4.2 Doob's Martingale Convergence Theorem 71
4.3 Uniform Integrability and L^1 Convergence of Martingales 73
4.4 Solutions 80

5. **Markov Chains** 85
5.1 First Examples and Definitions 86
5.2 Classification of States 101
5.3 Long-Time Behaviour of Markov Chains: General Case 108
5.4 Long-Time Behaviour of Markov Chains with Finite State Space 114
5.5 Solutions 119

6. **Stochastic Processes in Continuous Time** 139
6.1 General Notions 139
6.2 Poisson Process 140
6.2.1 Exponential Distribution and Lack of Memory 140
6.2.2 Construction of the Poisson Process 142
6.2.3 Poisson Process Starts from Scratch at Time t 145
6.2.4 Various Exercises on the Poisson Process 148
6.3 Brownian Motion 150
6.3.1 Definition and Basic Properties 151
6.3.2 Increments of Brownian Motion 153
6.3.3 Sample Paths 156
6.3.4 Doob's Maximal L^2 Inequality for Brownian Motion 159
6.3.5 Various Exercises on Brownian Motion 160
6.4 Solutions 161

7. **Itô Stochastic Calculus** 179
7.1 Itô Stochastic Integral: Definition 180
7.2 Examples 189
7.3 Properties of the Stochastic Integral 190
7.4 Stochastic Differential and Itô Formula 193
7.5 Stochastic Differential Equations 202
7.6 Solutions 209

Index 223

1 Review of Probability

In this chapter we shall recall some basic notions and facts from probability theory. Here is a short list of what needs to be reviewed:

1) Probability spaces, σ-fields and measures;
2) Random variables and their distributions;
3) Expectation and variance;
4) The σ-field generated by a random variable;
5) Independence, conditional probability.

The reader is advised to consult a book on probability for more information.

1.1 Events and Probability

Definition 1.1

Let Ω be a non-empty set. A *σ-field* $\mathcal{F}$ on Ω is a family of subsets of Ω such that

1) the empty set $\emptyset$ belongs to $\mathcal{F}$;
2) if A belongs to $\mathcal{F}$, then so does the complement $\Omega \setminus A$;

3) if $A_1, A_2, \ldots$ is a sequence of sets in $\mathcal{F}$, then their union $A_1 \cup A_2 \cup \cdots$ also belongs to $\mathcal{F}$.

Example 1.1

Throughout this course $\mathbb{R}$ will denote the set of real numbers. The family of *Borel sets* $\mathcal{F} = \mathcal{B}(\mathbb{R})$ is a σ-field on $\mathbb{R}$. We recall that $\mathcal{B}(\mathbb{R})$ is the smallest σ-field containing all intervals in $\mathbb{R}$.

Definition 1.2

Let $\mathcal{F}$ be a σ-field on Ω. A *probability measure* P is a function

$$P : \mathcal{F} \to [0,1]$$

such that

1) $P(\Omega) = 1$;
2) if $A_1, A_2, \ldots$ are pairwise disjoint sets (that is, $A_i \cap A_j = \emptyset$ for $i \neq j$) belonging to $\mathcal{F}$, then

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots .$$

The triple $(\Omega, \mathcal{F}, P)$ is called a *probability space*. The sets belonging to $\mathcal{F}$ are called *events*. An event A is said to occur *almost surely* (a.s.) whenever $P(A) = 1$.

Example 1.2

We take the unit interval $\Omega = [0,1]$ with the σ-field $\mathcal{F} = \mathcal{B}([0,1])$ of Borel sets $B \subset [0,1]$, and *Lebesgue measure* $P = \text{Leb}$ on $[0,1]$. Then $(\Omega, \mathcal{F}, P)$ is a probability space. Recall that Leb is the unique measure defined on Borel sets such that

$$\text{Leb}[a,b] = b - a$$

for any interval $[a,b]$. (In fact Leb can be extended to a larger σ-field, but we shall need Borel sets only.)

Exercise 1.1

Show that if $A_1, A_2, \ldots$ is an *expanding* sequence of events, that is,

$$A_1 \subset A_2 \subset \cdots ,$$

then

$$P(A_1 \cup A_2 \cup \cdots) = \lim_{n\to\infty} P(A_n).$$

Similarly, if $A_1, A_2, \ldots$ is a *contracting* sequence of events, that is,

$$A_1 \supset A_2 \supset \cdots ,$$

then

$$P(A_1 \cap A_2 \cap \cdots) = \lim_{n\to\infty} P(A_n).$$

Hint Write $A_1 \cup A_2 \cup \cdots$ as the union of a sequence of disjoint events: start with A_1, then add a disjoint set to obtain $A_1 \cup A_2$, then add a disjoint set again to obtain $A_1 \cup A_2 \cup A_3$, and so on. Now that you have a sequence of disjoint sets, you can use the definition of a probability measure. To deal with the product $A_1 \cap A_2 \cap \cdots$ write it as a union of some events with the aid of De Morgan's law.

Lemma 1.1 (Borel–Cantelli)

Let $A_1, A_2, \ldots$ be a sequence of events such that $P(A_1) + P(A_2) + \cdots < \infty$ and let $B_n = A_n \cup A_{n+1} \cup \cdots$. Then $P(B_1 \cap B_2 \cap \cdots) = 0$.

Exercise 1.2

Prove the Borel–Cantelli lemma above.

Hint $B_1, B_2, \ldots$ is a contracting sequence of events.

1.2 Random Variables

Definition 1.3

If $\mathcal{F}$ is a σ-field on Ω, then a function $\xi : \Omega \to \mathbb{R}$ is said to be *$\mathcal{F}$-measurable* if

$$\{\xi \in B\} \in \mathcal{F}$$

for every Borel set $B \in \mathcal{B}(\mathbb{R})$. If $(\Omega, \mathcal{F}, P)$ is a probability space, then such a function ξ is called a *random variable*.

Remark 1.1

A short-hand notation for events such as $\{\xi \in B\}$ will be used to avoid clutter. To be precise, we should write

$$\{\omega \in \Omega : \xi(\omega) \in B\}$$

in place of $\{\xi \in B\}$. Incidentally, $\{\xi \in B\}$ is just a convenient way of writing the inverse image $\xi^{-1}(B)$ of a set.

Definition 1.4

The σ-field $\sigma(\xi)$ *generated* by a random variable $\xi : \Omega \to \mathbb{R}$ consists of all sets of the form $\{\xi \in B\}$, where B is a Borel set in $\mathbb{R}$.

Definition 1.5

The σ-field $\sigma\{\xi_i : i \in I\}$ generated by a family $\{\xi_i : i \in I\}$ of random variables is defined to be the smallest σ-field containing all events of the form $\{\xi_i \in B\}$, where B is a Borel set in $\mathbb{R}$ and $i \in I$.

Exercise 1.3

We call $f : \mathbb{R} \to \mathbb{R}$ a *Borel function* if the inverse image $f^{-1}(B)$ of any Borel set B in $\mathbb{R}$ is a Borel set. Show that if f is a Borel function and ξ is a random variable, then the composition $f(\xi)$ is $\sigma(\xi)$-measurable.

Hint Consider the event $\{f(\xi) \in B\}$, where B is an arbitrary Borel set. Can this event be written as $\{\xi \in A\}$ for some Borel set A?

Lemma 1.2 (Doob–Dynkin)

Let ξ be a random variable. Then each $\sigma(\xi)$-measurable random variable η can be written as

$$\eta = f(\xi)$$

for some Borel function $f : \mathbb{R} \to \mathbb{R}$.

The proof of this highly non-trivial result will be omitted.

Definition 1.6

Every random variable $\xi : \Omega \to \mathbb{R}$ gives rise to a probability measure

$$P_\xi(B) = P\{\xi \in B\}$$

on $\mathbb{R}$ defined on the σ-field of Borel sets $B \in \mathcal{B}(\mathbb{R})$. We call P_ξ the *distribution* of ξ. The function $F_\xi : \mathbb{R} \to [0,1]$ defined by

$$F_\xi(x) = P\{\xi \le x\}$$

is called the *distribution function* of ξ.

Exercise 1.4

Show that the distribution function F_ξ is non-decreasing, right-continuous,

$$\lim_{x\to-\infty} F_\xi(x) = 0, \quad \lim_{x\to+\infty} F_\xi(x) = 1.$$

Hint For example, to verify right-continuity show that $F_\xi(x_n) \to F_\xi(x)$ for any creasing sequence x_n such that $x_n \to x$. You may find the results of Exercise useful.

Definition 1.7

If there is a Borel function $f_\xi : \mathbb{R} \to \mathbb{R}$ such that for any Borel set $B \subset \mathbb{R}$

$$P\{\xi \in B\} = \int_B f_\xi(x)\, dx,$$

then ξ is said to be a random variable with *absolutely continuous distribut* and f_ξ is called the *density* of ξ. If there is a (finite or infinite) sequence pairwise distinct real numbers $x_1, x_2, \ldots$ such that for any Borel set $B \subset \mathbb{R}$

$$P\{\xi \in B\} = \sum_{x_i \in B} P\{\xi = x_i\},$$

then ξ is said to have *discrete distribution* with values $x_1, x_2, \ldots$ and m $P\{\xi = x_i\}$ at x_i.

Exercise 1.5

Suppose that ξ has continuous distribution with density f_ξ. Show that

$$\frac{d}{dx} F_\xi(x) = f_\xi(x)$$

if f_ξ is continuous at x.

Hint Express $F_\xi(x)$ as an integral of f_ξ.

Exercise 1.6

Show that if ξ has discrete distribution with values $x_1, x_2, \ldots$, then F_ξ constant on each interval $(s,t]$ not containing any of the x_i's and has jumps size $P\{\xi = x_i\}$ at each x_i.

Hint The increment $F_\xi(t) - F_\xi(s)$ is equal to the total mass of the x_i's that belo to the interval $(s,t]$.

Definition 1.8

The *joint distribution* of several random variables $\xi_1, \ldots, \xi_n$ is a probability measure $P_{\xi_1,\ldots,\xi_n}$ on $\mathbb{R}^n$ such that

$$P_{\xi_1,\ldots,\xi_n}(B) = P\{(\xi_1,\ldots,\xi_n) \in B\}$$

for any Borel set B in $\mathbb{R}^n$. If there is a Borel function $f_{\xi_1,\ldots,\xi_n} : \mathbb{R}^n \to \mathbb{R}$ such that

$$P\{(\xi_1,\ldots,\xi_n) \in B\} = \int_B f_{\xi_1,\ldots,\xi_n}(x_1,\ldots,x_n)\, dx_1 \cdots dx_n$$

for any Borel set B in $\mathbb{R}^n$, then $f_{\xi_1,\ldots,\xi_n}$ is called the *joint density* of $\xi_1, \ldots, \xi_n$.

Definition 1.9

A random variable $\xi : \Omega \to \mathbb{R}$ is said to be *integrable* if

$$\int_\Omega |\xi|\, dP < \infty.$$

Then

$$E(\xi) = \int_\Omega \xi\, dP$$

exists and is called the *expectation* of ξ. The family of integrable random variables $\xi : \Omega \to \mathbb{R}$ will be denoted by L^1 or, in case of possible ambiguity, by $L^1(\Omega, \mathcal{F}, P)$.

Example 1.3

The *indicator function* 1_A of a set A is equal to 1 on A and 0 on the complement $\Omega \setminus A$ of A. For any event A

$$E(1_A) = \int_\Omega 1_A\, dP = P(A).$$

We say that $\eta : \Omega \to \mathbb{R}$ is a *step function* if

$$\eta = \sum_{i=1}^n \eta_i 1_{A_i},$$

where $\eta_1, \ldots, \eta_n$ are real numbers and $A_1, \ldots, A_n$ are pairwise disjoint events. Then

$$E(\eta) = \int_\Omega \eta\, dP = \sum_{i=1}^n \eta_i \int_\Omega 1_{A_i}\, dP = \sum_{i=1}^n \eta_i P(A_i).$$

Exercise 1.7

Show that for any Borel function $h : \mathbb{R} \to \mathbb{R}$ such that $h(\xi)$ is integrable

$$E(h(\xi)) = \int_\mathbb{R} h(x)\, dP_\xi(x).$$

Hint First verify the equality for step functions $h : \mathbb{R} \to \mathbb{R}$, then for non-negative ones by approximating them by step functions, and finally for arbitrary Borel functions by splitting them into positive and negative parts.

In particular, Exercise 1.7 implies that if ξ has an absolutely continuous distribution with density f_ξ, then

$$E(h(\xi)) = \int_{-\infty}^{+\infty} h(x) f_\xi(x)\, dx.$$

If ξ has a discrete distribution with (finitely or infinitely many) pairwise distinct values $x_1, x_2, \ldots$, then

$$E(h(\xi)) = \sum_i h(x_i) P\{\xi = x_i\}.$$

Definition 1.10

A random variable $\xi : \Omega \to \mathbb{R}$ is called *square integrable* if

$$\int_\Omega |\xi|^2\, dP < \infty.$$

Then the *variance* of ξ can be defined by

$$\mathrm{var}(\xi) = \int_\Omega (\xi - E(\xi))^2\, dP.$$

The family of square integrable random variables $\xi : \Omega \to \mathbb{R}$ will be denoted by $L^2(\Omega, \mathcal{F}, P)$ or, if no ambiguity is possible, simply by L^2.

Remark 1.2

The result in Exercise 1.8 below shows that we may write $E(\xi)$ in the definition of variance.

Exercise 1.8

Show that if ξ is a square integrable random variable, then it is integrable.

Hint Use the Schwarz inequality

$$[E(\xi\eta)]^2 \leq E(\xi^2)\,E(\eta^2) \tag{1.1}$$

with an appropriately chosen η.

Exercise 1.9

Show that if $\eta : \Omega \to [0,\infty)$ is a non-negative square integrable random variable, then

$$E(\eta^2) = 2\int_0^\infty tP(\eta > t)\,dt.$$

Hint Express $E(\eta^2)$ in terms of the distribution function $F_\eta(t)$ of η and then integrate by parts.

1.3 Conditional Probability and Independence

Definition 1.11

For any events $A, B \in \mathcal{F}$ such that $P(B) \neq 0$ the *conditional probability* of A given B is defined by

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

Exercise 1.10

Prove the *total probability formula*

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots$$

for any event $A \in \mathcal{F}$ and any sequence of pairwise disjoint events $B_1, B_2, \ldots \in \mathcal{F}$ such that $B_1 \cup B_2 \cup \cdots = \Omega$ and $P(B_n) \neq 0$ for any n.

Hint $A = (A\cap B_1) \cup (A \cap B_2) \cup \cdots$.

Definition 1.12

Two events $A, B \in \mathcal{F}$ are called *independent* if

$$P(A\cap B) = P(A)P(B).$$

In general, we say that n events $A_1, \ldots, A_n \in \mathcal{F}$ are *independent* if

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2})\cdots P(A_{i_k})$$

for any indices $1 \leq i_1 < i_2 < \cdots < i_k \leq n$.

Exercise 1.11

Let $P(B) \neq 0$. Show that A and B are independent events if and only if $P(A|B) = P(A)$.

Hint If $P(B) \neq 0$, then you can divide by it.

Definition 1.13

Two random variables ξ and η are called *independent* if for any Borel sets $A, B \in \mathcal{B}(\mathbb{R})$ the two events

$$\{\xi \in A\} \quad \text{and} \quad \{\eta \in B\}$$

are independent. We say that n random variables $\xi_1, \ldots, \xi_n$ are *independent* if for any Borel sets $B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R})$ the events

$$\{\xi_1 \in B_1\}, \ldots, \{\xi_n \in B_n\}$$

are independent. In general, a (finite or infinite) family of random variables is said to be *independent* if any finite number of random variables from this family are independent.

