Martingale Theory and Applications

Dr Nic Freeman

April 29, 2015

Contents

Chapter 1

Conditional Expectation

In this chapter we review probability spaces, introduce σ-fields and discuss the expectation of random variables. In particular, we introduce the concept of a random variable being measurable with respect to a given σ-field. We then introduce the conditional expectation of random variables and discuss its properties.

1.1 Probability spaces and σ-fields

Let Ω be a set. In probability theory, the symbol Ω is typically (and always, in this course) used to denote the *sample space*. Intuitively, we think of ourselves as conducting some random experiment, with an unknown outcome. The set Ω contains an $\omega \in \Omega$ for every possible outcome of the experiment.

Subsets of Ω correspond to collections of possible outcomes; such a subset is referred as an event. For instance, if we roll a dice we might take $\Omega = \{1, 2, 3, 4, 5, 6\}$ and the set $\{1, 3, 5\}$ is the event that our dice roll is an odd number.

Definition 1.1.1 *Let $\mathcal{F}$ be a set of subsets of Ω. We say $\mathcal{F}$ is a σ-field if it satisfies the following properties:*

1. $\emptyset \in \mathcal{F}$.
2. *if* $A \in \mathcal{F}$ *then* $\Omega \setminus A \in \mathcal{F}$.
3. *if* $A_1, A_2, \ldots \in \mathcal{F}$ *then* $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

The role of a σ-field is to choose which subsets of outcomes we are actually interested in. The power set $\mathcal{F} = \mathcal{P}(\Omega)$ is always a σ-field, meaning that every subset of Ω is an event. But $\mathcal{P}(\Omega)$ can be very big, and if our experiment is complicated, with many or even infinitely many possible outcomes, we might want to consider a smaller choice of $\mathcal{F}$ instead.

Sometimes we will need to deal with more than one σ-field at a time. A σ-field $\mathcal{G}$ such that $\mathcal{G} \subseteq \mathcal{F}$ is known as a *sub-σ-field* of $\mathcal{F}$.

We say that a subset $A \subseteq \Omega$ is *measurable*, or that it is an *event* (or *measurable event*), if $A \in \mathcal{F}$. To make to it clear which σ-field we mean to use in this definition, we sometimes write that an event is $\mathcal{F}$-measurable.

Example 1.1.2 *Some examples of experiments and the σ-fields we might choose for them are the following:*

- *We toss a coin, which might result in heads H or tails T. We take $\Omega = \{H, T\}$ and $\mathcal{F} = \big\{\{H,T\}, \{H\}, \{T\}, \emptyset\big\}$ to be the power set of Ω.*
- *We toss two coins, both of which might result in heads H or tails T. We take $\Omega = \{HH, TT, HT, TH\}$. However, we are only interested in the outcome that both coins are heads. We take $\mathcal{F} = \big\{\Omega, \Omega \setminus \{HH\}, \{HH\}, \emptyset\big\}$.*

There are natural ways to choose a σ-field, even if we think of Ω as just an arbitrary set. For example, $\mathcal{F} = \{\Omega, \emptyset\}$ is a σ-field. If A is a subset of Ω, then the set $\mathcal{F} = \{\Omega, A, \Omega \setminus A, \emptyset\}$ is a σ-field (check it!).

Given Ω and $\mathcal{F}$, the final ingredient of a probability space is a measure $\mathbb{P}$, which tells us how likely the events in $\mathcal{F}$ are to occur.

Definition 1.1.3 *A probability measure $\mathbb{P}$ is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ satisfying:*

1. $\mathbb{P}[\Omega] = 1$.
2. *If $A_1, A_2, \ldots \in \mathcal{F}$ are pair-wise disjoint, i.e. $A_i \cap A_j = \emptyset$ for all i, j such that $i \neq j$, then*

$$\mathbb{P}\left[\bigcup_{i=1}^{\infty} A_i\right] = \sum_{i=1}^{\infty} \mathbb{P}[A_i].$$

The second of these conditions if often called σ-additivity. Note that we needed Definition 1.1.1 to make sense of Definion 1.1.3, because we needed something to tell us that $\mathbb{P}\left[\bigcup_{i=1}^{\infty} A_i\right]$ was defined!

Definition 1.1.4 *A probability space is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F}$ is a σ algebra and $\mathbb{P}$ is a probability measure.*

We commented above that often we want to choose $\mathcal{F}$ to be smaller than $\mathcal{P}(\Omega)$, but we have not yet shown how to choose a suitably small $\mathcal{F}$. Fortunately, there is a general way of doing so, for which we need the following lemma.

Lemma 1.1.5 *Let I be any set and for each $i \in I$ let $\mathcal{F}_i$ be a σ-field. Then*

$$\mathcal{F} = \bigcap_{i \in I} \mathcal{F}_i \tag{1.1}$$

is a σ-field

Proof: We check the three conditions of Definition 1.1.1 for $\mathcal{F}$.

(1) Since each $\mathcal{F}_i$ is a σ-field, we have $\emptyset \in \mathcal{F}_i$. Hence $\emptyset \in \cap_i \mathcal{F}_i$.

(2) If $A \in \mathcal{F} = \cap_i \mathcal{F}_i$ then $A \in \mathcal{F}_i$ for each i. Since each $\mathcal{F}_i$ is a σ-field, $\Omega \setminus A \in \mathcal{F}_i$ for each i. Hence $\Omega \setminus A \in \cap_i \mathcal{F}_i$.

(3) If $A_j \in \mathcal{F}$ for all j, then $A_j \in \mathcal{F}_i$ for all i and j. Since each $\mathcal{F}_i$ is a σ-field, $\cup_j A_j \in \mathcal{F}_i$ for all i. Hence $\cup_j A_j \in \cap_i \mathcal{F}_i$. ∎

Corollary 1.1.6 *In particular, if $\mathcal{F}_1$ and $\mathcal{F}_2$ are σ-fields, so is $\mathcal{F}_1 \cap \mathcal{F}_2$.*

Now, suppose that we have our Ω and we have a finite or countable collection of $E_1, E_2, \ldots \subseteq \Omega$, which we want to be events. Let $\mathscr{F}$ be the set of all σ-fields that contain $E_1, E_2, \ldots$. We enumerate $\mathscr{F}$ as $\mathscr{F} = \{\mathcal{F}_i\,;\, i \in I\}$, and apply Lemma 1.1.5. We thus obtain a σ-field $\mathcal{F}$, which contains the events that we wanted.

The key point here is that $\mathcal{F}$ is the smallest σ-field that has $E_1, E_2, \ldots$ as events. To see why, note that by (1.1), $\mathcal{F}$ is contained inside any σ-field $\mathcal{F}'$ which has $E_1, E_2, \ldots$ as events.

Definition 1.1.7 *Let $E_1, E_2, \ldots$ be subsets of Ω. We write $\sigma(E_1, E_2, \ldots,)$ for the smallest σ-field containing $E_1, E_2, \ldots$.*

With $\Omega = \mathbb{R}$, the Borel σ-field is (by definition) the smallest σ-field containing all subintervals of $\mathbb{R}$. With Ω as any set, and $A \subseteq \Omega$, our example $\{\emptyset, A, \Omega \setminus A, \Omega\}$ is clearly $\sigma(A)$. In general, though, the point of Definition 1.1.7 is that we know useful σ-fields exist without having to construct them explicitly.

In the same style, if $\mathcal{F}_1, \mathcal{F}_2 \ldots$ are σ-fields then we write $\sigma(\mathcal{F}_1, \mathcal{F}_2, \ldots)$ for the smallest σ-algebra with respect to which all events in $\mathcal{F}_1, \mathcal{F}_2, \ldots$ are measurable.

From Definition 1.1.1 and 1.1.3 we can deduce the 'usual' properties of probability. For example:

- If $A \in \mathcal{F}$ then $\Omega \setminus A \in \mathcal{F}$, and since $\Omega = A \cup (\Omega \setminus A)$ we have $\mathbb{P}[\Omega \setminus A] = 1 - \mathbb{P}[A]$.
- If $A, B \in \mathcal{F}$ and $A \subseteq B$ then we can write $B = A \cup (B \setminus A)$, which gives us that $\mathbb{P}[B] = \mathbb{P}[B \setminus A] + P[A]$, which implies that $\mathbb{P}[A] \leq \mathbb{P}[B]$.

And so on. In this course we are concerned with applying probability theory rather than with relating its properties right back to the definition of a probability space; but you should realize that it is always possible to do so.

Definitions 1.1.1 and 1.1.3 both involve countable unions. Its convenient to be able to use countable intersections too, for which we need the following lemma.

Lemma 1.1.8 *Let $A_1, A_2, \ldots \in \mathcal{F}$, where $\mathcal{F}$ is a σ-field. Then $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$.*

PROOF: We can write

$$\bigcap_{i=1}^{\infty} A_i = \bigcap_{i=1}^{\infty} \Omega \setminus (\Omega \setminus A_i) = \Omega \setminus \left(\bigcup_{i=1}^{\infty} \Omega \setminus A_i \right).$$

Since $\mathcal{F}$ is a σ-field, $\Omega \setminus A_i \in \mathcal{F}$ for all i. Hence also $\bigcup_{i=1}^{\infty} \Omega \setminus A_i \in \mathcal{F}$, which in turn means that $\Omega \setminus (\bigcup_{i=1}^{\infty} \Omega \setminus A_i) \in \mathcal{F}$. ■

In general, *uncountable* unions and intersections of measurable sets need not be measurable. The reasons why we only allow countable unions/intersections in probability are complicated and beyond the scope of this course. Loosely speaking, the bigger we make $\mathcal{F}$, the harder it is to make a probability measure $\mathbb{P}$, because we need to define $\mathbb{P}[A]$ for all $A \in \mathcal{F}$ in a way that satisfies Definition 1.1.3. Allowing uncountable set operations would (in several natural situations, including the Borel σ-field on $\mathbb{R}$) result in $\mathcal{F}$ being so large that it would actually be impossible to find a suitable $\mathbb{P}$.

From now on, the symbols Ω, $\mathcal{F}$ and $\mathbb{P}$ always denote the three elements of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

1.2 Random Variables

Our probability space gives us a label $\omega \in \Omega$ for every possible outcome. Sometimes it is more convenient to think about a property of ω, rather than about Ω itself. For this, we use a random variable, $X : \Omega \to \mathbb{R}$. We write

$$X^{-1}(A) = \{\omega \in \Omega \,;\, X(\omega) \in A\},$$

for $A \subseteq \mathbb{R}$, which is called the *pre-image* of A under X.

Definition 1.2.1 *A function $X : \Omega \to \mathbb{R}$ is said to be $\mathcal{F}$-measurable if*

$$\textit{for all } a < b \in \mathbb{R},\ X^{-1}(a,b) \textit{ is } \mathcal{F}\textit{-measurable.}$$

If it is already clear which σ-field $\mathcal{F}$ should be used in the definition, which simply say that X is *measurable* or, equivalently, that X is a *random variable*. We will often shorten this to writing simply $X \in m\mathcal{F}$. The relationship to the notation you have used before in probability is that $\mathbb{P}[X \in A]$ means $\mathbb{P}[X^{-1}(A)]$, so as e.g. $\mathbb{P}\left[a < X < b\right] = \mathbb{P}[\omega \in \Omega \,;\, X(w) \in (a,b)]$.

The key point in Definition 1.2.1 is that, *when we choose how big we want our $\mathcal{F}$ to be, we are also choosing which functions $X : \Omega \to \mathbb{R}$ are random variables.* This will become very important to us later in the course.

For example, suppose we toss a coin twice, with $\Omega = \{HH, HT, TH, TT\}$ as in Example 1.1.2. If we chose $\mathcal{F} = \mathcal{P}(\Omega)$ then any subset of Ω is $\mathcal{F}$-measurable, and consequently any function $X : \Omega \to \mathbb{R}$ is $\mathcal{F}$-measurable. However, suppose we chose

$$\mathcal{F} = \big\{\Omega, \Omega \setminus \{HH\}, \{HH\}, \emptyset\big\}$$

(as we did in Example 1.1.1). Then if we look at function

$$X(\omega) = \text{ the total number of tails which occurred}$$

we have

$$X^{-1}(0,3) = \{HT, TH, TT\} \notin \mathcal{F}.$$

So X is not $\mathcal{F}$-measurable. However, the function

$$Y(\omega) = \begin{cases} 0 & \text{if both coins were heads} \\ 1 & \text{otherwise} \end{cases}$$

is $\mathcal{F}$-measurable; in fact we can list

$$Y^{-1}(a,b) = \begin{cases} \emptyset & \text{if } 0,1 \notin (a,b) \\ \{HH\} & \text{if } 0 \in (a,b), 1 \notin (a,b) \\ \Omega \setminus \{HH\} & \text{if } 0 \notin (a,b), 1 \in (a,b) \\ \Omega & \text{if } 0,1 \in (a,b). \end{cases} \tag{1.2}$$

The interaction between random variables and σ-fields can be summarised as follows:

$$\boxed{\sigma\text{-field } \mathcal{F}} \quad \leftrightarrow \quad \boxed{\text{which information we care about}}$$
$$\boxed{X \text{ is } \mathcal{F}\text{-measurable}} \quad \leftrightarrow \quad \boxed{X \text{ depends only on information that we care about}}$$

It is easiest to think of this in the finite setting, when the function $X : \Omega \to \mathbb{R}$ takes only finitely many values. Then, as you might already suspect from (1.2), to check if X is measurable its enough just to check if the subsets of Ω corresponding to the values that X actually takes are measurable. In fact, we can allow countably many values too. That is to say:

Lemma 1.2.2 *Let $X : \Omega \to \mathbb{R}$. Suppose that $\mathcal{Q} = \{x_1, \ldots, x_n\}$ is a finite subset of $\mathbb{R}$, and that $X(\omega) \in \mathcal{Q}$ for all $\omega \in \Omega$. Then*

$$X \text{ is } \mathcal{F}\text{-measurable} \;\Leftrightarrow\; \text{ for all } i \leq n,\; X^{-1}(x_i) \in \mathcal{F}.$$

PROOF: We prove each direction in turn.

($\Rightarrow$): Without loss of generality we assume that $x_1 < x_2 < \ldots < x_n$. Since $\{x_1, \ldots, x_n\}$ is finite, for each $j \leq n$ we can find $a, b \in \mathbb{R}$ such that

$$x_{j-1} < a < x_j < b < x_{j+1}.$$

Then $X^{-1}(x_j) = X^{-1}(a, b)$, which is $\in \mathcal{F}$ because X is $\mathcal{F}$-measurable.

