Discrete-time Stochastic Processes Module 2

MSc Financial Engineering

Table of Contents

1. Brief 2

2. Course Context 2

2.1 Course-level Learning Outcomes 3

2.2 Module Breakdown 4

3. Module 2: Stochastic Processes 5

3.1 Module-level Learning Outcomes 5

3.2 Transcripts and Notes 6

3.2.1 Notes: What is a Stochastic Process? 6

3.2.2 Transcript: What is a Stochastic Process? 9

3.2.3 Notes: Filtrations 12

3.2.4 Transcript: Filtrations 14

3.2.5 Notes: Stopping Times and Stopped Processes 16

3.2.6 Transcript: Exploring Stopping Times and Stopped Processes 18

3.2.7 Notes: Conditional Expectation 21

3.2.8 Notes: Martingales and Markov Processes 25

3.2.9 Transcript: An Example of Conditional Expectation 29

3.2.10 Notes: Problem Set 32

3.3 Additional Resources 37

3.4 Collaborative Review Task 38

1. Brief

This document contains the core content for Module 2 of Discrete-time Stochastic Processes, entitled Stochastic Processes. It consists of four lecture transcripts, five sets of notes, and a problem set.

2. Course Context

Discrete-time Stochastic Processes is the third course presented in the WorldQuant University (WQU) Master of Science in Financial Engineering (MScFE) program. The course introduces derivative pricing in discrete time. It begins with measure-theoretic probability and stochastic processes, with an emphasis on discrete-time martingales. These ideas are then applied to the pricing of derivatives in discrete time, followed by an introduction to interest rate and credit risk modeling.

2.1 Course-level Learning Outcomes

Upon completion of the Discrete-time Stochastic Processes course, you will be able to:

1. Understand the language of measure-theoretic probability.
2. Understand stochastic processes and their applications.
3. Understand the theory of discrete-time martingales.
4. Define trading strategies in discrete time.
5. Create replicating portfolios in discrete time.
6. Model stock price movements on a binomial tree.
7. Price and hedge European derivatives in discrete time.
8. Price and hedge exotic European derivatives in discrete time.
9. Price and hedge American derivatives on a binomial tree.
10. Construct a simple interest rate model on a tree.
11. Price interest rate derivatives on a tree.

2.2 Module Breakdown

The Discrete-time Stochastic Processes course consists of the following one-week modules:

1 Probability Theory
2 Stochastic Processes
3 Discrete Martingales
4 Trading in Discrete Time
5 The Binomial Model
6 American Options
7 An Introduction to Interest Rate Models

3. Module 2:

Stochastic Processes

Welcome to the second module of the Discrete-time Stochastic Processes course. This module covers the basics of stochastic processes from a probability theory perspective. Just like the previous module, the content of this module is very abstract but it also contains some concrete applications. Again, these notes can be thought of as a summary of important results in stochastic processes that are useful in mathematical finance.

3.1 Module-level Learning Outcomes

After completing this module, you will be able to:

1. Define stochastic processes rigorously.
2. Show whether a given stochastic process is adapted to a given filtration.
3. Define a stopping time and a stopped process.
4. Define and calculate the conditional expectation of a random variable.
5. Define a martingale and a Markov process.

3.2 Transcripts and Notes

3.2.1 Notes: What is a Stochastic Process?

A random variable X is a real-valued function of the outcome of a random experiment. Each outcome $\omega \in \Omega$ leads to exactly one value $X(\omega) \in \mathbb{R}$. We now want to take this a step further and consider the case when each outcome of the random experiment determines the values of a collection of random variables $\{X_t : t \in \mathbb{I}\}$. We will call such a collection of random variables a *stochastic process* and the set $\mathbb{I}$ the *index set*. We only consider $\mathbb{I} \subseteq [0, \infty)$ and think of each $t \in \mathbb{I}$ as a time point.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\emptyset \neq \mathbb{I} \subseteq [0, \infty)$ be an index set. A *stochastic process* $X = \{X_t : t \in \mathbb{I}\}$ is a collection of random variables. We call $\mathbb{I}$ the *index set*. If $\mathbb{I} \subseteq \mathbb{N}$, we call X a *discrete-time stochastic process*, while if $\mathbb{I} = [0, \infty)$ or $\mathbb{I} = [0, T]$ for some $T > 0$, we call X a *continuous-time stochastic process*. This course deals mainly with discrete-time stochastic processes.

Here are some real-world examples.

1 Let X_t be the value of the S&P 500 index at time t. Here, the index set could be taken to be continuous or discrete, if the prices are observed at discrete time intervals – e.g. daily. Each price X_t is a random variable, and one outcome ω determines the values, or prices, of X_t for all $t \in \mathbb{I}$.

2 In gambling, if X_n is your wealth (total earnings) after you have played the n^{th} game, then

$$X = \{X_n : n = 0,1,2,3, \dots\},$$

is a discrete-time stochastic process.

The important thing to note here is that one outcome $\omega \in \Omega$ must determine the values of all the random variables $\{X_t : t \in \mathbb{I}\}$. For a fixed outcome $\omega \in \Omega$, we call the function $t \mapsto X_t(\omega)$ a *sample path* of X. Thus, a stochastic process is a "sample path-valued" random variable. That is, we can think of a stochastic process as a function $X : \Omega \to \{\text{sample paths}\}$, that outputs a path (a function from $\mathbb{I}$ to $\mathbb{R}$) for each outcome, rather than one real number (as is the case for random variables).

Here's an example of the so-called *Renewal Process*. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and T_1, T_2, T_3, be a sequence of positive independent and identically distributed random variables in $\mathcal{L}^1$. Think of T_n as the time between the occurrence of the $(n-1)^{th}$ and n^{th} events (inter-arrival time). Define the arrival times $W_1, W_2, W_3, \ldots$ as

$$W_n \coloneqq \sum_{k=1}^{n} T_k, \quad n \geq 1.$$

The stochastic process $X = \{X_t : t \geq 0\}$, defined by

$$X_t \coloneqq \text{Number of } W_n\text{'s less than or equal to } t = \sum_{n=1}^{\infty} I_{\{W_n \leq t\}}, \ t \geq 0,$$

is called a *renewal process*. A Poisson process is a special case when each T_n has an exponential distribution with common parameter $\lambda > 0$.

Consider a gambling game where you start with an initial wealth of 5. A coin is tossed twice, and, for each toss, you either win 1 if the outcome is a head or lose 1 otherwise. Pick

$$\Omega = \{HH, HT, TH, TT\},\ \mathcal{F} = 2^{\Omega}\ \textit{and}\ \mathbb{P} = \frac{1}{4}\sum_{\omega\in\Omega} \delta_{\omega}.$$

Define $X_0(\omega) = 5$ for all $\omega \in \Omega$ and

$$X_1(\mathrm{HH}) = X_1(\mathrm{HT}) = 5 + 1 = 6,\ X_1(\mathrm{TH}) = X_1(\mathrm{TT}) = 5 - 1 = 4,$$
$$X_2(\mathrm{HH}) = 5 + 1 + 1 = 7,\ X_2(\mathrm{HT}) = 5 + 1 - 1 = 5,$$
$$X_2(\mathrm{TH}) = 5 - 1 + 1 = 5,\ X_2(\mathrm{TT}) = 5 - 1 - 1 = 3.$$

Then $X = \{X_0, X_1, X_2\}$ is a stochastic process, where X_t represents your wealth at times $t = 0,1,2$. The sample path corresponding to $\omega = \mathrm{HT}$, for instance, is

$$X(\mathrm{HT}) = \{X_0(\mathrm{HT}), X_1(\mathrm{HT}), X_2(\mathrm{HT})\} = \{5,6,5\}.$$

Here is a table summarizing the sample paths of X:

ω	$X_0(\omega)$	$X_1(\omega)$	$X_2(\omega)$
HH	5	6	7
HT	5	6	5
TH	5	4	5
TT	5	4	3

If the index set is $\mathbb{I} = [0, \infty)$ we say that a stochastic process X is continuous (resp. right continuous, left continuous) if each sample path is continuous (resp. right continuous, left continuous).

