Alma Mater Studiorum - Università di Bologna

Departement of Statistical Science
Curriculum Data Science

Stochastic Processes

Presented by:
Alberto Trashaj

Second edition 2022-2023

Abstract

Dear reader, thank you for your interest in my notes on stochastic processes. I have created these notes with care and attention to detail, attending every lecture and spending many hours outside of class to ensure their accuracy and clarity. While I do not wish to sell these notes, I would greatly appreciate any donation you may be willing to offer in exchange for their use. As a full-time student, attending these classes is a significant investment of time and resources for me, and your support would be greatly appreciated. Thank you for your consideration. If you have found these notes helpful, and would like to support my efforts in creating them, you can send a donation via PayPal to **trashajalberto@gmail.com**. Your contribution will go a long way in helping me continue to create and share high-quality educational materials. Thank you again for your support.

Contents

1 Introduction **1**
- 1.1 Definition . . . 1
  - 1.1.1 Remarks on S and T . . . 2
- 1.2 A process is a bivariate function . . . 3
  - 1.2.1 Another way to see what is a stochastic process . . . 3
- 1.3 Equality of the processes . . . 4
  - 1.3.1 Equivalent doesn't imply indistinguishable . . . 4
- 1.4 Stopping times . . . 5
- 1.5 Finite dimensional distributions . . . 6

2 Conditional expectation **7**
- 2.1 Conditional expectations . . . 7
  - 2.1.1 Best predictor? . . . 8
  - 2.1.2 Properties . . . 9

3 Martingales **11**
- 3.1 Discrete time: Martingales . . . 11
  - 3.1.1 Random walk . . . 12
  - 3.1.2 Connection between random walk and martingale . . . 13
- 3.2 Gambler's ruin problem . . . 15
- 3.3 Doob-Mayer decomposition . . . 16
  - 3.3.1 Convergence theorem of martingales . . . 18

4 Markov Processes - Markov Chains **19**
- 4.1 Definition . . . 19
- 4.2 Homogeneity . . . 19
- 4.3 Stationary sequence . . . 20
- 4.4 Recurrent chains . . . 21
- 4.5 Irreducible chains . . . 22
- 4.6 Generalization in k dimension . . . 23
- 4.7 Initial probability measure . . . 24
  - 4.7.1 Important remarks . . . 24
- 4.8 Aperiodic chains . . . 26
- 4.9 Ergodic chain . . . 26

5 Brownian motion **27**
5.1 Setting the table 27
5.2 How we define a brownian motion 27
5.2.1 Remarks 28
5.3 Brownian paths 28
5.3.1 Up to equivalence 28
5.4 Properties 29
5.5 Brownian motion as a martingale and calculation of covariance. . . . 29
5.6 Equivalent definitions of brownian processes 30
5.7 Standard Brownian Motion and Stopping Times 32

6 Poisson process **35**
6.1 The paths of a Poisson process 35
6.2 A counting process 36
6.3 Levy process 38
6.4 Exchangeable sequence 39
6.4.1 de Finetti's theorem 41

Chapter 1

Introduction

1.1 Definition

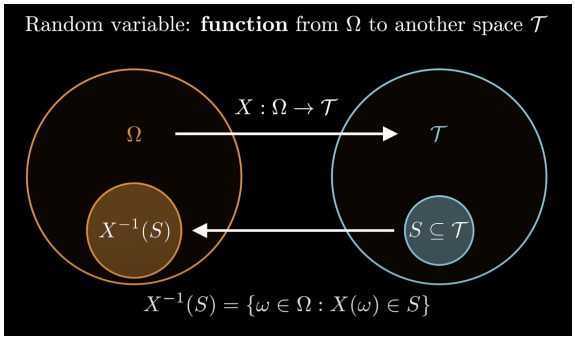
The first concept that we need to present in this course is the definition of stochastic process:

Definition 1.1.1 (Stochastic process)**.** A stochastic process is any collection of random variables.

So, quite simple, no?
Let's now briefly recall what is a random variables since it will be useful to understanding what we are talking about:
A random variable is a measurable map defined as follows:

$$X : \Omega \longrightarrow S$$

where we just recall that the triplet $(\Omega, \mathcal{A}, P)$ is the well known probability space and $(S, \mathcal{B})$ is the measurable space.
In particular, S is a set and B is a σ-field of subset of S.



It can be possible that you are not familiar with this concepts since some notion of the probability theory are needed, so if you have already problems with understanding what are we talking about pause the reading and try to comprehend the concept so far since they will be very useful for the rest of the course.

```
Disclaimer: in the repository
https://github.com/alberello00/stochastic_processes
you will find some notes of the course "Advanced probability"
which can be useful
to understanding some concepts.
```

So far we just briefly recalled the two main ingredients for defining a random variable: the other crucial ingredient that has to be mentioned when defining what is a stochastic process is the set T.
The set T can be any set, but the most natural and intuitive set that we can think of when we talk of stochastic processes is the **Time** set. But, later on we will define better what T actually is.
To give a formal definition we can repeat that a stochastic process X is any collection of random variables namely:

$$X = X_t : t \in T$$
$$\text{where } X_t : \Omega \longrightarrow S \text{ is a random variable } \forall t \in T$$

Another parentheses should be opened here: in fact, if one remember what it is a random variable can point out that a random variable has to be a measurable map, i.e. the random variables that are in our stochastic processes has to be measurable too:

$$X_t^{-1}(B) \in \mathcal{A} \; \forall B \in \mathcal{B}$$

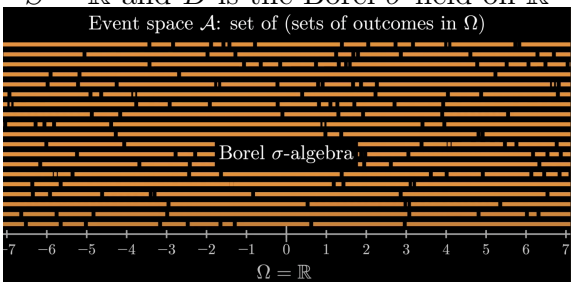
where

$$X_t^{-1}(B) = \{\omega \in \Omega : X_t(\omega) \in B\}$$

1.1.1 Remarks on S and T

- S is said to be the **state space** of the process: of course, the most important case is when the state space is equal to the real line, namely

$$S = \mathbb{R} \text{ and } B \text{ is the Borel } \sigma\text{-field on } \mathbb{R}$$



- T is an arbitrary set, which is usually called the **indexing set** of the **parameter space** of the process: a natural interpretation of this set can be thought as the *time* set.
- If T is a finite or countable set, X is said to be a discrete time process: for example, a sequence of random variables.
- If T is an interval of the real line $\mathbb{R}$, X is said to be a **continuous time process**

1.2 A process is a bivariate function

An important way to think of what is a stochastic process is that it is nothing else than a function of two variables:

$$X : \Omega \times T \longrightarrow S$$

So, we can write

$$X(\omega, t) \text{ or, equivalently, } X_t(\omega)$$

So, since a stochastic process can be thought as a function of two variables let's see what happens if we fix one of the two variables and let one be freely to change.

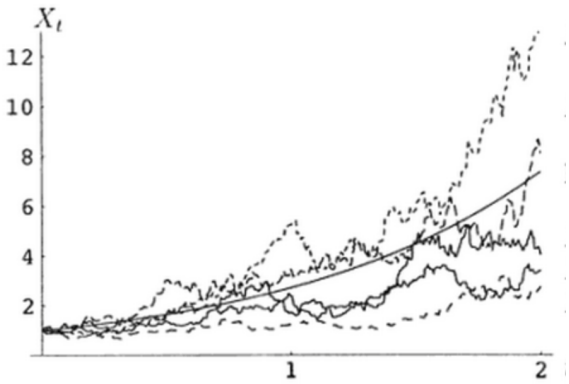
- Let fix $t \in T$:
  Then, X_t is still a random variable, namely a measurable function on Ω.
  Hence, we have

  $$X_t(\omega) \ \forall \omega \in \Omega$$

- Let fix $\omega \in \Omega$:
  Then, we obtain a function of t, namely:

  $$t \longrightarrow X_t(\omega)$$

  This function is usually called the **path** or **trajectory** of the process.
  For instance, if we choose the state process as $S = \mathbb{R}$ and $T = [0, \infty)$: the path is a function that maps from $[0, \infty)$ into $\mathbb{R}$



1.2.1 Another way to see what is a stochastic process

A process can be always regarded as a random function: it is sufficient to think of X as the map $\omega \longrightarrow$ to the path associated to ω, namely:

- $\omega \longrightarrow X(\omega, *)$

1.3 Equality of the processes

One can wonder when two processes can be regarded as "equal", "similar" or, in a more general sense, when two processes can be compared and see if they are in some sense indistinguishable.
In general, we can say that there are three level of "equality" of two processes:

1. $X \sim Y \Rightarrow (X_{t_1}, X_{t_2}, ..., X_{t_n}) \sim (Y_{t_1}, Y_{t_2}, ..., Y_{t_n})$
2. X is equal to $Y \longleftrightarrow P(X_t \neq Y_t) = 0 \ \forall \ t \in T$
3. X is indistinguishable from $Y \longleftrightarrow X_t(\omega) = Y_t(\omega) \ \forall \ t \in T$ and $\forall \ \omega \in A$ where $A \in \mathcal{A}$ is such that $P(A) = 1$

The natural question is if this three type of equality are in someway related:

We can state the following implications:
$$3 \Rightarrow 2 \Rightarrow 1$$
$$1 \nRightarrow 2 \nRightarrow 3$$

1.3.1 Equivalent doesn't imply indistinguishable

Usually to say that A doesn't imply B, it is sufficient to prove that A is true while B is false.
So, let's find an example where we have two processes that satisfies the second type of equality but not the third one.

Example 1.3.1. Let V be any random variable such that $V \geq 0$ (positive-definite) and $P(V = v) = 0 \ \forall \ v \geq 0$. For instance, we can take the absolute value of a standard Gaussian, so $V = |Z|$ where $Z \sim N(0, 1)$. Now, let's define two random process:

- $X(\omega, t) = 0 \ \forall \ t \in [0, +\infty) \ , \ \forall \ \omega \in \Omega$
- $$Y(\omega, t) = \begin{cases} 1, & \text{if } t = V(\omega). \\ 0, & \text{otherwise.} \end{cases} \tag{1.1}$$

Then, it's clear that X and Y are not indistinguishable since for $t = V(\omega)$ we get

$$Y(\omega, t) = 1 \text{ and } X(\omega, t) = 0, \text{ so } X(\omega, t) \neq Y(\omega, t)$$

However, with a close analysis, we can say that

$P(X_t \neq Y_t) = P(Y_t \neq 0)$ since $X_t = 0$ as assumption, and also $Y_t \neq 0$ when $t = V(\omega)$, therefore the $P(Y_t \neq 0) = P(V(\omega) = t)$ that we know it's equal to 0, since the probability of a point is 0 by the assumption that $P(V = v) = 0 \ \forall \ v \geq 0$

So, in conclusion we showed that X_t and Y_t are equivalent processes but they are not indistinguishable

1.4 Stopping times

Let's take a set $T = 0, 1, 2, ...$ and the usual sample space $(\Omega, \mathcal{A}, \mathcal{P})$; then we define the $filtration$ as an increasing sequence of sub-σ fields of $\mathcal{A}$, namely:

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset ... \subset \mathcal{A}.$$

Given this ingredients we can give a formal definition of the stopping time:

Definition 1.4.1. A stopping time is a map

$$T : \Omega \Rightarrow \{+\infty, 0, 1, 2, ...\}$$

such that $\{T = n\} \in \mathcal{F}_n \ \forall n \geq 0$.

The stopping time is generally known also as the stopping rule: in fact, a stopping time is a method for making decisions about whether to start or stop a procedure based on the current situation and previous occurrences.
In general, a σ-field $\mathcal{G} \subset \mathcal{A}$ may be used to describe the so called state of information. By now, it's not so clear what are we referring to when we talk about the "state of information" of $\mathcal{G}$, so let's make some examples:

Example 1.4.1.