($\Leftarrow$): Let $a < b \in \mathbb{R}$. Let $I = \{i\,;\, x_i \in (a, b)\}$. Note that I is finite. Then

$$X^{-1}(a, b) = \bigcup_{i \in I} X^{-1}(x_i)$$

Since each $X^{-1}(x_i) \in \mathcal{F}$, and $\mathcal{F}$ is a σ-field, $X^{-1}(a, b) \in \mathcal{F}$ too. ■

For example, in place of (1.2), using Lemma 1.2.2 we would only need to check that $Y^{-1}(0) \in \mathcal{F}$ and $Y^{-1}(1) \in \mathcal{F}$. If X takes many values, then $\sigma(X)$ can be very large, and Lemma 1.2.2 can save a lot of work.

There is a natural σ-algebra associated to each random variable:

Definition 1.2.3 *The σ-field generated by the random variable X, denoted $\sigma(X)$, is the smallest σ-field $\mathcal{F}$ such that X is $\mathcal{F}$-measurable.*

More generally, the σ-field $\sigma(X_1, X_2, \ldots)$ generated by a collection of random variables $X_1, X_2, \ldots$ is the smallest σ-field in which each X_n is measurable.

The intuition here is that $\sigma(X)$ is the minimal σ-field of events on which the random behaviour of X depends. By definition, $\sigma(X) = \sigma(X^{-1}(a, b)\,;\, a < b)$. If X takes only a finite or countable set of values $x_1, x_2, \ldots$, this means that $\sigma(X) = \sigma(X^{-1}(x_i)\,;\, i = 1, 2, \ldots)$.

Example 1.2.4 *Consider throwing a fair die. Let $\Omega = \{1, 2, 3, 4, 5, 6\}$, let $\mathcal{F} = \mathcal{P}(\Omega)$. Let*

$$X(\omega) = \begin{cases} 1 & \text{if } \omega = 1 \\ 3 & \text{if } \omega = 2 \\ 0 & \text{otherwise.} \end{cases}$$

Then $X(\omega) \in \{0, 1, 3\}$, and $X^{-1}(0) = \{3, 4, 5, 6\}$, $X^{-1}(1) = \{1\}$ and $X^{-1}(3) = \{2\}$. So $\sigma(X) = \sigma(\{1\}, \{2\}, \{3, 4, 5, 6\})$.

Lemma 1.2.5 *Let X be a random variable. Then X is $\sigma(X)$-measurable.*

Proof: Let $a < b \in \mathbb{R}$. Then, if $\mathcal{G}$ is any σ-field such that X is $\mathcal{G}$-measurable, $X^{-1}(a,b) \in \mathcal{G}$. Hence, by definition of $\sigma(X)$, $X^{-1}(a,b) \in \sigma(X)$. Since a, b were arbitrary, $X \in \sigma(X)$. ■

In particular, if $A \in \mathcal{F}$ then the indicator function of the set A, which we write as $\mathbb{1}_A$, satisfies $\mathbb{1}_A \in \sigma(A) \subseteq \mathcal{F}$.

Given a collection of random variables, it is useful to be able to construct other random variables from them. To do so we have the following proposition.

Proposition 1.2.6 *Let $\alpha \in \mathbb{R}$ and let $X, Y, X_1, X_2, \ldots$ be $\mathcal{F}$-measurable functions from $\Omega \to \mathbb{R}$. Then*

$$\alpha X, \quad X+Y, \quad XY, \quad \limsup_{n\to\infty} X_n, \quad \liminf_{n\to\infty} X_n, \tag{1.3}$$

are all $\mathcal{F}$-measurable. Further, if X_∞ given by $X_\infty(\omega) = \lim_{n\to\infty} X_n(\omega)$ exists almost surely, then X_∞ is $\mathcal{F}$-measurable.

Essentially, every natural way of combining random variables together leads to other random variables. Proposition 1.2.6 can usually be used to show this.

For example, suppose X is a random variable and let $Y = e^X$, which means that $Y(\omega) = \lim_{n\to\infty} \sum_{i=1}^n \frac{X(\omega)^i}{i!}$. Recall that we know from analysis that this limit exists since $e^x = \lim\limits_{n\to\infty} \sum_{i=1}^n \frac{x^i}{i!}$ exists for all $x \in \mathbb{R}$. Each of the partial sums $Y_n(\omega) = \sum_1^n \frac{X(\omega)^i}{i!}$ is a random variable (we could repeatedly use the first two parts of (1.3) to show this) and, since the limit exists, $Y(\omega) = \lim\limits_{n\to\infty} Y_n(\omega) = \limsup\limits_{n\to\infty} Y_n(\omega)$ is measurable.

Remark 1.2.7 $(\star)$ *Recall that if $\Omega = \mathbb{R}$, we can take $\mathcal{F} = \mathcal{B}(\mathbb{R})$ to be the Borel σ-field on $\mathbb{R}$. In this case, there are examples of functions $X : \mathbb{R} \to \mathbb{R}$ that are not $\mathcal{B}(\mathbb{R})$-measurable. But they are very difficult to find and they do not occur naturally.*

In general, if X is a random variable and $g : \mathbb{R} \to \mathbb{R}$ is any 'nice' deterministic function then $g(X)$ is also a random variable. This includes polynomials, all trig functions, all monotone functions, all piecewise linear functions, all integrals/derivatives, etc etc.

We can express the concept of independence, which you already know about for random variables, in terms of σ-fields. Recall that two events $E_1, E_2 \in \mathcal{F}$ are said to be independent if $\mathbb{P}[E_1]\mathbb{P}[E_2] = \mathbb{P}[E_1 \cap E_2]$.

Definition 1.2.8 *Sub-σ-fields $\mathcal{G}_1, \mathcal{G}_2$ of $\mathcal{F}$ are said to be* independent *if, whenever $G_i \in \mathcal{G}_i$, $i = 1, 2$, we have $\mathbb{P}(G_1 \cap G_2) = \mathbb{P}(G_1)\mathbb{P}(G_2)$.*

Random variables X_1 and X_2 are said to be independent if the σ-fields $\sigma(X_1)$ and $\sigma(X_2)$ are independent.

The extension of Definition 1.2.8 to sets of events/σ-fields/etc should be clear (and is not needed for this course), but remember that pairwise independence does not imply independence. Naturally, you should recall that $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ if X and Y are independent random variables (we will come to look at $\mathbb{E}$ shortly).

1.3 Two kinds of examples

In this section we consolidate our knowledge from the previous two sections by looking at two important contrasting examples.

1.3.1 Finite Ω

Let $n \in \mathbb{N}$, and let $\Omega = \{x_1, x_2, \dots x_n\}$ be a finite set. Let $\mathcal{F} = \mathcal{P}(\Omega)$, which is also a finite set. We have seen how it is possible to construct other σ-fields on Ω too. Since $\mathcal{F}$ contains every subset of Ω, any σ-field on Ω is a sub-σ-field of $\mathcal{F}$.

In this case we can define a probability measure on Ω by choosing a finite sequence $a_1, a_2, \dots, a_n$ such that each $a_i \in [0,1]$ and $\sum_1^n a_i = 1$. We set $\mathbb{P}[x_i] = a_i$. This naturally extends to defining $\mathbb{P}[A]$ for any subset $A \subseteq \Omega$, by setting

$$\mathbb{P}[A] = \sum_{\{i\,;\,x_i \in A\}} \mathbb{P}[x_i] = \sum_{\{i\,;\,x_i \in A\}} a_i. \tag{1.4}$$

It is hopefully obvious (and tedious to check) that, with this definition, $\mathbb{P}$ is a probability measure. Consequently $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.

All experiments with only finitely many outcomes fit into this category of examples. We have already seen several of them.

- Roll a biased die. Choose $\Omega = \{1,2,3,4,5,6\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ and define $\mathbb{P}$ by setting $\mathbb{P}[i] = \frac{1}{8}$ for $i = 1,2,3,4,5$ and $\mathbb{P}[6] = \frac{3}{8}$.
- Toss a coin twice, independently. Choose $\Omega = \{HH, TH, HT, TT\}$, $\mathcal{F} = \mathcal{P}(\Omega)$. Define $\mathbb{P}$ by setting $\mathbb{P}[**] = \frac{1}{4}$, where each instance of $*$ denotes either H or T.

For a sub-σ-field $\mathcal{G}$ of $\mathcal{F}$, the triplet $(\Omega, \mathcal{G}, \mathbb{P}_{\mathcal{G}})$ is also a probability space. Here $\mathbb{P}_{\mathcal{G}} : \mathcal{G} \to [0,1]$ simply means $\mathbb{P}$ restricted to $\mathcal{G}$, i.e. $\mathbb{P}_{\mathcal{G}}[A] = \mathbb{P}[A]$.

If $\mathcal{G} \neq \mathcal{F}$, some random variables $X : \Omega \to \mathbb{R}$ are $\mathcal{G}$-measurable and others are not. Intuitively, a random variable X is $\mathcal{G}$-measurable if we can deduce the value of $X(\omega)$ from knowing only, for all $G \in \mathcal{G}$, if $\omega \in G$. Each $G \in \mathcal{G}$ represents a piece of information that $\mathcal{G}$ gives us have access too (and this piece of information is whether or not $\omega \in \mathcal{G}$); if $\mathcal{G}$ gives us access to enough information then we can determine the value of $X(\omega)$ for all ω, in which case we say that X is $\mathcal{G}$-measurable.

Rigorously, to check if a given random variable is $\mathcal{G}$ measurable, we can either check each pre-image directly, or (usually better) use Lemma 1.2.2 and/or Proposition 1.2.6. To show that a given random variable X is not $\mathcal{G}$-measurable, we just need to find $a < b \in \mathbb{R}$ such that $X^{-1}(a,b) \notin \mathcal{G}$.

1.3.2 An example with infinite Ω

Now we flex our muscles a bit, and look at an example where Ω is infinite. We toss a coin infinitely many times, then $\Omega = \{H, T\}^{\mathbb{N}}$ and we write $\omega = (\omega_1, \omega_2, \dots)$ for a typical outcome. We define the random variables $X_n(\omega) = \omega_n$, so as X_n represents the results of the n^{th} throw. We take

$$\mathcal{F} = \sigma(X_1, X_2, \dots)$$

i.e. $\mathcal{F}$ is smallest σ-field with respect to which all the X_n are random variables. Then

$$\begin{aligned}
\sigma(X_1) &= \{\emptyset, \{H**\ast\dots\}, \{T**\ast\dots\}, \Omega\} \\
\sigma(X_1, X_2) &= \sigma(\{\{HH**\dots\}, \{TH**\dots\}, \{HT**\dots\}, \{TT**\dots\}\}) \\
&= \{\emptyset, \{HH**\dots\}, \{TH**\dots\}, \{HT**\dots\}, \{TT**\dots\}, \\
&\quad \{H**\ast\dots\}, \{T**\ast\dots\}, \{*H**\dots\}, \{*T**\dots\}, \left\{ \begin{matrix} HH \\ TT \end{matrix} \; **\dots, \right\}, \left\{ \begin{matrix} HT \\ TH \end{matrix} \; **\dots, \right\}, \\
&\quad \{HH**\dots\}^c, \{TH**\dots\}^c, \{HT**\dots\}^c, \{TT**\dots\}^c, \Omega\},
\end{aligned}$$

where $*$ means that it can take on either H or T, so $\{H**\ast\dots\} = \{\omega : \omega_1 = H\}$.

With the information available to us in $\sigma(X_1, X_2)$, we can distinguish between ω's where the first (or second) outcomes are different. But if two ω's have the same first and second outcomes, they fall into exactly the same subset(s) of $\sigma(X_1, X_2)$. Consequently, if a random variable depends on anything more than the first and second outcomes, it will not be $\sigma(X_1, X_2)$ measurable.

It is not immediately clear if we can define a probability measure on $\mathcal{F}$! Since Ω is uncountable, we cannot follow the scheme in Section 1.3.1 and define $\mathbb{P}$ in terms of $\mathbb{P}[\omega]$ for each individual $\omega \in \Omega$. Equation (1.4) simply would not make sense; there is no such thing as an uncountable sum.

To define a probability measure in this case requires a significant amount of machinery, from measure theory and Lebesgue integration. It is outside of the scope of this course. For our purposes, whenever we need to use an infinite Ω you will be *given* a probability measure and some of its helpful properties. For example, in this case there exists a probability measure $\mathbb{P} : \mathcal{F} \to [0, 1]$ such that

- $\mathbb{P}[X_n = H] = \mathbb{P}[X_n = T] = \frac{1}{2}$ for all $n \in \mathbb{N}$.
- Each X_n is independent.

From this, you can work with $\mathbb{P}$ without having to know how $\mathbb{P}$ was constructed. You don't even need to know exactly which subsets of Ω are in $\mathcal{F}$, because Proposition 1.2.6 gives you access to plenty of random variables.

Remark 1.3.1 ($\star$) *In this case $\mathcal{F}$ is much smaller than $\mathcal{P}(\Omega)$. If we tried to take $\mathcal{F} = \mathcal{P}(\Omega)$, we would (after some significant effort) discover that there is no probability measure $\mathbb{Q} : \mathcal{P}(\Omega) \to [0, 1]$ satisfying the two conditions we wanted above for $\mathbb{P}$. This is irritating, and we just have to live with it.*

1.4 Analysis background

In this section we survey some material that you should already know from previous courses, and give a rough outline of some results that are closely related to the present course but are part of other areas of mathematics.

1.4.1 Convergence of random variables

Recall the different modes of convergence of random variables: almost sure, in probability, in distribution, and in L^1.

Also, recall the definitions of limsup and liminf. Let $(x_n, n \in \mathbb{N})$ be a sequence of real numbers,

$$\begin{aligned} \limsup x_n &= \inf_m \left\{ \sup_{n \geq m} x_n \right\} \in [-\infty, \infty] \\ \liminf x_n &= \sup_m \left\{ \inf_{n \geq m} x_n \right\} \in [-\infty, \infty] \end{aligned}$$

Recall also that $x_n \to x \in \mathbb{R}$ if and only if $\limsup_n x_n = \liminf_n x_n = x$.