3.2.2 Transcript: What is a Stochastic Process?

Hi, in this video we introduce stochastic processes.

Remember that if we are given a probability space $(\Omega, \mathbb{F}, \mathbb{P})$, a function $X: \Omega \rightarrow R$ is called a random variable if X is measurable in the sense that the pre-image of every Borel set belongs to $\mathbb{F}$.

We're now going to take this a step further and consider a collection of random variables.

X is now a collection of random variables that are indexed by an index set $\mathbb{I}$. This index set can either be the positive real line, which is the set $[0, \infty]$, in which case we will call it a continuous-time stochastic process. On the other hand, it could be a discrete set, like the natural numbers which consist of $\{0,1,2 \dots\}$, and in this case we would call it a discrete-time stochastic process. A stochastic process is therefore a collection of random variables that are indexed by either this set of this form $[0, \infty]$ or the index set of the following form $\{0,1,2 \dots\}$, in which case it is discrete.

We may sometimes consider a case whereby this index set is restricted between $[0, T]$ where T is some positive real number $(T > 0)$. Let's look at some examples.

Firstly, if we let X_t be the value of the S&P500 index at time t then X_t can be thought of as a collection of random variables because at each time point we have no idea what the value of the S&P500 index will be. X_t is therefore a stochastic process. We can think of it as a continuous-time stochastic process if we observe the index at every time point between 0 and t or we can consider it as a discrete-time stochastic process if we observe it at certain points in time.

As a second example, let's consider a gambling game. If we let X_n be your wealth after the n^{th} game, then this sequence $\{X_n : n = 0,1,2 \ldots\}$ can be thought of as a stochastic process if we don't know the outcome of each game.

The important thing to note about a stochastic process is that the outcome ω must determine the values of all the random variables $X_t(\omega)$ whereas in a random variable the outcome determines the value of just one random variable. We will call this collection for each ω a "sample path", so it is a sample path of the stochastic process X. One way to think about it is that if, for instance, the index set is continuous like this, a stochastic process consists of different sample paths. Each outcome, for example ω_1, gives you a sample path. Another outcome, ω_2, gives the whole sample path and so it continues. Another way of thinking about a stochastic process is that it's a sample path valued random variable.

As an example, consider an experiment where two coins are tossed. This would be the sample space: $\{\text{HH,HT,TH,TT}\}$. You are playing a gambling game whereby your initial wealth is $\$5$. At time 1, if the outcome of the first toss is a head you get $\$1$ and if the outcome is a tail you lose $\$1$. $X_1(\omega)$ will either be $X_0(\omega) + 1$ if ω is $\{\text{HH,HT}\}$, which represents the outcome of the first toss being a head, or it will be $X_0(\omega) - 1$ if ω belongs to $\{\text{TH,TT}\}$, which represents the outcome of the first toss being a tail. X_2 represents the outcome of the second toss, which follows the same pattern. So, X_2 will be equal to $X_1(\omega) + 1$ if the outcome of the second toss is a head and ω belongs to $\{\text{HH,TH}\}$ and it will be $X_1(\omega) - 1$ if the outcome of the second toss is a tail and ω belongs to $\{\text{HT,TT}\}$. We can represent this stochastic process in a table that looks like this:

ω	$X_0(\omega)$	$X_1(\omega)$	$X_2(\omega)$
HH	5	6	7
HT	5	6	5
TH	5	4	5
TT	5	4	3

As an example of a sample path, if you pick ω to be $\{HT\}$ then the sample path corresponding to that is simply the set $\{5,6,5\}$. That's the sample path – the values of all the random variables when ω is $\{HT\}$.

Now that we've introduced stochastic processes, in the next video we are going to move on to filtrations.

3.2.3 Notes: Filtrations

Even though all the values of X are determined by one outcome, some of them can be determined without knowledge of the full outcome and are therefore known prior to others. In the example in the previous section, we do not need to know the full outcome of the experiment to determine X_1, since X_1 depends only on the outcome of the first toss. That is, X_1 is measurable with respect to a σ-algebra that is strictly smaller than $\mathcal{F}$.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathbb{I} \subseteq [0, \infty)$ be an index set. A *filtration* $\mathbb{F} = \{\mathcal{F}_t : t \in \mathbb{I}\}$ is an increasing collection of sub-σ-algebras of $\mathcal{F}$ – i.e. $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ for $s \leq t,\ s, t \in \mathbb{I}$. A stochastic process $X = \{X_t : t \in \mathbb{I}\}$ is *adapted* to a filtration $\mathbb{F}$ if for every $t \in \mathbb{I}$, X_t is $\mathcal{F}_t$-measurable.

Intuitively, a filtration contains the information about ω that is available to us at each time point. At time $t \in \mathbb{I}$, $\mathcal{F}_t$ consists of all the events that can be decided (whether or not they have occurred) by time t. The fact that $\mathbb{F}$ is increasing implies that "we do not forget". An adapted stochastic process is one whose values are known at the corresponding time point. We will call $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ a *filtered probability space.*

Associated with a stochastic process $X = \{X_t : t \in \mathbb{I}\}$ is a filtration $\mathbb{F}^X = \{\mathcal{F}_t^X : t \in \mathbb{I}\}$, called the *natural filtration* of X, defined by

$$\mathcal{F}_t^X \coloneqq \sigma(\{X_s : s \leq t, s \in \mathbb{I}, \}), \quad t \in \mathbb{I}.$$

This is the smallest filtration with respect to which X is adapted.

We continue with the example of two coin-tosses. Define $\mathbb{F} = \{\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2\}$ by

$$\mathcal{F}_0 = \{\emptyset, \Omega\}, \mathcal{F}_1 = \sigma(\{\{\mathrm{HH}, \mathrm{HT}\}, \{\mathrm{TH}, \mathrm{TT}\}\}), \mathcal{F}_2 = \mathcal{F} = 2^\Omega.$$

Then $\mathbb{F}$ is a filtration. The σ-algebra $\mathcal{F}_0$ contains no information, $\mathcal{F}_1$ contains information about the outcome of the first toss, and $\mathcal{F}_2$ contains information on both tosses. The natural filtration $\mathbb{F}^X$ also coincides with $\mathbb{F}$ in this case. Indeed,

$$\mathcal{F}_0^X = \sigma(X_0) = \{\emptyset, \Omega\},$$

$$\mathcal{F}_1^X = \sigma(\{X_0, X_1\}) = \sigma(X_1) = \sigma\left(\left\{X_1^{-1}(\{4\}, X_1^{-1}(\{6\}))\right\}\right) = \sigma(\{\{\text{HH,HT}\}, \{\text{TH,TT}\}\}), \text{ and}$$

$$\begin{aligned}\mathcal{F}_2^X &= \sigma(\{X_0, X_1, X_2\}) = \sigma(\{X_1^{-1}(4), X_1^{-1}(6), X_2^{-1}(3), X_2^{-1}(5), X_2^{-1}(7)\}) \\ &= \sigma(\{\{\text{TH,TT}\}, \{\text{HH,HT}\}, \{\text{TT}\}, \{\text{HT,TH}\}, \{\text{HH}\}\}) \\ &= \sigma(\{\{\text{HH}\}, \{\text{TH}\}, \{\text{TH}\}, \{\text{TT}\}\}) = 2^{\Omega}.\end{aligned}$$

