# Martingales

The unit will cover topics related to martingale convergence, possibly including materials on uniformly integrable martingales, and will be mostly restricted to discrete-time martingales. Chapters 3-4 are partially based on an earlier set of notes on martingales by Stas Volkov.

# 1 Background

## 1.1 Probability Theory Background

Recall:

**Definition 1** *A probability space is a triple* $(\Omega, \mathcal{F}, \mathbb{P})$.

$\Omega$ is the *sample space*, i.e. the set of all possible outcomes. $\mathcal{F}$ is a $\sigma$*-field* (also called $\sigma$-algebra), defined to satisfy these axioms:

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

The $\sigma$-field $\mathcal{F}$ consists of events whose probabilities we may want to calculate. These events (sets) are said to be $\mathcal{F}$*-measurable* (or simply measurable if it is clear what the underlying $\sigma$-field is). Note that by definition, an event is measurable, but sometimes we use the term "measurable event" to emphasize this fact. In particular, $\emptyset \in \mathcal{F}$ is known as the impossible event, $\Omega \in \mathcal{F}$ is known as the certain event, and intersections of events are also events. We require infinite unions of measurable sets to be measurable, because not requiring this will make life difficult if $\Omega$ is infinite, e.g. $[0,1]$ or $\mathbb{R}$.

**Examples of sample spaces**

- toss a coin once or twice: $\{H, T\}$
- toss a coin until a head turns up: $\{\omega_1, \omega_2, \ldots\}$. May wish to study the probability of requiring an even number of tries
- choose a number uniformly in $[0,1]$

**Examples of $\sigma$-fields**

- $\{\emptyset, \Omega\}$: absolutely useless
- $\{\emptyset, A, A^c, \Omega\}$: smallest $\sigma$-field containing $\{A\}$
- $\{0,1\}^{\Omega}$ (power set of $\Omega$, i.e. the set containing all subsets of $\Omega$), always a $\sigma$-field. Fine for finite sets, but not useful (too large) for infinite sets (for deep reasons beyond the scope of this unit)
- $\sigma(\mathcal{C})$, $\sigma$-field generated by a collection of events $\mathcal{C}$, smallest $\sigma$-field containing all events in $\mathcal{C}$, i.e. intersections of all $\sigma$-fields containing events of $\mathcal{C}$ (why is the intersection of two $\sigma$-fields still a $\sigma$-field?)
- *Borel* $\sigma$-field $\mathcal{B}(\mathbb{R})$, $\sigma$-field generated by open sets of $\mathbb{R}$ (same if generated by closed sets), contains $(a,b)$, $(a,b]$, $[a,b)$, $[a,b]$, etc.
- tossing a coin twice, $\mathcal{F}_{12} = \sigma(\{\{HH\}, \{HT\}, \{TH\}, \{TT\}\})$, $\mathcal{F}_1 = \sigma(\{\{HH, HT\}, \{TH, TT\}\})$

From the last example, one may be able to see vaguely that $\sigma$-field is a way to encode information about the experiment. One can only distinguish between two outcomes given the information available to us in the form of $\sigma$-field if they are in different subsets (which we call *atoms*, but we need to note that unions of atoms are still atoms and there may not be a "smallest" atom). If two outcomes are always in the same atom, then they cannot be distinguished

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

1. $\mathbb{P}(\emptyset) = 0, \mathbb{P}(\Omega) = 1$;

2. ($\sigma$-additivity) 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 axiom in the definition of $\mathcal{F}$ and $\mathbb{P}$ are related. A statement about outcomes is said to hold *almost surely*, or shortened to *a.s.*, if it has probability 1.

**Proposition 1** *If $F_n \in \mathcal{F}$ ($n \in \mathbb{N}$) and $\mathbb{P}(F_n) = 1$ for all $n$, then $\mathbb{P}(\bigcap_n F_n) = 1$.*

It is important that the intersection in the above proposition is *countable*. Uncountable intersections (unions, sums, etc) can lead to trouble.

**Definition 2** *A $\sigma$-field $\mathcal{G}$ is a sub-$\sigma$-field of the $\sigma$-field $\mathcal{F}$ if $\mathcal{G} \subset \mathcal{F}$.*

**Definition 3** *A* random variable *$X$ on $(\Omega, \mathcal{F}, \mathbb{P})$ is*

1. *a function $X : \Omega \to \mathbb{R}$*
2. *(measurability) with the property that $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$ for each $x \in \mathbb{R}$ (or we can write $X^{-1}((-\infty, x]) \in \mathcal{F}$).*

| | | |
|---|---|---|
| $\boxed{\sigma\text{-field } \mathcal{F}}$ | $\leftrightarrow$ | $\boxed{\text{“information”}}$ |
| $\boxed{\text{sub-}\sigma\text{-field } \mathcal{G} \subset \mathcal{F}}$ | $\leftrightarrow$ | $\boxed{\text{“subset of information”}}$ |
| $\boxed{X \text{ is } \mathcal{F}\text{-measurable}}$ | $\leftrightarrow$ | $\boxed{X \text{ can be decided based on information in } \mathcal{F}}$ |

At this point, one may wonder what the point of the measurability condition is. In addition to making all the mathematics work out without contradiction, the point to be emphasized for this unit is that it has to do with what kind of random variable can be defined with the information coded in $\mathcal{F}$. It is easier to think of this in the finite setting, i.e. when $\Omega$ is finite and the probability of any event can only be in $\{x_1, \ldots, x_n\}$. In this case, each atom in $\mathcal{F}$ can indeed be written as unions of minimal atoms (i.e. no proper subset of minimal atom is an atom). In addition, the measurability condition is now equivalent to

2' (measurability in the finite setting) $\{\omega \in \Omega : X(\omega) = x_i\} \in \mathcal{F}$ for each $x_i, \ldots, x_n$ (or we can write $X^{-1}(x_i) \in \mathcal{F}$).

If $X$ is $\mathcal{F}$-measurable and $\omega_1$ and $\omega_2$ are such that $x_1 := X(\omega_1) \neq X(\omega_2) =: x_2$, then $X^{-1}(x_1)$ and $X^{-1}(x_2)$ are two disjoint atoms. Therefore, if one can distinguish $\omega$'s using $X$, one can do the same with $\mathcal{F}$. In this sense, the information generated by $X$ is a subset of that contained in $\mathcal{F}$.

**Definition 4** *The $\sigma$-field* generated *by a collection of random variables $\{X_n : n \in \mathbb{N}\}$ is the smallest $\sigma$-field in which each $X_n$ is measurable, i.e. intersections of all $\sigma$-fiels in which each $X_n$ is measurable.*

**Fact.** Arbitrary intersections of $\sigma$-fields are still $\sigma$-fields.
**Examples of measurability**

- We toss a coin twice and let $\omega = (\omega_1, \omega_2) \in \{H, T\}^2$. Define $\mathcal{F}_{12} = \sigma(\{\{HH\}, \{HT\}, \{TH\}, \{TT\}\})$ and $\mathcal{F}_1 = \sigma(\{\{HH, HT\}, \{TH, TT\}\})$. Define for $n = 1, 2$ (and $\ldots$ for the next example),

$$X_n := \begin{cases} 1, & \text{if } \omega_n = H \\ 0, & \text{if } \omega_n = T \end{cases}.$$

Then $X_1^{-1}(1) = \{HH, HT\}$, $X_1^{-1}(0) = \{TH, TT\}$, $X_2^{-1}(1) = \{HH, TH\}$, $X_2^{-1}(0) = \{HT, TT\}$. So both $X_1$ and $X_2$ are $\mathcal{F}_{12}$-measurable, but only $X_1$ is $\mathcal{F}_1$-measurable (actually $\mathcal{F}_1$ is generated by $X_1$). Therefore $X_2$ is a random variable in $(\Omega, \mathcal{F}_{12})$ but not $(\Omega, \mathcal{F}_1)$. The reason is that $\mathcal{F}_1$ does not contain information necessary to distinguish between different outcomes of $X_2$.

- We toss a coin infinitely many times, then $\Omega = \{H, T\}^{\mathbb{N}}$ and we write $\omega = (\omega_1, \omega_2, \ldots)$ for a typical outcome. We define
$$\mathcal{F} = \sigma(\{\omega : \omega_n = W\} : n \in \mathbb{N}, W \in \{H, T\}),$$
i.e. $\mathcal{F}$ is generated by outcomes of each throw. So in particular, each $X_n$ is measurable and thus a random variable. Then
$$\begin{aligned}
\sigma(X_1) &= \{\emptyset, \{H *** \ldots\}, \{T *** \ldots\}, \Omega\} \\
\sigma(X_1, X_2) &= \sigma(\{\{HH ** \ldots\}, \{TH ** \ldots\}, \{HT ** \ldots\}, \{TT ** \ldots\}\}) \\
&= \{\emptyset, \{HH ** \ldots\}, \{TH ** \ldots\}, \{HT ** \ldots\}, \{TT ** \ldots\}, \\
&\quad \{H *** \ldots\}, \{T *** \ldots\}, \{*H ** \ldots\}, \{*T ** \ldots\}, \left\{ \begin{matrix} HH \\ TT \end{matrix} \; ** \ldots, \right\}, \left\{ \begin{matrix} HT \\ TH \end{matrix} \; ** \ldots, \right\}, \\
&\quad \{HH ** \ldots\}^c, \{TH ** \ldots\}^c, \{HT ** \ldots\}^c, \{TT ** \ldots\}^c, \Omega\},
\end{aligned}$$
where $*$ means that it can take on either $H$ or $T$, so $\{H *** \ldots\} = \{\omega : \omega_1 = H\}$. In $\sigma(X_1, X_2)$, $\{HH ** \ldots\}, \{TH ** \ldots\}, \{HT ** \ldots\}, \{TT ** \ldots\}$ are the minimal atoms. With the information available to us in $\sigma(X_1, X_2)$, we can distinguish between $\omega$'s where the first (or second) outcomes are different. But if two $\omega$'s have the same first and second outcomes, then they fall into the same atom and we cannot distinguish between them.