The Borel-Cantelli lemmas are a tool for understanding the tail behaviour (which may or may not extend to convergence) of a sequence of events E_n. The key definitions are

$$\begin{aligned} \{E_n \text{ i.o.}\} = \{E_n, \text{ infinitely often}\} &= \bigcap_m \bigcup_{n \geq m} E_n = \{\omega : \omega \in E_n \text{ for infinitely many } n\} \\ \{E_n \text{ e.v.}\} = \{E_n, \text{ eventually}\} &= \bigcup_m \bigcap_{n \geq m} E_n = \{\omega : \omega \in E_n \text{ for all sufficiently large } n\}. \end{aligned}$$

The Borel-Cantelli lemmas, respectively, give conditions under which the probability of $\{E_n \text{ i.o.}\}$ is either 0 or 1. To be precise:

Proposition 1.4.1 (First Borel-Cantelli Lemma) *Let $(E_n)_{n \in \mathbb{N}}$ be a sequence of events and suppose $\sum_{n=1}^{\infty} \mathbb{P}[E_n] < \infty$. Then $\mathbb{P}[E_n \text{ i.o.}] = 0$.*

Proposition 1.4.2 (Second Borel-Cantelli Lemma) *Let $(E_n)_{n \in \mathbb{N}}$ be a sequence of independent events and suppose that $\sum_{n=1}^{\infty} \mathbb{P}[E_n] = \infty$. Then $\mathbb{P}[E_n \text{ i.o.}] = 1$.*

1.4.2 Integration and Expectation

The expectation of a discrete random variable X that takes on values in $\{x_i : i \in \mathbb{N}\}$ is given by

$$\mathbb{E}(X) = \sum_{x_i} x_i \mathbb{P}(X = x_i).$$

For a continuous random variable, the expectation uses an integral against the probability density function. This is not very elegant as a definition, since it requires different formulae under only slightly different circumstances. With *Lebesgue integration*, the expectation can be defined elegantly using a single definition for both discrete and continuous (or other even more exotic) random variables. The detail of such a definition is beyond the scope of this course, but we will give a rough idea of how it is done, along with some properties of the construction.

Definition 1.4.3 ($\star$) *A simple random variable X is one that takes on finitely many positive values, i.e. $X = \sum_{k=1}^{n} a_k \mathbb{1}_{A_k}$ where $a_k \in (0,\infty)$ and $A_k \in \mathcal{F}$.*

Definition 1.4.4 ($\star$) *The definition of $\mathbb{E}$ comes in four incremental steps.*

1. ***Indicator functions:*** *for a measurable set A, define $\mathbb{E}(\mathbb{1}_A) := \mathbb{P}(A)$.*

2. ***Simple RVs:*** *Write $X = \sum_{k=1}^{n} a_k \mathbb{1}_{A_k}$ for measurable sets $A_1, \ldots, A_k$ and $a_1, \ldots, a_k > 0$. Define $\mathbb{E}[X] = \sum_{k=1}^{n} a_k \mathbb{P}[A_k]$. Note that this agrees with the 'usual' definition of $\mathbb{E}$ for discrete random variables.*

3. ***Non-negative RVs:*** *Find (somehow...) a sequence of simple random variables X_n such that $X_n \nearrow X$ almost surely and define $\mathbb{E}[X] =\uparrow \lim \mathbb{E}[X_n]$ (which might be $+\infty$).*

4. ***'Integrable' RVs:*** *Write $X = X^+ - X^-$ where X^+ and X^- are non-negative random variables. If both $E[X^+]$ and $E[X^-] < \infty$ then we define $\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-]$.*

This definition leaves many questions unanswered; for example several different approximating sequences could be used in the third step and might potentially define different values of $\mathbb{E}[X]$. However, it can be shown that they do not. The above procedure is well defined.

For purposes of this course, it is useful to have seen the above steps, but there is no need to remember them in detail. What you should know is: As a result of the above definition, $\mathbb{E}[X]$ is defined for all X such that (1) $\mathbb{E}[|X|] < \infty$, or (2) $X \geq 0$. You may assume that all the 'usual' properties of $\mathbb{E}$ hold (linearity, positivity, etc). Also,

Definition 1.4.5 *Let $p \in [1,\infty)$. We say say that $X \in L^p$ if $\mathbb{E}[|X|^p] < \infty$.*

You should also know the following convergence theorems. They provide conditions for showing L^1 convergence given a.s. convergence. They are some of the most useful tools in probability theory.

Proposition 1.4.6 (Fatou's Lemma) *If $X_n \geq 0$, then $\mathbb{E}[\liminf X_n] \leq \liminf \mathbb{E}[X_n]$.*

Proposition 1.4.7 (Monotone Convergence Theorem) *If $X_n \uparrow X$ a.s. and $X_n \geq 0$ for all n, then $\mathbb{E}(X_n) \uparrow \mathbb{E}(X)$.*

Proposition 1.4.8 (Dominated Convergence Theorem) *If $X_n \to X$ a.s., $|X_n| < Y$ for all n, and $\mathbb{E}[Y] < \infty$, then $\mathbb{E}[X_n] \to \mathbb{E}[X]$.*

A special case of the Dominated Convergence Theorem is the case where there is some (deterministic) $M \in \mathbb{R}$ such that $|X_n| \leq M$ almost surely for all n. In this case, which is sometimes known as the Bounded Convergence Theorem, we may simply take $Y = M \in \mathbb{R}$.

Lastly, we have an inequality which relates $\mathbb{E}$ to convex functions. Recall that a function $c : \mathbb{R} \to \mathbb{R}$ is convex if the set $\{(x,y) \in \mathbb{R}^2\,;\, y \geq c(x)\}$ is a convex set.

Proposition 1.4.9 (Jensen's Inequality) *Let $X \in L^1$, and suppose $c : \mathbb{R} \to \mathbb{R}$ is a convex function. Suppose also that $\mathbb{E}[c(X)] < \infty$. Then*

$$c(\mathbb{E}[X]) \leq \mathbb{E}[c(X)].$$

1.5 Conditional Expectation

Suppose X and Z are random variables that take on only finitely many values $\{x_1, \ldots, x_m\}$ and $\{z_1, \ldots, z_n\}$, respectively. In earlier courses, 'conditional expectation' was defined as follows:

$$\begin{array}{rcl} \mathbb{P}[X = x_i \,|\, Z = z_j] &=& \mathbb{P}[X = x_i, Z = z_j]/\mathbb{P}[Z = z_j] \\ \mathbb{E}[X \,|\, Z = z_j] &=& \sum_i x_i \mathbb{P}[X = x_i \,|\, Z = z_j] \\ Y = \mathbb{E}[X \,|\, Z] \text{ where:} && \text{if } Z(\omega) = z_j, \text{ then } Y(\omega) = \mathbb{E}[X \,|\, Z = z_j] \end{array} \tag{1.5}$$

You will also have seen a second definition, using probability density functions, for continuous random variables. These definition are problematic, for several reasons, chiefly (1) its not immediately clear (or simple to write down) how the two definitions interact and (2) we don't want to be restricted to handling only discrete or continuous random variables.

In this section, we define the conditional expectation of random variables using σ-fields. In this setting we are able to give a unified definition which is valid for general random variables. The definition is originally due to Kolmogorov (in 1933), and is sometimes referred to as Kolmogorov's conditional expectation. It is one of the most important concepts in modern probability theory.

Conditional expectation is a mathematical tool with the following function. We have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $X : \Omega \to \mathbb{R}$. However, $\mathcal{F}$ is large and we want to work with a sub-σ-algebra $\mathcal{G}$, instead. As a result, we want to have a random variable Y such that

1. Y is $\mathcal{G}$-measurable
2. Y is 'the best' way to approximate X with a $\mathcal{G}$-measurable random variable

The second statement on this wish-list does not fully make sense; there are many different ways in which we could compare X to potential Y.

Why might we want to do this? Imagine we are conducting an experiment in which we gradually gain information about the result X. This corresponds to gradually seeing a larger and larger $\mathcal{G}$, with access to more and more information. At all times we want to keep a prediction of what the future looks like, based on the currently available information. This prediction is Y.

It turns out there is *only one* natural way in which to realize our wish-list (which is convenient, and somewhat surprising). It is the following:

Theorem 1.5.1 (Conditional Expectation) *Let X be an L^1 random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{G}$ be a sub-σ-field of $\mathcal{F}$. Then there exists a random variable $Y \in L^1$ such that*

1. *Y is $\mathcal{G}$-measurable,*
2. *for every $G \in \mathcal{G}$, we have $\mathbb{E}[Y \mathbb{1}_G] = \mathbb{E}[X \mathbb{1}_G]$.*

Moreover, if $Y' \in L^1$ is a second random variable satisfying these conditions, $\mathbb{P}[Y = Y'] = 1$.

The first and second statements here correspond respectively to the items on our wish-list.

Definition 1.5.2 *We refer to Y as a version of the conditional expectation of X given $\mathcal{G}$.*

Since any two such Y are almost surely equal so we sometimes refer to Y simply as *the* conditional expectation of X. This is a slight abuse of notation, but it is commonplace and harmless.

Proof of Theorem 1.5.1 is beyond the scope of this course. Loosely speaking, there is an abstract recipe which constructs $\mathbb{E}[X|\mathcal{G}]$. It begins with the random variable X, and then averages out over all the information that is not accessible to $\mathcal{G}$, leaving only as much randomness as $\mathcal{G}$ can support, resulting in $\mathbb{E}[X|\mathcal{G}]$. In this sense the map $X \mapsto \mathbb{E}[X|\mathcal{G}]$ simplifies (i.e. reduces the amount of randomness in) X in a very particular way, to make it $\mathcal{G}$ measurable.

It is important to remember that $\mathbb{E}[X|\mathcal{G}]$ is (in general) a random variable. It is also important to remember that the two objects

$$\mathbb{E}[X|\mathcal{G}] \quad \text{and} \quad \mathbb{E}[X|Z=z]$$

are quite different. They are both useful. We will explore the connection between them in Section 1.5.1. Before doing so, let us look at a basic example.

Let X_1, X_2 be independent random variables such that $\mathbb{P}[X_i = 0] = \mathbb{P}[X_i = 1] = \frac{1}{2}$. Set $\mathcal{F}_1 = \sigma(X_1)$. We will show that

$$\mathbb{E}[X_1 + X_2|\sigma(X_1)] = X_1. \tag{1.6}$$

To do so, we should check that X_1 satisfies the two conditions in Theorem 1.5.1, with $X = X_1 + X_2$, $Y = X_1$ and $\mathcal{G} = \sigma(X_1)$. The first condition is immediate, since by Lemma 1.2.5 X_1 is $\sigma(X_1)$-measurable. To see the second condition, let $G \in \sigma(X_1)$. Then $\mathbb{1}_G \in \sigma(X_1)$ and $X_2 \in \sigma(X_2)$, which are independent, so 1_G and X_2 are independent. Hence

$$\begin{aligned}
\mathbb{E}[(X_1 + X_2)\mathbb{1}_G] &= \mathbb{E}[X_1\mathbb{1}_G] + \mathbb{E}[1_G]\mathbb{E}[X_2] \\
&= \mathbb{E}[X_1\mathbb{1}_G] + \mathbb{P}[G].0 \\
&= \mathbb{E}[X_1\mathbb{1}_G].
\end{aligned}$$

Thus, by Theorem 1.5.1 we have $X_1 = \mathbb{E}[X_1 + X_2|\sigma(X_1)]$ almost surely. The intuition for this, which is plainly visible in our calculation, is that X_2 is independent of $\sigma(X_1)$ so, thinking of conditional expectation as an operation which averages out all randomness in $X = X_1 + X_2$ that is not $\mathcal{G} = \sigma(X_1)$ measurable, we would average out X_2 completely i.e. $\mathbb{E}[X_2] = 0$.

1.5.1 Relationship to the naive definition

Conditional expectation extends the 'naive' definition of (1.5). Essentially the relationship is that

what you have previously called $\mathbb{E}[X|Z]$, we now call $\mathbb{E}[X|\sigma(Z)]$.

In this course we always write $\mathbb{E}[X|\sigma(Z)]$ to make clear that we are taking conditional expectation with respect to a σ-field. Naturally, the 'new' conditional expectation is much more general (and it is what the theory of martingales requires us to have), but we should still take the time to relate it to the naive definition.

To see the connection, we focus on the case where X, Z are random variables with finite sets of values $\{x_1, \dots, x_n\}$, $\{z_1, \dots, z_m\}$. Let Y be the naive version of conditional expectation defined in (1.5). That is,

$$Y(\omega) = \sum_j \mathbb{1}\{Z(\omega) = z_j\}\mathbb{E}[X|Z = z_j].$$

We can use Theorem 1.5.1 to check that, in fact, Y is a version of $\mathbb{E}[X|\sigma(Z)]$. We want to check that Y satisfies the two properties listed in Theorem 1.5.1.

- Since Z only takes finitely many values $\{z_1, \dots, z_m\}$, from the above equation we have that Y only takes finitely many values. These values are $\{y_1, \dots, y_m\}$ where $y_j = \mathbb{E}[X|Z = z_j]$. We note

$$\begin{aligned} Y^{-1}(y_j) &= \{\omega \in \Omega\,;\, Y(\omega) = \mathbb{E}[X|Z = z_j]\} \\ &= \{\omega \in \Omega\,;\, Z(\omega) = z_j\} \\ &= Z^{-1}(z_j) \in \sigma(Z). \end{aligned}$$

  Hence, by Lemma 1.2.2, Y is $\sigma(Z)$-measurable.

- We can calculate

$$\begin{aligned} \mathbb{E}[Y\mathbb{1}\{Z = z_j\}] &= y_j \mathbb{E}[\mathbb{1}\{Z = z_j\}] \\ &= y_j \mathbb{P}[Z = z_j] \\ &= \sum_i x_i \mathbb{P}[X = x_i | Z = z_j] \mathbb{P}[Z_j = z_j] \\ &= \sum_i x_i \mathbb{P}[X = x_i \text{ and } Z = z_j] \\ &= \sum_{i,j} x_i \mathbb{1}\{Z = z_j\} \mathbb{P}[X = x_i \text{ and } Z = z_j] \\ &= \mathbb{E}[X\mathbb{1}\{Z = z_j\}]. \end{aligned}$$

  ($\star$) Properly, to check that Y satisfies the second property in Theorem 1.5.1, we need to check $\mathbb{E}[Y\mathbb{1}_G] = \mathbb{E}[X\mathbb{1}_G]$ for a general $G \in \sigma(Z)$ and not just $G = \{Z = z_j\}$. However, for reasons beyond the scope of this course, in this case (thanks to the fact that Z is finite) its enough to consider only G of the form $\{Z = z_j\}$.

Therefore, we have $Y = \mathbb{E}[X|\sigma(Z)]$ almost surely. In this course we favour writing $\mathbb{E}[X|\sigma(Z)]$ instead of $\mathbb{E}[X|Z]$, to make it clear that we are looking at conditional expectation with respect to a σ-field. We could equally think of X_1 as being our best guess for $X_1 + X_2$, given only information in $\sigma(X_1)$, since $\mathbb{E}[X_2] = 0$. In general, guessing $\mathbb{E}[X|\mathcal{G}]$ is not so easy!

