and X's could depend on each other. The renewal-reward process is then defined as

$$R_t = \sum_{i=1}^{N_t} Y_i.$$

Theorem 3.7.6 (Renewal-Reward Theorem)**.**

$$\lim_{t\to+\infty} \frac{R_t}{t} = \nu\lambda \text{ } a.s. \text{ } and \lim_{t\to+\infty} \frac{\mathbb{E}[R_t]}{t} = \nu\lambda.$$

Proof. By the LLN and since $N_t \to +\infty$ a.s, we find

$$\frac{R_t}{N_t} = \frac{1}{N_t}\sum_{i=1}^{N_t} Y_i \xrightarrow{a.s.} \nu.$$

We have seen above that $N_t/t \xrightarrow{a.s.} \lambda$. Hence

$$\frac{R_t}{t} = \frac{R_t}{N_t}\frac{N_t}{t} \xrightarrow{a.s.} \nu\lambda.$$

To prove the second convergence, notice that

$$|R_t| \leq \sum_{i=1}^{N_t} |Y_i| \leq \sum_{i=1}^{N_t+1} |Y_i|.$$

The r.v. $N_t + 1$ is a stopping time for the sequence $(X_i)_{i\in\mathbb{N}}$ and hence, it is also a stopping time for the pairs $(X_i, Y_i)_{i\in\mathbb{N}}$. By Wald's equation

$$\mathbb{E}\left[\sum_{i=1}^{N_t+1} |Y_i|\right] = \mathbb{E}[|Y_1|]\mathbb{E}[N_t + 1] = \mathbb{E}[|Y_1|](\mathbb{E}[N_t] + 1).$$

Thus

$$\frac{\mathbb{E}\left[\sum_{i=1}^{N_t+1} |Y_i|\right]}{t} = \frac{\mathbb{E}[|Y_1|]\mathbb{E}[N_t] + \mathbb{E}[|Y_1|]}{t} \xrightarrow{t\to+\infty} \mathbb{E}[|Y_1|]\lambda.$$

Hence, the sequence $\frac{\sum_{i=1}^{N_t+1} |Y_i|}{t}$ is u.i. (by the Theorem 2.5.13) Therefore, the same is true for the sequence R_t/t. This concludes the proof again by Theorem 2.5.13. $\square$

Chapter 4

Gaussian Processes

Throughout this chapter, we fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

4.1 Multivariate Gaussian Variable

Consider random vectors $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$. We define

$$\mathbb{E}[X] = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_n]),$$
$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])^T].$$

We use the simplified notation: $\mathrm{Cov}(X) = \mathrm{Cov}(X, X)$.

Definition 4.1.1. The vector $X = (X_1, \ldots, X_n)$ is said to have a multivariate normal distribution if

$$X = DW + \mu,$$

where $\mu \in \mathbb{R}^n$, $D \in \mathbb{R}^{n \times k}$ and $W = (W_1, \ldots, W_k)$, with $W_i \overset{iid}{\sim} N(0, 1)$.

Theorem 4.1.2. *Assume X is multivariate normal.*

1. *X_i is normal with mean μ_i and variance $\sum_{j=1}^{k} d_{ij}^2$;*
2. *$Cov(X) = DD^T = \Sigma$;*
3. *If $C \in \mathbb{R}^{m \times n}$ and $d \in \mathbb{R}^m$, then $Y = CX + d$ is multivariate normal with mean $C\mu + d$ and covariance matrix $C\Sigma C^T$;*
4. *If $\det \Sigma > 0$, then*
$$f_X(x) = \frac{1}{\sqrt{(2\pi)^n |\det \Sigma|}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)};$$
5. *The distribution of X is completely determined by its means and covariance;*
6. *The components of X are uncorrelated if and only if they are independent.*

Notation: $X \sim N(\mu, \Sigma)$.

Theorem 4.1.3. *If*
$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix} \right)$$
and Σ_{YY} is positive definite, then

1. *$X \sim N(\mu_X, \Sigma_{XX})$ and $Y \sim N(\mu_Y, \Sigma_{YY})$;*
2. *$\mathbb{E}[X \mid Y] = \mu_X + \Sigma_{XX}\Sigma_{YY}^{-1}(Y - \mu_Y)$;*
3. *$X - \mathbb{E}[X \mid Y]$ is independent of Y and $\mathbb{E}[X \mid Y]$;*

4. *Cov*$(X - \mathbb{E}[X \mid Y] \mid Y) =$ *Cov*$(X - \mathbb{E}[X \mid Y]) = \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}$;

5. *Hence,* $X \mid Y \sim N(\mu_X + \Sigma_{XX}\Sigma_{YY}^{-1}(Y - \mu_Y), \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX})$.

Example 4.1.4. Assume $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ are univariate normal with $\mathrm{Cov}(X, Y) = \rho\sigma_X\sigma_Y$. Then

$$\mathbb{E}[X \mid Y] = \mu_X + \frac{\sigma_X^2}{\sigma_Y^2}(Y - \mu_Y) \text{ and } \mathrm{Var}(X \mid Y) = \sigma_X^2(1 - \rho^2).$$

4.2 Gaussian Processes

Definition 4.2.1 (Gaussian processes)**.** Let $\mathbb{X}$ be a some set, usually a subset of $\mathbb{R}^d$. A stochastic process $(f(x))_{x\in\mathbb{X}}$ is a *Gaussian process* if, for any $x_1, \ldots, x_n$ in $\mathbb{X}$, the random vector $(f(x_1), \ldots, f(x_n))$ is multivariate normal.

It is important to notice that a Gaussian process is completely determined by the function $\mu(x) = \mathbb{E}[f(x)]$ and $k(x, x') = \mathrm{Cov}(f(x), f(x'))$. Moreover, I would like to point out that the set $\mathbb{X}$ does not need to be ordered. It could be, for example, a subset of $\mathbb{R}^d$ (this is important for spatial-temporal stochastic processes).

Notation: $f \sim GP(m, k)$.

What are the function m and k that could be mean and covariance functions of a Gaussian process? The function m could be arbitrarily chosen. However, the function k must be such that, for any $x_1, \ldots, x_n \in \mathbb{X}$, the matrix

$$\begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \cdots & k(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n, x_1) & k(x_n, x_2) & \cdots & k(x_n, x_n) \end{bmatrix}$$

must be a valid covariance function. In fact, it needs to be symmetric and positive definite. Symmetry is easily attained by asking the kernel k to be symmetric: $k(x, x') = k(x', x)$. The positivity definite of the matrix is more subtle and outside the scope of these notes.

Example 4.2.2. An important example of a valid kernel is the *square exponential*

$$k(x, x') = \sigma_f e^{-\frac{1}{2\tau^2}\|x - x'\|^2}.$$

There are several other examples of kernels that might imply different behaviors for the Gaussian process. Moreover we can combine the kernel in several ways (sum, product, etc.).

A very important example of a Gaussian process is the Brownian motion that will examine thoroughly in the next chapter. Indeed, any process with independent, Gaussian increments is a Gaussian process.

4.3 Application in Machine Learning: Gaussian Processes Regression

We start by considering the Bayesian linear regression

$$y = X^T\theta + \epsilon,$$

where $y \in \mathbb{R}^m$, $\theta \in \mathbb{R}^n$, $\epsilon \in \mathbb{R}^m$ and $X^T \in \mathbb{R}^{m\times n}$. We assume the noise $\epsilon \sim N(0, \sigma^2 I)$, where σ^2 is known. Given the prior distribution $\theta \sim N(0, \Sigma)$ and a sample $S = \{(x_1, y_1), \ldots, (x_m, y_m)\}$ from the model above, we find (after a lot of calculations) that

$$y_* \mid x_*, S \sim N\left(\frac{1}{\sigma^2}x_*^T A^{-1}Xy,\ x_*^T A^{-1}x_* + \sigma^2\right),$$

where $A = \dfrac{1}{\sigma^2}XX^T + \Sigma^{-1}$. We are assuming (x_*, y_*) comes from the same model.

The Gaussian process regression comes as a generalization of the linear model above:

$$y = f(x) + \epsilon.$$

The prior information is now modeled by assuming $f \sim GP(0, k)$, for some valid covariance kernel. Assume as well that instead of just of test observation, we have $T = \{(x_1^*, y_1^*), \dots, (x_\ell^*, y_\ell^*)\}$ (S and T are mutually independent). Notice that, since $f \sim GP(0, k)$, we have

$$\begin{bmatrix} f(X) \\ f(X^*) \end{bmatrix} \mid X, X^* \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} K(X,X) & K(X,X^*) \\ K(X^*,X) & K(X^*,X^*) \end{bmatrix} \right),$$

where $f(X) = (f(x_1), \dots, f(x_n))$ and the other elements follow similar logic. Since the noise ϵ is assumed to be independent and Gaussian as in the linear model, we find that

$$\begin{bmatrix} Y \\ Y^* \end{bmatrix} \mid X, X^* \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} K(X,X) + \sigma^2 I & K(X,X^*) \\ K(X^*,X) & K(X^*,X^*) + \sigma^2 I \end{bmatrix} \right).$$

Therefore, we finally find that $Y^* \mid Y, X, X^*$ is Gaussian with:

$$\begin{aligned} \text{Mean: } & K(X^*, X)(K(X,X) + \sigma^2 I)^{-1} Y, \\ \text{Covariance: } & K(X^*, X^*) + \sigma^2 I - K(X^*, X)(K(X,X) + \sigma^2 I)^{-1} K(X, X^*) \end{aligned}$$

Chapter 5

General Stochastic Processes

5.1 Introduction

Definition 5.1.1. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. A *stochastic process* is a collection of random variables. Mathematically, $(X_t)_{t\in\mathcal{T}}$ is a stochastic process if X_t is $\mathcal{F}$-measurable, for all $t \in \mathcal{T}$.

There are three ways to see a stochastic process (taking values on $\mathcal{S}$):

1. For a fixed $t \in \mathcal{T}$, $X_t : \Omega \longrightarrow \mathcal{S}$, i.e. a random variable.
2. For a fixed ω, $X(\omega) : \mathcal{T} \longrightarrow \mathcal{S}$, i.e. a path.
3. $X : \mathcal{T} \times \Omega \longrightarrow \mathcal{S}$.

The first example of stochastic process is the well-known independent and identically distributed (iid) sequence. This is a very simple case where there is no dependence between the random variables. The classes of stochastic processes we have studied so far in these notes that weakens this assumption and allows more dependence between the r.v's are: Markov processes and Martingales. We have also allowed the time index set $\mathcal{T}$ to be discrete $\mathbb{N}$ or continuous $[0, +\infty)$.

In the next chapter, we will study a very important example of a Markov process and martingale: the Brownian motion. This is a building block for various applications. We then develop the well-known Itô 's stochastic calculus, that deals with functions of the Brownian motion. Before we venture in this area, let us state some general facts about stochastic processes. We will consider the case $\mathcal{T} = [0, +\infty)$ for simplicity of notation.

5.2 Equality of Processes

First we will discuss three notions of equality among stochastic processes.

Definition 5.2.1. Let $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ be two stochastic processes in $(\Omega, \mathcal{F}, \mathbb{P})$. We say that:

- Y is a modification of X if $X_t = Y_t$ a.s. for every $t \geq 0$ (i.e. $\mathbb{P}(X_t = Y_t) = 1$, for all $t \geq 0$);
- X and Y have the same finite-dimensional distributions (fdd) if, for every $n \in \mathbb{N}$ and $0 \leq t_1 < t_2 < \cdots < t_n$, we have $(X_{t_1}, \ldots, X_{t_n}) \sim (Y_{t_1}, \ldots, Y_{t_n})$;
- X and Y are indistinguishable if $\mathbb{P}(X_t = Y_t,\ \forall\, t \geq 0) = 1$.

It is clear that indistinguishability implies modification that implies they have the same fdd. However, the are not the same notion of equality. For instance, the process $Y_t = 1_{\tau=t}$, for a continuous r.v. τ is a modification of the process constant equals zero para it is not indistinguishable of it. Moreover, it is worth noticing that in order for two processes to have the same fdd they do not have to be defined in the same probability space.

We say that the sample paths of X are a.s. right(left)-continuous if $\Omega_0 = \{\omega \in \Omega\,;\ t \mapsto X_t(\omega) \text{ is right(left)-continuous})\}$ has probability 1.

Proposition 5.2.2. *Let X and Y processes with a.s right(left)-continuous sample paths. If Y is a modification of X, then X and Y are indistinguishable.*

5.3 Measurability

We have defined a stochastic process as a collection of r.v's meaning that X_t is $\mathcal{F}$-measurable for each t. It is important to add assumptions on the measurability with respect to t.

Definition 5.3.1. We say that X is measurable if the mapping $(t,\omega) \mapsto X_t(\omega)$ is measurable with respect to $\mathcal{B}([0,+\infty)) \otimes \mathcal{F}$.

This implies that the sample paths of X are Borel-measurable function of $t \in [0,+\infty)$.

We will consider now a family of sigma-algebras on $(\Omega, \mathcal{F})$ called a filtration: $(\mathcal{F}_t)_{t\geq 0}$. We require that $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$, for all $0 \leq s \leq t$. Given a process X, we define its natural filtration as

$$\mathcal{F}_t^X = \sigma(X_s \ ; \ s \leq t),$$

i.e. the smallest sigma-algebra such that X_s is measurable for every $s \leq t$. We understand $A \in \mathcal{F}_t^X$ as knowing whether A has occurred by time t if we observe X.

Additional assumptions on $(\mathcal{F}_t)_{t\geq 0}$ are required because of the following results:

Proposition 5.3.2. *Given a stochastic process $(X_t)_{t\geq 0}$, define $C_s = \{\omega \in \Omega \ ; \ t \mapsto X_t(\omega)$ is continuous in $[0,s]\}$. Then*

(i) If the sample paths of $(X_t)_{t\geq 0}$ are càdlàg[1], then $C_s \in \mathcal{F}_s^X$;

(ii) in the case where the sample paths are a.s. càdlàg, the C_s might not belong to $\mathcal{F}_s^X$, but if $(\mathcal{F}_t)_{t\geq 0}$ is filtration such that $\mathcal{F}_t^X \subset \mathcal{F}_t$ and $\mathcal{F}_s$ is complete under $\mathbb{P}$[2], then $C_s \in \mathcal{F}_s$.

Definition 5.3.3. We say that a filtration is complete if $\mathcal{F}_0$ contains all the null sets of $(\Omega, \mathcal{F}, \mathbb{P})$, i.e. all the sets $A \subset N \in \mathcal{F}$ such that $\mathbb{P}(N) = 0$. This implies that $\mathcal{F}_t$ contains all the null sets for every $t \geq 0$.

Another very important notation is being adapted to a filtration:

Definition 5.3.4. We say that X is adapted to $(\mathcal{F}_t)_{t\geq 0}$ if X_t is $\mathcal{F}_t$-measurable for every $t \geq 0$.

Clearly every process is adapted to its natural filtration. Moreover, it is easy to see that if X is adapted to $(\mathcal{F}_t)_{t\geq 0}$ and Y is a modification of X, then Y is also adapted to $(\mathcal{F}_t)_{t\geq 0}$ if this filtration is complete.

Being measurable and adapted does not guarantee that the mapping $(s,\omega) \mapsto X_s(\omega)$ will be measurable with respect to $\mathcal{B}([0,t]) \otimes \mathcal{F}_t$. We then consider the following definition:

Definition 5.3.5. We say that X progressively measurable with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$ if, for every $t \geq 0$, $(s,\omega) \mapsto X_s(\omega)$ is measurable with respect to $\mathcal{B}([0,t]) \otimes \mathcal{F}_t$.

The following result is the cornerstone of this section:

Theorem 5.3.6. *If the process X is measurable and adapted to $(\mathcal{F}_t)_{t\geq 0}$, then there exists a modification of X that is progressively measurable.*

This theorem makes the assumption of completeness of a filtration very important when considering general stochastic process. With additional assumptions, we do not have to consider a modification of the stochastic process.

Proposition 5.3.7. *If a stochastic process has right(left)-continuous sample paths and is adapted, then it is also progressively measurable.*

[1] right-continuous with finite left limits

[2] We say that a sigma-algebra $\mathcal{G} \subset \mathcal{F}$ is complete under $\mathbb{P}$ if $A \subset N \in \mathcal{G}$ and $\mathbb{P}(N) = 0$ implies that $A \in \mathcal{G}$.

Proof. Let us consider the right-continuous. The following is a very useful approximation of the sample paths of X. Fix $t > 0$ and $n \in \mathbb{N}$ and define

$$X_s^{(n)} = \sum_{k=1}^{2^n} X_{\frac{kt}{2^n}} 1_{\left(\frac{(k-1)t}{2^n}, \frac{kt}{2^n}\right]}(s)$$

with $X_0^{(n)} = X_0$. The process $X^{(n)}$ is clearly $\mathcal{B}([0,t]) \otimes \mathcal{F}_t$-measurable. Moreover, since the sample paths are right-continuous for every ω, then $X^{(n)}$ converges to X for every $(s, \omega) \in [0,t] \times \Omega$ and we conclude the result. $\square$

Remark 5.3.8. Measurability of X also holds under the assumption of the proposition (and it follows by the same argument).

Chapter 6

Brownian Motion

Throughout this chapter, we fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

6.1 Introduction

We will introduce the most important stochastic process: the Brownian motion. This process somehow has the same role, in the world of stochastic processes, as the Gaussian random variable in the world of probability distributions.

Definition 6.1.1 (Brownian motion)**.** A stochastic process $(B_t)_{t\geq 0}$ is called a *(standard) Brownian motion* in $[0,T]$ if

(i) $B_0 = 0$ a.s;

(ii) For any $0 \leq t_1 < \ldots < t_n$, the increments $B_{t_{i+1}} - B_{t_i}$, $i \in \{1, \ldots, n-1\}$, are independent;

(iii) For any $0 \leq s < t$, $B_t - B_s \sim N(0, t-s)$. In particular, $B_t \sim N(0,t)$;

(iv) $\mathbb{P}(\omega \in \Omega \ ; \ t \mapsto B_t(\omega) \text{ is continuous }) = 1$.

With simple modification, the definition above could consider the Brownian motion in $[0,T]$.

There are two main issues now: existence and uniqueness. Does a process satisfying the assumptions above exist? Is it unique? Before answering these questions, let us derive some properties of B following the definition above.

Properties (ii) and (iii) above completely defined the finite-dimensional distribution of B. Indeed, it is clearly a Gaussian process with $m(t) = 0$ and, if $s \geq t$,

$$\begin{aligned} k(s,t) &= \mathrm{Cov}(B_s, B_t) = \mathbb{E}[B_s B_t] = \mathbb{E}[(B_s - B_t + B_t)B_t] \\ &= \mathbb{E}[(B_s - B_t)B_t] + \mathbb{E}[B_t^2] = \mathbb{E}[(B_s - B_t)]\mathbb{E}[B_t] + t = t \end{aligned}$$

Similarly, if $s < t$, $k(s,t) = s$. Hence, $k(s,t) = s \wedge t$.
The next proposition shows that the only Gaussian process with $k(s,t) = s \wedge t$ and continuous paths is the Brownian motion.

Proposition 6.1.2. *Let $(X_t)_{t\in[0,T]}$ be a Gaussian process with $m(t) = 0$ and $k(s,t) = s \wedge t$, for any $s, t \in [0,T]$. Further assume that $(X_t)_{t\in[0,T]}$ has continuous sample paths and $X_0 = 0$ a.s. Then $(X_t)_{t\in[0,T]}$ is a Brownian motion.*

Proof. Since X is a Gaussian process, for any $t_1 < t_2 < \ldots t_n$ in $[0,T]$, the vector $(X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}})$ has n-dimensional multivariate normal distribution with mean zero and certain covariance matrix (this vector is the difference of two Gaussian vectors). The only thing we need to verify to prove that X is a Brownian motion is that it has independent increments, i.e. that the covariance matrix is diagonal. Indeed, for $i < j$,

$$\mathrm{Cov}(X_{t_i} - X_{t_{i-1}}, X_{t_j} - X_{t_{j-1}}) = \mathrm{Cov}(X_{t_i}, X_{t_j}) - \mathrm{Cov}(X_{t_i}, X_{t_{j-1}})$$

$$
\begin{aligned}
&- \operatorname{Cov}(X_{t_{i-1}}, X_{t_j}) + \operatorname{Cov}(X_{t_{i-1}}, X_{t_{j-1}}) \\
&= t_i - t_i - t_{i-1} + t_{i-1} = 0
\end{aligned}
$$

The same follows for $i > j$. $\square$

Example 6.1.3. From the discussion above, it is straightforward to see that the density of $(B_{t_1}, \ldots, B_{t_n})$ is giving by

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

Moreover, using the joint density above, we can derive the derive the density of B_s given $B_t = x$, for $t > s$,:

$$
B_s \mid B_t = x \sim N\left(\frac{s}{t}x,\ \frac{s}{t}(t-s)\right).
$$

Let us consider here the natural filtration of B: $\mathcal{F}_t = \sigma(B_s\ ;\ s \leq t)$.

Theorem 6.1.4 (Markov property)**.** *$(B_{t+s} - B_s)_{t\geq 0}$ is independent of $\mathcal{F}_s$.*

Proof. Notice that the theorem is equivalent of show that, for any $0 \leq t_1 < \cdots < t_n$ and $0 \leq s_1 < \cdots < s_m \leq s$, the vectors $(B_{t_1+s} - B_s, \ldots, B_{t_n+s} - B_s)$ and $(B_{s_1}, \ldots, B_{s_m})$ are independent, which is directly verified noticing that the Brownian motion has independent increments. $\square$

For the next theorem, we suggest the reader to read the first pages of Chapter 7.

Theorem 6.1.5 (Martingale property)**.** *The Brownian motion is a martingale with respect to $(\mathcal{F}_t)_{t\geq 0}$*

Proof. Adaptedness and integrability follows directly. The martingale property follows from: $\mathbb{E}[B_t - B_s] = 0$ and $B_t - B_s$ independent of $\mathcal{F}_s$. Indeed,

$$
\begin{aligned}
\mathbb{E}\left[B_t \mid \mathcal{F}_s\right] &= \mathbb{E}\left[B_s + (B_t - B_s) \mid \mathcal{F}_s\right] = \mathbb{E}\left[B_s \mid \mathcal{F}_s\right] + \mathbb{E}\left[(B_t - B_s) \mid \mathcal{F}_s\right] \\
&= B_s + \mathbb{E}[B_t - B_s] = B_s.
\end{aligned}
$$

$\square$

Remark 6.1.6. The Brownian motion generates several other interesting martingales. We suggest the reader prove that $X_t = B_t^2 - t$ and $Y_t = e^{\alpha B_t - \alpha^2 t/2}$ are martingales with respect to natural filtration of the Brownian motion.

Additionally, the Brownian motion is a strong Markov process and this implies the following:

Theorem 6.1.7 (Reflection principle)**.** *If τ is a stopping time wrt the Brownian filtration,*

$$
X_t = \begin{cases} B_t, & \text{if } t < \tau, \\ B_\tau - (B_t - B_\tau), & \text{if } t \geq \tau, \end{cases}
$$

is a Brownian motion.

This result could be used to find the joint distribution of B_t and $B_t^* = \max_{s\in[0,t]} B_s$. Indeed, define the following stopping time

$$
\tau_x = \min\{t\ ;\ B_t = x\}
$$

and notice

$$
\mathbb{P}(\tau \leq t, B_t > x + y) = \mathbb{P}(\tau \leq t, X_t < x - y) = \mathbb{P}(\tau \leq t, B_t < x - y),
$$

where the first equality follows from the fact $X_t = 2x - B_t$ for $t \geq \tau$ and the second equality follows from the reflection principle. Moreover, notice that

$$
\{\tau \leq t\} = \{B_t^* \geq x\}.
$$

Hence, for $y > 0$, we find

$$
\mathbb{P}(B_t > x + y) = \mathbb{P}(B_t^* \geq x, B_t > x + y) = \mathbb{P}(B_t^* \geq x, B_t < x - y),
$$

which implies that

$$\mathbb{P}(B_t < u, B_t^* \geq v) = 1 - \Phi\left(\frac{2v-u}{\sqrt{t}}\right).$$

Therefore,

$$\mathbb{P}(B_t < u, B_t^* < v) = \mathbb{P}(B_t < u) - \mathbb{P}(B_t < u, B_t^* \geq v) = \Phi\left(\frac{u}{\sqrt{t}}\right) - \Phi\left(\frac{u-2v}{\sqrt{t}}\right).$$

Finally, we find

$$f_{B_t,B_t^*}(u,v) = \frac{2(2v-u)}{\sqrt{2\pi t^3}} e^{-\frac{(2v-u)^2}{2t}},$$

for $(u,v) \in D = \{(u,v) \in \mathbb{R}^2 \ ; \ u \in \mathbb{R}, v \geq u^+\}$. On the other hand, notice that

$$\begin{cases} \mathbb{P}(B_t^* \geq x, B_t \leq x) = \mathbb{P}(B_t > x), \\ \mathbb{P}(B_t^* \geq x, B_t > x) = \mathbb{P}(B_t > x) \text{ (trivially)}. \end{cases}$$

Then, for any $x \geq 0$,

$$\mathbb{P}(B_t^* \geq x) = 2\mathbb{P}(B_t \geq x) = \mathbb{P}(|B_t| \geq x),$$

i.e. $B_t^* \sim |B_t|$, for every $t \geq 0$. However, as processes, $(B_t^*)_{t\in[0,T]}$ and $(|B_t|)_{t\in[0,T]}$ could not be more different. Additionally, we find

$$\mathbb{P}(B_t^* \leq x) = 2\Phi\left(\frac{x}{\sqrt{t}}\right) - 1.$$

Thus

$$f_{B_t^*}(x) = \frac{2}{\sqrt{t}}\phi\left(\frac{x}{\sqrt{t}}\right),$$

for $x \geq 0$. Equivalently, we can derive the probability density of τ_x:

$$\mathbb{P}(\tau_x > t) = \mathbb{P}(B_t^* < x) = 2\Phi\left(\frac{x}{\sqrt{t}}\right) - 1.$$

Therefore,

$$f_{\tau_x}(t) = \frac{x}{t^{3/2}}\phi\left(\frac{x}{\sqrt{t}}\right),$$

for $t \geq 0$. This implies the interesting, non-trivial fact

$$\mathbb{P}(\tau_x \geq t) \leq \int_t^{+\infty} \frac{x}{2s^{3/2}} ds = \frac{x}{\sqrt{t}},$$

where we have used the fact that $\phi(y) \leq 1/2$.

We finish this section with some additional properties of B. Their proofs are left as exercises.

Proposition 6.1.8. *Let $(B_t)_{t\in[0,+\infty)}$ be a Brownian motion.*

1. *(scaled process) For any $a > 0$, $X_t = \frac{1}{\sqrt{a}} B_{at}$;*
2. *(inverted process) $X_t = tB_{1/t}$, with $X_0 = 0$;*
3. *(translated process) $X_t = B_{t+s} - B_s$, for $t \geq 0$;*
4. *(reverted process) $X_t = B_T - B_{T-t}$, for $0 \leq t \leq T$;*
5. *(symmetric process) $X_t = -B_t$.*

are standard Brownian process.

Remark 6.1.9. One tricky part to prove in the proposition above is the continuity at 0 of inverted process. In order to do this, notice that, for any process Y_t, we have

$$\left\{\lim_{t\to 0^+} Y_t = 0\right\} = \bigcap_{m=1}^{+\infty} \bigcup_{n=1}^{+\infty} \left\{|Y_t| \leq \frac{1}{m}, \text{ for all } t \in \mathbb{Q} \cap (0, 1/n)\right\}.$$

Denote the set inside the intersection and the union by $A_{k,m}(Y)$. Notice that its probability depends only of the finite-dimensional distributions of Y (and away of 0). Since B and the inverted process B have the same finite-dimensional distributions away from 0, we find that $\mathbb{P}(A_{k,m}(B)) = \mathbb{P}(A_{k,m}(X))$. But it is clear that $\mathbb{P}(\lim_{t\to 0^+} B_t = 0) = 1$. Hence, the same is true for X.