  A useful way to model this experiment is by mapping each $\omega$ to a number $x$ in $[0, 1]$ (by mapping $H \mapsto 0$, $T \mapsto 1$, and $\omega_n \mapsto n^{\text{th}}$ digit in the binary expansion of $x$), then $\sigma(X_1)$ contains information about the first digit, $\sigma(X_1, X_2)$ contains information about the first two digits, etc. There is a slight problem with this mapping, since $(0, 1, 1, \ldots)$ and $(1, 0, 0, \ldots)$ map to the same number, but this is not really a problem over which one loses sleep. (why?) With $\sigma(X_1)$, we can only tell if $x$ is in $[0, 0.5]$ or $[0.5, 1]$. With $\sigma(X_1, X_2)$, we can tell if $x$ is in one of four intervals $[0, 0.25]$, $[0.25, 0.5]$, $[0.5, 0.75]$, $[0.75, 1]$, thus one has more detailed information. As one adds more and more $X_n$ in the $\sigma$-field, the distinguishable intervals become smaller and smaller.

- $\sigma(X)$ is a sub-$\sigma$-field of $\sigma(X, Y)$.

**Proposition 2** *Sums and products of random variables are random variables. A measurable function of a random variable is a random variable.*

This proposition is all but trivial if one does not need to consider measurability. All functions you can think of are measurable. It is really not easy to come up with non-measurable functions on $\mathbb{R}$. We will not prove this proposition. Instead, we give some intuitions in terms of information. If $X$ is $\mathcal{F}$-measurable, then any two outcomes $\omega_1$ and $\omega_2$ such that $X(\omega_1) \neq X(\omega_2)$ must fall into different atoms of $\mathcal{F}$. If $f(X(\omega_1)) \neq f(X(\omega_2))$, then certainly $X(\omega_1) \neq X(\omega_2)$, hence $\omega_1$ and $\omega_2$ are in different atoms of $\mathcal{F}$. This means that $\mathcal{F}$ has enough information to also distinguish between different values of the random variable $f(X)$, hence $f(X)$ is $\mathcal{F}$-measurable. The measurability of sums and products of random variables follows similar intuition: if $X(\omega_1) + Y(\omega_1) = Z(\omega_1) \neq Z(\omega_2) = X(\omega_2) + Y(\omega_2)$, then either $X(\omega_1) \neq X(\omega_2)$ or $Y(\omega_1) \neq Y(\omega_2)$, ...

**Definition 5** *Sub-$\sigma$-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 $\sigma$-fields $\sigma(X_1)$ and $\sigma(X_2)$ are independent. Events $E_1$ and $E_2$ are said to be independent if random variables $\mathbb{1}_{E_1}$ and $\mathbb{1}_{E_2}$ are independent.*

This definition of independence agrees with the classic definition, and crucially implies that $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ if $X$ and $Y$ are independent random variables (for the expectation we will define a bit later).

## 1.2 Analysis Background

In AP2, the expectation of a discrete random variable $X$ that takes on values in $\{x_i : i \in \mathbb{N}\}$ was defined as
$$\mathbb{E}(X) = \sum_{x_i} x_i \mathbb{P}(X = x_i).$$

And for a continuous random variable, the expectation uses integral against the probability density function. This is really not elegant, requiring different formulae under slightly different circumstances. With measure theory, 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 unit, but we can get some basic ideas on how it is done, which relies on the following crucial fact:

**Fact.** A simple random variable $Z$ is one that takes on finitely many values, i.e. $Z = \sum_{k=1}^{K} a_k \mathbb{1}_{A_k}$ for $A_k \in \mathcal{F}$. A random variable $X$ on $(\Omega, \mathcal{F}, \mathbb{P})$ can be approximated by simple random variables, i.e. there exists a sequence of simple random variables $X_n$ such that $X_n \to X$ a.s..

The definition is in 3 steps.

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

Step 2. Simple random variables $X$ that takes on only finitely many values $\{a_1, \ldots, a_n\}$, i.e. $X = \sum_{k=1}^{n} a_k \mathbb{1}_{A_k}$ for measurable sets $A_1, \ldots, A_k$: define $\mathbb{E}(X) := \sum_{k=1}^{n} a_k \mathbb{P}(A_k)$, which agrees with the "classic" definition.

Step 3. Arbitrary $X$: approximate $X$ by a sequence of simple random variables $X_n$ such that $X_n \leq X$ but $X_n \uparrow X$ as $n \to \infty$ a.s., and define $\mathbb{E}(X) :=\uparrow \lim \mathbb{E}(X_n)$.

Recall different modes of convergence of random variables: almost sure, in probability, in distribution, and $\mathcal{L}^1$ (also known as convergence in mean). Recall the definition of $\mathcal{L}^1$ and $\mathcal{L}^2$ spaces:

**Definition 6** *A random variable $X$ on $(\Omega, \mathcal{F}, \mathbb{P})$ is $\mathcal{L}^1$ if $\mathbb{E}(|X|) < \infty$, and $\mathcal{L}^2$ if $\mathbb{E}(|X|^2) < \infty$.*

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

$$\begin{aligned}
\limsup x_n &:= \inf_m \left\{ \sup_{n \geq m} x_n \right\} =\downarrow \lim_m \left\{ \sup_{n \geq m} x_n \right\} \in [-\infty, \infty] \\
\liminf x_n &:= \sup_m \left\{ \inf_{n \geq m} x_n \right\} =\uparrow \lim_m \left\{ \inf_{n \geq m} x_n \right\} \in [-\infty, \infty] \\
x_n \to x \in [-\infty, \infty] &\iff \limsup_n x_n = x = \liminf_n x_n
\end{aligned}$$

Analogously, for sequences of events $E_n$,

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

The following convergence theorems mostly provide conditions for $\mathcal{L}^1$ convergence given a.s. convergence. They are all very useful.

**Proposition 3 (Bounded Convergence Theorem)** *If $X_n \to X$ in probability, and $|X_n| \leq M$, then $\mathbb{E}(X_n) \to \mathbb{E}(X)$.*

**Proposition 4 (Fatou's Lemma)** *If $X_n \geq 0$, then $\liminf \mathbb{E}(X_n) \geq \mathbb{E}(\liminf X_n)$*

**Proposition 5 (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 6 (Dominated Convergence Theorem)** *If $X_n \to X$ a.s., $|X_n| < Y$ for all $n$, and $\mathbb{E}(Y) < \infty$, then $\mathbb{E}(X_n) \to \mathbb{E}(X)$.*

# 2 Conditional Expectation

Suppose $X$ and $Z$ are random variables that take on only finitely many values $\{x_1, \dots, x_m\}$ and $\{z_1, \dots, z_n\}$, respectively. From classic probability theory, we define condition probability and condition expectation as follows:

$$\begin{aligned}
\mathbb{P}(X = x_i \mid Z = z_j) &:= \mathbb{P}(X = x_i, Z = z_j)/\mathbb{P}(Z = z_j) \\
\mathbb{E}\,(X \mid Z = z_j) &:= \sum_i x_i \mathbb{P}(X = x_i \mid Z = z_j) \\
Y := \mathbb{E}\,(X \mid Z) \text{ such that:} & \qquad \text{if } Z(\omega) = z_j, \text{ then } Y(\omega) = \mathbb{E}\,(X \mid Z = z_j)
\end{aligned}$$

(Similarly, if $X$ and $Z$ are both continuous, we define $f_{X\,|\,Z}(x \,|\, z) := f_{X,Z}(x, z)/f_Z(z)$ for $f_Z(z) \neq 0$. This is again not elegant, for many reasons: (1) one needs different definitions for different type of random variables; (2) if $X$ is continuous but $Z$ is discrete, or the other way around, the definition will become very clumsy; (3) most importantly, it is not so easy to justify that these two definitions are really about the same thing, e.g. writing $f_{X\,|\,Z}(x \,|\, z) := \mathbb{P}(X = x, Z = z)/\mathbb{P}(Z = z)$ is nonsense since the RHS is 0/0.)

In this section, we define the conditional expectation (which is, as it turns out, easier to define than conditional probability[1]) in terms of measure theory and $\sigma$-fields. This definition, originally due to A. N. Kolmogorov, is of fundamental and central importance not only to the theory of martingales, but also to modern probability theory as a whole. Kolmogorov, due to his large contribution to modern probability theory (and many other fields of mathematics), of which this definition forms a small part, is now regarded as the founder of this discipline.

