Stochastic Calculus: An Introduction with Applications

Gregory F. Lawler

© 2014 Gregory F. Lawler
Last revised: Ocrober 10, 2023

ii

Contents

Introductory comments

This is an introduction to stochastic calculus. I will assume that the reader has had a post-calculus course in probability or statistics. For much of these notes this is all that is needed, but to have a deep understanding of the subject, one needs to know measure theory and probability from that perspective. My goal is to include discussion for readers with that background as well. I also assume that the reader can write simple computer programs either using a language like C++ or by using software such as Matlab or Mathematica.

> More advanced mathematical comments that can be skipped by the reader will be indented with a different font. Comments here will assume that the reader knows that language of measure-theoretic probability theory.

We will discuss some of the applications to finance but our main focus will be on the mathematics. Financial mathematics is a kind of applied mathematics, and I will start by making some comments about the use of mathematics in "the real world". The general paradigm is as follows.

- A mathematical model is made of some real world phenomenon. Usually this model requires simplification and does not precisely describe the real situation. One hopes that models are *robust* in the sense that if the model is not very far from reality then its predictions will also be close to accurate.

- The model consists of mathematical *assumptions* about the real world.

- Given these assumptions, one does mathematical analysis to see what they imply. The analysis can be of various types:

- Rigorous derivations of consequences.

- Derivations that are plausible but are not mathematically rigorous.

- Approximations of the mathematical model which lead to tractable calculations.

- Numerical calculations on a computer.

- For models that include randomness, *Monte Carlo simulations* using a (pseudo) random number generator.

- If the mathematical analysis is successful it will make predictions about the real world. These are then checked.

- If the predictions are bad, there are two possible reasons:

  - The mathematical analysis was faulty.

  - The model does not sufficiently reflect reality.

The user of mathematics does not always need to know the details of the mathematical analysis, but it is critical to understand the *assumptions* in the model. No matter how precise or sophisticated the analysis is, if the assumptions are bad, one cannot expect a good answer.

Chapter 1

Martingales in discrete time

A martingale is a mathematical model of a fair game. To understand the definition, we need to define *conditional expectation*. The concept of conditional expectation will permeate this book.

1.1 Conditional expectation

If X is a random variable, then its expectation, $\mathbb{E}[X]$ can be thought of as the best guess for X given no information about the result of the trial. A conditional expectation can be considered as the best guess given some but not total information.

Let $X_1, X_2, \ldots$ be random variables which we think of as a time series with the data arriving one at a time. At time n we have viewed the values $X_1, \ldots, X_n$. If Y is another random variable, then $E(Y \mid X_1, \ldots, X_n)$ is the best guess for Y given $X_1, \ldots, X_n$. We will assume that Y is an *integrable* random variable which means $\mathbb{E}[|Y|] < \infty$. To save some space we will write $\mathcal{F}_n$ for "the information contained in $X_1, \ldots, X_n$" and $E[Y \mid \mathcal{F}_n]$ for $E[Y \mid X_1, \ldots, X_n]$. We view $\mathcal{F}_0$ as no information. The best guess should satisfy the following properties.

- If we have no information, then the best guess is the expected value. In other words, $E[Y \mid \mathcal{F}_0] = \mathbb{E}[Y]$.

- The conditional expectation $E[Y \mid \mathcal{F}_n]$ should only use the information available at time n. In other words, it should be a function of

$X_1, \ldots, X_n$,

$$E[Y \mid \mathcal{F}_n] = \phi(X_1, \ldots, X_n).$$

We say that $E[Y \mid \mathcal{F}_n]$ is $\mathcal{F}_n$-measurable.

The definitions in the last paragraph are certainly vague. We can use measure theory to be precise. We assume that the random variables $Y, X_1, X_2, \ldots$ are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Here $\mathcal{F}$ is a σ-algebra or σ-field of subsets of Ω, that is, a collection of subsets satisfying

- $\emptyset \in \mathcal{F}$;
- $A \in \mathcal{F}$ implies that $\Omega \setminus A \in \mathcal{F}$;
- $A_1, A_2, \ldots \in \mathcal{F}$ implies that $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

The information $\mathcal{F}_n$ is the smallest sub σ-algebra $\mathcal{G}$ of $\mathcal{F}$ such that $X_1, \ldots, X_n$ are $\mathcal{G}$-measurable. The latter statement means that for all $t \in \mathbb{R}$, the event $\{X_j \le t\} \in \mathcal{F}_n$. The "no information" σ-algebra $\mathcal{F}_0$ is the trivial σ-algebra containing only $\emptyset$ and Ω.

The definition of conditional expectation is a little tricky, so let us try to motivate it by considering an example from undergraduate probability courses. Suppose that (X, Y) have a joint density

$$f(x, y), \quad 0 < x, y < \infty,$$

with marginal densities

$$f(x) = \int_{-\infty}^{\infty} f(x, y) \, dy, \quad g(y) = \int_{-\infty}^{\infty} f(x, y) \, dx.$$

The conditional density $f(y|x)$ is defined by

$$f(y|x) = \frac{f(x, y)}{f(x)}.$$

This is well defined provided that $f(x) > 0$, and if $f(x) = 0$, then x is an "impossible" value for X to take. We can write

$$\mathbb{E}[Y \mid X = x] = \int_{-\infty}^{\infty} y \, f(y \mid x) \, dy.$$

We can use this as the definition of conditional expectation in this case,

$$E[Y \mid X] = \int_{-\infty}^{\infty} y\, f(y \mid X)\, dy = \frac{\int_{-\infty}^{\infty} y\, f(X, y)\, dy}{f(X)}.$$

Note that $E[Y \mid X]$ is a random variable which is determined by the value of the random variable X. Since it is a random variable, we can take its expectation

$$
\begin{aligned}
\mathbb{E}\left[E[Y \mid X]\right] &= \int_{-\infty}^{\infty} \mathbb{E}[Y \mid X = x]\, f(x)\, dx \\
&= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} y\, f(y \mid x)\, dy\right] f(x)\, dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y\, f(x, y)\, dy\, dx \\
&= \mathbb{E}[Y].
\end{aligned}
$$

This calculation illustrates a basic property of conditional expectation. Suppose we are interested in the value of a random variable Y and we are going to be given data $X_1, \ldots, X_n$. Once we observe the data, we make our best prediction $E[Y \mid X_1, \ldots, X_n]$. If we average our best prediction given $X_1, \ldots, X_n$ over all the possible values of $X_1, \ldots, X_n$, we get the best prediction of Y. In other words,

$$\mathbb{E}[Y] = \mathbb{E}\left[E[Y \mid \mathcal{F}_n]\right].$$

More generally, suppose that A is an $\mathcal{F}_n$-measurable event, that is to say, if we observe the data $X_1, \ldots, X_n$, then we know whether or not A has occurred. An example of an $\mathcal{F}_4$-measurable event would be

$$A = \{X_1 \geq X_2, X_4 < 4\}.$$

Let 1_A denote the *indicator function* (or *indicator random variable*) associated to the event A,

$$1_A = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{if } A \text{ does not occur} \end{cases}.$$

Using similar reasoning, we can see that if A is $\mathcal{F}_n$-measurable, then

$$\mathbb{E}\left[Y\, 1_A\right] = \mathbb{E}\left[E[Y \mid \mathcal{F}_n]\, 1_A\right].$$

At this point, we have not derived this relation mathematically; in fact, we have not even defined the conditional expectation. Instead, we will use this reasoning to motivate the following definition.

Definition The *conditional expectation* $E[Y \mid \mathcal{F}_n]$ is the unique random variable satisfying the following.

- $E[Y \mid \mathcal{F}_n]$ is $\mathcal{F}_n$-measurable.

- For every $\mathcal{F}_n$-measurable event A,

$$\mathbb{E}\left[E[Y \mid \mathcal{F}_n]\, 1_A\right] = \mathbb{E}\left[Y\, 1_A\right].$$

We have used different fonts for the E of conditional expectation and the $\mathbb{E}$ of usual expectation in order to emphasize that the conditional expectation is a random variable. However, most authors use the same font for both leaving it up to the reader to determine which is being referred to.

Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and Y is an integrable random variable. Suppose $\mathcal{G}$ is a sub σ-algebra of $\mathcal{F}$. Then $E[Y \mid \mathcal{G}]$ is defined to be the unique (up to an event of measure zero) $\mathcal{G}$-measurable random variable such that if $A \in \mathcal{G}$,

$$\mathbb{E}\left[Y\, 1_A\right] = \mathbb{E}\left[E[Y \mid \mathcal{G}]\, 1_A\right].$$

Uniqueness follows from the fact that if Z_1, Z_2 are $\mathcal{G}$-measurable random dom variables with
$$\mathbb{E}\left[Z_1\, 1_A\right] = \mathbb{E}\left[Z_2\, 1_A\right]$$
for all $A \in \mathcal{G}$, then $\mathbb{P}\{Z_1 \neq Z_2\} = 0$. Existence of the conditional expectation can be deduced from the Radon-Nikodym theorem. Let $\mu(A) = \mathbb{E}\left[Y 1_A\right]$. Then μ is a (signed) measure on $(\Omega, \mathcal{G}, \mathbb{P})$ with $\mu \ll \mathbb{P}$, and hence there exists an L^1 random variable Z with $\mu(A) = \mathbb{E}\left[Z 1_A\right]$ for all $A \in \mathcal{G}$.

Although the definition does not give an immediate way to calculate the conditional expectation, in many cases one can compute it. We will give a number of properties of the conditional expectation most of which follow quickly from the definition.

Proposition 1.1.1. *Suppose $X_1, X_2, \ldots$ is a sequence of random variables and $\mathcal{F}_n$ denotes the information at time n. The conditional expectation $E[Y \mid \mathcal{F}_n]$ satisfies the following properties.*

- *If Y is $\mathcal{F}_n$-measurable, then $E[Y \mid \mathcal{F}_n] = Y$.*

- *If A is an $\mathcal{F}_n$-measurable event, then $\mathbb{E}\left[E[Y \mid \mathcal{F}_n] \, 1_A\right] = \mathbb{E}\left[Y \, 1_A\right]$. In particular,*
$$\mathbb{E}\left[E[Y \mid \mathcal{F}_n]\right] = \mathbb{E}[Y].$$

- *Suppose $X_1, \ldots, X_n$ are independent of Y. Then $\mathcal{F}_n$ contains no useful information about Y and hence*
$$E[Y \mid \mathcal{F}_n] = \mathbb{E}[Y].$$

- **Linearity***. If Y, Z are random variables and a, b are constants, then*
$$E[aY + bZ \mid \mathcal{F}_n] = a\, E[Y \mid \mathcal{F}_n] + b\, E[Z \mid \mathcal{F}_n]. \qquad (1.1)$$

- **Projection or Tower Property***. If $m < n$, then*
$$E\left[\, E[Y \mid \mathcal{F}_n] \mid \mathcal{F}_m \,\right] = E[Y \mid \mathcal{F}_m]. \qquad (1.2)$$

- *If Z is an $\mathcal{F}_n$-measurable random variable, then when conditioning with respect to $\mathcal{F}_n$, Z acts like a constant,*
$$E[YZ \mid \mathcal{F}_n] = Z\, E[Y \mid \mathcal{F}_n]. \qquad (1.3)$$

The proof of this proposition is not very difficult given our choice of definition for the conditional expectation. We will discuss only a couple of cases here, leaving the rest for the reader. To prove the linearity property, we know that $a\, E[Y \mid \mathcal{F}_n] + b\, E[Z \mid \mathcal{F}_n]$ is an $\mathcal{F}_n$-measurable random variable. Also if $A \in \mathcal{F}_n$, then linearity of expectation implies that

$$
\begin{aligned}
\mathbb{E}\left[1_A \left(a\, E[Y \mid \mathcal{F}_n] + b\, E[Z \mid \mathcal{F}_n]\right)\right] \\
= \quad & a\mathbb{E}\left[1_A \, E[Y \mid \mathcal{F}_n]\right] + b\,\mathbb{E}\left[1_A \, E[Z \mid \mathcal{F}_n]\right] \\
= \quad & a\mathbb{E}\left[1_A \, Y\right] + b\mathbb{E}\left[1_A \, Z\right] \\
= \quad & \mathbb{E}\left[1_A \left(aY + bZ\right)\right].
\end{aligned}
$$

Uniqueness of the conditional expectation then implies (1.1).

We first show the "constants rule" (1.3) for $Z = 1_A, A \in \mathcal{F}_n$, as follows. If $B \in \mathcal{F}_n$, then $A \cap B \in \mathcal{F}_n$ and

$$\mathbb{E}\left[1_B\, E(YZ \mid \mathcal{F}_n)\right] = \mathbb{E}\left[1_B\, E(1_A Y \mid \mathcal{F}_n)\right]$$

$$= \mathbb{E}\left[1_B\, 1_A\, Y\right] = \mathbb{E}\left[1_{A \cap B} Y\right] = \mathbb{E}\left[1_{A \cap B} E(Y \mid \mathcal{F}_n)\right]$$

$$= \mathbb{E}\left[1_B\, 1_A\, E(Y \mid \mathcal{F}_n)\right] = \mathbb{E}\left[1_B\, Z\, E(Y \mid \mathcal{F}_n)\right].$$

Hence $E(YZ \mid \mathcal{F}_n) = Z\, E(Y \mid \mathcal{F}_n)$ by definition. Using linearity, the rule holds for simple random variables of the form

$$Z = \sum_{j=1}^{n} a_j\, 1_{A_j}, \quad A_j \in \mathcal{F}_n.$$

We can then prove it for nonnegative Y by approximating from below by nonnegative simple random variables and using the monotone convergence theorem, and then for general Y by writing $Y = Y^+ - Y^-$. These are standard techniques in Lebesgue integration theory.

Example 1.1.1. Suppose that $X_1, X_2, \ldots$ are independent random variables with $\mathbb{E}[X_j] = \mu$ for each j. Let $S_n = X_1 + \cdots + X_n$, and let $\mathcal{F}_n$ denote the information contained in $X_1, \ldots, X_n$. Then if $m < n$,

$$
\begin{aligned}
E[S_n \mid \mathcal{F}_m] &= E[S_m \mid \mathcal{F}_m] + E[S_n - S_m \mid \mathcal{F}_m] \\
&= S_m + \mathbb{E}[S_n - S_m] \\
&= S_m + \mu\,(n - m).
\end{aligned}
$$

Example 1.1.2. In the same setup as Example 1.1.1 suppose that $\mu = 0$ and $\mathbb{E}[X_j^2] = \sigma^2$ for each j. Then if $m < n$,

$$
\begin{aligned}
E[S_n^2 \mid \mathcal{F}_m] &= E([S_m + (S_n - S_m)]^2 \mid \mathcal{F}_m) \\
&= E[S_m^2 \mid \mathcal{F}_m] + 2\, E[S_m(S_n - S_m) \mid \mathcal{F}_m] \\
&\quad + E[(S_n - S_m)^2 \mid \mathcal{F}_m].
\end{aligned}
$$

Since S_m is $\mathcal{F}_m$-measurable and $S_n - S_m$ is independent of $\mathcal{F}_m$,

$$E[S_m^2 \mid \mathcal{F}_m] = S_m^2,$$

$$E[S_m(S_n - S_m) \mid \mathcal{F}_m] = S_m \, E[S_n - S_m \mid \mathcal{F}_m] = S_m \, \mathbb{E}[S_n - S_m] = 0,$$

$$E[(S_n - S_m)^2 \mid \mathcal{F}_m] = \mathbb{E}[(S_n - S_m)^2] = \mathrm{Var}(S_n - S_m) = \sigma^2 \, (n - m),$$

and hence,

$$E[S_n^2 \mid \mathcal{F}_m] = S_m^2 + \sigma^2 \, (n - m).$$

Example 1.1.3. In the same setup as Example 1.1.1, let us also assume that $X_1, X_2, \ldots$ are identically distributed. We will compute $E[X_1 \mid S_n]$. Note that the information contained in the one data point S_n is less than the information contained in $X_1, \ldots, X_n$. However, since the random variables are identically distributed, it must be the case that

$$E[X_1 \mid S_n] = E[X_2 \mid S_n] = \cdots = E[X_n \mid S_n].$$

Linearity implies that

$$n \, E[X_1 \mid S_n] = \sum_{j=1}^{n} E[X_j \mid S_n] = E[X_1 + \cdots + X_n \mid S_n] = E[S_n \mid S_n] = S_n.$$

Therefore,

$$E[X_1 \mid S_n] = \frac{S_n}{n}.$$

It may be at first surprising that the answer does not depend on $\mathbb{E}[X_1]$.

Definition If $X_1, X_2, \ldots$ is a sequence of random variables, then the associated (discrete time) *filtration* is the collection $\{\mathcal{F}_n\}$ where $\mathcal{F}_n$ denotes the information in $X_1, \ldots, X_n$.

One assumption in the definition of a filtration, which may sometimes not reflect reality, is that information is never lost. If $m < n$, then everything known at time m is still known at time n. Sometimes a filtration is given starting at time $n = 1$ and sometimes starting at $n = 0$. If it starts at time $n = 1$, we define $\mathcal{F}_0$ to be "no information".

More generally, a *(discrete time) filtration* $\{\mathcal{F}_n\}$ is an increasing sequence of σ-algebras.

1.2   Martingales

A martingale is a model of a fair game. Suppose $X_1, X_2, \ldots$ is a sequence of random variables to which we associate the filtration $\{\mathcal{F}_n\}$ where $\mathcal{F}_n$ is the information contained in $X_1, \ldots, X_n$.

Definition A sequence of random variables $M_0, M_1, \ldots$ is called a *martingale with respect to the filtration* $\{\mathcal{F}_n\}$ if:

- For each n, M_n is an $\mathcal{F}_n$-measurable random variable with $\mathbb{E}[|M_n|] < \infty$.

- If $m < n$, then
$$E[M_n \mid \mathcal{F}_m] = M_m. \tag{1.4}$$

We can also write (1.4) as

$$E[M_n - M_m \mid \mathcal{F}_m] = 0.$$

If we think of M_n as the winnings of a game, then this implies that no matter what has happened up to time m, the expected winnings in the next $n - m$ games is 0. Sometimes one just says "$M_0, M_1, \ldots$ is a martingale" without reference to the filtration. In this case, the assumed filtration is $\mathcal{F}_n$, the information in $M_0, \ldots, M_n$. In order to establish (1.4) it suffices to show for all n,
$$E[M_{n+1} \mid \mathcal{F}_n] = M_n. \tag{1.5}$$

In order to see this, we can use the tower property (1.2) for conditional expectation to see that

$$E[M_{n+2} \mid \mathcal{F}_n] = E\left[\, E[M_{n+2} \mid \mathcal{F}_{n+1}] \mid \mathcal{F}_n\right] = E[M_{n+1} \mid \mathcal{F}_n] = M_n,$$

and so forth. Also note that if M_n is a martingale, then

$$\mathbb{E}[M_n] = \mathbb{E}\left[E[M_n \mid \mathcal{F}_0]\right] = \mathbb{E}[M_0].$$

Example 1.2.1. Suppose $X_1, X_2, \ldots$ are independent random variables with $\mathbb{E}[X_j] = 0$ for each j. Let $S_0 = 0$ and $S_n = X_1 + \cdots + X_n$. In the last section we showed that if $m < n$, then $E[S_n \mid \mathcal{F}_m] = S_m$. Hence, S_n is a martingale with respect to $\mathcal{F}_n$, the information in $X_1, \ldots, X_n$.

Example 1.2.2. Suppose $X_n, S_n, \mathcal{F}_n$ are as in Example 1.2.1 and also assume $\mathrm{Var}[X_j] = \mathbb{E}[X_j^2] = \sigma_j^2 < \infty$. Let

$$A_n = \sigma_1^2 + \cdots + \sigma_n^2,$$

$$M_n = S_n^2 - A_n,$$

where $M_0 = 0$. Then M_n is a martingale with respect to $\mathcal{F}_n$. To see this, we compute as in Example 1.1.2,

$$
\begin{aligned}
E[S_{n+1}^2 \mid \mathcal{F}_n] &= E[(S_n + X_{n+1})^2 \mid \mathcal{F}_n] \\
&= E[S_n^2 \mid \mathcal{F}_n] + 2E[S_n X_{n+1} \mid \mathcal{F}_n] + E[X_{n+1}^2 \mid \mathcal{F}_n] \\
&= S_n^2 + 2S_n \, E[X_{n+1} \mid \mathcal{F}_n] + \mathbb{E}[X_{n+1}^2] \\
&= S_n^2 + 2S_n \, \mathbb{E}[X_{n+1}] + \mathbb{E}[X_{n+1}^2] \\
&= S_n^2 + \sigma_{n+1}^2.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
E[M_{n+1} \mid \mathcal{F}_n] &= E[S_{n+1}^2 - A_{n+1} \mid \mathcal{F}_n] \\
&= S_n^2 + \sigma_{n+1}^2 - (A_n + \sigma_{n+1}^2) = M_n.
\end{aligned}
$$

There are various ways to view a martingale. One can consider M_n as the price of an asset (although we allow negative values of M_n) or as the winnings in a game. We can also consider

$$\Delta M_n = M_n - M_{n-1}$$

as either the change in the asset price or as the amount won in the game at time n. Negative values indicate drops in price or money lost in the game. The basic idea of stochastic integration is to allow one to change one's portfolio (in the asset viewpoint) or change one's bet (in the game viewpoint). However, we are not allowed to see the outcome before betting. We make this precise in the next example.

Example 1.2.3. Discrete stochastic integral. Suppose that $M_0, M_1, \ldots$ is a martingale with respect to the filtration $\mathcal{F}_n$. For $n \geq 1$, let $\Delta M_n = M_n - M_{n-1}$. Let B_j denote the "bet" on the jth game. We allow negative values of B_j which indicate betting that the price will go down or the game

will be lost. Let W_n denote the winnings in this strategy: $W_0 = 0$ and for $n \geq 1$,

$$W_n = \sum_{j=1}^{n} B_j \left[M_j - M_{j-1} \right] = \sum_{j=1}^{n} B_j \, \Delta M_j.$$

Let us assume that for each n there is a number $K_n < \infty$ such that $|B_n| \leq K_n$. We also assume that we cannot see the result of nth game before betting. This last assumption can be expressed mathematically by saying that B_n is $\mathcal{F}_{n-1}$-measurable. In other words, we can adjust our bet based on how well we have been doing. We claim that under these assumptions, W_n is a martingale with respect to $\mathcal{F}_n$. It is clear that W_n is measurable with respect to $\mathcal{F}_n$, and integrability follows from the estimate

$$
\begin{aligned}
\mathbb{E}[|W_n|] &\leq \sum_{j=1}^{n} \mathbb{E}[|B_j||M_j - M_{j-1}|] \\
&\leq \sum_{j=1}^{n} K_j \left(\mathbb{E}[|M_j|] + \mathbb{E}[|M_{j-1}|] \right) < \infty.
\end{aligned}
$$

Also,

$$
\begin{aligned}
E[W_{n+1} \mid \mathcal{F}_n] &= E[W_n + B_{n+1}(M_{n+1} - M_n) \mid \mathcal{F}_n] \\
&= E[W_n \mid \mathcal{F}_n] + E[B_{n+1}(M_{n+1} - M_n) \mid \mathcal{F}_n].
\end{aligned}
$$

Since W_n is $\mathcal{F}_n$-measurable, $E[W_n \mid \mathcal{F}_n] = W_n$. Also, since B_{n+1} is $\mathcal{F}_n$-measurable and M is a martingale,

$$E[B_{n+1}(M_{n+1} - M_n) \mid \mathcal{F}_n] = B_{n+1} \, E[M_{n+1} - M_n \mid \mathcal{F}_n] = 0.$$

Therefore,

$$E[W_{n+1} \mid \mathcal{F}_n] = W_n.$$

Example 1.2.3 demonstrates an important aspect of martingales. One cannot change a discrete-time martingale to a game in one's favor with a betting strategy in a finite amount of time. However, the next example shows that if we are allowed an infinite amount of time we can beat a fair game.

Example 1.2.4. Martingale betting strategy. Let $X_1, X_2, \ldots$ be independent random variables with

$$\mathbb{P}\{X_j = 1\} = \mathbb{P}\{X_j = -1\} = \frac{1}{2}. \qquad (1.6)$$

We will refer to such random variables as "coin-tossing" random variables where 1 corresponds to heads and -1 corresponds to tails. Let $M_0 = 0$, $M_n = X_1 + \cdots + X_n$. We have seen that M_n is a martingale. We will consider the following betting strategy. We start by betting \$1. If we win, we quit; otherwise, we bet \$2 on the next game. If we win the second game, we quit; otherwise we double our bet to \$4 and play. Each time we lose, we double our bet. At the time that we win, we will be ahead \$1. With probability one, we will eventually win the game, so this strategy is a way to beat a fair game. The winnings in this game can be written as

$$W_n = \sum_{j=1}^{n} B_j \, \Delta M_j = \sum_{j=1}^{n} B_j \, X_j,$$

where the bet $B_1 = 1$ and for $j > 1$,

$$B_j = 2^{j-1} \text{ if } X_1 = X_2 = \cdots = X_{j-1} = -1,$$

and otherwise $B_j = 0$. This is an example of a discrete stochastic integral as in the previous example, and hence, we know that W_n must be a martingale. In particular, for each n, $\mathbb{E}[W_n] = 0$. We can check this directly by noting that $W_n = 1$ unless $X_1 = X_2 = \cdots = X_n = -1$ in which case

$$W_n = -1 - 2^1 - 2^2 - \cdots - 2^{n-1} = -[2^n - 1].$$

This last event happens with probability $(1/2)^n$, and hence

$$\mathbb{E}[W_n] = 1 \cdot [1 - 2^{-n}] - [2^n - 1] \cdot 2^{-n} = 0.$$

However, we will eventually win which means that with probability one

$$W_\infty = \lim_{n \to \infty} W_n = 1,$$

and

$$1 = \mathbb{E}[W_\infty] > \mathbb{E}[W_0] = 0.$$

We have beaten the game (but it takes an infinite amount of time to guarantee it).

If the condition (1.4) is replaced with

$$E[M_n \mid \mathcal{F}_m] \geq M_m,$$

then the process is called a *submartingale*. If it is replaced with

$$E[M_n \mid \mathcal{F}_m] \leq M_m,$$

then it is called a *supermartingale*. In other words, games that are always in one's favor are submartingales and games that are always against one are supermartingales. (At most games in Las Vegas, one's winnings give a supermartingale.) Under this definition, a martingale is both a submartingale and a supermartingale. The terminology may seem backwards at first: submartingales get bigger and supermartingales get smaller. The terminology was set to be consistent with the related notion of subharmonic and superharmonic functions. Martingales are related to harmonic functions.

1.3   Optional sampling theorem

Suppose $M_0, M_1, M_2, \ldots$ is a martingale with respect to the filtration $\{\mathcal{F}_n\}$. In the last section we discussed the discrete stochastic integral. Here we will consider a particular case of a betting strategy where one bets 1 up to some time and then one bets 0 afterwards. Let T be the "stopping time" for the strategy. Then the winnings at time t is

$$M_0 + \sum_{j=1}^{n} B_j \left[M_j - M_{j-1}\right],$$

where $B_j = 1$ if $j \leq T$ and $B_j = 0$ if $j > T$. We can write this as

$$M_{n \wedge T},$$

where $n \wedge T$ is shorthand for $\min\{n, T\}$. The time T is random, but it must satisfy the condition that the betting rule is allowable.

Definition A nonnegative integer-valued random variable T is a *stopping time* with respect to the filtration $\{\mathcal{F}_n\}$ if for each n the event $\{T = n\}$ is $\mathcal{F}_n$-measurable.

The following theorem is a special case of the discrete stochastic integral. It restates the fact that one cannot beat a martingale in finite time. We call this the *optional sampling theorem*; it is also called the *optional stopping theorem*.

Theorem 1.3.1 (Optional Sampling Theorem I). *Suppose T is a stopping time and M_n is a martingale with respect to $\{\mathcal{F}_n\}$. Then $Y_n = M_{n \wedge T}$ is a martingale. In particular, for each n,*

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

If T is bounded, that is, if there exists $k < \infty$ such that $\mathbb{P}\{T \leq k\} = 1$, then

$$\mathbb{E}\left[M_T\right] = \mathbb{E}\left[M_0\right]. \tag{1.7}$$

The final conclusion (1.7) of the theorem holds since $\mathbb{E}[M_{n \wedge T}] = \mathbb{E}[M_T]$ for $n \geq k$. What if the stopping time T is not bounded but $\mathbb{P}\{T < \infty\} = 1$? Then, we cannot conclude (1.7) without further assumptions. To see this we need only consider the martingale betting strategy of the previous section. If we define

$$T = \min\{n : X_n = 1\} = \min\{n : W_n = 1\},$$

then with probability one $T < \infty$ and $W_T = 1$. Hence,

$$1 = \mathbb{E}\left[W_T\right] > \mathbb{E}\left[W_0\right] = 0.$$

Often one does want to conclude (1.7) for unbounded stopping times, so it is useful to give conditions under which it holds. Let us try to derive the equality and see what conditions we need to impose. First, we will assume that we stop, $\mathbb{P}\{T < \infty\} = 1$, so that M_T makes sense. For every $n < \infty$, we know that

$$\mathbb{E}[M_0] = \mathbb{E}[M_{n \wedge T}] = \mathbb{E}[M_T] + \mathbb{E}[M_{n \wedge T} - M_T].$$

If we can show that

$$\lim_{n \to \infty} \mathbb{E}\left[|M_{n \wedge T} - M_T|\right] = 0,$$

then we have (1.7). The random variable $M_{n \wedge T} - M_T$ is zero if $n \wedge T = T$, and

$$M_{n \wedge T} - M_T = 1\{T > n\} \left[M_n - M_T\right].$$

If $\mathbb{E}[|M_T|] < \infty$, then one can show that

$$\lim_{n \to \infty} \mathbb{E}\left[|M_T| \, 1\{T > n\}\right] = 0.$$

In the martingale betting strategy example, this term did not cause a problem since $W_T = 1$ and hence $\mathbb{E}[|W_T|] < \infty$.

If $\mathbb{P}\{T < \infty\} = 1$, then the random variables $X_n = |M_T| 1\{T > n\}$ converge to zero with probability one. If $\mathbb{E}\left[|M_T|\right] < \infty$, then we can use the dominated convergence theorem to conclude that

$$\lim_{n \to \infty} \mathbb{E}[X_n] = 0.$$

Finally, in order to conclude (1.7) we will make the hypothesis that the other term acts nicely.

Theorem 1.3.2 (Optional Sampling Theorem II). *Suppose T is a stopping time and M_n is a martingale with respect to $\{\mathcal{F}_n\}$. Suppose that $\mathbb{P}\{T < \infty\} = 1, \mathbb{E}\left[|M_T|\right] < \infty$, and for each n,*

$$\lim_{n \to \infty} \mathbb{E}\left[|M_n| 1\{T > n\}\right] = 0. \tag{1.8}$$

Then,

$$\mathbb{E}\left[M_T\right] = \mathbb{E}\left[M_0\right].$$

Let us check that the martingale betting strategy does *not* satisfy the conditions of the theorem (it better not since it does not satisfy the conclusion!) In fact, it does not satisfy (1.8). For this strategy, if $T > n$, then we have lost n times and $W_n = 1 - 2^n$. Also, $\mathbb{P}\{T > n\} = 2^{-n}$. Therefore,

$$\lim_{n \to \infty} \mathbb{E}\left[|W_n| 1\{T > n\}\right] = \lim_{n \to \infty} (2^n - 1) 2^{-n} = 1 \neq 0.$$

Checking condition (1.8) can be difficult in general. We will give one criterion which is useful.