This can be summarized in the following table:

Blocks of $\mathcal{F}_0$	Blocks of $\mathcal{F}_1$	Blocks of $\mathcal{F}_2$
$\{\{$HH, HT, TH, HH$\}\}$	$\{\{$HH, HT$\}$,$\{$TH, TT$\}\}$	$\{\{$HH$\}$,$\{$HT$\}$,$\{$TH$\}$,$\{$TT$\}\}$

3.2.4 Transcript: Filtrations

Hi, in this video we introduce filtrations.

So, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and consider X, which is a stochastic process – meaning that it's an index collection of random variables – even though the values of X are all determined by one outcome Ω. If you recall, the sample paths of X look like this: one outcome determines all the values of X. Some of them are known prior to others. For instance, if we take X_t to be the value of the S&P500 at time t, then the value of X_5, for instance, is known prior to the value of X_{10}. What we need therefore, is to have a way of collecting the information that is available at each point in time. To do so, we are going to introduce the notion of a *filtration*.

A filtration (or $\mathbb{F}$) is an increasing collection of σ-algebras. $\mathbb{F}$ can be written like this: $\mathbb{F} = \{F_t : t \in \mathbb{I}\}$. The idea behind this is that F_t represents the information we have that is available at time t and we call $\{\Omega, F, \mathbb{F}, \mathbb{P}\}$ a *filtered probability space*.

Now consider a filtered probability space $(\Omega, F, \mathbb{F}, \mathbb{P})$ and let X be a stochastic process that is indexed by $\mathbb{I}$. We say that X is adapted to the filtration $\mathbb{F}$ if X_t is F_t-measurable for each t. Intuitively, what that means is that the value of X_t is known at each time point if the information available to us is given by the filtration $\mathbb{F}$.

We define the natural filtration of X as the smallest filtration that makes X adapted. It is defined as follows: $\mathbb{F}^X$ is a collection of sub-σ-algebras that are defined as F_t^X which is the σ-algebra generated by X_s where s is less than or equal to t. It is the smallest σ-algebra that makes all of X_s measurable for s less than or equal to t. We call that the natural of filtration of X.

Let us look at an example:

ω	$X_0(\omega)$	$X_1(\omega)$	$X_2(\omega)$
HH	5	6	7
HT	5	6	5
TH	5	4	5
TT	5	4	3

Consider this experiment where we have HH, HT, TH, and TT – we are tossing two coins and we define the following stochastic process $X_0(\omega)$ as always equal to 5, $X_1(\omega)$ as $6, 6, 4, 4$ and $X_2(\omega)$ as $7, 5, 5, 3$. This is a stochastic process in discrete time and the index set is $\{0,1,2\}$. We want to find a natural filtration of this stochastic process.

F_0^X is a σ-algebra generated by $(\{X_0\})$ only and, since X_0 is constant, it does not distinguish between any of these elements. This will equate to the *trivial σ-algebra*.

F_1^X is a σ-algebra generated by $(\{X_0, X_1\})$ and, in this case, X_1 does not distinguish between HH and HT and it also does not distinguish between TH and TT. X_0, however, does not distinguish between any of the elements at all. Simply put, this σ-algebra generates HH and HT as one block and TH and TT as another block.

Finally, we have F_X^2 which is the σ-algebra generated by $(\{X_0, X_1, X_2\}$. While X_1 puts, for example, HH and HT together as a block, X_2 separates them, meaning that HH and HT would now be in separate blocks. The same goes for TH and TT. As a result, this σ-algebra is equal to the power set of Ω which is a set of all possible subsets.

The natural filtration of X, therefore, consists of these three σ-algebras which are all sub-σ-algebras of F, which is the powerset of Ω.

3.2.5 Notes: Stopping Times and Stopped Processes

Suppose you invest your money in an asset whose price is a stochastic process $X = \{X_t : t \in \mathbb{I}\}$. Let τ be a random variable that denotes the time when you decide to sell the asset (with $\tau = \infty$ if you never sell). So τ is a *random time*. If $\mathbb{F} = \{\mathcal{F}_t : t \in \mathbb{I}\}$ represents the information currently available at each time point, then for every $t \in \mathbb{I}$, the decision of whether or not to sell should only depend on the information available at time t, $\mathcal{F}_t$. That is, the event that you have stopped by time t, $\{\tau \leq t\}$, should belong to $\mathcal{F}_t$.

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space. A non-negative extended random variable $\tau : \Omega \to \mathbb{I} \cup \{\infty\} =: \bar{\mathbb{I}}$ is called a *stopping time* if $\{\tau \leq t\} \in \mathcal{F}_t$ for every $t \in \mathbb{I}$. If τ is a stopping time and $X = \{X_t : t \in \mathbb{I}\}$ is a stochastic process, we define the *stopped process* $X^\tau = \{X_t^\tau : t \in \mathbb{I}\}$ as $X_t^\tau := X_{\tau \wedge t}$ – i.e.,

$$X_t^\tau(\omega) := \begin{cases} X_t(\omega) & t < \tau(\omega) \\ X_{\tau(\omega)}(\omega) & t \geq \tau(\omega) \end{cases} \quad \omega \in \Omega.$$

We also define the random variable X_τ by

$$X_\tau(\omega) := \begin{cases} X_{\tau(\omega)}(\omega) & \tau(\omega) < \infty \\ 0 & \tau(\omega) = \infty. \end{cases}$$

In the coin toss example, say we decide to stop as soon as our wealth reaches 6. Then $\tau(\omega) := \inf\{t \in \mathbb{I} : X_t(\omega) = 6\}$, where we agree that $inf(\emptyset) =$. Then $\tau(\mathrm{HH}) = \tau(\mathrm{HT}) = 1, \tau(\mathrm{TH}) = \tau(\mathrm{TT}) = \infty$. The stopped process X^τ represents our wealth process if we employ this stopping strategy. We have $X_0^\tau = X_{\tau \wedge 0} = X_0 = 5, X_1^\tau = X_{\tau \wedge 1} = X_1$ since $\tau \geq 1$, and X_2^τ is given by

$$X_2^\tau(\mathrm{HH}) = X_{\tau(\mathrm{HH}) \wedge 2}(\mathrm{HH}) = X_{1 \wedge 2}(\mathrm{HH}) = X_1(\mathrm{HH}) = 6.$$

A similar argument gives

$$X_2^\tau(\mathrm{HT}) = X_1(\mathrm{HT}) = 6, \quad X_2^\tau(\mathrm{TH}) = \mathrm{X}_2(\mathrm{TH}) = 5, \quad X_2^\tau(\mathrm{TT}) = X_2(\mathrm{TT}) = 3.$$