Proposition 1.1

If two integrable random variables $\xi, \eta : \Omega \to \mathbb{R}$ are independent, then they are *uncorrelated*, i.e.

$$E(\xi\eta) = E(\xi)E(\eta),$$

provided that the product $\xi\eta$ is also integrable. If $\xi_1, \ldots, \xi_n : \Omega \to \mathbb{R}$ are independent integrable random variables, then

$$E(\xi_1\xi_2\cdots\xi_n) = E(\xi_1)\,E(\xi_2)\cdots E(\xi_n),$$

provided that the product $\xi_1\xi_2\cdots\xi_n$ is also integrable.

Definition 1.14

Two σ-fields $\mathcal{G}$ and $\mathcal{H}$ contained in $\mathcal{F}$ are called *independent* if any two events

$$A \in \mathcal{G} \quad \text{and} \quad B \in \mathcal{H}$$

are independent. Similarly, any finite number of σ-fields $\mathcal{G}_1, \ldots, \mathcal{G}_n$ contained in $\mathcal{F}$ are *independent* if any n events

$$A_1 \in \mathcal{G}_1, \ldots, A_n \in \mathcal{G}_n$$

are independent. In general, a (finite or infinite) family of σ-fields is said to be *independent* if any finite number of them are independent.

Exercise 1.12

Show that two random variables ξ and η are independent if and only if the σ-fields $\sigma(\xi)$ and $\sigma(\eta)$ generated by them are independent.

Hint The events in $\sigma(\xi)$ and $\sigma(\eta)$ are of the form $\{\xi \in A\}$, and $\{\eta \in B\}$, where A and B are Borel sets.

Sometimes it is convenient to talk of independence for a combination of random variables and σ-fields.

Definition 1.15

We say that a random variable ξ is *independent* of a σ-field $\mathcal{G}$ if the σ-fields

$$\sigma(\xi) \quad \text{and} \quad \mathcal{G}$$

are independent. This can be extended to any (finite or infinite) family consisting of random variables or σ-fields or a combination of them both. Namely, such a family is called *independent* if for any finite number of random variables $\xi_1, \ldots, \xi_m$ and σ-fields $\mathcal{G}_1, \ldots, \mathcal{G}_n$ from this family the σ-fields

$$\sigma(\xi_1), \ldots, \sigma(\xi_m), \mathcal{G}_1, \ldots, \mathcal{G}_n$$

are independent.

1.4 Solutions

Solution 1.1

If $A_1 \subset A_2 \subset \cdots$, then

$$A_1 \cup A_2 \cup \cdots = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \cdots,$$

where the sets $A_1, A_2 \setminus A_1, A_3 \setminus A_2, \ldots$ are pairwise disjoint. Therefore, by the definition of probability measure

$$\begin{aligned} P(A_1 \cup A_2 \cup \cdots) &= P(A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \cdots) \\ &= P(A_1) + P(A_2 \setminus A_1) + P(A_3 \setminus A_2) + \cdots \\ &= \lim_{n\to\infty} P(A_n). \end{aligned}$$

The last equality holds because the partial sums in the series above are

$$\begin{aligned} P(A_1) + P(A_2 \setminus A_1) + \cdots + P(A_n \setminus A_{n-1}) &= P(A_1 \cup \cdots \cup A_n) \\ &= P(A_n). \end{aligned}$$

If $A_1 \supset A_2 \supset \cdots$, then the equality

$$P(A_1 \cap A_2 \cap \cdots) = \lim_{n\to\infty} P(A_n)$$

follows by taking the complements of A_n and applying De Morgan's law

$$\Omega \setminus (A_1 \cap A_2 \cap \cdots) = (\Omega \setminus A_1) \cup (\Omega \setminus A_2) \cup \cdots.$$

Solution 1.2

Since B_n is a contracting sequence of events, the results of Exercise 1.1 imply that

$$\begin{aligned} P(B_1 \cap B_2 \cap \cdots) &= \lim_{n\to\infty} P(B_n) \\ &= \lim_{n\to\infty} P(A_n \cup A_{n+1} \cup \cdots) \\ &\leq \lim_{n\to\infty} (P(A_n) + P(A_{n+1}) + \cdots) \\ &= 0. \end{aligned}$$

The last equality holds because the series $\sum_{n=1}^{\infty} P(A_n)$ is convergent. The inequality above holds by the subadditivity property

$$P(A_n \cup A_{n+1} \cup \cdots) \leq P(A_n) + P(A_{n+1}) + \cdots.$$

It follows that $P(B_1 \cap B_2 \cap \cdots) = 0$.

Solution 1.3

If B is a Borel set in $\mathbb{R}$ and $f : \mathbb{R} \to \mathbb{R}$ is a Borel function, then $f^{-1}(B)$ is also a Borel set. Therefore

$$\{f(\xi) \in B\} = \{\xi \in f^{-1}(B)\}$$

belongs to the σ-field $\sigma(\xi)$ generated by ξ. It follows that the composition $f(\xi)$ is $\sigma(\xi)$-measurable.

Solution 1.4

If $x \le y$, then $\{\xi \le x\} \subset \{\xi \le y\}$, so

$$F_\xi(x) = P\{\xi \le x\} \le P\{\xi \le y\} = F_\xi(y).$$

This means that F_ξ is non-decreasing.

Next, we take any sequence $x_1 \ge x_2 \ge \cdots$ and put

$$\lim_{n\to\infty} x_n = x.$$

Then the events

$$\{\xi \le x_1\} \supset \{\xi \le x_2\} \supset \cdots$$

form a contracting sequence with intersection

$$\{\xi \le x\} = \{\xi \le x_1\} \cap \{\xi \le x_2\} \cap \cdots.$$

It follows by Exercise 1.1 that

$$F_\xi(x) = P\{\xi \le x\} = \lim_{n\to\infty} P\{\xi \le x_n\} = \lim_{n\to\infty} F_\xi(x_n).$$

This proves that F_ξ is right-continuous.

Since the events

$$\{\xi \le -1\} \supset \{\xi \le -2\} \supset \cdots$$

form a contracting sequence with intersection $\emptyset$ and

$$\{\xi \le 1\} \subset \{\xi \le 2\} \subset \cdots$$

form an expanding sequence with union Ω, it follows by Exercise 1.1 that

$$\lim_{x\to-\infty} F_\xi(x) = \lim_{n\to\infty} F_\xi(-n) = \lim_{n\to\infty} P\{\xi \le -n\} = P(\emptyset) = 0,$$

$$\lim_{x\to\infty} F_\xi(x) = \lim_{n\to\infty} F_\xi(n) = \lim_{n\to\infty} P\{\xi \le n\} = P(\Omega) = 1,$$

since F_ξ is non-decreasing.

Solution 1.5

If ξ has a density f_ξ, then the distribution function F_ξ can be written as

$$F_\xi(x) = P\{\xi \le x\} = \int_{-\infty}^{x} f_\xi(y)\,dy.$$

Therefore, if f_ξ is continuous at x, then F_ξ is differentiable at x and

$$\frac{d}{dx} F_\xi(x) = f_\xi(x).$$

Solution 1.6

If $s < t$ are real numbers such that $x_i \notin (s,t]$ for any i, then

$$F_\xi(t) - F_\xi(s) = P\{\xi \le t\} - P\{\xi \le s\} = P\{\xi \in (s,t]\} = 0,$$

i.e. $F_\xi(s) = F_\xi(t)$. Because F_ξ is non-decreasing, this means that F_ξ is constant on $(s,t]$. To show that F_ξ has a jump of size $P\{\xi = x_i\}$ at each x_i, we compute

$$\begin{aligned}\lim_{t\searrow x_i} F_\xi(t) - \lim_{s\nearrow x_i} F_\xi(s) &= \lim_{t\searrow x_i} P\{\xi \le t\} - \lim_{s\nearrow x_i} P\{\xi \le s\} \\ &= P\{\xi \le x_i\} - P\{\xi < x_i\} = P\{\xi = x_i\}.\end{aligned}$$

Solution 1.7

If h is a step function,

$$h = \sum_{i=1}^{n} h_i 1_{A_i},$$

where $h_1, \ldots, h_n$ are real numbers and $A_1, \ldots, A_n$ are pairwise disjoint Borel sets covering $\mathbb{R}$, then

$$\begin{aligned}E(h(\xi)) &= \sum_{i=1}^{n} h_i E(1_{A_i}(\xi)) = \sum_{i=1}^{n} h_i P\{\xi \in A_i\} \\ &= \sum_{i=1}^{n} h_i P_\xi(A_i) = \sum_{i=1}^{n} \int_{A_i} h(x)\,dP_\xi(x) = \int_{\mathbb{R}} h(x)\,dP_\xi(x).\end{aligned}$$

Next, any non-negative Borel function h can be approximated by a non-decreasing sequence of step functions. For such an h the result follows by the monotone convergence of integrals. Finally, this implies the desired equality for all Borel functions h, since each can be split into its positive and negative parts, $h = h^+ - h^-$, where $h^+, h^- \ge 0$.

Solution 1.8

By the Schwarz inequality (1.1) with $\eta = 1$, if ξ is square integrable, then

$$[E(|\xi|)]^2 = [E(1\,|\xi|)]^2 \le E(1^2)\,E(\xi^2) = E(\xi^2) < \infty,$$

i.e. ξ is integrable.

Solution 1.9

Let $F(t) = P\{\eta \le t\}$ be the distribution function of η. Then

$$E(\eta^2) = \int_0^\infty t^2 dF(t).$$

Since $P(\eta > t) = 1 - F(t)$, we need to show that

$$\int_0^\infty t^2\, dF(t) = 2\int_0^\infty t\,(1 - F(t))\, dt \tag{1.2}$$

First, let us establish a version of (1.2) with ∞ replaced by a finite number a. Integrating by parts, we obtain

$$\begin{aligned}\int_0^a t^2 dF(t) &= \int_0^a t^2 d(F(t) - 1)\\ &= t^2(F(t) - 1)\Big|_0^a - 2\int_0^a t(F(t) - 1)\, dt\\ &= -a^2(1 - F(a)) + 2\int_0^a t(1 - F(t))dt. \end{aligned}\tag{1.3}$$

We see that (1.2) follows from (1.3), provided that

$$a^2\,(1 - F(a)) \to 0, \quad \text{as } a \to \infty. \tag{1.4}$$

But

$$0 \le a^2\,(1 - F(a)) = a^2 P(\eta > a) \le (n+1)^2 P(\eta > n) \le 4n^2 P(\eta \ge n),$$

where n is the integer part of a, and

$$E(\eta^2) = \sum_{k=0}^{\infty} \int_{\{k \le \eta \le k+1\}} \eta^2 dP < \infty.$$

Hence,

$$n^2 P\,(\eta \ge n) \le \int_{\{\eta \ge n\}} \eta^2\, dP = \sum_{k=n}^{\infty} \int_{\{k \le \eta < k+1\}} \eta^2\, dP \to 0 \tag{1.5}$$

as $n \to \infty$, which proves (1.4).

Solution 1.10

Since $B_1 \cup B_2 \cup \cdots = \Omega$,

$$A = A \cap (B_1 \cup B_2 \cup \cdots) = (A \cap B_1) \cup (A \cap B_2) \cup \cdots ,$$

where

$$(A \cap B_i) \cap (A \cap B_j) = A \cap (B_i \cap B_j) = A \cap \emptyset = \emptyset.$$

By countable additivity

$$\begin{aligned} P(A) &= P\,((A \cap B_1) \cup (A \cap B_2) \cup \cdots)\\ &= P\,(A \cap B_1) + P\,(A \cap B_2) + \cdots\\ &= P\,(A|B_1)\,P(B_1) + P\,(A|B_2)\,P(B_2) + \cdots . \end{aligned}$$

Solution 1.11

If $P\,(B) \ne 0$, then A and B are independent if and only if

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

In turn, this equality holds if and only if $P\,(A) = P\,(A|B)$.

Solution 1.12

The σ-fields $\sigma\,(\xi)$ and $\sigma\,(\eta)$ consist, respectively, of events of the form

$$\{\xi \in A\} \quad \text{and} \quad \{\eta \in B\},$$

where A and B are Borel sets in $\mathbb{R}$. Therefore, $\sigma\,(\xi)$ and $\sigma\,(\eta)$ are independent if and only if the events $\{\xi \in A\}$, and $\{\eta \in B\}$ are independent for any Borel sets A and B, which in turn is equivalent to ξ and η being independent.

2
Conditional Expectation

Conditional expectation is a crucial tool in the study of stochastic processes. It is therefore important to develop the necessary intuition behind this notion, the definition of which may appear somewhat abstract at first. This chapter is designed to help the beginner by leading him or her step by step through several special cases, which become increasingly involved, culminating at the general definition of conditional expectation. Many varied examples and exercises are provided to aid the reader's understanding.

2.1 Conditioning on an Event

The first and simplest case to consider is that of the conditional expectation $E(\xi|B)$ of a random variable ξ given an event B.

Definition 2.1

For any integrable random variable ξ and any event $B \in \mathcal{F}$ such that $P(B) \neq 0$ the *conditional expectation* of ξ given B is defined by

$$E(\xi|B) = \frac{1}{P(B)} \int_B \xi \, dP.$$

Example 2.1

Three coins, 10p, 20p and 50p are tossed. The values of those coins that land heads up are added to work out the total amount ξ. What is the expected total amount ξ given that two coins have landed heads up?

Let B denote the event that two coins have landed heads up. We want to find $E(\xi|B)$. Clearly, B consists of three elements,

$$B = \{\mathrm{HHT}, \mathrm{HTH}, \mathrm{THH}\},$$

each having the same probability $\frac{1}{8}$. (Here H stands for heads and T for tails.) The corresponding values of ξ are

$$\begin{aligned}\xi(\mathrm{HHT}) &= 10 + 20 = 30,\\ \xi(\mathrm{HTH}) &= 10 + 50 = 60,\\ \xi(\mathrm{THH}) &= 20 + 50 = 70.\end{aligned}$$

Therefore

$$E(\xi|B) = \frac{1}{P(B)}\int_B \xi\, dP = \frac{1}{\frac{3}{8}}\left(\frac{30}{8} + \frac{60}{8} + \frac{70}{8}\right) = 53\tfrac{1}{3}.$$

Exercise 2.1

Show that $E(\xi|\Omega) = E(\xi)$.

Hint The definition of $E(\xi)$ involves an integral and so does the definition of $E(\xi|\Omega)$. How are these integrals related?

Exercise 2.2

Show that if

$$1_A(\omega) = \begin{cases} 1 & \text{for } \omega \in A \\ 0 & \text{for } \omega \notin A \end{cases}$$

(the *indicator function* of A), then

$$E(1_A|B) = P(A|B),$$

where

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

is the *conditional probability* of A given B.

Hint Write $\int_B 1_A\, dP$ as $P(A\cap B)$.

2.2 Conditioning on a Discrete Random Variable

The next step towards the general definition of conditional expectation involves conditioning by a discrete random variable η with possible values $y_1, y_2, \ldots$ such that $P\{\eta = y_n\} \neq 0$ for each n. Finding out the value of η amounts to finding out which of the events $\{\eta = y_n\}$ has occurred or not. Conditioning by η should therefore be the same as conditioning by the events $\{\eta = y_n\}$. Because we do not know in advance which of these events will occur, we need to consider all possibilities, involving a sequence of conditional expectations

$$E(\xi|\{\eta = y_1\}), E(\xi|\{\eta = y_2\}), \ldots.$$

A convenient way of doing this is to construct a new discrete random variable constant and equal to $E(\xi|\{\eta = y_n\})$ on each of the sets $\{\eta = y_n\}$. This leads us to the next definition.

Definition 2.2

Let ξ be an integrable random variable and let η be a discrete random variable as above. Then the *conditional expectation* of ξ given η is defined to be a random variable $E(\xi|\eta)$ such that

$$E(\xi|\eta)(\omega) = E(\xi|\{\eta = y_n\}) \quad \text{if } \eta(\omega) = y_n$$

for any $n = 1, 2, \ldots$.

Example 2.2

Three coins, 10p, 20p and 50p are tossed as in Example 2.1. What is the conditional expectation $E(\xi|\eta)$ of the total amount ξ shown by the three coins given the total amount η shown by the 10p and 20p coins only?