6.2 Existence By Series Approximation

Consider a Hilbert space $\mathbb{H}$ with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$ generated by this inner product. In this space, consider a sequence $(\phi_n)_{n\in\mathbb{N}} \subset \mathbb{H}$ of orthonormal functions, i.e. $\langle \phi_n, \phi_m \rangle = 0$, for $n \neq m$, and $\|\phi_n\| = 1$.

We further assume that this sequence is complete meaning that the set of its linear combinations is dense in $\mathbb{H}$. That means that, for any $f \in \mathbb{H}$, there exists a sequence of real number $(a_n)_{n\in\mathbb{N}}$ such that

$$f = \sum_{n=1}^{+\infty} a_n \phi_n.$$

One can further prove that $a_n = \langle f, \phi_n \rangle$ (use continuity and linearity of the inner product and then that $(\phi_n)_{n\in\mathbb{N}}$ is an othornormal basis). More precisely,

$$\lim_{N\to+\infty} \left\| f - \sum_{n=1}^{N} \langle f, \phi_n \rangle \phi_n \right\| = 0,$$

or $f_N \to f$ in $\mathbb{H}$, where

$$f_N = \sum_{n=1}^{N} \langle f, \phi_n \rangle \phi_n.$$

First, notice that

$$\|f\|^2 = \sum_{n=1}^{+\infty} \langle f, \phi_n \rangle^2.$$

This is called Parseval's identity and follows from

$$\left\| f - \sum_{n=1}^{N} \langle f, \phi_n \rangle \phi_n \right\| \geq \left| \|f\| - \left\| \sum_{n=1}^{N} \langle f, \phi_n \rangle \phi_n \right\| \right| = \left| \|f\| - \sqrt{\sum_{n=1}^{N} |\langle f, \phi_n \rangle|^2} \right|.$$

Moreover,

$$\langle f, g \rangle = \sum_{n=1}^{+\infty} \langle f, \phi_n \rangle \langle g, \phi_n \rangle.$$

This follows from the equation above and from computing $\|f - g\|^2$.

The example of Hilbert space we will use from now on is the space of square-integrable functions in $[0, 1]$, $L^2[0, 1]$ with

$$\langle f, g \rangle = \int_0^1 f(x)g(x)dx \text{ and } \|f\|^2 = \langle f, f \rangle = \int_0^1 f^2(x)dx.$$

Furthermore, in this case, taking $f = 1_{[0,s]}$ and $g = 1_{[0,t]}$ gives us

$$s \wedge t = \sum_{n=1}^{+\infty} \int_0^s \phi_n(x)dx \int_0^t \phi_n(x)dx.$$

Let us now consider the process

$$X_t = \sum_{n=1}^{+\infty} Z_n \langle 1_{[0,t]}, \phi_n \rangle = \sum_{n=1}^{+\infty} Z_n \int_0^t \phi_n(x)dx,$$

for $t \in [0, 1]$, where $(Z_n)_{n\in\mathbb{N}}$ is a sequence of iid $N(0, 1)$ random variables.

Theorem 6.2.1. *Fix $t \in [0, 1]$. Then*

$$X_t = \sum_{n=1}^{+\infty} Z_n \langle 1_{[0,t]}, \phi_n \rangle$$

converges in L^2 and a.s. The resulting process is a Gaussian with mean zero and covariance function $s \wedge t$. Moreover, X has independent increments and $X_t - X_s \sim N(0, t - s)$.

Proof. The convergence is guaranteed by the L^2 Martingale Convergence Theorem. Indeed, for a fixed $t \in [0,1]$, define

$$M_n^{(t)} = \sum_{k=1}^{n} Z_k \langle 1_{[0,t]}, \phi_k \rangle.$$

This discrete-time process is clearly a martingale. Moreover, by Parseval identity,

$$\mathbb{E}[(M_n^{(t)})^2] \leq \sum_{k=1}^{+\infty} \langle 1_{[0,t]}, \phi_k \rangle^2 = \|1_{[0,t]}\|^2 = t.$$

From this, it is straightforward to see that $\mathbb{E}[X_t] = 0$ and $\mathrm{Cov}(X_s, X_t) = s \wedge t$. The fact that X is a Gaussian process with independent increments follows directly when considering linear combinations of X at particular times. $\square$

The previous theorem shows that X is almost a Brownian motion, we just need to verify that the sample paths of X are a.s. continuous. Since the convergence we concluded are just in L^2 and a.s., it does not give us continuity. This is indeed the most complex property of this process. Originally, Wiener proved this claim by using the particular basis $\phi_n(t) = \sqrt{2}\cos(\pi n t)$. It turns out that the calculations are cumbersome. We will then follow Lévy's argument that uses the wavelet basis.

More precisely, the wavelet basis is formed by the Haar functions:

$$\psi_n(t) = 2^{j/2}\psi(2^j t - k),$$

where $n = 2^j + k$, $j \in \mathbb{N}$, $k = 0, \dots, 2^n - 1$ and $\psi(t) = 1_{[0,1/2]}(t) - 1_{(1/2,1]}(t)$. We also set $\psi_0 \equiv 1$. One might show that the Haar functions are indeed an orthonormal basis of $L^2([0,1])$.

Theorem 6.2.2. *The series*

$$\sum_{n=0}^{+\infty} Z_n \Delta_n(t),$$

where

$$\Delta_n(t) = \langle 1_{[0,t]}, \psi_n \rangle = \int_0^t \psi_n(s)ds,$$

converges uniformly in $[0,1]$ *and the limit is a standard Brownian motion*

Remark 6.2.3. The graph of the functions $t \mapsto \Delta_n(t)$ are isosceles triangles of height $2^{-j/2-1}$ and base $[k/2^j, (k+1)/2^j]$. Moreover, $\Delta_0(t) = t$.

The following lemma will be needed for the proof of Theorem 6.2.2.

Lemma 6.2.4. *Let* $(Z_n)_{n\in\mathbb{N}}$ *be a sequence of iid* $N(0,1)$ *r.v's. There exists a r.v.* C *with* $\mathbb{P}(C < +\infty) = 1$ *such that*

$$|Z_n| \leq C\sqrt{\log n}, \text{ for } n \geq 2.$$

Proof. For $x \geq 1$, notice that

$$\mathbb{P}(|Z_n| \geq x) = \sqrt{\frac{2}{\pi}} \int_x^{+\infty} e^{-y^2/2} dy \leq \sqrt{\frac{2}{\pi}} \int_x^{+\infty} y e^{-y^2/2} dy = \sqrt{\frac{2}{\pi}} e^{-x^2/2}.$$

Hence

$$\mathbb{P}(|Z_n| \geq \sqrt{2\alpha \log n}) \leq \sqrt{\frac{2}{\pi}} e^{-\alpha \log n} = \sqrt{\frac{2}{\pi}} n^{-\alpha},$$

which implies, by Borel-Cantelli lemma[1],

$$\mathbb{P}\left(|Z_n| \geq \sqrt{2\alpha \log n} \text{ i.o}\right) = 0,$$

for any $\alpha > 1$. Then, the r.v.

$$C = \sup_{n\geq 2} \frac{|Z_n|}{\sqrt{\log n}}$$

is finite a.s. completing the proof. $\square$

[1] The Borel-Cantelli lemma tells us the following: if $\sum_{n=1}^{+\infty} \mathbb{P}(A_n) < +\infty$, then $\mathbb{P}(A_n \text{ i.o}) = 0$, where $\{A_n \text{ i.o}\} = \cap_{n=1}^{+\infty} \cup_{k=n}^{+\infty} A_k = \limsup_{n\to+\infty} A_n$.

Proof of Theorem 6.2.2. From Theorem 6.2.1, we have to just prove the uniform convergence of the series. Since the partial sum is continuous in $[0,1]$, the result will clearly follow. To prove uniform convergence, notice it is enough to show that

$$\sum_{n=N}^{+\infty} |Z_n|\Delta_n(t) \overset{N\to+\infty}{\longrightarrow} 0$$

where the convergence does not depend on the particular $t \in [0,1]$. By Lemma 6.2.4, for any N,

$$\sum_{n=N}^{+\infty} |Z_n|\Delta_n(t) \le C \sum_{n=N}^{+\infty} \sqrt{\log n}\,\Delta_n(t).$$

Now, for $n \in [2^j, 2^{j+1})$, we have $\log n < j+1$ (since $\log 2 < 1$). Let $J \in \mathbb{N}$ such that $N \ge 2^J$. This implies

$$\begin{aligned}\sum_{n=N}^{+\infty} \sqrt{\log n}\,\Delta_n(t) &= \sum_{j=J}^{+\infty} \sum_{k=0}^{2^j-1} \sqrt{\log(2^j+k)}\,\Delta_{2^j+k}(t)\\ &\le \sum_{j=J}^{+\infty} \sqrt{\log(2^j+k)}\,2^{-j/2-1} < \sum_{j=J}^{+\infty} \sqrt{j+1}\;2^{-j/2-1} \overset{J\to+\infty}{\longrightarrow} 0.\end{aligned}$$

We have used the fact that for each t, $\Delta_{2^j+k}(t)$ is different than zero for just one k. □

This construction works for the interval $[0,1]$. One way to extend to the interval $[0,+\infty)$, the natural domain of the Brownian motion, is to paste independent Brownian motions:

$$B_t = \sum_{k=1}^{n} B_1^{(k)} + B_{t-n}^{n+1},$$

for $t \in [n-1,n)$. One could straightforwardly prove that B is a Brownian motion.

6.3 Path Properties

Definition 6.3.1. We say that $\phi : [0,1] \longrightarrow \mathbb{R}$ is α-Hölder continuous, with $\alpha \in (0,1)$, if there exists $c > 0$ such that

$$|\phi(t) - \phi(s)| \le c|t-s|^\alpha$$

for any $t, s \in [0,1]$.

Theorem 6.3.2 (Hölder continuous $\alpha < 1/2$)**.** *The paths of a Brownian motion are, almost surely, α-Hölder continuous with $\alpha < 1/2$.*

Proof. Notice that

$$B_t - B_s = \sum_{j=0}^{+\infty} D_j(t,s),$$

where

$$D_j(t,s) = \sum_{n=2^j}^{2^{j+1}-1} Z_n(\Delta_n(t) - \Delta_n(s)).$$

First notice that, for $n \in [2^j, 2^{j+1})$, we have

$$|Z_n| \le C\sqrt{j+1}.$$

Let us now analyze $|\Delta_n(t) - \Delta_n(s)|$. Since $\Delta_n \in [0, 2^{-j/2-1}]$, we have

$$|\Delta_n(t) - \Delta_n(s)| \le 2^{-j/2}.$$

Moreover, the slope of the triangles of Δ_n is equal $2^{j/2}$. This implies that

$$|\Delta_n(t) - \Delta_n(s)| \le 2^{j/2}|t-s|.$$

Finally, notice that, for any give u, there is at most one n between 2^j and $2^{j+1}-1$ such that $\Delta_n(u) \neq 0$. Hence,

$$|D_j(t,s)| \leq \begin{cases} C\sqrt{j+1}\; 2^{-j/2}, \\ C\sqrt{j+1}\; 2^{j/2}|t-s|. \end{cases}$$

Note that, for any $\alpha < 1/2$, we have $\sqrt{j+1}\; 2^{-j/2+\alpha j} \to 0$. Then, there exists $\tilde{c} > 0$ such that $\sqrt{j+1}\; 2^{-j/2} \leq \tilde{c} 2^{-\alpha j}$, for all $j \geq 0$. Hence

$$|D_j(t,s)| \leq \begin{cases} \tilde{C}\; 2^{-\alpha j}, \\ \tilde{C}\; 2^{-\alpha j} 2^j |t-s|, \end{cases}$$

where $\tilde{C}$ is different than $\tilde{c}$.

Then, for any k, we find

$$\begin{aligned} |B_t - B_s| &\leq \sum_{j=0}^{+\infty} |D_j(t,s)| = \sum_{j=0}^{k} |D_j(t,s)| + \sum_{j=k+1}^{+\infty} |D_j(t,s)| \\ &\leq \tilde{C}\left(|t-s| \sum_{j=0}^{k} 2^{-\alpha j} 2^j + \sum_{j=k+1}^{+\infty} 2^{-\alpha j} \right) \\ &= \tilde{C}\left(|t-s| \frac{2^{(1-\alpha)(k+1)} - 1}{2^{1-\alpha} - 1} + \frac{2^{-\alpha(k+1)}}{1 - 2^{-\alpha}} \right) \end{aligned}$$

Now, for any fixed t, s, there exists $k \in \mathbb{N}$ such that $2^{-k-1} \leq |t-s| \leq 2^{-k}$. Then,

$$\begin{aligned} |t-s| \frac{2^{(1-\alpha)(k+1)} - 1}{2^{1-\alpha} - 1} + \frac{2^{-\alpha(k+1)}}{1 - 2^{-\alpha}} &\leq 2^{-k} \frac{2^{(1-\alpha)(k+1)}}{2^{1-\alpha} - 1} + \frac{2^{-\alpha(k+1)}}{1 - 2^{-\alpha}} \\ &= \frac{2}{2^{1-\alpha} - 1} 2^{-\alpha(k+1)} + \frac{1}{1 - 2^{-\alpha}} 2^{-\alpha(k+1)} \leq c(\alpha)|t-s|^{\alpha}. \end{aligned}$$

Putting all together, we conclude the theorem. □

Theorem 6.3.3 (Not Hölder continuous $\alpha > 1/2$)**.** *The paths of a Brownian motion are, almost surely, not α-Hölder continuous with $\alpha > 1/2$.*

Proof. Fix $s \in [0,1]$ and define

$$G_s(\alpha, c, \varepsilon) = \{\omega \in \Omega \; ; \; |B_s(\omega) - B_t(\omega)| \leq c|s-t|^{\alpha}, \text{ for all } |t-s| < \varepsilon\}.$$

We want to prove that $\mathbb{P}(G_s(\alpha, c, \varepsilon)) = 0$. Consider $n \geq m$ and define

$$X_{n,k} = \max\{|B_{j/n} - B_{(j+1)/n}| \; ; \; k \leq j < k+m\},$$

for $0 \leq k \leq n-m$. Let n be large enough so that $m/n < \varepsilon$. Now, let k such that $s \in [k/n, (k+m)/n]$. Then, for any j, we have $\max\{|s - j/n|, |(j+1)/n - s|\} \leq m/n < \varepsilon$. Hence, for $\omega \in G_s(\alpha, c, \varepsilon)$, we find

$$|B_{j/n} - B_{(j+1)/n}| \leq 2c(m/n)^{\alpha} \Rightarrow X_{n,k} \leq 2c(m/n)^{\alpha}.$$

Therefore, for any m and for n such that $m/n < \varepsilon$, we find

$$\omega \in G_s(\alpha, c, \varepsilon) \Rightarrow \min_{k=0,\dots,n-m} X_{n,k}(\omega) \leq 2c(m/n)^{\alpha}.$$

Moreover

$$\begin{aligned} \mathbb{P}\left(\min_{k=0,\dots,n-m} X_{n,k} \leq 2c(m/n)^{\alpha} \right) &\leq n\mathbb{P}(X_{n,1} \leq 2c(m/n)^{\alpha}) \leq n\mathbb{P}(|B_{1/n}| \leq 2c(m/n)^{\alpha})^m \\ &= n\mathbb{P}(|Z| \leq 2cm^{\alpha} n^{1/2-\alpha})^m \leq n\left(\frac{4cm^{\alpha} n^{1/2-\alpha}}{\sqrt{2\pi}} \right)^m \\ &= \left(\frac{4cm^{\alpha}}{\sqrt{2\pi}} \right)^m n^{1+m(1/2-\alpha)}. \end{aligned}$$

Hence, we just have to choose m such that $1 + m(1/2 - \alpha) < 0$. Doing so implies that the right-hand side of the inequality above goes to 0. Thus, $\mathbb{P}(G_s(\alpha, c, \varepsilon)) = 0$. □

Corollary 6.3.4. *Almost surely, B is not differentiable in* $[0,1]$.

Proof. The proof follows easily from

$$D = \{\omega \in \Omega \ ; \ \text{ there exists } s \in [0,1] \text{ such that } B \text{ is differentiable at } s\} \subset \bigcup_{j=1}^{+\infty} \bigcup_{k=1}^{+\infty} G_s(1, j, 1/k).$$

Then $\mathbb{P}(D) = 0$. ☐

Some other properties of the paths of the Brownian motion are stated below.

Theorem 6.3.5. *Let $(B_t)_{t\in[0,+\infty)}$ be a Brownian motion. Then.*

1. *(Law of Large Numbers)* $\lim_{t\to+\infty} \frac{B_t}{t} = 0$ *a.s;*
2. *(monotonicity) almost surely, $t \mapsto B_t$ is not monotone in any interval;*
3. *(local maximum) the set of local maximum for $t \mapsto B_t(\omega)$ is countable and dense. All local maxima are strict;*
4. *(Law of Iterated Logarithm)* $\limsup_{t\to+\infty} \frac{B_t(\omega)}{\sqrt{2t \log\log t}} = 1$*;*
5. *(zeros) $\mathcal{Z}_\omega = \{t \ ; \ B_t(\omega) = 0\}$ has Lebesgue measure zero, is closed, unbounded and has no isolated points;*

Proof. 1. If X is the inverted process of Proposition 6.1.8, we have

$$\lim_{t\to+\infty} \frac{B_t}{t} = \lim_{t\to+\infty} X_{1/t} = X_0 = 0 \text{ a.s.}$$

2. Let $[a,b]$ be an interval where the Brownian motion is monotonic (increasing, for instance). Then, for any discretization of $[a,b]$, $a = a_0 < a_1 < \cdots < a_n = b$, we must have $B_{a_i} - B_{a_{i-1}} \geq 0$. The probability of this happening is, by independence of increments,

$$\mathbb{P}(B_{a_i} - B_{a_{i-1}} \geq 0, \text{ for all } i) = \frac{1}{2^n} \overset{n\to+\infty}{\longrightarrow} 0.$$

Hence, for any given interval $[a,b]$,

$$\mathbb{P}(B \text{ is monotonic in}[a,b]) = 0.$$

Now,

$$\mathbb{P}(B \text{ is monotonic in some interval}) = \mathbb{P}\left(\bigcup_{a<b\in\mathbb{Q}} B \text{ is monotonic in } [a,b] \right) = 0.$$

☐

Chapter 7

Continuous-Time Martingales

7.1 Definition

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration $(\mathcal{F}_t)_{t\geq 0}$ in this space. Remember that a filtration is a sequence of σ-algebras such that $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$, for $s \leq t$.

Definition 7.1.1 (Martingale)**.** We say $(X_t)_{t\geq 0}$ is a martingale (with respect to $(\mathcal{F}_t)_{t\geq 0}$ and $\mathbb{P}$) if:

(i) Adaptedness: X_t is $\mathcal{F}_t$-measurable, i.e. $(X_t)_{t\geq 0}$ is $(\mathcal{F}_t)_{t\geq 0}$-adapted;

(ii) Integrability: $\mathbb{E}[|X_t|] < +\infty$, for all $t \geq 0$;

(iii) Martingale property: $\mathbb{E}\left[X_t \mid \mathcal{F}_s\right] = X_s$, for any $0 \leq s \leq t$.

Moreover, we will say $(X_t)_{t\geq 0}$ is continuous if there exists $\Omega_0 \in \mathcal{F}$ with $\mathbb{P}(\Omega_0) = 1$ and such that, for any $\omega \in \Omega_0$, the function $t \mapsto X_t(\omega)$ is continuous.

Definition 7.1.2 (Usual Hypotheses)**.** A filtration $(\mathcal{F}_t)_{t\geq 0}$ is said to satisfy the *usual hypotheses* if it is:

(i) complete: all the $\mathbb{P}$-null sets are in $\mathcal{F}_0$;

(ii) and right-continuous: $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$.

Definition 7.1.3 (Filtration Generated by a Process)**.** Let N be the collection of subsets of Ω of $\mathbb{P}$-measure zero and consider the filtration $\sigma(X_s \ ; \ s \leq t)$ generated by the random variables X_s, for $s \leq t$. We then define $\mathcal{F}_t^X$ as the smallest σ-algebra that contains N and $\sigma(X_s \ ; \ s \leq t)$. We called this as the filtration generated by $(X_t)_{t\geq 0}$. For some class of processes, it is possible to prove that this filtration is right-continuous, satisfying then the usual hypotheses.

7.2 Stopping Times

Definition 7.2.1 (Stopping Time)**.** Fix a filtration $(\mathcal{F}_t)_{t\geq 0}$. A random variable $\tau : \Omega \longrightarrow [0, +\infty]$ is called a *stopping time* with respect to $(\mathcal{F}_t)_{t\geq 0}$ if

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

If the filtration is right-continuous, τ will be a stopping time if and only if

$$\{\tau < t\} \in \mathcal{F}_t, \ \forall \ t \geq 0.$$

Remark 7.2.2. We further define X_τ as

$$X_\tau(\omega) = X_t(\omega), \quad \text{if } \tau(\omega) = t,$$

when $\tau(\omega) < +\infty$.

Proposition 7.2.3. *Let $(X_t)_{t\geq 0}$ be a $\mathcal{F}$-adapted, continuous process. Define*

$$\tau_\xi = \inf\{t \geq 0 \ ; \ X_t = \xi\}.$$

Then, τ_ξ is a stopping time.

Proof. By the continuity of X, one can show that

$$\{\tau_\xi \leq t\} = \bigcap_{n=1}^{+\infty} \bigcup_{r\in\mathbb{Q}\cap[0,t]} \left\{X_r \in \left(\xi - \frac{1}{n}, \xi + \frac{1}{n}\right)\right\}.$$

Since X is adapted to $\mathcal{F}$, we can easily see that $\{\tau_\xi \leq t\} \in \mathcal{F}_t$, as desired. □

Theorem 7.2.4. *Let $(M_t)_{t\geq 0}$ be a continuous martingale with respect to $(\mathcal{F}_t)_{t\geq 0}$ that satisfies the usual hypotheses. If τ is a stopping time for $(\mathcal{F}_t)_{t\geq 0}$, then the process $X_t = M_{t\wedge\tau}$ is also a continuous martingale with respect to $(\mathcal{F}_t)_{t\geq 0}$.*

Proof. The idea is to use what we have proved in the discrete-time case. We need to prove that $\mathbb{E}[|X_t|] < +\infty$ and $\mathbb{E}[X_t \mid \mathcal{F}_s] = X_s$, $s \leq t$. For this, fix $s \leq t$ and $n \in \mathbb{N}$ and define $S(n) = \{s + (t-s)k/2^n \ ; \ k = 0, \dots, 2^n\}$ and

$$\tau_n = \min\{r \in S(n) \ ; \ r \geq \tau\}.$$

We can easily see that τ_n is a stopping time and τ_n converges to τ. Consider now the discrete-time martingale $(M_r)_{r\in S(n)}$. Hence $\mathbb{E}[|M_r|] \leq \mathbb{E}[|M_v|]$, for $v \leq r$ in $S(n)$. Since t and τ_n belong to $S(n)$ and $t \wedge \tau_n \leq t$, we find $\mathbb{E}[|M_{t\wedge\tau_n}|] \leq \mathbb{E}[|M_t|]$ (by the discrete-time stopping theorem). Moreover, by Fatou's lemma and the continuity of M,

$$\mathbb{E}[|M_{t\wedge\tau}|] \leq \liminf_{n\to+\infty} \mathbb{E}[|M_{t\wedge\tau_n}|] \leq \mathbb{E}[|M_t|] < +\infty.$$

To prove the martingale property, notice that

$$\mathbb{E}[M_{t\wedge\tau_n} \mid \mathcal{F}_s] = M_{s\wedge\tau_n}.$$

We have seen that $M_{t\wedge\tau_n} \xrightarrow{a.s.} M_{t\wedge\tau}$ (the same for s). In order to put the limit inside the conditional expectation sign, we will show that this sequence is uniform integrable. However, this follows directly from the same argument we provided above using the convex function φ from Lemma 2.5.16 and *(iii)* from Example 2.5.14. □

Remark 7.2.5. As an application of the previous result, let us consider a Brownian motion $(B_t)_{t\geq 0}$ and

$$\tau = \min\{t \geq= 0 \ ; \ B_t = -C \text{ or } B_t = D\},$$

for $C, D > 0$. Firstly, notice the following:

$$\mathbb{P}(\tau > n+1) \leq \mathbb{P}(|B_1| \leq C + D, |B_2 - B_1| \leq C + D, \dots, |B_{n+1} - B_n| \leq C + D) = (1-a)^n,$$

where $a = \mathbb{P}(|B_{n+1} - B_n| > C + D)$. This implies that τ is finite a.s. and has moments of all orders. Moreover, $(B_{t\wedge\tau})_{t\geq 0}$ is a martingale by the theorem above and it is bounded by $C + D$. Hence, by the DCT, we have

$$\mathbb{E}[B_\tau] = \lim_{t\to+\infty} \mathbb{E}[B_{t\wedge\tau}] = \lim_{t\to+\infty} \mathbb{E}[B_0] = 0.$$

On the other hand, we have

$$0 = \mathbb{E}[B_\tau] = D\mathbb{P}(B_\tau = D) - C(1 - \mathbb{P}(B_\tau = D)) \Rightarrow \mathbb{P}(B_\tau = D) = \frac{C}{C+D}.$$

Moreover, using the martingale $X_t = B_t^2 - t$ (and applying similar arguments as above), we find that

$$\mathbb{E}[\tau] = \mathbb{E}[B_\tau^2] = CD.$$

7.3 Further Results

Theorem 7.3.1 (Doob's Maximal Inequality). *Let $(M_t)_{t\geq 0}$ be a continuous, non-negative submartingale with respect to $(\mathcal{F}_t)_{t\geq 0}$. Then*

$$\lambda^p \mathbb{P}\left(\sup_{t\in[0,T]} M_t > \lambda\right) \leq \mathbb{E}[M_T^p],$$

for $\lambda > 0$ and $p \geq 1$. If $M_T \in L^p(\mathbb{P})$, for some $p > 1$,

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

Proof. Fix $n \in \mathbb{N}$ and define $S(n,T) = \{iT/2^n\ ;\ i \in \{0,\dots,2^n\}\}$. By the continuity of M,

$$\lim_{n\to+\infty} \sup_{t\in S(n,T)} M_t = \sup_{t\in[0,T]} M_t$$

and the convergence is monotonic increasingly. By Doob's Maximal Inequality in discrete time,

$$\lambda^p \mathbb{P}\left(\sup_{t\in S(n,T)} M_t > \lambda\right) \leq \mathbb{E}[M_T^p].$$

Then, by the continuity of probability,

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

Moreover, by Fatou's lemma and the result in discrete case,

$$\left\|\sup_{t\in[0,T]} M_t\right\|_p \leq \liminf_{n\to+\infty} \left\|\sup_{t\in S(n,T)} M_t\right\|_p \leq \frac{p}{p-1}\|M_T\|_p.$$

□

Theorem 7.3.2 (Martingale Convergence Theorem). *Let $(M_t)_{t\geq 0}$ be a continuous martingale such that $\|M_t\|_p \leq C < +\infty$, for all $t \geq 0$, for some $C > 0$ and $p \geq 1$. Then, there exists a r.v. M_∞ with $\|M_\infty\|_p \leq C$ and such that*

$$\lim_{t\to+\infty} M_t = M_\infty \ a.s.\ ,$$

and, if $p > 1$, $\|M_t - M_\infty\|_p \to 0$.