The random variable X_τ is given by

$$X_\tau(\mathrm{HH}) = X_1(\mathrm{HH}) = 6, \quad X_\tau(\mathrm{HT}) = X_1(\mathrm{HT}) = 6, \quad X_\tau(\mathrm{TH}) = 0 = X_\tau(\mathrm{TT}).$$

A table summary:

ω	$\tau(\omega)$	$X_0(\omega)$	$X_1(\omega)$	$X_2(\omega)$	$X_0^\tau(\omega)$	$X_1^\tau(\omega)$	$X_2^\tau(\omega)$	$X_\tau(\omega)$
HH	1	5	6	7	5	6	6	6
HT	1	5	6	5	5	6	6	6
TH	∞	5	4	5	5	4	5	0
TT	∞	5	4	3	5	4	3	0

In most instances, stopping times can be generated as follows. Let $B \in \mathcal{B}(\mathbb{R})$ be a Borel set and $X = \{X_t : t \in \mathbb{I}\}$ be a stochastic process. Define $\tau_B : \Omega \to \bar{\mathbb{I}}$ by

$$\tau_B(\omega) \coloneqq \inf\{t \in \mathbb{I} : X_t(\omega) \in B\}.$$

We call τ_B the *hitting time*.

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space, $\mathbb{I} \subseteq \mathbb{N}$ be a discrete index set and $X = \{X_t : t \in \mathbb{I}\}$ be a stochastic process. If $\mathrm{B} \in \mathcal{B}(\mathbb{R})$ and X is adapted to $\mathbb{F}$, then τ_B is an $\mathbb{F}$-stopping time.

If $\mathbb{I} \subseteq \mathbb{N}$ is discrete, then $\tau : \Omega \to \bar{\mathbb{I}}$ is a stopping time if and only if $\{\tau = n\} \in \mathcal{F}_n$ for each $n \in \mathbb{I}$.

3.2.6 Transcript: Exploring Stopping Times and Stopped Processes

Hi, in this video we're going to introduce *stopping times* and *stopped processes*. We are going to first define them and then look at an example.

Suppose you have invested in a share whose price is given by the stochastic process $X = \{X_t : t \in \mathbb{I}\}$. Let τ be the random time you decide to sell the share. So τ is the time that you sell the share X. What do I mean by random time? If you look at a sample path of the stochastic process $X(\omega)$ corresponding to the sample path ω, τ might be somewhere here. $\tau(\omega)$ represents the time you decide to sell the share.

As an example, you could define τ to be the first time that the share price reaches 150. This is random because it depends on the sample path under consideration. We're going to call τ a stopping time if it satisfies the following condition: $\{\tau \leq \ t\} \in \mathcal{F}_t$. This set here ($\{\omega \in \Omega : \tau(\omega) \leq t\}$) is the set of all outcomes for which $\tau \leq t$. Intuitively, what this means is that the decision of whether or not to sell by time t is known by time t because it belongs to the σ-algebra at time t, which is the filtration that contains the information available to us.

Similarly, we define a stopped process as simply the original stochastic process but then, once we stop, it becomes constant and that constant is equal to whatever the value is at the stopping time. This will be denoted by X^τ with the superscript, and it will be the collection of random variables $X^\tau = \{\mathrm{X}^\tau{}_t : t \in \mathbb{I}\}$ that are defined as $X^\tau{}_t \coloneqq X_t \wedge \tau(\omega)^{(\omega)}$ where this is a function of ω and this represents the minimum between the two. So, $x \wedge y$ is simply the minimum between x and y.

We also define the stopping random variable as $X_\tau{}^{(\omega)}$. This is just the value of the stochastic process at the stopping time itself. We are going to define it to be $X_{\tau(\omega)}{}^{(\omega)}$ if the stopping time is finite and it is zero otherwise. We need this condition since the way we define stopping times makes it possible for τ to take the value of infinity. If the stock price never reaches the level at 150 for instance, then $\tau(\omega)$ will be infinite for that sample path.

Let's look at an example now. Consider the experiment of tossing two coins where the outcomes are $\{$HH, HT, TH, and TT$\}$. We will define the same stochastic process that we looked at in previous videos.

σ	$X_0(\omega)$	$X_1(\omega)$	$X_2(\omega)$
HH	5	6	7
HT	5	6	5
TH	5	4	5
TT	5	4	3

Define $\tau(\omega) = \inf\{t \in \{0,1,2\}: X_t(\omega) = 6\}$. If we go back to the table, we can fill it in with the values of τ:

σ	$X_0(\omega)$	$X_1(\omega)$	$X_2(\omega)$	$\tau(\omega)$
HH	5	6	7	1
HT	5	6	5	1
TH	5	4	5	∞
TT	5	4	3	∞

Now the question is, is τ a stopping time with respect to the natural filtration of $X(\mathbb{F}^X)$? We found $\mathbb{F}^X$ in a previous video as a collection with $\mathbb{F}^X = \{\mathbb{F}_0{}^X, \mathbb{F}_1{}^X, \mathbb{F}_2{}^X\}$ where $\mathbb{F}_0{}^X$ is the trivial σ-algebra, $\mathbb{F}_1{}^X$ is a σ-algebra generated by $\{$HH,HT$\}$ as a block and $\{$TH,TT$\}$ as a block and $\mathbb{F}_2{}^X$ is equal to the powerset of ω.

To decide whether or not τ is a stopping time, we have to calculate the event $\{\tau \leq t\}$ and this should be in F_t for every t. To start, when $t = 0$, the event that $\{\tau \leq t\}$ is empty because τ is never less than or equal to 0. This belongs to ${\mathbb{F}_0}^X$. When $t = 1$, the event that $\{\tau \leq 1\}$, is equal to $\{\mathrm{HH,HT}\}$ which belongs to ${\mathbb{F}_1}^X$ because ${\mathbb{F}_1}^X$ is generated by this as a block. Finally, when $t = 2$, the event that $\{\tau \leq 2\}$, is exactly the same as before as $\{\mathrm{HH,HT}\}$ which obviously belongs to ${\mathbb{F}_2}^X$ because ${\mathbb{F}_2}^X$ is the powerset of Ω.

3.2.7 Notes: Conditional Expectation

In this chapter we put a rigorous foundation on another fundamental concept of probability: conditional expectation.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X \in m\mathcal{F}$ be a random variable. If $\mathcal{G} \subseteq \mathcal{F}$ is a sub-σ-algebra of $\mathcal{F}$, then any random variable Z that satisfies the following:

1. Z is $\mathcal{G}$-measurable;
2. Z satisfies the partial averaging property:
$$\int_A Z \, d\mathbb{P} = \int_A X \, d\mathbb{P}, \quad \text{for every } A \in \mathcal{G}$$

is called a *version of the conditional expectation* of X given $\mathcal{G}$.

We should really be saying *the* conditional expectation, but we can only do so after the following theorem.

Theorem 1

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X \in \mathcal{L}^1$ be a random variable. If $\mathcal{G} \subseteq \mathcal{F}$ is a sub-σ-algebra of $\mathcal{F}$, then a version of the conditional expectation of X given $\mathcal{G}$ exists and is unique $\mathbb{P}$-almost surely. We will denote it by $\mathbb{E}(X|\mathcal{G})$, and simply call it the (a.s. defined) conditional expectation of X given $\mathcal{G}$.

