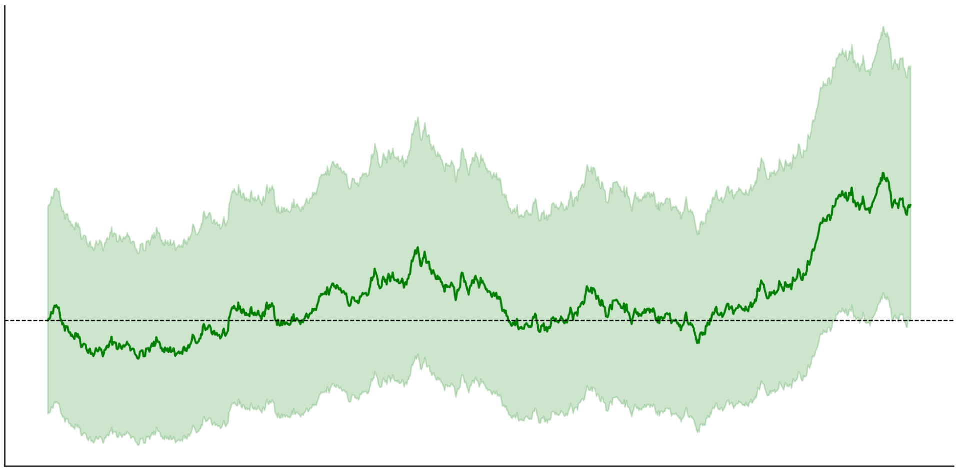
Introduction to Stochastic Modeling and Itô Calculus

Sébastien Haag



March 25, 2025

Contents

1 Review of Probability **3**
- 1.1 Events and Probability . . . 3
- 1.2 Random Variables . . . 5
- 1.3 Conditional Probability and Independence . . . 8

2 Conditional Expectation **9**
- 2.1 Conditioning on an Event . . . 9
- 2.2 Conditioning on a discrete random variable . . . 9
- 2.3 Conditioning on an Arbitrary Random Variable . . . 10
- 2.4 Conditioning on a σ-field . . . 11
- 2.5 General Properties . . . 11

3 Martingales in Discrete Time **15**
- 3.1 Sequences of Random Variables . . . 15
- 3.2 Filtrations . . . 16
- 3.3 Martingales . . . 17
- 3.4 Games of Chance . . . 17
- 3.5 Stopping Times . . . 18
- 3.6 Optional Stopping Theorem . . . 20

4 Martingale Inequalities and Convergence **21**
- 4.1 Doob's Martingale Inequalities . . . 21
- 4.2 Doob's Martingale Convergence Theorem . . . 25

5 Markov Chains **30**
- 5.1 First Examples and Definitions . . . 30
- 5.2 Classification of States . . . 37
- 5.3 Long-Time Behaviour of Markov Chains: General Case . . . 39
- 5.4 Long-Time Behaviour of Markov Chains with Finite State Space . . . 43

6 Stochastic Processes in Continious Time **46**
- 6.1 General Notions . . . 46
- 6.2 Poisson Process . . . 46
  - 6.2.1 Exponential Distribution and Lack of Memory . . . 46
  - 6.2.2 Construction of the Poisson Process . . . 47
  - 6.2.3 Poisson Process Starts from Scratch at Time t . . . 50
- 6.3 Brownian Motion . . . 53
  - 6.3.1 Definition and Basic Properties . . . 54
  - 6.3.2 Increments of Brownian Motion . . . 55

6.3.3 Sample path 57
6.3.4 Doob's Maximal L^2 Inequality for Brownian Motion 58

7 Itô Stochastic Calculus **60**
7.1 Itô Stochastic Integral: Definition 61
7.2 Properties of the Stochastic Integral 66
7.3 Stochastic Differential and Itô Formula 68
7.4 Stochastic Differential Equations 71

Chapter 1

Review of Probability

1.1 Events and Probability

Definition 1: Let Ω be a non-empty set. A σ-field $\mathcal{F}$ on Ω is a collection of subsets of Ω such that:

- $\emptyset \in \mathcal{F}$
- If $A \in \mathcal{F}$, then $\Omega \setminus A \in \mathcal{F}$
- If $(A_i)_{i=1}^n$ is a sequence of sets belonging to $\mathcal{F}$, then $\bigcup_{i=1}^n A_i \in \mathcal{F}$

$\mathcal{B}(\mathbb{R})$ denotes the Borel σ-field, which is the smallest σ-field containing all intervals of $\mathbb{R}$.

Definition 2: Let $\mathcal{F}$ be a σ-field on a set Ω. A probability measure $\mathbb{P}$ is a function $\mathbb{P} : \mathcal{F} \to [0;1]$ such that:

- $\mathbb{P}(\Omega) = 1$
- Si $(A_i)_{i=1}^n$ is a sequence of pairwise disjoint sets (i.e. $A_i \cap A_j = \emptyset, \forall i \neq j$) belonging to $\mathcal{F}$, then $P\left(\bigcup_{i=1}^\infty A_i\right) = \sum_{i=1}^\infty P(A_i)$ (Subadditivity).

The triplet $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. The sets belonging to $\mathcal{F}$ are called events. An event A is said to occur almost surely (a.s.) when $\mathbb{P}(A) = 1$.
Recall that the Lebesgue measure is the unique measure defined on Borel sets such that: $Leb([a;b]) = b - a$.

Lemma 1: (Borel Cantelli) Soit $A_1, A_2, ..., A_n$ a sequence of events in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{P}(A_1) + \mathbb{P}(A_2) + ... < \infty$ then the probability that infinitely many of the A_n occur is zero:

$$P\left(\limsup_{n\to\infty} A_n\right) = 0.$$

If the events A_n are independent and $\sum_{n=1}^\infty P(A_n) = \infty$ then:

$$P\left(\limsup_{n\to\infty} A_n\right) = 1.$$

Proof: Since

$$A = \bigcap_{m\in\mathbb{N}} \bigcup_{n\geq m} A_n$$

is a decreasing intersection, by the monotone convergence properties of a measure (finite),

$$0 \leq P(A) = \lim_{m\to\infty} P\left(\bigcup_{n\geq m} A_n\right)$$

Moreover, by subadditivity:

$$P\left(\bigcup_{n\geq m} A_n\right) \leq \sum_{n\geq m} P(A_n).$$

If the series $\sum_{n\in\mathbb{N}} P(A_n)$ converges, then its remainder tends to 0, which leads the conclusion.

An other simple proof consists in observing that, by monotone convergence,

$$\mathbb{E}\left[\sum_{n\in\mathbb{N}} \mathbf{1}_{A_n}\right] = \sum_{n\in\mathbb{N}} P(A_n).$$

Thus, if $\sum_{n\in\mathbb{N}} P(A_n) < \infty$, the series $\sum_{n\in\mathbb{N}} \mathbf{1}_{A_n}$ is almost surely finite. Hence, for P-almost every $\omega \in \Omega$, there are only finitely many A_n such that $\omega \in A_n$, which means that:

$$\omega \in A^c.$$

As a conclusion,

$$P(A^c) = 1.$$

For fixed integers $1 \leq m \leq N$, by the independence of the events A_n for $n \in \mathbb{N}$, and therefore also of A_n^c, we have:

$$P\left(\bigcup_{n=m}^{N} A_n\right) = 1 - P\left(\bigcap_{n=m}^{N} A_n^c\right).$$

By the independence of A_n^c, this simplifies to:

$$P\left(\bigcup_{n=m}^{N} A_n\right) = 1 - \prod_{n=m}^{N} (1 - P(A_n)).$$

For any real $u \geq 0$, it is easy to verify that:

$$1 - u \leq e^{-u}.$$

Applying this inequality to $u = P(A_n)$ for all n, we obtain :

$$\prod_{n=m}^{N} (1 - P(A_n)) \leq \exp\left(-\sum_{n=m}^{N} P(A_n)\right).$$

So,

$$1 \geq P\left(\bigcup_{n=m}^{N} A_n\right) \geq 1 - \exp\left(-\sum_{n=m}^{N} P(A_n)\right).$$

Since the series $\sum_{n\in\mathbb{N}} P(A_n)$ diverges, for any fixed integer m:

$$\lim_{N\to\infty} \sum_{n=m}^{N} P(A_n) = \infty.$$

Thus, by the previous inequality:

$$\lim_{N\to\infty} P\left(\bigcup_{n=m}^{N} A_n\right) = 1.$$

However, by monotone growth, the limit on the left-hand side is nothing other than:

$$P\left(\bigcup_{n\geq m} A_n\right).$$

Thus, for all $m \in \mathbb{N}$,

$$P\left(\bigcup_{n\geq m} A_n\right) = 1.$$

Moreover, we have seen in the proof of the Borel-Cantelli lemma that:

$$P(A) = \lim_{m\to\infty} P\left(\bigcup_{n\geq m} A_n\right).$$

Therefore, it is clear that :

$$P(A) = 1.$$

The proposition is thus proved.

1.2 Random Variables

Definition 3: If $\mathcal{F}$ is a σ-field on Ω, then a function $\xi : \Omega \to \mathbb{R}$ is said to be $\mathcal{F}$-measurable if:

$$\{\omega \in \Omega \mid \xi(\omega) \in B\} \in \mathcal{F}, \quad \forall B \in \mathcal{B}(\mathbb{R}),$$

where $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-field on $\mathbb{R}$. If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, then such a function is called a random variable.

Definition 4: The σ-field $\sigma(\xi)$ generated by a random variable ξ is the smallest σ-field containing all sets of the form $\{\xi \in B\}$, where B is a Borel set $\mathbb{R}$.

Definition 5: The σ-field $\sigma(\xi_i : i \in I)$ generated by a random variable $(\xi_i)_{i\in I}$ is the smallest σ-field containing all the Borel sets of the form $\{\xi_i \in B\}$, where B is a Borel set in $\mathbb{R}$ and $i \in I$.

Lemma 2: (Doob-Dinkyn) Let ξ be a random variable. Then, any random variable η that is $\sigma(\xi)$-measurable can be written is the form $\eta = f(\xi)$ for some Borel function $f : \mathbb{R} \to \mathbb{R}$.

Proof: Let f be a function $\sigma(T)/\mathcal{B}[0,1]$-measurable. First, note that, by the definition of $\sigma(T)$ as the collection of preimages of $\mathcal{A}'$-measurable sets under T, we know that if $A \in \sigma(T)$, then there exists a set $A' \in \mathcal{A}'$ such that:

$$A = T^{-1}(A').$$

Now, suppose that $f = \mathbf{1}_A$ is the indicator function of a set $A \in \sigma(T)$. If we identify a set $A' \in \mathcal{A}'$ such that:

$$A = T^{-1}(A'),$$

then the function $g = \mathbf{1}_{A'}$ satisfies the required condition. Since $A \in \sigma(T)$, such a set $A' \in \mathcal{A}'$ always exists. By linearity, this property extends to any simple measurable function f.
Now, let f be a measurable function but not necessarily simple. Moreover, f is the pointwise limit of an increasing sequence of positive simple functions f_n, that is:

$$f_n \geq 0.$$

The previous step ensures that:

$$f_n = g_n \circ T,$$

for some measurable function g_n.
The supremum:

$$g(x) = \sup_{n \geq 1} g_n(x)$$

exists on all Ω' and is measurable.
For all $x \in \operatorname{Im} T$, the sequence $g_n(x)$ is increasing, hence:

$$g|_{\operatorname{Im} T}(x) = \lim_{n \to \infty} g_n|_{\operatorname{Im} T}(x).$$

This shows that:

$$f = g \circ T.$$

Definition 6: Any random variable (v.a) $\xi : \Omega \to \mathbb{R}$ induces a probbaility measure $\mathbb{P}_\xi$ on $\mathbb{R}$, defined on the Borel σ-field $\mathcal{B}(\mathbb{R})$ by:

$$\mathbb{P}_\xi(B) = \mathbb{P}(\xi \in B), \quad \forall B \in \mathcal{B}(\mathbb{R}).$$

The measure $\mathbb{P}_\xi$ is called the distribution of ξ.
The function $F_\xi : \mathbb{R} \to [0,1]$ defined by:

$$F_\xi(x) = \mathbb{P}(\xi \leq x)$$

is called the cumulative distribution function (CDF) of ξ.

Definition 7: If ξ is a random variable and there exists a Borel function $f_\xi : \mathbb{R} \to \mathbb{R}$ such that, for every Borel set $B \subset \mathbb{R}$, we have:

$$\mathbb{P}(\xi \in B) = \int_B f_\xi(x)\, dx,$$

then ξ is said to have a absolutely continuous distribution, and f_ξ is called the probability density function (PDF) of ξ.
If ξ takes a (finite or infinite) sequence of distinct values $x_1, x_2, \dots, x_n$ and for every Borel set $B \subset \mathbb{R}$, we have:

$$\mathbb{P}(\xi \in B) = \sum_{x_i \in B} \mathbb{P}(\xi = x_i),$$

then ξ is said to have a discrete distribution, with values $\{x_1, x_2, \dots, x_n\}$ and probability masses $\mathbb{P}(\xi = x_i)$ at x_i.

Definition 8: The joint distribution of multiple random variables $\xi_1, \xi_2, \dots, \xi_n$ is a finite probability measure on $\mathbb{R}^n$ such that, for every Borel set $B \subset \mathbb{R}^n$, we have:

$$\mathbb{P}((\xi_1, \xi_2, \dots, \xi_n) \in B).$$

If there exists a Borel function $f_{\xi_1,\xi_2,\dots,\xi_n} : \mathbb{R}^n \to \mathbb{R}$ such that:

$$\mathbb{P}((\xi_1, \xi_2, \dots, \xi_n) \in B) = \int_B f_{\xi_1,\xi_2,\dots,\xi_n}(x_1, x_2, \dots, x_n)\, dx_1\, dx_2 \dots\, dx_n$$

for every Borel set B, then $f_{\xi_1,\xi_2,\dots,\xi_n}$ is called the joint density function of the random variables $\xi_1, \xi_2, \dots, \xi_n$.

Definition 9: A random variable $\xi : \Omega \to \mathbb{R}$ is said integrable if:

$$\int_\Omega |\xi|\, d\mathbb{P} < \infty.$$

In this case, the integral:

$$\mathbb{E}[\xi] = \int_\Omega \xi\, d\mathbb{P}$$

exists and is called the expectation of ξ.
The set of integrable random variables $\xi : \Omega \to \mathbb{R}$ is denoted:

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

Definition 10: A random variable $\xi : \Omega \to \mathbb{R}$ is said to be square integrable if:

$$\mathbb{E}[\xi^2] = \int_\Omega \xi^2\, d\mathbb{P} < \infty.$$

In this case, the variance of ξ is defined as:

$$\mathrm{Var}(\xi) = \mathbb{E}[\xi^2] - (\mathbb{E}[\xi])^2 = \int_\Omega (\xi - \mathbb{E}[\xi])^2\, d\mathbb{P}.$$

The set of all square integrable random variables is denoted:

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

1.3 Conditional Probability and Independence

Definition 11: For any events $A, B \in \mathcal{F}$ such that $\mathbb{P}(B) \neq 0$, the conditional probability of A given B is defined by:

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

Definition 12: Two events $A, B \in \mathcal{F}$ are said independent if:

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

More generally, n events $A_1, A_2, \dots, A_n$ are said to be independent if, for any subfamily $A_{i_1}, A_{i_2}, \dots, A_{i_k}$ (with $k \leq n$), we have:

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k}).$$

Definition 13: Two random variables ξ and η are said independent if, for every Borel set $A, B \in \mathcal{B}(\mathbb{R})$, the following events are independent:

$$\{\xi \in A\} \quad \text{et} \quad \{\eta \in B\}.$$

Similarly, n random variables $\xi_1, \xi_2, \dots, \xi_n$ are independent if, for all Borel sets $B_1, B_2, \dots, B_n \in \mathcal{B}(\mathbb{R})$, the events:

$$\{\xi_1 \in B_1\}, \quad \{\xi_2 \in B_2\}, \quad \dots, \quad \{\xi_n \in B_n\}$$

are independent.

Proposition 1: If two integrable random variables $\xi, \eta : \Omega \to \mathbb{R}$ are independent, then they are uncorrelated, meaning that:

$$\mathbb{E}[\xi\eta] = \mathbb{E}[\xi]\mathbb{E}[\eta],$$

provided that $\xi\eta$ is also integrable.

If $\xi_1, \xi_2, \dots, \xi_n : \Omega \to \mathbb{R}$ are independent and integrable, then:

$$\mathbb{E}[\xi_1\xi_2 \dots \xi_n] = \mathbb{E}[\xi_1]\mathbb{E}[\xi_2] \dots \mathbb{E}[\xi_n],$$

provided that $\xi_1\xi_2 \dots \xi_n$ is also integrable.

Définition 14 : Two σ-fields $\mathcal{G}$ and $\mathcal{H}$ included in $\mathcal{F}$ are said independent if every event $A \in \mathcal{G}$ and every event $B \in \mathcal{H}$ are independent. Similarly, a finite number of σ-fields $\mathcal{G}_1, \dots, \mathcal{G}_n$ included in $\mathcal{F}$ are said independent if, for any choice of events $A_1 \in \mathcal{G}_1, \dots, A_n \in \mathcal{G}_n$, this events are independent.

Definition 15: A random variable is said to be independent of a σ-field $\mathcal{G}$ if the σ-fields $\sigma(\xi)$ and $\mathcal{G}$ are independent.

A family is said independent if, for any finite subfamily consisting of random variables $\xi_1, \xi_2, \dots, \xi_m$ et de σ-fields $\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_n$, the following σ-fields are independent:

$$\sigma(\xi_1), \dots, \sigma(\xi_m), \mathcal{G}_1, \dots, \mathcal{G}_n.$$

Chapter 2

Conditional Expectation

The conditional expectation is a fundamental tool in the study of stochastic processes. It is therefore essential to develop a clear intuition behind this concept.

2.1 Conditioning on an Event

First case: Conditional expectation $E[\xi|B]$ of a random variable ξ given that an event B is realized.

Definition 1: For any integrable random variable ξ and any event $B \in \mathcal{F}$ such that $B \neq \emptyset$ and $P(B) > 0$, the conditional expectation of ξ given B is defined by:

$$\mathbb{E}[\xi|B] = \frac{1}{P(B)} \int_B \xi \, d\mathbb{P}.$$

2.2 Conditioning on a discrete random variable

The next step towards the general definition of conditional expectation is to condition on a discrete random variable η, taking possible values $y_1, y_2, ..., y_n$ such that $P(\eta = y_n) \neq \emptyset$, $\forall\ n$. Knowing the value of η is equivalent to know which event $\{\eta = y_n\}$ has occured. Conditioning on η is thus equivalent to conditioning on the events $\{\eta = y_n\}$. However, since we don't know pas in advance which event will occur, we must consider all the possibilities simultaneously by introducing a sequence of conditional expectations: $E[(\xi|\{\eta = y_1\}], E[(\xi|\{\eta = y_2\}], ..., E[(\xi|\{\eta = y_n\}]$. A practical way to proceed is to construct a new discrete random variable, that is constant and equal to $E[(\xi|\{\eta = y_n\}]$ on each set $\{\eta = y_n\}$.

Definition 2: Let ξ be a integrable random variable, and let η be a discrete random variable taking values $y_1, y_2, ..., y_n$. The conditional expectation of ξ given η is the random variable defined by: $E[\xi|\eta](\omega) = E[\xi|\{\eta = y_n\}]$ if $\eta(\omega) = y_n\ \forall\ n = 1, 2, ...$

Proposition 1: If ξ be an integrable random variable and η is a discrete random varaible, then:

- $E\left[\xi|\eta\right]$ is $\sigma\left(\eta\right)$-measurable.
- For any $A \in \sigma\left(\eta\right)$,
$$\int_A E\left[\xi|\eta\right] \, d\mathbb{P} = \int_A \xi \, d\mathbb{P}.$$

Proof: Suppose that η has pairwise distinct values $y_1, y_2, \ldots$. Then the events $\{\eta = y_1\}, \{\eta = y_2\}, \ldots$ are pairwise disjoint and cover Ω.
The σ-field $\sigma(\eta)$ is generated by these events; in fact, every $A \in \sigma(\eta)$ is a countable union of sets of the form $\{\eta = y_n\}$. Since $E[\xi|\eta]$ is constant on each of these sets, it must be $\sigma(\eta)$-measurable.
For each n, we have:

$$\int_{\{\eta=y_n\}} E[\xi|\eta]\ d\mathbb{P} = \int_{\{\eta=y_n\}} E[\xi|\{\eta = y_n\}]\ d\mathbb{P} = \int_{\{\eta=y_n\}} \xi\, d\mathbb{P}.$$

Since each $A \in \sigma(\eta)$ is a countable union of sets of the form $\{\eta = y_n\}$, which are pairwise disjoint, it follows that:

$$\int_A E[\xi \mid \eta]\ d\mathbb{P} = \int_A \xi\, d\mathbb{P}.$$

as required.

2.3 Conditioning on an Arbitrary Random Variable

Definition 3: Let ξ be an integrable random variable and let η be an arbitrary random variable. Then the conditional expectation of ξ given η is defined as a random variable $E[\xi|\eta]$ such that:

- $E[\xi|\eta]$ is $\sigma(\eta)$-measurable.
- For any $A \in \sigma(\eta)$,

$$\int_A E[\xi \mid \eta]\ d\mathbb{P} = \int_A \xi\, d\mathbb{P}.$$

Remark 1: We can also define the conditional probability of an event $A \in \mathfrak{I}$ given η by:

$$\mathbb{P}(A|\eta) = \mathbb{E}[\mathbf{1}_A|\eta],$$

where $\mathbf{1}_A$ is the indicator function.

Lemma 1: Let $(\Omega, \mathfrak{I}, \mathbb{P})$ be a probability space, and let $\mathcal{H}$ be a σ-field contained in $\mathfrak{I}$. If ξ is a $\mathcal{H}$-measurable random variable and for any $B \in \mathcal{H}$,

$$\int_B \xi\, d\mathbb{P} = 0,$$

then $\xi = 0$ almost surely.

Proof: Observe that $\mathbb{P}(\{\xi \geq \epsilon\}) = 0$ for any $\epsilon > 0$ because:

$$0 \leq \epsilon\, \mathbb{P}(\{\xi \geq \epsilon\}) = \int_{\{\xi\geq\epsilon\}} \epsilon\, d\mathbb{P} \leq \int_{\{\xi\geq\epsilon\}} \xi\, d\mathbb{P} = 0.$$

Similarly, $\mathbb{P}(\{\xi \leq -\epsilon\}) = 0$ for any $\epsilon > 0$ because:

$$0 \geq -\epsilon\, \mathbb{P}(\{\xi \leq -\epsilon\}) = \int_{\{\xi\leq-\epsilon\}} -\epsilon\, d\mathbb{P} \geq \int_{\{\xi\leq-\epsilon\}} \xi\, d\mathbb{P} = 0.$$

Thus,

$$\mathbb{P}(-\epsilon \leq \xi \leq \epsilon) = 1 \quad \forall \epsilon > 0.$$

Since this holds for all $\epsilon > 0$, it follows that $\xi = 0$ almost surely.

2.4 Conditioning on a σ-field

Proposition 2: If $\sigma(\eta) = \sigma(\eta')$, then

$$E[\xi|\eta] = E[\xi|\eta'] \quad \text{a.s.}$$

Proof: This is an immediate consequence of Lemma 1.

Definition 4: Let ξ be an integrable random variable on a probability space $(\Omega, \Im, \mathbb{P})$, and let $\mathcal{H}$ be a σ-field contained in $\Im$. Then the conditional expectation of ξ given $\mathcal{H}$ is defined as a random variable $E[\xi|\mathcal{H}]$ such that:

- $E[\xi|\mathcal{H}]$ is $\sigma(\mathcal{H})$-measurable.
- For any $A \in \mathcal{H}$,

$$\int_A E[\xi \mid \mathcal{H}]\, d\mathbb{P} = \int_A \xi\, d\mathbb{P}.$$

Proposition 3: $E[\xi|\mathcal{H}]$ exists and is unique in the sense that if $\xi = \xi'$ a.s., then

$$E[\xi|\mathcal{H}] = E[\xi'|\mathcal{H}] \quad \text{a.s.}$$

Theorem 1: (Radon-Nikodym) Let $(\Omega, \Im, \mathbb{P})$ be a probability space, and let $\mathcal{H}$ be a σ-field contained in $\Im$. Then, for any random variable ξ, there exists a $\mathcal{H}$-measurable random variable ζ such that:

$$\int_A \xi\, d\mathbb{P} = \int_A \zeta\, d\mathbb{P}$$

for each $A \in \mathcal{H}$.