Knowing the value of $Z$ is equivalent to partitioning $\Omega$ into $Z$-atoms and knowing which atom has been picked – no more detailed information is available to us.

| $\Omega$ | $G_1 = \{Z = z_1\}$ | $G_2 = \{Z = z_2\}$ | $\dots$ | $G_n = \{Z = z_n\}$ |
|---|---|---|---|---|

Define $\mathcal{G} = \sigma(Z)$. A set $G$ is in $\mathcal{G}$ if and only if it is a union of some (or all) of these atoms. A random variable $Y'$ is $\mathcal{G}$-measurable (which means the information carried by $Y'$ belongs in $\mathcal{G}$) if and only if it is constant on each atom in $\mathcal{G}$ (otherwise, using $Y'$, we can distinguish between different events within a single atom in $\mathcal{G}$). In particular, $Y = \mathbb{E}\,(X \mid Z)$ takes a constant value $\mathbb{E}\,(X \mid Z = z_j)$ on each $Z$-atom $G_j$, and is thus $\mathcal{G}$-measurable, which leads to the following calculation:

$$\begin{aligned}
\mathbb{E}\,(Y \mathbb{1}_{G_j}) &= \mathbb{E}\,(X \mid Z = z_j)\mathbb{P}(Z = z_j) = \sum_i x_i \mathbb{P}(X = x_i \mid Z = z_j)\mathbb{P}(Z = z_j) = \sum_i x_i \mathbb{P}(X = x_i, Z = z_j) \\
&= \sum_{i,j'} x_i \mathbb{1}_{j=j'} \mathbb{P}(X = x_i, Z = z'_j) = \mathbb{E}\,(X \mathbb{1}_{G_j}).
\end{aligned}$$

Intuition: $Y$ is taken to be constant on each $G_j$; this constant is, in a sense, the average of $X$ over $G_j$. This calculation can be generalised to all $G \in \mathcal{G}$, and thus we conclude that

$$\mathbb{E}\,(Y \mathbb{1}_G) = \mathbb{E}\,(X \mathbb{1}_G) \text{ for all } G \in \mathcal{G}.$$

Indeed, this is the defining equality of conditional expectation:

**Definition 7 (Definition of Conditional Expectation, Kolmogorov, 1933)** *Let $X$ be an $\mathcal{L}^1$ random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{G}$ be a sub-$\sigma$-field of $\mathcal{F}$. Then there exists a random variable $Y$ such that*

1. *$Y$ is $\mathcal{G}$-measurable,*
2. *$Y \in \mathcal{L}^1$,*
3. *for every $G \in \mathcal{G}$, we have $\mathbb{E}\,(Y \mathbb{1}_G) = \mathbb{E}\,(X \mathbb{1}_G)$.*

*If $\tilde{Y}$ is another random variable that satisfies these properties, then $\tilde{Y} = Y$ a.s.. The random variable $Y$ is called (a version of) the conditional expectation $\mathbb{E}\,(X \mid \mathcal{G})$ of $X$ given $\mathcal{G}$.*

We shall not prove the above is well defined in this set of notes, please refer to Williams for detail. As an example, in the experiment where the coin is tossed infinitely many times, the random variable $X_1$ partitions $\Omega$ into two atoms. The random variable $Y_1 := \mathbb{E}\,(X_1 + X_2 \mid X_1)$ is 0.5 if $X_1 = 0$ (which is the average of $X_1 + X_2$ over all outcomes in $\{0 * **\}$), and 1.5 if $X_1 = 1$.

[1] Conditional probability is rarely used in modern research in probability anyway.

| | | |
|---|---|---|
| $(\Omega, \sigma(X_1))$ | $\{0 * * * \ldots\}$ | $\{1 * * * \ldots\}$ |
| $Y_1$ | 0.5 | 1.5 |

But the random variable $(X_1, X_2)$ partitions $\Omega$ into four atoms. The random variable $Y_2 := \mathbb{E}(X_1 + X_2 \,|\, X_1, X_2)$ is completely determined by the conditioning.

| | | | | |
|---|---|---|---|---|
| $(\Omega, \sigma(X_1, X_2))$ | $\{00 * * \ldots\}$ | $\{10 * * \ldots\}$ | $\{01 * * \ldots\}$ | $\{11 * * \ldots\}$ |
| $Y_2$ | 0 | 1 | 1 | 2 |

The conditional expectation $Y = \mathbb{E}(X \,|\, \mathcal{G})$ defined above can be regarded as the best least-squares estimator of $X$ given the information in $\mathcal{G}$. This fact can be seen by the following calculation. Suppose $Y'$ is a random variable that is $\mathcal{G}$-measurable (i.e. it only uses information available in $\mathcal{G}$), then the square error of the estimator $Y'$ is

$$\mathbb{E}((X - Y')^2) = \mathbb{E}((X - Y + Y - Y')^2) = \mathbb{E}((X - Y)^2) + 2\mathbb{E}((X - Y)(Y - Y')) + \mathbb{E}((Y - Y')^2).$$

In the middle term above, $V := Y - Y'$ is $\mathcal{G}$-measurable, and thus we can take better and better approximations using $V_n = \sum_{k=1}^{m} a_k \mathbb{1}_{G_k}$, where $G_k \in \mathcal{G}$, such that $V_n \to V$ as $n \to \infty$. Then

$$\mathbb{E}((X - Y)V) = \lim_{n \to \infty} \sum_{k=1}^{n} a_k \mathbb{E}((X - Y)\mathbb{1}_{G_k}) = 0,$$

since $\mathbb{E}(X\mathbb{1}_{G_k}) = \mathbb{E}(Y\mathbb{1}_{G_k})$ for each $G_k \in \mathcal{G}$. (In order to turn this into a rigorous proof, one simply splits the relevant terms into positive and negative parts and then applies the monotone convergence theorem.) This show shat for any $\mathcal{G}$-measurable $Y'$,

$$\mathbb{E}((X - Y')^2) \geq \mathbb{E}((X - Y)^2).$$

**Proposition 7 (Properties of Conditional Expectation)** *(a)* $\mathbb{E}(\mathbb{E}(X \,|\, \mathcal{G})) = \mathbb{E}(X)$, $\mathbb{E}(X \,|\, \mathit{constant}) = \mathbb{E}(X)$, $\mathbb{E}(\mathbb{E}(X \,|\, Y) \,|\, Y) = \mathbb{E}(X \,|\, Y)$.
*(b) If* $X$ *is* $\mathcal{G}$*-measurable, then* $\mathbb{E}(X \,|\, \mathcal{G}) = X$.
*(c)* (Linearity) $\mathbb{E}(a_1 X_1 + a_2 X_2 \,|\, \mathcal{G}) = a_1 \mathbb{E}(X_1 \,|\, \mathcal{G}) + a_2 \mathbb{E}(X_2 \,|\, \mathcal{G})$.
*(d)* (Positivity) *If* $X \geq 0$*, then* $\mathbb{E}(X|\mathcal{G}) \geq 0$.
*(e)* (MON) *If* $0 \leq X_n \uparrow X$*, then* $\mathbb{E}(X_n \,|\, \mathcal{G}) \uparrow \mathbb{E}(X_n \,|\, \mathcal{G})$.
*(f)* (FATOU) *If* $0 \leq X_n$*, then* $\mathbb{E}(\liminf X_n \,|\, \mathcal{G}) \leq \liminf \mathbb{E}(X_n \,|\, \mathcal{G})$.
*(g)* (DOM) *If* $|X_n| \leq V$*,* $X_n \to X$*, and* $\mathbb{E}(V) < \infty$*, then* $\mathbb{E}(X_n \,|\, \mathcal{G}) \to \mathbb{E}(X \,|\, \mathcal{G})$.
*(h)* (JENSEN) *If* $f$ *is convex, then* $\mathbb{E}(f(X) \,|\, \mathcal{G}) \geq f(\mathbb{E}(X \,|\, \mathcal{G}))$.
*(i)* (Tower) *If* $\mathcal{H} \subset \mathcal{G}$ *then* $\mathbb{E}(\mathbb{E}(X \,|\, \mathcal{G}) \,|\, \mathcal{H}) = \mathbb{E}(X \,|\, \mathcal{H})$.
*(j)* (Taking out what is known) *If* $Z$ *is* $\mathcal{G}$ *measurable, then* $\mathbb{E}(ZX \,|\, \mathcal{G}) = Z\mathbb{E}(X \,|\, \mathcal{G})$.
*(k)* (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:`(a) homework. (b) and (i) are almost immediate from definition. For (j), let $Y = \mathbb{E}(X \,|\, \mathcal{G})$ and $G \in \mathcal{G}$, then it is sufficient to establish that

$$\mathbb{E}(ZX\mathbb{1}_G) = \mathbb{E}(\mathbb{E}(ZX \,|\, \mathcal{G})\mathbb{1}_G) = \mathbb{E}(Z\mathbb{E}(X \,|\, \mathcal{G})\mathbb{1}_G) = \mathbb{E}(ZY\mathbb{1}_G),$$

where the first equality is the definition of $\mathbb{E}(ZX \,|\, \mathcal{G})$. We use standard (for measure theory) machinary for this, then it suffices to establish the above for $Z = \mathbb{1}_{G'}$, i.e.

$$\mathbb{E}(X\mathbb{1}_{G'}\mathbb{1}_G) = \mathbb{E}(Y\mathbb{1}_{G'}\mathbb{1}_G),$$

which follows from the definition of $Y$ since $\mathbb{1}_{G'}\mathbb{1}_G = \mathbb{1}_{G' \cap G}$ and $G' \cap G \in \mathcal{G}$. ■

# 3 Martingale Definitions and the Optional Stopping Theorem

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability triple as usual.