Proof. We will first consider the case when $p > 1$. Notice that if we restrict ourselves to the case $t = n \in \mathbb{N}$, by the Martingale Convergence Theorem in discrete time, there exists $M_\infty \in L^p$ such that $M_n \to M_\infty$ a.s. Moreover, for any $n \in \mathbb{N}$, we have, for $t \geq n$,

$$|M_t - M_\infty| \leq |M_n - M_\infty| + \sup_{s\geq n}|M_s - M_n|,$$

which implies

$$\limsup_{t\to+\infty} |M_t - M_\infty| \leq \lim_{n\to+\infty} \sup_{s\geq n}|M_s - M_n|. \tag{7.1}$$

Furthermore, $(M_s - M_n)_{s\geq n}$ is a continuous martingale, which gives us, by Doob's Maximal Inequality, for any $\lambda > 0$,

$$\mathbb{P}\left(\sup_{s\in[n,N]} |M_s - M_n| > \lambda\right) \leq \frac{1}{\lambda^p}\mathbb{E}[|M_N - M_n|^p].$$

Since $M_N \xrightarrow{L^p} M_\infty$, we find

$$\mathbb{P}\left(\sup_{s\in[n,+\infty)} |M_s - M_n| > \lambda\right) \leq \frac{1}{\lambda^p}\mathbb{E}[|M_\infty - M_n|^p].$$

Letting $n \to +\infty$, we conclude that, for all $\lambda > 0$,

$$\mathbb{P}\left(\lim_{n\to+\infty} \sup_{s\in[n,+\infty)} |M_s - M_\infty| > \lambda\right) = 0.$$

Hence, by Equation (7.1),

$$\mathbb{P}\left(\lim_{t\to+\infty} |M_t - M_\infty| = 0\right) = 1.$$

This is the almost sure convergence. To prove convergence in L^p, notice

$$\|M_t - M_\infty\|_p \leq \|M_t - M_n\|_p + \|M_n - M_\infty\|_p.$$

Since $X_t = |M_t - M_n|$ is a submartingale, for any $m > t$, we have

$$\|M_t - M_n\|_p \leq \|M_m - M_n\|_p \Rightarrow \limsup_{t\to+\infty} \|M_t - M_\infty\|_p \leq \|M_t - M_n\|_p + \sup_{m\geq n} \|M_m - M_n\|_p.$$

The results then follows from the Martingale Convergence Theorem in discrete time that gives us the fact that $M_n \xrightarrow{L^p} M_\infty$ (and hence both terms on the right go to zero).

Let us now consider the case $p = 1$. For this, consider the stopping time $\tau_n = \inf\{t > 0 \ ; \ |M_t| \geq n\}$. By Theorem 7.2.4, $M_{t\wedge\tau_n}$ is a continuous martingale and by the definition of τ_n, this process is bounded (by n). Hence, by the previous part of this proof, $M_{t\wedge\tau_n}$ converges a.s. when $t \to +\infty$. Moreover, for any $\omega \in \{\tau_n = +\infty\}$ (but a set of probability zero), $M_t = M_{t\wedge\tau_n}$ and then M_t converges a.s. when $t \to +\infty$. We will now verify that

$$\mathbb{P}\left(\bigcup_{n=1}^{+\infty} \{\tau_n = +\infty\}\right) = 1. \tag{7.2}$$

Before doing so, notice that this is enough to conclude the proof of the case $p = 1$. Additionally, Fatou's lemma guarantees that $\mathbb{E}[|M_\infty|] \leq C$. Now, by Doob's maximal inequality, we have, for any $s > 0$:

$$\mathbb{P}\left(\sup_{0\leq t\leq s} |M_t| > \lambda\right) \leq \frac{1}{\lambda}\mathbb{E}[|M_t|] \leq \frac{C}{\lambda} \Rightarrow \mathbb{P}\left(\sup_{0\leq t<+\infty} |M_t| > \lambda\right) \leq \frac{C}{\lambda}.$$

Thus,

$$\mathbb{P}(\tau_n = +\infty) = \mathbb{P}\left(\sup_{0\leq t<+\infty} |M_t| \leq n\right) \geq 1 - \frac{C}{n}.$$

Since $\tau_n \leq \tau_{n+1}$, we have concluded (7.2). □

7.4 Local Martingales

Localization is very useful technique in Stochastic Calculus and it will be useful later.

Definition 7.4.1 (Local Martingale)**.** A stochastic process $(M_t)_{t\geq0}$, adapted to the filtration $(\mathcal{F}_t)_{t\geq0}$, is a *local martingale* if there exists a non-decreasing sequence of stopping times $(\tau_n)_{n\in\mathbb{N}}$ satisfying

(i) $\tau_n \to +\infty$ a.s, when $n \to +\infty$;

(ii) for each n, $M_t^{(n)} = M_{t\wedge\tau_n}$ is a (true) martingale with respect to $(\mathcal{F}_t)_{t\geq0}$.

Let us state some results regarding local martingales to better understand them:

Proposition 7.4.2. *If $(M_t)_{t\geq0}$ is a local martingale and τ is a stopping time, then $(M_{t\wedge\tau})_{t\geq0}$ is also a local martingale.*

Proof. Let $(\tau_n)_{n\in\mathbb{N}}$ be a localizing sequence for M. Define $Y_t = M_{t\wedge\tau}$ and notice $Y_{t\wedge\tau_n} = M_{(t\wedge\tau_n)\wedge\tau} = M^{(n)}_{t\wedge\tau}$. Since τ is a stopping time and $M^{(n)}$ is a martingale, then $(Y_{t\wedge\tau_n})_{t\geq0}$ is also a martingale, concluding the proof. □

Proposition 7.4.3. *If $(M_t)_{t\geq 0}$ is a local martingale and there exists an integrable r.v. Y such that $|M_t| \leq Y$, for any $t \geq 0$, then M is a true martingale.*

Proof. Let $(\tau_n)_{n\in\mathbb{N}}$ be a localizing sequence for M. Hence, $M_{t\wedge\tau_n} \xrightarrow{a.s.} X_t$, as $n \to +\infty$, and by assumption, $|M_{t\wedge\tau_n}| \leq Y$. By the Dominated Convergence Theorem:

$$M_s = \lim_{n\to+\infty} M_{s\wedge\tau_n} = \lim_{n\to+\infty} \mathbb{E}\left[M_{t\wedge\tau_n} \mid \mathcal{F}_s\right] = \mathbb{E}\left[M_t \mid \mathcal{F}_s\right].$$

Therefore, M is a martingale. □

Proposition 7.4.4. *If $(M_t)_{t\in[0,T]}$ is a positive local martingale, then $(M_t)_{t\in[0,T]}$ is a supermartingale. Now, if $M_0 \in L^1$ and*

$$\mathbb{E}[M_T] = \mathbb{E}[M_0],$$

then $(M_t)_{t\in[0,T]}$ is a true martingale

Proof. Let us repeat the argument on the proof above. By Fatou's lemma (since M is positive):

$$M_s = \lim_{n\to+\infty} M_{s\wedge\tau_n} = \lim_{n\to+\infty} \mathbb{E}\left[M_{t\wedge\tau_n} \mid \mathcal{F}_s\right] \geq \mathbb{E}\left[M_t \mid \mathcal{F}_s\right].$$

Then, M is a supermartingale. This implies $\mathbb{E}[M_s] \geq \mathbb{E}[M_t]$, which gives

$$\mathbb{E}[M_0] \geq \mathbb{E}[M_s] \geq \mathbb{E}[M_t] \geq \mathbb{E}[M_T].$$

Assuming $\mathbb{E}[M_T] = \mathbb{E}[M_0]$, we conclude that $\mathbb{E}[M_s]$ is constant. This implies (together with positivity) that M is a martingale. Indeed, if $M_s > \mathbb{E}[M_t \mid \mathcal{F}_s]$ on some $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$, then taking expectation on both sides and using the fact M is positive, we would conclude $\mathbb{E}[M_s] > \mathbb{E}[M_t]$, which is a contradiction to $\mathbb{E}[M_s]$ being constant. □

Part III

Stochastic Calculus

Chapter 8

Itô Integration

8.1 Quadratic Variation of the Brownian Motion

The usual Riemann-Stieltjes-Lebesgue integration,

$$\int_0^t \phi(s)dg(s),$$

assumes that g is of finite variation:

$$V_1(g,[0,t]) = \sup_\pi \sum_{i=0}^{n-1} |g(t_{i+1}) - g(t_i)| < +\infty,$$

where $\pi = \{0 = t_0 < t_1 < \cdots < t_n = t\}$ is any partition of $[0,t]$.

The problem with integration with respect to a Brownian motion $(B_t)_{t\geq 0}$ is not the fact that B is a stochastic process (i.e. a random function), but that B has unbounded variation. To prove this, let us compute the quadratic variation of B:

$$V_2(B,[0,t]) = \lim_{\|\pi\|\to 0} \sum_{i=0}^{n-1} |B_{t_{i+1}} - B_{t_i}|^2,$$

where $\|\pi\| = \max_i |t_{i+1} - t_i|$ and the convergence is in the sense of in probability.

Theorem 8.1.1.

$$V_2(B,[0,t]) = t.$$

Proof. Define $X_n = \sum_{i=0}^{n-1} |B_{t_{i+1}} - B_{t_i}|^2$. We will show that $X_n \to t$ in L^2. In particular, this implies the convergence in probability. Notice that

$$\mathbb{E}[X_n] = \sum_{i=0}^{n-1} \mathbb{E}[|B_{t_{i+1}} - B_{t_i}|^2] = \sum_{i=0}^{n-1} t_{i+1} - t_i = t.$$

Hence, if we prove that $\mathrm{Var}(X_n) \to 0$, we are done. Notice the increments $B_{t_{i+1}} - B_{t_i}$ are independent and normally distributed with mean 0 and variance $t_{i+1} - t_i$. Then

$$\mathrm{Var}((B_{t_{i+1}} - B_{t_i})^2) = 2\mathrm{Var}(B_{t_{i+1}} - B_{t_i})^2 = 2(t_{i+1} - t_i)^2.$$

We have used the fact that $\mathrm{Var}(Z^2) = 2\mathrm{Var}(Z)^2$, if $Z \sim N(0,\sigma^2)$. Therefore,

$$\mathrm{Var}(X_n) = \sum_{i=0}^{n-1} \mathrm{Var}((B_{t_{i+1}} - B_{t_i})^2) = 2\sum_{i=0}^{n-1} (t_{i+1} - t_i)^2 \leq 2\|\pi\| t.$$

Since $\|\pi\| \to 0$, we have our result. $\square$

Remark 8.1.2. One can extend the result above to deal with convergence in L^p, for any $p \geq 2$. However, to get the limit pathwise (i.e. in the almost sure sense), one needs to further restrict the choice of partitions: $\|\pi\| = o(1/(\log n)^{1/2})$. In particular, we may take the dyadic partition, $t_i = it/2^n$. Furthermore, knowing the realization of the Brownian path, one can build a partition (not satisfying the prior condition) that gives infinite quadratic variation. In particular, we cannot naively extend the definition of first variation to quadratic variation, i.e. taking the *sup*.

Remark 8.1.3. Some simple properties of the quadratic variation:

$$V_2(X, [a, b]) = V_2(X, [a, c]) + V_2(X, [c, b]),$$

$$V_2(\alpha + \beta X, [a, b]) = \beta^2 V_2(X, [a, b]).$$

Proposition 8.1.4. *If $(X_t)_{t\geq 0}$ is continuous and of finite variation, then the quadratic variation is 0.*

Proof. By direct computation

$$\begin{aligned}\sum_{i=0}^{n-1} |X_{t_{i+1}} - X_{t_i}|^2 &\leq \sup_i |X_{t_{i+1}} - X_{t_i}| \left(\sum_{i=0}^{n-1} |X_{t_{i+1}} - X_{t_i}|\right) \\ &\leq V_1(X, [0, t]) \sup_i |X_{t_{i+1}} - X_{t_i}|.\end{aligned}$$

The continuity of paths implies that the supremum above converges to 0 and the proof is complete. □

Corollary 8.1.5.

$$V_1(B, [0, t]) = +\infty \text{ a.s, for any } t > 0.$$

Proof. This follows directly by the proposition above and from the fact that the quadratic variation of B is finite and strictly positive. □

8.2 The Itô Integral

$$\int_0^t \phi(s) dg(s)$$

As we have said, the usual integration theory can be generalized "at most" to functions of bounded variation. However, there is a natural balance between the assumptions on f and g. It is well-known that in the usual framework, the integral above can be approximated with Riemann sums where ϕ can be evaluated at any point in the partition interval $[t_i, t_{i+1}]$:

$$\int_0^t \phi(s) dg(s) \approx \sum_{i=0}^{n-1} \phi(t_i^*)(g(t_{i+1}) - g(t_i)),$$

for $t_i^* \in [t_i, t_{i+1}]$. Therefore, the idea is to play with this balance in order to generalized the integral to Brownian motion paths. Two successful approaches are the Young Integral and the Rough Path Theory. Here we will pursue the most important integration theory for the Brownian motion due to Itô . It is essentially an L^2 theory, but the main restriction is that the integrand is evaluated only at t_i.

This is in sync with the financial applications. Indeed, the Riemann sums, in Finance, should be understand as the value of an strategy that trades the quantity ϕ of an asset with price g. Then, since the trade should be performed with the information available at the beginning of the period and the earnings will depend on the return of the asset in this period. In symbols,

$$\sum_{i=0}^{n-1} \phi(t_i)(g(t_{i+1}) - g(t_i)).$$

Other notions of stochastic integral are defined by considering a different point, like t_{i+1} (Backward Itô Integral) and $\frac{t_i+t_{i+1}}{2}$ (Stratonovich Integral).

We will now construct the Itô integral:

$$\int_0^T \phi(\omega, t)dB_t,$$

for a class of functions ϕ. In what follows, we will fix a final time $T > 0$, a filtration $(\mathcal{F}_t)_{t\geq 0}$ and a Brownian motion $(B_t)_{t\geq 0}$ with respect to this filtration.

8.2.1 Simple Functions

We say $\phi : \Omega \times [0,T] \longrightarrow \mathbb{R}$ is simple, and write $\phi \in \mathcal{H}_0^2[0,T]$, if

$$\phi(\omega, t) = \sum_{i=0}^{n-1} a_i(\omega) 1_{(t_i, t_{i+1}]}(t),$$

where $0 = t_0 < t_1 < \cdots < t_n = T$ and $a_i \in \mathcal{F}_{t_i}$ with $\mathbb{E}[a_i^2] < +\infty$. We would like the integral with respect to B to satisfy two fundamental properties: linearity and

$$\int_s^u 1 \ dB_t = B_u - B_s.$$

In this case, there is only one choice for the integral of $f \in \mathcal{H}_0^2[0,T]$:

$$\int_0^T \phi(\omega, t)dB_t = \sum_{i=0}^{n-1} a_i(\omega)(B_{t_{i+1}}(\omega) - B_{t_i}(\omega)).$$

In operators language, we have define the integral operator $I : \mathcal{H}_0^2[0,T] \longrightarrow L^2$.

We would like to extend this operator to a wider class of functions:

$$\mathcal{H}^2[0,T] = \left\{ \phi \in \mathcal{M} \ ; \ \mathbb{E}\left[\int_0^T \phi^2(\omega, t)dt\right] < +\infty \right\},$$

where $\mathcal{M}$ is the space of functions $\phi : \Omega \times [0,T] \longrightarrow \mathbb{R}$ such that ϕ is measurable with respect to $\mathcal{F} \otimes \mathcal{B}([0,T])$ and adapted with respect to $(\mathcal{F}_t)_{t\geq 0}$. Define also the norm in $\mathcal{H}^2$:

$$\|\phi\|_{\mathcal{H}^2}^2 = \mathbb{E}\left[\int_0^T \phi^2(\omega, t)dt\right] = \int_0^T \mathbb{E}\left[\phi^2(\omega, t)\right] dt.$$

To consider this extension, the idea is to consider the denseness of $\mathcal{H}_0^2$ in $\mathcal{H}^2$ and then define I in $\mathcal{H}^2$ as an approximation of I in $\mathcal{H}_0^2$. To do this we need to control the norm of $\|I(\phi)\|_{L^2}$ with the norm of $\|\phi\|_{\mathcal{H}^2}$. This is done in the next lemma. Just a reminder: $\|I(\phi)\|_{L^2} = \mathbb{E}[I(\phi)^2]^{1/2}$.

Lemma 8.2.1 (Itô Isometry in $\mathcal{H}_0^2$)**.**

$$\|I(\phi)\|_{L^2} = \|\phi\|_{\mathcal{H}^2}, \ \forall \ \phi \in \mathcal{H}_0^2.$$

Then, the mapping $I : \mathcal{H}_0^2[0,T] \longrightarrow L^2$ is continuous.

Proof. It follows from direct computation:

$$\phi^2(\omega, t) = \sum_{i=0}^{n-1} a_i^2(\omega) 1_{(t_i, t_{i+1}]}(t),$$

and then

$$\|\phi\|_{\mathcal{H}^2}^2 = \mathbb{E}\left[\int_0^T \phi^2(\omega, t)dt\right] = \sum_{i=0}^{n-1} \mathbb{E}[a_i^2](t_{i+1} - t_i).$$

Now,

$$I(\phi)^2 = \sum_{i,j=0}^{n-1} a_i a_j (B_{t_{i+1}} - B_{t_i})(B_{t_{j+1}} - B_{t_j})$$

and $\mathbb{E}[a_i a_j (B_{t_{i+1}} - B_{t_i})(B_{t_{j+1}} - B_{t_j})]$ can be easily computed using the fact that $B_{t_{j+1}} - B_{t_j} \sim N(0, t_{j+1} - t_j)$ and that, if $i < j$, for example, then a_i, a_j and $B_{t_{i+1}} - B_{t_i}$ are independent of $B_{t_{j+1}} - B_{t_j}$. Hence, by the independence of a_i and the increment $B_{t_{i+1}} - B_{t_i}$, we have

$$\begin{aligned}\mathbb{E}[I(\phi)^2] &= \sum_{i=0}^{n-1} \mathbb{E}[a_i^2 (B_{t_{i+1}} - B_{t_i})^2] = \sum_{i=0}^{n-1} \mathbb{E}[a_i^2]\mathbb{E}[(B_{t_{i+1}} - B_{t_i})^2] \\ &= \sum_{i=0}^{n-1} \mathbb{E}[a_i^2](t_{i+1} - t_i),\end{aligned}$$

proving the lemma. □

Lemma 8.2.2. *$\mathcal{H}_0^2$ is dense in $\mathcal{H}^2$, i.e. for any $\phi \in \mathcal{H}^2$, there exists $(\phi_n)_{n\in\mathbb{N}}$ in $\mathcal{H}_0^2$ such that $\|\phi - \phi_n\|_{\mathcal{H}^2} \to 0$, as $n \to +\infty$.*

Proof. It easily follows from the density of simple function in $L^2(\Omega)$. □

8.2.2 The Integral

Definition 8.2.3 (Itô Integral)**.** For any $\phi \in \mathcal{H}^2$, we define $I(\phi)$ as the limit of $I(\phi_n)$, where $(\phi_n)_{n\in\mathbb{N}}$ belongs to $\mathcal{H}_0^2$ and $\phi_n \to \phi$ in $\mathcal{H}^2$. We call $I(\phi)$ Itô Integral of f with respect to B and write

$$I(\phi) = \int_0^T \phi(\omega, t) dB_t.$$

Theorem 8.2.4. *The Itô Integral is well defined in $\mathcal{H}^2$.*

Proof. We will firstly prove that the limit of $I(\phi_n)$ exists. Since $(\phi_n)_{n\in\mathbb{N}}$ converges in $\mathcal{H}^2$, then this sequence is Cauchy in in $\mathcal{H}^2$. By Itô Isometry in $\mathcal{H}_0^2$, it is clear that $(I(\phi_n))_{n\in\mathbb{N}}$ is Cauchy in L^2. Since this space is complete, there exists a random variable in it that is the limit of $(I(\phi_n))_{n\in\mathbb{N}}$. This r.v. is $I(\phi)$.

We need to further verify that this definition of $I(\phi)$ does not depend on the sequence $(\phi_n)_{n\in\mathbb{N}}$. Thus, consider $(\phi_n')_{n\in\mathbb{N}}$ in $\mathcal{H}_0^2$ such that $\|\phi - \phi_n'\|_{\mathcal{H}^2} \to 0$ and write $\psi = I(\phi)$. Then,

$$\begin{aligned}\|\psi - I(\phi_n')\|_{L^2} &\le \|\psi - I(\phi_n)\|_{L^2} + \|I(\phi_n) - I(\phi_n')\|_{L^2} \\ &= \|\psi - I(\phi_n)\|_{L^2} + \|I(\phi_n - \phi_n')\|_{L^2} \\ &= \|\psi - I(\phi_n)\|_{L^2} + \|\phi_n - \phi_n'\|_{\mathcal{H}^2} \overset{n\to+\infty}{\longrightarrow} 0.\end{aligned}$$

Hence, the choice of the sequence do not matter in the definition of $I(\phi)$. □

The Itô integral operator I is then linear and continuous in $\mathcal{H}^2$. Moreover, Itô Isometry also holds in $\mathcal{H}^2$.

Theorem 8.2.5 (Itô Isometry)**.**

$$\|I(\phi)\|_{L^2} = \|\phi\|_{\mathcal{H}^2},\ \forall\ \phi \in \mathcal{H}^2.$$

Proof. Consider $(\phi_n)_{n\in\mathbb{N}}$ in $\mathcal{H}_0^2$ such that $\phi_n \to \phi$ in $\mathcal{H}^2$. Then, $\|\phi_n\|_{\mathcal{H}^2} \to \|\phi\|_{\mathcal{H}^2}$, when $n \to +\infty$. Since $I(\phi_n)$ converges to $I(\phi)$ in L^2, we have $\|I(\phi_n)\|_{L^2} \to \|I(\phi)\|_{L^2}$. The result then follows from Itô Isometry in $\mathcal{H}_0^2$. □

Example 8.2.6 (An Explicit Computation)**.** Let us compute

$$\int_0^T B_t dB_t.$$

If B were of finite variation, Riemann integration would show that

$$\int_0^T B_t dB_t = \frac{B_T^2}{2}.$$

But that is WRONG. The quadratic variation plays a very important role here. Define $\phi(\omega,t) = B_t(\omega)$, which clearly belongs $\mathcal{H}^2$. Let us consider $\phi_n \in \mathcal{H}^2_0$ such $\phi_n \to \phi$ in $\mathcal{H}^2$:

$$\phi_n(\omega,t) = \sum_{i=0}^{n-1} B_{t_i}(\omega) 1_{(t_i,t_{i+1}]}(t),$$

where $t_i = iT/n$. Let us verify that $\phi_n \to \phi$ as $n \to +\infty$. Notice

$$\begin{aligned}
(\phi(\omega,t) - \phi_n(\omega,t))^2 &= \left(B_t(\omega) - \sum_{i=0}^{n-1} B_{t_i}(\omega) 1_{(t_i,t_{i+1}]}(t) \right)^2 \\
&= \left(\sum_{i=0}^{n-1} (B_t(\omega) - B_{t_i}(\omega)) 1_{(t_i,t_{i+1}]}(t) \right)^2 \\
&= \sum_{i=0}^{n-1} (B_t(\omega) - B_{t_i}(\omega))^2 1_{(t_i,t_{i+1}]}(t)
\end{aligned}$$

Hence

$$\begin{aligned}
\|\phi - \phi_n\|_{\mathcal{H}^2} &= \mathbb{E}\left[\int_0^T (\phi(\omega,t) - \phi_n(\omega,t))^2 dt \right] \\
&= \mathbb{E}\left[\int_0^T \sum_{i=0}^{n-1} (B_t(\omega) - B_{t_i}(\omega))^2 1_{(t_i,t_{i+1}]}(t) dt \right] \\
&= \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \mathbb{E}\left[(B_t(\omega) - B_{t_i}(\omega))^2 \right] dt \\
&= \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} (t - t_i) dt = \sum_{i=0}^{n-1} \frac{1}{2}(t_{i+1} - t_i)^2 = \frac{T^2}{2n}
\end{aligned}$$

We now have to compute $I(\phi_n)$ and understand the limit of these r.v's. Notice that

$$I(\phi_n) = \sum_{i=0}^{n-1} B_{t_i}(B_{t_{i+1}} - B_{t_i})$$

and

$$B_{t_i}(B_{t_{i+1}} - B_{t_i}) = \frac{1}{2}(B^2_{t_{i+1}} - B^2_{t_i}) - \frac{1}{2}(B_{t_{i+1}} - B_{t_i})^2.$$

Hence,

$$I(\phi_n) = \frac{1}{2}B_T^2 - \frac{1}{2}\sum_{i=0}^{n-1}(B_{t_{i+1}} - B_{t_i})^2,$$

since $t_n = T$. Letting $n \to +\infty$ and noticing the second term in the right-hand side converges to quadratic variation of B in the interval $[0,T]$, which is equal to T, we conclude

$$\lim_{n\to+\infty} I(\phi_n) = \frac{1}{2}B_T^2 - \frac{T}{2}.$$

Therefore,

$$\int_0^T B_t dB_t = \frac{B_T^2}{2} - \frac{T}{2}.$$

8.2.3 Itô Integral as a Process

We have fixed the final time T and integrated $\phi \in \mathcal{H}^2$ against B from 0 to T. We would like now to integrate up to t and understand the time evolution of this stochastic process. One could think that we just need to consider the operator I we have defined above applied to the function $\phi\, 1_{[0,t]} \in \mathcal{H}^2$. However, our definition of I is as a element of $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Doing what we have just considered would create r.v's

$I(\phi\ 1_{[0,t]})$, for each $t \in [0,T]$, but they would be defined almost surely, i.e. for each $t \in [0,T]$, there exists $A_t \in \mathcal{F}$ with $\mathbb{P}(A_t) = 1$ and where $I(\phi\ 1_{[0,t]})$ is the Itô integral of ϕ from 0 to t. Since the set $[0,T]$ is uncountable, we cannot guarantee that the intersection of the A_t's, which is the set where all these r.v's are defined as the Itô integral, has probability one. Something else must be done.

Theorem 8.2.7. *For any $\phi \in \mathcal{H}^2[0,T]$, there exists a continuous martingale $(X_t)_{t\in[0,T]}$ with respect to the filtration generated by the Brownian motion and such that*

$$\mathbb{P}\left(X_t = I(\phi\ 1_{[0,t]})\right) = 1,\ \forall\ t \in [0,T].$$

Remark 8.2.8. Notice that, for any $\phi \in \mathcal{H}^2$, $\phi\ 1_{[0,t]} \in \mathcal{H}^2$.