- $\mathcal{G} = \{\varnothing, \Omega\}$ contains null information since, by definition, we always know that the $\varnothing$ is always $False$ and the Ω is always $True$.
- $\mathcal{G} = \{\varnothing, \Omega\, A, A^c\}$ in this set the only information we have is whether or not A is $True$.

Now, I think that another example should clarify better what we mean by "state of the information":

Example 1.4.2. Let's take $\Omega = \{1, 2, 3, 4, 5, 6\}$ and define a σ-algebra $\mathcal{F} = \{\{1\},\{2, 3\}\{4, 5\}\{\varnothing\}, \{\Omega\}\}$: now, we can see that, for example, the subset $\{1, 2\}$ is not contained, so our probabilistic model can't make conclusion in this last set, so the "information" is contained only in the set $\mathcal{F}$. Therefore, the only "probabilistic" conclusion we can make are only in the σ-field we constructed before; there will no exists random variable that maps the subset $\{1, 2\}$ on the interval $[0, 1]$.

T should be regarded as the first time when something we are interested actually happens: so, when we say that $T = n$ we are implicitly saying that at the time n something happened for the first time.
Remember when we defined the set where T takes values? Yes, at first can seems strange to include the $+\infty$ but now it should be more clear that saying $T = +\infty$ means that something does not happen.

Example 1.4.3. (X_n) is a sequence of real random variable and $A \in \mathcal{B}(\mathbb{R})$;
$T = inf\{n : X_n \in A\}$, so the first time n such that $X_n \in A$.

- $T = n = X_j \notin A \ \forall j < n, \ X_n = A$
- $T = +\infty = X_n \notin A \ \forall n$

Example 1.4.4 (Casinò)**.** Suppose you and your friends are playing at a roulette table in a casino, and you want to come up with a strategy to maximize your winnings. You decide that you will keep playing until a certain point in time, at which you will stop and walk away with your winnings. The catch is that you don't know ahead of time when this stopping time will occur - it could happen after 10 minutes of playing, or it could happen after several hours.
However, you have a secret weapon: a fortune teller who has told you that the stopping time will occur at the 10th round of the game. In other words, you will keep playing for 10 rounds, and then stop.
In this example, the stopping time is the $10th$ round of the game. This is a stopping time because it is determined by information that is available at or before the $10th$ round of the game. Specifically, the information that you have received from the fortune teller is available to you before the $10th$ round, and therefore the event $T = 10$ (where T is the stopping time) belongs to the sigma-algebra $\mathcal{F}_{10}$.

1.5 Finite dimensional distributions

Definition 1.5.1. Let X be a process indexed by T $\forall n \geq 1$ $\forall t_1, ..., t_n \in T$ we have a n-dimensional random vector $(X_{t_1}, X_{t_2}, ..., X_{t_n})$: the distributions of $(X_{t_1}, X_{t_2}, ..., X_{t_n})$ for all $n \geq 1$ and $\forall t_1, ..., t_n \in T$ are called finite dimensional distribution.

Example 1.5.1. Suppose that $T = \{1, 2\}$ and $X = \{X_1, X_2\}$: $X_1 \sim$ Binomial and $X_2 \sim$ Poisson. Then, for obvious reasons, $(X_1, X_2) \sim N$ it's not possible and such process fail to exist.

Usually, in application we choose the finite dimensional distributions and we look for a process having such a finite dimensional distribution: the problem here arise since this process may not exist.
However there are some theorem usually called *consistency* theorems which provides conditions on the finite dimensional distributions under which the processes (with such finite dimensional distribution) exists.

Chapter 2

Conditional expectation

2.1 Conditional expectations

In order to give a proper definition of conditional expectations, three ingredients are necessary:

1. A sample space, $(\Omega, \mathcal{A}, \mathcal{P})$
2. A sub-σ-field of $\mathcal{A}$: $\mathcal{G} \subset \mathcal{A}$
3. A real random variable X such that $E|X| < +\infty$.

Now, let's define a conditional expectation:

Definition 2.1.1. A conditional expectation of X given $\mathcal{G}$ is any real random variable $V : \Omega \Rightarrow \mathbb{R}$ such that:

1. $E(|V|) < +\infty$
2. V is $\mathcal{G}$-measurable
3. $E(\mathbb{1}_{\mathbb{A}} \cdot X) = E(\mathbb{1}_{\mathbb{A}} \cdot V)\ \forall A \in \mathcal{G}$

Where, on the third point we have,

$$X \cdot \mathbb{1}_{\mathbb{A}} = \begin{cases} \text{X, on } A \\ 0, \text{on } A^c. \end{cases} \tag{2.1}$$

In particular, since $\Omega \in \mathcal{G}$, we obtain:

$$E(X) = E(X \cdot \mathbb{1}) = E(V \cdot \mathbb{1}) = E(V)$$

We need to make a second remark on the second condition stated in the definition: a random variable V is said to be $\mathcal{G}$-measurable if

$$V^{-1}(B) \in \mathcal{G}\ \forall B \in \mathcal{B}(\mathbb{R})$$

In other terms, V is a random variable not only in the "big" probability space $(\Omega, \mathcal{A}, \mathcal{P})$ but also in the "small" probability space $(\Omega, \mathcal{G}, \mathcal{P})$.
For instance, take the set $\mathcal{G} = \{\varnothing, \Omega\}$ (null information), then if X is $\mathcal{G}$-measurable, it must be that

$$X^{-1}(B) = \Omega \text{ or } X^{-1}(B) = \varnothing \ \forall B \in \mathcal{B}(\mathbb{R})$$

Hence, with the set taken the only $\mathcal{G}$-measurable random variables are the constants. At the opposite extreme, if $\mathcal{G} = \mathcal{A}$, then every random variable is G-measurable.
However, the interpretation of $\mathcal{G}$-measurability is under the information $\mathcal{G}$, so a $\mathcal{G}$-measurable random variable reduces to a constant: suppose we have a random variable X that is $\mathcal{G}$-measurable for some sub-σ-field $\mathcal{G}$. Intuitively, this means that X behaves like a constant, because its value can be determined from the information in $\mathcal{G}$: let's see with an example why this is true.

Example 2.1.1. Fix $A \in \mathcal{A}$ and define

- $\mathcal{G} = \{\varnothing, \Omega, A, A^c\}$
- $X = 3\mathbb{1}_{\mathbb{A}} - 2\mathbb{1}_{\mathbb{A}^c}$

Then, X is $\mathcal{G}$-measurable:

$$X^{-1}(B) = \begin{cases} \Omega, \text{if } -2, 3 \in \mathcal{B} \\ \varnothing, \text{if } -2, 3 \notin \mathcal{B} \\ A, \text{if } 3 \in \mathcal{B} \text{ and } -2 \notin \mathcal{B} \\ A^c, \text{if } 3 \notin \mathcal{B} \text{ and } -2 \in \mathcal{B} \end{cases} \tag{2.2}$$

But, under the information of $\mathcal{G}$, we know whether A is $True$ or $False$, thus X becomes a constant.

Theorem 2.1.1 (Theorem)**.** A conditional expectation V always exists, and it is almost surely unique, namely, if V_1 and V_2 are both conditional expectations, then $P(V_1 \neq V_2) = 0$ or, equivalently, $P(V_1 = V_2) = 1$

From now on, to denote a conditional expectation of X given $\mathcal{G}$, we adopt the notation $E(X|\mathcal{G})$, namely $V = E(X|\mathcal{G})$.
The interpretation is that the expected value is our prediction of X under the information $\mathcal{G}$: in the special case where $V = E(X)$, the natural interpretation is that $E(X)$ is our prediction of Xwithout any specific information.
Now, we should be more clear on why $E(X|\mathcal{G})$ should be $\mathcal{G}$-measurable: in fact, our prediction of X under the information $\mathcal{G}$ should be something which only depends on the information $\mathcal{G}$.

2.1.1 Best predictor?

In general, to make a prediction, we are not only interested in finding a good predictor, but it would be better to find the **best** predictor: don't worry, we already find it! Yes, $E(X|\mathcal{G})$ is the best prediction we can make in the following sense:

- Suppose $E(X^2) < +\infty$: then,

$$E\{(X - (E(X|\mathcal{G}))^2\} = \min E\{(X - Z)^2\}$$

We can think of Z as a predictor of X under the information $\mathcal{G}$

2.1.2 Properties

There are 7 useful properties of the conditional expectation:

1. Linearity: $E(aX + bY|\mathcal{G}) = aE(X|\mathcal{G}) + bE(Y|\mathcal{G})$
2. Positivity: if $X \geq 0 \rightarrow E(X|\mathcal{G}) \geq 0$
3. Constant: $E(constant|\mathcal{G}) = constant$
4. Measurability: if Z is $\mathcal{G}$-measurable, then $E(X \cdot Z|\mathcal{G}) = Z \cdot E(X|\mathcal{G})$
   It follows that, if X is $\mathcal{G}$-measurable, then
   $$E(X|\mathcal{G}) = E(X \cdot 1|\mathcal{G}) = X \cdot E(1|\mathcal{G}) = X \cdot 1 = X$$
5. Independence: X is said to be independent from $\mathcal{G}$, by definition, if
   $$P(A, X \in B) = P(A) \cdot P(X \in B) \forall A \in \mathcal{G} \forall B \in \mathcal{B}(\mathbb{R})$$
   If X is independent from $\mathcal{G}$, then $E(X|\mathcal{G}) = E(X)$
6. Chain rule: if $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{A}$, then $E(X|\mathcal{G}_1) = E(E(X|\mathcal{G}_2)|\mathcal{G}_1)$.
   For instance if we know that $E(X|\mathcal{G}_2) = 0$, then
   $E(X|\mathcal{G}_1) = E(E(X|\mathcal{G}_2)|\mathcal{G}_1) = E(0|\mathcal{G}_1) = 0$
7. Conditional probability: if $A \in \mathcal{A}$, we can define
   $$P(A|\mathcal{G}) = E(\mathbb{1}_{\mathbb{A}}|\mathcal{G})$$

   An example is the following:

Example 2.1.2. If A and $B \in \mathcal{A}$ and $P(B) > 0$,

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

Despite this condition is well-known, a nice connection between this elementary definition and the general notion of conditional probability can be done.
In fact, the elementary definition is actually a special case of the general definition of probability conditional to a σ-field $\mathcal{G}$: let us take the following σ-field $\mathcal{G} = \{\varnothing, \Omega, B, B^c\}$
It can be shown that if $f : \Omega \rightarrow \mathbb{R}$ is $\mathcal{G}$-measurable, then:

$$f = \begin{cases} \alpha, \text{if } B \\ \beta, \text{if } B^c \end{cases} \tag{2.3}$$

for some constants α and β.
Hence,

$$P(A|\mathcal{G}) = \alpha \cdot \mathbb{1}_{\mathbb{B}} + \beta \cdot \mathbb{1}_{\mathbb{B}^c},$$

but,

$$E\{P(A|\mathcal{G}) \cdot \mathbb{1}_{\mathbb{D}}\} = E\{\mathbb{1}_{\mathbb{A}} \cdot \mathbb{1}_{\mathbb{D}}\} \; \forall D \in \mathcal{G}.$$

So, taking $D = B$, we get

$$E\{\mathbb{1}_{\mathbb{A}} \cdot \mathbb{1}_{\mathbb{B}}\} = E\{1_{A \bigcap B}\} = P(A \cap B)$$

Hence,

$$E\{P(A|\mathcal{G}) \cdot \mathbb{1}_{\mathbb{B}}\} = E\alpha \cdot \mathbb{1}_{\mathbb{B}} = \alpha \cdot E\{\mathbb{1}_{\mathbb{B}}\} = \alpha \cdot P(B)$$

It follows that,

$$\alpha = \tfrac{P(A \bigcap B)}{P(B)} = P(A|B)$$

Similarly, one obtain

$$\beta = P(A|B^c), \text{ provided that } P(B^c) > 0$$

In short,

$$P(A|\mathcal{G}) = P(A|B) \cdot \mathbb{1}_{\mathbb{B}} + P(A|B^c) \cdot \mathbb{1}_{\mathbb{B}^c}$$