**Definition 8** *An increasing family of $\sigma$-fields $\mathcal{F}_n$ is* filtration *if $\mathcal{F}_{12} \subset \mathcal{F}_1 \subset \mathcal{F}_2 \ldots \subset \mathcal{F}$. A process $X = (X_n : n \geq 0)$ is* adapted *if each $X_n$ is $\mathcal{F}_n$ measurable. The $\sigma$-field $\mathcal{F}_\infty := \sigma(\cup_n \mathcal{F}_n)$ is $\subset \mathcal{F}$.*

Intuition: $\mathcal{F}_n$ is an increasing sequence of $\sigma$-fields and carries information about the process up to and including time $n$. Usually, $\mathcal{F}_n$ is taken to be the natural filtration $\mathcal{F}_n = \sigma(X_0, X_1, \ldots, X_n)$. If $X$ is adapted to $\mathcal{F}_n$, then we know the values of $X_0, \ldots, X_n$ at time $n$ using information in $\mathcal{F}_n$. Refer to the second example of measurability.

**Definition 9** *A process $M$ is* martingale *(relative to $(\{\mathcal{F}_n\}, \mathbb{P})$)*

1. *if $M_n$ is adapted,*
2. *$\mathbb{E}(|M_n|) < \infty$,*
3. *$\mathbb{E}(M_n \,|\, \mathcal{F}_{n-1}) = M_{n-1}$.*

Submartingale*: $=$ replaced by $\geq$.* supermartingale*: $=$ replaced by $\leq$.*

Remark: WLOG, we can assume $M_0 = 0$ when studying martingales.
**Examples**:

1. $S_n := X_1 + \cdots + X_n$, $\mathcal{F}_n = \sigma(X_1, X_2, \ldots, X_n)$, where $\mathbb{E}(X_k) = 0$ and $\mathbb{E}(|X_k|) < \infty$ for all $k$. For example, $X_k = \pm 1$ dependent on the outcome of a (fair/unfair) coin toss. Can think of $S_n$ as your net winning after $n$ plays of the a certain game, if one bets exactly £1 on each play.
2. $M_n := X_1 X_2 \ldots X_n$ where $X_1, X_2, \ldots$ are independent nonnegative random variables with $\mathbb{E}(X_k) = 1$; Calculate: $\mathbb{E}(M_n \,|\, \mathcal{F}_{n-1}) = \mathbb{E}(M_{n-1} X_n \,|\, \mathcal{F}_{n-1}) = M_{n-1} \mathbb{E}(X_n \,|\, \mathcal{F}_{n-1}) = M_{n-1} \mathbb{E}(X_n) = M_{n-1}$.
3. Given a filtration $\mathcal{F}_n$ and a random variable $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, $M_n := \mathbb{E}(X \,|\, \mathcal{F}_n)$ is a martingale. Calculate: $\mathbb{E}(\mathbb{E}(X \,|\, \mathcal{F}_n) \,|\, \mathcal{F}_{n-1}) = \mathbb{E}(X \,|\, \mathcal{F}_{n-1}) = M_{n-1}$. An interesting questions here is whether $M_n \to M_\infty := \mathbb{E}(X \,|\, F_\infty)$ a.s.. $M_\infty$ can be thought of as the best prediction we can ever make.

**Definition 10** *A process $C = (C_n : n \in \mathbb{N})$ is said to be* previsible *if it is $\mathcal{F}_{n-1}$ measurable. The* martingale transform *of $X$ by $C$ is defined by*

$$(C \circ X)_n := \sum_{k=1}^{n} C_k \Delta X_k = \sum_{k=1}^{n} C_k (X_k - X_{k-1}).$$

Remark: $X$ does not have to be a martingale in the above definition. $C \circ X$ is the discrete time analogue of the stochastic integral $\int C \, dX$. Indeed, $C \circ X$ is comparable to the Riemann integral if $X$ is a deterministic sequence. $(C \circ X)_n$ can be thought of as the winning after $n$ plays of the game, where at play $k$ a bet of $C_k$ is made (that's why it needs to be $\mathcal{F}_{n-1}$ measurable) and $X_k - X_{k-1}$ is the outcome of game $k$.

**Theorem 1** *(a) If $M$ is a martingale, $C$ is previsible and bounded, then $(C \circ M)_n$ is also a martingale, zero at $n = 0$;*
*(b) if $X$ is a supermartingale and $C$ is previsible, bounded and non-negative, then $C \circ X$ is also a supermartingale with $\mathbb{E}((C \circ X)_n) \leq 0$.*

`Proof:`(a) Write $Y = C \circ M$. $Y$ is obviously adapted, and since $|C| \leq c$ for some $c$, we have

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

Since $C_n$ is $\mathcal{F}_{n-1}$-measurably, we have

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

Hence $Y$ is a martingale.
(b) Obvious since for a supermartingale $X$, $\mathbb{E}(X_n - X_{n-1} \,|\, \mathcal{F}_{n-1}) \leq 0$. ■

**Definition 11** *A map $T : \Omega \to \{0, 1, 2, \ldots, \infty\}$ is called a* stopping time *or* optional random variable *if $\{T \le n\}$ is $\mathcal{F}_n$ measurable. The* stopped process *is written $X^T = (X_{T\wedge n} : n \in \mathbb{Z}^+)$.*

Remark: This requirement is equivalent to $\{T = n\}$ is $\mathcal{F}_n$ measurable. $T$ can be thought of as the time when you decide to stop the game. In order to decide whether to stop the game at time $n$, you can only consider information up to and including time $n$.
**Example**: $T = \inf\{n \ge 0 : X_n \in B\}$ is a stopping time, $L = \sup\{n : n \le 10, X_n \in B\}$ is not.

**Theorem 2** *Let $X_n$ be a martingale (resp. supermartingale), $T$ a stopping time. Then $X^T$ is also such. Hence $\mathbb{E}(X_{n\wedge T}) = \mathbb{E}(X_0)$ (resp. $\mathbb{E}(X_{n\wedge T}) \le \mathbb{E}(X_0)$).*

Note: cannot omit "$n\wedge$", e.g. simple symmetric random walk (SSRW) $X$ on $\mathbb{Z}$ is a martingale. Let $T := \inf\{n : X_n = 1\}$. Then it is well known that $\mathbb{P}(T < \infty) = 1$. But $\mathbb{E}(X_{n\wedge T}) = \mathbb{E}(X_0)$ and $1 = \mathbb{E}(X_T) = \mathbb{E}(X_0) = 0$.
`Proof:`Let $C_n := \mathbb{1}_{T\ge n}$. It is previsible as $\{C_n = 0\} = \{T \le n-1\}$. Moreover,

$$(C \circ X)_n = \sum_{k=1}^{n} \mathbb{1}_{n\le T}(X_k - X_{k-1}) = \sum_{k=1}^{n\wedge T}(X_k - X_{k-1}) = X_{T\wedge n} - X_0.$$

Hence $C \circ X$ is a martingale (resp. supermartingale) by Theorem 1. ■

**Theorem 3 (Doob's Optional Stopping Theorem)** *Let $X_n$ be martingale (resp. supermartingale), $T$ a stopping time. Then $\mathbb{E}(X_T) = \mathbb{E}(X_0)$ (resp. "$\le$") if one of these conditions holds:*

*a. $T$ bounded;*

*b. $X$ bounded, $T$ finite a.s.;*

*c. $\mathbb{E}(T)$ finite, jumps of $X_n$ bounded uniformly.*

`Proof:`We shall prove this for the supermartingale case. Observe that $\mathbb{E}(X_{n\wedge T} - X_0) \le 0$. For (a), we take $n = \sup T$ and the conclusion follows. For (b), we use the Bounded Convergence Theorem and let $n \to \infty$. For (c), we observe that

$$|X_{n\wedge T} - X_0| = \left|\sum_{k=1}^{n\wedge T}(X_k - X_{k-1})\right| \le T(\sup_n |X_n - X_{n-1}|).$$

Since $\mathbb{E}(T(\sup_n |X_n - X_{n-1}|)) < \infty$, we can use the Dominated Convergence Theorem to obtain the desired conclusion. ■

**Corollary 1** *If $M$ is a martingale with bounded increments (i.e. $\exists K$, $|M_n - M_{n-1}| \le K$ for all $n$), $C$ is previsible and bounded, $T$ is a stopping time such that $\mathbb{E}(T) < \infty$, then*

$$\mathbb{E}[(C \circ M)_T] = 0.$$

**Example (Asymmetric simple random walk)**: Let $S_n = X_1 + \ldots + X_n$ with $\mathbb{P}(X_i = 1) = p > 1/2$ and $\mathbb{P}(X_i = -1) = q = 1 - p$. We define $M_n = (q/p)^{S_n}$. Since $M_n = \prod_{i=1}^{n}(q/p)^{X_i}$ and $\mathbb{E}((q/p)^{X_i}) = p\frac{q}{p} + q\frac{p}{q} = 1$, $M$ is a martingale (w.r.t. the filtration $\mathcal{F}_n = \sigma(S_1, \ldots, S_n) = \sigma(X_1, \ldots, X_n)$).