Theorem 1.3.3 (Optional Sampling Theorem III). *Suppose T is a stopping time and M_n is a martingale with respect to $\{\mathcal{F}_n\}$. Suppose that $\mathbb{P}\{T < \infty\} = 1, \mathbb{E}\left[|M_T|\right] < \infty$, and that there exists $C < \infty$ such that for each n,*

$$\mathbb{E}\left[|M_{n \wedge T}|^2\right] \leq C. \tag{1.9}$$

Then,

$$\mathbb{E}\left[M_T\right] = \mathbb{E}\left[M_0\right].$$

To prove this theorem, first note that with probability one,

$$|M_T|^2 = \lim_{n\to\infty} |M_{T\wedge n}|^2 \, 1\{T \le n\},$$

and hence by the Hölder inequality and the monotone convergence theorem,

$$\mathbb{E}\left[|M_T|\right]^2 \le \mathbb{E}\left[|M_T|^2\right] = \lim_{n\to\infty} \mathbb{E}\left[|M_{T\wedge n}|^2| \, 1\{T \le n\}\right] \le C.$$

We need to show that (1.9) implies (1.8). If $b > 0$, then for every n,

$$\mathbb{E}\left[|M_n| \, 1\{|M_n| \ge b, T > n\}\right] \le \frac{\mathbb{E}[|M_{n\wedge T}|^2]}{b} \le \frac{C}{b}.$$

Therefore,

$$
\begin{aligned}
\mathbb{E}\left[|M_n| \, 1\{T > n\}\right] &= \mathbb{E}\left[|M_n| \, 1\{T > n, |M_n| \ge b\}\right] \\
&\quad + \mathbb{E}\left[|M_n| \, 1\{T > n, |M_n| < b\}\right] \\
&\le \frac{C}{b} + b\,\mathbb{P}\{T > n\}.
\end{aligned}
$$

Hence,

$$\limsup_{n\to\infty} \mathbb{E}\left[|M_n| \, 1\{T > n\}\right] \le \frac{C}{b} + b \lim_{n\to\infty} \mathbb{P}\{T > n\} = \frac{C}{b}.$$

Since this holds for every $b > 0$ we get (1.8).

Example 1.3.1. Gambler's ruin for random walk. Let $X_1, X_2, \ldots$ be independent, coin-tosses as in (1.6) and let

$$S_n = 1 + X_1 + \cdots + X_n.$$

S_n is called *simple (symmetric) random walk* starting at 1. We have shown that S_n is a martingale. Let $K > 1$ be a positive integer and let T denote the first time n such that $S_n = 0$ or $S_n = K$. Then $M_n = S_{n\wedge T}$ is a martingale. Also $0 \le M_n \le K$ for all n, so (1.9) is satisfied. We can apply the optional sampling theorem to deduce that

$$1 = M_0 = \mathbb{E}[M_T] = 0 \cdot \mathbb{P}\{M_T = 0\} + K \cdot \mathbb{P}\{M_T = K\}.$$

By solving, we get

$$\mathbb{P}\{M_T = K\} = \frac{1}{K}.$$

This relation is sometimes called the *gambler's ruin estimate* for the random walk. Note that

$$\lim_{K \to \infty} \mathbb{P}\{M_T = K\} = 0.$$

If we consider 1 to be the starting stake of a gambler and K to be the amount held by a casino, this shows that with a fair game, the gambler will almost surely lose. If $\tau = \min\{n : S_n = 0\}$, then the last equality implies that $\mathbb{P}\{\tau < \infty\} = 1$. The property that the walk always returns to the origin is called *recurrence*.

Example 1.3.2. Let $S_n = X_1 + \cdots + X_n$ be simple random walk starting at 0. We have seen that

$$M_n = S_n^2 - n$$

is a martingale. Let J, K be positive integers and let

$$T = \min\{n : S_n = -J \text{ or } S_n = K\}.$$

As in Example 1.3.1, we have

$$0 = \mathbb{E}[S_0] = \mathbb{E}[S_T] = [1 - \mathbb{P}\{S_T = K\}] \cdot (-J) + \mathbb{P}\{S_T = K\} \cdot K,$$

and solving gives

$$\mathbb{P}\{S_T = K\} = \frac{J}{J + K}.$$

In Exercise 1.13 it is shown that there exists $C < \infty$ such that for all n $\mathbb{E}[M_{n \wedge T}^2] \le C$. Hence we can use Theorem 1.3.3 to conclude that

$$0 = \mathbb{E}[M_0] = \mathbb{E}[M_T] = \mathbb{E}[S_T^2] - \mathbb{E}[T].$$

Moreover,

$$\begin{aligned}
\mathbb{E}[S_T^2] &= J^2 \, \mathbb{P}\{S_T = -J\} + K^2 \, \mathbb{P}\{S_T = K\} \\
&= J^2 \, \frac{K}{J + K} + K^2 \, \frac{J}{J + K} = JK.
\end{aligned}$$

Therefore,

$$\mathbb{E}[T] = \mathbb{E}[S_T^2] = JK.$$

In particular, the expected amount of time for the random walker starting at the origin to get distance K from the origin is K^2.

Example 1.3.3. As in Example 1.3.2, let $S_n = X_1 + \cdots + X_n$ be simple random walk starting at 0. Let

$$T = \min\{n : S_n = 1\}, \quad T_J = \min\{n : S_n = 1 \text{ or } S_n = -J\}.$$

Note that $T = \lim_{J \to \infty} T_J$ and

$$\mathbb{P}\{T = \infty\} = \lim_{J \to \infty} \mathbb{P}\{S_{T_J} = -J\} = \lim_{J \to \infty} \frac{1}{J+1} = 0.$$

Therefore, $\mathbb{P}\{T < \infty\} = 1$, although Example 1.3.2 shows that for every J,

$$\mathbb{E}[T] \geq \mathbb{E}[T_J] = J,$$

and hence $\mathbb{E}[T] = \infty$. Also, $S_T = 1$, so we do not have $\mathbb{E}[S_0] = \mathbb{E}[S_T]$. From this we can see that (1.8) and (1.9) are not satisfied by this example.

1.4 Martingale convergence theorem

The martingale convergence theorem describes the behavior of a martingale M_n as $n \to \infty$.

Theorem 1.4.1 (Martingale Convergence Theorem). *Suppose M_n is a martingale with respect to $\{\mathcal{F}_n\}$ and there exists $C < \infty$ such that $\mathbb{E}\left[|M_n|\right] \leq C$ for all n. Then there exists a random variable M_∞ such that with probability one*

$$\lim_{n \to \infty} M_n = M_\infty.$$

It does *not* follow from the theorem that $\mathbb{E}[M_\infty] = \mathbb{E}[M_0]$. For example, the martingale betting strategy satisfies the conditions of the theorem since

$$\mathbb{E}\left[|W_n|\right] = (1 - 2^{-n}) \cdot 1 + (2^n - 1) \cdot 2^{-n} \leq 2.$$

However, $W_\infty = 1$ and $W_0 = 0$.

We will prove the martingale convergence theorem. The proof uses a well-known financial strategy — buy low, sell high. Suppose $M_0, M_1, \ldots$ is a martingale such that

$$\mathbb{E}\left[|M_n|\right] \leq C < \infty,$$

for all n. Suppose $a < b$ are real numbers. We will show that it is impossible for the martingale to fluctuate infinitely often below a and above b. Define a sequence of stopping times by

$$S_1 = \min\{n : M_n \leq a\}, \quad T_1 = \min\{n > S_1 : M_n \geq b\},$$

and for $j > 1$,
$$S_j = \min\{n > T_{j-1} : M_n \leq a\},$$
$$T_j = \min\{n > S_j : M_n \geq b\}.$$

We set up the discrete stochastic integral

$$W_n = \sum_{k=0}^{n} B_k \left[M_k - M_{k-1} \right],$$

with $B_n = 0$ if $n - 1 < S_1$ and

$$B_n = 1 \quad \text{if } S_j \leq n - 1 < T_j,$$

$$B_n = 0 \quad \text{if } T_j \leq n - 1 < S_{j+1}.$$

In other words, every time the "price" drops below a we buy a unit of the asset and hold onto it until the price goes above b at which time we sell. Let U_n denote the number of times by time n that we have seen a fluctuation; that is,

$$U_n = j \quad \text{if} \quad T_j < n \leq T_{j+1}.$$

We call U_n the number of *upcrossings* by time n. Every upcrossing results in a profit of at least $b - a$. From this we see that

$$W_n \geq U_n (b - a) + (M_n - a).$$

The term $a - M_n$ represents a possible loss caused by holding a share of the asset at the current time. Since W_n is a martingale, we know that $\mathbb{E}[W_n] = \mathbb{E}[W_0] = 0$, and hence

$$\mathbb{E}[U_n] \leq \frac{\mathbb{E}[a - M_n]}{b - a} \leq \frac{|a| + \mathbb{E}[|M_n|]}{b - a} \leq \frac{|a| + C}{b - a}.$$

This holds for every n, and hence

$$\mathbb{E}[U_\infty] \leq \frac{|a| + C}{b - a} < \infty.$$

In particular with probability one, $U_\infty < \infty$, and hence there are only a finite number of fluctuations. We now allow a, b to run over all rational numbers to see that with probability one,

$$\liminf_{n\to\infty} M_n = \limsup_{n\to\infty} M_n.$$

Therefore, the limit

$$M_\infty = \lim_{n\to\infty} M_n$$

exists. We have not yet ruled out the possibility that M_∞ is $\pm\infty$, but it is not difficult to see that if this occurred with positive probability, then $\mathbb{E}[|M_n|]$ would not be uniformly bounded.

To illustrate the martingale convergence theorem, we will consider another example of a martingale called *Polya's urn*. Suppose we have an urn with red and green balls. At time $n = 0$, we start with one red ball and one green ball. At each positive integer time we choose a ball at random from the urn (with each ball equally likely to be chosen), look at the color of the ball, and then put the ball back in with another ball of the same color. Let R_n, G_n denote the number of red and green balls in the urn after the draw at time n so that

$$R_0 = G_0 = 1, \qquad R_n + G_n = n + 2,$$

and let

$$M_n = \frac{R_n}{R_n + G_n} = \frac{R_n}{n+2}$$

be the fraction of red balls at this time. Let $\mathcal{F}_n$ denote the information in the data $M_1, \ldots, M_n$, which one can check is the same as the information in $R_1, R_2, \ldots, R_n$. Note that the probability that a red ball is chosen at time n depends only on the number (or fraction) of red balls in the urn before choosing. It does not depend on what order the red and green balls were put in. This is an example of the *Markov property*. This concept will appear a number of times for us, so let us define it.

Definition A discrete time process $Y_0, Y_1, Y_2, \ldots$ is called *Markov* if for each n, the conditional distribution of

$$Y_{n+1}, Y_{n+2}, \ldots$$

given $Y_0, Y_1, \ldots, Y_n$ is the same as the conditional distribution given Y_n. In other words, the only thing about the past and present that is useful for predicting the future is the current value of the process.

We can describe the rule of Polya's urn by

$$\mathbb{P}\{R_{n+1} = R_n + 1 \mid \mathcal{F}_n\} = 1 - \mathbb{P}\{R_{n+1} = R_n \mid \mathcal{F}_n\} =$$

$$\mathbb{P}\{R_{n+1} = R_n + 1 \mid M_n\} = \frac{R_n}{n+2} = M_n.$$

We claim that M_n is a martingale with respect to $\mathcal{F}_n$. To check this,

$$
\begin{aligned}
E\left[M_{n+1} \mid \mathcal{F}_n\right] &= E\left[M_{n+1} \mid M_n\right] \\
&= M_n \frac{R_n + 1}{n+3} + [1 - M_n] \frac{R_n}{n+3} \\
&= \frac{R_n(R_n + 1)}{(n+2)(n+3)} + \frac{(n + 2 - R_n)R_n}{(n+2)(n+3)} \\
&= \frac{R_n(n+3)}{(n+2)(n+3)} = M_n.
\end{aligned}
$$

Since $\mathbb{E}[|M_n|] = \mathbb{E}[M_n] = \mathbb{E}[M_0] = 1/2$, this martingale satisfies the conditions of the martingale convergence theorem. (In fact, the same argument shows that every martingale that stays nonnegative satisfies the conditions.) Hence, there exists a random variable M_∞ such that with probability one,

$$\lim_{n\to\infty} M_n = M_\infty.$$

It turns out that the random variable M_n is really random in the sense that it has a nontrivial distribution. In Exercise 1.11 you will show that for each n, the distribution of M_n is uniform on

$$\left\{ \frac{1}{n+2}, \frac{2}{n+2}, \ldots, \frac{n+1}{n+2} \right\},$$

and from this it is not hard to see that M_∞ has a uniform distribution on $[0, 1]$. You will also be asked to simulate this process to see what happens. There is a lot of randomness in the first few draws to see what fraction of red balls the urn will settle down to. However, for large n this ratio changes very little; for example, the ratio after 2000 draws is very close to the ratio after 4000 draws.

While Polya's urn seems like a toy model, it arises in a number of places. We will give an example from Bayesian statistics. Suppose that we perform independent trials of an experiment where the probability of success for each experiment is θ (such trials are called *Bernoulli trials*). Suppose that we do not know the value of θ, but want to try to deduce it by observing trials. Let $X_1, X_2, \ldots$ be independent random variables with

$$\mathbb{P}\{X_j = 1\} = 1 - \mathbb{P}\{X_j = 0\} = \theta.$$

The (strong) law of large numbers implies that with probability one,

$$\lim_{n \to \infty} \frac{X_1 + \cdots + X_n}{n} = \theta. \tag{1.10}$$

Hence, if were able to observe infinitely many trials, we could deduce θ exactly.

Clearly, we cannot deduce θ with 100% assurance if we see only a finite number of trials. Indeed, if $0 < \theta < 1$, there is always a chance that the first n trials will all be failures and there is a chance they will all be successes. The Bayesian approach to statistics is to assume that θ is a random variable with a certain *prior distribution*. As we observe the data we update to a *posterior distribution*. We will assume we know nothing initially about the value and choose the prior distribution to the uniform distribution on $[0, 1]$ with density

$$f_0(\theta) = 1, \quad 0 < \theta < 1.$$

Suppose that after observing n trials, we have had $S_n = X_1 + \cdots + X_n$ successes. If we know θ, then the distribution of S_n is binomial,

$$\mathbb{P}\{S_n = k \mid \theta\} = \binom{n}{k} \theta^k (1 - \theta)^{n-k}.$$

We use a form of the Bayes rule to update the density

$$f_{n,k}(\theta) := f_n(\theta \mid S_n = k) = \frac{\mathbb{P}\{S_n = k \mid \theta\}}{\int_0^1 \mathbb{P}\{S_n = k \mid x\}\, dx} = C_{n,k}\, \theta^k (1 - \theta)^{n-k},$$

where $C_{n,k}$ is the appropriate constant so that $f_{n,k}$ is a probability density. This is the *beta density* with parameters $k+1$ and $n-k+1$. The probability of

a success on the $(n+1)$st trial given that $S_n = k$ is the conditional expectation of θ given $S_n = k$. A little computation which we omit shows that

$$\mathbb{E}[\theta \mid S_n = k] = \int_0^1 \theta \, f_{n,k}(\theta) \, d\theta = \frac{k+1}{n+2} = \frac{S_n + 1}{n+2}.$$

These are exactly the transition probabilities for Polya's urn if we view $S_n + 1$ as the number of red balls in the urn (S_n is the number of red balls added to the urn). The martingale convergence theorem can now be viewed as the law of large numbers (1.10) for θ. Even though we do not initially know the value of θ (and hence treat it as a random variable) we know that the conditional value of θ given $\mathcal{F}_n$ approaches θ.

Example 1.4.1. We end with a simple example where the conditions of the martingale convergence theorem do not apply. Let $S_n = X_1 + \cdots + X_n$ be simple symmetric random walk starting at the origin as in the previous section. Then one can easily see that $\mathbb{E}[|S_n|] \to \infty$. For this example, with probability one

$$\limsup_{n \to \infty} S_n = \infty,$$

$$\liminf_{n \to \infty} S_n = -\infty.$$

1.5   Square integrable martingales

Definition A martingale M_n is called *square integrable* if for each n, $\mathbb{E}[M_n^2] < \infty$.

Note that this condition is not as strong as (1.9). We do not require that there exists a $C < \infty$ such that $\mathbb{E}[M_n^2] \leq C$ for each n. Random variables X, Y are *orthogonal* if $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. Independent random variables are orthogonal, but orthogonal random variables need not be independent. If $X_1, \ldots, X_n$ are pairwise orthogonal random variables with mean zero, then $\mathbb{E}[X_j X_k] = 0$ for $j \neq k$ and by expanding the square we can see that

$$\mathbb{E}\left[(X_1 + \cdots + X_n)^2\right] = \sum_{j=1}^n \mathbb{E}[X_j^2].$$

This can be thought of as a generalization of the Pythagorean theorem $a^2 + b^2 = c^2$ for right triangles. The increments of a martingale are not necessarily

independent, but for square integrable martingales they are orthogonal as we now show.

Proposition 1.5.1. *Suppose that M_n is a square integrable martingale with respect to $\{\mathcal{F}_n\}$. Then if $m < n$,*

$$\mathbb{E}\left[(M_{n+1} - M_n)(M_{m+1} - M_m)\right] = 0.$$

Moreover, for all n,

$$\mathbb{E}[M_n^2] = \mathbb{E}[M_0^2] + \sum_{j=1}^{n} \mathbb{E}\left[(M_j - M_{j-1})^2\right].$$

Proof. If $m < n$, then $M_{m+1} - M_m$ is $\mathcal{F}_n$-measurable, and hence

$$E\left[(M_{n+1} - M_n)(M_{m+1} - M_m) \mid \mathcal{F}_n\right]$$

$$= (M_{m+1} - M_m) E[M_{n+1} - M_n \mid \mathcal{F}_n] = 0.$$

Hence

$$\mathbb{E}\left[(M_{n+1} - M_n)(M_{m+1} - M_m)\right]$$

$$= \mathbb{E}\left[E\left[(M_{n+1} - M_n)(M_{m+1} - M_m) \mid \mathcal{F}_n\right]\right] = 0.$$

Also, if we set $M_{-1} = 0$,

$$
\begin{aligned}
M_n^2 &= \left[M_0 + \sum_{j=1}^{n}(M_j - M_{j-1})\right]^2 \\
&= M_0^2 + \sum_{j=1}^{n}(M_j - M_{j-1})^2 + \sum_{j \neq k}(M_j - M_{j-1})(M_k - M_{k-1}).
\end{aligned}
$$

Taking expectations of both sides gives the second conclusion. □

The natural place to discuss the role of orthogonality in the study of square integrable martingales is $L^2 = L^2(\Omega, \mathcal{F}, \mathbb{P})$, the space of square integrable random variables. This is a (real) Hilbert space under the inner product

$$(X, Y) = \mathbb{E}[XY].$$

Two mean zero random variables are orthogonal if and only if $(X, Y) = 0$. The conditional expectation has a nice interpretation in L^2. Suppose Y is a square integrable random variable and $\mathcal{G}$ is a sub- σ-algebra. Then $L^2(\Omega, \mathcal{G}, \mathbb{P})$ is a closed subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$ and the conditional expectation $\mathbb{E}[Y \mid \mathcal{G}]$ is the same as the Hilbert space projection onto the subspace. It can also be characterized as the $\mathcal{G}$-measurable random variable Z that minimizes the mean-squared error

$$\mathbb{E}[(Y - Z)^2].$$

The reason L^2 rather than L^p for other values of p is so useful is because of the inner product which gives the idea of orthogonality.

1.6   Integrals with respect to random walk

Suppose that $X_1, X_2, \ldots$ are independent, identically distributed random variables with mean zero and variance σ^2. The two main examples we will use are:

- Binomial or coin-tossing random variables,

$$\mathbb{P}\{X_j = 1\} = \mathbb{P}\{X_j = -1\} = \frac{1}{2},$$

  in which case $\sigma^2 = 1$.

- Normal increments where $X_j \sim N(0, \sigma^2)$. We write $Z \sim N(\mu, \sigma^2)$ if Z has a normal distribution with mean μ and variance σ^2.

Let $S_n = X_1 + \cdots + X_n$ and let $\{\mathcal{F}_n\}$ denote the filtration generated by $X_1, \ldots, X_n$. A sequence of random variables $J_1, J_2, \ldots$ is called *predictable (with respect to $\{\mathcal{F}_n\}$)* if for each n, J_n is $\mathcal{F}_{n-1}$-measurable. Recall that this is the condition that makes J_n allowable "bets" on the martingale in the sense of the discrete stochastic integral.

Suppose $J_1, J_2, \ldots$ is a predictable sequence with $\mathbb{E}[J_n^2] < \infty$ for each n. The integral of J_n with respect to S_n is defined by

$$Z_n = \sum_{j=1}^{n} J_j \, X_j = \sum_{j=1}^{n} J_j \, \Delta S_j.$$

There are three important properties that the integral satisfies.

- **Martingale property**. The integral Z_n is a martingale with respect to $\{\mathcal{F}_n\}$. We showed this in Section 1.2.

- **Linearity**. If J_n, K_n are predictable sequences and a, b constants, then $aJ_n + bK_n$ is a predictable sequence and

$$\sum_{j=1}^{n}(aJ_j + bK_j)\,\Delta S_j = a\sum_{j=1}^{n} J_j\,\Delta S_j + b\sum_{j=1}^{n} K_j\,\Delta S_j.$$

This is immediate.

- **Variance rule**

$$\mathrm{Var}\left[\sum_{j=1}^{n} J_j\,\Delta S_j\right] = \mathbb{E}\left[(\sum_{j=1}^{n} J_j\,\Delta S_j)^2\right] = \sigma^2\sum_{j=1}^{n}\mathbb{E}\left[J_j^2\right].$$

To see this we first use the orthogonality of martingale increments to write

$$\mathbb{E}\left[(\sum_{j=1}^{n} J_j\,\Delta S_j)^2\right] = \sum_{j=1}^{n}\mathbb{E}\left[J_j^2 X_j^2\right].$$

Since J_j is $\mathcal{F}_{j-1}$-measurable and X_j is independent of $\mathcal{F}_{j-1}$, we can see that

$$\begin{aligned}
\mathbb{E}\left[J_j^2 X_j^2\right] &= \mathbb{E}\left[E[J_j^2 X_j^2 \mid \mathcal{F}_{j-1}]\right]\\
&= \mathbb{E}\left[J_j^2\, E[X_j^2 \mid \mathcal{F}_{j-1}]\right]\\
&= \mathbb{E}\left[J_j^2\, \mathbb{E}[X_j^2]\right] = \sigma^2\,\mathbb{E}[J_j^2].
\end{aligned}$$

1.7 A maximal inequality

There is another result about martingales that we will use.

Theorem 1.7.1. *Suppose Y_n is a nonnegative submartingale with respect to $\{\mathcal{F}_n\}$, and*

$$\overline{Y}_n = \max\{Y_0, Y_1, \ldots, Y_n\}.$$

Then for every $a > 0$,

$$\mathbb{P}\{\overline{Y}_n \geq a\} \leq a^{-1}\,\mathbb{E}[Y_n].$$

Proof. Let T denote the smallest integer k such that $Y_k \geq a$. Then

$$\{\overline{Y}_n \geq a\} = \bigcup_{k=0}^{n} A_k, \quad A_k = \{T = k\}.$$

Note that A_k is $\mathcal{F}_k$-measurable. Since Y_n is a submartingale,

$$\mathbb{E}\left[Y_n 1_{A_k}\right] = \mathbb{E}\left[E(Y_n \mid \mathcal{F}_k) 1_{A_k}\right] \geq \mathbb{E}\left[Y_k 1_{A_k}\right].$$

By summing over k, we see that

$$\mathbb{E}\left[Y_n\right] \geq \mathbb{E}\left[Y_n 1\{\overline{Y}_n \geq a\}\right] = \sum_{k=0}^{n} \mathbb{E}\left[Y_n 1_{A_k}\right]$$

$$\geq \sum_{k=0}^{n} \mathbb{E}\left[Y_k 1_{A_k}\right] = \mathbb{E}\left[Y_T 1\{\overline{Y}_n \geq a\}\right] \geq a\,\mathbb{P}\{\overline{Y}_n \geq a\}. \quad \square$$

Corollary 1.7.2. *If M_n is a square integrable martingale with respect to $\{\mathcal{F}_n\}$ and*

$$\overline{M}_n = \max\left\{|M_0|, \ldots, |M_n|\right\},$$

then for every $a > 0$,

$$\mathbb{P}\left\{\overline{M}_n \geq a\right\} \leq a^{-2}\,\mathbb{E}\left[M_n^2\right].$$

Proof. In Exercise 1.15, it is shown that M_n^2 is a submartingale, and we can use the previous theorem.

1.8   Exercises

Exercise 1.1. Suppose we roll two dice, a red and a green one, and let X be the value on the red die and Y the value on the green die. Let $Z = XY$.

1. Let $W = E(Z \mid X)$. What are the possible values for W? Give the distribution of W.

2. Do the same exercise for $U = E(X \mid Z)$.

3. Do the same exercise for $V = E(Y \mid X, Z)$

Exercise 1.2. Suppose we roll two dice, a red and a green one, and let X be the value on the red die and Y the value on the green die. Let $Z = X/Y$.

1. Find $E[(X + 2Y)^2 \mid X]$.

2. Find $E[(X + 2Y)^2 \mid X, Z]$.

3. Find $E[X + 2Y \mid Z]$.

4. Let $W = E[Z \mid X]$. What are the possible values for W? Give the distribution of W.

Exercise 1.3. Suppose $X_1, X_2, \ldots$ are independent random variables with

$$\mathbb{P}\{X_j = 2\} = 1 - \mathbb{P}\{X_j = -1\} = \frac{1}{3}.$$

Let $S_n = X_1 + \cdots + X_n$ and let $\mathcal{F}_n$ denote the information in $X_1, \ldots, X_n$.

1. Find $\mathbb{E}[S_n], \mathbb{E}[S_n^2], \mathbb{E}[S_n^3]$.

2. If $m < n$, find

$$E[S_n \mid \mathcal{F}_m], \quad E[S_n^2 \mid \mathcal{F}_m], \quad E[S_n^3 \mid \mathcal{F}_m].$$

3. If $m < n$, find $E[X_m \mid S_n]$.

Exercise 1.4. Repeat Exercise 1.3 assuming that

$$\mathbb{P}\{X_j = 3\} = \mathbb{P}\{X_j = -1\} = \frac{1}{2}.$$

Exercise 1.5. Suppose $X_1, X_2, \ldots$ are independent random variables with

$$\mathbb{P}\{X_j = 1\} = \mathbb{P}\{X_j = -1\} = \frac{1}{2}.$$

Let $S_n = X_1 + \cdots + X_n$. Find

$$E\left(\sin S_n \mid S_n^2\right).$$

Exercise 1.6. In this exercise, we consider simple, nonsymmetric random walk. Suppose $1/2 < q < 1$ and $X_1, X_2, \ldots$ are independent random variables with

$$\mathbb{P}\{X_j = 1\} = 1 - \mathbb{P}\{X_j = -1\} = q.$$

Let $S_0 = 0$ and $S_n = X_1 + \cdots + X_n$. Let $\mathcal{F}_n$ denote the information contained in $X_1, \ldots, X_n$.

1. Which of these is S_n: martingale, submartingale, supermartingale (more than one answer is possible)?

2. For which values of r is $M_n = S_n - rn$ a martingale?

3. Let $\theta = (1 - q)/q$ and let
$$M_n = \theta^{S_n}.$$

   Show that M_n is a martingale.

4. Let a, b be positive integers, and
$$T_{a,b} = \min\{j : S_j = b \text{ or } S_j = -a\}.$$

   Use the optional sampling theorem to determine
$$\mathbb{P}\left\{S_{T_{a,b}} = b\right\}.$$

5. Let $T_a = T_{a,\infty}$. Find
$$\mathbb{P}\{T_a < \infty\}.$$

Exercise 1.7. Suppose two people want to play a game in which person A has probability $2/3$ of winning. However, the only thing that they have is a fair coin which they can flip as many times as they want. They wish to find a method that requires only a finite number of coin flips.

1. Give one method to use the coins to simulate an experiment with probability $2/3$ of success. The number of flips needed can be random, but it must be finite with probability one.

2. Suppose $K < \infty$. Explain why there is no method such that with probability one we flip the coin at most K times.

Exercise 1.8. Repeat the last exercise with $2/3$ replaced by $1/\pi$.

Exercise 1.9. Let $X_1, X_2, \ldots$ be independent, identically distributed random variables with
$$\mathbb{P}\{X_j = 2\} = \frac{1}{3}, \quad \mathbb{P}\{X_j = \frac{1}{2}\} = \frac{2}{3}.$$
Let $M_0 = 1$ and for $n \geq 1$, $M_n = X_1 X_2 \cdots X_n$.

1. Show that M_n is a martingale.

2. Explain why M_n satisfies the conditions of the martingale convergence theorem.

3. Let $M_\infty = \lim_{n \to \infty} M_n$. Explain why $M_\infty = 0$. (Hint: there are at least two ways to show this. One is to consider $\log M_n$ and use the law of large numbers. Another is to note that with probability one M_{n+1}/M_n does not converge.)

4. Use the optional sampling theorem to determine the probability that M_n ever attains a value as large as 64.

5. Does there exist a $C < \infty$ such that $\mathbb{E}[M_n^2] \le C$ for all n?

Exercise 1.10. Let $X_1, X_2, \ldots$ be independent, identically distributed random variables with

$$\mathbb{P}\{X_j = 1\} = q, \quad \mathbb{P}\{X_j = -1\} = 1 - q.$$

Let $S_0 = 0$ and for $n \ge 1$, $S_n = X_1 + X_2 + \cdots + X_n$. Let $Y_n = e^{S_n}$.

1. For which value of q is Y_n a martingale?

2. For the remaining parts of this exercise assume q takes the value from part 1. Explain why Y_n satisfies the conditions of the martingale convergence theorem.

3. Let $Y_\infty = \lim_n Y_n$. Explain why $Y_\infty = 0$. (Hint: there are at least two ways to show this. One is to consider $\log Y_n$ and use the law of large numbers. Another is to note that with probability one Y_{n+1}/Y_n does not converge.)

4. Use the optional sampling theorem to determine the probability that Y_n ever attains a value greater than 100.

5. Does there exist a $C < \infty$ such that $\mathbb{E}[Y_n^2] \le C$ for all n?

Exercise 1.11. This exercise concerns Polya's urn and has a computing/simulation component. Let us start with one red and one green ball as in the lecture and let M_n be the fraction of red balls at the nth stage.

1. Show that the distribution of M_n is uniform on the set

$$\left\{ \frac{1}{n+2}, \frac{2}{n+2}, \dots, \frac{n+1}{n+2} \right\}.$$

   (Use mathematical induction, that is, note that it is obviously true for $n = 0$ and show that if it is true for n then it is true for $n + 1$.)

2. Write a short program that will simulate this urn. Each time you run the program note the fraction of red balls after 600 draws and after 1200 draws. Compare the two fractions. Then, repeat this twenty times.

Exercise 1.12. Consider the martingale betting strategy as discussed in Section 1.2. Let W_n be the "winnings" at time n, which for positive n equals either 1 or $1 - 2^n$.

1. Is W_n a square integrable martingale?

2. If $\Delta_n = W_n - W_{n-1}$ what is $\mathbb{E}[\Delta_n^2]$?

3. What is $\mathbb{E}[W_n^2]$?

4. What is $E(\Delta_n^2 \mid \mathcal{F}_{n-1})$?

Exercise 1.13. Suppose $S_n = X_1 + \cdots + X_n$ is simple random walk starting at 0. For any K, let
$$T = \min\{n : |S_n| = K\}.$$

- Explain why for every j,

$$\mathbb{P}\{T \leq j + K \mid T > j\} \geq 2^{-K}.$$

- Show that there exists $c < \infty, \alpha > 0$ such that for all j,

$$\mathbb{P}\{T > j\} \leq c e^{-\alpha j}.$$

   Conclude that $\mathbb{E}[T^r] < \infty$ for every $r > 0$.

- Let $M_n = S_n^2 - n$. Show there exists $C < \infty$ such that for all n,

$$\mathbb{E}\left[M_{n \wedge T}^2\right] \leq C.$$

Exercise 1.14. Suppose that $X_1, X_2, \ldots$ are independent random variables with $\mathbb{E}[X_j] = 0$, $\operatorname{Var}[X_j] = \sigma_j^2$, and suppose that

$$\sum_{n=1}^{\infty} \sigma_n^2 < \infty.$$

Let $S_0 = 0$ and $S_n = X_1 + \cdots + X_n$ for $n > 0$. Let $\mathcal{F}_n$ denote the information contained in $X_1, \ldots, X_n$.

- Show that S_n is a martingale with respect to $\{\mathcal{F}_n\}$.

- Show that there exists $C < \infty$ such that for all n, $\mathbb{E}[S_n^2] \leq C$.

- Show that with probability one the limit

$$S_\infty = \lim_{n \to \infty} S_n,$$

  exists.

- Show that

$$\mathbb{E}[S_\infty] = 0, \quad \operatorname{Var}[S_\infty] = \sum_{n=1}^{\infty} \sigma_n^2.$$

Exercise 1.15.

