CHAPTER 1

Martingales: Basic Properties and Stopping Theory

1. The framework

Martingales are a class of real-valued *stochastic processes* in discrete time. The mathematical model for a stochastic process has two ingredients,

- a *probability space* $(\Omega, \mathcal{A}, \mathbb{P})$ consisting of a set Ω endowed with
  - a σ-algebra $\mathcal{A} \subset \mathcal{P}(\Omega)$,
  - and a probability measure $\mathbb{P} : \mathcal{A} \to [0, 1]$,
- a sequence $X = \{X_1, X_2, X_3, \ldots\}$ of random variables, modelled as measurable mappings $X_n : \Omega \to \mathbb{R}$.

The σ-algebra $\mathcal{A}$ consists of all events, i.e. the outcomes of the random experiment, which can be observed. The fact that $X_1, X_2, X_3, \ldots$ are measurable ensures that we can assign probabilities, for example, to the following events,

- $\big\{X_n \in A\big\}$, for $A \subset \mathbb{R}$ an open (or even Borel) set,
- $\big\{X_n \in A \text{ for infinitely many } n\big\} = \bigcap_{m=1}^\infty \bigcup_{n=m}^\infty \{X_n \in A\}$,
- $\big\{\lim_{n\to\infty} X_n = c\big\} = \bigcap_{m=1}^\infty \bigcup_{n=m}^\infty \bigcap_{j=n}^\infty \big\{|X_j - c| < \frac{1}{m}\big\}$.

The exact form of the underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is irrelevant in most cases, at the end of the day it is just a tool to consistently assign probabilities to events, which involve several of the random variables $X_1, X_2, X_3, \ldots$.

2. Information and filtrations

If we have *information* about the outcome of a random event, this can be described by a subset $\mathcal{F} \subset \mathcal{A}$ consisting of all the events about which we know whether they have happened or not. It is easy to check that this subset must satisfy the axioms of a σ-algebra. In other words,

$$\boxed{\text{Information can be modelled by a sub-}\sigma\text{-algebra } \mathcal{F} \subset \mathcal{A}.}$$

Examples (1) Suppose $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{A} = \mathcal{P}(\Omega)$ and $\mathbb{P}$ given by $\mathbb{P}\{i\} = 1/6$ for all $i \in \Omega$. This is the model for throwing a die. The information whether *an even number* was thrown is encoded in the σ-algebra $\mathcal{F} = \big\{\emptyset, \{2, 4, 6\}, \{1, 3, 5\}, \Omega\big\}$.

(2) Suppose $X : \Omega \to \{1, \ldots, n\}$ and $Y : \Omega \to \{1, \ldots, m\}$ are two random variables on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. If we have observed the value for Y we know whether the events $\{Y = k\} = Y^{-1}\{k\}$, $k = 1, \ldots, m$, have taken place and thus we have received information encoded in the σ-algebra

$$\sigma(Y) := \big\{Y^{-1}(A) \, : A \subset \{1, \ldots, m\}\big\}.$$

(3) In our example of the stochastic process X, if we have observed $X_1, X_2, \ldots, X_n$ we let $\mathcal{F}_n := \sigma(X_1, \ldots, X_n)$ be the *smallest σ-algebra that makes the random variables $X_1, X_2, \ldots, X_n : \Omega \to \mathbb{R}$ measurable.* This σ-algebra represents the information available from observing the process up to time n.

In the last example we have seen that a stochastic process X gives rise to a whole nested sequence of σ-algebras $\mathcal{F}_n, n \geq 0$, which model the increase of information obtained from observing a longer and longer part of the process. This situation is so important that it deserves a definition of its own.

Definition A sequence $\{\mathcal{F}_n : n \in \mathbb{N}\}$ of σ-algebras such that $\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{A}$, for every $n \in \mathbb{N}$ is called a *filtration*. As seen in example (3) any stochastic process gives rise to a filtration $\{\mathcal{F}_n : n \in \mathbb{N}\}$, which is called *the natural filtration* induced by the process.

3. Conditional expectation

The second step is, given information, how does this change the probability of events? Note first that it is obvious that probabilities change when additional information is present. For example in Example (1), if we know that an even number was thrown, the new (conditional) probability that three was thrown is zero, and the probabilities that two, four or six were thrown must go up accordingly.

Let us look more closely at Example (2). Suppose that $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space and X and Y are random variables on this space. We assume for simplicity that both random variables take on finitely many real values $\{1, \ldots, n\}$ resp. $\{1, \ldots, m\}$ each with positive probability. We suppose that X and Y are *not* independent. How does knowledge about the outcome of X influence the outcome of Y? This can be described by means of *conditional probabilities*

$$\mathbb{P}\big\{X = i \,\big|\, Y = j\big\} := \frac{\mathbb{P}\big\{X = i \text{ and } Y = j\big\}}{\mathbb{P}\{Y = j\}}.$$

One way to look at this is the following: if we have observed the event $\{Y = j\}$ this changes our perception of the random variable X, it is now defined *on a different probability space*, on which only those ω can occur, which satisfy $Y(\omega) = j$. The new space still consists of the set Ω and the σ-algebra $\mathcal{A}$, but the probability measure is now $\mathbb{P}\{\cdot \,|\, Y = j\}$, given by

$$\mathbb{P}\big\{A \,\big|\, Y = j\big\} := \frac{\mathbb{P}\big(A \cap \{Y = j\}\big)}{\mathbb{P}\{Y = j\}}.$$

The random variable X is still be defined on the space, but its distribution has changed and is now the conditional distribution given $Y = j$. Its expectation is now

$$\mathbb{E}\big\{X \,\big|\, Y = j\big\} := \sum_{k=1}^{n} k\, \mathbb{P}\big\{X = k \,\big|\, Y = j\big\}.$$

This *conditional expectation* depends on j and can be interpreted as a *random variable* on Ω namely by

$$\mathbb{E}\big\{X \,\big|\, Y\big\} : \Omega \to \mathbb{R}, \qquad \mathbb{E}\big\{X \,\big|\, Y\big\}(\omega) = \mathbb{E}\big\{X \,\big|\, Y = Y(\omega)\big\}.$$

The random variable $\mathbb{E}\{X|Y\}$ has the following properties :

(1) it is measurable with respect to the σ-algebra $\sigma(Y)$,

(2) for every $A \subset \{1, \ldots, m\}$ we have $\displaystyle\int_{Y^{-1}(A)} \mathbb{E}\{X|Y\}(\omega) d\mathbb{P}(\omega) = \int_{Y^{-1}(A)} X(\omega) d\mathbb{P}(\omega)$.

The first property means that it is constant on each set $Y^{-1}\{j\}$ for $j = 1, \ldots, m$. The second property follows from the calculation,

$$\int_{Y^{-1}(A)} \mathbb{E}\{X|Y\}(\omega)\, d\mathbb{P}(\omega) = \sum_{j \in A} \mathbb{E}\{X|Y = j\}\mathbb{P}\{Y = j\} = \sum_{j \in A} \sum_{k=1}^{n} k\mathbb{P}\{X = k \,|\, Y = j\}\mathbb{P}\{Y = j\}$$

$$= \sum_{j \in A} \sum_{k=1}^{n} k\, \mathbb{P}\{X = k \text{ and } Y = j\} = \int \sum_{j \in A} \sum_{k=1}^{n} k\, 1_{\{X(\omega)=k\}} 1_{\{Y(\omega)=j\}}\, d\mathbb{P}(\omega)$$

$$= \int X(\omega) \sum_{k=1}^{n} 1_{\{X(\omega)=k\}} \sum_{j \in A} 1_{\{Y(\omega)=j\}}\, d\mathbb{P}(\omega) = \int_{Y^{-1}(A)} X(\omega)\, d\mathbb{P}(\omega)\,.$$

In this example the additional information we had consisted of the knowledge of the cell $Y^{-1}\{j\} \subset \Omega$ in which the random ω was to be found.

What we need is an extension of the notion of conditional expectation to general σ-algebras representing information. We have formulated the properties of the conditional expectation in the special case in such a way that it can be used as a basis for a new definition.

Theorem 1.1. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\mathcal{F} \subset \mathcal{A}$ a sub-σ-algebra and X a random variable with $\mathbb{E}|X| < \infty$. Then there is a random variable $Z = \mathbb{E}\{X|\mathcal{F}\}$ with $\mathbb{E}\big|Z\big| < \infty$ called* the conditional expectation of X given $\mathcal{F}$ *with the following two properties:*

1) $\mathbb{E}\{X|\mathcal{F}\}$ *is $\mathcal{F}$-measurable,*

2) *for all $F \in \mathcal{F}$,* $\displaystyle\int_F \mathbb{E}\{X|\mathcal{F}\}\, d\mathbb{P} = \int_F X\, d\mathbb{P}$.

Any two random variables Z satisfying these two conditions coincide almost surely.

If $\mathcal{F} = Y^{-1}(\mathcal{A}')$ is the σ-algebra of preimages of all Borel sets under a random variable Y, one also writes $\mathbb{E}\{X|Y\}$ for $\mathbb{E}\{X|\mathcal{F}\}$ and says this is the conditional expectation of X given Y.

By choosing $F = \Omega$ in the last property, we see that $\mathbb{E}\big\{\mathbb{E}\{X|\mathcal{F}\}\big\} = \mathbb{E}X$.

Handout 1 gives a proof of almost sure uniqueness and existence for conditional expectations. More can be found in Williams, Chapter 9. We now formulate some important properties of conditional expectations, which can be proved from the definition and uniqueness.

Theorem 1.2. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\mathcal{F} \subset \mathcal{A}$ a sub-σ-algebra and X, Y random variables with $\mathbb{E}|X|, \mathbb{E}|Y| < \infty$. Then the conditional expectations have the following properties.*

Linearity: *For all a, b real, almost surely,*

$$\mathbb{E}\big\{aX + bY \mid \mathcal{F}\big\} = a\mathbb{E}\{X \mid \mathcal{F}\} + b\mathbb{E}\{Y \mid \mathcal{F}\}\,.$$

Positivity: *If $X \geq 0$, then $\mathbb{E}\{X|\mathcal{F}\} \geq 0$ almost surely.*

Monotone Convergence: *If $0 \leq X_n \uparrow X$, then $\mathbb{E}\{X_n|\mathcal{F}\} \uparrow \mathbb{E}\{X|\mathcal{F}\}$ almost surely.*

Fatou: *If* $0 \leq X_n$ *and* $\mathbb{E}\{X_n|\mathcal{F}\} < \infty$, *then*

$$\mathbb{E}\{\liminf_{n\to\infty} X_n \mid \mathcal{F}\} \leq \liminf_{n\to\infty} \mathbb{E}\{X_n \mid \mathcal{F}\} \text{ almost surely.}$$

Dominated Convergence: *If there is a random variable* Z *such that* $\mathbb{E}Z < \infty$ *and* $|X_n| \leq Z$ *for all* n, *and if* $X_n \to X$ *almost surely, then* $\mathbb{E}\{X_n|\mathcal{F}\} \to \mathbb{E}\{X|\mathcal{F}\}$.

As an EXERCISE you should try to prove at least two of the statements.

The next set of properties is of more probabilistic nature. Try to get an intuition for the meaning of the statement.

THEOREM 1.3. *Let* $(\Omega, \mathcal{A}, \mathbb{P})$ *be a probability space,* $\mathcal{F} \subset \mathcal{A}$ *a sub-*σ*-algebra and* X *a random variable with* $\mathbb{E}|X| < \infty$. *Then every conditional probability* $\mathbb{E}\{X|\mathcal{F}\}$ *has the following properties.*

Tower property: *If* $\mathcal{G} \subset \mathcal{F}$ *is a sub-*σ*-algebra, then* $\mathbb{E}\Big\{\mathbb{E}\{X|\mathcal{F}\}\Big|\mathcal{G}\Big\} = \mathbb{E}\{X|\mathcal{G}\}$ *almost surely.*

Taking out what is known, TOWIK: *If* Z *is* $\mathcal{F}$*-measurable and bounded, then* $\mathbb{E}\{ZX|\mathcal{F}\} = Z\mathbb{E}\{X|\mathcal{F}\}$ *almost surely. If* X *itself is* $\mathcal{F}$*-measurable, then* $\mathbb{E}\{X|\mathcal{F}\} = X$ *almost surely.*

Independence: *If* X *is independent of* $\mathcal{F}$, *then* $\mathbb{E}\{X \mid \mathcal{F}\} = \mathbb{E}\{X\}$ *almost surely.*

Proof: By linearity we can assume that $X \geq 0$ in all parts of the proof.