If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $X, Y \in m\mathcal{F}$ are random variables, then we define

$$\mathbb{E}(X|Y) := \mathbb{E}\big(X\big|\sigma(Y)\big),$$

if it exists.

Here is the interpretation. You have conducted a random experiment whose random outcome $\omega \in \Omega$ is not known. The σ-algebra $\mathcal{F}$ consists of all the events for which we can decide whether or not they have occurred once the experiment is conducted. (Note that this information might still not enable us to uniquely determine ω.) The random variable X is $\mathcal{F}$-measurable, which means that its value can be determined only from the knowledge of the occurrence or non-occurrence of the events in $\mathcal{F}$ (without knowing ω). Suppose that at the current stage we are only given partial information contained in the σ-algebra $\mathcal{G} \subseteq \mathcal{F}$. Then $\mathbb{E}(X|\mathcal{G})$ is the $\mathcal{G}$-measurable random variable that best approximates X (in some sense). That is, the best estimator for X that can be determined from the information in $\mathcal{G}$.

Suppose a fair coin is tossed twice and X represents the number of heads in the two tosses. Choose $\Omega = \{\text{HH,HT,TH,TT}\}, \mathcal{F} = 2^{\Omega}$ and $\mathbb{P} = \sum_{\omega \in \Omega} \frac{1}{4} \delta_{\omega}$. Now let $\mathcal{G} \subseteq \mathcal{F}$ be the σ-algebra that contains information about the outcome of the first toss. Then $\mathcal{G} = \sigma(\{\{\text{HH,HT}\}, \{\text{TH,TT}\}\})$. We want to find $\mathbb{E}(X|\mathcal{G})$. Since this random variable must be $\mathcal{G}$-measurable, it must be constant on the blocks that generate $\mathcal{G}$; that is,

$$\mathbb{E}(X|\mathcal{G}) = aI_{\{\text{HH,HT}\}} + bI_{\{\text{TH,TT}\}}$$

for some $a, b \in \mathbb{R}$. To find a and b, we use the partial averaging property. If $A = \{\text{HH,HT}\}$, then

$$\int_A \mathbb{E}(X|\mathcal{G}) d\mathbb{P} = a\mathbb{P}(\{\text{HH,HT}\}) = \int_A X \, d\mathbb{P} = \frac{1}{4}\big(X(\text{HH}) + X(\text{HT})\big) = \frac{3}{4}.$$

Hence $a = \frac{3}{2}$. A similar calculation on $A = \{\text{TH,TT}\}$ gives $b = \frac{1}{2}$.

Here is a more intuitive approach to finding $\mathbb{E}(X|\mathcal{G})$. The (constant) value of the random variable $\mathbb{E}(X|\mathcal{G})$ on the block $\{\text{HH,HT}\}$ is given by the weighted average of X over that block:

$$\mathbb{E}(X|\mathcal{G})(\omega) = X(\mathrm{HH}) \times \frac{\mathbb{P}(\{\mathrm{HH}\})}{\mathbb{P}(\{\mathrm{HH,HT}\})} + X(\mathrm{HT}) \times \frac{\mathbb{P}(\{\mathrm{HT}\})}{\mathbb{P}(\{\mathrm{HH,HT}\})} = \frac{3}{2}, \quad \omega \in \{\mathrm{HH,HT}\}.$$

Similarly,

$$\mathbb{E}(X|\mathcal{G})(\omega) = X(\mathrm{TH}) \times \frac{\mathbb{P}(\{\mathrm{TH}\})}{\mathbb{P}(\{\mathrm{TH,TT}\})} + X(\mathrm{TT}) \times \frac{\mathbb{P}(\{\mathrm{TT}\})}{\mathbb{P}(\{\mathrm{TH,TT}\})} = \frac{1}{2}, \quad \omega \in \{\mathrm{TH,TT}\}.$$

So, the random variable $\mathbb{E}(X|\mathcal{G})$ can be determined only from the outcome of the first toss. It is equal to $\frac{3}{2}$ if the first toss is a head and $\frac{1}{2}$ if the first toss is a tail. In contrast, X cannot be determined from this information: if, for instance, the first toss results in a head, the value of X could be 1 or 2 with equal probability. The value of $\mathbb{E}(X|\mathcal{G})$ on this scenario is simply the weighted average of these possibilities.

Consider the probability space $([0,1], \mathcal{B}([0,1]), \mathbb{P} = \lambda_1)$. Define $X, Y: [0,1] \to \mathbb{R}$ by

$$X(\omega) := \omega^2, \quad Y(\omega) = (\omega + 1) I_{[0,1/2]}(\omega), \quad \omega \in [0,1].$$

We want $\mathbb{E}(X|Y) = \mathbb{E}\big(X\big|\sigma(Y)\big)$. First, we have

$$\sigma(Y) = \sigma(\{\{Y \leq c\}: c \in \mathbb{R}\}) = \sigma\left(\left\{[0,c]: 0 \leq c \leq \frac{1}{2}\right\}\right)$$
$$= \left\{B \subseteq [0,1]: B \in \mathcal{B}\left(\left[0,\frac{1}{2}\right]\right) \text{ or } B = \left(\frac{1}{2},1\right] \cup B',\ B' \in \mathcal{B}\left(\left[0,\frac{1}{2}\right]\right)\right\}.$$

Thus, a necessary condition for a function Z to be $\sigma(Y)$-measurable is that it must be constant on $\left(\frac{1}{2},1\right]$. Hence $\mathbb{E}(X|Y)(\omega) = a$ for $\omega \in \left(\frac{1}{2},1\right]$. To find a, we use partial averaging

$$\int_{\left(\frac{1}{2},1\right]} \mathbb{E}(X|Y)\,d\mathbb{P} = a\mathbb{P}\left(\left(\frac{1}{2},1\right]\right) = \int_{\left(\frac{1}{2},1\right]} \omega^2 d\,\lambda_1 = \frac{7}{24}$$

$$\Longrightarrow a = \frac{\frac{7}{24}}{\mathbb{P}\left(\left(\frac{1}{2},1\right]\right)} = \frac{7}{12}.$$

For $\omega \in \left[0,\frac{1}{2}\right]$, $X(\omega) = \omega^2 = (Y(\omega) - 1)^2$, hence X is $\sigma(Y)$-measurable on that interval. So, we have

$$\mathbb{E}(X|Y)(\omega)\begin{cases} X(\omega) & \omega \in \left[0,\frac{1}{2}\right] \\ \frac{7}{12} & \omega \in \left(\frac{1}{2},1\right]. \end{cases}$$

Here are some useful properties of the conditional expectation. We say that two σ-algebras $\mathcal{F}$ and $\mathcal{G}$ are *independent* if for any $A \in \mathcal{F}$ and $B \in \mathcal{G}$, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X, Y \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. If $\mathcal{G} \subseteq \mathcal{F}$ is a sub-σ-algebra of $\mathcal{F}$, then:

1 (Total Expectation) $\mathbb{E}\big(\mathbb{E}(X|\mathcal{G})\big) = \mathbb{E}(X)$.

2 If X is $\mathcal{G}$-measurable, then $\mathbb{E}(X|\mathcal{G}) = X$ a.s..

3 If $X \geq 0$, then $\mathbb{E}(X|\mathcal{G}) \geq 0$ a.s..

4 $\mathbb{E}(aX + bY|\mathcal{G}) = a\mathbb{E}(X|\mathcal{G}) + b\mathbb{E}(Y|\mathcal{G})$ a.s., $a, b \in \mathbb{R}$.