2.5 General Properties

Proposition 4: Conditional expectation has the following properties:

- **Linearity:**

$$E[a\xi + b\zeta|\mathcal{H}] = aE[\xi|\mathcal{H}] + bE[\zeta|\mathcal{H}].$$

- **Taking expectations:**

$$E[E[\xi \mid \mathcal{H}]] = E[\xi].$$

- **Multiplication by a measurable function:**

$$E[\xi\zeta|\mathcal{H}] = \xi E[\zeta|\mathcal{H}] \quad \text{if } \xi \text{ is } \mathcal{H}\text{-measurable.}$$

- **Independence property:**

$$E[\xi|\mathcal{H}] = E[\xi] \quad \text{if } \xi \text{ is independent of } \mathcal{H}.$$

- **Tower property (Iterated Expectation):**

$$E[E[\xi|\mathcal{H}]|\mathcal{F}] = E[\xi|\mathcal{F}] \quad \text{if } \mathcal{F} \subset \mathcal{H}.$$

- **Positivity:** If $\xi \geq 0$, then
$$E\left[\xi|\mathcal{H}\right] \geq 0.$$

Proof: 1) The conditional expectation $E\left[X|\mathcal{H}\right]$ is a random variable measurable with respect to $\mathcal{H}$ that satisfies the fundamental property:

$$\forall A \in \mathcal{H}, \quad E\left[1_A X\right] = E\left[1_A E\left[X|\mathcal{H}\right]\right].$$

By the definition of conditional expectation, we must verify that for any $A \in \mathcal{H}$:

$$E[1_A(a\xi + b\zeta)] = E[1_A Y].$$

Using the linearity of standard expectation:

$$E\left[1_A\left(a\xi + b\zeta\right)\right] = aE\left[1_A\xi\right] + bE\left[1_A\zeta\right].$$

From the definition of $E\left[\xi|\mathcal{H}\right]$ and $E\left[\zeta|\mathcal{H}\right]$, we have:

$$E\left[1_A\xi\right] = E\left[1_A E\left[\xi|\mathcal{H}\right]\right],$$

$$E\left[1_A\zeta\right] = E\left[1_A E\left[\zeta|\mathcal{H}\right]\right].$$

Thus,
$$\begin{aligned} aE\left[1_A\xi\right] + bE\left[1_A\zeta\right] &= aE\left[1_A E\left[\xi|\mathcal{H}\right]\right] + bE\left[1_A E\left[\zeta|\mathcal{H}\right]\right]. \\ &= E[1_A(aE\left[\xi|\mathcal{H}\right] + bE\left[\zeta|\mathcal{H}\right])]. \end{aligned}$$

Now, if we define:
$$Y' = aE\left[\xi|\mathcal{H}\right] + bE\left[\zeta|\mathcal{H}\right],$$

we obtain:
$$E\left[1_A Y\right] = E\left[1_A Y'\right], \quad \forall A \in \mathcal{H}.$$

By the uniqueness of conditional expectation, we conclude that:

$$E\left[a\xi + b\zeta|\mathcal{H}\right] = aE\left[\xi|\mathcal{H}\right] + bE\left[\zeta|\mathcal{H}\right].$$

2) This follows by putting $A = \Omega$ in Proposition 3. Also, property 2) is a special case of property 5) when $\mathcal{H} = \{\emptyset, \Omega\}$.

3) We need to show that for any set $A \in \mathcal{H}$, the following holds:

$$E[1_A E[\xi\zeta \mid \mathcal{H}]] = E[1_A\xi\zeta].$$

By the linearity and integration property of conditional expectation:

$$E\left[1_A\xi\zeta\right] = E\left[\xi 1_A\zeta\right].$$

Since ξ is $\mathcal{H}$-measurable, it follows that $1_A\xi$ is also $\mathcal{H}$-measurable. By the pulling out measurable factors property of conditional expectation:

$$E\left[1_A\xi\zeta\right] = E\left[1_A\xi E[\zeta \mid \mathcal{H}]\right].$$

Moreover, from the integration property of conditional expectation, we know that:

$$E\left[1_A\xi E[\zeta \mid \mathcal{H}]\right] = E\left[1_A E\left[\xi\zeta|\mathcal{H}\right]\right].$$

Since this holds for all $A \in \mathcal{H}$, by the uniqueness of conditional expectation, we conclude:

$$E\left[\xi\zeta|\mathcal{H}\right] = \xi E\left[\zeta|\mathcal{H}\right].$$

4) Now, let $B \in \mathcal{H}$. By the definition of conditional expectation,

$$E\left[\xi|B\right] = \frac{E\left[\xi 1_B\right]}{\mathbb{P}(B)}.$$

But since ξ and B are independent,

$$E\left[\xi 1_B\right] = E\left[\xi\right]\mathbb{P}(B).$$

Thus,

$$E\left[\xi|B\right] = \frac{E\left[\xi\right]\mathbb{P}(B)}{\mathbb{P}(B)} = E\left[\xi\right].$$

5) We need to prove that $Y = E\left[E\left[\xi|\mathcal{H}\right]|\mathcal{F}\right]$ satisfies the definition of the conditional expectation of ξ with respect to $\mathcal{F}$, that is:

$$E\left[1_B Y\right] = E\left[1_B \xi\right], \quad \forall B \in \mathcal{F}.$$

By the definition of Y:

$$E\left[1_B Y\right] = E[1_B E[E[\xi \mid \mathcal{H}] \mid \mathcal{F}]].$$

Now, using the fundamental property of conditional expectation:

$$E\left[1_B E\left[\xi|\mathcal{H}\right]\right] = E\left[1_B \xi\right], \quad \forall B \in \mathcal{H}.$$

Since $\mathcal{F} \subset \mathcal{H}$, we apply the definition once more:

$$E[1_B E[E[\xi \mid \mathcal{H}] \mid \mathcal{F}]] = E[1_B E[\xi \mid \mathcal{H}]].$$

Thus,

$$E\left[1_B Y\right] = E\left[1_B \xi\right].$$

By the uniqueness of conditional expectation, this proves that:

$$Y = E\left[\xi|\mathcal{F}\right].$$

6) Since $\xi \geq 0$, we know that for any measurable set $A \in \mathcal{H}$:

$$E\left[1_A \xi\right] \geq 0.$$

By the definition of conditional expectation,

$$E[1_A E[\xi \mid \mathcal{H}]] = E[1_A \xi] \geq 0.$$

This holds for all $A \in \mathcal{H}$, which means that the function $E[\xi \mid \mathcal{H}]$ cannot take negative values on sets with positive probability. Thus, we conclude:

$$E\left[\xi|\mathcal{H}\right] \geq 0 \quad \text{a.s.}$$

Theorem 2: (Jensen's Iniequality) Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function, and let ξ be an integrable random variable on a probability space $(\Omega, \mathfrak{I}, \mathbb{P})$ such that $\varphi(\xi)$ is also integrable. Then:

$$\varphi(E[\xi \mid \mathcal{H}]) \leq E[\varphi(\xi) \mid \mathcal{H}] \quad \text{a.s.}$$

for any σ-field $\mathcal{H}$ on Ω contained in $\mathcal{F}$.

Proof: We can consider $(\Omega, \mathfrak{I}, \mathbb{P})$ as a probability space. In this case, if $f \in L^1(\Omega, \mathfrak{I}, \mathbb{P})$, and if $\varphi : \mathbb{R} \to \mathbb{R}$ is convex, then φ is continuous and right differentiable. It satisfies the inequality:

$$\varphi(y) \geq \varphi(c) + (y-c)\varphi'_d(c), \quad \forall y, c \in]a, b[.$$

Applying this inequality to our random variable $f(\omega)$ for all $\omega \in \Omega$, and choosing $c = E[f]$, we obtain:

$$\varphi(f(\omega)) \geq \varphi(E[f]) + (f(\omega) - E[f])\varphi'_d(E[f]).$$

Since $E[f]$ is a constant, we take the expectation on both sides:

$$E[\varphi(f(\omega))] \geq E[\varphi(E[f]) + (f(\omega) - E[f])\varphi'_d(E[f])].$$

Using the linearity of expectation, we separate the terms:

$$E[\varphi(f(\omega))] \geq \varphi(E[f]) + E[(f(\omega) - E[f])\varphi'_d(E[f])].$$

Since $E[f(\omega) - E[f]] = 0$, the second term disappears, leaving:

$$E[\varphi(f)] \geq \varphi(E[f]).$$

Chapter 3

Martingales in Discrete Time

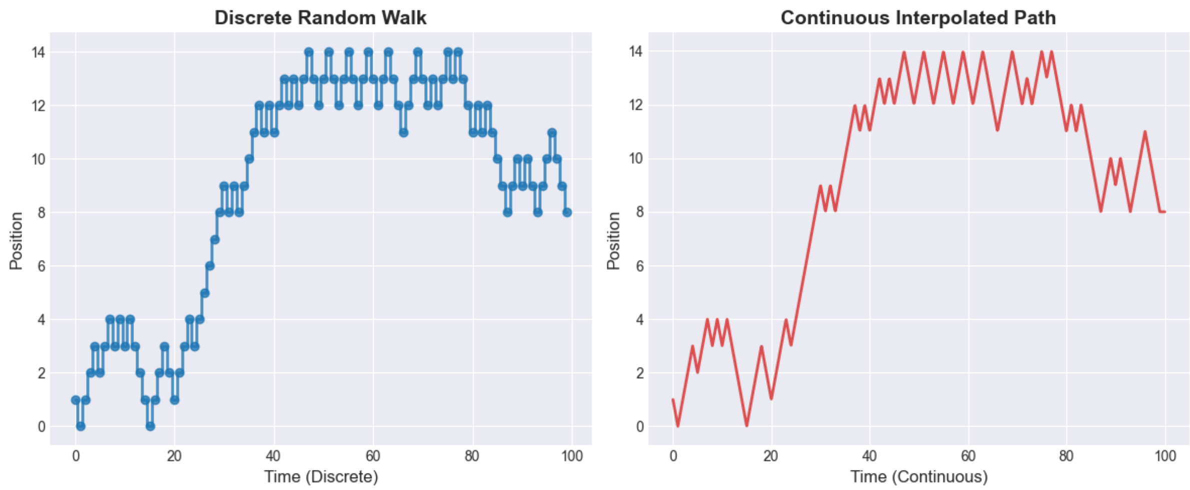
3.1 Sequences of Random Variables

A sequence $\xi_1, \xi_2, ..., \xi_n$ of random variables is typically used as a mathematical model of the outcomes of a series of random phenomena, such as the value of the FTSE All-share Index at the London Stock Exchange. The random variables in such a sequence are indexed by whole number, which are customarily referred to a discrete time.

Definition 1: The sequence of numbers $\xi_1(\omega), \xi_2(\omega), ..., \xi_n(\omega)$ for any fixed $\omega \in \Omega$ is called a sample path.

```
import numpy as np
import matplotlib.pyplot as plt

plt.style.use("seaborn-v0_8-darkgrid")

# Simulation parameters
n = 100
t = np.arange(n)
w = np.linspace(0, n, n*10)

# Generate a random walk: at each step, move +1 or -1 randomly
steps = np.random.choice([-1, 1], size=n)
discrete_path = np.cumsum(steps)  # Cumulative sum to track position over time

# Use linear interpolation to create a smooth continuous version of the path
continuous_path = np.interp(w, t, discrete_path)

fig, axes = plt.subplots(1, 2, figsize=(12, 5))

# First plot: Discrete random walk
axes[0].step(t, discrete_path, where='mid', marker='o', linestyle='-',
             color='#1f77b4', alpha=0.8, linewidth=2)
axes[0].set_title("Discrete Random Walk", fontsize=14, fontweight='bold')
axes[0].set_xlabel("Time (Discrete)", fontsize=12)
axes[0].set_ylabel("Position", fontsize=12)
axes[0].grid(True)
```

```
# Second plot: Continuous interpolated path
axes[1].plot(w, continuous_path, color='#d62728', linestyle='-',
             alpha=0.8, linewidth=2)
axes[1].set_title("Continuous Interpolated Path", fontsize=14,
↪ fontweight='bold')
axes[1].set_xlabel("Time (Continuous)", fontsize=12)
axes[1].set_ylabel("Position", fontsize=12)
axes[1].grid(True)

plt.tight_layout()

plt.savefig("random_walk_plot.png", dpi=300, bbox_inches='tight')

plt.show()
```



3.2 Filtrations

As the time n increases, so does our knowledge about what has happened in the past. This can be modelled by a filtration as defined below.

Definition 2: A sequence of σ-fields $\mathcal{F}_1, \mathcal{F}_2, ..., \mathcal{F}_n$ on Ω such that:

$$\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}$$

is called a filtration.
Here, $\mathcal{F}_n$ represents our knowledge at time n. It contains all events A such that at time n it is possible to decide whether A has occured or not. As n increases, there will be more such events A, i.e. the family $\mathcal{F}_n$ representing our knowledge will become larger.

Definition 3: We say that a sequence of random variables $\xi_1, \xi_2, ..., \xi_n$ is adapted to a filtration $\mathcal{F}_1, \mathcal{F}_2, ..., \mathcal{F}_n$ if ξ_n is $\mathcal{F}_n$-measurable for each $n = 1, 2, ...$

3.3 Martingales

Definition 4: A sequence $\xi_1, \xi_2, ..., \xi_n$ of random variables is called a martingale with respect to a filtration $\mathcal{F}_1, \mathcal{F}_2, ..., \mathcal{F}_n$ if:

1. ξ_n is integrable for each $n = 1, 2, \ldots$;
2. (ξ_n) is adapted to $\mathcal{F}_1, \mathcal{F}_2, \ldots$;
3. $\mathbb{E}[\xi_{n+1} \mid \mathcal{F}_n] = \xi_n$ a.s. for each $n = 1, 2, \ldots$.

Definition 5: A sequence $\xi_1, \xi_2, ..., \xi_n$ of random variables is called a supermartingale (submartingale) with respect to a filtration $\mathcal{F}_1, \mathcal{F}_2, ..., \mathcal{F}_n$ if:

1. ξ_n is integrable for each $n = 1, 2, \ldots$;
2. (ξ_n) is adapted to $\mathcal{F}_1, \mathcal{F}_2, \ldots$;
3. $E[\xi_{n+1} \mid \mathcal{F}_n] \leq \xi_n$ (respectively, $E[\xi_{n+1} \mid \mathcal{F}_n] \geq \xi_n$) a.s. for each $n = 1, 2, \ldots$.

3.4 Games of Chance

Definition 6: A gambling strategy $\alpha_1, \alpha_2, ...$ (with respect to a filtration $\mathcal{F}_1, \mathcal{F}_2, ..., \mathcal{F}_n$) is a sequence of random variables such that α_n is $\mathcal{F}_{n-1}$-measurable for each $n = 1, 2, ...$, where $\mathcal{F}_0 = \{\emptyset, \Omega\}$. If you follow a strategy $\alpha_1, \alpha_2, ...$ then your total winnings after n games will be:

$$\zeta_n = \alpha_1 \eta_1 + ... + \alpha_n \eta_n = \alpha_1(\xi_1 - \xi_0) + ... + \alpha_n(\xi_n - \xi_{n-1})$$

We also put $\zeta_0 = 0$ for covenience.

Proposition 1: Let $\alpha_1, \alpha_2, ...$ be a gambling strategy.

1. If $\alpha_1, \alpha_2, \ldots$ is a bounded sequence and $\xi_0, \xi_1, \xi_2, \ldots$ is a martingale, then $\zeta_0, \zeta_1, \zeta_2, \ldots$ is a martingale.
2. If $\alpha_1, \alpha_2, \ldots$ is a non-negative bounded sequence and $\xi_0, \xi_1, \xi_2, \ldots$ is a supermartingale, then $\zeta_0, \zeta_1, \zeta_2, \ldots$ is a supermartingale.
3. If $\alpha_1, \alpha_2, \ldots$ is a non-negative bounded sequence and $\xi_0, \xi_1, \xi_2, \ldots$ is a submartingale, then $\zeta_0, \zeta_1, \zeta_2, \ldots$ is a submartingale.

Proof: Because α_n and ζ_{n-1} are $\mathcal{F}_{n-1}$-measurable, we can take them out of the expectation conditioned on $\mathcal{F}_{n-1}$. Thus, we obtain:

$$E[\zeta_n \mid \mathcal{F}_{n-1}] = E[\zeta_{n-1} + \alpha_n(\xi_n - \xi_{n-1}) \mid \mathcal{F}_{n-1}] = \zeta_{n-1} + \alpha_n(E[\xi_n \mid \mathcal{F}_{n-1}] - \xi_{n-1})$$

If ξ_n is a martingale, then:

$$\alpha_n(E[\xi_n \mid \mathcal{F}_{n-1}] - \xi_{n-1}) = 0,$$

which proves assertion 1. If ξ_n is a supermartingale and $\alpha_n \geq 0$, then:

$$\alpha_n(E[\xi_n \mid \mathcal{F}_{n-1}] - \xi_{n-1}) \leq 0,$$

proving assertion 2. Finally, assertion 3 follows because:

$$\alpha_n(E[\xi_n \mid \mathcal{F}_{n-1}] - \xi_{n-1}) \geq 0,$$

if ξ_n is a submartingale and $\alpha_n \geq 0$.

3.5 Stopping Times

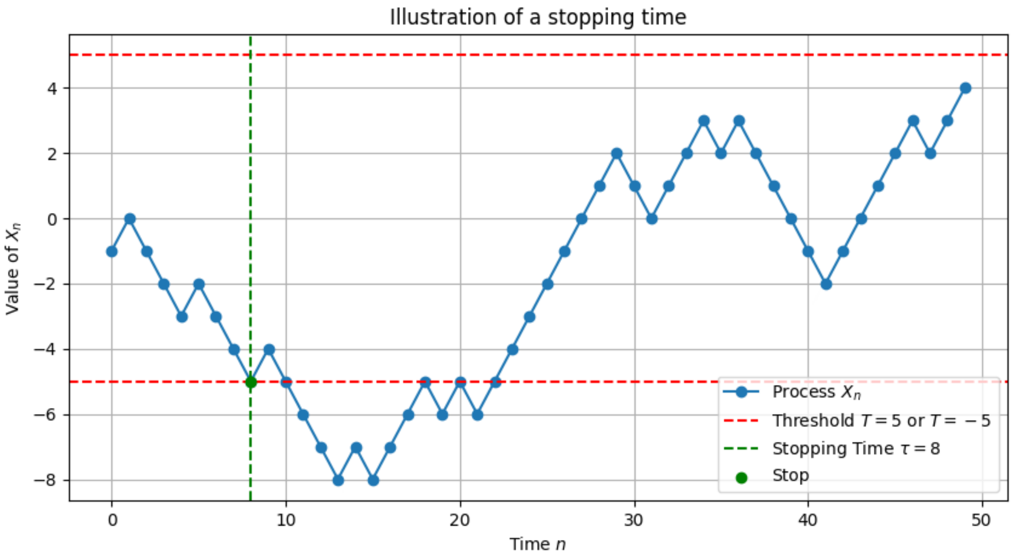
In many games of chance one usually has the option to quit at any time. The number of rounds played before quitting the game will be denoted by τ. τ is assumed to be a random variable with values in the set $\{1, 2, ...\} \cup \{\infty\}$. At each step n we should be able to decide whether to stop playing or not, i.e. whether or not $\tau = n$. Therefore the event $\{T = n\}$ should be in the σ-field $\mathcal{F}_n$ representing our knowledge at time n.

Definition 7: A random variable τ with values in the set $\{1, 2, ...\} \cup \{\infty\}$ is called a stopping time (with respect to a filtration $\mathcal{F}_n$) if for each $n = 1, 2, ...$:

$$\{\tau = n\} \in \mathcal{F}_n$$

```
import numpy as np
import matplotlib.pyplot as plt

# Definition of parameters
np.random.seed(42)
n_steps = 50
threshold = 5

# Generate a random walk
X = np.cumsum(np.random.choice([-1, 1], size=n_steps))
tau = np.where(np.abs(X) >= threshold)[0]

# Determine the stopping time
if len(tau) > 0:
    stopping_time = tau[0]
else:
    stopping_time = n_steps

# Plot the process
plt.figure(figsize=(10, 5))
plt.plot(range(n_steps), X, marker='o', linestyle='-', label='Process $X_n$')
plt.axhline(y=threshold, color='r', linestyle='--', label='Threshold $T=5$ or
↪  $T=-5$')
plt.axhline(y=-threshold, color='r', linestyle='--')
plt.axvline(x=stopping_time, color='g', linestyle='--', label=f"Stopping Time
↪  $\\tau={stopping_time}$")
```

```
plt.scatter(stopping_time, X[stopping_time], color='g', zorder=3, label='Stop')

plt.xlabel('Time $n$')
plt.ylabel('Value of $X_n$')
plt.title('Illustration of a stopping time')
plt.legend()
plt.grid()
plt.show()
```



Definition 8: We call $\xi_{\min\{\tau,n\}}$ the sequence stopped at τ. If is often denoted by ξ_n^τ. Thus, for each $\omega \in \Omega$:

$$\xi_n^\tau(\omega) = \xi_{\min\{\tau(\omega),n\}}(\omega)$$

Proposition 2: Let τ be a stopping time.

1. If ξ_n is a martingale, then so is $\xi_{\min\{\tau,n\}}$.
2. If ξ_n is a supermartingale, then so is $\xi_{\min\{\tau,n\}}$.
3. If ξ_n is a submartingale, then so is $\xi_{\min\{\tau,n\}}$.

Proof: This is in fact a consequence of Proposition 1. Given a stopping time τ, we put:

$$\alpha_n = \begin{cases} 1, & \text{si } \tau \geq n, \\ 0, & \text{si } \tau < n. \end{cases}$$

We claim that α_n is a gambing strategy (that is α_n is $\mathcal{F}_{n-1}$-measurable). This is because the inverse image $\{\alpha_n \in B\}$ of any Borel set $B \subset \mathbb{R}$ is equal to:

$$\emptyset \in \mathcal{F}_{n-1}$$

if $0, 1 \notin B$, or to

$$\Omega \in \mathcal{F}_{n-1}$$

if $0, 1 \in B$, or to

$$\{\alpha_n = 1\} = \{\tau \geq n\} = \{\tau > n - 1\} \in \mathcal{F}_{n-1}$$

if $1 \in B$ and $0 \notin B$, or to

$$\{\alpha_n = 0\} = \{\tau < n\} = \{\tau \leq n - 1\} \in \mathcal{F}_{n-1}$$

if $1 \notin B$ and $0 \in B$. For this gambling strategy:

$$\xi_{\min\{\tau,n\}} = \alpha_1(\xi_1 - \xi_0) + ... + \alpha_n(\xi_n - \xi_{n-1}).$$

3.6 Optional Stopping Theorem

Theorem 1: Let ξ_n be a martingale and τ a stopping time with respect to a filtration $\mathcal{F}_n$ such that the following conditions hold:

1. If $\tau < \infty$ a.s.,
2. If ξ_t is integrable,
3. If $E[\xi_n \mathbf{1}_{\{\tau>n\}}] \to 0$ as $n \to \infty$.

Then,

$$E[\xi_\tau] = E[\xi_1].$$

Proof: Because

$$\xi_\tau = \xi_{\min\{\tau,n\}} + (\xi_\tau - \xi_n)\mathbf{1}_{\{\tau>n\}}$$

it follows that

$$E[\xi_\tau] = E[\xi_{\min\{\tau,n\}}] + E[(\xi_\tau)\mathbf{1}_{\{\tau>n\}}] - E[(\xi_n)\mathbf{1}_{\{\tau>n\}}]$$

Since $\xi_{\min\{\tau,n\}}$ is a martingale by Proposition 2, the first term on the right-hand side is equal to:

$$E[\xi_{\min\{\tau,n\}}] = E[\xi_1]$$

The last term tends to zero by assumption 2. The middle term

$$E[(\xi_\tau)\mathbf{1}_{\{\tau>n\}}] = \sum_{k=n+1}^{\infty} E[\xi_k \mathbf{1}_{\{\tau=k\}}]$$

tends to zero as $n \to \infty$ because the series:

$$E[(\xi_\tau)] = \sum_{k=1}^{\infty} E[\xi_k \mathbf{1}_{\{\tau=k\}}]$$

is convergent by 2. It follows that:

$$E[\xi_\tau] = E[\xi_1].$$

as required.