Proof. Consider $(\phi_n)_{n\in\mathbb{N}}$ in $\mathcal{H}_0^2$ such that $\phi_n \to \phi$ in $\mathcal{H}^2$. It is easy to see that $1_{[0,t]}\ \phi_n \to 1_{[0,t]}\ \phi$ in $\mathcal{H}^2$ as well and hence, $I(\phi_n 1_{[0,t]}) \to I(\phi 1_{[0,t]})$ in L^2. We will then study the properties of $X_t^{(n)} = I(\phi_n 1_{[0,t]})$. Let us write

$$\phi_n(\omega,s) = \sum_{i=0}^{n-1} a_i^{(n)}(\omega) 1_{(t_i,t_{i+1}]}(s),$$

and notice, if k is such that $t_k < t \le t_{k+1}$,

$$\begin{aligned}\phi_n(\omega,s)1_{[0,t]}(s) &= \sum_{i=0}^{n-1} a_i^{(n)}(\omega)1_{(t_i,t_{i+1}]\cap[0,t]}(s)\\ &= a_k^{(n)}(\omega)1_{(t_k,t]}(s) + \sum_{i=0}^{k-1} a_i^{(n)}(\omega)1_{(t_i,t_{i+1}]}(s).\end{aligned}$$

Then,

$$X_t^{(n)} = I(\phi_n 1_{[0,t]}) = a_k^{(n)}(B_t - B_{t_k}) + \sum_{i=0}^{k-1} a_i^{(n)}(B_{t_{i+1}} - B_{t_i})$$

It is simple to prove that $(X_t^{(n)})_{t\in[0,T]}$ is a continuous martingale adapted to the Brownian filtration. Let us now prove there exists a subsequence $(n_k)_{k\in\mathbb{N}}$ such that $X_t^{(n_k)} \to X_t$ uniformly in $[0,T]$ a.s. For this, fix $n \ge m$, and apply Doob's maximal inequality to the continuous submartingale $|X_t^{(n)} - X_t^{(m)}|$ (with $p = 2$):

$$\begin{aligned}\mathbb{P}\left(\sup_{0\le t\le T}|X_t^{(n)} - X_t^{(m)}| \ge \varepsilon\right) &\le \frac{1}{\varepsilon^2}\mathbb{E}\left[|X_T^{(n)} - X_T^{(m)}|^2\right] = \frac{1}{\varepsilon^2}\mathbb{E}[|I(\phi_n) - I(\phi_m)|^2]\\ &= \frac{1}{\varepsilon^2}\mathbb{E}[I(\phi_n - \phi_m)^2] = \frac{1}{\varepsilon^2}\|\phi_n - \phi_m\|^2_{\mathcal{H}^2}.\end{aligned}$$

Since $\phi_n \to \phi$ in $\mathcal{H}^2$, we may find a subsequence $(n_k)_{k\in\mathbb{N}}$ such that

$$\|\phi_{n_{k+1}} - \phi_{n_k}\|_{\mathcal{H}^2} \le 2^{-3k/2}.$$

Taking $\varepsilon = 2^{-k}$, $n = n_k$ and $m = n_{k+1}$, we find

$$\mathbb{P}\left(\sup_{0\le t\le T}|X_t^{(n_{k+1})} - X_t^{(n_k)}| \ge 2^{-k}\right) \le 2^{-k}.$$

Consider then

$$A_k = \left\{\sup_{0\le t\le T}|X_t^{(n_{k+1})} - X_t^{(n_k)}| \ge 2^{-k}\right\}.$$

By the Borel-Cantelli lemma,

$$\mathbb{P}(A_k \text{ i.o. }) = 0,$$

which implies that there exists Ω_0 with probability 1 and a r.v. N, finite in Ω_0, and such that, for $\omega \in \Omega_0$,

$$\sup_{0\le t\le T}|X_t^{(n_{k+1})}(\omega) - X_t^{(n_k)}(\omega)| < 2^{-k},\ \forall\ k \ge N(\omega).$$

Since the sequence 2^{-k} is summable, the sequence $(X^{(n_k)}(\omega))_{k\in\mathbb{N}}$ is Cauchy in the supremum norm of $C([0,T])$. Since $C[0,T]$ is complete under this norm, there exists $X(\omega) \in C[0,T]$ such that $X^{(n_k)}(\omega) \to$

$X(\omega)$ under the sup norm. This implies that $X(\omega)$ is continuous in t. Additionally, by Itô Isometry, the sequence $(X_t^{(n_k)})_{k\in\mathbb{N}}$ is Cauchy in L^2, and hence, $(X_t^{(n_k)})_{k\in\mathbb{N}}$ converges to X_t in L^2 as well. This clearly implies that X is a martingale. This proves almost all the claims of this theorem. For the last one, notice that $I(\phi_n \; 1_{[0,t]}) \to I(\phi \; 1_{[0,t]})$ in L^2 (by Itô Isometry). Since $X_t^{(n_k)} = I(\phi_{n_k} \; 1_{[0,t]})$ and by the (a.s.) uniqueness of the limit in L^2, we have $X_t = I(\phi \; 1_{[0,t]})$ a.s. □

Itô Isometry works also for conditional expectation:

Proposition 8.2.9. *For any $\phi \in \mathcal{H}^2$ and $0 \le s \le t$, we have*

$$\mathbb{E}\left[\left(\int_s^t \phi(\omega,u)dB_u\right)^2 \middle| \mathcal{F}_s\right] = \mathbb{E}\left[\int_s^t \phi^2(\omega,u)du \middle| \mathcal{F}_s\right].$$

Proof. By definition of conditional expectation, we have to prove that, for any $A \in \mathcal{F}_s$,

$$\mathbb{E}\left[1_A\left(\int_s^t \phi(\omega,u)dB_u\right)^2\right] = \mathbb{E}\left[1_A \int_s^t \phi^2(\omega,u)du\right].$$

Let us define then $\tilde{\phi} = \phi \; 1_A \; 1_{(s,t]}$. Clearly, $\tilde{\phi} \in \mathcal{H}^2$, which implies that $\|\tilde{\phi}\|_{\mathcal{H}^2} = \|I(\tilde{\phi})\|_{L^2}$. Moreover,

$$I(\tilde{\phi}) = \int_0^T \phi(\omega,u) \; 1_A(\omega) \; 1_{(s,t]}(u)dB_u = 1_A(\omega)\int_s^t \phi(\omega,u)dB_u.$$

Additionally,

$$\|\tilde{\phi}\|_{\mathcal{H}^2} = \mathbb{E}\left[\int_0^T \phi^2(\omega,u) \; 1_A(\omega) \; 1_{(s,t]}(u)du\right] = \mathbb{E}\left[1_A(\omega)\int_s^t \phi^2(\omega,u)du\right],$$

from where the result follows. □

Remark 8.2.10. Define

$$M_t = \left(\int_0^t \phi(\omega,u)dB_u\right)^2 - \int_0^t \phi^2(\omega,u)du,$$

for $\phi \in \mathcal{H}^2$. The proposition above shows that M is a martingale.

Proposition 8.2.11. *Let ϕ and ψ in $\mathcal{H}^2$ and consider a stopping time τ. Assume*

$$\phi(\omega,s) = \psi(\omega,s) \text{ a.s. if } s \le \tau(\omega).$$

Then, the stochastic integrals

$$\int_0^t \phi(\omega,s)dB_s = \int_0^t \psi(\omega,s)dB_s \text{ for } t \le \tau(\omega).$$

Proof. We start by assuming that ϕ is bounded and $\psi \equiv 0$. The idea is to prove for $\phi \in \mathcal{H}_0^2$ bounded and then take the limit. For $\phi \in \mathcal{H}_0^2$, we may write:

$$\phi(\omega,s) = \sum_{i=0}^{n-1} a_i(\omega)1_{(t_i,t_{i+1}]}(s),$$

where $a_i \in \mathcal{F}_{t_i}$ and $0 = t_0 < \cdots < t_n = T$. Since $\phi(\omega,s) = 0$ for $s \le \tau(\omega)$, we must have $a_i(\omega) = 0$ for i such $t_i < \tau(\omega)$. Notice now that

$$\int_0^t \phi(\omega,s)dB_s = \sum_{i=0}^{k-1} a_i(\omega)(B_{t_{i+1}} - B_{t_i}) + a_k(\omega)(B_t - B_{t_k}),$$

where k is chosen such that $t \in (t_k, t_{k+1}]$. By the property of a, we have $a_i(\omega) = 0$, $i = 0,\ldots,k$, for ω s.t. $t \le \tau(\omega)$. Then, the Itô integral of ϕ is zero for $t \le \tau(\omega)$.

□

8.2.4 Localization

The first important generalization is to consider $\phi : \Omega \times [0,T] \longrightarrow \mathbb{R}$ measurable, adapted and satisfying the weaker integrability condition

$$\mathbb{P}\left(\int_0^T \phi^2(\omega,t)dt < +\infty\right) = 1.$$

We write $\phi \in \mathcal{L}^2_{loc}[0,T]$. Clearly, $\mathcal{H}^2[0,T] \subset \mathcal{L}^2_{loc}[0,T]$.

One of the main reasons to consider this space is to able to integrate continuous functions of the Brownian motion.

Definition 8.2.12 (Localizing Sequence in $\mathcal{H}^2$)**.** We say $(\nu_n)_{n\in\mathbb{N}}$ is a localizing sequence for ϕ if , for any $n \in \mathbb{N}$,

(i) ν_n is a stopping time;

(ii) $\nu_n \leq \nu_{n+1}$;

(iii) $\phi_n(\omega,t) = \phi(\omega,t)\ 1_{\{t\leq\nu_n\}}(\omega) \in \mathcal{H}^2[0,T]$;

(iv) $\mathbb{P}\left(\bigcup_{n\in\mathbb{N}}\{\nu_n = T\}\right) = 1$.

Theorem 8.2.13. *For any $\phi \in \mathcal{L}^2_{loc}[0,T]$, the sequence*

$$\tau_n(\omega) = \inf\left\{t \in [0,T]\ ;\ \int_0^t \phi^2(\omega,s)ds \geq n\ or\ t \geq T\right\}$$

is a localizing sequence for ϕ.

Proof. Firstly, notice that $\tau_{n+1} \geq \tau_n$. Additionally,

$$\bigcup_{n\in\mathbb{N}}\{\tau_n = T\} = \left\{\int_0^T \phi^2(\omega,t)dt < +\infty\right\}.$$

Now remember $\phi_n(\omega,t) = \phi(\omega,t)1_{\{\tau_n\geq t\}}(\omega)$ and note that

$$\int_0^T \phi_n^2(\omega,t)dt = \int_0^{\tau_n} \phi^2(\omega,t)dt \leq n.$$

Hence, $\|\phi_n\|^2_{\mathcal{H}^2} \leq n$. Therefore, $(\tau_n)_{n\in\mathbb{N}}$ is a localizing sequence for ϕ. □

Theorem 8.2.14. *The Itô integral of ϕ, denoted by X, is defined as the limit of the continuous martingales*

$$X_t^{(n)} = \int_0^t \phi(\omega,s)\ 1_{\{s\leq\tau_n\}}dB_s.$$

This limit exists and is well defined.

Proof. We need to prove that the limit of $X_t^{(n)}$ exists and it is well defined. First, note that $X_t^{(n)} = X_t^{(m)}$ for $m \leq n$ and on the set $\{\tau_m \geq t\}$. This follows from Theorem 8.2.11 and from the fact $\tau_m \leq \tau_n$. Then, define $N = \min\{n\ ;\ \tau_n = T\}$ and note that $\mathbb{P}(N < +\infty) = 1$. Moreover, consider $\Omega_0 \in \mathcal{F}$ with $\mathbb{P}(\Omega_0) = 1$ such that, for any $\omega \in \Omega_0$, $t \mapsto X_t^{(n)}(\omega)$ is continuous for every $n \in \mathbb{N}$. Define then $\Omega_1 = \{N < +\infty\} \cap \Omega_0$. Clearly, Ω_1 has probability 1. Let $X_t(\omega) = X_t^{(N(\omega))}(\omega)$. Thus $(X_t)_{t\in[0,T]}$ is continuous and it is the limit of $X_t^{(n)}$ with probability, for every $t \in [0,T]$.

□

Remark 8.2.15. One could prove that the limit that defines the Itô integral above is independent of the localizing sequence.

Proposition 8.2.16. *If $\phi \in \mathcal{L}^2_{loc}$ and τ is a stopping time, then*

$$\int_0^{\tau \wedge t} \phi(\omega, s) dB_s = \int_0^t \phi(\omega, s) 1_{\{s \le \tau(\omega)\}} dB_s$$

This proposition follows by approximating the stopping time with discrete-valued stopping times. Moreover, Theorem 8.2.11 also holds in $\mathcal{L}^2_{loc}$.

Proposition 8.2.17. *Let ϕ and ψ in $\mathcal{L}^2_{loc}$ and consider a stopping time τ. Assume*

$$\phi(\omega, s) = \psi(\omega, s) \text{ a.s. if } s \le \tau(\omega).$$

Then, the stochastic integrals

$$\int_0^t \phi(\omega, s) dB_s = \int_0^t \psi(\omega, s) dB_s \text{ for } t \le \tau(\omega).$$

We have seen that the stochastic integral of function in $\mathcal{H}^2$ are martingales. What happens when we consider functions in $\mathcal{L}^2_{loc}$?

Theorem 8.2.18. *For any $\phi \in \mathcal{L}^2_{loc}[0, T]$, there exists a local martingale $(X_t)_{t \in [0,T]}$ such that*

$$\mathbb{P}\left(X_t = \int_0^t \phi(\omega, s) dB_s \right) = 1,$$

for any $t \in [0, T]$.

Proof. It follows from the fact $\phi_n \in \mathcal{H}^2$ and from Proposition 8.2.16. □

8.2.5 Riemann Sums

It is very easy to verify that if $f : \mathbb{R} \longrightarrow \mathbb{R}$ continuous, then $f(B) \in \mathcal{L}^2_{loc}[0, T]$. Additionally, we have the following useful representation:

Theorem 8.2.19. *If $f : \mathbb{R} \longrightarrow \mathbb{R}$ continuous and $t_i = iT/n$, $i = 0, \dots, n$, then*

$$\int_0^T f(B_t) dB_t = \lim_{n \to +\infty} \sum_{i=0}^{n-1} f(B_{t_i})(B_{t_{i+1}} - B_{t_i}),$$

where the convergence above is in probability.

Proof. Notice that $(f(B_t))_{t \in [0,T]} \in \mathcal{L}^2_{loc}$ and that

$$\tau_M = \inf\{t \ge 0 \ ; \ |B_t| \ge M \text{ or } t \ge T\}$$

is a localizing sequence for this process (since f is continuous). Moreover, let f_M be continuous with compact support and such that $f_M(x) = f(x)$, for all $|x| \le M$. Notice that $f_M(B) \in \mathcal{H}^2$ and that

$$\phi_n(\omega, t) = \sum_{i=0}^{n-1} f_M(B_{t_i}) 1_{(t_i, t_{i+1}]}(t)$$

belongs to $\mathcal{H}^2_0$ and converges to $f_M(B)$ in $\mathcal{H}^2$. Hence

$$\int_0^T f_M(B_t) dB_t = \lim_{n \to +\infty} \sum_{i=0}^{n-1} f_M(B_{t_i})(B_{t_{i+1}} - B_{t_i}). \tag{8.1}$$

For $\omega \in \{\tau_M = T\}$, we have $f(B_{t_i}) = f_M(B_{t_i})$ for any i. Then, by Theorem 8.2.11, we also have

$$\int_0^T f_M(B_t) dB_t = \int_0^T f(B_t) dB_t \text{ on } \{\tau_M = T\}.$$

These are all the ingredients needed to prove the desired result. Indeed, for any $\varepsilon > 0$, define

$$A_n(\varepsilon) = \left\{ \left| \sum_{i=0}^{n-1} f(B_{t_i})(B_{t_{i+1}} - B_{t_i}) - \int_0^T f(B_s)dB_s \right| \geq \varepsilon \right\}.$$

By the definition of convergence in probability, we need to prove $\mathbb{P}(A_n(\varepsilon)) \to 0$, as $n \to +\infty$. Note now

$$\mathbb{P}(A_n(\varepsilon)) \leq \mathbb{P}(\tau_M < T) + \mathbb{P}(A_n(\varepsilon) \cap \{\tau_M = T\}).$$

Moreover, $\mathbb{P}(\tau_M < T) \to 0$, as $M \to +\infty$. The second term goes to zero because we can replace f by f_M, apply Chebyshev's inequality and use the fact that $f_M(B) \in \mathcal{H}^2$ and (8.1). □

8.2.6 Gaussian Integrals

Consider $\phi : [0, T] \longrightarrow \mathbb{R}$ continuous and define

$$X_t = \int_0^t \phi(s)dB_s$$

One can show that $(X_t)_{t\in[0,T]}$ is a Gaussian process with mean 0 and covariance function

$$\mathrm{Cov}(X_t, X_s) = \int_0^{s\wedge t} \phi^2(u)du.$$

Moreover, if we take the sequence of partitions $t_i = iT/n$ and t_i^* any point in $[t_i, t_{i+1}]$, we have

$$\int_0^T \phi(s)dB_s = \lim_{n\to+\infty} \sum_{i=0}^{n-1} \phi(t_i^*)(B_{t_{i+1}} - B_{t_i}),$$

where the convergence is in probability.

Chapter 9

Itô Formula

The Fundamental Theorem of Calculus tells us that if $f \in C^1(\mathbb{R})$, then

$$f(x) = f(0) + \int_0^x f'(y)dy.$$

In general, if g is of finite variation, we have

$$f(g(t)) = f(g(0)) + \int_0^t f'(g(s))dg(s).$$

How this theorem could be extended to deal with function of finite *quadratic* variation?

9.1 The Brownian Case

Theorem 9.1.1 (Itô Formula)**.** *If $f \in C^2(\mathbb{R})$, then*

$$f(B_t) = f(B_0) + \int_0^t f'(B_s)dB_s + \frac{1}{2}\int_0^t f''(B_s)ds \ \ a.s. \tag{9.1}$$

Proof. There are many ways to prove this results, some shorter and some longer. We will follow a more straightforward way hoping it will make the reader better understand the main ideas.

We firstly assume that f has compact support. This implies that f, f' and f'' are bounded. Moreover, Taylor formula tells us that

$$f(y) - f(x) = (y-x)f'(x) + \frac{1}{2}(y-x)^2 f''(x) + r(x,y),$$

where

$$r(x,y) = \int_x^y (y-s)(f''(s) - f''(x))ds,$$

and notice that

$$|r(x,y)| \leq (y-x)^2 h(x,y),$$

with h uniformly continuous, bounded and satisfying $h(x,x) = 0$.

Consider now the partition of $[0,t]$, $t_i = it/n$, where $t > 0$ is fixed. Notice

$$f(B_t) - f(B_0) = \sum_{i=0}^{n-1}(f(B_{t_{i+1}}) - f(B_{t_i})).$$

Then

$$f(B_{t_{i+1}}) - f(B_{t_i}) = f'(B_{t_{i-1}})(B_{t_{i+1}} - B_{t_i})$$

$$+\frac{1}{2}f''(B_{t_i})(B_{t_{i+1}}-B_{t_i})^2+r(B_{t_{i+1}},B_{t_i}).$$

Let us define:

$$\begin{aligned}
A_n &= \sum_{i=0}^{n-1} f'(B_{t_i})(B_{t_{i+1}}-B_{t_i}),\\
B_n &= \frac{1}{2}\sum_{i=0}^{n-1} f''(B_{t_i})(B_{t_{i+1}}-B_{t_i})^2,\\
C_n &= \sum_{i=0}^{n-1} r(B_{t_{i+1}},B_{t_i}).
\end{aligned}$$

Notice that, since we are assuming f has compact support, $f'(B_s)\in\mathcal{H}^2$ and hence

$$\lim_{n\to+\infty} A_n = \int_0^t f'(B_s)dB_s,$$

where the convergence above is in probability. Consider now

$$\begin{aligned}
B_n = &\frac{1}{2}\sum_{i=0}^{n-1} f''(B_{t_i})(t_{i+1}-t_i)\\
&+\frac{1}{2}\sum_{i=0}^{n-1} f''(B_{t_i})\left((B_{t_{i+1}}-B_{t_i})^2-(t_{i+1}-t_i)\right)
\end{aligned}$$

By the continuity (in s) of $f''(B_s(\omega))$,

$$\lim_{n\to+\infty}\frac{1}{2}\sum_{i=0}^{n-1} f''(B_{t_i})(t_{i+1}-t_i)=\frac{1}{2}\int_0^t f''(B_s)ds,$$

where the convergence above is almost surely (since $s\mapsto B_s$ is continuous almost surely). Let us denote the second term of B_n by $\tilde{B}_n$. Then

$$\begin{aligned}
\mathbb{E}[\tilde{B}_n^2]=\frac{1}{4}\sum_{i,j=0}^{n-1}&\mathbb{E}[f''(B_{t_i})\left((B_{t_{i+1}}-B_{t_i})^2-(t_{i+1}-t_i)\right)\\
&f''(B_{t_j})\left((B_{t_{j+1}}-B_{t_j})^2-(t_{j+1}-t_j)\right)]
\end{aligned}$$

Without loss of generality, assume $i<j$. Then $f''(B_{t_i})\left((B_{t_{i+1}}-B_{t_i})^2-(t_{i+1}-t_i)\right)f''(B_{t_j})$ is independent of $(B_{t_{j+1}}-B_{t_j})^2-(t_{j+1}-t_j)$, which has zero expectation. Hence,

$$\begin{aligned}
&\mathbb{E}[f''(B_{t_i})\left((B_{t_{i+1}}-B_{t_i})^2-(t_{i+1}-t_i)\right)\\
&f''(B_{t_{j-1}})\left((B_{t_j}-B_{t_{j-1}})^2-(t_j-t_{j-1})\right)]=0,
\end{aligned}$$

which implies

$$\begin{aligned}
\mathbb{E}[\tilde{B}_n^2]&=\frac{1}{4}\sum_{i=0}^{n-1}\mathbb{E}[(f''(B_{t_i}))^2\left((B_{t_{i+1}}-B_{t_i})^2-(t_{i+1}-t_i)\right)^2]\\
&=\frac{1}{4}\sum_{i=0}^{n-1}\mathbb{E}[(f''(B_{t_i}))^2]\mathbb{E}[\left((B_{t_{i+1}}-B_{t_i})^2-(t_{i+1}-t_i)\right)^2]\\
&\le\frac{1}{4}\|f''\|_\infty^2\sum_{i=0}^{n-1}\mathbb{E}[\left((B_{t_{i+1}}-B_{t_i})^2-(t_{i+1}-t_i)\right)^2]\\
&=\frac{1}{4}\|f''\|_\infty^2\sum_{i=0}^{n-1}\mathbb{E}[\left((B_{t_{i+1}}-B_{t_i})^4-2(B_{t_{i+1}}-B_{t_i})^2(t_{i+1}-t_i)+(t_{i+1}-t_i)^2\right)]
\end{aligned}$$

$$= \frac{1}{4}\|f''\|_\infty^2 \sum_{i=0}^{n-1} \left(3(t_{i+1}-t_i)^2 - 2(t_{i+1}-t_i)^2 + (t_{i+1}-t_i)^2\right)$$
$$= \frac{1}{4}\|f''\|_\infty^2 \sum_{i=0}^{n-1} 2(t_{i+1}-t_i)^2 = \|f''\|_\infty^2 \frac{t^2}{2n}.$$

Moreover, $\mathbb{E}[\tilde{B}_n] = 0$. Hence, $\tilde{B}_n \to 0$ in L^2 as $n \to +\infty$. This implies convergence in probability. Let us now verify that $C_n \to 0$ in probability. By the Cauchy-Schwarz inequality,

$$\begin{aligned}\mathbb{E}[|C_n|] &\leq \sum_{i=0}^{n-1} \mathbb{E}[|r(B_{t_i}, B_{t_{i+1}})|] \leq \sum_{i=0}^{n-1} \mathbb{E}[(B_{t_{i+1}} - B_{t_i})^2 h(B_{t_{i+1}}, B_{t_i})] \\ &\leq \sum_{i=0}^{n-1} \mathbb{E}[(B_{t_{i+1}} - B_{t_i})^4]^{1/2} \mathbb{E}[h(B_{t_{i+1}}, B_{t_i})^2]^{1/2} \\ &= \sum_{i=0}^{n-1} \sqrt{3}(t_{i+1}-t_i)\mathbb{E}[h(B_{t_{i+1}}, B_{t_i})^2]^{1/2}.\end{aligned}$$

Let us analyze the term $\mathbb{E}[h(B_{t_{i+1}}, B_{t_i})^2]^{1/2}$. Since h is uniformly continuous and $h(x,x) = 0$, for any fixed $\varepsilon > 0$, there exists $\delta > 0$ such that $|x-y| < \delta \Rightarrow |h(x,y)| \leq \varepsilon$. Thus, for any $\varepsilon > 0$,

$$\begin{aligned}\mathbb{E}[h(B_{t_{i+1}}, B_{t_i})^2] &= \mathbb{E}[h(B_{t_{i+1}}, B_{t_i})^2 \mid |B_{t_{i+1}} - B_{t_i}| < \delta]\ \mathbb{P}(|B_{t_{i+1}} - B_{t_i}| < \delta) \\ &\quad + \mathbb{E}[h(B_{t_{i+1}}, B_{t_i})^2 \mid |B_{t_{i+1}} - B_{t_i}| \geq \delta]\ \mathbb{P}(|B_{t_{i+1}} - B_{t_i}| \geq \delta) \\ &< \varepsilon^2 + \|h\|_\infty^2 \mathbb{P}(|B_{t_{i+1}} - B_{t_i}| \geq \delta) \leq \varepsilon^2 + \|h\|_\infty^2 \frac{\mathbb{E}[|B_{t_{i+1}} - B_{t_i}|^2]}{\delta^2} \\ &= \varepsilon^2 + \|h\|_\infty^2 \frac{t}{n\delta^2}.\end{aligned}$$

Hence,

$$\mathbb{E}[|C_n|] \leq n\ \sqrt{3}\frac{t}{n}\left(\varepsilon^2 + \|h\|_\infty^2 \frac{t}{n\delta^2}\right)^{1/2},$$

which tells us

$$\limsup_{n\to+\infty} \mathbb{E}[|C_n|] \leq \sqrt{3}t\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, it is clear that $C_n \to 0$ in L^1, which also implies convergence in probability.

Therefore, we have proved that, for any $t \geq 0$ fixed, the sequences $(A_n)_{n\in\mathbb{N}}$, $(B_n)_{n\in\mathbb{N}}$ and $(C_n)_{n\in\mathbb{N}}$ converge to the desired limits in probability. This means that we can choose a subsequence $(n_k)_{k\in\mathbb{N}}$ such that all three subsequences converge to the same limits almost surely. That is, for any $t \geq 0$, there exists a set Ω_t with probability 1 such that the Itô formula (9.1) is verified for $\omega \in \Omega_t$. Define also Λ as the set where the stochastic integral $\int_0^t f'(B_s)dB_s$ and the Brownian motion B are continuous. Clearly, $\mathbb{P}(\Lambda) = 1$. Define then

$$\Omega_0 = \Lambda \cap (\cap_{t\in\mathbb{Q}}\Omega_t),$$

which has also probability 1. Then, for $\omega \in \Omega_0$, the Itô formula holds for any $t \in \mathbb{Q}$ and both sides are continuous as a function of time. Hence, by the density of $\mathbb{Q}$ in $\mathbb{R}$, the Itô formula holds for any $t \geq 0$ almost surely.