1.5.2 Properties of conditional expectation

In all but the easiest cases, calculating conditional expectations explicitly from Theorem 1.5.1 is not feasible. Instead, we are able to work with them via a set of useful properties, provided by the following proposition.

Proposition 1.5.3 *Let $\mathcal{G}, \mathcal{H}$ be sub-σ-fields of $\mathcal{F}$ and $X \in L^1$.*
(a) $\mathbb{E}[\mathbb{E}[X \,|\, \mathcal{G}]] = \mathbb{E}[X]$ *and, in particular* $\mathbb{E}[X \,|\, \{\emptyset, \Omega\}] = \mathbb{E}[X]$*,*
(b) If X is $\mathcal{G}$-measurable, then $\mathbb{E}[X \,|\, \mathcal{G}] = X$*.*
(Linearity) $\mathbb{E}[a_1X_1 + a_2X_2 \,|\, \mathcal{G}] = a_1\mathbb{E}[X_1 \,|\, \mathcal{G}] + a_2\mathbb{E}[X_2 \,|\, \mathcal{G}]$.
(Positivity) *If $X \geq 0$, then* $\mathbb{E}[X|\mathcal{G}] \geq 0$.
(Taking out what is known) *If Z is $\mathcal{G}$ measurable, then* $\mathbb{E}[ZX \,|\, \mathcal{G}] = Z\mathbb{E}[X \,|\, \mathcal{G}]$.
(Tower) *If $\mathcal{H} \subset \mathcal{G}$ then* $\mathbb{E}[\mathbb{E}[X \,|\, \mathcal{G}] \,|\, \mathcal{H}] = \mathbb{E}[X \,|\, \mathcal{H}]$.
(Independence) *If $\mathcal{H}$ is independent of $\sigma(X, \mathcal{G})$, then* $\mathbb{E}[X \,|\, \sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[X \,|\, \mathcal{G}]$

PROOF: We will prove the first three properties; proof of the others is beyond the scope of the course.

(a): See Problem Sheet 1.

(b): We check the two conditions of Theorem 1.5.1. Clearly $\mathbb{E}[X|\mathcal{G}]$ is $\mathcal{G}$-measurable. Also, for any $G \in \mathcal{G}$ Proposition 1.2.6 tells us that $\mathbb{E}[X|\mathcal{G}]\mathbb{1}_G$ is $\mathcal{G}$-measurable. Since $\mathbb{E}[X\mathbb{1}_G] = \mathbb{E}[X\mathbb{1}_G]$, we are done.

(Linearity): Again, we check the two conditions of Theorem 1.5.1. By Proposition 1.2.6, $a_1\mathbb{E}[X_1 \,|\, \mathcal{G}] + a_2\mathbb{E}[X_2 \,|\, \mathcal{G}]$ is $\mathcal{G}$-measurable. Then, by linearity of $\mathbb{E}$, for any $G \in \mathcal{G}$,

$$\begin{aligned}\mathbb{E}\big[(a_1\mathbb{E}[X_1 \,|\, \mathcal{G}] + a_2\mathbb{E}[X_2 \,|\, \mathcal{G}])\mathbb{1}_G\big] &= a_1\mathbb{E}[\mathbb{E}[X_1|\mathcal{G}]\mathbb{1}_G] + a_2\mathbb{E}[\mathbb{E}[X_2|\mathcal{G}]\mathbb{1}_G] \\ &= a_1\mathbb{E}[X_1\mathbb{1}_G] + a_2\mathbb{E}[X_2\mathbb{1}_G] \\ &= \mathbb{E}[(a_1X_1 + A_2X_2)\mathbb{1}_G]\end{aligned}$$

Here, the second line follows by the same method as we used in (b), and the third line follows from the second by linearity of $\mathbb{E}$. ■

Note that Property (b) is really a special case of the 'taking out what is known' rule (take $Z = 1$). It is also useful to note a special case of the independence property: take $\mathcal{G} = \{\emptyset, \Omega\}$ and, using (a), we have that if X is independent of $\mathcal{H}$, $\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[X]$.

Remark 1.5.4 *Although we have not proved the last four properties in Proposition 1.5.3, they are intuitive properties for conditional expectation to have.*

For example, in the taking out what is known property, we can think of Z as already being simple enough to be $\mathcal{G}$ measurable, so we'd expect that taking conditional expectation with respect to $\mathcal{G}$ doesn't need to affect it.

In the tower property for $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]$, we start with X, simplify it to be $\mathcal{G}$ measurable and simplify it to be $\mathcal{H}$ measurable. But since $\mathcal{H} \subseteq \mathcal{G}$, we might as well have just simplified X enough to be $\mathcal{H}$ measurable in a single step, which would be $\mathbb{E}[X|\mathcal{H}]$.

Positivity is clearly intuitive. You might like to think of an explanation, in the same non-rigorous style, for the independence property.

The conditional expectation $Y = \mathbb{E}[X \,|\, \mathcal{G}]$ is the 'best least-squares estimator' of X, based on the information available in $\mathcal{G}$. We can state this rigorously and use our toolkit from Proposition 1.5.3 prove it. This demonstrates another way in which Y is 'the best' $\mathcal{G}$-measurable approximation to X.

Lemma 1.5.5 *Let $\mathcal{G}$ be a sub-σ-field of $\mathcal{F}$. Let X be an $\mathcal{F}$-measurable random variable and let $Y = \mathbb{E}[X|\mathcal{G}]$. Suppose that Y' is a $\mathcal{G}$-measurable, random variable. Then*

$$\mathbb{E}[(X - Y)^2] \le \mathbb{E}[(X - Y')^2].$$

PROOF: We note that

$$\begin{aligned}\mathbb{E}[(X - Y')^2] &= \mathbb{E}[(X - Y + Y - Y')^2] \\ &= \mathbb{E}[(X - Y)^2] + 2\mathbb{E}[(X - Y)(Y - Y')] + \mathbb{E}[(Y - Y')^2]. \qquad (1.7)\end{aligned}$$

In the middle term above, we can write

$$\begin{aligned}\mathbb{E}[(X-Y)(Y-Y')] &= \mathbb{E}[\mathbb{E}[(X-Y)(Y-Y')|\mathcal{G}]] \\ &= \mathbb{E}[(Y-Y')\mathbb{E}[X-Y|\mathcal{G}]] \\ &= \mathbb{E}[(Y-Y')(\mathbb{E}[X|\mathcal{G}]-Y)].\end{aligned}$$

Here, in the first step we used the 'tower' rule, in the second step we used Proposition 1.2.6 to tell us that $Y - Y'$ is $\mathcal{G}$-measurable, followed by the 'taking out what is known' rule. In the final step we used linearity and property (b). Since $\mathbb{E}[X|\mathcal{G}] = Y$ almost surely, we obtain that $\mathbb{E}[(X-Y)(Y-Y')] = 0$. Hence, since $\mathbb{E}[(Y-Y')^2] \geq 0$, from (1.7) we obtain $\mathbb{E}[(X-Y')^2] \geq \mathbb{E}[(X-Y)^2]$. ■

The equivalents of the convergence theorems for $\mathbb{E}$ and Jensen's inequality also hold for conditional expectation.

Proposition 1.5.6 *Let $\mathcal{G}$ be a sub-σ-field of $\mathcal{F}$ and let (X_n) be random variables.*
(MON) *If $0 \leq X_n \uparrow X$ almost surely, then $\mathbb{E}[X_n \,|\, \mathcal{G}] \uparrow \mathbb{E}[X_n \,|\, \mathcal{G}]$.*
(FATOU) *If $0 \leq X_n$, then $\mathbb{E}[\liminf X_n \,|\, \mathcal{G}] \leq \liminf \mathbb{E}[X_n \,|\, \mathcal{G}]$.*
(DOM) *If $|X_n| \leq V$, $X_n \to X$ almost surely, and $\mathbb{E}[V] < \infty$, then $\mathbb{E}[X_n \,|\, \mathcal{G}] \to \mathbb{E}[X \,|\, \mathcal{G}]$.*
(JENSEN) *If f is convex, then $\mathbb{E}[f(X) \,|\, \mathcal{G}] \geq f[\mathbb{E}(X \,|\, \mathcal{G})]$.*

Proof of Proposition 1.5.6 is beyond the scope of this course.

Chapter 2

Martingales

In this chapter we introduce martingales. We prove the Optional Stopping Theorem for martingales, and use it to analyse some common stochastic processes.

As usual, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We adopt the convention that, whenever we say 'X is a random variable', we mean that X is $\mathcal{F}$-measurable.

We refer to a sequence of random variables $(S_n)_{n=0}^{\infty}$ as a stochastic process. In this course we only deal with discrete time stochastic processes. We say that a stochastic process (S_n) is *bounded* if there exists $c \in \mathbb{R}$ such that $|S_n(\omega)| \leq c$ for all n, ω.

2.1 Filtrations and martingales

We have previously discussed the idea of gradually learning more and more information about the outcome of some experiment, through seeing the information visible from gradually larger σ-fields. We formalize this concept as follows.

Definition 2.1.1 *A countable set of σ-fields $(\mathcal{F}_n)_{n=0}^{\infty}$ is known as a filtration if $\mathcal{F}_0 \subset \mathcal{F}_1 \ldots \subset \mathcal{F}$.*

Definition 2.1.2 *We say that a stochastic process $X = (X_n)$ is adapted to the filtration $(\mathcal{F}_n)$ if, for all n, X_n is $\mathcal{F}_n$ measurable.*

We should think of the filtration $\mathcal{F}_n$ as telling us which information we have access too at time $t = n$. Thus, an adapted process is a process whose random value we know at all times $n \in \mathbb{N}$.

We are now ready to give the definition of a martingale.

Definition 2.1.3 *A process $M = (M_n)_{n=0}^{\infty}$ is martingale if*

1. *if (M_n) is adapted,*
2. *$M_n \in L^1$ for all n,*
3. *$\mathbb{E}[M_{n+1} \,|\, \mathcal{F}_{n-1}] = M_n$ almost surely, for all n.*

We say that M is a submartingale if, instead of 3, we have $\mathbb{E}[M_n \,|\, \mathcal{F}_{n-1}] \geq M_{n-1}$ almost surely. We say that M is a supermartingale if, instead of 3, we have $\mathbb{E}[M_n \,|\, \mathcal{F}_{n-1}] \leq M_{n-1}$ almost surely.

Remark 2.1.4 *The second condition in Definition 2.1.3 is needed for the third to make sense.*

A martingale is the mathematical idealization of a fair game. It is best to understand what we mean by this through an example.

Let (X_n) be a sequence of i.i.d. random variables such that $\mathbb{P}[X_i = 1] = \mathbb{P}[X_i = -1] = \frac{1}{2}$. Define $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ and $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Then $(\mathcal{F}_n)$ is a filtration. Define

$$S_n = \sum_{i=1}^{n} X_n$$

(and $S_0 = 0$). We claim that S_n is a martingale. To see this, we check the three properties in the definition. (1) Since $X_1, X_2, \ldots, X_n \in \sigma(X_1, \ldots, X_n)$ we have that $S_n \in \mathcal{F}_n$ for all $n \in \mathbb{N}$. (2) Since $|S_n| \leq n$ for all $n \in \mathbb{N}$, $\mathbb{E}[|S_n|] \leq n$ for all n, so $S_n \in L^1$ for all n. (3) We have

$$\begin{aligned}
\mathbb{E}\left[S_{n+1} \,|\, \mathcal{F}_{n-1}\right] &= \mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n] + \mathbb{E}[S_n \,|\, \mathcal{F}_n] \\
&= \mathbb{E}[X_{n+1}] + S_n \\
&= S_n.
\end{aligned}$$

Here, in the first line we used the linearity of conditional expectation. To deduce the second line we used the relationship between independence and conditional expectation (for the first term) and the tower rule (for the second term). To deduce the final line we used that $\mathbb{E}[X_{n+1}] = 0$.

We can think of S_n as a game in the following way. At each time $n = 1, 2, \ldots$ we toss a coin. We the n^{th} round if the coin is heads, and lose if it is tails. Each time we win we score 1, each time we lose we score -1. Thus, S_n is our score after n rounds.

At time n we have seen the result of rounds $1, 2, \ldots, n$, so the information we currently have access to is given by $\mathcal{F}_n$. This means that at time n we know $S_1, \ldots, S_n$. But we don't know S_{n+1}, because S_{n+1} is not $\mathcal{F}_n$-measurable. However, using our current information we can make our best guess at what S_{n+1} will be, which naturally is $\mathbb{E}[S_{n+1}|\mathcal{F}_n]$. Since the game is fair, in the future, on average we do not expect to win more than we lose, that is $\mathbb{E}[S_{n+1}|\mathcal{F}_n] = S_n$.

Of course, S_n is a simple random walk. In this chapter we will see many examples of martingales, and we will gradually build up an intuition for how to recognize a martingale. There is, however, one easy sufficient (but not necessary) condition under which we can recognize that a stochastic process is not a martingale.

Lemma 2.1.5 *Let $(\mathcal{F}_n)$ be a filtration and suppose that $X = (X_n)$ is an $\mathcal{F}_n$-martingale. Then*

$$\mathbb{E}[X_n] = \mathbb{E}[X_1]$$

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

Proof: We have $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n$. Taking expectations and using property (a) of Proposition 1.5.3, we have $\mathbb{E}[X_{n+1}] = \mathbb{E}[X_n]$. The result follows by a trivial induction. ■

Suppose, now, that (X_n) is an i.i.d. sequence of random variables such that $\mathbb{P}[X_i = 2] = \mathbb{P}[X_i = -1] = \frac{1}{2}$. Note that $\mathbb{E}[X_n] > 0$. Define S_n and $\mathcal{F}_n$ as before. Now, $\mathbb{E}[S_n] = \sum_1^n \mathbb{E}[X_n]$, which is not constant, so S_n is not a martingale.

However, as before, S_n is $\mathcal{F}_n$-measurable, and $|S_n| \leq 2n$ so $S_n \in L^1$, essentially as before. We have

$$\begin{aligned}
\mathbb{E}[S_{n+1}|\mathcal{F}_n] &= \mathbb{E}[X_{n+1}|\mathcal{F}_n] + \mathbb{E}[S_n|\mathcal{F}_n] \\
&= \mathbb{E}[X_{n+1}] + S_n \\
&\geq S_n.
\end{aligned}$$

Hence S_n is a submartingale.

Note that submartingales, on average, increase, whereas supermartingale, on average, decrease. As first glance this is the wrong way round. There is a good reason for this, but we won't cover it in this course.