Chapter 4

Martingale Inequalities and Convergence

It turns out that a large class of martingales can be represented in the form:

$$\xi_n = E[\xi \mid \mathcal{F}_n] \tag{4.1}$$

where $\xi = \lim \xi_n$ is an integrable random variable and $\mathcal{F}_1, \mathcal{F}_2, ..., \mathcal{F}_n$ is the filtration generated by $\xi_1, \xi_2, ...\xi_n$.

4.1 Doob's Martingale Inequalities

Proposition 1: (Doob's Maximal Inequality) Suppose that ξ_n, $n \in \mathbb{N}$, is a non-negative submartingale (with respect to a filtration $\mathcal{F}_n$. Then for any $\lambda > 0$:

$$\lambda P\left(\max_{k\leq n} \xi_k \geq \lambda\right) \leq E\left[\xi_n \mathbf{1}_{\{\max_{k\leq n} \xi_k \geq \lambda\}}\right]$$

Proof: We put $\xi_n^* = \max_{k\leq n} \xi_k$ for brevity. For $\lambda > 0$ let us define:

$$\tau = \min\{k \leq n : \xi_k \geq \lambda\},$$

if there is a $k \leq n$ such that $\xi_k \geq \lambda$, and $\tau = n$ otherwise. Then τ is a stopping time such txhat $\tau \leq n$ a.s. Since ξ_n is a submartingale,

$$E[\xi_n] \geq E[\xi_\tau].$$

But

$$E[\xi_\tau] = E[\xi_\tau \mathbf{1}_{\{\xi_n^* \geq \lambda\}}] + E[\xi_\tau \mathbf{1}_{\{\xi_n^* < \lambda\}}]$$

Observe that if $\xi_n^* \geq \lambda$, then $\xi_\tau \geq \lambda$. Moreover, if $\xi_n^* < \lambda$, then $\tau = n$, and so $\xi_\tau = \xi_n$. Therefore:

$$E[\xi_n] \geq E[\xi_\tau] \geq \lambda P(\xi_n^* \geq \lambda) + E[\xi_\tau \mathbf{1}_{\{\xi_n^* < \lambda\}}],$$

It follows that

$$\lambda P(\xi_n^* \geq \lambda) \leq E[\xi_n] - E[\xi_\tau \mathbf{1}_{\{\xi_n^* < \lambda\}}] = E[\xi_\tau \mathbf{1}_{\{\xi_n^* \geq \lambda\}}],$$

completing the proof.

Theorem 1: (Doob's Maximal $\mathcal{L}^2$ Inequality) Suppose that ξ_n, $n \in \mathbb{N}$, is a non-negative square integrable submartingale (with respect to a filtration $\mathcal{F}_n$. Then,

$$E\left[\left|\max_{k\leq n} \xi_k\right|^2\right] \leq 4E\left[|\xi_n|^2\right] \tag{4.2}$$

Proof: Put $\xi_n^* = \max_{k\leq n} \xi_k$. By the Fubini theorem and finally the Cauchy-Schwarz inequality:

$$\begin{aligned}
E\left[|\xi_n|^2\right] &= 2\int_0^\infty tP(\xi_n^* > t)\, dt \leq 2\int_0^\infty E\left[\xi_n \mathbf{1}_{\{\xi_n^* \geq t\}}\right] dt \\
&= 2\int_0^\infty \left(\int_{\{\xi_n^* \geq t\}} \xi_n \, dP\right) dt = 2\int_\Omega \xi_n \left(\int_0^{\xi_n^*} dt\right) dP \\
&= 2\int_\Omega \xi_n \xi_n^* \, dP = 2E[\xi_n \xi_n^*] \leq 2\left(E\left[|\xi_n|^2\right]\right)^{1/2} \left(E\left[|\xi_n^*|^2\right]\right)^{1/2}
\end{aligned}$$

Dividing by $\left(E\left[|\xi_n^*|^2\right]\right)^{1/2}$, we get (4.2).

Definition 1: Given an adapted sequence of random variable $\xi_1, \xi_2, ..., \xi_n$ and two real numbers $a < b$, we define a gambling strategy $\alpha_1, \alpha_2, ...$ by putting $\alpha_1 = 0$ and for $n = 1, 2, ...$,

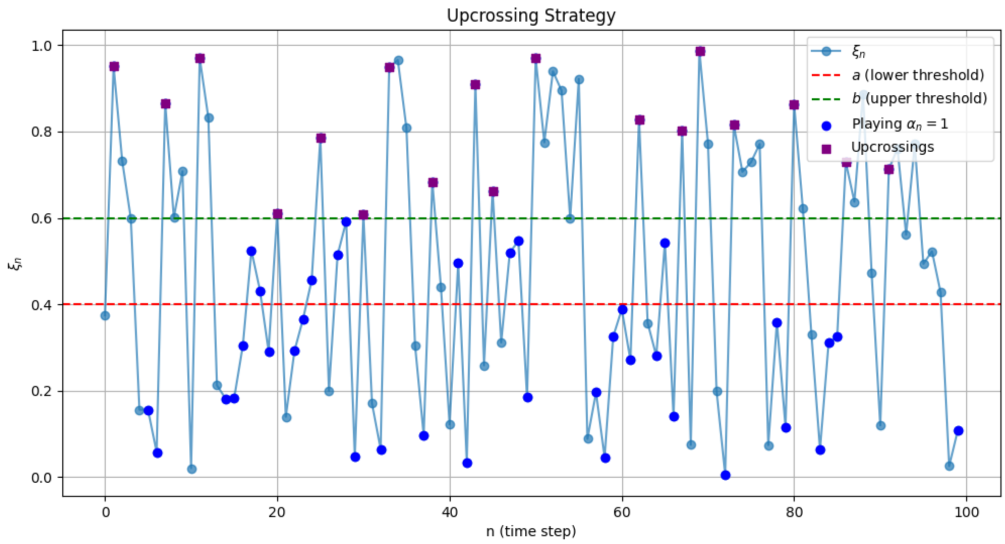
$$\alpha_{n+1} = \begin{cases} 1, & \text{if } \alpha_n = 0 \text{ and } \xi_n < a, \\ 1, & \text{if } \alpha_n = 1 \text{ and } \xi_n \leq b, \\ 0, & \text{otherwise.} \end{cases}$$

It will be called the *upcrossing strategy*. Each $k = 1, 2, ...$ such that $\alpha_k = 1$ and $\alpha_{k+1} = 0$ will be called an *upcrossing* of the interval $[a, b]$. The upcrossings form a (finite or infinite) increasing sequence $u_1 < u_2 < ....$ The number of upcrossings made up to time n, that is, the largest k such that $u_k \leq n$ will be denoted by $U_n[a, b]$ (we put $U_n[a, b] = 0$ if no such k exists).

The meaning of the above definition is this. Initially, we refrain from playing the game and wait until ξ_n becomes less than a.As soon as tis happens, we start playing unit stakes at each round of the game and continue until ξ_n becomes greater than b. At this stage we refrain from playing again, wait until ξ_n becomes less than a, and so on. The strategy α_n is defined in such a way that $\alpha_n = 0$ whenever we refrain from playing the n game, and $\alpha_n = 1$ otherwise. During each run of consecutive games with $\alpha_n = 1$ the process ξ_n crosses the interval $[a, b]$, starting below a and finishing above b. This is what is meant by an upcrossing. Observe that each upcrossing will increase our total winnings by at least $b - a$. For convenience, we identify each upcrossing with its last step k, such that $\alpha_k = 1$ and $\alpha_{k+1} = 0$. A typical sample path of the upcrossings strategy is shown in the figure below.

```python
import numpy as np
import matplotlib.pyplot as plt

# Parameters
n = 100
```

```python
a, b = 0.4, 0.6
np.random.seed(42)

# Generate a sequence of random variables $\xi_n$
xi = np.random.uniform(0, 1, n)

# Initialize the strategy $\alpha_n$
alpha = np.zeros(n, dtype=int)
upcrossings = []

# Apply the upcrossing strategy
for i in range(1, n):
    if alpha[i - 1] == 0 and xi[i - 1] < a:
        alpha[i] = 1
    elif alpha[i - 1] == 1 and xi[i - 1] <= b:
        alpha[i] = 1
    else:
        alpha[i] = 0

    if alpha[i - 1] == 1 and alpha[i] == 0:
        upcrossings.append(i - 1)

plt.figure(figsize=(12, 6))
plt.plot(xi, label=r'$\xi_n$', marker='o', linestyle='-', alpha=0.7)
plt.axhline(y=a, color='r', linestyle='--', label=r'$a$ (lower threshold)')
plt.axhline(y=b, color='g', linestyle='--', label=r'$b$ (upper threshold)')
plt.scatter(np.where(alpha == 1), xi[alpha == 1], color='blue', label=r'Playing
↪   $\alpha_n=1$', zorder=3)
plt.scatter(upcrossings, xi[upcrossings], color='purple', marker='s',
↪   label='Upcrossings', zorder=4)

plt.xlabel('n (time step)')
plt.ylabel(r'$\xi_n$')
plt.title('Upcrossing Strategy')
plt.legend()
plt.grid()
plt.show()
```



Lemma 1: (Upcrossings Inequality) If $\xi_1, \xi_2, \dots$ is a supermartingale and $a < b$, then

$$(b-a)E[U_n[a,b]] \leq E[(\xi_n - a)^-]$$

By x^- we denote the negative part of a real number x, i.e. $x^- = \max\{0, -x\}$.

Proof: Let

$$\zeta_n = \alpha_1(\xi_1 - \xi_0) + \dots + \alpha_n(\xi_n - \xi_{n-1})$$

be the total winnings at step $n = 1, 2, \dots$ if the upcrossings strategy is followed. It will be convenient to put $\zeta_0 = 0$. By Proposition 1(one cannot beat the system using a gambling strategy) ζ_n is a supermartingale. Let us fix an n and put $k = U_n[a,b]$, so that $0 < u_1 < u_2 < \dots < u_k \leq n$.
Clearly, each upcrossing increases the total winnings by $b - a$,

$$\zeta_{u_i} - \zeta_{u_{i-1}} \geq b - a$$

for $i = 1, \dots, k$. (We put $u_0 = 0$ for simplicity.) Moreover,

$$\zeta_n - \zeta_{u_k} \geq -(\xi_n - a)^-.$$

It follows that,

$$\zeta_n \geq (b-a)U_n[a,b] - (\xi_n - a)^-.$$

taking the expectation on both sides, we get

$$E[\zeta_n] \geq (b-a)E[U_n[a,b]] - E[(\xi_n - a)^-].$$

But ζ_n is a supermartingale, so $0 = E[\zeta_1] \geq E[\zeta_n]$, which proves the Upcrossings Inequality.

4.2 Doob's Martingale Convergence Theorem

Theorem 2: (Doob's Martingale Convergence Theorem) Suppose that $\xi_1, \xi_2, \ldots$ is a supermartingale (with respect to a filtration $\mathcal{F}_1, \mathcal{F}_2, \ldots$ such that

$$\sup_n \mathbb{E}[|\xi_n|] < \infty$$

Then, there is an integrable random variable ξ such that:

$$\lim_{n\to\infty} \xi_n = \xi \quad \text{a.s.}$$

Proof: By the Upcrossings Inequality

$$\mathbb{E}[U_n[a,b]] \leq \frac{\mathbb{E}[(\xi_n - a)^-]}{b-a} \leq \frac{M + |a|}{b-a} < \infty$$

where

$$M = \sup_n \mathbb{E}[|\xi_n|] < \infty.$$

Since $U_n[a,b]$ is a non-decreasing sequence, it follows that

$$\mathbb{E}\left[\lim_{n\to\infty} U_n[a,b]\right] = \lim_{n\to\infty} \mathbb{E}[U_n[a,b]] \leq \frac{M+|a|}{b-a} < \infty$$

This implies that

$$P\left(\lim_{n\to\infty} U_n[a,b] < \infty\right) = 1$$

for any $a < b$. Since the set of all pairs of rational numbers $a < b$ is countable, the event

$$A = \bigcap_{a<b,\, a,b\in\mathbb{Q}} \left\{\lim_{n\to\infty} U_n[a,b] < \infty\right\} \tag{4.3}$$

has probability 1. (The intersection of countably many events has probability 1 if each of these events has probability 1.)
We claim that sequence ξ_n converges a.s. to a limit ξ. Consider the set

$$B = \left\{\liminf_{n\to\infty} \xi_n < \limsup_{n\to\infty} \xi_n\right\} \subset \Omega$$

on which the sequence ξ_n fails to converge. Then for any $\omega \in B$ there are rational numbers a, b such that

$$\liminf_{n\to\infty} \xi_n(\omega) < a < b < \limsup_{n\to\infty} \xi_n(\omega)$$

implying that $\lim_{n\to\infty} U_n[a,b](\omega) = \infty$. This meant that B and the event A in (4.3) are disjoint, so $P(B) = 0$, since $P(A) = 1$, which proves the claim. It remains to show that the limit ξ is an integrable random variable. By Fatou's Lemma:

$$\begin{aligned}\mathbb{E}[|\xi|] &= \mathbb{E}\left[\liminf_{n\to\infty} |\xi_n|\right]\\ &\leq \liminf_{n\to\infty} \mathbb{E}[|\xi_n|]\\ &< \sup_n \mathbb{E}[|\xi_n|] < \infty.\end{aligned}$$

This complete the proof.

Remark 1: In particular, the theorem is valid for martingales because every martingale is a supermartingale. It is also valid for submartingales, since ξ_n is a submartingale if and only if $-\xi_n$ is a supermartingale.

Definition 2: A sequence $\xi_1, \xi_2, \ldots$ of random variables is called *uniformly integrable* if for every $\epsilon > 0$ there exists an $M > 0$ such that:

$$\int_{\{|\xi_n|>M\}} |\xi_n| \, dP < \epsilon$$

for all $n = 1, 2, \ldots$.

Proposition 2: Uniform integrability is a necessary condition for a sequence $\xi_1, \xi_2, \ldots$ of integrable random variables to converge in L^1.

Proof: Suppose that $\xi_n \to \xi$ in L^1, i.e. $E[|\xi_n - \xi|] \to 0$. We take any $\epsilon > 0$. There is an integer N such that:

$$n \geq N \implies E[|\xi_n - \xi|] < \frac{\epsilon}{2}$$

By Lemma 2 there is a $\delta > 0$ such that

$$P(A) < \delta \implies \int_A |\xi| \, dP < \frac{\epsilon}{2}.$$

Taking a smaller $\delta > 0$ if necessary, we also have

$$P(A) < \delta \implies \int_A |\xi_n| \, dP < \epsilon \quad \text{pour } n = 1, 2, \ldots, N.$$

We claim that there is $M > 0$ such that

$$P\left(|\xi_n| > M\right) < \delta$$

for all n. Indeed, since

$$\begin{aligned}\mathbb{E}[|\xi_n|] &\geq \int_{\{|\xi_n|>M\}} |\xi_n| \, dP \\ &\geq MP(|\xi_n| > M).\end{aligned}$$

it suffices to take

$$M = \frac{1}{\delta} \sup_n \mathbb{E}[|\xi_n|]$$

(Because the sequence ξ_n converges in L^1, it is bounded in L^1, so the supremum is $< \infty$.) Now, since $P(|\xi_n| > M) < \delta$,

$$\begin{aligned}\int_{\{|\xi_n|>M\}} |\xi_n| \, dP &\leq \int_{\{|\xi_n|>M\}} |\xi| \, dP + \int_{\{|\xi_n|>M\}} |\xi_n - \xi| \, dP \\ &\leq \int_{\{|\xi_n|>M\}} |\xi| \, dP + \mathbb{E}[|\xi_n - \xi|] \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.\end{aligned}$$

for any $n > N$ and

$$\int_{\{|\xi_n|>M\}} |\xi_n| \, dP < \epsilon$$

for any $n = 1, ..., N$, completing the proof.

Lemma 2: If ξ is integrable, then for every $\epsilon > 0$ there is a $\delta > 0$ such that

$$P(A) < \delta \implies \int_A |\xi| \, dP < \frac{\epsilon}{2}.$$

Proof: Let $\epsilon > 0$. Since ξ is integrable, there is an $M > 0$ such that:

$$\int_{\{|\xi|>M\}} |\xi| \, dP < \frac{\epsilon}{2}$$

Now,

$$\begin{aligned}
\int_A |\xi| \, dP &= \int_{A\cap\{|\xi|\leq M\}} |\xi| \, dP + \int_{A\cap\{|\xi|>M\}} |\xi| \, dP \\
&\leq \int_A M \, dP + \int_{\{|\xi|>M\}} |\xi| \, dP \\
&< MP(A) + \frac{\epsilon}{2}.
\end{aligned}$$

Let $\delta = \frac{\epsilon}{2M}$. Then

$$P(A) < \delta \implies \int_A |\xi| \, dP < \epsilon.$$

as required.

Theorem 3: Every uniformly integrable supermartingale (submartingale) ξ_n converges in L^1.

Proof: The sequence ξ_n is bounded in L^1, so it satisfies the conditions of Theorem 2 (Doob's Martingale Convergence Theorem). Therefore, there is an integrable random variable ξ such that $\xi_n \to \xi$ a.s. We can assume without loss of generality that $\xi = 0$ (since $\xi_n - \xi$ can be taken in place of ξ_n). That is to say,

$$P\left(\lim_{n\to\infty} \xi_n = 0\right) = 1$$

It follows that $\xi_n \to 0$ in probability, i.e. for any $\epsilon > 0$

$$P\left(|\xi_n| > \epsilon\right) \to 0 \quad \text{quand } n \to \infty$$

This is because by Fatou's lemma

$$\begin{aligned}
\lim_{n\to\infty} \sup P(|\xi_n| > \epsilon) &\leq P\left(\limsup_{n\to\infty}\{|\xi_n| > \epsilon\}\right) \\
&\leq P\left(\Omega \setminus \{\lim_{n\to\infty} \xi_n = 0\}\right) \\
&= 0.
\end{aligned}$$

Let $\epsilon > 0$. By uniform integrability there is an $M > 0$ such that

$$\int_{\{|\xi_n|>M\}} |\xi_n| \, dP \leq \frac{\epsilon}{3}$$

for all n. Since $\xi_n \to 0$ in probability, there is an integer N such that if $n > N$, then

$$P\left(|\xi_n| > \frac{\epsilon}{3}\right) < \frac{\epsilon}{3M}$$

We can assume without loss of generality that $M > \frac{\epsilon}{3}$. Then

$$\begin{aligned}
\mathbb{E}[|\xi_n|] &= \int_{\{|\xi_n|>M\}} |\xi_n| \, dP + \int_{\{M \geq |\xi_n| > \frac{\epsilon}{3}\}} |\xi_n| \, dP + \int_{\{\frac{\epsilon}{3} \geq |\xi_n|\}} |\xi_n| \, dP \\
&\leq \frac{\epsilon}{3} + MP(|\xi_n| > \frac{\epsilon}{3}) + \frac{\epsilon}{3} P(\frac{\epsilon}{3} \geq |\xi_n|) \\
&< \epsilon.
\end{aligned}$$

for all $n > N$. This proves that $E[|\xi_n|] \to 0$, that is, $\xi_n \to 0$ in L^1.

Theorem 4: Let ξ_n be a uniformly integrable martingale. Then,

$$\xi_n = E[\xi \mid \mathcal{F}_n],$$

where $\xi = \lim_{n\to\infty} \xi_n$ is the limit of ξ_n in L^1 and $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$ is the filtration generated by ξ_n.

Proof: For any $m > n$, $E[\xi_m \mid \mathcal{F}_n] = \xi_n$, i.e. for any $A \in \mathcal{F}_n$,

$$\int_A \xi_m \, dP = \int_A \xi_n \, dP.$$

Let n be an arbitrary integer and let $A \in \mathcal{F}_n$. For any $m > n$,

$$\begin{aligned}
\left| \int_A (\xi_n - \xi) \, dP \right| &= \left| \int_A (\xi_m - \xi) \, dP \right| \\
&\leq \int_A |\xi_m - \xi| \, dP \\
&\leq \mathbb{E}[|\xi_m - \xi|] \to 0.
\end{aligned}$$

as $m \to \infty$. It follows that

$$\int_A \xi_n \, dP = \int_A \xi \, dP$$

for any $A \in \mathcal{F}_n$, so $\xi_n = E[\xi \mid \mathcal{F}_n]$.

Theorem 5: (Kolmogorov's 0-1 Law) Let $\eta_1, \eta_2, \ldots$ be a sequence of independent random variables. We define the *tail σ-field*

$$\mathcal{T} = \bigcap_{n \geq 1} \mathcal{T}_n$$

where $\mathcal{T}_n = \sigma(\eta_n, \eta n + 1, \ldots)$. Then

$$P(A) = 0 \text{ or } 1$$

for any $A \in \mathcal{T}$.

Proof: Take any $A \in \mathcal{T}$ and define,

$$\xi_n = \mathbb{E}\left[\mathbf{1}_A \mid \mathcal{F}_n\right],$$

where $\mathcal{F}_n = \sigma(\eta_1, \dots, \eta_n)$. ξ_n is a uniformly integrable martingale, so $\xi_n \to \xi$ in L^1. By Theorem 4,

$$\mathbb{E}\left[\xi \mid \mathcal{F}_n\right] = \mathbb{E}\left[\mathbf{1}_A \mid \mathcal{F}_n\right]$$

for all n. Both $\xi = \lim_{n\to\infty} \xi_n$ and $\mathbf{1}_A$ are measurable with respect to the σ-field

$$\mathcal{F}_\infty = \sigma(\eta_1, \eta_2, \dots)$$

The family $\mathcal{G}$ consisting of all sets $B \in \mathcal{F}$ such that $\int_B \xi \, dP = \int_B \mathbf{1}_A \, dP$ is a σ-field containing $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \dots$. As a result, $\mathcal{G}$ contains the σ-field $\mathcal{F}_\infty$ generated by the family $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \dots$. By Lemma 1, it follows that $\xi = \mathbf{1}_A$ a.s. Since η_n is a sequence of independent random variables, the σ-field $\mathcal{F}_n$ and $\mathcal{T}_{n+1}$ are independent. Because $\mathcal{T} \subset \mathcal{T}_{n+1}$, the σ-field $\mathcal{F}_n$ and $\mathcal{T}$ are independent. Being $\mathcal{T}$-measurable, $\mathbf{1}_A$ is therefore independent of $\mathcal{F}_n$ for any n. This means that:

$$\begin{aligned} \xi_n &= \mathbb{E}\left[\mathbf{1}_A \mid \mathcal{F}_n\right] \\ &= \mathbb{E}\left[\mathbf{1}_A\right] \\ &= P(A) \quad \text{p.s.} \end{aligned}$$

Therefore the limit $\lim_{n\to\infty} \xi_n = \xi$ is also constant and equal to $P(A)$ a.s. This means that $P(A) = \mathbf{1}_A$ a.s., so $P(A) = 0$ or 1.

Chapter 5

Markov Chains

5.1 First Examples and Definitions

In some homes the use of the telephone can become quite a sensitive issue. Suppose that if the phone is free during some period of time, say the nth minute, then with probability p, where $0 < p < 1$, it will be busy during the next minute. If the phone has been busy during the nth minute, it will become free during the next minute with probability q, where $0 < q < 1$. Assume that the phone is free in the 0th minute. We would like to answer the following two questions.

- What is the probability x_n that the telephone will be free in the nth minute?
- What is $\lim_{n\to\infty} x_n$, if it exists?