Usually, we write $E(X|Y)$ to denote $E(X|\sigma(y))$, namely $E(X|Y) = E(X|\sigma(y))$: $\sigma(y)$ is the least field which makes Y measurable.
In short, when we have a random variable Y, $\sigma(Y)$ is the smallest σ-field making Y measurable.
Let's make another (more simple) example:

Example 2.1.3. Let Y, X be *iid* with $E|X| < +\infty$. What about $E(X|X+Y)$?
It can be shown that

$$E(X|X+Y) = \tfrac{X+Y}{2}$$

because

- $\frac{X+Y}{2}$ has the mean
- $\frac{X+Y}{2}$ is measurable with respect to $\sigma(X+Y)$
- $E(X \cdot \mathbb{1}_{\mathbb{A}}) = E(\frac{X+Y}{2} \cdot \mathbb{1}_{\mathbb{A}})$

Chapter 3

Martingales

3.1 Discrete time: Martingales

Definition 3.1.1. Given a filtration $\mathfrak{F}_0 \subset \mathfrak{F}_1 \subset ... \subset \mathcal{A}$ and the sequence of $x_0, x_1, x_2, ...$ real random variables: we say that (X_n) is a martingale with respect to $\mathfrak{F}_n$ if three condition are satisfied:

1. $E|X_n| < +\infty$
2. X_n is $\mathfrak{F}_n$-measurable
3. $E(X_{n+1}|\mathfrak{F}_n) = X_n$

Some remarks:

- If $E(X_{n+1}|\mathfrak{F}_n) \geq X_n$, then (X_n) is called sub-martingale.
- If $E(X_{n+1}|\mathfrak{F}_n) \leq X_n$ then (X_n) is called a super-martingale.
  Obviously a martingale is both a sub-martingale and a super-martingale.
- Let $\mathfrak{F}_n{}^* = \sigma(X_0, X_1, ..., X_n)$, this is the least sigma field that makes $(X_0, X_1, ..., X_n)$ measurable: therefore, (X_n) is still a martingale with respect to $(\mathfrak{F}_n{}^*)$.
  In fact, since $\mathfrak{F}_n{}^* \subset \mathfrak{F}_n$, one obtains

$$E(X_{n+1}|\mathfrak{F}_n{}^*) = E(E(X_{n+1}|\mathfrak{F}_\mathfrak{N})|\mathfrak{F}_n{}^*) = E(X_n|\mathfrak{F}_n{}^*) = X_n$$

  In the last expression we used respectively the chain rule, the definition of conditional expectation and the notion of measurability.

A simple theorem about martingales is described above:

Theorem 3.1.1. (X_n) is a martingale is and only if $\exists \subset \in \mathcal{R}$ and a Z_n such that $X_n = c + \sum_{i=0}^{n} Z_i$.
For instance, $E|Z_i| < +\infty$, Z_i is $\mathfrak{F}_\mathfrak{i}$ measurable and $E(Z_{i+1}|\mathfrak{F}_\mathfrak{i}) = 0$

A intuitive interpretation could be the following: imagine you are at the casinò, and Z_i represent the outcome of the i-th trial = the win at day i. X_n is the amount of money in your pocket after you play $(n+1)$ times.
If the game is **fair**, $E(Z_{i+1}|\mathfrak{F}_i) = 0$

Let's make an example:

Example 3.1.1. If Z_i is any independent sequence of real random variables such that

$$E(Z_i) = 0 \forall i, \text{ then}$$

$X_n = c + \sum_0^n Z_i$ is a proper martingale.
In fact, letting $\mathfrak{F}_n = \sigma(X_0, X_1, ..., X_n) = \sigma(Z_0, ..., Z_n)$, one obtain:

$E(Z_{i+1}|\mathfrak{F}_i) = E(Z_{i+1}|Z_0, ..., Z_i) =$ by independence $= E(Z_{i+1})$
and by assumption, $E(Z_{i+1}) = 0$

3.1.1 Random walk

A random walk is a process where an object or variable takes successive random steps in some space, such as a one-dimensional number line or a two-dimensional plane. The steps can be in any direction and can be positive or negative.
A simple example of a random walk is a gambler who starts with some amount of money and repeatedly flips a fair coin. If the coin lands heads, the gambler wins some amount of money, and if the coin lands tails, the gambler loses some amount of money. The gambler's wealth then follows a random walk as they make successive bets.

Definition 3.1.2. A sequence (X_n) is a random walk if:

$$X_n = X_0 + \sum_{i=1}^{n} Z_i, \tag{3.1}$$

where (Z_i) is a sequence of independent and identically distributed (i.i.d.) random variables, and X_0 is independent of (Z_i).

In this notation, the sum goes from $i = 1$ to n to indicate that the first term in the sum is Z_1, and X_0 is assumed to be a fixed constant or a random variable that is independent of (Z_i).

A very popular example of a random walk is the following:

Example 3.1.2.
- $P(X_0 \in Z) = 1$
- $P(Z_i = 1) = P(Z_i = -1) = \frac{1}{2}$

when $\mathcal{Z} = \{..., -2, -1, 0, 1, 2, ...\}$ is the set of integers.

3.1.2 Connection between random walk and martingale

A random walk is a martingale if and only if the following criteria are met:

- $E(|X_0|) < +\infty$
- $E(|Z_1|) < +\infty$
- $E(|Z_1|) = 0$

If X_n is a martingale and τ a stopping time, then $X_{(\tau \wedge n)}$ is still a martingale.

[Notion alert: $a \wedge b = \min(a, b)$]

In particular $X_{(\tau \wedge n)}$ is called stopped sequence:

$$X_{(\tau \wedge n)} = \begin{cases} X_n \text{ if } n \leq \tau \\ X_\tau \text{ if } n > \tau \end{cases} \tag{3.2}$$

Then the natural interpretation would be that when the event of interest occurs, the sequence became

$$X_1, X_2, ..., X_\tau, X_\tau, ...$$

In order to say that this is still a martingale, we have to show that

$$E(X_{\tau \wedge (n+i)} | \mathfrak{F}_\mathfrak{n}) = X_{\tau \wedge (n)}$$

Proof :
$E(X_{\tau \wedge (n+i)} | \mathfrak{F}_\mathfrak{n}) = E(X_\tau \mathbb{1}_{(\tau \leq \mathbb{n})} + X_{n+1} \mathbb{1}_{\tau > \mathbb{n}} | \mathfrak{F}_\mathfrak{n})$

$= E(X_\tau \mathbb{1}_{\tau \leq \mathbb{n}} | \mathfrak{F}_\mathfrak{n}) + E(X_{n+1} \mathbb{1}_{\tau > \mathbb{n}} | \mathfrak{F}_\mathfrak{n})$

The first term and the second term (since this one is $\mathfrak{F}_\mathfrak{n}$-measurable) can be rewritten in the following way:

$= E(\sum_{i=0}^{n} X_i \cdot \mathbb{1}_{\tau = \mathbb{i}}) + \mathbb{1}_{\tau > \mathbb{n}} \cdot E(X_{n+1} | \mathfrak{F}_\mathfrak{n})$

Moreover, using the definition of martingale,

$= \sum_{i=0}^{n} E(X_i \cdot \mathbb{1}_{\tau = \mathbb{i}} | \mathfrak{F}_\mathfrak{n}) + \mathbb{1}_{\tau > \mathbb{n}} \cdot \mathbb{X}_\mathbb{n}$

$= \sum_{i=0}^{n} X_i \cdot \mathbb{1}_{(\tau = \mathbb{i})} + \mathbb{1}_{\tau > \mathbb{n}} \cdot X_n = X_{\tau \wedge n}$

Another example follows:

Example 3.1.3. Let's take $X_n = \sum_{i=0}^{n} Z_i$ and $X_0 = 0$, where (Z_i) is *iid* and $E(Z_1) = 0$, $E(Z_1^2) = 1$.
Incidentally, in this case, X_n is both a martingale and a random walk.
Moreover $Y_n = X_n^2 - n$ is still a martingale. Let's prove this last point:
letting $\mathfrak{F}_\mathfrak{n} = \sigma(X_0, ..., X_n)$ we obtain

$$E(Y_{n+1}|\mathfrak{F}_\mathfrak{n}) = E(X_{n+1}^2 - (n+1)|\mathfrak{F}_\mathfrak{n})$$

And since $(n+1)$ is a constant

$$= E((X_n + Z_{n+1})^2|\mathfrak{F}_\mathfrak{n}) - (n+1)$$

by developing the square,

$$= E(X_n^2 + Z_{n+1}^2 + 2X_nZ_{n+1}|\mathfrak{F}_\mathfrak{n}) - (n+1)$$

applying the properties of martingales,

$$= X_n^2 + 2X_n \cdot E(Z_{n+1}|\mathfrak{F}_\mathfrak{n}) + E(Z_{n+1}^2|\mathfrak{F}_\mathfrak{n}) - (n+1)$$

$$= X_n^2 - (n+1) + 2X_n \cdot E(Z_{n+1}) + E(Z_{n+1}^2)$$

and by applying the assumptions

$$= X_n^2 - n = Y_n$$

Incidentally, it's possible to note that

$$X_n^2 = Y_n + n$$

is not a martingale.
In fact, if (M_n) is a martingale, then

$$E(M_{n+1}) = E(E(M_{n+1}|\mathfrak{F}_\mathfrak{n})) = E(M_n)$$

Hence, for every M_n, we get

$$E(M_0) = E(M_1) = ...$$

but, if we take the example before we can show that,

$$E(Y_n + n) = E(Y_0) + n = E(Y_{n+1} + (n+1)) \neq E(Y_n + n)$$

hence

$$E(Y_0) + (n+1) \neq E(Y_0) + n$$

For any real random variable X such that $E(|X|) < +\infty$ and any σ-field $\mathfrak{G} \subset A$, we have that

$$E(X) = E(E(X|\mathfrak{G}))$$

3.2 Gambler's ruin problem

In order to model this situation, let $X_0 = 0$ and $X_n = \sum_1^n Z_i$ where Z_i is iid and $P(Z_1 = 1) = P(Z_1 = -1) = \frac{1}{2}$.
Moreover, we suppose that player 1 has $a > 0$ euros and player 2 has $b > 0$ euros.
There are three possibilities

1. The game does not end, namely

$$-a < X_n < b \ \forall n$$

   It can be shown that $P(-a < X_n < b \forall n) = 0$. (For now we can take this fact as given)

2. The winner is player 1, namely

$$X_n = b \text{ happens before } X_n = -a$$

3. The winner is player 2, namely

$$X_n = -a \text{ happens before } X_n = b$$

Let $\tau =$ time when the game ends $= \inf n : X_n = -a or X_n = b$. Now, let's evaluate

$$E(X_\tau) = -a \cdot P(X_\tau = -a) + b \cdot P(X_\tau = b)$$

by denoting the last term with α, we get

$$= -a \cdot (1 - \alpha) + b \cdot \alpha$$

So, it's easy to show that

$$E(X_\tau) + a = \alpha(a + b)$$

and

$$\alpha = \frac{E(X_\tau)+a}{a+b}$$

where α is the probability that player 1 is the winner. In the end, we note that X_n is a martingale since $E(Z_1) = 0$, so that $X_{\tau \wedge n}$ is still a martingale.
Therefore, since $P(\tau < \infty) = 1$ and $n \to +\infty$ at a certain point

$$\tau = \min \tau, n$$

and so we get that

$$E(X_\tau) = E(\lim_{n\to\infty} X_{\tau\wedge n})$$

and by some mathematical fact, which we take as given, we have

$$= \lim_{n\to\infty} E(X_{\tau\wedge n})$$

and since $(X_{\tau\wedge n})$ is a martingale and $E(X_{\tau\wedge n}) = E(X_{\tau\wedge 0}) \forall n$ we conclude

$$= E(X_{\tau\wedge 0}) = E(X_0) = E(0) = 0$$

So, in conclusion $\alpha = \frac{a}{a+b}$.