Remark 2.1.6 *Our definition of a filtration and a martingale both make sense if we look at only a finite set of times $n = 1, \dots, N$. We sometimes also use the terms filtration and martingale in this situation.*

We end this section with two more important examples of martingales. You can check the conditions yourself.

Example 2.1.7 *Let (X_n) be a sequence of i.i.d. random variables such that $\mathbb{E}[X_n] = 1$ for all n, and there exists $c \in \mathbb{R}$ such that $|X_n| \leq c$ for all n. Define $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$. Then*

$$M_n = \prod_{i=1}^{n} X_n$$

is a martingale.

Example 2.1.8 *Let $Z \in L^1$ be a random variable and let $(\mathcal{F}_n)$ be a filtration. Then*

$$M_n = \mathbb{E}[Z|\mathcal{F}_n]$$

is a martingale.

2.2 The Optional Stopping Theorem

It is inconvenient to continually write '$(\mathcal{F}_n)$ martingale' instead of simply 'martingale', and in many situations it is clear which filtration we mean to use. We will often omit the '$(\mathcal{F}_n)$' for this reason.

Definition 2.2.1 *A quadruplet $(\Omega, \mathcal{F}, (\mathcal{F}_n), \mathbb{P})$ is said to be a filtered space if $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space equipped with a filtration $(\mathcal{F}_n)$.*

If a filtration and probability space are not explicitly specified, it is a standard convention in martingale theory to assume that we are working over a filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_n), \mathbb{P})$, and that $(\mathcal{F}_n)$ is the filtration with respect to which our martingales are martingales.

The following definition is similar to the definition of a filtration.

Definition 2.2.2 *A process $C = (C_n)_{n=1}^{\infty}$ is said to be previsible if, for all n, C_n is $\mathcal{F}_{n-1}$ measurable.*

The difference is that, in an adpated process (X_n), the value of X becomes fully known to us, using information from $\mathcal{F}_n$, at time n. In a previsible process (C_n), the value of C_n becomes known at time $n-1$.

If M is a stochastic process and C is previsible process, we define the *martingale transform* of S by M

$$(C \circ M)_n = \sum_{i=1}^{n} C_i(M_i - M_{i-1}).$$

Here, by convention, we set $(C \circ M)_0 = 0$.

Remark 2.2.3 ($\star$) *The process $C \circ M$ is the discrete time analogue of the stochastic integral $\int C \; dM$, which is covered in the fourth year stochastic analysis course. In fact, if S is a deterministic sequence, $C \circ S$ is a discrete analogue of the Lebesgue (or Riemann) integral.*

If M is a martingale, the process $(C \circ M)_n$ can be thought of as our winnings after n plays of a game. Here, at round i, a bet of C_i is made, and the change to our resulting wealth is $C_i(M_i - M_{i-1})$. For example, if $C_i \equiv 1$ and M_n is the simple random walk $M_n = \sum_1^n X_i$ then $M_i - M_{i-1} = X_{i-1}$, so we win/lose each round with even chances; we bet 1 on each round, if we win we get our money back doubled, if we lose we get nothing back.

Theorem 2.2.4 *If M is a martingale and C is previsible and bounded, then $(C \circ M)_n$ is also a martingale.*

Similarly, if M is a supermartingale (resp. submartingale), and C is previsible, bounded and non-negative, then $(C \circ M)_n$ is also a supermartingale martingale (resp. submartingale).

PROOF: Let M be a martingale. Write $Y = C \circ M$. We have $C_n \in \mathcal{F}_{n-1}$ and $X_n \in \mathcal{F}_n$, so Proposition 1.2.6 implies that $Y_n \in m\mathcal{F}_n$. Since $|C| \leq c$ for some c, we have

$$\mathbb{E}|Y_n| \leq \sum_{k=1}^{n} \mathbb{E}|C_k(M_k - M_{k-1})| \leq c \sum_{k=1}^{n} \mathbb{E}|M_k| + \mathbb{E}|M_{k-1}| < \infty.$$

So $Y_n \in L^1$. Since C_n is $\mathcal{F}_{n-1}$-measurable, by linearity of conditional expectation, the taking out what is known rule and the martingale property of M, we have

$$\begin{aligned}\mathbb{E}[Y_n \,|\, \mathcal{F}_{n-1}] &= \mathbb{E}[Y_{n-1} + C_n(M_n - M_{n-1}) \,|\, \mathcal{F}_{n-1}] \\ &= Y_{n-1} + C_n\mathbb{E}[M_n - M_{n-1} \,|\, \mathcal{F}_{n-1}] \\ &= Y_{n-1} + C_n(\mathbb{E}[M_n|\mathcal{F}_{n-1}] - M_{n-1}) \\ &= Y_{n-1}.\end{aligned}$$

Hence Y is a martingale.

The argument is easily adapted to prove the second statement, e.g. for a supermartingale M, $\mathbb{E}[M_n - M_{n-1} \,|\, \mathcal{F}_{n-1}] \leq 0$. Note that in these cases it is important that C is non-negative. ■

Definition 2.2.5 *A map $T : \Omega \to \{0, 1, 2, \ldots, \infty\}$ is called a $(\mathcal{F}_n)$ stopping time if, for all n, $\{T = n\}$ is $\mathcal{F}_n$ measurable.*

Equivalently, we could say that T is a stopping time if $\{T \leq n\}$ is $\mathcal{F}_n$ measurable for all n. To see why, recall the definition of a σ-algebra and note that

$$\{T \leq n\} = \bigcup_{i \leq n} \{T = i\}, \quad \{T = n\} = \{T \leq n\} \setminus \{T \leq n-1\}.$$

A stopping time is a random time with the property that, if we have only information from $\mathcal{F}_n$ accessible to us at time n, we are able to decide at any n whether or not T has already happened.

The simplest example of a stopping time is a constant random variable; if $t \in \mathbb{R}$ and $T = t$ then $\{T = n\} = \{t = n\}$, which is empty if $t \neq n$ and equal to Ω if $t = n$. However, in general stopping times are useful because they allow us to describe the random behaviour of stochastic processes.

Example 2.2.6 *Let $S_n = \sum_{i=1}^n X_i$ be the simple symmetric random walk, with $\mathbb{P}[X_i = 1] = \mathbb{P}[X_i = -1] = \frac{1}{2}$. Let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Then, for any $a \in \mathbb{N}$, the time*

$$T = \inf\{n \geq 0\,;\, S_n = a\},$$

which is the first time S_n takes the value a, is a stopping time. It is commonly called the hitting time of a. To see that T is a stopping time we note that

$$\begin{aligned}\{T = n\} &= \{S_n = a\} \cap \{S_i \neq a \textit{ for all } i < a\} \\ &= S_n^{-1}(a) \cap \left(\bigcap_{i=0}^{n-1} \Omega \setminus S_n^{-1}(a) \right).\end{aligned}$$

Since S_n is $\mathcal{F}_n$ measurable (by Proposition 1.2.6), the above equation shows that $\{T = n\} \in \mathcal{F}_n$.

Lemma 2.2.7 *Let S and T be stopping times with respect to the filtration $(\mathcal{F}_n)$. Then $S \wedge T$ is also a $(\mathcal{F}_n)$ stopping time.*

PROOF: Problem sheet 2, question 4. ■

If T is a stopping time and M is a stochastic process, we define M^T to be the process

$$M_n^T = M_{n \wedge T}.$$

Here $a \wedge b$ denotes the minimum of a and b. To be precise, this means that $M_n^T(\omega) = M_{n\wedge T(\omega)}(\omega)$ for all $\omega \in \Omega$. In Example 2.2.6, S^T would be the random walk S which is stopped (i.e. never moves again) when (if!) it reaches state a.

Lemma 2.2.8 *Let M_n be a martingale (resp. supmartingale, supermartingale) and let T be a stopping time. Then M^T is also a martingale (resp. supmartingale, supermartingale).*

PROOF: Let $C_n := \mathbb{1}\{T \geq n\}$. Note that $\{C_n = 0\} = \{T \leq n-1\}$ and $\{C_n = 1\} = \Omega \setminus \{C_n = 0\}$, so C_n is $\mathcal{F}_{n-1}$ measurable by Proposition 1.2.2. Hence (C_n) is a previsible process. Moreover,

$$(C \circ M)_n = \sum_{k=1}^{n} \mathbb{1}_{k \leq T}(M_k - M_{k-1}) = \sum_{k=1}^{n\wedge T}(M_k - M_{k-1}) = M_{T\wedge n} - M_{T\wedge 0}.$$

Hence, if M is a martingale (resp. submartingale, supermartingale), $C \circ M$ is also a martingale (resp. submartingale, supermartingale) by Theorem 2.2.4. ∎

Theorem 2.2.9 (Doob's Optional Stopping Theorem) *Let M be martingale (resp. submartingale, supermartingale) and let T be a stopping time. Then*

$$\mathbb{E}[X_T] = \mathbb{E}[X_0]$$

(resp. $\geq$, $\leq$) if any one of the following conditions hold:

a. T is bounded.

b. M is bounded and $\mathbb{P}[T < \infty] = 1$.

c. $\mathbb{E}[T] < \infty$ and there exists $c \in \mathbb{R}$ such that $|M_n - M_{n-1}| \leq c$ for all n.

PROOF: We'll prove this for the supermartingale case. The submartingale case then follows by considering $-M$, and the martingale case follows since martingales are both supermartingales and submartingales.

Note that

$$\mathbb{E}[M_{n\wedge T} - M_0] \leq 0,$$

because M^T is a supermartingale, by Lemma 2.2.8. For (a), we take $n = \sup_\omega T(\omega)$ and the conclusion follows. For (b), we use the Bounded Convergence Theorem and let $n \to \infty$ (note that $n \wedge T(\omega)$ is a.s. eventually equal to $T(\omega)$ as $n \to \infty$, so $M_{n\wedge T} \to M_T$ a.s.). For (c), we observe that

$$|X_{n\wedge T} - X_0| = \left|\sum_{k=1}^{n\wedge T}(X_k - X_{k-1})\right| \leq T \sup_{n\in\mathbb{N}} |X_n - X_{n-1}|.$$

Since $\mathbb{E}[T(\sup_n |X_n - X_{n-1}|)] \leq c\mathbb{E}[T] < \infty$, we can use the Dominated Convergence Theorem as $n \to \infty$ to obtain the result. ∎

2.3 Examples I

We can now use the optional stopping theorem to tell us about hitting probability and expected hitting times of various stochastic processes. In this section we focus on the simple random walk. Many more examples (branching processes, urn processes, shuffling processes, etc) appear on the problem sheets. In Section 2.3.3 we will explain how to calculate (in principle) the length of time it would take for a monkey to type out the complete works of Shakespeare.

2.3.1 Asymmetric simple random walks

Let $(X_i)_{i=1}^{\infty}$ be a sequence of i.i.d. random variables. Let $p+q=1$ with $p,q\in[0,1]$, $p\neq q$ and suppose that

$$\mathbb{P}[X_i=1]=p,\quad \mathbb{P}[X_i=-1]=q.$$

Set $\mathcal{F}_n=\sigma(X_1,\ldots,X_n)$ and note that $(\mathcal{F}_n)$ is a filtration. The asymmetric simple random walk is the stochastic process

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

We define

$$M_n=(q/p)^{S_n}.$$

We will show that M_n is a martingale. Since $X_i\in m\mathcal{F}_n$ for all $i\leq n$, by Proposition 1.2.6 we have $M_n\in m\mathcal{F}_n$. We have $|X_i|\leq 1$ so $|M_n|\leq (q/p)^n$, which implies that $S_n\in L^1$ for all n. Moreover,

$$\begin{aligned}\mathbb{E}[M_{n+1}|\mathcal{F}_n]&=(q/p)^{S_n}\mathbb{E}[(q/p)^{X_{n+1}}|\mathcal{F}_n]\\&=(q/p)^{S_n}\mathbb{E}[(q/p)^{X_{n+1}}]\\&=(q/p)^{S_n}=M_n.\end{aligned}$$

Here we use the taking out what is known rule, followed by the fact that X_{n+1} is independent of $\mathcal{F}_n$ and the relationship between conditional expectation and independence. To deduce the final step we use that $\mathbb{E}[(q/p)^{X_{n+1}}]=p(q/p)^1+q(q/p)^{-1}=p+q=1$.

Let $T_a=\inf\{n:S_n=a\}$ and $T=T_a\wedge T_b$ for integer $a<0<b$. We can show that T_a is a stopping time by noting that

$$\{T_a\leq n\}=\bigcup_{i=0}^{n}\{S_n\leq a\}.$$

Similarly, T_b is a stopping time and it follows from Lemma 2.2.7 that T is also a stopping time.

Since $T\wedge n$ is a bounded stopping time (by Lemma 2.2.7), we have the condition (a) of the optional stopping theorem for (M_n) and $T\wedge n$. So

$$1=M_0=\mathbb{E}[M_{T\wedge n}].$$

We will show as part of question 2 on problem sheet 3 that $\mathbb{P}[T<\infty]=1$, which implies that $n\wedge T$ is eventually equal to T as $n\to\infty$. Hence $M_{n\wedge T}\to M_T$ almost surely as $n\to\infty$. Since $M_{T\wedge n}$ is bounded between $(q/p)^a$ and $(q/p)^b$, by the bounded convergence theorem, $\mathbb{E}[M_{T\wedge n}]\to\mathbb{E}[M_T]$ as $n\to\infty$, so that

$$1=\mathbb{E}[M_T].$$

Since $\mathbb{P}[T < \infty] = 1$, we have that $\mathbb{P}[T = T_a] + \mathbb{P}[T = T_b] = 1$. By partitioning the expectation on whether or not $\{T = T_a\}$ (which is equal to $\{T_a < T_b\} = \{S_T = a\}$), we can deduce

$$1 = \mathbb{P}[T = T_a] \left(\frac{q}{p}\right)^a + \mathbb{P}[T = T_b]) \left(\frac{q}{p}\right)^b.$$

Solving these two linear equations (recall that $p \neq q$) gives that

$$\mathbb{P}[T = T_a] = \frac{(q/p)^b - 1}{(q/p)^b - (q/p)^a}. \tag{2.1}$$

Let us now assume that

$$p > q.$$

Heuristically, this would mean we expect our random walk to drift upwards, towards ∞. Of course we could analyse the case $q < p$ by symmetry.