Denote by A_n the event the event that the phone is free during the nth minute and let $B_n = \Omega \setminus A_n$ be its complement, i.e. the event that the phone is busy during the nth minute. The conditions of the example give us:

$$P(B_{n+1} \mid A_n) = p, \tag{5.1}$$
$$P(A_{n+1} \mid B_n) = q. \tag{5.2}$$

We also assume that $P(A_0) = 1$, i.e. $x_0 = 1$. Using this notation, we have $x_n = P(A_n)$. Then the total probability formula, together with (5.1)-(5.2) imply that:

$$\begin{aligned} x_{n+1} &= P(A_{n+1}) \\ &= P(A_{n+1} \mid A_n)P(A_n) + P(A_{n+1} \mid B_n)P(B_n) \\ &= (1-p)x_n + q(1-x_n) \\ &= q + (1-p-q)x_n. \end{aligned} \tag{5.3}$$

It's a bit tricky to find an explicit formula for x_n. To do so we suppose first that the sequence $\{x_n\}$ is convergent, i.e.

$$\lim_{n\to\infty} x_n = x. \tag{5.4}$$

The elementary properties of limits and equation (5.3), i.e. $x_{n+1} = q + (1-p-q)x_n$, yield

$$x = q + (1-p-q)x. \tag{5.5}$$

The unique solution to the last equation is:

$$x = \frac{q}{q+p}. \tag{5.6}$$

In particular,

$$\frac{q}{(p+q)} = q + (1-p-q)\frac{q}{p+q} \tag{5.7}$$

Subtracting (5.7) from (5.3), we infer that

$$x_{n+1} - \frac{q}{q+p} = (1-p-q)\left(x_n - \frac{q}{q+p}\right) \tag{5.8}$$

Thus, $\{x_n - \frac{q}{q+p}\}$ is a geometric sequence and therefore, for all $n \in \mathbb{N}$,

$$x_n - \frac{q}{q+p} = (1-p-q)^n \left(x_0 - \frac{q}{q+p}\right).$$

Hence, by taking into account the initial condition $x_0 = 1$, we have

$$\begin{aligned} x_n &= \frac{q}{q+p} + \left(x_0 - \frac{q}{q+p}\right)(1-p-q)^n \\ &= \frac{q}{q+p} + \frac{p}{q+p}(1-p-q)^n. \end{aligned} \tag{5.9}$$

Let us point out that although we have used the assumption (5.4) to derive (5.8), the proof of the latter is now complete. Indeed, having proven (5.9), we can show that the assumption (5.4) is indeed satisfied. This is because the conditions $0 < p, q < 1$ imply that $|1-p-| < 1$, and so $(1-p-q)^n \to 0$ as $n \to \infty$. Thus, (5.4) holds. This provides an answer to the second part of the example, i.e. $\lim_{n\to\infty} x_n = \frac{q}{p+q}$.

Remark 1: The formula (5.3) can be written in a compact form by using vector and matrix notation. First of all, since $x_n + y_n = 1$, we get

$$\begin{aligned} x_{n+1} &= (1-p)x_n + qy_n, \\ y_{n+1} &= px_n + (1-q)y_n. \end{aligned}$$

Hence, the matrix version takes the form

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} 1-p & q \\ p & 1-q \end{bmatrix} \begin{bmatrix} x_n \\ y_n \end{bmatrix}.$$

The situation described in Example 5.1 is quite typical. Often the probability of a certain event at time $n+1$ depends only on what happens at time n, but not further into the past. Example 5.1 provides us with a simple case of a Markov chain.

Definition 1: Suppose that S is a finite or a countable set. Suppose also that a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is given. An S-valued sequence of random variables ξ_n, $n \in \mathbb{N}$, is called an S-valued *Markov chain* or *Markov chain* on S if for all $n \in \mathbb{N}$ and all $s \in S$

$$P(\xi_{n+1} = s \mid \xi_0, \ldots, \xi_n) = P(\xi_{n+1} = s \mid \xi_n). \tag{5.10}$$

Here, $\mathbb{P}(\xi_{n+1} = s \mid \xi_n)$ is the conditional probability of the event $\{\xi_{n+1} = s\}$ with respect to the random variable ξ_n, or equivalently, with respect to the σ-field $\sigma(\xi_n)$ generated by ξ_n. Similarly, $\mathbb{P}(\xi_{n+1} = s \mid \xi_0, \ldots, \xi_n)$ is the conditional probability of the event $\{\xi_{n+1} = s\}$ with respect to the σ-field $\sigma(\xi_0, \ldots, \xi_n)$ generated by $\xi_0, \ldots, \xi_n$. Property (5.10) will usually be referred to as the *Markov property* of the Markov chain $(\xi_n)_{n\in\mathbb{N}}$. The set S is called the *state space*, and the elements of S are called *states*.

Definition 2: An S-valued Markov chain ξ_n, $n \in \mathbb{N}$, is called *time homogeneous* or *homogeneous* if for all $n \in \mathbb{N}$ and all $i, j \in S$

$$P(\xi_{n+1} = j \mid \xi_n = i) = P(\xi_1 = j \mid \xi_0 = i). \tag{5.11}$$

The number $P(\xi_1 = j \mid \xi_0 = i)$ is denoted by $p(j \mid i)$ and called the *transition probability* from state i to state j. The matrix $P = [p(j \mid i)]_{j,i\in S}$ is called the *transition matrix* of the chain ξ_n.

Definition 3: $A = [a_{ji}]_{i,j\in S}$ is called a stochastic matrix if

1. $a_{ji} \geq 0$, for all $i, j \in S$.
2. The sum of the entries in each column is 1, i.e.,

$$\sum_{j\in S} a_{ji} = 1, \quad \text{for any } i \in S.$$

A is called a *double stochastic matrix* if both A and its transpose A^t are stochastic matrices.

Proposition 2: Show that a stochastic matrix is doubly stochastic if and only if the sum of the entries in each row is 1, i.e. $\sum_{j\in S} a_{ji} = 1, \quad \text{for any } j \in S$.

Proof: Put $A^t = [b_{ij}]$. Then, by definition of the transposed matrix, $b_{ij} = a_{ji}$. Therefore, A^t is a stochastic matrix if and only if

$$\sum_{i\in S} a_{ji} = \sum_{i\in S} b_{ij} = 1$$

Definition 4: The *n-step transition matrix* of a Markov chain ξ_n with transition probabilities $p(j \mid i), j, i \in S$ is the matrix P_n with entries

$$P_n(j \mid i) = P(\xi_n = j \mid \xi_0 = i). \tag{5.12}$$

Proposition 3: (Chapman-Kolmogorov equation) Suppose that $\xi_n, n \in \mathbb{N}$, is an S-valued Markov chain with n-step transition probabilities $p_n(j \mid i)$. Then for all $k, n \in \mathbb{N}$

$$p_{n+k}(j \mid i) = \sum_{s\in S} p_n(j \mid s) p_k(s \mid i), \quad i, j \in S. \tag{5.13}$$

Proof: Let P and P_n be, respectively, the transition probability matrix and the n-step transition probability matrix. Since $p_n(j \mid i)$ are the entries of P_n, we only need to show that $P_n = P^n$ for all $n \in \mathbb{N}$. This can be done by induction. The assertion is clearly true for $n = 1$. Suppose that $P_n = P^n$. Then, for $i, j \in S$, by the total probability formula and the Markov property (5.10)

$$\begin{aligned}
p_{n+1}(j \mid i) &= P(\xi_{n+1} = j \mid \xi_0 = i) \\
&= \sum_{s \in S} P(\xi_{n+1} = j \mid \xi_0 = i, \xi_n = s) P(\xi_n = s \mid \xi_0 = i) \\
&= \sum_{s \in S} P(\xi_{n+1} = j \mid \xi_n = s) P(\xi_n = s \mid \xi_0 = i) \\
&= \sum_{s \in S} p(j \mid s) p_n(s \mid i).
\end{aligned}$$

which proves that $P_{n+1} = PP_n$

Proposition 4: For all $p \in (0, 1)$

$$P(\xi_n = i \mid \xi_0 = i) \to 0, \quad \text{as } n \to \infty. \tag{5.14}$$

Proof: To begin with, we shall consider the case $p \neq \frac{1}{2}$. When $j = i$, we have:

$$\begin{aligned}
P(\xi_n = j \mid \xi_0 = i) &= \binom{n}{\frac{n+j-i}{2}} p^{\frac{n+j-i}{2}} q^{\frac{n-j+i}{2}} \\
&= \begin{cases} \frac{(2k)!}{(k!)^2} (pq)^k, & \text{if } n = 2k, \\ 0, & \text{if } n \text{ is odd.} \end{cases}
\end{aligned}$$

Then, denoting $a_k = \frac{(2k)!}{(k!)^2}(pq)^k$, we have:

$$\frac{a_{k+1}}{a_k} = pq \frac{(2k+1)(2k+2)}{(k+1)^2} \to 4pq < 1.$$

Hence, $a_k \to 0$. Thus, $P(\xi_{2k} = i \mid \xi_0 = i) \to 0$. The result follows, since $P(\xi_{2k+1} = i \mid \xi_0 = i) = 0 \to 0$. This argument does not work for $p = \frac{1}{2}$ because $4pq = 1$. In this case we shall need the Stirling formula

$$k! \sim \sqrt{2\pi k} \left(\frac{k}{e}\right)^k, \quad \text{as } k \to \infty. \tag{5.15}$$

Here we use the standard convention: $a_n \sim b_n$ whenever $\frac{a_n}{b_n} \to 1$ lorsque $n \to \infty$. By (5.15),

$$a_k \sim \frac{\sqrt{4\pi k}}{2\pi k} (\frac{2k}{e})^{2k} (\frac{e}{k})^{2k} (pq)^k = \frac{1}{\sqrt{\pi k}} \to 0, \quad \text{as } k \to \infty.$$

Let us note that the second method works in the first case too. However, in the first case there is no need for anything as sophisticated as the Stirling formula.

Proposition 5: The probability that the random walk ξ_n ever returns to the starting point is $1 - |p - q|$.

Proof: Suppose that $\xi_0 = 0$ and denote by $f_0(n)$ the probability that the process returns to 0 at time n for the first time, i.e.,

$$f_0(n) = P(\xi_n = 0 \mid \xi_i \neq 0,\ i = 1, \ldots, n-1).$$

If also $p_0(n) = P(\xi_n = 0)$ for any $n \in \mathbb{N}$, then we can prove that

$$\sum_{n=1}^{\infty} p_0(n) = \sum_{n=0}^{\infty} p_0(n) \sum_{n=1}^{\infty} f_0(n). \tag{5.16}$$

Since all the numbers involved are non-negative, in order to prove (5.16) we need only to show that

$$p_0(n) = \sum_{k=1}^{n} f_0(k) p_0(n-k), \quad \text{for } n \geq 1.$$

The total probability formula and the Markov property (5.10) yield

$$\begin{aligned}
p_0(n) &= \sum_{k=1}^{n} P(\xi_n = 0, \xi_k = 0, \xi_i \neq 0,\ i = 1, \dots, k-1) \\
&= \sum_{k=1}^{n} P(\xi_k = 0, \xi_i \neq 0,\ i = 1, \dots, k-1) \\
&= \sum_{k=1}^{n} P(\xi_k = 0, \xi_i \neq 0,\ i = 1, \dots, k-1) P(\xi_n = 0 \mid \xi_k = 0) \\
&= \sum_{k=1}^{n} f_0(k) p_0(n-k).
\end{aligned}$$

Having proved (5.16), we are going to make use of it. First, we notice that the probability that the process will ever return to 0 equals $\sum_{n=1}^{\infty} f_0(n)$. Next, from (5.16), we infer that $P(\exists n \geq 1 : \xi_n = 0) = \sum_{n=1}^{\infty} f_0(n) = 1 - \left(\sum_{n=0}^{\infty} p_0(n)\right)^{-1} = 1 - \left(\sum_{k=0}^{\infty} p_0(2k)\right)^{-1}$. Since $p_0(2k) = \frac{(2k)!}{(k!)^2}(pq)^k$ and

$$\sum_{k=0}^{\infty} \binom{2k}{k} x^k = (1-4x)^{-1/2}, \quad |x| < \frac{1}{4} \tag{5.17}$$

it follows, that for $p \neq \frac{1}{2}$

$$P(\exists n \geq 1 : \xi_n = 0) = 1 - (1-4pq)^{1/2} = 1 - |p-q|. \tag{5.18}$$

since, recalling that $q = 1-p$, we have $1 - 4pq = 1 - 4p + 4p^2 = (1-2p)^2 = (q-p)^2$. The case $p = \frac{1}{2}$ is more delicate and we shall not pursue this topic here. Let us only remark that the case $p = \frac{1}{2}$ needs a special treatment as in Proposition 4.

Proposition 6: The probability of survival $P_m = \frac{\lambda^m e^{-\lambda}}{m!}, \quad m \in \mathbb{N}$ equals 0 if $\lambda \leq 1$, and $1 - \hat{r}^k$ if $\lambda > 1$, where k is the initial Vugiel population and $\hat{r} \in (0,1)$ is a solution to

$$r = e^{(r-1)\lambda} \tag{5.19}$$

Proof: We denote by $\varphi(i)$, $i \in \mathbb{N}$, the probability of dying out subject to the condition $\xi_0 = i$. Hence, if $A = \{\xi_n = 0 \text{ for } n \in \mathbb{N}\}$, then

$$\varphi(i) = P(A \mid \xi_0 = i). \tag{5.20}$$

Obviously, $\varphi(0) = 1$, and the total probability formula together with the Markov property (5.10) imply that for each $i \in \mathbb{N}$,

$$\begin{aligned}\varphi(i) &= \sum_{j=0}^{\infty} P(A \mid \xi_0 = i, \xi_1 = j) P(\xi_1 = j \mid \xi_0 = i) \\ &= \sum_{j=0}^{\infty} P(A \mid \xi_1 = j) P(\xi_1 = j \mid \xi_0 = i) \\ &= \sum_{j=0}^{\infty} \varphi(j) p(j \mid i).\end{aligned}$$

Therefore, the sequence $\varphi(i)$, $i \in \mathbb{N}$, is bounded (by 1 from above and by 0 from below) and satisfies the following system of equations:

$$\begin{aligned}\varphi(i) &= \sum_{j=0}^{\infty} \varphi(j) p(j \mid i), i \in \mathbb{N}, \\ \varphi(0) &= 1.\end{aligned} \tag{5.21}$$

So far, we have not used any particular distribution of X_j. From now on, we shall assume that the X_j have the Poisson distribution. Hence, we have $p(j \mid i) = \frac{(i\lambda)^j e^{-i\lambda}}{j!}$. It is not an easy problem to find a solution to (5.30), even in this special case. $\varphi(i)$ represents the probability that the population will die out, given that initially there were i individuals. Since we assume that reproduction of different individuals is independent, it is reasonable to make the following Ansatz:

$$\varphi(i) = A[\varphi(1)]^i, \quad i \in \mathbb{N}. \tag{5.22}$$

For some $A > 0$. Although it is possible to prove this Ansatz, we shall not do so here. Note that the boundary condition $\varphi(0) = 1$ implies that $A = 1$. Substituting (5.22) (with $A = 1$ and $r := \varphi(1)$) into (5.21), we get

$$\begin{aligned}r &= \sum_{j=0}^{\infty} r^j \frac{(i\lambda)^j}{j!} e^{-i\lambda} \\ &= e^{-i\lambda} \sum_{j=0}^{\infty} \frac{(i\lambda r)^j}{j!} \\ &= e^{-i\lambda} e^{i\lambda r}.\end{aligned}$$

Hence, r should satisfy

$$r = e^{(r-1)\lambda}.$$

Since the function $g(r) = e^{(r-1)\lambda}$, $r \in [0, 1]$, is convex, there exist at most two solutions to the equation (5.19). Obviously, one of them is $r = 1$. A bit of analysis, not included here, shows the following: 1) If $\lambda \leq 1$, then the only solution to (5.19) in $[0, 1]$ is $r = 1$. 2) If $\lambda > 1$, then there exists a second solution $f \in (0, 1)$ of the equation (5.19). In case 1) the situation is simple. We have $\varphi(i) = 1$ for all i, and thus the probability of extinction is 1 for any initial number of individuals. Case 2) is slightly more involved. The first question we need to address is which of the two solutions of (5.19) gives the correct value of $\varphi(1)$? Recall that $p_k = \frac{\lambda^k}{k!}e^{-\lambda}$. Define

$$F(x) = \sum_{k=0}^{\infty} p_k x^k = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} x^k = e^{\lambda(x-1)}, \quad |x| \le 1. \tag{5.23}$$

Since $P(\xi_1 = 0 \mid \xi_0 = 1) = p_0$ and

$$\begin{aligned} P(\xi_2 = 0 \mid \xi_0 = 1) &= \sum_{i=0}^{\infty} P(\xi_2 = 0 \mid \xi_1 = i) P(\xi_1 = i \mid \xi_0 = 1) \\ &= \sum_{i=0}^{\infty} (p_0)^i p_i \\ &= F(p_0) = F(F(0)), \end{aligned}$$

we guess that the following holds:

$$P(\xi_n = 0 \mid \xi_0 = 1) = F^{(n)}(0). \tag{5.24}$$

where $F^{(n)}$ is the n-fold composition of F. To prove (5.24), it is enough to prove it for n, while assuming it holds for $n-1$. We have

$$\begin{aligned} P(\xi_n = 0 \mid \xi_0 = 1) &= \sum_{i=0}^{\infty} P(\xi_n = 0 \mid \xi_1 = i) P(\xi_1 = i \mid \xi_0 = 1). \\ &= \sum_{i=0}^{\infty} p_i P(\xi_{n-1} = 0 \mid \xi_0 = i) \\ &= \sum_{i=0}^{\infty} p_i \left[P(\xi_{n-1} = 0 \mid \xi_0 = 1)\right]^i \\ &= \sum_{i=0}^{\infty} p_i \left[F^{(n-1)}(0)\right]^i \\ &= F(F^{(n-1)}(0)) = F^{(n)}(0). \end{aligned}$$

Since the event $\{\xi_n = 0\}$ is contained in the events $\{\xi_{n+1} = 0\}$ for all $n \in \mathbb{N}$, we have

$$\begin{aligned} \varphi(1) &= P(\xi_n = 0 \text{ for some } n \in \mathbb{N} \mid \xi_0 = 1) \\ &= \lim_{n\to\infty} P(\xi_n = 0 \mid \xi_0 = 1), \end{aligned}$$

by the Lebesgue monotone convergence theorem. Therefore, we infer that

$$\varphi(1) = \lim_{n\to\infty} F^{(n)}(0).$$

With $F^{(0)}(x) = x$, we only need to show that

$$F^{(n)}(0) \le \hat{r}, \quad n \in \mathbb{N}. \tag{5.25}$$

Indeed, once the inequality (5.35) is proven, we infer that $\varphi(1) \le \hat{r}$, and thus $\varphi(1) = r$. We shall prove (5.25) by induction. It is obviously valid for $n = 0$, so we need to study the inductive step. We have

$$F^{(n)}(0) = F(F^{(n-1)}(0)) \le F(\hat{r}) = \hat{r},$$

Since F is increasing, we conclude that in the case $\lambda > 1$, q the population will become extinct with positive probability. In the simplest example of the binomial distribution case, i.e., when $N = 1$, equations (5.21) become

$$\varphi(i) = \sum_{j=0}^{i} \varphi(j) \binom{i}{j} p^j (1-p)^{i-j}, \quad i \in \mathbb{N}.$$

Since $\varphi(0) = 1$, $\varphi(1)$ satisfies

$$\varphi(1) = q + p\varphi(1),$$

with $q = 1 - p$. Hence, trivially, $\varphi(1) = 1$. Then, by induction, one proves that $\varphi(i) = 1$. Therefore, whatever the initial number of individuals, extinction of the species is certain.

5.2 Classification of States

In what follows we fix an S-valued Markov chain with transition matrix $P = [p(j|i)]_{j,i\in S}$, where S is a non-empty and at most countable set.

Definition 5: A state i is called, *recurrent* if the process ξ_n will eventually return to i given that it starts at i, i.e.

$$\mathbb{P}(\xi_n = 1 \text{ for some } n \geq 1 \mid \xi_0 = i) = 1 \tag{5.26}$$

If the condition (5.26) is not satisfied, then the state i is called transient.

Theorem 1: Show that for a random walk on $\mathbb{Z}$ with parameters $p \in (0, 1)$, the state 0 is recurrent if and only if $p = \frac{1}{2}$. Show that the same holds if 0 is replaced by any other state $i \in \mathbb{Z}$.

Proof: We know that $\mathbb{P}(\xi_n = i \text{ for some } n \geq 1 \mid \xi_0 = i) = 1 - |p - q|$ for any $i \in \mathbb{Z}$.

Definition 6: We say that a state i communicates with a state j if with positive probability the chain will visit the state j having started at i, i.e.

$$\mathbb{P}(\xi_n = j \text{ for some } n \geq 0 \mid \xi_0 = i) > 0 \tag{5.27}$$

If i communicates with j, then we shall write $i \to j$. We say that the state i intercommunicates with a state j, and write $i \leftrightarrow j$, if $i \to j$ and $j \to i$.

Theorem 2: A state $j \in S$ is recurrent if and only if

$$\mathbb{P}\left(\xi_n = j \text{ for infinitely many } n \mid \xi_0 = j\right) = 1,$$

and is transient if and only if

$$\mathbb{P}\left(\xi_n = j \text{ for infinitely many } n \mid \xi_0 = j\right) = 0.$$

Definition 7: For an S-valued Markov chain ξ_n, $n \in \mathbb{N}$, a state $i \in S$ is called *null-recurrent* if it is recurrent and it mean recurrence time m_i defined by

$$m_i := \sum_{n=0}^{\infty} n f_n(i \mid i) \tag{5.28}$$

equals ∞. A state $i \in S$ is called positive-recurrent if it is recurrent and its mean recurrence time m_i is finite.

Remark 4: One can show that for a random walk on $\mathbb{Z}$ the state 0 is recurrent if and only if $p = \frac{1}{2}$, i.e. if and only if the random walk is symmetric. In the following problem we shall try to answer if 0 is a null-recurrent or positive-recurrent state (when $p = \frac{1}{2}$).

Definition 8: Suppose that ξ_n, $n \in \mathbb{N}$, is a Markov chain on a state space S. Let $i \in S$. We say that i is a periodic state if and only if the greatest common divisor (gcd) of all $n \in \mathbb{N}^*$, where $\mathbb{N}^* = 1, 2, 3, ...$, such that $p_n(i|i) > 0$ is ≥ 2. Otherwise, the state i is called aperiodic. In both cases, the gdc is denoted by $d(i)$ and is called the period of the state i. Thus, i is periodic if and only if $d(i) \geq 2$. A state i which is positive recurrent and aperiodic is called *ergodic*.

Proposition 7: Suppose that $i, j \in S$ and $i \leftrightarrow j$. Show that

1. i is transient if and only if j is;
2. i is recurrent if and only if j is;
3. i is null-recurrent if and only if j is;
4. i is positive-recurrent if and only if j is;
5. i is periodic if and only if j is, in which case $d(i) = d(j)$;
6. i is ergodic if and only if j is.

Proof: It is enough to show properties 1), 4) and 5). Since $i \leftrightarrow j$ one can find $n, m \in \mathbb{N}$ such that $p_m(j \mid i) > 0$ and $p_n(i \mid j) > 0$. Hence $\epsilon := p_m(j \mid i)p_n(i \mid j)$ is positive. Let us take $k \in \mathbb{N}$. Then by the Chapman-Kolmogorov equations

$$p_{m+k+n}(j \mid j) = \sum_{r,s \in S} p_m(j \mid s)\, p_k(s \mid r)\, p_n(r \mid j) \geq p_m(j \mid i)\, p_k(i \mid i)\, p_n(i \mid j) = \epsilon\, p_k(i \mid i).$$

By symmetry

$$p_{m+k+n}(i \mid i) = \sum_{r,s \in S} p_n(i \mid s)\, p_k(s \mid r)\, p_m(r \mid i) \geq p_n(i \mid j)\, p_k(j \mid j)\, p_m(j \mid i) = \epsilon\, p_k(j \mid j).$$

Hence, the series $\sum_k p_k(i \mid i)$ and $\sum_k p_k(j \mid j)$ are simultaneously convergent or divergent. To prove 5) it is enough to show that

$$d(i) \leq d(j)$$

Using the first inequality derived above, we have

$$p_{n+k+m}(i \mid i) \geq \epsilon p_k(j \mid j)$$

for all $k \in \mathbb{N}$. From this inequality we can draw two conclusions:

(a) $d(i) \mid n + m$, since by taking $k = 0$ we get $p_{n+m}(i \mid i) > 0$;

(b) if $p_k(j \mid j) > 0$, then $P_{n+k+m}(i \mid i) > 0$.

From (a) and (b), we can see that $d(i) \mid k$ provided that $p_k(j \mid j) > 0$. This proves what is required.

Definition 9: Suppose that ξ_n, $n \in \mathbb{N}$, is a Markkv chain on a countable state space S.

1. A set $C \subset S$ is called *closed* if once the chain enters C it will never leave it, i.e.,

$$P\left(\xi_k \in S \setminus C \text{ for some } k \geq n \,|\, \xi_n \in C\right) = 0 \tag{5.29}$$

2. A set $C \subset S$ is called *irreducible* if any two elements i, j of C intercommunicate, i.e., for all $i, j \in C$, there exists an $n \in \mathbb{N}$ such that

$$p_n(j \mid i) > 0.$$

Theorem 3: Suppose that $\{\xi_n\}_{n\in\mathbb{N}}$ is a Markov chain on a countable state space S. Then

$$S = T \cup \bigcup_{j=1}^{N} C_j \tag{5.30}$$

(disjoint union), where T is the set of all transient states in S, and each C_j is a closed irreducible set of recurrent states.