3.3 Doob-Mayer decomposition

Let $f : \mathbb{R} \to \mathbb{R}$ be a function such that

$$E(|f(X_n)|) < +\infty \ \forall n$$

Then $f(X_n)$ is a sub-martingale provided

- X_n is a martingale and f is convex or,
- X_n is a sub-martingale and f is convex and increasing

Here, what we are trying to say is that, take X_n and you are interested to know if $f(X_n)$ is a martingale or a sub-martingale.

For instance, if X_n is a martingale, then, for example, X_n^2 or $|X_n|$ are sub-martingales.

Definition 3.3.1 (Doob-Meyer decomposition)**.** (X_n) is a sub-martingale if and only if it's possible to write the sub-martingale as follows:

- $X_n = M_n + A_n$

where M_n is a martingale and A_n is an increasing sequence, namely

$$0 = A_0 \leq A_1 \leq A_2 \leq ...$$

and A_n is $\mathfrak{F}_{n-1}$-measurable ("predictability").
Also, $E(A_n) < +\infty$.
Furthermore, the sequences A_n and M_n are almost surely unique.

Example 3.3.1. Suppose $X_n = \sum_1^n Z_i$, $X_0 = 0$ where (Z_i) is *iid* with $E(Z_i) = 0$ and $E(Z_i^2) = 1$. X_n is a martingale but X_n^2 is not a martingale.
However X_n^2 is a sub-martingale. Moreover, since $Y_n = X^2 - n$ is martingale, we can say that $M_n = Y_n = X^2 - n$ and $A_n = n$

To prove this theorem we have to show both the implication (from left to right and to right to left):

- Let first explain the "easy implication":

$$E(X_{n+1}|\mathfrak{F}_n) = E(M_{n+1} + A_{n+1}|\mathfrak{F}_n)$$

$$= E(M_{n+1}|\mathfrak{F}_n) + E(A_{n+1}|\mathfrak{F}_n)$$

since the first term is a martingale and the second is $\mathfrak{F}_n$-measurable

$$= M_n + A_{n+1}$$

and since it is an increasing sequence

$$\geq M_n + A_n = X_n$$

- Now we need to show the other way round, the "harder implication":
  Let $A_0 = 0$ and $A_n = \sum_{j=0}^{n-1}\{E(X_{j+1}|\mathfrak{F}_j) - X_j\}\ \forall n > 0$.
  Now that we defined A_n, we have to show that A_n is increasing and predictable:
  A_n is predictable because it involves only $X_0, X_1, ..., X_{n-1}$ and $E(X_1|\mathfrak{F}_0), ..., E(X_n|\mathfrak{F}_n - 1)$ and all such random variables are measurable with respect to $\mathfrak{F}_{n-1}$. Hence, A_n is $\mathfrak{F}_{n-1}$-measurable.
  Moreover, A_n is increasing as

$$A_{n+1} = \sum_{j=0}^{n}\{E(X_{j+1})|\mathfrak{F}_j - X_j\}$$

$$= A_n + \{E(X_{n+1}|\mathfrak{F}_n) - X_n\} \geq A_n$$

  Remind me to explain better this steps please.
  Next, let's define $M_n = X_n - A_n$:
  it remains only to see that M_n is a martingale.
  In fact,

$$E(M_{n+1}|\mathfrak{F}_n) = E(X_{n+1} - A_{n+1}|\mathfrak{F}_n)$$

$$= E(X_{n+1}|\mathfrak{F}_n) - E(A_{n+1}|\mathfrak{F}_n)$$

$$= E(X_{n+1}|\mathfrak{F}_n) - A_{n+1}$$

$$= E(X_{n+1}|\mathfrak{F}_n) - \{A_n + E(X_{n+1}|\mathfrak{F}) - X_n\}$$

$$= X_n - A_n$$

Now, let's prove that A_n and M_n are actually almost surely unique:
to prove this fact suppose

$$M_n + A_n = X_n = M_n^{'} + A_n^{'}$$

Then, $M_n - M_n^{'}$ is a martingale and $A_n{}' - A_n$ is $\mathfrak{F}_{n-1}$-measurable: thus,

$$M_{n-1} - M_{n-1}^{'} = E(M_n - M_n^{'}|\mathfrak{F}_{n-1}) = E(A_n^{'} - A_n|\mathfrak{F}_{n-1})$$

$$= A_n^{'} - A_n$$

Hence,

$$A_n^{'} - A_n = M_{n-1} - M_{n-1}^{'}$$

Thus, for $n < 1$,

$$A_1^{'} - A = M_0 - M_0^{'} = X_0 - X_0 = 0$$

Hence $A_1^{'} = A_1$, and by induction $A_n^{'} = A_n\ \forall n$: since

$$M_n + A_n = X_n = M_n^{'} + A_n^{'}$$

we obtain that $M_n = M_n^{'}\ \forall n$

3.3.1 Convergence theorem of martingales

Theorem 3.3.1. Let X_n be a submartingale. If $\sup_n E[|X_n|] < \infty$, then:

1. X_n converges almost surely (a.s.) to some limit X: $X_n \xrightarrow{a.s.} X$
2. If X_n is uniformly integrable, then X_n converges a.s. and in L^1 to X: $X_n \xrightarrow{a.s., L_1} X$

- Remark
  - X_n is said to be uniformly integrable if:

    $$\lim_{c\to+\infty} \sup_n E\{|X_n| \cdot \mathbb{1}_{|X_n|>c}\} = 0$$

    If the $\sup_n E\{|X_n|^{1+\epsilon}\}$ is finite for some $\epsilon > 0 \Rightarrow X_n$ is uniformly integrable. The contrary is not true.
    It is also possible to show that if X_n is uniformly integrable $\Rightarrow \sup_n E|X_n| < +\infty$

Let's show an example where the $\sup_n E|X_n| < +\infty$ but X_n is not uniformly integrable.
Let X_n be such that

$$P(X_n = 0) = \tfrac{n-1}{n} \text{ and } P(X_n = n) = \tfrac{1}{n}$$

Then, $E|X_n| = n \cdot P(X_n = n) + 0 \cdot P(X_n = 0) = n \cdot \frac{1}{n} = 1$.
Hence, $\sup_n E|X_n| = \sup_n 1 = 1 < +\infty$.
Fix $c > 0$, then $\sup_n E(|X_n| \cdot \mathbb{1}_{||X_n>c})$

$$= \sup_n [E\{\mathbb{1}_{X_n=0} \cdot \mathbb{1}_{|X_n|>c} \cdot (|X_n|)\} + E\{\mathbb{1}_{X_n=n} \cdot \mathbb{1}_{|X_n|>c} \cdot (|X_n|)] \tag{3.3}$$

$$= \sup_n E\{\mathbb{1}_{(X_n=0)} \cdot \mathbb{1}_{(X_n=c)} \cdot n\}$$

$$= \sup_n n \cdot E\{\mathbb{1}_{(X_n=n)}\}$$

$$= \sup_n n \cdot \tfrac{1}{n} = 1$$

then

$$\lim_{c\to+\infty} \sup_n E(|X_n| \cdot \mathbb{1}_{|X_n|>c}) = 1 \neq 0$$

Has to be completed

Chapter 4

Markov Processes - Markov Chains

4.1 Definition

Definition 4.1.1. A sequence $X_0, X_1, X_2, ...$ of real i.i.d. random variables to be a markov chain with respect to a filtration $\mathfrak{F}_0 \subset \mathfrak{F}_1 \subset ... \subset \mathfrak{A}$ if two conditions are satisfied:

1. X_n is $\mathfrak{F}_n$-measurable $\forall$ n
2. $P(X_{n+1} \in A|\mathfrak{F}_n) = P(X_{n+1} \in A|X_n)\ \forall n \geq 0\ \forall A \in \mathfrak{B}(\mathfrak{R})$

Roughly speaking, we are saying that P(future|Past and present) = P(future|present)

4.2 Homogeneity

Definition 4.2.1. Let $\alpha(x, A) = P(X_{n+1} \in A|X_n = x)\ \forall x \in \mathbb{R},\ A \in \mathfrak{B}(\mathbb{R})$ and $n \geq 0$
$\Downarrow$
then the sequence (X_n) is homogeneous $\leftrightarrow \alpha_n = \alpha_0\ \forall n \geq 0$.

If (X_n) is an homogeneous markov chain, the probability distribution of the sequence is determined by:

- The probability distribution of X_0, usually called initial distribution;
- The conditional distribution α_0, usually called transition kernel.

If we write$\alpha = \alpha_0$ and we call Π the probability distribution of X_0, we have, by applying the definition, that the

$$P(X_0 \in A_0, X_1 \in A_1) = E_x\{P(X_1 \in A_1|X_0 = x) \cdot \mathbb{1}(\mathbb{x})_{A_0}\}$$

$$= E_x\alpha(x, A_1) \cdot \mathbb{1}_{x \in A_0}$$

$$= \sum_x \alpha(x_j, A_1) \cdot \mathbb{1}_{x \in A_0} \cdot \Pi(x) \text{ if } X_0 \text{ is discrete.}$$

Instead, if X_0 is absolutely continuous,

$$P(X_0 \in A_0, X_1 \in A_1) = \int \alpha(x_j, A_1) \cdot \mathbb{1}_{x \in A_0} \cdot f(x), \text{ where } f \text{ is the density of } X_0$$

Suppose for instance, that X_n is discrete for every n: then,

$$P(X_0 = x_0, X_1 = x_1, X_2 = x_2) = P(X_2 = x_2 | X_0 = x_0, X_1 = x_1) \cdot P(X_0 = x_0, X_1 = x_1)$$

$$= P(X_2 = x_2 | X_0 = x_0, X_1 = x_1) \cdot P(X_1 = x_1 | X_0 = x_0) \cdot P(X_0 = x_0)$$

$$= P(X_2 = x_2 | X_1 = x_1) \cdot P(X_1 = x_1 | X_0 = x_0) \cdot P(X_0 = x_0)$$

$$= \alpha(x_1, \{x_2\}) \cdot \alpha(x_0, \{x_1\}) \cdot \Pi(x_0)$$

4.3 Stationary sequence

In general, a sequence (X_n) is said to be stationary if

$$(X_1, X_2, X_3, ...) \sim (X_0, X_1, X_2, X_3, ...).$$

If (X_n) is stationary, then $X_n \sim X_0 \ \forall n \geq 0$, but not conversely.
Instead, if X_n is an homogeneous markov chain, then

$$(X_n) \text{ is stationary} \Leftrightarrow X_n \sim X_0 \ \forall n \geq 0$$

Example 4.3.1. Any random walk is an homogeneous markov chain: in fact, let $X_n = X_0 + \sum_1^n Z_i$ where

- X_0 is independent of (Z_i)
- (Z_i) is iid

Let $\mathfrak{F}_\mathfrak{n} = \sigma(X_0, X_1, x_2, ...X_n)$, Then,

$$P(X_{n+1} \in A | X_0 = x_0, ..., X_n = x_n) = P(X_n + Z_{n+1} \in A | X_0 = x_0, ..., X_n = x_n)$$

$$= P(x_n + Z_{n+1} \in A | X_0 = x_0, ..., X_n = x_n)$$

and since independence

$$= P(x_n + Z_{n+1} \in A)$$

and since they are identically distributed

$$= P(X_n + Z_1 \in A)$$

So, similarly

$$P(X_{n+1} \in A | X_n = x_n) = P(x_n + Z_1 \in A)$$

In this case

$$\alpha(x, A) = P(X_1 \in A | X_0 = x) = P(x + Z_1 \in A)$$

In particular, let us consider the symmetric random walk on $\mathbb{Z}$:

$$X_0 \in \mathbb{Z} \text{ and } P(Z_1 = 1) = P(Z_1 = -1) = \tfrac{1}{2}$$

In this case $\alpha(i, \{j\}) = P(X_1 = j | X_0 = i)$, that can be written as

$$= P(Z_1 + i = j) = P(Z_1 = j - i)$$

$$= \begin{cases} \frac{1}{2} \text{ if } j - i = 1 \text{ or } j - i = -1 \\ 0 \text{ otherwise} \end{cases} \tag{4.1}$$

If you have reach this point, well done! Don't give up, we are almost at the half of our journey!!