Clearly, η is a discrete random variable with four possible values: 0, 10, 20 and 30. We find the four corresponding conditional expectations in a similar way as in Example 2.1:

$$E(\xi|\{\eta = 0\}) = 25, \quad E(\xi|\{\eta = 10\}) = 35,$$
$$E(\xi|\{\eta = 20\}) = 45, \quad E(\xi|\{\eta = 30\}) = 55.$$

Therefore

$$E(\xi|\eta)(\omega) = \begin{cases} 25 & \text{if } \eta(\omega) = 0, \\ 35 & \text{if } \eta(\omega) = 10, \\ 45 & \text{if } \eta(\omega) = 20, \\ 55 & \text{if } \eta(\omega) = 30. \end{cases}$$

Example 2.3

Take $\Omega = [0,1]$ with the σ-field of Borel sets and P the Lebesgue measure on $[0,1]$. We shall find $E(\xi|\eta)$ for

$$\xi(x) = 2x^2, \qquad \eta(x) = \begin{cases} 1 & \text{if } x \in [0,\frac{1}{3}], \\ 2 & \text{if } x \in (\frac{1}{3},\frac{2}{3}), \\ 0 & \text{if } x \in (\frac{2}{3},1]. \end{cases}$$

Clearly, η is discrete with three possible values $1,2,0$. The corresponding events are

$$\begin{aligned} \{\eta = 1\} &= [0,\tfrac{1}{3}], \\ \{\eta = 2\} &= (\tfrac{1}{3},\tfrac{2}{3}), \\ \{\eta = 0\} &= (\tfrac{2}{3},1]. \end{aligned}$$

For $x \in [0,\frac{1}{3}]$

$$E(\xi|\eta)(x) = E(\xi|[0,\tfrac{1}{3}]) = \frac{1}{\frac{1}{3}} \int_0^{\frac{1}{3}} 2x^2 dx = \frac{2}{27}.$$

For $x \in (\frac{1}{3},\frac{2}{3})$

$$E(\xi|\eta)(x) = E(\xi|(\tfrac{1}{3},\tfrac{2}{3})) = \frac{1}{\frac{1}{3}} \int_{\frac{1}{3}}^{\frac{2}{3}} 2x^2 dx = \frac{14}{27}.$$

And for $x \in (\frac{2}{3},1]$

$$E(\xi|\eta)(x) = E(\xi|(\tfrac{2}{3},1]) = \frac{1}{\frac{1}{3}} \int_{\frac{2}{3}}^{1} 2x^2 dx = \frac{38}{27}.$$

The graph of $E(\xi|\eta)$ is shown in Figure 2.1 together with those of ξ and η.

Exercise 2.3

Show that if η is a constant function, then $E(\xi|\eta)$ is constant and equal to $E(\xi)$.

Hint The event $\{\eta = c\}$ must be $\emptyset$ or Ω for any $c \in \mathbb{R}$.

Exercise 2.4

Show that

$$E(1_A|1_B)(\omega) = \begin{cases} P(A|B) & \text{if } \omega \in B \\ P(A|\Omega \setminus B) & \text{if } \omega \notin B \end{cases}$$

for any B such that $1 \neq P(B) \neq 0$.

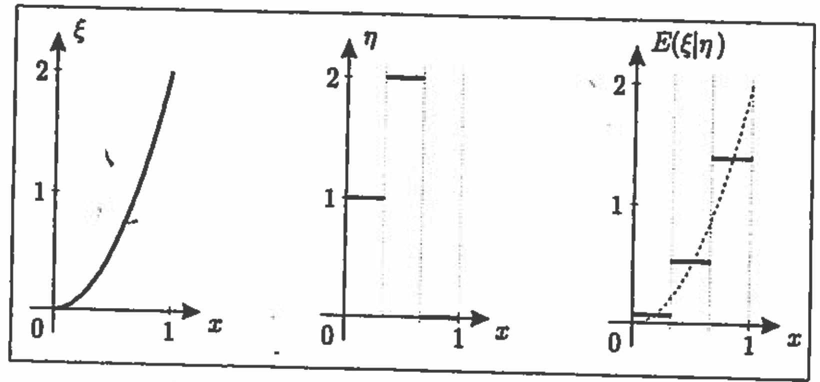

Figure 2.1. The graph of $E(\xi|\eta)$ in Example 2.3

Hint How many different values does 1_B take? What are the sets on which these values are taken?

Exercise 2.5

Assuming that η is a discrete random variable, show that

$$E(E(\xi|\eta)) = E(\xi).$$

Hint Observe that

$$\int_B E(\xi|\eta)\, dP = \int_B \xi\, dP$$

for any event B on which η is constant. The desired equality can be obtained by covering Ω by countably many disjoint events of this kind.

Proposition 2.1

If ξ is an integrable random variable and η is a discrete random variable, then

1) $E(\xi|\eta)$ is $\sigma(\eta)$-measurable;

2) For any $A \in \sigma(\eta)$

$$\int_A E(\xi|\eta)\, dP = \int_A \xi\, dP. \tag{2.1}$$

Proof

Suppose that η has pairwise distinct values $y_1, y_2, \ldots$. Then the events

$$\{\eta = y_1\}, \{\eta = y_2\}, \ldots$$

are pairwise disjoint and cover Ω. The σ-field $\sigma(\eta)$ is generated by these events, in fact every $A \in \sigma(\eta)$ is a countable union of sets of the form $\{\eta = y_n\}$. Because $E(\xi|\eta)$ is constant on each of these sets, it must be $\sigma(\eta)$-measurable.

For each n we have

$$\int_{\{\eta=y_n\}} E(\xi|\eta)\,dP = \int_{\{\eta=y_n\}} E(\xi|\{\eta=y_n\})\,dP = \int_{\{\eta=y_n\}} \xi\,dP.$$

Since each $A \in \sigma(\eta)$ is a countable union of sets of the form $\{\eta = y_n\}$, which are pairwise disjoint, it follows that

$$\int_A E(\xi|\eta)\,dP = \int_A \xi\,dP,$$

as required. □

2.3 Conditioning on an Arbitrary Random Variable

Properties 1) and 2) in Proposition 2.1 provide the key to the definition of the conditional expectation given an arbitrary random variable η.

Definition 2.3

Let ξ be an integrable random variable and let η be an arbitrary random variable. Then the *conditional expectation* of ξ given η is defined to be a random variable $E(\xi|\eta)$ such that

1) $E(\xi|\eta)$ is $\sigma(\eta)$-measurable;
2) For any $A \in \sigma(\eta)$

$$\int_A E(\xi|\eta)\,dP = \int_A \xi\,dP.$$

Remark 2.1

We can also define the *conditional probability* of an event $A \in \mathcal{F}$ given η by

$$P(A|\eta) = E(1_A|\eta),$$

where 1_A is the indicator function of A.

Do the conditions of Definition 2.3 characterize $E(\xi|\eta)$ uniquely? The lemma below implies that $E(\xi|\eta)$ is defined to within equality on a set of full measure. Namely,

$$\text{if } \xi = \xi' \text{ a.s., then } E(\xi|\eta) = E(\xi'|\eta) \text{ a.s.} \tag{2.2}$$

The existence of $E(\xi|\eta)$ will be discussed later in this chapter.

Lemma 2.1

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathcal{G}$ be a σ-field contained in $\mathcal{F}$. If ξ is a $\mathcal{G}$-measurable random variable and for any $B \in \mathcal{G}$

$$\int_B \xi\,dP = 0,$$

then $\xi = 0$ a.s.

Proof

Observe that $P\{\xi \geq \varepsilon\} = 0$ for any $\varepsilon > 0$ because

$$0 \leq \varepsilon P\{\xi \geq \varepsilon\} = \int_{\{\xi\geq\varepsilon\}} \varepsilon\,dP \leq \int_{\{\xi\geq\varepsilon\}} \xi\,dP = 0.$$

The last equality holds, since $\{\xi \geq \varepsilon\} \in \mathcal{G}$. Similarly, $P\{\xi \leq -\varepsilon\} = 0$ for any $\varepsilon > 0$. As a consequence,

$$P\{-\varepsilon < \xi < \varepsilon\} = 1$$

for any $\varepsilon > 0$.

Let us put

$$A_n = \left\{-\tfrac{1}{n} < \xi < \tfrac{1}{n}\right\}.$$

Then $P(A_n) = 1$ and

$$\{\xi = 0\} = \bigcap_{n=1}^{\infty} A_n.$$

Because the A_n form a contracting sequence of events, it follows that

$$P\{\xi = 0\} = \lim_{n\to\infty} P(A_n) = 1,$$

as required. □

One difficulty involved in Definition 2.3 is that no explicit formula for $E(\xi|\eta)$ is given. If such a formula is known, then it is usually fairly easy to verify conditions 1) and 2). But how do you find it in the first place? The examples and exercises below are designed to show how to tackle this problem in concrete cases.

Example 2.4

Take $\Omega = [0,1]$ with the σ-field of Borel sets and P the Lebesgue measure on $[0,1]$. We shall find $E(\xi|\eta)$ for

$$\xi(x) = 2x^2, \qquad \eta(x) = \begin{cases} 2 & \text{if } x \in [0,\frac{1}{2}), \\ x & \text{if } x \in [\frac{1}{2},1]. \end{cases}$$

Here η is no longer discrete and the general Definition 2.3 should be used.

First we need to describe the σ-field $\sigma(\eta)$. For any Borel set $B \subset [\frac{1}{2},1]$ we have

$$B = \{\eta \in B\} \in \sigma(\eta)$$

and

$$[0,\tfrac{1}{2}) \cup B = \{\eta \in B\} \cup \{\eta = 2\} \in \sigma(\eta).$$

In fact sets of these two kinds exhaust all elements of $\sigma(\eta)$. The inverse image $\{\eta \in C\}$ of any Borel set $C \subset \mathbb{R}$ is of the first kind if $2 \notin C$ and of the second kind if $2 \in C$.

If $E(\xi|\eta)$ is to be $\sigma(\eta)$-measurable, it must be constant on $[0,\frac{1}{2})$ because η is. If for any $x \in [0,\frac{1}{2})$

$$\begin{aligned} E(\xi|\eta)(x) &= E(\xi|[0,\tfrac{1}{2})) \\ &= \frac{1}{P([0,\frac{1}{2}))}\int_{[0,\frac{1}{2})} \xi(x)\,dx \\ &= \frac{1}{\frac{1}{2}}\int_0^{\frac{1}{2}} 2x^2\,dx \\ &= \frac{1}{6}, \end{aligned}$$

then

$$\int_{[0,\frac{1}{2})} E(\xi|\eta)(x)\,dx = \int_{[0,\frac{1}{2})} \xi(x)\,dx,$$

i.e. condition 2) of Definition 2.3 will be satisfied for $A = [0,\frac{1}{2})$.

Moreover, if $E(\xi|\eta) = \xi$ on $[\frac{1}{2},1]$, then of course

$$\int_B E(\xi|\eta)(x)\,dx = \int_B \xi(x)\,dx$$

for any Borel set $B \subset [\frac{1}{2},1]$.

Therefore, we have found that

$$E(\xi|\eta)(x) = \begin{cases} \frac{1}{6} & \text{if } x \in [0,\frac{1}{2}), \\ 2x^2 & \text{if } x \in [\frac{1}{2},1]. \end{cases}$$

Because every element of $\sigma(\eta)$ is of the form B or $[0,\frac{1}{2}) \cup B$, where $B \subset [\frac{1}{2},1]$ is a Borel set, it follows immediately that conditions 1) and 2) of Definition 2.3 are satisfied. The graph of $E(\xi|\eta)$ is presented in Figure 2.2 along with those of ξ and η.

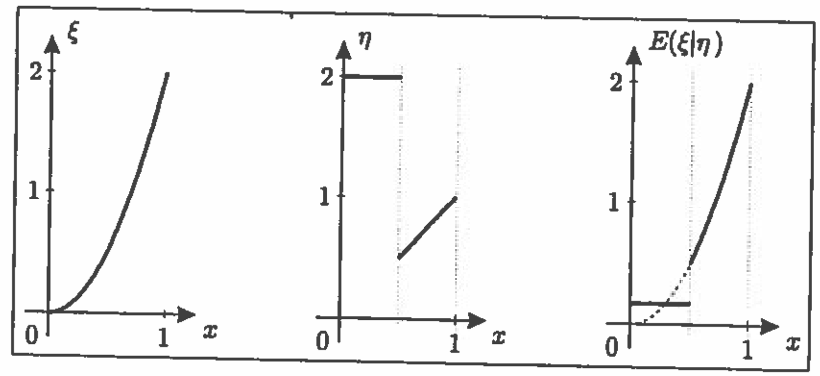

Figure 2.2. The graph of $E(\xi|\eta)$ in Example 2.4

Exercise 2.6

Let $\Omega = [0,1]$ with Lebesgue measure as in Example 2.4. Find the conditional expectation $E(\xi|\eta)$ if

$$\xi(x) = 2x^2, \qquad \eta(x) = 1 - |2x - 1|.$$

Hint First describe the σ-field generated by η. Observe that η is symmetric about $\frac{1}{2}$. What does it tell you about the sets in $\sigma(\eta)$? What does it tell you about $E(\xi|\eta)$ if it is to be $\sigma(\eta)$-measurable? Does it need to be symmetric as well? For any A in $\sigma(\eta)$ try to transform $\int_A \xi\,dP$ to make the integrand symmetric.

Exercise 2.7

Let Ω be the unit square $[0,1] \times [0,1]$ with the σ-field of Borel sets and P the Lebesgue measure on $[0,1] \times [0,1]$. Suppose that ξ and η are random variables on Ω with joint density

$$f_{\xi,\eta}(x,y) = x + y$$

for any $x, y \in [0,1]$, and $f_{\xi,\eta}(x,y) = 0$ otherwise. Show that

$$E(\xi|\eta) = \frac{2+3\eta}{3+6\eta}.$$

Hint It suffices (why?) to show that for any Borel set B

$$\int_{\{\eta\in B\}} \xi\, dP = \int_{\{\eta\in B\}} \frac{2+3\eta}{3+6\eta}\, dP.$$

Try to express each side of this equality as an integral over the square $[0,1]\times[0,1]$ using the joint density $f_{\xi,\eta}(x,y)$.

Exercise 2.8

Let Ω be the unit square $[0,1]\times[0,1]$ with Lebesgue measure as in Exercise 2.7. Find $E(\xi|\eta)$ if ξ and η are random variables on Ω with joint density

$$f_{\xi,\eta}(x,y) = \frac{3}{2}(x^2+y^2)$$

for any $x, y \in [0,1]$, and $f_{\xi,\eta}(x,y) = 0$ otherwise.

Hint This is slightly harder than Exercise 2.7 because here we have to derive a formula for the conditional expectation. Study the solution to Exercise 2.7 to find a way of obtaining such a formula.

Exercise 2.9

Let Ω be the unit disc $\{(x,y): x^2+y^2 \le 1\}$ with the σ-field of Borel sets and P the Lebesgue measure on the disc normalized so that $P(\Omega) = 1$, i.e.

$$P(A) = \frac{1}{\pi}\int\int_A dx\, dy$$

for any Borel set $A \subset \Omega$. Suppose that ξ and η are the projections onto the x and y axes,

$$\xi(x,y) = x, \qquad \eta(x,y) = y$$

for any $(x,y) \in \Omega$. Find $E(\xi^2|\eta)$.

Hint What is the joint density of ξ and η? Use this density to transform the integral

$$\int_{\{\eta\in B\}} \xi^2\, dP$$

for an arbitrary Borel set B so that the integrand becomes a function of η. How is this function of η related to $E(\xi^2|\eta)$?

2.4 Conditioning on a σ-Field

We are now in a position to make the final step towards the general definition of conditional expectation. It is based on the observation that $E(\xi|\eta)$ depends only on the σ-field $\sigma(\eta)$ generated by η, rather than on the actual values of η.

Proposition 2.2

If $\sigma(\eta) = \sigma(\eta')$, then $E(\xi|\eta) = E(\xi|\eta')$ a.s. (Compare this with (2.2).)