Let $T_x = \inf\{n : S_n = x\}$ and $T = T_a \wedge T_b$ for $a < 0 < b$. It can be shown that $T < \infty$ (for a specific case of this proof, see the next example). By the optional stopping theorem,

$$1 = M_0 = \mathbb{E}(M_{T\wedge n}).$$

Now $M_{T\wedge n}$ is bounded between $(q/p)^a$ and $(q/p)^b$, so by the bounded convergence theorem, $\mathbb{E}(M_{T\wedge n}) \to \mathbb{E}(M_T)$ as $n \to \infty$, so that

$$1 = \mathbb{E}(M_T) = \mathbb{P}(T_a < T_b)(\frac{q}{p})^a + (1 - \mathbb{P}(T_a < T_b))(\frac{q}{p})^b.$$

Solving the above yields

$$\mathbb{P}(T_a < T_b) = \frac{(q/p)^b - 1}{(q/p)^b - (q/p)^a}.$$

This also means for $a < 0 < b$,

$$\mathbb{P}(T_a < \infty) = \mathbb{P}(\min_n S_n < a) = (q/p)^{-a},\ \mathbb{P}(T_b < \infty) = 1,$$

so that one can also calculate explicitly $\mathbb{E}\,(\min_n S_n)$, which is $> -\infty$. We can also use the martingale method to calculate $\mathbb{E}\,(T_b)$. Since $X_n = S_n - (p-q)n$ is also a martingale, the optional stopping theorem implies that

$$0 = \mathbb{E}\,(S_{T_b \wedge n} - (p-q)(T_b \wedge n)).$$

Now $S_{T_b\wedge n}$ is bounded above by $b$ and below by $\min_m S_m$, both of which have finite expectation, so by the dominated convergence theorem (careful), $\mathbb{E}\,(S_{T_b\wedge n}) \to \mathbb{E}\,(S_{T_b})$ as $n \to \infty$. The monotone convergence theorem implies that $\mathbb{E}\,(T_b \wedge n) \to \mathbb{E}\,(T_b)$. Hence

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

**Example (SSRW)**: Let $S_n = X_1 + \ldots + X_n$ be a SSRW, then $S_n$ is a martingale (w.r.t. the filtration $\mathcal{F}_n = \sigma(S_1, \ldots, S_n) = \sigma(X_1, \ldots, X_n)$). Set

$$T := \inf\{n : S_n = 1\}.$$

We hope to calculate the distribution of $T$.

Define

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

then $M_n^{(\theta)} := \prod_{i=1}^n (e^{\theta X_i}/\cosh\theta)$. Since $\mathbb{E}\,(e^{\theta X_i}/\cosh\theta) = \frac{1}{2}(e^\theta + e^{-\theta})/\cosh\theta = 1$, $M^{(\theta)}$ is a martingale. By the optional stopping theorem,

$$1 = \mathbb{E}\,(M_{T\wedge n}^{(\theta)}) = \mathbb{E}\,(e^{\theta S_{T\wedge n}}/(\cosh\theta)^{T\wedge n}).$$

In order to apply the bounded convergence theorem, we make two observations: (1) $e^{\theta S_{T\wedge n}}$ is bounded above by $e^\theta$ since $S_{T\wedge n}$ is 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

$$1 = \mathbb{E}\,(M_T^{(\theta)}) = \mathbb{E}\,(e^{\theta S_T}/(\cosh\theta)^T) = \mathbb{E}\,(e^\theta/(\cosh\theta)^T).$$

Hence

$$\mathbb{E}\,((\cosh\theta)^{-T}) = e^{-\theta}$$

for $\theta > 0$ and $T \in [0, \infty]$. If $T = \infty$, then $(\cosh\theta)^{-T} = 0$ for $\theta > 0$. But if $T < \infty$, then $(\cosh\theta)^{-T} \uparrow 1$ as $\theta \downarrow 0$. So bounded convergence theorem as $\theta \downarrow 0$ implies

$$\mathbb{P}(T < \infty) = \mathbb{E}\,(\mathbb{1}_{T<\infty}) = 1.$$

Hence we can ignore the possiblity $T = \infty$.

Put $\alpha = 1/\cosh\theta = 2/(e^\theta + e^{-\theta})$, then $e^{-\theta} = \frac{1}{\alpha}(1 - \sqrt{1-\alpha^2})$ (the equation $\alpha = 2/(x + 1/x)$ has two roots, of which this is one; the other root leads to a function that cannot be a probability generating function), hence

$$\mathbb{E}\,(\alpha^T) = \frac{1}{\alpha}(1 - \sqrt{1-\alpha^2}) = \sum_n \alpha^n \mathbb{P}(T = n).$$

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

$$\mathbb{P}(T = 2m-1) = (-1)^{m+1}\binom{1/2}{m}.$$

**Example (ABRACADABRA)**: At each times $1, 2, 3, \ldots$, a monkey types a capital letter (out of a choice of 26) at random (i.e. independently of the letters he has typed previously). We would like to find out how long one expects to wait before the phrase ABRACADABRA to come out of the typewriter.

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 $n^{\text{th}}$ gambler (hence $M_k^{(n)} = 0$ for $k < n$ since the $n^{\text{th}}$ 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 $\mathbb{E}(T) < \infty$. Doob's optional stopping theorem implies

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

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

# 4 Martingale Convergence

In this chapter, we are concerned with almost sure convergence of supermartingales (of which martingales are a special case) as $n \to \infty$. We will deal with $\mathcal{L}^1$ convergence in the next chapter. We first define upcrossing $U_N[a,b]$ of the interval $[a,b]$ by $X_1, \ldots, X_N$. (Refer to Figure 11.1 in Williams). Let $C_1 = \mathbb{1}_{X_0<a}$ and recursively define

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

Then

$$(C \circ X)_N = \sum_{k=1}^{N} C_k \Delta X_k \ge (b-a)U_N[a,b] - (X_N - a)^-.$$

**Lemma 1 (Doob's Upcrossing Lemma)** *Let $X$ be a supermartingale. Then*

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

`Proof:`Since $C$ is previsible, bounded (by $N$) and non-negative, we can conclude that $C \circ X$ is a supermartingale. Thus $\mathbb{E}\left((C\circ X)_N\right) \le 0$, from which the desired conclusion follows. ■

Define $U_\infty[a,b] = \lim_{N\uparrow\infty} U_N[a,b]$. The following can be proved using the Monotone Convergence Theorem.

**Corollary 2** *Suppose $X$ is a supermartingale bounded in $\mathcal{L}^1$ (i.e. $\sup_n \mathbb{E}\left(|X_n|\right) < \infty$). Then*

$$(b-a)\mathbb{E}\left(U_\infty[a,b]\right) \le |a| + \sup_n \mathbb{E}\left(|X_n|\right) < \infty,$$

*which implies that*

$$P(U_\infty[a,b] = \infty) = 0.$$

**Theorem 4 (Martingale Convergence Theorem)** *Suppose $X$ is a supermartingale bounded in $\mathcal{L}^1$. Then $X_\infty := \lim_{n\to\infty} X_n$ exists and is finite almost surely.*

`Proof:`Set $\Lambda_{a,b} = \{\omega : \liminf X_n(\omega) < a < b < \limsup X_n(\omega)\}$. We observe that $\Lambda_{a,b} \subset \{U_\infty[a,b] = \infty\}$, which has probability 0 by the corollary above. But since

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

we conclude that

$$\mathbb{P}(X_n \text{ converges to some } X_\infty \in [-\infty,+\infty]) = 1.$$

Now Fatou's Lemma shows that $\mathbb{E}\left(|X_\infty|\right) = \mathbb{E}\left(\liminf |X_n|\right) \le \liminf \mathbb{E}\left(|X_n|\right) \le \sup \mathbb{E}\left(|X_n|\right) < \infty$, whence $\mathbb{P}(|X| < \infty) = 1$. ■

Remark: might be not true $X_n \to X$ in $\mathcal{L}^1$ (e.g. see critical branching process below).

**Corollary 3** *A non-negative supermartingale $X$ converges a.s.. The limit $X_\infty$ satisfies $\mathbb{E}\left(X_\infty\right) \le \mathbb{E}\left(X_0\right)$.*

`Proof:`Trivial since $\mathbb{E}\left(|X_n|\right) = \mathbb{E}\left(X_n\right) \le \mathbb{E}\left(X_0\right)$. ■

The following example shows that one can only obtain $\le$ in the above corollary even if $X$ is a martingale.
**Example**: Let $S_n = 1 + X_1 + \ldots + X_n$ be a SSRW started at 1, $T := T_0$ be the first time $S_n$ hits 0, and $Y_n = S_{T\wedge n}$. Then $Y$ is a non-negative martingale, hence it converges to a limit $Y_\infty < \infty$ a.s.. Indeed, $Y_\infty \equiv 0$. Therefore we have $\mathbb{E}\left(Y_n\right) = \mathbb{E}\left(Y_0\right) = 1$ but $\mathbb{E}\left(Y_\infty\right) = 0$.
**Example (branching process)**: Let $X_i^n$, $n, i \ge 0$, be i.i.d. nonnegative integer-valued random variables. Define a sequence $(Z_n; n \ge 0)$ by $Z_0 = 1$ 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. The idea is that $Z_n$ represents the number of individuals in the $n^{\text{th}}$ generation, each of whom gives birth to an i.i.d. number of children. The distribution of $X_i^n$ is known as the *offspring distribution*.

