A Primer on Probability and Stochastic Processes

Martin V. Day[1]

June 30, 2007

[1]day@math.vt.edu

Contents

Preface

The careful study of stochastic processes, including Brownian motion and stochastic differential equations, is usually a subject one only approaches after acquiring a solid foundation in measure theory and functional analysis. However, because it has become an important tool in theoretical finance, there are many students who need an understanding of this material, but who don't have the background necessary for a traditional mathematically rigorous development. Numerous introductory texts on mathematical finance try to provide some loose intuitive explanation of these ideas. Sometimes, however, those explanations are so vague as to be of little use, and of no help to those students who want to pursue a more careful study or application in the future.

This primer has grown out of my efforts to provide a more accurate introduction for such students. It is intended to provide a quick introduction to some basic ideas from probability theory and stochastic processes, including martingales and stochastic integrals. I believe that it is possible to communicate the main ideas relatively accurately, if one is willing to suppress the mathematician's impulse to leave no detail overlooked and no assertion unproven. For instance, it is possible to understand what probability measures and σ-algebras are, without stating or proving the Caratheodory extension and monotone class theorems. We simply need to ask students to accept some assertions without proof and let some technical details go unresolved, just as we do in a freshman's first exposure to calculus. I do cut plenty of corners, such as ignoring exceptions on sets of measure zero, and leaving things like progressive measurability unexplained. I will prove almost nothing rigorously. Thus what we offer here is an overview or user's guide. It is not meant to be quoted as a reference[1]. Even so, we will try to be as accurate as possible, within the scope of our discussion. The student who does go on to study these topics more carefully will find, I hope, that the presentation here is mathematically justifiable after all.

Chapters 1—4 concern the ideas needed for discrete time models (e.g. binomial trees). Chapters 5—7 will discuss Brownian motion and stochastic calculus for continuous time processes. A few problems are included at the end of each chapter. They are labeled with a "P" (to distinguish them from problems in other course materials.) For instance problem P.6.C is problem C from Chapter 6 of this primer. When we refer to "the text" we mean the book [Bö] by Björk which is the main text for our course, Math 5725.

Martin Day; Blacksburg, Virginia, June 2004

[1] If in some future work of your own you need to cite a source for any of the material described here, use some of the references listed at the end of this primer, or some of the many other fine texts available.

Chapter 1

The Basics

1.1 Mathematical Representation of Randomness

We are used to thinking about *deterministic* mathematical quantities. By this we mean an object x (representing a real number, or a matrix, or a function, or ...) which has a fixed or definite value. Even if x is the solution of some equation which we have not yet solved, we still view it as having a value which is prescribed by the equation, although we may not have discovered what the actual value is yet. But when we talk about a quantity X as being *random* the fundamental idea is that its value is not fixed or determined at all, but could take one of many possible values: $X = x_1$, or $X = x_2$, or .... Eventually one of the possible values of X will occur as the *realization* of X, but we view the mechanism that determines the realized value to be unpredictable. Until the realization actually occurs we can only ascribe probabilities to the various possibilities. Examples might be the flip of a coin, next year's total rainfall, or tomorrow's closing price for a share of IBM. All of these are things that we would call random, whose status changes from unknowable to known with the passage of time. Although there may be trends and past information which strongly influence the probabilities that we ascribe to the various possible realizations, we can not say with certainty what the actual value of X will be until the event in question actually happens. Such a random quantity X is typically called a *random variable*. The theory of probability provides concrete mathematical structures that allow our ideas about random quantities and statements about probabilities to be given precise meanings, so that theoretical statements can be proven with mathematical rigor.

We need to deal with situations in which there are many different random variables which may be related to each other in complicated ways. Our mathematical formulation begins with a "master set" Ω whose elements ω account for all conceivable outcomes, one (or more) for each possibility. Each $\omega \in \Omega$ is what we might call a possible *state of the world* – each ω determines specific values for all random variables under consideration. Thus random variables are actually functions defined on Ω. For instance $X : \Omega \to \mathbb{R}$ is the function which specifies the specific value $X(\omega)$ which is realized by X for each each possible state of the world ω. The other basic ingredient is an assignment of probabilities to subsets $A \subseteq \Omega$, usually indicated $P(A)$.

For a simple illustration suppose we want to describe a situation in which all the randomness boils down to the outcome of flipping two coins, C_1 and C_2. We could take $\Omega = \{a, b, c, d\}$ where

a corresponds to C_1=H and C_2=H,

b corresponds to C_1=H and C_2=T,

c corresponds to C_1=T and C_2=H, and

d corresponds to C_1=T and C_2=T.

Then Ω accounts for all the possible states of the world. We would let C_1 be the function described by $C_1(a) = C_1(b) =$ H and, $C_1(c) = C_1(d) =$ T. The definition of C_2 in a similar manner is should be clear. Thus we have two random variables $C_i : \Omega \to \{H, T\}$. If the coins are assumed "fair" (i.e. heads and tails

are equally likely, and neither coin's outcome influences the other's) then we would assign probabilities by $P(A) = \frac{1}{4} \times$(the number of elements in A).

This is all very simple when the number of different possibilities that we need to account for is finite (or countable). But as soon as we allow random variables for which any real number is a possible value (normally distributed random variables, for instance) then Ω will be at least as large as the real numbers $\mathbb{R}$, and often much larger. This raises difficult issues for our intent to assign probabilities $P(A)$ to subsets $A \subseteq \Omega$. For most situations it turns out to be *impossible* to define $P(A)$ for *all* $A \subseteq \Omega$ without violating some of the basic rules that probabilities should satisfy (see below). The resolution of this is that we can only define $P(A)$ for certain subsets A. The collection of those $A \subseteq \Omega$ for which $P(A)$ is defined will be what we call a *σ-algebra*. We will usually denote a σ-algebra with a script letter, like $\mathcal{F}$. Together $\mathcal{F}$ and P must satisfy a number of properties, listed below. The point at the moment is that unless $A \in \mathcal{F}$ we view $P(A)$ as undefined.

1.2 Properties of Probability Measures, σ-Algebras, and Random Variables

We list here the essential properties that a *probability space* $(\Omega, \mathcal{F}, P)$ must have. These are the combined definitions of a *probability measure* P and a *σ-algebra* $\mathcal{F}$ of subsets of Ω.

- $\Omega \in \mathcal{F}$, and $P(\Omega) = 1$.
- $\emptyset \in \mathcal{F}$, and $P(\emptyset) = 0$.
- For any $A \in \mathcal{F}$, $P(A)$ is a real number between 0 and 1.
- If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$ also, and $P(A^c) = 1 - P(A)$.
- If $A_1, A_2, \ldots$ is a sequence of sets in $\mathcal{F}$ then $\cup_1^\infty A_i \in \mathcal{F}$. If in addition the A_i are disjoint (i.e. $A_i \cap A_j = \emptyset$ when $i \neq j$), then $P(\cup_1^\infty A_i) = \sum_1^\infty P(A_i)$.

Regarding a random variable $X : \Omega \to \mathbb{R}$, we require the following property, which we describe by saying that X is *$\mathcal{F}$-measurable*:

- For every real number c, the set $\{\omega \in \Omega : \ X(\omega) \leq c\}$ is in $\mathcal{F}$.

There is a significant body of theory about the existence of σ-algebras with probability measures defined on them. This is the subject of *measure theory*[1]. We will not concern ourselves with such technical issues – we will just take for granted the existence of the $(\Omega, \mathcal{F}, P)$ that we need. However we do need to have some understanding of σ-algebras for a different reason. They are essential in understanding how we work with conditional probabilities and conditional expectations, i.e. situations in which we have partial but incomplete information about the state of the world. We will say more about this in §2.1 below.

We may choose $\mathcal{F}$ in various ways, depending on the setting. In our simple example two fair coins above, $\mathcal{F}$ can simply consist of the collection of all subsets of $\Omega = \{a, b, c, d\}$. If $\Omega = \mathbb{R}$ the standard choice of $\mathcal{F}$ is the collection $\mathcal{B}(\mathbb{R})$ of the *Borel sets*. This is the smallest σ-algebra of subsets of $\mathbb{R}$ which includes the half-open intervals $(-\infty, c]$. Every open set is in $\mathcal{B}(\mathbb{R})$ as well as every closed set. If fact every set you can explicitly write down, is a Borel set. However one can prove that there exist subsets of $\mathbb{R}$ which are not Borel sets.

The measurability property of a random variable X says that

$$\{\omega : \ X(\omega) \in A\} \in \mathcal{F},$$

whenever $A = (-\infty, c]$. It can be proven that this automatically extends to any set $A \subseteq \mathbb{R}$ which is a Borel set. Thus it remains true for any open or closed set A, not just intervals as explicitly required above.

[1] You can learn about measure theory in Math 5225 or Math/Stat 6105

1.3 A More Complicated Example

Here is an example with more complexity[2]. Take $\Omega = [0, 1)$, the unit interval on the real line. Each $\omega \in \Omega$ has a decimal representation:

$$\omega = .d_1 d_2 d_3 \dots d_n \dots \tag{1.1}$$

(Some numbers have more than one representation, for instance $.5000\dots = .49999\dots$ In such cases we agree always to use the representation with repeating 0 s, not repeating 9 s.) This defines a whole sequence of random variables $X_n : \Omega \to \{0, 1, \dots, 9\}$ given by $X_n(\omega) = d_n$ in the decimal representation (1.1) of ω. For a probability measure P on Ω we will use the usual length of an interval:

$$P([a, b)) = b - a, \text{ assuming } [a, b) \subset [0, 1). \tag{1.2}$$

Now we face one of the technical problems of measure theory. We have only said what $P(A)$ is if A is an interval. But what if A is not an interval? We have to identify a collection $\mathcal{F}$ that has all the properties listed above and includes all intervals, and we need to extend the definition of $P(A)$ to all such $A \in \mathcal{F}$ in a way that satisfies all the properties listed above and maintains (1.2) for the intervals themselves. The resulting P is called *Lebesgue measure* on $[0, 1)$ and the sets in $\mathcal{F}$ are the Borel sets[3] in $[0, 1)$, mentioned above. We will simply be satisfied with the assurance that that all these technical details can be worked out.

Now lets think about what we have created. Consider one of the random variables, say X_2. This is a random variable whose value can be any of the single digits $0, 1, \dots 9$. To calculate the probability that it takes a specific value, say $P(X_2 = 3)$, we first find the associated set of ω:

$$A = \{\omega : \ X_2(\omega) = 3\} = [.03, .04) \cup [.13, .14) \cup \dots [.93, .94)$$

and then compute $P(A)$:

$$\begin{aligned} P(X_2 = 3) &= P([.03, .04) \cup [.13, .14) \cup \dots [.93, .94)) \\ &= .01 + .01 + \dots + .01 = .1 \end{aligned}$$

Although the sets are more complicated in general, you should be able to convince yourself that $P(X_n = j) = .1$ for all $n \geq 1$ and every $j = 0, \dots, 9$. Thus each X_n has the probabilistic structure of a fair "10-sided dice". But the story doesn't end here. We can specify the values of any number of the X_i simultaneously and ask for the probability. For instance if each of $j_1, \dots j_m$ is a single digit, then we would calculate

$$P(X_1 = j_1, \, X_2 = j_2, \dots, X_m = j_m)$$

by identifying the set B of ω so that $X_1(\omega) = j_1$ and $X_2(\omega) = j_2 \dots$ and $X_m(\omega) = j_m$, and then calculating $P(B)$. After a little thought you should see that

$$B = [b, b + 10^{-m}),$$

where b is the number $.j_1 j_2 \dots j_m$. Thus $P(B) = 10^{-m} = (.1)^m$, which is what we expect if the $X_1, \dots, X_m$ are "independent", i.e. don't influence each other's outcomes.

Next suppose we want to know the probability that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i = 4.5,$$

a possibility that depends on the full sequence of digits in (1.1), not just the first few. This defines a set

$$L = \{\omega \in \Omega : \ \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i(\omega) = 4.5\}.$$

[2]This example is adapted from [Bi], where it was based on dyadic expansions and coin tossing. We have converted it to decimal expansions, since they are more familiar.

[3]In treatments of real analysis, such as Math 5225, a larger collection $\mathcal{F}$ called the *Lebesgue sets* is typically used.

Although L is not an interval or even a union of intervals, it does turn out that $L \in \mathcal{F}$ (i.e. is a Borel set) and that $P(L) = 1$, which is what we might have expected, since 4.5 is the average of the possible values $0, \ldots, 9$. (That $P(L) = 1$ is consequence of the Strong Law of Large Numbers, stated in Chapter 4 below.)

There are many different ways to assign probabilities to the sets $A \subseteq [0, 1)$ which are in $A \in \mathcal{F}$. Another would be to take

$$\tilde{P}([a, b)) = \int_a^b f(x)\, dx,$$

where $f(x)$ can be any (Riemann integrable) function $f(x) \geq 0$ on $[0, 1)$ with $\int_0^1 f(x)\, dx = 1$. For instance take $f(x) = 2x$. This leads to a different probability measure $\tilde{P}$ on the same Ω and $\mathcal{F}$. With respect to this measure the X_i will no longer behave like independent fair dice. For instance,

$$\tilde{P}(X_1 = 0) = .01, \quad \tilde{P}(X_2 = 0) = .091, \quad \text{and } \tilde{P}(X_1 = 0 \text{ and } X_2 = 0) = .0001. \tag{1.3}$$

(You can check my calculations; see Problem P.1.A.) We have not changed what the random variables X_i, as functions of $\omega \in \Omega$. We have only changed the assignment of probabilities to sets of ω. But this does change how we would describe the X_i.

1.4 Standard Constructions and Quantities

If Ω is something simple and explicit, like our examples so far, we can often define $P(A)$ directly. For instance if $\Omega = \{1, 2, 3 \ldots\}$, the positive integers, we can define a P using any infinite series $\sum_1^\infty p_i = 1$ with $p_i \geq 0$:

$$P(A) = \sum_{i \in A} p_i.$$

Or, if Ω is a set on which we know how to integrate (e.g. $\Omega \subseteq \mathbb{R}$ or $\Omega \subseteq \mathbb{R}^n$ for instance) and $f(\omega) \geq 0$ is an integrable function with $\int_\Omega f(\omega)\, d\omega = 1$, then we can start with a definition like that of $\tilde{P}$ at the end of the last section, using f as a *probability density function*:

$$P(A) = \int_A f(\omega)\, d\omega, \tag{1.4}$$

and appeal to measure theory to help us get it extended to some suitably large σ-algebra $\mathcal{F}$ of subsets. You should observe that in this case knowing the value of $P(\{\omega\})$ for each singleton set $A = \{\omega\}$ is *not* enough to determine $P(A)$ for infinite sets A. Regardless of the choice of density f in (1.4) we will find $P(\{\omega\}) = 0$. But for larger sets (such as intervals) $P(A)$ genuinely depends on the choice of f. This is one reason why in general we must assign probabilities to sets $A \subseteq \Omega$, not simply to individual $\omega \in \Omega$.

Whatever $(\Omega, \mathcal{F}, P)$ is, there are some standard probabilistic constructions and definitions that we need to understand. The first is the idea of the *distribution* μ_X of a random variable X. This refers to the assignment of probabilities to sets of possible values of X (as opposed to subsets of Ω): for a (Borel) subset $C \subseteq \mathbb{R}$,

$$\mu_X(C) = P(X \in C).$$

We would calculate this from P by $\mu_X(C) = P(A)$ where $A = \{\omega \in \Omega : \ X(\omega) \in C\}$. Figure 1.1 is an effort to illustrate this.

For instance in our example with a sequence of 10-sided dice, each of the random variables $X_1, X_2, \ldots$ has what we would describe as the *uniform distribution* on $\{0, 1, 2, \ldots 9\}$. Also, the X_i all have the same distribution, even though they are quite different as functions $X_i : \Omega \to \{0, 1, \ldots 9\}$. This makes an important point: the distribution of a random variable tells you what the probabilities of the outcomes of that single random variable X are, but it does *not* tell you what the underlying Ω is, how $X : \Omega \to \mathbb{R}$ is actually defined, or how its probabilities are related to those of other random variables.

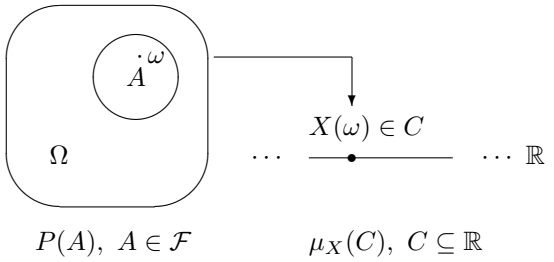


Figure 1.1: The distribution of a random variable

If we say that Y is a *standard normal* random variable[4], that means that its distribution is described by

$$\mu_Y([a,b]) = P(a \le Y \le b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx. \tag{1.5}$$

The function $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ appearing above is the standard normal density function[5]. By knowing it (even though we might not know Ω, $\mathcal{F}$, or P) we can calculate probabilities of things which depend only on the value of Y.

Also associated with a random variable X are several numbers that describe its distribution: its *mean*, *expectation*, or *expected value*

$$m = E[X];$$

its *variance*

$$\sigma^2 = E[(X-m)^2];$$

its *standard deviation*

$$\sigma = \sqrt{E[(X-m)^2]};$$

its *moments*

$$m_k = E[X^k], \quad k = 1, 2, \ldots;$$

its *characteristic function*

$$\hat{\mu}_X(s) = E[e^{isX}], \quad s \in \mathbb{R};$$

its *distribution function*

$$F_X(x) = P(X \le x);$$

and others which we need not mention here. Notice that all of these are expected values of some function of X: $E[g(X)]$ for some $g(x)$. So we will discuss expectations in general.

If we have an explicit description of the distribution of a random variable, then there are explicit formulas for the expectation $E[g(X)]$. For our examples above,

$$E[g(X_2)] = \sum_{j=0}^{9} g(j)/10 \quad \text{in the setting of (1.1) and (1.2),}$$

$$E[g(Y)] = \int_{-\infty}^{\infty} g(y)\phi(y)\, dy \quad \text{in the setting of (1.5).}$$

(There is nothing special about the standard normal density ϕ here. It can be replaced by whatever the appropriate density for Y might be.) However there is a generalized definition of the expected value that does not depend on the specific form of the distribution. This is important theoretically because there do

[4]A common notation for this is to say that Y has an $N[0,1]$ distribution. More generally Y has an $N[\mu,\sigma]$ distribution if the density function in (1.5) is replaced by $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$.

[5]Note that ϕ is a density for the distribution of Y, not for P itself as in (1.4).

exist random variables for which neither summations or Riemann integrals are adequate to describe their expectations. The general definition of the expectation of a random variable (or a function of) is the same as the definition of the measure-theoretic integral

$$E[g(X)] = \int_\Omega g(X(\omega))\, dP(\omega)$$

that you would learn about in Math 5225, or 6105. We can pursue that no further here. For our purposes it is enough to know that there is a generalized notion of integral (over Ω with respect to P) that defines $E[X] = \int_\Omega X(\omega)\, dP$, and obeys all the usual properties of integrals that you are used to. If $X, Y : \Omega \to \mathbb{R}$, are both random variables (and therefore $\mathcal{F}$-measurable), then

- $E[aX + bY] = aE[X] + bE[Y]$ for any constants a, b.
- If $A \in \mathcal{F}$ and $1_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$, then $E[1_A] = P(A)$.
- $X(\omega) \le Y(\omega)$ for all ω implies $E[X] \le E[Y]$.
- $|E[X]| \le E[|X|]$.

There is one other qualification to the definition of $E[X]$ which concerns the possibility that an ambiguous $\infty - \infty$ somehow arises in the calculation. To avoid this we will insist that X be *integrable*, which means that

$$E[|X|] < \infty.$$

(It turns out that if $X \ge 0$ then $E[X]$ can always be defined, possibly with value $+\infty$, so $E[|X|]$ always makes sense.) We should emphasize that $E[X]$ depends on both the random variable X *and* the probability measure P. Changing either will change the value of $E[X]$, or even make it undefined!