Proof

This is an immediate consequence of Lemma 2.1. □

Because of Proposition 2.2 it is reasonable to talk of conditional expectation given a σ-field. The definition below differs from Definition 2.3 only by using an arbitrary σ-field $\mathcal{G}$ in place of a σ-field $\sigma(\eta)$ generated by a random variable η.

Definition 2.4

Let ξ be an integrable random variable on a probability space $(\Omega, \mathcal{F}, P)$, and let $\mathcal{G}$ be a σ-field contained in $\mathcal{F}$. Then the ***conditional expectation*** of ξ given $\mathcal{G}$ is defined to be a random variable $E(\xi|\mathcal{G})$ such that

1) $E(\xi|\mathcal{G})$ is $\mathcal{G}$-measurable;
2) For any $A \in \mathcal{G}$
$$\int_A E(\xi|\mathcal{G})\, dP = \int_A \xi\, dP. \tag{2.3}$$

Remark 2.2

The ***conditional probability*** of an event $A \in \mathcal{F}$ given a σ-field $\mathcal{G}$ can be defined by

$$P(A|\mathcal{G}) = E(1_A|\mathcal{G}),$$

where 1_A is the indicator function of A.

The notion of conditional expectation with respect to a σ-field extends conditioning on a random variable η in the sense that

$$E(\xi|\sigma(\eta)) = E(\xi|\eta),$$

where $\sigma(\eta)$ is the σ-field generated by η.

Proposition 2.3

$E(\xi|\mathcal{G})$ exists and is unique in the sense that if $\xi = \xi'$ a.s., then $E(\xi|\mathcal{G}) = E(\xi'|\mathcal{G})$ a.s.

Proof

Existence and uniqueness follow, respectively, from Theorem 2.1 below and Lemma 2.1. □

Theorem 2.1 (Radon–Nikodym)

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathcal{G}$ be a σ-field contained in $\mathcal{F}$. Then for any random variable ξ there exists a $\mathcal{G}$-measurable random variable ζ such that

$$\int_A \xi \, dP = \int_A \zeta \, dP$$

for each $A \in \mathcal{G}$.

The Radon–Nikodym theorem is important from a theoretical point of view. However, in practice there are usually other ways of establishing the existence of conditional expectation, for example, by finding an explicit formula, as in the examples and exercises in the previous section. The proof of the Radon–Nikodym theorem is beyond the scope of this course and is omitted.

Exercise 2.10

Show that if $\mathcal{G} = \{\emptyset, \Omega\}$, then $E(\xi|\mathcal{G}) = E(\xi)$ a.s.

Hint What random variables are $\mathcal{G}$-measurable if $\mathcal{G} = \{\emptyset, \Omega\}$?

Exercise 2.11

Show that if ξ is $\mathcal{G}$-measurable, then $E(\xi|\mathcal{G}) = \xi$ a.s.

Hint The conditions of Definition 2.4 are trivially satisfied by ξ if ξ is $\mathcal{G}$-measurable.

Exercise 2.12

Show that if $B \in \mathcal{G}$, then

$$E\left(E\left(\xi|\mathcal{G}\right)|B\right) = E\left(\xi|B\right).$$

Hint The conditional expectation on either side of the equality involves an integral over B. How are these integrals related to one another?

2.5 General Properties

Proposition 2.4

Conditional expectation has the following properties:

1) $E(a\xi + b\zeta|\mathcal{G}) = aE(\xi|\mathcal{G}) + bE(\zeta|\mathcal{G})$ (*linearity*);
2) $E(E(\xi|\mathcal{G})) = E(\xi)$;
3) $E(\xi\zeta|\mathcal{G}) = \xi E(\zeta|\mathcal{G})$ if ξ is $\mathcal{G}$-measurable (*taking out what is known*);
4) $E(\xi|\mathcal{G}) = E(\xi)$ if ξ is independent of $\mathcal{G}$ (*an independent condition drops out*);
5) $E(E(\xi|\mathcal{G})|\mathcal{H}) = E(\xi|\mathcal{H})$ if $\mathcal{H} \subset \mathcal{G}$ (*tower property*);
6) If $\xi \geq 0$, then $E(\xi|\mathcal{G}) \geq 0$ (*positivity*).

Here a, b are arbitrary real numbers, ξ, ζ are integrable random variables on a probability space $(\Omega, \mathcal{F}, P)$ and $\mathcal{G}, \mathcal{H}$ are σ-fields on Ω contained in $\mathcal{F}$. In 3) we also assume that the product $\xi\zeta$ is integrable. All equalities and the inequalities in 6) hold P-a.s.

Proof

1) For any $B \in \mathcal{G}$

$$\begin{aligned} \int_B (aE(\xi|\mathcal{G}) + bE(\zeta|\mathcal{G})) \, dP &= a\int_B E(\xi|\mathcal{G}) \, dP + b\int_B E(\zeta|\mathcal{G}) \, dP \\ &= a\int_B \xi \, dP + b\int_B \zeta \, dP \\ &= \int_B (a\xi + b\zeta) \, dP. \end{aligned}$$

By uniqueness this proves the desired equality.

2) This follows by putting $A = \Omega$ in (2.3). Also, 2) is a special case of 5) when $\mathcal{H} = \{\emptyset, \Omega\}$.

3) We first verify the result for $\xi = 1_A$, where $A \in \mathcal{G}$. In this case

$$\begin{aligned} \int_B 1_A E(\eta|\mathcal{G}) \, dP &= \int_{A\cap B} E(\eta|\mathcal{G}) \, dP \\ &= \int_{A\cap B} \eta \, dP \\ &= \int_B 1_A \eta \, dP \end{aligned}$$

for any $B \in \mathcal{G}$, which implies that

$$1_A E(\eta|\mathcal{G}) = E(1_A \eta|\mathcal{G})$$

by uniqueness. In a similar way we obtain the result if ξ is a $\mathcal{G}$-measurable step function,

$$\xi = \sum_{j=1}^{m} a_j 1_{A_j},$$

where $A_j \in \mathcal{G}$ for $j = 1, \ldots, m$. Finally, the result in the general case follows by approximating ξ by $\mathcal{G}$-measurable step functions.

4) Since ξ is independent of $\mathcal{G}$, the random variables ξ and 1_B are independent for any $B \in \mathcal{G}$. It follows by Proposition 1.1 (independent random variables are uncorrelated) that

$$\begin{aligned}\int_B E(\xi)\, dP &= E(\xi)E(1_B) \\ &= E(\xi 1_B) \\ &= \int_B \xi\, dP,\end{aligned}$$

which proves the assertion.

5) By Definition 2.4

$$\int_B E(\xi|\mathcal{G})\, dP = \int_B \xi\, dP$$

for every $B \in \mathcal{G}$, and

$$\int_B E(\xi|\mathcal{H})\, dP = \int_B \xi\, dP$$

for every $B \in \mathcal{H}$. Because $\mathcal{H} \subset \mathcal{G}$ it follows that

$$\int_B E(\xi|\mathcal{G})\, dP = \int_B E(\xi|\mathcal{H})\, dP$$

for every $B \in \mathcal{H}$. Applying Definition 2.4 once again, we obtain

$$E(E(\xi|\mathcal{G})|\mathcal{H}) = E(\xi|\mathcal{H}).$$

6) For any n we put

$$A_n = \left\{ E(\xi|\mathcal{G}) \le -\frac{1}{n} \right\}.$$

Then $A_n \in \mathcal{G}$. If $\xi \ge 0$ a.s., then

$$0 \le \int_{A_n} \xi\, dP = \int_{A_n} E(\xi|\mathcal{G})\, dP \le -\frac{1}{n} P(A_n),$$

which means that $P(A_n) = 0$. Because

$$\{E(\xi|\mathcal{G}) < 0\} = \bigcup_{n=1}^{\infty} A_n$$

it follows that

$$P\{E(\xi|\mathcal{G}) < 0\} = 0,$$

completing the proof. □

The next theorem, which will be stated without proof, involves the notion of a convex function, such as $\max(1, x)$ or e^{-x}, for example. In this course the theorem will be used mainly for $|x|$, which is also a convex function. In general, we call a function $\varphi : \mathbb{R} \to \mathbb{R}$ *convex* if for any $x, y \in \mathbb{R}$ and any $\lambda \in [0,1]$

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

This condition means that the graph of φ lies below the cord connecting the points with coordinates $(x, \varphi(x))$ and $(y, \varphi(y))$.

Theorem 2.2 (Jensen's Inequality)

Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function and let ξ be an integrable random variable on a probability space $(\Omega, \mathcal{F}, P)$ such that $\varphi(\xi)$ is also integrable. Then

$$\varphi(E(\xi|\mathcal{G})) \le E(\varphi(\xi)|\mathcal{G}) \quad \text{a.s.}$$

for any σ-field $\mathcal{G}$ on Ω contained in $\mathcal{F}$.

2.6 Various Exercises on Conditional Expectation

Exercise 2.13

Mrs. Jones has made a steak and kidney pie for her two sons. Eating more than a half of it will give indigestion to anyone. While she is away having tea with a neighbour, the older son helps himself to a piece of the pie. Then the younger son comes and has a piece of what is left by his brother. We assume that the size of each of the two pieces eaten by Mrs. Jones' sons is random and uniformly distributed over what is currently available. What is the expected size of the remaining piece given that neither son gets indigestion?

Hint All possible outcomes can be represented by pairs of numbers, the portions of the pie consumed by the two sons. Therefore Ω can be chosen as a subset of the plane. Observe that the older son is restricted only by the size of the pie, while the younger one is restricted by what is left by his brother. This will determine the shape of Ω. Next introduce a probability measure on Ω consistent with the conditions of the exercise. This can be done by means of a suitable density over Ω. Now you are in a position to compute the probability that neither son will get indigestion. What is the corresponding subset of Ω? Finally, define a random variable on Ω representing the portion of the pie left by the sons and compute the conditional expectation.

Exercise 2.14

As a probability space take $\Omega = [0,1)$ with the σ-field of Borel sets and the Lebesgue measure on $[0,1)$. Find $E(\xi|\eta)$ if

$$\xi(x) = 2x^2, \qquad \eta(x) = \begin{cases} 2x & \text{for } 0 \le x < \frac{1}{2}, \\ 2x-1 & \text{for } \frac{1}{2} \le x < 1. \end{cases}$$

Hint What do events in $\sigma(\eta)$ look like? What do $\sigma(\eta)$-measurable random variables look like? If you devise a neat way of describing these, it will make the task of finding $E(\xi|\eta)$ much easier. You will need to transform the integrals in condition 2) of Definition 2.3 to find a formula for the conditional expectation.

Exercise 2.15

Take $\Omega = [0,1]$ with the σ-field of Borel sets and P the Lebesgue measure on $[0,1]$. Let

$$\eta(x) = x(1-x)$$

for $x \in [0,1]$. Show that

$$E(\xi|\eta)(x) = \frac{\xi(x) + \xi(1-x)}{2}$$

for any $x \in [0,1]$.

Hint Observe that $\eta(x) = \eta(1-x)$. What does it tell you about the σ-field generated by η? Is $\frac{1}{2}(\xi(x) + \xi(1-x))$ measurable with respect to this σ-field? If so, it remains to verify condition 2) of Definition 2.3.

Exercise 2.16

Let ξ, η be integrable random variables with joint density $f_{\xi,\eta}(x,y)$. Show that

$$E(\xi|\eta) = \frac{\int_{\mathbb{R}} x\, f_{\xi,\eta}(x,\eta)\, dx}{\int_{\mathbb{R}} f_{\xi,\eta}(x,\eta)\, dx}.$$

Hint Study the solutions to Exercises 2.7 and 2.8.

Remark 2.3

If we put

$$f_{\xi,\eta}(x|y) = \frac{f_{\xi,\eta}(x,y)}{f_\eta(y)},$$

where

$$f_\eta(y) = \int_{\mathbb{R}} f_{\xi,\eta}(x,y)\, dx$$

is the density of η, then by the result in Exercise 2.16

$$E(\xi|\eta) = \int_{\mathbb{R}} x\, f_{\xi,\eta}(x|\eta)\, dx.$$

We call $f_{\xi,\eta}(x|y)$ the *conditional density* of ξ given η.

Exercise 2.17

Consider $L^2(\mathcal{F}) = L^2(\Omega, \mathcal{F}, P)$ as a Hilbert space with scalar product

$$L^2(\mathcal{F}) \times L^2(\mathcal{F}) \ni (\xi, \zeta) \mapsto E(\xi\zeta) \in \mathbb{R}.$$

Show that if ξ is a random variable in $L^2(\mathcal{F})$ and $\mathcal{G}$ is a σ-field contained in $\mathcal{F}$, then $E(\xi|\mathcal{G})$ is the orthogonal projection of ξ onto the subspace $L^2(\mathcal{G})$ in $L^2(\mathcal{F})$ consisting of $\mathcal{G}$-measurable random variables.

Hint Observe that condition 2) of Definition 2.4 means that $\xi - E(\xi|\mathcal{G})$ is orthogonal (in the sense of the scalar product above) to the indicator function 1_A for any $A \in \mathcal{G}$.

2.7 Solutions

Solution 2.1

Since $P(\Omega) = 1$ and $\int_\Omega \xi\, dP = E(\xi)$,

$$E(\xi|\Omega) = \frac{1}{P(\Omega)} \int_\Omega \xi\, dP = E(\xi).$$

Solution 2.2

By Definition 2.1

$$\begin{aligned} E(1_A|B) &= \frac{1}{P(B)} \int_B 1_A\, dP \\ &= \frac{1}{P(B)} \int_{A\cap B} dP \end{aligned}$$

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

Solution 2.3

Since η is constant, it has only one value $c \in \mathbb{R}$, for which

$$\{\eta = c\} = \Omega.$$

Therefore $E(\xi|\eta)$ is constant and equal to

$$E(\xi|\eta)(\omega) = E(\xi|\{\eta = c\}) = E(\xi|\Omega) = E(\xi)$$

for each $\omega \in \Omega$. The last equality has been verified in Exercise 2.1.

Solution 2.4

The indicator function 1_B takes two values 1 and 0. The sets on which these values are taken are

$$\{1_B = 1\} = B, \qquad \{1_B = 0\} = \Omega \setminus B.$$

Thus, for $\omega \in B$

$$E(1_A|1_B)(\omega) = E(1_A|B) = P(A|B),$$

as in Exercise 2.2. Similarly, for $\omega \in \Omega \setminus B$

$$E(1_A|1_B)(\omega) = E(1_A|\Omega \setminus B) = P(A|\Omega \setminus B).$$

Solution 2.5

First we observe that

$$\int_B E(\xi|B)\, dP = \int_B \left(\frac{1}{P(B)} \int_B \xi\, dP\right) dP = \int_B \xi\, dP \tag{2.4}$$

for any event B.

Since η is discrete, it has countably many values $y_1, y_2, \ldots$. We can assume that these values are pairwise distinct, i.e. $y_i \neq y_j$ if $i \neq j$. The sets $\{\eta = y_1\}, \{\eta = y_2\}, \ldots$ are then pairwise disjoint and they cover the whole space Ω. Therefore, by (2.4)

$$\begin{aligned} E(E(\xi|\eta)) &= \int_\Omega E(\xi|\eta)\, dP \\ &= \sum_n \int_{\{\eta = y_n\}} E(\xi|\{\eta = y_n\})\, dP \end{aligned}$$

$$\begin{aligned} &= \sum_n \int_{\{\eta = y_n\}} \xi\, dP \\ &= \int_\Omega \xi\, dP \\ &= E(\xi). \end{aligned}$$

Solution 2.6

First we need to describe the σ-field $\sigma(\eta)$ generated by η. Observe that η is symmetric about $\frac{1}{2}$,

$$\eta(x) = \eta(1 - x)$$

for any $x \in [0,1]$. We claim that $\sigma(\eta)$ consists of all Borel sets $A \subset [0,1]$ symmetric about $\frac{1}{2}$, i.e. such that

$$A = 1 - A. \tag{2.5}$$

Indeed, if A is such a set, then $A = \{\eta \in B\}$, where

$$B = \{2x : x \in A \cap [0, \tfrac{1}{2}]\}$$

is a Borel set, so $A \in \sigma(\eta)$. On the other hand, if $A \in \sigma(\eta)$, then there is a Borel set B in $\mathbb{R}$ such that $A = \{\eta \in B\}$. Then

$$\begin{aligned} x \in A &\Leftrightarrow \eta(x) \in B \\ &\Leftrightarrow \eta(1 - x) \in B \\ &\Leftrightarrow 1 - x \in A, \end{aligned}$$

so A satisfies (2.5).