4.4 Recurrent chains

So far we considered a general homogeneous markov chain: from now on, we deal with an important special case, namely we assume X_n is discrete $\forall n \geq 0$.
In this case, we have

$$P(X_n \in S, \forall n \geq 0) = 1$$

where S is a finite or countable set, called the **state space** of the chain and $x \in S$ is called **state**.
A state x can be **recurrent** or **transient**:

- is said to be **recurrent** if $P(\exists n > 0 \text{ with } X_n = x | X_0 = x) = 1$
- is said to be **transient** if $P(\exists n > 0 \text{ with } X_n = x | X_0 = x) < 1$

Let's recall the definition of a stopping time:

$$\tau_x = \inf\{n > 0 : X_n = x\} \begin{cases} \text{first time } n > 0 \text{ such that } X_n = x \\ +\infty \text{ if } X_n \neq x \forall n > 0 \end{cases} \tag{4.2}$$

Now, we can show that if x is recurrent, then

- x is null recurrent if

$$E(\tau_x | X_0 = x) = +\infty$$

- x is positive recurrent if

$$E(\tau_x | X_0 = x) < +\infty$$

4.5 Irreducible chains

An homogeneous markov chain is **irreducible** if $\forall x, y \in S \exists m \geq 1 and n \geq 1$ such that

- $P(X_n = y | X_0 = x) > 0$
- $P(X_n = x | X_0 = y) > 0$

revise the notation above. Now, let α be the transition kernel:

$$\alpha(x, A) = P(X_1 \in A | X_0 = x) \ \forall x in \mathbb{R} \forall A \in \mathbb{B}(\mathbb{R})$$

In this case, since S is countable we will write

$$\alpha(x, y) \text{ instead of } \alpha(x, \{y\}) \ \forall x, y \in S$$

Hence, $\alpha(x, y) = P(X_1 = y | X_0 = x)$ In this notation, let us suppose $S = \mathbb{Z}$ and $\alpha(n, n+1) = 1$ In this case, $\alpha(n, n+1)$ represents the probability of transitioning from state n to state $n+1$ in one step, given that the current state is n. Since $\alpha(n, n+1) = 1$, it means that the chain can only move to state $n+1$ from state n and cannot stay in state n or move to any other state in one step.
Next, suppose

- $S = \{a, b\}$
- $\alpha(a, b) = \alpha(b, a)$

if the chain is irreducible, all states are of the same type. If S is finite we can conclude that $\exists$ a positive recurrent state but $\nexists$ any null reccurent state.

Theorem 4.5.1. Let $x \in S$. Then,

$$x \text{ is recurrent} \Leftrightarrow \sum_{n=1}^{+\infty} P(X_n = x | X_0 = x) = +\infty \Leftrightarrow$$
$$P(X_n = x \text{ for infinetly many n } | X_0 = x) = 1$$

The first equivalence follows from the definition of recurrent state, which is that starting from x, the chain returns to x with probability 1. This means that the probability of returning to x in n steps, given that the chain started at x is $P(X_n = x | X_0 = x)$. So if x is recurrent, then the probability of ever returning to x is infinite, which implies that the sum of these probabilities over all n is infinite as well.

Example 4.5.1. Let (X_n) be the symmetric random walk on $\mathbb{Z}$, namely $X_n = X_0 + \sum_i^n Z_i$, X_0 is independent of (Z_i), and Z_i is i.i.d.
$P(X_0 \in Z) = 1$,
$P(Z_1 = 1) = P(Z_1 = -1) = \frac{1}{2}$.
This chain is irreducible:
let us consider $\sum_{n=1}^{+\infty} P(X_n = 0 | X_0 = 0)$

$$= \sum_{n=1}^{+\infty} P(X_{2n} = 0 | X_0 = 0)$$

where the probability above can be rewritten as follow:
$P(X_{2n} = 0|X_0 = 0) = P(\sum Z_i = 0|X_0 = 0) = P(\sum Z_i = 0) =$
$P(\sum Z_i + 2n = 2n) = P(\sum \frac{(Z_i+1)}{2} = n)$
so that

$$= P(\sum_1^{2n} Z_i = 0)$$
$$= \sum_{n=1}^{+\infty} P(\frac{\sum Z_i+1}{2} = n)$$
$$\sum P(Bin(2n, \tfrac{1}{2}) = n)$$
$$= \sum_{n=1}^{+\infty} \binom{2n}{n} (\tfrac{1}{2})^{2n}$$

It can be shown that

$$\binom{2n}{n} (\tfrac{1}{2})^{2n} = \frac{c_n}{\sqrt{n}}$$

where $C_n \to c$ with $c \in (0, +\infty)$, and since $\sum_n \frac{1}{\sqrt{n}} = +\infty$ (since it is a divergent series):
therefore,

$$= \sum_{n=1}^{+\infty} \binom{2n}{n} (\tfrac{1}{2})^{2n} = +\infty$$

In order to prove that 0 is null recurrent, toward a contradiction, we assume that

$$E(\tau_0|X_0 = 0) < +\infty$$

In this case

$$Var(X_{\tau_0}|X_0) = Var(\sum_{i=1}^{\tau_i} Z_i|X_0 = 0) = E(\tau_0|X_0 = 0) > 0$$

So $Var(X_{\tau_0}|X_0 = 0) > 0$ but this is a contradiction since $X_{\tau_0=0}$.
Therefore, it must be that

$$E(\tau_0|X_0 = 0) = +\infty \Leftrightarrow 0 \text{ is null recurrent.}$$

4.6 Generalization in k dimension

It is possible to define a symmetric random walk in

$$\mathbb{Z}^k = \mathbb{Z} \times ... \times \mathbb{Z} = \{(J_1, ..., J_k) : J_i \in Z \forall i\}$$

Definition 4.6.1. Let $X_n = X_0 + \sum_{i=1}^n Z_i$, (Z_i) i.i.d., X_0 is independent of (Z_i), $P(X_0 \in \mathbb{Z}^k) = 1$ and

$$P(Z_1 = x) = \tfrac{1}{2k} \ \forall x \bigcup_{i=1}^k \{y : y_j = 0 \forall j \neq i, y_i \in \{-1; 1\}\}$$

Note that if $k = 1$, we end in the simple case that we already studied.
If $k = 2$, insert image it is clearly irreducible so that all states are of the same type (all transient or all recurrent). This chain is null recurrent if $k = 1$ or $k = 2$ but it is transient if $k > 2$.

The example is describing a generalization of a random walk in one dimension to a random walk in k dimensions, where k is a positive integer. In this case, the state space is $\mathbb{Z}^k$, which is the Cartesian product of k copies of $\mathbb{Z}$.

The definition specifies that the random walk is symmetric, which means that the probability of moving in any given direction is the same. The step sizes are drawn from a distribution that is uniform over the $2k$ nearest neighbors in each direction.

4.7 Initial probability measure

A probability measure π on S is called an "inital" probability measure because its merely regards the distribution of X_0: in particular, an initial probability Π is said to be a stationary initials if

$$X_0 \sim \pi \Rightarrow X_1 \sim \pi$$

Hence, if π is stationary one obtains:
$X_0 \sim X_1 \Leftrightarrow (X_n)$ is stationary.
Recall that $P(X_1 = a) = \sum_{x \in S} P(X_1 = a | X_0 = x) \cdot P(X_0 = x)$: so,

$$\pi \text{ is stationary} \Leftrightarrow P(X_1 = a) = P(X_0 = a) \ \forall a \in S$$

$$\Updownarrow$$

$$\pi(a) = \sum_{x \in S} \alpha(x, a)\pi(x) \ \forall a \in S$$

An important question that arise is the following:

$$\text{Given the transition kernel } \alpha \ \exists \text{ a stationary probability } \pi?$$

We can answer to this question as follows:

$$\text{There } \exists \text{ a stationary probability } \pi$$

$$\Updownarrow$$

$$\exists \text{ a positive recurrent state.}$$

4.7.1 Important remarks

- If S is finite, there $\exists$ a positive recurrent state while $\nexists$ any null recurrent state. Hence if S is finite, there $\exists$ a stationary probability π
- If the chain is irreducible, a stationary probability (provided it exists) it's unique. Moreover

  $$\pi(a) = \frac{1}{E(\tau_a | X_0 = a)}$$

  where $\tau_a = inf\{n > 0 : X_n = a\}$

Let's make some examples:

Example 4.7.1. Let $S = \{a, b\}$ and $\alpha(a, b) = \alpha(b, a) = 1$: who is the initial probability?
The chain is irreducible, positive recurrent and the only stationary probability is

$$\pi(a) = \pi(b) = \frac{1}{2}$$

In fact, suppose to get

$$\sum_{x\in S}\pi(x)\alpha(x,a) = \pi(a)\alpha(a,a) + \pi(b)\alpha(b,a)$$
$$= 0 + \pi(b)\alpha(b,a)$$
$$= \tfrac{1}{2}\cdot 1 = \tfrac{1}{2} = \pi(a)$$

Example 4.7.2. If (X_n) is the symmetric random walk on $\mathbb{Z}$, then there $\nexists$ positive recurrent state and thus $\nexists$ any stationary probability.
However, suppose we let

$$\pi(x) = 1 \ \forall x \in S$$

Then

$$\sum_{x\in S}\pi(x)\cdot\alpha(x,a) = \pi(a-1)\cdot\alpha(a-1,\alpha)+\pi(a+1)\cdot\alpha(a+1,\alpha)+\sum_{x\in S}\pi(x)\alpha(x,a)$$

where the last term is 0, so that

$$= \pi(a-1)\alpha(a-1,\alpha) + \pi(a+1)\alpha(a+1,\alpha)$$
$$= 1 = \pi(a)$$

In other terms, π meets the equation

$$\pi(a) = \sum_{x\in S}\pi(x)\alpha(x,a) \ \forall a \in S$$

However, π is not a probability, as

$$\pi(s) = \sum_{x\in S}\pi(x) = \sum_{x\in S} 1 = +\infty$$

Example 4.7.3. Let X_n be the number of balls in the first box.
Thus

$$S = \{0, 1, ..., r\}$$

And the kernel

$$\alpha(i,j) = P(X_1 = j | X_0 = i)$$
$$= \begin{cases} 1 - \frac{i}{r} \text{ if } j = i+1 \\ \frac{i}{r} \text{ if } j = i-1 \\ 0 \text{ otherwise} \end{cases} \tag{4.3}$$

In this case, S is finite and the chain is irreducible, so $\exists!$ a stationary probability, namely

$$\pi(x) = \frac{\binom{r}{x}}{2^r} \ \forall x \in S$$

Note

$$\sum_{x\in S} = \frac{\sum_{j=0}^{r}\binom{r}{x}\cdot 1^j\cdot 1^{r-j}}{2^r}$$

remember Newton binomial?

$$= \frac{(1+i)^r}{2^r} = \frac{2^r}{2^r} = 1$$

This proves that π is a probability measure.
It remains to see that

$$\pi(a) = \sum_{x\in S}\alpha(x,a)\cdot\pi(x) \ \forall a \in S$$

4.8 Aperiodic chains

The last result that we are going to state is the definition of aperiodic chains.

Definition 4.8.1. A state $a \in S$ is said to be aperiodic if

$$\{n > 0 : P(X_n = a | X_0 = a) = 0\}$$

is a finite set.

For instance if $S = \{a, b\}$, $\alpha(a, b) = \alpha(b, a) = 1$, both a and b are not aperiodic. Similarly, if X_n is the symmetric random walk on $\mathbb{Z}$, all states are not aperiodic. In conclusion, if the chain is irreducible, then either all states are aperiodic or all states are not aperiodic.

4.9 Ergodic chain

Definition 4.9.1. The chain is said to be ergodic if $\exists$ a probability π on S such that $\pi(a) = \lim_{n \to +\infty} P(X_n = a | X_0 = x)$ $\forall x \in S$.

In this case, it can be shown that π is the unique stationary probability.
Note that, if the chain is egodic, for large n, the chain "forgets" the starting state x.