Proof: Let $R = S \setminus T$ denote the set of all recurrent states. If $i \leftrightarrow j$, then both i and j belong either to T or to R. It follows that the interconnection relation $\leftrightarrow$ restricted to R is an equivalence relation as well. Therefore,

$$R = \bigcup_{j=1}^{N} C_j, \quad C_j = [s_j],\; s_j \in R.$$

Here, N denotes the number of different equivalence classes. Since, by definition, each C_j is an irreducible set, we only need to show that it is closed. Indeed, if $i \in C_k$ and $i \to j$, then $i \leftrightarrow j$, and so $j \in C_k$.

5.3 Long-Time Behaviour of Markov Chains: General Case

For convenience, we shall denote the countable state space S by $\{1, 2, 3, \ldots\}$ when S is an infinite set, and by $\{1, 2, \ldots, n\}$ when S is finite.

Proposition 8: Let $P = [p(j \mid i)]$ be the transition matrix of a Markov chain with state space S. Suppose that for all $i, j \in S$,

$$\lim_{n\to\infty} p_n(j \mid i) =: \pi_j. \tag{5.31}$$

(In particular, the limit is independent of i.) Then:

1. $\sum_j \pi_j \leq 1$;
2. $\sum_i p(j \mid i)\, \pi_i = \pi_j$;

3. either $\sum_j \pi_j = 1$, or $\pi_j = 0$ for all $j \in S$.

Proof: To begin with, let us assume that S is finite with m elements. Using the Chapman-Kolmogorov equations (5.13), we have

$$\begin{aligned}
\sum_{j\in S} \pi_j &= \sum_{j=1}^m \pi_j \\
&= \sum_{j=1}^m \left(\lim_{n\to\infty} p_n(j \mid i) \right), \\
&= \lim_{n\to\infty} \sum_{j=1}^m p_n(j \mid i) = \lim_{n\to\infty} 1 = 1.
\end{aligned}$$

since $\sum_{j=1}^m p_n(j \mid i) = 1$ for any $n \in \mathbb{N}$. This proves (1) and (3) simultaneously. Moreover, it shows that the second alternative in (3) can never occur. To prove (2), we argue in a similar way. Let us fix $j \in S$ and (as an auxiliary index) $k \in S$. Then,

$$\begin{aligned}
\sum_{i=1}^m p(j \mid i)\, \pi_i &= \sum_{i=1}^m \left(\lim_{n\to\infty} p(j \mid i) p_n(i \mid k) \right) \\
&= \lim_{n\to\infty} \sum_{i=1}^m p(j \mid i)\, p_n(i \mid k) \\
&= \lim_{n\to\infty} p_{n+1}(j \mid k) = \pi_j,
\end{aligned}$$

since $\sum_{i=1}^m p(j \mid i)\, p_n(i \mid k) = p_{n+1}(j \mid k)$ by the Chapman-Kolmogorov equations. When the set S is infinite, we cannot just repeat the above argument. The reason is quite simple: in general, the two operations lim and $\sum$ cannot be interchanged. They can when the sum is finite, and we used this fact above. But if S is infinite, then the situation is more subtle. One possible solution to this difficulty is contained in the following version of Fatou's lemma.

Lemma 1: (Fatou) Suppose that $a_j(n) \geq 0$ for $j, n \in \mathbb{N}$. Then

$$\sum_j \liminf_{n\to\infty} a_j(n) \leq \liminf_{n\to\infty} \sum_j a_j(n). \tag{5.32}$$

Moreover, if $a_j(n) \leq b_j$ for all $j, n \in \mathbb{N}$ and $\sum_j b_j < \infty$, then

$$\limsup_{n\to\infty} \sum_j a_j(n) \leq \sum_j \limsup_{n\to\infty} a_j(n). \tag{5.33}$$

Proof: Using the fact that for a convergent sequence lim and lim inf coincide, by the Fatou lemma we have

$$\begin{aligned}\sum_{j\in S}\pi_j &= \sum_{j=1}^{\infty}\pi_j\\ &= \sum_{j=1}^{\infty}\left(\lim_{n\to\infty}p_n(j\mid i)\right)\\ &\le \liminf_{n\to\infty}\sum_{j=1}^{\infty}p_n(j\mid i)\\ &= \liminf_{n\to\infty}1 = 1,\end{aligned}$$

since, as before, $\sum_{j=1}^{\infty}p_n(j\mid i)=1$ for any $n\in\mathbb{N}$. This proves (1). A similar argument shows (2). Indeed, with $j\in S$ and $k\in S$ fixed, by the Chapman-Kolmogorov equations and the Fatou lemma we have

$$\sum_{i=1}^{\infty}\pi_i\,p(j\mid i)=\sum_{i=1}^{\infty}\left(\lim_{n\to\infty}p_n(i\mid k)\right)p(j\mid i)\le\liminf_{n\to\infty}\sum_{i=1}^{\infty}p_n(i\mid k)\,p(j\mid i)=\liminf_{n\to\infty}p_{n+1}(j\mid k)=\pi_j.$$

To complete the proof of (2), suppose that for some $k\in S$

$$\sum_{i=1}^{\infty}p(k\mid i)\,\pi_i<\pi_k.$$

Then, since $\sum_{j\in S}\pi_j=\sum_{j\neq k}\pi_j+\pi_k$, and using the part of (2) already proven, we have

$$\begin{aligned}\sum_{j=1}^{\infty}\pi_j &> \sum_{j\neq k}\left(\sum_{i}^{\infty}p(j\mid i)\,\pi_i\right)+\sum_{i}^{\infty}p(k\mid i)\,\pi_i\\ &= \sum_{j=1}^{\infty}\sum_{i}^{\infty}p(j\mid i)\,\pi_i\\ &= \sum_{i=1}^{\infty}\sum_{j}^{\infty}p(j\mid i)\,\pi_i\\ &= \sum_{i=1}^{\infty}\pi_i\left(\sum_{j}^{\infty}p(j\mid i)\right)\\ &= \sum_{i=1}^{\infty}\pi_i.\end{aligned}$$

We used the fact that $\sum_j p(j\mid i)=1$ together with (5.65). This contradiction proves (2). In order to verify (3), observe that by iterating (2), we obtain

$$\sum_i p_n(j\mid i)\,\pi_i=\pi_j.$$

Hence,

$$\pi_j=\lim_{n\to\infty}\sum_i p_n(j\mid i)\,\pi_i$$

And,

$$\sum_i \lim_{n\to\infty} p_n(j \mid i)\, \pi_i = \sum_i \pi_j\, \pi_i = \pi_j \sum_i \pi_i.$$

Therefore, the product $\pi_j\,(\sum_i \pi_i - 1)$ is equal to 0 for all $j \in S$. As a result, (3) follows. Indeed, if $\sum_i \pi_i \neq 1$, then $\pi_j = 0$ for all $j \in S$.

Definition 10: A probability measure $\mu := \sum_{j\in S} \mu_j\, \delta_j$ is an invariant measure of a Markov chain $\{\xi_n\}_{n\in\mathbb{N}}$ with transition probability matrix $P = [p(j \mid i)]$ if for all $n \in \mathbb{N}$ and all $j \in S$

$$\sum_{i\in S} p_n(j \mid i)\, \mu_i = \mu_j.$$

Theorem 4: Suppose that $\{\xi_n\}_{n\in\mathbb{N}}$ is a Markov chain on a state space $S = T \cup C$, where T is the set of all transient states and C is a closed irreducible set of recurrent states. Then, there exists an invariant measure if and only if each element of C is positive-recurrent. Moreover, if this is the case, then the invariant measure is unique and it is given by $\mu = \sum_i \mu_i\, \delta_i$, where

$$\mu_i = \frac{1}{m_i},$$

with m_i being the mean recurrence time of the state i (see equation (5.28)).

Remark 5: If $C = \bigcup_{j=1}^{N} C_j$, where each C_j is a closed irreducible set of recurrent states, then the above result holds, except for the uniqueness part. In fact, if each element of some C_j is positive-recurrent, then there exists an invariant measure μ_j supported on C_j. Moreover, μ_j is the unique invariant measure with support in C_j. In the special case when each element of C is positive-recurrent, every invariant measure μ is a convex combination of the invariant measures μ_j, with $j \in \{1, \ldots, N\}$.

Theorem 5: Suppose that $\{\xi_n\}_{n\in\mathbb{N}}$ is a Markov chain with state space S. Let $j \in S$ be a recurrent state.

1. If j is aperiodic, then
$$p_n(j \mid j) \longrightarrow \frac{1}{m_j} \tag{5.34}$$
Moreover, for any $i \in S$,
$$p_n(j \mid i) \longrightarrow \frac{F_{ji}(1)}{m_j} \tag{5.35}$$
where $F_{ji}(1)$ is the probability that the chain will ever visit state j if it starts at i, and m_j is the mean recurrence time of state j.

2. If j is a periodic state of period $d \geq 2$, then
$$p_{nd}(j \mid j) \longrightarrow \frac{d}{m_j}. \tag{5.36}$$

Definition 11: A Markov chain $\{\xi_n\}_{n\in\mathbb{N}}$ with state space S is called *ergodic* if each $i \in S$ is ergodic, i.e., each state $i \in S$ is positive recurrent and aperiodic.

Theorem 6: Suppose that $\{\xi_n\}_{n\in\mathbb{N}}$ is an irreducible aperiodic Markov chain with state space S. Then $\{\xi_n\}_{n\in\mathbb{N}}$ is ergodic if and only if it has a unique invariant measure.

Proof: We shall deal with the "only if" part. Suppose that $\pi = \sum_j \pi_j \, \delta_j$ is the unique invariant measure of the chain. Then $\pi_j > 0$ for some $j \in S$. Since $\sum_i p_n(j \mid i) \, \pi_i = \pi_j$, by the Fatou lemma,

$$\sum_i \lim_{n\to\infty} p_n(j \mid i) \, \pi_i \geq \limsup_{n\to\infty} \sum_i p_n(j \mid i) \, \pi_i = \pi_j.$$

Hence, there exists some $i \in S$ such that

$$\lim_{n\to\infty} p_n(j \mid i) \, \pi_i > 0.$$

Therefore, $\lim_{n\to\infty} p_n(j \mid i) > 0$, which, in view of Theorem 5, implies that $m_j < \infty$. Thus, j is an ergodic state and, since the chain is irreducible, all states are ergodic as well.

5.4 Long-Time Behaviour of Markov Chains with Finite State Space

Theorem 7: Suppose that S is finite and the transition matrix $P = [p(j \mid i)]$ of a Markov chain on S satisfies the condition

$$\exists n_0 \in \mathbb{N}, \ \exists \varepsilon > 0 \text{ such that } p_{n_0}(j \mid i) \geq \varepsilon, \quad \forall i, j \in S. \tag{5.37}$$

Then, the following limit exists for all $i, j \in S$ and is independent of i:

$$\lim_{n\to\infty} p_n(j \mid i) = \pi_j. \tag{5.38}$$

The numbers π_j satisfy

$$\pi_j > 0 \quad \text{for all } j \in S, \quad \text{and} \quad \sum_{j\in S} \pi_j = 1. \tag{5.39}$$

Conversely, if a sequence of numbers π_j, $j \in S$, satisfies conditions (5.38)–(5.39), then assumption (5.37) is also satisfied.

Proof: Denote the matrix $P^{n_0} = [p_{n_0}(j \mid i)]$ by $Q = [q(j \mid i)]$. Then the process $\eta_k = \{\xi_{kn_0}\}_{k\in\mathbb{N}}$ is a Markov chain on S with transition probability matrix Q satisfying (5.37) with n_0 equal to 1. Note that $p_{kn_0}(j \mid i) = q_k(j \mid i)$ due to the Chapman-Kolmogorov equations. Suppose that the properties (5.38)–(5.39) hold true for Q. In particular, $\lim_{k\to\infty} p_{kn_0}(j \mid i) = \pi_j$ exists and is independent of i. We claim that they are also true for the original matrix P. Obviously, one only needs to check condition (5.38). The Chapman-Kolmogorov equations (and the fact that S is finite) imply that for any $r = 1, \ldots, n_0 - 1$

$$p_{kn_0+r}(j \mid i) = \sum_{s\in S} p_{kn_0}(j \mid s) \, p_r(s \mid i) \to \sum_{s\in S} \pi_j \, p_r(s \mid i) = \pi_j \sum_{s\in S} p_r(s \mid i) = \pi_j.$$

Therefore, by a simple result in calculus, according to which, if for a sequence a_n, $n \in \mathbb{N}$, there exists a natural number n_0 such that for each $r \in \{0, 1, \ldots, n_0 - 1\}$ the limit $\lim_{k\to\infty} a_{kn_0+r}$ exists and is r-independent, then the sequence a_n is convergent to the common limit of those subsequences, we infer that (5.38) is satisfied. In what follows we shall assume that (5.37) holds with $n_0 = 1$. Let us put $p_0(j \mid i) = \delta_{ji}$ and for $j \in S$

$$m_n(j) := \min_{i \in S} p_n(j \mid i), \quad M_n(j) := \max_{i \in S} p_n(j \mid i).$$

Observe that $M_0(j) = 1$ and $m_0(j) = 0$ for all $j \in S$. From the Chapman-Kolmogorov equations, it follows that the sequence $M_n(j)$, $n \in \mathbb{N}$, is decreasing, while the sequence $m_n(j)$, $n \in \mathbb{N}$, is increasing. Indeed, since $\sum_k p(k \mid i) = 1$,

$$p_{n+1}(j \mid i) = \sum_{k \in S} p_n(j \mid k)\, p(k \mid i) \geq \min_k p_n(j \mid k) \sum_{k \in S} p(k \mid i) = \min_k p_n(j \mid k) = m_n(j).$$

Hence, by taking the minimum over all $i \in S$, we arrive at

$$m_{n+1}(j) = \min_i p_{n+1}(j \mid i) \geq m_n(j).$$

Similarly,

$$p_{n+1}(j \mid i) = \sum_{k \in S} p_n(j \mid k)\, p(k \mid i) \leq \max_k p_n(j \mid k) \sum_{k \in S} p(k \mid i) = \max_k p_n(j \mid k) = M_n(j).$$

Hence, by taking the maximum over all $i \in S$, we obtain

$$M_{n+1}(j) = \max_i p_{n+1}(j \mid i) \leq M_n(j).$$

Since $M_n(j) \geq m_n(j)$, the sequences $M_n(j)$ and $m_n(j)$ are bounded from below and from above, respectively. As a consequence, they both have limits. To show that the limits coincide, we shall prove that

$$\lim_{n \to \infty} \left(M_n(j) - m_n(j)\right) = 0. \tag{5.40}$$

For $n \geq 0$ we have

$$\begin{aligned} p_{n+1}(j \mid i) &= \sum_{s \in S} p_n(j \mid s)\, p(s \mid i) \qquad (5.41)\\ &= \sum_{s \in S} p_n(j \mid s)\, [p(s \mid i) - \varepsilon p_n(s \mid j)] + \varepsilon \sum_{s \in S} p_n(j \mid s)\, p_n(s \mid j) \\ &= \sum_{s \in S} p_n(j \mid s)\, [p(s \mid i) - \varepsilon p_n(s \mid j)] + \varepsilon p_{2n}(j \mid j) \end{aligned}$$

by the Chapman–Kolmogorov equations. The expression in square brackets is ≥ 0. Indeed, by assumption (5.37), $p(s \mid i) \geq \varepsilon$ and $p_n(s \mid j) \leq 1$. Therefore,

$$\begin{aligned} p_{n+1}(j \mid i) &\geq \min_{s \in S} p_n(j \mid s) \sum_{s \in S} [p(s \mid i) - \varepsilon p_n(s \mid j)] + \varepsilon p_{2n}(j \mid j) \\ &= (1 - \varepsilon) m_n(j) + \varepsilon p_{2n}(j \mid j). \qquad (5.42) \end{aligned}$$

By taking the minimum over $i \in S$, we arrive at

$$m_{n+1}(j) \geq (1 - \varepsilon) m_n(j) + \varepsilon p_{2n}(j \mid j). \tag{5.43}$$

Recycling the above argument, we obtain a similar inequality for the sequence $M_n(j)$:

$$M_{n+1}(j) \leq (1-\varepsilon)M_n(j) + \varepsilon p_{2n}(j \mid j). \tag{5.44}$$

Thus, by subtracting (5.43) from (5.44), we get

$$M_{n+1}(j) - m_{n+1}(j) \leq (1-\varepsilon)\left(M_n(j) - m_n(j)\right). \tag{5.45}$$

Hence, by induction

$$M_n(j) - m_n(j) \leq (1-\varepsilon)^n, \quad n \in \mathbb{N}.$$

This proves (5.40). Denote by π_j the common limit of $M_n(j)$ and $m_n(j)$. Then (5.38) follows from (5.40). Indeed, if $i, j \in S$, then

$$m_n(j) \leq p_n(j \mid i) \leq M_n(j).$$

To prove that $\pi_j > 0$, let us recall that $m_n(j)$ is an increasing sequence and $m_1(j) \geq \varepsilon$ by (5.37). We infer that $\pi_j \geq \varepsilon$.

Theorem 8: Suppose that the transition matrix $P = [p(j \mid i)]$ of a Markov chain $(\xi_n)_{n\in\mathbb{N}}$ satisfies assumption (5.37). Show that there exists a unique invariant measure μ. Moreover, for some $A > 0$ and $\alpha < 1$, we have

$$|p_n(j \mid i) - \pi_j| \leq A\alpha^n, \quad \text{for all } i, j \in S, \; n \in \mathbb{N}. \tag{5.46}$$

Proof: Put

$$\epsilon = \frac{1}{2}\min_j \pi_j.$$

Since $p_n(j \mid i) \to \pi_j$ for all $i, j \in S$, there exists $n_0 \in \mathbb{N}$ such that $p_k(j \mid i) \geq \epsilon$ for all $k \geq n_0$ and $(i,j) \in S^2$. Putting $k = n_0$ proves that (5.37) is satisfied. Let us observe that we have used only two facts: $\pi_j > 0$ for all $j \in S$, and $p_n(j \mid i) \to \pi_j$ for all $i, j \in S$.

Chapter 6

Stochastic Processes in Continious Time

6.1 General Notions

Definition 1: A stochastic process is a family of random variables $\xi(t)$ parametrized by $t \in T$, where $T \subset \mathbb{R}$. When $T = \{1, 2, \dots\}$, we shall say that $\xi(t)$ is a stochastic process in *discrete time* (i.e., a sequence of random variables). When T is an interval in $\mathbb{R}$ (typically $T = [0, \infty)$), we shall say that $\xi(t)$ is a stochastic process in *continuous time*. For every $\omega \in \Omega$, the function

$$T \ni t \mapsto \xi(t, \omega)$$

is called a *path* (or *sample path*) of $\xi(t)$.

Definition 2: A family $\{\mathcal{F}_t\}_{t \in T}$ of σ-fields on Ω parametrized by $t \in T$, where $T \subset \mathbb{R}$, is called a *filtration* if

$$\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$$

for any $s, t \in T$ such that $s \leq t$.

Definition 3: A stochastic process $\xi(t)$ parametrized by $t \in T$ is called a martingale (submartingale, supermartingale) with respect to a filtration $\{\mathcal{F}_t\}$ if

- $\xi(t)$ is integrable for each $t \in T$;
- $\xi(t)$ is $\mathcal{F}_t$-measurable for each $t \in T$ (in which case we say that $\xi(t)$ is adapted to $\mathcal{F}_t$);
- $\mathbb{E}[\xi(t) \mid \mathcal{F}_s] = \xi(s)$ (respectively, $\geq$ or $\leq$) for every $s, t \in T$ such that $s \leq t$.

6.2 Poisson Process

6.2.1 Exponential Distribution and Lack of Memory

Definition 4: We say that a random variable η has the exponential distribution of rate $\lambda > 0$ if

$$\mathbb{P}(\eta > t) = e^{-\lambda t} \quad \text{for all } t \geq 0.$$

For example, the emissions of particles by a sample of radioactive material (or calls made at a telephone exchange) occur at random times. The probability that no particle is emitted (no call is made) up to time t is known to decay exponentially as t increases. That is to say, the time η of the first emission has the exponential distribution:

$$\mathbb{P}(\eta > t) = e^{-\lambda t}.$$

6.2.2 Construction of the Poisson Process

Let $\eta_1, \eta_2, \ldots$ be a sequence of independent random variables, each having the exponential distribution with rate λ. For example, the times between emissions of radioactive particles (or between phone calls at a telephone exchange) have this property. We define:

$$\xi_n = \eta_1 + \cdots + \eta_n,$$

which can be interpreted as the time of the nth emission (or nth call). For convenience, we also set $\xi_0 = 0$. The number of emissions (or calls) up to time $t \geq 0$ is the integer n such that $T_{n+1} > t \geq T_n$. In other words, this number is given by:

$$\max\{n : T_n \leq t\}.$$

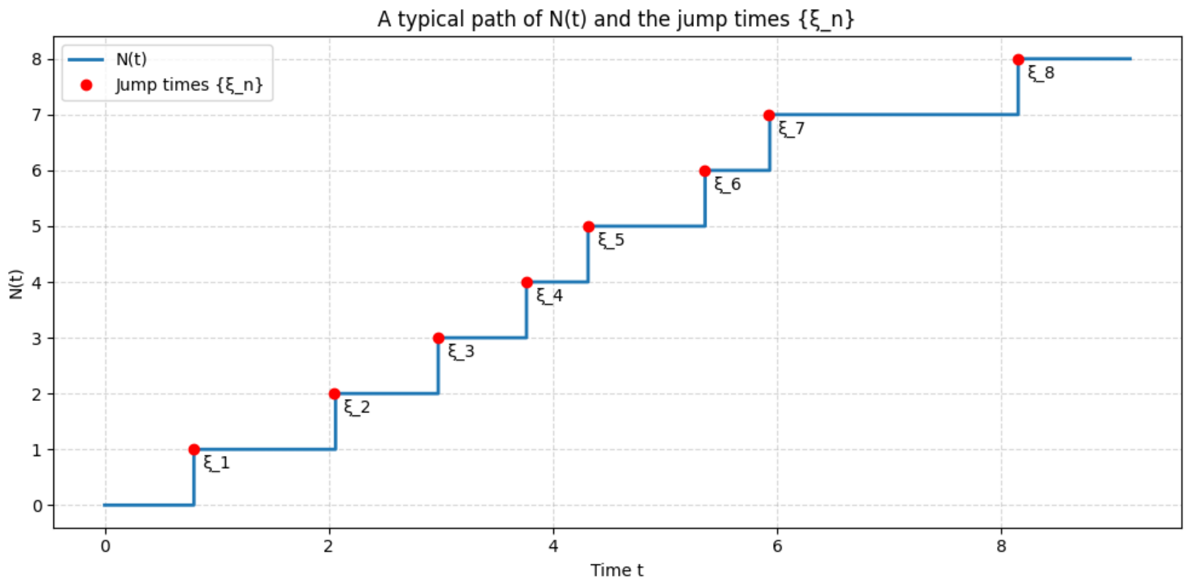
Definition 5: We say that $N(t)$, where $t \geq 0$, is a *Poisson process* if

$$N(t) = \max\{n : t \geq \xi_n\}.$$

Thus, $N(t)$ can be regarded as the number of particles emitted (or calls made) up to time t. It is an example of a stochastic process in continuous time. A typical path of $N(t)$ is shown in the figure below. It begins at the origin, $N(0) = 0$ (no particles emitted at time 0), and is right-continuous, non-decreasing, and piecewise constant with jumps of size 1 at the times ξ_n.

```python
import matplotlib.pyplot as plt
import numpy as np

# Simuler un exemple de processus de Poisson
np.random.seed(0)
n_jumps = 8
jump_times = np.cumsum(np.random.exponential(scale=1.0, size=n_jumps))
t_max = jump_times[-1] + 1
t_values = np.linspace(0, t_max, 1000)
N_t = np.searchsorted(jump_times, t_values)

# Tracer la trajectoire typique
plt.figure(figsize=(10, 5))
plt.step(t_values, N_t, where='post', label='N(t)', linewidth=2)
plt.plot(jump_times, np.arange(1, n_jumps + 1), 'ro', label='Jump times
↪  {xi_n}')

for i, jt in enumerate(jump_times):
    plt.annotate(f'xi_{i+1}', (jt, i + 0.5), textcoords="offset points",
    ↪  xytext=(5, 5), fontsize=10)
```

```
plt.title("A typical path of N(t) and the jump times {xi_n}")
plt.xlabel("Time t")
plt.ylabel("N(t)")
plt.grid(True, linestyle='--', alpha=0.5)
plt.legend()
plt.tight_layout()
plt.show()
```