For the *tower property* note that the left hand side is $\mathcal{G}$ measurable and, for every $G \in \mathcal{G} \subset \mathcal{F}$,

$$\int_G \mathbb{E}\Big\{\mathbb{E}\{X|\mathcal{F}\}\Big|\mathcal{G}\Big\} d\mathbb{P} = \int_G \mathbb{E}\{X|\mathcal{F}\} d\mathbb{P} = \int_G X d\mathbb{P}\,.$$

Hence the integrand on the left hand side satisfies all the properties of conditional expectation $\mathbb{E}\{X|\mathcal{G}\}$ and must be a conditional expectation of X given $\mathcal{G}$.

To check *TOWIK*, we observe that $Z\mathbb{E}\{X|\mathcal{F}\}$ is $\mathcal{F}$-measurable and integrable. We just have to show, for every $F \in \mathcal{F}$,

$$\int_F ZX\, d\mathbb{P} = \int_F Z\mathbb{E}\{X|\mathcal{F}\}\, d\mathbb{P}\,,$$

then it follows that $Z\mathbb{E}\{X|\mathcal{F}\}$ is a conditional expectation of ZX given $\mathcal{F}$. The given equality holds for $Z = 1_H$, $H \in \mathcal{F}$, by definition and follows for bounded $\mathcal{F}$-measurable Z by an application of the monotone class theorem.

The particular case mentioned at the end of TOWIK holds, because in the given situation, X itself satisfies the conditions of a conditional expectation of X given $\mathcal{F}$.

For the *independence property* observe that the constant function $\mathbb{E}\{X\}$ is $\mathcal{F}$-measurable. Recall that, if X is independent of $\mathcal{F}$ this means that X is independent of every $\mathcal{F}$-measurable random variable Y, and hence $\mathbb{E}\{XY\} = \mathbb{E}\{X\}\mathbb{E}\{Y\}$. In particular, for $F \in \mathcal{F}$, we get

$$\int_F X(\omega)\, d\mathbb{P}(\omega) = \mathbb{E}\{1_F X\} = \mathbb{E}\{1_F\}\, \mathbb{E}\{X\} = \mathbb{P}(F)\mathbb{E}\{X\} = \int_F \mathbb{E}\{X\}\, d\mathbb{P}(\omega)\,.$$

This proves that the constant function $\mathbb{E}\{X\}$ is a conditional probability of X given $\mathcal{F}$. ■

4. Martingales and super-martingales

In this chapter we get to know the processes which correspond to fair games, the martingales. Symmetric random walk and the critical Galton-Watson process (with expected offspring number one) turn out to be important examples. The word *martingale* originally denotes a special gambling strategy, indicating the connection to fair games.

Let $\{X_n : n \geq 0\}$ be a stochastic process in discrete time and let $\{\mathcal{F}(n) : n \geq 0\}$ be a *filtration*, i.e. an increasing sequence

$$\mathcal{F}(0) \subset \mathcal{F}(1) \subset \mathcal{F}(2) \subset \ldots \subset \mathcal{A}$$

of σ-algebras. We call the process $\{X_n\}$ an $\{\mathcal{F}(n)\}$-*adapted* process, if

$$X_n \text{ is } \mathcal{F}(n)\text{-measurable, for all } n \geq 0,$$

In most cases we consider the natural filtration. Clearly, every process is adapted to its natural filtration.

Definition
A discrete time process $\{X_n : n \geq 0\}$ is called a *martingale* relative to the filtration $\{\mathcal{F}(n)\}$ if

- $\{X_n\}$ is $\{\mathcal{F}(n)\}$-adapted,
- $\mathbb{E}\{|X_n|\} < \infty$ for all n, and
- $\mathbb{E}\{X_n \mid \mathcal{F}(n-1)\} = X_{n-1}$ almost surely, for all $n \geq 1$.

If we have just $\leq$ in the last condition, then $\{X_n\}$ is called a *supermartingale*, if $\geq$ holds in the last condition, then $\{X_n\}$ is called a *submartingale* .

Let us convince ourselves, before starting the discussion of martingales, that there are plenty of interesting **examples**.

(1) Random walk Suppose that $X_1, X_2, \ldots$ are independent random variables with $\mathbb{E}|X_n| < \infty$ and $\mathbb{E}X_n = 0$. Let $S_n = \sum_{k=1}^n X_k$ be the partial sums and $\mathcal{F}(n)$ be the natural filtration of the $\{S_n\}$. Observe that this is also the natural filtration for the $\{X_n\}$. Then

$$\mathbb{E}\{S_n \mid \mathcal{F}(n-1)\} = \mathbb{E}\Big\{ \sum_{k=1}^{n-1} X_k + X_n \,\Big|\, \mathcal{F}(n-1)\Big\} = \sum_{k=1}^{n-1} X_k + \mathbb{E}X_n = S_{n-1} \text{ a.s.}$$

Hence the *random walk* $\{S_n\}$ is a martingale.

(2) Random product Suppose that $X_1, X_2, \ldots$ are independent nonnegative random variables with $\mathbb{E}X_n = 1$. Let $M_n = \prod_{k=1}^n X_k$. Let $\mathcal{F}(n)$ be the natural filtration. Then,

$$\mathbb{E}\big\{ M_n \,\big|\, \mathcal{F}(n-1)\big\} = M_{n-1}\mathbb{E}\big\{ X_n \big| \, \mathcal{F}(n-1)\big\} = M_{n-1}\mathbb{E}X_n = M_{n-1}\,.$$

Hence $\{M_n\}$ is a martingale.

(3) Accumulating data about a random variable Let the filtration $\{\mathcal{F}(n)\}$ be arbitrary and X an integrable random variable. Define $X_n = \mathbb{E}\{X|\mathcal{F}(n)\}$, one should interpret X_n as the data accumulated about X at time n. The martingale property of $\{X_n\}$ follows from the tower property of conditional expectation

$$\mathbb{E}\{X_n \mid \mathcal{F}(n-1)\} = \mathbb{E}\{\mathbb{E}\{X \mid \mathcal{F}(n)\} \mid \mathcal{F}(n-1)\} = \mathbb{E}\{X \mid \mathcal{F}(n-1)\} = X_{n-1}\,.$$

(4) Galton Watson process The *Galton Watson process* $\{X_n\}$ is constructed in the usual way. Choose an offspring distribution on the nonnegative integers characterized by the probability masses $(p_0, p_1, p_2, \ldots)$. There are independent random variables Y_k^n, $k, n \geq 1$, with $\mathbb{P}\{Y_k^n = i\} = p_i$, and define $\{X_n\}$ recursively by $X_0 = 1$ and

$$X_{n+1} = \sum_{k=1}^{X_n} Y_k^{n+1} \text{ for all } n \geq 0.$$

We assume that $\mu := \sum_{k=0}^{\infty} k p_k$, the expected offspring number, is finite. Let

$$\mathcal{F}(n) = \sigma(Y_k^m : m \leq n, k \geq 1)$$

and note that $\{X_n\}$ is adapted to the filtration $\{\mathcal{F}(n)\}$. Then

$$\mathbb{E}\big\{ X_{n+1} \,|\, \mathcal{F}(n) \big\} = \sum_{k=1}^{\infty} \mathbb{E}\big\{ 1_{k \leq X_n} Y_k^{n+1} \,|\, \mathcal{F}(n) \big\} = \sum_{k=1}^{\infty} 1_{k \leq X_n} \, \mathbb{E}\big\{ Y_k^{n+1} \big\} = \mu \, X_n.$$

where we have taken out what is known and used independence. Hence $\{X_n\}$ is a martingale if and only if the expected number of offspring is $\mu = 1$, and a supermartingale if $\mu \leq 1$.

(5) Polya's Urn At time 0, an urn contains one black ball and one white ball. At each time $n = 1, 2, \ldots$, a ball is chosen at random from the urn and is replaced together wit a new ball of the same colour. 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 chosen by time n. Let

$$M_n = \frac{B_n + 1}{n + 2}$$

be the proportion of black balls in the urn after time n. We argue that $\{M_n\}$ is a martingale in the natural filtration $\{\mathcal{F}(n)\}$ of $\{B_n\}$.

We consider a sequence of random variables $\{X_n\}$, where X_n takes the value 1 if a black ball is chosen at time n, and zero otherwise. Then $B_n = X_1 + \cdots + X_n$. We have $\mathbb{P}\{X_1 = 1\} = \mathbb{P}\{X_1 = 0\} = 1/2$ and

$$\mathbb{P}\big\{ X_n = 1 \,\big|\, B_{n-1} = k \big\} = \frac{k+1}{n+1} \text{ for } 0 \leq k \leq n-1, n \geq 2.$$

In terms of conditional expectations this means that

$$\mathbb{E}\big\{ X_n \,\big|\, \mathcal{F}(n-1) \big\} = \frac{B_{n-1} + 1}{n + 1} \text{ for } n \geq 2.$$

Now, for $n \geq 1$,

$$\begin{aligned}
\mathbb{E}\big\{ M_n \,|\, \mathcal{F}(n-1) \big\} &= \mathbb{E}\Big\{ \frac{B_{n-1} + X_n + 1}{n+2} \,\Big|\, \mathcal{F}(n-1) \Big\} \\
&= \frac{B_{n-1} + 1}{n+2} + \mathbb{E}\Big\{ \frac{X_n}{n+2} \,\Big|\, \mathcal{F}(n-1) \Big\} \\
&= \frac{B_{n-1} + 1}{n+2} + \frac{B_{n-1} + 1}{(n+1)(n+2)} = \frac{B_{n-1} + 1}{n+1} = M_{n-1}.
\end{aligned}$$

Let us discuss some consequences of the definition. First note that, for every martingale $\{X_n\}$, from the tower property of conditional expectation, for all $m < n$,

$$\mathbb{E}\{X_n|\mathcal{F}(m)\} = \mathbb{E}\{\mathbb{E}\{X_n|\mathcal{F}(n-1)\}\,|\,\mathcal{F}(m)\} = \mathbb{E}\{X_{n-1}|\mathcal{F}(m)\} = \ldots = X_m\,.$$

Taking expectations gives,

$$\mathbb{E}\{X_n\} = \mathbb{E}\{X_0\} \text{ for all } n.$$

It is immediate from the definition that, for every martingale $\{X_n\}$,

$$\mathbb{E}\{X_n - X_{n-1}|\mathcal{F}(n-1)\} = 0\,. \tag{4.1}$$

Considering $\{X_n\}$ to be the capital of a gambler at time n, this can be interpreted as saying that the game is *fair*, the expected profit in each step is 0. In the supermartingale case this is ≤ 0 and the game is unfavourable. Let us explore this interpretation a bit more.

Suppose that $\{C_n : n \geq 1\}$ is your stake on game n. You have to base your decision on C_n on the history of the game up to time $n-1$. Formally, C_n has to be $\mathcal{F}(n-1)$-measurable. We use this to **define:**

A process $\{C_n : n \geq 1\}$ is called *previsible* if C_n is $\mathcal{F}(n-1)$-measurable.