We are ready to find $E(\xi|\eta)$. If it is to be $\sigma(\eta)$-measurable, it must be symmetric about $\frac{1}{2}$, i.e.

$$E(\xi|\eta)(x) = E(\xi|\eta)(1 - x)$$

for any $x \in [0,1]$. For any $A \in \sigma(\eta)$ we shall transform the integral below so as to make the integrand symmetric about $\frac{1}{2}$:

$$\begin{aligned} \int_A 2x^2\, dx &= \int_A x^2\, dx + \int_A x^2\, dx \\ &= \int_A x^2\, dx + \int_{1-A} (1 - x)^2\, dx \\ &= \int_A x^2\, dx + \int_A (1 - x)^2\, dx \\ &= \int_A \left(x^2 + (1 - x)^2\right) dx. \end{aligned}$$

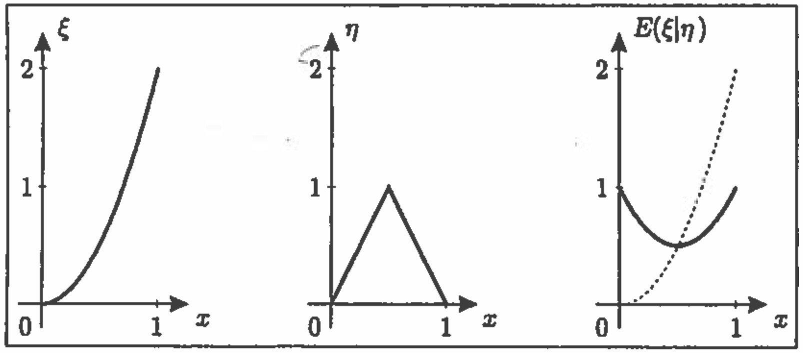

Figure 2.3. The graph of $E(\xi|\eta)$ in Exercise 2.6

It follows that

$$E(\xi|\eta)(x) = x^2 + (1-x)^2.$$

The graphs of ξ, η and $E(\xi|\eta)$ are shown in Figure 2.3.

Solution 2.7

Since

$$\{\eta \in B\} = [0,1] \times B$$

for any Borel set B, we have

$$\begin{aligned}\int_{\{\eta\in B\}} \xi\, dP &= \int_B \int_{\mathbb{R}} x\, f_{\xi,\eta}(x,y)\, dx\, dy \\ &= \int_B \left(\int_{[0,1]} x(x+y)\, dx \right) dy \\ &= \int_B \left(\frac{1}{3} + \frac{1}{2} y \right) dy\end{aligned}$$

and

$$\begin{aligned}\int_{\{\eta\in B\}} \frac{2+3\eta}{3+6\eta}\, dP &= \int_B \int_{\mathbb{R}} \frac{2+3y}{3+6y} f_{\xi,\eta}(x,y)\, dx\, dy \\ &= \int_B \frac{2+3y}{3+6y} \left(\int_{[0,1]} (x+y)\, dx \right) dy \\ &= \int_B \left(\frac{1}{3} + \frac{1}{2} y \right) dy.\end{aligned}$$

Because each event in $\sigma(\eta)$ is of the form $\{\eta \in B\}$ for some Borel set B, this means that condition 2) of Definition 2.3 is satisfied. The random variable $\frac{2+3\eta}{3+6\eta}$

is $\sigma(\eta)$-measurable, so condition 1) holds too. It follows that

$$E(\xi|\eta) = \frac{2+3\eta}{3+6\eta}.$$

Solution 2.8

We are looking for a Borel function $F : \mathbb{R} \to \mathbb{R}$ such that for any Borel set B

$$\int_{\{\eta\in B\}} \xi\, dP = \int_{\{\eta\in B\}} F(\eta)\, dP. \tag{2.6}$$

Then $E(\xi|\eta) = F(\eta)$ by Definition 2.3.

We shall transform both integrals above using the joint density $f_{\xi,\eta}(x,y)$ in much the same way as in the solution to Exercise 2.7, except that here we do not know the exact form of $F(y)$. Namely,

$$\begin{aligned}\int_{\{\eta\in B\}} \xi\, dP &= \int_B \int_{\mathbb{R}} x\, f_{\xi,\eta}(x,y)\, dx\, dy \\ &= \frac{3}{2} \int_B \left(\int_{[0,1]} x\left(x^2+y^2\right) dx \right) dy \\ &= \frac{3}{2} \int_B \left(\frac{1}{4} + \frac{1}{2} y^2 \right) dy\end{aligned}$$

and

$$\begin{aligned}\int_{\{\eta\in B\}} F(\eta)\, dP &= \int_B \int_{\mathbb{R}} F(y)\, f_{\xi,\eta}(x,y)\, dx\, dy \\ &= \frac{3}{2} \int_B F(y) \left(\int_{[0,1]} \left(x^2+y^2\right) dx \right) dy \\ &= \frac{3}{2} \int_B F(y) \left(\frac{1}{3} + y^2 \right) dy.\end{aligned}$$

Then, (2.6) will hold for any Borel set B if

$$F(y) = \frac{\frac{1}{4} + \frac{1}{2}y^2}{\frac{1}{3} + y^2} = \frac{3+6y^2}{4+12y^2}.$$

It follows that

$$E(\xi|\eta) = F(\eta) = \frac{3+6\eta^2}{4+12\eta^2}.$$

Solution 2.9

We are looking for a Borel function $F : \mathbb{R} \to \mathbb{R}$ such that for any Borel set $B \subset \mathbb{R}$

$$\int_{\{\eta\in B\}} \xi^2\, dP = \int_{\{\eta\in B\}} F(\eta)\, dP. \tag{2.7}$$

Then, by Definition 2.3 we shall have $E(\xi^2|\eta) = F(\eta)$.

Let us transform both sides of (2.7). To do so we observe that the random variables ξ and η have uniform joint distribution over the unit disc $\Omega = \{(x,y) : x^2 + y^2 \le 1\}$, with density

$$f_{\xi,\eta}(x,y) = \frac{1}{\pi}$$

if $x^2 + y^2 \le 1$, and $f_{\xi,\eta}(x,y) = 0$ otherwise. It follows that

$$\begin{aligned}\int_{\{\eta\in B\}} \xi^2\,dP &= \int_B \int_{\mathbb{R}} x^2 f_{\xi,\eta}(x,y)\,dx\,dy \\ &= \frac{1}{\pi}\int_B \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} x^2\,dx\,dy \\ &= \frac{2}{3\pi}\int_B \left(1-y^2\right)^{3/2} dy\end{aligned}$$

and

$$\begin{aligned}\int_{\{\eta\in B\}} F(\eta)\,dP &= \int_B \int_{\mathbb{R}} F(y)\, f_{\xi,\eta}(x,y)\,dx\,dy \\ &= \frac{1}{\pi}\int_B F(y) \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} dx\,dy \\ &= \frac{2}{\pi}\int_B F(y)\left(1-y^2\right)^{1/2} dy.\end{aligned}$$

If (2.7) is to be satisfied for all Borel sets B, then

$$F(y) = \frac{1}{3}\left(1-y^2\right).$$

This means that

$$E(\xi^2|\eta)(x,y) = F(\eta(x,y)) = F(y) = \frac{1}{3}\left(1-y^2\right)$$

for any (x,y) in Ω.

Solution 2.10

If $\mathcal{G} = \{\emptyset, \Omega\}$, then any constant random variable is $\mathcal{G}$-measurable. Since

$$\int_\Omega \xi\,dP = E(\xi) = \int_\Omega E(\xi)\,dP$$

and

$$\int_\emptyset \xi\,dP = 0 = \int_\emptyset E(\xi)\,dP,$$

it follows that $E(\xi|\mathcal{G}) = E(\xi)$ a.s., as required.

Solution 2.11

Because the trivial identity

$$\int_A \xi\,dP = \int_A \xi\,dP$$

holds for any $A \in \mathcal{G}$ and ξ is $\mathcal{G}$-measurable, it follows that $E(\xi|\mathcal{G}) = \xi$ a.s.

Solution 2.12

By Definition 2.3

$$\int_B E(\xi|\mathcal{G})\,dP = \int_B \xi\,dP,$$

since $B \in \mathcal{G}$. It follows that

$$\begin{aligned}E(E(\xi|\mathcal{G})|B) &= \frac{1}{P(B)}\int_B E(\xi|\mathcal{G})\,dP \\ &= \frac{1}{P(B)}\int_B \xi\,dP \\ &= E(\xi|B).\end{aligned}$$

Solution 2.13

The whole pie will be represented by the interval $[0,1]$. Let $x \in [0,1]$ be the portion consumed by the older son. Then $[0, 1-x]$ will be available to the younger one, who takes a portion of size $y \in [0, 1-x]$. The set of all possible outcomes is

$$\Omega = \{(x,y) : x, y \ge 0, x + y \le 1\}.$$

The event that neither of Mrs. Jones' sons will get indigestion is

$$A = \left\{(x,y) \in \Omega : x, y < \frac{1}{2}\right\}.$$

These sets are shown in Figure 2.4. If x is uniformly distributed over $[0,1]$ and y is uniformly distributed over $[0, 1-x]$, then the probability measure P over Ω with density

$$f(x,y) = \frac{1}{1-x}$$

will describe the joint distribution of outcomes (x,y), see Figure 2.5.

Now we are in a position to compute

$$\begin{aligned}P(A) &= \int_A f(x,y)\,dx\,dy \\ &= \int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}} \frac{1}{1-x}\,dx\,dy \\ &= \ln\sqrt{2}.\end{aligned}$$

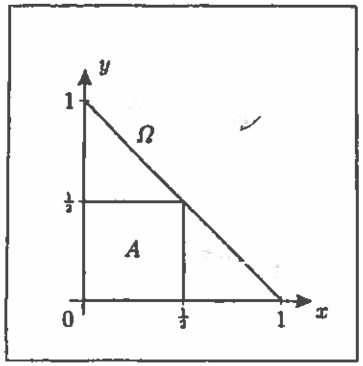

Figure 2.4. The sets Ω and A in Exercise 2.13

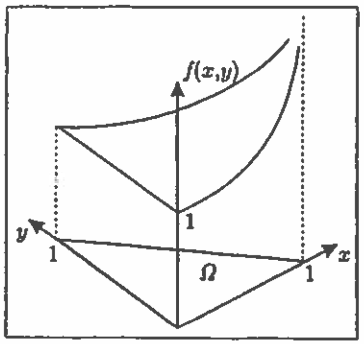

Figure 2.5. The density $f(x,y)$ in Exercise 2.13

The random variable

$$\xi(x,y) = 1 - x - y$$

defined on Ω represents the size of the portion left by Mrs. Jones' sons. Finally, we find that

$$\begin{aligned} E(\xi|A) &= \frac{1}{P(A)}\int_A (1-x-y)\, f(x,y)\, dx\, dy \\ &= \frac{1}{\ln\sqrt{2}}\int_0^{\frac{1}{2}}\int_0^{\frac{1}{2}} \frac{1-x-y}{1-x}\, dx\, dy \\ &= \frac{1-\ln\sqrt{2}}{\ln 4}. \end{aligned}$$

Solution 2.14

The σ-field $\sigma(\eta)$ generated by η consists of sets of the form $B \cup \left(B+\frac{1}{2}\right)$ for some Borel set $B \subset [0,\frac{1}{2})$. Thus, we are looking for a $\sigma(\eta)$-measurable random variable ζ such that for each Borel set $B \subset [0,\frac{1}{2})$

$$\int_{B\cup(B+\frac{1}{2})} \xi(x)\, dx = \int_{B\cup(B+\frac{1}{2})} \zeta(x)\, dx. \tag{2.8}$$

Then $E(\xi|\eta) = \zeta$ by Definition 2.3.

Transforming the integral on the left-hand side, we obtain

$$\begin{aligned} \int_{B\cup(B+\frac{1}{2})} \xi(x)\, dx &= \int_B 2x^2\, dx + \int_{B+\frac{1}{2}} 2x^2\, dx \\ &= \int_B 2x^2\, dx + \int_B 2\left(x+\tfrac{1}{2}\right)^2 dx \\ &= 2\int_B \left(x^2 + \left(x+\tfrac{1}{2}\right)^2\right) dx. \end{aligned}$$

For ζ to be $\sigma(\eta)$-measurable it must satisfy

$$\zeta(x) = \zeta\left(x+\tfrac{1}{2}\right) \tag{2.9}$$

for each $x \in [0,\frac{1}{2})$. Then

$$\begin{aligned} \int_{B\cup(B+\frac{1}{2})} \zeta(x)\, dP &= \int_B \zeta(x)\, dx + \int_{B+\frac{1}{2}} \zeta(x)\, dx \\ &= \int_B \zeta(x)\, dx + \int_B \zeta\left(x+\tfrac{1}{2}\right) dx \\ &= \int_B \zeta(x)\, dx + \int_B \zeta(x)\, dx \\ &= 2\int_B \zeta(x)\, dx. \end{aligned}$$

If (2.8) is to hold for any Borel set $B \subset [0,\frac{1}{2})$, then

$$\zeta(x) = x^2 + \left(x+\tfrac{1}{2}\right)^2$$

for each $x \in [0,\frac{1}{2})$. The values of $\zeta(x)$ for $x \in [\frac{1}{2},1)$ can be obtained from (2.9). It follows that

$$E(\xi|\eta)(x) = \zeta(x) = \begin{cases} x^2 + \left(x+\frac{1}{2}\right)^2 & \text{for } 0 \le x < \frac{1}{2}, \\ \left(x-\frac{1}{2}\right)^2 + x^2 & \text{for } \frac{1}{2} \le x < 1. \end{cases}$$

The graphs of ξ, η and $E(\xi|\eta)$ are shown in Figure 2.6.

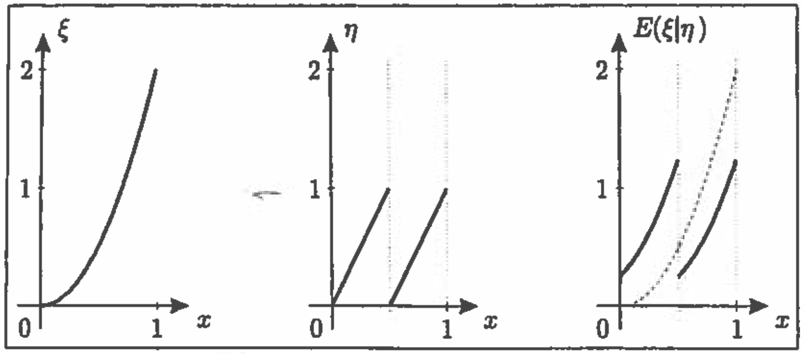

Figure 2.6. The graph of $E(\xi|\eta)$ in Exercise 2.14

Solution 2.15

Since $\eta(x) = \eta(1-x)$, the σ-field $\sigma(\eta)$ consists of Borel sets $B \subset [0,1]$ such that

$$B = 1 - B,$$

where $1 - B = \{1 - x : x \in B\}$. For any such B

$$\begin{aligned}
\int_B \xi(x)\,dx &= \frac{1}{2}\int_B \xi(x)\,dx + \frac{1}{2}\int_B \xi(x)\,dx \\
&= \frac{1}{2}\int_B \xi(x)\,dx + \frac{1}{2}\int_{1-B} \xi(1-x)\,dx \\
&= \frac{1}{2}\int_B \xi(x)\,dx + \frac{1}{2}\int_B \xi(1-x)\,dx \\
&= \int_B \frac{\xi(x)+\xi(1-x)}{2}\,dx.
\end{aligned}$$

Because $\frac{1}{2}(\xi(x) + \xi(1-x))$ is $\sigma(\eta)$-measurable, it follows that

$$E(\xi|\eta)(x) = \frac{\xi(x)+\xi(1-x)}{2}.$$

Solution 2.16

We are looking for a Borel function $F(y)$ such that

$$\int_{\{\eta\in B\}} \xi\,dP = \int_{\{\eta\in B\}} F(\eta)\,dP$$

for any Borel set B in $\mathbb{R}$. Because $F(\eta)$ is $\sigma(\eta)$-measurable and each event in $\sigma(\eta)$ can be written as $\{\eta \in B\}$ for some Borel set B, this will mean that $E(\xi|\eta) = F(\eta)$.