Theorem 4.9.1. If the chain is aperiodic, irreducible and positive recurrent then it is egodic.

Chapter 5

Brownian motion

5.1 Setting the table

Let $X = \{X_t : t \geq 0\}$ be a real process indexed by the non-negative reals (i.e. $S = \mathbb{R}$ and $T = [0; +\infty)$).
In this framework a filtration is any collection of σ-fields $\{\mathfrak{F}_t : t \geq 0\}$ such that

$$\mathfrak{F}_s \subset \mathfrak{F}_t \subset \mathfrak{A} \ \forall 0 \leq s \leq t$$

Definition 5.1.1. X is said to be adapted to $(\mathfrak{F}_t) \iff X_t$ is $\mathfrak{F}_t$-measurable $\forall t \geq 0$

Then, we say that X is a martingale $\Updownarrow$

- X is adapted to $\mathfrak{F}_t$
- $E|X_t| < +\infty \ \forall t \geq 0$
- $E(X_t|\mathfrak{F}_s) = X_s \ \forall 0 \leq s < t$

Definition 5.1.2. X has independent increments $\Updownarrow$ The increment $(X_t - X_s)$ is independent of $\mathfrak{F}_s \ \forall 0 \leq s < t$

To understand this notation suppose that

$$\mathfrak{F}_t = \sigma(X_u : u \leq t)$$

So, if X has independent increments we say that X has stationary increments $\Updownarrow$ $(X_t - X_s) \sim (X_{t-s} - X_0) \ \forall 0 \leq s < t$

5.2 How we define a brownian motion

Definition 5.2.1. A process $X = \{X_t : t \geq 0\}$ is said to be a brownian motion with respect to a filtration $\mathfrak{F}_t$ if

- X is adapted to $(\mathfrak{F}_t)$
- X has independent increments
- $X_0 = 0$ *a.s.* and $X_t \sim N(0, \sigma^2 \cdot t) \ \forall t > 0$

5.2.1 Remarks

- If $\sigma^2 = 1$, X is called a standard brownian motion.
- As a consequence of the definition, a brownian motion has stationary increments.

Let's dive in the second result: fix $0 \leq s < t$ and $X_t = X_s + X_t - X_s$.
Since the increments are independent X_s is independent of $(X_t - X_s)$, hence

$$\phi_{X_t} = \phi_{X_s} \cdot \phi_{X_{t-s}}$$

$$\phi_{X_{t-s}}(u) = \frac{\phi_{X_t}(u)}{\phi_{X_s}(u)} = \frac{e^{-\frac{1}{2}t\sigma^2}}{e^{-\frac{1}{2}s\sigma^2}} = e^{-\frac{1}{2}(t-s)\sigma^2}$$

which is the characteristic function of a $N(0, \sigma^2(t-s))$.
This proves that

$$X_t - X_s \sim X_{t-s} = X_{t-s} - X_0$$

To summarize, if X is a Brownian motion with variance σ^2, then for any $0 \leq s < t$, we have $X_t - X_s \sim N(0, \sigma^2(t-s))$, which implies that the distribution of the increment $X_t - X_s$ depends only on the time difference $t - s$ and not on the specific values of s and t.

5.3 Brownian paths

Definition 5.3.1. A path of brownian motion is a function from $[0; +\infty)$ into $\mathbb{R}$.

Such function satisfies several properties, but due to time reason we will mention only two of them. But before explaining them we need to specify what we mention with the term **up to equivalence**.

5.3.1 Up to equivalence

Say that X satisfy a property up to equivalence of $\exists$ a process Z such that

- $P(X_t \neq Z_t) = 0 \ \forall t$
- Z satisfy the property

Saying that X satisfies a property up to equivalence means that there exists another process Z which is essentially the same as X in the sense that they behave in the same way almost surely. The first bullet point mentioned ensures that X and Z are equivalent in terms of their values at all times, while the second bullet point ensures that Z also satisfies the same property that X does.

5.4 Properties

Essentially, it can be shown that a brownian motion X has continous (1) and nowhere differentiable (2) paths up to equivalence.

1. Continuous paths: The paths of X are continuous functions of time. More precisely, for any fixed $\omega \in \Omega$, the path $t \mapsto X_t(\omega)$ is a continuous function from $[0, +\infty)$ to $\mathbb{R}$.

2. Nowhere differentiable paths: The paths of X are almost surely nowhere differentiable. This means that for almost every $\omega \in \Omega$, the path $t \mapsto X_t(\omega)$ is not differentiable at any point in $[0, +\infty)$. In other words, the graph of the path has infinite slope at every point, and the path looks jagged and irregular when zoomed in at any scale.

5.5 Brownian motion as a martingale and calculation of covariance.

A brownian motion is a martingale: in fact, if $0 \leq s < t$,

$$E(X_t|\mathfrak{F}_s) = E(X_s + X_t - X_s|\mathfrak{F}_s)$$

$$= X_s + E(X_t - X_s|\mathfrak{F}_s) = X_s + E(X_t - X_s)$$

And since X_t and X_s has both mean 0

$$= X_s + 0$$

Moreover, let's compute the covariance between X_t and X_s:

$$Cov(X_t, X_s) = E(X_s X_t) - E(X_s)E(X_t)$$

$$= E(X_s X_t) - 0 \cdot 0$$

$$= E\{X_s \cdot [X_s + X_t - X_s]\}$$

$$= E(X_s^2) + E(X_s(X_t - X_s))$$

and since the increments are independent

$$= \sigma^2 \cdot s + E(X_s) \cdot E(X_t - X_s)$$

where the last two terms are clearly 0, so that

$$= \sigma^2 \cdot s$$

In conclusion,

$$Cov(X_t, X_s) = \sigma^2 \cdot s \text{ if } 0 \leq s < t$$

$$Cov(X_t, X_s) = \sigma^2(s \wedge t) \ \forall s, t \geq 0$$

5.6 Equivalent definitions of brownian processes

- Let X be a processes such that $X_0 = 0$ *a.s.*
  Then X is a brownian motion if and only if:

  1. X is gaussian
  2. $E(X_t) = 0$
  3. $E(X_s X_t) = \sigma^2(s \wedge t)\ \forall s, t \geq 0$

  $\Updownarrow$
  X has independent and stationary increments and continuous paths up to equivalence
  $\Updownarrow$
  X is a martingale, $(X_t^2 - \sigma^2 \cdot t)$, is a martingale and X has continuous paths up to equivalence.

Remarks

- A process Y is said to be *Gaussian* if its finite dimensional distribution are Normal, namely

$$(Y_{t_1}, Y_{t_2}, ..., Y_{t_n}) \sim N\ \forall n \geq 1\ \forall t_1, ..., t_n$$

- In the third definition we said that $(X_t^2 - \sigma^2 \cdot t)$ is a martingale: so, as an exercise, let's prove that a process $Y_t = X_t^2 - \sigma^2 \cdot t$ is a martingale.
  In fact, if $0 \leq s < t$,

$$E(Y_t|\mathfrak{F}_s) = E(X_t^2 - \sigma^2 \cdot t|\mathfrak{F}_s)$$

since $-\sigma^2 \cdot t$ is a constant

$$= -\sigma^2 \cdot t + E[(X_t^2|\mathfrak{F}_s)]$$

$$= -\sigma^2 \cdot t + E[(X_s + X_t - X_s)^2|\mathfrak{F}_s]$$

by developing the square

$$= -\sigma^2 \cdot t + E[X_s^2 + (X_t - X_s)^2 + 2X_s(X_t - X_s)|\mathfrak{F}_s]$$

then since the expectation of the sum is the sum of the expectations

$$= -\sigma^2 \cdot t + X_s^2 + E[(X_t - X_s)^2|\mathfrak{F}_s] + 2X_s E[(X_t - X_s|\mathfrak{F}_s)]$$

since we assumed independence we can drop the conditioning

$$= -\sigma^2 \cdot t + X_s^2 + E[(X_t - X_s)^2] + 2X_s E[X_t - X_s]$$

now, it's possible to see that the last term is 0 and the second term is equal to $\sigma^2(t-s)$, so

$$= X_s^2 - \sigma \cdot s = Y_s$$

Example 5.6.1. Given a standard brownian motion X, define $Y_0 = 0$ and $Y_t = t \cdot X_{\frac{1}{t}}$ $\forall t > 0$.
Then Y is still a standard brownian motion: we have to prove that

1. Y is gaussian,
2. $E(Y_t) = 0$,
3. $E(Y_s \cdot Y_t) = s \wedge t$.

1. Gaussian: $\forall n \geq 1$ $\forall t_1, ..., t_n$
   The process can be represented as a vector:

$$\begin{pmatrix} Y_{t_1} \\ Y_{t_2} \\ \vdots \\ Y_{t_n} \end{pmatrix} = \begin{pmatrix} t_1 \cdot X_{\frac{1}{t_1}} \\ \vdots \\ t_n \cdot X_{\frac{1}{t_n}} \end{pmatrix}$$

   and can be decomposed in the vector matrix product:

$$= \begin{pmatrix} t_1 & \dots & 0 \\ \dots & t_2 & \dots \\ 0 & \dots & t_n \end{pmatrix} \cdot \begin{pmatrix} X_{\frac{1}{t_1}} \\ \vdots \\ X_{\frac{1}{t_n}} \end{pmatrix}$$

   Where the last vector is a Normal vector since X by assumption (recall the definition of the standard brownian motion): then, Y_t is still Normal since it is a linear transformation of a Normal vector.

2. Mean = 0:

$$E(Y_t) = E(t \cdot X_{\frac{1}{t}}) = t \cdot E(X_{\frac{1}{t}}) = 0$$

   because $E(X_{\frac{1}{t}})$ is equal to 0 since it is a standard brownian motion.

3. Covariance:

$$E(Y_s \cdot Y_t) = E(s \cdot t \cdot X_{\frac{1}{s}} \cdot X_{\frac{1}{t}}) = s \cdot t \cdot E(X_{\frac{1}{s}} X_{\frac{1}{t}}) = s \cdot t \cdot (\tfrac{1}{s} \wedge \tfrac{1}{t})$$

$$= \tfrac{s \cdot t}{\max(s,t)} = \tfrac{\min(s,t) \cdot \max(s,t)}{\max(s,t)} = s \wedge t$$

5.7 Standard Brownian Motion and Stopping Times

Let X be a standard brownian motion and $a > 0$ is a constant.