Your winnings in game n are then $C_n(X_n - X_{n-1})$ and the *total winnings up to time* n are given by

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

The process $\{(C \bullet X)_n\}$ is called the *martingale transform* of $\{X_n\}$ by $\{C_n\}$. The big question is now: Can you choose $\{C_n\}$ such that your expected total winnings are positive? A positive answer to this question would be the most useful result of this lecture, however, we can prove:

THEOREM 1.4 (You can't beat the system). *Let $\{C_n\}$ be a previsible process, which is bounded, i.e. there exists $C > 0$ such that $|C_n(\omega)| \leq C$ for all $n \geq 1$ and $\omega \in \Omega$. Then, if $\{X_n\}$ is a martingale, so is $\{(C \bullet X)_n\}$. Moreover, $(C \bullet X)_0 = 0$ and hence $\mathbb{E}\{(C \bullet X)_n\} = 0$ for all $n \geq 0$.*

Proof: It is clear that $(C \bullet X)_n$ is integrable (as C_n is bounded) and by definition the process $\{Y_n\} = \{(C \bullet X)_n\}$ is adapted to the filtration $\{\mathcal{F}(n)\}$ and starts in 0. We calculate

$$\mathbb{E}\{Y_n|\mathcal{F}(n-1)\} = \mathbb{E}\Big\{\sum_{k=1}^{n} C_k(X_k - X_{k-1})\Big|\mathcal{F}(n-1)\Big\}$$

$$= \sum_{k=1}^{n-1} C_k(X_k - X_{k-1}) + C_n\mathbb{E}\{X_n - X_{n-1}|\mathcal{F}(n-1)\} = Y_{n-1}\,.$$

This proves it all. ■

Remarks:

- The proof also shows, if also $C_n \geq 0$ and $\{X_n\}$ is a supermartingale, so is $\{(C \bullet X)_n\}$.
- In a continuous time setting the equivalent of the martingale transform is the *stochastic integral*, an advanced topic of great relevance.

5. Stopping Times and Doob's Optional Stopping Theorem

We now study stopping times for martingales. The intuition is that a stopping time is a random time such that the knowledge about a random process at time n suffices to determine whether the stopping time has happened at time n or not, in other words, at any time you know whether you have reached the stopping time or not.

Suppose that $\{\mathcal{F}(n)\}$ is a filtration on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. A map $T : \Omega \to \mathbb{N} \cup \{\infty\}$ is called a *stopping time* if

$$\{T \le n\} = \{\omega : T(\omega) \le n\} \in \mathcal{F}(n) \text{ for all } n < \infty .$$

One can equivalently require $\{T = n\} \in \mathcal{F}(n)$ for all $n < \infty$. This is easy to see, as the first definition implies $\{T = n\} = \{T \le n\} \setminus \{T \le n-1\} \in \mathcal{F}(n)$, and the second definition implies $\{T \le n\} = \bigcup_{k=0}^{n} \{T = k\} \in \mathcal{F}(n)$.

Interpreting a martingale $\{X_n\}$ as a fair game we interpret the stopping times as those instances when we can quit playing (and obtain your winnings or pay your losses). If we follow this strategy and play unit stakes up to time T and then quit playing, the stake process C is

$$C_n = \mathbf{1}_{\{n \le T\}} = 1 - \mathbf{1}_{\{T \le n-1\}},$$

which is $\mathcal{F}(n-1)$-measurable and hence $\{C_n\}$ is previsible. The winnings process is

$$(C \bullet X)_n = \sum_{k=1}^{n} C_k (X_k - X_{k-1}) = X_{T \wedge n} - X_0 ,$$

We define X^T, the *process X stopped at T*, as

$$X_n^T(\omega) = X_{T(\omega) \wedge n}(\omega) .$$

Recall that $X^T - X_0$ is the martingale transform of X by the (bounded) stake process C defined above. Theorem 1.4 can be applied and yields:

THEOREM 1.5 (Elementary stopping theorem). *If X is a martingale and T a stopping time, then X^T is a martingale. In particular,*

$$\mathbb{E}\{X_{T \wedge n}\} = \mathbb{E}\{X_0\} \textit{ for all } n .$$

If X is a supermartingale, then so is X^T and we still have

$$\mathbb{E}\{X_{T \wedge n}\} \le \mathbb{E}\{X_0\} \textit{ for all } n .$$

Example: We look at a *simple symmetric random walk.* Let $Y_1, Y_2, \ldots$ be a sequence of independent random variables with distribution $\mathbb{P}\{Y_n = 1\} = \mathbb{P}\{Y_n = -1\} = \frac{1}{2}$, and let

$$X_n = \sum_{k=1}^{n} Y_k .$$

We have seen that $\{X_n\}$ is a martingale (with respect to the natural filtration) and we have studied this process a little in previous probability lectures. Let

$$T = \inf\{n \ge 0 : X_n = 1\} .$$

Clearly, T is a stopping time. Probably, you already know that $\mathbb{P}\{T < \infty\} = 1$ from lectures about stochastic processes. The theorem states that

$$\mathbb{E}\{X_{T\wedge n}\} = \mathbb{E}\{X_0\} \text{ for all } n.$$

However, if $T < \infty$ almost surely, we have $X_T = 1$ almost surely and hence

$$1 = \mathbb{E}X_T \neq \mathbb{E}X_0 = 0\,.$$

In some sense this means that you *can* beat the system if you have an infinite amount of time and credit.

After this example one would very much like to see conditions which make sure that for nice martingales and stopping times we have $\mathbb{E}X_T = \mathbb{E}X_0$. This is the content of Doob's optional stopping theorem.

THEOREM 1.6 (Doob's optional stopping theorem). *Let T be a stopping time and X a martingale. Then X_T is integrable and*

$$\mathbb{E}\{X_T\} = \mathbb{E}\{X_0\}\,,$$

if one of the following conditions hold:

(1) *T is bounded (i.e. there is N such that $T(\omega) < N$ for all ω),*
(2) *T is almost surely finite and X is bounded, (i.e. there is a real K such that $|X_n(\omega)| < K$ for all n and ω),*
(3) *$\mathbb{E}\{T\} < \infty$ and there is $K > 0$ such that, for all n and ω, $|X_n(\omega) - X_{n-1}(\omega)| \leq K$.*

If $\{X_n\}$ is a super-martingale and either one of the three previous conditions or

(4) *X is nonnegative and T almost surely finite.*

holds, then X_T is integrable and $\mathbb{E}\{X_T\} \leq \mathbb{E}\{X_0\}$.

Proof: We assume that $\{X_n\}$ is a supermartingale. Then $\{X_{T\wedge n}\}$ is a supermartingale and, in particular, integrable, and

$$\mathbb{E}\{X_{T\wedge n} - X_0\} \leq 0\,.$$

For (1) the result follows by choosing $n = N$. For (2) let $n \to \infty$ and use dominated convergence. For (3) we observe that

$$|X_{T\wedge n} - X_0| = \Big|\sum_{k=1}^{T\wedge n}(X_k - X_{k-1})\Big| \leq KT\,.$$

By assumption KT is an integrable function and we can use dominated convergence again. For (4) we use Fatou's lemma to see

$$\mathbb{E}\{X_T\} = \mathbb{E}\{\liminf_{n\to\infty} X_{T\wedge n}\} \leq \liminf_{n\to\infty} \mathbb{E}\{X_{T\wedge n}\} = \mathbb{E}\{X_0\}\,.$$

The statement for martingales follows by applying the previous to the supermartingales $\{X_n\}$ and $\{-X_n\}$ separately. ∎

In the situation of our example these conditions must fail. A glance at the third condition gives a striking corollary.

Corollary 1.7. *For a simple random walk the expected first hitting time of level* 1 *is infinite.*

6. Examples and Applications

6.1. The first run of three sixes. A fair die is is thrown independently at each time step. A gambler wins a fixed amount of money as soon as the first run of three consecutive sixes appears. What is the mean number of throws of the dice until the gambler wins for the first time?

Let $X_1, X_2, \dots$ be the outcomes of the throws, i.e. an i.i.d. sequence of random variables with $\mathbb{P}\{X_i = k\} = 1/6$ for every $k \in \{1, \dots, 6\}$. Write $\mathcal{F}(n) = \sigma(X_1, \dots, X_n)$. Let T be the first time three consecutive sixes appear. Clearly, T is a stopping time, and we are looking for $\mathbb{E}\{T\}$.

We argue that $\mathbb{E}\{T\} < \infty$. Indeed, in order to observe $T = k$ we need to have one number different from six in every tuple $(X_{3m+1}, X_{3m+2}, X_{3m+3})$ for $3m + 3 < k$. The probability of this is $1 - 1/6^3 = 215/216$ and the events are independent, for each m. Hence

$$\mathbb{P}\{T = k\} \leq \left(\frac{215}{216}\right)^{\frac{k-3}{3}},$$

and we infer that $\mathbb{E}\{T\} = \sum_{k=1}^{\infty} k\mathbb{P}\{T = k\}$ converges. Note that this argument is based on a rough upper bound and does not enable us to find $\mathbb{E}\{T\}$.

We make the following thought experiment. Just before each time n a gambler appears on the scene and bets £1 that the nth throw will show six. If he loses, he leaves, if he wins he receives £6, all of which he bets on the event that the $(n + 1)$st throw shows six again. Again if he loses, he leaves, and otherwise he bets all his £36 on a six in the third throw, and so forth. As the game is fair, at any time $n \geq 3$ the expected winnings should be equal to the total money spent by the gamblers up to time n, which is n. As T is a stopping time satisfying condition (3) of Doob's optional stopping theorem, this should hold at time T as well, hence

$$\mathbb{E}\{T\} = 6 + 6^2 + 6^3 = 258.$$

Indeed, $\mathbb{E}\{T\}$ is the expected money spent by the gamblers, and at time T the last gambler has won £6, the one before has won £36 and the one before that has won the highest prize of £216. All other gamblers have lost their stakes.

To make this rigorous, we let S_n be the total stakes of all players at time n (i.e. $S_n = 1 + 6 + \cdots + 6^k$ if at time n we are in a run of k sixes) and let $M_n = S_n - n$, in particular $M_0 = 1$. Then $\{M_n\}$ is a martingale, as

$$\mathbb{E}\{M_{n+1} \mid \mathcal{F}(n)\} = (5/6)(1 - (n + 1)) + (1/6)(6S_n + 1 - (n + 1)) = S_n - n = M_n.$$

We look at the stopped martingale M^T. We have seen that $\mathbb{E}\{T\} < \infty$ and $|M_n^T - M_{n-1}^T| \leq 260$, hence by Doob's Optional Stopping Theorem, Part (3), we have

$$1 = \mathbb{E}\{M_0\} = \mathbb{E}\{M_T\} = 1 + 6 + 6^2 + 6^3 - \mathbb{E}\{T\}.$$

Remark: This problem can also be solved with Markov chain methods, but the argument is much less elegant and requires a lengthy calculation.

6.2. The gambler's ruin problem. Let $X_1, X_2, \dots$ be a sequence of i.i.d. random variables with $\mathbb{P}\{X_i = 1\} = p$, and $\mathbb{P}\{X_i = -1\} = q$, where $0 < p = 1 - q < 1$ and $p > q$. Let $\mathcal{F}(n) = \sigma(X_1, \dots, X_n)$ and pick integers $0 < a < b$. Let $S_n = a + X_1 + \cdots + X_n$, the intuition is that S_n represents the capital of a gambler, starting with a capital a in a favourable game. Suppose that the gambler stops playing when he is bankrupt or when his capital reaches b. The corresponding stopping time is

$$T = \inf\{n \in \mathbb{N} : S_n = 0 \text{ or } S_n = b\}.$$

We want to find the probability of ruin $\mathbb{P}\{S_T = 0\}$, the expected time $\mathbb{E}\{T\}$ when the game ends, and the expected capital $\mathbb{E}\{S_T\}$ at this time.

We first show that T satisfies $\mathbb{E}\{T\} < \infty$. This is easy with the same method as in the previous example: if $T = k$, any tuple $(X_{nb+1}, \dots, X_{(n+1)b})$ with $(n+1)b < k$ contains at least one step down. This can be used to observe that $\mathbb{P}\{T = k\}$ decreases exponentially fast.