Just as we often want to calculate integrals over specific sets, e.g. $\int_a^b g(t)\, dt$ or $\int_{[a,b]} g(t)\, dt$, we sometimes want to do the same thing with an expectation. The usual notation is $E[X; A]$, where $A \subseteq \Omega$ (and $A \in \mathcal{F}$). There are a couple equivalent ways to express the same quantity:

$$\begin{aligned} E[X; A] &= E[X \cdot 1_A] \quad \text{(see above for the notation } 1_A) \\ &= \int X(\omega) 1_A(\omega)\, dP(\omega) \\ &= \int_A X(\omega)\, dP(\omega) \end{aligned} \tag{1.6}$$

Notice that

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

is simply a new random variable that agrees with X on A but has been set to 0 outside A. We should add to the list properties of expectations the usual "additivity" with respect to A in $E[X; A]$:

- $E[X; A \cup B] = E[X; A] + E[X; B]$, provided $A, B \in \mathcal{F}$ and $A \cap B = \emptyset$.

1.5 Problems

P.1.A Check the calculations in (1.3).

P.1.B A normal random variable X with mean μ and standard deviation σ (i.e. with distribution $N[\mu, \sigma]$) is one whose distribution is described by the density function

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Verify the following calculations:

$$\begin{aligned}
&\int_{-\infty}^{\infty} xp(x)\,dx = \mu; \quad \text{mean of } X \\
&\int_{-\infty}^{\infty} x^2 p(x)\,dx = \sigma^2 + \mu^2; \quad \text{second moment} \\
&\int_{-\infty}^{\infty} e^{\theta x} p(x)\,dx = e^{\frac{1}{2}\sigma^2\theta^2 + \mu\theta}; \quad \text{moment generating function.}
\end{aligned} \tag{1.7}$$

You are to work to integrals out by hand. Do not appeal to tables, formulas that you may have learned elsewhere, or software. The point is for you to verify that these formulas for the normal density are indeed correct. You *may* take for granted that $\int_{-\infty}^{\infty} p(x)\,dx = 1$ for all $\sigma > 0$ and all μ. The *standard normal distribution function* (associated with $\mu = 0$ and $\sigma = 1$) is

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx \tag{1.8}$$

Express the distribution function for X (arbitrary μ and σ) in terms of Φ.

P.1.C Suppose X has distribution $N[\mu, \sigma]$ as in P.1.B. Consider the random variable

$$S = S_0 e^X,$$

where S_0 is a positive constant. (This is often called a log-normal random variable. The Black-Scholes model of stock prices that we will consider in Chapter 3 will produce stock price random variables with this kind of distribution.)

1. Show that
$$E[S] = S_0 e^{\mu + \frac{1}{2}\sigma^2}.$$
You may use the formulas of problem P.1.B.

2. Find a probability density function for S, i.e. a function $f(s)$ so that for all real numbers c
$$P[S \le c] = \int_{-\infty}^{c} f(s)\,ds.$$

3. Using $S_0 = 1$, $\mu = 2$, and $\sigma = 1/2$ produce a plot of $f(s)$ for $-1 \le s \le 20$.

P.1.D A *Cauchy* random variable is one whose distribution is given by a density of the form

$$f(x) = \frac{1}{\pi} \frac{u}{u^2 + x^2},$$

for some parameter $u > 0$. Explain why a Cauchy random variable does not have finite moments. What is the distribution function of such a random variable?

Chapter 2

Conditioning

2.1 Time and Knowledge: Partial Information

In the context we have described so far, our point of view is that initially we don't know which of the possible states of the world $\omega \in \Omega$ will be realized but we can ascribe probabilities $P(A)$ to sets of possibilities. Later, one particular ω will emerge as the realized state of the world, and so a specific value $X(\omega)$ will be realized for a random variable X. However we often need to deal with situations where the realization of ω occurs progressively over time, not all at once. We deal with this using σ-algebras and conditional expectations.

For starters, imagine that we have two random variables Z and W defined on the same probability space $(\Omega, \mathcal{F}, P)$. Initially (at time 0) we know nothing about which state of the world will be realized. At time $s > 0$ we are told the value of Z, but not W. Later, at time $t > s$, the rest of the realization occurs and the value of W becomes known. For instance suppose that at time s we learn that $Z = 9$. Then we know that any ω with $Z(\omega) \neq 9$ is no longer a possibility. Whatever state of the world ω is ultimately realized, our information that $Z = 9$ tells us that the possibilities are now limited to those in

$$C_9 = \{\omega \in \Omega : \ Z(\omega) = 9\}.$$

Within C_9 we may still find that W can take many different possible values, and at time s we do not know which of them the actual realization will be. So at time s we can still only ascribe probabilities to different subsets $A \subset \Omega$ ($A \in \mathcal{F}$), and expected values to random variables like W whose value is still uncertain, but now we would ascribe *different* probabilities and expected values because of the partial knowledge we have from $Z = 9$. Moreover, our new probabilities and expectations at time s will depend on what value of Z was actually revealed at time s; we would compute different numbers if we had learned that $Z = 3$ instead of $Z = 9$. So the probabilities and expectations that we would compute based on what we learn about Z at time s will also depend on the actual value of $Z(\omega)$, and so describe some function $\phi(Z(\omega))$ retaining some dependence on ω. What we are beginning to describe are *conditional probabilities* and *conditional expectations*.

Perhaps the most unnatural part of this topic is the use of σ-algebras to describe a specified amount of information. To introduce this consider another random variable which is a function of Z, $R(\omega) = Z(\omega)^2$ for instance. When time s arrives and the value of Z is realized, then we will also know with certainty the realized value of R. We might say that the information we obtain by knowing $Z(\omega)$ is sufficient for us to know R as well. Another way to say it is that a set of ω described in terms of the value of $R(\omega)$ can also be described in terms of $Z(\omega)$. For instance

$$A = \{\omega : \ R(\omega) \leq c\}$$

is the same as

$$A = \begin{cases} \{\omega : \ Z(\omega) \in [-\sqrt{c}, \sqrt{c}]\} & \text{if } 0 \leq c, \\ \{\omega : \ Z(\omega) \in \emptyset\} & \text{if } c < 0. \end{cases}$$

The subsets of Ω that can be descirbed in terms of $Z(\omega)$, i.e. the sets of the form

$$A = \{\omega : \ Z(\omega) \in B\}$$

for some (Borel) set $B \subseteq \mathbb{R}$, form a σ-algebra. This is the σ-algebra *generated* by Z, often denoted $\sigma(Z)$. We will refer to it as $\mathcal{G}$ for the time being. What we have said above is that if a random variable like $R = Z^2$ is a function of Z, then R is $\mathcal{G}$ measurable. It turns out that the converse is true, namely if R is $\mathcal{G}$-measurable, then R is a function of Z: $R(\omega) = f(Z(\omega))$ for some function $f(\cdot)$. In other words, for R to be determined by the information we obtain by knowing Z is the same as saying the R is measurable with respect to the σ-algebra $\mathcal{G} = \sigma(Z)$. This may seem like a cumbersome way to say that $R(\omega) = f(Z(\omega))$, but it is the point of view that allows the rules for manipulating conditional expectations to be expressed in succinct form; see Section 2.3 below.

To understand how the σ-algebra $\mathcal{G}$ is different from $\mathcal{F}$ you might think of it this way. Z partitions Ω into clumps which we will call "*Z-cells*[1]", each cell consisting of those ω for which $Z(\omega)$ has a single common value. For instance our C_9 above is one such cell. Any random variable like our $R = Z^2$, whose value is known to us once the value of Z is known, must be constant on each such cell. A set of the form

$$A = \{\omega : \ R(\omega) \leq c\}$$

must be a union of the cells associated with Z. Each Z-cell

$$C_x = \{\omega \in \Omega : \ Z(\omega) = x\}$$

must be entirely contained in A (if $x^2 \leq c$) or else completely disjoint from A (if $x^2 > c$). For R to be a random variable, sets of the form A must be in our σ-algebra $\mathcal{F}$. But to be a random variable *with the additional property that its value is determined by the value of* Z, the sets A must not only be in $\mathcal{F}$ but must also be a union of the Z-cells. I.e. all sets of the form of A above must belong to a more limited collection of sets than $\mathcal{F}$. This more limited collection of sets is our σ-algebra, $\mathcal{G}$. We write $\mathcal{G} \subseteq \mathcal{F}$ to indicate that $\mathcal{G}$ is a more limited collection of sets. The intuitive idea is that a set A is in $\mathcal{G}$ if $A \in \mathcal{F}$ and A can be written as a union of Z-cells.

The σ-algebra $\mathcal{G}$ is the mathematical object that describes for us exactly what information we will have about states of the world at time s. At time s we will know which of the Z-cells the final state of the world will be taken from. For any subset $A \in \mathcal{G}$, at time s we will be able to give a definitive "yes" or "no" answer to whether the realized state of the world will be found in A.

A simple example may help. Suppose Ω consists of all pairs $\omega = (i, j)$ where i and j are integers between 1 and 6, and for any $A \subseteq \Omega$, $P(A) = \frac{1}{36} \times$ (the number of elements in A). We will let $\mathcal{F}$ consist of all subsets of Ω. Let $Z((i, j)) = i+j$ and $W((i, j)) = i-j$. In Figure 2.1 we have illustrated Ω as a 6×6 rectangular grid of dots in the plane, with i horizontally and j vertically. The Z-cells are just the backward diagonals of the grid, separated by lines in the figure. What we called C_9 above is $\{(3, 6), (4, 5), (5, 4), (6, 3)\}$, the boxed-in backward diagonal in the figure. The sets in $\mathcal{G} = \sigma(Z)$ are just those that are unions of some selection of the backward diagonals. The main (forward) diagonal $D = \{(1, 1), (2, 2), \ldots (6, 6)\}$ is an example of a set which *is* in $\mathcal{F}$ but *not* in $\mathcal{G}$.

Continuing with this simple example, lets go on to the issue of revising the probabilities and expectations based on the value of Z that is revealed to us at time s. Suppose for instance that at time s we find that $Z = 9$. Now this tells us that the state of the world ω must belong to our Z-cell C_9, but beyond that we can't yet say which $\omega \in C_9$ it actually is. What if I am now, at time s, asked for the probability of $A = \{\omega : \ W(\omega) \geq 3\}$? The $\omega \in A$ are the solid dots of the figure, another example of a set which does *not* beling to $\mathcal{G}$. We want to take advantage of the information that $Z(\omega) = 9$ provides and calculate a more informed probability of $W \geq 3$. What we are asking for here is the value of the *conditional probability* usually denoted $P(W \geq 3 | \ Z = 9)$. The standard definition is

$$P(W \geq 3 | \ Z = 9) = \frac{P(W \geq 3 \text{ and } Z = 9)}{P(Z = 9)} = \frac{1/36}{4/36} = \frac{1}{4}. \tag{2.1}$$

Thus $P(W \geq 3 | \ Z = 9)$ represents the *fraction* of the probability of $Z = 9$ that corresponds to $W \geq 3$. (For comparison, $P(W \geq 3) = 1/6$.)

[1] The terminology of "cells" or is *not* standard - you won't find it used anywhere else but here because it is not quite adequate as a definition to base a rigorous development on. But it is very convenient for the intuitive understanding that we are trying to develop.

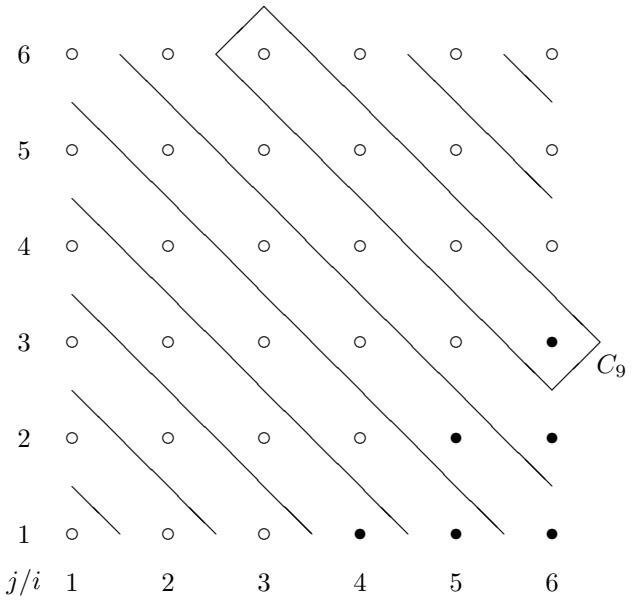


Figure 2.1: Example, with grading for $\mathcal{G}$-sets

If we change the value for Z in these calculations, the value of the conditional probability will change. For instance you can check that

$$P(W \geq 3|\ Z = 8) = \frac{1}{5} \text{ and } P(W \geq 3|\ Z = 10) = 0.$$

Comparing this to (2.1) we see that $P(W \geq 3|\ Z = k)$ does depend on k. If we are asked for $P(W \geq 3|\ Z)$ (without a sepcification of k), we should say that the answer depends on the value of $Z(\omega)$. If $Z = 9$ the conditional probability is $1/4$. If $Z = 10$ the conditional probability is 0. We can put all the possibilities together by saying

$$P(W \geq 3|\ Z) = \begin{cases} 0 & \text{if } Z(\omega) \leq 4 \\ \frac{1}{4} & \text{if } Z(\omega) = 5 \\ \frac{1}{5} & \text{if } Z(\omega) = 6 \\ \frac{1}{3} & \text{if } Z(\omega) = 7 \\ \frac{1}{5} & \text{if } Z(\omega) = 8 \\ \frac{1}{4} & \text{if } Z(\omega) = 9 \\ 0 & \text{if } Z(\omega) \geq 10 \end{cases}$$

Viewed this way, $P(W \geq 3|\ Z)$ is a function of ω, another random variable! We really should write $P(W \geq 3|\ Z)(\omega)$. You might think of it this way:

$$\omega \mapsto Z(\omega) = k \mapsto P(W \geq 3|\ Z = k) = P(W \geq 3|\ Z)(\omega).$$

Moreover, $P(W \geq 3|\ Z)$ is a $\mathcal{G}$-measurable random variable because its value is determined by the value of $Z(\omega)$. The usual notation is to write $P(W \geq 3|\ \mathcal{G})$ instead of $P(W \geq 3|\ Z)$, or $P(W \geq 3|\ \mathcal{G})(\omega)$ when we want to make the ω-dependence clear. Again, we view $\mathcal{G}$ as the object that actually specifies the information we have by knowing the value of Z.

As a prelude to the general definition of the next section, lets look at what we have said another way. Let Y be the random variable we called $P(W \geq 3|\ \mathcal{G})$ above:

$$Y(\omega) = P(W \geq 3|\ Z = k) \text{ if } Z(\omega) = k.$$

Then Y is a $\mathcal{G}$-measurable random variable, which means it takes constant values $v_k = Y(\omega)$ on each of the cells $\omega \in C_k$. Now (2.1) says that v_9 is the value determined by the formula $v_9 P(Z = 9) = P(W \geq 3 \text{ and } Z = 9)$, or written another way, and for any value of k,

$$E[Y; C_k] = v_k P(C_k) = P(\{W \geq 3\} \cap C_k).$$

Since every $A \in \mathcal{G}$ is a union of some selection of C_k, we see that

$$E[Y; A] = P(\{W \geq 3\} \cap A) \text{ for every } A \in \mathcal{G}.$$

This formula and the fact that Y is $\mathcal{G}$-measurable identify the random variable Y that we call the *conditional probability* $P(W \geq 3 |\ \mathcal{G})$.

A *conditional expectation* of W, or more generally of $g(W)$, would be calculated using the conditional probabilities:

$$\begin{aligned} E[g(W) |\ Z = 9] &= \sum_k g(k) P(W = k |\ Z = 9) \\ &= \sum_k g(k) P(W = k \text{ and } Z = 9)/P(Z = 9) \\ &= E[g(W); C_9]/P(Z = 9) = (g(3) + g(1) + g(-1) + g(-3)) \cdot \frac{1}{36} / \frac{1}{9}. \end{aligned}$$

Letting m_9 denote this value, the above can be summarized as $E[m_9; C_9] = E[g(W); C_9]$. If we calculate $m_k = E[g(W) |\ Z = k]$ for the other values of k and put them together as a single random variable,

$$E[g(W) |\ \mathcal{G}](\omega) = m_k \text{ for that } k \text{ with } k = Z(\omega),$$

then just as above, $Y = E[g(W) |\ \mathcal{G}]$ is the unique $\mathcal{G}$-measurable random variable with the property that

$$E[Y; A] = E[g(W); A] \text{ for every } A \in \mathcal{G}. \tag{2.2}$$

Notice that if we take

$$g(k) = \begin{cases} 1 & \text{for } k \geq 3 \\ 0 & \text{for } k < 3 \end{cases},$$

Then $E[g(W); A] = P(\{W \geq 3\} \cap A)$ so that our conditional probability is in fact a special case of our conditional expectation. In general $P(B |\ \mathcal{G}) = E[1_B |\ \mathcal{G}]$, so we will just talk about conditional expectations from now on.

It may seem unnatural to you to view $P(A |\ \mathcal{G})$ or $E[W |\ \mathcal{G}]$ as random variables. But it turns out that this is the point of view that allows the ideas to be generalized to settings where the above simple constructions fail. For instance if $P(Z = 9) = 0$, which would happen if Z were a standard normal random variable for instance, then the formulas we used above make no sense. Different formulas can be used in some circumstances. But the definition of $E[\cdot |\ \mathcal{G}]$ based on (2.2) turns out to be the correct formulation in all contexts, and is the natural setting in which to list the essential properties that we need to know about manipulating conditional expectations.

2.2 Definition and Calculation of Conditional Expectations

Suppose that X is an (integrable) random variable on $(\Omega, \mathcal{F}, P)$, and $\mathcal{G} \subseteq \mathcal{F}$ is a smaller σ-algebra. The general definition of $E[X |\ \mathcal{G}]$ is the following.

- $E[X |\ \mathcal{G}]$ is a $\mathcal{G}$-measurable random variable.
- $E[X |\ \mathcal{G}]$ has the property that for any set $A \in \mathcal{G}$, the following holds:

$$E[\, E[X |\ \mathcal{G}]\, ; A] = E[X; A].$$

I like to think of $E[X|\ \mathcal{G}]$ as a "smoothed out" version of X – within each $\mathcal{G}$-cell it has been replaced by a constant, an averaged value of X over that cell. That makes it $\mathcal{G}$-measurable (intuitively). However this averaging only takes place within individual $\mathcal{G}$-cells. The second property above is just (2.2) again. Note that using $A = \Omega$ in particular,

$$E[E[X|\ \mathcal{G}]] = E[X].$$

It is possible (measure theory again) to prove that $E[X|\ \mathcal{G}]$ always exists. The practical question is how we actually calculate it. Of course that depends on the particulars of $(\Omega, \mathcal{F}, P)$ and $\mathcal{G}$. We can state some formulas for special cases. You should try to view these as expressions of the idea of smoothing or averaging X within $\mathcal{G}$-cells.

Suppose we have a setting like the above involving two random variables Z and X each assuming only a finite number of possible values. Suppose

$$P(Z = i \text{ and } X = j) = p_{ij}$$

and, as before let $\mathcal{G} = \sigma(Z)$, the σ-algebra associated with knowing only Z. For ω in the Z-cell

$$C_i = \{\omega :\ Z(\omega) = i\}$$

we would calculate

$$E[g(X)|\ \mathcal{G}](\omega) = \frac{\sum_j g(j)\, p_{ij}}{\sum_j p_{ij}}.$$

This agrees with $E[g(X); C_i]/P(C_i)$, which is the way we thought of it before. There is a helpful way to see what we are doing in terms of the probabilities p_{ij} themselves. Let

$$\bar{p}_i = \sum_j p_{ij} = P(Z = i), \text{ and}$$
$$p_{j|i} = \frac{p_{ij}}{\bar{p}_i} = P(X = j|\ Z = i), \text{ assuming } \bar{p}_i \neq 0.$$

Then for all i, j pairs we can write

$$p_{ij} = p_{j|i}\bar{p}_i.$$

Notice that $\sum_i \bar{p}_i = 1$, and $\sum_j p_{j|i} = 1$ for each i. We can give a complete formula (i.e. not just for one C_i at a time) as

$$E[g(X)|\ \mathcal{G}] = \sum_j g(j) p_{j|Z(\omega)}. \tag{2.3}$$

In other words by finding an appropriate factorization of p_{ij} we can write a simplified formula.