- Suppose Y is a random variable and ϕ is a convex function, that is, if $0 \leq \lambda \leq 1$,

$$\phi(\lambda x + (1 - \lambda) y) \leq \lambda \phi(x) + (1 - \lambda) \phi(y).$$

  Suppose that $\mathbb{E}[|\phi(Y)|] < \infty$. Show that $E(\phi(Y) \mid X) \geq \phi(E(Y \mid X))$. (Hint: you will need to review Jensen's inequality.)

- Show that if M_n is a martingale with respect to $\{\mathcal{F}_n\}$ and $r \geq 1$, then $Y_n = |M_n|^r$ is a submartingale.

Chapter 2

Brownian motion

2.1   Limits of sums of independent variables

We will discuss two major results about sums of random variables that we
hope the reader has seen. They both discuss the limit distribution of

$$X_1 + X_2 + \cdots + X_n$$

where $X_1, X_2, \ldots, X_n$ are independent, identically distributed random vari-
ables.

Suppose $X_1, X_2, \ldots$ have mean μ and variance $\sigma^2 < \infty$. Let

$$Z_n = \frac{(X_1 + \cdots + X_n) - n\mu}{\sigma \sqrt{n}}.$$

We let Φ denote the standard normal distribution function,

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

While this function cannot be written down explicitly, the numerical values
are easily accessible in tables and computer software packages.

Theorem 2.1.1 (Central Limit Theorem). *As $n \to \infty$, the distribution of
Z_n approaches a standard normal distribution. More precisely, if $a < b$, then*

$$\lim_{n \to \infty} \mathbb{P}\{a \leq Z_n \leq b\} = \Phi(b) - \Phi(a).$$

This amazing theorem states that no matter what distribution we choose for the X_j, then as long as the distribution has a finite variance, the scaled random variables approach a normal distribution. This is what is referred to in physics literature as a *universality* result and in the mathematics literature as an *invariance principle*. The assumption that the random variables are identically distributed can be relaxed somewhat. We will not go into details here, but whenever a quantity can be written as a sum of independent (or at least not too dependent) increments, *all of which are small compared to the sum*, then the limiting distribution is normal. This is why it is often reasonable to assume normal distributions in nature.

To see a complete proof of Theorem 2.1.1, consult any book on measure-theoretic probability. We will sketch part of the proof which shows how the normal distribution arises. Without loss of generality, we can assume that $\mu = 0, \sigma^2 = 1$ for otherwise we consider $Y_j = (X_j - \mu)/\sigma$. Let $\phi(t) = \mathbb{E}[e^{itX_j}]$ denote the characteristic function of the increments. Then, ϕ can be expanded near the origin,

$$
\begin{aligned}
\phi(t) &= 1 + i\,\mathbb{E}\left[X_j\right] t + \frac{(i)^2}{2}\,\mathbb{E}\left[X_j^2\right] t^2 + o(t^2) \\
&= 1 - \frac{t^2}{2} + o(t^2),
\end{aligned}
$$

where $o(t^2)$ denotes a function such that $|o(t^2)|/t^2 \to 0$ as $t \to 0$. Using the independence of the X_j, we see that the characteristic function of Z_n is

$$
\phi_{Z_n}(t) = \phi(t/\sqrt{n})^n = \left[1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right]^n \longrightarrow e^{-t^2/2}.
$$

The right-hand side is the characteristic function of a standard normal random variable.

When one considers sums of independent random variables where a few terms contribute the bulk of the sum, then one does not expect to get normal limits. The prototypical example of a nonnormal limit is the Poisson distribution. Suppose $\lambda > 0$ and $X_1^{(n)}, X_2^{(n)}, \ldots, X_n^{(n)}$ are independent random variables each with

$$
\mathbb{P}\left\{X_j^{(n)} = 1\right\} = 1 - \mathbb{P}\left\{X_j^{(n)} = 0\right\} = \frac{\lambda}{n}.
$$

Let

$$Y_n = X_1^{(n)} + \cdots + X_n^{(n)},$$

and note that for each n, $\mathbb{E}[Y_n] = \lambda$. Recall that a random variable Y has a *Poisson distribution with mean* λ if for each nonnegative integer k,

$$\mathbb{P}\{Y = k\} = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Theorem 2.1.2 (Convergence to the Poisson distribution). *As $n \to \infty$, the distribution of Y_n approaches a Poisson distribution with mean λ. More precisely, for every nonnegative integer k,*

$$\lim_{n \to \infty} \mathbb{P}\{Y_n = k\} = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Proof. For each n, Y_n has a binomial distribution with parameters n and λ/n, and hence

$$\lim_{n \to \infty} \mathbb{P}\{Y_n = k\}$$

$$= \lim_{n \to \infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{\lambda^k}{k!} \lim_{n \to \infty} \frac{n(n-1)\cdots(n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^{-k} \left(1 - \frac{\lambda}{n}\right)^n$$

Since we are fixing k and letting $n \to \infty$, one can see that

$$\lim_{n \to \infty} \frac{n(n-1)\cdots(n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^{-k} = 1,$$

and in calculus one learns the limit

$$\lim_{n \to \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}.$$

$\square$

We will be taking limits of processes as the time increment goes to zero. The kind of limit we expect will depend on the assumptions. When we take limits of random walks with finite variance, then we are in the regime of the

central limit theorem, and we will expect normal distributions. Also, because all of the increments are small with respect to the sum, the limit process will have paths that are continuous. Here the limit is *Brownian motion* which we discuss in this section.

In the Poisson case, the limit distribution will not have continuous paths but rather will be a *jump process*. The prototypical case is the Poisson process with intensity λ. In this case, N_t denotes the number of occurrences of an event by time t. The function $t \mapsto N_t$ takes on nonnegative integer values and the jumps are always of size $+1$. It satisfies the following conditions.

- For each $s < t$, the random variable $N_t - N_s$ has a Poisson distribution with parameter $\lambda(t - s)$.

- For each $s < t$, the random variable $N_t - N_s$ is independent of the random variables $\{N_r : 0 \le r \le s\}$.

We discuss this further in Section 6.2.

2.2   Multivariate normal distribution

Although the normal or Gaussian distribution is a little inconvenient in the sense that the distribution function cannot be computed exactly, there are many other aspects that make the distribution very convenient. In particular, when dealing with many variables, assuming a *joint or multivariate* normal distribution makes computations tractable. In this section we will give the basic definitions. Roughly speaking, the basic assumption is that if $(X_1, \ldots, X_n, Y)$ have a joint normal distribution then not only does each variable have a normal distribution but also, the conditional distribution of Y given $X_1, \ldots, X_n$ is normal with mean $E[Y|X_1, \ldots, X_n]$ and a variance that depends on the joint distribution but not on the observed data points. There are a number of equivalent ways to define a joint normal distribution. We will use the following.

Definition A finite sequence of random variables $(X_1, \ldots, X_n)$ has a *joint (or multivariate) normal (or Gaussian)* distribution if they are linear combinations of independent standard normal random variables. In other words, if there exist independent random variables $(Z_1, \ldots, Z_m)$, each $N(0, 1)$, and constants m_j, a_{jk} such that for $j = 1, \ldots, n$,

$$X_j = m_j + a_{j1} Z_1 + a_{j2} Z_2 + \cdots + a_{jm} Z_m.$$

Clearly $\mathbb{E}[X_j] = m_j$. Let us consider the case of mean-zero (also called *centered*) joint normals, in which case the equation above can be written in matrix form

$$\mathbf{X} = A\,\mathbf{Z},$$

where

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_m \end{pmatrix},$$

and A is the $n \times m$ matrix with entries a_{jk}. Each X_j is a normal random variable with mean zero and variance

$$\mathbb{E}[X_j^2] = a_{j1}^2 + \cdots + a_{jm}^2.$$

More generally, the covariance of X_j and X_k is given by

$$\mathrm{Cov}(X_j, X_k) = \mathbb{E}[X_j X_k] = \sum_{l=1}^{m} a_{jl} a_{kl}.$$

Let $\Gamma = AA^T$ be the $n \times n$ matrix whose entries are

$$\Gamma_{jk} = \mathbb{E}[X_j X_k].$$

Then Γ is called the *covariance matrix.*

We list some important properties. Assume $(X_1, \ldots, X_n)$ has a joint normal distribution with mean zero and covariance matrix Γ.

- Each X_j has a normal distribution. In fact, if $b_1, \ldots, b_n$ are constants, then

$$b_1\,X_1 + \cdots + b_n\,X_n,$$

has a normal distribution. We can see this easily since we can write the sum above as a linear combination of the independent normals $Z_1, \ldots, Z_m$.

- The matrix Γ is symmetric, $\Gamma_{jk} = \Gamma_{kj}$. Moreover, it is *positive semi-definite* which means that if $\mathbf{b} = (b_1, \ldots, b_n)$ is a vector in $\mathbb{R}^n$, then

$$\mathbf{b} \cdot \Gamma \mathbf{b} = \sum_{j=1}^{n} \sum_{k=1}^{n} \Gamma_{jk} b_j b_k \geq 0. \qquad (2.1)$$

(If the $\geq$ is replaced with > 0 for all $\mathbf{b} = (b_1, \ldots, b_n) \neq (0, \ldots, 0)$, then the matrix is called *positive definite*.) The inequality (2.1) can be derived by noting that the left-hand side is the same as

$$\mathbb{E}\left[(b_1 X_1 + \cdots + b_n X_n)^2\right],$$

which is clearly nonnegative.

- If Γ is a positive semidefinite, symmetric matrix, then it is the co-variance matrix for a joint normal distribution. The proof of this fact, which we omit, uses linear algebra to deduce that there exists an $n \times n$ matrix A with $AA^T = \Gamma$. (The A is not unique.)

- The distribution of a mean-zero joint normal is determined by the co-variance matrix Γ.

In order to show that the covariance matrix Γ determines the dis-tribution of a mean-zero joint normal, we compute the characteristic function. Suppose that $\Gamma = AA^T$ where A is $n \times n$. Using the indepen-dence of $Z_1, \ldots, Z_n$ and the characteristic function of the standard normal, $\mathbb{E}[e^{itZ_k}] = e^{-t^2/2}$, we see that the characteristic function of

$(X_1, \ldots, X_n)$ is

$$
\begin{aligned}
\phi(\theta_1, \ldots, \theta_n) &= \mathbb{E}\left[\exp\left\{i(\theta_1 X_1 + \ldots + \theta_n X_n)\right\}\right] \\
&= \mathbb{E}\left[\exp\left\{i\sum_{j=1}^{n}\theta_j \sum_{k=1}^{n} a_{jk} Z_k\right\}\right] \\
&= \mathbb{E}\left[\exp\left\{i\sum_{k=1}^{n} Z_k\left(\sum_{j=1}^{n}\theta_j a_{jk}\right)\right\}\right] \\
&= \prod_{k=1}^{n}\mathbb{E}\left[\exp\left\{iZ_k\left(\sum_{j=1}^{n}\theta_j a_{jk}\right)\right\}\right] \\
&= \exp\left\{-\frac{1}{2}\sum_{k=1}^{n}\left(\sum_{j=1}^{n}\theta_j a_{jk}\right)^2\right\} \\
&= \exp\left\{-\frac{1}{2}\sum_{k=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}\theta_j \theta_i a_{jk} a_{lk}\right\} \\
&= \exp\left\{-\frac{1}{2}\theta A A^T \theta^T\right\} \\
&= \exp\left\{-\frac{1}{2}\theta \Gamma \theta^T\right\}
\end{aligned}
$$

where we write $\theta = (\theta_1, \ldots, \theta_n)$. Even though we used A, which is not unique, in our computation, the answer only involves Γ. Since the characteristic function determines the distribution, the distribution depends only on the covariance matrix.

- If Γ is invertible, then $(X_1, \ldots, X_n)$ has a density. We write it in the case that the random variables have mean $\mathbf{m} = (m_1, \ldots, m_n)$,

$$
f(x_1, \ldots, x_n) = f(\mathbf{x}) =
$$

$$
\frac{1}{(2\pi)^{n/2}\sqrt{\det \Gamma}}\exp\left\{-\frac{(\mathbf{x} - \mathbf{m}) \cdot \Gamma^{-1}(\mathbf{x} - \mathbf{m})^T}{2}\right\}.
$$

Sometimes this density is used as a definition of a joint normal. The formula for the density looks messy, but note that if $n = 1$, $\mathbf{m} = m$, $\Gamma = [\sigma^2]$, then the right-hand side is the density of a $N(m, \sigma^2)$ random variable.

- If (X_1, X_2) have a mean-zero joint normal density, and $\mathbb{E}(X_1 X_2) = 0$, then X_1, X_2 are independent random variables. To see this let $\sigma_j^2 = \mathbb{E}[X_j^2]$. Then the covariance matrix of (X_1, X_2) is the diagonal matrix with diagonal entries σ_j^2. If (Z_1, Z_2) are independent $N(0, 1)$ random variables and $Y_1 = \sigma_1 Z_1, Y_2 = \sigma_2 Z_2$, then by our definition (Y_1, Y_2) are joint normal with the same covariance matrix. Since the covariance matrix determines the distribution, X_1, X_2 must be independent,

It is a special property about joint normal random variables that orthogonality implies independence. In our construction of Brownian motion, we will use a particular case, that we state as a lemma.

Proposition 2.2.1. *Suppose X, Y are independent $N(0, 1)$ random variables and*

$$Z = \frac{X + Y}{\sqrt{2}}, \qquad W = \frac{X - Y}{\sqrt{2}}.$$

Then Z and W are independent $N(0, 1)$ random variables.

Proof. By definition (Z, W) has a joint normal distribution and Z, W clearly have mean 0. Using $\mathbb{E}[X^2] = \mathbb{E}[Y^2] = 1$ and $\mathbb{E}[XY] = 0$, we get

$$\mathbb{E}[Z^2] = 1, \quad \mathbb{E}[W^2] = 1, \quad \mathbb{E}[ZW] = 0.$$

Hence the covariance matrix for (Z, W) is the identity matrix and this is the covariance matrix for independent $N(0, 1)$ random variables.   □

2.3   Limits of random walks

Brownian motion can be viewed as the limit of random walk as the time and space increments tend to zero. It is necessary to be careful about how the limit is taken. Suppose $X_1, X_2, \ldots$ are independent random variables with $\mathbb{P}\{X_j = 1\} = \mathbb{P}\{X_j = -1\} = 1/2$ and let

$$S_n = X_1 + \cdots + X_n$$

be the corresponding random walk. Here we have chosen time increment $\Delta t = 1$ and space increment $\Delta x = 1$. Suppose we choose $\Delta t = 1/N$ where N is a large integer. Hence, we view the process at times

$$\Delta t, \, 2\Delta t, \, 3\Delta t, \cdots,$$

and at each time we have a jump of $\pm \Delta x$. At time $1 = N\Delta t$, the value of the process is

$$W_1^{(N)} = \Delta x \, (X_1 + \cdots + X_N).$$

We would like to choose Δx so that $\mathrm{Var}[W_1^{(N)}] = 1$, and since

$$
\begin{aligned}
\mathrm{Var}\left[\Delta x \, (X_1 + \cdots + X_N)\right] &= (\Delta x)^2 \left[\mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_N)\right] \\
&= (\Delta x)^2 \, N,
\end{aligned}
$$

we see that we need to choose

$$\Delta x = \sqrt{1/N} = \sqrt{\Delta t}.$$

We also know from the central limit theorem, that if N is large, then the distribution of

$$\frac{X_1 + \cdots + X_N}{\sqrt{N}},$$

is approximately $N(0, 1)$.

Brownian motion could be defined formally as the limit of random walks, but there are subtleties in describing the kind of limit. In the next section, we define it directly using the idea of "continuous random motion". However, the random walk intuition is useful to retain.

2.4 Brownian motion

Brownian motion or the *Wiener process* is a model of random continuous motion. We will start by making the assumptions that underlie the phrase "random continuous motion". Let $B_t = B(t)$ be the value at time t. For each t, B_t is a random variable.[1] A collection of random variables indexed by time is called a *stochastic process*. We can view the process in two different ways:

- For each t, there is a random variable B_t, and there are correlations between the values at different times.

---

[1]In this book and usually in the financial world, the terms Brownian motion and Wiener process are synonymous. However, in the scientific world, the word Brownian motion is often used for a physical process for which what we will describe is one possible mathematical model. The term Wiener process always refers to the model we define. The letter W_t is another standard notation for Brownian motion/Wiener process and is more commonly used in financial literature. We will use both B_t and W_t later in the book when we need two notations.

- The function $t \mapsto B(t)$ is a random function. In other words, it is a random variable whose value is a function.

There are three major assumptions about the random variables B_t.

- **Stationary increments**. If $s < t$, then the distribution of $B_t - B_s$ is the same as that of $B_{t-s} - B_0$.

- **Independent increments**. If $s < t$, the random variable $B_t - B_s$ is independent of the values B_r for $r \le s$.

- **Continuous paths**. The function $t \mapsto B_t$ is a continuous function of t.

We often assume $B_0 = 0$ for convenience, but we can also take other initial conditions. All of the assumptions above are very reasonable for a model of random continuous motion. However, it is not obvious that these are enough assumptions to characterize our process uniquely. It turns out that they do up to two parameters. One can prove (see Theorem 6.8.4), that if B_t is a process satisfying the three conditions above, then the distribution of B_t for each t must be normal. Suppose B_t is such a process, and let m, σ^2 be the mean and variance of B_1. If $s < t$, then independent, identically distributed increments imply that

$$\mathbb{E}[B_t] = \mathbb{E}[B_s] + \mathbb{E}[B_t - B_s] = \mathbb{E}[B_s] + \mathbb{E}[B_{t-s}],$$

$$\mathrm{Var}[B_t] = \mathrm{Var}[B_s] + \mathrm{Var}[B_t - B_s] = \mathrm{Var}[B_s] + \mathrm{Var}[B_{t-s}].$$

Using this relation, we can see that $\mathbb{E}[B_t] = tm, \mathrm{Var}[B_t] = t\sigma^2$. At this point, we have only shown that *if* a process exists, then the increments must have a normal distribution. We will show that such a process exists. It will be convenient to put the normal distribution in the definition.

Definition A stochastic process B_t is called a *(one-dimensional) Brownian motion* with *drift m* and *variance (parameter)* σ^2 starting at the origin if it satisfies the following.

- $B_0 = 0$.

- For $s < t$, the distribution of $B_t - B_s$ is normal with mean $m(t - s)$ and variance $\sigma^2(t - s)$.

- If $s < t$, the random variable $B_t - B_s$ is independent of the values B_r for $r \leq s$.

- With probability one, the function $t \mapsto B_t$ is a continuous function of t.

If $m = 0, \sigma^2 = 1$, then B_t is called a *standard Brownian motion.*

Recall that if Z has a $N(0,1)$ distribution and $Y = \sigma Z + m$, then Y has a $N(m, \sigma^2)$ distribution. Given that it is easy to show the following.

- If B_t is a standard Brownian motion and

$$Y_t = \sigma\, B_t + mt,$$

then Y_t is a Brownian motion with drift m and variance σ^2.

Indeed, one just checks that it satisfies the conditions above. Hence, in order to establish the *existence* of Brownian motion, it suffices to construct a standard Brownian motion.

There is a mathematical challenge in studying stochastic processes indexed by continuous time. The problem is that the set of positive real numbers is *uncountable*, that is, the elements cannot be enumerated $t_1, t_2, \ldots$. The major axiom of probability theory is the fact that if $A_1, A_2, \ldots$ is a *countable* sequence of disjoint events, then

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

This rule does not hold for uncountable unions. An example that we have all had to deal with arises with continuous random variables. Suppose, for instance, that Z has a $N(0,1)$ distribution. Then for each $x \in \mathbb{R}$,

$$\mathbb{P}\{Z = x\} = 0.$$

However,

$$1 = \mathbb{P}\{Z \in \mathbb{R}\} = \mathbb{P}\left[\bigcup_{x \in \mathbb{R}} A_x\right],$$

where A_x denotes the event $\{Z = x\}$. The events A_x are disjoint, each with probability zero, but it is not the case that

$$\mathbb{P}\left[\bigcup_{x\in\mathbb{R}} A_x\right] = \sum_{x\in\mathbb{R}} \mathbb{P}(A_x) = 0.$$

In constructing Brownian motion, we use the fact that if $g : [0, \infty) \to \mathbb{R}$ is a *continuous* function and we know the value of g on a countable, dense set, such as the *dyadic rationals*

$$\left\{ \frac{k}{2^n} : k = 0, 1, \ldots; n = 0, 1, \ldots \right\},$$

then we know the value at every t. Indeed, we need only find a sequence of dyadic rationals t_n that converge to t, and let

$$g(t) = \lim_{t_n \to t} g(t_n).$$

This is fine if *a priori* we know the function g is continuous. Our strategy for constructing Brownian motion will be:

- First define the process B_t when t is a dyadic rational.

- Prove that with probability one, the function $t \mapsto B_t$ is continuous on the dyadics (this is the hardest step, and we need some care in the definition of continuity).

- Extend B_t to other t by continuity.

The next section shows that one can construct a Brownian motion. The reader can skip this section and just have faith that such a process exists.

2.5   Construction of Brownian motion

We will show how to construct Brownian motion. For ease, we will consider $B_t, 0 \le t \le 1$. Once we know how to construct this, we can take a countable collection of such processes and connect them appropriately to get $B_t, 0 \le t < \infty$. We start with a probability space

$(\Omega, \mathcal{F}, \mathbb{P})$ which is large enough so that it contains a countable collection of independent standard normal random variables. Let

$$\mathcal{D}_n = \left\{ \frac{k}{2^n} : k = 0, 1, \ldots, 2^n \right\},$$

denote the dyadic rationals in $[0, 1]$ with denominator 2^n and let $\mathcal{D} = \cup_n \mathcal{D}$. Since $\mathcal{D}$ is a countable set, we can assume that our collection of $N(0, 1)$ random variables

$$\{Z_q : q \in \mathcal{D}\}$$

is indexed by $\mathcal{D}$.

We will use the independent random variables $\{Z_q\}$ to define the Brownian motion $B_q, q \in \mathcal{D}$. We start by defining B_0, B_1, and then $B_{1/2}$, and then $B_{1/4}$ and $B_{3/4}$, and so forth, by always subdividing our intervals. We start with $B_0 = 0$ and we let $B_1 = Z_1$ which is clearly $N(0, 1)$. We then define

$$B_{1/2} = \frac{B_1}{2} + \frac{Z_{1/2}}{2},$$

and hence

$$B_1 - B_{1/2} = \frac{B_1}{2} - \frac{Z_{1/2}}{2}.$$

We think of the definition of $B_{1/2}$ as being $E[B_{1/2} \mid B_1]$ plus some independent randomness. Using Proposition 2.2.1, we can see that $B_{1/2}$ and $B_1 - B_{1/2}$ are independent random variables, each $N(0, 1/2)$. We continue this splitting. If $q = (2k+1)/2^{n+1} \in \mathcal{D}_{n+1} \setminus \mathcal{D}_n$, we define

$$B_q = B_{k2^{-n}} + \frac{B_{(k+1)2^{-n}} - B_{k2^{-n}}}{2} + \frac{Z_q}{2^{(n+2)/2}}.$$

This formula looks a little complicated, but we can again view this as

$$B_q = E[B_q \mid B_{k2^{-n}}, B_{(k+1)2^{-n}}] + (\text{independent randomness}),$$

where the variance of the new randomness is chosen appropriately. By repeated use of Proposition 2.2.1, we can see that for each n, the random variables

$$\left\{ B_{k2^{-n}} - B_{(k-1)2^{-n}} : k = 1, \ldots, 2^n \right\}$$

are independent, each with $N(0, 2^{-n})$ distribution. From this we can see that $\{B_q : q \in \mathcal{D}\}$ satisfies the first three properties in the definition of a Brownian motion.

We will define B_t for other t by continuity. To do this we prove a theorem that states that the Brownian motion restricted to $\mathcal{D}$ has *uniformly continuous* paths. In other words, we show that with probability one, for every $\epsilon > 0$, there exists $\delta > 0$ such that if $|s - t| \leq \delta$ and $s, t \in \mathcal{D}$, then $|B_s - B_t| < \epsilon$. Equivalently, if we define the random variable

$$K_n = \sup \left\{ |B_s - B_t| : s, t \in \mathcal{D}, |s - t| \leq 2^{-n} \right\},$$

then with probability one, $K_n \to 0$ as $n \to \infty$. Note that K_n is the supremum of a countable number of random variables and hence is a well-defined random variable (perhaps taking the value ∞). We will prove a much stronger version of this.

Proposition 2.5.1. *With probability one,*

$$\limsup_{n \to \infty} \frac{2^{n/2}}{\sqrt{n}} K_n < \infty. \tag{2.2}$$

In particular, $K_n \to 0$.

In order to prove this proposition, it is easier to consider another sequence of random variables

$$J_n = \max_{j=1,\dots,2^n} Y(j, n)$$

where $Y(j, n)$ equals

$$\sup \left\{ |B_q - B_{(j-1)2^{-n}}| : q \in \mathcal{D}, (j-1)2^{-n} \leq q \leq j2^{-n} \right\}.$$

A simple argument using the triangle inequality shows that $K_n \leq 3J_n$. It turns out J_n is easier to analyze. For any $\epsilon > 0$,

$$\mathbb{P}\{J_n \geq \epsilon\} \leq \sum_{j=1}^{2^n} \mathbb{P}\{Y(j, n) \geq \epsilon\} = 2^n \, \mathbb{P}\{Y(1, n) \geq \epsilon\}.$$

Also the distribution of

$$Y(1, n) = \sup \left\{ |B_q| : q \in \mathcal{D}, q \leq 2^{-n} \right\}.$$

is the same as that of $2^{-n/2}Y$ where

$$Y = Y(1,0) = \sup\{|B_q| : q \in \mathcal{D}\}.$$

Using this we see that

$$\mathbb{P}\{J_n \geq C\sqrt{n}\,2^{-n/2}\} \leq 2^n\,\mathbb{P}\{Y \geq C\sqrt{n}\}.$$

We will show below that if $C > \sqrt{2\log 2}$, then

$$\sum_{n=1}^{\infty} 2^n\,\mathbb{P}\{Y \geq C\sqrt{n}\} < \infty. \tag{2.3}$$

The Borel-Cantelli lemma then implies that with probability one, the event $\{J_n \geq 2^{-n/2}n\}$ happens for only finitely many values of n. In particular,

$$\lim_{n\to\infty} 2^{n/2}\,n^{-1}\,J_n = 0,$$

which implies (2.2). To get our estimate, we will need the following lemma which is a form of the "reflection principle" for Brownian motion.

Lemma 2.5.2. *For every $a \geq 4$,*

$$\mathbb{P}\{Y > a\} \leq 4\,\mathbb{P}\{B_1 \geq a\} \leq e^{-a^2/2}.$$

To prove this lemma, let

$$Y_n = \max\{|B_q| : q \in \mathcal{D}_n\}.$$

Then

$$\mathbb{P}\{Y > a\} = \lim_{n\to\infty} \mathbb{P}\{Y_n > a\},$$

and hence it suffices to prove the inequality for each n. Fix n and let A_k be the event that

$$|B_{k2^{-n}}| > a, \quad |B_{j2^{-n}}| \leq a,\ j = 1,\ldots,k-1.$$

Then the event $\{Y_n > a\}$ can be written as a disjoint union

$$\{Y_n > a\} = \bigcup_{k=1}^{2^n} A_k.$$

The important observation is that if $|B_{k2^{-n}}| > a$, then with probability at least $1/2$, we have $|B_1| > a$. Indeed, if $B_{k2^{-n}}$ and $B_1 - B_{k2^{-n}}$ have the same sign, which happens with probability $1/2$, then $|B_1| \geq |B_{k2^{-n}}|$. Therefore,

$$\mathbb{P}\left[A_k \cap \{|B_1| > a\}\right] \geq \frac{1}{2}\,\mathbb{P}(A_k),$$

and hence

$$
\begin{aligned}
\mathbb{P}\{|B_1| > a\} &= \sum_{k=1}^{2^n} \mathbb{P}\left[A_k \cap \{|B_1| > a\}\right] \\
&\geq \frac{1}{2} \sum_{k=1}^{2^n} \mathbb{P}\left[A_k\right] \\
&= \frac{1}{2}\,\mathbb{P}\{Y_n > a\}.
\end{aligned}
$$

Here we use the fact that the events A_k are disjoint. Therefore, since B_1 has a $N(0,1)$ distribution,

$$\mathbb{P}\{Y_n > a\} \leq 2\,\mathbb{P}\{|B_1| > a\} = 4\,\mathbb{P}\{B_1 > a\} = \frac{4}{\sqrt{2\pi}} \int_a^\infty e^{-x^2/2}\,dx.$$

To estimate the integral for $a \geq 4$, we write

$$
\begin{aligned}
\frac{4}{\sqrt{2\pi}} \int_a^\infty e^{-x^2/2}\,dx &\leq 2 \int_a^\infty e^{-ax/2}\,dx \\
&= \frac{4}{a}\,e^{-a^2/2} \leq e^{-a^2/2}.
\end{aligned}
$$

This proves Lemma 2.5.2. If we apply the estimate with $a = C\sqrt{n}$, then we see that for large n,

$$\mathbb{P}\{Y > C\sqrt{n}\} \leq e^{-C^2 n/2} = 2^{-n\beta}, \quad \beta = \frac{C^2}{2\log 2}.$$

In particular, if $\beta > 1$, then (2.3) holds and this gives (2.2) with probability one.

The hard work is finished, with the proposition, we can now *define* B_t for $t \in [0,1]$ by

$$B_t = \lim_{t_n \to \infty} B_{t_n}$$

where $t_n \to t, t_n \in \mathcal{D}$. One needs to check that this satisfies the conditions for a Brownian motion. This is relatively straightforward and we omit it.

There is a similar approach to prove (2.2) that does not need Lemma 2.5.2. Let

$$\Delta_n = \max\left\{|B_{j/2^n} - B_{(j-1)/2^n}| : j = 1, 2, \ldots, 2^n\right\}.$$

Estimating as above we have

$$\mathbb{P}\{\Delta_n \geq C\sqrt{n}\,2^{-n/2}\} \leq \sum_{j=1}^{2^n} \mathbb{P}\left\{|B_{j/2^n} - B_{(j-1)/2^n}| \geq C\sqrt{n}\,2^{-n/2}\right\}$$
$$\leq 2^n\,\mathbb{P}\{|B_{2^{-n}}| \geq C\sqrt{n}\,2^{-n/2}\}$$
$$= 2^n\,\mathbb{P}\{|B_1| \geq C\sqrt{n}\}$$
$$\leq e^{n\log 2}\,e^{-C^2 n/2},$$

at least for n sufficiently large. Hence if $C > \sqrt{2\log 2}$, the Borel-Cantelli lemma tells us that with probability one for all n sufficiently large, $\Delta_n \leq C\sqrt{n}\,2^{-n/2}$. We now use the deterministic estimate

$$K_n \leq 2\sum_{j=n+1}^{\infty} \Delta_j, \tag{2.4}$$

to see that if C^* is large enough, then for all n sufficiently large,

$$K_n \leq 2C\sum_{j=n+1}^{\infty} \sqrt{j}\,2^{-j/2} \leq C^*\sqrt{n}\,2^{-n/2}.$$

The estimate (2.4) is obtained from the triangle inequality and the following observation. Suppose $s, t \in \mathcal{D}$ with $s \leq t \leq s + 2^{-n}$. Let q be the smallest number in $\mathcal{D}_n$ with $s \leq q$. Then $|s-q| \leq 2^{-n}, |t-q| \leq 2^{-n}$. If $s \in \mathcal{D}$ then $s \in \mathcal{D}_m$ for some m, and we can write s uniquely as

$$s = q - \sum_{j=n+1}^{m} \frac{a_j}{2^j}$$

where $a_j \in \{0, 1\}$. Using the triangle inequality we see that

$$|B_s - B_q| \leq \sum_{k=n+1}^{m} |B_{s_k} - B_{s_{k-1}}| \leq \sum_{k=n+1}^{m} \Delta_k \leq \sum_{k=n+1}^{\infty} \Delta_k$$

where $s_n = q$ and

$$s_k = q - \sum_{j=n+1}^{k} \frac{a_j}{2^j} \in \mathcal{D}_k.$$

Similarly, $|B_t - B_q| \leq \sum_{k=n+1}^{\infty} \Delta_k$.