Recall from the tutorials that

$$M_n = \left(\frac{q}{p}\right)^{S_n} \text{ and } N_n = S_n - n(p-q)$$

define martingales. We apply Theorem 1.6 (3) to $\{M_n\}$, noting that $|M_n^T - M_{n+1}^T| \le 1$, and get

$$\left(\frac{q}{p}\right)^a = \mathbb{E}\{M_0\} = \mathbb{E}\{M_T\} = \mathbb{P}\{S_T = 0\} + \left(\frac{q}{p}\right)^b \mathbb{P}\{S_T = b\}.$$

Note that $\mathbb{P}\{S_T = b\} = 1 - \mathbb{P}\{S_T = 0\}$. Hence

$$\mathbb{P}\{S_T = 0\} = \frac{\left(\frac{q}{p}\right)^a - \left(\frac{q}{p}\right)^b}{1 - \left(\frac{q}{p}\right)^b}.$$

From this we get

$$\mathbb{E}\{S_T\} = b\mathbb{P}\{S_T = b\} = b \times \frac{1 - \left(\frac{q}{p}\right)^a}{1 - \left(\frac{q}{p}\right)^b}.$$

Applying Theorem 1.6 (3) to $\{N_n\}$, noting that $|N_n - N_{n+1}| \le 1 + p + q$, we get

$$a = \mathbb{E}\{N_0\} = \mathbb{E}\{N_T\} = \mathbb{E}\{S_T - T(p-q)\} = \mathbb{E}\{S_T\} - \mathbb{E}\{T\}(p-q),$$

hence

$$\mathbb{E}\{T\} = \frac{1}{p-q}(\mathbb{E}\{S_T\} - a) = \frac{b}{p-q}\frac{1 - \left(\frac{q}{p}\right)^a}{1 - \left(\frac{q}{p}\right)^b} - \frac{a}{p-q}.$$

CHAPTER 2

Convergence Theorems for Martingales

1. Doob's Martingale Convergence Theorem

Doob's famous forward convergence theorem gives a sufficient condition for the almost sure convergence of martingales $\{X_n\}$ to a limiting random variable.

THEOREM 2.1 (Martingale Convergence Theorem). *Let $\{X_n\}$ be a supermartingale, which is bounded in L^1, i.e. there is $K > 0$ such that $\mathbb{E}|X_n| \le K$ for all n. Then there exists a real-valued random variable X on the same probability space such that*

$$\lim_{n\to\infty} X_n = X \text{ almost surely.}$$

Important remark: Note that if X_n is nonnegative, we have

$$\mathbb{E}|X_n| = \mathbb{E}\{X_n\} = \mathbb{E}\{X_0\} := K$$

and thus X_n is automatically bounded in L^1 and $\lim_{n\to\infty} X_n = X$ exists.

The idea of the proof is to count the number of upcrossings of each interval. Basically, if $\{X_n\}$ did not converge, it oscillates and hence crosses some interval $[a, b]$ infinitely often. Then we can run an anticyclic betting strategy: Whenever we are below a, we put unit stakes on an increase and once we are above b we have won an amount of $b - a$. But we can't beat the system....

Formally, pick two numbers $a < b$ and let $I = [a, b]$. The number $U_N[a, b]$ of upcrossings of $[a, b]$ made by $\{X_n\}$ up to time N is defined as the largest integer k such that there are integers

$$0 \le s_1 < t_1 < s_2 < t_2 < \cdots < s_k < t_k \le N$$

with $X_{s_i} < a$ and $X_{t_i} > b$ for all $1 \le i \le k$. Clearly, $U_N[a, b]$ is a random variable. The following lemma is the key idea of the proof.

LEMMA 2.2 (Doob's upcrossing lemma). *Let $\{X_n\}$ be a supermartingale and $a < b$. Then, for all N,*

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

Proof: We discuss this in the language of fair games. Suppose that we bet on the game $\{X_n\}$ with the following anti-cyclic strategy: Initially our stake C_0 is zero. We leave it like this until $X_n < a$ when we choose unit stake $C_{n+1} = 1$. We play unit stakes until $X_n > b$ when we stop and choose $C_{n+1} = 0$, keep it until $X_n < a$, and so fourth. Formally,

$$C_1 = 1_{\{X_0<a\}} \text{ and } C_n = 1_{\{C_{n-1}=1\}}1_{\{X_{n-1}\le b\}} + 1_{\{C_{n-1}=0\}}1_{\{X_{n-1}<a\}}\,.$$

C_n is clearly previsible and we study the martingale transform

$$Y_n = (C \bullet X)_n = \sum_{k=1}^{n} C_k (X_k - X_{k-1}) \, .$$

Recalling our strategy, we observe that

$$Y_N \geq (b-a) U_N[a,b] - (X_N - a)^- \, .$$

Each upcrossing increases the value of Y_n by at least $b-a$ and the last term is responsible for a possible unfinished upcrossing at the end. As C is previsible, nonnegative and bounded, Y is a supermartingale and

$$0 = \mathbb{E}\{Y_0\} \geq \mathbb{E}\{Y_N\} \geq (b-a)\mathbb{E}\{U_N[a,b]\} - \mathbb{E}\{(X_N - a)^-\} \, .$$

This is the claimed inequality. ∎

From this we can observe that the expected number of upcrossings is bounded if $\mathbb{E}\{(X_N - a)^-\}$ is bounded.

Lemma 2.3. *Let $\{X_n\}$ be a supermartingale, which is bounded in L^1. Then, for the increasing limit*

$$U[a,b] := \lim_{N \to \infty} U_N[a,b]$$

we have

$$(b-a)\mathbb{E}\{U[a,b]\} \leq |a| + \sup_n \mathbb{E}|X_n| < \infty \, .$$

In particular, $\mathbb{P}\{U[a,b] = \infty\} = 0$.

Proof: Recall that, by the upcrossing lemma,

$$(b-a)\mathbb{E}\{U_N[a,b]\} \leq \mathbb{E}\{(X_N - a)^-\} \leq |a| + \mathbb{E}|X_N| \leq |a| + \sup_n \mathbb{E}|X_n| \, .$$

Now let $N \to \infty$ using monotone convergence. ∎

We are almost done, look at the event

$$M[a,b] := \{\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n\}$$

and observe that if X_n neither converges nor diverges to $\pm\infty$, there are rationals $a < b$ such that event $M[a,b]$ takes place. But if $M[a,b]$ takes place, then $U[a,b] = \infty$, which has probability zero. Taking the union over the countable collection of rationals $a < b$ we obtain that almost surely $\{X_n\}$ converges to a possibly infinite random variable X. But, by Fatou's lemma,

$$\mathbb{E}|X| = \mathbb{E}\{\liminf_{n\to\infty} |X_n|\} \leq \liminf_{n\to\infty} \mathbb{E}|X_n| \leq K < \infty \, ,$$

hence $|X| < \infty$ almost surely. This finishes the proof of Doob's martingale convergence theorem.

Examples. **(a)** Let $\{X_n\}$ be a simple, symmetric random walk. Then $\{X_n\}$ is a martingale, which does not converge. Namely, by the recurrence of simple random walks, both level 0 and level 1 are reached infinitely often, contradicting convergence. Clearly, $\{X_n\}$ is not L^1-bounded.

(b) If $\mathcal{F}(0) \subset \mathcal{F}(1) \subset \ldots$ is a filtration and X is a nonnegative random variable, then we infer from the martingale convergence theorem that

$$\lim_{n\to\infty} \mathbb{E}\{X|\mathcal{F}(n)\}$$

exists almost surely. Under a suitable condition, more precisely if X is measurable with repect to the smallest σ-algebra containing $\bigcup \mathcal{F}(n)$, this limit agrees with X.

(c) We now look at the Galton-Watson process $\{X_n\}$ offspring distribution given by $(p_0, p_1, \ldots)$. Assume that the mean offspring number is

$$\sum_{n=0}^{\infty} np_n = 1 \, .$$

As $\{X_n\}$ is a nonnegative martingale it converges almost surely to a random variable $X \geq 0$. Because $\{X_n\}$ is integer-valued we must have $X_n = X$ for sufficiently large n almost surely. Now assume $p_1 < 1$ (in the case $p_1 = 1$ trivially $X_n = 1$ for all n). For every $k > 0$ we then have, for any K,

$$\mathbb{P}\{X_n = k \text{ for all } n \geq K\} \leq \lim_{M\to\infty} \prod_{n=K+1}^{M} \mathbb{P}\{X_n = k \,|\, X_{n-1} = k\} = 0 \, ,$$

hence $X = 0$ almost surely. In other words, *the critical Galton-Watson process becomes extinct in finite time.* However, recall that $\mathbb{E}\{X_n\} = 1$ for all n. The convergence in the martingale convergence theorem does not hold for the expectations.

2. Uniformly integrable martingales

The key to the difference between the two examples is the question when the almost sure convergence in the martingale convergence theorem can be replaced by L^1-convergence. We first study a general criterion for L^1-convergence of an arbitrary sequence of random variables. Recall that a sequence $\{X_n\}$ of random variables converges in L^1 to X iff $\mathbb{E}|X_n - X| \to 0$. This also implies that $\mathbb{E}\{X_n\} \to \mathbb{E}\{X\}$ by the triangle inequality.

Definition: A sequence $\{X_n\}$ of random variables is called *uniformly integrable* if, for every $\varepsilon > 0$ there is $K \geq 0$ such that

$$\int_{\{|X_n|>K\}} |X_n| \, d\mathbb{P} < \varepsilon \text{ for all } n \, .$$

THEOREM 2.4.

- (a) *Every bounded sequence of random variables is uniformly integrable.*
- (b) *If a sequence of random variables is dominated by an integrable, nonnegative random variable Y, i.e. if $|X_n| \leq Y$ for all n, then the sequence is uniformly integrable.*
- (c) *Let $p \geq 1$. A sequence is L^p-bounded if* $\sup_n \mathbb{E}\{|X_n|^p\} < \infty$. *Every sequence, which is L^p-bounded for some $p > 1$ is uniformly integrable.*
- (d) *Every uniformly integrable sequence is L^1-bounded.*
- (e) *There are L^1-bounded sequences, which are not uniformly integrable.*

Proof: Clearly, (a) follows from (b), which we now prove. Observe that,

$$\int_{\{|X_n|>K\}} |X_n| \, d\mathbb{P} \le \int_{\{|Y|>K\}} |Y| \, d\mathbb{P} .$$

Now, using monotone convergence,

$$\mathbb{E}|Y| = \liminf_{K\to\infty} \int 1_{\{|Y|\le K\}} |Y| \, d\mathbb{P} .$$

Hence $\lim_{K\to\infty} \int_{\{|Y|>K\}} |Y| \, d\mathbb{P} = 0$ and this implies uniform integrability of our family.
We come to (c) and suppose that $\sup_n \mathbb{E}\{|X_n|^p\} < C$ for some $p > 1$. Observe that, for every K,

$$\int_{\{|X_n|>K\}} |X_n| \, d\mathbb{P} \le K^{1-p} \int_{\{|X_n|>K\}} |X_n|^p \, d\mathbb{P} \le K^{1-p} C .$$

Choosing K large gives uniform integrability. (d) is easy, since for $\varepsilon = 1$ we find $K > 0$ with

$$\mathbb{E}|X_n| \le \int_{\{|X_n|>K\}} |X_n| \, d\mathbb{P} + \int_{\{|X_n|\le K\}} |X_n| \, d\mathbb{P} \le 1 + K .$$

Finally, to prove (e), suppose that U is uniformly distributed on $(0, 1)$ and let

$$X_n = n 1_{\{U \le 1/n\}} \ge 0 .$$

Then $\mathbb{E}|X_n| = 1$ and hence the family is L^1-bounded, but

$$\int_{\{|X_n|>K\}} |X_n| \, d\mathbb{P} = 1 \text{ for all } n > K ,$$

and hence the family cannot be uniformly integrable. ∎

Recall that every almost surely convergent sequence also converges *in probability*, which means that

$$\lim_{n\to\infty} \mathbb{P}\{|X_n - X| > \varepsilon\} = 0 \text{ for all } \varepsilon > 0 .$$

We can now state the key theorem of this section.