Let us transform the two integrals above using the joint density of ξ and η:

$$\begin{aligned}
\int_{\{\eta\in B\}} \xi\,dP &= \int_B\int_{\mathbb{R}} x\, f_{\xi,\eta}(x,y)\,dx\,dy \\
&= \int_B \left(\int_{\mathbb{R}} x\, f_{\xi,\eta}(x,y)\,dx\right) dy
\end{aligned}$$

and

$$\begin{aligned}
\int_{\{\eta\in B\}} F(\eta)\,dP &= \int_B\int_{\mathbb{R}} F(y)\, f_{\xi,\eta}(x,y)\,dx\,dy \\
&= \int_B F(y) \left(\int_{\mathbb{R}} f_{\xi,\eta}(x,y)\,dx\right) dy.
\end{aligned}$$

If these two integrals are to be equal for each Borel set B, then

$$F(y) = \frac{\int_{\mathbb{R}} x\, f_{\xi,\eta}(x,y)\,dx}{\int_{\mathbb{R}} f_{\xi,\eta}(x,y)\,dx}.$$

It follows that

$$E(\xi|\eta) = F(\eta) = \frac{\int_{\mathbb{R}} x\, f_{\xi,\eta}(x,\eta)\,dx}{\int_{\mathbb{R}} f_{\xi,\eta}(x,\eta)\,dx}.$$

Solution 2.17

We denote by ζ the orthogonal projection of ξ onto the subspace $L^2(\mathcal{G}) \subset L^2(\mathcal{F})$ consisting of $\mathcal{G}$-measurable random variables. Thus, $\xi - \zeta$ is orthogonal to $L^2(\mathcal{G})$, that is,

$$E[(\xi-\zeta)\gamma] = 0$$

for each $\gamma \in L^2(\mathcal{G})$. But for any $A \in \mathcal{G}$ the indicator function 1_A belongs to $L^2(\mathcal{G})$, so

$$E[(\xi-\zeta)1_A] = 0.$$

Therefore

$$\int_A \xi\,dP = E(\xi 1_A) = E(\zeta 1_A) = \int_A \zeta\,dP$$

for any $A \in \mathcal{G}$. This means that $\zeta = E(\xi|\mathcal{G})$.

3
Martingales in Discrete Time

3.1 Sequences of Random Variables

A sequence $\xi_1, \xi_2, \ldots$ of random variables is typically used as a mathematical model of the outcomes of a series of random phenomena, such as coin tosses or the value of the FTSE All-Share Index at the London Stock Exchange on consecutive business days. The random variables in such a sequence are indexed by whole numbers, which are customarily referred to as ***discrete time***. It is important to understand that these whole numbers are not necessarily related to the physical time when the events modelled by the sequence actually occur. Discrete time is used to keep track of the order of events, which may or may not be evenly spaced in physical time. For example, the share index is recorded only on business days, but not on Saturdays, Sundays or any other holidays. Rather than tossing a coin repeatedly, we may as well toss 100 coins at a time and count the outcomes.

Definition 3.1

The sequence of numbers $\xi_1(\omega), \xi_2(\omega), \ldots$ for any fixed $\omega \in \Omega$ is called a ***sample path***.

A sample path for a sequence of coin tosses is presented in Figure 3.1 ($+1$ stands for heads and -1 for tails). Figure 3.2 shows the sample path of the FTSE All-Share Index up to 1997. Strictly speaking the pictures should con-

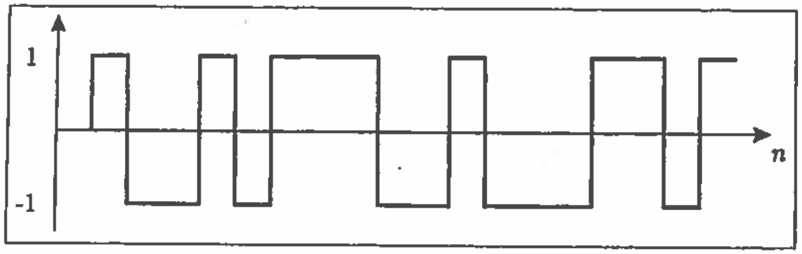

Figure 3.1. Sample path for a sequence of coin tosses

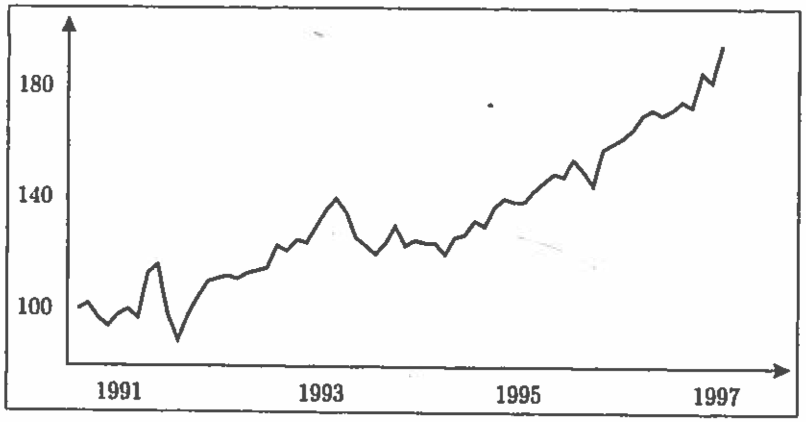

Figure 3.2. Sample path representing the FTSE All-Share Index up to 1997

sist of dots, representing the values $\xi_1(\omega), \xi_2(\omega), \ldots$, but it is customary to connect them by a broken line for illustration purposes.

3.2 Filtrations

As the time n increases, so does our knowledge about what has happened in the past. This can be modelled by a filtration as defined below.

Definition 3.2

A sequence of σ-fields $\mathcal{F}_1, \mathcal{F}_2, \ldots$ on Ω such that

$$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}$$

is called a *filtration*.

Here $\mathcal{F}_n$ represents our knowledge at time n. It contains all events A such that at time n it is possible to decide whether A has occurred or not. As n increases, there will be more such events A, i.e. the family $\mathcal{F}_n$ representing our knowledge will become larger. (The longer you live the wiser you become!)

Example 3.1

For a sequence $\xi_1, \xi_2, \ldots$ of coin tosses we take $\mathcal{F}_n$ to be the σ-field generated by $\xi_1, \ldots, \xi_n$,

$$\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n).$$

Let

$$A = \{\text{the first 5 tosses produce at least 2 heads}\}.$$

At discrete time $n = 5$, i.e. once the coin has been tossed five times, it will be possible to decide whether A has occurred or not. This means that $A \in \mathcal{F}_5$. However, at $n = 4$ it is not always possible to tell if A has occurred or not. If the outcomes of the first four tosses are, say,

tails, tails, heads, tails,

then the event A remains undecided. We will have to toss the coin once more to see what happens. Therefore $A \notin \mathcal{F}_4$.

This example illustrates another relevant issue. Suppose that the outcomes of the first four coin tosses are

tails, heads, tails, heads.

In this case it is possible to tell that A has occurred already at $n = 4$, whatever the outcome of the fifth toss will be. It does not mean, however, that A belongs to $\mathcal{F}_4$. The point is that for A to belong to $\mathcal{F}_4$ it must be possible to tell whether A has occurred or not after the first four tosses, *no matter what the first four outcomes are*. This is clearly not so in the example in hand.

Exercise 3.1

Let $\xi_1, \xi_2, \ldots$ be a sequence of coin tosses and let $\mathcal{F}_n$ be the σ-field generated by $\xi_1, \ldots, \xi_n$. For each of the following events find the smallest n such that the event belongs to $\mathcal{F}_n$:

$A = \{$the first occurrence of heads is preceeded by no more than 10 tails$\}$,
$B = \{$there is at least 1 head in the sequence $\xi_1, \xi_2, \ldots\}$,
$C = \{$the first 100 tosses produce the same outcome$\}$,
$D = \{$there are no more than 2 heads and 2 tails among the first 5 tosses$\}$.

Hint To find the smallest element in a set of numbers you need to make sure that the set is non-empty in the first place.

Suppose that $\xi_1, \xi_2, \dots$ is a sequence of random variables and $\mathcal{F}_1, \mathcal{F}_2, \dots$ is a filtration. A priori, they may have nothing in common. However, in practice the filtration will usually contain the knowledge accumulated by observing the outcomes of the random sequence, as in Example 3.1. The condition in the definition below means that $\mathcal{F}_n$ contains everything that can be learned from the values of $\xi_1, \dots, \xi_n$. In general, it may and often does contain more information.

Definition 3.3

We say that a sequence of random variables $\xi_1, \xi_2, \dots$ is *adapted* to a filtration $\mathcal{F}_1, \mathcal{F}_2, \dots$ if ξ_n is $\mathcal{F}_n$-measurable for each $n = 1, 2, \dots$.

Example 3.2

If $\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n)$ is the σ-field generated by $\xi_1, \dots, \xi_n$, then $\xi_1, \xi_2, \dots$ is adapted to $\mathcal{F}_1, \mathcal{F}_2, \dots$.

Exercise 3.2

Show that

$$\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n)$$

is the smallest filtration such that the sequence $\xi_1, \xi_2, \dots$ is adapted to $\mathcal{F}_1, \mathcal{F}_2, \dots$. That is to say, if $\mathcal{G}_1, \mathcal{G}_2, \dots$ is another filtration such that $\xi_1, \xi_2, \dots$ is adapted to $\mathcal{G}_1, \mathcal{G}_2, \dots$, then $\mathcal{F}_n \subset \mathcal{G}_n$ for each n.

Hint For $\sigma(\xi_1, \dots, \xi_n)$ to be contained in $\mathcal{G}_n$ you need to show that $\xi_1, \dots, \xi_n$ are $\mathcal{G}_n$-measurable.

3.3 Martingales

The concept of a martingale has its origin in gambling, namely, it describes a fair game of chance, which will be discussed in detail in the next section. Similarly, the notions of submartingale and supermartingale defined below are related to favourable and unfavourable games of chance. Some aspects of gambling are inherent in the mathematics of finance, in particular, the theory of financial derivatives such as options. Not surprisingly, martingales play a crucial role there. In fact, martingales reach well beyond game theory and appear

in various areas of modern probability and stochastic analysis, notably, in diffusion theory. First of all, let us introduce the basic definitions and properties in the case of discrete time.

Definition 3.4

A sequence $\xi_1, \xi_2, \dots$ of random variables is called a *martingale* with respect to a filtration $\mathcal{F}_1, \mathcal{F}_2, \dots$ if

1) ξ_n is integrable for each $n = 1, 2, \dots$;
2) $\xi_1, \xi_2, \dots$ is adapted to $\mathcal{F}_1, \mathcal{F}_2, \dots$;
3) $E(\xi_{n+1}|\mathcal{F}_n) = \xi_n$ a.s. for each $n = 1, 2, \dots$.

Example 3.3

Let $\eta_1, \eta_2, \dots$ be a sequence of independent integrable random variables such that $E(\eta_n) = 0$ for all $n = 1, 2, \dots$. We put

$$\begin{aligned} \xi_n &= \eta_1 + \dots + \eta_n, \\ \mathcal{F}_n &= \sigma(\eta_1, \dots, \eta_n). \end{aligned}$$

Then ξ_n is adapted to the filtration $\mathcal{F}_n$, and it is integrable because

$$\begin{aligned} E(|\xi_n|) &= E(|\eta_1 + \dots + \eta_n|) \\ &\leq E(|\eta_1|) + \dots + E(|\eta_n|) \\ &< \infty. \end{aligned}$$

Moreover,

$$\begin{aligned} E(\xi_{n+1}|\mathcal{F}_n) &= E(\eta_{n+1}|\mathcal{F}_n) + E(\xi_n|\mathcal{F}_n) \\ &= E(\eta_{n+1}) + \xi_n \\ &= \xi_n, \end{aligned}$$

since η_{n+1} is independent of $\mathcal{F}_n$ ('and independent condition drops out') and ξ_n is $\mathcal{F}_n$-measurable ('taking out what is known'). This means that ξ_n is a martingale with respect to $\mathcal{F}_n$.

Example 3.4

Let ξ be an integrable random variable and let $\mathcal{F}_1, \mathcal{F}_2, \dots$ be a filtration. We put

$$\xi_n = E(\xi|\mathcal{F}_n)$$

for $n = 1, 2, \dots$.

Then ξ_n is $\mathcal{F}_n$-measurable,

$$|\xi_n| = |E(\xi|\mathcal{F}_n)| \le E(|\xi|\,|\mathcal{F}_n),$$

which implies that

$$E(|\xi_n|) \le E(E(|\xi|\,|\mathcal{F}_n)) = E(|\xi|) < \infty,$$

and

$$E(\xi_{n+1}|\mathcal{F}_n) = E(E(\xi|\mathcal{F}_{n+1})|\mathcal{F}_n) = E(\xi|\mathcal{F}_n) = \xi_n,$$

since $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ (the tower property of conditional expectation). Therefore ξ_n is a martingale with respect to $\mathcal{F}_n$.

Exercise 3.3

Show that if ξ_n is a martingale with respect to $\mathcal{F}_n$, then

$$E(\xi_1) = E(\xi_2) = \cdots .$$

Hint What is the expectation of $E(\xi_{n+1}|\mathcal{F}_n)$?

Exercise 3.4

Suppose that ξ_n is a martingale with respect to a filtration $\mathcal{F}_n$. Show that ξ_n is a martingale with respect to the filtration

$$\mathcal{G}_n = \sigma(\xi_1, \dots, \xi_n).$$

Hint Observe that $\mathcal{G}_n \subset \mathcal{F}_n$ and use the tower property of conditional expectation.

Exercise 3.5

Let ξ_n be a symmetric random walk, that is,

$$\xi_n = \eta_1 + \cdots + \eta_n,$$

where $\eta_1, \eta_2, \dots$ is a sequence of independent identically distributed random variables such that

$$P\{\eta_n = 1\} = P\{\eta_n = -1\} = \frac{1}{2}$$

(a sequence of coin tosses, for example). Show that $\xi_n^2 - n$ is a martingale with respect to the filtration

$$\mathcal{F}_n = \sigma(\eta_1, \dots, \eta_n).$$

Hint You want to transform $E\left(\xi_{n+1}^2 - (n+1)\,|\mathcal{F}_n\right)$ to obtain $\xi_n^2 - n$. Write

$$\begin{aligned} \xi_{n+1}^2 &= (\xi_n + \eta_{n+1})^2 \\ &= \eta_{n+1}^2 + 2\eta_{n+1}\xi_n + \xi_n^2 \end{aligned}$$

and observe that ξ_n is $\mathcal{F}_n$-measurable, while η_{n+1} is independent of $\mathcal{F}_n$. To transform the conditional expectation you can 'take out what is known' and use the fact that 'an independent condition drops out'. Do not forget to verify that $\xi_n^2 - n$ is integrable and adapted to $\mathcal{F}_n$.

Exercise 3.6

Let ξ_n be a symmetric random walk and $\mathcal{F}_n$ the filtration defined in Exercise 3.5. Show that

$$\zeta_n = (-1)^n \cos(\pi\xi_n)$$

is a martingale with respect to $\mathcal{F}_n$.