Now consider a different situation in which there is a density for the joint distribution of (Z, X). By this we mean that there is a density function $f(z, x)$ which allows us to calculate probabilities associated with the pair (Z, X) as follows:

$$P(Z \leq a \text{ and } X \leq b) = \int_{-\infty}^{b}\int_{-\infty}^{a} f(z, x)\, dz\, dx.$$

Observe that we cannot define $E[X|\ Z = z]$ like we did before, using $E[X; Z = z]/P[Z = z]$, because

$$P[Z = z] = \int_{z}^{z}\int_{-\infty}^{\infty} f(z, x)\, dx\, dz = 0.$$

We would be dividing by 0, so the ratio makes no sense. However there is a formula for this case, analogous to the one above. If we can factor $f(z, x)$ as a product

$$f(z, x) = \bar{f}(z) f(x|z) \tag{2.4}$$

of the *marginal* density $\bar{f}(z)$ for Z alone,

$$P(Z \leq a) = \int_{-\infty}^{a} \bar{f}(z)\, dz,$$

and a so-called *conditional density* $f(x|z)$, which is a density in x for each value of z individually,

$$\int_{-\infty}^{\infty} f(x|z)\,dx = 1 \quad \text{for each } z,$$

then conditional expectations with respect to $\mathcal{G} = \sigma(Z)$ are computed by

$$E[g(X)|\ \mathcal{G}](\omega) = \int g(x) f(x|Z(\omega))\,dx. \tag{2.5}$$

Usually $\bar{f}(z)$ and $f(x|z)$ can be computed from the original $f(z,x)$.

$$\bar{f}(z) = \int f(z,x)\,dx$$
$$f(x|z) = \frac{f(z,x)}{\bar{f}(z)}.$$

(If $\bar{f}(z) = 0$ occurs then it is a little more complicated.)

What makes (2.5) correct? Its comforting that it follows a similar pattern to (2.3), but that's not logically sound reason. The reason (2.5) is correct is that it satisfies the definition we stated at the begining of this section, as you will verify in problem 2.B. The formula (2.5) generalizes to conditional expectations of any function of the pair (Z,X); just integrate out the W dependence using $f(x|Z(\omega))$:

$$E[g(Z,X)|\ \mathcal{G}](\omega) = \int g(Z(\omega),x) f(x|Z(\omega))\,dx. \tag{2.6}$$

2.3 Properties of Conditional Expectations

There are a number of general properties of conditional expectations that allow us to work with them efficiently. These are all provable as theorems, based on the definition we gave above. Most of them are generalized versions of the properties of expectation that we listed earlier. (In fact the standard expectation $E[X]$ is the same as the conditional expectation $E[X|\ \mathcal{G}_0]$ using the "trivial" σ-algebra $\mathcal{G}_0 = \{\emptyset, \Omega\}$. The only $\mathcal{G}_0$-measurable random variables are constants.) Here are the essential properties. We assume X, Y are both integrable random variables.

1. If c is a constant, $E[c|\ \mathcal{G}] = c$.
2. if $X(\omega) \le Y(\omega)$ for (almost) all ω, then $E[X|\ \mathcal{G}](\omega) \le E[Y|\ \mathcal{G}](\omega)$ for (almost) all ω.
3. For any two constants α, β,
$$E[\alpha X + \beta Y|\ \mathcal{G}] = \alpha E[X|\ \mathcal{G}] + \beta E[Y|\ \mathcal{G}]$$
4. $|E[X|\ \mathcal{G}]| \le E[|X|\,|\ \mathcal{G}]$.
5. If Y is already $\mathcal{G}$-measurable (and XY is integrable) then
$$E[XY|\ \mathcal{G}] = Y E[X|\ \mathcal{G}].$$
6. (Tower Law) If $\mathcal{H} \subseteq \mathcal{G}$ is a third even smaller σ-algebra, then
$$E[X|\ \mathcal{H}] = E[\,E[X|\ \mathcal{G}]\,|\ \mathcal{H}].$$

(Note that the last two of these are Proposition 4.5 of page 42 of the text.) We should qualify all the statements of equality above. Looking back at the definition of $E[X|\ \mathcal{G}]$, if $\mathcal{G}$ contains some nonempty sets $N \in \mathcal{G}$ with $P(N) = 0$, then the definition does not uniquely determine $E[X|\ \mathcal{G}]$. We could alter $E[X|\ \mathcal{G}]$ for those $\omega \in N$ without disrupting either part of the definition. For that reason statements about conditional expectations can only be made for almost all ω, meaning they are allowed to fail on a set of probability 0. We wrote this in in item 2 above, but neglected in in the others. This is another technicality we would need to be more careful about if this were a rigorous treatment, but which we will ignore for our present purposes.

2.4 Independence

Conditional expectations give us one way to make precise statements about how the value of one random variable Z ($\mathcal{G} = \sigma(Z)$) determines the value of another. If Y is completely determined for us once we know the value of Z, this means that Y is $\mathcal{G}$-measurable, and so by the properties above

$$Y = E[Y|\ \mathcal{G}].$$

Intuitively this says that there is no randomness in Y beyond what is already present in Z, so there is no randomness for the conditional expectation to smooth out. On the other end of the scale are those random variables X which are *independent* of Z. The definition of this is that for all (Borel) sets $C, D \subseteq \mathbb{R}$,

$$P(X \in C \text{ and } Z \in D) = P(X \in C)\, P(Z \in D).$$

The set $A = \{w \in \Omega :\ Z(\omega) \in D\}$ is a typical set in $\mathcal{G}$, so the above could be rephrased $P(\{\omega :\ X \in C\} \cap A) = P(\{\omega :\ X \in C\})\, P(A)$ for all $A \in \mathcal{G}$. In other words

$$P(X \in C|\ \mathcal{G})(\omega) = P(X \in C) \text{ for all } \omega \in \Omega.$$

We could say "X is independent of $\mathcal{G}$," instead of "X is independent of Z." The idea is that knowing the value of Z tells us nothing at all about X; the probabilites of different X-outcomes remain unchanged by knowledge of Z. In terms of expectations this means

$$E[X|\ \mathcal{G}](\omega) = E[X], \quad \text{constant with respect to } \omega. \tag{2.7}$$

More generally, for any function $h(x)$, the independence of X from $\mathcal{G}$ means

$$E[h(X)|\ \mathcal{G}](\omega) = E[h(X)].$$

Independence of X and Z has many consequences, among them

$$E[XZ] = E[X]\, E[Z], \tag{2.8}$$

$$P(X \in A \text{ and } Z \in B) = P(X \in A) \cdot P(Z \in B). \tag{2.9}$$

These are both consequences of (2.7) and the general definition (page 11) and properties of conditional expectations above. Another useful consequence of independence is that for a function $h(x, y)$ (assuming $h(X, Y)$ is integrable),

$$E[h(X, Z)] = \iint h(x, z) f_X(x) f_Z(z)\, dx\, dz, \tag{2.10}$$

provided f_X and f_Z are densities for the distributions of X and Y respectively: i.e.

$$P[X \le b] = \int_{-\infty}^{b} f_X(x)\, dx, \quad P[Z \le b] = \int_{-\infty}^{b} f_Z(z)\, dz.$$

This simply says that the joint density for (X, Z) is the product of the individual or marginal densities for X and Z.

2.5 A Separation Result

The features of conditional expectations involving random variables that are measurable with respect to the conditioning σ-field and random variables that are independent of it can be combined into a separation result that is often useful. Suppose we are interested in $E[Y|\ \mathcal{G}]$, and we can find a way to express the random variable Y as some function of some random variables $X = (X_1, \ldots, X_n)$ which are independent of $\mathcal{G}$ and some other random variables $Z = (Z_1, \ldots, Z_m)$ which are measurable with respect to $\mathcal{G}$:

$$Y = h(X, Z).$$

Then, assuming Y is integrable,

$$E[Y|\ \mathcal{G}] = g(Z)$$

where $g(z)$ is the function obtained by replacing the random variable Z by the conventional variable z in the expectation:

$$g(z) = E[h(X, z)].$$

In other words the result of the conditional expectation is essentially to "integrate out" the X dependence, leaving the Z dependence intact.

2.6 Problems

P.2.A In the example of figure 2.1, compute $E[W^2|\ Z = k]$ for all possible values of k.

P.2.B Consider the situation where we have a joint density function $f(z, w)$ for a pair Z, W and assume that we can factor f as described in (2.4). Explain why the right side of (2.5) has the properties required by the definition (page 11) of the left side.

For the next two problems, please read the opening paragraph of the next chapter for the definition of $\mathcal{F}_t$ $(t = 1, 2)$, the σ-algebra of information avaliable at time t.

P.2.C Consider the tree illustrated in Figure 2.2 below. Note that this is *not* a binary tree, because the number of branches from each node is not always 2. Since it is not recombinant we *can* use the set of final

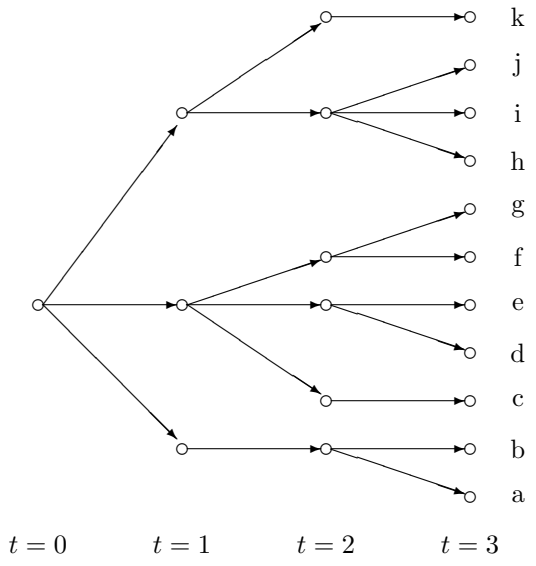


Figure 2.2: Tree for Problem G

nodes $\Omega = \{a, b, c, d, \ldots, k\}$ as the set of all possible states of the world. Associated with the intermediate times $t = 1$ and $t = 2$ are σ-algebras $\mathcal{F}_1$ and $\mathcal{F}_2$ of those subsets of Ω which are measurable at time t. Although there are too many sets in these σ-algebras to give a complete listing[2], the "indivisible cells" which are used to make up sets in $\mathcal{F}_i$ can be listed. What are the indivisible cells for $\mathcal{F}_1$? What about for $\mathcal{F}_2$?

Consider the following two random variables defined on Ω:

$$X(\omega) = \begin{cases} 1 & \text{if } \omega \in \{c, k\} \\ 0 & \text{if } \omega \in \{a, b, h, i, j\} \\ -1 & \text{if } \omega \in \{d, e, f, g\} \end{cases} \qquad Y(\omega) = \begin{cases} 1 & \text{if } \omega \in \{a, b\} \\ 0 & \text{if } \omega \in \{h, i, j, k\} \\ -1 & \text{if } \omega \in \{c, d, e, f, g\} \end{cases}$$

[2] $\mathcal{F}_2$ contains 64 different sets.

Which of these are $\mathcal{F}_1$-measurable? Which are $\mathcal{F}_2$-measurable?

P.2.D Consider the tree of Figure 2.4, page 19 of the text. Assume that the probabilities of all the up-jumps are $\frac{3}{4}$ and for the down-jumps, $\frac{1}{4}$. Because the tree is recombinant we can not use the set of final nodes as Ω. One way to describe a complete set of states of the world would be to let Ω consist of the set of all triples $\omega = (i, j, k)$ where each of i, j, k is ± 1, indicating whether the transition was up or down from $t = 0$, $t = 1$ and $t = 2$ respectively. For example, for the state of the world $\omega = (+1, -1, -1)$ would correspond to the sequence of stock prices $S_0 = 80$, $S_1 = 120$, $S_2 = 60$, $S_3 = 30$. Using this description, define a new random variable X on Ω according to $X(\omega) = i + 3j + 5k$, if $\omega = (i, j, k)$. Using the probabilities specified above, calculate each of $E[X|\ \mathcal{F}_1]$ and $E[X|\ \mathcal{F}_2]$ and express them as functions on Ω.

Chapter 3

Stochastic Processes: Martingales, Markov Chains and Trees

A *stochastic process* is essentially a time-dependent random variable X_t. We usually view this as a collection of random variables, one for each possible time value t. For now we will only consider a finite sequence of possible time values: $t = 0, 1, \ldots, n$. So we have random variables X_0, X_1, $\ldots$, X_n whose values are realized one at a time. We will know what X_0 is at the outset, $t = 0$. (Typically X_0 is a constant, with no randomness at all.) We will learn the value of X_1 at $t = 1$. At $t = 2$ we will find out what X_2 is, and so forth. These are all defined on the same underlying set Ω of possible states of the world. As time progresses we learn progressively more about the ultimate state of the world $\omega \in \Omega$. So there are σ-algebras $\mathcal{F}_0$, $\mathcal{F}_1$, $\ldots$, $\mathcal{F}_n$ that account for the information we have at the successive times. For smaller values of t, $\mathcal{F}_t$ will be "coarser", i.e. its cells will be fewer and bigger. As we move from $t = k$ to $t = k + 1$ we obtain more information about ω and will be able to make more refined statements about where ω is or is not. A single cell in $\mathcal{F}_k$ might be the union of several smaller cells in $\mathcal{F}_{k+1}$. So the σ-algebras are nested:

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}_n.$$

A collection of nested σ-algebras such as this is called a *filtration*. The filtration tells us what information about the state of the world is available at different times. The random variables whose values will be known at time $t = k$ are precisely the $\mathcal{F}_k$-measurable random variables, such as X_k in particular. When it is said that the stochastic process X_t is *adapted* to the filtration $(\mathcal{F}_t)$ this simply means that X_t is $\mathcal{F}_t$-measurable for each t.

Our various trees are all special cases of a particular type of stochastic process called a *Markov chain*. The idea is that there is a set of possible states $\{1, \ldots, m\}$ (like the individual nodes in the trees of Section 2.2). At each time $t = k$ the process X_t is located at one of the states ($X_k = i$ for instance), and jumps to another ($X_{k+1} = j$ for instance) at the next time with a specified *transition probability*

$$X_k = i \to X_{k+1} = j \quad \text{with probability } p_{i,j}.$$

We require that $0 \leq p_{i,j} \leq 1$ and $\sum_j p_{i,j} = 1$ for each i. These are really conditional probabilities:

$$P(X_{k+1} = j \mid X_k = i) = p_{i,j}.$$

To write this as a conditional statement with respect to $\mathcal{F}_k$,

$$P(X_{k+1} = j \mid \mathcal{F}_k) = p_{X_k,j},$$

or more generally,

$$E[f(X_{k+1}) \mid \mathcal{F}_k] = \sum_j f(j) p_{X_k,j}. \tag{3.1}$$

Notice that the $p_{i,j}$ here play the same role as the $p_{j|i}$ in (2.3).

To complete the description of a Markov chain, we also need to know the *initial distribution*, namely the probabilities

$$p_i = P(X_0 = i).$$

With these in hand, the probability that the chain proceeds through any specific sequence of values is computed (from the tower law):

$$P(X_0 = i_0,\, X_1 = i_1,\, \ldots X_n = i_n) = p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}.$$

Trees are special in that X_k can never visit the same node twice. This means $p_{i,j} > 0$ only for those nodes j immediately to the right of i in the tree. Binary trees have the additional property that $p_{i,j} > 0$ for only two nodes j, given i. For a node j in a recombinant tree, like those on pages 20—23 of the text, there can be more than one node i with $p_{i,j} > 0$, but for nonrecombinant trees there is a unique i with $p_{i,j} > 0$. As far as choosing a set of states of the world Ω to describe a tree, if the tree is not recombinant, then the set of final nodes will do fine. This is because for a nonrecombinant tree knowing the final node determines all the previous nodes it must have passed through to get there. But for a recombinant tree this would not be adequate. For the tree of Figure 2.5, page 21 we might take Ω to be the set of triples $\omega = (i, j, k)$ where each of i, j, k is ± 1 indicating whether the transition is up or down as the process moves to time $t = 1, 2, 3$ respectively.

3.1 Martingales

One of the most important types of stochastic process for us is a *martingale*. This refers to a stochastic process M_k with the property that for each time value $k = 0, 1, \ldots$

$$E[M_{k+1} \mid \mathcal{F}_k] = M_k. \tag{3.2}$$

(It is assumed that each M_k is integrable.) Because of the tower law for conditional expectations,

$$E[M_{k+2} \mid \mathcal{F}_k] = E[E[M_{k+2} \mid \mathcal{F}_{k+1}] \mid \mathcal{F}_k] = E[M_{k+1} \mid \mathcal{F}_k] = M_k.$$

In general, it follows that for any $s < t$,

$$E[M_t \mid \mathcal{F}_s] = M_s.$$

If there is a largest time, $t = T$ then the values of M_t for $t < T$ are all calculated from M_T: $M_t = E[M_T \mid \mathcal{F}_t]$. Also note that (3.2) implies that M_t is $\mathcal{F}_t$ measurable, so that being adapted to the filtration $(\mathcal{F}_t)$ is implicit in the definition (3.2).

3.2 Doob's Inequality

The fact that the values of a martingale M_t are all related to each other through conditional expectations means that inequalities and formulas are possible which take advantage of that relationship. One particularly important inequality is *Doob's L_2 martingale inequality*:

$$E[(\max_{1 \le k \le m} |M_k|)^2] \le 4E[M_m^2]. \tag{3.3}$$

(It is assumed that $E[M_m^2] < \infty$.) This will be important for us when we construct stochastic integrals with respect to Brownian motion later on.

3.3 Stopping Times and Optional Stopping

The word "martingale" comes from a French word referring to gambling strategies. Martingales do indeed arise naturally in the context of gambling. Imagine that a gambler starts with $\$M_0$ in cash. Lets say he is playing a simple game in which he bets on the outcome of some random variable X_1. If he places a bet of size $\$w_1$ (which we assume he does not recover - it is the price he pays to play) then his winnings will be

$\$w_1X_1$. For this to be a "fair" game, we assume that $E[X_1] = 1$. After any winnings from this wager have been paid, he has $\$M_1$ where $M_1 = M_0 + w_1(X_1 - 1)$. If he does this repeatedly, wagering w_1, w_2, ... on games with outcomes X_1, X_2, ... then the history of his fortune will be described by M_t, $t = 0, 1, \ldots$ where in general

$$M_n = M_{n-1} + w_n(X_n - 1).$$

Assuming that the X_t are independent with $E[X_t] = 1$, and that the wagers w_i are determined only on the basis of the information available *before* the i^{th} game is played, M_t will be a martingale.

Now suppose our gambler has a strategy: he will play until the first time that his fortunes reach some some level, and then he will quit. For instance suppose he plans to play until the first time $M_t \geq 1000$. Let τ refer to the time t when this first occurs. The value of τ will depend on the sequence of outcomes X_t (and the sizes of his wagers w_t). τ is a time-valued random variable, but it has some special structure, because its value does not anticipate the future. I.e. we can always decide whether to stop now or not based only on what has happened up to now. The mathematical way to say this is that for each t,

$$\{\omega \in \Omega : \ \tau \leq t\} \in \mathcal{F}_t.$$

(Note that a consequence of this is that $\{\tau = t\}, \{\tau > t\} \in \mathcal{F}_t$ as well.) A time-valued random variable with this property is called a *stopping time* (sometimes a *Markov time*). If we use a stopping time to stop a martingale, the result will again be a martingale. For instance the gambler's fortunes with his stopping rule in place are

$$M_t^\tau = \begin{cases} M_t & \text{if } t < \tau \\ M_\tau & \text{if } \tau \leq t \end{cases} = 1_{\{\tau > t\}} M_t + \sum_{k=0}^{t} 1_{\{\tau = k\}} M_k = M_{t \wedge \tau}$$

I.e. once $\tau = t$ he places no more wagers so his fortune remains fixed after τ. If M_t is a martingale, and τ is a stopping time, then M_t^τ will also be a martingale. (This is a simple calculation with conditional expectations; see Problem P.3.A.) As a consequence, assuming $\mathcal{F}_0 = \{\emptyset, \Omega\}$, so that M_0 is a constant,

$$M_0 = M_0^\tau = E[M_n^\tau]$$

for any n. The implication is that there is no betting and stopping strategy that will guarantee a profit in a fixed amount of time. (If you were allowed to continue betting indefinitely the situation gets more complicated – there are "doubling strategies" that guarantee that $M_t \geq \$1000$ eventually does occur, but it may be a quite a while before that actually happens and in the meantime you would probably be forced to abandon your strategy by house rules such as limits on the sizes of bets or how much credit ($M_t < 0$) the house will give you to keep playing.)