THEOREM 2.5 (Uniform Integrability Theorem). *Suppose that $\{X_n\}$ is a uniformly integrable sequence of random variables, which converges in probability to a random variable X. Then the sequence converges also in L^1.*

Proof: By Fatou's Lemma $\mathbb{E}|X| \le \liminf \mathbb{E}|X_n| < \infty$. For every $K \ge 0$ we define the cutoff-function

$$\varphi_K(x) = \begin{cases} K & \text{if } x > K \\ x & \text{if } |x| \le K \\ -K & \text{if } x < -K. \end{cases}$$

By uniform integrability one can choose K such that, for all n,

$$\mathbb{E}\{|\varphi_K(X_n) - X_n|\} < \frac{\varepsilon}{3} \text{ and } \mathbb{E}\{|\varphi_K(X) - X|\} < \frac{\varepsilon}{3} .$$

Since

$$|\varphi_K(X) - \varphi_K(X_n)| \le |X - X_n|$$

we infer that $\{\varphi_K(X_n)\}$ converges in probability to $\{\varphi_K(X)\}$. Hence there is N such that

$$\mathbb{P}\Big\{ |\varphi_K(X_n) - \varphi_K(X)| > \frac{\varepsilon}{6} \Big\} < \frac{\varepsilon}{12K} \text{ for all } n \geq N .$$

Then, for $n \geq N$,

$$\begin{aligned}
\mathbb{E}\{|\varphi_K(X_n) - \varphi_K(X)|\} &\leq \int_{\{|\varphi_K(X_n)-\varphi_K(X)|>\varepsilon/6\}} |\varphi_K(X_n) - \varphi_K(X)|\, d\mathbb{P} \\
&\quad + \int_{\{|\varphi_K(X_n)-\varphi_K(X)|\leq\varepsilon/6\}} |\varphi_K(X_n) - \varphi_K(X)|\, d\mathbb{P} \\
&\leq 2K\mathbb{P}\Big\{ |\varphi_K(X_n) - \varphi_K(X)| > \frac{\varepsilon}{6} \Big\} + \frac{\varepsilon}{6} \leq \varepsilon/3 .
\end{aligned}$$

We thus get, from the triangle inequality, for all $n \geq N$,

$$\mathbb{E}\{|X_n - X|\} \leq \mathbb{E}\{|X_n - \varphi_K(X_n)|\} + \mathbb{E}\{|\varphi_K(X_n) - \varphi_K(X)|\} + \mathbb{E}\{|X - \varphi_K(X)|\} \leq \varepsilon .$$

This completes the proof. ∎

Note that, by Theorem 2.4(b), the uniform integrability theorem is a generalization of the dominated convergence theorem.

We go back to the study of martingales. Let $\{\mathcal{F}(n)\}$ be a filtration and $\{X_n\}$ a martingale with respect to this filtration. The next theorem shows that every uniformly integrable martingale is of the form that data is accumulated about a random variable X.

THEOREM 2.6 (Convergence for uniformly integrable martingales). *Suppose that the martingale $\{X_n\}$ is uniformly integrable. Then there is an almost surely finite random variable X such that*

$$\lim_{n\to\infty} X_n = X \text{ almost surely and in } L^1.$$

Moreover, for every n, $X_n = \mathbb{E}\{X|\mathcal{F}(n)\}$.

Proof: Because $\{X_n\}$ is uniformly integrable, it is in particular L^1-bounded and thus, by the martingale convergence theorem, almost surely convergent to a real-valued random variable X. By the previous theorem this convergence holds also in the L^1-sense. To check the last assertion, we verify the two properties of conditional expectation. $\mathcal{F}(n)$-measurability of X_n is clear by definition, so let $F \in \mathcal{F}(n)$. For all $m \geq n$ we have, by the martingale property,

$$\int_F X_m \, d\mathbb{P} = \int_F X_n \, d\mathbb{P} .$$

We let $m \to \infty$. Then

$$\Big| \int_F X_m \, d\mathbb{P} - \int_F X \, d\mathbb{P} \Big| \leq \int |X_m - X| \, d\mathbb{P} \to 0 ,$$

hence we obtain, as required,

$$\int_F X \, d\mathbb{P} = \int_F X_n d\mathbb{P} .$$

∎

Example: **Exponential increase of a rabbit population.**

Suppose that $\{X_n\}$ is a Galton-Watson process. Recall that in the *critical case* $\mu = 1$ we have seen that the process *dies* almost surely in finite time. Now we give conditions that the process *grows exponentially* with positive probability. Assume that the offspring distribution (given by the sequence $\{p_n\}$) has

- mean $\mu = \sum_{n=0}^{\infty} np_n > 1$ (*supercritical case*),
- positive and finite variance $\sigma^2 = \sum_{n=0}^{\infty} n^2 p_n - \mu^2$.

Recall that we have seen earlier that

$$\mathbb{E}\big\{ X_{n+1} \,\big|\mathcal{F}(n)\big\} = \mu X_n,$$

and hence $M_n = X_n/\mu^n$ defines a martingale. As $M_k \geq 0$ there exists a random variable $M \geq 0$ with

$$\lim_{n\to\infty} M_n = M \text{ almost surely.}$$

We want to show that $\mathbb{P}\{M > 0\} > 0$. For this purpose we show that $\{M_n\}$ is uniformly integrable. If this holds we have

$$\mathbb{E}M = \lim_{n\to\infty} \mathbb{E}M_n = 1\,,$$

hence $M \geq 0$ cannot be identically zero so that $\mathbb{P}\{M > 0\} > 0$. To check uniform integrability it suffices to check L^2-boundedness (see Theorem 2.4c). We have, almost surely,

$$\begin{aligned}
\mathbb{E}\{M_n^2|\mathcal{F}(n-1)\} &= \mathbb{E}\{M_{n-1}^2|\mathcal{F}(n-1)\} + \mathbb{E}\{2M_{n-1}(M_n - M_{n-1})|\mathcal{F}(n-1)\} \\
&\quad + \mathbb{E}\{(M_n - M_{n-1})^2|\mathcal{F}(n-1)\} \\
&= M_{n-1}^2 + \mathbb{E}\{(M_n - M_{n-1})^2|\mathcal{F}(n-1)\}\,.
\end{aligned}$$

To compute the second term observe

$$\mathbb{E}\{(M_n - M_{n-1})^2|\mathcal{F}(n-1)\} = \mu^{-2n}\mathbb{E}\{(X_n - \mu X_{n-1})^2|\mathcal{F}(n-1)\}$$

and that

$$\mathbb{E}\Big\{ (X_n - \mu X_{n-1})^2 \,|\, \mathcal{F}(n-1)\Big\} = \mathbb{E}\Big\{ \Big(\sum_{k=1}^{X_{n-1}} (Y_k^n - \mu)\Big)^2 \Big|\mathcal{F}(n-1)\Big\}$$

$$= \sum_{k=1}^{\infty}\sum_{\ell=1}^{\infty} 1\{k \leq X_{n-1}\}\, 1\{\ell \leq X_{n-1}\}\; \mathbb{E}\big\{ (Y_k^n - \mu)(Y_\ell^n - \mu) \,|\, \mathcal{F}(n-1)\big\}$$

$$= \sum_{k=1}^{X_{n-1}} \mathbb{E}\big\{ (Y_k^n - \mu)^2\big\} = X_{n-1}\sigma^2\,.$$

Combining all terms we get,

$$\mathbb{E}\{M_n^2\} = \mathbb{E}\{M_{n-1}^2\} + (\sigma^2/\mu^{2n})\mathbb{E}\{X_{n-1}\}\,.$$

Now, $\mathbb{E}\{X_{n-1}\} = \mu^{n-1}\mathbb{E}\{M_{n-1}\} = \mu^{n-1}$. Hence,

$$\mathbb{E}\{M_n^2\} = \mathbb{E}\{M_{n-1}^2\} + \frac{\sigma^2}{\mu^{n+1}}\,,$$

which implies by induction,

$$\mathbb{E}\{M_n^2\} \leq 1 + \sigma^2 \sum_{k=1}^{\infty} \frac{1}{\mu^{k+1}} < \infty,$$

and we infer that the martingale $\{M_n\}$ is L^2-bounded. Altogether we have shown that, almost surely, for all large n,

$$X_n = M_n \mu^n \geq \frac{M}{2} \mu^n,$$

so that if $M > 0$ the Galton Watson process increases exponentially and the event $\{M > 0\}$ has positive probability.

3. Lévy's Upward and Downward Theorems

Theorem 2.5 shows that every uniformly integrable martingale is of the type that data about some (hidden) random variable is accumulated (see Example 3). Conversely, every martingale of this type is uniformly integrable and convergent to the hidden variable. This is the content of Lévy's upward theorem.

THEOREM 2.7 (Lévy's upward theorem). *Let X be an integrable random variable and $\{\mathcal{F}(n)\}$ be a filtration such that X is measurable with respect to the σ-algebra generated by the union of the $\mathcal{F}(n)$. Define $X_n = \mathbb{E}\{X|\mathcal{F}(n)\}$. Then $\{X_n\}$ is a uniformly integrable martingale and*

$$\lim_{n\to\infty} X_n = X \text{ almost surely and in } L^1.$$

The key to the proof of the theorem is the following lemma.

LEMMA 2.8. *Let X be an integrable random variable and $\{\mathcal{F}(n)\}$ be a sequence of sigma-algebras. If $X_n = \mathbb{E}\{X|\mathcal{F}(n)\}$, the sequence $\{X_n\}$ is uniformly integrable.*

Proof: Let $\varepsilon > 0$ be given. Choose $\delta > 0$ such that, for $F \in \mathcal{A}$,

$$\mathbb{P}(F) < \delta \text{ implies } \int_F |X|\, d\mathbb{P} < \varepsilon.$$

This is possible, since otherwise we could find a sequence $F(n) \in \mathcal{A}$ of events with $\mathbb{P}(F(n)) < 2^{-n}$ and $\int_{F(n)} |X| d\mathbb{P} \geq \varepsilon$. Then look at the event H that infinitely many of the events $F(n)$ happen. We have $\mathbb{P}(H) = 0$ as

$$\mathbb{P}(H) \leq \mathbb{P}\Big(\bigcup_{n\geq m} F(n)\Big) \leq \sum_{n\geq m} \mathbb{P}\big(F(n)\big) \leq \sum_{n\geq m} 2^{-n} = 2^{-m+1} \stackrel{m\to\infty}{\longrightarrow} 0.$$

At the same time

$$\int_H |X|\, d\mathbb{P} = \int \limsup_{n\to\infty} \mathbf{1}_{F(n)}|X|\, d\mathbb{P} \geq \limsup_{n\to\infty} \int \mathbf{1}_{F(n)}|X|\, d\mathbb{P} \geq \varepsilon,$$

using Fatou's lemma. This is a contradiction.

Having δ at our disposal, we choose K larger than $\mathbb{E}|X|/\delta$. By considering positive and negative part of X separately, we obtain, almost surely,

$$\Big|\mathbb{E}\{X|\mathcal{F}(n)\}\Big| \leq \mathbb{E}\Big\{|X| \,\Big|\, \mathcal{F}(n)\Big\}.$$

Hence,

$$K\mathbb{P}\{|\mathbb{E}\{X|\mathcal{F}(n)\}| > K\} = \mathbb{E}\big\{K1_{\{|\mathbb{E}\{X|\mathcal{F}(n)\}|>K\}}\big\} \le \mathbb{E}\Big|\mathbb{E}\{X|\mathcal{F}(n)\}\Big|$$
$$\le \mathbb{E}\Big\{\mathbb{E}\{|X|\,|\mathcal{F}(n)\}\Big\} = \mathbb{E}|X|\,,$$

which implies

$$\mathbb{P}\{|\mathbb{E}\{X|\mathcal{F}(n)\}| > K\} \le \frac{\mathbb{E}|X|}{K} < \delta\,.$$

Note that this event is in $\mathcal{F}(n)$. We obtain, from the definition of conditional expectation, for all n,

$$\int_{\{|\mathbb{E}\{X|\mathcal{F}(n)\}|>K\}} |\mathbb{E}\{X|\mathcal{F}(n)\}|\,d\mathbb{P} \le \int_{\{|\mathbb{E}\{X|\mathcal{F}(n)\}|>K\}} \mathbb{E}\{|X|\,|\mathcal{F}(n)\}\,d\mathbb{P}$$
$$= \int_{\{|\mathbb{E}\{X|\mathcal{F}(n)\}|>K\}} |X|\,d\mathbb{P} < \varepsilon\,.$$