Let $\mu = \mathbb{E}\left(X_i^n\right)$, then $Z_n/\mu^n$ is a martingale by the following calculation:

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

Now we will establish that the process $Z$ dies out (i.e. it becomes 0 eventually) if $\mu \leq 1$. For $\mu < 1$, we observe that

$$\mathbb{P}(Z_n > 0) \leq \mathbb{E}\left(Z_n \mathbb{1}_{Z_n \geq 0}\right) = \mathbb{E}\left(Z_n\right) = \mu^n.$$

Since $\mu < 1$, $\sum_{n=1}^{\infty} \mathbb{P}(Z_n > 0) < \infty$, therefore by the Borel-Cantelli lemma, $\mathbb{P}(Z_n > 0 \ i.o.) = 0$, i.e. $Z_n = 0$ eventually a.s.. Note that we did not use martingale theory for this.

For $\mu = 1$ (critical branching process), the situation is slightly trickier and we shall use the martingale convergence theorem. In this case, $Z_n$ is a non-negative martingale, therefore it converges to a limit $Z_\infty$ a.s.. Since $Z$ is integer valued, we must have $Z_n = Z_\infty$ for all sufficiently large $n$ a.s.. Therefore it must be that $Z_\infty \equiv 0$. (Note that again, $\mathbb{E}\left(Z_n\right) = 1$ for all $n$, but the $Z_\infty \equiv 0$.)

`Proof:` The number of particles $Z_n$ is a non-negative integer-valued martingale, so $Z_n = Z_\infty$ for large $n$. The only possibility is $Z_n = 0$ for large $n$. ■

## 5 Uniformly Integrable Martingales

**Definition 12** *A class $\mathcal{C}$ of random variables is said to be* uniformly integrable (UI) *if given $\varepsilon > 0$, there exists $K \in [0, \infty)$ such that*

$$\mathbb{E}\left(|X| \mathbb{1}_{|X|>K}\right) < \varepsilon, \ \forall X \in \mathcal{C}.$$

*$M$ is said to be a UI martingale if $M$ is a martingale and the family $(M_n : n \in \mathbb{Z}^+)$ is UI.*

A UI family $(X_n)$ is bounded in $\mathcal{L}^1$ (i.e. $\exists K, \mathbb{E}\left(|X_n|\right) < K$ for all $n$), by the following calculation:

$$\mathbb{E}\left(|X_n|\right) = \mathbb{E}\left(|X_n| \mathbb{1}_{|X_n| \geq K}\right) + \mathbb{E}\left(|X_n| \mathbb{1}_{|X_n| < K}\right) \leq 1 + K.$$

The converse is not true, i.e. a family of random variables bounded in $\mathcal{L}^1$ is not necessarily UI, by the following couter-example.
**Example**: Let $\Omega = [0, 1]$ with uniform probability measure. Define $X_n = n \mathbb{1}_{(0,1/n)}$. Then $\mathbb{E}\left(|X_n|\right) = 1$, but for any $K > 0$, we have for $n > K$,

$$\mathbb{E}\left(|X_n| \mathbb{1}_{|X_n| \geq K}\right) = n\mathbb{P}((0, 1/n)) = 1.$$

Here, $X_n \to 0$, but $\mathbb{E}\left(X_n\right) \not\to 0$.
An important example of a UI family is the following

**Theorem 5** *Let $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. Then the class*

$$\{\mathbb{E}\left(X \mid \mathcal{G}\right) : \mathcal{G} \textit{ is a sub-}\sigma\textit{-field of } \mathcal{F}\}$$

*is uniformly integrable.*

`Proof:`Let $\epsilon$ be given. By a standard result (see e.g. Lemma 13.1 in Williams), we can choose $\delta > 0$ so that, for $F \in \mathcal{F}$, $\mathbb{P}(F) < \delta$ implies that $\mathbb{E}\left( \mid X \mid \mathbb{1}_F\right) < \epsilon$. We also choose $K$ so that $\mathbb{E}\left(|X|\right) < K\delta$.
Now let $\mathcal{G}$ be a sub-$\sigma$-field of $\mathcal{F}$ and $Y = \mathbb{E}\left(X \mid \mathcal{G}\right)$. Then by Jensen's inequality,

$$|Y| \leq \mathbb{E}\left(|X| \mid \mathcal{G}\right). \tag{1}$$

Hence $\mathbb{E}\left(|Y|\right) \leq \mathbb{E}\left(|X|\right)$ and

$$K\mathbb{P}(|Y| > K) \leq \mathbb{E}\left(|Y|\right) \leq \mathbb{E}\left(|X|\right),$$

so that

$$\mathbb{P}(|Y| > K) < \delta.$$

By (1),

$$\mathbb{E}\left(|Y| \mathbb{1}_{|Y| \geq K}\right) \leq \mathbb{E}\left(|X| \mathbb{1}_{|Y| \geq K}\right) < \epsilon.$$

Hence $\{\mathbb{E}\left(X \mid \mathcal{G}\right) : \mathcal{G}$ is a sub-$\sigma$-field of $\mathcal{F}\}$ is uniformly integrable. ∎

The reason we study UI martingales is a sequence of UI random variables $X_n$ that converges to $X$ in probability will also converge in $\mathcal{L}^1$ if and only if $(X_n)$ is a UI class. We state the following theorem without proof, which mainly involves some analysis.

**Theorem 6** *Let $(X_n)$ be a sequence in $\mathcal{L}^1$ and let $X_\infty \in \mathcal{L}^1$. Then $X_n \to X_\infty$ in $\mathcal{L}^1$ (i.e. $\mathbb{E}\left(|X_n - X_\infty|\right) \to 0$), if and only if the following two conditions hold:*

1. *$X_n \to X_\infty$ in probability,*
2. *the sequence $(X_n)$ is UI.*

The above theorem guarantees that a UI martingale $X_n$ that converges almost surely also have the corresponding convergence property for expectations.

**Corollary 4** *Let $M$ be a UI martingale on $(\Omega, \mathcal{F}, \mathbb{P})$. Then $M_\infty := \lim_{n\to\infty} M_n$ exists a.s. and furthermore, as $n \to \infty$,*

$$\mathbb{E}\left(|M_n - M_\infty|\right) \to 0.$$

*Moreover, $M_n = \mathbb{E}\left(M_\infty \mid \mathcal{F}_n\right)$.*

`Proof:`The $\mathcal{L}^1$ convergence to $X_\infty$ is immediate by Theorem 6. Only the last part is nontrivial. For that, we take $F \in \mathcal{F}_n$ and $r \geq n$. The martingale property of $M$ implies that

$$\mathbb{E}\,(M_r \mathbb{1}_F) = \mathbb{E}\,(M_n \mathbb{1}_F). \tag{2}$$

But

$$|\mathbb{E}\,(M_r \mathbb{1}_F) - \mathbb{E}\,(M_\infty \mathbb{1}_F)| \leq \mathbb{E}\,(|M_r - M_\infty| \mathbb{1}_F) \leq \mathbb{E}\,(|M_r - M_\infty|),$$

which converges to 0 as $r \to \infty$. Plugging this into (2) yields that $\mathbb{E}\,(M_\infty \mathbb{1}_F) = \mathbb{E}\,(M_n \mathbb{1}_F)$, which implies the desired property since this holds for all $F \in \mathcal{F}_n$. ■

**Theorem 7 (Levy's 'Upward' Theorem)** *Let $\xi$ be an $\mathcal{L}^1$ random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ and define $M_n := \mathbb{E}\,(\xi \,|\, \mathcal{F}_n)$. Then $M$ is a UI martingale and $M_n \to \eta := \mathbb{E}\,(\xi \,|\, \mathcal{F}_\infty)$ almost surely and in $\mathcal{L}^1$.*

`Proof:`We know $M$ is a martingale because of the tower property. It is UI by Theorem 5. Hence $M_\infty := \lim_{n\to\infty} M_n$ exists a.s. and in $\mathcal{L}^1$. It only remains to show that $M_\infty = \eta$. We first note that both $M_\infty$ and $\eta$ are $\mathcal{F}_\infty$ measurable, hence in order to show that $M_\infty = \eta$ a.s., it suffices to show that for all $F \in \mathcal{F}_\infty$,

$$\mathbb{E}\,(M_\infty \mathbb{1}_F) = \mathbb{E}\,(\eta \mathbb{1}_F).$$

Since $\mathcal{F}_\infty = \sigma(\cup_n \mathcal{F}_n)$, it suffices to show that for all $n$ and $F \in \mathcal{F}_n$, the above identitiy holds. This follows from the following calculation:

$$\mathbb{E}\,(\eta \mathbb{1}_F) = \mathbb{E}\,(\xi \mathbb{1}_F) = \mathbb{E}\,(M_n \mathbb{1}_F) = \mathbb{E}\,(M_\infty \mathbb{1}_F),$$

the last equality due to Corollary 4. (Aside: how to show the following: if $X$ and $Y$ are both $\mathcal{F}$ measurable and $\mathbb{E}\,(X\mathbb{1}_F) = \mathbb{E}\,(Y\mathbb{1}_F)$ for all $F \in \mathcal{F}$, then $X = Y$ a.s.. Since $0 = \mathbb{E}\,((X-Y)\mathbb{1}_{X-Y>0}) \geq \sum_{n=1}^\infty \frac{1}{n} \mathbb{1}_{1/n \leq X-Y < 1/(n-1)}$, we must have $\mathbb{P}(1/n \leq X - Y < 1/(n-1)) = 0$ for all $n \geq 1$. Hence $\mathbb{P}(X > Y) = 0$, and similarly $\mathbb{P}(X < Y) = 0$. ■