Obviously, buying/selling strategies for stocks have some close connections to this. American options in particular give the holder the right to exercise the option at a time of his/her choosing, i.e. at a stopping time. So stopping times will be relevant in the study of more complicated financial products.

3.4 Stochastic Integration in Discrete Time

There is a "calculus of stochastic processes" that is based on using stochastic integrals with respect to a martingale M_t to find formulas connecting different stochastic processes with each other. The basic idea is that given a martingale M_k we can build a new martingale by adding its increments

$$\Delta M_k = M_k - M_{k-1}$$

in new combinations. The martingale property (3.2) says that

$$E[\Delta M_k \mid \mathcal{F}_{k-1}] = 0.$$

Suppose we have a starting value N_0, which is $\mathcal{F}_0$ measurable (e.g. a constant), and some other stochastic process ϕ_k. Now build a new stochastic process N_k according to

$$N_k = N_0 + \phi_1(M_1 - M_0) + \ldots + \phi_k(M_k - M_{k-1}) \tag{3.4}$$

$$= N_0 + \sum_{i=1}^{k} \phi_i \Delta M_i. \tag{3.5}$$

If ϕ_k is adapted, so that ϕ_k is $\mathcal{F}_k$-measurable for each k, then N_k will likewise be $\mathcal{F}_k$-measurable because it is constructed from other $\mathcal{F}_k$-measurable quantities. Now consider whether N_k might actually be another martingale. Notice that

$$N_{k+1} = N_k + \phi_{k+1}\Delta M_{k+1}.$$

Using the rules for conditional expectations

$$\begin{aligned} E[N_{k+1} \mid \mathcal{F}_k] &= E[N_k + \phi_{k+1}\Delta M_{k+1} \mid \mathcal{F}_k] \\ &= N_k + E[\phi_{k+1}\Delta M_{k+1} \mid \mathcal{F}_k] \\ &= N_k + \phi_{k+1}E[\Delta M_{k+1} \mid \mathcal{F}_k], \quad \textit{provided } \phi_{k+1} \text{ is } \mathcal{F}_k\text{-measurable} \\ &= N_k, \end{aligned}$$

so that N_k is another martingale. There are some technical conditions that are needed to justify this. For instance something needs to be said to insure that $\phi_k\Delta M_k$ is integrable. For us it will be enough to say they are bounded (which is always true if there are only a finite number of possible states of the world in Ω). The really critical assumption however is the one we used in the third line above, that instead of being $\mathcal{F}_{k+1}$-measurable (which is all adapted asks for), ϕ_{k+1} should be $\mathcal{F}_k$-measurable. In other words the random variable ϕ_{k+1} should determined by the information available one time-step early, at $t = k$. Such a stochastic process ϕ_k is called a *previsible* stochastic process. I like to think of ϕ_{k+1} as associated with the time *interval* between $t = k$ and $t = k + 1$; previsible means ϕ_{k+1} should be known at the start of that interval. The basic point then is that if ϕ_k is a previsible stochastic process, and M_k is a martingale, then

$$N_k = N_0 + \sum_{i=1}^{k} \phi_i \Delta M_i$$

defines a new martingale. This is the discrete version of the stochastic integral construction that we will talk more about later.

3.5 No-Arbitrage Pricing and Martingales for Binomial Trees

In this section we want to re-express what we know about binomial tree models in the language of martingales, and rearrange our formulas in ways which anticipate those of the Black-Scholes model to come.

Consider a binomial tree as discussed in Chapter 2 of the text. This consists of a collection of nodes which we will denote by k. Each node is associated with a particular time. There is one node $k = 0$ at time $t = 0$; two nodes $k = 1, 2$ at time $t = 1$ and so forth. (In general there will be $t + 1$ nodes at time t.) Lets say the possible times are $t = 0, 1, 2, \ldots T$. From node k at time t there are two possible transitions as we move to time $t + 1$: we can go "up" to a node k^u or "down" to a node k^d. (Of course k^u and k^d are among the possible $t + 1$ nodes.) These happen with probabilities p_k^u and p_k^d respectively. (Assume $0 < p_,^u p_k^d$ and $p_k^u + p_k^d = 1$.)

Let X_t denote the node we reach at time t. This is an example of a Markov chain as described above. The initial distribution has $p_0 = 1$ and all other $p_k = 0$. The transition probabilities are

$$p_{k,j} = \begin{cases} p_k^u & \text{if } j = k^u \\ p_k^d & \text{if } j = k^d \\ 0 & \text{otherwise.} \end{cases}$$

In general a Markov chain can return to the same node multiple times, but in our binomial tree every node is associated with a specific time, and the only transitions allowed are those which move to nodes for the next time. So we can't return to nodes from an earlier time.

To set this up formally we can take

$$\Omega = \{\pm 1\} \times \{\pm 1\} \times \cdots \times \{\pm 1\} \ (T \text{ times}),$$

so that a typical $\omega \in \Omega$ consists of a seqence of T ± 1s: $\omega = (+1, -1, \ldots, 1)$, which identifies a path through the tree by specifying whether we follow the upward or downward branch at each step. In other words

knowing ω determinds the specific sequence of nodes $X_0(\omega), X_1(\omega), \ldots, X_T(\omega)$ the chain will follow. The probability of each ω is the product of the associated transition probabilities. For instance, suppose $T = 3$ and consider $\omega = (+1, -1, -1)$. This corresponds to

$$X_0(\omega) = 0,\ X_1(\omega) = 2,\ X_2(\omega) = 4,\ X_3(\omega) = 7,$$

and would have probability

$$P(\{\omega\}) = p_{0,2}\, p_{2,4}\, p_{4,7} = p_0^u p_2^d p_4^d.$$

$\mathcal{F}$ would consist of all subsets of Ω. The cells in $\mathcal{F}_1$ are subsets in which the first term of ω is fixed. The cells in $\mathcal{F}_2$ are the subsets in which the first two terms of ω are fixed, and so forth. $\mathcal{F}_0 = \{\emptyset, \Omega\}$

At each node k we have a stock price $s(k)$. In the text we assumed that $s(k^u)$ and $s(k^d)$ were related to $s(k)$ by factors u and d:

$$s(k^u) = u\, s(k), \quad s(k^d) = d\, s(k).$$

The stock price process S_t is just $s(X_t)$, i.e. the value of the price associated with the node where the chain is located at time t.

Our discussion of the absence of arbitrage in a binomial tree involved the existence of "martingale probabilities" q_k^u, q_k^d for each branch determined by

$$s(k) = \frac{s(k^u)q_k^u + s(k^d)q_k^d}{1+R}. \tag{3.6}$$

From these $q_k^{u,d}$ we determine an alternate set of transition probabilities $q_{k,j}$ for our Markov chain, by again using $q_{k,j} = 0$ for the transitions not in our our tree, and the same initial distribution $q_0 = 1$, $q_{k\neq 0} = 0$. In these terms, (3.6) can be expressed as

$$s(k) = \sum_j \frac{s(j)}{1+R} q_{k,j},$$

for all k except the final ($t = T$) nodes. Associated with these $q_{k,j}$ is a probability measure Q on Ω, just like P for the original $p_{k,j}$. In terms of Q, (3.6) together with (3.1) says that for $t < T$,

$$\begin{aligned} E^Q[S_{t+1}/(1+R)|\ \mathcal{F}_t] &= E^Q[s(X_{t+1})/(1+R)|\ \mathcal{F}_t] \\ &= s(X_t) \\ &= S_t. \end{aligned}$$

With $B_t = (1+R)^t$, so that $\frac{1}{B_{t+1}} = \frac{1}{B_t}\frac{1}{1+R}$, we can write this as

$$E^Q[S_{t+1}/B_{t+1}|\ \mathcal{F}_t] = S_t/B_t. \tag{3.7}$$

In other words S_t/B_t is a Q-martingale.

In our discussion of binomial tree models we observed that, under the hypothesis $p_k^{u,d} > 0$, the absence of arbitrage condition $d^k < 1 + R < u^k$ is equivalent to $q_k^{u,d} > 0$. This means that the transitions $k \to j$ for which $p_{k,j} > 0$ are the same as those for which $q_{k,j} > 0$. In other words, the sequences of states X_0, X_1, …, X_T which have probability 0 according to P are precisely the same as the ones which have probability 0 according to Q. This can be stated succinctly as

$$\text{For all } A \in \mathcal{F},\ P(A) = 0 \text{ if and only if } Q(A) = 0. \tag{3.8}$$

Two probability measures P and Q on $(\Omega, \mathcal{F})$ for which (3.8) holds are called *equivalent*[1]. So all our discussion of binomial trees can be summarized as follows.

[1]This does not mean they are they same. $P(A)$ and $Q(A)$ do not need to be equal except when they are both 0. To be equivalent in this sense only means the same events have probability 0 under both measures. This is what we have been described previously by saying "equivalent positivity properties." The usual terminology is simply "equivalent." We will return to this idea in Section 7.1.

The binomial tree is free of arbitrage if and only if there is a probability measure Q which is equivalent to P and such that S_t/B_t is a martingale with respect to Q.

Next let's consider self-financing portfolios in terms of martingales. Supposing that $d < u$ we know that given any "terminal" random variable $X = \Phi(S_T)$ we can find a self-financing portfolio $h_t = (x_t, y_t)$ which replicates X. Recall that h_t specifies the holdings over the time interval $t-1 \to t$. In particular x_t and y_t are $\mathcal{F}_{t-1}$-measurable, i.e. they are *previsible* processes.

To anticipate the expression we will use in continuous time, lets define z_t by $x_t = z_t B_{t-1}$. That is, z_t is the number of bonds (or shares of the bond account) held for $t-1 \to t$. Since $B_t = (1+R)^t$ we can write

$$\begin{aligned} V^h_{(t-1)+} &= x_t + y_t S_{t-1} = z_t B_{t-1} + y_t S_{t-1} \\ V^h_{t-} &= x_t(1+R) + y_t S_t = z_t B_t + y_t S_t \\ V^h_{t+} &= x_{t+1} + y_{t+1} S_t = z_{t+1} B_t + y_{t+1} S_t. \end{aligned}$$

The self-financing property is that $V^h_{t-} = V^h_{t+}$. This can be expressed several ways:

$$\begin{aligned} 0 &= B_t \Delta z_{t+1} + S_t \Delta y_{t+1} \ (t = 0, \dots, T-1) \\ 0 &= B_{t-1} \Delta z_t + S_{t-1} \Delta y_t \ (t = 1, \dots, T) \\ 0 &= \Delta z_t + M_{t-1} \Delta y_t \ (t = 1, \dots, T), \end{aligned}$$

where $M_t = S_t/B_t$ is our basic Q-martingale. Notice that a general "product rule" for backward differences is

$$\Delta(f_t g_t) = f_t g_t - f_{t-1} g_{t-1} + [f_t g_{t-1} - f_t g_{t-1}] = f_t \Delta g_t + g_{t-1} \Delta f_t.$$

(See problem P.6.A for a more symmetric form.) So in general $V^h_t = z_t B_t + y_t S_t$ has differences

$$\begin{aligned} \Delta V^h_t &= z_t \Delta B_t + B_{t-1} \Delta z_t + y_t \Delta S_t + S_{t-1} \Delta y_t \\ &= z_t \Delta B_t + y_t \Delta S_t + [B_{t-1} \Delta z_t + S_{t-1} \Delta y_t]. \end{aligned}$$

So another way to express the self-financing property is that $V^h_t = z_t B_t + y_t S_t$ satisfies

$$\Delta V^h_t = z_t \Delta B_t + y_t \Delta S_t,$$

for $t = 1, \dots, T$. This expresses the property that all the changes to V^h_t result from the changes in B_t and S_t, not the changes in z_t and y_t.

Now consider $N_t = V^h_t/B_t = z_t + y_t M_t$. By the same difference product rule,

$$\begin{aligned} \Delta N_t &= \Delta z_t + y_t \Delta M_t + M_{t-1} \Delta y_t \\ &= y_t \Delta M_t, \end{aligned}$$

by the self-financing identity $0 = \Delta z_t + M_{t-1} \Delta y_t$. Thus

$$N_t = N_0 + \sum_{i=1}^{t} y_i \Delta M_i$$

is a discrete stochastic integral of our Q-martingale M_t with previsible integrand y_t. Consequently $N_t = V^h_t/B_t$ is also a Q-martingale. This is not news. The point is that the relationship of V^h_t/B_t to S_t/B_t is a stochastic integral construction. The martingale property $N_t = E^Q[N_T | \ \mathcal{F}_t]$ and the replicating property $V^h_t = X$ translate into the *risk-neutral* or *martingale pricing formula*:

$$\begin{aligned} \pi(t, X) &= B_t E^Q[X/B_T | \ \mathcal{F}_t] \\ \pi(0, X) &= E^Q[X/B_T], \end{aligned}$$

the last line because $B_0 = 1$. This formulation makes no reference to h; everything is in terms of Q. (However, the reasoning for why these formulas must give the market price depends very much on the existence of h. This is the issue of completeness.)

3.6 Problems

P.3.A Suppose that M_t, for $t = 0, 1, 2, \ldots n$, is a martingale with respect to a filtration $\{\mathcal{F}_t\}$, and that $\tau : \Omega \to \{0, 1, \ldots n\}$ is a stopping time. Recall our definition of M_t^τ. Show that M_t^τ is a martingale by verifying

$$E[M_{t+1}^\tau \mid \mathcal{F}_t] = M_t^\tau$$

for each $t = 0, \ldots, n-1$. [Hint: write $M_{t+1}^\tau = M_t^\tau + 1_{\{\tau > t\}} \cdot (M_{t+1} - M_t)$.] Show that as a consequence

$$E[M_n^\tau] = E[M_0] = E[M_n].$$

If M_0 is just a constant, then all three of these are $= M_0$.

Chapter 4

Various Results about Limits

There are some important concepts and theorems about limits of random variables that we should mention. Suppose $X_1, X_2, \ldots$ is a sequence of random variables, all defined on $(\Omega, \mathcal{F}, P)$. There are various different senses in which we can say $\lim_{n\to\infty} X_n = Y$, Y being another random variable. We can say $X_n \to Y$ *almost surely*, which means that

$$\lim_{n\to\infty} X_n(\omega) = Y(\omega)$$

for each $\omega \in \Omega$, excluding a set $N \in \mathcal{F}$ of exceptions which has zero probability: $P(N) = 0$.

A different form of convergence is *mean-square* or L_2 convergence:

$$E[|Y - X_n|^2] \to 0.$$

This does not imply $X_n(\omega) \to Y(\omega)$ for any particular $\omega \in \Omega$ (although it does for some subsequence X_{n_k}). It does however identify the random variable Y uniquely. *This will be important for stochastic integrals below.*

Weaker yet is the notion of *convergence in distribution*, written $X_n \Rightarrow Y$. This means only that the probabilities associated with X_n converge to those associated with Y:

$$P(X_n \le a) \to P(Y \le a)$$

for all $a \in \mathbb{R}$ with $P(Y = a) = 0$. This does *not* imply that $X_n(\omega) \to Y(\omega)$ for any ω at all; see Problem P.4.A.

Suppose $X_1, X_2, \ldots$ is a sequence of independent, identically distributed random variables. This means two things. First that they are all independent of each other. The easiest way to define this is to say that given any sequence $c_1, c_2, \ldots, c_n$ of real numbers

$$P(X_1 \le c_1 \text{ and } X_2 \le c_2 \text{ and } \ldots \text{ and } X_n \le c_n) = P(X_1 \le c_1) \cdot P(X_2 \le c_2) \cdots P(X_n \le c_n).$$

To say they are identically distributed simply means that they all have the same distribution:

$$P(X_k \le c) \quad \text{is the same for all } k.$$

The important classical results we want to record concerns the convergence of the partial averages of the X_i:

$$S_n = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n}\sum_1^n X_i.$$

Most people seem to intuitively realize that S_n ought to converge in some sense to the mean $m = E[X_k]$ (assumed finite). Kolmogorov's famous *Strong Law of Large Numbers* says that this convergence is almost sure: for all ω excepting a set of probability 0,

$$\frac{1}{n}\sum_1^n X_i(\omega) \to m.$$

The Lindberg-Lévy Theorem, also known as the *Central Limit Theorem* involves convergence in distribution: assuming that both the mean $m = E[X_k]$ and variance $\sigma^2 = E[(X_k - m)^2]$ are finite,

$$\frac{1}{\sigma\sqrt{n}} \sum_{1}^{n} (X_i - m)$$

converges in distribution to a standard normal random variable. In other words, for any pair of real numbers $a \leq b$

$$P(a \leq \frac{1}{\sigma\sqrt{n}} \sum_{1}^{n} (X_i - m) \leq b) \to \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx.$$

Perhaps you can see a hint of things to come here. If we want to move from discrete time values $t = 1, 2, \ldots$ to continuous time $0 \leq t$ by trying to cram lots of small independent changes into a fixed time interval, we will most likely find ourselves dealing with the normal density function. We will see how true this is when we talk about Brownian motion in the next chapter.

4.1 Problems

P.4.A Let X_n be the random variables of our 10-sided dice example, Section 1.3. Show that $X_n \Rightarrow X_1$ (convergence in distribution) but that the probability that $\lim_{n\to\infty} X_n$ exists is 0.

Chapter 5

Continuous Time

5.1 Introduction to Continuous-time Processes

Up to this point we have been considering stochastic processes indexed by a discrete time parameter: X_k where $k = 0, 1, 2, \ldots$. Now we want to use a continuous time parameter t where $0 \leq t \leq T$. The basic mathematical structure is the same as before. There should be an underlying set Ω consisting of all possible states of the world $\omega \in \Omega$. Each X_t is a function that assigns an outcome $X_t(\omega)$ to each ω, $X_t : \Omega \to \mathbb{R}$. For each $0 \leq t \leq T$ there should be a σ-algebra $\mathcal{F}_t$ which describes the information about the state of the world available at time t. Intuitively this means that the $\mathcal{F}_t$-measurable random variables are precisely those values will be known to us at time t. In particular X_s should be $\mathcal{F}_t$-measurable for all $s \leq t$. These should be nested, $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s \leq t$. As before this is described by saying that the stochastic process X_t is *adapted* to the filtration of $\mathcal{F}_t$. Note that the text writes "$X(t)$" while we use "X_t". Both are really functions of two variables: $X(t, \omega)$ or $X_t(\omega)$, but the reference to $\omega \in \Omega$ is typically omitted.

An alternate point of view is that a stochastic process X_t assigns to each ω a function of t: $f(t) = X_t(\omega)$. So you might think of a continuous-time stochastic process as a function-valued random variable. Naturally we will want to carry out calculus-like operations on this function of t. Were we had things like $\sum_{i=1}^{k} \phi_i \Delta M_i$ before we will now want to work with things like $\int_0^t \phi_s \, dM_s$. We will want to describe stochastic processes with formulas like

$$X_t = X_0 + \int_0^t \mu(t, X_t) \, dt + \int_0^t \sigma(t, X_t) \, dM_t.$$

See (3.6) in the text for instance. We have a fair bit of explaining to do about the second integral above. The point at the moment is that such operations typically require that X_t and M_t satisfy additional regularity properties in their dependence on t, such as continuity or differentiability. Previously we were concerned only with how each $X_t(\omega)$ depends on ω, one t at a time. That's what "adapted" refers to. Now we need to worry about the nature of the dependence of $X_t(\omega)$ on *both* t and ω together. These joint measurability issues are intricate and take more sophisticated techniques and ideas to handle adequately. We will have to simply gloss over most of those technicalities. We will be as accurate as possible in our statements of results, but remember that this is a user's guide or overview, not a rigorous treatment.