2.6   Understanding Brownian motion

Let B_t be a standard Brownian motion starting at the origin. By doing simulations, one can see that although the paths of the process are continuous they are very rough. To do simulations, we must discretize time. If we let Δt be chosen, then we will sample the values

$$B_0, B_{\Delta t}, B_{2\Delta t}, B_{3\Delta t}, \ldots$$

The increment $B_{(k+1)\Delta t} - B_{k\Delta t}$ is a normal random variable with mean 0 and variance Δt. If $N_0, N_1, N_2, \ldots$ denote independent $N(0,1)$ random variables (which can be generated on a computer), we set

$$B_{(k+1)\Delta t} = B_{k\Delta t} + \sqrt{\Delta t}\, N_k,$$

which we can write as

$$\Delta B_{k\Delta t} = B_{(k+1)\Delta t} - B_{k\Delta t} = \sqrt{\Delta t}\, N_k.$$

The relationship $|\Delta B_t| = |B_{t+\Delta t} - B_t| \approx \sqrt{\Delta t}$ is fundamental for understanding Brownian motion. As $\Delta t \to 0$, $\sqrt{\Delta t} \to 0$, which is consistent with the continuity of the paths. However, we claim that the paths do not have derivatives. To see why this should be true, recall that the definition of the derivative at time t would be

$$\lim_{\Delta t \to 0} \frac{B_{t+\Delta t} - B_t}{\Delta t}.$$

As $\Delta t \to 0$, the absolute value of the numerator is of order $\sqrt{\Delta t}$ which is much larger than Δt for small Δt. Hence, we do not expect the limit to exist. In fact, the following stronger statement is true.

Theorem 2.6.1. *With probability one, the function $t \mapsto B_t$ is nowhere differentiable.*

Note carefully the statement of this theorem. It is a stronger statement than the similar statement: "for every t, with probability one, the derivative does not exists at t". As a comparison, consider the following two statements:

- With probability one, for every t, $B_t \neq 1$.

- For every t, with probability one, $B_t \neq 1$.

Since B_t has a $N(0, t)$ distribution, the second of these statements is obviously true. However, the first statement is false. To see this, note that $\mathbb{P}\{B_1 > 1\}$ is greater than 0, and if $B_0 = 0, B_1 > 1$, then continuity implies that there exists $t \in (0, 1)$ such that $B_t = 1$. Here again we have the challenge of dealing with uncountably many events of probability 0. Even though for each t, $\mathbb{P}\{B_t = 1\} = 0$,

$$\mathbb{P}\left[\bigcup_{t \in [0,1]} \{B_t = 1\} \right] > 0.$$

While nondifferentiable paths may seem surprising, a little thought about our assumptions will show why we could not expect to have differentiable paths. Suppose that B_t were differentiable at t. Then, we could determine the derivative by observing $B_s, 0 \leq s \leq t$. This would then tell us something about the increment $B_{t+\Delta t} - B_t$ for $\Delta t > 0$. However, our assumptions tell us that $B_{t+\Delta t} - B_t$ is independent of $B_s, 0 \leq s \leq t$.

We will make a more precise statement about the relationship $|\Delta B_t| = |B_{t+\Delta t} - B_t| \approx \sqrt{\Delta t}$. If $\alpha > 0$, then a function $f : [0, 1] \to \mathbb{R}$ is *Hölder continuous of order α* if there exists $C < \infty$ such that for all $0 \leq s, t \leq 1$,

$$|f(s) - f(t)| \leq C \, |s - t|^\alpha.$$

The smaller α is the easier it is to be Hölder continuous of order α. Note that differentiable functions are Hölder continuous of order 1 since

$$|f(s) - f(t)| \approx |f'(t)| \, |s - t|.$$

Theorem 2.6.2. *With probability one, for all $\alpha < 1/2$, B_t is Hölder continuous of order α but it is not Hölder continuous of order $1/2$.*

We will be using Brownian motion and functions of Brownian motions to model prices of assets. In all of the Brownian models, the functions will have Hölder exponent $1/2$.

We will prove Theorem 2.6.1 for $0 \leq t \leq 1$. We start by making an observation about differentiable functions. Suppose B_t were differentiable at some $0 \leq t \leq 1$ with derivative r. Then there would exist δ such that for $|t - s| \leq \delta$,

$$|B_t - B_s| \leq 2|r(t - s)|.$$

Hence, we could find a positive integer $M < \infty$ such that for all sufficiently large integers n, there exists $k \leq n$ such that $Y_{k,n} \leq M/n$, where $Y_{k,n}$ is

$$\max\left\{ \left| B\left(\frac{k+1}{n}\right) - B\left(\frac{k}{n}\right) \right|, \right.$$

$$\left. \left| B\left(\frac{k+2}{n}\right) - B\left(\frac{k+1}{n}\right) \right|, \left| B\left(\frac{k+3}{n}\right) - B\left(\frac{k+2}{n}\right) \right| \right\}.$$

Let $Y_n = \min\{Y_{k,n} : k = 0, 1, \ldots, n-1\}$ and let A_M be the event that for all n sufficiently large, $Y_n \leq M/n$. For each positive integer M,

$$\begin{aligned}
\mathbb{P}\{Y_{k,n} \leq M/n\} &= [\mathbb{P}\{|B(1/n)| \leq M/n\}]^3 \\
&= \left[\mathbb{P}\{n^{-1/2}|B_1| \leq M/n\}\right]^3 \\
&= \left[\int_{|x| \leq M/\sqrt{n}} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy\right]^3 \\
&\leq \left[\frac{2M}{\sqrt{n}} \frac{1}{\sqrt{2\pi}}\right]^3 \leq \frac{M^3}{n^{3/2}},
\end{aligned}$$

and hence,

$$\mathbb{P}\{Y_n \leq M/n\} \leq \sum_{k=0}^{n-1} \mathbb{P}\{Y_{k,n} \leq M/n\} \leq \frac{M^3}{n^{1/2}} \longrightarrow 0.$$

This shows that $\mathbb{P}(A_M) = 0$ for each M, and hence

$$\mathbb{P}\left[\bigcup_{M=1}^{\infty} A_M\right] = 0.$$

But our first remark shows that the event that B_t is differentiable at some point is contained in $\cup_M A_M$.

Theorem 2.6.2 is a restatement of (2.2).

2.6.1 Brownian motion as a continuous martingale

The definition of a martingale in continuous time is essentially the same as in discrete time. Suppose we have an increasing filtration $\{\mathcal{F}_t\}$ of information and integrable random variables M_t such that for each t, M_t is $\mathcal{F}_t$-measurable. (We say that M_t is *adapted* to the filtration if M_t is $\mathcal{F}_t$-measurable for each t.) Then, M_t is a *martingale* with respect to $\{\mathcal{F}_t\}$ if for each $s < t$,

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

When one writes an equality as above there is an implicit "up to an event of probability zero". In discrete time this presents no problem because there are only a countable number of pairs of times (s, t) and hence there can be only a countable number of sets of measure zero. For continuous time, there are instances where some care is needed but we will not worry about this at the moment. As in the discrete case, if the filtration is not mentioned explicitly then one assumes that $\mathcal{F}_t$ is the information contained in $\{M_s : s \leq t\}$. In that case, if B_t is a standard Brownian motion and $s < t$,

$$E[B_t \mid \mathcal{F}_s] = E[B_s \mid \mathcal{F}_s] + E[B_t - B_s \mid \mathcal{F}_s] = B_s + \mathbb{E}[B_t - B_s] = B_s. \quad (2.5)$$

Often we will have more information at time t than the values of the Brownian motion so it is useful to extend our definition of Brownian motion. We say that B_t is Brownian motion with respect to the filtration $\{\mathcal{F}_t\}$ if each B_t is $\mathcal{F}_t$-measurable and B_t satisfies the conditions to be a Brownian motion with the third condition being replaced by

- If $s < t$, the random variable $B_t - B_s$ is independent of $\mathcal{F}_s$.

In other words, although we may have more information at time s than the value of the Brownian motion, there is nothing useful for predicting the future increments. Under these conditions, (2.5) holds and B_t is a martingale with respect to $\{\mathcal{F}_t\}$.

If $M_s, 0 \le s \le t$ is a martingale, then by definition, for each $s \le t$,

$$E(Y \mid \mathcal{F}_s) = M_s$$

where $Y = M_t$. Conversely, if Y is an integrable random variable that is measurable with respect to $\mathcal{F}_t$, we can define a martingale $M_s, 0 \le s \le t$ by

$$M_s = \mathbb{E}(Y \mid \mathcal{F}_s).$$

Indeed, if we define M_s as above and $r < s$, then the tower property for conditional expectation implies that

$$E(M_s \mid \mathcal{F}_r) = E(E(Y \mid \mathcal{F}_s) \mid \mathcal{F}_r) = E(Y \mid \mathcal{F}_r) = M_r.$$

A martingale M_t is called a *continuous martingale* if with probability one the function $t \mapsto M_t$ is a continuous function. The word continuous in continuous martingale refers not just to the fact that time is continuous but also to the fact that the paths are continuous functions of t. One can have martingales in continuous time that are not continuous martingales. One example is to let N_t be a Poisson process with rate λ as in Section 2.1 and

$$M_t = N_t - \lambda t.$$

Then using the fact that the increments are independent we see that for $s < t$,

$$
\begin{aligned}
E[M_t \mid \mathcal{F}_s] &= E[M_s \mid \mathcal{F}_s] + E[N_t - N_s \mid \mathcal{F}_s] - \lambda(t - s) \\
&= M_s + \mathbb{E}\left[N_t - N_s\right] - \lambda(t - s) = M_s.
\end{aligned}
$$

Brownian motion with zero drift ($m = 0$) is a continuous martingale with one parameter σ. As we will see, the *only* continuous martingales are essentially Brownian motion where one allows the σ to vary with time. The factor σ will be the analogue of the "bet" from the discrete stochastic integral.

A continuous time filtration on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a collection of sub-σ-algebras $\{\mathcal{F}_t\}$ of $\mathcal{F}$ such that if $s < t$, then $\mathcal{F}_s \subset \mathcal{F}_t$. It has been found that it is useful to make some additional technical assumptions. First, we assume that $\mathcal{F}$ is *complete* under $\mathbb{P}$, that is, it contains all the nulls sets. A set $A' \subset \Omega$ is a *null set (for $\mathbb{P}$)* if $A' \subset A$ for some $A \in \mathcal{F}$ with $\mathbb{P}(A) = 0$. This is convenient for then

one can prove that an event has probability zero by showing that it is contained in an event of probability zero. If $\mathcal{F}$ is not complete, one can complete it by considering the collection of subsets $A \cup A'$ where $A \in \mathcal{F}$ and A' is a null set. In the case of Lebesgue measure on the Borel subsets of $\mathbb{R}$, the completion is the set of Lebesgue measurable sets.

When using filtrations $\{\mathcal{F}_t\}$ one often makes further assumptions:

- *Right continuity.* For each t

$$\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s.$$

  It is easy to check that the right-hand side is a σ-algebra containing $\mathcal{F}_t$. and it is often denoted by $\mathcal{F}_{t+}$. Right continuity states that $\mathcal{F}_t = \mathcal{F}_{t+}$. If the original filtration was not right continuous, we can replace it with the filtration $\{\mathcal{F}_{t+}\}$.

- *(Strong) completeness.* We assume that $\mathcal{F}_t$ contains all the null sets of $\mathcal{F}$ (note that this is stronger than saying that $\mathcal{F}_t$ is complete). If this does not hold initially, we enlarge our σ-algebra to include all events of the form $A \cup A'$ where $A \in \mathcal{F}_t$ and A' is a null set.

If we start with Brownian motion and let $\mathcal{F}_t$ be the σ-algebra generated by $\{B_s : s \leq t\}$, then it is not true that $\mathcal{F}_t = \mathcal{F}_{t+}$. However, it is almost true in the sense that every set in $\mathcal{F}_{t+}$ can be written as $A \cup A'$ where $A \in \mathcal{F}_t$ and A' is a null set. This is a consequence of the Blumenthal $0 - 1$ law that states that the σ-algebra $\mathcal{F}_{0+}$ contains only events of probability zero or one. We will not prove this here but it is related to the Kolmogorov $0 - 1$ law.

At this point one may not appreciate why one wants to make these assumptions, but we will not try to motivate them further.

2.6.2 Brownian motion as a Markov process

A continuous time process X_t is called *Markov* if for every t, the conditional distribution of $\{X_s : s \geq t\}$ given $\{X_r : r \leq t\}$ is the same as the conditional distribution given X_t. In other words, the future of the process is conditionally independent of the past given the present value.

Brownian motion is a Markov process. Indeed, if B_t is a Brownian motion with parameters (m, σ^2), and

$$Y_s = B_{t+s}, \quad 0 \leq s < \infty,$$

then the conditional distribution of $\{Y_s\}$ given $\mathcal{F}_t$ is that of a Brownian motion with initial condition $Y_0 = B_t$. Indeed, if

$$\hat{B}_s = B_{t+s} - B_t,$$

then $\hat{B}_s$ is a Brownian motion that is independent of $\mathcal{F}_t$. There is a stronger notion of this called the *strong Markov property* that we will discuss in Section 2.7.

2.6.3   Brownian motion as a Gaussian process

A process $\{X_t\}$ is called a *Gaussian process* if each finite subcollection

$$(X_{t_1}, \ldots, X_{t_n})$$

has a joint normal distribution. Recall that to describe a joint normal distribution one needs only give the means and the covariances. Hence the *finite-dimensional distributions* of a Gaussian process are determined by the numbers

$$m_t = \mathbb{E}[X_t], \quad \Gamma_{st} = \mathrm{Cov}(X_s, X_t).$$

If B_t is a standard Brownian motion and $t_1 < t_2 < \cdots < t_n$, then we can write $B_{t_1}, \ldots, B_{t_n}$ as linear combinations of the independent standard normal random variables

$$Z_j = \frac{B_{t_j} - B_{t_{j-1}}}{\sqrt{t_j - t_{j-1}}}, \quad j = 1, \ldots, n.$$

Hence B_t is a Gaussian process with mean zero. If $s < t$,

$$
\begin{aligned}
\mathbb{E}[B_s B_t] &= \mathbb{E}\left[B_s\left(B_s + B_t - B_s\right)\right] \\
&= \mathbb{E}[B_s^2] + \mathbb{E}[B_s(B_t - B_s)] \\
&= s + \mathbb{E}[B_s]\,\mathbb{E}[B_t - B_s] = s,
\end{aligned}
$$

which gives the general rule

$$\mathrm{Cov}(B_s, B_t) = \min\{s, t\}.$$

The description of Brownian motion as a Gaussian process describes only the finite-dimensional distributions but our definition includes some aspects that depend on more than finite-dimensional distributions. In particular, one cannot tell from the finite-dimensional distributions alone whether or not the paths are continuous.

2.6.4 Brownian motion as a self-similar process

If one looks at a small piece of a Brownian motion and blows it up, then the blown-up picture looks like a Brownian motion provided that the dilation uses the appropriate scaling. We leave the derivation as Exercise 2.12.

Theorem 2.6.3. *Suppose B_t is a standard Brownian motion and $a > 0$. Let*

$$Y_t = \frac{B_{at}}{\sqrt{a}}.$$

Then Y_t is a standard Brownian motion.

The key here is that if time is scaled by a factor of a, then space must be scaled by a factor of $1/\sqrt{a}$. The proof of this theorem is not difficult; one needs only show that Y_t satisfies the conditions to be a standard Brownian motion. One can check that the scaling is right by computing

$$\mathrm{Var}[Y_t] = \mathrm{Var}[B_{at}/\sqrt{a}] = a^{-1}\,\mathrm{Var}[B_{at}] = a^{-1}\,(at) = t.$$

2.7 Computations for Brownian motion

We will discuss some methods for computing probabilities for Brownian motions. For ease, we will assume that B_t is a standard Brownian motion starting at the origin with respect to a filtration $\{\mathcal{F}_t\}$. If we are interested in probabilities about the Brownian motion at one time t, we need only use the normal distribution. Often, it is easier to scale to the standard normal. For example,

$$
\begin{aligned}
\mathbb{E}\left[|B_t|\right] = \mathbb{E}\left[t^{1/2}|B_1|\right] &= \frac{t^{1/2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}|x|\,e^{-x^2/2}\,dx \\
&= \sqrt{\frac{2t}{\pi}}\int_0^{\infty} x\,e^{-x^2/2}\,dx \\
&= \sqrt{\frac{2t}{\pi}},
\end{aligned}
$$

and

$$\begin{aligned}
\mathbb{P}\{B_t \geq r\} = \mathbb{P}\{\sqrt{t}\, B_1 \geq r\} \; &= \; \mathbb{P}\{B_1 \geq r/\sqrt{t}\} \\
&= \; 1 - \Phi(r/\sqrt{t}) \\
&= \; \int_{r/\sqrt{t}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx,
\end{aligned}$$

where Φ denotes the distribution function for a standard normal.

   If we are considering probabilities for a finite number of times, we can use the joint normal distribution. Often it is easier to use the Markov property as we now illustrate. We will compute

$$\mathbb{P}\{B_1 > 0, B_2 > 0\}.$$

The events $\{B_1 > 0\}$ and $\{B_2 > 0\}$ are not independent; we would expect them to be positively correlated. We compute by considering the possibilities at time 1,

$$\begin{aligned}
\mathbb{P}\{B_1 > 0, B_2 > 0\} \; &= \; \int_0^{\infty} \mathbb{P}\{B_2 > 0 \mid B_1 = x\}\, d\mathbb{P}\{B_1 = x\} \\
&= \; \int_0^{\infty} \mathbb{P}\{B_2 - B_1 > -x\}\, \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \\
&= \; \int_0^{\infty} \int_{-x}^{\infty} \frac{1}{2\pi} e^{-(x^2+y^2)/2}\, dy\, dx \\
&= \; \int_0^{\infty} \int_{-\pi/4}^{\pi/2} \frac{e^{-r^2/2}}{2\pi}\, d\theta\, r\, dr = \frac{3}{8}.
\end{aligned}$$

One needs to review polar coordinates to do the fourth equality. Note that

$$\mathbb{P}\{B_2 > 0 \mid B_1 > 0\} = \frac{\mathbb{P}\{B_1 > 0, B_2 > 0\}}{\mathbb{P}\{B_1 > 0\}} = \frac{3}{4},$$

which confirms our intuition that the events are positively correlated.

   For more complicated calculations, we need to use the *strong Markov property*. We say that a random variable T taking values in $[0, \infty]$ is a *stopping time (with respect to the filtration $\{\mathcal{F}_t\}$)* if for each t, the event $\mathbb{P}\{T \leq t\}$ is $\mathcal{F}_t$-measurable. In other words, the decision to stop can use the information up to time t but cannot use information about the future values of the Brownian motion.

- If $x \in \mathbb{R}$ and
$$T = \min\{t : B_t = x\},$$

    then T is a stopping time.

- Constants are stopping times.

- If S, T are stopping times then
$$S \wedge T = \min\{S, T\}$$

    and
$$S \vee T = \max\{S, T\}$$

are stopping times.

Theorem 2.7.1 (Strong Markov Property). *If T is a stopping time with* $\mathbb{P}\{T < \infty\} = 1$ *and*
$$Y_t = B_{T+t} - B_T,$$

then Y_t is a standard Brownian motion. Moreover, Y is independent of

$$\{B_t : 0 \leq t \leq T\}.$$

Let us apply this theorem, to prove a very useful tool for computing probabilities.

Proposition 2.7.2 (Reflection Principle). *If B_t is a standard Brownian motion with $B_0 = 0$, then for every $a > 0$,*

$$\mathbb{P}\left\{\max_{0 \leq s \leq t} B_s \geq a\right\} = 2\mathbb{P}\{B_t > a\} = 2\left[1 - \Phi(a/\sqrt{t})\right].$$

To derive the reflection principle, let

$$T_a = \min\left\{s : B_s \geq a\right\} = \min\left\{s : B_s = a\right\}.$$

The second equality holds because B_s is a continuous function of s. Then

$$\mathbb{P}\left\{\max_{0 \leq s \leq t} B_s \geq a\right\} = \mathbb{P}\{T_a \leq t\} = \mathbb{P}\{T_a < t\}.$$

The second equality uses the fact that $\mathbb{P}\{T_a = t\} \leq \mathbb{P}\{B_t = a\} = 0$. Since $B_{T_a} = a$,

$$\begin{aligned} \mathbb{P}\{B_t > a\} &= \mathbb{P}\{T_a < t, B_t > a\} \\ &= \mathbb{P}\{T_a < t\}\,\mathbb{P}\{B_t - B_{T_a} > 0 \mid T_a < t\}. \end{aligned}$$

We now appeal to the Strong Markov Property to say that

$$\mathbb{P}\{B_t - B_{T_a} > 0 \mid T_a < t\} = 1/2.$$

This gives the first equality of the proposition and the second follows from

$$\mathbb{P}\{B_t > a\} = \mathbb{P}\{B_1 > a/\sqrt{t}\} = 1 - \Phi(a/\sqrt{t}).$$

Example 2.7.1. Let $a > 0$ and let $T_a = \inf\{t : B_t = a\}$. The random variable T_a is called a *passage time*. We will find the density of T_a. To do this, we first find its distribution function

$$F(t) = \mathbb{P}\{T_a \leq t\} = \mathbb{P}\left\{\max_{0 \leq s \leq t} B_s \geq a\right\} = 2\left[1 - \Phi(a/\sqrt{t})\right].$$

The density is obtained by differentiating

$$f(t) = F'(t) = -2\Phi'\left(\frac{a}{\sqrt{t}}\right)\left(-\frac{a}{2t^{3/2}}\right) = \frac{a}{t^{3/2}\sqrt{2\pi}}\,e^{-\frac{a^2}{2t}}, \quad 0 < t < \infty.$$

Example 2.7.2. We will compute

$$q(r, t) = \mathbb{P}\left\{B_s = 0 \text{ for some } r \leq s \leq t\right\}.$$

The scaling rule for Brownian motion (Theorem 2.6.3) shows that $q(r, t) = q(1, t/r)$, so it suffices to compute $q(t) = q(1, 1 + t)$. Let A be the event that $B_s = 0$ for some $1 \leq s \leq 1 + t$. The Markov property for Brownian motion and symmetry imply that

$$\begin{aligned} q(t) &= \int_{-\infty}^{\infty} \mathbb{P}[A \mid B_1 = r]\,d\mathbb{P}\{B_1 = r\} \\ &= \int_{-\infty}^{\infty} \mathbb{P}[A \mid B_1 = r]\left[\frac{1}{\sqrt{2\pi}}\,e^{-r^2/2}\,dr\right] \\ &= \sqrt{\frac{2}{\pi}}\int_0^{\infty} \mathbb{P}[A \mid B_1 = r]\,e^{-r^2/2}\,dr. \end{aligned}$$

The reflection principle and symmetry imply that

$$\mathbb{P}[A \mid B_1 = r] = \mathbb{P}\left\{ \min_{1 \le s \le 1+t} B_s \le 0 \mid B_1 = r \right\} =$$

$$\mathbb{P}\left\{ \max_{0 \le s \le t} B_s \ge r \right\} = 2\mathbb{P}\{B_t \ge r\} = 2\left[1 - \Phi(r/\sqrt{t}) \right].$$

Combining this we get

$$q(t) = \int_{-\infty}^{\infty} 2\left[1 - \Phi(r/\sqrt{t}) \right] \frac{1}{\sqrt{2\pi}} e^{-r^2/2} \, dr.$$

This integral can be computed with polar coordinates. We just give the answer

$$q(t) = 1 - \frac{2}{\pi} \arctan \frac{1}{\sqrt{t}}.$$

If the filtration $\{\mathcal{F}_t\}$ is right continuous then the condition $\{T \le t\} \in \mathcal{F}_t$ for all t is equivalent to the condition that $\{T < t\} \in \mathcal{F}_t$ for all t. Indeed,

$$\{T \le t\} = \bigcap_{n=1}^{\infty} \left\{ T < t + \frac{1}{n} \right\} \in \mathcal{F}_{t+},$$

$$\{T < t\} = \bigcup_{n=1}^{\infty} \left\{ T \le t - \frac{1}{n} \right\} \in \mathcal{F}_t.$$

If T is a stopping time, the σ-algebra $\mathcal{F}_T$ is defined to be the set of events $A \in \mathcal{F}$ such that for each t, $A \cap \{T \le t\} \in \mathcal{F}_t$. (It is not hard to show that this is a σ-algebra.) We think of $\mathcal{F}_T$ as the information available up to time T. If $\{T < \infty\}$ let

$$Y_t = B_{T+t} - B_T,$$

and let $\mathcal{G}_T$ denote the σ-algebra generated by $\{Y_t : t \ge 0\}$. We can state the strong Markov property as:

- $\{Y_t : t \ge 0\}$ is a standard Brownian motion.

- The σ-algebras $\mathcal{F}_T$ and $\mathcal{G}_T$ are independent, that is, if $A \in \mathcal{F}_T, A' \in \mathcal{G}_T$, then $\mathbb{P}(A \cap A') = \mathbb{P}(A)\,\mathbb{P}(A')$.

To prove this, one first considers the case where T takes values in $\{k2^{-n} : k = 0, 1, 2, \ldots\}$. By splitting up separately into the events $\{T = k2^{-n}\}$, one can use the usual Markov property to prove the result. For more general stopping times T we approximate T by T_n where $T_n = k2^{-n}$ on the event $\{(k-1)2^{-n} < T \leq k2^{-n}\}$. It is important that we approximate "from above" in order to guarantee that the T_n are stopping times. The continuity of the Brownian paths implies that

$$Y_t = \lim_{n \to \infty} \left[B_{T_n + t} - B_{T_n} \right],$$

and this can be used to conclude the independence of $\mathcal{F}_T$ and $\mathcal{G}_T$.

2.8 Quadratic variation

In the next chapter we will come across quantities such as

$$Q_n = \sum_{j=1}^{n} \left[B\left(\frac{j}{n}\right) - B\left(\frac{j-1}{n}\right) \right]^2,$$

where B_t is a standard Brownian motion. We can write Q_n as

$$\frac{1}{n} \sum_{j=1}^{n} Y_j,$$

where

$$Y_j = Y_{j,n} = \left[\frac{B\left(\frac{j}{n}\right) - B\left(\frac{j-1}{n}\right)}{1/\sqrt{n}} \right]^2.$$

The random variables $Y_1, \ldots, Y_n$ are independent, each with the distribution of Z^2 where Z is a standard normal. In particular,

$$\mathbb{E}\left[Y_j\right] = \mathbb{E}\left[Z^2\right] = 1, \quad \mathbb{E}\left[Y_j^2\right] = \mathbb{E}\left[Z^4\right] = 3.$$

(One can use integration by parts to calculate $\mathbb{E}\left[Z^4\right]$ or one could just look it up somewhere.) Hence $\mathrm{Var}[Y_j] = \mathbb{E}\left[Y_j^2\right] - \mathbb{E}\left[Y_j\right]^2 = 2$, and

$$\mathbb{E}\left[Q_n\right] = \frac{1}{n} \sum_{j=1}^{n} \mathbb{E}\left[Y_j\right] = 1, \quad \mathrm{Var}\left[Q_n\right] = \frac{1}{n^2} \sum_{j=1}^{n} \mathrm{Var}\left[Y_j\right] = \frac{2}{n}.$$

As $n \to \infty$, the variance of Q_n tends to zero and the random variable approaches a constant random variable of 1.

Similarly, for any time t, we let

$$Q_n(t) = \sum_{j \leq tn} \left[B\left(\frac{j}{n}\right) - B\left(\frac{j-1}{n}\right) \right]^2.$$

As $n \to \infty$, the variance of $Q_n(t)$ goes to zero and it approaches a constant random variable with value t.

Definition If X_t is a process, the *quadratic variation* is defined by

$$\langle X \rangle_t = \lim_{n \to \infty} \sum_{j \leq tn} \left[X\left(\frac{j}{n}\right) - X\left(\frac{j-1}{n}\right) \right]^2,$$

where the sum is over all j with $j/n \leq t$.

The calculation above shows that the quadratic variation of a standard Brownian motion is the constant process $\langle B \rangle_t = t$. Suppose W_t is a Brownian motion with drift m and variance σ^2. Then we can write $W_t = \sigma B_t + mt$ where B_t is a standard Brownian motion. Fix t and write

$$\sum \left[W\left(\frac{j}{n}\right) - W\left(\frac{j-1}{n}\right) \right]^2 = \sigma^2 \sum \left[B\left(\frac{j}{n}\right) - B\left(\frac{j-1}{n}\right) \right]^2$$

$$+ \frac{2\sigma m}{n} \sum \left[B\left(\frac{j}{n}\right) - B\left(\frac{j-1}{n}\right) \right] + \sum \frac{m^2}{n^2},$$

where in each case the sum is over $j \leq tn$. As $n \to \infty$,

$$\sigma^2 \sum \left[B\left(\frac{j}{n}\right) - B\left(\frac{j-1}{n}\right) \right]^2 \longrightarrow \sigma^2 \langle B \rangle_t = \sigma^2 t,$$

$$\frac{2\sigma m}{n} \sum \left[B\left(\frac{j}{n}\right) - B\left(\frac{j-1}{n}\right) \right] \sim \frac{2\sigma m}{n} B_t \longrightarrow 0,$$

$$\sum \frac{m^2}{n^2} \sim \frac{tnm^2}{n^2} \longrightarrow 0.$$

We have established the following.

Theorem 2.8.1. *If W_t is a Brownian motion with drift m and variance σ^2, then $\langle W \rangle_t = \sigma^2 \, t$.*

The important facts are that the quadratic variation is not random and that it depends on the variance but not on the mean. It may seem silly at this point to give a name and notation to a quantity which is almost trivial for Brownian motion, but in the next chapter we will deal with processes for which the quadratic variation is not just a linear function of time.

If $0 = t_0 < t_1 < \cdots < t_n = t$, we call the times a *partition* of $[0, t]$. We will write Π for partitions and write

$$\|\Pi\| = \max_{j=1,\ldots,n} \{t_j - t_{j-1}\}.$$

For any partition of $[0, t]$, we define

$$Q(t; \Pi) = \sum_{j=1}^{n} [B(t_j) - B(t_{j-1})]^2.$$

Computing as above, we see that

$$\mathbb{E}[Q(t; \Pi)] = \sum_{j=1}^{n} (t_j - t_{j-1}) = t,$$

$$\begin{aligned}
\text{Var}[Q(t; \Pi)] &= \sum_{j=1}^{n} \text{Var}\left([B(t_j) - B(t_{j-1})]^2\right) \\
&= 2 \sum_{j=1}^{n} (t_j - t_{j-1})^2 \\
&\leq 2\|\Pi\| \sum_{j=1}^{n} (t_j - t_{j-1}) = 2\|\Pi\| t.
\end{aligned}$$

Theorem 2.8.2. *Suppose B is a standard Brownian motion, $t > 0$, and Π_n is a sequence of partitions of the form*

$$0 = t_{0,n} < t_{1,n} < \cdots < t_{l_n,n} = t,$$

with $\|\Pi_n\| \to 0$. Then $Q(t; \Pi_n) \to t$ in probability. Moreover, if

$$\sum_{n=1}^{\infty} \|\Pi_n\| < \infty, \tag{2.6}$$

then with probability one $Q(t; \Pi_n) \to t$.

Proof. Using Chebyshev's inequality, we see that for each integer k,

$$\mathbb{P}\left\{|Q(t;\Pi_n) - t| > \frac{1}{k}\right\} \leq \frac{\text{Var}[Q(t;\Pi_n)]}{(1/k)^2} \leq 2k^2\|\Pi_n\|t,$$

and the right-hand side goes to zero as $n \to \infty$. This gives convergence in probability. If (2.6) holds as well, then

$$\sum_{n=1}^{\infty} \mathbb{P}\left\{|Q(t;\Pi_n) - t| > \frac{1}{k}\right\} < \infty,$$

and hence by the Borel-Cantelli lemma, with probability one, for all n sufficiently large,

$$|Q(t;\Pi_n) - t| \leq \frac{1}{k}. \quad \square$$

It is important to note the order of the quantifiers in this theorem. Let $t = 1$. The theorem states that for every sequence $\{\Pi_n\}$ satisfying (2.6), $Q(1;\Pi_n) \to 1$. The event of measure zero on which convergence does not hold depends on the sequence of partitions. Because there are an uncountable number of such sequences we cannot conclude from the theorem that with probability one, for all sequences $\{\Pi_n\}$ satisfying (2.6), that $Q(1;\Pi_n) \to 1$. In fact, the latter statement is false. Let us give an example. Let us start with the dyadic partition Π_n of $[0,1]$ with 2^n intervals of length 2^{-n}. We will now subdivide some, but not all, of these intervals into two equal pieces. Suppose $[s,t]$ is one of these intervals and let r denote the midpoint. We will subdivide $[s,t]$ into $[s,r]$ and $[r,t]$ if

$$(B_r - B_s)^2 + (B_t - B_s)^2 > (B_t - B_s)^2,$$

and we will retain the interval $[s,t]$ otherwise. This defines a partition $\tilde{\Pi}_n$ with $\|\tilde{\Pi}_n\| \leq \|\Pi_n\|$. It has been chosen so that $Q(1;\tilde{\Pi}_n) \geq Q(1;\Pi_n)$, and one can show (see Exercise 2.5) that

$$\lim_{n\to\infty} Q(1;\tilde{\Pi}_n) = \mathbb{E}\left[\max\{B_{1/2}^2 + (B_1 - B_{1/2})^2, B_1^2\}\right] > 1.$$

This does not contradict the theorem because the partitions $\tilde{\Pi}_n$ depend on the realization of the Brownian motion.