We have assumed at the beginning that f has compact support. Let us remove this assumption. For any $f \in C^2(\mathbb{R})$, there exists $f_M \in C^2(\mathbb{R})$ with compact support such that $f(x) = f_M(x)$, for any $|x| \leq M$. Moreover,

$$f_M(B_t) = f_M(B_0) + \int_0^t f_M'(B_s)dB_s + \frac{1}{2}\int_0^t f_M''(B_s)ds.$$

Define now

$$\tau_M = \min\{t \geq 0\ ;\ |B_t| \geq M\},$$

and notice

$$f_M(B_s(\omega)) = f(B_s(\omega)),\ \forall\ \omega \text{ s.t. } s \leq \tau_M(\omega).$$

Hence,

$$f(B_t) = f(B_0) + \int_0^t f'(B_s)dB_s + \frac{1}{2}\int_0^t f''(B_s)ds \text{ in } \{t \le \tau_M\}.$$

We clearly have $\tau_M \to +\infty$ a.s. when $M \to +\infty$. Therefore, Itô Formula holds a.s. for any $t \ge 0$ and any $f \in C^2(\mathbb{R})$. □

Theorem 9.1.2 (Time-Dependent Itô Formula)**.** *If* $f \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R})$*, then*

$$\begin{aligned} f(t, B_t) = f(0, B_0) + \int_0^t \frac{\partial f}{\partial t}(s, B_s)ds + \int_0^t \frac{\partial f}{\partial x}(s, B_s)dB_s \\ + \frac{1}{2}\int_0^t \frac{\partial^2 f}{\partial x^2}(s, B_s)ds \end{aligned} \tag{9.2}$$

A more concise notation for Itô Formula is the *differential notation.* If $X_t = f(t, B_t)$, then

$$dX_t = df(t, B_t) = \frac{\partial f}{\partial t}(t, B_t)dt + \frac{\partial f}{\partial x}(t, B_t)dB_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t, B_t)dt.$$

9.2 Applications

Example 9.2.1 ("Solving" Itô integrals)**.** One of the most direct applications of Itô formula is to "solve" Itô integral. It is not exactly solving because we will most certainly find a formula for the Itô integral in terms of a regular Riemann (or Lebesgue) integral. Let us start with an example that we know very well:

$$\int_0^t B_s dB_s$$

We then look for a function F such that $F'(x) = x$. Clearly $F(x) = x^2/2$ and we find

$$\frac{B_t^2}{2} = \int_0^t B_s dB_s + \frac{1}{2}\int_0^t 1ds = \int_0^t B_s dB_s + \frac{t}{2}.$$

This gives the well-known formula:

$$\int_0^t B_s dB_s = \frac{B_t^2}{2} - \frac{t}{2}.$$

In general, we have

$$\int_0^t f(s, B_s)dB_s = F(t, B_t) - F(0, B_0) - \int_0^t \frac{\partial F}{\partial t}(s, B_s)ds - \frac{1}{2}\int_0^t \frac{\partial f}{\partial x}(s, B_s)ds,$$

where F is such that $\frac{\partial F}{\partial x} = f$.

Theorem 9.2.2 (Martingale Conditions)**.** *If* $f \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R})$ *satisfies*

$$\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2} = 0, \tag{9.3}$$

then $X_t = f(t, B_t)$ *is a (local) martingale. Moreover, if* $\frac{\partial f}{\partial x}(t, B_t)$ *belongs to* $\mathcal{H}^2[0, T]$*, then* X *is a true martingale in* $[0, T]$*.*

Proof. The PDE implies

$$f(t, B_t) = f(0, B_0) + \int_0^t \frac{\partial f}{\partial x}(s, B_s)dB_s.$$

Since $\frac{\partial f}{\partial x}$ is continuous, it is easy to see that the Itô integral above is a local martingale (the integrand is in $\mathcal{L}^2_{loc}$ since it is continuous in t a.s.). Moreover, if the integrand is in $\mathcal{H}^2$, we have seen that the Itô integral will be a true martingale. □

Example 9.2.3. $M_t = e^{\alpha B_t - \frac{\alpha^2}{2}t}$ is a martingale with respect to the Brownian filtration. Define $f(t, x) = e^{\alpha x - \frac{\alpha^2}{2}t}$ and notice that f satisfies Equation (9.3).

Example 9.2.4 (First Feynman-Kac Theorem)**.** Let $u \in C^{1,2}([0,T]\times\mathbb{R})$ be the solution of the following PDE:

$$\frac{\partial u}{\partial t} - \frac{1}{2}\frac{\partial^2 u}{\partial x^2} = 0$$

with final condition $u(0,x) = g(x)$. Let us prove the following representation for u:

$$u(t,x) = \mathbb{E}[g(x+B_t)].$$

Define, from the solution of the PDE above, the process

$$X_s = u(t-s, x+B_s),$$

where t is fixed and $0 \le s \le t$. By Itô formula, we find

$$\begin{aligned} dX_s &= -\frac{\partial u}{\partial t}(t-s,x+B_s)ds + \frac{\partial u}{\partial x}(t-s,x+B_s)dB_s + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}(t-s,x+B_s)ds \\ &= \frac{\partial u}{\partial x}(t-s,x+B_s)dB_s. \end{aligned}$$

Then X is a (local) martingale. Under some integrability assumptions, we have

$$\mathbb{E}[X_t] = \mathbb{E}[X_0]$$

This can be seen as a martingale interpolation between times 0 and t. Therefore

$$\mathbb{E}[g(x+B_t)] = u(t,x).$$

Example 9.2.5 (The Brownian motion with a drift)**.** Define $X_t = \mu t + \sigma B_t$ and $\tau = \inf\{t\ ;\ X_t = A \text{ or } X_t = -B\}$, with $A, B > 0$. In other to compute $\mathbb{P}(X_\tau = A)$ we will look for $h : [-B, A] \longrightarrow \mathbb{R}$ such that $M_t = h(X_t)$ is a bounded martingale. If we can make $h(A) = 1$ and $h(-B) = 0$, we would have $\mathbb{P}(X_\tau = A) = h(0)$. Indeed, we would have

$$\begin{aligned} h(0) = h(X_0) &= \mathbb{E}[M_0] = \mathbb{E}[M_\tau] \\ &= h(A)\mathbb{P}(X_\tau = A) + h(-B)\mathbb{P}(X_\tau = B) = \mathbb{P}(X_\tau = A). \end{aligned}$$

Consider now $f(t,x) = h(\mu t + \sigma x)$. We want h such that $\partial_t f + 1/2\partial_{xx} f = 0$. However,

$$\begin{aligned} \frac{\partial f}{\partial t} &= \mu h'(\mu t + \sigma x), \\ \frac{\partial^2 f}{\partial x^2} &= \sigma^2 h''(\mu t + \sigma x) \Rightarrow h''(z) = -\frac{2\mu}{\sigma^2}h'(z) \end{aligned}$$

Then

$$h'(z) = Ce^{-2\mu z/\sigma^2} \Rightarrow h(z) = -\frac{C\sigma^2}{2\mu}e^{-2\mu z/\sigma^2} + D.$$

Choosing C and D so $h(A) = 1$ and $h(-B) = 0$ gives us

$$\begin{aligned} C &= -\frac{2\mu}{\sigma^2(e^{-2\mu A/\sigma^2} - e^{2\mu B/\sigma^2})}, \\ D &= -\frac{e^{2\mu B/\sigma^2}}{e^{-2\mu A/\sigma^2} - e^{2\mu B/\sigma^2}} \end{aligned}$$

Hence,

$$h(z) = \frac{e^{-2\mu z/\sigma^2} - e^{2\mu B/\sigma^2}}{e^{-2\mu A/\sigma^2} - e^{2\mu B/\sigma^2}} \Rightarrow \mathbb{P}(X_\tau = A) = h(0) = \frac{1 - e^{2\mu B/\sigma^2}}{e^{-2\mu A/\sigma^2} - e^{2\mu B/\sigma^2}}.$$

Moreover, letting $B \to +\infty$ and assuming $\mu < 0$, we find

$$\mathbb{P}\left(\sup_{t>0} X_t \ge A\right) = \lim_{B\to+\infty} \mathbb{P}(X_\tau = A) = e^{2\mu A/\sigma^2}.$$

The case $\mu \ge 0$, we find

$$\mathbb{P}\left(\sup_{t>0} X_t \ge A\right) = 1.$$

9.3 The General Case

Definition 9.3.1 (Itô Process)**.** A process $(X_t)_{t\in[0,T]}$ is called an *Itô process* if

$$X_t = X_0 + \int_0^t \mu(\omega, s)ds + \int_0^t \sigma(\omega, s)dB_s,$$

for all $t \in [0,T]$, where $\mu : \Omega \times [0,T] \longrightarrow \mathbb{R}$ and $\sigma : \Omega \times [0,T] \longrightarrow \mathbb{R}$ measurable, adapted to $(\mathcal{F}_t)_{t\in[0,T]}$ and such that

$$\mathbb{P}\left(\int_0^T (|\mu(\omega, s)| + |\sigma(\omega, s)|^2)ds < +\infty\right) = 1.$$

μ is called the drift and σ the volatility of the Itô process.

Theorem 9.3.2 (Itô Formula)**.** *If $f \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R})$ and $(X_t)_{t\in[0,T]}$ is an Itô process with drift μ and volatility σ, then*

$$\begin{aligned} f(t, X_t) = f(0, X_0) &+ \int_0^t \frac{\partial f}{\partial t}(s, X_s)ds + \int_0^t \frac{\partial f}{\partial x}(s, X_s)\mu(\omega, s)ds \\ &+ \int_0^t \frac{\partial f}{\partial x}(s, X_s)\sigma(\omega, s)dB_s + \frac{1}{2}\int_0^t \frac{\partial^2 f}{\partial x^2}(s, X_s)\sigma^2(\omega, s)ds \end{aligned} \tag{9.4}$$

Remark 9.3.3 (Notation)**.** The following rules can be applied when dealing with the differential notation

$\cdot$	dt	dB_t
dt	0	0
dB_t	0	dt

We can then write in differential notation

$$df(t, X_t) = \frac{\partial f}{\partial t}(t, X_t)dt + \frac{\partial f}{\partial x}(t, X_t)dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t, X_t)dX_t \cdot dX_t.$$

Example 9.3.4. Let us consider the Geometric Brownian Motion: $dX_t = \mu X_t dt + \sigma X_t dB_t$. Define $Y_t = \log X_t$ and let us use Itô formula to find dY_t:

$$dY_t = \frac{1}{X_t}dX_t - \frac{1}{2X_t^2}dX_t \cdot dX_t = \left(\mu - \frac{\sigma^2}{2}\right)dt + \sigma dB_t.$$

Since the right-hand side does not depend on Y, we find

$$Y_t = Y_0 + \left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t.$$

Therefore,

$$X_t = X_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t}.$$

Similar technique could be applied to find the solution of other Stochastic Differential Equation (SDE), as we will see in the next chapter.

Example 9.3.5. Another interesting example is the well-known Ornstein-Uhlenbeck SDE: $dX_t = \kappa(m - X_t) + \sigma dB_t$. This is one example of stochastic model that presents mean reversion. To find X_t, we apply Itô formula to $Y_t = e^{\kappa t}X_t$ to find

$$dY_t = \kappa Y_t dt + e^{\kappa t}dX_t = \kappa m e^{\kappa t}dt + \sigma e^{\kappa t}dB_t.$$

Hence

$$Y_t = Y_0 + \kappa m \int_0^t e^{\kappa s}ds + \int_0^t \sigma e^{\kappa s}dB_s,$$

which gives us

$$X_t = e^{-\kappa t}X_0 + m(1 - e^{-\kappa t}) + \sigma \int_0^t e^{-\kappa(t-s)}dB_s.$$

Notice that

$$X_t \mid X_0 \sim N\left(e^{-\kappa t}X_0 + m(1 - e^{-\kappa t}), \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa t})\right).$$

9.3.1 Fokker-Planck Equation

As an application of the general Itô formula above, we will derive the Fokker-Planck PDE. For this, consider the *diffusion*:

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t,$$

with $X_0 = x_0$. Let us assume that X_t has a probability density denoted by $m(t, \cdot)$. Hence, if we consider $\phi \in C_c^\infty([0, T) \times \mathbb{R})$ (notice this implies $\phi(T, x) = 0$). By Itô formula, we have

$$\begin{aligned}\phi(T, X_T) - \phi(0, x_0) = &\int_0^T \left(\frac{\partial \phi}{\partial t}(t, X_t) + \mu(t, X_t)\frac{\partial \phi}{\partial x}(t, X_t) + \frac{1}{2}\sigma^2(t, X_t)\frac{\partial^2 \phi}{\partial x^2}(t, X_t) \right) dt \\ &+ \int_0^T \sigma(t, X_t)\frac{\partial \phi}{\partial x}(t, X_t)dB_t.\end{aligned}$$

Hence, taking expectation and using the definition of $m(t, \cdot)$:

$$-\phi(0, x_0) = \int_0^T \int_{\mathbb{R}} \left(\frac{\partial \phi}{\partial t}(t, x) + \mu(t, x)\frac{\partial \phi}{\partial x}(t, x) + \frac{1}{2}\sigma^2(t, x)\frac{\partial^2 \phi}{\partial x^2}(t, x) \right) m(t, x)dxdt.$$

Integrating by parts, we find:

$$\begin{aligned}
\int_{\mathbb{R}} \int_0^T \frac{\partial \phi}{\partial t}(t, x)m(t, x)dtdx &= -\int_{\mathbb{R}} \overset{\delta_{x_0}}{\cancel{m(0, x)}}\phi(0, x)dx - \int_{\mathbb{R}} \int_0^T \frac{\partial m}{\partial t}(t, x)\phi(t, x)dtdx, \\
\int_0^T \int_{\mathbb{R}} \frac{\partial \phi}{\partial x}(t, x)\mu(t, x)m(t, x)dtdx &= -\int_{\mathbb{R}} \int_0^T \frac{\partial (m\mu)}{\partial x}(t, x)\phi(t, x)dtdx, \\
\int_0^T \int_{\mathbb{R}} \frac{\partial^2 \phi}{\partial x^2}(t, x)\sigma^2(t, x)m(t, x)dtdx &= \int_{\mathbb{R}} \int_0^T \frac{\partial^2 (m\sigma^2)}{\partial x^2}(t, x)\phi(t, x)dtdx.
\end{aligned}$$

Since ϕ is arbitrary, putting all together we find the Fokker-Planck PDE for m:

$$-\frac{\partial m}{\partial t}(t, x) - \frac{\partial (m\mu)}{\partial x}(t, x) + \frac{\partial^2 (m\sigma^2)}{\partial x^2}(t, x) = 0,$$

with $m(0, x) = \delta_{x_0}(x)$.

9.4 Quadratic Variation

Let us remember again what is the quadratic variation of a process.

Definition 9.4.1. Consider a stochastic process $(X_t)_{t \geq 0}$ and let $\pi = \{t_0 < t_1 \cdots < t_n\}$ be a partition of $[0, t]$, for $t > 0$. We define the π-quadratic variation of $(X_t)_{t \geq 0}$ as the random variable:

$$Q_\pi(X_t) = \sum_{i=0}^{n-1} (X_{t_{i+1}} - X_{t_i})^2.$$

Define also $\|\pi\| = \max_{i=1,\dots,n} |t_{i+1} - t_i|$. The *quadratic variation of* X is defined by the following limit, when it exists,

$$\langle X \rangle_t = \lim_{\|\pi\| \to 0} Q_\pi(X_t),$$

for any $t > 0$, where the convergence is in probability.

Proposition 9.4.2. *Let $(a_t)_{t \geq 0}$ be a stochastic process, measurable in $\Omega \times [0, +\infty)$, with*

$$\mathbb{P}\left(\int_0^t |a_s|ds < +\infty \right) = 1,$$

for any $t > 0$. Then, the quadratic variation of the process $(A_t)_{t \geq 0}$ given by

$$A_t = \int_0^t a_s ds$$

exists and it is 0.

Proof. Fix $t > 0$ and consider a partition of $[0, t]$: $\pi = \{t_0, \dots, t_n\}$. Notice that, if we define $\phi(\omega, u) = \int_0^u |a_s(\omega)|ds$, then, $\phi(\omega, \cdot) : [0, t] \longrightarrow \mathbb{R}$ is a continuous function in a compact interval, for almost all ω. Hence, it is uniformly continuous. This means that, for any $\omega \in \Omega_0$ (with $\mathbb{P}(\Omega_0) = 1$) and $\varepsilon > 0$, there exists $\delta > 0$, such that

$$\|\pi\| < \delta \Rightarrow \max_{i=0,\dots,n-1} \left\{ \int_{t_i}^{t_{i+1}} |a_s(\omega)|ds \right\} < \varepsilon.$$

Then,

$$\begin{aligned} Q_\pi(A_t(\omega)) &= \sum_{i=0}^{n-1} \left(\int_{t_i}^{t_{i+1}} a_s(\omega)ds \right)^2 \\ &\leq \max_{i=0,\dots,n-1} \left\{ \int_{t_i}^{t_{i+1}} |a_s(\omega)|ds \right\} \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} |a_s(\omega)|ds \\ &< \varepsilon \int_0^t |a_s(\omega)|ds, \end{aligned}$$

if $\|\pi\| < \delta$. Therefore, $Q_\pi(A_t(\omega))$ converges almost surely to 0, when $\|\pi\| \to 0$. □

Let us remember the result we derived in the previous chapter, Proposition 8.2.9:

Proposition. *Let $\phi \in \mathcal{H}^2[0, T]$ and define*

$$Z_t = \int_0^t \phi(\omega, s)dB_s \text{ and } M_t = Z_t^2 - \int_0^t \phi^2(\omega, s)ds.$$

Then, $(Z_t)_{t\in[0,T]}$ and $(M_t)_{t\in[0,T]}$ are continuous martingales with respect to the Brownian filtration.

Proposition 9.4.3. *Let $\phi \in \mathcal{H}^2[0, T]$ such that*

$$\int_0^T \phi^2(\omega, t)dt \leq C,$$

for some $C > 0$. If the process

$$Z_t = \int_0^t \phi(\omega, s)dB_s$$

is bounded (by C as well), then the quadratic variation of $(Z_t)_{t\in[0,T]}$ exists and

$$\langle Z \rangle_t = \int_0^t \phi^2(\omega, s)ds.$$

Proof. Define

$$\Delta_\pi = Q_\pi(Z_t) - \int_0^t \phi^2(\omega, s)ds = \sum_{i=0}^{n-1} d_i,$$

with $d_i = (Z_{t_{i+1}} - Z_{t_i})^2 - \int_{t_i}^{t_{i+1}} \phi^2(\omega, s)ds$. By direct computation, one can easily see that $\mathbb{E}\left[d_i \mid \mathcal{F}_{t_i}\right] = 0$. This means that $\mathbb{E}[d_i d_j] = 0$ and

$$\mathbb{E}[\Delta_\pi^2] = \sum_{i=0}^{n-1} \mathbb{E}[d_i^2] \leq \sum_{i=0}^{n-1} \mathbb{E}[(Z_{t_{i+1}} - Z_{t_i})^4] + \mathbb{E}\left[\left(\int_{t_i}^{t_{i+1}} \phi^2(\omega, s)ds \right)^2 \right].$$

From the proof Proposition 9.4.2,

$$\sum_{i=0}^{n-1} \left(\int_{t_i}^{t_{i+1}} \phi^2(\omega, s)ds \right)^2 \leq \max_{i=0,\dots,n-1} \left(\int_{t_i}^{t_{i+1}} \phi^2(\omega, s)ds \right) \underbrace{\int_0^T \phi^2(\omega, t)dt}_{\leq C} \overset{\|\pi\|\to 0}{\longrightarrow} 0.$$

By the DCT, we conclude

$$\sum_{i=0}^{n-1} \mathbb{E}\left[\left(\int_{t_i}^{t_{i+1}} \phi^2(\omega, s)ds \right)^2 \right] \overset{\|\pi\|\to 0}{\longrightarrow} 0.$$

Moreover,

$$\sum_{i=0}^{n-1}(Z_{t_{i+1}} - Z_{t_i})^4 \leq \underbrace{\left(\max_{i=0,\ldots,n-1}(Z_{t_{i+1}} - Z_{t_i})^2\right)}_{X_\pi} Q_\pi(Z_t).$$

Notice that X_π is bounded (by $4C^2$) and converges to 0 when $\|\pi\| \to 0$. Hence, by DCT, $\mathbb{E}[X_\pi^2] \to 0$, when $\|\pi\| \to 0$. Additionally, one can show that $\mathbb{E}[Q_\pi(Z_t)^2]$ is bounded in L^2 (since it is the quadratic variation of bounded martingale). Hence, by Cauchy-Schwarz

$$\mathbb{E}\left[\sum_{i=0}^{n-1}(Z_{t_{i+1}} - Z_{t_i})^4\right] \overset{\|\pi\|\to 0}{\longrightarrow} 0.$$

Therefore,

$$\langle Z \rangle_t = \int_0^t \phi^2(\omega, s)ds.$$

□

Corollary 9.4.4. *Consider the cross-variation of A given in Proposition 9.4.2 and Z given by the proposition above. Then*

$$\sum_{i=0}^{n-1}(A_{t_{i+1}} - A_{t_{i-1}})(Z_{t_{i+1}} - Z_{t_i}) \overset{\|\pi\|\to 0}{\longrightarrow} 0.$$

Proof. The proof follow easily from the Cauchy-Schwarz inequality:

$$\sum_{i=0}^{n-1}(A_{t_{i+1}} - A_{t_{i-1}})(Z_{t_{i+1}} - Z_{t_i}) \leq Q_\pi(A)^{1/2} Q_\pi(Z)^{1/2}.$$

□

Theorem 9.4.5. *Let $(X_t)_{t\in[0,T]}$ be an Itô process with drift μ and volatility σ. Then*

$$\langle X \rangle_t = \int_0^t \sigma^2(\omega, s)ds.$$

In differential equation, we have

$$d\langle X \rangle_t = \sigma^2(\omega, t)dt = dX_t \cdot dX_t.$$

Proof. Consider the τ_M defined as

$$\tau_M = \inf\left\{t \leq T\ ;\ \int_0^t |\mu(\omega, s)|ds \geq M \text{ or } \int_0^t \sigma^2(\omega, s)ds \geq M \right. \\ \left. \text{or } \left|\int_0^t \sigma(\omega, s)dB_s\right| \geq M\right\}$$

First notice that $X_{\cdot\wedge\tau_M}$ is also an Itô process but with drift $\mu\, 1_{\{\tau_M \geq t\}}$ and volatility $\sigma\, 1_{\{\tau_M\geq t\}}$. Moreover, X satisfies the propositions and corollary above. This implies that

$$\langle X_{\cdot\wedge\tau_M} \rangle_t = \int_0^t \sigma^2(\omega, s) 1_{\{\tau_M \geq s\}} ds.$$

Since $\tau_M \to T$, when $M \to +\infty$, we conclude that

$$\langle X \rangle_t = \int_0^t \sigma^2(\omega, s)ds,$$

as desired. □

Theorem 9.4.6 (Itô Formula Revisited)**.** *If $f \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R})$ and $(X_t)_{t\in[0,T]}$ is an Itô process, then*

$$df(t, X_t) = \frac{\partial f}{\partial t}(t, X_t)dt + \frac{\partial f}{\partial x}(t, X_t)dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t, X_t)d\langle X \rangle_t.$$

9.5 Multidimensional Itô Calculus

We say $B_t = (B_t^{(1)}, \dots, B_t^{(d)})$ is a Brownian motion in $\mathbb{R}^d$ if each $B^{(i)}$ is a Brownian motion and $B^{(i)}$ and $B^{(j)}$ are independent for any $i \neq j$. The Itô integral in $\mathbb{R}^d$ is defined as

$$\int_0^t \phi(\omega, s) \cdot dB_s = \sum_{i=1}^d \int_0^t \phi^{(i)}(\omega, s) dB_s^{(i)}.$$

Moreover, if $f \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^d)$, we have

$$df(t, B_t) = \frac{\partial f}{\partial t}(t, B_t)dt + \nabla f(t, B_t) \cdot dB_t + \frac{1}{2}\Delta f(t, B_t)dt,$$

where ∇f is the gradient and Δf is the Laplacian of f. Hence, $f(t, B_t)$ will be a local martingale if

$$\frac{\partial f}{\partial t}(t, x) + \frac{1}{2}\Delta f(t, x) = 0.$$

Remark 9.5.1 (Correlated Brownian motions)**.** Let B and $B^{\perp}$ be two independent Brownian motions and $\rho : \Omega \times [0, T] \longrightarrow [-1, 1]$ an adapted stochastic process taking values in $[-1, 1]$. Define then

$$W_t = \rho(\omega, t)B_t + \sqrt{1 - \rho(\omega, t)^2} B_t^{\perp}.$$

One can easily prove that W is itself a Brownian motion. Moreover, using the box calculus, we find

$$dW_t dB_t = \rho_t dt$$

Theorem 9.5.2 (Itô Formula in Two Dimensions)**.** *If $f \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^2)$, $(X_t)_{t\in[0,T]}$ and $(Y_t)_{t\in[0,T]}$ are Itô processes, then*

$$\begin{aligned} df(t, X_t, Y_t) &= \frac{\partial f}{\partial t}(t, X_t, Y_t)dt + \frac{\partial f}{\partial x}(t, X_t, Y_t)dX_t + \frac{\partial f}{\partial y}(t, X_t, Y_t)dY_t \\ &+ \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t, X_t, Y_t)dX_t \cdot dX_t + \frac{\partial^2 f}{\partial x \partial y}(t, X_t, Y_t)dX_t \cdot dY_t + \frac{1}{2}\frac{\partial^2 f}{\partial y^2}(t, X_t, Y_t)dY_t \cdot dY_t \end{aligned}$$

For instance, assume

$$\begin{cases} dX_t = \mu(\omega, t)dt + \sigma(\omega, t)dB_t, \\ dY_t = a(\omega, t)dt + b(\omega, t)dW_t, \end{cases}$$

where B and W are correlated Brownian motions with $dB_t dW_t = \rho(\omega, t)dt$. This implies

$$\begin{aligned} &df(t, X_t, Y_t) \\ &= \left(\frac{\partial f}{\partial t} + \mu(\omega, t)\frac{\partial f}{\partial x} + a(\omega, t)\frac{\partial f}{\partial y} + \frac{1}{2}\sigma^2(\omega, t)\frac{\partial^2 f}{\partial x^2} + \frac{1}{2}b^2(\omega, t)\frac{\partial^2 f}{\partial y^2} + \sigma(\omega, t)b(\omega, t)\rho(\omega, t)\frac{\partial^2 f}{\partial x \partial y} \right) dt \\ &+ \frac{\partial f}{\partial x}\sigma(\omega, t)dB_t + \frac{\partial f}{\partial y}b(\omega, t)dW_t. \end{aligned}$$

Remark 9.5.3 (Cross-Variation)**.** The cross-variation of X and Y along the partition π is given by

$$Q_\pi(X_t, Y_t) = \sum_{i=0}^{n-1} (X_{t_{i+1}} - X_{t_{i-1}})(Y_{t_{i+1}} - Y_{t_i}).$$

Notice that

$$Q_\pi(X_t, Y_t) = \frac{1}{4}(Q_\pi(X_t + Y_t) - Q_\pi(X_t - Y_t)).$$

Hence, if $X + Y$ and $X - Y$ have quadratic variation, then $Q_\pi(X_t, Y_t)$ converges in probability when $\|\pi\| \to 0$. In this case, we say that there exists the *cross-variation of X and Y* and we denote it by $\langle X, Y \rangle$. Moreover,

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

If X and Y are Itô processes with volatility σ and b respectively, we can use the formula above to show

$$\langle X, Y \rangle_t = \int_0^t \sigma(\omega, s)b(\omega, s)\rho(\omega, s)ds$$

9.6 Levy's Characterization of Brownian Motion

We have seen that the Brownian motion is a martingale starting at 0, with continuous paths and $\langle X\rangle_t = t$. The next result shows the converse of this fact.