Define

$$\tau = \inf\{t \geq 0 : X_t = a\}$$

$$= \begin{cases} +\infty \text{ if } X_t \neq a \forall t \geq 0 \\ \text{first time } t \text{ such that } X_t = a \text{ otherwise} \end{cases}$$

So, τ is a stopping time and now we are interested in evaluating its distribution. Obviously, $P(\tau \leq t) = 0$ for $t \leq 0$. So, by letting $t > 0$, we can compute the distribution as follows:

$$P(\tau \leq t) = P(\tau \leq t, X_t = a) + P(\tau \leq t, X_t > a) + P(\tau \leq t, X_t < a)$$

where the first term

$$P(\tau \leq t, X_t > a) \leq P(X_t = a) = 0$$

since X_t is an absolutely continuous random variable.
Moreover, it can be shown that

$$P(\tau \leq t, X_t > a) = P(\tau \leq t, X_t < a)$$

but we take this fact as given. So trivially,

$$P(\tau \leq t) = 2P(\tau \leq t, X_t > a).$$

Now, since X can be taken to have continuous paths, and $X_0 = 0$, then

- $X_t > a \Rightarrow \exists\ s \in (0, t)$ such that

$$X_s = a \Rightarrow \tau \leq s < t$$

Hence,

$$P(\tau \leq t) = 2P(X_t > a) = 2P(\tfrac{X_t}{\sqrt{t}} > \tfrac{a}{\sqrt{t}})$$

$$= 2P(N(0,1) > \tfrac{a}{\sqrt{t}})$$

$$= 2[1 - \phi(\tfrac{a}{\sqrt{t}})]$$

where ϕ denote the c.d.f of $N(0,1)$. So far, we supposed that $a > 0$. Now, suppose that $a < 0$:

$$\tau = \inf t \geq 0 : X_t = a$$

$$= \inf\{t \geq 0 : -X_t = |a|\}$$

Since $-X$ is still a standard brownian motion, we can repeat the previous calculation with $|a|$ in the place of a, and we end up with

$$P(\tau \leq t) = 2[1 - \phi(\tfrac{|a|}{\sqrt{t}})]$$

So, in general,

$$P(\tau \leq t) = 2[1 - \phi(\tfrac{|a|}{\sqrt{t}})] \ \forall t > 0 \ \forall a \neq 0$$

As a consequence

$$P(\tau \leq +\infty) = \lim_{t \to +\infty} P(\tau < t)$$

$$= \lim_{t \to +\infty} 2[1 - \phi(\tfrac{|a|}{\sqrt{t}})]$$

$$= 2[1 - \phi(0)]$$

$$= 2[1 - \tfrac{1}{2}] = 1$$

After realizing the above fact, a natural question is how much time X needs to reach the level a?
To answer this question, we evaluate $E(\tau)$. The density of τ is:

$$f(t) = \tfrac{d}{dt} P(\tau \leq t)$$

$$= \tfrac{d}{dt}[2(1 - \phi(\tfrac{|a|}{\sqrt{t}}))]$$

$$= -2\phi'(\tfrac{|a|}{\sqrt{t}})(\tfrac{-|a|}{t^{\frac{3}{2}}}) \cdot (-\tfrac{1}{2})$$

$$= \tfrac{|a|}{t^{\frac{3}{2}}} \cdot e^{-\frac{1}{2}\frac{a^2}{t}} \cdot \tfrac{1}{\sqrt{2\pi}}$$

where the last exponential factor can be regarded as the density of a $N(0,1)$ evaluated at the point $\frac{|a|}{\sqrt{t}}$.

Hence, $E(\tau) = \int_0^{+\infty} t \cdot f(t)dt = \frac{|a|}{\sqrt{2\pi}} \cdot \int_0^{+\infty} \frac{e^{-\frac{a^2}{2t}}}{\sqrt{t}} dt \geq \frac{|a|}{\sqrt{2\pi}} \int_c^{+\infty} \frac{e^{-\frac{a^2}{2t}}}{\sqrt{t}} dt.$

Since $e^{-\frac{a^2}{2t}} \to 1$ as $t \to +\infty$, we can take c such that

$$e^{-\frac{a^2}{2t}} \geq \tfrac{1}{2} \ \forall t \geq c.$$

Hence,

$$E(\tau) \geq \tfrac{|a|}{\sqrt{2\pi}} \cdot \int_c^{+\infty} \tfrac{1}{2} \tfrac{1}{\sqrt{t}} dt = +\infty$$

In short, $\tau < +\infty$ a.s. but $E(\tau) = +\infty$.
We can interpret the result as follows: $\forall a \neq 0$, sooner or later, the X- paths attains the level a, but the time τ needed to reach a is very big.
In fact, the expectation of such time, namely $E(\tau)$, is infinite.
In a sense, using the language of the Markov chain, a is a "null recurrent state".

Theorem 5.7.1. Next, let X be a standard brownian motion. Fix $t > 0$ and for each n, partition the interval from $[0, t]$ in k_n sub-intervals: so that, $0 = t_0^n < ... < t_{k_n}^n = t$. Then, if $\max_s(t_j^n - t_{j-1}^n) \to 0$

$$\sum_{j=1}^{k_n}(X_{t_j^n} - X_{t_{j-1}^n})^2 \xrightarrow{L_2} t$$

In other terms, the sum of the squares of the increments goes to the length t of the interval.
For instance, one can take

$$t_j^n = \frac{j}{n}t \ \forall n \geq 1 \ \forall j = 1, ..., n$$

Then,

$$\sum_{j=1}^{k_n}(X_{t_{\frac{j}{n}}} - X_{t_{\frac{j-1}{n}}})^2 \xrightarrow{L_2} t$$

In order to prove this theorem recall that in general

$$Y_n \xrightarrow{L_2} b \Rightarrow E\{(Y_n - b)^2\} \to 0$$

Since

$$E\{(Y_n - b)^2\} = E\{(Y_n - E(Y_n)) + (E(Y_n) - b)^2\}$$
$$= Var(Y_n) + E[(Y_n) - b]^2 + 2(E(Y_n - b)) \cdot E(Y_n - E(Y_n))$$

And since the last term is 0, one obtain that

$$Y_n \xrightarrow{L_2} b \Leftrightarrow E(Y_n) \to b$$

and, moreover,

$$Var(Y_n) \to 0$$

In other terms, Y_n converges to a constant b in the L_2 sense if and only if $E(Y_n) \to b$ and $Var(Y_n) \to 0$.
Hence, letting $b = t$ and $Y_n = \sum_{j=1}^{k_n}(X_{t_j^n} - X_{t_{j-1}^n})^2$ we have to show that

$$E(Y_n) \to t \text{ and } Var(Y_n) \to 0.$$

In fact,

$$E(Y_n) = \sum_{j=1}^{k_n} E\{(X_{t_j^n} - X_{t_{j-1}^n})^2\}$$
$$= \sum_{j=1}^{k_n}(t_j^n - t_{j-1}^n)$$
$$= (t_1^n - t_0^n) + (t_2^n - t_1^n) + ... + (t_{k_n}^n - t_{k_{n-1}}^n)$$

It remains only to show that $Var(Y_n) \to 0$:
in fact

$$Var(Y_n) = Var(\sum_j (X_{t_j{}^n} - X_{t_{j-1}{}^n})^2)$$

and since they have independent increments

$$= \sum_s Var((X_{t_j{}^n} - X_{t_{j-1}{}^n})^2))$$

but we can also state that

$$\leq \sum_j E\{(X_{t_j^n} - X_{t_{j-1}^n})^4\}$$

which is equal to

$$= \sum_j (t_j)^2$$

Chapter 6

Poisson process

Definition 6.0.1. Let $\mathfrak{F}_t$ be a filtration, then, a process $N = \{N_t : t \geq 0\}$ indexed by the non-negatives $[0; +\infty)$ is said to be a Poisson process if the following properties are meeted:

- N is adapted to $\mathfrak{F}_t$, namely N_t is $\mathfrak{F}_t$-measurable $\forall t \geq 0$
- N has independent increments
- $N_0 = 0$ a.s. and $N_t \sim Poisson(\lambda t)$ where $\lambda > 0$ is a parameter.

Hence, in a sense, the definition of a poisson process is very similar to that of a brownian motion. Moreover, as in case of a brownian motion, this definition implies that the process N has stationary increments.
In fact, arguing as in case of brownian motion, for $0 \leq s < t$, we write $N_t = N_s + N_t - N_s$ and we know that N_s is independent of $N_t - N_s$.
Hence,

$$\phi_{N_t} = \phi_{N_s} \cdot \phi_{N_t - N_s}.$$

It follows that

$$\phi_{N_t - N_s}(u) = \frac{\phi_{N_t}(u)}{\phi_{N_s}(u)}$$

$$= \frac{e^{\lambda t(e^{iu}-1)}}{e^{\lambda s(e^{iu}-1)}}$$

Therefore, $N_t - N_s \sim Poisson(\lambda(t-s))$, namely, $N_t - N_s \sim N_{t-s} - N_0$.

6.1 The paths of a Poisson process

For what concerns the paths of N, we can say that such paths satisfies the following properties:

- increasing
- right continuous

- $N_0 = 0$
- piece-wise constant
- unitary jumps

Finally, if the filtration $\mathfrak{F}_t$ is $\mathfrak{F}_t = \sigma(N_s : s \leq t)$ (such filtration is the least σ-field which makes N adapted), the following characterization is available:

Theorem 6.1.1. Let X be a process with $X_0 = 0$ a.s. Then, X is a poisson process if and only if it has stationary, independent increments and X path's satisfy the properties up to equivalence.

Thus, if $X_0 = 0$ a.s. and X has stationary and independent increments, then X is a brownian motion if its paths and X.... to be completed

6.2 A counting process

In order to recall the properties listed before, it is convenient to recall that a Poisson process is actually a counting process.
Let us discuss more deeply:

- First define the following stopping times:

$$\tau_0 = 0,\ \tau_1 = \inf\{t \geq 0 : N_t = 1\},\ \tau_2 = \inf\{t \geq 0 : N_t = 2\},\ ...,$$
$$\tau_n = \inf\{t \geq 0 : N_t = n\}$$

So, τ_n is the first time such that $N_t = n$.
It can be shown that the sequence

$$(\tau_n - \tau_{n-1} : n = 1, 2, ...) \text{ is i.i.d and } \tau_1 - \tau_0 = \tau_1 \sim exp(\lambda)$$

Let us prove that $\tau_1 \sim exp(\lambda)$.
In fact, if $t > 0$,

$$P(\tau_1 > t) = P(N_t = 0) = P[N(\lambda \cdot t) = 0] = \exp^{-\lambda t}$$

This is equivalent to say that

$$\tau_1 \sim exp(\lambda).$$

In the previous calculation we said that $P(\tau_1 > t) = P(N_t = 0)$ since $\{\tau_1 > t\} = \{N_t = 0\}$.
If $N_t \neq 0$, then (since the jumps of N are unitary) it must be $\tau_1 \leq t$! Remember that τ_1 is the first time t such that $N_t \neq 0$.
Hence, $\tau_1 = \tau_1 - 0 = \tau_1 - \tau_0 \sim exp(\lambda)$.
Moreover, it can be shown that

- $\tau_n - \tau_{n-1} \sim \tau_1\ \forall n$;
- $\{(\tau_n - \tau_{n-1}) | n > 1\}$ are independent.

Hence,

$$N_t = \begin{cases} 0 \text{ if } t \in [0, \tau_1) \\ 1 \text{ if } t \in [\tau_1, \tau_2) \\ 2 \text{ if } t \in [\tau_2, \tau_3) \end{cases} \quad \Leftrightarrow N_t = j \Leftrightarrow t \in [\tau_j; \tau_{j+1}) \forall j = 0, 1, 2, ...$$

N_t then can be interpreted as follows:

$$N_t = \text{n}^\circ \text{ of arrivals in } [0, t]$$

so, if it wasn't clear, in this sense the process can be regarded as counting process. For instance, let say that we are counting the number of visits to a web-site: in this case, an arrival is regarded as a visit to a web-site and the random variable N_t is the number of visits.
If we have in mind this interpretation of N, the properties stated before become more clear and easy to remember.

Example 6.2.1. Given $0 \leq s < t$, $P(0 \text{ arrivals in } [0, s] | j \text{ arrivals in } [0, t])$

$$= P(N_s = 0 | N_t = j) = \frac{P(N_s=0, N_j=j)}{P(N_t=j)}$$

$$= \frac{P(N_s=0, N_t - N_s=j)}{P(N_t=j)}$$

and since independent increments

$$= \frac{P(N_s=0)P(N_t - N_s=j)}{P(N_t=j)}$$

$$= \frac{\exp^{-\lambda s} \cdot exp^{-\lambda(t-s)} \cdot \frac{[\lambda(t-s)]^j}{j!}}{\exp^{-\lambda t} \frac{(\lambda t)^j}{j!}}$$

can be simplified in

$$= (\frac{(t-s)}{t})^j = (1 - \frac{1}{t})^j$$

And finally, let's make the last two remarks:

1. Contrary to the brownian motion, the poisson process N is not a martingale. However, the transformed process $\tilde{N}_t = N_t - \lambda t$ is a martingale: in fact, if $0 \leq s < t$,

$$E(\tilde{N}_t | \mathfrak{F}_s) = E(N_t - \lambda t | \mathfrak{F}_s)$$

$$= -\lambda t + E(N_s + N_t - N_s | \mathfrak{F}_s)$$

$$= -\lambda t + N_s + E(N_t - N_s | \mathfrak{F}_s)$$

$$= -\lambda t + N_s + E(N_t - N_s)$$

The transformed process $\tilde{N}$ is usually called compensated poisson process.