Sometimes it is convenient to fix a sequence of partitions, say partitions with dyadic intervals. Let

$$Q_n(t) = \sum_{j < t2^n} \left[B\left(\frac{j+1}{2^n}\right) - B\left(\frac{j}{2^n}\right)\right]^2.$$

Since the dyadic rationals are countable, the theorem implies that with probability one for every dyadic rational t, $Q_n(t) \to t$. However, since $t \mapsto Q_n(t)$ is increasing, we can conclude that with probability one, for every $t0$, $Q_n(t) \to t$.

2.9 Multidimensional Brownian motion

In finance one is often interested in considering the value of many assets at the same time. Multidimensional Brownian motion B_t is random continuous motion in d-dimensional space. It can be viewed as d one-dimensional Brownian motions

$$B_t = \left(B_t^1, \ldots, B_t^d \right),$$

with perhaps some correlations. If we make the same assumptions that we made in the one-dimensional case (independent, identically distributed increments and continuous paths), then one can show that the increments must be multivariate normal random variables. We use this fact in our definition.

Definition The d-dimensional process

$$B_t = \left(B_t^1, \ldots, B_t^d \right),$$

is called a *d-dimensional Brownian motion starting at the origin with drift* $\mathbf{m} = (m_1, \ldots, m_d) \in \mathbb{R}^d$ *and $d \times d$ covariance matrix Γ with respect to the filtration $\{\mathcal{F}_t\}$* if each B_t is $\mathcal{F}_t$-measurable and the following holds.

- $B_0 = 0$.

- If $s < t$, the distribution of $B_t - B_s$ is joint normal with mean $(t-s)\mathbf{m}$ and covariance matrix $(t-s)\Gamma$.

- If $s < t$, the random vector $B_t - B_s$ is independent of $\mathcal{F}_s$.

- With probability one, the function $t \mapsto B_t$ is continuous.

In particular, each B_t^k is a Brownian motion with drift m_k and variance Γ_{kk} with respect to the filtration $\{\mathcal{F}_t\}$. If $\mathbf{m} = 0$ and $\Gamma = I$, then $B_t^1, \ldots, B_t^d$ are independent standard Brownian motions and B_t is called *standard d-dimensional Brownian motion*.

Definition If X_t, Y_t are processes defined on the same probability space, the *covariation (process)* is defined by

$$\langle X, Y \rangle_t = \lim_{n \to \infty} \sum_{j \leq tn} \left[X \left(\frac{j}{n} \right) - X \left(\frac{j-1}{n} \right) \right] \left[Y \left(\frac{j}{n} \right) - Y \left(\frac{j-1}{n} \right) \right],$$

where the sum is over all j with $j/n \leq t$.

If $X = Y$, then the covariation is the same as the quadratic variation, that is

$$\langle X, X \rangle_t = \langle X \rangle_t.$$

If $B_t = (B_t^1, \ldots, B_t^d)$ is a Brownian motion with drift 0 and covariance matrix Γ, then

$$\mathbb{E} \left[(B_t^i - B_s^i)(B_t^k - B_s^k) \right] = (t - s) \, \Gamma_{ik}.$$

As in the quadratic variation, the drift does not contribute to the covariation. We state the following result which is proved in the same way as for quadratic variation.

Theorem 2.9.1. *If B_t is a d-dimensional Brownian motion with drift* **m** *and covariance matrix Γ, then*

$$\langle B^i, B^k \rangle_t = \Gamma_{ik} \, t.$$

In particular, if $B_t = (B_t^1, \ldots, B_t^d)$ is a standard Brownian motion in $\mathbb{R}^d$, then the components are independent and

$$\langle B^i, B^k \rangle_t = 0, \quad i \neq k.$$

2.10 Heat equation and generator

Even if one's interest in Brownian motion comes from other applications, it is useful to consider the diffusion of heat. Heat flow can be viewed by considering a large number of independent "heat particles" each doing random continuous motion. This viewpoint leads to a deterministic partial differential equation (PDE) that describes the evolution of the temperature. Imagine for the moment that the temperature on the line is determined by the density of heat particles. Let $p_t(x)$ denote the temperature at x at time t. If the heat

particles are moving independently and randomly then we can assume that they are doing Brownian motions. If we also assume that

$$\int_R p_t(x)\, dx = 1,$$

then reasonable to see $p_t(x)$ as the probability density for Brownian motion.

2.10.1   One dimension

We start by taking advantage of what we already know. Assume B_t is a standard Brownian motion starting at the origin and let $p_t(x)$ denote the density of B_t. Since $B_t \sim N(0, t)$, we know that

$$p_t(x) = \frac{1}{\sqrt{2\pi t}}\, e^{-\frac{x^2}{2t}}. \tag{2.7}$$

We view this as a function of two variables t, x. If we are interested in the position at time $s + t$, we can use the Markovian nature of Brownian motion to first observe the position at time s and then to consider what happens in the next time interval of length t. This leads to the *Chapman-Kolmogorov equation*

$$p_{s+t}(x) = \int_{-\infty}^{\infty} p_s(y)\, p_t(x - y)\, dy. \tag{2.8}$$

A straightforward calculation, which we omit, shows that if $p_t(x)$ is given by (2.7), then $p_t(x)$ satisfies (2.8). We emphasize that (2.8) uses the Markovian property: if we are interested in the position at time $t + s$ and are given the information up to time s, the only information that is relevant is the position at time s.

   We will give a heuristic derivation of the PDE that describes the evolution of $p_t(x)$. For a number of our heuristic derivations we will use a *binomial approximation* where we view the Brownian motion as satisfying

$$\mathbb{P}\left\{B_{t+\Delta t} = B_t + \Delta x\right\} = \mathbb{P}\left\{B_{t+\Delta t} = B_t - \Delta x\right\} = \frac{1}{2},$$

where $\Delta x = \sqrt{\Delta t}$. In the approximation, to be at x at time $t + \Delta t$ one must be at $x \pm \Delta x$ at time t which gives

$$p_{t+\Delta t}(x) \approx \frac{1}{2}\, p_t(x - \Delta x) + \frac{1}{2}\, p_t(x + \Delta x).$$

Using $\Delta t = (\Delta x)^2$, this implies

$$\frac{p_{t+\Delta t}(x) - p_t(x)}{\Delta t} = \frac{p_t(x + \Delta x) + p_t(x - \Delta x) - 2\,p_t(x)}{2\,(\Delta x)^2}. \qquad (2.9)$$

We now let $\Delta t \to 0$ in (2.9). The definition of the partial derivative implies that

$$\lim_{\Delta t \to 0} \frac{p_{t+\Delta t}(x) - p_t(x)}{\Delta t} = \partial_t p_t(x).$$

The right-hand side of (2.9) is a little more complicated. Since such limits are fundamental in our study, let us give two ways to compute the limit. One way is to write the right-hand side as

$$\frac{1}{2\Delta x} \left[\frac{p_t(x + \Delta x) - p_t(x)}{\Delta x} - \frac{p_t(x) - p_t(x - \Delta x)}{\Delta x} \right].$$

Waving our hands, we say this is about

$$\frac{1}{2} \frac{\partial_x p_t(x) - \partial_x p_t(x - \Delta x)}{\Delta x},$$

and now we have a difference quotient for the first derivatives in which case the limit should be

$$\frac{1}{2}\, \partial_{xx} p_t(x).$$

Another, essentially equivalent, method to evaluate the limit is to write $f(x) = p_t(x)$ and expand in a Taylor series about x,

$$f(x + \epsilon) = f(x) + f'(x)\,\epsilon + \frac{1}{2}\, f''(x)\,\epsilon^2 + o(\epsilon^2),$$

where $o(\epsilon^2)$ denotes a term such that

$$\lim_{\epsilon \to 0} \frac{o(\epsilon^2)}{\epsilon^2} = 0.$$

Then we see that

$$f(x + \epsilon) + f(x - \epsilon) - 2f(x) = f''(x)\,\epsilon^2 + o(\epsilon^2),$$

and hence the limit of the right-hand side of (2.9) is $\partial_{xx} p_t(x)/2$. We have derived the *heat equation*

$$\partial_t p_t(x) = \frac{1}{2}\, \partial_{xx}\, p_t(x).$$

While we have been a bit sketchy on details, one could start with this equation and note that p_t as defined in (2.7) satisfies this. This is the solution given that $B_0 = 0$, that is, when the "initial density" $p_0(x)$ is the "delta function at 0". (The delta function, written $\delta(\cdot)$ is the probability density of the probability distribution that gives measure one to the point 0. This is not really a density, but informally we write

$$\delta(0) = \infty, \qquad \delta(x) = 0, \quad x \neq 0,$$

$$\int \delta(x)\, dx = 1.$$

These last equations do not make mathematical sense, but they give a workable heuristic definition.)

If the Brownian motion has variance σ^2, one binomial approximation is

$$\mathbb{P}\left\{B_{t+\Delta t} = B_t + \sigma\, \Delta x\right\} = \mathbb{P}\left\{B_{t+\Delta t} = B_t - \sigma\, \Delta x\right\} = \frac{1}{2},$$

where $\Delta x = \sqrt{\Delta t}$. The factor σ is put in so that

$$\mathrm{Var}\left[B(t + \Delta t) - B(t)\right] = \sigma^2\, \Delta t.$$

We can use the same argument to derive that this density should satisfy the heat equation

$$\partial_t p_t(x) = \frac{\sigma^2}{2}\, \partial_{xx} p_t(x).$$

The coefficient σ^2 (or in some texts $(\sigma^2/2)$) is referred to as the *diffusion coefficient*. One can check that a solution to this equation is given by

$$p_t(x) = \frac{1}{\sqrt{2\pi\sigma^2 t}}\, \exp\left\{-\frac{x^2}{2\sigma^2 t}\right\}.$$

When the Brownian motion has drift m, the equation gets another term. To see what the term should look like, let us first consider the case of deterministic linear motion, that is, motion with drift m but no variance. Then if $p_t(x)$ denotes the density at x at time t, we get the relationship

$$p_{t+\Delta t}(x) = p_t(x - m\, \Delta t),$$

since particles at x at time $t + \Delta t$ must have been at $x - m \Delta t$ at time t. This gives

$$\frac{p_{t+\Delta t}(x) - p_t(x)}{\Delta t} = \frac{p_t(x - m\,\Delta t) - p_t(x)}{\Delta t}$$
$$= -m\,\frac{p_t(x) - p_t(x - m\,\Delta t)}{m\Delta t}.$$

Letting $\Delta t \to 0$, we get the equation for linear motion at rate m,

$$\partial_t p_t(x) = -m\,\partial_x p_t(x).$$

This is a first-order equation which means it contains only first derivatives.

If B_t is Brownian motion with drift m and variance σ^2, then we can do a similar argument and show that the density $p_t(x)$ satisfies the second order PDE

$$\partial_t p_t(x) = -m\,\partial_x p_t(x) + \frac{\sigma^2}{2}\,\partial_{xx} p_t(x).$$

As a check, one can consider the appropriate density,

$$p_t(x) = \frac{1}{\sqrt{2\pi\sigma^2 t}}\,\exp\left\{ -\frac{(x - mt)^2}{2\sigma^2 t} \right\},$$

and show that it satisfies this equation.

Before summarizing, we will change the notation slightly. Let $p_t(x, y)$ denote the density of B_t given $B_0 = x$. Under this notation $p_t(x) = p_t(0, x)$. We will define an operator on functions. A *(linear) operator* is a function L from (a linear subspace of) functions to functions satisfying $L(af + bg) = aL(f) + bL(g)$ where a, b are constants and f, g are functions. We say that a function $f : \mathbb{R} \to \mathbb{R}$ is C^2 if it is twice differentiable and the derivatives are continuous functions of x. For any m, σ^2, we define an operator on C^2 functions by

$$L^* f(x) = -m\,f'(x) + \frac{\sigma^2}{2}\,f''(x).$$

If f is a function of more variables, such as $f_t(x, y)$ we write L_x^* to indicate the operator acting on the variable x,

$$L_x^* f_t(x, y) = -m\,\partial_x f_t(x, y) + \frac{\sigma^2}{2}\,\partial_{xx} f_t(x, y).$$

Theorem 2.10.1. *Suppose B_t is a Brownian motion with drift m and variance σ^2. Then the transition density $p_t(x, y)$ satisfies the heat equation*

$$\partial_t p_t(x, y) = L_y^* p_t(x, y)$$

with initial condition $p_0(x, \cdot) = \delta_x(\cdot)$. Here L^* is the operator on functions

$$L^* f(y) = -m\, f'(y) + \frac{\sigma^2}{2}\, f''(y).$$

We think of $p_t(x, y)$ as "the probability of being at y at time t given that $B_0 = x$". For driftless Brownian motion, this is the same as the probability of being at x given that one was at y. However, the reversal of a Brownian motion with drift m should be a Brownian motion with drift $-m$. This gives the following.

Theorem 2.10.2. *Suppose B_t is a Brownian motion with drift m and variance σ^2. Then the transition density $p_t(x, y)$ satisfies the heat equation*

$$\partial_t p_t(x, y) = L_x p_t(x, y)$$

with initial condition $p_0(\cdot, y) = \delta_y(\cdot)$. Here L is the operator on functions

$$L f(x) = m\, f'(x) + \frac{\sigma^2}{2}\, f''(x).$$

The operator L will be more important to us than the operator L^* which is why we give it the simpler notation.

Let H denote the real Hilbert space $L^2(\mathbb{R})$ with inner product

$$(f, g) = \int_{-\infty}^{\infty} f(x)\, g(x) dx.$$

The operators L, L^* are not defined on all of H but they can be defined on a dense subspace, for example, the set of C^2 functions all of whose derivatives decay rapidly at infinity. The operator L^* is the *adjoint* of L which means,

$$(L^* f, g) = (f, L g).$$

One can verify this using the following relations that are obtained by integration by parts:

$$\int_{-\infty}^{\infty} f(x)\, g'(x)\, dx = -\int_{-\infty}^{\infty} f'(x)\, g(x)\, dx,$$

$$\int_{-\infty}^{\infty} f(x)\, g''(x)\, dx = \int_{-\infty}^{\infty} f''(x)\, g(x)\, dx.$$

Suppose B_0 has an initial density f. Then the density of B_t is given by

$$f_t(y) = P_t^* f(y) := \int_{-\infty}^{\infty} f(x)\, p_t(x,y)\, dx.$$

This is the solution to the equation

$$\partial_t f_t(y) = L_y^* f_t(y)$$

with initial condition $f_0(y) = f(y)$. We can write the heat equation as a derivative for operators,

$$\partial_t P_t^* = L^* P_t^*.$$

2.10.2 Expected value at a future time

Suppose B_t is a Brownian motion with drift m and variance σ^2, and let f be a function on $\mathbb{R}$. For example, if we consider B_t to be the price of a stock and f to be the worth of a call option at strike price S at time t, then

$$f(x) = (x - S)_+ = \begin{cases} x - S & \text{if } x \geq S \\ 0 & \text{if } x < S \end{cases}. \tag{2.10}$$

Let $\phi(t, x)$ be the expected value of $f(B_t)$ given that $B_0 = x$. We will write this as

$$\phi(t, x) = \mathbb{E}^x\left[f(B_t)\right] = \mathbb{E}\left[f(B_t) \mid B_0 = x\right].$$

Then

$$\phi(t, x) = \int_{-\infty}^{\infty} f(y)\, p_t(x, y)\, dy.$$

The time derivative can be computed by interchanging differentiation and integration and using the rule from the previous section.

$$
\begin{aligned}
\partial_t \phi(t, x) &= \partial_t \int_{-\infty}^{\infty} f(y)\, p_t(x, y)\, dy \\
&= \int_{-\infty}^{\infty} f(y)\, \partial_t p_t(x, y)\, dy \\
&= \int_{-\infty}^{\infty} f(y)\, L_x p_t(x, y)\, dy \\
&= L_x \int_{-\infty}^{\infty} f(y)\, p_t(x, y)\, dy = L_x \phi(t, x).
\end{aligned}
$$

We will give another derivation of this equation, and while we are at it we will do the multivariate case. Suppose $B_t = (B_t^1, \ldots, B_t^d)$ is a d-dimensional Brownian motion with drift $m = (m_1, \ldots, m_d)$ and covariance matrix Γ. We assume that the Brownian motion is truly d-dimensional, or, in other words, that Γ is a nonsingular matrix. Let $p_t(x, y)$ be the transition probability. If $f : \mathbb{R}^d \to \mathbb{R}$ is a bounded function, define $P_t f$ by

$$
P_t f(x) = \mathbb{E}^x \left[f(B_t) \right] = \int f(y)\, p_t(x, y)\, dy.
$$

Here $x = (x_1, \ldots, x_d), y = (y_1, \ldots, y_d) \in \mathbb{R}^d$ and $dy = dy_1 \cdots dy_d$.

Definition The *(infinitesimal) generator* of a Markov process X_t is the operator L on functions defined by

$$
Lf(x) = \lim_{t \to 0} \frac{P_t f(x) - f(x)}{t},
$$

where

$$
P_t f(x) = \mathbb{E}^x \left[f(X_t) \right].
$$

Theorem 2.10.3. *If B_t is a d-dimensional Brownian motion with drift $m = (m_1, \ldots, m_d)$ and covariance matrix Γ, then the infinitesimal generator is given by*

$$
Lf(x) = m \cdot \nabla f(x) + \frac{1}{2} \sum_{j=1}^{d} \sum_{k=1}^{d} \Gamma_{jk}\, \partial_{jk} f(x).
$$

In particular, if f is a function and

$$\phi(t, x) = P_t f(x) = \mathbb{E}^x[f(B_t)],$$

then ϕ satisfies

$$\partial_t \phi(t, x) = L_x \phi(t, x), \quad t > 0,$$

with initial condition $\phi(0, x) = f(x)$.

We will give a quick derivation, assuming f is C^2 and $t = x = 0$. The second-order Taylor approximation of f around the origin is

$$f(\epsilon) = f(0) + \sum_{j=1}^{d} b_j \, \epsilon_j + \frac{1}{2} \sum_{j=1}^{d} \sum_{k=1}^{d} a_{jk} \, \epsilon_j \, \epsilon_k + o(|\epsilon|^2),$$

where $\epsilon = (\epsilon_1, \ldots, \epsilon_d)$ and

$$b_j = \partial_j f(0), \quad a_{jk} = \partial_{jk} f(0).$$

In particular,

$$f(B_t) - f(B_0) = \sum_{j=1}^{d} b_j \, B_t^j + \frac{1}{2} \sum_{j=1}^{d} \sum_{k=1}^{d} a_{jk} \, B_t^j \, B_t^k + o(|B_t|^2).$$

We know that

$$\mathbb{E}\left[B_t^j\right] = m_j \, t, \quad \mathbb{E}\left(B_t^j \, B_t^k\right) = \Gamma_{jk} \, t,$$

and the expectation of the "error" term $o(|B_t|^2)$ is $o(t)$. Hence,

$$\frac{d}{dt} \mathbb{E}\left[f(B_t)\right] \big|_{t=0+} = \lim_{t \downarrow 0} \frac{\mathbb{E}[f(B_t) - f(B_0)]}{t} = Lf(0).$$

Suppose f is a measurable function that does not grow too fast at infinity. For example, we assume that

$$e^{-\alpha|x|^2} f(x) \to 0, \quad |x| \to \infty,$$

for every $\alpha > 0$. Then

$$\int_{-\infty}^{\infty} |f(y)| \, p_t(x, y) \, dy < \infty,$$

and hence

$$\phi(t,x) := P_t f(x) = \int_{-\infty}^{\infty} f(y)\, p_t(x,y)\, dy < \infty$$

is well defined for all t. Moreover, for $t > 0$, the interchange of integration

$$\partial_t \int_{-\infty}^{\infty} f(y)\, p_t(x,y)\, dy = \int_{-\infty}^{\infty} f(y)\, \partial_t p_t(x,y)\, dy$$

can be justified (say by the dominated convergence theorem). Similarly, integrals with respect to x can be taken to show that $\phi(t,\cdot)$ is C^{∞} in x and

$$L_x \phi(t,x) = \int_{-\infty}^{\infty} f(y)\, L_x p_t(x,y)\, dy$$

Therefore, for $t > 0$, $x \mapsto \phi(t,x)$ is C^2 and we can take the right derivative with respect to time,

$$\lim_{s \downarrow 0} \frac{\phi(t+s,x) - \phi(t,x)}{s} = L_x \phi(t,x),$$

using the argument as above. Although this argument only computes the right derivative, since for fixed x, $\phi(t,x)$ and $L_x\phi(t,x)$ are continuous functions of t, we can conclude that

$$\partial_t \phi(t,x) = L_x \phi(t,x), \quad t > 0.$$

We cannot expect this to hold at $t = 0$, but we can state that if f is continuous at x, then

$$\lim_{t \downarrow 0} \phi(t,x) = f(x).$$

A simple call option works a little differently. Suppose that B_t is a one-dimensional Brownian motion with parameters m, σ^2. A simple call option at time $T > 0$ with strike price S allows the owner to buy a share of the stock at time T for price S. If the price at time T is B_T, then the value of the option is $f(B_T) = (B_T - S)_+$ as in (2.10). We are specifying the value of the function at time T rather than at time 0. However, we can use our work to give a PDE for the expected value of the option. If $t < T$, then the expected value of this option, given that $B_t = x$ is

$$F(t,x) = \mathbb{E}\left[F(B_T) \mid B_t = x\right] = \mathbb{E}^x\left[F(B_{T-t})\right] = \phi(T-t,x).$$

Since

$$\partial_t F(t, x) = -\partial_t \phi(T - t, x) = -L_x \phi(T - t, x) = -L_x F(t, x),$$

we get the following.

- If f is a function, $T > 0$ and

$$F(t, x) = \mathbb{E}[f(B_T) \mid B_t = x],$$

  then for $t < T$, F satisfies the *backwards heat equation*

$$\partial_t F(t, x) = -L_x F(t, x),$$

  with terminal condition $F(T, x) = f(x)$.

As in the one-dimensional case, we can find the operator associated to the transition density. If we run a Brownian motion with drift m and covariance matrix Γ backwards, we get the same covariance matrix but the drift becomes $-m$. Therefore

- For $t > 0$, the transition probability $p_t(x, y)$ satisfies the equation

$$\partial_t p_t(x, y) = L_y^* p_t(x, y),$$

  where

$$L^* f(y) = -m \cdot \nabla f(y) + \frac{1}{2} \sum_{j=1}^d \sum_{k=1}^d \Gamma_{jk} \, \partial_{jk} f(y).$$

The equations

$$\partial_t p_t(x, y) = L_x p_t(x, y), \quad \partial_t p_t(x, y) = L_y^* p_t(x, y)$$

are sometimes called the *Kolmogorov backwards* and *forwards equations*, respectively. The name comes from the fact that they can be derived from the Chapman-Kolmogorov equations by writing

$$p_{t+\Delta t}(x, y) = \int p_{\Delta t}(x, z) \, p_t(z, y) \, dz,$$

$$p_{t+\Delta t}(x, y) = \int p_t(x, z) \, p_{\Delta t}(z, y) \, dz,$$

respectively. The forward equation is also known as the *Fokker-Planck equation*. We will make more use of the backwards equation.

2.11   Exercises

Exercise 2.1. Let Z_1, Z_2, Z_3 be independent $N(0, 1)$ random variables. Let

$$X_1 = Z_1 + Z_3, \quad X_2 = Z_2 + 4Z_3, \quad X_3 = 2Z_1 - 2Z_2 + rZ_3$$

where r is a real number.

1. Explain why $\mathbf{X} = (X_1, X_2, X_3)$ has a joint normal distribution.

2. Find the covariance matrix for $\mathbf{X}$ (in terms of r).

3. For what values of r are X_1 and X_3 independent random variables?

4. For what values of r does the random vector X have a density in $\mathbb{R}^3$?

Exercise 2.2. Let B_t be a standard Brownian motion. Find the following probabilities. If you cannot give the answer precisely give it up to at least three decimal places using a table of the normal distribution.

1. $\mathbb{P}\{B_3 \geq 1/2\}$

2. $\mathbb{P}\{B_1 \leq 1/2, B_3 > B_1 + 2\}$

3. $\mathbb{P}(E)$ where E is the event that the path stays below the line $y = 6$ up to time $t = 10$.

4. $\mathbb{P}\{B_4 \leq 0 \mid B_2 \geq 0\}$.

Exercise 2.3. Let B_t be a standard Brownian motion. For each positive integer n, let

$$J_n = \sum_{j=1}^{n} \left[B\left(\frac{j}{n}\right) - B\left(\frac{j-1}{n}\right) \right]^2.$$

1. Find the mean and variance of the random variable J_n.

2. Prove the following "weak law of large numbers": if $\epsilon > 0$, then

$$\lim_{n \to \infty} \mathbb{P}\left\{ |J_n - 1| > \epsilon \right\} = 0.$$

   Hint: this uses Chebyshev's inequality — look it up if this is not familiar to you.

In the next exercise, you can use the following computation. If $X \sim N(0, 1)$, then the moment generating function of X is given by

$$m(t) = \mathbb{E}\left[e^{tX}\right] = e^{t^2/2}.$$

Exercise 2.4. Suppose B_t is a standard Brownian motion and let $\mathcal{F}_t$ be its corresponding filtration. Let

$$M_t = e^{\sigma B_t - \frac{\sigma^2 t}{2}}.$$

Show that M_t is a martingale with respect to $\mathcal{F}_t$. In other words, show that if $s < t$, then

$$E(M_t \mid \mathcal{F}_s) = M_s.$$

Exercise 2.5. Let B_t be a standard Brownian motion and let

$$Y = \max\left\{ B_1^2 + (B_2 - B_1)^2, \ B_2^2 \right\}.$$

- Show that

$$Y = B_2^2 + 2 B_1 (B_1 - B_2) \, \mathbb{1}\{B_1(B_1 - B_2) \geq 0\}.$$

- Find $\mathbb{E}[Y]$.

- Show that for every $k < \infty$, $\mathbb{E}\left[Y^k\right] < \infty$.

Exercise 2.6. Let B_t be a standard Brownian motion and let $\{\mathcal{F}_t\}$ denote the usual filtration. Suppose $s < t$. Compute the following.

1. $E[B_t^2 \mid \mathcal{F}_s]$

2. $E[B_t^3 \mid \mathcal{F}_s]$

3. $E[B_t^4 \mid \mathcal{F}_s]$

4. $E[e^{4B_t - 2} \mid \mathcal{F}_s]$

Exercise 2.7. Let B_t be a standard Brownian motion and let

$$Y(t) = t \, B(1/t).$$

1. Is $Y(t)$ a Gaussian process?

2. Compute $\text{Cov}(Y(s), Y(t))$.

3. Does $Y(t)$ have the distribution of a standard Brownian motion?

Exercise 2.8. If $f(t), 0 \le t \le 1$ is a continuous function, define the $(3/2)$-variation up to time one to be

$$Q = \lim_{n \to \infty} \sum_{j=1}^{n} \left| f\left(\frac{j}{n}\right) - f\left(\frac{j-1}{n}\right) \right|^{3/2}.$$

What is Q if

1. f is a nonconstant, continuously differentiable function on $\mathbb{R}$?

2. f is a Brownian motion?

Exercise 2.9. Suppose B_t is a standard Brownian motion. For the functions $\phi(t, x), 0 < t < 1, -\infty < x < \infty$, defined below, give a PDE satisfied by the function.

1. $\phi(t, x) = \mathbb{P}\{B_t \ge 0 \mid B_0 = x\}$.

2. $\phi(t, x) = \mathbb{E}[B_1^2 \mid B_t = x]$.

3. Repeat the two examples above if B_t is a Brownian motion with drift m and variance σ^2.

Exercise 2.10. Suppose B_t is a standard Brownian motion and

$$M_t = \max_{0 \le s \le t} B_s.$$

1. Explain why M_t has the same distribution as $\sqrt{t} \, M_1$.

2. Find the density of M_1.

3. Find $\mathbb{E}[M_t]$.

4. If $a > 0$, find $\mathbb{E}[M_1 \, 1\{M_1 \ge a\}]$.

5. Let $\mathcal{F}_t$ be the information in $\{B_s : 0 \le s \le t\}$. Find $E[M_2 \mid \mathcal{F}_1]$.

Exercise 2.11. Write a program that will simulate a standard Brownian motion using step size $\Delta t = .01$.

1. Graph at least one realization of the Brownian motion to produce a "pretty picture" of a Brownian path.

2. Use simulations to estimate the following probability:

$$\mathbb{P}\{B_t \leq 2 \text{ for all } t \leq 1\}.$$

   Run the simulation enough times to get a good estimate for the probability. Use the reflection principle to calculate the actual probability and compare the result.

Exercise 2.12. Prove Theorem 2.6.3.

Exercise 2.13. Suppose $Z_t, t \in \mathcal{D}$ are defined and there exists $c < \infty$, $\alpha < \infty$, $\beta > 0$ such that for all $s, t \in \mathcal{D}$

$$\mathbb{E}\left[|Z_t - Z_s|^\alpha\right] \leq c \, |t - s|^{1+\beta}.$$

Let $\epsilon = \beta/(2\alpha)$.

- Show that

$$\mathbb{P}\{\exists s \in \mathcal{D}_n : |Z_{s+2^{-n}} - Z_s| \geq 2^{-n\epsilon}\} \leq c \, 2^{-n\beta/2}.$$

- Show that with probability one, for all n sufficiently large and all $s \in \mathcal{D}_n$,
$$|Z_{s+2^{-n}} - Z_s| \leq 2^{-n\epsilon}.$$

- Show that with probability one, there exists C such that for all $s, t \in \mathcal{D}$,
$$|Z_s - Z_t| \leq C \, |s - t|^\epsilon.$$

Chapter 3

Stochastic integration

3.1 What is stochastic calculus?

Before venturing into stochastic calculus, it will be useful to review the basic ideas behind the usual differential and integral calculus. The main deep idea of calculus is that one can find values of a function knowing the rate of change of the function. For example, suppose $f(t)$ denotes the position of a (one-dimensional) particle at time t, and we are given the rate of change

$$df(t) = C(t, f(t)) \, dt,$$

or as it is usually written,

$$\frac{df}{dt} = f'(t) = C(t, f(t)).$$

In other words, at time t the graph of f moves for an infinitesimal amount of time along a straight line with slope $C(t, f(t))$. This is an example of a differential equation, where the rate depends both on time and position. If we are given an initial condition $f(0) = x_0$, then (under some hypotheses on the rate function C, see the end of Section 3.5), the function is defined and can be given by

$$f(t) = x_0 + \int_0^t C(s, f(s)) \, ds.$$

If one is lucky, then one can do the integration and find the function exactly. If one cannot do this, one can still use a computer to approximate the solution.

The simplest, but still useful, technique is *Euler's method* where one chooses a small increment Δt and then writes

$$f((k+1)\Delta t) = f(k\Delta t) + \Delta t\, C(k\Delta t, f(k\Delta t)).$$

Stochastic calculus is similar, except that one adds randomness to the change. We will make sense of equations such as

$$dX_t = m(t, X_t)\, dt + \sigma(t, X_t)\, dB_t, \tag{3.1}$$

where B_t is a standard Brownian motion. This is an example of a *stochastic differential equation (SDE)*. We should read this equation as stating that at time t, X_t is evolving like a Brownian motion with drift $m(t, X_t)$ and variance $\sigma(t, X_t)^2$. Solving such equations exactly is much harder than for usual calculus, and we often have to resort to numerical methods. One technique is the *stochastic Euler method* which allows us to do *Monte Carlo simulations* of the process. This is easy enough to describe so we give the formula now

$$X((k+1)\Delta t) =$$

$$X(k\Delta t) + \Delta t\, m(k\Delta t, X(k\Delta t)) + \sqrt{\Delta t}\, \sigma(k\Delta_t, X(k\Delta t))\, N_k,$$

where N_k is a $N(0,1)$ random variable.