This finishes our proof. ■

Proof of the Upward Theorem. We already know that $\{X_n\}$ is a martingale (see example section) and that it is uniformly integrable. Hence there is a random variable Y such that

$$\lim_{n\to\infty} X_n = Y \text{ almost surely and in } L^1.$$

We have to show that $X = Y$ almost surely. We may assume that $X \ge 0$, which also implies $X_n \ge 0$ and hence $Y \ge 0$ almost surely. Observe that

$$\mathbb{E}\{X\} = \mathbb{E}\{X_n\} \to \mathbb{E}\{Y\}\,.$$

Define probability measures P and Q on $\mathcal{A}$ (defined to be the σ-algebra generated by the union of $\mathcal{F}(n)$) by

$$P(A) = \frac{1}{\mathbb{E}\{X\}}\int_A X\,d\mathbb{P} \qquad \text{and} \qquad Q(A) = \frac{1}{\mathbb{E}\{X\}}\int_A Y\,d\mathbb{P}\,.$$

Now $\mathcal{F} := \bigcup \mathcal{F}(n)$ is a π-system which generates $\mathcal{A}$. If $A \in \mathcal{F}(n)$, then, for all $m \ge n$,

$$P(A) = \frac{1}{\mathbb{E}\{X\}}\int_A X\,d\mathbb{P} = \frac{1}{\mathbb{E}\{X\}}\int_A X_m\,d\mathbb{P}\,.$$

Also,

$$\lim_{m\to\infty}\frac{1}{\mathbb{E}\{X\}}\int_A X_m\,d\mathbb{P} = \frac{1}{\mathbb{E}\{X\}}\int_A Y\,d\mathbb{P} = Q(A)\,,$$

where the first equality follows, as before, from L^1-convergence. The uniqueness theorem now yields $P(A) = Q(A)$ for all $A \in \mathcal{A}$. Let $A = \{X > Y\} \in \mathcal{A}$. Then $\mathbb{P}(A) > 0$ would imply, by definition, $P(A) > Q(A)$, which is a contradiction. Hence, $\mathbb{P}(A) = 0$, which means $X \le Y$ almost surely. In the same manner one can show $X \ge Y$ almost surely, and this finishes the proof of the Upward Theorem.

THEOREM 2.9 (Lévy's downward theorem). *Suppose that* $\{\mathcal{G}(-n) : n \geq 0\}$ *is a collection of* σ*-algebras such that*

$$\mathcal{G}(-\infty) := \bigcap_{k=0}^{\infty} \mathcal{G}(-k) \subset \cdots \subset \mathcal{G}(-n) \subset \cdots \subset \mathcal{G}(-2) \subset \mathcal{G}(-1) .$$

Let X *be an integrable random variable and define*

$$X_{-n} = \mathbb{E}\{X|\mathcal{G}(-n)\} .$$

Then

$$\lim_{n\to\infty} X_{-n} = \mathbb{E}\{X|\mathcal{G}(-\infty)\} \text{ almost surely and in } L^1.$$

Proof: Fix a positive integer N. We look at the filtration $\{\mathcal{F}(n)\}$ given by

$$\mathcal{F}(n) = \mathcal{G}((n-N) \wedge (-1))$$

and the adapted process $Y_n = X((n-N) \wedge (-1))$. Because, by the tower property,

$$\begin{aligned} \mathbb{E}\{Y_n \,|\, \mathcal{F}(n-1)\} &= \mathbb{E}\Big\{ \mathbb{E}\{X \,|\, \mathcal{G}((n-N) \wedge (-1))\} \,\Big|\, \mathcal{G}((n-1-N) \wedge (-1)) \Big\} \\ &= \mathbb{E}\Big\{ X \,\Big|\, \mathcal{G}((n-1-N) \wedge (-1)) \Big\} = Y_{n-1} , \end{aligned}$$

this is indeed a martingale. We obtain, from the upcrossing lemma,

$$(b-a)\mathbb{E}\{U_N[a,b]\} \leq \mathbb{E}\{(X_{-1}-a)^-\} .$$

Letting $N \to \infty$ shows that, almost surely, the total number of downcrossings of $[a,b]$ by the process $\{X_{-n}\}$ is finite. Now one has this simultaneously for all rationals $a < b$ and we can argue for $\{X_{-n}\}$ as in the martingale convergence theorem. Hence, $\lim X_{-n} = X_{-\infty}$ exists almost surely. By Lemma 2.8 the sequence is even uniformly integrable and hence convergence holds in L^1. To see that $X_{-\infty} = \mathbb{E}\{X|\mathcal{G}(-\infty)\}$, first observe that, for all m and $G \in \mathcal{G}(-\infty) \subset \mathcal{G}(-m)$,

$$\int_G X \, d\mathbb{P} = \int_G X_{-m} \, d\mathbb{P} ,$$

and then let $m \to \infty$ to see that $X_{-\infty}$ satisfies the conditions of a conditional probability of X given $\mathcal{G}(-\infty)$. ■

4. The Doob Decomposition Theorem

How far, one wonders, is a general adapted process X from being a martingale? Can one define a *drift*, which can be subtracted from X to turn it into a martingale? The answer is, under some mild conditions, affirmative, as the following important result shows.

THEOREM 2.10 (Doob Decomposition Theorem). *Let* $\{\mathcal{F}(n) : n \geq 0\}$ *be a filtration and* $\{X_n : n \geq 0\}$ *an adapted process with* $\mathbb{E}|X_n| < \infty$ *for all* $n \geq 0$*. The there exist*

- *a martingale* $\{M_n : n \geq 0\}$ *with* $M_0 = 0$*, and*
- *a previsible process* $\{A_n : n \geq 1\}$ *with* $A_0 = 0$*,*

such that

$$X_n = X_0 + M_n + A_n \text{ for all } n \geq 0. \tag{4.1}$$

Moreover, any processes M and A, which satisfy this for a given X coincide almost surely. X is a submartingale if and only if the process A is increasing, i.e. $\mathbb{P}\{A_n \leq A_{n+1} \forall n\} = 1$.

Proof: It is extremely simple. We look at the uniqueness statement first. Suppose that A, M are as above. Then, almost surely,

$$\mathbb{E}\big\{ X_n - X_{n-1} \,\big|\, \mathcal{F}_{n-1} \big\} = \mathbb{E}\big\{ M_n - M_{n-1} \,\big|\, \mathcal{F}_{n-1} \big\} + \mathbb{E}\big\{ A_n - A_{n-1} \,\big|\, \mathcal{F}_{n-1} \big\} = A_n - A_{n-1},$$

using that $\{A_n\}$ is previsible and $\{M_n\}$ a martingale. This gives an explicit formula for A_n, namely

$$A_n = \sum_{k=1}^{n} \mathbb{E}\big\{ X_k - X_{k-1} \,\big|\, \mathcal{F}(k-1) \big\} \text{ almost surely.}$$

This implies uniqueness of $\{A_n\}$ and hence of $\{M_n\}$. Conversely, this formula can be used to define $\{A_n\}$ from $\{X_n\}$. Then $\{A_n\}$ is clearly previsible, and the display above shows that $M_n = X_n - A_n - X_0$ is a martingale. From this definition of A_n it is also obvious that the submartingale property of $\{X_n\}$ is equivalent to the fact that $\{A_n\}$ is increasing almost surely. ■

We now discuss a particularly interesting case of the Doob decomposition. Let $\{M_n : n \geq 0\}$ be a martingale with $M_0 = 0$ and $\mathbb{E}\{M_n^2\} < \infty$ for all $n \geq 0$. You can convince yourself easily that in general $X_n = M_n^2$ is *not* a martingale, but a submartingale. $\{M_n^2 : n \geq 0\}$ has an (essentially unique) Doob decomposition

$$M_n^2 = N_n + A_n, \text{ for } n \geq 0,$$

with $N_0 = A_0 = 0$. The increasing process A is frequently denoted $\langle M \rangle$ and is called the angle-brackets (or increasing) process associated with the martingale M. We denote by

$$\langle M \rangle_\infty = \lim_{n\to\infty} \langle M \rangle_n.$$

The next result includes convergence statements not included in the set-up of the martingale convergence theorem.

Theorem 2.11.

(a) $\lim_{n\to\infty} M_n$ *exists (and is finite) almost surely on* $\{\langle M \rangle_\infty < \infty\}$.
(b) $\lim_{n\to\infty} M_n / \langle M \rangle_n = 0$ *almost surely on* $\{\langle M \rangle_\infty = \infty\}$.

In the proof we will use the follownig little lemma.

Lemma 2.12. *If $\{M_n : n \geq 1\}$ is a martingale with $\mathbb{E}M_n^2 < \infty$, then*

$$\mathbb{E}\big\{ (M_n - M_{n-1})^2 \,\big|\, \mathcal{F}_{n-1} \big\} = \mathbb{E}\big\{ M_n^2 - M_{n-1}^2 \,\big|\, \mathcal{F}_{n-1} \big\}.$$

Proof: Using the binomial formula, taking out what is known, we get

$$\begin{aligned}\mathbb{E}\big\{ (M_n - M_{n-1})^2 \,\big|\, \mathcal{F}_{n-1} \big\} &= \mathbb{E}\big\{ M_n^2 \,\big|\, \mathcal{F}_{n-1} \big\} - 2M_{n-1}\mathbb{E}\big\{ M_n \,\big|\, \mathcal{F}_{n-1} \big\} + \mathbb{E}\big\{ M_{n-1}^2 \,\big|\, \mathcal{F}_{n-1} \big\} \\ &= \mathbb{E}\big\{ M_n^2 - M_{n-1}^2 \,\big|\, \mathcal{F}_{n-1} \big\},\end{aligned}$$

■

We now prove Theorem 2.11. We assume that M_n is in L^2, but not that M is L^2-bounded, so that convergence does not follow from the martingale convergence theorem.

Proof of (a) Because $A = \langle M \rangle$ is previsible, for every $k \geq 1$,

$$S := S(k) := \inf \{ n \geq 0 : A_{n+1} > k \}$$

is a stopping time. We now show that $\langle M^S \rangle = \langle M \rangle^S$. Note that

$$\{A_\infty < \infty\} = \bigcup_{k=1}^{\infty} \{S(k) = \infty\}. \tag{4.2}$$

For any fixed k, the stopped process A^S is previsible because, for any Borel set $B \subset \mathbb{R}$,

$$\{A_n^S \in B\} = \bigcup_{r=0}^{n-1} \{S = r, A_r \in B\} \cup \Big(\{A_n \in B\} \cap \{S \leq n-1\}^c \Big).$$

Hence $(M^S)^2 - A^S = (M^2 - A)^S = N^S$ is a martingale and this implies $\langle M^S \rangle = A^S$.

However, the process A^S is bounded by k, and

$$\mathbb{E}\{(M_n^S)^2\} = \mathbb{E}\{A_n^S\} + \mathbb{E}\{N_n^S\} = \mathbb{E}\{A_n^S\} \leq k,$$

hence M^S is L^2-bounded. We infer that $\lim_{n\to\infty} M_n^{S(k)}$ exists almost surely. By (4.2) we have on $\{\langle M \rangle_\infty < \infty\}$ that there exists a k with $S(k) = \infty$, which completes the proof of (a).

Proof of (b) Define a process $\{W_n : n \geq 0\}$ by $W_0 = 0$ and

$$W_n := \sum_{k=1}^{n} \frac{M_k - M_{k-1}}{1 + A_k} = (Y \bullet M)_n,$$

where $Y_n = 1/(1 + A_n)$ is a bounded, previsible process. By Theorem 1.4, W is a martingale. Moreover, almost surely, using Lemma 2.12,

$$\begin{aligned}
\mathbb{E}\big\{ (W_n - W_{n-1})^2 | \mathcal{F}_{n-1} \big\} &= \frac{1}{(1 + A_n)^2} \mathbb{E}\big\{ (M_n - M_{n-1})^2 | \mathcal{F}_{n-1} \big\} \\
&= \frac{1}{(1 + A_n)^2} \mathbb{E}\big\{ M_n^2 - M_{n-1}^2 | \mathcal{F}_{n-1} \big\} = \frac{1}{(1 + A_n)^2} (A_n - A_{n-1}) \\
&\leq \frac{(1 + A_n) - (1 + A_{n-1})}{(1 + A_n)(1 + A_{n-1})} = \frac{1}{1 + A_{n-1}} - \frac{1}{1 + A_n},
\end{aligned}$$

hence, using Lemma 2.12 once more,

$$\mathbb{E}\langle W \rangle_n = \mathbb{E} W_n^2 = \sum_{k=1}^{n} \mathbb{E}\{W_k^2 - W_{k-1}^2\} = \sum_{k=1}^{n} \mathbb{E}\{(W_k - W_{k-1})^2\} \leq 1.$$

We infer, using Fatou's Lemma, that $\mathbb{E}\langle W \rangle_\infty = \mathbb{E}\big\{ \liminf_{n\to\infty} \langle W \rangle_n \big\} \leq \liminf_{n\to\infty} \mathbb{E}\big\{ \langle W \rangle_n \big\} \leq 1$, and hence $\langle W \rangle_\infty < \infty$ almost surely, so that by part (a) $\lim_{n\to\infty} W_n$ exists a.s. By Kronecker's lemma we infer that, almost surely,

$$\lim_{n\to\infty} \frac{M_n}{1 + A_n} = \lim_{n\to\infty} \frac{\sum_{k=1}^{n} M_k - M_{k-1}}{1 + A_n} = 0,$$

which implies the result.