5 If Y is $\mathcal{G}$-measurable, then $\mathbb{E}(YX|\mathcal{G}) = Y\mathbb{E}(X|\mathcal{G})$ a.s..

6 (Tower Property) If $\mathcal{H} \subseteq \mathcal{G}$, then $\mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H}) = \mathbb{E}(X|\mathcal{H})$ a.s..

7 If $\sigma(X)$ and $\mathcal{G}$ are independent, then $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X)$.

3.2.8 Notes: Martingales and Markov Processes

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(E, \mathcal{E})$ be a measurable space. A collection of random elements $X = \{X_t : \Omega \to E : t \in \mathbb{I}\}$ is called an E*-valued stochastic process*. The special case when $\{E, \mathcal{E}\} \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ gives the usual stochastic process discussed previously.

An E-valued stochastic process X is called a *Markov process with state space* E if for every bounded measurable function $f : \{E, \mathcal{E}\} \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$,

$$\mathbb{E}(f(X_t) | \mathcal{F}_s^X) = g_t(X_s) \quad s \leq t \text{ for } g_t : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})).$$

In simple terms, this means that the conditional distribution of X_t given all information about X up to and including time $s \leq t$ depends only on X_s.

An example of a Markov process is the Poisson process discussed previously.

We say that a process X has independent increments if for every $s < t$, $X_t - X_s$ is independent of $\mathcal{F}_s^X$. It is easy to see that a process with independent increments is a Markov process.

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space and $X = \{X_t : t \in \mathbb{I}\}$ be a stochastic process. We call X a *martingale* (or an $(\mathbb{F}, \mathbb{P})$-martingale) if

1 X is adapted to $\mathbb{F}$;

2 $\mathbb{E}(|X_t|) < \infty$ for all $t \in \mathbb{I}$;

3 $\mathbb{E}(X_t | \mathcal{F}_s) = X_s$ $\mathbb{P}$-a.s. for $s \leq t$.

We call X a *sub-martingale (resp. super-martingale)* if it satisfies 1, 2, and

3′ $\mathbb{E}(X_t | \mathcal{F}_s) \geq X_s$ (resp. $\mathbb{E}(X_t | \mathcal{F}_s) \leq X_s$) $\mathbb{P}$ a.s. for $s \leq t$.

A martingale is a driftless process — the best estimator of any future value is the current one. A submartingale drifts upwards, while a supermartingale drifts downwards. Furthermore, X is a martingale if and only if it is both a submartingale and a supermartingale.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $N = \{N_t : t \geq 0\}$ be a Poisson process with rate $\lambda > 0$. Let $\mathbb{F} = \mathbb{F}^N$ be the natural filtration of N. Then N is a submartingale. Indeed, N is adapted to $\mathbb{F}^N$, and $\mathbb{E}(|N_t|) = \mathbb{E}(N_t) = \lambda t < \infty$ for each $t \geq 0$. We also have (for $s < t$)

$$\mathbb{E}(N_t | \mathcal{F}_s^N) = \mathbb{E}(N_s + N_t - N_s | \mathcal{F}_s^N) = N_s + \lambda(t - s) \geq N_s.$$

On the other hand, the *compensated Poisson process* $M = \{N_t - \lambda t : t \geq 0\}$ is a martingale.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X = (X_n)_{n=1}^{\infty}$ be a sequence of i.i.d random variables in $\mathcal{L}^1$ with $\mathbb{E}(X_n) = \mu \ \forall n \in \mathbb{N}^+$. Fix $x \in \mathbb{R}$ and define the random walk $S = \{S_n : n \in \mathbb{N}\}$ by

$$S_0 = x, \quad S_n = \sum_{k=1}^{n} X_k \,, n \geq 1$$

We pick $\mathbb{F} = \{\mathcal{F}_n : \mathrm{n} \in \mathbb{N}\}$ as follows:

$$\mathcal{F}_0 \coloneqq \{\emptyset, \Omega\}, \quad \mathcal{F}_n = \sigma(\{X_1, \dots, X_n\}), \quad n \geq 1.$$

It is easy to see that conditions 1 and 2 are both satisfied. We also have for $m \leq n$,

$$\mathbb{E}(S_n | \mathcal{F}_m) = \mathbb{E}\left(S_m + \sum_{k=m+1}^{n} X_k | \mathcal{F}_m\right) = S_m + (n - m)\mu.$$

Thus, S is a submartingale if $\mu \geq 0$, a supermartingale if $\mu \leq 0$ and a martingale if $\mu = 0$.

In a similar fashion to the previous example, if $X_k \geq 0$ with $\mathbb{E}(X_k) = \mu$, then $S = \{S_n : n \in \mathbb{N}\}$ defined by $S_0 := 1, S_n := \prod_{k=1}^{n} X_k$ for $n \geq 1$ is a martingale if and only if $\mu = 1$. It is a submartingale if $\mu \geq 1$ and a supermartingale if $\mu \leq 1$.

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered space and $X = \{X_n : n \in \mathbb{N}\}$ be a stochastic process. Then X is a martingale if and only if X is adapted to $\mathbb{F}$, $X_n \in \mathcal{L}^1$ for each $n \in \mathbb{N}$ and

$$E(X_{n+1} | \mathcal{F}_n) = X_n \quad \mathbb{P} - \text{a.s.}.$$

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered space and $X = \{X_t : t \in \mathbb{I}\}$ be a stochastic process.

1 If X is a martingale, then $\mathbb{E}(X_t) = \mathbb{E}(X_s) = \mathbb{E}(X_0)$ for every $s, t \in \mathbb{I}$. That is, X has a constant mean.
2 If X is a submartingale, then $\mathbb{E}(X_t) \geq \mathbb{E}(X_s) \geq \mathbb{E}(X_0)$ for every $s \leq t \in \mathbb{I}$. That is, X has an increasing mean.
3 If X is a supermartingale, then $\mathbb{E}(X_t) \leq \mathbb{E}(X_s) \leq \mathbb{E}(X_0)$ for every $s \leq t \in \mathbb{I}$. That is, X has a decreasing mean.

A function $\varphi: \mathbb{R} \to \mathbb{R}$ is *convex* if for any $x, y \in \mathbb{R}$ and $\lambda \in \{0,1\}$,

$$\varphi(\lambda x + (1 - \lambda)y) \leq \lambda\varphi(x) + (1 - \lambda)\varphi(y).$$

Examples of convex functions include $\varphi(x) = x^2, e^x, -\ln x, |x|$.

Jensen's inequality: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X \in \mathcal{L}^1$. If $\varphi: \mathbb{R} \to \mathbb{R}$ is a convex function such that $\varphi(X) \in \mathcal{L}^1$, and $\mathcal{G} \subseteq \mathcal{F}$ is a sub-σ-algebra of $\mathcal{F}$, then

$$\mathbb{E}(\varphi(X) | \mathcal{G}) \geq \varphi\big(\mathbb{E}(X | \mathcal{G})\big) \quad \mathbb{P} - a.s..$$

In particular,

$$\mathbb{E}\big(\varphi(X)\big) \geq \varphi\big(\mathbb{E}(X)\big).$$

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space and $X = \{X_t : t \in \mathbb{I}\}$ be a martingale. If φ is a convex function such that $\varphi(X_t) \in \mathcal{L}^1$ for every $t \in \mathbb{I}$, then $\{\varphi(X_t) : t \in \mathbb{I}\}$ is a submartingale.