Theorem 9.6.1 (Levy's Martingale Characterization of Brownian motion)**.** *Let $(X_t)_{t\in[0,T]}$ be a stochastic process with continuous paths, $X_0 = 0$ and such that $(X_t)_{t\in[0,T]}$ is a martingale and $\langle X\rangle_t = t$. Then $(X_t)_{t\in[0,T]}$ is a Brownian motion.*

Proof. Notice that $\langle X\rangle_t = t$ implies that $(X_t^2 - t)_{t\in[0,T]}$ is a martingale. Fix $0 \le t \le T$ and notice that, for $s \ge t$,

$$\mathbb{E}[X_s - X_t \mid \mathcal{F}_t] = 0 \text{ and } \mathbb{E}[(X_s - X_t)^2 \mid \mathcal{F}_t] = s - t.$$

By Itô formula, we find

$$(X_s - X_t)^n = n\int_t^s (X_u - X_t)^{n-1} dX_u + \frac{1}{2}n(n-1)\int_s^t (X_u - X_t)^{n-2} du.$$

Taking expectation:

$$\mathbb{E}[(X_s - X_t)^n \mid \mathcal{F}_t] = \frac{1}{2}n(n-1)\int_s^t \mathbb{E}[(X_u - X_t)^{n-2} \mid \mathcal{F}_t] du.$$

Hence, by induction, we see that

$$\mathbb{E}[(X_s - X_t)^{2n+1} \mid \mathcal{F}_t] = 0 \text{ and } \mathbb{E}[(X_s - X_t)^{2n} \mid \mathcal{F}_t] = \frac{(2n)!}{2^n n!}(s-t)^n.$$

Therefore

$$\begin{aligned}\mathbb{E}[e^{\alpha(X_s - X_t)} \mid \mathcal{F}_t] &= \sum_{n=0}^{+\infty} \frac{\alpha^n}{n!}\mathbb{E}[(X_s - X_t)^n \mid \mathcal{F}_t] = \sum_{n=0}^{+\infty} \frac{\alpha^{2n}}{(2n)!}\frac{(2n)!}{2^n n!}(s-t)^n \\ &= \sum_{n=0}^{+\infty} \frac{(\alpha^2(s-t)/2)^n}{n!} = e^{\frac{\alpha^2}{2}(s-t)}.\end{aligned}$$

This implies that

$$X_s - X_t \mid \mathcal{F}_t \sim N(0, t-s).$$

This concludes the proof. □

9.7 Burkeholder-Davis-Gundy Inequality

In this section, we will prove a useful generalization of Doob's Maximal Inequality.

Theorem 9.7.1 (Burkeholder-Davis-Gundy Inequality)**.** *Consider $\sigma \in L^2_{loc}[0,T]$ such that*

$$\|\sigma\|_{p,2}^p = \mathbb{E}\left[\left(\int_0^T \sigma_t^2 dt\right)^{\frac{p}{2}}\right] < +\infty,$$

for $p > 0$. Then, there exists universal constants $0 < c_p < C_p$ such that

$$c_p\|\sigma\|_{p,2} \le \|M_T^*\|_p \le C_p\|\sigma\|_{p,2}, \tag{9.5}$$

where $M_T^* = \sup_{t\in[0,T]} M_t$.

Chapter 10

Stochastic Differential Equations

10.1 Introduction

The results presented in this chapter are usually valid in a multi-dimensional setting. However, to focus on the main ideas, we will present them in the one-dimensional case.

The goal of this chapter is to study the following equation:

$$X_t = \eta + \int_0^t \mu_s(X_s)ds + \int_0^t \sigma_s(X_s)dB_s, \ 0 \le t \le T, \quad \text{a.s.} \tag{10.1}$$

This is called a Stochastic Differential Equation, or SDE.

As of now, we have studied particular cases of this equation:

- μ and σ independent of X;
- giving a specific formula for X depending on the Brownian motion B and finding the dynamics of it;
- finding a smart transformation that simplifies the SDE to the first situation.

Let us exemplify the last item with the following important example:

$$dX_t = \mu_t X_t dt + \sigma_t X_t dB_t.$$

To solve this SDE, one just need to consider the transformation $\log X_t$. A simple application of Itô formula gives us

$$d\log X_t = \frac{1}{X_t}dX_t - \frac{1}{2X_t^2}(dX_t)^2 = \left(\mu_t - \frac{1}{2}\sigma_t^2\right)dt + \sigma_t dW_t,$$

which implies

$$X_t = \eta \exp\left\{\int_0^t \left(\mu_s + \frac{1}{2}\sigma_s^2\right)ds + \int_0^t \sigma_s dB_s\right\}.$$

Let us now study the linear case: $\mu_s(x) = \mu_s^1 x + \mu_s^0$ and $\sigma_s(x) = \sigma_s^1 x + \sigma_s^0$ and then move to the general non-linear case (with Lipschitz coefficients).

The cases $\mu_s^1 = \sigma_s^1 \equiv 0$ and $\mu_s^0 = \sigma_s^0 \equiv 0$ are trivially dealt with (explicit and Itô formula with log). Let us then consider the fully linear case. Define

$$d\Phi_t = (-\mu_t^1 + (\sigma_t^1)^2)\Phi_t dt - \sigma_t^1 \Phi_t dB_t, \tag{10.2}$$

with $\Phi_0 = 1$, and notice that

$$\begin{aligned} d(\Phi_t X_t) &= \Phi_t dX_t + X_t d\Phi_t + dX_t d\Phi_t \\ &= \left(\Phi_t(\mu_t^1 X_t + \mu_t^0) + X_t(-\mu_t^1 + (\sigma_t^1)^2)\Phi_t - \sigma_t^1 \Phi_t(\sigma_t^1 X_t + \sigma_t^0)\right) dt \end{aligned}$$

$$\begin{aligned}&+ \Phi_t(\sigma_t^1 X_t + \sigma_t^0)dB_t - X_t\sigma_t^1\Phi_t dB_t \\ &= (\mu_t^0 - \sigma_t^1\sigma_t^0)\Phi_t dt + \sigma_t^0\Phi_t dB_t.\end{aligned}$$

Hence, under suitable conditions, $\Phi_t > 0$ and then:

$$X_t = \frac{1}{\Phi_t}\left(\eta + \int_0^t (\mu_s^0 - \sigma_s^1\sigma_s^0)\Phi_s ds + \int_0^t \sigma_s^0\Phi_s dB_s\right).$$

Example 10.1.1. A very interesting example is the Ornstein-Uhlenbeck (OU) process:

$$dX_t = \kappa(m - X_t)dt + \sigma dB_t.$$

This process has the mean-reversion property. Indeed, the drift is positive when $X_t < m$ and negative otherwise. Applying the above discussion for X, gives us $d\Phi_t = \kappa\Phi_t dt$, with $\Phi_0 = 1$. This implies that $\Phi_t = e^{\kappa t}$. Hence

$$X_t = e^{-\kappa t}\eta + m(1 - e^{-\kappa t}) + e^{-\kappa t}\sigma\int_0^t e^{\kappa s}dB_s.$$

Since the Itô integral of a deterministic integrand is Gaussian, the process X is Gaussian as well. Moreover,

$$X_t \sim N\left(e^{-\kappa t}\eta + m(1 - e^{-\kappa t}),\ \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa t})\right).$$

We conclude that the stationary distribution of X is $N\left(m, \dfrac{\sigma^2}{2\kappa}\right)$.

Remark 10.1.2. In what follows, we will use several inequalities and we list them below:

- Young: $2ab \le a^2 + b^2$;
- ε-Young: $2ab \le \varepsilon a^2 + \varepsilon^{-1}b^2$;
- Gronwall: $u(t) \le \alpha(t) + \int_0^t \beta(s)u(s)ds \Rightarrow u(t) \le \alpha(t)e^{\int_0^t \beta(s)ds}$, if α is increasing.
- Doob maximal: if M is a martingale, then $\mathbb{E}\left[\sup_{0\le t\le T} M_t^2\right] \le 4\mathbb{E}[M_T^2]$.

10.2 Existence and Uniqueness of Solution

In this section, we will study general results for the SDE (10.1). The assumptions that will be present throughout the section are:

Assumption 10.2.1.

(i) μ and σ are progressively measurable;

(ii) μ and σ are uniformly Lipschitz in x, i.e. there exists $L \ge 0$ such that

$$|\mu_t(x_1) - \mu_t(x_2)| + |\sigma_t(x_1) - \sigma_t(x_2)| \le L|x_1 - x_2|,$$

for all x_1, x_2.

(iii) $\eta \in L^2$ is measurable with respect to $\mathcal{F}_0$ and $\mu_t^0 = \mu_t(0)$ and $\sigma_t^0 = \sigma_t(0)$ are sufficiently integrable.

Notice that the assumptions *(ii)* and *(iii)* above guarantee that μ and σ grow at most linearly in x:

$$|\mu_t(x)| \le |\mu_t^0| + L|x| \text{ and } |\sigma_t(x)| \le |\sigma_t^0| + L|x|.$$

We will first assume there exists a solution to (10.1) and study its properties:

Theorem 10.2.2. *Let (η, μ, σ) satisfy Assumption 10.2.1 and let X be a solution in $\mathcal{H}^2$. Then, there exists a constant $C > 0$ depending only on T and L such that*

$$\mathbb{E}\left[\sup_{0\le t\le T}|X_t|^2\right] \le C\,\mathbb{E}\left[|\eta|^2 + \left(\int_0^T |\mu_t^0|dt\right)^2 + \int_0^T |\sigma_t^0|^2 dt\right].$$

Proof. Since X is a solution of (10.1), we have

$$\sup_{0\leq t\leq T}|X_t|\leq|\eta|+\int_0^T|\mu_t(X_t)|dt+\sup_{0\leq t\leq T}\left|\int_0^T\sigma_t(X_t)dB_t\right|$$

Taking the squares, applying Young inequality, Doob's maximal inequality for martingales, using the linear growth inequality, applying Cauchy-Schwartz inequality (for the time integral) and choosing the biggest constant, we find

$$\mathbb{E}\left[\sup_{0\leq t\leq T}|X_t|^2\right]\leq CI_0^2+C\,\mathbb{E}\left[\left(\int_0^T X_t dt\right)^2\right] \tag{10.3}$$

$$\leq CI_0^2+C\,\mathbb{E}\left[\int_0^T X_t^2 dt\right] \tag{10.4}$$

$$\leq CI_0^2+C\sup_{0\leq t\leq T}\mathbb{E}[X_t^2], \tag{10.5}$$

where I_0^2 is the right-hand bound. Notice now that

$$dX_t^2=(2X_t\mu_t(X_t)+\sigma_t^2(X_t))dt+2X_t\sigma_t(X_t)dB_t.$$

Taking expectation, we find

$$\begin{aligned}\mathbb{E}[X_t^2]&=\mathbb{E}\left[\eta^2+\int_0^t(2X_s\mu_s(X_s)+\sigma_s^2(X_s))ds\right]\\&\leq C\int_0^t\mathbb{E}[X_s^2]ds+2\mathbb{E}\left[\left(\sup_{0\leq s\leq T}|X_s|\right)\int_0^T|\mu_s^0|ds\right]+CI_0^2\end{aligned}$$

By Gronwall's inequality, we have

$$\mathbb{E}[X_t^2]\leq C\mathbb{E}\left[\left(\sup_{0\leq s\leq T}|X_s|\right)\int_0^T|\mu_s^0|ds\right]+CI_0^2.$$

We now use a version of Young's inequality that is very important: $2ab\leq\varepsilon a^2+\varepsilon^{-1}b^2$. This version gives us more freedom to choose which term should be more important. Hence

$$\mathbb{E}[X_t^2]\leq C\varepsilon\mathbb{E}\left[\sup_{0\leq s\leq T}|X_s|^2\right]+C\varepsilon^{-1}\mathbb{E}\left[\left(\int_0^T|\mu_s^0|ds\right)^2\right]+CI_0^2.$$

Notice that the integral term above is present in I_0^2 and that $C\varepsilon^{-1}$ is larger than C. Hence, plugging this in Equation (10.3), we find

$$\mathbb{E}\left[\sup_{0\leq t\leq T}|X_t|^2\right]\leq C\varepsilon\mathbb{E}\left[\sup_{0\leq s\leq T}|X_s|^2\right]+C\varepsilon^{-1}I_0^2$$

or

$$\mathbb{E}\left[\sup_{0\leq t\leq T}|X_t|^2\right]\leq\frac{C\varepsilon^{-1}}{(1-C\epsilon)}I_0^2=CI_0^2,$$

where ε is such that $1-C\varepsilon>0$. $\square$

How does the solution X depend on η, μ and σ? The next result answers this question.

Theorem 10.2.3. *Let $(\eta^{(i)},\mu^{(i)},\sigma^{(i)})$ be coefficients satisfying Assumption 10.2.1, for $i=1,2$ and consider $X^{(i)}$ a solution in $\mathcal{H}^2$ of SDE (10.1) with corresponding coefficients. Then*

$$\mathbb{E}\left[\sup_{0\leq t\leq T}|\Delta X_t|^2\right]\leq C\mathbb{E}\left[\Delta\eta^2+\left(\int_0^T|\Delta\mu_t(X_t^1)|dt\right)^2+\int_0^T|\Delta\sigma_t(X_t^1)|^2dt\right],$$

where Δ means the difference between $i=1$ and $i=2$, i.e. $\Delta X=X^{(1)}-X^{(2)}$.

Proof. Notice

$$\Delta X_t = \Delta\eta + \int_0^t (\Delta\mu_s(X_s^1) + \alpha_s \Delta X_s)ds + \int_0^t (\Delta\sigma_s(X_s^1) + \beta_s \Delta X_s)dB_s,$$

where

$$\alpha_s = \frac{\mu_t^{(2)}(X_t^{(1)}) - \mu_t^{(2)}(X_t^{(2)})}{\Delta X_t} 1_{\{\Delta X_t \neq 0\}} \text{ and } \beta_s = \frac{\sigma_t^{(2)}(X_t^{(1)}) - \sigma_t^{(2)}(X_t^{(2)})}{\Delta X_t} 1_{\{\Delta X_t \neq 0\}}.$$

are bounded by L. Hence, ΔX is a solution of the (linear) SDE above and the result follows from the previous theorem. □

The next corollary is a direct implication of the result above.

Corollary 10.2.4. *If (η, μ, σ) satisfy Assumption 10.2.1 and the SDE (10.1) has a solution, then it is unique.*

The main result of this chapter is the following theorem:

Theorem 10.2.5. *If (η, μ, σ) satisfy Assumption 10.2.1, then there exists a unique solution of SDE (10.1).*

Proof. Uniqueness was already stated in the corollary above. We will see two different arguments for the proof of existence.

Global argument: This follows the usual argument from ODE, the so-called Picard Iteration. Define $X_t^{(0)} = \eta$ and

$$X_t^{(n+1)} = \eta + \int_0^t \mu_s(X_s^{(n)})ds + \int_0^t \sigma_s(X_s^{(n)})dB_s. \tag{10.6}$$

One might prove, by induction, that $\mathbb{E}[\sup_{0\le t\le T}(X_t^{(n)})^2] < +\infty$. Moreover, defining $\Delta X^{(n)} = X^{(n)} - X^{(n-1)}$, one can easily conclude

$$\Delta X_t^{(n+1)} = \int_0^t \alpha_s^{(n)} \Delta X_s^{(n)} ds + \int_0^t \beta_s^{(n)} \Delta X_s^{(n)} dB_s,$$

where $\alpha^{(n)}$ and $\beta^{(n)}$ are defined in a similar fashion as in the proof of the theorem above and are bounded by L. Applying Itô formula for $e^{-\lambda t}(\Delta X_t^{(n+1)})^2$, we find

$$\begin{aligned}
\lambda \mathbb{E}\left[\int_0^T e^{-\lambda t}(\Delta X_t^{(n+1)})^2 dt\right] &\le \mathbb{E}\left[e^{-\lambda T}(\Delta X_T^{(n+1)})^2 + \lambda \int_0^T e^{-\lambda t}(\Delta X_t^{(n+1)})^2 dt\right] \\
&= \mathbb{E}\left[\int_0^T e^{-\lambda t}\left(2\Delta X_t^{(n+1)}\alpha_t^{(n)}\Delta X_t^{(n)} + (\beta_t^{(n)}\Delta X_t^{(n)})^2\right)dt\right] \\
&\le \mathbb{E}\left[\int_0^T e^{-\lambda t}\left((\Delta X_t^{(n+1)})^2 + 2L^2(\Delta X_t^{(n)})^2\right)dt\right]
\end{aligned}$$

where we have used Young's inequality. Define

$$D_n = \mathbb{E}\left[\int_0^T e^{-\lambda t}(\Delta X_t^{(n)})^2 dt\right]$$

and choose $\lambda = 1 + 8L^2$. Then

$$D_{n+1} \le \frac{1}{4} D_n \le \cdots \le \frac{C}{4^{n+1}}.$$

Hence, multiplying and dividing by $e^{-\lambda t}$:

$$\mathbb{E}\left[\int_0^T (\Delta X_t^{(n)})^2 dt\right] \le C D_n \le \frac{C}{4^n}.$$

This clearly implies that X^n is a Cauchy sequence with respect to the $\mathcal{H}^2$ norm. Completeness guarantees that existence of the limit of X^n. Letting $n \to +\infty$ in Equation (10.6) gives us the result.

Local argument: We start by squaring both sides of the dynamics of ΔX^{n+1} and taking expectation (using again Young's inequality):

$$\begin{aligned}
\mathbb{E}[(\Delta X_t^{(n+1)})^2] &= \mathbb{E}\left[2\int_0^t \alpha_s^{(n)}\Delta X_s^{(n)}\Delta X_s^{(n+1)}ds + \int_0^t (\beta_s^{(n)})^2(\Delta X_s^{(n)})^2 ds\right] \\
&\leq \mathbb{E}\left[\int_0^t 2L|\Delta X_s^{(n)}\Delta X_s^{(n+1)}|ds + L^2\int_0^t (\Delta X_s^{(n)})^2 ds\right] \\
&\leq \mathbb{E}\left[\int_0^t \left(2L^2T|\Delta X_s^{(n)}|^2 + \frac{1}{2T}|\Delta X_s^{(n+1)}|^2\right)ds + L^2\int_0^t (\Delta X_s^{(n)})^2 ds\right] \\
&\leq L^2(2T+1)\mathbb{E}\left[\int_0^T (\Delta X_s^{(n)})^2 ds\right] + \frac{1}{2T}\mathbb{E}\left[\int_0^T (\Delta X_s^{(n+1)})^2 ds\right].
\end{aligned}$$

Then

$$\mathbb{E}\left[\int_0^T (\Delta X_s^{(n+1)})^2 ds\right] \leq \frac{1}{2}L^2(2T+1)T\mathbb{E}\left[\int_0^T (\Delta X_s^{(n)})^2 ds\right].$$

Therefore, if T is small enough (i.e. such that $\frac{1}{2}L^2(2T+1)T < 1/4$), we can follow the same methods of the Global argument and conclude the existence of the solution. In the general case, we have to consider a discretization of the interval $[0,T]$ such that each sub-interval has size δ that satisfies $\frac{1}{2}L^2(2\delta+1)\delta < 1/4$. □

10.3 Properties of the SDEs

We start with a comparison result where the dimension of the process X must be 1.

Theorem 10.3.1 (Comparison Theorem). *Let $(\eta^{(i)}, \mu^{(i)}, \sigma)$ be coefficients satisfying Assumption 10.2.1, with $i = 1, 2$. Notice σ does not change with i. Let $X^{(i)}$ be the unique solution with coefficients $(\eta^{(i)}, \mu^{(i)}, \sigma)$. If $\eta^{(1)} \leq \eta^{(2)}$ and $\mu_t^{(1)}(x) \leq \mu_t^{(2)}(x)$ for all $x \in \mathbb{R}$, then $X_t^{(1)} \leq X_t^{(2)}$, for all $t \in [0,T]$.*

Proof. Notice that

$$\Delta X_t = \Delta\eta + \int_0^t (\Delta\mu_t(X_t^{(1)}) + \alpha_s\Delta X_s)ds + \int_0^t \beta_s\Delta X_s dB_s,$$

where α and β are bounded by L. This is a linear SDE for ΔX. Then

$$\Delta X_t = \frac{1}{\Phi_t}\left(\Delta\eta + \int_0^t \Phi_s\Delta\mu_s(X_s^{(1)})ds\right),$$

where Φ is given by Equation (10.2). Notice that $\Phi_t > 0$, $\Delta\eta \leq 0$ and $\Delta\mu_t(x) \leq 0$. Then $\Delta X_t \leq 0$, as desired. □

Theorem 10.3.2 (Stability Theorem). *Let $(\eta^{(n)}, \mu^{(n)}, \sigma^{(n)})$ and (η, μ, σ) be coefficients satisfying Assumption 10.2.1, with their correspondent unique solution being $X^{(n)}$ and X. Let $\Delta X^{(n)} = X^{(n)} - X$ (and similar for the other processes). If*

$$\lim_{n\to+\infty}\mathbb{E}\left[(\Delta\eta^{(n)})^2 + \left(\int_0^T |\Delta\mu_t^{(n)}(0)|dt\right) + \int_0^T (\Delta\sigma_t^{(n)}(0))^2 dt\right] = 0,$$

$\Delta\mu_t^{(n)}(x) \longrightarrow 0$ and $\Delta\sigma_t^{(n)}(x) \longrightarrow 0$, then

$$\lim_{n\to+\infty}\mathbb{E}\left[\sup_{0\leq t\leq T}(\Delta X_t^{(n)})^2\right] = 0.$$

Proof. Apply Theorem 10.2.3, sum and subtract $\Delta\mu_t^{(n)}(0)$ and $\Delta\sigma_t^{(n)}(0)$, use Lipschitz and finally the Dominated Covergence Theorem. □

10.4 Markov SDEs

Fix $t \in [0, T]$ and denote:

$$X_s = x + \int_0^s \mu(r, X_r)dr + \int_0^s \sigma(r, X_r)dB_r, \tag{10.7}$$

$$\mathcal{X}_s^{t,\eta} = \eta + \int_t^s \mu(r, \mathcal{X}_r^{t,\eta})dr + \int_t^s \sigma(r, \mathcal{X}_r^{t,\eta})dB_r, \tag{10.8}$$

$$X_s^{t,x} = x + \int_t^s \mu(r, X_r^{t,x})dr + \int_t^s \sigma(r, X_r^{t,x})dB_r, \tag{10.9}$$

where (η, μ, σ) satisfies Assumption 10.2.1 and η is $\mathcal{F}_t$-measurable. Moreover, by uniqueness of solutions, $X = X^{0,x}$, $\mathcal{X}^{t,x} = X^{t,x}$ and $X^{t,x} = \mathcal{X}^{r, X_r^{t,x}}$.

The next lemma follows directly from Theorem 10.2.3.

Lemma 10.4.1. *The function $x \mapsto X^{t,x}$ is Lipschitz continuous:*

$$\mathbb{E}\left[\sup_{t \le s \le T} |X_s^{t,x_1} - X_s^{t,x_2}|^2\right] \le C|x_1 - x_2|^2.$$

Define $\mathcal{F}_s^t = \sigma(B_r - B_t \ ; \ t \le r \le s)$.

Lemma 10.4.2. *There exists a version of $X^{t,x}$ for each x such that the function $(s, \omega) \mapsto X_s^{t,x}(\omega)$ is $\mathcal{F}^t$-progressively measurable. In particular, $X^{t,x}$ is independent of $\mathcal{F}_t$.*

Proof. The proof follows from convergence arguments using Picard iteration. □

Theorem 10.4.3. *Under the usual assumptions,*

$$\mathcal{X}_s^{t,\eta}(\omega) = X_s^{t,\eta(\omega)}(\omega).$$

This implies X is Markovian.

Proof. Let's prove first the result in the case where

$$\eta = \sum_{i=1}^{+\infty} x_i 1_{A_i},$$

where $A_i \in \mathcal{F}_t$ is a partition of Ω. The general result will follow from a limit argument. Define $\widetilde{X}$ as below and notice

$$\begin{aligned}
\widetilde{X}_s(\omega) &= \sum_{i=1}^{+\infty} X_s^{t,x_i}(\omega) 1_{A_i}(\omega) \\
&= \sum_{i=1}^{+\infty} \left(x_i + \int_t^s \mu(r, X_r^{t,x_i})dr + \int_t^s \sigma(r, X_r^{t,x_i})dB_r \right) 1_{A_i}(\omega) \\
&= \eta + \int_t^s \sum_{i=1}^{+\infty} \mu(r, X_r^{t,x_i}) 1_{A_i}(\omega)dr + \int_t^s \sum_{i=1}^{+\infty} \sigma(r, X_r^{t,x_i}) 1_{A_i}(\omega)dB_r \\
&= \eta + \int_t^s \mu(r, \widetilde{X}_r)dr + \int_t^s \sigma(r, \widetilde{X}_r)dB_r.
\end{aligned}$$

Hence, by uniqueness of solution, $\widetilde{X} = \mathcal{X}^{t,\eta}$. Now, if $\eta \in L^2(\mathcal{F}_t)$, there exists $\eta_n \in L^2(\mathcal{F}_t)$ such that η_n take only countable many values and $|\eta_n - \eta| \le 1/n$. By the previous argument,

$$\mathcal{X}_s^{t,\eta_n}(\omega) = X_s^{t,\eta_n(\omega)}(\omega).$$

If we denote $x = \eta(\omega)$ and $x_n = \eta_n(\omega)$, then

$$\mathbb{E}\left[\sup_{t \le s \le T} |X_s^{t,x} - X_s^{t,x_n}|^2 \;\middle|\; \mathcal{F}_t\right] = \mathbb{E}\left[\sup_{t \le s \le T} |X_s^{t,x} - X_s^{t,x_n}|^2\right] \le C|x - x_n|.$$

Then, by tower property, we find

$$\mathbb{E}\left[\sup_{t\le s\le T}|X_s^{t,\eta}-X_s^{t,\eta_n}|^2\right]\le C\mathbb{E}[|\eta-\eta_n|^2]\le\frac{C}{n^2}.$$

On the other hand, it is clear that

$$\mathbb{E}\left[\sup_{t\le s\le T}|\mathcal{X}_s^{t,\eta}-\mathcal{X}_s^{t,\eta_n}|^2\right]\le C\mathbb{E}[|\eta-\eta_n|^2]\le\frac{C}{n^2}.$$

Therefore,

$$\mathbb{E}\left[\sup_{t\le s\le T}|\mathcal{X}_s^{t,\eta}-X_s^{t,\eta}|^2\right]\le\frac{2C}{n^2}.$$

Since inequality above holds for every n, the letting $n\to+\infty$, we conclude the desired result.

Let us now prove that X is then a Markov process, i.e. we need to verify that for any ϕ measurable and bounded, there exists ψ such that

$$\mathbb{E}[\phi(X_s)\mid\mathcal{F}_t]=\psi(t,X_t).$$

Indeed, we have proved that

$$X_s(\omega)=\mathcal{X}_s^{t,X_t}(\omega)=X_s^{t,X_t(\omega)}(\omega).$$

Hence

$$\mathbb{E}[\phi(X_s)\mid\mathcal{F}_t]=\mathbb{E}[\phi(X_s^{t,X_t})\mid\mathcal{F}_t]=\psi(t,X_t),$$

by Proposition 2.1.12. □

Chapter 11

Backward Stochastic Differential Equations

11.1 Introduction

A Backward SDE is a stochastic differential equation where the final value is known (albeit random). Its study started during the 1970s in the stochastic optimal control theory. Naively, one could think that integration the usual SDE from t to T would give the desired object:

$$X_t = X_T - \int_t^T \mu_s(X_s)ds - \int_t^T \sigma_s(X_s)dB_s.$$

The problem is that the right-hand side is $\mathcal{F}_T$-measurable and the left-hand side must be $\mathcal{F}_t$-measurable. We cannot allow σ to be given. In fact, it is necessary to allow that the volatility process to be part of the solution. More precisely, we have

$$Y_t = \xi + \int_t^T f_s(Y_s, Z_s)ds - \int_t^T Z_s dB_s, \tag{11.1}$$

where $\xi \in L^2(\mathcal{F}_T)$. Notice that ξ is the final condition for Y: $Y_T = \xi$. The function f is called the generator of the BSDE. A solution for this BSDE is a pair (Y_t, Z_t), *adapted to* $\mathcal{F}$, and that satisfies the equation above.