2. A very useful version of the poisson process in application is the so call spatial poisson process. To define this process, let's fix a bounded subset $H \subset R^n$. Then, given a class T of Borel subsets of H, a spatial poisson process is a process

$$N = \{N(A) : A \in T\}$$

such that

- $N(A_1)...N(A_K)$ are independent random variables if $A_1, ..., A_k$ are pairwise disjoint.
- $N(A) \sim Poisson(\lambda \cdot m(A))$ where m is the Lebsgue measure on R^n

To fix ideas, suppose $n = 2$: in this case, H is merely a bounded subset of the plane. Moreover, T is a collection of Borel subset of H, say

$$T = \{A_1, A_2, A_3, A_4\}$$

Example 6.2.2. Let's take again $n = 2$ where $H =$ Italy and $T = \{$the italian regions$\}$ $= \{A_1, ..., A_{20}\}$ and $N(A)$ is the number of cases of a certain disease in A $\forall A \in T$ In this case, however, a spatial Poisson process could be unsuitable. In fact, the first proposition (i.e. independent increments) looks not very realistic, for instance, if $A =$ Emilia-Romagna and $B =$ Tuscany, then $A \bigcap B = \varnothing$ but it is not reasonable to assume $N(A)$ is independent of $N(B)$.

6.3 Levy process

Next, a general class of process which includes the brownian motion and the poisson process as a special cases is the class of Levy process.

Definition 6.3.1. A process $X = (X_t : t \geq 0)$, indexed by $[0; +\infty)$ is a Levy process with respect to a filtration $(\mathfrak{F}_t)$ if:

- X is adapted to $\mathfrak{F}_t$
- X has stationary and independent increments
- $X_0 = 0$
- $X_s \xrightarrow{p} X_t$ if $s \to t$

As already noted the brownian motion is a Levy process. Similarly, the poisson process is a Levy process. In fact, given $\epsilon > 0$, due to Čebyšëv inequality

$$P(|X_s - X_t| > \epsilon) \leq \tfrac{1}{\epsilon} E(|X_s - X_t|)$$

and if $t \geq s$

$$= \tfrac{E(X_t - X_s)}{\epsilon} = \tfrac{\lambda(t-s)}{\epsilon} \text{ as } s \to t.$$

Example 6.3.1. Another example of the Levy process is the compounded poisson process which is defined as follows:

$$X_t = \mathbb{1}_{N_t>0} \cdot \sum_{j=1}^{N_t} Z_j$$

where N is a poisson process, (Z_j) is an *iid* sequence of real random variables and N is independent of (Z_j).
Such a process has the same jump times of the poisson process N, but now the jumps are not unitary.
In fact, the jumps are given by the random variables Z_j. Obviously, if $Z_j = 1\ \forall j$, the compunded Poisson process reduces to the arrival Poisson process.

6.4 Exchangeable sequence

Definition 6.4.1. Let $(X_n : n \geq 1)$ a sequence of real random variables.
Then, (X_n) is exchangeable if

$$(X_{\pi_1}, X_{\pi_2}, ..., X_{\pi_n}) \sim (X_1, X_2, ..., X_n)\ \forall n \geq 1 \text{ and } \forall \text{ permutation } (\pi_1, ..., \pi_n) \text{ of } (1, ..., n)$$

In other terms, for each $n \geq 1$, the distribution of the vector $(X_1, ..., X_n)$ is invariant under permutation. So, for instance $(X_1, X_2) \sim (X_2, X_1)$.
An *iid* sequence is a special case of exchangeable sequence: in fact, if (X_n) is *iid* and $(\pi_1, ..., \pi_n)$ is a permutation of $(1, ..., n)$. Then

$$P(X_{\pi_1} \in A_1, ..., X_{\pi_n} \in A_n) =$$

$$= P(X_{\pi_1} \in A_1) \cdot \dots \cdot P(X_{\pi_n} \in A_n)$$

$$= P(X_1 \in A_1) \cdot P(X_2 \in A_2) \cdot ... \cdot P(X_n \in A_n)$$

$$= P(X_1 \in A_1, ..., X_n \in A_n)$$

Namely, $(X_{\pi_1} ... X_{\pi_n}) \sim (X_1, ..., X_n)$ if the sequence is *iid*.
However an exchangeable sequence need not to be *iid*.
Let's define in fact the following sequence:

$$X_1 = Z,\ X_2 = Z,\ X_3 = Z,\ ...$$

where Z is any non degenerate random variable. Then (X_n) is obviously not independent.
For instance, since

$$X_1 = Z \text{ and } X_2 = Z,\ X_1 \text{ is not independent of } X_2.$$

However, this sequence is obviously exchangeable.
In fact, since all random variable are equal to Z, nothing happens if we make a permutation.
In fact

$$(X_{\pi_1}, X_{\pi_2}, ..., X_{\pi_n}) = (Z, Z, Z, ..., Z)$$

$$= (X_1, X_2, ..., X_n).$$

Thus we can say that

$$iid \Rightarrow \text{exchangeable}.$$

but

$$iid \nLeftarrow \text{exchangeable}.$$

Similarly,

$$iid \Rightarrow \text{exchangeable} \Rightarrow \text{Stationary}$$

but

$$iid \nLeftarrow \text{exchangeable} \nLeftarrow \text{Stationary}$$

Suppose in fact (X_n) is exchangeable, then

$$(X_2, ..., X_n, X_{n+1}, X_1) \sim (X_1, X_2, ..., X_n, X_{n+1})$$

In particular this implies

$$(X_2, ..., X_n, X_{n+1}) \sim (X_1, X_2, ..., X_n) \ \forall n$$

which in term implies

$$(X_2, X_3, X_4, ...) \sim (X_1, X_2, X_3, ...)$$

Hence, as obtained above,

$$\text{exchangeable} \Rightarrow \text{stationary}$$

It remains to make an example where (X_n) is stationary but not exchangeable.

Example 6.4.1. Let $X_1 = Z, X_2 = -Z, X_3 = Z, X_4 = -Z$ and so on, where Z is any random variable such that $P(Z = 0) < 1$ and $Z \sim -Z$.
Such a sequence is not exchangeable: for instance,
$P(X_1 = X_3) = 1$ since by definition $X_1 = X_3 = Z$.
$P(X_1 = X_2) = P(Z = 0) < 1$ since $X_1 = Z$ and $X_2 = -Z$ Hence, $(X_1, X_3) \nsim (X_1, X_2)$ namely, (X_n) is not exchangeable.
However (X_n) is stationary

$$(X_1, X_2, X_3, ...) = (Z, -Z, Z, -Z, ...)$$

$$(X_2, X_3, ...) = (-Z, Z, -Z, ...)$$

Hence, both sequence can be regarded as follows: at first, one select the value of the first random variable and then the sequence goes on in a deterministic way.
Namely, after choosing X_1, we have that

$$X_2 = -X_1, X_3 = X_1, X_4 = -X_1, ...$$

Similarly, after choosing X_2, we have

$$X_3 = -X_2, X_4 = X_2, X_5 = -X_2, ...$$

Hence, we have that

$$(X_2, X_3, ...) \sim (X_1, X_2, ...)$$

$\Updownarrow$
$X_2 \sim X_1$ and $X_2 \sim X_1$ is true since $X_1 = Z, X_2 = -Z$ and Z is such that $Z \sim -Z$

6.4.1 de Finetti's theorem

The most important result concerning exchangeable sequence is the de Finetti's theorem.

Definition 6.4.2. Let X_n be a sequence of real random variables.
Then X_n is exchangeable $\Updownarrow$ $P(X_1 \in A_1, ..., X_n \in A_n) = E_Q\{Q(A_1), ..., Q(A_n)\}$ $\forall n \geq 1$ $\forall A_1, ..., A_n$.

The notation $E_Q\{Q(A_1), ..., Q(A_n)\}$ should be meant as follows: first, we choose a probability measure Q on $\mathfrak{B}(\mathbb{R})$ and we compute $Q(A_1), ..., Q(A_n)$, and then we take the expectation with respect to Q.
In other terms, the underlying idea is that if (X_n) is exchangeable, in order to evaluate $P(X_1 \in A_1, ..., X_n \in A_n)$ we first evaluate such a probability where the assumption (X_n) is *iid* with common distribution Q, and subsequentially we take the expectation with respect to Q.
In order to evaluate

$$P(X_1 \in A_1, ..., X_n \in A_n)$$

we first assume that (X_n) is *iid* with common distribution Q.
In this case, $P(X_1 \in A_1, ..., X_n \in A_n) = P(X_1 \in A_1) \cdot ... \cdot P(X_n \in A_n)$
$= Q(A_1) \cdot ... \cdot Q(A_n)$. Subsequentially, we take the expectation with respect to Q and we finally obtain

$$P(X_1 \in A_1, ..., X_n \in A_n) = E_Q\{Q(A_1)...Q(A_n)\}.$$

So, de Finetti's theorem states that any exchangeable sequence can be regarded in this way.

Next, if (X_n) is exchangeable and $E({X_1}^2) < +\infty$, then

$$Cov(X_i, X_j) = Cov(X_1, X_2) \geq 0$$

In order to prove this fact, not that

$$0 \leq Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i) + \sum_{i=1}^n Cov(X_i, X_j)$$

Since X_n is exchangeable, $Var(X_i) = Var(X_1)$ $\forall i$ and

$$Cov(X_i, X_j) = Cov(X_1, X_2) \ \forall i \neq j$$

Hence, we get

$$0 \leq n \cdot Var(X_1) + n(n-1)Cov(X_1, X_2)$$

and by re-arranging the terms

$$Cov(X_1, X_2) \geq \frac{-Var(X_1)}{n-1} \ \forall n$$

which leads to say that

$$Cov(X_1, X_2) \geq 0.$$

A nice consequence of de Finetti's theorem is the strong law of large numbers.
If the (X_n) is exchangeable and $E(|X_1| < +\infty)$, then

$$\bar{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} V$$

where V is some real random variable.
Note that V is not necessary degenerate, let's prove it.
$P(\bar{X_n}$ converges as $n \to +\infty) = E_Q\{P_Q(\bar{X_n}$ converges as$n \to +\infty)\}$ where P_Q denotes the probability under which X_n is iid with common distribution Q.
In fact, $P_Q(\bar{X_n}$ converges as $n \to +\infty) = 1$ by the classical strong law of large numbers, since, under P_Q, the sequence (X_n) is iid and $E_Q(|X_n|) < +\infty$.

Example 6.4.2. Polya sequence:
At time 0, the urn contains $a > 0$ white balls and $b > 0$ black balls.
At time 1, we select a ball and then replace it together with other k balls of the same colour of the extracted ball and so on.
Let $X_n = \mathbb{1}($ white ball at time n $)$
Then, $P(X_{n+1} = 1|X_1, ..., X_n) = \frac{a+k\cdot\sum_{i=1}^{n} X_i}{a+b+n\cdot k}$ In fact, $\sum_{i=1}^{n} X_i = \sum_{i=1}^{n}$ white ball at time i = number of white balls extracted in the first n trials.
Hence,

$$P(X_1 = x_1, ..., X_n = x_n) = P(X_n = x_n|X_1 = x_1, ..., X_{n-1} = x_{n-1}) \cdot P(X_1 = x_1, ..., X_{n-1} = x_{n-1})$$

$$= ... = P(X_n = x_n|X_1 = x_1, ..., X_{n-1} = x_{n-1}) \cdot P(X_{n-1} = x_{n-1}|X_1 = x_1, ..., X_n = x_n) \cdot P(X_2 = x_2|X_1 = x_1) \cdot P(X_1 = x_1)$$

and after replacing these conditional probabilities, we find that

$$P(X_1 = x, ..., X_n = x_n)$$

depends only on n and $\sum_{i=1}^{n} X_i =$ number of white balls in the first n trial Hence, for each permutation $(\pi_1, ..., \pi_n)$ of (1, ..., n), we obtain

$$P(X_{\pi_1} = x_1, X_{\pi_2} = x_2, ..., X_{\pi_n} = x_n)$$

$$= P(X_1 = x_1, ..., X_n = x_n), \text{ namely, the sequence } (X_n) \text{ is exchangeable.}$$