5.2 Brownian Motion

The single most important stochastic process in continuous time is Brownian motion. First, here is the definition[1]. A stochastic process W_t, $0 \leq t \leq T$, defined on some $(\Omega, \mathcal{F}, P)$ is called *Brownian motion* if the following are satisfied:

1) $W_0(\omega) = 0$ for (almost) all $\omega \in \Omega$;

2) For each pair $0 \leq s < t \leq T$, $W_t - W_s$ is independent of all W_u, $0 \leq u \leq s$;

[1] This is the same as Definition 4.1 in the text.

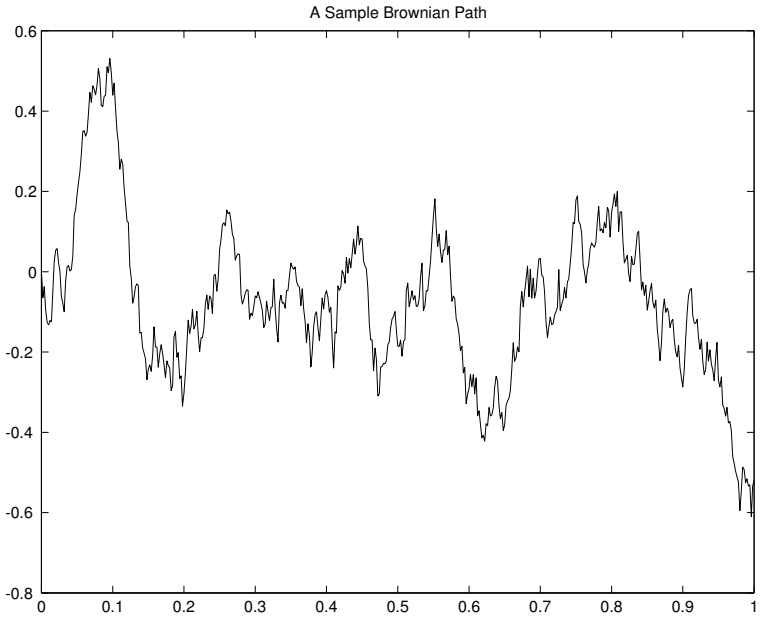
3) For each pair $0 \le s < t \le T$, $W_t - W_s$ has a normal distribution with mean 0 and variance $t - s$.

4) $W_t(\omega)$ is continuous in t for (almost) all $\omega \in \Omega$.

There are a number of other equivalent characterizations, but this is most natural for us.

Several observations should be made about this definition. First, the definition does not prescribe what the underlying $(\Omega, \mathcal{F}, P)$ is. Just as there can be many standard normal random variables, using different $(\Omega, \mathcal{F}, P)$ or different ways to associate $\omega \in \Omega$ with an outcome $Y(\omega)$ having the desired distribution, so there can be many Brownian motions. To say a stochastic process W_t is a Brownian motion is to prescribe what the probabilities of various events associated with W_t are, not what the underlying $(\Omega, \mathcal{F}, P)$ is.

Next consider part 4). For W_t to be continuous in t means that its value at one s is determined (as a limit) from its values at other nearby t. But for $s < t$, W_t and W_s differ by something that is independent of W_s. This makes part 4) rather amazing – that it is possible to have so many independent normal random variables fitted together to make a continuous function is remarkable. That is one reason Brownian motion is such a fascinating stochastic process. On the other hand, W_t is *only* continuous. You can see from the picture[2] below that a typical Brownian path is very irregular, although it is continuous. We will say more about this in §5.4.1.



Brownian motion can be obtained from a limit ($n \to \infty$) of increasingly refined binomial trees. Talking through this construction may help your appreciation of what Brownian motion is as well as why it might arise in applications. The idea is to consider a tree in which the time increment $\delta t = \frac{1}{n}$ is very small, and the size δx of the up/down jump between steps is likewise very small (exactly how small will be specified soon). To cover a time interval $0 \le t \le T$ with steps of size $\delta t = \frac{1}{n}$ it will take a total of $N = nT$ steps [3]. If $t = \frac{k}{n}$ for some integer $k \le N$, $W_t^{(n)}$ will be the state of our tree-process after k steps of size δt each. After one more time step we will have $W_{t+\delta t}^{(n)} = W_t^{(n)} \pm \delta x$, each with probability $\frac{1}{2}$. If we fix the value of t, then $W_t^{(n)}$ is the result of nt individual up/down jumps each of size δx. Thus as $n \to \infty$ but t remains fixed, $W_t^{(n)}$ will be the combined effect of a large number of independent random up/down jumps each of very small size. If we choose δx correctly we will be able to use the limit laws described in Chapter 4 to obtain the distribution of $W_t = \lim_{n\to\infty} W_t^{(n)}$.

[2] An approximate Brownian path as above is simple to generate with Matlab. Simply specify a value for `n` and then enter the command `plot(0:1/n:1,[0,cumsum(randn(1,n))/sqrt(n)])`.

[3] If nT is not an integer this presents a problem. The same problem occurs in the next paragraph if t is not a multiple of $1/n$. However by interpolating between the multiples of $1/n$ we can work around the difficulty. (5.1) below is the kind of thing we need to to use. We will not go through all the details. We can just assume T is an integer.

To make this look more like the the limit laws, let $X_1, X_2, \ldots$ be an infinite sequence of independent random variables each with $P(X_k = +1) = P(X_k = -1) = \frac{1}{2}$. X_k specifies whether the jump between $t = (k-1)\delta t$ and $t = k\delta t$ is up ($X_k = +1$) or down ($X_k = -1$). Thus we can write

$$W_t^{(n)} = W_{k\delta t}^{(n)} = \delta x \sum_{i=1}^{nt} X_i,$$

provided t is a multiple of $\delta t = \frac{1}{n}$. Now we want to pick δx so that we can apply one of our limit theorems to find $\lim_{n\to\infty} W_t^{(n)}$. One natural guess would be to try $\delta x = \frac{1}{n}$. Then the Strong Law of Large Numbers would apply:

$$W_t^{(n)} = \frac{1}{n}\sum_{i=1}^{nt} X_i = t \cdot \left(\frac{1}{nt}\sum_{i=1}^{nt} X_i\right) \to t \cdot m = 0,$$

where $m = E[X_i] = 0$. This is not very useful – we want a stochastic process in the limit, something with genuine random behavior. For that we need the Central Limit Theorem instead of the Law of Large Numbers. For this to work we take

$$\delta x = \frac{1}{\sqrt{n}}.$$

Then we get

$$W_t^{(n)} = \frac{1}{\sqrt{n}}\sum_{i=1}^{nt} X_i = \sqrt{t} \cdot \left(\frac{1}{\sqrt{nt}}\sum_{i=1}^{nt} X_i\right) \Rightarrow \sqrt{t}\, Y,$$

where Y is a standard normal random variable. So $W_t = \lim_{n\to\infty} W_t^{(n)}$ should be a normal random variable with mean 0 and variance t:

$$\begin{aligned} P(a \le W_t \le b) &= \lim_{n\to\infty} P(a \le W_t^{(n)} \le b) \\ &= P(a \le \sqrt{t}\, Y \le b) \\ &= P(\frac{a}{\sqrt{t}} \le Y \le \frac{b}{\sqrt{t}}) \\ &= \int_{a/\sqrt{t}}^{b/\sqrt{t}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\, dy \\ &= \int_a^b \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}\, dx. \end{aligned}$$

Strictly speaking, $W_t^{(n)}$ is only defined if $t = \frac{k}{n} = k\delta t$ for some k. As n changes the t-values for which $W_t^{(n)}$ is defined also change. This is awkward, but we can remedy it by using linear interpolation to define $W_t^{(n)}$ for t between multiples of δt. In other words we "connect the dots" to get the graph of $W_t^{(n)}$. This makes $W_t^{(n)}$ defined for *all* $0 \le t \le T$, and a continuous function of t, anticipating Part 4) of the definition. If t is not a multiple of δt, its difference from $W^{(n)}$ at the nearest multiple will be small: if $(m-1)\delta t \le t \le m\delta t$ then

$$\begin{aligned} |W_t^{(n)} - W_{(m-1)\delta t}^{(n)}| &\le \delta x \\ |W_t^{(n)} - W_{m\delta t}^{(n)}| &\le \delta x. \end{aligned} \tag{5.1}$$

We won't go through the details, but the upshot is that in the limit as $n \to \infty$ we can proceed as if all t were multiples of δt; the discrepancy is negligible as $n \to \infty$. (We have presumed this already by overlooking the possibility of t not a multiple of δk above.)

The W_t resulting from the above construction is Brownian motion, if the limit is understood in the appropriate sense. There are a number of technical issues about the construction that a rigorous treatment would need to address. We will mention some of them in the next subsection. But for now lets compare the W_t of our construction to the requirements of the definition of Brownian motion. Part 1) of the definition

is natural from our construction; $W_0^{(n)} = 0$ in our binary tree, so it is obvious that $W_t = 0$ in the limit. For part 2), consider $s = m\delta t < t = k\delta t$. Then $W_t^{(n)} - W_s^{(n)}$ depends on $X_{m+1}, \ldots, X_k$, while all $W_u^{(n)}$ for $u \le s$ depend only on $X_1, \ldots, X_m$. Since the X_i are all independent of each other, this makes it clear that $W_u^{(n)}$ for $u \le s$ is independent of $W_t^{(n)} - W_s^{(n)}$. The independence passes through to the limit as $n \to \infty$. Part 3) is a modest generalization of our earlier calculation:

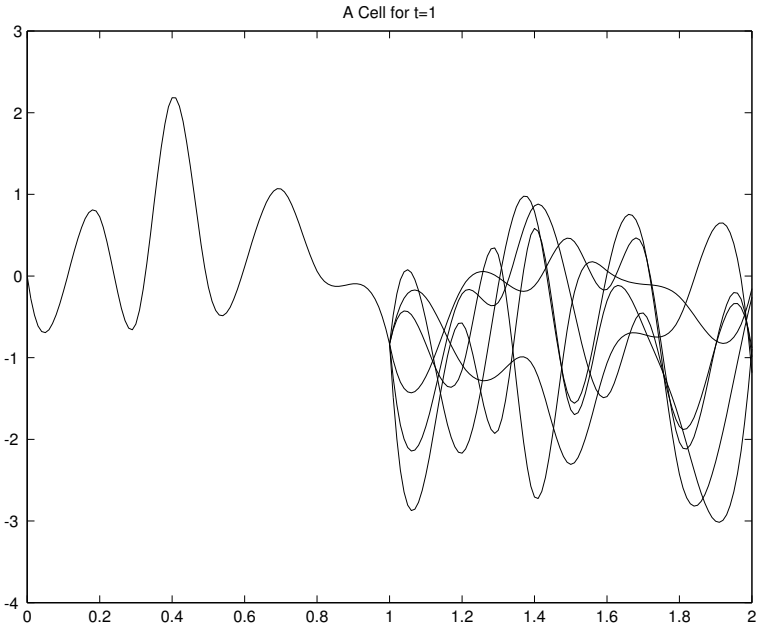
$$\begin{aligned} W_t^{(n)} - W_s^{(n)} &= \delta x \sum_{i=ns+1}^{nt} X_i \\ &= \sqrt{t-s}\left(\frac{1}{\sqrt{n(t-s)}} \sum_{i=1}^{n(t-s)} X_{ns+i}\right) \\ &\Rightarrow \sqrt{t-s}\, Y, \end{aligned}$$

where, by the Central Limit Theorem, Y is a standard normal random variable. So $W_t - W_s = \sqrt{t-s}\, Y$ is normal with mean 0 and variance $t - s$, as stated.

5.3 Theoretical Foundations

If you think carefully about our construction above, you will realize that we never constructed a stochastic process W_t at all! We only described what the *probabilities* associated with $\lim_{n\to\infty} W_t^{(n)}$ would be if the limit exists. There remains a fundamental question of whether there really does exist a stochastic process with all the properties of our definition. Such a process was first contemplated in the context of financial applications in the Ph.D. thesis of Bachelier [Ba] more than 100 years ago. The proof that it actually exists was first achieved by N. Wiener in 1923. As we have pointed out, the definition of Brownian motion does not determine the underlying Ω. However the most natural choice is simply the collection $\Omega = C([0,T])$ of all continuous functions $\omega : [0,T] \to \mathbb{R}$. Then $W_t(\omega) = \omega(t)$. (In this context, the P that Wiener proved to exist is called *Wiener measure*.) Should you want to pursue the issue issue of existence, or any of the many other results stated below, there are many possible references you could consult, such as [GS, Fr, Mc, RW].

We haven't yet said anything about the filtration $\mathcal{F}_t$. Of course W_t needs to be $\mathcal{F}_t$-measurable. The other essential feature is that the forward increments $W_t - W_s$ need to be independent of all $\mathcal{F}_s$-measurable random variables. Given a Brownian motion W_t defined on some (Ω, P), one way to obtain a suitable filtration is to take $\mathcal{F}_t$ to be the smallest σ-algebra with respect to which W_s is measurable for all $s \le t$. Intuitively this consists of those subsets of Ω which can be described in terms of $W_s(\omega)$, $s \le t$. (This is what the text is trying to describe as $\mathcal{F}_t^W$ in Definition 4.2 on page 39.) If you take Ω as the set of continuous paths $C([0,T])$ itself, then a set $A \in \mathcal{F}_t$ would consist of a union of "t-cells", where each t-cell consists of a bundle of continuous functions all having a common first part on $[0,t]$ followed by all possible continuous extensions on $[t,T]$. We have tried to illustrate such a cell in the following picture.



Another theoretical issue is whether the construction $W_t = \lim_{n\to\infty} W_t^{(n)}$ that we described above is really valid in some sense. The answer to this is provided by what is called *Donsker's Invariance Principle*. The answer is "yes" in a certain sense of convergence in distribution for continuous function-valued random variables. It says that any "continuous operation" applied to the paths of $W_t^{(n)}$ will converge (in distribution) to the same operation applied to the paths of W_t. For instance

$$\int_0^T W_t^{(n)}\, dt \Rightarrow \int_0^T W_t\, dt.$$

To make this precise we would need to explain what we mean by a "continuous operation" — again something that would take us too far afield.

5.4 Properties

Brownian motion has many remarkable and important properties. We summarize just a few of them below.

5.4.1 Irregularity

The definition of Brownian motion W_t says that the *sample paths*, $t \mapsto W_t$ are continuous, for all (or at least almost all) $\omega \in \Omega$. However, as the picture on page 27 suggests, they are rather ragged functions. For one thing, they are never differentiable in t:

$$P(\{\omega :\ \frac{d}{dt}W_t(\omega) \text{ exits for some } t\}) = 0.$$

So we can never talk about W_t' in the usual sense of $' = \frac{d}{dt}$. It is important to remember this when we encounter "dW_t" in expressions below; it will *not* mean $W_t'\, dt$ as you might expect from change of variable calculations in calculus.

Another aspect of the "raggedness" of Brownian paths concerns the set of crossing times of a specified level a:

$$T_a(\omega) = \{t \geq 0 :\ W_t(\omega) = a\}.$$

Although this is a random set, depending on ω, it does have certain properties with probability 1. Given any $t_a \in T_a$ and any $\epsilon > 0$, T_a will contain infinitely many t in $(t_a, t_a + \epsilon)$. In words, W_t oscillates so

frantically in the vertical direction that once it hits a level a it will hit it again infinitely many times in the next split-second. It is also true (with probability one) that T_a is unbounded. No matter where W_t is, it will always return to a again sometime after t.

Given the level a, we can define the *first* time that $W_t = a$ as

$$\tau_a = \inf\{t > 0 : \ W_t = a\}.$$

This is a time-valued random variable, and is a stopping time as defined in Section 3.3: $\{\omega : \ \tau_a \leq t\} \in \mathcal{F}_t$ for each t. Intuitively this is because whether $\tau_a \leq t$ or not is something we can answer by knowing just the values of W_s for $s \leq t$. (But beware random times defined in terms of the last time something happens. These will generally *not* be stopping times, because they are statements about the future rather than the past!)

5.4.2 Scaling

It is sometimes said that "Brownian motion looks the same on any scale." This is only true when interpreted correctly. If we take a nice, differentiable function $f(t)$ (lets say $f(0) = 0$ for simplicity) and look at its graph under a magnifying glass, we will see the graph of $y = \frac{1}{c}f(cs)$ where $\frac{1}{c}$ is the magnification factor. We are rescaling time and space by the same factor; a point (s, y) on the graph of $y = \frac{1}{c}f(cs)$ corresponds to a point $(t, x) = (cs, cy)$ on the graph of $x = f(t)$. For a differentiable function f with $f(0) = 0$, under high magnification (c close to 0) we will see essentially the graph of the tangent line at 0: $y = f'(0)s$.

Brownian motion will "look the same" only if we rescale time and space *differently*, namely if points (t, x) from the graph of $x = W_t$ are sent to points (s, y) on the rescaled graph so that $(t, x) = (cs, \sqrt{c}y)$. This corresponds to looking at the graph of $y = \frac{1}{\sqrt{c}}W_{cs}$. The basic fact is that, for any constant $c > 0$, if W_t is a Brownian motion, then $\tilde{W}_s = \frac{1}{\sqrt{c}}W_{cs}$ is also a Brownian motion, with s as the time variable. One can check the parts of the definition of Brownian motion directly to confirm this. As a consequence, if we look at the graph of Brownian motion under a magnifying class, what we see will not look like another Brownian motion, but $\frac{1}{\sqrt{c}}$ times Brownian motion. I.e. the vertical axis will be enlarged and the ragged nature of the graph will seem more pronounced.

There are other transformations of Brownian motion that result in new Brownian motions:

- $-W_t$,
- $W_{t+t_0} - W_{t_0}$, for any $t_0 \geq 0$;
- $tW_{1/t}$.

The last one is particularly interesting because it reverses the direction of the time axis (but nonlinearly)!

5.4.3 Calculating Expected Values

If W_t, $0 \leq t \leq T$ is a Brownian motion defined on $(\Omega, \mathcal{F}, P)$ and $A \in \mathcal{F}_T$ is a set that truly depends on W_t for infinitely many t-values, it can be quite challenging to determine the probability $P(A)$. However for probabilities determined only by the values of $W_{t_1}, W_{t_2}, \ldots, W_{t_n}$, for a pre-determined set of times $t_1, t_2, \ldots, t_n$, there is a formula. This will be all we need for simple conditional expectation calculations.

The basic pattern emerges if we calculate $E[\phi(W_s, W_t)]$ where ϕ is some function of two variables and $0 \leq s < t$. The key is to write $W_t = W_s + \Delta W$, where $\Delta W = W_t - W_s$. We know W_s and ΔW are *independent* normal random variables, both with mean 0 and variances, s and $t - s$ respectively. The density of W_s is $\frac{1}{\sqrt{2\pi s}}e^{-u^2/2s}$ and that of ΔW is $\frac{1}{\sqrt{2\pi(t-s)}}e^{-v^2/2(t-s)}$. So we can proceed using (2.10) as follows:

$$\begin{aligned} E[\phi(W_s, W_t)] &= E[\phi(W_s, W_s + \Delta W)] \\ &= \int\int \phi(u, u+v)\frac{1}{\sqrt{2\pi s}}e^{-u^2/2s} \ \frac{1}{\sqrt{2\pi(t-s)}}e^{-v^2/2(t-s)}\, dv\, du. \end{aligned}$$

Now make the change of variable to $w = v + u$ in the inside (dv) integral. This produces

$$E[\phi(W_s, W_t)] = \int\int \phi(u,w) \frac{1}{\sqrt{2\pi s}} e^{-u^2/2s} \; \frac{1}{\sqrt{2\pi(t-s)}} e^{-(w-u)^2/2(t-s)} \, dw \, du.$$

The function multiplying $\phi(u,w)$ inside the integral is the *joint* density of (W_s, W_t). To make this a little less cumbersome to write, lets introduce the *transition density* for Brownian motion:

$$p(s,x;t,y) = \frac{1}{\sqrt{2\pi(t-s)}} e^{-(y-x)^2/2(t-s)}.$$

Then our formula above is simply

$$E[\phi(W_s, W_t)] = \int\int \phi(u,w) \, p(0,0;s,u) p(s,u;t,w) \, dw \, du.$$

The same idea, worked out for more than two W_{t_i}, $0 < t_1 < t_2 < \ldots < t_n$, leads to the formula

$$E[\phi(W_{t_1}, W_{t_2}, \ldots, W_{t_n})] = \int \cdots \int \phi(x_1, x_2, \ldots, x_n) \prod_{i=1}^{n} p(t_{i-1}, x_{i-1}; t_i, x_i) \, dx_n \ldots dx_1, \tag{5.2}$$

with $t_0 = 0$ and $x_0 = 0$. Thus $\prod_{i=1}^n p(t_{i-1}, x_{i-1}; t_i, x_i)$ is the joint density of $(W_{t_1}, \ldots, W_{t_n})$.