It is important to notice two things: being adapted to $\mathcal{F}$ is the most important requirement. If we would drop it, the BSDE would have trivial solutions like $Y_t = \xi$ and $Z_t = 0$ (in the case of $f_s^0 = f_s(0,0) \equiv 0$). Moreover, if $\xi \in \mathcal{F}_0$, this previous argument would hold and the trivial solution would be valid. Therefore, it is paramount to have ξ measurable with respect to $\mathcal{F}_T$ and the pair (Y, Z) adapted to $\mathcal{F}$.

Notice as well that BSDE (11.1) can also be written as an SDE (with unknown initial value):

$$Y_t = Y_0 - \int_0^t f_s(Y_s, Z_s)ds + \int_0^t Z_s dB_s, \tag{11.2}$$

or

$$dY_t = -f_t(Y_t, Z_t)dt + Z_t dB_t.$$

The assumptions we will consider in the results of this chapter are listed below:

Assumption 11.1.1.

(i) f is progressively measurable in all variables;

(ii) f is uniformly Lipschitz in (y, z), i.e. there exists $L \geq 0$ such that

$$|f_t(y_1, z_1) - f_t(y_2, z_2)| \leq L(|y_1 - y_2| + |z_1 - z_2|),$$

for all y_1, y_2, z_1, z_2.

(iii) $\xi \in L^2(\mathcal{F}_T)$ and $f_t^0 = f_t(0,0)$ is sufficiently integrable.

We note that the assumptions above imply the linear growth of f:

$$|f_t(y,z)| \leq |f_t^0| + L(|y| + |z|).$$

11.2 Martingale Representation Theorem

The building block of the theory we will study in this chapter is the so-called Martingale Representation Theorem. This result answers the following question: is every martingale a stochastic integral with respect to the Brownian motion? More precisely, if M is a martingale with respect to $\mathcal{F}$, does it exist σ such that

$$M_t = M_0 + \int_0^t \phi_s dB_s,$$

for a $\mathcal{F}$-Brownian motion? We know that stochastic integrals are (local) martingales. In general, the answer to our quesiton is no as one can see in the following example.

Example 11.2.1. Let B and W be independent Brownian motions. Since both are martingales, assume, by contradiction, that there exists ϕ such that

$$W_t = \int_0^t \phi_s dB_s.$$

Notice that

$$dW_t^2 = d\left(W_t \int_0^t \phi_s dB_s\right) = W_t\phi_t dB_t + W_t dW_t + dW_t\phi_t dB_t = W_t\phi_t dB_t + W_t dW_t,$$

which is a (local) martingale. On the other hand, by Itô formula,

$$dW_t^2 = 2W_t dW_t + dt,$$

which cannot be a (local) martingale.

The main issue with the example above is that W is not adapted to the filtration *generated* by B, which we denote by $\mathcal{F}_t^B$.

Theorem 11.2.2. *Let $\xi \in L^2(\mathcal{F}_T^B)$, then there exists a unique $\phi \in \mathcal{H}^2$ adapted to $\mathcal{F}^B$ such that*

$$\xi = \mathbb{E}[\xi] + \int_0^T \phi_t dB_t.$$

Before proving, we notice the following important corollary (proof left as exercise):

Corollary 11.2.3 (Martingale Representation Theorem)**.** *If M is a square-integrable martingale adapted to $\mathcal{F}^B$, then there exists $\sigma \in \mathcal{H}^2$ adapted to $\mathcal{F}^B$ such that*

$$M_t = M_0 + \int_0^t \sigma_s dB_s.$$

Proof of Theorem 11.2.2. We start by proving uniqueness. Assume there exists $\phi \in \mathcal{H}^2$ adapted to $\mathcal{F}^B$ such that

$$\xi = \mathbb{E}[\xi] + \int_0^T \phi_t dB_t.$$

Then

$$\int_0^T (\phi_t - \sigma_t) dB_t = 0.$$

By Itô 's isometry we find that $\phi = \sigma$. To prove existence, we will proceed in several steps

Step 1: Assume $\xi = g(B_T)$, where g is bounded, Borel measurable function. Define $u(t,x) = \mathbb{E}[g(x + B_T - B_t)]$ and notice

$$u(t,x) = \int_{\mathbb{R}} g(z)p(T-t, x-z)dz,$$

where $p(t,x)$ is the Gaussian density with mean 0 and variance t. One can easily check that

$$\frac{\partial p}{\partial t} - \frac{1}{2}\frac{\partial^2 p}{\partial x^2} = 0.$$

This implies that $u \in C^{1,2}$ is bounded with bounded derivatives and satisfy

$$\frac{\partial u}{\partial t} + \frac{1}{2}\frac{\partial^2 u}{\partial x^2} = 0,$$

with $u(T,x) = g(x)$. Hence, if we define $X_t = u(t, B_t)$, one can easily check that, by Itô formula,

$$X_T = X_0 + \int_0^T \frac{\partial u}{\partial x}(t, B_t)dB_t,$$

which is the martingale representation in this case.

Step 2: Assume $\xi = g(B_{t_1}, \dots, B_{t_n})$, with $0 = t_1 < \cdots < t_n = T$. Apply Step 2 in each sub-interval $[t_{i-1}, t_i]$. More precisely, define $g_n(x_1, \dots, x_n) = g(x_1, \dots, x_n)$. By Step 2, there exists σ^n in $[t_{n-1}, t_n]$ such that

$$g_n(B_{t_1}, \dots, B_{t_n}) = \mathbb{E}[g(B_{t_1}, \dots, B_{t_n}) \mid \mathcal{F}^B_{t_{n-1}}] + \int_{t_{n-1}}^{t_n} \phi_s^{(n)} dB_s.$$

Define

$$g_{n-1}(x_1, \dots, x_{n-1}) = \mathbb{E}[g(x_1, \dots, x_{n-1}, x_{n-1} + B_{t_n} - B_{t_{n-1}})].$$

By backward induction, it is easily seen that

$$g_i(B_{t_1}, \dots, B_{t_i}) = g_{i-1}(B_{t_1}, \dots, B_{t_{i-1}}) + \int_{t_{i-1}}^{t_i} \phi_s^{(i)} dB_s,$$

with

$$g_{i-1}(x_1, \dots, x_{i-1}) = \mathbb{E}[g_i(x_1, \dots, x_{i-1}, x_{i-1} + B_{t_i} - B_{t_{i-1}})].$$

Hence, it is clear that the martingale representation for ξ works with

$$\sigma_s = \sum_{i=1}^n \phi_s^{(i)} 1_{[t_{i-1}, t_i]}(s).$$

Step 3: Assume now that $\xi \in L^\infty(\mathcal{F}^B_T)$. Define $\mathcal{F}^{(n)}_T$ the sigma-algebra generated by $B_{t_0^{(n)}}, \dots, B_{t_n^{(n)}}$, with $t_i^{(n)} = iT/n$, and consider $\xi_n = \mathbb{E}[\xi \mid \mathcal{F}^{(n)}_T]$. Thus, there exists g_n Borel measurable and bounded such that $\xi_n = g_n(B_{t_0^{(n)}}, \dots, B_{t_n^{(n)}})$. By Step 3, there exists the martingale representation for ξ_n. Noticing that $\mathcal{F}^B_T$ is generated by $\cup_{n=1}^{+\infty} \mathcal{F}^{(n)}_T$ (by the continuity of B), one might conclude the martingale representation for ξ.

Step 4: The general case where $\xi \in L^2(\mathcal{F}^B_T)$, just consider ξ_n the truncation of ξ in the interval $[-n, n]$. □

From now on the filtration we will consider will be the Brownian filtration and we will denote it by $\mathcal{F}$.

11.2.1 Application: Stochastic Filtering

Let us assume that

$$dX_t = \mu_Z dt + \sigma dB_t,$$

where $Z \in \{z_1, z_2\}$ is a random variable, independent of B. Consider the prior probability of z_1 being $p_0 = \mathbb{E}[1_{\{Z=z_1\}} \mid \mathcal{F}_0]$ and posterior probability

$$p_t = \mathbb{E}[1_{\{Z=z_1\}} \mid \mathcal{F}_t],$$

where $\mathcal{F}_t$ is the filtration generated by X. Define then the innovation process

$$W_t = B_t + \frac{\mu_Z - \bar{\mu}(t)}{\sigma} t,$$

where $\bar{\mu}(t) = p_t\mu_{z_1} + (1-p_t))\mu_{z_2}$. Hence

$$dX_t = \bar{\mu}(t)dt + \sigma dW_t.$$

As we will see in the Girsanov theory chapter, there exists a probability measure $\mathbb{Q}$ such that W is a Brownian motion. Moreover, notice that $(p_t)_{t\geq 0}$ is a martingale. By the Martingale Representation Theorem, we have

$$p_t = p_0 + \int_0^t \phi_s dW_s.$$

In order to find an explicit expression for ϕ, let $(\beta_t)_{t\geq 0}$ be any adapted process in $\mathcal{H}^2$. Then

$$\mathbb{E}\left[\int_0^t \beta_s dW_s (p_t - p_0)\right] = \mathbb{E}\left[\int_0^t \beta_s\phi_s ds\right].$$

On the other hand, we find

$$\begin{aligned}
\mathbb{E}\left[\int_0^t \beta_s dW_s (p_t - p_0)\right] &= \mathbb{E}\left[\int_0^t \beta_s dW_s p_t\right] = \mathbb{E}\left[\int_0^t \beta_s dW_s \mathbb{E}[1_{\{Z=z_1\}} \mid \mathcal{F}_t]\right] \\
&= \mathbb{E}\left[\int_0^t \beta_s dW_s 1_{\{Z=z_1\}}\right] = \mathbb{E}\left[\int_0^t \beta_s\left(dB_s + \frac{\mu_Z - \bar{\mu}(t)}{\sigma}ds\right)1_{\{Z=z_1\}}\right] \\
&= \mathbb{E}\left[\int_0^t \beta_s\left(dB_s + \frac{\mu_{z_1} - \bar{\mu}(t)}{\sigma}ds\right)1_{\{Z=z_1\}}\right] \\
&= \mathbb{E}\left[\int_0^t \beta_s\left(dB_s + (1-p_s)\frac{\mu_{z_1} - \mu_{z_2}}{\sigma}ds\right)1_{\{Z=z_1\}}\right] \\
&= \mathbb{E}\left[\int_0^t \beta_s dB_s 1_{\{Z=z_1\}}\right] + \mathbb{E}\left[\int_0^t \beta_s(1-p_s)\frac{\mu_{z_1} - \mu_{z_2}}{\sigma}ds 1_{\{Z=z_1\}}\right] \\
&= \mathbb{E}\left[\int_0^t \beta_s dB_s\right]\mathbb{E}\left[1_{\{Z=z_1\}}\right] + \int_0^t \mathbb{E}\left[\beta_s(1-p_s)\frac{\mu_{z_1} - \mu_{z_2}}{\sigma}1_{\{Z=z_1\}}\right]ds \\
&= \int_0^t \mathbb{E}\left[\beta_s(1-p_s)\frac{\mu_{z_1} - \mu_{z_2}}{\sigma}p_s\right]ds.
\end{aligned}$$

Since this is true for any β, we find that

$$\phi_t = \frac{\mu_{z_1} - \mu_{z_2}}{\sigma}p_t(1-p_t).$$

Therefore, we have found the dynamics for p:

$$dp_t = \frac{\mu_{z_1} - \mu_{z_2}}{\sigma}p_t(1-p_t)dW_t.$$

11.3 Linear BSDE

We start by studying the simplest BSDE:

$$Y_t = \xi + \int_t^T f_s^0 ds - \int_t^T Z_s dB_s, \tag{11.3}$$

where $\xi \in L^2(\mathcal{F}_T)$ and f_s^0 is adapted to $\mathcal{F}$.

To construct a solution for the Linear BSDE, consider the following:

$$Y_t = \mathbb{E}\left[\xi + \int_t^T f_s^0 ds \,\middle|\, \mathcal{F}_t\right].$$

Y is not a martingale, but the process M below is:

$$M_t = Y_t + \int_0^t f_s^0 ds = \mathbb{E}\left[\xi + \int_0^T f_s^0 ds \,\middle|\, \mathcal{F}_t\right].$$

Hence, by assuming integrability of ξ and f^0, by the Martingale Representation Theorem, there exists a process $Z \in \mathcal{H}^2$ such that

$$M_t = M_0 + \int_0^t Z_s dB_s.$$

Equivalently to the equation above, we have

$$M_T = M_t + \int_t^T Z_s dB_s \Leftrightarrow \xi + \int_0^T f_s^0 ds = Y_t + \int_0^t f_s^0 ds + \int_t^T Z_s dB_s.$$

Hence, Equation (11.3) is satisfied.

Let us now consider the general *linear* BSDE:

$$Y_t = \xi + \int_t^T (\alpha_s Y_s + \beta_s Z_s + f_s^0) ds - \int_t^T Z_s dB_s, \tag{11.4}$$

To find the solution of this equation, consider the familiar process

$$d\Phi_t = \alpha_t \Phi_t dt + \beta_t \Phi_t dB_t. \tag{11.5}$$

Moreover, as we have seen, the dynamics of Y are given by $dY_t = -(\alpha_t Y_t + \beta_t Z_t + f_t^0)dt + Z_t dB_t$ and then

$$d(\Phi_t Y_t) = -\Phi_t f_t^0 dt + \Phi_t (Y_t \beta_t + Z_t) dB_t.$$

This BSDE is in the form of (11.3) with Z given by $\Phi_t(Y_t\beta_t + Z_t)$. Therefore there exists a unique solution given by

$$Y_t = \frac{1}{\Phi_t} \mathbb{E}\left[\Phi_T \xi + \int_t^T \Phi_s f_s^0 ds \,\middle|\, \mathcal{F}_t \right]. \tag{11.6}$$

11.4 Existence and Uniqueness

We start this section by defining the following norm:

$$\|(Y, Z)\|^2 = \mathbb{E}\left[\sup_{0 \le t \le T} |Y_t|^2 + \int_0^T |Z_t|^2 dt \right].$$

Theorem 11.4.1. *Let (ξ, f) satisfy Assumption 11.1.1 and let (Y, Z) be a solution in $\mathcal{H}^2 \times \mathcal{H}^2$ of the BSDE (11.1). Then, there exists a constant $C > 0$ depending only on T and L such that*

$$\|(Y, Z)\|^2 \le C \; \mathbb{E}\left[|\xi|^2 + \left(\int_0^T |f_t^0| dt \right)^2 \right].$$

Proof. This proof follows the same reasoning as in the SDE case. First, notice

$$|Y_t| \le |\xi| + \int_t^T (|f_s^0| + L(|Y_s| + |Z_s|)) ds + \left| \int_t^T Z_s dB_s \right|.$$

Similar to the SDE case, we conclude

$$\mathbb{E}\left[\sup_{0 \le t \le T} |Y_t|^2 \right] \le C I_0^2 + C \mathbb{E}\left[\int_0^T (Y_t^2 + Z_t^2) dt \right],$$

where I_0^2 is the right-hand bound. Moreover, by Itô's formula,

$$dY_t^2 = 2Y_t dY_t + (dY_t)^2 = -2Y_t f_t(Y_t, Z_t) dt + 2Y_t Z_t dB_t + Z_t^2 dt.$$

Then

$$Y_t^2 + \int_t^T Z_s^2 ds = \xi^2 + 2 \int_t^T Y_s f_s(Y_s, Z_s) ds - 2 \int_t^T Y_s Z_s dB_s.$$

Taking expectation, we find

$$\begin{aligned}\mathbb{E}\left[Y_t^2+\int_t^T Z_s^2 ds\right] &= \mathbb{E}\left[\xi^2+2\int_t^T Y_s f_s(Y_s,Z_s)ds\right] \leq \\ &\leq \mathbb{E}\left[\xi^2+C\int_t^T |Y_s|(|f_s^0+|Y_s|+|Z_s|)ds\right] \\ &\leq \mathbb{E}\left[\xi^2+C\left(\sup_{0\leq t\leq T}|Y_t|\right)\int_t^T |f_s^0|ds+C\int_t^T(|Y_s|^2+|Y_sZ_s|)ds\right] \\ &\leq \mathbb{E}\left[\xi^2+C\left(\sup_{0\leq t\leq T}|Y_t|\right)\int_t^T |f_s^0|ds+C\int_t^T|Y_s|^2ds+\frac{1}{2}\int_t^T|Z_s|ds\right],\end{aligned}$$

where we have used Young's inequality once again. Hence

$$\mathbb{E}\left[Y_t^2+\frac{1}{2}\int_t^T Z_s^2 ds\right] \leq \mathbb{E}\left[\xi^2+C\left(\sup_{0\leq t\leq T}|Y_t|\right)\int_0^T |f_s^0|ds+C\int_t^T|Y_s|^2ds\right], \tag{11.7}$$

and thus

$$\mathbb{E}\left[Y_t^2\right] \leq \mathbb{E}\left[\xi^2+C\left(\sup_{0\leq t\leq T}|Y_t|\right)\int_0^T |f_s^0|ds\right]+C\int_t^T \mathbb{E}[|Y_s|^2]ds.$$

By Gronwall inequality, we have

$$\mathbb{E}\left[Y_t^2\right] \leq C\mathbb{E}\left[\xi^2+C\left(\sup_{0\leq t\leq T}|Y_t|\right)\int_0^T |f_s^0|ds\right],$$

which implies

$$\sup_{0\leq t\leq T}\mathbb{E}\left[Y_t^2\right] \leq C\mathbb{E}\left[\xi^2+C\left(\sup_{0\leq t\leq T}|Y_t|\right)\int_0^T |f_s^0|ds\right],$$

Using the equation above and the inequality in (11.7), we have

$$\mathbb{E}\left[\int_0^T |Z_s|^2ds\right] \leq C\mathbb{E}\left[\xi^2+C\left(\sup_{0\leq t\leq T}|Y_t|\right)\int_0^T |f_s^0|ds\right].$$

By the generalized Young inequality ($2ab\leq\varepsilon a^2+\varepsilon^{-1}b^2$), we conclude

$$\sup_{0\leq t\leq T}\mathbb{E}\left[Y_t^2\right]+\mathbb{E}\left[\int_0^T |Z_s|^2ds\right] \leq C\varepsilon\mathbb{E}\left[\sup_{0\leq t\leq T}|Y_t|^2\right]+C\varepsilon^{-1}I_0^2. \tag{11.8}$$

Therefore,

$$\mathbb{E}\left[\sup_{0\leq t\leq T}|Y_t|^2\right] \leq C\varepsilon\mathbb{E}\left[\sup_{0\leq t\leq T}|Y_t|^2\right]+C\varepsilon^{-1}I_0^2.$$

Choosing $\varepsilon=1/(2C)$, we find

$$\mathbb{E}\left[\sup_{0\leq t\leq T}|Y_t|^2\right] \leq CI_0^2.$$

Using Equation (11.8) completes the proof. □

Theorem 11.4.2. *Let (ξ^i,f^i) satisfy Assumption 11.1.1 and let (Y^i,Z^i) be a solution in $\mathcal{H}^2\times\mathcal{H}^2$ of the BSDE with coefficients (ξ^i,f^i), for $i=1,2$. Then, there exists a constant $C>0$ depending only on T and L such that*

$$\|(\Delta Y,\Delta Z)\|^2 \leq C\ \mathbb{E}\left[|\Delta\xi|^2+\left(\int_0^T|\Delta f_t(Y_t^1,Z_t^1)|dt\right)^2\right],$$

where $\Delta Y=Y^1-Y^2$ and similarly for the other variables.

Proof. Similarly to what we have done in the SDE case, we have

$$\begin{aligned}\Delta Y_t &= \Delta\xi + \int_t^T (f_s^1(Y_s^1, Z_s^1) - f_s^2(Y_s^2, Z_s^2))ds - \int_t^T \Delta Z_s dB_s \\ &= \Delta\xi + \int_t^T (\Delta f_s^1(Y_s^1, Z_s^1) + \alpha_s \Delta Y_s + \beta_s \Delta Z_s)ds - \int_t^T \Delta Z_s dB_s,\end{aligned}$$

where

$$\alpha_s = \frac{f_t^2(Y_t^1, Z_t^1) - f_t^2(Y_t^2, Z_t^1)}{\Delta Y_t} 1_{\{\Delta Y_t \neq 0\}} \text{ and } \beta_s = \frac{f_t^2(Y_t^2, Z_t^1) - f_t^2(Y_t^2, Z_t^2)}{\Delta Z_t} 1_{\{\Delta Z_t \neq 0\}}$$

are bounded by L. The result then follows from previous theorem. □

As before, uniqueness follows directly from the result above.

Corollary 11.4.3. *If (ξ, f) satisfy Assumption 11.1.1 and the BSDE (11.1) has a solution, then it is unique.*

Theorem 11.4.4. *If (ξ, f) satisfy Assumption 11.1.1, then there exists a unique solution of BSDE (11.1).*

Proof. We will proceed using the local approach from the proof in the SDE case. Define $Y_t^0 \xi$ and $Z_t^0 = 0$ and

$$Y_t^n = \xi + \int_t^T f_s(Y_s^{n-1}, Z_s^{n-1})ds - \int_t^T Z_s^n dB_s.$$

Define now $\Delta Y^n = Y^n - Y^{n-1}$, $\Delta Z^n = Z^n - Z^{n-1}$ and note that

$$\Delta Y_t^n = \int_t^T (\alpha_s^{n-1} \Delta Y_s^{n-1} + \beta_s^{n-1} \Delta Z_s^{n-1})ds - \int_t^T \Delta Z_s^n dB_s,$$

where α^n and β^n are bounded by L. Applying Itô formula for $(\Delta Y^n)^2$ and taking expectation, we find

$$\begin{aligned}\mathbb{E}\left[(\Delta Y_t^n)^2 + \int_t^T (\Delta Z_s^n)^2 ds\right] &= 2\mathbb{E}\left[\int_t^T \Delta Y_s^n (\alpha_s^{n-1} \Delta Y_s^{n-1} + \beta_s^{n-1} \Delta Z_s^{n-1})ds\right] \\ &\leq 2L\mathbb{E}\left[\int_0^T |\Delta Y_s^n|(|\Delta Y_s^{n-1}| + |\Delta Z_s^{n-1}|)ds\right].\end{aligned} \tag{11.9}$$

Integrating from 0 to T, we have

$$\begin{aligned}\mathbb{E}\left[\int_0^T (\Delta Y_t^n)^2 dt\right] &\leq 2LT\mathbb{E}\left[\int_0^T |\Delta Y_s^n|(|\Delta Y_s^{n-1}| + |\Delta Z_s^{n-1}|)ds\right] \\ &\leq 2LT\mathbb{E}\left[\int_0^T (|\Delta Y_s^n|^2 + \frac{1}{2}|\Delta Y_s^{n-1}|^2 + \frac{1}{2}|\Delta Z_s^{n-1}|^2)ds\right].\end{aligned}$$

Thus, if T is such that $1 - 2LT > 0$, we have

$$\mathbb{E}\left[\int_0^T (\Delta Y_t^n)^2 dt\right] \leq \frac{LT}{1-2LT}\mathbb{E}\left[\int_0^T (|\Delta Y_s^{n-1}|^2 + |\Delta Z_s^{n-1}|^2)ds\right].$$

Furthermore, set $t = 0$ in Equation (11.9):

$$\begin{aligned}\mathbb{E}\left[\int_0^T (\Delta Z_s^n)^2 ds\right] &\leq 2\mathbb{E}\left[\int_0^T \Delta Y_s^n (\alpha_s^{n-1} \Delta Y_s^{n-1} + \beta_s^{n-1} \Delta Z_s^{n-1})ds\right] \\ &\leq \mathbb{E}\left[\int_0^T \left(8|\alpha_s^{n-1} \Delta Y_s^n|^2 + \frac{1}{8}|\Delta Y_s^{n-1}|^2 + 8|\beta_s^{n-1} \Delta Y_s^n|^2 + \frac{1}{8}|\Delta Z_s^{n-1}|^2\right) ds\right] \\ &\leq 16L^2\mathbb{E}\left[\int_0^T |\Delta Y_s^n|^2 ds\right] + \frac{1}{8}\mathbb{E}\left[\int_0^T \left(|\Delta Y_s^{n-1}|^2 + |\Delta Z_s^{n-1}|^2\right) ds\right]\end{aligned}$$

$$\leq \left(\frac{16L^3T}{1-2LT}+\frac{1}{8}\right)\mathbb{E}\left[\int_0^T \left(|\Delta Y_s^{n-1}|^2+|\Delta Z_s^{n-1}|^2\right)ds\right]$$

Hence, choosing T small enough so that $\dfrac{16L^3T}{1-2LT}<\dfrac{1}{8}$, we conclude

$$\mathbb{E}\left[\int_0^T ((\Delta Y_t^n)^2+(\Delta Z_t^n)^2)dt\right]\leq C\mathbb{E}\left[\int_0^T (|\Delta Y_t^{n-1}|^2+|\Delta Z_t^{n-1}|^2)dt\right],$$

with $C<1/4$. Following the same arguments in the SDE case, there exists the limit (Y,Z) in the norm $\|(\cdot,\cdot)\|$ if T is small enough. Set this value for T to be δ. In the general case, divide the interval $[0,T]$ in sub-intervals with mesh size δ and use the argument above in each sub-interval. It is important that δ does not depend on the final condition ξ in the argument above. □

11.5 Properties of the BSDEs

Theorem 11.5.1 (Comparison Theorem). *Assume (ξ^i,f^i) satisfies Assumptions 11.1.1 and let (Y_i,Z_i) be the unique solution to the respective BSDE, for $i=1,2$. If $\xi^1\leq\xi^2$ and $f^1\leq f^2$, then $Y^1\leq Y^2$.*

Proof. Define $\Delta Y=Y^1-Y^2$ (similar to other variables). As we have seen

$$\Delta Y_t=\Delta\xi+\int_t^T(\alpha_s\Delta Y_s+\beta_s\Delta Z_s+\Delta f_s(Y_s^2,Z_s^2))ds-\int_t^T\Delta Z_s dB_s.$$

This is a linear BSDE for $(\Delta Y,\Delta Z)$. By formula (11.6), we find

$$\Delta Y_t=\frac{1}{\Phi_t}\mathbb{E}\left[\Phi_T\Delta\xi+\int_t^T\Phi_s\Delta f_s(Y_s^2,Z_s^2)ds\ \middle|\ \mathcal{F}_t\right],$$

thus the result follows. □

Theorem 11.5.2 (Stability Theorem). *Assume (ξ^n,f^n) and (ξ,f) satisfy Assumptions 11.1.1 and let (Y^n,Z^n) and (Y,Z) be the unique solution to the respective BSDE. Denote $\Delta Y^n=Y^n-Y$ (similar to the other variables). If*

$$\lim_{n\to+\infty}\mathbb{E}\left[(\Delta\xi^n)^2+\left(\int_0^T|\Delta f_t^n(0,0)|dt\right)^2\right]=0$$

and $\Delta f^n\longrightarrow 0$, then

$$\lim_{n\to+\infty}\|(\Delta Y^n,\Delta Z^n)\|=0.$$

Proof. Apply Theorem 11.4.1, sum and subtract $\Delta f_t^n(0)$, use Lipschitz and finally the Dominated Convergence Theorem. □

11.6 Markov BSDEs

We firstly consider the Forward-Backward SDE (FBSDE):

$$X_t=x+\int_0^t\mu(s,X_s)ds+\int_0^T\sigma(s,X_s)dB_s, \tag{11.10}$$

$$Y_t=g(X_T)+\int_t^T f(s,X_s,Y_s,Z_s)ds-\int_t^T Z_s dB_s. \tag{11.11}$$

Notice that μ, σ and f are deterministic (not stochastic processes as in the section before). Under the following assumptions, the system above is well-posed (it has a unique solution and it depends continuously on its parameters):

Assumption 11.6.1.