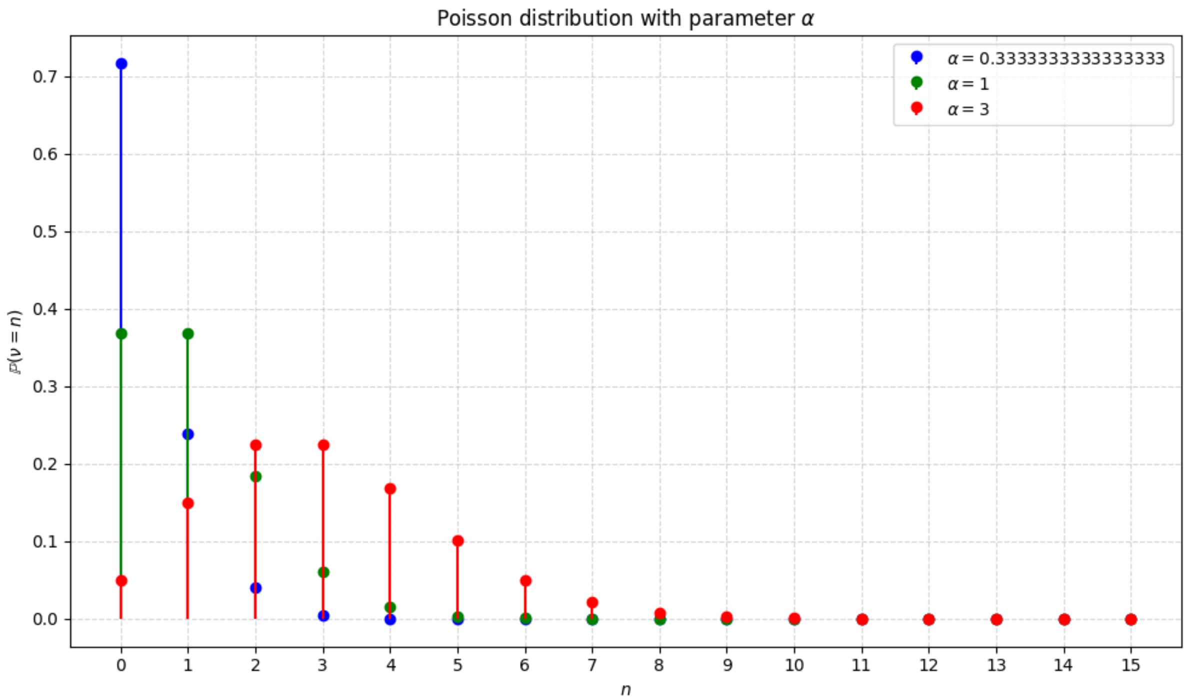
Definition 6: A random variable ν has the Poisson distribution with parameter $\alpha > 0$ if, for any $n = 0, 1, 2, \ldots$,

$$\mathbb{P}(\nu = n) = \frac{\alpha^n e^{-\alpha}}{n!}.$$

The probabilities $\mathbb{P}(\nu = n)$ for various values of a are shown in the figure below.

```
import matplotlib.pyplot as plt
import numpy as np
from scipy.stats import poisson

# Poisson's Parameters
alphas = [1/3, 1, 3]
colors = ['blue', 'green', 'red']
n_max = 15  # Valeur maximale de n
n_values = np.arange(0, n_max + 1)

plt.figure(figsize=(10, 6))
for alpha, color in zip(alphas, colors):
    pmf = poisson.pmf(n_values, mu=alpha)
    markerline, stemlines, baseline = plt.stem(n_values, pmf, basefmt=" ",
    ↪  label=fr'$\alpha = {alpha}$')
```

```
    plt.setp(markerline, color=color)
    plt.setp(stemlines, color=color)

plt.title("Poisson distribution with parameter $\\alpha$")
plt.xlabel("$n$")
plt.ylabel(r"$\mathbb{P}(\nu = n)$")
plt.xticks(n_values)
plt.grid(True, linestyle="--", alpha=0.5)
plt.legend()
plt.tight_layout()
plt.show()
```



Proposition 1: $N(t)$ has the Poisson distribution with parameter λt, that is,

$$\mathbb{P}(N(t) = n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}, \quad n = 0, 1, 2, \ldots$$

Proof: First of all, observe that

$$\{N(t) < n\} = \{\xi_n > t\}.$$

It suffices to compute the probability of this event for any n, because

$$\begin{aligned} \mathbb{P}(N(t) = n) &= \mathbb{P}(N(t) < n + 1) - \mathbb{P}(N(t) < n) \\ &= \mathbb{P}(\xi_{n+1} > t) - \mathbb{P}(\xi_n > t). \end{aligned} \tag{6.1}$$

We shall prove by induction on n that

$$\mathbb{P}(\xi_n > t) = e^{-\lambda t} \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!}. \tag{6.2}$$

For $n = 1$,

$$\mathbb{P}(\xi_1 > t) = \mathbb{P}(\eta_1 > t) = e^{-\lambda t}.$$

Next, suppose that (6.2) holds for some n. Then, expressing ξ_{n+1} as the sum of the independent random variables ξ_n and η_{n+1}, we compute:

$$\begin{aligned}
\mathbb{P}(\xi_{n+1} > t) &= \mathbb{P}(\xi_n + \eta_{n+1} > t) \\
&= \mathbb{P}(\eta_{n+1} > t) + \mathbb{P}(\xi_n > t - \eta_{n+1},\ \eta_{n+1} \le t) \\
&= e^{-\lambda t} + \int_0^t \mathbb{P}(\xi_n > t - s)\, f_{\eta_{n+1}}(s)\, ds,
\end{aligned}$$

where $f_{\eta_{n+1}}(s) = \lambda e^{-\lambda s}$ is the density of η_{n+1}. Using the induction hypothesis:

$$\mathbb{P}(\xi_n > t - s) = e^{-\lambda(t-s)} \sum_{k=0}^{n-1} \frac{(\lambda(t-s))^k}{k!},$$

so the integral becomes:

$$\int_0^t \mathbb{P}(\xi_n > t - s)\, \lambda e^{-\lambda s} ds = \lambda e^{-\lambda t} \sum_{k=0}^{n-1} \frac{\lambda^k}{k!} \int_0^t (t-s)^k ds.$$

Now, apply the change of variable $u = t - s$, so that when $s = 0$, $u = t$, and when $s = t$, $u = 0$, hence:

$$\int_0^t (t-s)^k ds = \int_0^t u^k du = \frac{t^{k+1}}{k+1}.$$

Therefore, the integral becomes:

$$\lambda e^{-\lambda t} \sum_{k=0}^{n-1} \frac{\lambda^k}{k!} \cdot \frac{t^{k+1}}{k+1} = e^{-\lambda t} \sum_{k=0}^{n-1} \frac{(\lambda t)^{k+1}}{(k+1)!}.$$

We now add $e^{-\lambda t}$ from earlier:

$$\mathbb{P}(\xi_{n+1} > t) = e^{-\lambda t} + e^{-\lambda t} \sum_{k=0}^{n-1} \frac{(\lambda t)^{k+1}}{(k+1)!} = e^{-\lambda t} \sum_{k=0}^{n} \frac{(\lambda t)^k}{k!}.$$

This completes the induction step, so formula (6.2) holds for all $n \in \mathbb{N}^*$.

6.2.3 Poisson Process Starts from Scratch at Time t

Imagine that you are to take part in an experiment to count the emissions of a radioactive particle. Unfortunately, in the excitement of proving the lack of memory property, you forget about the engagement and arrive late to find that the experiment has already been running for time t, and you have missed the first $N(t)$ emissions. Determined to make the best of it, you start counting right away, so at time $t + s$ you will have registered

$N(t+s)-N(t)$ emissions. It will now be necessary to discuss $N(t+s)-N(t)$ instead of $N(s)$ in your report. What are the properties of $N(t+s)-N(t)$? Perhaps you can guess something from the physical picture? After all, a sample of radioactive material will keep emitting particles no matter whether anyone cares to count them or not. So the moment when someone starts counting does not seem important. You can expect $N(t+s)-N(t)$ to behave in a similar way as $N(s)$. And because radioactive emissions have no memory of the past, $N(t+s)-N(t)$ should be independent of $N(t)$. To study this conjecture, recall the construction of a Poisson process $N(t)$ based on a sequence of independent random variables $\eta_1, \eta_2, \ldots$, all having the same exponential distribution. We shall try to represent $N(t+s)-N(t)$ in a similar way. Let us put

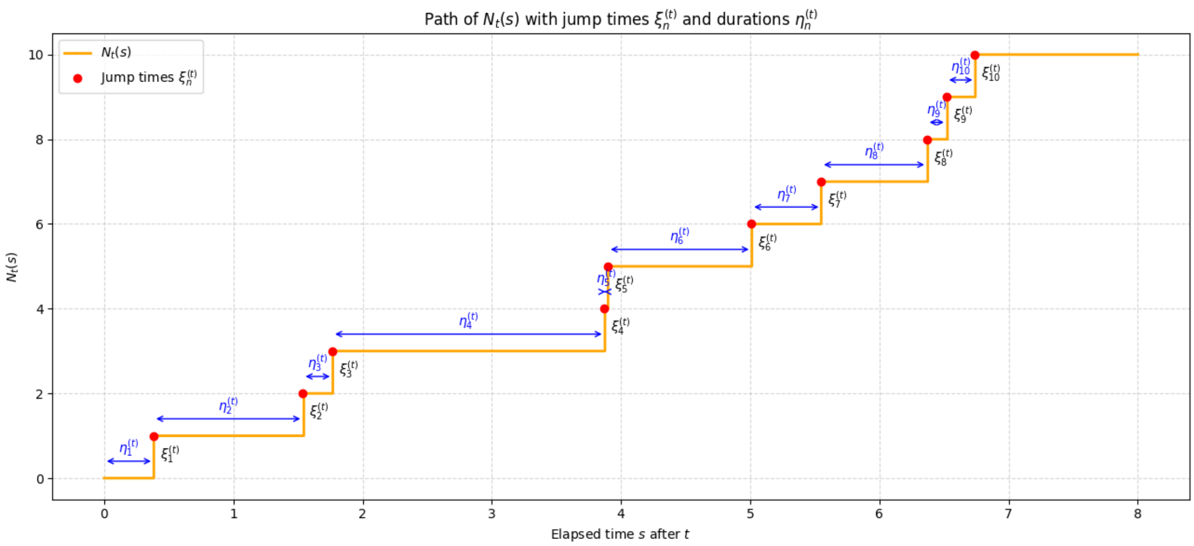
$$\eta_1^{(t)} := \xi_{N(t)+1} - t, \quad \eta_n^{(t)} := \eta_{N(t)+n}, \quad n = 2, 3, \ldots,$$

(see the figure below). These are the times between the jumps of $N(t+s)-N(t)$. Then we define

$$\xi_n^{(t)} := \eta_1^{(t)} + \cdots + \eta_n^{(t)}, \qquad N_t(s) := \max\{n : \xi_n^{(t)} \leq s\}.$$

```
import matplotlib.pyplot as plt
import numpy as np

# Parameters
np.random.seed(1)
lambda_param = 1.0
n_total = 20
eta = np.random.exponential(scale=1/lambda_param, size=n_total)
xi = np.cumsum(eta)
t_obs = 4.5

# N(t)
N_t = np.searchsorted(xi, t_obs)
eta_t = eta[N_t:]
xi_t = np.cumsum([xi[N_t] - t_obs] + list(eta_t[1:]))

s_max = 8
s_values = np.linspace(0, s_max, 1000)
N_t_s = np.searchsorted(xi_t, s_values)

plt.figure(figsize=(13, 6))
plt.step(s_values, N_t_s, where='post', label=r'$N_t(s)$', linewidth=2,
↪  color='orange')
plt.plot(xi_t, np.arange(1, len(xi_t)+1), 'ro', label=r'Jump times
↪  $\xi^{(t)}_n$')

for i, jump in enumerate(xi_t):
    plt.annotate(fr'$\xi^{{(t)}}_{{{i+1}}}$', (jump, i + 0.3),
                 textcoords="offset points", xytext=(5, 5), fontsize=10)

for i, (start, end) in enumerate(zip([0] + list(xi_t[:-1]), xi_t)):
    if end <= s_max:
        y_pos = i + 0.4
```

```
        plt.annotate("", xy=(start, y_pos), xytext=(end, y_pos),
                     arrowprops=dict(arrowstyle="<->", color='blue'))
        plt.text((start + end)/2, y_pos + 0.2,
                 fr'$\eta^{{(t)}}_{{{i+1}}}$', ha='center', color='blue',
                 ↪  fontsize=10)

plt.title(r"Path of $N_t(s)$ with jump times $\xi^{(t)}_n$ and durations
 ↪  $\eta^{(t)}_n$")
plt.xlabel("Elapsed time $s$ after $t$")
plt.ylabel(r"$N_t(s)$")
plt.grid(True, linestyle='--', alpha=0.5)
plt.legend()
plt.tight_layout()
plt.show()
```



Theorem 1: For any fixed $t \geq 0$,

$$N^t(s) = N(t+s) - N(t), \quad s \geq 0,$$

is a Poisson process independent of $N(t)$, with the same probability law as $N(s)$. That is to say, for any $s, t \geq 0$, the increment $N(t+s) - N(t)$ is independent of $N(t)$ and has the same probability distribution as $N(s)$. This assertion can be generalized to several increments, resulting in the following important theorem.

Theorem 2: For any $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$, the increments

$$N(t_1), \quad N(t_2) - N(t_1), \quad N(t_3) - N(t_2), \quad \ldots, \quad N(t_n) - N(t_{n-1})$$

are independent and have the same probability distribution as

$$N(t_1), \quad N(t_2 - t_1), \quad N(t_3 - t_2), \quad \ldots, \quad N(t_n - t_{n-1}).$$

Proof: From Theorem 1, it follows immediately that each increment $N(t_i) - N(t_{i-1})$ has the same distribution as $N(t_i - t_{i-1})$ for $i = 1, \ldots, n$. It remains to prove independence. This can be done by induction. The case when $n = 2$ is covered by Theorem 1. Now suppose that independence holds for n increments of a Poisson process for some $n \geq 2$. Take any sequence

$$0 \leq t_1 \leq t_2 \leq \cdots \leq t_n \leq t_{n+1}.$$

By the induction hypothesis,

$$N(t_{n+1}) - N(t_n), \quad \ldots, \quad N(t_2) - N(t_1)$$

are independent, since they can be regarded as increments of the process

$$N^{t_1}(s) := N(t_1 + s) - N(t_1),$$

which is a Poisson process by Theorem 1. By the same theorem, these increments are also independent of $N(t_1)$. It follows that the $n + 1$ random variables

$$N(t_{n+1}) - N(t_n), \quad \ldots, \quad N(t_2) - N(t_1), \quad N(t_1)$$

are independent, completing the proof.

Definition 7: We say that a stochastic process $\xi(t)$, where $t \in T$, has *independent increments* if

$$\xi(t_1) - \xi(t_0), \quad \xi(t_2) - \xi(t_1), \quad \ldots, \quad \xi(t_n) - \xi(t_{n-1})$$

are independent for any $t_0 < t_1 < \cdots < t_n$ such that $t_0, t_1, \ldots, t_n \in T$.

Definition 8: A stochastic process $\xi(t)$, where $t \in T$, is said to have *stationary increments* if for any $s, t \in T$, the probability distribution of $\xi(t + h) - \xi(s + h)$ is the same for each h such that $s + h$, $t + h \in T$. Theorem 2 implies that the Poisson process has stationary independent increments. The result in the next exercise is also a consequence of Theorem 2.

6.3 Brownian Motion

Imagine a cloud of smoke in completely still air. In time, the cloud will spread over a large volume, the concentration of smoke varying in a smooth manner. However, if a single smoke particle is observed, its path turns out to be extremely rough due to frequent collisions with other particles. This exemplifies two aspects of the same phenomenon called *diffusion*: erratic particle trajectories at the microscopic level, giving rise to a very smooth behaviour of the density of the whole ensemble of particles. The Wiener process $W(t)$ defined below is a mathematical device designed as a model of the motion of individual diffusing particles. In particular, its paths exhibit similar erratic behaviour to the trajectories of real smoke particles. Meanwhile, the density $f_{W(t)}$ of the random variable $W(t)$ is very smooth, given by the exponential function

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

which is a solution of the diffusion equation

$$\frac{\partial f}{\partial t} = \frac{1}{2} \frac{\partial^2 f}{\partial x^2},$$

and can be interpreted as the density at time t of a cloud of smoke issuing from a single point source at time 0. The Wiener process $W(t)$ is also associated with the name of the British botanist Robert Brown, who around 1827 observed the random movement of pollen particles in water. We shall study mainly the one-dimensional Wiener process, which can be thought of as the projection of the position of a smoke particle onto one of the axes of a coordinate system. Apart from describing the motion of diffusing particles, the Wiener process is widely applied in mathematical models involving various noisy systems, for example, the behaviour of asset prices at the stock exchange. If the noise in the system is due to a multitude of independent random changes, then the Central Limit Theorem predicts that the net result will have the normal distribution, a property shared by the increments $W(t) - W(s)$ of the Wiener process. This is one of the main reasons for the widespread use of $W(t)$ in mathematical models.

6.3.1 Definition and Basic Properties

Definition 9: The Wiener process (or Brownian motion) is a stochastic process $W(t)$ with values in $\mathbb{R}$ defined for $t \in [0, \infty)$ such that:

1) $W(0) = 0$ almost surely;
2) the sample paths $t \mapsto W(t)$ are almost surely continuous;
3) for any finite sequence of times $0 < t_1 < \cdots < t_n$ and Borel sets $A_1, \ldots, A_n \subset \mathbb{R}$,

$$\mathbb{P}(W(t_1) \in A_1, \ldots, W(t_n) \in A_n) = \int_{A_1} \cdots \int_{A_n} p(t_1, 0, x_1)\, p(t_2 - t_1, x_1, x_2) \ldots p(t_n - t_{n-1}, x_{n-1}, x_n)\, dx_1 \ldots dx_n.$$

where the function

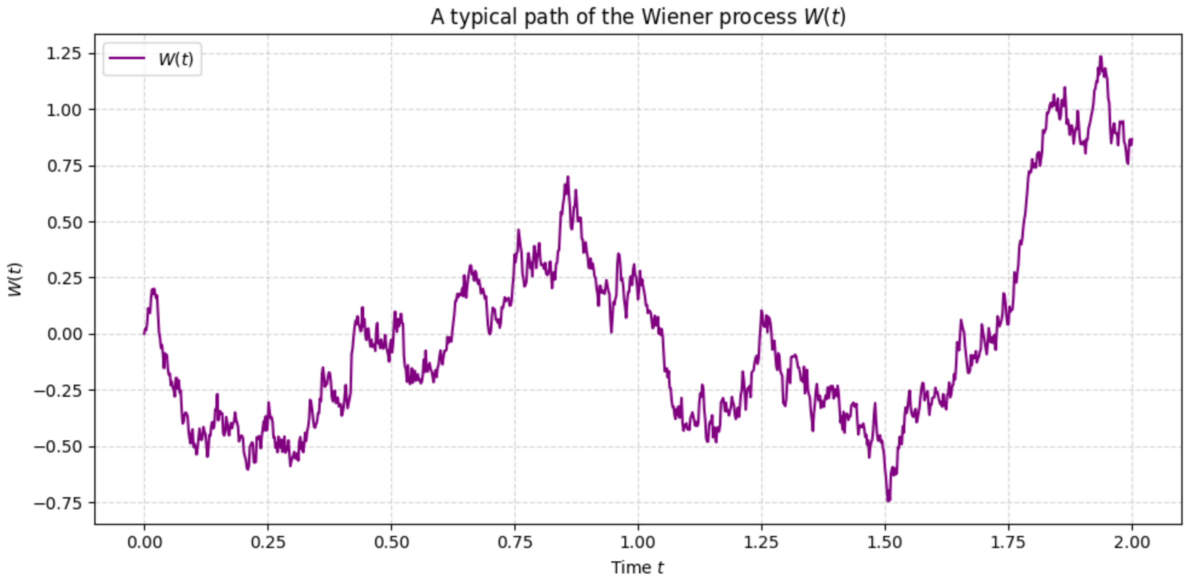
$$p(t, x, y) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(y-x)^2}{2t}\right), \tag{6.3}$$

defined for any $x, y \in \mathbb{R}$ and $t > 0$, is called the *transition density.*

A typical sample path of the Wiener process is shown in the figure below.

```python
import matplotlib.pyplot as plt
import numpy as np

# Parameters
np.random.seed(42)
T = 2.0
N = 1000
dt = T / N
t = np.linspace(0, T, N+1)

dW = np.random.normal(0, np.sqrt(dt), size=N)
W = np.insert(np.cumsum(dW), 0, 0)

plt.figure(figsize=(10, 5))
plt.plot(t, W, label=r"$W(t)$", color="purple")
```

```
plt.title(r"A typical path of the Wiener process $W(t)$")
plt.xlabel(r"Time $t$")
plt.ylabel(r"$W(t)$")
plt.grid(True, linestyle='--', alpha=0.5)
plt.legend()
plt.tight_layout()
plt.show()
```



Definition 10: We call $W(t) = (W_1(t), \ldots, W_n(t))$ an n-dimensional Wiener process if $W_1(t), \ldots, W_n(t)$ are independent $\mathbb{R}$-valued Wiener processes.

6.3.2 Increments of Brownian Motion

Proposition 2: For any $0 \leq s < t$, the increment $W(t) - W(s)$ has the normal distribution with mean 0 and variance $t - s$.

Proof: By condition 3) of Definition 9, the joint density of $W(s)$ and $W(t)$ is

$$f_{W(s),W(t)}(x, y) = p(s, 0, x)\, p(t - s, x, y)$$

Hence, for any Borel set A,

$$\begin{aligned}
\mathbb{P}(W(t)-W(s)\in A) &= \iint_{\{(x,y):y-x\in A\}} p(s,0,x)\,p(t-s,x,y)\,dx\,dy \\
&= \int_{-\infty}^{+\infty} p(s,0,x)\left(\int_{\{y:y-x\in A\}} p(t-s,x,y)\,dy\right)dx \\
&= \int_{-\infty}^{+\infty} p(s,0,x)\left(\int_A p(t-s,x,x+u)\,du\right)dx \\
&= \int_{-\infty}^{+\infty} p(s,0,x)\left(\int_A p(t-s,0,u)\,du\right)dx \\
&= \left(\int_A p(t-s,0,u)\,du\right)\left(\int_{-\infty}^{+\infty} p(s,0,x)\,dx\right) \\
&= \int_A p(t-s,0,u)\,du.
\end{aligned}$$

But $f(u)=p(t-s,0,u)$ is the density of the normal distribution with mean 0 and variance $t-s$, which proves the claim.

Corollary 1: Proposition 2 implies that $W(t)$ has stationary increments.

Proposition 3: For any $0=t_0\le t_1\le\cdots\le t_n$, the increments

$$W(t_1)-W(t_0),\quad W(t_2)-W(t_1),\quad \ldots,\quad W(t_n)-W(t_{n-1})$$

are independent.

Proof: From Proposition 2 we know that the increments of $W(t)$ have the normal distribution. Because normally distributed random variables are independent if and only if they are uncorrelated, it suffices to verify that

$$\mathbb{E}\left[(W(u)-W(t))\,(W(s)-W(r))\right]=0$$

for any $0\le r\le s\le t\le u$. But:

$$\begin{aligned}
\mathbb{E}\left[(W(u)-W(t))\,(W(s)-W(r))\right] &= \mathbb{E}[W(u)W(s)]-\mathbb{E}[W(u)W(r)] \\
&\quad -\mathbb{E}[W(t)W(s)]+\mathbb{E}[W(t)W(r)] \\
&= s-r-s+r=0,
\end{aligned}$$

as required.

Corollary 2: For any $0\le s<t$, the increment $W(t)-W(s)$ is independent of the σ-field $\mathcal{F}_s=\sigma\{W(r):0\le r\le s\}$.

Proof: By Proposition 3, the random variables $W(t)-W(s)$ and $W(r)-W(0)=W(r)$ are independent if $0\le r\le s\le t$. Because the σ-field $\mathcal{F}_s$ is generated by such $W(r)$, it follows that $W(t)-W(s)$ is independent of $\mathcal{F}_s$.

Theorem 3: A stochastic process $W(t)$, $t\ge 0$, is a Wiener process if and only if the following conditions hold:

1. $W(0)=0$ almost surely;
2. the sample paths $t\mapsto W(t)$ are continuous almost surely;

3. $W(t)$ has stationary independent increments;

4. the increment $W(t) - W(s)$ has the normal distribution with mean 0 and variance $t - s$ for any $0 \le s < t$.