Note that $T_b \geq b$, since it takes b steps to reach height b from height 0. Thus, as $b \to \infty$,

$$\mathbb{1}\{T = T_a\} = \mathbb{1}\{T_a < T_b\} \to \mathbb{1}\{T_a < \infty\}$$

almost surely. Applying the bounded convergence theorem to (2.1) as $b \to \infty$, we have $(q/p)^b \to 0$ and thus obtain

$$\mathbb{P}[T_a < \infty] = (q/p)^{-a}.$$

Since $\{T_a < \infty\} = \{\min_n S_n < a\}$, this allows us to explicitly calculate $\mathbb{E}[\min_n S_n]$.

$$\begin{aligned}
\mathbb{E}[\min_n S_n] &= -\sum_{x=0}^{\infty} x\mathbb{P}[\min_n S_n = -x] \\
&= -\sum_{x=1}^{\infty} x\left((q/p)^{x-1} - (q/p)^x\right) \\
&= -\sum_{x=1}^{\infty} x\left(\frac{q}{p} - 1\right)(q/p)^x.
\end{aligned}$$

In particular we have $\mathbb{E}[\min_n S_n] > -\infty$, so $\min_n S_n \in L^1$.

Similarly, from (2.1) we have

$$\mathbb{P}[T_b = T] = \frac{1 - (q/p)^a}{(q/p)^b - (q/p)^a}$$

and letting $a \to -\infty$ (with justification similar to above, which is left to you) we have $(q/p)^a \to \infty$ and thus obtain that $\mathbb{P}[T_b < \infty] = 1$.

We can also use martingales to calculate $\mathbb{E}[T_b]$. Since $X_n = S_n - (p-q)n$ is also a martingale (I leave it to you to prove this), and $T_b \wedge n$ is a bounded stopping time, the optional stopping theorem implies that $0 = \mathbb{E}[S_{T_b \wedge n} - (p-q)(T_b \wedge n)]$, which we rearrange to

$$\mathbb{E}[S_{T_b \wedge n}] = (p-q)\mathbb{E}[T_b \wedge n].$$

We take limits as $n \to \infty$ on both sides of this equation. Since $\mathbb{P}[T_b < \infty] = 1$, as $n \to \infty$ we have $T_b \wedge n = T_b$ eventually, so $T_b \wedge n \nearrow T_b$ almost surely and the monotone convergence theorem implies that $\mathbb{E}[T_b \wedge n] \to \mathbb{E}[T_b]$. Note that $S_{T_b \wedge n}$ is bounded above by b and below by $\min_m S_m$, both of which are in L^1, so by the dominated convergence theorem, $\mathbb{E}[S_{T_b \wedge n}] \to \mathbb{E}[S_{T_b}] = b$ as $n \to \infty$. Hence

$$\mathbb{E}[T_b] = b/(p-q).$$

2.3.2 Simple symmetric random walks

Let $(X_i)_{i=1}^{\infty}$ be a sequence of i.i.d. random variables where

$$\mathbb{P}[X_i = 1] = \mathbb{P}[X_i = -1] = \frac{1}{2}.$$

Set $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ and note that $(\mathcal{F}_n)$ is a filtration. The asymmetric simple random walk is the stochastic process

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

We have already seen that S_n is a martingale. There are also some non-obvious martingales associated to S_n.

Let $\theta \in \mathbb{R}$ and define

$$M_n^{(\theta)} := \frac{e^{\theta S_n}}{(\cosh\theta)^n},$$

and note that $M_n^{(\theta)} := \prod_{i=1}^{n}(e^{\theta X_i}/\cosh\theta)$. Since $X_i \in m\mathcal{F}_n$ for all $i \le n$, $S_n \in m\mathcal{F}_n$ for all n by Proposition 1.2.6. Since $|S_n| \le n$ we have $|M_n| \le \frac{e^{\theta n}}{(\cosh\theta)^n} < \infty$, hence $M_n \in L^1$. We have also that

$$\begin{aligned}
\mathbb{E}[S_{n+1}|\mathcal{F}_n] &= \left(\prod_{i=1}^{n} \frac{e^{\theta X_i}}{\cosh\theta}\right) \mathbb{E}\left[\frac{e^{\theta X_{n+1}}}{\cosh\theta}\Big|\mathcal{F}_n\right] \\
&= M_n \mathbb{E}\left[\frac{e^{\theta X_{n+1}}}{\cosh\theta}\right] \\
&= M_n.
\end{aligned}$$

Here we use the taking out what is known rule, the fact that X_{n+1} is independent of $\mathcal{F}_n$ and the relationship between conditional expectation and independence. To deduce the final line we note that $\mathbb{E}[e^{\theta X_i}/\cosh\theta] = \frac{1}{2}(e^{\theta} + e^{-\theta})/\cosh\theta = 1$.

Let

$$T = \inf\{n\,;\, S_n = 1\},$$

which we have seen is a stopping time in Example 2.2.6. It is not obvious whether $\mathbb{P}[T = \infty]$ is equal to or greater than zero, but with the help of M_n and the optional stopping theorem we can show:

Lemma 2.3.1 *It holds that* $\mathbb{P}[T < \infty] = 1$.

Proof: By Lemma 2.2.7, $T \wedge n$ is a stopping time, and since $T \wedge n \le n$ it is a bounded stopping time. We therefore have condition (a) of the optional stopping theorem and can apply it to deduce that

$$\mathbb{E}[M_0] = 1 = \mathbb{E}\left[M_{T\wedge n}^{(\theta)}\right].$$

We now apply the bounded convergence theorem to let $n \to \infty$ in the rightmost term. To do so we make two observations: (1) $M_{T\wedge n} = e^{\theta S_{T\wedge n}}/(\cosh\theta)^n$ is bounded above by e^{θ} since $\cosh\theta \ge 1$ and $S_{T\wedge n} \in (-\infty, 1]$; (2) as $n \to \infty$, $M_{T\wedge n}^{(\theta)} \to M_T^{(\theta)}$, where the latter is defined to be 0 if $T = \infty$. So bounded convergence theorem implies

$$1 = \mathbb{E}\left[M_T^{(\theta)}\right].$$

Noting that $M_T^{(\theta)} = e^{\theta S_T}/(\cosh\theta)^T$ and $S_T = 1$, we thus have

$$\mathbb{E}[(\cosh\theta)^{-T}] = e^{-\theta} \tag{2.2}$$

for $\theta > 0$.

If $T = \infty$, then $(\cosh\theta)^{-T} = 0$ for all $\theta \neq 0$. If $T < \infty$, then $(\cosh\theta)^{-T} \to 1$ as $\theta \to 0$. Noting that $0 \leq (\cosh\theta)^{-T} \leq 1$, we can apply the bounded convergence theorem and let $\theta \to 0$, with the result that

$$\mathbb{E}[\mathbb{1}\{T < \infty\}] = \lim_{\theta\to 0} e^{-\theta} = 1.$$

Hence $\mathbb{P}[T < \infty] = 1$. ■

With Lemma 2.3.1 in hand, we can go even further and work out the distribution of T. Put $\alpha = 1/\cosh\theta = 2/(e^\theta + e^{-\theta})$, and then some simple algebra tells us that $e^{-\theta} = \frac{1}{\alpha}(1 - \sqrt{1-\alpha^2})$. Hence, from (2.2),

$$\mathbb{E}[\alpha^T] = \mathbb{E}[(\cosh\theta)^{-T}] = e^{-\theta} = \frac{1}{\alpha}(1 - \sqrt{1-\alpha^2}) = \sum_n \alpha^n \mathbb{P}[T = n].$$

The final two terms of the above equation give us a formula for the moment generating function of T. Since for arbitrary $\alpha \in \mathbb{C}$,

$$(1+x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k,$$

where $\binom{\alpha}{k} := \alpha(\alpha-1)\ldots(\alpha-k+1)/k!$, we have (after a short calculation) that

$$\mathbb{P}[T = 2m-1] = (-1)^{m+1}\binom{1/2}{m}, \tag{2.3}$$

Note that $\mathbb{P}[T = 2m] = 0$, because $2m$ is even and 1 is odd.

Remark 2.3.2 $(\star)$ *Strictly speaking, to justify obtaining* (2.3) *by comparing coefficients of* α^n *in the generating function we should use e.g. the continuity theorem for moment generating functions.*

Even though we now know that $\mathbb{P}[T < \infty] = 1$, we are not able to apply optional stopping to (S_n) and T; because (S_n) is unbounded and we do not know if $\mathbb{E}[T] < \infty$. In fact, $\mathbb{E}[T] = \infty$, so the optional stopping theorem does not apply here. But, we can use the optional stopping theorem to *deduce* that $\mathbb{E}[T] = \infty$, as follows.

Suppose, for a contradiction, that $\mathbb{E}[T] < \infty$. Then we could apply the optional stopping theorem to (S_n) and T using the condition (c). Hence $\mathbb{E}[S_0] = 0 = \mathbb{E}[S_T]$. But, $\mathbb{P}[T < \infty] = 1$ which means, by definition of T, that $S_T = 1$ almost surely. This is a contradiction so we must have $\mathbb{E}[T] = \infty$.

2.3.3 Monkeys

At each time $1, 2, 3, \ldots$, a monkey types a single capital letter (out of a choice of 26), chosen independently of the letters it has previously typed. We would like to find out how long one expects to wait before the monkey types the word ABRACADABRA.

Just before each time $n = 1, 2 \ldots$, a new gambler arrives and bets £1 that

$$\text{the } n^{\text{th}} \text{ letter will be "A".}$$

If he loses, he leaves. If he wins, he receives £26, all of which he bets on the event that

$$\text{the } (n+1)^{\text{th}} \text{ letter will be "B".}$$

If he loses, he leaves. If he wins, he bets his whole fortune of £26^2 that

$$\text{the } (n+2)^{\text{th}} \text{ letter will be "R",}$$

and so through the sequence ABRACADABRA. Let $M^{(n)}$ be the winning of the n^{th} gambler (hence $M_k^{(n)} = 0$ for $k < n$ since the n^{th} gambler has not even started gambling before time n), then each $M^{(n)}$ is a martingale, and so is $M_n := \sum_{k=1}^{n} M_n^{(k)}$. Furthermore, M has uniformly bounded increments. Let T be the first time by which the monkey has produced the consecutive sequence ABRACADABRA, then $\mathbb{E}[T] < \infty$.

At time T, ABRACADABRA has just been typed. All gamblers except those who started at times $T-11, T-4$ and $T-1$ have lost their £1. Thus, the optional stopping theorem implies that

$$0 = \mathbb{E}[M_T] = \mathbb{E}\left[\sum_{n=1}^{T} M_T^{(n)}\right] = \mathbb{E}[(26^{11}-1) + (26^4-1) + (26-1) + (-1)(T-3)],$$

hence $\mathbb{E}(T) = 26^{11} + 26^4 + 26$.

Remark 2.3.3 *If, instead, we that require our monkey types out the complete works of Shakespeare, it will take a very long time.*

2.4 The Martingale Convergence Theorem

In this section, we are concerned with almost sure convergence of supermartingales (M_n) as $n \to \infty$. Naturally, martingales are a special case and submartingales can be handled through multiplying by -1. In particular, we will prove the (first version of the) martingale convergence theorem. We will look at L^1 convergence of supermartingales in Section 2.6.

Let (M_n) be a stochastic process and fix $a < b$. We define $U_N[a,b]$ to be the number of *upcrossings* made in the interval $[a,b]$ by $M_1, \ldots, M_N$. That is, $U_n[a,b]$ is the largest k such there exists

$$0 \le s_1 < t_2 < \ldots < s_k < t_k \le N \quad \text{such that} \quad M_{s_i} \le a, M_{t_i} > b \text{ for all } i = 1, \ldots, k.$$

Studying upcrossings is key to establishing almost sure convergence of supermartingales. To see why upcrossings are important, note that if $(c_n) \subseteq \mathbb{R}$ is a (deterministic) sequence and $c_n \to c$, for some $c \in \mathbb{R}$, then there is no interval $[a,b]$ $a < b$ such that $(c_n)_{n=1}^{\infty}$ makes infinitely many upcrossings of $[a,b]$; if there was then (c_n) would oscillate and couldn't converge.

Lemma 2.4.1 (Doob's Upcrossing Lemma) *Let M be a supermartingale. Then*

$$(b-a)\mathbb{E}(U_N[a,b]) \le \mathbb{E}((M_N - a)^-).$$

Proof: Let $C_1 = \mathbb{1}\{M_0 < a\}$ and recursively define

$$C_n = \mathbb{1}\{C_{n-1} = 1, M_{n-1} \le b\} + \mathbb{1}\{C_{n-1} = 0, M_{n-1} < a\}.$$

The behaviour of C_n is that, when X enters the region below a, C_n starts taking the value 1. It will continue to take the value 1 until M enters the region above b, at which point C_n will start taking the value 0. It will continue to take the value 0 until M enters the region below a, and so on. Hence,

$$(C \circ M)_N = \sum_{k=1}^{N} C_k(X_{k+1} - X_k) \ge (b-a)U_N[a,b] - (M_N - a)^-.$$

That is, each upcrossing of $[a,b]$ by X picks up at least $(b-a)$; the final term corresponds to an upcrossing that X might have started but not finished.

Note that C is previsible, bounded and non-negative. Hence, by Theorem 2.2.4 we have that $C \circ X$ is a supermartingale. Thus $\mathbb{E}((C \circ M)_N) \le 0$, which proves the given result. ∎

Note that $U_N[a,b]$ is an increasing function of N, and define $U_\infty[a,b]$ by

$$U_\infty[a,b](\omega) = \lim_{N \uparrow \infty} U_N[a,b](\omega).$$

With this definition, $U_\infty[a,b]$ could potentially be infinite, but we can prove that it is not. Recall that a stochastic process (X_n) is bounded in L^p if $\sup_n \mathbb{E}[|X_n|^p] < \infty$.

Lemma 2.4.2 *Suppose M is a supermartingale that is bounded in L^1. Then $P[U_\infty[a,b] = \infty] = 0$.*

Proof: From Lemma 2.4.1 we have

$$(b-a)\mathbb{E}[U_N[a,b]] \le |a| + \sup_{n \in \mathbb{N}} \mathbb{E}|M_n| < \infty.$$

Hence, by the bounded convergence theorem we have

$$(b-a)\mathbb{E}[U_\infty[a,b]] \leq |a| + \sup_{n\in\mathbb{N}} \mathbb{E}|M_n| < \infty,$$

which implies that $\mathbb{P}[U_\infty[a,b] < \infty] = 1$. ■

Essentially, Lemma 2.4.2 says that the paths of M cannot oscillate indefinitely. This is the crucial ingredient of the martingale convergence theorem.