The basic conditional expectation formula that we will need is

$$E[g(W_t) \mid \mathcal{F}_s] = h(W_s), \quad s < t,$$

where $h(x)$ is the function determined from $g(y)$ by

$$h(x) = \int g(y) p(s,x;t,y) \, dy. \tag{5.3}$$

You should view this as a version of (2.5) above: $p(s,x;t,y)$ plays the role of $f(x|z)$, and the role of $\bar{f}(z)$ is taken by $\prod_{i=1}^{n-1} p(t_{i-1}, x_{i-1}; t_i, x_i)$. (Think of z as the vector $(x_1, \ldots, x_{n-1})$.) The formula (2.6) generalizes likewise: $E[g(W_s, W_t) \mid \mathcal{F}_s] = h(W_s)$ where

$$h(x) = \int g(x,y) p(x,s;y,t) \, dy. \tag{5.4}$$

There is another approach to identifying probabilities and expectations associated with Brownian motion. We will say more about it when we talk about Chapter 4 of the text, but we can give a hint of it here. If you check, you will find that for all $t < T$ and $y \in \mathbb{R}$

$$\frac{\partial}{\partial t} p(t,x;T,y) + \frac{1}{2} \frac{\partial^2}{\partial x^2} p(t,x;T,y) = 0.$$

From here it should be no surprize that for a bounded continuous function $\phi(\cdot)$,

$$u(t,x) = \int p(t,x;T,y) \phi(y) \, dy$$

is the solution of the partial differential equation

$$u_t(t,x) + \frac{1}{2} u_{xx}(t,x) = 0 \text{ for } 0 \le t < T \text{ with } u(T,x) = \phi(x).$$

From what we said above,

$$E[\phi(W_T) \mid \mathcal{F}_t] = u(t, W_t).$$

In brief, the conditional probability given $\mathcal{F}_t$ is given by a function u of t, W_t which is the solution of a partial differential equation involving $u_t + \frac{1}{2} u_{xx}$. This connection between conditional expectations and PDEs runs quite deep. We have only scratched the surface. It comes out again in the Black-Scholes formula: what is defined as a conditional expectation is determined by solving a PDE.

5.4.4 Martingale

Brownian motion W_t is a martingale. This is simple to check using the conditional expectation formula (5.4) above. Suppose $0 \le s < t \le T$. Then

$$\begin{aligned} E[W_t \mid \mathcal{F}_s] &= \int y\, p(s, W_s; t, y)\, dy \\ &= W_s. \end{aligned}$$

Essentially this is another way of saying that $W_t - W_s$ has mean 0 and is independent of $\mathcal{F}_s$.

5.5 Problems

In the following you may find it helpful to take advantage of the facts about normal random variables from Problem P.1.B. Another vital fact about normal random variables is that if you take two of them, Z_1 and Z_2, which are independent, both normally distributed (but with possibly different means and variances), then a linear combination of them, $Z = c_1 Z_1 + c_2 Z_2$ for any constants c_i, is also a normal random variable. (This property makes the normal distribution what what called a *stable* distribution. Most probability distributions do not have this property.)

P.5.A In this problem you will carry out computational experiments to demonstrate some of the assertions we have made about the sequence $X_1, X_2, \ldots$ of independent random variables with common distribution

$$P(X_i = \pm 1) = \frac{1}{2}$$

as used in the construction of Brownian motion. You can do the required calculations with Matlab, Mathematica, Excel, or with any other software tool that you choose, but you *won't* want to attempt them by hand.

Start by using a random number generator to produce a realization $X_1, \ldots, X_{1000}$ of the first 1000 terms of such a sequence.[4] Please don't turn in the list of all 1000 values, but you should document how you generated it.

1. To consider the convergence of $\frac{1}{n}\sum_{i=1}^n X_i$ as $n \to \infty$, plot a graph of the values of $\frac{1}{n}\sum_{i=1}^n X_i$ for $n = 1, \ldots 1000$. Repeat for a second realization of the X_i. Discuss what your graphs illustrate about this convergence, and whether/how this agrees with what we said in class and in the Primer.

2. To consider the convergence of $\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$, plot a graph of the values of $\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$ for $n = 1, \ldots 1000$. Repeat for a second realization of the X_i. Discuss what your graphs illustrate about this convergence, and whether/how this agrees with what we said in class and in the Primer.

3. In this part you will consider the convergence of the *distribution* of $\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$ as $n \to \infty$. Here we are *not* computing and plotting a particular realization of $\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$, but the values of the *probabilities* associated with it. Consider $Z_i = \frac{1}{2}(1 + X_i)$ - these are independent random variables taking the values 0 or 1 with probability $\frac{1}{2}$ each. First explain why

$$\sum_{i=1}^n X_i = k \text{ is equivalent to } \sum_{i=1}^n Z_i = m$$

[4] In Matlab, `y=rand(1,1000);` will produce an array $y = [Y_1, \ldots, Y_{1000}]$ of random values, each uniformly distributed on $[0, 1]$, which means that $0 \le Y_i \le 1$ and for $0 \le a < b \le 1$, $P(a \le Y_i \le b) = b - a$. (An equivalent in Mathematica is `y=Table[Random[], i, 1000]`. In Excel you can accomplish the same thing by copying `=RAND()` into 1000 cells.) If you round Y_i off to the nearest integer (`round(...)` in Matlab, `Round[...]` in Mathematica, `=ROUND(...,0)` in Excel) the result will be random variables taking only the values 0 and 1, each with probability $\frac{1}{2}$. Now So $X_i = 2\,\text{round}(Y_i) - 1$ will produce the random values we want.

where $k = 2m - n$, and why the probability of this is given by

$$p_m = \frac{n!}{m!(n-m)!2^n}.$$

Thus the possible values of $\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$ are $x = \frac{2m-n}{\sqrt{n}}$ for $m = 0, \ldots n$, occurring with respective probabilities p_m. Now we want express the distribution of $\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$ with an approximate density $g_n(x)$ so that

$$P(\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i \le c) = \int_{-\infty}^c g_n(x)\,dx.$$

This is not true for all c, because $\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$ actually only takes a finite number of possible values. But we can find a $g_n(x)$ so that this works for all c of the particular form $c = \frac{2m-n}{\sqrt{n}}$ – take the value of p_m associated with $x = \frac{2m-n}{\sqrt{n}}$ and spread it out over the interval down to the next lowest such value of x, namely $\frac{2(m-1)-n}{\sqrt{n}}$, by giving $g_n(\cdot)$ the value $p_m/\frac{2}{\sqrt{n}}$ on that interval, so that

$$\int_{\frac{2(m-1)-n}{\sqrt{n}}}^{\frac{2m-n}{\sqrt{n}}} g_n(s)\,ds = p_m.$$

According to the Central Limit Theorem, what would you expect $\int_a^b g_n(s)\,ds$ to do as $n \to \infty$, at least if a, b can be assumed to have the same form as c? If we plot the values of $g_n(x) = p_m\sqrt{n}/2$ with respect to $x = \frac{2m-n}{\sqrt{n}}$ for $m = 0 \ldots n$, for a large value of n, what might you expect to see?

Take $n = 50$ and try it. Plot the values[5] of $g_n(x) = p_m\sqrt{n}/2$ with respect to the values of $x = \frac{2m-n}{\sqrt{n}}$ for $m = 0 \ldots n$ and compare what you get to a graph of what you speculated that it would converge to. (I don't recommend trying a value of n much larger than $n = 50$ unless you are careful about how you compute p_m, to avoid numerical overflow problems.)

4. For $n = 1000$ plot $W_t^{(n)} = \frac{1}{\sqrt{n}}\sum_{i=1}^{nt} X_i$ with respect to $t = \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}$. Repeat for a second realization of the X_i. Explain how this plot is different from what you did in part 2 above.

P.5.B Show, for any $0 < s < t$, $tW_{1/t} - sW_{1/s}$ is a normally distributed random variable with mean 0 and variance $t - s$.

P.5.C Show that $N_t = W_t^2 - t$ and (for any $\theta \in \mathbb{R}$) $M_t = e^{\theta W_t - \frac{1}{2}\theta^2 t}$ are both martingales.

[5] In Matlab, `nchoosek(n,m)` will give the value of the binomial coefficient $\frac{n!}{m!(n-m)!}$. In Mathematica, use `Binomial[n,m]`. In Excel, `=BINOMDIST(m,n,.5,FALSE)` will give the value of p_m, *including* the power of 2.

Chapter 6

Stochastic Integration

We will encounter many expressions of the form

$$\int_0^t \psi_s \, ds + \int_0^t \phi_s \, dW_s,$$

where ϕ_s and ψ_s are stochastic processes, adapted to $\mathcal{F}_t$. The $\int \psi_s \, ds$ is a conventional Riemann integral (as in calculus). $\int \phi_s \, dW_s$ is the *stochastic integral* of Itô, which we want to define.

6.1 About Integration in General

All integrals are defined in terms of limits. For instance the standard Riemann integral is defined by forming partitions $0 = s_0 < s_1 < s_2 < \ldots < s_n = t$ of $[0, t]$, with $s_i = i\,\Delta s$, $\Delta s = t/n$. For each partition pick some *sample points* $s_i \le \bar{s}_{i-1} \le s_i$ and then take the limit of the Riemann sums:

$$\int_0^t f(s) \, ds = \lim_{n\to\infty} \sum_{i=1}^{n} f(\bar{s}_i)\Delta s.$$

If $f(s)$ is nice, continuous for instance, then the limit is guaranteed to exist. As a function of the upper limit

$$g(t) = \int_0^t f(s) \, ds$$

is continuous. If $\psi_t(\omega)$ is a continuous, adapted stochastic process then $X_t(\omega) = \int_0^t \psi_s(\omega) \, ds$ will also be a continuous stochastic process, adapted to $\mathcal{F}_t$. Nothing but conventional integration is involved here.

If $h(s)$ is another function we can generalize this definition to define $\int_0^t f(s) \, dh(s)$ by replacing Δs with $\Delta h(s_i) = h(s_i) - h(s_{i-1})$. I.e. h is used to assign a ''size" to the intervals $[s_{i-1}, s_i]$. This leads to the definition of the Riemann-Stieltjes integral. If h is continuously differentiable, $\int_0^t f(s) \, dh(s)$ works out the be the same as $\int_0^t f(s)h'(s) \, ds$. But the definition still works for discontinuous h provided h has bounded variation:

$$\sum_{i=1}^{n} |h(s_i) - h(s_{i-1})| \le B < \infty$$

regardless of the partition of $[0, t]$.

We would like to define $\int_0^t \phi_s \, dW_s$ as a limit of Riemann sums. Since W_t is a martingale our construction (3.5) provides a nice precedent. But there is a big difficulty: Brownian paths have infinite (not finite) variation!

$$\sum_{i=1}^{n} |W_{s_i} - W_{s_{i-1}}| \to \infty \text{ as } n \to \infty.$$

This is a consequence of the following important result.

Lemma (Quadratic Variation)**.** *For $n = 2^m$, let $t_k^{(m)} = k\frac{T}{2^m}$. With probability 1,*

$$\lim_{m\to\infty} \sum_{i=1}^{2^m} (W_{t_i^{(m)}} - W_{t_{i-1}^{(m)}})^2 \to T.$$

This lemma is quite believable if you check that the sum $\sum(\Delta W_{t_i})^2$ in the limit has mean $= T$ and variance $= 2T\frac{T}{2^m} \to 0$. Also, notice that this implies our assertion about infinite variation, because

$$\sum_{i=1}^{2^m} (W_{t_i^{(m)}} - W_{t_{i-1}^{(m)}})^2 \le \Big(\max |W_{t_i^{(m)}} - W_{t_{i-1}^{(m)}}|\Big) \sum_{i=1}^{2^m} |W_{t_i^{(m)}} - W_{t_{i-1}^{(m)}}|.$$

The term with the maximum $\to 0$ since W_s is continuous. So the last term on the right must $\to \infty$.

The Quadratic Variation Lemma reveals another difficulty with attempting to define $\int_0^T \phi_t\, dW_t$ in the conventional way. Suppose we tried to define $\int_0^T W_t\, dW_t$ as a limit of Riemann sums

$$\lim_{n\to\infty} \sum W_{\bar{t}_i}(W_{t_i} - W_{t_{i-1}}),$$

where as usual any choice of $t_{i-1} \le \bar{t}_i \le t_i$ is to be allowed. If we compare the left sum ($\bar{t}_i = t_{i-1}$) to the right sum ($\bar{t}_i = t_i$), we find that

$$\text{right sum} \;-\; \text{left sum} \;= \sum (W_{t_i} - W_{t_{i-1}})^2 \to T.$$

Thus the left and right sums lead to different limits! This shows again that any definition of $\int_0^T \phi_t\, dW_t$ must be rather different than that of the Riemann integral $\int_0^T \phi_t\, dt$.

What will rescue all this is that the martingale property of W_t will provide an alternative to bounded variation. This comes out in (6.3) below, which comes from Doob's inequality, which is a martingale property. But we are getting ahead of ourselves.

6.2 Stochastic Integrals of Simple Integrands

The definition of stochastic integrals begins by considering a type of integrand ϕ_t for which we can see clearly what we want $\int_0^T \phi_t\, dW_t$ to be. We will call a stochastic process ϕ_t *simple* if it is piecewise constant, adapted and square-integrable. This means that there exist some $0 = t_0 < t_1 < \ldots < t_n = T$ and random variables $X_1, \ldots, X_n$ so that

- $\phi_t = X_i$ for t in $(t_{i-1}, t_i]$,
- each X_i is $\mathcal{F}_{t_{i-1}}$-measurable,
- each $E[X_i^2] < \infty$.

For such a ϕ_t we define

$$\int_0^T \phi_t\, dW_t = \sum_{i=1}^{n} X_i(W_{t_i} - W_{t_{i-1}}).$$

Notice that if ϕ_t is simple on $[0, T]$ then its restriction to any $[0, t]$ with $t \le T$ is also simple. So $\int_0^t \phi_s\, dW_s$ is defined for all $0 \le t \le T$. The properties of

$$M_t = \int_0^t \phi_s\, dW_s$$

as a stochastic process in its own right are very important:

1. $M_0 = 0$;

2. M_t is adapted and continuous in t;

3. M_t is a martingale;

4. $M_t^2 - \int_0^t \phi_u^2\, du$ is a martingale.

Suppose that t falls in the interval $t_k < t \le t_{k+1}$. Then we can write

$$M_t = \sum_{i=1}^{k-1} X_i(W_{t_i} - W_{t_{i-1}}) \quad + \quad X_k(W_t - W_{t_k}).$$

From this you should be able to convince yourself that M_t is indeed continuous in t, because W_t is. What about the martingale property? Considered at just at the discrete set of times t_i, we see that $N_i = M_{t_i}$ is a martingale, because it is just an instance of discrete stochastic integration as in (3.4). To check that $M_s = E[M_t|\ \mathcal{F}_s]$ when s and t are not among the t_i, observe that we can just insert them into the list of the t_i by adding some duplicate copies of some of the X_i, thereby re-expressing ϕ_t as a simple process with an enlarged set of t_i that now *does* include both s and t.

Now lets look at the fourth assertion above. (This is the most important part!) If $0 \le s < t \le T$, we want to show that

$$E[M_t^2 - \int_0^t \phi_u^2\, du \,|\, \mathcal{F}_s] = M_s^2 - \int_0^s \phi_u^2\, du$$

Since M_s and $\int_0^s \phi_u^2\, du$ are $\mathcal{F}_s$-measurable,

$$M_s^2 - \int_0^s \phi_u^2\, du = E[M_s^2 - \int_0^s \phi_u^2\, du \,|\, \mathcal{F}_s]$$

So what we need to show is that

$$E[M_t^2 - M_s^2 \,|\, \mathcal{F}_s] = E[\int_s^t \phi_u^2\, du \,|\, \mathcal{F}_s]. \tag{6.1}$$

As above, we can always add a few extra partition points to the t_i without altering the definition of M_t, so lets include s and t. Suppose then that

$$s = t_{k-1} < \cdots < t_m = t.$$

With our usual increment notation we can write

$$M_t^2 - M_s^2 = \sum_{i=k}^{m} \Delta(M_{t_i}^2). \tag{6.2}$$

Now observe several facts. First is the "discrete product rule" (Problem P.6.A):

$$\Delta(M_{t_i}^2) = 2M_{t_{i-1}}\Delta M_{t_i} + (\Delta M_{t_i})^2.$$

Next, calculate $E[\,\cdot \,|\, \mathcal{F}_s]$ of each term on the right in (6.2) using the tower rule: $E[\,\cdot \,|\, \mathcal{F}_s] = E[\,E[\,\cdot \,|\, \mathcal{F}_{t_{i-1}}]\,|\, \mathcal{F}_s]$. Since $E[\Delta M_{t_i} \,|\, \mathcal{F}_{t_{i-1}}] = 0$ (because M_t is a martingale), we find that

$$E[2M_{t_{i-1}}\Delta M_{t_i} \,|\, \mathcal{F}_{t_{i-1}}] = 2M_{t_{i-1}}E[\Delta M_{t_i} \,|\, \mathcal{F}_{t_{i-1}}] = 0.$$

Therefore

$$E[2M_{t_{i-1}}\Delta M_{t_i} \,|\, \mathcal{F}_s] = 0.$$

Since $(\Delta M_{t_i})^2 = X_i^2(\Delta W_{t_i})^2$, and X_i is $\mathcal{F}_{t_{i-1}}$-measurable,

$$E[(\Delta M_{t_i})^2 \,|\, \mathcal{F}_{t_{i-1}}] = X_i^2 E[(\Delta W_{t_i})^2 \,|\, \mathcal{F}_{t_{i-1}}] = X_i^2 \Delta t_i.$$

Therefore

$$E[\Delta(M_{t_i}^2) \,|\, \mathcal{F}_s] = E[X_i^2 \Delta t_i \,|\, \mathcal{F}_s].$$

Now we can put these pieces together:

$$E[M_t^2 - M_s^2 \mid \mathcal{F}_s] = E[\sum_{i=k}^{m} X_i^2 \Delta t_i \mid \mathcal{F}_s].$$

But now just notice that since ϕ_t is simple,

$$\sum_{i=k}^{m} X_i^2 \Delta t_i = \int_s^t \phi_u^2 \, du.$$

Writing this in above, we are done!

A consequence of what we just showed is that for a simple process ϕ_t

$$E\left[\left(\int_0^T \phi_t \, dW_t\right)^2\right] = E\left[\int_0^T \phi_t^2 \, dt\right].$$

If we combine this with Doob's L_2 martingale inequality (3.3), we find that

$$E\left[\max_{0\le t\le T}\left|\int_0^t \phi_u \, dW_u\right|^2\right] \le 4E\left[\int_0^t \phi_t^2 \, dt\right]. \tag{6.3}$$

This is significant for our goal of defining stochastic integrals for more general ϕ_t. What we hope to do is define $\int_0^T \phi_t \, dW_t$ for a non-simple ϕ_t by writing it as a limit, $\phi_t = \lim_{n\to\infty} \phi_t^{(n)}$ where each $\phi_t^{(n)}$ is simple, and then take

$$\int_0^T \phi_t \, dW_t = \lim_{n\to\infty} \int_0^T \phi_t^{(n)} \, dW_t$$

as the definition. The inequality (6.3) suggests the sense in which we would want the $\phi_t^{(n)}$ to converge to ϕ_t: if

$$E[\int_0^T (\phi_t - \phi_t^{(n)})^2 \, dt] \to 0. \tag{6.4}$$

then we would have that $\int_0^t \phi_u^{(n)} \, dW_u$ converges *uniformly* in t for (almost) all $\omega \in \Omega$ along a subsequence, which is gratifying because it tells us that the limit will be continuous in t.