Theorem 4: (Lévy's martingale characterization) Let $W(t)$, $t \ge 0$, be a stochastic process and let $\mathcal{F}_t = \sigma(W(s),\ s \le t)$ be the filtration generated by it. Then $W(t)$ is a Wiener process if and only if the following conditions hold:

1. $W(0) = 0$ almost surely;

2. the sample paths $t \mapsto W(t)$ are continuous almost surely;

3. $W(t)$ is a martingale with respect to the filtration $\mathcal{F}_t$;

4. $|W(t)|^2 - t$ is a martingale with respect to $\mathcal{F}_t$.

6.3.3 Sample path

Let

$$0 = t_0^n < t_1^n < \cdots < t_n^n = T,$$

where

$$t_i^n = \frac{iT}{n}$$

be a partition of the interval $[0, T]$ into n equal parts. We denote by

$$\Delta_i^n W = W(t_{i+1}^n) - W(t_i^n)$$

the corresponding increments of the Wiener process $W(t)$.

Definition 11: The variation of a function $f : [0, T] \to \mathbb{R}$ is defined to be

$$\limsup_{\Delta t \to 0} \sum_{i=0}^{n-1} |f(t_{i+1}) - f(t_i)|,$$

where $t = (t_0, t_1, \dots, t_n)$ is a partition of $[0, T]$, i.e., $0 = t_0 < t_1 < \cdots < t_n = T$, and where

$$\Delta t = \max_i |t_{i+1} - t_i|.$$

Theorem 5: The variation of the paths of $W(t)$ is infinite almost surely.

Proof: Consider the sequence of partitions $t^n = (t_0^n, t_1^n, \dots, t_n^n)$ of $[0, T]$ into n equal parts. Then

$$\sum_{i=0}^{n-1} |\Delta_i^n W|^2 \le \left(\max_{i=0,\dots,n-1} |\Delta_i^n W| \right) \sum_{i=0}^{n-1} |\Delta_i^n W|.$$

Since the paths of $W(t)$ are a.s. continuous on $[0, T]$,

$$\lim_{n\to\infty} \max_{i=0,\dots,n-1} |\Delta_i^n W| = 0 \quad \text{a.s.}$$

There is a subsequence $t^{n_k} = (t_0^{n_k}, t_1^{n_k}, \ldots, t_{n_k}^{n_k})$ of partitions such that

$$\lim_{k\to\infty} \sum_{i=0}^{n_k-1} |\Delta_i^{n_k} W|^2 = T \quad \text{a.s.}$$

This is because every sequence of random variables converging in L^2 has a subsequence that converges a.s. It follows that

$$\lim_{k\to\infty} \sum_{i=0}^{n_k-1} |\Delta_i^{n_k} W| = \infty \quad \text{a.s.},$$

while

$$\lim_{k\to\infty} \Delta t^{n_k} = \lim_{k\to\infty} \frac{T}{n_k} = 0,$$

which proves the theorem. Theorem 5 has important consequences for the theory of stochastic integrals presented in the next chapter. This is because an integral of the form

$$\int_0^T f(t)\, dW(t)$$

cannot be defined pathwise (that is, separately for each $\omega \in \Omega$) as a Riemann–Stieltjes integral if the paths have infinite variation. It turns out that an intrinsically stochastic approach will be needed to tackle such integrals; see Chapter 7.

Theorem 6: With probability 1, the Wiener process $W(t)$ is non-differentiable at any $t \geq 0$.

6.3.4 Doob's Maximal L^2 Inequality for Brownian Motion

The inequality proved in this section is necessary to study the properties of stochastic integrals in the next chapter. It can be viewed as an extension of Doob's maximal L^2 inequality in Theorem 1 in Chapter 4 to the case of continuous time. In fact, in the result below, the Wiener process can be replaced by any square-integrable martingale $\xi(t)$, $t \geq 0$, with almost surely continuous paths.

Theorem 7: (Doob's maximal L^2 inequality) For any $t > 0$,

$$\mathbb{E}\left[\max_{s\leq t} |W(s)|^2\right] \leq 4\,\mathbb{E}\left[|W(t)|^2\right]. \tag{6.4}$$

Proof: For $t > 0$ and $n \in \mathbb{N}$, we define

$$M_k^n = |W(t_k)|, \quad \text{where } t_k = \frac{kt}{2^n}, \quad 0 \leq k \leq 2^n. \tag{6.5}$$

Then, by Jensen's inequality, M_k^n, $k = 0, \ldots, 2^n$, is a non-negative square-integrable submartingale with respect to the filtration $\mathcal{F}_k^n = \mathcal{F}_{t_k}$, so by Theorem 1 in Chapter 4,

$$\mathbb{E}\left(\max_{0\leq k\leq 2^n} |M_k^n|^2\right) \leq 4\,\mathbb{E}|M_{2^n}^n|^2 = 4\,\mathbb{E}|W(t)|^2.$$

Since $W(t)$ has almost surely continuous paths,

$$\lim_{n\to\infty} \max_{0\le k\le 2^n} |M_k^n|^2 = \max_{0\le s\le t} |W(s)|^2 \quad \text{a.s.}$$

Moreover, since $M_k^n = M_{2k}^{n+1}$, the sequence $\sup_{0\le k<2^n} |M_k^n|^2$ is increasing. Hence, by the Lebesgue monotone convergence theorem, $\max_{s\le t} |W(s)|^2$ is an integrable function, and

$$\mathbb{E}\left(\max_{0\le s\le t} |W(s)|^2\right) = \lim_{n\to\infty} \mathbb{E}\left(\max_{0\le k\le 2^n} |M_k^n|^2\right) \le 4\,\mathbb{E}|W(t)|^2,$$

completing the proof.

Chapter 7

Itô Stochastic Calculus

One of the first applications of the Wiener process was proposed by Bachelier, who around 1900 wrote a ground-breaking paper on the modelling of asset prices at the Paris Stock Exchange. Of course, Bachelier could not have called it the Wiener process, but he used what in modern terminology amounts to $W(t)$ as a description of the market fluctuations affecting the price $X(t)$ of an asset. Namely, he assumed that infinitesimal price increments $dX(t)$ are proportional to the increments $dW(t)$ of the Wiener process,

$$dX(t) = \sigma\, dW(t),$$

where σ is a positive constant. As a result, an asset with initial price $X(0) = x$ would be worth

$$X(t) = x + a\, W(t)$$

at time t. This approach was ahead of Bachelier's time, but it suffered from one serious flaw: for any $t > 0$ the price $X(t)$ can be negative with non-zero probability. Nevertheless, for short times it works well enough, since the probability is negligible. But as t increases, so does the probability that $X(t) < 0$, and the model departs from reality. To remedy the flaw, it was observed that investors work in terms of their potential gain or loss $dX(t)$ in proportion to the invested sum $X(t)$. Therefore, it is in fact the relative price $dX(t)/X(t)$ of an asset that reacts to the market fluctuations, i.e., should be proportional to $dW(t)$,

$$dX(t) = \sigma\, X(t)\, dW(t). \tag{7.1}$$

What is the precise mathematical meaning of this equality? Formally, it resembles a differential equation, but this immediately leads to a difficulty because the paths of $W(t)$ are nowhere differentiable. A way around the obstacle was found by Itô in the 1940s. In his hugely successful theory of stochastic integrals and stochastic differential equations, Itô gave a rigorous meaning to equations such as (7.1) by writing them as integral equations involving a new kind of integral. In particular, (7.1) can be written as

$$X(t) = x + \sigma \int_0^t X(s)\, dW(s),$$

where the integral with respect to $W(t)$ on the right-hand side is called the *Itô stochastic integral* and will be defined in the next section. While at first sight one would expect the solution to this equation to be $xe^{W(t)}$, in fact it turns out to be

$$X(t) = x\, e^{W(t)}\, e^{-\frac{t}{2}},$$

The intriguing additional factor $e^{-t/2}$ is due to the non-differentiability of the paths of the Wiener process. Clearly, if $x > 0$, then $X(t) > 0$ for all $t \geq 0$, as required in the model of asset prices. In the following sections we shall learn how to transform and compute stochastic integrals and how to solve stochastic differential equations. Throughout this chapter, $W(t)$ will denote a Wiener process adapted to a filtration $\mathcal{F}_t$ and L^2 will be the space of square-integrable random variables.

7.1 Itô Stochastic Integral: Definition

We shall follow a construction resembling that of the Riemann integral. First, the integral will be defined for a class of piecewise constant processes called *random step processes*. Then it will be extended to a larger class by approximation. There are, however, at least two major differences between the Riemann and Itô integrals. One is the type of convergence. The approximations of the Riemann integral converge in $\mathbb{R}$, while the Itô integral will be approximated by sequences of random variables converging in L^2. The other difference is this: the Riemann sums approximating the integral of a function $f : [0, T] \to \mathbb{R}$ are of the form

$$\sum_{j=0}^{n-1} f(s_j)(t_{j+1} - t_j),$$

where $0 = t_0 < t_1 < \cdots < t_n = T$ and s_j is an arbitrary point in $[t_j, t_{j+1}]$ for each j. The value of the Riemann integral does not depend on the choice of the points $s_j \in [t_j, t_{j+1})$. In the stochastic case, the approximating sums will have the form

$$\sum_{j=0}^{n-1} f(s_j) \big(W(t_{j+1}) - W(t_j)\big).$$

It turns out that the limit of such approximations *does* depend on the choice of the intermediate points $s_j \in [t_j, t_{j+1}]$. In the next exercise, we take $f(t) = W(t)$ and consider two different choices of intermediate points.

Definition 1: We shall call $f(t)$, $t \geq 0$, a *random step process* if there exists a finite sequence of times $0 = t_0 < t_1 < \cdots < t_n$ and square-integrable random variables $\eta_0, \eta_1, \ldots, \eta_{n-1}$ such that

$$f(t) = \sum_{j=0}^{n-1} \eta_j \, \mathbf{1}_{[t_j, t_{j+1})}(t), \tag{7.2}$$

where η_j is $\mathcal{F}_{t_j}$-measurable for $j = 0, 1, \ldots, n-1$. The set of random step processes will be denoted by $\mathcal{M}^2_{\text{step}}$. Observe that the assumption that the η_j are $\mathcal{F}_{t_j}$-measurable ensures that $f(t)$ is adapted to the filtration $\mathcal{F}_t$. The assumption that the η_j are square integrable ensures that $f(t)$ is square integrable for each t. Also, $\mathcal{M}^2_{\text{step}}$ is a vector space, that is, $af + bg \in \mathcal{M}^2_{\text{step}}$ for any $f, g \in \mathcal{M}^2_{\text{step}}$ and $a, b \in \mathbb{R}$.

Definition 2: The stochastic integral of a random step process $f \in \mathcal{M}^2_{\text{step}}$ of the form (7.2) is defined by

$$I(f) = \sum_{j=0}^{n-1} \eta_j \left(W(t_{j+1}) - W(t_j)\right). \tag{7.3}$$

Proposition 1: For any random step process $f \in \mathcal{M}^2_{\text{step}}$, the stochastic integral $I(f)$ is a square-integrable random variable, i.e., $I(f) \in L^2$, such that

$$\mathbb{E}\left[|I(f)|^2\right] = \mathbb{E}\left[\int_0^\infty |f(t)|^2\, dt\right].$$

Proof: Let us denote the increment $W(t_{j+1}) - W(t_j)$ by $\Delta_j W$ and $t_{j+1} - t_j$ by $\Delta_j t$ for brevity. Then

$$\mathbb{E}[\Delta_j W] = 0 \quad \text{and} \quad \mathbb{E}[(\Delta_j^2 W)] = \Delta_j t.$$

First, we shall compute the expectation of

$$|I(f)|^2 = \left(\sum_{j=0}^{n-1} \eta_j \Delta_j W\right)^2 = \sum_{j=0}^{n-1} \eta_j^2 (\Delta_j^2 W) + 2 \sum_{0 \le k < j \le n-1} \eta_j \eta_k \Delta_j W \Delta_k W.$$

Since η_j and $\Delta_j W$ are independent, we have

$$\mathbb{E}[\eta_j^2 (\Delta_j^2 W)] = \mathbb{E}[\eta_j^2] \cdot \mathbb{E}[(\Delta_j^2 W)] = \mathbb{E}[\eta_j^2]\, \Delta_j t.$$

If $k < j$, then $\eta_j \eta_k \Delta_k W$ and $\Delta_j W$ are independent, so

$$\mathbb{E}[\eta_j \eta_k \Delta_j W \Delta_k W] = \mathbb{E}[\eta_j \eta_k \Delta_k W] \cdot \mathbb{E}[\Delta_j W] = 0.$$

Therefore,

$$\mathbb{E}(|I(f)|^2) = \sum_{j=0}^{n-1} \mathbb{E}[\eta_j^2]\, \Delta_j t.$$

It follows that $I(f) \in L^2$, since $\eta_0, \eta_1, \dots, \eta_{n-1} \in L^2$. On the other hand,

$$|f(t)|^2 = \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} \eta_j \eta_k\, \mathbf{1}_{[t_j, t_{j+1})}(t)\, \mathbf{1}_{[t_k, t_{k+1})}(t) = \sum_{j=0}^{n-1} \eta_j^2\, \mathbf{1}_{[t_j, t_{j+1})}(t),$$

implying that

$$\mathbb{E}\left[\int_0^\infty |f(t)|^2\, dt\right] = \sum_{j=0}^{n-1} \mathbb{E}[\eta_j^2]\, \Delta_j t.$$

This means that

$$\mathbb{E}[|I(f)|^2] = \mathbb{E}\left[\int_0^\infty |f(t)|^2\, dt\right],$$

as required.

Definition 3: We denote by $\mathcal{M}^2$ the class of stochastic processes $f(t)$, $t \ge 0$, such that

$$\mathbb{E}\left(\int_0^\infty |f(t)|^2\, dt\right) < \infty,$$

and there exists a sequence $f_1, f_2, \dots \in \mathcal{M}^2_{\text{step}}$ of random step processes such that

$$\lim_{n\to\infty} \mathbb{E}\left(\int_0^\infty |f(t) - f_n(t)|^2\, dt\right) = 0. \tag{7.4}$$

In this case, we shall say that the sequence of random step processes $f_1, f_2, \ldots$ approximates f in $\mathcal{M}^2$.

Definition 4: We call $I(f) \in L^2$ the *Itô stochastic integral* (from 0 to ∞) of $f \in \mathcal{M}^2$ if

$$\lim_{n\to\infty} \mathbb{E}\left(|I(f) - I(f_n)|^2\right) = 0 \tag{7.5}$$

for any sequence $f_1, f_2, \dots \in \mathcal{M}^2_{\text{step}}$ of random step processes that approximates f in $\mathcal{M}^2$, i.e., such that condition (7.4) is satisfied. We shall also write

$$\int_0^\infty f(t)\, dW(t)$$

in place of $I(f)$.

Proposition 2: For any $f \in \mathcal{M}^2$, the stochastic integral $I(f) \in L^2$ exists, is unique (as an element of L^2, i.e., up to almost sure equality), and satisfies

$$\mathbb{E}\left[|I(f)|^2\right] = \mathbb{E}\left[\int_0^\infty |f(t)|^2\, dt\right]. \tag{7.6}$$

Proof: It will be convenient to write

$$\|f\|_{\mathcal{M}^2} := \sqrt{\mathbb{E}\left[\int_0^\infty |f(t)|^2\, dt\right]} \quad \text{and} \quad \|\eta\|_{L^2} := \sqrt{\mathbb{E}[\eta^2]}$$

for any $f \in \mathcal{M}^2$ and $\eta \in L^2$. These are norms on $\mathcal{M}^2$ and L^2, respectively. Let $f_1, f_2, \dots \in \mathcal{M}^2_{\text{step}}$ be a sequence of random step processes approximating $f \in \mathcal{M}^2$, i.e., satisfying (7.4), which can be written as

$$\lim_{n\to\infty} \|f - f_n\|_{\mathcal{M}^2} = 0.$$

We claim that $I(f_1), I(f_2), \ldots$ is a Cauchy sequence in L^2. Indeed, for any $\varepsilon > 0$, there exists N such that $\|f - f_n\|_{\mathcal{M}^2} < \varepsilon/2$ for all $n > N$. By Proposition 7.1,

$$\|I(f_m) - I(f_n)\|_{L^2} = \|I(f_m - f_n)\|_{L^2} = \|f_m - f_n\|_{\mathcal{M}^2} \leq \|f - f_m\|_{\mathcal{M}^2} + \|f - f_n\|_{\mathcal{M}^2} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

for any $m, n > N$, which proves the claim. Because L^2 with the norm $\|\cdot\|_{L^2}$ is a complete space (in fact, a Hilbert space), every Cauchy sequence in L^2 has a limit. It follows that $I(f_1), I(f_2), \ldots$ has a limit in L^2 for any sequence f_n of random step processes approximating f. It remains to show that the limit is the same for all such sequences. Suppose that $f_1, f_2, \ldots$ and $g_1, g_2, \ldots$ are two sequences of random step processes approximating f. Then the interlaced sequence $f_1, g_1, f_2, g_2, \ldots$ also approximates f, so the sequence $I(f_1), I(g_1), I(f_2), I(g_2), \ldots$ has a limit in L^2. But then all of its subsequences, in particular $I(f_n)$ and $I(g_n)$, must have the same limit, which we denote by $I(f)$. We have thus shown that

$$\lim_{n\to\infty} \|I(f) - I(f_n)\|_{L^2} = 0,$$

i.e., relation (7.5) holds for any sequence f_n of random step processes approximating f. Finally, by Proposition 7.1,

$$\|I(f_n)\|_{L^2} = \|f_n\|_{\mathcal{M}^2}$$

for each n, since the f_n are random step processes. Taking the limit as $n \to \infty$, we obtain

$$\|I(f)\|_{L^2} = \|f\|_{\mathcal{M}^2},$$

which is exactly equality (7.6).

Definition 5: For any $T > 0$, we shall denote by $\mathcal{M}^2_T$ the space of all stochastic processes $f(t)$, $t \geq 0$, such that

$$\mathbf{1}_{[0,T)} f \in \mathcal{M}^2.$$

The Itô stochastic integral (from 0 to T) of $f \in \mathcal{M}^2_T$ is defined by

$$I_T(f) := I\left(\mathbf{1}_{[0,T)} f\right). \tag{7.7}$$

We shall also write

$$\int_0^T f(t)\, dW(t)$$

in place of $I_T(f)$.

Theorem 1: Let $f(t)$, $t \geq 0$, be a stochastic process with almost surely continuous paths adapted to the filtration $\mathcal{F}_t$. Then:

1. $f \in \mathcal{M}^2$, i.e., the Itô integral $I(f)$ exists, whenever

$$\mathbb{E}\left(\int_0^\infty |f(t)|^2\, dt\right) < \infty. \tag{7.8}$$

2. $f \in \mathcal{M}^2_T$, i.e., the Itô integral $I_T(f)$ exists, whenever

$$\mathbb{E}\left(\int_0^T |f(t)|^2\, dt\right) < \infty. \tag{7.9}$$

Proof:

1. Suppose that $f(t)$, $t \geq 0$, is an adapted process with a.s. continuous paths. If condition (7.8) holds, then define

$$f_n(t) = \begin{cases} n \int_{t_k}^{t_{k+1}} f(s)\, ds & \text{for } t \in [t_k, t_{k+1}), \quad k = 0, 1, \ldots, n^2 - 1, \\ 0 & \text{otherwise,} \end{cases} \tag{7.10}$$

where $t_k = \frac{k}{n}$ for $k = 0, \ldots, n^2$. Then (f_n) is a sequence of random step processes in $\mathcal{M}^2_{\text{step}}$. Observe that for any $k = 0, 1, \ldots, n^2 - 1$, by Jensen's inequality,

$$\int_{t_k}^{t_{k+1}} |f_n(t)|^2\, dt = \int_{t_k}^{t_{k+1}} \left| n \int_{t_k}^{t_{k+1}} f(s)\, ds \right|^2 dt \leq n \int_{t_k}^{t_{k+1}} |f(s)|^2\, ds.$$

We claim that

$$\lim_{n \to \infty} \int_0^\infty |f(t) - f_n(t)|^2\, dt = 0 \quad \text{a.s.} \tag{7.11}$$

This will imply that

$$\lim_{n\to\infty} \mathbb{E}\left(\int_0^\infty |f(t) - f_n(t)|^2\, dt\right) = 0,$$

by the dominated convergence theorem and condition (7.8), because

$$\int_0^\infty |f(t) - f_n(t)|^2\, dt \le 2\int_0^\infty \left(|f(t)|^2 + |f_n(t)|^2\right) dt \le 4\int_0^\infty |f(t)|^2\, dt.$$

The last inequality follows since

$$\int_0^\infty |f_n(t)|^2\, dt \le \int_0^\infty |f(t)|^2\, dt \quad \text{a.s.}$$

for any n, by taking the sum from $k = 0$ to ∞ in the estimate from (7.11). To verify the claim, observe that

$$\begin{aligned}
\int_0^\infty |f(t) - f_n(t)|^2\, dt &= \int_0^N |f(t) - f_n(t)|^2\, dt + \int_N^\infty |f(t) - f_n(t)|^2\, dt \\
&\le \int_0^N |f(t) - f_n(t)|^2\, dt + 2\int_N^\infty \left(|f(t)|^2 + |f_n(t)|^2\right)\, dt \\
&\le \int_0^N |f(t) - f_n(t)|^2\, dt + 4\int_{N-1}^\infty |f(t)|^2\, dt \quad \text{a.s.}
\end{aligned}$$

The last inequality holds because

$$\int_N^\infty |f_n(t)|^2\, dt \le \sum_{k=nN}^\infty \int_{t_k}^{t_{k+1}} |f(t)|^2\, dt \le \int_{N-1}^\infty |f(t)|^2\, dt \quad \text{a.s.}$$

The claim follows because

$$\lim_{N\to\infty} \int_{N-1}^\infty |f(t)|^2\, dt = 0 \quad \text{a.s.}$$

by condition (7.8), and

$$\lim_{n\to\infty} \int_0^N |f(t) - f_n(t)|^2\, dt = 0 \quad \text{a.s.}$$

for any fixed N, by the continuity of the paths of f. The above shows that the sequence $f_n \in \mathcal{M}^2_{\text{step}}$ approximates f in the sense of Definition 3, so $f \in \mathcal{M}^2$.

2) If f satisfies (7.9) for some $T > 0$, then $\mathbf{1}_{[0,T)}f$ satisfies (7.8). Since f is adapted and has almost surely continuous paths, $\mathbf{1}_{[0,T)}f$ is also adapted, and its paths are almost surely continuous, except perhaps at T. But the lack of continuity at a single point T does not affect the argument in part 1), so $\mathbf{1}_{[0,T)}f \in \mathcal{M}^2$. This in turn implies that $f \in \mathcal{M}^2_T$, completing the proof.

Definition 6: A stochastic process $f(t)$, $t \ge 0$, is called *progressively measurable* if for any $t \ge 0$, the mapping $(s, \omega) \mapsto f(s, \omega)$ is measurable from $[0, t] \times \Omega$ equipped with the σ-field $\mathcal{B}([0, t]) \otimes \mathcal{F}_t$ into $\mathbb{R}$. Here, $\mathcal{B}([0, t]) \otimes \mathcal{F}_t$ is the product σ-field on $[0, t] \times \Omega$, that is, the smallest σ-field containing all sets of the form $A \times B$, where $A \subset [0, t]$ is a Borel set and $B \in \mathcal{F}_t$.

Theorem 2:

1. The space $\mathcal{M}^2$ consists of all progressively measurable stochastic processes $f(t)$, $t \geq 0$, such that
$$\mathbb{E}\left(\int_0^\infty |f(t)|^2 \, dt\right) < \infty.$$

2. The space $\mathcal{M}_T^2$ consists of all progressively measurable stochastic processes $f(t)$, $t \geq 0$, such that
$$\mathbb{E}\left(\int_0^T |f(t)|^2 \, dt\right) < \infty.$$

7.2 Properties of the Stochastic Integral

The basic properties of the Itô integral are summarized in the theorem below.