Theorem 2.4.3 (Martingale Convergence Theorem I) *Suppose M is a supermartingale bounded in L^1. Then the almost sure limit $M_\infty = \lim_{n\to\infty} M_n$ exists and $\mathbb{P}[|M_\infty| < \infty] = 1$.*

PROOF: Define

$$\Lambda_{a,b} = \{\omega : \liminf_n M_n(\omega) < a < b < \limsup_n M_n(\omega)\}.$$

We observe that $\Lambda_{a,b} \subset \{U_\infty[a,b] = \infty\}$, which has probability 0 by Lemma 2.4.2. But since

$$\{\omega : M_n(\omega) \text{ does not converge to a limit in } [-\infty,\infty]\} = \bigcup_{a,b\in\mathbb{Q}} \Lambda_{a,b},$$

we have that $\liminf_n M_n = \limsup_n M_n$ almost surely, so

$$\mathbb{P}[M_n \text{ converges to some } M_\infty \in [-\infty,+\infty]] = 1.$$

To finish, by Fatou's Lemma,

$$\mathbb{E}[|M_\infty|] = \mathbb{E}[\liminf_{n\to\infty} |M_n|] \leq \liminf_{n\to\infty} \mathbb{E}[|M_n|] \leq \sup_n \mathbb{E}[|M_n|] < \infty,$$

which means that $\mathbb{P}[|M_\infty| < \infty] = 1$. ■

Note that if M_n is a non-negative supermartingale then we have $\mathbb{E}[|M_n|] = \mathbb{E}[M_n] \leq \mathbb{E}[M_0]$, so M is automatically bounded in L^1 and the martingale convergence theorem applies.

2.5 Examples II

The martingale convergence theorem can tell us about the long term behaviour of stochastic processes.

2.5.1 Galton-Watson processes

Let X_i^n, where $n, i \geq 1$, be i.i.d. nonnegative integer-valued random variables. Define a sequence $(Z_n))$ by $Z_1 = 1$ and

$$Z_{n+1} = \begin{cases} X_1^{n+1} + \ldots + X_{Z_n}^{n+1}, & \text{if } Z_n > 0 \\ 0, & \text{if } Z_n = 0 \end{cases}$$

Then Z is known as a branching process, or Galton-Watson process. Typically, we think of Z_n as representing the number of individuals in the n^{th} generation of some population, each of whom gives birth to an i.i.d. number of children in the $(n+1)^{th}$ generation. The (common) distribution of the X_i^n is known as the *offspring distribution*.

Let $\mu = \mathbb{E}(X_i^n)$, and let $\mathcal{F}_n = \sigma(X_{m,i}\,;\, i \in \mathbb{N}, m \leq n)$. Then $Z_n/\mu^n \in m\mathcal{F}_n$ (I leave it for you to prove this), and it is easily seen that $\mathbb{E}[Z_{n+1}] = \mu\mathbb{E}[Z_n]$ so as $\mathbb{E}[|Z_n|] = \mathbb{E}[Z_n] = \mu^{n-1}$ for all n. Also,

$$\begin{aligned}
\mathbb{E}[Z_{n+1} \,|\, \mathcal{F}_n] &= \sum_{k=1}^{\infty} \mathbb{E}[Z_{n+1}\mathbb{1}\{Z_n = k\} \,|\, \mathcal{F}_n] \\
&= \sum_{k=1}^{\infty} \mathbb{E}\left[\left(X_1^{n+1} + \ldots + X_k^{n+1}\right)\mathbb{1}\{Z_n = k\} \,|\, \mathcal{F}_n\right] \\
&= \sum_{k=1}^{\infty} \mathbb{1}\{Z_n = k\}\mathbb{E}\left[X_1^{n+1} + \ldots + X_k^{n+1} \,|\, \mathcal{F}_n\right] \\
&= \sum_{k=1}^{\infty} k\mu\mathbb{1}\{Z_n = k\} \\
&= \mu Z_n.
\end{aligned}$$

Next, we will show that the process Z dies out (i.e. it becomes 0 eventually) if $\mu \leq 1$. We will look at the case $\mu > 1$ on Problem Sheet 3.

For $\mu < 1$, we note that

$$\mathbb{P}[Z_n > 0] \leq \mathbb{E}[Z_n\mathbb{1}\{Z_n \geq 0\}] = \mathbb{E}[Z_n] = \mu^n.$$

Since $\mu < 1$, we have $\sum_{n=1}^{\infty} \mathbb{P}[Z_n > 0] < \infty$, therefore by the first Borel-Cantelli lemma, $\mathbb{P}[Z_n > 0 \text{ i.o.}] = 0$, or in other words $\mathbb{P}[Z_n = 0 \text{ e.v.}] = 1$. Note that we did not use martingales for this!

For $\mu = 1$, the situation is more delicate and we will need the martingale convergence theorem. In this case, Z_n itself is a non-negative martingale, hence by the martingale convergence theorem there exists Z_∞ such that $Z_n \to Z_\infty$ almost surely. Since Z is integer valued, we must have $Z_n = Z_\infty$ eventually (i.e. for all ω, n there exists $N(\omega)$ such that $Z_n(\omega) = Z_\infty(\omega)$ for all $n \geq N(\omega)$). We now have two cases to consider:

- If the offspring distribution is not deterministic, by definition of Z it must be that $Z_\infty \equiv 0$.

Note that in this case $\mathbb{E}[Z_n] = 1$ for all n, we have just sown that $Z_n \to 0$ as $n \to \infty$. This is similar to the picture that emerged for the simple symmetric random walk, stopped at 1, in Section 2.3.2; we had that $S_{n\wedge T} \to S_t = 1$ almost surely but $\mathbb{E}[S_{n\wedge T}] = 0$ for all n.

- If the offspring distribution is deterministic, then since $\mathbb{E}[X_i^n] = 1$ we must have $X_i^n = 1$, which in turn means that $Z_n = 1$ almost surely for all n. Hence $Z_n \to Z_\infty = 1$ almost surely.

2.5.2 An Urn Process

On Q3 of Problem Sheet 2, we looked at the following urn process. At time 0, an urn contains 1 black ball and 1 white ball. At each time $n = 1, 2, 3, \dots,$, a ball is chosen from the urn and returned to the urn. At the same time, a new ball of the same colour as the chosen ball is added to the urn. Just after time n, there are $n + 2$ balls in the urn, of which $B_n + 1$ are black, where B_n is the number of black balls added into the urn at or before time n.

Let
$$M_n = \frac{B_n + 1}{n + 2}$$
be the proportion of balls in the urn that are black, at time n. Note that $M_n \in [0, 1]$. We showed that M_n was a martingale. Note that M_n is non-negative, hence by the martingale convergence theorem there exists a real valued random variable M_∞ such that $M_n \to M_\infty$ almost surely. In fact, we showed as part of the problem sheet that $\lim_n \mathbb{P}[M_n \le p] = p$ for all $p \in [0, 1]$. Hence, M_∞ is uniformly distributed on $[0, 1]$.

2.6 Uniformly Integrable Martingales

The martingale convergence theorem, as stated in Theorem 2.4.3, doesn't provide any information about what the limit M_∞ is. In some cases, like that of Section 2.5.2, we can calculate the limit explicitly, but often it is not possible to do so. In these cases we can still ask simple questions about M_∞, such as if M_∞ is almost surely zero, or not. In this section we look at one way to show that, under suitable extra conditions, $\mathbb{P}[M_\infty > 0] > 0$. To do so, we will prove a version of the martingale convergence theorem which, under a suitable extra condition, guarantees that $M_n \to M_\infty$ in L^1.

Definition 2.6.1 *A set $\mathcal{C}$ of random variables is said to be uniformly integrable if, for all $\varepsilon > 0$, there exists $K \in [0, \infty)$ such that for all $X \in \mathcal{C}$,*

$$\mathbb{E}[|X|\mathbb{1}\{|X| > K\}] < \varepsilon.$$

A martingale (or, more generally, a stochastic process) (M_n) is said to be uniformly integrable if $\{M_n : n = 0, 1, \dots\}$ is uniformly integrable.

The reason we are interested in uniformly integrable martingales is that, roughly speaking, uniformly integrability is what is needed to upgrade convergence in probability into L^1 convergence. To be precise we have the following theorem, which we state without proof.

Proposition 2.6.2 *Let (X_n) be a sequence of L^1 random variables and let $X_\infty \in L^1$. Then, $X_n \to X_\infty$ in L^1 if and only if the following two conditions hold:*

1. *$X_n \to X_\infty$ in probability,*
2. *(X_n) is uniformly integrable.*

Recall that a set $\mathcal{C}$ of random variables is bounded in L^1 if there exists $K \in \mathbb{R}$ such that $\mathbb{E}[|X|] \le K$ for all $X \in \mathcal{C}$. Uniform integrability is a stronger condition that L^1 boundedness:

Lemma 2.6.3 *Let $\mathcal{C}$ be a set of uniformly integrable random variables. Then $\mathcal{C}$ is bounded in L^1.*

Proof: Let $K \in \mathbb{N}$ be such that, for all $X \in \mathcal{C}$, $\mathbb{E}\big[|X|\mathbb{1}\{|X| \ge K\}\big] < 1$. Then, for all $X \in \mathcal{C}$,

$$\mathbb{E}[|X|] = \mathbb{E}\big[|X|\mathbb{1}\{|X| \ge K\}\big] + \mathbb{E}\big[|X|\mathbb{1}\{|X_n| < K\}\big] \le 1 + K$$

and the proof is complete. ■

The converse of Lemma 2.6.3 is not true. For example, consider the set $\mathcal{C} = \{X_n\,;\, n \in \mathbb{N}\}$ where X_n has distribution $\mathbb{P}[X_n = n] = \frac{1}{n}$ and $\mathbb{P}[X_n = 0] = 1 - \frac{1}{n}$. Then $X_n = |X_n|$, $\mathbb{E}[|X_n|] = 1$ for all n, so as $\mathcal{C}$ is bounded in L^1, but

$$E[X_n\mathbb{1}\{X_n > K\}] = n\frac{1\{K > n\}}{n} = \mathbb{1}\{K > n\}$$

so $\mathcal{C}$ is not uniformly integrable.

An important class of examples of uniformly integrable random variables is given by the following lemma.

Lemma 2.6.4 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X \in L^1$. Then the set*

$$\mathcal{C} = \{\mathbb{E}[X \mid \mathcal{G}] : \mathcal{G} \text{ is a sub-}\sigma\text{-field of } \mathcal{F}\}$$

is uniformly integrable.

PROOF: ($\star$) Let $\epsilon > 0$. By a 'standard' result (see e.g. Lemma 13.1 of 'Probability with Martingales' by David Williams), choose $\delta > 0$ so that, for all $F \in \mathcal{F}$, if $\mathbb{P}(F) < \delta$ then $\mathbb{E}[\,|\,X\,|\,\mathbb{1}_F] < \epsilon$. Choose also some $K \in \mathbb{R}$ such that $\mathbb{E}[|X|] < K\delta$.

Let $\mathcal{G}$ be a sub-σ-field of $\mathcal{F}$ and write $Y = \mathbb{E}[X \mid \mathcal{G}]$. Then, by Jensen's inequality, using that $|\cdot|$ is a convex function,

$$|Y| \leq \mathbb{E}[|X| \mid \mathcal{G}]. \tag{2.4}$$

Hence $\mathbb{E}[|Y|] \leq \mathbb{E}[|X|]$ and by Markov's inequality we have

$$K\mathbb{P}[|Y| > K] \leq \mathbb{E}[|Y|] \leq \mathbb{E}[|X|].$$

Hence, by definition of K we have $\mathbb{P}[|Y| > K] < \delta$. By (2.4), the tower rule and the definition of δ,

$$\mathbb{E}[|Y|\mathbb{1}\{|Y| \geq K\}] \leq \mathbb{E}[|X|\mathbb{1}\{|Y| \geq K\}] < \epsilon.$$

Since Y and ϵ were arbitrary, $\mathcal{C}$ is uniformly integrable. ■

We are now ready to prove the second version of the martingale convergence theorem, which extends Theorem 2.4.3; it requires the extra condition of uniform integrability and provides the extra conclusion of L^1 convergence.

Theorem 2.6.5 (Martingale Convergence Theorem II) *Let M be a uniformly integrable martingale. Then there exists a real valued random variable M_∞ such that $M_n \to M_\infty$ almost surely and in L^1. Moreover, $M_n = \mathbb{E}[M_\infty \mid \mathcal{F}_n]$.*

PROOF: The existence of M_∞ and the almost sure convergence $M_n \to M_\infty$ are immediate from the (first version of the) Martingale Convergence Theorem. Since almost sure convergence implies convergence in probability, this means that $M_n \to M_\infty$ in probability. By Fatou's Lemma,

$$\mathbb{E}[|M_\infty|] \leq \liminf_{n\to\infty} \mathbb{E}[|M_n|]$$

and, by Lemma 2.6.3, we have that $M_\infty \in L^1$. Hence, we can apply Proposition 2.6.2 and deduce that $M_n \to M_\infty$ in L^1.

To prove the second statement, take $F \in \mathcal{F}_n$ and $m \geq n$. Then, the taking out what is known rule and the martingale property of M implies that that

$$\mathbb{E}[M_m \mathbb{1}_F] = \mathbb{E}[\mathbb{E}[M_m|\mathcal{F}_n]\mathbb{1}_F] = \mathbb{E}[M_n \mathbb{1}_F]. \tag{2.5}$$

But note that

$$|\mathbb{E}[M_m \mathbb{1}_F] - \mathbb{E}[M_\infty \mathbb{1}_F]| \leq \mathbb{E}[|M_m - M_\infty|\mathbb{1}_F] \leq \mathbb{E}[|M_m - M_\infty|],$$

which, by the first part of the proof, converges to 0 as $m \to \infty$. Hence, letting $m \to \infty$ in (2.5) we obtain

$$\mathbb{E}[M_\infty \mathbb{1}_F] = \mathbb{E}[M_n \mathbb{1}_F].$$

Since M_n is $\mathcal{F}_n$ measurable, by definition of conditional expectation, $\mathbb{E}[M_\infty|\mathcal{F}_n] = M_n$. ■

When Theorem 2.6.5 holds, we have $M_n \to M_\infty$ in L^1, and hence also $\mathbb{E}[M_n] \to \mathbb{E}[M_\infty]$. Because (M_n) is a martingale, we have $\mathbb{E}[M_\infty] = \mathbb{E}[M_0]$, and if $\mathbb{E}[M_\infty] \neq 0$ then we can conclude that $\mathbb{P}[M_\infty \neq 0] > 0$.

Note that, as we saw in Section 2.3.2, if (M_n) is not uniformly integrable then we might not have $\mathbb{E}[M_n] \to \mathbb{E}[M_\infty]$.