**Example**: Let $Z_1, Z_2, \ldots$ be i.i.d. random variables with mean 0 and $\mathbb{E}\,(|Z_i|) < \infty$. Let $\theta$ be an independent random variable with finite mean. Let $Y_i = Z_i + \theta$ and $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$. The distribution of $\theta$ is called the *prior distribution*, and $\mathbb{P}(\theta \in \cdot \,|\, \mathcal{F}_n)$ is called the *posterior distribution* after $n$ observations. Then we can use martingale techniques to show that $\mathbb{E}\,(\theta \,|\, \mathcal{F}_n) \to \theta$ a.s. as $n \to \infty$.
`Proof:`$M_n := \mathbb{E}\,(\theta \,|\, \mathcal{F}_n)$ is a UI martingale and therefore converges to some $M_\infty$ both a.s. and in $\mathcal{L}^1$, and furthermore $M_\infty = \mathbb{E}\,(\theta \,|\, \mathcal{F}_\infty)$, which is $= \theta$ if we can show that $\theta$ is $\mathcal{F}_\infty$ measurable. This is indeed the case since for an arbitrary $a$,

$$\{\theta \leq a\} = \bigcap_{k=1}^\infty \left\{ \frac{1}{n} \sum_{i=1}^n Y_i \leq a + \frac{1}{k} \ \textit{eventually} \right\} \in \mathcal{F}_\infty.$$

Hence $M_\infty = \theta$ a.s.. ■

UI Martingales can be used to give a simple proof of the Kolmogorov's 0-1 law (it is also possible to prove this just using measure theory), which concerns events in the tail $\sigma$-field $\mathcal{T}$, defined as

$$\mathcal{T}_n := \sigma(X_{n+1}, X_{n+2}, \ldots), \ \mathcal{T} := \bigcap_n \mathcal{T}_n.$$

The tail $\sigma$-field is a sub-$\sigma$-field of $\mathcal{F}_\infty$ and contains many important events about the limiting behaviour of the sequence $X_n$, e.g. (you will show this in the homework)

$$\begin{aligned}
F_1 &:= \{\omega : \lim_n X_n(\omega) \ \textit{exists}\} \\
F_2 &:= \{\omega : \sum_n X_n(\omega) \ \textit{converges}\} \\
F_3 &:= \{\omega : \frac{1}{n} \sum_{k=1}^n X_k(\omega) \leq c\}.
\end{aligned}$$

Examples of random variables measurable in $\mathcal{T}$ include:

$$\liminf X_k, \ \limsup \frac{1}{n} \sum_{k=1}^n X_k.$$

Note that $\lim X_k$ is not necessary measurable in $\mathcal{T}$, since it may not even exist.

**Theorem 8 (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 $\sigma$-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 it is independent of $\mathcal{T}_n$ (two $\sigma$-fields $\mathcal{G}_1$ and $\mathcal{G}_2$ are independent if any $G_1 \in \mathcal{G}_1$ and $G_2 \in \mathcal{G}_2$ are independent). Let $F \in \mathcal{T} \subset \mathcal{F}_\infty$ and $\eta = \mathbb{1}_F$. Then Levy's Upward Theorem says

$$\eta = \mathbb{E}\,(\eta \,|\, \mathcal{F}_\infty) = \lim \mathbb{E}\,(\eta \,|\, \mathcal{F}_n).$$

But $\eta$ is also independent of every $\mathcal{F}_n$, hence

$$\mathbb{E}\,(\eta \,|\, \mathcal{F}_n) = \mathbb{E}\,(\eta) = \mathbb{P}(F).$$

Hence $\eta = \mathbb{P}(F)$. Since $\eta$ can only take on values 0 or 1, the result follows. ■

**Corollary 5** *Let $X_1, X_2, \ldots$ be an independent sequence and $\mathcal{T}$ be the tail $\sigma$-field. Suppose $Y$ is a random variable measurable in $\mathcal{T}$, then there exists $k \in [-\infty, \infty]$ such that $\mathbb{P}(Y = k) = 1$.*

`Proof`: Let $k := \inf\{y : \mathbb{P}(Y \leq y) = 1\}$ (infimum of an empty set is $+\infty$). Then

$$\mathbb{P}(Y \leq y) = \begin{cases} 0, & \text{if } y < k \\ 1, & \text{if } y \geq k \end{cases}.$$

■

**Examples**: If $X_1, X_2, \ldots$ is a sequence of independent random variables, Kolmogorov's 0-1 law implies that

$$\mathbb{P}(\sum X_n \textit{ converges}) = 0 \textit{ or } 1.$$

If $S_n = X_1 + \ldots + X_n$, then

$$Z_1 := \liminf_n \frac{S_n}{n},\ Z_2 := \limsup_n \frac{S_n}{n}$$

are almost surely constant.

## 6 L2 Martingales

A martingale $M$ is said to be bounded in $\mathcal{L}^2$ if $\sup_n \mathbb{E}(M_n^2) < \infty$. It is often easier to work in $\mathcal{L}^2$ because of the following Pythagorean formula

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

This can be proved using the fact that for $0 < s < t$, $\mathbb{E}(M_t - M_s \mid \mathcal{F}_s) = 0$.

**Theorem 9** *Let $M$ be a martingale for which $M_n \in \mathcal{L}^2$ for all $n$. Then $M$ is an $\mathcal{L}^2$ martingale if and only if $\sum_{k=1}^{\infty} \mathbb{E}[(M_k - M_{k-1})^2] < \infty$. In this case, $M_n \to M_\infty$ almost surely and in $\mathcal{L}^2$.*

## 7 Inequalities for martingales

Set $Z_n^* = \sup_{1\le k\le n} Z_k$ .

**Lemma 2 (Doob's submartingale inequality)** *Let $Z \geq 0$ sub-mart, $c > 0$. Then*

$$c\mathbb{P}(Z_n^* \geq u) \leq \mathbb{E}(Z_n \mathbb{1}_{Z_n^* \geq u}) \leq \mathbb{E} Z_n$$

`Proof:`Let $A_k = \{Z_0, \ldots, Z_{k-1} < u \leq Z_k\}$, $A_0 = \{Z_0 \geq u\}$, that is $A_k = \{W = k\}$ where $W = \inf\{k \geq 0 : \ Z_k \geq u\}$. Note $A_k \in \mathcal{F}_k$, $Z \geq u$ on $A_k$. Hence $\mathbb{E}(Z_n \mathbb{1}_{A_k}) \geq \mathbb{E}(Z_k \mathbb{1}_{A_k}) \geq u\mathbb{P}(A_k)$. Sum $\sum_{k=1}^{n}$ and note $A_k$'s are disjoint and sum up to $\{\sup_{k\le n} Z_k \geq u\}$. ■

Young inequality: if $a, b > 0$, $p^{-1} + q^{-1} = 1$, then $ab \leq \frac{1}{p}a^p + \frac{1}{q}b^q$.
`Proof:`Since log is concave $\log(ab) = \frac{1}{p}\log a^p + \frac{1}{q}\log b^q \leq \log\left(\frac{1}{p}a^p + \frac{1}{q}b^q\right)$. ■

Let $\|X\|_p = \sqrt[p]{\mathbb{E}|X|^p}$.
Hölder inequality: if $p, q > 1$: $1/p + 1/q = 1$, then $\mathbb{E}|XY| \leq \|X\|_p \|Y\|_q$
`Proof:`Using Young inequality,

$$\frac{\mathbb{E}|XY|}{\|X\|_p\|Y\|_q} = \mathbb{E}\left(\frac{|X|}{\|X\|_p}\frac{|Y|}{\|Y\|_q}\right) \leq \mathbb{E}\left(\frac{|X|^p}{p(\|X\|_p)^p} + \frac{|Y|^q}{q(\|Y\|_q)^q}\right) = \frac{1}{p} + \frac{1}{q} = 1$$

■

**Lemma 3** *$X, Y \geq 0$ r.vs. with $u\mathbb{P}(X \geq u) \leq \mathbb{E}(Y\mathbb{1}_{X\geq u})$ $\forall u > 0$, and $p, q$ as above. Then*

$$\|X\|_p \leq q\|Y\|_p$$

`Proof:`Using Fubini a few times and that $p/(p-1) = q$, and Hölder, we have

$$\mathbb{E} X^p = \int_0^\infty pu^{p-1}\mathbb{P}(X \geq u)du \leq \int_0^\infty pu^{p-2}\mathbb{E}(Y\mathbb{1}_{X\geq u})du = \mathbb{E}(Y\int_0^X pu^{p-2}dx) = q\mathbb{E}(X^{p-1}Y) \leq q\|Y\|_p\|X^{p-1}\|_q$$

Since $(p-1)q = p$, $\|X^{p-1}\|_q = \mathbb{E}(X^p)^{1/q}$, hence above is equivalent to lemma since $1 - 1/q = 1/p$. ■

**Theorem 10 (Doob inequality)** *Let $p, q$ as before. Let $Z \geq 0$ sub-mart bounded in $\mathcal{L}^p$. Set $Z^* = \sup_k Z_k = \lim Z_n^*$. Then $\mathbb{Z}^*$ is in $\mathcal{L}^p$ and*

$$(*)\ \|Z_n^*\|_p \leq q\|Z_n\|_p \ \Rightarrow\ \|Z^*\|_p \leq q\sup_k \|Z_k\|_p$$

*Moreover, $Z_\infty = \lim Z_n$ exists and*

$$(**)\ \|Z_\infty\|_p = \sup_k \|Z_k\|_p = \lim_{k\uparrow\infty} \|Z_k\|_p$$

`Proof:`From two Lemmas above, $\|Z_n^*\|_p \leq q\|Z_n\|_p \leq \sup_k q\|Z_k\|_p$. Now (*) follows from Monotone CT. Next, $-Z$ is super-mart bounded in $\mathcal{L}^p$ and hence in $\mathcal{L}^1$ whence the limit exists. ■

# 8 Martingale criteria from the "red book"

Define Markov chain $X_n$ on countable state space $S$, recurrence, transience.