Theorem 3: The following properties hold for any $f, g \in \mathcal{M}_T^2$, any $a, \beta \in \mathbb{R}$, and any $0 \leq s < t$:

1. **Linearity:**
$$\int_s^t (\alpha f(r) + \beta g(r)) \, dW(r) = \alpha \int_s^t f(r) \, dW(r) + \beta \int_s^t g(r) \, dW(r).$$

2. **Isometry:**
$$\mathbb{E}\left(\left|\int_s^t f(r) \, dW(r)\right|^2\right) = \mathbb{E}\left(\int_s^t |f(r)|^2 \, dr\right).$$

3. **Martingale property:**
$$\mathbb{E}\left(\int_s^t f(r) \, dW(r) \,\middle|\, \mathcal{F}_s\right) = \int_0^s f(r) \, dW(r).$$

Proof:

1. If f and g belong to $\mathcal{M}_T^2$, then $\mathbf{1}_{[0,t)} f$ and $\mathbf{1}_{[0,t)} g$ belong to $\mathcal{M}^2$, so there exist sequences $f_1, f_2, \ldots$ and $g_1, g_2, \ldots$ in $\mathcal{M}_{\text{step}}^2$ approximating $\mathbf{1}_{[0,t)} f$ and $\mathbf{1}_{[0,t)} g$. It follows that $\mathbf{1}_{[0,t)}(\alpha f + \beta g)$ can be approximated by the sequence $\alpha f_n + \beta g_n$. So,
$$I(\alpha f_n + \beta g_n) = \alpha I(f_n) + \beta I(g_n)$$
for each n. Taking the limit in L^2 on both sides as $n \to \infty$, we obtain
$$I(\mathbf{1}_{[0,t)}(\alpha f + \beta g)) = \alpha I(\mathbf{1}_{[0,t)} f) + \beta I(\mathbf{1}_{[0,t)} g),$$
which proves linearity.

2. This follows by approximating $\mathbf{1}_{[0,t)} f$ by random step processes in $\mathcal{M}_{\text{step}}^2$ and using Proposition 1 (the Itô isometry):
$$\mathbb{E}\left(\left|\int_0^t f(r) \, dW(r)\right|^2\right) = \mathbb{E}\left(\int_0^t |f(r)|^2 \, dr\right).$$

3. If f belongs to $\mathcal{M}_T^2$, then $\mathbf{1}_{[0,t)} f \in \mathcal{M}^2$. Let $f_1, f_2, \dots$ be a sequence of processes in $\mathcal{M}^2_{\text{step}}$ approximating $\mathbf{1}_{[0,t)} f$. Hence,

$$\mathbb{E}\left(I(\mathbf{1}_{[0,t)} f_n) \,\middle|\, \mathcal{F}_s\right) = I(\mathbf{1}_{[0,s)} f_n). \tag{7.12}$$

for each n. Taking the limit in L^2 on both sides as $n \to \infty$, we obtain

$$\mathbb{E}\left(I(\mathbf{1}_{[0,t)} f) \,\middle|\, \mathcal{F}_s\right) = I(\mathbf{1}_{[0,s)} f),$$

which proves the martingale property.

Indeed, observe that $\mathbf{1}_{[0,s)} f_1, \mathbf{1}_{[0,s)} f_2, \dots$ is a sequence in $\mathcal{M}^2_{\text{step}}$ approximating $\mathbf{1}_{[0,s)} f$, so

$$I(\mathbf{1}_{[0,s)} f_n) \longrightarrow I(\mathbf{1}_{[0,s)} f) \quad \text{in } L^2 \text{ as } n \to \infty.$$

Similarly, $\mathbf{1}_{[0,t)} f_1, \mathbf{1}_{[0,t)} f_2, \dots$ is also a sequence in $\mathcal{M}^2_{\text{step}}$ approximating $\mathbf{1}_{[0,t)} f$, which implies that the same holds for the left-hand side of equation (7.12). The lemma below implies that

$$\mathbb{E}\left(I(\mathbf{1}_{[0,t)} f_n) \,\middle|\, \mathcal{F}_s\right) \to \mathbb{E}\left(I(\mathbf{1}_{[0,t)} f) \,\middle|\, \mathcal{F}_s\right) \text{ in } L^2 \text{ as } n \to \infty.$$

completing the proof.

Lemma 1: If ξ and $\xi_1, \xi_2, \dots$ are square integrable random variables such that $\xi_n \to \xi$ in L^2 as $n \to \infty$, then

$$\mathbb{E}[\xi_n \mid \mathcal{G}] \longrightarrow \mathbb{E}[\xi \mid \mathcal{G}] \quad \text{in } L^2 \text{ as } n \to \infty$$

for any σ-field $\mathcal{G}$ on Ω such that $\mathcal{G} \subseteq \mathcal{F}$.

Proof: By Jensen's inequality (see Theorem 2),

$$|\mathbb{E}[\xi_n \mid \mathcal{G}] - \mathbb{E}[\xi \mid \mathcal{G}]|^2 = |\mathbb{E}[\xi_n - \xi \mid \mathcal{G}]|^2 \le \mathbb{E}\left[|\xi_n - \xi|^2 \mid \mathcal{G}\right].$$

This implies that

$$\mathbb{E}\left(|\mathbb{E}[\xi_n \mid \mathcal{G}] - \mathbb{E}[\xi \mid \mathcal{G}]|^2\right) \le \mathbb{E}\left(\mathbb{E}\left[|\xi_n - \xi|^2 \mid \mathcal{G}\right]\right) = \mathbb{E}\left[|\xi_n - \xi|^2\right] \longrightarrow 0 \quad \text{as } n \to \infty.$$

In the next theorem, we consider the stochastic integral $\int_0^t f(s)\,dW(s)$ as a function of the upper integration limit t. Similarly to the Riemann integral, it is natural to ask whether this defines a continuous function of t. The answer to this question involves the notion of a *modification* of a stochastic process.

Definition 7: Let $\xi(t)$ and $\zeta(t)$ be stochastic processes defined for $t \in T$, where $T \subset \mathbb{R}$. We say that the processes are *modifications* (or *versions*) of one another if

$$\mathbb{P}\left\{\xi(t) = \zeta(t)\right\} = 1 \quad \text{for all } t \in T. \tag{7.13}$$

Remark 1: If $T \subset \mathbb{R}$ is a countable set, then condition (7.13) is equivalent to the condition

$$\mathbb{P}\left\{\xi(t) = \zeta(t) \text{ for all } t \in T\right\} = 1.$$

However, this is not necessarily so if T is uncountable. The following result is stated without proof.

Theorem 4: Let $f(s)$ be a process belonging to $\mathcal{M}_T^2$ and let

$$\xi(t) := \int_0^t f(s)\, dW(s)$$

for every $t \geq 0$. Then there exists an adapted modification $\zeta(t)$ of $\xi(t)$ with almost surely continuous paths. This modification is unique up to almost sure equality. From now on, we shall always identify $\int_0^t f(s)\, dW(s)$ with the adapted modification having almost surely continuous paths. This convention works beautifully together with Theorem 1 whenever there is a need to show that a stochastic integral can be used as the integrand of another stochastic integral, i.e., belongs to $\mathcal{M}_T^2$ for some $T \geq 0$.

7.3 Stochastic Differential and Itô Formula

Any continuously differentiable function $x(t)$ such that $x(0) = 0$ satisfies the formula:

$$x(T)^2 = 2 \int_0^T x(t)\, dx(t).$$

$$x(T)^3 = 3 \int_0^T x(t)^2\, dx(t),$$

where $dx(t)$ is simply understood as shorthand for $x'(t)\, dt$, and the integrals on the right-hand side are Riemann integrals. Similar formulae have been obtained for the Wiener process:

$$W(T)^2 = \int_0^T dt + 2 \int_0^T W(t)\, dW(t),$$

$$W(T)^3 = 3 \int_0^T W(t)\, dt + 3 \int_0^T W(t)^2\, dW(t).$$

Here, the stochastic integrals resemble the corresponding expressions for a smooth function $x(t)$, but there are also the intriguing terms $\int_0^T dt$ and $3 \int_0^T W(t)\, dt$. The formulae for $W(T)^2$ and $W(T)^3$ are examples of the much more general *Itô formula*, a crucial tool for transforming and computing stochastic integrals. Terms such as $\int_0^T dt$ and $3 \int_0^T W(t)\, dt$, which have no analogues in the classical calculus of smooth functions, are a feature inherent in the Itô formula and are referred to as the *Itô correction*. The class of processes appearing in the Itô formula is defined as follows.

Definition 8: A stochastic process $\xi(t)$, $t \geq 0$, is called an *Itô process* if it has almost surely continuous paths and can be represented as

$$\xi(T) = \xi(0) + \int_0^T a(t)\, dt + \int_0^T b(t)\, dW(t) \quad \text{a.s.} \tag{7.14}$$

where $b(t)$ is a process belonging to $\mathcal{M}_T^2$ for all $T > 0$, and $a(t)$ is a process adapted to the filtration $\mathcal{F}_t$ such that

$$\int_0^T |a(t)|\, dt < \infty \quad \text{a.s.} \tag{7.15}$$

for all $T \geq 0$. The class of all adapted processes $a(t)$ satisfying (7.15) for some $T > 0$ will be denoted by $\mathcal{L}^1_T$. For an Itô process ξ, it is customary to write (7.14) in the differential form:

$$d\xi(t) = a(t)\,dt + b(t)\,dW(t) \tag{7.16}$$

and to call $d\xi(t)$ the *stochastic differential* of $\xi(t)$. This is known as the *Itô differential notation.* It should be emphasized that the stochastic differential has no well-defined mathematical meaning on its own, and should always be understood in the context of the rigorous equation (7.14). The Itô differential notation is an efficient way of writing this equation, rather than an attempt to give a precise mathematical meaning to the differential itself.

Theorem 5:(Itô formula, simplified version) Suppose that $F(t, x)$ is a real-valued function with continuous partial derivatives $F'_t(t, x)$, $F'_x(t, x)$, and $F''_{xx}(t, x)$ for all $t \geq 0$ and $x \in \mathbb{R}$. We also assume that the process $F'_x(t, W(t))$ belongs to $\mathcal{M}^2_T$ for all $T \geq 0$. Then $F(t, W(t))$ is an Itô process such that

$$F(T, W(T)) - F(0, W(0)) = \int_0^T \left(F'_t(t, W(t)) + \tfrac{1}{2} F''_{xx}(t, W(t))\right) dt + \int_0^T F'_x(t, W(t))\,dW(t) \tag{7.17}$$

In differential notation, this formula can be written as

$$dF(t, W(t)) = \left(F'_t(t, W(t)) + \tfrac{1}{2} F''_{xx}(t, W(t))\right) dt + F'_x(t, W(t))\,dW(t). \tag{7.18}$$

Proof: First, we shall prove the Itô formula under the assumption that F and the partial derivatives F_t and F_{xx} are bounded by some $C > 0$. Consider a partition $0 = t_0 < t_1 < \cdots < t_n = T$, where $t_i = i\Delta t$, of $[0, T]$ into n equal parts. We denote the increments $W(t_{i+1}) - W(t_i)$ by $\Delta_i W$ and $t_{i+1} - t_i$ by Δt. We also write W_i instead of $W(t_i)$ for brevity. According to the Taylor formula, there is a point $\theta_i \in [W(t_i), W(t_{i+1})]$ and a point $\tau_i \in [t_i, t_{i+1}]$ such that

$$F(T, W(T)) - F(0, W(0)) = \sum_{i=0}^{n-1} \left(F(t_{i+1}, W_{i+1}) - F(t_i, W_i)\right)$$

$$= \sum_{i=0}^{n-1} \left(F(t_{i+1}, W_{i+1}) - F(t_i, W_{i+1}) + F(t_i, W_{i+1}) - F(t_i, W_i)\right)$$

$$= \sum_{i=0}^{n-1} F_t(\tau_i, W_{i+1})\Delta t + \sum_{i=0}^{n-1} F_x(t_i, W_i)\Delta_i W + \frac{1}{2}\sum_{i=0}^{n-1} F_{xx}(t_i, \theta_i)(\Delta_i W)^2.$$

Regrouping and rewriting:

$$F(T, W(T)) - F(0, W(0)) = \sum_{i=0}^{n-1} F_t(\tau_i, W_{i+1})\Delta t + \frac{1}{2}\sum_{i=0}^{n-1} F_{xx}(t_i, W_i)\Delta t + \sum_{i=0}^{n-1} F_x(t_i, W_i)\Delta_i W$$

$$+ \frac{1}{2}\sum_{i=0}^{n-1} F_{xx}(t_i, W_i)\left((\Delta_i W)^2 - \Delta t\right) + \frac{1}{2}\sum_{i=0}^{n-1} \left(F_{xx}(t_i, \theta_i) - F_{xx}(t_i, W_i)\right)(\Delta_i W)^2.$$

We now analyze each term in this sum:

- **Step 1.** By uniform continuity of F_t on compact sets and continuity of $W(t)$, we have:
$$\sum_{i=0}^{n-1} F_t(\tau_i, W_{i+1})\Delta t \longrightarrow \int_0^T F_t(t, W(t))\,dt \quad \text{a.s.}$$

- **Step 2.** Similarly, we get:
$$\sum_{i=0}^{n-1} F_{xx}(t_i, W_i)\Delta t \longrightarrow \int_0^T F_{xx}(t, W(t))\,dt \quad \text{a.s.}$$

- **Step 3.** Define $f_n(t) := \sum_{i=0}^{n-1} F_x(t_i, W_i)\mathbf{1}_{[t_i,t_{i+1})}(t)$. Then $f_n \in \mathcal{M}^2_{\text{step}}$ and $f_n(t) \to F_x(t, W(t))$ in L^2 by dominated convergence. Therefore,
$$\sum_{i=0}^{n-1} F_x(t_i, W_i)\Delta_i W \longrightarrow \int_0^T F_x(t, W(t))\,dW(t) \quad \text{in } L^2.$$

- **Step 4.** Since F_{xx} is bounded by $C > 0$ and $\mathbb{E}[(\Delta_i W)^2 - \Delta t] = 0$, we get:
$$\mathbb{E}\left[\left(\sum_{i=0}^{n-1} F_{xx}(t_i, W_i)\left((\Delta_i W)^2 - \Delta t\right)\right)^2\right] \le C^2 \sum_{i=0}^{n-1} \mathbb{E}\left[(\Delta_i W)^2 - \Delta t\right]^2 = \mathcal{O}(\tfrac{1}{n}) \to 0.$$

- **Step 5.** Since F_{xx} is continuous and $W(t)$ is almost surely continuous and bounded on $[0, T]$, the uniform difference $|F_{xx}(t_i, \theta_i) - F_{xx}(t_i, W_i)| \to 0$ a.s. Moreover, $\sum(\Delta_i W)^2 \to T$ in L^2, hence the full term tends to 0 in L^2.

Thus, we obtain:
$$F(T, W(T)) = F(0, W(0)) + \int_0^T \left(F_t(t, W(t)) + \tfrac{1}{2}F_{xx}(t, W(t))\right) dt + \int_0^T F_x(t, W(t))\,dW(t).$$

To remove the assumption that F_t and F_{xx} are bounded, let $\varphi_n \in C^\infty(\mathbb{R})$ be such that

$\varphi_n(x) = 1$ for $x \in [-n, n]$, and $\varphi_n(x) = 0$ for $x \notin [-n-1, n+1]$. Define:
$$F_n(t, x) := \varphi_n(x)F(t, x).$$

Each F_n has bounded derivatives and satisfies the assumptions of the theorem, so the Itô formula holds for F_n. Let $A_n := \left\{\sup_{t\in[0,T]} |W(t)| < n\right\}$. Then $F(t, x) = F_n(t, x)$ on A_n, so the formula holds on A_n. Finally, by Doob's maximal L^2 inequality (Theorem 6.7),
$$n^2(1 - \mathbb{P}(A_n)) \le \mathbb{E}\left[\sup_{t\in[0,T]} |W(t)|^2\right] \le 4\mathbb{E}[W(T)^2] = 4T.$$

Hence $\mathbb{P}(A_n) \to 1$ as $n \to \infty$, and the result holds almost surely by passing to the limit.

Theorem 6:(Itô formula, general case) Let $\xi(t)$ be an Itô process as above. Suppose that $F(t, x)$ is a real-valued function with continuous partial derivatives $F'_t(t, x)$, $F'_x(t, x)$

and $F''_{xx}(t,x)$ for all $t \geq 0$ and $x \in \mathbb{R}$. We also assume that the process $b(t)F'_x(t,\xi(t))$ belongs to $\mathcal{M}^2_T$ for all $T \geq 0$. Then $F(t,\xi(t))$ is an Itô process such that

$$dF(t,\xi(t)) = \left(F'_t(t,\xi(t)) + F'_x(t,\xi(t))a(t) + \tfrac{1}{2}F''_{xx}(t,\xi(t))b(t)^2\right)dt + F'_x(t,\xi(t))b(t)\,dW(t). \tag{7.20}$$

A convenient way to remember the Itô formula is to write down the Taylor expansion for $F(t,x)$ up to second-order partial derivatives, substitute $\xi(t)$ for x and replace $d\xi(t)$ using the expression from the Itô process definition:

$$d\xi(t) = a(t)\,dt + b(t)\,dW(t),$$

and then apply the so-called *Itô multiplication table*:

$$dt \cdot dt = 0, \quad dW(t) \cdot dt = 0, \quad dt \cdot dW(t) = 0, \quad dW(t) \cdot dW(t) = dt.$$

This informal procedure gives:

$$\begin{aligned}
dF &= F'_t\,dt + F'_x\,d\xi + \tfrac{1}{2}F''_{tt}\,dt\,dt + F'_{tx}\,dt\,d\xi + \tfrac{1}{2}F''_{xx}\,d\xi\,d\xi \\
&= F'_t\,dt + F'_x(a\,dt + b\,dW) \\
&\quad + \tfrac{1}{2}F''_{tt}\,dt\,dt + F''_{tx}\,dt(a\,dt + b\,dW) + \tfrac{1}{2}F''_{xx}(a\,dt + b\,dW)^2 \\
&= F'_t\,dt + F'_x(a\,dt + b\,dW) + \tfrac{1}{2}F''_{xx}b^2\,dt \\
&= \left(F'_t + F'_x a + \tfrac{1}{2}F''_{xx}b^2\right)dt + F'_x b\,dW.
\end{aligned}$$

which is precisely equation (7.20). Here, for brevity, we have omitted the explicit arguments $(t,\xi(t))$ and (t) in all functions.

7.4 Stochastic Differential Equations

This section will be devoted to stochastic differential equations of the form

$$d\xi(t) = f(\xi(t))\,dt + g(\xi(t))\,dW(t).$$

Solutions will be sought in the class of Itô processes $\xi(t)$ with almost surely continuous paths. As in the theory of ordinary differential equations, we need to specify an initial condition:

$$\xi(0) = \xi_0.$$

Here, ξ_0 can be a fixed real number or, more generally, a random variable. Being an Itô process, $\xi(t)$ must be adapted to the filtration $\mathcal{F}_t$ of $W(t)$, so ξ_0 must be $\mathcal{F}_0$-measurable.

Definition 9: An Itô process $\xi(t)$, $t \geq 0$, is called a solution of the initial value problem

$$d\xi(t) = f(\xi(t))\,dt + g(\xi(t))\,dW(t), \quad \xi(0) = \xi_0$$

if ξ_0 is an $\mathcal{F}_0$-measurable random variable, the processes $f(\xi(t))$ and $g(\xi(t))$ belong, respectively, to $\mathcal{L}^1_T$ and $\mathcal{M}^2_T$, and

$$\xi(T) = \xi_0 + \int_0^T f(\xi(t))\,dt + \int_0^T g(\xi(t))\,dW(t) \quad \text{a.s.} \tag{7.23}$$

for all $T \geq 0$.

Remark 2: In view of this definition, the notion of a stochastic differential equation is a fiction. In fact, only stochastic integral equations of the form (7.23) have a rigorous mathematical meaning. However, it proves convenient to use stochastic differentials informally and talk of stochastic differential equations, in order to draw on the analogy with ordinary differential equations. This analogy will be employed to solve some stochastic differential equations later on in this section. The existence and uniqueness theorem below resembles that in the theory of ordinary differential equations, where it is also crucial for the right-hand side of the equation to be Lipschitz continuous as a function of the solution.

Theorem 7: Suppose that f and g are Lipschitz continuous functions from $\mathbb{R}$ to $\mathbb{R}$, i.e., there exists a constant $C > 0$ such that for any $x, y \in \mathbb{R}$,

$$|f(x) - f(y)| \leq C|x - y|, \quad |g(x) - g(y)| \leq C|x - y|.$$

Moreover, let ξ_0 be an $\mathcal{F}_0$-measurable square integrable random variable. Then the initial value problem

$$d\xi(t) = f(\xi(t))\, dt + g(\xi(t))\, dW(t), \tag{7.24}$$
$$\xi(0) = \xi_0 \tag{7.25}$$

has a solution $\xi(t)$, $t \geq 0$, in the class of Itô processes. The solution is unique in the sense that if $\eta(t)$, $t \geq 0$, is another Itô process satisfying (7.24) and (7.25), then the two processes are identical almost surely, that is,

$$\mathbb{P}\{\xi(t) = \eta(t) \text{ for all } t \geq 0\} = 1.$$

Proof (Outline): Let us fix $T > 0$. We are looking for a process $\xi \in \mathcal{M}_T^2$ such that

$$\xi(s) = \xi_0 + \int_0^s f(\xi(t))\, dt + \int_0^s g(\xi(t))\, dW(t) \quad \text{a.s.} \tag{7.26}$$

for all $s \in [0, T]$. Once we have shown that such a $\xi \in \mathcal{M}_T^2$ exists, to obtain a solution to the stochastic differential equation (7.24) with initial condition (7.25), it suffices to take a modification of ξ with a.s. continuous paths, which exists by Theorem 7.A. To show that a solution to the stochastic integral equation (7.26) exists, we shall employ the Banach fixed point theorem in $\mathcal{M}_T^2$ with the norm

$$\|\xi\|_\lambda = \left(\mathbb{E}\left[\int_0^T e^{-\lambda t}|\xi(t)|^2\, dt\right]\right)^{1/2}, \tag{7.27}$$

which turns $\mathcal{M}_T^2$ into a complete normed vector space. The number $\lambda > 0$ should be chosen large enough, as explained below. To apply the fixed point theorem, define $P : \mathcal{M}_T^2 \to \mathcal{M}_T^2$ by

$$P(\xi)(s) = \xi_0 + \int_0^s f(\xi(t))\, dt + \int_0^s g(\xi(t))\, dW(t), \quad \text{for any } \xi \in \mathcal{M}_T^2 \text{ and } s \in [0, T]. \tag{7.28}$$

We claim that P is a strict contraction, i.e.,

$$\|P(\xi) - P(\eta)\|_\lambda \leq C\|\xi - \eta\|_\lambda, \quad \text{for some } C < 1 \text{ and all } \xi, \eta \in \mathcal{M}_T^2. \tag{7.29}$$

Then, by the Banach fixed point theorem, P has a unique fixed point $\xi = P(\xi)$, which is the desired solution to (7.26). It remains to verify that P is indeed a strict contraction. We split P as $P = P_1 + P_2$, where:

$$P_1(\xi)(s) = \int_0^s f(\xi(t))\,dt, \quad P_2(\xi)(s) = \int_0^s g(\xi(t))\,dW(t).$$

We estimate each term separately. For P_1, the Lipschitz continuity of f and Jensen's inequality imply that:

$$\|P_1(\xi) - P_1(\eta)\|_\lambda^2 = \mathbb{E}\left[\int_0^T e^{-\lambda s}\left|\int_0^s (f(\xi(t)) - f(\eta(t)))\,dt\right|^2 ds\right] \le \frac{C^2}{\lambda}\|\xi - \eta\|_\lambda^2.$$

Similarly, for P_2, using the isometry property of the Itô integral:

$$\|P_2(\xi) - P_2(\eta)\|_\lambda^2 = \mathbb{E}\left[\int_0^T e^{-\lambda s}\left|\int_0^s (g(\xi(t)) - g(\eta(t)))\,dW(t)\right|^2 ds\right]$$

$$= \mathbb{E}\left[\int_0^T e^{-\lambda s}\int_0^s |g(\xi(t)) - g(\eta(t))|^2 dt\,ds\right] = \mathbb{E}\left[\int_0^T |g(\xi(t)) - g(\eta(t))|^2 \int_t^T e^{-\lambda s} ds\,dt\right]$$

$$\le \frac{C^2}{\lambda}\|\xi - \eta\|_\lambda^2.$$

Thus, for the full operator P, we have:

$$\|P(\xi) - P(\eta)\|_\lambda \le \left(\frac{2C^2}{\lambda}\right)^{1/2}\|\xi - \eta\|_\lambda.$$

Choosing $\lambda > 2C^2$ ensures that P is a strict contraction. Some technical details remain, but this establishes the core idea of the proof.