The standard approach to calculus is to define the derivative, then define the integral, and then give the relationship between the two. One could also define the integral first. In stochastic calculus, the derivative is not so easily defined, so in order to give a mathematical formulation we concentrate on defining the stochastic integral. We will say that X_t is a solution to (3.1) if

$$X_t = X_0 + \int_0^t m(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s.$$

The ds integral is the usual integral from calculus; the integrand $m(s, X_s)$ is random, but that does not give any problem in defining the integral. The main task will be to give precise meaning to the second term, and more generally to

$$\int_0^t A_s\, dB_s.$$

There are several approaches to stochastic integration. The approach we give, which is most commonly used in mathematical finance, is that of the *Itô integral*.

3.2 Stochastic integral

Let B_t be a standard (one-dimensional) Brownian motion with respect to a filtration $\{\mathcal{F}_t\}$. We want to define the process

$$Z_t = \int_0^t A_s \, dB_s.$$

We think of Z_t as a Brownian motion which at time s has variance A_s^2. In analogy to the discrete stochastic integral, we think of A_s as the amount "bet" or "invested" at time s with negative bets indicating bets that the process will go down. Also, as in the discrete case, we want to restrict our betting strategies to those that cannot look into the future. For continuous time processes for which changes are happening infinitesimally it is somewhat trickier to make the last condition precise. We will start by doing some easy cases and then define more complicated cases by taking limits.

3.2.1 Review of Riemann integration

Let us review the definition of the usual Riemann integral from calculus. Suppose $f(t)$ is a continuous function and we wish to define

$$\int_0^1 f(t) \, dt.$$

We partition $[0, 1]$ into small intervals, say

$$0 = t_0 < t_1 < \cdots < t_n = 1,$$

and approximate $f(t)$ by a *step function*

$$f_n(t) = f(s_j), \quad t_{j-1} < t \le t_j,$$

where s_j is some point chosen in $[t_{j-1}, t_j]$. We define

$$\int_0^1 f_n(t) \, dt = \sum_{j=1}^n f(s_j) \, (t_j - t_{j-1}).$$

It is a theorem that if we take a sequence of partitions such that the maximum length of the intervals $[t_{j-1}, t_j]$ goes to zero, then the limit

$$\int_0^1 f(t) \, dt = \lim_{n \to \infty} \int_0^1 f_n(t) \, dt$$

exists and is independent of the choice of sequence of partitions or the choice of s_j. This is the *definition* of the left-hand side. One later proves the important facts, for example, the fundamental theorem of calculus,

$$\int_a^b f'(t)\, dt = f(b) - f(a).$$

3.2.2   Integration of simple processes

The analogue of a step function for the stochastic integral is a *simple process*. This corresponds to a betting strategy that allows one to change bets only at a prescribed finite set of times. To be more precise, a process A_t is a simple process if there exist times

$$0 = t_0 < t_1 < \cdots < t_n < \infty$$

and random variables $Y_j, j = 0, 1, \ldots, n$ that are $\mathcal{F}_{t_j}$-measurable such that

$$A_t = Y_j, \quad t_j \le t < t_{j+1}.$$

Here we set $t_{n+1} = \infty$. Since Y_j is $\mathcal{F}_{t_j}$-measurable, A_t is $\mathcal{F}_t$-measurable. We also assume that $\mathbb{E}[Y_j^2] < \infty$ for each j. If A_t is a simple process we define

$$Z_t = \int_0^t A_s\, dB_s,$$

by

$$Z_{t_j} = \sum_{i=0}^{j-1} Y_i \left[B_{t_{i+1}} - B_{t_i} \right],$$

and, more generally,

$$Z_t = Z_{t_j} + Y_j \left[B_t - B_{t_j} \right] \quad \text{if } t_j \le t \le t_{j+1},$$

$$\int_r^t A_s\, dB_s = Z_t - Z_r.$$

There are four important properties of the stochastic integral of simple processes which we give in the next proposition. The reader should compare these with the properties of integration with respect to random walk in Section 1.6.

Proposition 3.2.1. *Suppose B_t is a standard Brownian motion with respect to a filtration $\{\mathcal{F}_t\}$, and A_t, C_t are simple processes.*

- **Linearity.** *If a, b are constants, then $aA_t + bC_t$ is also a simple process and*

$$\int_0^t (aA_s + bC_s) \, dB_s = a \int_0^t A_s \, dB_s + b \int_0^t C_s \, dB_s.$$

  Moreover, if $0 < r < t$,

$$\int_0^t A_s \, dB_s = \int_0^r A_s \, dB_s + \int_r^t A_s \, dB_s.$$

- **Martingale property.** *The process*

$$Z_t = \int_0^t A_s \, dB_s$$

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

- **Variance rule**. *Z_t is square integrable and*

$$\operatorname{Var}[Z_t] = \mathbb{E}\left[Z_t^2\right] = \int_0^t \mathbb{E}[A_s^2] \, ds.$$

- **Continuity**. *With probability one, the function $t \mapsto Z_t$ is a continuous function.*

Let us discuss the proof of the proposition. Linearity follows immediately from the definition, and continuity follows from the continuity of Brownian motion. To show that Z_t is a martingale, we need to show that

$$E(Z_t \mid \mathcal{F}_s) = Z_s \quad \text{if} \quad s < t. \tag{3.2}$$

We will do this in the case $t = t_j, s = t_k$ for some $j > k$ and leave the other cases for the reader. In this case

$$Z_s = \sum_{i=0}^{k-1} Y_i \left[B_{t_{i+1}} - B_{t_i}\right],$$

and

$$Z_t = Z_s + \sum_{i=k}^{j-1} Y_i \left[B_{t_{i+1}} - B_{t_i}\right].$$

Since $E(Z_s \mid \mathcal{F}_s) = Z_s$, we see that

$$E(Z_t \mid \mathcal{F}_s) = Z_s + \sum_{i=k}^{j-1} E\left[Y_i\left[B_{t_{i+1}} - B_{t_i}\right] \mid \mathcal{F}_s\right].$$

For $k \leq i \leq j-1$, we have $t_i \geq s$ and hence

$$E\left[Y_i\left[B_{t_{i+1}} - B_{t_i}\right] \mid \mathcal{F}_s\right] = E\left[E\left(Y_i\left[B_{t_{i+1}} - B_{t_i}\right] \mid \mathcal{F}_{t_i}\right) \mid \mathcal{F}_s\right].$$

Since Y_i is $\mathcal{F}_{t_i}$-measurable and $B_{t_{i+1}} - B_{t_i}$ is independent of $\mathcal{F}_{t_i}$,

$$E\left(Y_i\left[B_{t_{i+1}} - B_{t_i}\right] \mid \mathcal{F}_{t_i}\right) = Y_i\, E\left(B_{t_{i+1}} - B_{t_i} \mid \mathcal{F}_{t_i}\right) = Y_i\, \mathbb{E}[B_{t_{i+1}} - B_{t_i}] = 0.$$

This gives (3.2).

We will prove the variance rule for $t = t_j$ in which case

$$Z_t^2 = \sum_{i=0}^{j-1}\sum_{k=0}^{j-1} Y_i\left[B_{t_{i+1}} - B_{t_i}\right] Y_k\left[B_{t_{k+1}} - B_{t_k}\right].$$

If $i < k$,

$$\mathbb{E}\left[Y_i\left[B_{t_{i+1}} - B_{t_i}\right] Y_k\left[B_{t_{k+1}} - B_{t_k}\right]\right]$$
$$= \mathbb{E}\left[E\left(Y_i\left[B_{t_{i+1}} - B_{t_i}\right] Y_k\left[B_{t_{k+1}} - B_{t_k}\right] \mid \mathcal{F}_{t_k}\right)\right].$$

The random variables $Y_i, Y_k, B_{t_{i+1}} - B_{t_i}$ are all $\mathcal{F}_{t_k}$-measurable while $B_{t_{k+1}} - B_{t_k}$ is independent of $\mathcal{F}_{t_k}$, and hence

$$E\left(Y_i\left[B_{t_{i+1}} - B_{t_i}\right] Y_k\left[B_{t_{k+1}} - B_{t_k}\right] \mid \mathcal{F}_{t_k}\right)$$
$$= Y_i\left[B_{t_{i+1}} - B_{t_i}\right] Y_k\, E\left(B_{t_{k+1}} - B_{t_k} \mid \mathcal{F}_{t_k}\right)$$
$$= Y_i\left[B_{t_{i+1}} - B_{t_i}\right] Y_k\, \mathbb{E}\left[B_{t_{k+1}} - B_{t_k}\right] = 0.$$

Arguing similarly for $i > k$, we see that

$$\mathbb{E}[Z_t^2] = \sum_{i=0}^{j-1} \mathbb{E}\left[Y_i^2\left(B_{t_{i+1}} - B_{t_i}\right)^2\right].$$

(We have just reproduced the argument showing orthogonality of martingale increments.) Since Y_i^2 is $\mathcal{F}_{t_i}$-measurable and $(B_{t_{i+1}} - B_{t_i})^2$ is independent of $\mathcal{F}_{t_i}$, we get

$$\begin{aligned}
E\left[Y_i^2\left(B_{t_{i+1}} - B_{t_i}\right)^2 \mid \mathcal{F}_{t_i}\right] &= Y_i^2\, E\left[\left(B_{t_{i+1}} - B_{t_i}\right)^2 \mid \mathcal{F}_{t_i}\right] \\
&= Y_i^2\, \mathbb{E}\left[\left(B_{t_{i+1}} - B_{t_i}\right)^2\right] \\
&= Y_i^2\left(t_{i+1} - t_i\right),
\end{aligned}$$

and hence,

$$\mathbb{E}\left[Y_i^2\left(B_{t_{i+1}} - B_{t_i}\right)^2\right] = \mathbb{E}\left[E\left(Y_i^2\left(B_{t_{i+1}} - B_{t_i}\right)^2 \mid \mathcal{F}_{t_i}\right)\right]$$
$$= \mathbb{E}[Y_i^2]\left(t_{i+1} - t_i\right).$$

The function $s \mapsto \mathbb{E}[A_s^2]$ is a step function that takes on the value $\mathbb{E}[Y_i^2]$ for $t_i \le s < t_{i+1}$. Therefore,

$$\mathbb{E}[Z_t^2] = \sum_{i=0}^{j-1} \mathbb{E}[Y_i^2]\left(t_{i+1} - t_i\right) = \int_0^t \mathbb{E}[A_s^2]\, ds.$$

3.2.3 Integration of continuous processes

In this section we will assume that A_t is a process with continuous paths that is adapted to the filtration $\{\mathcal{F}_t\}$. Recall that this implies that each A_t is $\mathcal{F}_t$-measurable. To integrate these processes we use the following approximation result.

Lemma 3.2.2. *Suppose A_t is a process with continuous paths, adapted to the filtration $\{\mathcal{F}_t\}$. Suppose also that there exists $C < \infty$ such that with probability one $|A_t| \le C$ for all t. Then there exists a sequence of simple processes $A_t^{(n)}$ such that for all t,*

$$\lim_{n \to \infty} \int_0^t \mathbb{E}\left[|A_s - A_s^{(n)}|^2\right]\, ds = 0. \tag{3.3}$$

Moreover, for all n, t, $|A_t^{(n)}| \le C$.

We will prove this lemma for $t = 1$. let

$$A_t^{(n)} = A(j, n), \quad \frac{j}{n} \le t < \frac{j+1}{n},$$

where $A(0, n) = A_0$ and for $j \ge 1$,

$$A(j, n) = n \int_{(j-1)/n}^{j/n} A_s\, ds.$$

It suffices to prove (3.3) for each fixed value of t and for ease we will choose $t = 1$. By construction, the $A_t^{(n)}$ are simple processes satisfying $|A_t^{(n)}| \leq C$. Since (with probability one) the function $t \mapsto A_t$ is continuous, we have

$$A_t^{(n)} \to A_t,$$

and hence by the bounded convergence theorem applied to Lebesgue measure,

$$\lim_{n \to \infty} Y_n = 0,$$

where

$$Y_n = \int_0^1 [A_t^{(n)} - A_t]^2 \, dt. \tag{3.4}$$

Since the random variables $\{Y_n\}$ are uniformly bounded, this implies

$$\lim_{n \to \infty} \mathbb{E} \left[\int_0^1 [A_t^{(n)} - A_t]^2 \, dt \right] = \lim_{n \to \infty} \mathbb{E}[Y_n] = 0.$$

Given the lemma, we can define

$$\int_0^t A_s \, dB_s,$$

for a bounded, continuous process A_s as follows. Find a sequence of simple process $A_s^{(n)}$ satisfying (3.3). Then it can be shown that for each t there is an square integrable random variable Z_t such that

$$Z_t = \lim_{n \to \infty} \int_0^t A_s^{(n)} \, dB_s.$$

We define

$$\int_0^t A_s \, dB_s = Z_t.$$

The integral satisfies four properties which should start becoming familiar.

Proposition 3.2.3. *Suppose B_t is a standard Brownian motion with respect to a filtration $\{\mathcal{F}_t\}$, and A_t, C_t are bounded, adapted processes with continuous paths.*

- **Linearity.** *If a, b are constants, then*

$$\int_0^t (aA_s + bC_s) \, dB_s = a \int_0^t A_s \, dB_s + b \int_0^t C_s \, dB_s.$$

Also, if $r < t$,

$$\int_0^t A_s \, dB_s = \int_0^r A_s \, dB_s + \int_r^t A_s \, dB_s.$$

- **Martingale property.** *The process*

$$Z_t = \int_0^t A_s \, dB_s$$

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

- **Variance rule**. *Z_t is square integrable and*

$$\text{Var}[Z_t] = \mathbb{E}\left[Z_t^2\right] = \int_0^t \mathbb{E}[A_s^2] \, ds.$$

- **Continuity.** *With probability one, $t \mapsto Z_t$ is a continuous function.*

For fixed t, the existence of Z_t follows from the estimate (3.3) and the completeness of L^2. In this case, the convergence is in L^2. However, there are issues dealing with the fact that there are an uncountable number of times. This can be handled similarly to the construction of Brownian motion. For convenience, we restrict to $0 \leq t \leq 1$.

We first consider $t \in \mathcal{D}$ and define $Z_t, t \in \mathcal{D}$ as the L^2 limit which is defined for $t \in \mathcal{D}$ up to a single event of probability zero. Suppose A_t is a simple process with $|A_t| \leq C$. Then (Exercise 3.1) one can show that

$$\mathbb{E}[(Z_t - Z_s)^4] \leq 4 \, C^4 \, |t - s|^2.$$

By Fatou's lemma, this bound will also hold for the limit process. This estimate and an argument first due to Kolmogorov suffice to give continuity. We leave the argument as Exercise 2.13.

Let $A_t^{(n)}$ be a sequence of simple processes and let $Z_t^{(n)}, Z_t$ denote the corresponding stochastic integrals. Let

$$\|A^{(n)} - A\|^2 = \mathbb{E}\left[(Z_1^{(n)} - Z_1)^2\right] = \int_0^1 \mathbb{E}\left[(A_t^{(n)} - A_t)^2\right]\, dt,$$

and let

$$Q_n = \max_{0 \leq t \leq 1} |Z_t^{(n)} - Z_t|.$$

If we view $Z^{(n)}, Z$ as random variables taking values in $C[0,1]$, the set of continuous functions on $[0,1]$ with the supremum norm, then Q_n is the distance between $Z^{(n)}$ and Z. The next proposition establishes convergence with probability one in this space.

Proposition 3.2.4. *If*

$$\sum_{n=1}^{\infty} \|A^{(n)} - A\|^2 < \infty, \tag{3.5}$$

then with probability one, $Q_n \to 0$.

Proof. By the Borel-Cantelli lemma, it suffices to show that for every $\epsilon > 0$,

$$\sum_{n=1}^{\infty} \mathbb{P}\{Q_n > \epsilon\} < \infty.$$

Continuity of $Z_t^{(n)}$ and Z_t implies that

$$\mathbb{P}\{Q_n > \epsilon\} = \lim_{m \to \infty} \mathbb{P}\left\{\max_{t \in \mathcal{D}_m} |Z_t^{(n)} - Z_t| > \epsilon\right\}.$$

For fixed m, the process $M_t = Z_t^{(n)} - Z_t, t \in \mathcal{D}_m$, is a discrete time martingale and Corollary 1.7.2 implies that

$$\mathbb{P}\left\{\max_{t \in \mathcal{D}_m} |M_t| > \epsilon\right\} \leq \frac{\mathbb{E}[M_1^2]}{\epsilon^2}.$$

Therefore,

$$\mathbb{P}\{Q_n > \epsilon\} \leq \frac{\mathbb{E}[M_1^2]}{\epsilon^2} = \epsilon^{-2}\|A^{(n)} - A_t\|^2. \quad \square$$

Let us emphasize the order of quantifiers in the proposition. Given a sequence $A^{(n)}$ of approximating simple processes satisfying (3.5), we get convergence with probability one. It is not true that with probability one, we get convergence for every sequence. See an analogous discussion in Section 2.8 about the quadratic variation.

If A_t is adapted with continuous paths, but not necessarily bounded, we can still define

$$Z_t = \int_0^t A_s \, dB_s$$

as follows. For each $n < \infty$, let $T_n = \min\{t : |A_t| = n\}$ and let $A_s^{(n)} = A_{s \wedge T_n}$. Then $A_s^{(n)}$ is a bounded, continuous process and

$$Z_t^{(n)} = \int_0^t A_s^{(n)} \, dB_s.$$

is well defined. We define

$$Z_t = \lim_{n \to \infty} Z_t^{(n)}.$$

The existence of this limit is easy to establish. Indeed, let

$$K_t = \max_{0 \le s \le t} |A_s|.$$

Then for $n \ge K_t$, $A_s^{(n)} = A_s$ for $0 \le s \le t$, and hence

$$Z_t^{(n)} = Z_t, \quad t \ge K_t.$$

There is a subtlety here. Since K_t is a random variable, the n that one needs to choose may depend on the path. Although Z_t is well defined in the limit and satisfies linearity and continuity, we will see in Section 4.1 that it might not satisfy the martingale property. Let us not worry about this at the moment. We will state the following.

Proposition 3.2.5. *Suppose B_t is a standard Brownian motion with respect to a filtration $\{\mathcal{F}_t\}$, and A_t, C_t are adapted processes with continuous paths.*

- **Linearity.** *If a, b are constants and $r < t$, then*

$$\int_0^t (aA_s + bC_s) \, dB_s = a \int_0^t A_s \, dB_s + b \int_0^t C_s \, dB_s,$$

$$\int_0^t A_s \, dB_s = \int_0^r A_s \, dB_s + \int_r^t A_s \, dB_s.$$

- **Variance rule**. *The variance of Z_t satisfies*

$$\mathrm{Var}[Z_t] = \mathbb{E}\left[Z_t^2\right] = \int_0^t \mathbb{E}[A_s^2]\, ds.$$

  It is possible for both sides to equal infinity. However, if $\mathrm{Var}[Z_t] < \infty$ for all t, then Z_t is a square integrable martingale.

- **Continuity** *With probability one, $t \mapsto Z_t$ is a continuous function.*

We can relax the condition of continuous paths. We say that the paths are piecewise continuous if with probability one on each finite interval the paths are continuous except a finite number of points, say $0 < t_1 < t_2 < \cdots < t_n = t$. For a piecewise continuous function with discontinuities at times $t_1, t_2, \ldots$, we define

$$\int_0^t A_s\, dB_s = \int_0^{t_1} A_s\, dB_s + \cdots + \int_{t_{n-1}}^{t_n} A_s\, dB_s. \tag{3.6}$$

The value of this integral does not depend on how we define A_t at the discontinuity. In this book, the process A_t will have continuous or piecewise continuous paths although the integral can be extended to more general processes. Note that the simple processes have piecewise continuous paths. One important case that arises comes from a stopping time. Suppose T is a stopping time with respect to $\{\mathcal{F}_t\}$. Then if A_t is a continuous, adapted process and

$$Z_t = \int_0^t A_s\, dB_s,$$

then

$$Z_{t \wedge T} = \int_0^{t \wedge T} A_s\, dB_s = \int_0^t A_{s,T}\, dB_s,$$

where $A_{s,T}$ denotes the piecewise continuous process,

$$A_{s,T} = \left\{ \begin{array}{ll} A_s & s < T \\ 0 & s \geq T \end{array} \right. .$$

In other words, stopping a stochastic integral is the same as changing the bet to zero.

Continuity of paths is much more than is needed for the stochastic integral to exist. If A_t is a uniformly bounded, adapted process, and we let

$$A_t^{(n)} = n \int_{t-\frac{1}{n}}^{t} A_s \, ds, \tag{3.7}$$

then $A_t^{(n)}$ is an adapted continuous process. This requires some assumptions on the process A_t so that integrals as in (3.4) and (3.7) exists, at least as Lebesgue integrals, and so that we can use Fubini's theorem to interchange expectation and integrals on the line. The condition to guarantee this is called *progressive measurablility*, but we will not go into details here. If the A_t are uniformly bounded, then the Lebesgue density theory implies that $A_t^{(n)} \to A_t$ for Lebesgue almost every t. Hence, with probability one,

$$\int_0^1 [A_t^{(n)} - A_t]^2 \, dt \longrightarrow 0,$$

and since the processes are bounded the convergence is also in L^2. For such processes we can define the stochastic integral as above as an L^2-limit.

We will write the stochastic differential

$$dX_t = A_t \, dB_t,$$

to mean that X_t satisfies

$$X_t = X_0 + \int_0^t A_s \, dB_s.$$

This has been made mathematically precise by the definition of the integral. Intuitively, we think of X_t as a process that at time t evolves like a Brownian motion with zero drift and variance A_t^2. This is well defined for any adapted, continuous process A_t, and X_t is a continuous function of t. In particular, if ϕ is a bounded continuous function, then we can hope to solve the equation

$$dX_t = \phi(X_t) \, dB_t.$$

Solving such an equation can be difficult (see the end of Section 3.5), but simulating such a process is straightforward using the stochastic Euler rule:

$$X_{t+\Delta t} = X_t + \phi(X_t) \sqrt{\Delta t} \, N,$$

where $N \sim N(0, 1)$.

The rules of stochastic calculus are not the same as those of usual calculus. As an example, consider the integral

$$Z_t = \int_0^t B_s \, dB_s.$$

Although B_t is not a bounded process, it is continuous, adapted and

$$\int_0^t \mathbb{E}[B_s^2] \, ds = \int_0^t s \, ds = \frac{t^2}{2} < \infty,$$

so Z_t is a square integrable martingale. If one naively followed the rules of calculus, one might hope that

$$Z_t = \frac{1}{2} \left[B_t^2 - B_0^2 \right] = \frac{B_t^2}{2}.$$

However, a quick check shows that this cannot be correct. The left-hand side is a martingale with $Z_0 = 0$ and hence

$$\mathbb{E}[Z_t] = 0.$$

However,

$$\mathbb{E}\left[B_t^2 / 2 \right] = t/2 \neq 0.$$

In the next section we will derive the main tool for calculating integrals, *Itô's formula or lemma*. Using this we will show that, in fact,

$$Z_t = \frac{1}{2} [B_t^2 - t]. \tag{3.8}$$

This is a very special case. In general, one must do more than just subtract the expectation. One thing to note about the solution — for $t > 0$, the random variable on the right-hand side of (3.8) does not have a normal distribution. Even though stochastic integrals are defined as limits of normal increments, the "betting" factor A_t can depend on the past and this allows one to get non-normal random variables. If the integrand $A_t = f(t)$ is nonrandom, then the integral

$$\int_0^t f(s) \, dB_s,$$

is really a limit of normal random variables and hence has a normal distribution (see Exercise 3.8).

If

$$Z_t = \int_0^t A_s \, dB_s,$$

then the *quadratic variation* of Z is defined by

$$\langle Z \rangle_t = \lim_{n \to \infty} \sum_{j \leq nt} \left[Z\left(\frac{j}{n}\right) - Z\left(\frac{j-1}{n}\right) \right]^2.$$

Theorem 3.2.6. *If A_t is an adapted process with continuous or piecewise continuous paths and*

$$Z_t = \int_0^t A_s \, dB_s,$$

then

$$\langle Z \rangle_t = \int_0^t A_s^2 \, ds.$$

Note that if $\sigma > 0$ is a constant and $Z_t = \int_0^t \sigma \, dB_s$, then Z is a Brownian motion with variance parameter σ^2 for which we have already seen that $\langle Z \rangle_t = \sigma^2 t$. The quadratic variation $\langle Z \rangle_t$ should be viewed as the "total amount of randomness" or the "total amount of betting" that has been done up to time t. For Brownian motion, the betting rate stays constant and hence the quadratic variation grows linearly. For more general stochastic integrals, the current bet depends on the results of the games up to that point and hence is a random variable. The quadratic variation at time t is a random variable with mean

$$\mathbb{E}\left[\langle Z \rangle_t\right] = \mathbb{E}\left[\int_0^t A_s^2 \, ds\right] = \int_0^t \mathbb{E}\left[A_s^2\right] \, ds.$$

An alternative way to define the quadratic variation $\langle Z \rangle_t$ for square integrable martingales is as the unique increasing process such that

$$M_t = Z_t^2 - \langle Z \rangle_t$$

is a martingale. If A_t is a simple process, it is not hard (Exercise 3.4) to show that M_t as defined above is a martingale. More generally, one

can establish this by taking limits. To show that this characterization defines $\langle Z \rangle$ uniquely, we need the following proposition. The total variation of a function $f : [0, \infty) \to \mathbb{R}$ is defined as

$$V_t(f) = \sup \sum_{j=1}^{n} |f(t_{j-1}) - f(t_j)| \leq K,$$

where the supremum is over all partitions

$$0 = t_0 < t_1 < t_2 < \cdots < t_n = t.$$

We say f has *bounded variation (locally)* if $V_t(f) < \infty$ for all t. Increasing functions clearly have bounded variation, and it is not too hard to see that the difference of increasing functions has bounded variation. Also, if f is continuous with bounded variation,

$$\lim_{t \downarrow 0} V_t(f) = 0.$$

The next proposition shows that nontrivial continuous martingales never have paths of bounded variation.

Proposition 3.2.7. *Suppose M_t is a square integrable martingale with respect to $\{\mathcal{F}_t\}$ with $M_0 = 0$ such that with probability one, the paths of M_t are continuous with bounded variation. Then $M_t = 0$ for every t.*

Proof. Since M_t has continuous paths, it suffices to show that $\mathbb{P}\{M_t = 0\} = 1$ for every rational t. The argument is the same for all t, so let us assume $t = 1$. We first make the stronger assumption that $V_1(M) \leq K < \infty$. Let

$$Q_n = \sum_{j=1}^{n} \left[M\left(\frac{j}{n}\right) - M\left(\frac{j-1}{n}\right) \right]^2,$$

$$\delta_n = \max_{j=1,\ldots,n} \left\{ \left| M\left(\frac{j}{n}\right) - M\left(\frac{j-1}{n}\right) \right| \right\},$$

and note that

$$Q_n \leq \delta_n \sum_{j=1}^{n} \left| M\left(\frac{j}{n}\right) - M\left(\frac{j-1}{n}\right) \right| \leq \delta_n K.$$

Orthogonality of martingale increments implies that $\mathbb{E}[M_1^2] = \mathbb{E}[Q_n]$. Continuity of the paths implies that with probability one $\delta_n \downarrow 0$ and hence $Q_n \to 0$. Since $Q_n \leq \delta_n K \leq K^2$, we can use the bounded convergence theorem to conclude that

$$\mathbb{E}\left[M_1^2\right] = \lim_{n\to\infty} \mathbb{E}[Q_n] = 0.$$

If $V_1(M)$ is not bounded, we can consider the stopping times

$$T_K = \inf\{t : V_t(M) = K\}.$$

The argument above gives $\mathbb{E}\left[M_{1\wedge T_K}^2\right] = 0$ for each K, and hence with probability one for each K, $M_{1\wedge T_K} = 0$. But $M_1 = \lim_{K\to\infty} M_{1\wedge T_K}$.
$\square$

Using the proposition, we see that if Y_t is an increasing adapted process such that $Z_t^2 - Y_t$ is a martingale, then

$$M_t - (Z_t^2 - Y_t) = Y_t - \langle Z \rangle_t,$$

is a continuous martingale with paths of bounded variation. Therefore, $Y_t = \langle Z \rangle_t$.

3.3 Itô's formula

Itô's formula is the fundamental theorem of stochastic calculus. Before stating it, let us review the fundamental theorem of ordinary calculus. Suppose f is a C^1 function. (We recall that a function if C^k if it has k continuous derivatives.) Then we can expand f in a Taylor approximation,

$$f(t + s) = f(t) + f'(t)\, s + o(s),$$

where $o(s)$ denotes a term such that $o(s)/s \to 0$ as $s \to 0$. We write f as a telescoping sum:

$$f(1) = f(0) + \sum_{j=1}^{n} \left[f\left(\frac{j}{n}\right) - f\left(\frac{j-1}{n}\right) \right].$$

The Taylor approximation gives

$$f\left(\frac{j}{n}\right) - f\left(\frac{j-1}{n}\right) = f'\left(\frac{j-1}{n}\right)\frac{1}{n} + o\left(\frac{1}{n}\right).$$

Hence,

$$f(1) = f(0) + \lim_{n\to\infty}\sum_{j=1}^{n} f'\left(\frac{j-1}{n}\right)\frac{1}{n} + \lim_{n\to\infty}\sum_{j=1}^{n} o\left(\frac{1}{n}\right).$$

The first limit on the right-hand side is the Riemann sum approximation of the definite integral and the second limit equals zero since $n\,o(1/n) \to 0$. Therefore,

$$f(1) = f(0) + \int_0^1 f'(t)\,dt.$$

Itô's formula is similar but it requires considering both first and second derivatives.

Theorem 3.3.1 (Itô's formula I). *Suppose f is a C^2 function and B_t is a standard Brownian motion. Then for every t,*

$$f(B_t) = f(B_0) + \int_0^t f'(B_s)\,dB_s + \frac{1}{2}\int_0^t f''(B_s)\,ds.$$

The theorem is often written in the differential form

$$df(B_t) = f'(B_t)\,dB_t + \frac{1}{2}f''(B_t)\,dt.$$

In other words, the process $Y_t = f(B_t)$ at time t evolves like a Brownian motion with drift $f''(B_t)/2$ and variance $f'(B_t)^2$. Note that $f'(B_t)$ is a continuous adapted process so the stochastic integral is well defined.