Hint You want to transform $E((-1)^{n+1}\cos(\pi\xi_{n+1})|\mathcal{F}_n)$ to obtain $(-1)^n\cos(\pi\xi_n)$. Use a similar argument as in Exercise 3.5 to achieve this. But, first of all, make sure that ζ_n is integrable and adapted to $\mathcal{F}_n$.

Definition 3.5

We say that $\xi_1, \xi_2, \dots$ is a ***supermartingale*** (***submartingale***) with respect to a filtration $\mathcal{F}_1, \mathcal{F}_2, \dots$ if

1) ξ_n is integrable for each $n = 1, 2, \dots$;

2) $\xi_1, \xi_2, \dots$ is adapted to $\mathcal{F}_1, \mathcal{F}_2, \dots$;

3) $E(\xi_{n+1}|\mathcal{F}_n) \le \xi_n$ (respectively, $E(\xi_{n+1}|\mathcal{F}_n) \ge \xi_n$) a.s. for each $n = 1, 2, \dots$.

Exercise 3.7

Let ξ_n be a sequence of square integrable random variables. Show that if ξ_n is a martingale with respect to a filtration $\mathcal{F}_n$, then ξ_n^2 is a submartingale with respect to the same filtration.

Hint Use Jensen's inequality with convex function $\varphi(x) = x^2$.

3.4 Games of Chance

Suppose that you take part in a game such as the roulette, for example. Let $\eta_1, \eta_2, \dots$ be a sequence of integrable random variables, where η_n are your *winnings (or losses) per unit stake* in game n. If your stake in each game is one, then your total winnings after n games will be

$$\xi_n = \eta_1 + \cdots + \eta_n.$$

We take the filtration

$$\mathcal{F}_n = \sigma(\eta_1, \ldots, \eta_n)$$

and also put $\xi_0 = 0$ and $\mathcal{F}_0 = \{\emptyset, \Omega\}$ for notational simplicity.

If $n-1$ rounds of the game have been played so far, your accumulated knowledge will be represented by the σ-field $\mathcal{F}_{n-1}$. The game is fair if

$$E(\xi_n|\mathcal{F}_{n-1}) = \xi_{n-1},$$

that is, you expect that your fortune at step n will on average be the same as at step $n-1$. The game will be favourable to you if

$$E(\xi_n|\mathcal{F}_{n-1}) \geq \xi_{n-1},$$

and unfavourable to you if

$$E(\xi_n|\mathcal{F}_{n-1}) \leq \xi_{n-1}$$

for $n = 1, 2, \ldots$. This corresponds to ξ_n being, respectively, a martingale, a submartingale, or a supermartingale with respect to $\mathcal{F}_n$, see Definitions 3.4 and 3.5.

Suppose that you can vary the stake to be α_n in game n. (In particular, α_n may be zero if you refrain from playing the nth game; it may even be negative if you own the casino and can accept other people's bets.) When the time comes to decide your stake α_n, you will know the outcomes of the first $n-1$ games. Therefore it is reasonable to assume that α_n is $\mathcal{F}_{n-1}$-measurable, where $\mathcal{F}_{n-1}$ represents your knowledge accumulated up to and including game $n-1$. In particular, since nothing is known before the first game, we take $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

Definition 3.6

A *gambling strategy* $\alpha_1, \alpha_2, \ldots$ (with respect to a filtration $\mathcal{F}_1, \mathcal{F}_2, \ldots$) is a sequence of random variables such that α_n is $\mathcal{F}_{n-1}$-measurable for each $n = 1, 2, \ldots$, where $\mathcal{F}_0 = \{\emptyset, \Omega\}$. (Outside the context of gambling such a sequence of random variables α_n is called *previsible*.)

If you follow a strategy $\alpha_1, \alpha_2, \ldots$, then your *total winnings* after n games will be

$$\begin{aligned} \zeta_n &= \alpha_1\eta_1 + \cdots + \alpha_n\eta_n \\ &= \alpha_1(\xi_1 - \xi_0) + \cdots + \alpha_n(\xi_n - \xi_{n-1}). \end{aligned}$$

We also put $\zeta_0 = 0$ for convenience.

The following proposition has important consequences for gamblers. It means that a fair game will always turn into a fair one, no matter which gambling strategy is used. If one is not in a position to wager negative sums of money (e.g. to run a casino), it will be impossible to turn an unfavourable game into a favourable one or vice versa. You cannot beat the system! The boundedness of the sequence α_n means that your available capital is bounded and so is your credit limit.

Proposition 3.1

Let $\alpha_1, \alpha_2, \ldots$ be a gambling strategy.

1) If $\alpha_1, \alpha_2, \ldots$ is a bounded sequence and $\xi_0, \xi_1, \xi_2, \ldots$ is a martingale, then $\zeta_0, \zeta_1, \zeta_2, \ldots$ is a martingale (a fair game turns into a fair one no matter what you do);
2) If $\alpha_1, \alpha_2, \ldots$ is a non-negative bounded sequence and $\xi_0, \xi_1, \xi_2, \ldots$ is a supermartingale, then $\zeta_0, \zeta_1, \zeta_2, \ldots$ is a supermartingale (an unfavourable game turns into an unfavourable one).
3) If $\alpha_1, \alpha_2, \ldots$ is a non-negative bounded sequence and $\xi_0, \xi_1, \xi_2, \ldots$ is a submartingale, then $\zeta_0, \zeta_1, \zeta_2, \ldots$ is a submartingale (a favourable game turns into a favourable one).

Proof

Because α_n and ζ_{n-1} are $\mathcal{F}_{n-1}$-measurable, we can take them out of the expectation conditioned on $\mathcal{F}_{n-1}$ ('taking out what is known', Proposition 2.4). Thus, we obtain

$$\begin{aligned} E(\zeta_n|\mathcal{F}_{n-1}) &= E(\zeta_{n-1} + \alpha_n(\xi_n - \xi_{n-1})|\mathcal{F}_{n-1}) \\ &= \zeta_{n-1} + \alpha_n(E(\xi_n|\mathcal{F}_{n-1}) - \xi_{n-1}). \end{aligned}$$

If ξ_n is a martingale, then

$$\alpha_n(E(\xi_n|\mathcal{F}_{n-1}) - \xi_{n-1}) = 0,$$

which proves assertion 1). If ξ_n is a supermartingale and $\alpha_n \geq 0$, then

$$\alpha_n(E(\xi_n|\mathcal{F}_{n-1}) - \xi_{n-1}) \leq 0,$$

proving assertion 2). Finally, assertion 3) follows because

$$\alpha_n(E(\xi_n|\mathcal{F}_{n-1}) - \xi_{n-1}) \geq 0$$

if ξ_n is a submartingale and $\alpha_n \geq 0$. □

3.5 Stopping Times

In roulette and many other games of chance one usually has the option to quit at any time. The number of rounds played before quitting the game will be denoted by τ. It can be fixed, say, to be $\tau = 10$ if one decides in advance to stop playing after 10 rounds, no matter what happens. But in general the decision whether to quit or not will be made after each round depending on the knowledge accumulated so far. Therefore τ is assumed to be a random variable with values in the set $\{1, 2, \ldots\} \cup \{\infty\}$. Infinity is included to cover the theoretical possibility (and a dream scenario of some casinos) that the game never stops. At each step n one should be able to decide whether to stop playing or not, i.e. whether or not $\tau = n$. Therefore the event that $\tau = n$ should be in the σ-field $\mathcal{F}_n$ representing our knowledge at time n. This gives rise to the following definition.

Definition 3.7

A random variable τ with values in the set $\{1, 2, \ldots\} \cup \{\infty\}$ is called a ***stopping time*** (with respect to a filtration $\mathcal{F}_n$) if for each $n = 1, 2, \ldots$

$$\{\tau = n\} \in \mathcal{F}_n.$$

Exercise 3.8

Show that the following conditions are equivalent:

1) $\{\tau \leq n\} \in \mathcal{F}_n$ for each $n = 1, 2, \ldots$;
2) $\{\tau = n\} \in \mathcal{F}_n$ for each $n = 1, 2, \ldots$.

Hint Can you express $\{\tau \leq n\}$ in terms of the events $\{\tau = k\}$, where $k = 1, \ldots, n$? Can you express $\{\tau = n\}$ in terms of the events $\{\tau \leq k\}$, where $k = 1, \ldots, n$?

Example 3.5 (First hitting time)

Suppose that a coin is tossed repeatedly and you win or lose £1, depending on which way it lands. Suppose that you start the game with, say, £5 in your pocket and decide to play until you have £10 or you lose everything. If ξ_n is the amount you have at step n, then the time when you stop the game is

$$\tau = \min\{n : \xi_n = 10 \text{ or } 0\},$$

and is called the ***first hitting time*** (of 10 or 0 by the random sequence ξ_n). It is a stopping time in the sense of Definition 3.7 with respect to the filtration $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$. This is because

$$\{\tau = n\} = \{0 < \xi_1 < 10\} \cap \cdots \cap \{0 < \xi_{n-1} < 10\} \cap \{\xi_n = 10 \text{ or } 0\}.$$

Now each of the sets on the right-hand side belongs to $\mathcal{F}_n$, so their intersection does too. This proves that

$$\{\tau = n\} \in \mathcal{F}_n$$

for each n, so τ is a stopping time.

Exercise 3.9

Let ξ_n be a sequence of random variables adapted to a filtration $\mathcal{F}_n$ and let $B \subset \mathbb{R}$ be a Borel set. Show that the ***time of first entry*** of ξ_n into B,

$$\tau = \min\{n : \xi_n \in B\}$$

is a stopping time.

Hint Example 3.5 covers the case when $B = (-\infty, 0] \cup [10, \infty)$. Extend the argument to an arbitrary Borel set B.

Let ξ_n be a sequence of random variables adapted to a filtration $\mathcal{F}_n$ and let τ be a stopping time (with respect to the same filtration). Suppose that ξ_n represents your winnings (or losses) after n rounds of a game. If you decide to quit after τ rounds, then your total winnings will be ξ_τ. In this case your winnings after n rounds will in fact be $\xi_{\tau \wedge n}$. Here $a \wedge b$ denotes the smaller of two numbers a and b,

$$a \wedge b = \min(a, b),$$

Definition 3.8

We call $\xi_{\tau \wedge n}$ the sequence ***stopped*** at τ. It is often denoted by ξ_n^τ. Thus, for each $\omega \in \Omega$

$$\xi_n^\tau(\omega) = \xi_{\tau(\omega) \wedge n}(\omega).$$

Exercise 3.10

Show that if ξ_n is a sequence of random variables adapted to a filtration $\mathcal{F}_n$, then so is the sequence $\xi_{\tau \wedge n}$.

Hint For any Borel set B express $\{\xi_{\tau \wedge n} \in B\}$ in terms of the events $\{\xi_k \in B\}$ and $\{\tau = k\}$, where $k = 1, \ldots, n$.

We already know that it is impossible to turn a fair game into an unfair one, an unfavourable game into a favourable one, or vice versa using a gambling strategy. The next proposition shows that this cannot be achieved using a stopping time either (essentially, because stopping is also a gambling strategy).

Proposition 3.2

Let τ be a stopping time.

1) If ξ_n is a martingale, then so is $\xi_{\tau\wedge n}$.

2) If ξ_n is a supermartingale, then so is $\xi_{\tau\wedge n}$.

3) If ξ_n is a submartingale, then so is $\xi_{\tau\wedge n}$.

Proof

This is in fact a consequence of Proposition 3.1. Given a stopping time τ, we put

$$\alpha_n = \begin{cases} 1 & \text{if } \tau \geq n, \\ 0 & \text{if } \tau < n. \end{cases}$$

We claim that α_n is a gambling strategy (that is, α_n is $\mathcal{F}_{n-1}$-measurable). This is because the inverse image $\{\alpha_n \in B\}$ of any Borel set $B \subset \mathbb{R}$ is equal to

$$\emptyset \in \mathcal{F}_{n-1}$$

if $0, 1 \notin B$, or to

$$\Omega \in \mathcal{F}_{n-1}$$

if $0, 1 \in B$, or to

$$\{\alpha_n = 1\} = \{\tau \geq n\} = \{\tau > n-1\} \in \mathcal{F}_{n-1}$$

if $1 \in B$ and $0 \notin B$, or to

$$\{\alpha_n = 0\} = \{\tau < n\} = \{\tau \leq n-1\} \in \mathcal{F}_{n-1}$$

if $1 \notin B$ and $0 \in B$. For this gambling strategy

$$\xi_{\tau\wedge n} = \alpha_1(\xi_1 - \xi_0) + \cdots + \alpha_n(\xi_n - \xi_{n-1}).$$

Therefore Proposition 3.1 implies assertions 1), 2) and 3) above. □

Example 3.6

(You could try to beat the system if you had unlimited capital and unlimited time.) The following gambling strategy is called 'the martingale'. (Do not confuse this with the general definition of a martingale earlier in this section.) Suppose a coin is flipped repeatedly. Let us denote the outcomes by $\eta_1, \eta_2, \ldots$, which can take values $+1$ (heads) or -1 (tails). You wager £1 on heads. If you win, you quit. If you lose, you double the stake and play again. If you win this time round, you quit. Otherwise you double the stake once more, and so on. Thus, your gambling strategy is

$$\alpha_n = \begin{cases} 2^{n-1} & \text{if } \eta_1 = \cdots = \eta_{n-1} = \text{tails}, \\ 0 & \text{otherwise.} \end{cases}$$

Let us put

$$\zeta_n = \eta_1 + 2\eta_2 + \cdots + 2^{n-1}\eta_n$$

and consider the stopping time

$$\tau = \min\{n : \eta_n = \text{heads}\}.$$

Then $\zeta_{\tau\wedge n}$ will be your winnings after n rounds. It is a martingale (check it!).

It can be shown that $P\{\tau < \infty\} = 1$ (heads will eventually appear in the sequence $\eta_1, \eta_2, \ldots$ with probability one). Therefore it makes sense to consider ζ_τ. This would be your total winnings if you were able to continue to play the game no matter how long it takes for the first heads to appear. It would require unlimited time and capital. If you could afford these, you would be bound to win eventually because $\zeta_\tau = 1$ identically, since

$$-1 - 2 - \cdots - 2^{n-1} + 2^n = 1$$

for any n.

Exercise 3.11

Show that if a gambler plays 'the martingale', his expected loss just before the ultimate win is infinite, that is,

$$E(\zeta_{\tau-1}) = -\infty.$$

Hint What is the probability that the game will terminate at step n, i.e. that $\tau = n$? If $\tau = n$, what is $\zeta_{\tau-1}$ equal to? This will give you all possible values of $\zeta_{\tau-1}$ and their probabilities. Now compute the expectation of $\zeta_{\tau-1}$.

3.6 Optional Stopping Theorem

If ξ_n is a martingale, then, in particular,

$$E(\xi_n) = E(\xi_1)$$

for each n. Example 3.6 shows that $E(\xi_\tau)$ is not necessarily equal to $E(\xi_1)$ for a stopping time τ. However, if the equality

$$E(\xi_\tau) = E(\xi_1)$$

does hold, it can be very useful. The Optional Stopping Theorem provides sufficient conditions for this to happen.

Theorem 3.1 (Optional Stopping Theorem)

Let ξ_n be a martingale and τ a stopping time with respect to a filtration $\mathcal{F}_n$ such that the following conditions hold:

1) $\tau < \infty$ a.s.,
2) ξ_τ is integrable,
3) $E(\xi_n 1_{\{\tau>n\}}) \to 0$ as $n \to \infty$.