6.3 An Example

As an example of the above strategy for defining stochastic integrals of non-simple ϕ_t, we will work out the case of $\phi_t = W_t$:

$$\int_0^t W_s \, dW_s.$$

First we need a sequence $\phi_t^{(n)}$ of simple processes approximating W_t in the sense of (6.4). For that we just make a piecewise-constant approximation by freezing W_t at the start of each time interval of length $\frac{1}{n}$:

$$\phi_s^{(n)} = W_{\frac{k}{n}} \text{ for } \frac{k}{n} < s \le \frac{k+1}{n}.$$

The notation "$\lfloor ns \rfloor/n$" is sometimes used to refer to the value $\frac{k}{n}$ for which $\frac{k}{n} \le s < \frac{k+1}{n}$. Using it we can write $\phi_s^{(n)} = W_{\lfloor ns \rfloor/n}$. Each $\phi_s^{(n)}$ is a simple process. Next we need to check that $E[\int_0^t (W_s - \phi_s^{(n)})^2\, ds] \to 0$.

$$\begin{aligned} E[\int_0^t (W_s - \phi_s^{(n)})^2\, ds] &= \int_0^t E[(W_s - \phi_s^{(n)})^2]\, ds \\ &= \int_0^t E[(W_s - W_{\lfloor ns \rfloor/n})^2]\, ds \\ &= \int_0^t s - \lfloor ns \rfloor/n\, ds \\ &\le \int_0^t \frac{1}{n}\, ds = \frac{t}{n} \to 0. \end{aligned}$$

Therefore our strategy leads to

$$\begin{aligned} \int_0^t W_s\, dW_s &= \lim_{n\to\infty} \int_0^t \phi_s^{(n)}\, dW_s \\ &= \lim_{n\to\infty} \left(\sum_1^N W_{\frac{k-1}{n}} \Delta W_{\frac{k}{n}} \quad + W_{\frac{N}{n}} (W_t - W_{\frac{N}{n}}) \right), \end{aligned}$$

where $\Delta W_{\frac{k}{n}} = W_{\frac{k}{n}} - W_{\frac{k-1}{n}}$ and N is the integer for which $\frac{N}{n} \le t < \frac{N+1}{n}$. Now, to evaluate the sum, the discrete product rule says that

$$W_{\frac{k-1}{n}} \Delta W_{\frac{k}{n}} = \frac{1}{2}[\Delta(W_{\frac{k}{n}}^2) - (\Delta W_{\frac{k}{n}})^2].$$

Writing $\Delta W_t = W_t - W_{\frac{N}{n}}$ to abbreviate the last term, we have

$$\begin{aligned} \int_0^t \phi_s^{(n)}\, dW_s &= \frac{1}{2} \left\{ \sum_1^N [\Delta(W_{\frac{k}{n}}^2) - (\Delta W_{\frac{k}{n}})^2] + \Delta(W_t^2) - (\Delta W_t)^2 \right\} \\ &= \frac{1}{2} \left\{ W_t^2 - W_0^2 - \left(\sum_1^N (\Delta W_{\frac{k}{n}})^2 + (\Delta W_t)^2 \right) \right\}. \end{aligned}$$

The quadratic variation lemma now tells us the result of taking $n \to \infty$ on both sides (at least along a subsequence):

$$\int_0^t W_s\, dW_s = \frac{1}{2}(W_t^2 - t). \tag{6.5}$$

Contrast this with the conventional integral $\int_0^t s\, ds = \frac{1}{2}t^2$.

6.4 The General Definition and Properties

The approach to defining $\int_0^t \phi_u\, dW_u$ in general is to find a sequence of simple $\phi_t^{(n)}$ which converge to the desired ϕ_t in the sense of (6.4), and then define (in L^2)

$$\int_0^t \phi_t\, dW_t = \lim_{n\to\infty} \int_0^t \phi_t^{(n)}\, dW_t, \quad \text{in } L^2. \tag{6.6}$$

There are a number of technical issues associated with this definition. First, for what ϕ_t does a simple approximating sequence of $\phi_t^{(n)}$ exist? It is necessary that

$$E\left[\int_0^T \phi_u^2\, du \right] < \infty, \tag{6.7}$$

but there is also a more subtle measurability requirement. Just being adapted is not enough. ϕ_t needs to satisfy a more stringent requirement of being *progressively measurable*. (Björk passes over this technicality.) We won't try to explain what that is, except to say that if ϕ_t is adapted and continuous in t, then it *is* progressively measurable. Then one needs to be precise about the sense of convergence in (6.6), and show that the limit is the same no matter what simple approximating sequence of $\phi_t^{(n)}$ is used. Having resolved those issues, one would proceed to prove that the resulting $M_t = \int_0^t \phi_u \, dW_u$ has the same properties 1 — 4 that we listed above in the case of a simple ϕ_t.

1. $M_0 = 0$;
2. M_t is adapted and continuous in t;
3. M_t is a martingale;
4. $M_t^2 - \int_0^t \phi_u^2 \, du$ is a martingale.

This includes the assertions of Proposition 4.4, page 41 of the text. In addition we should point out that if ϕ_t and ψ_t both satisfy (6.7), and c_1, c_2 are any constants then

$$\int_0^t c_1\phi_u + c_2\psi_u \, dW_u = c_1 \int_0^t \phi_u \, dW_u + c_2 \int_0^t \psi_u \, dW_u,$$

as you would expect from any integral.

There is usually a second extension of the definition of stochastic integrals, in which the requirement (6.7) is relaxed to ask only that

$$\int_0^T \phi_u^2 \, du < \infty \tag{6.8}$$

for (almost) all $\omega \in \Omega$; see (4.15) in the text. Then $M_t = \int_0^t \phi_u \, dW_u$ is still defined and continuous, but the two martingale properties 2 and 3 (i.e. (4.12) and (4.13) in the text) may fail. M_t becomes what is called a *local martingale*. This makes $\int_0^t \phi_u \, dW_u$ defined whenever ϕ_t is adapted and continuous in t for instance, but if we need to use the martingale properties 3 or 4 of M_t, we usually need to find a way to verify (6.7).

6.5 Itô's Formula

Although the Riemann integral of calculus is defined by a limiting procedure, we rarely use the definition itself for calculations. Instead most calculations are based on the Chain Rule: if $x'(t) = \phi(t)$ and $f(x)$ is a continuously differentiable function, then

$$\frac{d}{dt} f(x(t)) = f'(x(t))\phi(t)$$

so that

$$f(x(b)) = f(x(a)) + \int_a^b f'(x(t))\phi(t) \, dt.$$

For stochastic integrals likewise: most calculations are not based on the definition described above, but on Itô's formula, which plays the role of the Chain Rule for stochastic integrals. There are several versions of this important formula; see Propositions 4.11 and 4.16 in the text. We will not attempt to prove it, but will try to motivate it with some special cases.

The stochastic processes we want to work with are all ones that can be written as sums of Riemann and stochastic integrals, i.e. X_t for which

$$X_t = X_0 + \int_0^t \mu_s \, ds + \int_0^t \sigma_s \, dW_s,$$

for some progressively measurable stochastic processes μ_t and σ_t. This is what is meant when we write

$$dX_t = \mu_t \, dt + \sigma_t \, dW_t.$$

In this notation, the conventional Chain Rule says that if $dX_t = \mu_t\,dt$ and $f(x)$ is a continuously differentiable function, then $df(X_t) = f'(X_t)\mu_t\,dt$. Itô's formula will provide the same kind of result for cases in which dX_t includes a dW_t term as well.

To see a simple instance of Itô's formula, consider two stochastic processes defined by

$$X_t = \mu_X t + \sigma_X W_t, \quad Y_t = \mu_Y t + \sigma_Y W_t,$$

where $\mu_X, \mu_Y, \sigma_X, \sigma_Y$ are constants. In other words

$$dX_t = \mu_X\,dt + \sigma_X\,dW_t, \quad dY_t = \mu_Y\,dt + \sigma_Y\,dW_t. \tag{6.9}$$

Now consider the new stochastic process obtained by simply multiplying these together:

$$X_tY_t = \sigma_X\sigma_Y\,W_t^2 + (\sigma_X\mu_Y + \sigma_Y\mu_X)\,tW_t + \mu_X\mu_Y\,t^2.$$

By our example (6.5) above we know that

$$dW_t^2 = dt + 2W_t\,dW_t.$$

You also know from Problem P.6.C that

$$d(tW_t) = W_t\,dt + t\,dW_t.$$

The Chain Rule itself says that $d(t^2) = 2t\,dt$. Putting these together, it follows that

$$\begin{aligned} d(X_tY_t) &= \sigma_X\sigma_Y[dt + 2W_t\,dW_t] + (\sigma_X\mu_Y + \sigma_Y\mu_X)[W_t\,dt + t\,dW_t] + \mu_X\mu_Y\,2tdt \\ &= (\mu_X t + \sigma_X W_t)(\mu_Y\,dt + \sigma_Y\,dW_t) + (\mu_Y t + \sigma_Y W_t)(\mu_X\,dt + \sigma_X\,dW_t) + \sigma_X\sigma_Y\,dt \\ &= X_t\,dY_t + Y_t\,dX_t + \sigma_X\sigma_Y\,dt. \end{aligned}$$

The first two terms in the last line resemble the usual product rule, but the extra dt term on the end is perhaps not what we might have guessed. However the above *does* follow from a *stochastic product rule*:

$$d(X_tY_t) = X_t\,dY_t + Y_t\,dX_t + dX_t dY_t, \tag{6.10}$$

where we compute $dX_t dY_t$ using (6.9) and the basic differential product formulas

$$(dt)^2 = 0, \quad dt \cdot dW_t = 0, \quad (dW_t)^2 = dt. \tag{6.11}$$

We have verified (6.10) assuming $\mu_X, \mu_Y, \sigma_X, \sigma_Y$ are constants in (6.9). It turns out that (6.10) remains true if we replace these constants by stochastic processes[1] $\mu_t^X, \mu_t^Y, \sigma_t^X, \sigma_t^Y$.

Notice that both our formulas $dW_t^2 = dt + 2W_t\,dW_t$ and $d(tW_t) = W_t\,dt + t\,dW_t$ are examples of (6.10). Applying (6.10), we can build up many more stochastic integration/differential formulas. For instance,

$$\begin{aligned} dW_t^3 &= d(W_t\,W_t^2) \\ &= W_t\,d(W_t^2) + W_t^2\,dW_t + d(W_t^2)\cdot dW_t \\ &= W_t[dt + 2W_t\,dW_t] + W_t^2\,dW_t + [dt + 2W_t\,dW_t]\cdot dW_t \\ &= 3W_t\,dt + 3W_t^2\,dW_t. \end{aligned}$$

In general you can show by induction that if X_t has a stochastic differential, $dX_t = \mu_t\,dt + \sigma_t\,dW_t$, then for any positive integer n,

$$dX_t^n = nX_t^{n-1}\,dX_t + \frac{n(n-1)}{2}X_t^{n-2}\,(dX_t)^2. \tag{6.12}$$

The left side is $df(X_t)$ where $f(x) = x^n$. The two functions of X_t appearing on the right side are $f'(x) = nx^{(n-1)}$ and $\frac{1}{2}f''(x) = \frac{n(n-1)}{2}x^{(n-2)}$. This is an instance of *Itô's formula for functions of a single variable*:

$$df(X_t) = f'(X_t)\,dX_t + \frac{1}{2}f''(X_t)\,(dX_t)^2, \tag{6.13}$$

[1] This is provided the processes are progressively measurable and $\int_0^T(\sigma_t^X)^2\,dt$, and $\int_0^T(\sigma_t^Y)^2\,dt$ are finite.

provided $f''(x)$ is continuous. (We have essentially derived it for polynomials. The general case can be proved by finding a sequence of polynomials that converge to $f(x)$ in an appropriate sense.) You can check that for $f(x) = x^2$, (6.13) reproduces our formula for $d(W_t^2)$.

Suppose $f(x_1, \ldots, x_n)$ is a function of several variables, with all its second order partial derivatives continuous, and we have several stochastic processes $X_t^{(i)}$ each with stochastic differentials. For brevity, lets write $f(X_t^{(\cdot)})$ for $f(X_t^{(1)}, \ldots, X_t^{(n)})$. *Itô's formula for functions of several variables* says that

$$df(X_t^{(\cdot)}) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(X_t^{(\cdot)})\, dX_t^{(i)} + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_t^{(\cdot)})\, (dX_t^{(i)}\, dX_t^{(i)}). \tag{6.14}$$

If you use our power rule (6.12) with the stochastic product rule (6.10), you can check that (6.14) holds for the particular case of $f(x, y) = x^n y^m$. Several comments are needed here:

- If you compare (6.14) with Proposition 3.18 in the text, you will notice that we don't have a $\frac{\partial f}{\partial t}$ term. However if we just add an extra component $X_t^{(0)} = t$ in (6.14), we recover the text's version.
- The double sum $\sum_{i,j=1}^{n}$ is a sum of a sum:

$$\sum_{i,j=1}^{n} \ldots = \sum_{i=1}^{n} \left[\sum_{j=1}^{n} \ldots \quad \right].$$

  This means that the "mixed" second order partial derivatives ($i \neq j$) will each occur *twice*. For instance $\frac{\partial^2 f}{\partial x_1 \partial x_2}$ occurs once for $i = 1, j = 2$ and again for $i = 2, j = 1$. The "diagonal" terms ($i = j$) only occur once.
- The stochastic integrals involved in Itô's formula for $df(X_t^{(\cdot)})$ are only promised to be in the sense of (6.8). This means that if we want to use the martingale property for any of the $\int_0^t \ldots dW_t$ that arise, then we will need to find a way to verify (6.7) directly, which will depend on knowing something about the $\frac{\partial f}{\partial x_i}$ in particular.

As an example, consider

$$S_t = e^{(\alpha - \frac{1}{2}\sigma^2)t + \sigma W_t}.$$

Then $S_t = f(X_t)$ where $f(x) = e^x$ and $dX_t = (\alpha - \frac{1}{2}\sigma^2)\, dt + \sigma\, dW_t$. Itô's formula tells us that

$$\begin{aligned} dS_t &= e^{X_t} dX_t + \frac{1}{2} e^{X_t}\, (dX_t)^2 \\ &= S_t \left[(\alpha - \frac{1}{2}\sigma^2)\, dt + \sigma\, dW_t \right] + \frac{1}{2}\sigma^2 S_t\, dt \\ &= \alpha S_t\, dt + \sigma S_t\, dW_t. \end{aligned}$$

(This will describe the stock price in the basic Black-Scholes model.) We would like to know that the stochastic integral $\int_0^t \sigma S_u\, dW_u$ is a martingale. For that we would like to verify that the following is finite:

$$\begin{aligned} E\left[\int_0^T (\sigma S_t)^2\, dt \right] &= \sigma^2 \int_0^T E[e^{2(\alpha - \frac{1}{2}\sigma^2)t + 2\sigma W_t}]\, dt \\ &= \sigma^2 \int_0^T e^{2(\alpha - \frac{1}{2}\sigma^2)t} E[e^{2\sigma W_t}]\, dt \\ &= \sigma^2 \int_0^T e^{2(\alpha - \frac{1}{2}\sigma^2)t} e^{\frac{1}{2}(2\sigma)^2 t}\, dt < \infty. \end{aligned}$$

(To obtain the last line we have used (1.7) with $\theta = 2\sigma$.) We have been thinking of α and σ as constants through this calculation. But what we have said generalizes if α_t and σ_t are progressively measurable

processes, with some additional hypotheses to help us through the $E[\int_0^T (\ldots)^2\,dt] < \infty$ calculation above. If α_t and σ_t are bounded for instance, then everything works out fine.

Notice that if we take $\sigma = \theta$ and $\alpha = 0$ in S_t, then S_t reduces to $N_t = e^{\theta W_t - \frac{1}{2}\theta^2 t}$, our familiar exponential martingale. Also, $dN_t = \theta N_t\,dW_t$, which explains why we should expect it to be a martingale. For this case the Novikov Condition in Section 7.2 below provides an alternative to the second moment calculation above.

As a second example, we will compute the stochastic differential for

$$Y_t = \sin(W_t^2)W_t.$$

The simplest approach would be to apply Itô's formula for functions of a single variable to $f(x) = x\sin(x^2)$. However we want to illustrate (6.14). With

$$f(x_1, x_2) = \sin(x_1)x_2,$$

and

$$\begin{aligned} X_t^{(1)} &= W_t^2, \quad dX_t^{(1)} = 2W_t\,dW_t + dt; \\ X_t^{(2)} &= W_t, \quad dX_t^{(2)} = dW_t, \end{aligned}$$

we have

$$Y_t = f(X_t^{(1)}, X_t^{(2)}),$$

to which we apply Itô's formula. We will use f_{x_i} and f_{x_i,x_j} to denote the various partial derivatives. These will always be assumed to be evaluated at $X_t^{(1)}, X_t^{(2)}$ in the dY_t calculation below. Itô's formula says that

$$dY_t = f_{x_1}\,dX_t^{(1)} + f_{x_2}\,dX_t^{(2)} + \frac{1}{2}\left[f_{x_1,x_1}(dX_t^{(1)})^2 + 2f_{x_1,x_2}dX_t^{(1)}dX_t^{(2)} + f_{x_2,x_2}(dX_t^{(2)})^2\right].$$

Here are the various pieces:

$$\begin{aligned} f_{x_1} &= \cos(x_1)x_2 \\ f_{x_2} &= \sin(x_1) \\ f_{x_1x_1} &= -\sin(x_1)x_2 \\ f_{x_1x_2} &= \cos(x_1) \\ f_{x_2x_2} &= 0 \\ (dX_t^{(1)})^2 &= 4W_t^2\,dt \\ dX_t^{(1)}dX_t^{(2)} &= 2W_t\,dt \end{aligned}$$

Now assemble the pieces and make the substitutions for $X_t^{(1)}$ and $X_t^{(2)}$ to obtain

$$\begin{aligned} dY_t &= \cos(W_t^2)W_t\,(2W_t\,dW_t + dt) + \sin(W_t^2)\,dW_t + \frac{1}{2}\left[-\sin(W_t^2)W_t\,4W_t^2 + 2\cos(W_t^2)2W_t\right]dt \\ &= \left[\cos(W_t^2)2W_t^2 + \sin(W_t^2)\right]\,dW_t + \left[3\cos(W_t^2)W_t - 2\sin(W_t^2)W_t^3\right]\,dt. \end{aligned}$$

6.6 Martingale Representation Theorem

There are two additional results about stochastic integration that we need to mention. First is what we might call the *uniqueness of stochastic differentials*. If

$$\int_0^t \alpha_t\,dt + \int_0^t \sigma_t\,dW_t = \int_0^t \bar{\alpha}_t\,dt + \int_0^t \bar{\sigma}_t\,dW_t,$$

or, said another way if $\alpha_t\,dt + \sigma_t\,dW_t = \bar{\alpha}_t\,dt + \bar{\sigma}_t\,dW_t$, can we be sure that $\alpha_t = \bar{\alpha}_t$ and $\sigma_t = \bar{\sigma}_t$? The answer is as we would hope, essentially "yes." The only possible differences between $\alpha_t, \bar{\alpha}_t$ and between $\sigma_t, \bar{\sigma}_t$ are ones that do not affect their integrals. So for all practical purposes, stochastic differentials *are* unique.