To see where the formula comes from, let us assume for ease that $t = 1$. Let us expand f in a second-order Taylor approximation,

$$f(x+y) = f(x) + f'(x)\,y + \frac{1}{2}f''(x)\,y^2 + o(y^2),$$

where $o(y^2)/y^2 \to 0$ as $y \to 0$. Similarly to above, we write

$$f(B_1) - f(B_0) = \sum_{j=1}^{n}\left[f\left(B_{j/n}\right) - f\left(B_{(j-1)/n}\right)\right].$$

Using the Taylor approximation, we write

$$f\left(B_{j/n}\right) - f\left(B_{(j-1)/n}\right)$$

$$= f'\left(B_{(j-1)/n}\right)\Delta_{j,n} + \frac{1}{2}f''\left(B_{(j-1)/n}\right)\Delta_{j,n}^2 + o(\Delta_{j,n}^2),$$

where

$$\Delta_{j,n} = B_{j/n} - B_{(j-1)/n}.$$

Hence $f(B_1) - f(B_0)$ is equal to the sum of the following three limits:

$$\lim_{n\to\infty}\sum_{j=1}^{n} f'\left(B_{(j-1)/n}\right)\left[B_{j/n} - B_{(j-1)/n}\right], \tag{3.9}$$

$$\lim_{n\to\infty}\frac{1}{2}\sum_{j=1}^{n} f''\left(B_{(j-1)/n}\right)\left[B_{j/n} - B_{(j-1)/n}\right]^2, \tag{3.10}$$

$$\lim_{n\to\infty}\sum_{j=1}^{n} o\left(\left[B_{j/n} - B_{(j-1)/n}\right]^2\right). \tag{3.11}$$

The increment of the Brownian motion satisfies $\left[B_{j/n} - B_{(j-1)/n}\right]^2 \approx 1/n$. Since the sum in (3.11) looks like n terms of smaller order than $1/n$ the limit equals zero. The limit in (3.9) is a simple process approximation to a stochastic integral and hence we see that the limit is

$$\int_0^1 f'(B_t)\, dB_t.$$

If f'' were constant, say b, then the limit in (3.10) would be

$$\lim_{n\to\infty}\frac{b}{2}\sum_{j=1}^{n}\left[B_{j/n} - B_{(j-1)/n}\right]^2 = \frac{b}{2}\langle B\rangle_1 = \frac{b}{2}.$$

This tells us what the limit should be in general. Let $h(t) = f''(B_t)$ which is a continuous function. For every $\epsilon > 0$, there exists a step function $h_\epsilon(t)$ such that $|h(t) - h_\epsilon(t)| < \epsilon$ for every t. For fixed ϵ, we can consider each interval on which h_ϵ is constant to see that

$$\lim_{n\to\infty}\sum_{j=1}^{n} h_\epsilon(t)\left[B_{j/n} - B_{(j-1)/n}\right]^2 = \int_0^1 h_\epsilon(t)\, dt.$$

Also,

$$\left| \sum_{j=1}^{n} [h(t) - h_\epsilon(t)] \left[B_{j/n} - B_{(j-1)/n} \right]^2 \right| \leq \epsilon \sum_{j=1}^{n} \left[B_{j/n} - B_{(j-1)/n} \right]^2 \to \epsilon.$$

Therefore, the limit in (3.10) is the same as

$$\lim_{\epsilon \to 0} \frac{1}{2} \int_0^1 h_\epsilon(t)\, dt = \frac{1}{2} \int_0^1 h(t)\, dt = \frac{1}{2} \int_0^1 f''(B_t)\, dt.$$

Example 3.3.1. Let $f(x) = x^2$. Then $f'(x) = 2x$, $f''(x) = 2$, and hence

$$B_t^2 = B_0^2 + \int_0^t f'(B_s)\, dB_s + \frac{1}{2} \int_0^t f''(B_s)\, ds = 2 \int_0^t B_s\, dB_s + t.$$

This gives the equation

$$\int_0^t B_s\, dB_s = \frac{1}{2} [B_t^2 - t],$$

which we mentioned in the last section. This example is particularly easy because f'' is constant. If f'' is not constant, then $f''(B_t)$ is a random variable, and the dt integral is not so easy to compute.

Example 3.3.2. Let $f(x) = e^{\sigma x}$ where $\sigma > 0$. Then $f'(x) = \sigma e^{\sigma x} = \sigma f(x)$, $f''(x) = \sigma^2 e^{\sigma x} = \sigma^2 f(x)$. Let $X_t = f(B_t) = e^{\sigma B_t}$. Then

$$dX_t = f'(B_t)\, dB_t + \frac{1}{2} f''(B_t)\, dt = \sigma X_t\, dB_t + \frac{\sigma^2}{2} X_t\, dt.$$

The process X_t is an example of *geometric Brownian motion* which we discuss in more detail in the next section.

The derivation of Itô's formula given above is essentially correct but ignores some technical details. For ease assume $t = 1$ and assume we have a partition Π of the form

$$0 = t_0 < t_1 < \cdots < t_n = 1,$$

with mesh $\|\Pi\| = \max\{t_j - t_{j-1}\}$. Then we write

$$f(B_1) = f(B_0) + \sum_{j=1}^{n} \left[f(B_{t_j}) - f(B_{t_{j-1}}) \right].$$

Taylor's theorem implies that

$$\frac{m(j,n)}{2} \left[B_{t_j} - B_{t_{j-1}} \right]^2$$
$$\leq \quad f(B_{t_j}) - f(B_{t_{j-1}}) - f'(B_{t_{j-1}}) \left[B_{t_j} - B_{t_{j-1}} \right]$$
$$\leq \quad \frac{M(j,n)}{2} \left[B_{t_j} - B_{t_{j-1}} \right]^2,$$

where $m(j,n), M(j,n)$ denote the minimum and maximum of $f''(x)$ for x on the interval with endpoints $B_{t_{j-1}}$ and B_{t_j}. Hence if we let

$$Q^1(\Pi) = \sum_{j=1}^{n} f'(B_{t_{j-1}}) \left[B_{t_j} - B_{t_{j-1}} \right],$$

$$Q_+^2(\Pi) = \sum_{j=1}^{n} \frac{M(j,n)}{2} \left[B_{t_j} - B_{t_{j-1}} \right]^2,$$

$$Q_-^2(\Pi) = \sum_{j=1}^{n} \frac{m(j,n)}{2} \left[B_{t_j} - B_{t_{j-1}} \right]^2,$$

we have
$$Q_-^2(\Pi) \leq f(B_1) - f(B_0) - Q^1(\Pi) \leq Q_+^2(\Pi).$$

We now suppose that we have a sequence of partitions Π_n of the form

$$0 = t_{0,n} < t_{1,n} < \cdots < t_{k_n,n} = 1,$$

with $\sum \|\Pi_n\| < \infty$. Then we have seen that with probability one, for all $0 < s < t < 1$,

$$\lim_{n \to \infty} \sum_{s \leq t_{j,n} < t} \left[B_{t_{j,n}} - B_{t_{j-1,n}} \right]^2 = t - s.$$

Using the continuity of B_t and f'', we can see that on the event that this is true, we also have

$$\lim_{n \to \infty} Q_-^2(\Pi_n) = \lim_{n \to \infty} Q_+^2(\Pi_n) = \frac{1}{2} \int_0^1 f''(B_s) \, ds.$$

We now assume for the moment that there exists $K < \infty$ such that $|f''(x)| \le K$ for all x. This happens, for example, if f has compact support. Then

$$|f'(B_s) - f'(B_{t_{j-1,n}})| \le K \, |B_s - B_{t_{j-1,n}}|.$$

Let $A_t = f'(B_t)$ and let $A_t^{(n)}$ be the simple process that equals $f'(B_{t_{j-1,n}})$ for $t_{j-1,n} \le t < t_{j,n}$. For $t_{j-1,n} \le s < t_{j,n}$,

$$\mathbb{E}([A_t - A_t^{(n)}]^2) \le K^2 \, \mathbb{E}([B_s - B_{t_{j-1,n}}]^2) = K^2 \, [s - t_{j-1,n}] \le K^2 \, \|\Pi_n\|.$$

Therefore,

$$\int_0^1 \mathbb{E}([A_t - A_t^{(n)}]^2) \, dt \le K^2 \, \|\Pi_n\|.$$

In particular, we get the following.

Proposition 3.3.2. *Suppose that f'' is bounded and $\{\Pi_n\}$ is a sequence of partitions with $\sum \|\Pi_n\| < \infty$. Let*

$$Y_t^{(n)} = f(B_0) + \sum_{j=1}^{k_n} f'(B_{t_{j-1,n}}) \, [B_{t_{j,n}} - B_{t_{j-1,n}}]$$

$$+ \sum_{j=1}^{k_n} \frac{f''(B_{t_{j-1,n}})}{2} \, [B_{t_{j,n}} - B_{t_{j-1,n}}]^2.$$

Then with probability one

$$\lim_{n \to \infty} \max_{0 \le t \le 1} |f(B_t) - Y_t^{(n)}| = 0.$$

If f'' is not bounded, we can still prove Itô's formula by using stopping times. This general procedure is often called *localization*. For each K, let

$$T_K = \inf\{t : |f''(B_t)| \ge K\}.$$

Then, our argument shows that with probability one for each K,

$$f(B_{t \wedge T_K}) = \int_0^{t \wedge T_K} f'(B_s) \, dB_s + \frac{1}{2} \int_0^{t \wedge T_K} f''(B_s) \, ds.$$

However, with probability one $T_K \to \infty$ as $K \to \infty$, and hence this gives us a formula for $f(B_t)$.

Suppose that f is defined only on an interval $I = (a, b)$ and $B_0 \in I$. Let

$$T_\epsilon = \inf\{t : B_t \leq a + \epsilon \text{ or } B_t \geq b - \epsilon\}, \quad T = T_0.$$

We can apply Itô's formula to conclude for all t and all $\epsilon > 0$.

$$f(B_{t \wedge T_\epsilon}) = f(B_0) + \int_0^{t \wedge T_\epsilon} f'(B_s)\, dB_s + \frac{1}{2} \int_0^{t \wedge T_\epsilon} f''(s)\, ds.$$

This is sometimes written shorthand as

$$df(B_t) = f'(B_t)\, dB_t + \frac{1}{2} f''(B_t)\, dt, \quad 0 \leq t < T.$$

The general strategy for proving the generalizations of Itô's formula that we give in the next couple of sections is the same, and we will not give the details.

3.4 More versions of Itô's formula

We first extend Itô's formula to functions that depend on time as well as position.

Theorem 3.4.1 (Itô's Formula II)**.** *Suppose $f(t, x)$ is a function that is C^1 in t and C^2 in x. If B_t is a standard Brownian motion, then*

$$f(t, B_t) = f(0, B_0) + \int_0^t \partial_x f(s, B_s)\, dB_s$$

$$+ \int_0^t \left[\partial_s f(s, B_s) + \frac{1}{2} \partial_{xx} f(s, B_s) \right] ds,$$

or in differential form,

$$df(t, B_t) = \partial_x f(t, B_t)\, dB_t + \left[\partial_t f(t, B_t) + \frac{1}{2} \partial_{xx} f(t, B_t) \right] dt.$$

This is derived similarly to the first version except when we expand with a Taylor polynomial around x we get another term:

$$f(t + \Delta t, x + \Delta x) - f(t, x) =$$

$$\partial_t f(t,x)\,\Delta t + o(\Delta t) + \partial_x f(t,x)\,\Delta x + \frac{1}{2}\partial_{xx} f(t,x)\,(\Delta x)^2 + o((\Delta x)^2).$$

If we set $\Delta t = 1/n$ and write a telescoping sum for $f(1, B_1) - f(0, B_0)$ we get terms as (3.9)– (3.11) as well as two more terms:

$$\lim_{n\to\infty} \sum_{j=1}^{n} \partial_t f((j-1)/n, B_{(j-1)/n})\,(1/n), \tag{3.12}$$

$$\lim_{n\to\infty} \sum_{j=1}^{n} o(1/n). \tag{3.13}$$

The limit in (3.13) equals zero, and the sum in (3.12) is a Riemann sum approximation of a integral and hence the limit is

$$\int_0^1 \partial_t f(t, B_t)\,dt.$$

Example 3.4.1. Let $f(t,x) = e^{at+bx}$. Then

$$\partial_t f(t,x) = a\,f(t,x), \quad \partial_x f(t,x) = b\,f(t,x), \quad \partial_{xx} f(t,x) = b^2\,f(t,x).$$

Therefore, if $X_t = f(t, B_t) = \exp\{at + bB_t\}$,

$$
\begin{aligned}
dX_t &= \left[\partial_t f(t, B_t) + \frac{1}{2}\partial_{xx} f(t, B_t)\right] dt + \partial_x f(t, B_t)\,dB_t \\
&= \left(a + \frac{b^2}{2}\right) X_t\,dt + b\,X_t\,dB_t.
\end{aligned}
$$

Definition A process X_t is a *geometric Brownian motion* with *drift m* and *volatility* σ if it satisfies the SDE

$$dX_t = m\,X_t\,dt + \sigma\,X_t\,dB_t = X_t\,[m\,dt + \sigma\,dB_t], \tag{3.14}$$

where B_t is a standard Brownian motion.

Example 3.4.1 shows that if B_t is a standard Brownian motion and

$$X_t = X_0 \exp\left\{\left(m - \frac{\sigma^2}{2}\right) t + \sigma B_t\right\}, \tag{3.15}$$

then X_t is a geometric Brownian motion with parameters (m, σ). Even though we have an exact expression (3.15) for geometric Brownian motion, it is generally more useful to think of it in terms of its SDE (3.14).

Geometric Brownian motion is more natural than usual Brownian motion for modeling prices of assets such as stock. It measures changes in terms of fractions or percentages of the current price rather than the listed price per share. In particular, the latter quantity includes a rather arbitrary unit "share" which does not appear if one models with geometric Brownian motion.

The geometric Brownian motion (3.15) is what is sometimes called a "strong" solution to the SDE (3.14). (All of the solutions to SDEs that we discuss in this book are strong solutions.) We will not give the exact definition, but roughly speaking, if one uses the same Brownian motion B_t in both places, one gets the same function. Let us explain this in terms of simulation. Suppose a small Δt is chosen. Then we can define $B_{k\Delta t}$ by $B_0 = 0$ and

$$B_{k\Delta t} = B_{(k-1)\Delta t} + \sqrt{\Delta t}\, N_k, \qquad (3.16)$$

where $N_1, N_2, \ldots$ is a sequence of independent $N(0,1)$ random variables. Using the *same* sequence, we could define an approximate solution to (3.14) by choosing $X_0 = e^0 = 1$ and

$$X_{k\Delta t} = X_{(k-1)\Delta t} + X_{(k-1)\Delta t}\left[m\,\Delta t + \sigma\sqrt{\Delta t}\, N_j\right]. \qquad (3.17)$$

If Δt is small, this should be very close to

$$Y_{k\Delta t} = \exp\left\{\left(m - \frac{\sigma^2}{2}\right)(k\Delta t) + \sigma\, B_{k\Delta t}\right\}. \qquad (3.18)$$

In Exercise 3.9 you are asked to do a simulation to compare (3.17) and (3.18).

To do formal[1] calculations in usual calculus, one writes down differentials and throws away all terms that are of smaller order than dt. In stochastic calculus, one can go far computing similarly using the rules

$$(dB_t)^2 = dt, \quad (dB_t)(dt) = 0, \quad (dt)^2 = 0.$$

---

[1]Mathematicians use the word formal to refer to calculations that look correct "in form", but for which not all the steps have been justified. When they construct proofs, they often start with formal calculations and then go back and justify the steps.

In some of our derivations below, we will use this kind of argument. For example, a formal derivation of Itô's formula II can be given as

$$df(t, B_t) = \partial_t f(t, B_t)\, dt + \partial_x f(t, B_t)\, dB_t + \frac{1}{2}\, \partial_{xx} f(t, B_t)\, (dB_t)^2$$

$$+ o(dt) + o((dt)(dB_t)) + o((dB_t)^2).$$

By setting $(dB_t)^2 = dt$ and setting the last three terms equal to zero we get the formula.

Suppose that X_t satisfies

$$dX_t = R_t\, dt + A_t\, dB_t, \tag{3.19}$$

which we recall is equivalent to

$$X_t = X_0 + \int_0^t R_s\, ds + \int_0^t A_s\, dB_s.$$

Whenever we write such equations we will assume implicitly that R_t, A_t are adapted processes with piecewise continuous paths. As before, we define the quadratic variation by

$$\langle X \rangle_t = \lim_{n \to \infty} \sum_{j \le tn} \left[X\left(\frac{j}{n}\right) - X\left(\frac{j-1}{n}\right) \right]^2.$$

As in the case for Brownian motion, the drift term does not contribute to the quadratic variation,

$$\langle X \rangle_t = \left\langle \int A\, dB \right\rangle_t = \int_0^t A_s^2\, ds.$$

This is often written in the differential form

$$d\langle X \rangle_t = A_t^2\, dt.$$

To see that R_t does not appear, we can do the formal calculation

$$
\begin{aligned}
(dX_t)^2 &= (R_t\, dt + A_t\, dB_t)^2 \\
&= R_t^2\, (dt)^2 + 2\, R_t\, A_t\, (dt)(dB_t) + A_t^2\, (dB_t)^2 = A_t^2\, dt.
\end{aligned}
$$

If X_t satisfies (3.19) and H_t is another adapted process we define

$$\int_0^t H_s \, dX_s = \int_0^t H_s \, R_s \, ds + \int_0^t H_s \, A_s \, dB_s.$$

If we view this in terms of simulation then

$$Y_t = \int_0^t H_s \, dX_s$$

can be simulated by

$$\Delta Y_t = H_t \, \Delta X_t = H_t \left[X_{t+\Delta t} - X_t \right] = H_t \left[R_t \, \Delta t + A_t \, \sqrt{\Delta t} \, N \right],$$

where $N \sim N(0,1)$.

Theorem 3.4.2 (Itô's formula III). *Suppose X_t satisfies (3.19) and $f(t,x)$ is C^1 in t and C^2 in x. Then*

$$
\begin{aligned}
df(t, X_t) &= \partial_t f(t, X_t) \, dt + \partial_x f(t, X_t) \, dX_t + \frac{1}{2} \partial_{xx} f(t, X_t) \, d\langle X \rangle_t \\
&= \left[\partial_t f(t, X_t) + R_t \, \partial_x f(t, X_t) + \frac{A_t^2}{2} \partial_{xx} f(t, X_t) \right] dt \\
&\quad + A_t \, \partial_x f(t, X_t) \, dB_t.
\end{aligned}
$$

Example 3.4.2. Suppose X is a geometric Brownian motion satisfying

$$dX_t = m \, X_t \, dt + \sigma \, X_t \, dB_t.$$

Let $f(t,x) = e^{-t} x^3$. Then

$$\partial_t f(t,x) = -e^{-t} x^3 = -f(t,x),$$

$$\partial_x f(t,x) = 3x^2 \, e^{-t} = \frac{3}{x} f(t,x), \quad \partial_{xx} f(t,x) = 6x \, e^{-t} = \frac{6}{x^2} f(t,x),$$

and

$$
\begin{aligned}
df(t, X_t) &= \partial_t f(t, X_t) \, dt + \partial_x f(t, X_t) \, dX_t + \frac{1}{2} \partial_{xx} f(t, X_t) \, d\langle X \rangle_t \\
&= \left[\partial_t f(t, X_t) + m \, X_t \, \partial_x f(t, X_t) + \frac{\sigma^2 X_t^2}{2} \, \partial_{xx} f(t, X_t) \right] dt \\
&\quad + \sigma \, X_t \, \partial_x f(t, X_t) \, dB_t \\
&= \left(-1 + 3m + \frac{6\sigma^2}{2} \right) f(t, X_t) \, dt + 3 \, \sigma \, f(t, X_t) \, dB_t.
\end{aligned}
$$

$$d[e^{-t} X_t^3] = 3 e^{-t} X_t^3 \left[\left(-\frac{1}{3} + m + \sigma^2 \right) dt + \sigma \, dB_t \right].$$

Example 3.4.3. The *exponential SDE* is

$$dX_t = A_t\, X_t\, dB_t \quad X_0 = x_0.$$

We claim that the solution is

$$X_t = x_0\, \exp\left\{ \int_0^t A_s\, dB_s - \frac{1}{2} \int_0^t A_s^2\, ds \right\}. \tag{3.20}$$

To see this, let

$$Y_t = \int_0^t A_s\, dB_s - \frac{1}{2} \int_0^t A_s^2\, ds,$$

which satisfies

$$dY_t = -\frac{A_t^2}{2}\, dt + A_t\, dB_t, \quad d\langle Y \rangle_t = A_t^2\, dt.$$

If $f(x) = x_0\, e^x$, then $f(x) = f'(x) = f''(x)$ and

$$df(Y_t) = f'(Y_t)\, dY_t + \frac{1}{2} f''(Y_t)\, d\langle Y \rangle_t = A_t\, X_t\, dB_t.$$

We can compare (3.20) to the solution to the deterministic exponential equation

$$f'(t) = r(t)\, f(t)\, dt, \quad f(0) = x_0,$$

which is

$$f(t) = x_0\, \exp\left\{ \int_0^t r(s)\, ds \right\}.$$

Itô's formula requires that the function $f(t,x)$ be C^1 in t and C^2 in x. In applications, one often has functions that are C^2 only for x in an interval of the reals. In this case, we can use Itô's formula until the (random) time that the process leave the interval.

Theorem 3.4.3 (Itô's formula III, local form). *Suppose X_t satisfies (3.19) with $a < X_0 < b$, and $f(t,x)$ is C^1 in t and C^2 in x for $a < x < b$. Let $T = \inf\{t : X_t = a \text{ or } X_t = b\}$. Then if $t < T$,*

$$f(t, X_t) = f(0, X_0) + \int_0^t A_s\, \partial_x f(s, X_s)\, dB_s$$

$$+ \int_0^t \left[\partial_s f(s, X_s) + R_s\, \partial_x f(s, X_s) + \frac{A_s^2}{2} \partial_{xx} f(s, X_s) \right] ds.$$

We write the conclusion of the theorem in differential form by

$$df(t, X_t) = \dot{f}(t, X_t)\, dt + f'(t, X_t)\, dX_t + \frac{1}{2}\, f''(t, X_t)\, d\langle X \rangle_t \quad t < T.$$

Example 3.4.4. Suppose that B_t is a standard Brownian motion, and $Y_t = t/B_t^2$. Let $T = \inf\{t : B_t = 0\}$. Then we can use Itô's formula to see that for $t < T$,

$$
\begin{aligned}
dY_t &= \dot{f}(t, B_t)\, dt + f'(t, B_t)\, dB_t + \frac{1}{2}\, f''(t, B_t)\, dt \\
&= [B_t^{-2} + 3tB_t^{-4}]\, dt - 2tB_t^{-3}\, dB_t.
\end{aligned}
$$

To prove the local form of Itô's formula, we let $\epsilon > 0$ and $T_\epsilon = \inf\{t : X_t \le a + 2\epsilon \text{ or } X_t \ge b - 2\epsilon\}$. We can find a C^∞ function ϕ_ϵ such that $\phi_\epsilon(x) = 0$ if $x < a + \epsilon$ or $x > b - \epsilon$ and such that $\phi(x) = 1$ for $a + 2\epsilon \le x \le b - 2\epsilon$. Let $f_\epsilon(t, x) = \phi_\epsilon(x)\, f(t, x)$. Then f_ϵ satisfies the conditions of Itô's formula and

$$f_\epsilon(t, X_t) - f_\epsilon(0, X_0) = \int_0^t A_s\, f'_\epsilon(s, X_s)\, dB_s$$

$$+ \int_0^t \left[\dot{f}_\epsilon(s, X_s) + R_s\, f'_\epsilon(s, X_s) + \frac{A_s^2}{2} f''_\epsilon(s, X_s) \right]\, ds.$$

If $t < T_\epsilon$, then both sides of this equation are not changed if we replace f_ϵ with f.

3.5 Diffusions

Geometric Brownian motion is an example of a time-homogeneous diffusion. We say that X_t is a *diffusion (process)* if it is a solution to an SDE of the form

$$dX_t = m(t, X_t)\, dt + \sigma(t, X_t)\, dB_t, \tag{3.21}$$

where $m(t, x), \sigma(t, x)$ are functions. It is called *time-homogeneous* if the functions do not depend on t,

$$dX_t = m(X_t)\, dt + \sigma(X_t)\, dB_t.$$

Diffusion processes are Markov processes. Given all the information in $\mathcal{F}_t$, the only important information needed to give the conditional distribution of $X_s, s \geq t$ is X_t. Simulations of diffusions can be done using the stochastic Euler rule:

$$X_{t+\Delta t} = X_t + m(t, X_t)\,\Delta t + \sigma(t, X_t)\,\sqrt{\Delta t}\,N,$$

where $N \sim N(0, 1)$. The study of diffusions leads to partial differential equations.

We define the generator $L = L^0$ of a (time homogeneous) Markov process X_t by

$$Lf(x) = \lim_{t \downarrow 0} \frac{\mathbb{E}^x[f(X_t)] - f(x)}{t}.$$

We will use Itô's formula to understand the generator of the diffusion X_t. For ease, we will assume that m and σ are bounded continuous functions. If $f : \mathbb{R} \to \mathbb{R}$ is a C^2 function with bounded first and second derivatives, then Itô's formula gives

$$
\begin{aligned}
df(X_t) &= f'(X_t)\,dX_t + \frac{1}{2}\,f''(X_t)\,d\langle X \rangle_t \\
&= \left[m(t, X_t)\,f'(X_t) + \frac{\sigma^2(t, X_t)}{2}\,f''(X_t) \right]\,dt \\
&\quad + f'(X_t)\,\sigma(t, X_t)\,dB_t.
\end{aligned}
$$

That is,

$$
f(X_t) - f(X_0) = \int_0^t \left[m(s, X_s)\,f'(X_s) + \frac{\sigma^2(s, X_s)}{2}\,f''(X_s) \right]\,ds
$$

$$
+ \int_0^t f'(X_s)\,\sigma(s, X_s)\,dB_s.
$$

The second term on the right-hand side is a martingale (since the integrand is bounded) and has expectation zero, so the expectation of the right-hand side is

$$t\,\mathbb{E}[Y_t],$$

where

$$
Y_t = \frac{1}{t} \int_0^t \left[m(s, X_s)\,f'(X_s) + \frac{\sigma^2(s, X_s)}{2}\,f''(X_s) \right]\,ds.
$$

The fundamental theorem of calculus implies that

$$\lim_{t \downarrow 0} Y_t = m(0, X_0) \, f'(X_0) + \frac{\sigma^2(0, X_0)}{2} \, f''(X_0).$$

Moreover, the random variables Y_t are uniformly bounded which allows us to take the limit inside the expectation,

$$Lf(x) = \lim_{t \downarrow 0} \frac{\mathbb{E}^x[f(X_t)] - f(x)}{t} = m(0, x) \, f'(x) + \frac{\sigma^2(0, x)}{2} \, f''(x).$$

Similarly, if we define

$$L^t f(x) = \lim_{s \downarrow 0} \frac{\mathbb{E}[f(X_{t+s}) - f(X_t) \mid X_t = x]}{s},$$

we get

$$L^t f(x) = m(t, x) \, f'(x) + \frac{\sigma^2(t, x)}{2} \, f''(x).$$

The assumption that m, σ are bounded is stronger than we would like (for example, the differential equation (3.14) for geometric Brownian motion does not satisfy this), but one can ease this by stopping the path before it gets too large. We will not worry about this now.

In this section we assumed we had a solution to the SDE (3.21). This leads naturally to the question: given functions $m(t, x), \sigma(t, x)$, can we find a process X_t satisfying (3.21)?

Consider first the deterministic equation

$$y'(t) = F(t, y(t)), \quad y(0) = y_0. \tag{3.22}$$

We will assume that F is uniformly Lipschitz, that is, there exists $\beta < \infty$ such that for all t, x, y,

$$|F(t, x) - F(t, y)| \leq \beta \, |x - y|. \tag{3.23}$$

We can construct a solution to (3.22) valid for $0 \leq t \leq t_0$ by using the following method called *Picard iteration*. We start with the initial function

$$y_0(t) = y_0,$$

and recursively define

$$y_k(t) = y_0 + \int_0^t F(s, y_{k-1}(s)) \, ds.$$

Note that

$$|y_1(t) - y_0(t)| \le \int_0^t |F(s, y_0)| \, ds \le Ct,$$

where $C = C(y_0, t_0) = \max_{0 \le s \le t_0} |F(s, y_0)|$, and for $k \ge 1$,

$$
\begin{aligned}
|y_{k+1}(t) - y_k(t)| &\le \int_0^t |F(s, y_k(s)) - F(s, y_{k-1}(s))| \, ds \\
&\le \beta \int_0^t |y_k(s) - y_{k-1}(s)| \, ds
\end{aligned}
$$

Using this estimate and induction, we see that

$$|y_{k+1}(t) - y_k(t)| \le \frac{\beta^k \, C \, t^{k+1}}{(k+1)!}, \quad 0 \le t \le t_0.$$

In particular, the limit

$$y(t) = \lim_{k \to \infty} y_k(t),$$

exists and

$$|y_k(t) - y(t)| \le C \sum_{j=k}^{\infty} \frac{\beta^j \, t^{j+1}}{(j+1)!}.$$

If we let

$$\tilde{y}(t) = y_0 + \int_0^t F(s, y(s)) \, ds,$$

then for each k,

$$
\begin{aligned}
|\tilde{y}(t) - y_{k+1}(t)| &\le \int_0^t |F(s, y(s)) - F(s, y_k(s))| \, ds \\
&\le \beta \int_0^t |y(s) - y_k(s)| \, ds,
\end{aligned}
$$

and hence $\tilde{y} = y$, and y satisfies (3.22). This argument assumed that (3.23) holds for all s, t, x, y, and we put no restriction on t_0. If instead we had such an estimate only for times near zero and x, y near y_0, we could establish the same result for some (but not necessarily all) positive values of t_0.

Let us now consider (3.21) where the functions $m(t,x)$ and $\sigma(t,x)$ both satisfy (3.23). For ease, choose $t_0 = 1$ and define recursively

$$X_t^0 = y_0, \quad 0 \le t \le 1,$$

$$X_t^{(k+1)} = y_0 + \int_0^t m(s, X_s^{(k)})\, ds + \int_0^t \sigma(s, X_s^{(k)})\, dB_s,$$

so that

$$\mathbb{E}\left[|X_t^{(k+1)} - X_t^k|^2\right] \le 2\mathbb{E}\left[\left(\int_0^t \beta\, |X_s^{(k)} - X_s^{(k-1)}|\, ds\right)^2\right]$$

$$+ 2\,\mathbb{E}\left[\left(\int_0^t [\sigma(s, X_s^{(k)}) - \sigma(s, X_s^{(k-1)})]\, dB_s\right)^2\right].$$

Using the Hölder inequality on the ds integral we see that for $t \le 1$,

$$\mathbb{E}\left[\left(\int_0^t \beta\, |X_s^{(k)} - X_s^{(k-1)}|\, ds\right)^2\right]$$

$$\le \;\; \mathbb{E}\left[\beta^2\, t \int_0^t |X_s^{(k)} - X_s^{(k-1)}|^2\, ds\right]$$

$$\le \;\; \beta^2 \int_0^t \mathbb{E}\left[|X_s^{(k)} - X_s^{(k-1)}|^2\right]\, ds.$$

The variance rule for the stochastic integral gives

$$\mathbb{E}\left[\left(\int_0^t [\sigma(s, X_s^{(k)}) - \sigma(s, X_s^{(k-1)})]\, dB_s\right)^2\right]$$

$$= \;\; \int_0^t \mathbb{E}\left([\sigma(s, X_s^{(k)}) - \sigma(s, X_s^{(k-1)})]^2\right)\, ds$$

$$\le \;\; \beta^2 \int_0^t \mathbb{E}\left[|X_s^{(k)} - X_s^{(k-1)}|^2\right]\, ds.$$

Combining these estimates, we see that there exists $\lambda < \infty$ such that for all $0 \le t \le 1$ and positive integers k,

$$\mathbb{E}\left[|X_t^{(k+1)} - X_t^k|^2\right] \le \frac{\lambda^{k+1}\, t^{k+1}}{(k+1)!}.$$

We then define $X_t = \lim_k X_t^{(k)}$; as before, we do this for dyadic t as an L^2 limit and extend it to other t by continuity. This gives a solution.

If the Lipschitz condition holds only locally, then we can use this argument to show that there exists a stopping time T with $\mathbb{P}\{T > 0\} = 1$ such that $X_t, 0 \le t < T$ satisfies the equation. In the random case the time T for which the solution is valid is a random variable.

3.6   Covariation and the product rule

Suppose X_t, Y_t satisfy

$$dX_t = H_t \, dt + A_t \, dB_t, \quad dY_t = K_t \, dt + C_t \, dB_t. \tag{3.24}$$

Here B_t is the same standard Brownian motion in both equations. As before, the covariation process is defined by

$$\langle X, Y \rangle_t = \lim_{n \to \infty} \sum_{j \le tn} \left[X\left(\frac{j}{n}\right) - X\left(\frac{j-1}{n}\right) \right] \left[Y\left(\frac{j}{n}\right) - Y\left(\frac{j-1}{n}\right) \right].$$

Using the formal rules for stochastic calculus we get

$$
\begin{aligned}
[dX_t]\,[dY_t] &= [H_t \, dt + A_t \, dB_t]\,[K_t \, dt + C_t \, dB_t] \\
&= A_t \, C_t \, dt + O((dt)^2) + O((dt)(dB_t)).
\end{aligned}
$$

This indeed shows what happens and we get

$$\langle X, Y \rangle_t = \int_0^t A_s \, C_s \, ds,$$

or in differential form,

$$d\langle X, Y \rangle_t = A_t \, C_t \, dt.$$

The product rule for the usual derivative can be written in differential form as

$$d(fg) = f \, dg + g \, df = fg' \, dt + f'g \, dt.$$

It can be obtained formally by writing

$$
\begin{aligned}
d(fg) &= f(t + dt)\, g(t + dt) - f(t)\, g(t) \\
&= [f(t + dt) - f(t)]\, g(t + dt) + f(t)\,[g(t + dt) - g(t)] \\
&= (df)g + (dg)f + (df)(dg).
\end{aligned}
$$

Since $df = f'dt, dg = g'dt$ and $df\, dg = O((dt)^2)$, we get the usual product formula.