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space and $H \in \mathcal{L}^1$. For any index set $\mathbb{I} \subseteq [0, \infty)$, define $X = \{X_t : t \in \mathbb{I}\}$ by $X_t \coloneqq \mathbb{E}(H | \mathcal{F}_t)$ for each $t \in \mathbb{I}$. Then X is an $(\mathbb{F}, \mathbb{P})$-martingale.

3.2.9 Transcript: An Example of Conditional Expectation

Hi, in this video we define the conditional expectation of a random variable with respect to a σ-algebra.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. What that means is that $\mathcal{G}$ in itself is a σ-algebra and all the elements of $\mathcal{G}$ are also elements of $\mathcal{F}$.

If X is a random variable, then what a conditional expectation of X given $\mathcal{G}$ represents is the best $\mathcal{G}$-measurable estimator of X. In other words, it is the random variable that best approximates X and is also $\mathcal{G}$-measurable because X itself need not be $\mathcal{G}$-measurable.

Let's look at a number line. Consider $\mathcal{F}$ the σ-algebra. The random variable X is $\mathcal{F}$-measurable, meaning that if we know the information or the events that are in $\mathcal{F}$ then we know the value of X. However, it may not be $\mathcal{G}$-measurable. The conditional expectation of X given $\mathcal{G}$ provides the best estimator that is $\mathcal{G}$-measurable and approximates X in the most efficient way. We will see what that means later on.

The intuition about conditional expectation is captured by the following definition: a random variable Z is a version of the conditional expectation of X given $\mathcal{G}$ if it satisfies the following two conditions:

1 The first one is that Z must be $\mathcal{G}$-measurable itself, and
2 The second condition, which is what captures the intuition of approximation, is called a partial averaging property, and says that the integral over any A of Z is the same as the integral over A of X for every A in $\mathcal{G}$.

Now, Z exists if $E(|X|) < \infty$. So, if the expected value of X is finite, one can show, using the Radon-Nikodym theorem, that the version of the conditional expectation of X given $\mathcal{G}$ does exist. In that case Z is also unique and we denote it by writing $Z = E(X|\mathcal{G})$, which is a random variable that is defined almost surely.

Let's look at an example. Consider the experiment of tossing two coins, where the corresponding sample space is $\Omega = \{\text{HH, HT, TH, TT}\}$ and let X be the number of heads of the two tosses. We will take as the σ-algebra the powerset of Ω and as a probability measure, we'll take a weighted sum of Dirac measures where each sample point has equal weight. All this is saying is that each sample point has probability $\frac{1}{4}$. Let $\mathcal{G}$ be the σ-algebra that represents the information we have by observing the first toss. $\mathcal{G}$ is generated by the following two blocks: $(\{\{\text{HH,HT}\}, \{\text{TH,TT}\}\})$. $\{\text{HH,HT}\}$ corresponds to the event that the first toss is a head and $\{\text{TH,TT}\}$ corresponds to the event that the first is a tail.

The question is: what is the conditional expectation of X given $\mathcal{G}$? That's what we want to calculate. Now, first note that in order for this to be $\mathcal{G}$-measureable it has to be constant on these two blocks. This can be written a times the indicator of $\{\text{HH,HT}\}$ plus b times the indicator of $\{\text{TH,TT}\}$ and all we need to do is try and find what a and b are. For that we will use a partial averaging property. So, we take a set in $\mathcal{G}$ and we have to evaluate the integral of the conditional expectation with respect to the measure $\mathbb{P}$ and do the same evaluating the integral of X with respect to the measure $\mathbb{P}$. So, if we evaluate the integral of $\{\text{HH,HT}\}$ of the conditional expectation with respect to $\mathbb{P}$, we will get a times the integral over $\{\text{HH,HT}\}$ of the indicator of $\{\text{HH,HT}\}$ with respect to $\mathbb{P}$, which is equal to a times the probability of $\{\text{HH,HT}\}$ and that is $a\frac{1}{2}$. Written in full:

$$\int_{\{TH,TT\}} E(X|\mathcal{G})d\mathbb{P} = a\int_{\{HH,HT\}} I_{\{HH,HT\}}d\mathbb{P} = a\mathbb{P}(\{HH,HT\}) = a\frac{1}{2}.$$

Similarly, if we evaluate another integral of $\{\text{TH,TT}\}$ of the conditional expectation X given $\mathbb{P}$, that will give us $b\frac{1}{2}$. Written in full:

$$\int_{\{TH,TT\}} E(X|\mathcal{G})d\mathbb{P} = b\frac{1}{2}.$$

Those are the partial averages with respect to the conditional expectation.

Now, we do the same for X. Evaluate the integral of $\{\mathrm{HH},\ \mathrm{HT}\}^X$ and you'll see that the answer in that case is $\frac{3}{4}$. When you calculate the integral over $\{\mathrm{TH},\ \mathrm{TT}\}^X$ with respect to $\mathbb{P}$, the answer is $\frac{1}{4}$. We then equate $\frac{3}{4}$ to a times $\frac{1}{2}$ and that gives us a is equal to $\frac{3}{2}$. And, by solving this, we can get that b is equal to $\frac{1}{2}$.

Those are the values of a and b, which implies that the conditional expectation X given $\mathcal{G}$ is just simply equal to $\frac{3}{2}$ times the indicator of $\{\mathrm{HH,HT}\}$ plus $\frac{1}{2}$ times the indicator of $\{\mathrm{TH,TT}\}$. Written in full:

$$E(X|\mathcal{G}) = \frac{3}{2} I_{\{HH,HT\}} + \frac{1}{2} I_{\{TH,TT\}}.$$

We can sketch the conditional expectation on a graph together with X. X of (HH) is 2, (HT) is 1, (TH) is 1, and (TT) is 0. We can then do the same for the conditional expectation: it's $\frac{3}{2}$ for both (TH) and (TT). So, what the conditional expectation does is groups (HH) and (HT) together and gives them one value which is the average of the two. It also groups (TH) and (TT) together and gives them one value which is the average of the two.

3.2.10 Notes: Problem Set

Problem 1

Let $\Omega = \{a, b, c, d\}, \mathcal{F} = 2^{\Omega}, \mathcal{F}_0 = \{\emptyset, \Omega\}, \mathcal{F}_1 = \sigma(\{a, b\}, \{c, d\}), \mathcal{F}_2 = \mathcal{F}$ and $X = \{X_0, X_1, X_2\}$, be defined as

ω	$X_0(\omega)$	$X_1(\omega)$	$X_2(\omega)$
a	1	3	4
b	1	3	5
c	1	2	3
d	1	2	1

Let $\tau: \Omega \to \{0,1,2,\infty\}$ be $\tau(\omega) \coloneqq \inf\{n \in \{0,1,2\}: X_n(\omega) \geq 3\}$. Then what is $\big(\tau(a), \tau(b), \tau(c), \tau(d)\big)$?