**Theorem 11** *Irreducible, aperiodic, homogeneous MC is recurrent iff* $\exists\ f : A \mapsto \mathbb{R}^+$ *and a* <u>*finite*</u> *set* $A \subset S$ *such that*

$$\mathbb{E}\left(f(X_{n+1}) - f(X_n)\,|\,X_n = s\right) \leq 0 \text{ if } s \notin A$$

*and* $f(s) \to \infty$.

`Proof:`(⇒) Start chain at $X_0 = \alpha \notin A$, and let $\tau_\alpha$ be the time of first entry into $A$, and set $Y_n = f(X_{n\wedge\tau_\alpha})$. $Y_n \geq 0$ is supermartingale, hence converges a.s. to $Y = \lim_{n\to\infty} Y_n$, moreover from Fatou $\mathbb{E}\,Y \leq Y_0 = f(\alpha)$. If the chain were <u>not</u> recurrent, there would exist an $\alpha$ such that $\mathbb{P}(\forall n\ X_n \notin A$ and $f(X_n) \to \infty\,|\,X_0 = \alpha) > 0$. This implies $\mathbb{P}(Y_n \to \infty) > 0$ leading to a contradiction. Hence MC is recurrent.

(⇐) Let us prove the existence of $f$ in case MC is recurrent. Suppose $S = \{0, 1, 2 \dots\}$. Set 0 be the absorbing state: $p_{00} = 1$. Let

$$\phi(i, n) = \mathbb{P}(\exists k :\ X_k \in \{n, n+1, \dots\}\,|\,X_0 = i).$$

Recurrence implies that for each $i$ we have $\phi(i, n) \to 0$ as $n \to \infty$. For each $k$ set $n_k$ s.t. $\sup_{i\in\{1,..,k\}} \phi(i, n_k) < 2^{-k}$. Observe that for a fixed $n$

$$\mathbb{E}\left(\phi(X_{m+1}, n) - \phi(X_m, n)\,|\,X_m = j\right) \leq 0 \text{ for } j \neq 0.$$

Define now $f(i) = \sum_{k=1}^{\infty} \phi(i, n_k)$ which verifies the conditions. ■

**Theorem 12** *MC is transient iff* $\exists\ f : A \mapsto \mathbb{R}^+$ *and a set* $A \subset S$ *such that*

$$\begin{aligned} \mathbb{E}\left(f(X_{n+1}) - f(X_n)\,|\,X_n = s\right) &\leq 0 \text{ if } s \notin A \\ f(\alpha^*) &< \inf_{x\in A} f(x) \text{ for at least one } \alpha^* \notin A \qquad (*) \end{aligned}$$

*and* $f(s) \to \infty$.

`Proof:`(⇒) Let $\tau_{\alpha^*}$ same as before (first entry into $A$ time) and suppose chain is <u>not</u> transient, then $\mathbb{P}(\tau_{\alpha^*} < \infty) = 1$, hence $X_{n\wedge\tau_{\alpha^*}} \to X_{\tau_{\alpha^*}}$ a.s., and by Fatou's we get $\mathbb{E}\,f(X_{\tau_{\alpha^*}}) \leq \mathbb{E}\,f(X_0) = f(\alpha^*)$ since $f(X_{n\wedge\tau_{\alpha^*}})$ is supermartingale. This contradicts $(*)$ however.

(⇐) Fix any $\alpha_0 \in S$ and set $A = \{\alpha_0\}$. Because of transience, there is an $\alpha^*$ s.t. $\mathbb{P}(\forall n\ X_n \neq \alpha_0\,|\,X_0 = \alpha^*) > 0$. Define

$$f(s) = \begin{cases} 1, & \text{if } s = \alpha_0, \\ \mathbb{P}(\text{ever to reach } \alpha_0\,|\,X_0 = s), & \text{othewise.} \end{cases}$$

Then $\mathbb{E}\left(f(X_{n+1}) - f(X_n)\,|\,X_n = s\right) = 0$ unless $s = \alpha_0$, and $f(\alpha^*) < 1$. Hence $f$ is what we need. ■

# 9 School lecture on monkey typing Shakespeare

Monkey typing the complete works of Shakespeare (WS) amounts to typing a particular sequence of $N$ symbols on a typewriter (with $M$ possible characters). A monkey types a symbol per unit of time, producing an infinite sequence of i.i.d. r.v.'s with values $\{1, 2, \ldots, M\}$. Prob(the monkey eventually types out WS) $= 1$.

Simpler example: coin (fair) tossing, 0 (head) or 1 (tail), outcome of tosses $X_0, X_1, \ldots, X_i$ is a "random variable", $\mathbb{P}(X_i = 0) = \mathbb{P}(X_i = 1) = 1/2$, notion of independence, $\mathbb{P}(X_0 = 0, X_1 = 0) = \ldots$,

Now we count the number of tosses $T$ required for 1 (tail) to appear, $\mathbb{P}(T = 1) = 1/2, \mathbb{P}(T = 2) = 1/4, \ldots, \sum_{n=1}^{\infty} \mathbb{P}(T = n) = 1$, therefore the probability that eventually a tail appears is 1. What if the coin is not fair, e.g. $\mathbb{P}(X_i = 0) = 1 - p, \mathbb{P}(X_i = 1) = p$, then $\mathbb{P}(T = 1) = p, \mathbb{P}(T = 2) = (1 - p)p, \mathbb{P}(T = 3) = (1 - p)^2 p, \ldots, \sum_{n=1}^{\infty} \mathbb{P}(T = n) = \sum_{n=1}^{\infty} (1 - p)^{n-1} p = p \frac{1}{1-(1-p)} = 1$, so whether the coin is fair or not makes not difference. How can show that Prob(the monkey eventually types out WS) $= 1$?

A slightly more difficult question: how long do I need to wait before the first tail appears in a sequence of coin tosses? Notion of expected value (long-run average), e.g. $\mathbb{E}(X_0) = 1/2$ for a fair coin, $\mathbb{E}(X_0) =$ for an unfair coin, $\frac{1}{n} \sum_{k=1}^{n} X_k$ approaches the expected value of $X_0$ as $n$ becomes large. What about the expected time of waiting for a tail for a fair coin: $\mathbb{E}(Y) = \sum_{n=1}^{\infty} n \mathbb{P}(Y = n) = \sum_n n \frac{1}{2^n} = \ldots = 2$.

At each toss of the fair coin, I bet £1 on tail, let $X_n$ be my winning on the $n$th toss, then $\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = 1/2$ and $\mathbb{E}(X_n) = 0$ (fair game). Let $Y_n = X_1 + \ldots + X_n$, then $\mathbb{E}(Y_n) = 0$ for all $n$, i.e. $Y$ is a martingale. A more intriguing result is that $\mathbb{E}(Y_T) = 0$. This amounts for stopping the game when the first tail appears. And

$$0 = \mathbb{E}(Y_T) = \mathbb{E}(\sum_{n=1}^{T} X_n) = \mathbb{E}((-1)(T - 1) + 1),$$

which means $\mathbb{E}(T) = 2$.

What if the coin is not fair? Bet £1 on tail each time, but be compensated $\pounds(1 - 2p)$ after each toss. Then

$$0 = \mathbb{E}(Y_T) = \mathbb{E}(\sum_{n=1}^{T} X_n) = \mathbb{E}((-1 + 1 - 2p)(T - 1) + 1 + 1 - 2p),$$

which means $\mathbb{E}(T) = 1/p$.

A slightly trickier question: if I toss a fair coin repeatedly, how long do I expect to wait until the first TT appears?

$$0 = \mathbb{E}(M_T) = \mathbb{E}\left(\sum_{n=1}^{T} M_T^{(n)}\right) = \mathbb{E}((2^2 - 1) + (2 - 1) + (-1)(T - 2)),$$

so $\mathbb{E}(T) = 6$.

**Example (ABRACADABRA)**: At each times $1, 2, 3, \ldots$, a monkey types a capital letter (out of a choice of 26) at random (i.e. independently of the letters he has typed previously). We would like to find out how long one expects to wait before the phrase ABRACADABRA to come out of the typewriter.

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

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

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

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

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

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

and so through the sequence ABRACADABRA. Let $M^{(n)}$ be the winning of the $n^{\text{th}}$ gambler (hence $M_k^{(n)} = 0$ for $k < n$ since the $n^{\text{th}}$ 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 $\mathbb{E}(T) < \infty$. Doob's optional stopping theorem implies

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

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