Then

$$E(\xi_\tau) = E(\xi_1).$$

Proof

Because

$$\xi_\tau = \xi_{\tau\wedge n} + (\xi_\tau - \xi_n)\, 1_{\{\tau>n\}},$$

it follows that

$$E(\xi_\tau) = E(\xi_{\tau\wedge n}) + E(\xi_\tau 1_{\{\tau>n\}}) - E(\xi_n 1_{\{\tau>n\}}).$$

Since $\xi_{\tau\wedge n}$ is a martingale by Proposition 3.2, the first term on the right-hand side is equal to

$$E(\xi_{\tau\wedge n}) = E(\xi_1).$$

The last term tends to zero by assumption 3). The middle term

$$E(\xi_\tau 1_{\{\tau>n\}}) = \sum_{k=n+1}^{\infty} E(\xi_k 1_{\{\tau=k\}})$$

tends to zero as $n \to \infty$ because the series

$$E(\xi_\tau) = \sum_{k=1}^{\infty} E(\xi_k 1_{\{\tau=k\}})$$

is convergent by 2). It follows that $E(\xi_\tau) = E(\xi_1)$, as required. □

Example 3.7 (Expectation of the first hitting time for a random walk)

Let ξ_n be a symmetric random walk as in Exercise 3.5 and let K be a positive integer. We define the first hitting time (of $\pm K$ by ξ_n) to be

$$\tau = \min\{n : |\xi_n| = K\}.$$

By Exercise 3.9 τ is a stopping time. By Exercise 3.5 we know that $\xi_n^2 - n$ is a martingale. If the Optional Stopping Theorem can be applied, then

$$E\left(\xi_\tau^2 - \tau\right) = E\left(\xi_1^2 - 1\right) = 0.$$

This allows us to find the expectation

$$E(\tau) = E(\xi_\tau^2) = K^2,$$

since $|\xi_\tau| = K$.

Let us verify conditions 1)–3) of the Optional Stopping Theorem.

1) We shall show that $P\{\tau = \infty\} = 0$. To this end we shall estimate $P\{\tau > 2Kn\}$. We can think of $2Kn$ tosses of a coin as n sequences of $2K$ tosses. A necessary condition for $\tau > 2Kn$ is that no one of these n sequences contains heads only. Therefore

$$P\{\tau > 2Kn\} \le \left(1 - \frac{1}{2^{2K}}\right)^n \to 0$$

as $n \to \infty$. Because $\{\tau > 2Kn\}$ for $n = 1, 2, \ldots$ is a contracting sequence of sets (i.e. $\{\tau > 2Kn\} \supset \{\tau > 2K(n+1)\}$), it follows that

$$\begin{aligned} P\{\tau = \infty\} &= P\left(\bigcap_{n=1}^{\infty}\{\tau > 2Kn\}\right) \\ &= \lim_{n\to\infty} P\{\tau > 2Kn\} = 0, \end{aligned}$$

completing the argument.

2) We need to show that

$$E\left(\left|\xi_\tau^2 - \tau\right|\right) < \infty.$$

Indeed,

$$\begin{aligned}E(\tau) &= \sum_{n=1}^{\infty} nP\{\tau = n\} \\ &= \sum_{n=0}^{\infty}\sum_{k=1}^{2K}(2Kn+k)\,P\{\tau = 2Kn+k\} \\ &\le \sum_{n=0}^{\infty}\sum_{k=1}^{2K} 2K(n+1)\,P\{\tau > 2Kn\} \\ &\le 4K^2\sum_{n=0}^{\infty}(n+1)\left(1-\frac{1}{2^{2K}}\right)^n \\ &< \infty,\end{aligned}$$

since the series $\sum_{n=1}^{\infty}(n+1)\,q^n$ is convergent for any $q \in (-1,1)$. Here we have recycled the estimate for $P\{\tau > 2Kn\}$ used in 2). Moreover, $\xi_\tau^2 = K^2$, so

$$\begin{aligned}E\left(|\xi_\tau^2 - \tau|\right) &\le E\left(\xi_\tau^2\right) + E(\tau) \\ &= K^2 + E(\tau) \\ &< \infty.\end{aligned}$$

3) Since $\xi_n^2 \le K^2$ on $\{\tau > n\}$,

$$E\left(\xi_n^2 1_{\{\tau>n\}}\right) \le K^2 P\{\tau > n\} \to 0$$

as $n \to \infty$. Moreover,

$$E\left(n 1_{\{\tau>n\}}\right) \le E\left(\tau 1_{\{\tau>n\}}\right) \to 0$$

as $n \to \infty$. Convergence to 0 holds because $E(\tau) < \infty$ by 2) and $\{\tau > n\}$ is a contracting sequence of sets with intersection $\{\tau = \infty\}$ of measure zero. It follows that

$$E\left(\left(\xi_n^2 - n\right)1_{\{\tau>n\}}\right) \to 0,$$

as required.

Exercise 3.12

Let ξ_n be a symmetric random walk and $\mathcal{F}_n$ the filtration defined in Exercise 3.5. Denote by τ the smallest n such that $|\xi_n| = K$ as in Example 3.7. Verify that

$$\zeta_n = (-1)^n \cos\left[\pi\left(\xi_n + K\right)\right]$$

is a martingale (see Exercise 3.6). Then show that ζ_n and τ satisfy the conditions of the Optional Stopping Theorem and apply the theorem to find $E[(-1)^\tau]$.

Hint The equality $\zeta_\tau = (-1)^\tau$ is a key to computing $E[(-1)^\tau]$ with the aid of the Optimal Stopping Theorem. The first two conditions of this theorem are either obvious in the case in hand or have been verified elsewhere in this chapter. To make sure that condition 3) holds it may be helpful to show that

$$\left|E(\zeta_n 1_{\{\tau>n\}})\right| \le P\{\tau > n\}.$$

Use Jensen's inequality with convex function $\varphi(x) = |x|$ to estimate the left-hand side. Do not forget to verify that ζ_n is a martingale in the first place.

3.7 Solutions

Solution 3.1

A belongs to $\mathcal{F}_{11}$, but not to $\mathcal{F}_{10}$. The smallest n is 11.
B does not belong to $\mathcal{F}_n$ for any n. There is no smallest n such that $B \in \mathcal{F}_n$.
C belongs to $\mathcal{F}_{100}$, but not to $\mathcal{F}_{99}$. The smallest n is 100.
Since $D = \emptyset$, it belongs to $\mathcal{F}_n$ for each $n = 1, 2, \ldots$. Here the smallest n is 1.

Solution 3.2

Because the sequence of random variables $\xi_1, \xi_2, \ldots$ is adapted to the filtration $\mathcal{G}_1, \mathcal{G}_2, \ldots$, it follows that ξ_n is $\mathcal{G}_n$-measurable for each n. But

$$\mathcal{G}_1 \subset \mathcal{G}_2 \subset \cdots,$$

so $\xi_1, \ldots, \xi_n$ are $\mathcal{G}_n$-measurable for each n. As a consequence,

$$\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n) \subset \mathcal{G}_n$$

for each n.

Solution 3.3

Taking the expectation on both sides of the equality

$$\xi_n = E(\xi_{n+1}|\mathcal{F}_n),$$

we obtain

$$E(\xi_n) = E(E(\xi_{n+1}|\mathcal{F}_n)) = E(\xi_{n+1})$$

for each n. This proves the claim.

Solution 3.4

The random variables ξ_n are integrable because ξ_n is a martingale with respect to $\mathcal{F}_n$. Since $\mathcal{G}_n$ is the σ-field generated by $\xi_1, \dots, \xi_n$, it follows that ξ_n is adapted to $\mathcal{G}_n$. Finally, since $\mathcal{G}_n \subset \mathcal{F}_n$,

$$\begin{aligned} \xi_n &= E(\xi_n|\mathcal{G}_n) \\ &= E(E(\xi_{n+1}|\mathcal{F}_n)|\mathcal{G}_n) \\ &= E(\xi_{n+1}|\mathcal{G}_n) \end{aligned}$$

by the tower property of conditional expectation (Proposition 2.4). This proves that ξ_n is a martingale with respect to $\mathcal{G}_n$.

Solution 3.5

Because

$$\xi_n^2 - n = (\eta_1 + \cdots + \eta_n)^2 - n$$

is a function of $\eta_1, \dots, \eta_n$, it is measurable with respect to the σ-field $\mathcal{F}_n$ generated by $\eta_1, \dots, \eta_n$, i.e. $\xi_n^2 - n$ is adapted to $\mathcal{F}_n$. Since

$$|\xi_n| = |\eta_1 + \cdots + \eta_n| \le |\eta_1| + \cdots |\eta_n| = n,$$

it follows that

$$E(|\xi_n^2 - n|) \le E(\xi_n^2) + n \le n^2 + n < \infty,$$

so $\xi_n^2 - n$ is integrable for each n. Because

$$\xi_{n+1}^2 = \eta_{n+1}^2 + 2\eta_{n+1}\xi_n + \xi_n^2,$$

where ξ_n and ξ_n^2 are $\mathcal{F}_n$-measurable and η_{n+1} is independent of $\mathcal{F}_n$, we can use Proposition 2.4 ('taking out what is known' and 'independent condition drops out') to obtain

$$\begin{aligned} E(\xi_{n+1}^2|\mathcal{F}_n) &= E(\eta_{n+1}^2|\mathcal{F}_n) + 2E(\eta_{n+1}\xi_n|\mathcal{F}_n) + E(\xi_n^2|\mathcal{F}_n) \\ &= E(\eta_{n+1}^2) + 2\xi_n E(\eta_{n+1}) + \xi_n^2 \\ &= 1 + \xi_n^2. \end{aligned}$$

This implies that

$$E(\xi_{n+1}^2 - n - 1|\mathcal{F}_n) = \xi_n^2 - n,$$

so $\xi_n^2 - n$ is a martingale.

Solution 3.6

Being a function of ξ_n, the random variable ζ_n is $\mathcal{F}_n$-measurable for each n, since ξ_n is. Because $|\zeta_n| \le 1$, it is clear that ζ_n is integrable. Because η_{n+1} is independent of $\mathcal{F}_n$ and ξ_n is $\mathcal{F}_n$-measurable, it follows that

$$\begin{aligned} E(\zeta_{n+1}|\mathcal{F}_n) &= E\left((-1)^{n+1}\cos[\pi(\xi_n + \eta_{n+1})]|\mathcal{F}_n\right) \\ &= (-1)^{n+1}E\left(\cos(\pi\xi_n)\cos(\pi\eta_{n+1})|\mathcal{F}_n\right) \\ &\quad -(-1)^{n+1}E\left(\sin(\pi\xi_n)\sin(\pi\eta_{n+1})|\mathcal{F}_n\right) \\ &= (-1)^{n+1}\cos(\pi\xi_n)E\left(\cos(\pi\eta_{n+1})\right) \\ &\quad -(-1)^{n+1}\sin(\pi\xi_n)E\left(\sin(\pi\eta_{n+1})\right) \\ &= (-1)^n\cos(\pi\xi_n) \\ &= \zeta_n, \end{aligned}$$

using the formula

$$\cos(\alpha + \beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta.$$

To compute $E(\cos(\pi\eta_{n+1}))$ and $E(\sin(\pi\eta_{n+1}))$ observe that $\eta_{n+1} = 1$ or -1 and

$$\begin{aligned} \cos\pi &= \cos(-\pi) = -1, \\ \sin\pi &= \sin(-\pi) = 0. \end{aligned}$$

It follows that ζ_n is a martingale with respect to the filtration $\mathcal{F}_n$.

Solution 3.7

If ξ_n is adapted to $\mathcal{F}_n$, then so is ξ_n^2. Since $\xi_n = E(\xi_{n+1}|\mathcal{F}_n)$ for each n and $\varphi(x) = x^2$ is a convex function, we can apply Jensen's inequality (Theorem 2.2) to obtain

$$\xi_n^2 = [E(\xi_{n+1}|\mathcal{F}_n)]^2 \le E(\xi_{n+1}^2|\mathcal{F}_n)$$

for each n. This means that ξ_n^2 is a submartingale with respect to $\mathcal{F}_n$.

Solution 3.8

1)⇒2). If τ has property 1), then

$$\{\tau \le n\} \in \mathcal{F}_n$$

and

$$\{\tau \le n-1\} \in \mathcal{F}_{n-1} \subset \mathcal{F}_n,$$

so

$$\{\tau = n\} = \{\tau \le n\} \setminus \{\tau \le n-1\} \in \mathcal{F}_n.$$

2)⇒1). If τ has property 2), then

$$\{\tau = k\} \in \mathcal{F}_k \subset \mathcal{F}_n$$

for each $k = 1, \ldots, n$. Therefore

$$\{\tau \le n\} = \{\tau = 1\} \cup \cdots \cup \{\tau = n\} \in \mathcal{F}_n.$$

Solution 3.9

If

$$\tau = \min\{n : \xi_n \in B\},$$

then for any n

$$\{\tau = n\} = \{\xi_1 \notin B\} \cap \cdots \cap \{\xi_{n-1} \notin B\} \cap \{\xi_n \in B\}.$$

Because B is a Borel set, each of the sets on the right-hand side belongs to the σ-field $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$, and their intersection does too. This proves that $\{\tau = n\} \in \mathcal{F}_n$ for each n, so τ is a stopping time.

Solution 3.10

Let $B \subset \mathbb{R}$ be a Borel set. We can write

$$\{\xi_{\tau\wedge n} \in B\} = \{\xi_n \in B, \tau > n\} \cup \bigcup_{k=1}^{n} \{\xi_k \in B, \tau = k\},$$

where

$$\{\xi_n \in B, \tau > n\} = \{\xi_n \in B\} \cap \{\tau > n\} \in \mathcal{F}_n$$

and for each $k = 1, \ldots, n$

$$\{\xi_k \in B, \tau = k\} = \{\xi_k \in B\} \cap \{\tau = k\} \in \mathcal{F}_k \subset \mathcal{F}_n.$$

It follows that for each n

$$\{\xi_{\tau\wedge n} \in B\} \in \mathcal{F}_n,$$

as required.

Solution 3.11

The probability that 'the martingale' terminates at step n is

$$P\{\tau = n\} = \frac{1}{2^n}$$

($n-1$ tails followed by heads at step n). Therefore

$$\begin{aligned} E(\zeta_{\tau-1}) &= \sum_{n=1}^{\infty} \zeta_{n-1} P\{\tau = n\} \\ &= \sum_{n=1}^{\infty} (-1 - 2 - \cdots - 2^{n-2}) \frac{1}{2^n} \\ &= -\sum_{n=1}^{\infty} \frac{2^{n-1} - 1}{2^n} = -\infty. \end{aligned}$$

Solution 3.12

The proof that ζ_n is a martingale is almost the same as in Exercise 3.6. We need to verify that ζ_n and τ satisfy conditions 1)–3) of the Optional Stopping Theorem.

Condition 1) has in fact been verified in Example 3.7.

Condition 2) holds because $|\zeta_\tau| \le 1$, so $E(|\zeta_\tau|) \le 1 < \infty$.

To verify condition 3) observe that $|\zeta_n| \le 1$ for all n, so

$$\begin{aligned} \left|E(\zeta_n 1_{\{\tau > n\}})\right| &\le E(|\zeta_n| 1_{\{\tau > n\}}) \\ &\le E(1_{\{\tau > n\}}) \\ &= P\{\tau > n\}. \end{aligned}$$

The family of events $\{\tau > n\}$, $n = 1, 2, \ldots$ is a contracting one with intersection $\{\tau = \infty\}$. It follows that

$$\left|E(\zeta_n 1_{\{\tau > n\}})\right| \le P\{\tau > n\} \searrow P\{\tau = \infty\}$$

as $n \to \infty$. But

$$P\{\tau = \infty\} = 0$$

by 1), completing the proof.

The Optional Stopping Theorem implies that

$$E(\zeta_\tau) = E(\zeta_1)$$

Because $\xi_\tau = K$ or $-K$, we have

$$\zeta_\tau = (-1)^\tau \cos[\pi(K + \xi_\tau)] = (-1)^\tau.$$

Let us compute

$$\begin{aligned} E(\zeta_1) &= -\frac{1}{2}(\cos[\pi(1 + K)] + \cos[\pi(-1 + K)]) \\ &= \cos(\pi K) = (-1)^K. \end{aligned}$$

It follows that

$$E[(-1)^\tau] = (-1)^K.$$