The second question is exactly what random variables Y can be written as stochastic integrals

$$Y = \int_0^T \phi_t \, dW_t$$

for some ϕ_t satisfying (6.7)? We already know that it is necessary that $E[Y^2] < \infty$ and $E[Y] = 0$. The converse is true if, in addition, we assume that $Y(\omega)$ depends *only* on the values of $W_t(\omega)$, $0 \le t \le T$. The precise way to say this is that Y is measurable with respect to the σ-algebra $\mathcal{F}_T^W$ of Definition 3.1; see also §5.3 above. The thing that must be ruled out is that Y might depend on ω in some way that is not discernible in the values of $W_t(\omega)$. As long as $E[Y^2] < \infty$, $E[Y] = 0$, and Y is measurable in terms of the W_t, $0 \le t \le T$ alone then there will be a ϕ_t with $Y = \int_0^T \phi_t \, dW_t$. This is often called the *Martingale Representation Theorem* for Brownian motion, because if $Y = M_T - M_0$ where M_t is a martingale (with respect to $\mathcal{F}_t^W$), then $M_T = \int_0^T \phi_u \, dW_u + M_0$ which implies that $M_t = \int_0^t \phi_u \, dW_u + M_0$ for all $0 \le t \le T$. In terms of stochastic differentials,

$$dM_t = 0 \, dt + \phi_t \, dW_t.$$

This is very useful, because it makes recognizing martingales quite easy in terms of their stochastic differentials. Any process X_t, $0 \le t \le T$, with a stochastic differential

$$dX_t = \alpha_t \, dt + \sigma_t \, dW_t$$

is a martingale only if $\alpha_t = 0$. Conversely if $\alpha_t = 0$ then X_t will be a martingale if we can somehow verify integrability: $E[|X_T|] < \infty$. A sufficient condition is that $E[\int_0^T \sigma_s^2 \, ds] < \infty$. Indeed that implies square integrability because $E[X_T^2] = E[\int_0^T \sigma_s^2 \, ds] < \infty$.

6.7 Problems

P.6.A If $x_0, x_1, \ldots$ is a sequence of real numbers, let $\Delta x_i = x_i - x_{i-1}$ denote the usual backward difference ($i \ge 1$). If x_i, y_i are two such sequences, verify the "discrete product rule":

$$\Delta(x_i y_i) = x_{i-1} \Delta y_i + y_{i-1} \Delta x_i + (\Delta x_i)(\Delta y_i).$$

P.6.B Check the mean and variance calculations mentioned in the first sentence after the Quadratic Variation Lemma.

P.6.C Compute the stochastic integral $\int_0^T t \, dW_t$. (For each n describe a partition $t_i^{(n)}$ of $[0, T]$. Let $\phi_t^{(n)} = t_{k-1}^{(n)}$ on $(t_{k-1}^{(n)}, t_k^{(n)}]$. Explain why $E[\int_0^T (\phi_t^{(n)} - t)^2 \, dt] \to 0$. Use Problem P.6.A to rewrite $\int_0^T \phi_t^{(n)} \, dW_t$ and take the limit to determine $\int_0^T t \, dW_t$.)

P.6.D Use Itô's formula to verify that $e^{\frac{1}{2}\theta^2 t} \sin(\theta W_t)$ is a martingale for any $\theta \in \mathbb{R}$.

P.6.E

a) Verify the following formula, for each integer $n \ge 1$:

$$\int_0^t s^n \, dW_s = t^n W_t - \int_0^t n s^{n-1} W_s \, ds$$

b) Find a similar formula for $\int_0^t W_s^n \, dW_s$.

P.6.F Write out the induction argument to verify (6.12). Verify the statement following (6.14) for $f(x, y) = x^n y^m$.

P.6.G For ordinary integration the n-fold iterated integrals of $f(t) = 1$ are the familiar monomials from Taylor series:

$$\int_0^t \int_0^{s_n} \int_0^{s_{n-1}} \cdots \int_0^{s_2} 1 \, ds_1 \cdots ds_{n-1} \, ds_n = t^n/n! \, .$$

The analogous formula for stochastic integrals is different. Verify that the following formulas are correct and that each is square integrable.

$$\int_0^t \int_0^{s_2} 1 \, dW_{s_1} dW_{s_2} = \frac{1}{2}(W_t^2 - t)$$

$$\int_0^t \int_0^{s_3} \int_0^{s_2} 1 \, dW_{s_1} dW_{s_2} dW_{s_3} = \frac{1}{6}(W_t^3 - 3tW_t)$$

$$\int_0^t \int_0^{s_4} \int_0^{s_3} \int_0^{s_2} 1 \, dW_{s_1} dW_{s_2} dW_{s_3} dW_{s_4} = \frac{1}{24}(W_t^4 - 6tW_t^2 + 3t^2)$$

P.6.H Use the last formula of P.6.G to compute $E[W_t^4]$. You should get the same result as Example 4.12, page 51, of the text.

Chapter 7

Change of Measure and the Black-Scholes Model

In our binomial tree calculations we saw that our approach to pricing a financial option could be thought of as changing the original probability measure P to a new probability measure Q which made the discounted stock price, S_t/B_t a martingale. This approach works out nicely in the continuous time setting too, but it takes some more effort to explain.

7.1 Motivation and Simple Example

To set the stage, lets first consider making a change of probability measure in a simple setting: $\Omega = \mathbb{R}$ and $X(\omega) = \omega$. Suppose P and Q are probability themeasures on Ω which assign probabilities

$$P([a,b]) = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}\, dx \text{ and}$$
$$Q([a,b]) = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$$

to each interval $[a,b]$. With respect to P, X is a normal random variable with mean 0 and variance σ^2. But if considered with respect to Q, then X has mean μ. (The variance is still σ^2 with respect to Q, but we could have made it something different had we wanted to.) The point is that by changing from P to Q we change the probabilities associated with X, so that its distribution changes, even though $X : \Omega \to \mathbb{R}$ has not changed at all. We can describe this change of measure in terms of a special random variable ζ:

$$\zeta = e^{(X^2-(X-\mu)^2)/2\sigma^2} = e^{(2\mu X-\mu^2)/2\sigma^2}. \tag{7.1}$$

Every Q-expectation can be written as a P-expectation with an extra factor of ζ thrown in:

$$E^Q[\phi(X)] = E^P[\phi(X)\,\zeta].$$

(If you write out both sides as integrals you will see that they are the same.) In particular, for any (measurable) set $A \subseteq \Omega$,

$$Q[A] = E^Q[1; A] = E^P[\zeta; A].$$

In general, if we start with a probability measure P on some $(\Omega, \mathcal{F})$ and a random variable $\zeta \geq 0$ with $E^P[\zeta] = 1$, then we can construct a new probability measure Q on the same $(\Omega, \mathcal{F})$ by the formula $Q[A] = E^P[\zeta; A]$. When two probability measures are related to each other in this way we say that Q is *absolutely continuous* with respect to P (usually written $Q \ll P$), and ζ is called the *Radon-Nikodym derivative* of Q with respect to P:

$$\zeta = \frac{dQ}{dP}.$$

(This notation makes the connection between P and Q look natural in terms of the notation one would use in real analysis: $\int_A dQ = \int_A \frac{dQ}{dP}\, dP$.) Note that if $P(A) = 0$ then we must also have $Q(A) = E^P[\zeta; A] = 0$. The Radon-Nikodym Theorem says that the converse of this is true: if $Q(A) = 0$ whenever $P(A) = 0$, then $Q \ll P$. When both $Q \ll P$ and $P \ll Q$ are true then we say the P and Q are *equivalent*. (But this does *not* mean that they are the same in any sense.)

Next, suppose W_t is a Brownian motion defined on $(\Omega, \mathcal{F}_t, P)$. We want to do something like the above in this context – change from P to a new probability measure Q with respect to which W_t changes from Brownian motion to something else. If we want to put W_t in the role of X above, then we should take $\sigma^2 = t$. Then if we take $\mu = \gamma t$ (γ some constant) then we get something familiar from (7.1):

$$\zeta_t = e^{\gamma W_t - \frac{1}{2}\gamma^2 t},$$

our basic exponential martingale. If $\frac{dQ}{dP} = \zeta_t$ then with respect to Q, W_t would have mean γt and variance t, i.e. it would be as if $W_t = \gamma t + \tilde{W}_t$ where $\tilde{W}_t$ is a Brownian motion with respect to Q. Now you might object because ζ_t is different for each t value, so it seems we are using a different Q for each t. But the fact that ζ_t is a martingale (with respect to P) resolves that: for any $t \le T$ we have

$$\begin{aligned} E^P[\phi(W_t)\zeta_t] &= E^P[\phi(W_t)E^P[\zeta_T|\ \mathcal{F}_t]] \\ &= E^P[E^P[\phi(W_t)\zeta_T|\ \mathcal{F}_t]] \\ &= E^P[\phi(W_t)\zeta_T]. \end{aligned}$$

So we can use Q with $\frac{dQ}{dP} = \zeta_T$ for all $t \le T$. W_t will be normal with mean γt and variance t, so that $W_t - \gamma t = \tilde{W}_t$ has mean 0 and variance t.

Certainly $\tilde{W}_t$ is continuous in t. So $\tilde{W}_t$ will truly be a Brownian motion with respect to Q if we can show that $\tilde{W}_t - \tilde{W}_s$ is independent of all $\tilde{W}_u$ for $u \le s$, when considered with respect to Q. To this end, suppose Y is some random variable that is a function of $\tilde{W}_u$ for $u \le s$, i.e. Y is $\mathcal{F}_s$ measurable, and consider

$$E^Q[Yf(\tilde{W}_t - \tilde{W}_s)] = E^P[Yf(\tilde{W}_t - \tilde{W}_s)\zeta_t].$$

Write $\zeta_t = \zeta_s e^{\gamma(W_t - W_s) - \frac{1}{2}\gamma^2(t-s)}$ so that

$$Yf(\tilde{W}_t - \tilde{W}_s)\zeta_t = Y\zeta_s \cdot \left[f(W_t - W_s - \gamma(t-s))e^{\gamma(W_t - W_s) - \frac{1}{2}\gamma^2(t-s)} \right]$$

The factors outside the brackets are functions of W_u, $u \le s$ and the quantity in the brackets is a function only of $W_t - W_s$. So since W_t is a Brownian motion with respect to P the expectation splits:

$$E^P[Yf(\tilde{W}_t - \tilde{W}_s)\zeta_t] = E^P[Y\zeta_s]\, E^P[f(W_t - W_s - \gamma(t-s))e^{\gamma(W_t - W_s) - \frac{1}{2}\gamma^2(t-s)}].$$

For the first of these, $E^P[Y\zeta_s] = E^Q[Y]$. For the second, notice that simply applying the above with $Y = 1$ tells us that

$$E^Q[f(\tilde{W}_t - \tilde{W}_s)] = E^P[f(W_t - W_s - \gamma(t-s))e^{\gamma(W_t - W_s) - \frac{1}{2}\gamma^2(t-s)}].$$

So we have shown that

$$E^Q[Yf(\tilde{W}_t - \tilde{W}_s)] = E^Q[Y]\, E^Q[f(\tilde{W}_t - \tilde{W}_s)].$$

This establishes the independence of $\tilde{W}_t - \tilde{W}_s$ from $\tilde{W}_u$, all $u \le s$ under Q and confirms that (with respect to Q) $\tilde{W}_t$ is indeed a Brownian motion, as we have been anticipating!

7.2 The Girsanov Formula and the Novikov condition

What we have just worked out is an example of the Cameron-Martin-Girsanov change of measure formula for Brownian motion.

Theorem (Girsanov Formula)**.** *Suppose that W_t, $0 \le t \le T$ is a Brownian motion with respect to (Ω, P) and that θ_t is a progressively measurable stochastic process for which*

$$\zeta_t = \exp\left(\int_0^t \theta_s \, dW_s - \frac{1}{2}\int_0^t \theta_s^2 \, ds\right)$$

is a martingale. Let Q be the probability measure defined by $Q(A) = E^P[\zeta_T; A]$ for $A \in \mathcal{F}_T$. Then Q is equivalent to P and

$$\tilde{W}_t = W_t - \int_0^t \theta_s \, ds$$

is a Brownian motion with respect to Q, for $0 \le t \le T$.

The case we worked through in the preceding section is $\theta_s = \gamma$ (constant). We knew that ζ_t was indeed a martingale from our earlier Itô formula calculations. In the generality of the theorem it can be difficult to verify that ζ_t is indeed a martingale. You can use Itô's formula as before to calculate that

$$d\zeta_t = \theta_t \zeta_t \, dW_t,$$

but then we would need a way to verify something like $E[\int \theta_s^2 \zeta_s^2 \, ds] < \infty$, which is hard in general. The following provides a sufficient condition which is frequently checkable.

Theorem (Novikov Condition)**.** *Under the hypotheses of the Girsanov formula, ζ_t will be a martingale if*

$$E^P[e^{\frac{1}{2}\int_0^T \theta_s^2 \, ds}] < \infty.$$

When working with the change of measure described by Girsanov's formula it is important to remember that what was a stochastic integral with respect to P becomes the sum of stochastic and Riemann integrals with respect to Q, because $dW_t = \theta_t \, dt + d\tilde{W}_t$:

$$\int_0^t \phi_t \, dW_t = \int_0^t \phi_t \theta_t \, dt + \int_0^t \phi_t \, d\tilde{W}_t.$$

The left side is the natural expression to use with respect to P; the right is the naural form with respect to Q. What was a martingale with respect to P is now a martingale $+ \int_0^t \phi_t \theta_t \, dt$ with respect to Q.

7.3 The Black-Scholes Formula

We now have all the ingredients to work out the risk-neutral or martingale measure Q for the standard Black-Scholes model. Assume that with respect to P, $\bar{W}_t$ is a Brownian motion[1] and the stock and bond prices are given respectively by

$$S_t = S_0 e^{(\alpha - \frac{1}{2}\sigma^2)t + \sigma \bar{W}_t}, \quad B_t = B_0 e^{rt}.$$

We want to change to a new probability measure Q that makes $M_t = S_t/B_t$ a martingale. After writing $M_t = M_0 \exp((\alpha - r - \frac{1}{2}\sigma^2)t + \sigma \bar{W}_t)$ we see that

$$\begin{aligned} dM_t &= \sigma M_t \, d\bar{W}_t + M_t(\alpha - r) \, dt \\ &= \sigma M_t [d\bar{W}_t + \frac{1}{\sigma}(\alpha - r) \, dt] \\ &= \sigma M_t [d\bar{W}_t - \theta \, dt], \end{aligned}$$

where $-\theta = \frac{1}{\sigma}(\alpha - r)$. So if we define Q by $\frac{dQ}{dP} = \zeta_t$ where $\zeta_t = \exp(\theta \bar{W}_t - \frac{1}{2}\theta^2 t)$ with the θ we just identified, then with respect to Q we will have a new Brownian motion $W_t = \bar{W}_t - \theta t$ and for M_t we can write

$$dM_t = \sigma M_t dW_t.$$

[1] We are following the text's notation from §6.5 here. $\bar{W}_t$ is the *original* Brownian motion with respect to P, so that the new Q-Brownian motion can have the cleaner notation W_t, since it is the one we really want to work with.

It follows that M_t is the Q-martingale $\exp(\sigma W_t - \frac{1}{2}\sigma^2 t)$. (This can be double-checked by substituting $\bar{W}_t = W_t + \theta t$ in the expression for M_t above.) Q is the desired risk-neutral probability measure.

Next consider how S_t appears with respect to Q. We can work directly from $S_t = S_0 \exp[(\alpha - \frac{1}{2}\sigma^2)t + \sigma\bar{W}_t]$. With respect to Q we know that $W_t = \bar{W}_t - \theta t$ is a Brownian motion, with $\theta = \frac{-1}{\sigma}(\alpha - r)$. Therefore

$$(\alpha - \frac{1}{2}\sigma^2)t + \sigma\bar{W}_t = ((\alpha - \frac{1}{2}\sigma^2) + \sigma\theta)t + \sigma W_t = (r - \frac{1}{2}\sigma^2)t + \sigma W_t.$$

So with respect to the Q-Brownian motion W_t we have

$$\begin{aligned} S_t &= S_0 e^{(\alpha - \frac{1}{2}\sigma^2)t + \sigma\bar{W}_t} \\ &= S_0 e^{(r - \frac{1}{2}\sigma^2)t + \sigma W_t} \end{aligned}$$

This was the X_t we used in our calculation (in class) to solve the Black-Scholes PDE:

$$dX_t = rX_t\,dt + \sigma X_t\,dW_t.$$

When we did our calculation in class, X_t was a new stochastic process considered under the original P. Here we see that $X_t = S_t$ considered under Q. An alternate derivation is to manipulate the stochastic differentials.

$$\begin{aligned} dS_t &= \alpha S_t\,dt + \sigma S_t\,d\bar{W}_t \\ &= \alpha S_t\,dt + \sigma S_t\,(dW_t + \theta\,dt) \\ &= (\alpha + \sigma\theta)S_t\,dt + \sigma S_t\,dW_t \\ &= rS_t\,dt + \sigma S_t\,dW_t. \end{aligned}$$

Our former calculation

$$F(t, X_t) = e^{-r(T-t)} E^P\left[\Phi(X_t)|\ \mathcal{F}_t\right]$$

is the same as

$$F(t, S_t) = e^{-r(T-t)} E^Q\left[\Phi(S_t)|\ \mathcal{F}_t\right],$$

or

$$\frac{F(t, S_t)}{B_t} = E^Q\left[\frac{\Phi(S_t)}{B_T}|\ \mathcal{F}_t\right].$$

We see that the same Q that makes S_t/B_t a martingale also makes $\frac{F(t,S_t)}{B_t}$ a martingale.

Bibliography

[Bi] P. Billingsley, **Probability and Measure** (second edition), J. Wiley, New York, 1986. *This is a general text on the measure-theoretic formulation of probability theory. Most of the results we have referred to can be found there.*

[Bö] Tomas Björk, **Arbitrage Theory in Continuous Time**, Oxford Univ. Press, Oxford, 1998. *This is the text for our course.*

[Pf] P. E. Pfiffer, **Concepts of Probability Theory**, Dover, 1978. *This little book provides a slower introduction to the measure-theoretic formulation of probability than [Bi]. It develops most of the ideas we have summarized above (except martingales) at a slower pace and with illustrative examples.*

[Ba] L. Bachelier, Theorie de la speculation, *Ann. Sci. Ecole Norm. Sup.* 17 (1900), pp. 21–86. [English translation in **The Random Character of Stock Market Process**, P. H. Coonter (ed.), MIT Press, Cambridge MA, 1964, pp. 17–78.] *This is included as a reference for its historical interest only.*

[GS] I. I. Gihman and A. V. Skorokhod, **Stochastic Differential Equations**, Springer, Berlin, 1968. *This is one of the classics. Its treatment of Itô's formula based on what we have called the stochastic product rule is unique. Most authors follow the approach outlined in the text instead.*

[Fr] A. Friedman, **Stochastic Differential Equations and Applications** (2 vol.s), Academic Press, New York, 1975. *One of the more readable introductions to the subject. Although the subtleties of progressive measurability are overlooked, it is still a nice treatment of most of what we have summarized.*

[Mc] H. P. McKean, **Stochastic Integrals**, Academic Press, New York, 1969. *This is a very nice concise introduction to aspects of Brownian motion and stochastic integration, but a fair amount of detail is left for the reader to fill in. It is written for mathematicians with a solid background in analysis.*

[Mi] T. Mikosch, **Elementary Stochastic Calculus, with Finance in View**, World Scientific, Singapore, 1998. *This summary was written more recently, with the purpose of providing a concise summary of the mathematical background specifically for those interested in financial applications.*

[MR] M. Musiela and M. Rutkowski, **Martingale Methods in Financial Modeling : Theory and Applications**, Springer, New York, 2005. *Provides an extensive overview of the application of martingale methods to many topics in mathematical finance, with a huge bibliography.*

[RW] L. C. G. Rogers and D. Williams, **Diffusions, Markov Processes and Martingales** (2 vol.s), Cambridge Univ. Press, 2000. *Of the references cited here, this is the most complete and comprehensive. It covers a lot of material we haven't even touched on.*

Index