CHAPTER 3

Examples and Applications

1. Martingale proofs and improvements of classical probability theorems

The Kolmogorov 0-1-law is an extremely useful tool to see that certain events necessarily have probability zero or one.

THEOREM 3.1. *Suppose that $\{X_n\}$ is a sequence of independent and identically distributed random variables. Define*

$$\mathcal{T}_n = \sigma(X_{n+1}, X_{n+2}, \ldots), \text{ and } \mathcal{T} = \bigcap_{n=1}^{\infty} \mathcal{T}_n.$$

$\mathcal{T}$ is called the tail-σ-*algebra. Every event $A \in \mathcal{T}$ has $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.*

Examples of tail events $A \in \mathcal{T}$ are

- $\{X_n \in A_n$ for infinitely many $n\}$,
- $\{\lim_{n\to\infty} X_n$ exists $\}$, or $\{\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i = \mu\}$.

Proof: Let $\mathcal{F}_n = \sigma(X_1, X_2, \ldots, X_n)$ and $\mathcal{F}_\infty = \sigma(X_1, X_2, \ldots)$. $\mathcal{F}_n$ and $\mathcal{T}_n$ are independent. Now let $A \in \mathcal{T}$ and $X = 1_A$. Then X is bounded and $\mathcal{F}_\infty$-measurable. Hence Lévy's upward theorem gives

$$X = \lim_{n\to\infty} \mathbb{E}\{X \mid \mathcal{F}_n\} \text{ almost surely.}$$

As X is $\mathcal{T}_n$-measurable, it is independent of $\mathcal{F}_n$, hence the right hand side equals $\mathbb{E}\{X\} = \mathbb{P}(A)$. The result follows, because the left hand side only takes on the values zero or one. ■

We now give a very short proof of Kolmogorov's strong law of large numbers under minimal moment conditions. Martingale theory will serve as the major tool.

THEOREM 3.2. *Suppose that $\{X_n\}$ is a sequence of independent and identically distributed integrable random variables. Then*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n X_k = \mu \textit{ almost surely and in } L^1,$$

where μ is the common expectation of the X_n.

Proof: Abbreviate $S_n = \sum_{k=1}^n X_k$. Let $\mathcal{G}(-n)$ be the σ-algebra generated by the random variables $S_n, S_{n+1}, S_{n+2}, \ldots$ and $\mathcal{G}(-\infty)$ the intersection of all these algebras. From Problem Sheet 2 we know that

$$\mathbb{E}\{X_1 \,|\, \mathcal{G}(-n)\} = \frac{S_n}{n} \text{ almost surely.}$$

Hence, by Lévy's downward theorem,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k = \lim_{n\to\infty} \mathbb{E}\{X_1 | \mathcal{G}(-n)\} = \mathbb{E}\{X_1 | \mathcal{G}(-\infty)\} \text{ almost surely and in } L^1 .$$

Now the limit is measurable with respect to the tail σ-algebra $\mathcal{T}$ of the sequence $\{X_n\}$ and, as this has only events of probability zero or one (by Kolmogorov's 0-1-law), we have $\mathbb{E}\{X_1 | \mathcal{G}(-\infty)\} = \mu$ almost surely for some constant value μ. Integrating over Ω gives

$$\mu = \int \mathbb{E}\{X_1 \mid \mathcal{G}(-\infty)\}\, d\mathbb{P} = \int X_1 \, d\mathbb{P} = \mathbb{E}\{X_n\} \text{ for all } n,$$

and this finishes the proof. ∎

As an application of the Doob decomposition we discuss a strengthening of the Borel-Cantelli Lemma.

Theorem 3.3 (Borel-Cantelli Lemma). *Let $\{E_n : n \geq 0\}$ be a sequence of events with $E_n \in \mathcal{F}_n$. Define $X_n = \mathbb{P}\{E_n \mid \mathcal{F}_{n-1}\}$. Then, almost surely,*

(a) *If $\sum_{n=1}^{\infty} X_n < \infty$, then only finitely many of the events in $\{E_n : n \geq 0\}$ occur.*
(b) *If $\sum_{n=1}^{\infty} X_n = \infty$, then infinitely many of the events in $\{E_n : n \geq 0\}$ occur.*

Remark: For the classical Borel-Cantelli Lemma note the following.

(a) $\mathbb{E}\{X_k\} = \mathbb{P}(E_k)$ and hence $\sum_{n=1}^{\infty} \mathbb{P}(E_n) < \infty$ implies $\sum_{n=1}^{\infty} X_n < \infty$ almost surely.
(b) If the events $\{E_n : n \geq 0\}$ are independent, then let $\mathcal{F}_n = \sigma(E_1, \ldots, E_n)$ and note that $X_k = \mathbb{P}\{E_k\}$.

Proof: Let $Z_n = \sum_{k=1}^{n} 1_{E_k}$ be the number of events which occur up to time n, and $Y_n = \sum_{k=1}^{n} X_k$. Then $M_n = Z_n - Y_n$ is a martingale, and $Z_n = M_n + Y_n$ is the Doob decomposition of the submartingale $\{Z_n : n \geq 0\}$. Now we have,

$$\langle M \rangle_n = \sum_{k=1}^{n} X_k(1 - X_k) \leq Y_n, \text{ almost surely.}$$

Indeed the first equality follows by checking that the difference of M_n^2 and the right hand side is a martingale.

$$\begin{aligned}
\mathbb{E}\Big\{ M_n^2 - \sum_{k=1}^{n} X_k(1 - X_k) \,\Big|\, \mathcal{F}_{n-1} \Big\} &= M_{n-1}^2 + \mathbb{E}\big\{ (1_{E_n} - X_n)^2 \,\big|\, \mathcal{F}_{n-1} \big\} - \sum_{k=1}^{n} X_k(1 - X_k) \\
&= M_{n-1}^2 - \sum_{k=1}^{n-1} X_k(1 - X_k) \text{ almost surely.}
\end{aligned}$$

First suppose that $Y_\infty := \sum_{n=1}^{\infty} X_n < \infty$, then $\langle M \rangle_\infty < \infty$ and we infer from Lemma 2.11(a) that $\lim M_n$ exists and is finite. Hence $\sum_{k=1}^{\infty} 1_{E_k} < \infty$. Now suppose that $Y_\infty = \infty$, but $\langle M \rangle_\infty < \infty$. Then also $\lim M_n$ exists and is finite, hence $\sum_{k=1}^{\infty} 1_{E_k} = \infty$. Finally, suppose $\langle M \rangle_\infty = \infty$. Then, by Lemma 2.11(b), we have $\lim_{n\to\infty} M_n / \langle M \rangle_n = 0$, hence $\lim_{n\to\infty} M_n / Y_n = 0$ and hence $\lim_{n\to\infty} Z_n / Y_n = 1$. This implies $\sum_{k=1}^{\infty} 1_{E_k} = \lim_{n\to\infty} Z_n = \infty$. ∎

2. Martingales in optimisation problems: The Bellman optimality principle

Suppose you are playing a favourable game such that your winnings $X_1, X_2, \dots, X_N$ per unit stake are i.i.d. random variables with

$$\mathbb{P}\{X_n = 1\} = p, \mathbb{P}\{X_n = -1\} = q := 1 - p \text{ for } \tfrac{1}{2} < p < 1.$$

Starting with an initial capital Y_0 you want to choose your stakes C_n in such a way that the expected interest

$$\mathbb{E}\Big\{ \log\Big(\frac{Y_N}{Y_0}\Big)\Big\}$$

is maximized. Let $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$. Here we obviously assume

- $Y_n = Y_{n-1} + C_n X_n$ is your capital at time n,
- your stakes C_n must be previsible and $C_n \le Y_{n-1}$.

What is the best strategy and the expected interest when this stratregy is used?

To answer this question we let $R_n = C_n / Y_{n-1}$ be the proportion of your captial you put in the nth game. Then $R = (R_1, \dots, R_N)$ represents a strategy.

LEMMA 3.4. *Let R be any (previsible) strategy and let*

$$\alpha = p\log(2p) + q\log(2q)$$

be the entropy of (p, q) relative to $(1/2, 1/2)$. Then

$$M_n = \log\Big(\frac{Y_n}{Y_0}\Big) - n\alpha,\ M_0 = 0$$

defines a supermartingale.

Proof: We have

$$\mathbb{E}\big\{ M_n \,|\, \mathcal{F}_{n-1}\big\} = \mathbb{E}\Big\{ \log\Big(\frac{Y_n}{Y_0}\Big) - n\alpha \Big| \mathcal{F}_{n-1}\Big\} = \log\Big(\frac{Y_{n-1}}{Y_0}\Big) - n\alpha + \mathbb{E}\Big\{ \log\Big(\frac{Y_n}{Y_{n-1}}\Big) \Big| \mathcal{F}_{n-1}\Big\}.$$

The last expression is

$$\mathbb{E}\Big\{ \log\Big(\frac{Y_n}{Y_{n-1}}\Big) \,\Big|\, \mathcal{F}_{n-1}\Big\} = \mathbb{E}\big\{ \log\big(1 + R_n X_n\big) \,\big|\, \mathcal{F}_{n-1}\big\} = p\log\big(1 + R_n\big) + q\log\big(1 - R_n\big).$$

We wish to show that this is $\le \alpha$, which is equivalent to

$$p\log\Big(\frac{1 + R_n}{2p}\Big) + q\log\Big(\frac{1 - R_n}{2q}\Big) \le 0.$$

As log is a concave function we get

$$p\log\Big(\frac{1 + R_n}{2p}\Big) + q\log\Big(\frac{1 - R_n}{2q}\Big) \le \log\Big(p\frac{1 + R_n}{2p} + q\frac{1 - R_n}{2q}\Big) = 0,$$

as claimed. Altogether we have shown

$$\mathbb{E}\big\{ M_n \,|\, \mathcal{F}_{n-1}\big\} \le \log\Big(\frac{Y_{n-1}}{Y_0}\Big) - n\alpha + \alpha = M_{n-1}.$$

This is the supermartingale property. ∎

The last proof also indicates for which strategy we get a martingale.

LEMMA 3.5. *Let $R = R^* = (2p-1, \ldots, 2p-1)$. Then $(M_n : n \geq 0)$ is a martingale.*

Proof: We have equality in the proof of Lemma 3.4 if we choose R_n such that

$$p \log\Big(\frac{1+R_n}{2p}\Big) + q \log\Big(\frac{1-R_n}{2q}\Big) = 0.$$

This is the case if (and only if) $1 + R_n = 2p$ and $1 - R_n = 2q$, and both these equations are equivalent to $R_n = 2p - 1$. ■

To answer the initially asked question we now note from Lemma 3.4 that, for any strategy R,

$$\mathbb{E}\Big\{ \log\Big(\frac{Y_N}{Y_0}\Big)\Big\} \leq N\alpha,$$

whereas for the strategy $R^* = (2p-1, \ldots, 2p-1)$ we have

$$\mathbb{E}\Big\{ \log\Big(\frac{Y_N}{Y_0}\Big)\Big\} = N\alpha,$$

which makes this strategy optimal.