Solution:

In this example we have four different possible realizations of ω, that is $\Omega = \{a, b, c, d\}$. We can interpret each realization $\{a, b, c, d\}$ as a potential path that the discrete stochastic process can follow. In this case, $X = \{X_0, X_1, X_2\}$, is only allowed to follow four different paths, as you can see in the table:

ω	$X_0(\omega)$	$X_1(\omega)$	$X_2(\omega)$
a	1	3	4
b	1	3	5
c	1	2	3
d	1	2	1

For instance, in the path a, the realization of $X(a) = \{X_0, X_1, X_2\}$ is $X(a) = \{1,3,4\}$. As the problem defined the stopping time as:

$$\tau(\omega) \coloneqq \inf\{n \in \{0,1,2\}: X_n(\omega)\},$$

we just have to find the value of t (take into account that we have only three times in this problem $\{0,1,2\}$) such that $X_t(\omega) \geq 3$, and repeat the process for each path, $\{a, b, c, d\}$, in order to compute $\big(\tau(a), \tau(b), \tau(c), \tau(d)\big)$.

Let focus in the first path, a. The realization for this path (from the table) is: $X(a) = \{1,3,4\}$. It is easy to see that: $X_0(a) = 1$ does not hold the condition that $X_t(a) \geq 3$. On the other hand, $X_1(a) = 3$ holds the condition, thus $t = 1$ is the minimum time (inf) such that $X_t(\omega) \geq 3$. Thus, we can conclude that $\tau(a) = 1$.

For paths b and c, we will get (following the same procedure) $\tau(b) = 1$ and $\tau(c) = 2$.

Finally, considering path d, we observe that there isn't any time $X_t(d)$ such that $X_t(d) \geq 3$. As the problem tell us that $\tau: \Omega \to \{0,1,2,\infty\}$, we can conclude that $\tau(d) = \infty$.

To sum up, the solution is equal to: $\big(\tau(a), \tau(b), \tau(c), \tau(d)\big) = (1,1,2,\infty)$.

Problem 2

Let $\Omega = \{a, b, c, d\}, \mathcal{F} = 2^\Omega, \mathcal{F}_0 = \{\emptyset, \Omega\}, \mathcal{F}_1 = \sigma\big(\{\{a, c\}, \{b, d\}\}\big), \mathcal{F}_2 = \mathcal{F}$ and $X = \{X_0, X_1, X_2\}$, τ be defined as

ω	$X_0(\omega)$	$X_1(\omega)$	$X_2(\omega)$	$\tau(\omega)$
a	1	4	4	1
b	2	3	5	∞
c	1	2	3	1
d	1	1	1	∞

Then what is $X_\tau(d)$?

Solution:

First of all, let's summarize the theory (from the lecture notes) that we are going to need to solve this problem. We know that if τ is a stopping time and $X = \{X_t : t \in \mathbb{I}\}$ is a stochastic process, we can define the *stopped process* $X^\tau = \{X_t^\tau : t \in \mathbb{I}\}$ as $X_t^\tau := X_{\tau \wedge t}$ – i.e.,

$$X_t^\tau(\omega) := \begin{cases} X_t(\omega) & t < \tau(\omega) \\ X_{\tau(\omega)}(\omega) & t \geq \tau(\omega) \end{cases} \quad \omega \in \Omega.$$

We also define the random variable X_τ by\]

$$X_\tau(\omega) := \begin{cases} X_{\tau(\omega)}(\omega) & \tau(\omega) < \infty \\ 0 & \tau(\omega) = \infty. \end{cases}$$

As the problem only asks about path d, we can focus only on $\omega = d$. As $\tau(d) = \infty$ we can already conclude (see the definition of the random variable X_τ above) that $X_\tau(d) = 0$.

Problem 3

Let $\Omega = \{a, b, c, d\}, \mathcal{F} = 2^\Omega, \mathcal{F}_0 = \{\emptyset, \Omega\}, \mathcal{F}_1 = \sigma(\{\{a, b\}, \{c, d\}\}), \mathcal{F}_2 = \mathcal{F}$ and $X = \{X_0, X_1, X_2\}$ be a martingale defined as

ω	$\mathbb{P}(\{\omega\})$	$X_0(\omega)$	$X_1(\omega)$	$X_2(\omega)$
a	$\frac{1}{4}$	2	3	4
b	$\frac{1}{4}$	2	α	2
c	$\frac{1}{4}$	2	2	2
d	$\frac{1}{4}$	2	1	0

Then what is α?

Solution:

We know the definition of a martingale from the theory.

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered space and $X = \{X_n : n \in \mathbb{N}\}$ be a stochastic process. Then X is a martingale if and only if X is adapted to $\mathbb{F}$, $X_n \in \mathcal{L}^1$ for each $n \in \mathbb{N}$ and

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n \quad \mathbb{P} - a.s..$$

Moreover, we also know that the condition above is equivalent (more practical) to the following proposition.

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered space and $X = \{X_t : t \in \mathbb{I}\}$ be a stochastic process. Then, if X is a martingale, then $\mathbb{E}(X_t) = \mathbb{E}(X_s) = \mathbb{E}(X_0)$ for every $s, t \in \mathbb{I}$. That is, X has a constant mean.

Following the above proposition, we can solve the problem as follows. First, we compute the $\mathbb{E}(X_0)$ which does not depend on α:

$$\mathbb{E}(X_0) = \frac{1}{4} * 2 + \frac{1}{4} * 2 + \frac{1}{4} * 2 + \frac{1}{4} * 2 = 2.$$

Now we have to find α such that $\mathbb{E}(X_1) = 2$. It is easy to see that $\alpha = 2$ solving the following equation:

$$\frac{1}{4} * 3 + \frac{1}{4} * \alpha + \frac{1}{4} * 2 + \frac{1}{4} * 1 = 2.$$

Thus, the solution is $\alpha = 2$.

3.3 Additional Resources

Shiryaev, A.N., 1996. *Probability*. Graduate Texts in Mathematics, R.P. Boas, Ed. Springer.

Roger, L.C.G. & Williams, D., 1987. *Diffusions, Markov processes and martingales*. Cambridge University Press.

Williams, D., 1991. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press.

3.4 Collaborative Review Task

In this module, you are required to complete a collaborative review task, which is designed to test your ability to apply and analyze the knowledge you have learned during the week.

Question

Let $\Omega = \{a, b, c, d\}$ and

ω	$\mathbb{P}(\{\omega\})$	$X_0(\omega)$	$X_1(\omega)$	$X_2(\omega)$	$Y(\omega)$
a	$\frac{1}{9}$	12	18	36	12
b	$\frac{2}{9}$	12	18	9	15
c	$\frac{1}{6}$	12	9	12	16
d	$\frac{1}{2}$	12	9	8	8

Let $\mathbb{I} = \{0,1,2\}$ and define $\tau: \Omega \to \mathbb{I} \cup \{\infty\}$ by

$$\tau(\omega) = inf\{n \in \mathbb{I}: X_n(\omega) < 10\}.$$

1. Find $\mathbb{F}^X = \{\mathcal{F}_n^X : n \in I\} = \{\mathcal{F}_0^X, \mathcal{F}_1^X, \mathcal{F}_2^X\}$, the natural filtration of X.
2. Calculate $\tau(\omega)$ for each $\omega \in \Omega$.
3. Show that τ is a stopping time with respect to $\mathbb{F}^X$.

4 Is X a martingale with respect to $\mathbb{F}^X$?

5 Define $Z = \{Z_0, Z_1, Z_2\}$ by $Z_n = E(Y|\mathcal{F}_n^X)$ for $n \in \mathbb{I}$. Show that Z is an $\mathbb{F}^X$-martingale.

6 Find the stopped process Z^τ.