(i) $\mu(\cdot,0)$, $\sigma(\cdot,0)$ and $f(\cdot,0,0,0)$ are bounded;

(ii) μ,σ,f and g are uniformly Lipschitz in (x,y,z);

Following the arguments from Section 10.4, we define

$$Y_s^{t,x} = g(X_T^{t,x}) - \int_t^T f(s, X_s^{t,x}, Y_s^{t,x}, Z_s^{t,x})ds + \int_t^T Z_s^{t,x} dB_s,$$
$$\mathcal{Y}_s^{t,\eta} = g(\mathcal{X}_T^{t,\eta}) - \int_t^T f(s, \mathcal{X}_s^{t,\eta}, \mathcal{Y}_s^{t,\eta}, \mathcal{Z}_s^{t,\eta})ds + \int_t^T \mathcal{Z}_s^{t,\eta} dB_s.$$

Define $\mathcal{F}^t = \sigma(B_r - B_t \ ; \ t \leq r \leq T)$. The next result is proved using similar techniques as in Section 10.4:

Theorem 11.6.2. *Under Assumption 11.6.1, we have*

(i) There exists a version of $(Y^{t,x}, Z^{t,x})$ for each x the function $(s,\omega) \mapsto (Y_s^{t,x}(\omega), Z_s^{t,x}(\omega))$ is $\mathcal{F}^t$-progressively measurable. In particular, $(Y^{t,x}, Z^{t,x})$ is independent of $\mathcal{F}_t$.

(ii) For any η $\mathcal{F}_t$-measurable,

$$(\mathcal{Y}_s^{t,\eta}(\omega), \mathcal{Z}_s^{t,\eta}(\omega)) = (Y_s^{t,\eta(\omega)}(\omega), Z_s^{t,\eta(\omega)}(\omega)).$$

(iii) (X,Y,Z) is Markovian.

Therefore, define

$$u(t,x) = Y_t^{t,x}. \tag{11.12}$$

The function u is possibly random. However, since $Y_t^{t,x}$ is at the same time measurable with respect to $\mathcal{F}_t$ and independent of $\mathcal{F}_t$, we conclude that $u(t,x)$ is deterministic. Additionally, because $Y_t = Y_t^{0,x} = \mathcal{Y}_t^{t,X_t} = Y_t^{t,X_t}$, we finally find

$$u(t,X_t) = Y_t.$$

11.7 Quadratic Generator

As of now, we have considered BSDEs with generator that are Lipschitz continuous in (y,z). It is possible to consider more general generator. For instance, let's analyze the following case:

$$dY_t = -\frac{1}{2} Z_t^2 dt + Z_t dB_t.$$

It makes sense to consider:

$$de^{Y_t} = e^{Y_t} Z_t dB_t,$$

which gives us

$$e^{\xi} - e^{Y_t} = \int_t^T e^{Y_s} Z_s dB_s.$$

Hence, assuming the stochastic integral above is a true martingale, we have

$$Y_t = \log(\mathbb{E}[e^{\xi} \mid \mathcal{F}_t]).$$

This shows us that square-integrability of ξ is not sufficient anymore. The next theorem (that we state without proof) is a version of the existence and uniqueness theorem for the quadratic case:

Theorem 11.7.1 (Kobylanski, AoP, 2000)**.** *Assume ξ is bounded and there exists $C > 0$ such that, for all (y,z),*

$$|f_t(y,z)| \leq C + C|z|^2,$$
$$\left| \frac{\partial f_t}{\partial z}(y,z) \right| \leq C + C|z|,$$
$$\frac{\partial f_t}{\partial y}(y,z) \leq C + C|z|^2.$$

Then the BSDE with generator f and final condition ξ has a unique solution (Y,Z), with Y bounded and Z square-integrable.

Chapter 12

Girsanov Theory

12.1 Introduction

The following very simple example shows the importance of the result we will study in the sections to follow. Consider $X \sim N(0,1)$ and suppose we want to compute $\mathbb{E}[f(X)]$. The Monte Carlo approach is given by

$$\mathbb{E}[f(X)] \approx \frac{1}{n}\sum_{i=1}^{n} f(X_i),$$

where $X_i \overset{iid}{\sim} N(0,1)$. This approach works very well, but some cases might be computationally challenging. Suppose, for instance, $f(x) = 1_{\{x>30\}}$. In this case, very few simulations of X will be bigger than 30. In fact, one can show that the if N denote the number of draws until first non-zero $f(X_i)$, then $\mathbb{E}[N] > 10^{100}$. However, there is a trick, called importance sampling. Notice that

$$\begin{aligned}\mathbb{E}[f(X)] &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} f(x)e^{-x^2/2}dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} f(x)e^{-(x-\mu)^2/2}e^{+\mu^2/2-\mu x}dx \\ &= \mathbb{E}_\mu[f(X)e^{-\mu X+\mu^2/2}],\end{aligned}$$

where $\mathbb{E}_\mu$ is the expectation assuming $X \sim N(\mu,1)$. Therefore

$$\mathbb{E}[f(X)] \approx \frac{1}{n}\sum_{i=1}^{n} g(Y_i),$$

where $Y_i \overset{iid}{\approx} N(\mu,1)$ and $g(x) = f(x)e^{-\mu^2/2-\mu x}$. In the example we were considering, we should take $\mu = 30$.

Consider now the process $X_t = B_t + \mu t$. Notice that X also has independent increments. Moreover, we can always consider the increments transformation:

$$\mathbb{E}[f(X_{t_1},\ldots,X_{t_n})] = \mathbb{E}[g(X_{t_1}, X_{t_2}-X_{t_1},\ldots,X_{t_n}-X_{t_{n-1}})],$$

where $g(y_1,y_2,\ldots,y_n) = f(y_1,y_1+y_2,\ldots,y_1+y_2+\cdots+y_n)$. Moreover, we know that $(X_{t_1}, X_{t_2} - X_{t_1},\ldots,X_{t_n}-X_{t_{n-1}})$ is a multivariate Gaussian with mean $\mu(t_i - t_{i-1})$ and covariance $(t_i-t_{i-1})1_{\{i=j\}}$, with $t_0 = 0$. Using $x_0 = 0$, we can analyze the probability density of this random vector as:

$$\begin{aligned}&\prod_{i=1}^{n}\exp\left\{-\frac{((x_i-x_{i-1})-\mu(t_i-t_{i-1}))^2}{2(t_i-t_{i-1})}\right\} \\ &= \prod_{i=1}^{n}\exp\left\{-\frac{((x_i-x_{i-1})^2}{2(t_i-t_{i-1})}+\mu(x_i-x_{i-1})-\frac{1}{2}\mu^2(t_i-t_{i-1})\right\} \\ &= \left(\prod_{i=1}^{n}\exp\left\{-\frac{((x_i-x_{i-1})^2}{2(t_i-t_{i-1})}\right\}\right)e^{\mu x_n-\frac{1}{2}\mu^2 t_n}.\end{aligned}$$

Notice that the product term is the density of the increments of the Brownian motion. Therefore, we conclude

$$\mathbb{E}[f(X_{t_1},\dots,X_{t_n})] = \mathbb{E}[f(B_{t_1},\dots,B_{t_n})e^{\mu B_{t_n}-\mu^2 t_n/2}] = \mathbb{E}[f(B_{t_1},\dots,B_{t_n})M_{t_n}].$$

We have found the well-known martingale $M_t = e^{\mu B_t - \mu^2 t/2}$.

We will now analyze the maximum of a Brownian motion with drift using the formula above. Define $S(n) = \{iT/n \ ; \ 0 \le i \le n\}$ and notice that, by the continuity of the Brownian motion,

$$\max_{t\in[0,T]} B_t + \mu t = \lim_{n\to+\infty} \max_{t\in S(n)} B_t + \mu t.$$

By the computation we have performed above, we have

$$\mathbb{P}\left(\max_{t\in S(n)} B_t + \mu t < x\right) = \mathbb{E}\left[1_{\{\max_{t\in S(n)} B_t < x\}} M_T\right].$$

Hence, taking $n \to +\infty$, we find

$$\mathbb{P}\left(\max_{t\in[0,T]} B_t + \mu t < x\right) = \mathbb{E}\left[1_{\{B_T^* < x\}} M_T\right].$$

Therefore, we may analyze the hitting time of a line for the Brownian motion. Define

$$\tau_L = \min\{t \ ; \ B_t = a + bt\}.$$

Then

$$\tau_L > t \Leftrightarrow B_s - bs < a, \text{ for all } s \in [0,t).$$

Hence

$$\mathbb{P}(\tau_L > t) = \mathbb{P}\left(\max_{s\in[0,t]} B_s - bs < a\right) = \mathbb{E}\left[1_{\{B_t^* < a\}} e^{-bB_t - \frac{1}{2}b^2 t}\right].$$

In order to compute the density of τ_L one just needs to use the joint density of (B_t, B_t^*).

12.2 Change of Measure and Martingales

Let $(M_t)_{t\in[0,T]}$ be a martingale with $M_t > 0$ and $\mathbb{E}[M_T] = M_0 = 1$. We therefore can define a probability measure $\mathbb{P}_M$ that is equivalent to $\mathbb{P}$ as:

$$\mathbb{P}_M(A) = \mathbb{E}[M_T 1_A], \text{ for } A \in \mathcal{F}_T.$$

The proof of the next lemma is straightforward.

Lemma 12.2.1. *$\xi \in L^1(\mathbb{P}_M)$ if and only if $\xi M_T \in L^1(\mathbb{P})$. Moreover,*

$$\mathbb{E}_{\mathbb{P}_M}[\xi] = \mathbb{E}[M_T \xi].$$

We can also analyze conditional expectation when we change measure.

Lemma 12.2.2. *If $\xi \in L^1(\mathbb{P})$, then*

$$\mathbb{E}_{\mathbb{P}_M}[\xi \mid \mathcal{F}_t] = \frac{1}{M_t}\mathbb{E}[M_T \xi \mid \mathcal{F}_t].$$

Proof. Let $A \in \mathcal{F}_t$ and notice that, since M is a $\mathbb{P}$-martingale,

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_M}[(M_t)^{-1}\mathbb{E}[M_T\xi \mid \mathcal{F}_t]1_A] &= \mathbb{E}[M_T(M_t)^{-1}\mathbb{E}[M_T\xi \mid \mathcal{F}_t]1_A] \\ &= \mathbb{E}[M_t(M_t)^{-1}\mathbb{E}[M_T\xi \mid \mathcal{F}_t]1_A] \\ &= \mathbb{E}[\mathbb{E}[M_T\xi \mid \mathcal{F}_t]1_A] = \mathbb{E}[M_T\xi 1_A] = \mathbb{E}_{\mathbb{P}_M}[\xi 1_A] \end{aligned}$$

□

This result implies the next one:

Lemma 12.2.3. *Let $(X_t)_{t\in[0,T]}$ be a stochastic process such that $X_t \in L^1(\mathbb{P}_M)$, for all $t \in [0,T]$. Hence, X is a $\mathbb{P}_M$-martingale if and only if MX is a $\mathbb{P}$-martingale.*

Moreover, since $M > 0$, we have the following result

Lemma 12.2.4. *$\mathbb{P}_M \sim \mathbb{P}$, i.e. $\mathbb{P}_M$ is equivalent to $\mathbb{P}$, which means $\mathbb{P}_M(A) = 0 \Leftrightarrow \mathbb{P}(A) = 0$.*

Proof. This follows from the fact $M_T > 0$ $\mathbb{P}$-a.s. and the definition of $\mathbb{P}_M$. □

12.3 Stochastic Exponential

We fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a Brownian space $(B_t)_{t\in[0,T]}$ in it. Consider $\theta \in \mathcal{L}^2_{loc}[0,T]$ and define

$$M^\theta_t = \exp\left\{\int_0^t \theta_s dB_s - \frac{1}{2}\int_0^t \theta_s^2 ds\right\}. \tag{12.1}$$

A simple application of Itô formula implies

$$M^\theta_t = 1 + \int_0^t \theta_s M^\theta_s dB_s. \tag{12.2}$$

Hence, M^θ is a local martingale. The next lemma shows a condition that makes M a true martingale.

Lemma 12.3.1. *Assume θ is bounded. Then*

$$\mathbb{E}\left[\sup_{t\in[0,T]} |M^\theta_t|^p\right] < +\infty,$$

for all $p \geq 1$. Then, M^θ is indeed a martingale.

Proof. Define

$$X_t = \int_0^t \theta_s dB_s.$$

By the BDG inequality, see Theorem 9.7.1, for any $n \in \mathbb{N}$,

$$\mathbb{E}\left[\sup_{t\in[0,T]} |X_t|^n\right] < +\infty.$$

Moreover, by Itô formula:

$$X_t^{2n} = 2n\int_0^t X_s^{2n-1}\theta_s dB_s + n(2n-1)\int_0^t X_s^{2n-2}\theta_s^2 ds.$$

Let C_0 be the constant that bounds θ. Since the Itô integral above is a martingale, we have

$$\mathbb{E}[|X_t|^{2n}] = n(2n-1)\mathbb{E}\left[\int_0^t X_s^{2n-2}\theta_s^2 ds\right] \leq C_0^2 n(2n-1)\int_0^t \mathbb{E}[X_s^{2n-2}]ds.$$

This implies, by induction and Cauchy-Schwarz, that there exists $C_1 > 0$ such that $\mathbb{E}[|X_t|^n] \leq C_1^n$. Therefore

$$(M^\theta_t)^p = \exp\left\{pX_t - \frac{p}{2}\int_0^t \theta_s^2 ds\right\} \leq e^{pX_t} = \sum_{n=0}^{+\infty} \frac{p^n X_t^n}{n!},$$

which implies

$$\mathbb{E}[|M^\theta_t|^p] \leq \sum_{n=0}^{+\infty} \frac{p^n \mathbb{E}[|X_t|^n]}{n!} \leq e^{pC_1} < +\infty.$$

Using Equation (12.2) and the BDG inequality, we conclude

$$\mathbb{E}\left[\sup_{t\in[0,T]} |M^\theta_t|^p\right] < +\infty,$$

for all $p \geq 1$. □

Therefore, at least for bounded θ, we may define the following probability $\mathbb{P}_\theta = \mathbb{P}_{M^\theta}$. There is a weaker condition that assures that M^θ is a proper martingale; it is called Novikov condition:

Theorem 12.3.2 (Novikov Condition)**.** *Let $\theta \in \mathcal{L}^2_{loc}[0,T]$ such that*

$$\mathbb{E}\left[\exp\left\{\frac{1}{2}\int_0^T \theta_t^2 dt\right\}\right] < +\infty. \tag{12.3}$$

12.4 Girsanov Theorem

In order o prove the main theorem of this chapter, we need the result discussed in Section 9.6.
We shall now prove the main result of the chapter.

Theorem 12.4.1 (Girsanov's Theorem - First Version)**.** *Let θ be a bounded stochastic process. Define*

$$B_t^\theta = B_t - \int_0^t \theta_s ds.$$

Then B^θ is a Brownian motion under $\mathbb{P}_\theta$.

Proof. By Itô formula,

$$\begin{aligned} d(M_t^\theta B_t^\theta) &= M_t^\theta(\theta_t B_t^\theta + 1)dB_t, \\ d(M_t^\theta((B_t^\theta)^2 - t)) &= M_t^\theta(2B_t^\theta + ((B_t^\theta)^2 - t)\theta_t)dB_t. \end{aligned}$$

Hence, $M_t^\theta B_t^\theta$ and $M_t^\theta((B_t^\theta)^2 - t)$ are martingales under $\mathbb{P}$, which implies that B_t^θ and $(B_t^\theta)^2 - t$ are martingales under $\mathbb{P}_\theta$. By Levy's Characterization of the Brownian motion, we conclude that B^θ is a $\mathbb{P}_\theta$-Brownian motion. □

Theorem 12.4.2 (Girsanov's Theorem - Final Version)**.** *Consider X given by*

$$X_t = x + \int_0^t \mu_s ds + \int_0^t \sigma_s dB_s.$$

Assume $\sigma_t > 0$. Let $(\beta_t)_{t\in[0,T]}$ be a stochastic process such that $\theta_t = (\beta_t - \mu_t)/\sigma_t$ is a bounded stochastic process. Define

$$B_t^\theta = B_t - \int_0^t \theta_s ds.$$

Then B^θ is a Brownian motion under $\mathbb{P}_\theta$ and

$$X_t = x + \int_0^t \beta_s ds + \int_0^t \sigma_s dB_s^\theta.$$

Proof. We just need to verify the dynamics for X under $\mathbb{P}_\theta$:

$$\begin{aligned} dX_t &= \beta_t dt + \sigma_t dB_t^\theta = \beta_t dt + \sigma_t(dB_t - \theta_t dt) \\ &= \beta_t dt + \sigma_t dB_t - (\beta_t - \mu_t)dt = \mu_t dt + \sigma_t dB_t. \end{aligned}$$

□

Example 12.4.3 (Black–Scholes model)**.** Consider a process $(S_t)_{t\in[0,T]}$ under a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. This process models the price of a stock. In Mathematical Finance, we call the probability measure $\mathbb{P}$ the historical or physical probability measure. A risk-neutral probability measure is any probability measure $\mathbb{Q}$ in $(\Omega, \mathcal{F})$ such that $\mathbb{Q} \sim \mathbb{P}$ and the discounted stock price $(e^{-rt}S_t)_{t\in[0,T]}$ is a (local) bQ-martingale. Here r denotes the risk-neutral interest rate. It means that there exists the possibility to invest in the bank account that yields r. This is denoted by the process $(B_t)_{t\in[0,T]}$ that follows

$$dB_t = rB_t dt,$$

with $B_0 = 1$. We assume that S follows the following dynamics

$$dS_t = \mu S_t dt + \sigma S_t dB_t.$$

This is called the Black–Scholes model. The connection of these processes and a market model (with admissible trading strategies and other necessary objects) are left to other more complete exposition of the theory.

Let us characterize the measure $\mathbb{Q}$ completely using the Girsanov theory. Notice that the discounted stock price, which we denote by X, follows

$$dX_t = (\mu - r)X_t dt + \sigma X_t dB_t.$$

Hence, if the define

$$W_t = B_t - \frac{r-\mu}{\sigma}t,$$

then W is a Brownian motion under the measure

$$\mathbb{Q}(A) = \mathbb{E}[M_T^\theta 1_A],$$

where $\theta = \frac{r-\mu}{\sigma}$. Notice that

$$M_T^\theta = e^{(r-\mu)B_T/\sigma - \frac{1}{2}(r-\mu)^2 T/\sigma^2}.$$

Under this measure, X follows the dynamics

$$dX_t = \sigma X_t dW_t.$$

Consequently, S follows

$$dS_t = rS_t dt + \sigma S_t dW_t$$

under $\mathbb{Q}$.

Chapter 13

Feynman-Kac Representations

13.1 Introduction - the Brownian motion case

This chapter is devoted to the relation between stochastic equations and partial differential equations. We have seen in Section 9.2 that the famous heat equation is connected to the Brownian motion in a very special way. More precisely, we have the following theorem that was proved in the aforesaid section:

Theorem 13.1.1. *If u is the unique classical, bounded solution of the PDE:*

$$\frac{\partial u}{\partial t} - \frac{1}{2}\frac{\partial^2 u}{\partial x^2} = 0,$$

with $u(0,x) = g(x)$, then

$$u(t,x) = \mathbb{E}[g(x+B_t)].$$

The first simple and important generalization is to consider the PDE:

$$\frac{\partial u}{\partial t}(t.,x) - \frac{1}{2}\frac{\partial^2 u}{\partial x^2}(t,x) - q(x)u(t,x) = 0.$$

We then consider the following interpolating martingale (where t is fixed):

$$M_s = u(t-s, x+B_s)e^{\int_0^s q(x+B_r)dr}.$$

Notice that $M_0 = u(t,x)$ and $M_t = g(x+B_t)e^{\int_0^t q(x+B_r)dr}$. Hence, if we prove that M is indeed a martingale, we will find the desired representation:

$$u(t,x) = \mathbb{E}\left[g(x+B_t)e^{\int_0^t q(x+B_r)dr}\right].$$

To verify that M is a martingale, we simply apply Itô formula to find dM_s:

$$\begin{aligned} dM_s &= -\frac{\partial u}{\partial t}(t-s,x+B_s)e^{\int_0^s q(x+B_r)dr}ds + q(x+B_s)u(t-s,x+B_s)e^{\int_0^s q(x+B_r)dr}ds \\ &+ \frac{\partial u}{\partial x}(t-s,x+B_s)e^{\int_0^s q(x+B_r)dr}dB_s + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}(t-s,x+B_s)e^{\int_0^s q(x+B_r)dr}ds \\ &= \frac{\partial u}{\partial x}(t-s,x+B_s)e^{\int_0^s q(x+B_r)dr}dB_s. \end{aligned}$$

Hence, M is a (local) martingale and a true one because of the assumptions on u and q (usually, we require they are bounded). We have then conclude the following theorem:

Theorem 13.1.2. *If u is the unique classical, bounded solution of the PDE:*

$$\frac{\partial u}{\partial t}(t,x) - \frac{1}{2}\frac{\partial^2 u}{\partial x^2}(t,x) - q(x)u(t,x) = 0,$$

with $u(0,x) = g(x)$ and if q is bounded, then

$$u(t,x) = \mathbb{E}\left[g(x+B_t)e^{\int_0^t q(x+B_r)dr}\right].$$

Example 13.1.3. Let us compute

$$\mathbb{E}\left[e^{\int_0^t B_s ds}\right].$$

One might take the path of finding the distribution of the integral of the Brownian motion (which is Gaussian with mean zero and variance $t^3/3$). We will consider a different approach here. Define

$$u(t,x) = \mathbb{E}\left[e^{\int_0^t q(x+B_r)dr}\right] \text{ and } \phi(t) = \mathbb{E}\left[e^{\int_0^t B_s ds}\right],$$

where $q(y) = y$. Then

$$u(t,x) = e^{tx}\phi(t)$$

and u satisfies the PDE

$$\frac{\partial u}{\partial t}(t,x) - \frac{1}{2}\frac{\partial^2 u}{\partial x^2}(t,x) - xu(t,x) = 0,$$

with $\phi(0) = 1$. Hence, we find that

$$xu(t,x) + e^{tx}\phi'(t) - \frac{1}{2}t^2 e^{tx}\phi(t) - xu(t,x) = 0 \Rightarrow \phi'(t) = \frac{1}{2}t^2\phi(t).$$

Therefore,

$$\mathbb{E}\left[e^{\int_0^t B_s ds}\right] = \phi(t) = e^{t^3/6}.$$

13.2 Connections with Markov SDEs

If instead of the heat equation, we had a general parabolic PDE (still in dimension one):

$$\frac{\partial u}{\partial t}(t,x) - \frac{1}{2}\sigma^2(x)\frac{\partial^2 u}{\partial x^2}(t,x) - \mu(x)\frac{\partial u}{\partial x}(t,x) - q(x)u(t,x) = 0,$$

what would be the Feynman-Kac representation? It is very straightforward to see that under boundedness assumptions, the representation becomes:

$$u(t,x) = \mathbb{E}\left[g(X_t^{0,x})e^{\int_0^t q(X_r^{0,x})dr}\right],$$

where $X^{0,x}$ is defined as

$$dX_t^{0,x} = \mu(X_t^{0,x})dt + \sigma(X_t^{0,x})dB_t,$$

with $X_0 = x$.

Another example we will find very often in the Quantitative Finance literature is the final value problem

$$\frac{\partial u}{\partial t}(t,x) + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2 u}{\partial x^2}(t,x) + \mu(t,x)\frac{\partial u}{\partial x}(t,x) - q(t,x)u(t,x) = 0,$$

with $u(T,x) = g(x)$. Define $X^{t,x}$ as we have done in Section 10.4:

$$dX_s^{t,x} = \mu(s, X_s^{t,x})ds + \sigma(s, X_s^{t,x})dB_s,$$

with $X_t^{t,x} = x$. Then, under the same regularity assumptions as before, we find

$$u(t,x) = \mathbb{E}\left[g(X_T^{t,x})e^{-\int_t^T q(s,X_s^{t,x})ds}\right].$$

Example 13.2.1 (Black–Scholes)**.** As we have seen in the previous chapter, there exists a probability measure, called risk neutral and denoted by $\mathbb{Q}$, such that the stock price S follows the dynamics

$$dS_t = rS_t dt + \sigma S_t dW_t,$$

where W is a $\mathbb{Q}$-Brownian motion. An European derivative on S with payoff g and maturity T is a financial contract that pays its holder $g(S_T)$ at time T. If this contract is liquid, we could argue that its discounted price should also be a $\mathbb{Q}$-martingale. Hence, if we denote the price by $(V_t)_{t\in[0,T]}$, we must have

$$e^{-rt}V_t = \mathbb{E}_{\mathbb{Q}}[e^{-rT}V_T \mid \mathcal{F}_t],$$

where $(\mathcal{F}_t)_{t\in[0,T]}$ is the natural filtration of S. The formula above can be rewritten as

$$V_t = \mathbb{E}_{\mathbb{Q}}[e^{-r(T-t)}g(S_T) \mid \mathcal{F}_t].$$

Using the fact that S is Markovian, we find that $V_t = V(t, S_t)$ where the function V is given by

$$V(t, S) = \mathbb{E}[e^{-r(T-t)}g(S_T) \mid S_t = S].$$

By the Feynman-Kac formula, if we assume V is smooth, it must solve the following PDE

$$\frac{\partial V}{\partial t}(t, S) + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}(t, S) + rS\frac{\partial V}{\partial S}(t, S) - rV(t, x) = 0,$$

with $V(T, S) = g(S)$. This is called Black–Scholes PDE. The most important payoff function is $g(S) = (S - K)^+$, which describes a call option with strike K. This case (and some others) can be solved in closed form to find the famous Black–Scholes formula.

13.3 Connections with Markov BSDEs - Nonlinear Feynman-Kac Representation

We have concluded in Section 11.6 that $Y_t = u(t, X_t)$, where Y and X are given by

$$\begin{aligned} X_t &= x + \int_0^t \mu(s, X_s)ds + \int_0^T \sigma(s, X_s)dB_s, \\ Y_t &= g(X_T) + \int_t^T f(s, X_s, Y_s, Z_s)ds - \int_t^T Z_s dB_s. \end{aligned}$$

If u is sufficiently smooth, we may apply Itô formula and find

$$du(t, X_t) = \left(\frac{\partial u}{\partial t} + \mu(t, X_t)\frac{\partial u}{\partial x} + \frac{1}{2}\sigma^2(t, X_t)\frac{\partial^2 u}{\partial x^2}\right)dt + \sigma(t, X_t)\frac{\partial u}{\partial x}dB_t.$$

Remember that

$$dY_t = -f(t, X_t, Y_t, Z_t)dt + Z_t dB_t.$$

Matching the dt and dB_t parts we find that

$$\frac{\partial u}{\partial t} + \mu(t, X_t)\frac{\partial u}{\partial x} + \frac{1}{2}\sigma^2(t, X_t)\frac{\partial^2 u}{\partial x^2} + f(t, X_t, Y_t, Z_t) = 0 \text{ and } \sigma(t, X_t)\frac{\partial u}{\partial x}(t, X_t) = Z_t.$$

Therefore, if u is a classical solution of the PDE

$$\frac{\partial u}{\partial t} + \mu(t, x)\frac{\partial u}{\partial x} + \frac{1}{2}\sigma^2(t, x)\frac{\partial^2 u}{\partial x^2} + f\left(t, x, u, \frac{\partial u}{\partial x}\sigma\right) = 0,$$

with $u(T, x) = g(x)$, we have

$$Y_t = u(t, X_t) \text{ and } Z_t = \sigma(t, X_t)\frac{\partial u}{\partial x}(t, X_t).$$