The next theorem provides a converse for Theorem 2.6.5.

Theorem 2.6.6 (Levy's Upward Theorem) *Let ξ be an L^1 random variable, let $(\mathcal{F}_n)$ be a filtration and define $M_n := \mathbb{E}[\xi \,|\, \mathcal{F}_n]$. Then M is a uniformly integrable martingale.*

Further, $M_n \to \mathbb{E}[\xi \,|\, \mathcal{F}_\infty]$ almost surely and in $\mathcal{L}^1$, as $n \to \infty$, where $\mathcal{F}_\infty = \sigma(\mathcal{F}_n\,;\, n \in \mathbb{N})$.

Proof: We have $M_n \in m\mathcal{F}_n$ and $M_n \in L^1$ for all n, by definition of conditional expectation. By the tower property,

$$\mathbb{E}[M_{n+1}|\mathcal{F}_n] = \mathbb{E}[\mathbb{E}[\xi|\mathcal{F}_{n+1}]|\mathcal{F}_n] = \mathbb{E}[\xi|\mathcal{F}_n] = M_n.$$

Hence (M_n) is a martingale. By Lemma 2.6.4, (M_n) is uniformly integrable. Hence, by the (second version of the) Martingale Convergence Theorem, we have that there exists M_∞ such that $M_n \to M_\infty$ almost surely and in L^1. It remains only to show that $M_\infty = \mathbb{E}[\xi|\mathcal{F}_\infty]$.

Both M_∞ and $\eta = \mathbb{E}[\xi|\mathcal{F}_\infty]$ are $\mathcal{F}_\infty$ measurable. Therefore, by definition of conditional expectation, in order to show that $M_\infty = \eta$ a.s it suffices to show that

$$\mathbb{E}[M_\infty \mathbb{1}_F] = \mathbb{E}[\eta \mathbb{1}_F]. \tag{2.6}$$

for all $F \in \mathcal{F}_\infty$.

($\star$) Since $\mathcal{F}_\infty = \sigma(\cup_n \mathcal{F}_n)$, by a standard result from measure theory (see Lemma 1.6(b) in 'Probability with Martingales' by David Williams) it suffices to show that (2.6) holds for all $F \in \cup_n \mathcal{F}_n$. To see this, fix $F \in \mathcal{F}_n$, and note that by the tower property of conditional expectation we have

$$\mathbb{E}[\eta \mathbb{1}_F] = \mathbb{E}[\xi \mathbb{1}_F] = \mathbb{E}[M_n \mathbb{1}_F] = \mathbb{E}[M_\infty \mathbb{1}_F],$$

where the last equality holds by Theorem 2.6.5. ■

2.7 Two applications of uniformly integrable martingales

In this section we look at two very different ways in which uniformly integrable martingales can be used.

2.7.1 Learning from noisy observations

Let $Z_1, Z_2, \dots$ be i.i.d. L^1 random variables with mean 0. Let θ be an L^1 random variable, independent of all the Z_i. Let

$$Y_i = \theta + Z_i$$

and $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$.

We think of θ as a random value that we would like to estimate. Unfortunately, all we can see are so-called noisy samples $\theta + Z_i$, where Z_i is said to be the noise. A fundamental problem in statistics is producing good estimates of the distribution of θ from a sequence of such samples.

Often, the distribution of θ is called the *prior distribution*, and $\mathbb{P}[\theta \in \cdot \,|\, \mathcal{F}_n]$ is called the *posterior distribution* after n observations. We can use martingale techniques to show that

$$\mathbb{E}[\theta \,|\, \mathcal{F}_n] \to \theta \text{ almost surely} \tag{2.7}$$

as $n \to \infty$. Therefore, we know that it is possible to estimate θ, using the information that comes from the observations (Y_i).

To see (2.7), by Lemma 2.6.4 we have that $M_n := \mathbb{E}[\theta \,|\, \mathcal{F}_n]$ is a UI martingale and hence, by Theorem 2.6.6, converges to some M_∞ both a.s. and in $\mathcal{L}^1$. Further, $M_\infty = \mathbb{E}[\theta \,|\, \mathcal{F}_\infty]$, which is equal to θ if we can show that θ is $\mathcal{F}_\infty$ measurable. This is indeed the case since for an arbitrary a,

$$\{\theta \le a\} = \bigcap_{k=1}^{\infty} \left\{ \frac{1}{n} \sum_{i=1}^{n} Y_i \le a + \frac{1}{k} \text{ e.v.} \right\} \in \mathcal{F}_\infty. \tag{2.8}$$

Hence $M_\infty = \theta$ almost surely, and we are done.

Remark 2.7.1 ($\star$) *How we would best estimate θ using the observations Y_i is another matter. But this is a question for statisticians; there are highly sophisticated methods for computing (i.e. using a computer!) estimates of the distribution of $\mathbb{E}[\theta|\mathcal{F}_n]$, which in turn become better and better estimates for the distribution of θ as $n \to \infty$.*

2.7.2 Kolmogorov's 0-1 law

Uniformly integrable martingales can be used to give a short proof of the Kolmogorov's 0-1 law. Let (X_i) be a sequence of random variables, and define $\mathcal{F}_n = \sigma(X_i\,;\, i \le n)$.

Definition 2.7.2 *The tail σ-field $\mathcal{T}$ of (X_n) is*

$$\mathcal{T} = \bigcap_n \mathcal{T}_n \qquad \textit{where} \qquad \mathcal{T}_n = \sigma(X_{n+1}, X_{n+2}, \dots).$$

Note that $\mathcal{T}$ is a σ-field by Lemma 1.1.5. If (X_n) is adapted to $(\mathcal{F}_n)$, then the tail σ-field contains many important events involving the limiting behaviour of X_n as $n \to \infty$. For example, for any $N \in \mathbb{N}$ we have

$$E_1 = \{\omega\,;\, \lim_n X_n(\omega) \text{ exists}\} = \{\omega\,;\, \lim_n X_{n+N}(\omega) \text{ exists}\} \in \mathcal{T}_N,$$

which implies that $E_1 \in \mathcal{T}$. Using similar methods, we could show that the sets

$$E_2 = \left\{ \omega\,;\, \sum_{n=1}^{\infty} X_n \text{ converges} \right\},$$
$$E_3 = \left\{ \limsup_{n\to\infty} X_n \le c \right\} \text{ (or } \liminf\text{)},$$
$$E_4 = \left\{ \limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_n \le c \right\} \text{ (or } \liminf\text{)},$$

and many more, are elements of $\mathcal{T}$. Heuristically, any event which depends only on the tail behaviour of (X_n) will be $\mathcal{T}$ measurable. Note that as a consequence $\limsup_n X_n$ (or $\liminf$) and $\limsup_n \frac{1}{n}\sum_1^n X_i$ (or $\liminf$) are $\mathcal{T}$ measurable.

It is easily seen that $\mathcal{T} \subseteq \mathcal{F}_\infty = \sigma(\mathcal{F}_n\,;\, n \in \mathbb{N})$, and since

$$\mathcal{T} \subseteq \bigcap_{n \ge N} \mathcal{T}_n = \sigma(X_{N+1}, X_{N+2}, \ldots)$$

we have that $\mathcal{T}$ is independent of $\mathcal{F}_N$ for all $N \in \mathbb{N}$.

Theorem 2.7.3 (Kolmogorov's 0-1 law) *Let $X_1, X_2, \ldots$ be a sequence of independent random variables and $\mathcal{T}$ be the corresponding tail σ-field. If $F \in \mathcal{T}$, then $\mathbb{P}[F] = 0$ or 1.*

PROOF: Define $\mathcal{F}_n := \sigma(X_1, \ldots, X_n)$, then $\mathcal{F}_n$ is independent of $\mathcal{T}_n$. Let $F \in \mathcal{T} \subset \mathcal{F}_\infty$ and $\eta = \mathbb{1}_F$. By Levy's Upward Theorem,

$$\eta = \mathbb{E}[\eta \,|\, \mathcal{F}_\infty] = \lim_{n\to\infty} \mathbb{E}[\eta \,|\, \mathcal{F}_n].$$

Since η is independent of every $\mathcal{F}_n$, $\mathbb{E}[\eta \,|\, \mathcal{F}_n] = \mathbb{E}[\eta] = \mathbb{P}[F]$. Hence $\eta = \mathbb{P}(F)$, almost surely. Since η can only take on values 0 or 1, the result follows. ■

Corollary 2.7.4 *Let $X_1, X_2, \ldots$ be an i.i.d. sequence of random variables and let $\mathcal{T}$ be the associated tail σ-field. Suppose Y is a $\mathcal{T}$ measurable random variable. Then there exists $c \in [-\infty, \infty]$ such that $\mathbb{P}[Y = k] = c$.*

PROOF: By Kolmogorov's 0-1 law we have that $\mathbb{P}[Y \le y] \in \{0, 1\}$ for all $y \in \mathbb{R}$. Note that$y \mapsto \mathbb{P}[Y \le y]$ is an increasing function, we have of Y and set $c := \inf\{y : \mathbb{P}[Y \le y] = 1\}$. Here, as usual, the infimum of the empty set is $+\infty$. Then

$$\mathbb{P}[Y \le y] = \begin{cases} 0, & \text{if } y < c \\ 1, & \text{if } y \ge c \end{cases}.$$

Hence, $\mathbb{P}[Y = c] = 1$. ■

2.8 L^2 Martingales

A martingale M is said to be bounded in L^2 if $\sup_n \mathbb{E}[M_n^2] < \infty$. In this section we briefly look at the properties of martingales that are bounded in L^2.

Remark 2.8.1 ($\star$) *It is often easier to work in L^2 than in L^1, primarily because L^2 is a Hilbert space whereas L^1 is not.*

Since $\mathbb{E}[M_n]^2 \le \mathbb{E}[M_n^2]$ (which follows from Jensen's inequality, or the Cauchy-Schwarz inequality), a set of random variables that is bounded in L^2 is automatically bounded in L^1. The following lemma shows that L^2 bounded sets of random variables are also uniformly integrable. In view of Lemma 2.6.3, we can therefore describe the relationship between these various conditions as:

$$L^2 \text{ bounded} \quad \Rightarrow \quad \text{uniformly integrable} \quad \Rightarrow \quad L^1 \text{ bounded.}$$

Lemma 2.8.2 *Let $\mathcal{C}$ be a set of random variables which is bounded in L^2. Then $\mathcal{C}$ is uniformly integrable.*

PROOF: Let $M \in \mathbb{R}$ be such that $\mathbb{E}[|X|^2] \le M$ for all $X \in \mathcal{C}$. Let $\epsilon > 0$. Then, for any $K \in \mathbb{N}$ we have

$$\mathbb{E}\left[|X|\mathbb{1}\{|X| \ge K\}\right] \le \frac{1}{K}\mathbb{E}\big[K|X|\mathbb{1}\{|X| \ge K\}\big] \le \frac{1}{K}\mathbb{E}[X^2] \le \frac{M}{K}.$$

So, if we choose $K \in \mathbb{N}$ such that $\frac{M}{K} < \epsilon$, and then for all $X \in \mathcal{C}$ we have $\mathbb{E}\left[|X|\mathbb{1}\{|X| \ge K\}\right] < \epsilon$. Hence $\mathcal{C}$ is uniformly integrable. ∎

Let (M_n) be a martingale such that $M_n \in L^2$ for all n. Note that we are not (yet) assuming that (M_n) is L^2 bounded. Using the linearity of conditional expectation, the taking out what is known rule, and the martingale property of M_n, we have

$$\begin{aligned} \mathbb{E}[(M_{n+1} - M_n)^2|\mathcal{F}_n] &= \mathbb{E}[M_{n+1}^2 - 2M_{n+1}M_n + M_n^2|\mathcal{F}_n]` \\ &= \mathbb{E}[M_{n+1}^2|\mathcal{F}_n] - 2M_n\mathbb{E}[M_{n+1}|\mathcal{F}_n] + M_n^2 \\ &= \mathbb{E}[M_{n+1}^2 - M_n^2|\mathcal{F}_n]. \end{aligned} \tag{2.9}$$

The above identity makes martingales in L^2 very tractable, from the point of view of doing calculations; it is rare to be able to treat $x \mapsto x^2$ as though it was a linear function! Breaking the symmetry of the above formula, we can rearrange to obtain

$$\mathbb{E}[M_{n+1}^2|\mathcal{F}_n] = M_n^2 + \mathbb{E}[(M_{n+1} - M_n)^2|\mathcal{F}_n].$$

We can use this identity to prove a useful condition for L^2 boundedness.

Lemma 2.8.3 *Let (M_n) be a martingale such that $M_n \in L^2$ for all n. Then*

$$(M_n) \textit{ is bounded in } L^2 \textit{ if and only if } \sum_{i=1}^{\infty} \mathbb{E}[(M_i - M_{i-1})^2] < \infty.$$

PROOF: Note that

$$M_n^2 = M_0^2 + \sum_{i=1}^{n} M_i^2 - M_{i-1}^2$$

and from (2.9) we have that

$$\mathbb{E}[(M_i - M_{i-1})^2] = \mathbb{E}[M_i^2 - M_{i-1}^2].$$

Putting these two equations together, we obtain

$$\mathbb{E}[M_n^2] = \mathbb{E}[M_0^2] + \sum_{i=1}^{n} \mathbb{E}[(M_i - M_{i-1})^2].$$

The right hand side is an increasing function of n, hence so is the left hand side. Thus,

$$\sup_{n\in\mathbb{N}} \mathbb{E}[M_n^2] = \mathbb{E}[M_0^2] + \sum_{i=1}^{\infty} \mathbb{E}[(M_i - M_{i-1})^2].$$

The stated result follows immediately. ■

To finish, we state a third version of the martingale convergence theorem, which applies to L^2 bounded martingales.

Theorem 2.8.4 (Martingale Convergence Theorem III) *Let (M_n) be a martingale which is bounded in L^2. Then there exists a real valued random variable M_∞ such that $M_n \to M_\infty$ almost surely and in L^2.*

The proof of this theorem is beyond the scope of the course. Recall that we were able to use the second version of the Martingale Convergence Theorem (which asserted L^1 convergence) to deduce information about the limit M_∞. Having a stronger mode of convergence, namely L^2 convergence, potentially allows us to deduce even more information about M_∞. In particular, when Theorem 2.8.4 holds, we have

$$\mathrm{var}[M_n] \to \mathrm{var}[M_\infty],$$

which might allow us to determine whether or not M_∞ is deterministic; of course M_∞ is deterministic if and only if $\mathrm{var}[M_\infty] = 0$.