3. Martingales in optimisation problems: The secretary problem

The following optimisation problem is known as the *secretary problem*: N candidates present themselves for a job interview. The ith candidate's suitability for the job is X_i and the X_i are independent random variables, uniformly distributed on $[0,1]$. The boss interviews each in turn and can determine the value of X_i perfectly. He must immediately decide whether to accept or reject the candidate, no recall of rejected candidates is possible.

The problem is that the boss has to find a stopping time T which maximises $\mathbb{E}X_T$. We now use martingale theory to solve this problem.

THEOREM 3.6. *The stopping time $T^* = \inf\{n > 0 : X_n > \alpha_n\}$, for $\alpha_N = 0$, and*

$$\alpha_{n-1} = \frac{1}{2} + \frac{\alpha_n^2}{2} \text{ for } 1 \leq n \leq N,$$

maximises $\mathbb{E}X_T$.

Proof: We proceed in four steps. The *first* step is to show that, for any $0 \leq \alpha \leq 1$, we have

$$\mathbb{E}\{X_n \vee \alpha\} = \frac{1}{2} + \frac{\alpha^2}{2}. \tag{3.1}$$

To verify (3.1) just note that

$$\begin{aligned} \mathbb{E}\{X_n \vee \alpha\} &= \int_0^1 x \vee \alpha \, dx \\ &= \int_0^\alpha \alpha \, dx + \int_\alpha^1 x \, dx \\ &= \alpha^2 + \frac{1}{2} - \frac{\alpha^2}{2} = \frac{1}{2} + \frac{\alpha^2}{2}. \end{aligned}$$

The *second* step is to show that, for *any* stopping time T, the process Y defined by

$$Y_0 = \alpha_0, \text{ and } Y_n = (X_{T\wedge n}) \vee \alpha_n \text{ for } n \geq 1,$$

is a supermartingale. Indeed, on the event $\{T \leq n-1\}$, we have

$$\mathbb{E}\{Y_n \,|\, \mathcal{F}(n-1)\} = \mathbb{E}\{X_{T\wedge n} \vee \alpha_n \,|\, \mathcal{F}(n-1)\} = X_T \vee \alpha_n \leq X_T \vee \alpha_{n-1} = Y_{n-1},$$

using that α_n is decreasing. On the event $\{T > n-1\}$, we have

$$\mathbb{E}\{Y_n \,|\, \mathcal{F}(n-1)\} = \mathbb{E}\{X_n \vee \alpha_n\} = \frac{\alpha_n^2}{2} + \frac{1}{2} = \alpha_{n-1} \leq Y_{n-1},$$

which shows the supermartingale property.

As a *third* step we show that for $T = T^*$ the process Y is a martingale. Indeed, on the event $\{T^* \leq n-1\}$, we have, from above,

$$\mathbb{E}\{Y_n \,|\, \mathcal{F}(n-1)\} = X_{T^*} \vee \alpha_n = X_{T^*},$$

as $X_{T^*} > \alpha_{T^*} \geq \alpha_{n-1} \geq \alpha_n$. Note that $Y_{n-1} = X_{T^*} \vee \alpha_{n-1} = X_{T^*}$. On the event $\{T^* \geq n\}$ we have, as before,

$$\mathbb{E}\{Y_n \,|\, \mathcal{F}(n-1)\} = \mathbb{E}\{X_n \vee \alpha_n\} = \frac{\alpha_n^2}{2} + \frac{1}{2} = \alpha_{n-1} = Y_{n-1},$$

which completes the third step.

Finally, as a *fourth* step, we show that, for any stopping time T, we have $\mathbb{E}X_T \leq \mathbb{E}X_{T^*}$. For this we use Doob's optimal stopping theorem (note that all stopping times are bounded), to see that, for arbitrary stopping times,

$$\mathbb{E}\{X_T\} \leq \mathbb{E}\{X_T \vee \alpha_T\} = \mathbb{E}\{Y_T\} \leq \mathbb{E}\{Y_0\} = \alpha_0,$$

and for the special choice T^*,

$$\mathbb{E}\{X_{T^*}\} = \mathbb{E}\{X_{T^*} \vee \alpha_{T^*}\} = \mathbb{E}\{Y_{T^*}\} = \mathbb{E}\{Y_0\} = \alpha_0.$$

This completes the proof. ■

For more sophisticated examples, see the Mabinogion sheep problem, Chapter 15.3 in Williams, and references quoted there.

4. Noisy observations: a simple filtering problem

A random number Θ is chosen uniformly between 0 and 1, and a coin with probability Θ of heads is minted. The coin is tossed repeatedly, and the outcomes are $X_1, X_2, \ldots$ with $X_n = 1$ denoting heads in the nth toss.

Let B_n denote the number of heads in the first n tosses and $\mathcal{F}(n) = \sigma(X_1 \ldots, X_n)$. The aim of this section is to find out how, by observing $X_1, X_2, \ldots$, we accumulate information about the hidden random variable Θ, i.e. the coin. The idea here is that we cannot observe the random variable Θ directly, but we try to recover it from the coin tosses $X_1, X_2, X_3, \ldots$, which represent a *noisy observation*. Here the term noisy refers to the fact that other random effects (noise) interfer with the randomness of Θ. The problem is to

(a) find out whether the value of Θ can be recovered by observing the coin tosses,

(b) find the random variable $\mathbb{E}\{\Theta \,|\, \mathcal{F}_n\}$.

The first problem can be solved by means of the strong law of large numbers. Consider $\Theta = \theta$ as fixed. Then, almost surely,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k = \mathbb{E}_\theta X_1 = \theta.$$

Making Θ random, we get

$$\mathbb{P}\Big\{ \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k = \Theta \Big\} = \int_0^1 \mathbb{P}_\theta \Big\{ \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k = \theta \Big\} \, d\theta = 1, \tag{4.1}$$

giving us the means to get Θ from the *infinite* sequence $X_1, X_2, X_3, \ldots$.

For the second problem we have to look deeper into the structure of the problem. The solution can be given using martingale theory.

THEOREM 3.7. *$\mathbb{E}\{\Theta \,|\mathcal{F}(n)\} = \frac{B_n+1}{n+2}$ almost surely.*

Proof: To give an overview over the proof, we *first* show that, for all $k = 0, \ldots, n$, we have

(a) $\mathbb{P}\{B_n = k\} = \frac{1}{n+1}$,
(b) $\mathbb{P}\{B_n = k, B_{n-1} = k-1\} = \frac{k}{(n+1)n}$.

Using this we show as a *second* step that $M_n = \frac{B_n+1}{n+2}$ defines a martingale with respect to $\{\mathcal{F}(n) : n \geq 0\}$. From this the result follows almost immediately, as we shall see.

To verify the *first* step we use the formula

$$\int_0^1 d\theta \, \theta^k (1-\theta)^{n-k} = \frac{k!(n-k)!}{(n+1)!} \text{ for } 0 \leq k \leq n. \tag{4.2}$$

This can be checked by induction using partial integration. For (a), using (4.2),

$$\begin{aligned} \mathbb{P}\{B_n = k\} &= \int_0^1 d\theta \, \mathbb{P}_\theta\{B_n = k\} = \int_0^1 d\theta \, \theta^k (1-\theta)^{n-k} \binom{n}{k} \\ &= \frac{k!(n-k)!}{(n+1)!} \, \frac{n!}{k!(n-k)!} = \frac{1}{n+1}. \end{aligned}$$

For (b) we see similarly,

$$\begin{aligned} \mathbb{P}\{B_n = k, B_{n-1} = k-1\} &= \int_0^1 d\theta \, \mathbb{P}_\theta\{B_n = k, B_{n-1} = k-1\} \\ &= \int_0^1 d\theta \, \theta\theta^{k-1} (1-\theta)^{n-k} \binom{n-1}{k-1} = \frac{k!(n-k)!}{(n+1)!} \, \frac{(n-1)!}{(k-1)!(n-k)!} = \frac{k}{n(n+1)}. \end{aligned}$$

Recall that the event $\{B_n = k\}$ means that there are exactly k heads in the first n tosses. We infer that

$$\mathbb{P}\{B_n = k \,|\, B_{n-1} = k-1\} = \tfrac{k}{n+1} \text{ and } \mathbb{P}\{B_n = k-1 \,|\, B_{n-1} = k-1\} = \tfrac{n+1-k}{n+1}.$$

Because all paths $B_1, \dots, B_{n-1}$ with the same terminal value B_{n-1} are equally likely (as they have the same numbers of heads and tails) we infer that

$$\mathbb{P}\{B_n = B_{n-1} + 1 \,|\, \mathcal{F}(n-1)\} = \tfrac{1+B_{n-1}}{n+1} \text{ and } \mathbb{P}\{B_n = B_{n-1} \,|\, \mathcal{F}(n-1)\} = \tfrac{n-B_{n-1}}{n+1}.$$

Hence,

$$\mathbb{E}\{B_n \,|\mathcal{F}(n-1)\} = (B_{n-1}+1)\,\tfrac{1+B_{n-1}}{n+1} + B_{n-1}\,\tfrac{n-B_{n-1}}{n+1} = B_{n-1} + \tfrac{1+B_{n-1}}{n+1}.$$

Let $M_0 = 1/2$ and

$$M_n = \frac{B_n + 1}{n+2}, \qquad \text{for } n = 1, 2, \dots.$$

Then $\{M_n : n \geq 0\}$ is a martingale. Indeed,

$$\mathbb{E}\{M_n \,|\mathcal{F}_{n-1}\} = \frac{1}{n+2}\Big(1 + B_{n-1} + \frac{1+B_{n-1}}{n+1}\Big) = \frac{1}{n+2}\Big(\frac{n+2}{n+1} + B_{n-1}\frac{n+2}{n+1}\Big) = M_{n-1}.$$

This completes the proof of the *second* step. Finally, note that $0 \leq M_n \leq 1$. By Theorem 2.6 the martingale $\{M_n : n \geq 0\}$ converges almost surely and in L^1 to a finite limit random variable M and, moreover,

$$M_n = \mathbb{E}\big\{ M \,\big|\, \mathcal{F}_n \big\}.$$

But, using (4.1), almost surely,

$$M = \lim_{n\to\infty} M_n = \lim_{n\to\infty} \frac{B_n}{n} = \Theta.$$

Hence $\mathbb{E}\big\{\Theta \,\big|\, \mathcal{F}_n\big\} = \frac{B_n+1}{n+2}$. ∎

This and related problems are classical in filtering theory, more sophisticated examples of noisy observations are discussed in Williams, Chapter 15.

5. The second hearts problem

In a deck of 52 cards, well-shuffled, we turn the cards from the top until the first hearts appears. If we turn one more card, what is the probability that this card shows hearts again?

Let X_n be the proportion of hearts remaining in the deck after the nth card is turned. Let Y_n be the indicator of the event that the nth card is hearts, and let

$$\mathcal{F}(n) = \sigma(Y_0, \dots, Y_n) = \sigma(X_0, \dots, X_n).$$

LEMMA 3.8. *$\{X_n : 0 \leq n \leq 51\}$ is a martingale.*

Proof: Indeed, for $1 \leq n \leq 51$, we have

$$\begin{aligned}\mathbb{E}\{X_n \,|\, \mathcal{F}(n-1)\} &= \frac{(53-n)X_{n-1} - 1}{52-n} X_{n-1} + \frac{(53-n)X_{n-1}}{52-n}(1 - X_{n-1}) \\ &= \frac{(53-n)X_{n-1}^2 - X_{n-1} + (53-n)(X_{n-1} - X_{n-1}^2)}{52-n} = X_{n-1}.\end{aligned}$$

∎

Now let $T = \min\{n : Y_n = 1\}$ be the first time hearts appear. Note that, given X_T, the probability that the $T+1$st card is again hearts is exactly X_T. Hence the unconditional probability that the $T+1$st card is hearts is $\mathbb{E}\{X_T\}$. As T is a bounded stopping time, by Doob's optional stopping theorem, we have that

$$\mathbb{E}\{X_T\} = \mathbb{E}\{X_0\} = \frac{1}{4}.$$

The probability that the $T+1$st card is again hearts is therefore exactly $\frac{1}{4}$.