If X_t, Y_t are process as above, then when we take the stochastic differential $d(X_tY_t)$ we get in the same way

$$d(X_tY_t) = X_t\, dY_t + Y_t\, dX_t + (dX_t)\,(dY_t).$$

However, the $(dX_t)\,(dY_t)$ term does not vanish, but rather equals $d\langle X, Y\rangle_t$. This gives the stochastic product rule.

Theorem 3.6.1 (Stochastic product rule). *If X_t, Y_t satisfy (3.24), then*

$$d(X_tY_t) = X_t\, dY_t + Y_t\, dX_t + d\langle X, Y\rangle_t.$$

In other words,

$$
\begin{aligned}
X_tY_t &= X_0\, Y_0 + \int_0^t X_s\, dY_s + \int_0^t Y_s\, dX_s + \int_0^t d\langle XY\rangle_s \\
&= X_0\, Y_0 + \int_0^t [X_s\, K_s + Y_s\, H_s + A_s\, C_s]\, ds \\
&\quad + \int_0^t [X_s\, C_s + Y_s\, A_s]\, dB_s.
\end{aligned}
$$

Example 3.6.1. Suppose B_t is a standard Brownian motion and X_t is the geometric Brownian motion satisfying

$$dX_t = m\, X_t\, dt + \sigma\, X_t\, dB_t.$$

Then in the notation of (3.24), $H_t = m\, X_t$ and $A_t = \sigma\, X_t$. If we set $Y_t = B_t$, then $K_t = 0, C_t = 1$. Therefore,

$$
\begin{aligned}
d(B_tX_t) &= B_t\, dX_t + X_t\, dB_t + d\langle B, X\rangle_t \\
&= B_t\, [m\, X_t\, dt + \sigma\, X_t\, dB_t] + X_t\, dB_t + \sigma\, X_t\, dt \\
&= X_t\, [(m\, B_t + \sigma)\, dt + (\sigma\, B_t + 1)\, dB_t].
\end{aligned}
$$

3.7 Several Brownian motions

Up to this point, we have integrated with respect to a single Brownian motion. Extending the definitions to several Brownian motions is straightforward.

Suppose that $B_t^1, \ldots, B_t^d$ are independent Brownian motions with respect to the filtration $\{\mathcal{F}_t\}$, and that $A_t^1, \ldots, A_t^d$ are adapted processes. We write

$$dX_t = H_t \, dt + \sum_{j=1}^{d} A_t^j \, dB_t^j,$$

to mean

$$X_t = X_0 + \int_0^t H_s \, ds + \sum_{j=1}^{d} \int_0^t A_s^j \, dB_s^j.$$

The covariation process can be computed if we remember the rule

$$\langle B^i, B^j \rangle = 0, \quad i \neq j.$$

In particular, if

$$Y_t = Y_0 + \int_0^t K_s \, ds + \sum_{j=1}^{d} \int_0^t C_s^j \, dB_s^j,$$

then

$$d\langle X, Y \rangle_t = \sum_{j=1}^{d} A_t^j \, C_t^j \, dt.$$

Itô's formula for several processes can be obtained by expanding in a Taylor approximation. As before, we need the first order term in the time variable and all second order terms in the space variables. We state the theorem; the proof is essentially the same as for the previous version. The reader may note that this version includes all the previous versions as special cases, so we call this the final form. If $f : [0, \infty) \times \mathbb{R}^d \to \mathbb{R}$ is a function, we write

$$\dot{f}(t, \mathbf{x}) = \partial_t f(t, \mathbf{x}), \quad \partial_j f(t, \mathbf{x}) = \partial_{x_j} f(t, \mathbf{x}), \quad \partial_{jk} f(t, \mathbf{x}) = \partial_{x_j x_k} f(t, \mathbf{x}).$$

If $\mathbf{X}_t = (X_t^1, \ldots, X_t^n)$ we write

$$\nabla f(t, \mathbf{X}_t) \cdot d\mathbf{X}_t = \sum_{k=1}^{n} \partial_k f(t, \mathbf{X}_t) \, dX_t^k.$$

Theorem 3.7.1 (Itô's formula, final form)**.** *Suppose $B_t^1, \ldots, B_t^d$ are independent standard Brownian motions, and $X_t^1, \ldots, X_t^n$ are processes satisfying*

$$dX_t^k = H_t^k \, dt + \sum_{i=1}^{d} A_t^{i,k} \, dB_t^i.$$

Suppose $f(t, \mathbf{x}), t \geq 0, \mathbf{x} \in \mathbb{R}^n$, is a function that is C^1 in t and C^2 in $\mathbf{x} = (x_1, \ldots, x_n)$. Then, if $\mathbf{X}_t = (X_t^1, \ldots, X_t^n)$,

$$df(t, \mathbf{X}_t) = \dot{f}(t, \mathbf{X}_t)\, dt + \nabla f(t, \mathbf{X}_t) \cdot d\mathbf{X}_t$$

$$+ \frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} \partial_{jk} f(t, \mathbf{X}_t)\, d\langle X^j, X^k \rangle_t.$$

In other words,

$$f(t, \mathbf{X}_t) - f(0, \mathbf{X}_0) = \sum_{i=1}^{d} \int_0^t \left[\sum_{k=1}^{n} \partial_k f(s, \mathbf{X}_s)\, A_s^{i,k} \right] dB_s^i$$

$$+ \int_0^t \left[\dot{f}(s, \mathbf{X}_s) + \left(\sum_{k=1}^{n} \partial_k f(s, \mathbf{X}_s)\, H_s^k \right) \right.$$

$$\left. + \left(\frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{n} \sum_{k=1}^{n} \partial_{jk} f(s, \mathbf{X}_s)\, A_s^{i,j}\, A_s^{i,k} \right) \right] dt$$

If $f(t, x)$ is C^2 only in an open subset $U \subset \mathbb{R}^n$, we can give a local form of this theorem as in Theorem 3.4.3 to see that the conclusion holds until the first time that $\mathbf{X}_t$ leaves U. Since it comes up often, we state a particular case of this theorem. The *Laplacian* of a function $f : \mathbb{R}^d \to \mathbb{R}$ is defined by

$$\nabla^2 f(\mathbf{x}) = \sum_{j=1}^{d} \partial_{jj} f(\mathbf{x}).$$

Other standard notations for ∇^2 are Δ and $\nabla \cdot \nabla$. In the statement below the gradient ∇ and the Laplacian ∇^2 are taken with respect to the $\mathbf{x}$ variable only.

Theorem 3.7.2. *Suppose $B_t = (B_t^1, \ldots, B_t^d)$ is a standard Brownian motion in $\mathbb{R}^d$. Then if $f : [0, \infty) \times \mathbb{R}^d$ is C^1 in t and C^2 in $\mathbf{x} \in \mathbb{R}^d$, then*

$$df(t, B_t) = \nabla f(t, B_t) \cdot dB_t + \left[\dot{f}(t, B_t) + \frac{1}{2} \nabla^2 f(t, B_t) \right] dt$$

A function f is *harmonic* if $\nabla^2 f = 0$. There is a close relationship between Brownian motion, martingales, and harmonic functions that we discuss in Chapter 8.

3.8   Exercises

Exercise 3.1. Suppose A_t is a simple process with $|A_t| \leq C$ for all t. Let

$$Z_t = \int_0^t A_s \, dB_s.$$

Show that

$$\mathbb{E}\left[Z_t^4\right] \leq 4\,C^4\,t^2.$$

Hint: We may assume $C = 1$. By conditioning with respect to $\mathcal{F}_t$ show that if $s, t > 0$,

$$\mathbb{E}\left[Z_{t+s}^4\right] \leq \mathbb{E}\left[Z_t^4\right] + s^2 + 2\sqrt{\mathbb{E}[Z_t^4]}\,s.$$

Exercise 3.2. Use Itô's formula to find the stochastic differential $f(t, B_t)$ where B_t is a standard Brownian motion and

1. $f(t, x) = \sin x$;

2. $f(t, x) = e^{-t}\,(x/t)^2$

3. Repeat these exercises for $f(t, X_t)$ where

$$dX_t = X_t\,[m\,dt + \sigma\,dB_t].$$

Exercise 3.3. Suppose an asset follows the following geometric SDE,

$$dX_t = 4\,X_t\,dt + X_t\,dB_t.$$

1. Write the exact solution of this equation. In other words, find X_t as a function of B_t.

2. Suppose $X_0 = 2$. What is the probability that $X_1 > 8$?

3. Suppose $X_0 = 1$. What is the probability that $X_2 < 6$?

4. Suppose $X_0 = 4.4565$. What is the probability that $X_t < 0$ for some $2 < t < 5$?.

Exercise 3.4. Show that if B_t is a standard Brownian motion, A_t is a simple process, and

$$Z_t = \int_0^t A_s \, dB_s,$$

then

$$M_t = Z_t^2 - \langle Z \rangle_t$$

is a martingale.

Exercise 3.5. Suppose that two assets X_t, Y_t follow the SDEs

$$dX_t = X_t \left[\mu_1 \, dt + \sigma_1 \, dB_t \right],$$

$$dY_t = Y_t \left[\mu_2 \, dt + \sigma_2 \, dB_t \right],$$

where B_t is a standard Brownian motion. Suppose also that $X_0 = Y_0 = 1$.

1. Let $Z_t = X_t Y_t$. Give the SDE satisfied by Z_t; in other words write an expression for dZ_t in terms of $Z_t, \mu_1, \mu_2, \sigma_1, \sigma_1$.

2. Does there exist a function $f : \mathbb{R} \to \mathbb{R}$ such that $f(X_t) = B_t$ for all t?

3. Does there exist a function $g : \mathbb{R} \to \mathbb{R}$ such that $g(Z_t) = B_t$ for all t?

Exercise 3.6. Suppose B_t is a standard Brownian motion and X_t satisfies

$$dX_t = X_t^2 \, dt + X_t \, dB_t.$$

For each of the following find A_t, C_t such that

$$d\langle Y \rangle_t = A_t \, dt, \quad d\langle Y, X \rangle_t = C_t \, dt.$$

1. $Y_t = B_t$.

2. $Y_t = X_t^3$.

3.
$$Y_t = \exp \left\{ \int_0^t (X_s^2 + 1) \, ds \right\}.$$

Exercise 3.7. Consider the geometric SDE

$$dX_t = X_t [-2 \, dt + 2 \, dB_t].$$

Use the stochastic Euler's method to do Monte Carlo simulations for $X_t, 0 \leq t \leq 2$ assuming $X_0 = 1$. Use $\Delta t = .01$.

1. Graph at least one of your simulations.

2. Use the simulation to estimate the probability that $X_2 \geq 3$. Do at least 1000 simulations.

3. Compute the exact probability (up to at least three decimal places) that $X_2 \geq 3$ and compare this to your simulations.

Exercise 3.8. Suppose $f(t)$ is a (nonrandom) continuously differentiable function of t and B_t is a standard Brownian motion. Justify the integration by parts formula

$$\int_r^s f(t)\, dB_t = f(s)\, B_s - f(r)\, B_r - \int_r^s B_t\, f'(t)\, dt.$$

(Hint: write the left-hand side as a limit of simple process approximations and do "summation by parts" on the approximations.) Explain why $\int_r^s f(t)\, dB_t$ has a normal distribution.

Exercise 3.9. Let $m = 1, \sigma = 2, \Delta t = 1/1000$ and simulate geometric Brownian motion

$$dX_t = X_t\, dt + 2\, X_t\, dB_t, \quad X_0 = 1,$$

using both (3.17) and (3.18). Be sure that the same sequence of normal random variables $N_1, N_2, \ldots$ is used for (3.17) and (3.16). Run the simulation at least 20 times and compare the values of $X_{1/4}, X_{1/2}, X_{3/4}, X_1$ obtained from the two formulas.

Exercise 3.10. Suppose that X_t satisfies the SDE

$$dX_t = X_t\, [(1/2)\, dt + dB_t], \quad X_0 = 2$$

Let

$$M = \max_{0 \leq t \leq 1} X_t.$$

1. Find the density of M.

2. Find the probability that $M \geq 4$.

Hint: Write $X_t = f(B_t)$ for an appropriate function.

Chapter 4

More stochastic calculus

4.1 Martingales and local martingales

In the last chapter we defined

$$Z_t = \int_0^t A_s \, dB_s,$$

where B_t is a standard Brownian motion and A_s is a continuous or piecewise continuous process. If

$$\int_0^t \mathbb{E}[A_s^2] \, ds < \infty$$

for each t, then Z_t is a square integrable martingale. However, our next example will show that if this inequality does not hold, the stochastic integral might not be a martingale. Our example is a modification of the martingale betting strategy from Section 1.2. In that example, we beat a fair game by doubling our bet until we won. We recall that if W_n denotes the winnings after n plays, then $\mathbb{E}[W_n] = 0$, but with probability one $W_\infty = 1$. In order to beat the game, we had to allow for arbitrarily large numbers of plays of the game until we won. In continuous time, one can do the same idea, but adapt it so that we guarantee to win by time 1.

Example 4.1.1. (Martingale betting strategy revisited) We will consider

$$Z_t = \int_0^t A_s \, dB_s$$

with a betting strategy A_s that changes only at times $t_0 < t_1 < t_2 < \cdots < 1$ where

$$t_n = 1 - 2^{-n}.$$

We start by setting $A_t = 1$ for $0 \le t < 1/2$. Then $Z_{1/2} = B_{1/2}$. Note that

$$\mathbb{P}\{Z_{1/2} \ge 1\} = \mathbb{P}\{B_{1/2} \ge 1\} = \mathbb{P}\{B_1 \ge \sqrt{2}\} = 1 - \Phi(\sqrt{2}) =: q > 0.$$

If $Z_{1/2} \ge 1$, we stop, that is, we set $A_t = 0$ for $1/2 \le t \le 1$. If $Z_{1/2} < 1$, let $x = 1 - Z_{1/2} > 0$. Define a by the formula

$$\mathbb{P}\{a[B_{3/4} - B_{1/2}] \ge x\} = q.$$

We set $A_t = a$ for $1/2 \le t < 3/4$. Note that we only need to know $Z_{1/2}$ to determine a and hence a is $\mathcal{F}_{1/2}$-measurable. Also, if $a[B_{3/4} - B_{1/2}] \ge x$, then

$$Z_{3/4} = \int_0^{3/4} A_s \, dB_s = Z_{1/2} + a[B_{3/4} - B_{1/2}] \ge 1.$$

Therefore,

$$\mathbb{P}\{Z_{3/4} \ge 1 \mid Z_{1/2} < 1\} = q,$$

and hence

$$\mathbb{P}\{Z_{3/4} < 1\} = (1 - q)^2.$$

If $Z_{3/4} \ge 1$, we set $A_t = 0$ for $3/4 \le t \le 1$. Otherwise, we proceed as above. At each time t_n we adjust the bet so that

$$\mathbb{P}\{Z_{t_{n+1}} \ge 1 \mid Z_{t_n} < 1\} = q,$$

and hence

$$\mathbb{P}\{Z_{t_n} < 1\} \le (1 - q)^n.$$

Using this strategy, with probability one $Z_1 \ge 1$, and hence, $\mathbb{E}[Z_1] \ge 1$. Therefore, Z_t is not a martingale. Our choice of strategy used discontinuous bets, but it is not difficult to adapt this example so that $t \mapsto A_t$ is a continuous function except at the one time at which the bet changes to zero.

Although the stochastic integral may not be a martingale, it is almost a martingale in the sense that one needs to make the bets arbitrarily large to find a way to change the expectation. The next definition makes precise the idea of a process that is a martingale "if it is stopped before the values get big".

Definition A continuous process M_t adapted to the filtration $\{\mathcal{F}_t\}$ is called a *local martingale* on $[0, T)$ if there exists an increasing sequence of stopping times

$$\tau_1 \le \tau_2 \le \tau_3 \le \cdots$$

such that with probability one

$$\lim_{j \to \infty} \tau_j = T,$$

and for each j,

$$M_t^{(j)} = M_{t \wedge \tau_j},$$

is a martingale.

In the case of the stochastic integral

$$Z_t = \int_0^t A_s \, dB_s,$$

we let $\{\tau_j\}$ be the stopping times,

$$\tau_j = \inf \left\{ t : \langle Z \rangle_t = \int_0^t A_s^2 \, ds = j \right\}.$$

Then for each j, $M_t^{(j)}$ is a square integrable martingale. Therefore, Z_t is a local martingale on $[0, T)$ where

$$T = \inf \left\{ t : \int_0^t A_s^2 \, ds = \infty \right\}.$$

The stochastic integral Z_t is not defined after time T. Note that if $s \mapsto A_s$ is continuous for all $s \in [0, \infty)$, then $T = \infty$.

More generally, suppose that

$$dX_t = R_t \, dt + A_t \, dB_t.$$

If $R_t \ne 0$, then X_t is not a martingale. In other words, if X_t is to be a martingale we need $R_t \equiv 0$. However, as we have just shown, it is possible for $R_t \equiv 0$ but X_t to not be a martingale. It is a local martingale in this case. In an abuse of terminology, it is standard to refer to $A_t \, dB_t$ as the "martingale part" of dX_t even though one should say the "local martingale part".

For continuous martingales, we have the optional sampling theorem which states under certain circumstances that one cannot beat a fair game. The next theorem includes two versions which are the most useful for applications.

Theorem 4.1.1 (Optional Sampling Theorem). *Suppose Z_t is a continuous martingale and T is a stopping time, both with respect to the filtration $\{\mathcal{F}_t\}$.*

- *If $M_t = Z_{t \wedge T}$, then M_t is a continuous martingale with respect to $\{\mathcal{F}_t\}$. In particular, $\mathbb{E}[Z_{t \wedge T}] = \mathbb{E}[Z_0]$.*

- *Suppose there exists $C < \infty$ such that for all t, $\mathbb{E}[Z_{t \wedge T}^2] \leq C$. Then if $\mathbb{P}\{T < \infty\} = 1$, $\mathbb{E}[Z_T] = \mathbb{E}[Z_0]$.*

Example 4.1.2. Suppose Z_t is a continuous martingale with $Z_0 = 0$. Suppose that $a, b > 0$ and let

$$T = \inf\{t : Z_t = -a \text{ or } Z_t = b\}.$$

Suppose that $\mathbb{P}\{T < \infty\} = 1$, which happens, for example, if Z_t is a standard Brownian motion. Then $Z_{t \wedge T}$ is a bounded martingale and hence

$$0 = \mathbb{E}[Z_0] = \mathbb{E}[Z_T] = -a \, \mathbb{P}\{Z_T = a\} + b \, \mathbb{P}\{Z_T = b\}.$$

By solving, we get

$$\mathbb{P}\{Z_T = b\} = \frac{a}{a+b},$$

which is the gambler's ruin estimate for continuous martingales.

Results about continuous martingales can be deduced from corresponding results about discrete-time martingales. If Z_t is a continuous martingale and

$$D_n = \left\{ \frac{k}{2^n} : k = 0, 1, \ldots \right\},$$

then $Z_t, t \in D_n$ is a discrete-time martingale. If T_n is a stopping time taking values in D_n, then $Z_{t \wedge T_n}$ is also a martingale.

For more general T, we define T_n by

$$T_n = \frac{k}{2^n} \quad \text{if} \quad \frac{k-1}{2^n} \leq T < \frac{k}{2^n}, \quad k = 1, 2, \ldots, n2^n,$$

$$T_n = n \quad \text{if} \quad T \geq n,$$

Suppose $s < t$ and let s_n, t_n denote the smallest element of D_n greater than or equal to s, t, respectively. If A is an $\mathcal{F}_s$-measurable event, then

it is also $\mathcal{F}_{s_n}$ measurable, and since $Z_{r \wedge T_n}, r \in D_n$ is a discrete-time martingale,

$$\mathbb{E}\left[1_A Z_{s_n \wedge T_n}\right] = \mathbb{E}\left[1_A Z_{t_n \wedge T_n}\right].$$

Since Z has continuous paths, we know that as $n \to \infty$,

$$Z_{s_n \wedge T_n} \to Z_{s \wedge T}, \quad Z_{t_n \wedge T_n} \to Z_{t \wedge T}.$$

In general, if X_n is a sequence of random variables converging to X, we cannot conclude that $\mathbb{E}[X_n] \to \mathbb{E}[X]$. However, this will hold if the sequence $\{X_n\}$ is *uniformly integrable*, that is, if for every $\epsilon > 0$, there exists $K < \infty$ such that for all n,

$$\mathbb{E}\left[|X_n|\, 1\{|X_n| \ge K\}\right] < \epsilon.$$

(We leave this as an exercise or the reader can check a book on measure-theoretic probability.)

Lemma 4.1.2. *For every positive integer m, the collection of random variables $\{Z_{T_n}\}$ where n ranges over all positive integers, T ranges over all stopping times with $T \le m$, and T_n is defined as above, is uniformly integrable.*

Proof. Let T be a stopping time with $T \le m$ and n a positive integer. Note that $N_t = |Z_t|, t \in D_n$ is a submartingale. Let $K > 0$ and let $\tau_n = \tau_{n,K}$ be the first $t \in D_n$ such that $N_t K$. If $t \in D_n$, let $A_t = A_{t,n}$ be the event $\{\tau_n = t\}$. Since N_t is a discrete-time submartingale, if $t \le m$,

$$\mathbb{E}\left[N_m\, 1_{A_t}\right] \ge \mathbb{E}\left[N_{T_n}\, 1_{A_t}\right].$$

By summing over $t \in D_n, t \le m$, we see that

$$\mathbb{E}\left[N_{T_n}\, J_{n,K}\right] \le \mathbb{E}\left[N_m\, J_{n,K}\right].$$

where $J_{n,K} = 1\{\tau_n \le m\}$. As $n \to \infty$. the random variables $J_{n,K}$ converge monotonically to

$$J_{\infty,K} = 1\left\{\max_{0 \le s \le t} N_s > K\right\}.$$

Therefore, for each $K < \infty$,

$$\mathbb{E}\left[N_{T_n}\, 1\{N_{T_n} > K\}\right] \le \mathbb{E}\left[N_{m \wedge T_n}\, J_{\infty,K}\right] \le \mathbb{E}\left[|Z_m|\, J_{\infty,K}\right].$$

Note that the right-hand side is independent of T and n. As $K \to \infty$, $J_{\infty,K} \to 0$, and since Z_m is integrable,

$$\lim_{K \to \infty} \mathbb{E}\left[|Z_m| \, J_{\infty,K}\right] = 0.$$

$\square$

Using the lemma, we now conclude

$$\mathbb{E}\left[1_A \, Z_{s \wedge T}\right] = \mathbb{E}\left[1_A \, Z_{t \wedge T}\right].$$

A sufficient condition for uniform integrability of $\{X_n\}$ is the existence of $C < \infty$ such that $\mathbb{E}[X_n^2] \leq C$ for all n. Indeed,

$$\mathbb{E}\left[|X_n| \, 1\{|X_n| \geq K\}\right] \leq K^{-1} \, \mathbb{E}\left[|X_n|^2 \, 1\{|X_n| \geq K\}\right] \leq C/K.$$

In particular, if $\mathbb{E}[Z_{t \wedge T}^2] \leq C$ for all t and $\mathbb{P}\{T < \infty\} = 1$, then $Z_{n \wedge T} \to Z_T$ with probability one and uniform integrability implies that

$$\mathbb{E}[Z_T] = \lim_{n \to \infty} \mathbb{E}[Z_{n \wedge T}] = \lim_{n \to \infty} \mathbb{E}[Z_0] = \mathbb{E}[Z_0].$$

Two other theorems that extend almost immediately to the continuous martingale setting are the following. The proofs are essentially the same as for their discrete counterparts.

Theorem 4.1.3. *(Martingale Convergence Theorem) Suppose Z_t is a continuous martingale and there exists $C < \infty$ such that $\mathbb{E}\left[|Z_t|\right] \leq C$ for all t. Then there exists a random variable Z_∞ such that with probability one*

$$\lim_{t \to \infty} Z_t = Z_\infty.$$

Theorem 4.1.4. *(Maximal inequality). Suppose Z_t is a continuous square integrable martingale, and let*

$$N_t = \max_{0 \leq s \leq t} |Z_t|.$$

Then for every $a > 0$,

$$\mathbb{P}\{N_t \geq a\} \leq \frac{\mathbb{E}[Z_t^2]}{a^2}.$$

4.2 An example: the Bessel process

The Bessel process with parameter a is the solution to the SDE

$$dX_t = \frac{a}{X_t} \, dt + dB_t, \quad X_0 = x_0 > 0.$$

If we choose $a = (d-1)/2$, this is called the Bessel-d process and is related to d-dimensional Brownian motion (see Exercise 4.4). Let $T_\epsilon = \inf\{t : X_t \leq \epsilon\}$. There is no problem finding a solution to this equation for $t < T_\epsilon$, and hence it is well defined for $t < T$ where

$$T = T_0 = \inf\{t : X_t = 0\}.$$

At time T the equation is ill-defined so we will not consider the process for $t > T$. If $a > 0$, then when X_t gets close to 0, there is a strong drift away from the origin. It is not obvious whether or not this is strong enough to keep the diffusion from reaching the origin.

Suppose that $0 < r < R < \infty$ and let

$$\tau = \tau(r, R) = \inf\{t : X_t = r \text{ or } X_t = R\}.$$

For $r \leq x \leq R$, let

$$\phi(x) = \mathbb{P}\{X_\tau = R \mid X_0 = x\}.$$

We will use Itô's formula to compute ϕ. Note that $\phi(r) = 0, \phi(R) = 1$. Let J denote the indicator function of the event $\{X_\tau = R\}$ and let

$$M_t = E[J \mid \mathcal{F}_t].$$

The tower property for conditional expectation implies that M_t is a martingale; indeed, if $s < t$,

$$E[M_t \mid \mathcal{F}_s] = E[\, E(J \mid \mathcal{F}_t) \mid \mathcal{F}_s] = E[J \mid \mathcal{F}_s] = M_s.$$

The Markovian nature of the diffusion X implies that

$$E[J \mid \mathcal{F}_t] = \phi(X_{t \wedge \tau}).$$

In other words, if $\tau \leq t$, then we already know whether or not $\{X_\tau = R\}$. However, if $\tau > t$, the only useful information in $\mathcal{F}_t$ for predicting if $X_\tau = R$ is X_t, and the conditional probability is the probability that this occurs given

that the process started at X_t. The upshot of this reasoning is that $\phi(X_{t \wedge \tau})$ must be a martingale. Itô's formula (see Theorem 3.4.3) gives for $t < T$,

$$
\begin{aligned}
d\phi(X_t) &= \phi'(X_t)\,dX_t + \frac{1}{2}\,\phi''(X_t)\,d\langle X\rangle_t \\
&= \left[\frac{a\,\phi'(X_t)}{X_t} + \frac{\phi''(X_t)}{2}\right]\,dt + \phi'(X_t)\,dB_t.
\end{aligned}
$$

If this is to be a martingale, the dt term must vanish at all times. The way to guarantee this is to choose ϕ to satisfy the ordinary differential equation (ODE)

$$
x\,\phi''(x) + 2a\,\phi'(x) = 0.
$$

Solving such equations is standard (and this one is particularly easy for one can solve the first-order equation for $g(x) = \phi'(x)$), and the solutions are of the form

$$
\phi(x) = c_1 + c_2\,x^{1-2a}, \quad a \neq \frac{1}{2},
$$

$$
\phi(x) = c_1 + c_2\,\log x, \quad a = \frac{1}{2},
$$

where c_1, c_2 are arbitrary constants. The boundary conditions $\phi(r) = 0, \phi(R) = 1$ determine the constants giving

$$
\phi(x) = \frac{x^{1-2a} - r^{1-2a}}{R^{1-2a} - r^{1-2a}}, \quad a \neq \frac{1}{2}, \tag{4.1}
$$

$$
\phi(x) = \frac{\log x - \log r}{\log R - \log r}, \quad a = \frac{1}{2}. \tag{4.2}
$$

We now answer the question that we posed.

Proposition 4.2.1. *If $a \geq 1/2$, then $\mathbb{P}\{T = \infty\} = 1$, that is, with probability one the Bessel process never reaches zero. If $a < 1/2$, then $\mathbb{P}\{T < \infty\} = 1$.*

Proof. Assume $X_0 = x < R$ and let $\tau(r, R)$ be defined as above. If $T < \infty$, then there must be $R < \infty$ such that $X_{\tau(r,R)} = r$ for all $r > 0$. Using the form of the probability in (4.1) and (4.2), we can see that

$$
\lim_{r \to 0} \mathbb{P}\{X_{\tau(r,R)} = r\} = \begin{cases} 0 & \text{if } a \geq 1/2 \\ 1 - (x/R)^{1-2a} & \text{if } a < 1/2 \end{cases}
$$

$\square$

The alert reader will note that we cheated a little bit in our derivation of ϕ because we assumed that ϕ was C^2. After assuming this, we obtained a differential equation and found what ϕ should be. To finish a proof, we can start with ϕ as defined in (4.1) or (4.2), and use Itô's formula to show that $M_t = \phi(X_{t\wedge\tau})$ is a continuous martingale. Since it is also bounded, we can use the optional sampling theorem to conclude that

$$\phi(x_0) = \mathbb{E}[M_0] = \mathbb{E}[M_\tau] = \mathbb{P}\{X_\tau = R\}.$$

This kind of argument is often done in stochastic calculus where one first assumes sufficient smoothness of a a function in order to use Itô's formula and obtain a differential equation for the function. Once the differential equation is solved, one can go back and redo the Itô's formula calculation rigorously.

4.3 Feynman-Kac formula

Suppose that the price of a stock follows a geometric Brownian motion

$$dX_t = m\, X_t\, dt + \sigma\, X_t\, dB_t. \tag{4.3}$$

Suppose that at some future time T we have an option to buy a share of the stock at price S. We will exercise the option only if $X_T \geq S$ and the value of the option at time T is $F(X_T)$ where

$$F(x) = (x - S)_+ = \max\{x - S, 0\}.$$

Suppose that there is an inflation rate of r so that x dollars at time t in future is worth only $e^{-rt}x$ in current dollars. Let $\phi(t, x)$ be the expected value of this option at time t, measured in dollars at time t, given that the current price of the stock is x,

$$\phi(t, x) = \mathbb{E}\left[e^{-r(T-t)}F(X_T) \mid X_t = x\right]. \tag{4.4}$$

The Feynman-Kac formula gives a PDE for this quantity.

Since it is not any harder, we generalize and assume that X_t satisfies the SDE

$$dX_t = m(t, X_t) \, dt + \sigma(t, X_t) \, dB_t, \quad X_0 = x_0,$$

and that there is a payoff $F(X_T)$ at some future time T. Suppose also that there is an inflation rate $r(t, x)$ so that if R_t denotes the value at time t of R_0 dollars at time 0,

$$dR_t = r(t, X_t) \, R_t \, dt,$$

$$R_t = R_0 \, \exp\left\{ \int_0^t r(s, X_s) \, ds \right\}.$$

If $\phi(t, x)$ denote the expected value of the payoff in time t dollars given $X_t = x$, then

$$\phi(t, x) = \mathbb{E}\left[\exp\left\{ -\int_t^T r(s, X_s) \, ds \right\} F(X_T) \mid X_t = x \right]. \qquad (4.5)$$

We will use Itô's formula to derive a PDE for ϕ under the assumption that ϕ is C^1 in t and C^2 in x.

Let

$$M_t = E\left[R_T^{-1} F(X_T) \mid \mathcal{F}_t \right].$$

The tower property for conditional expectation implies that if $s < t$,

$$E[M_t \mid \mathcal{F}_s] = E[E(M_T \mid \mathcal{F}_t) \mid \mathcal{F}_s] = E[M_T \mid \mathcal{F}_s] = M_s.$$

In other words, M_t is a martingale. Since R_t is $\mathcal{F}_t$-measurable, we see that

$$M_t = R_t^{-1} E\left[\exp\left\{ -\int_t^T r(s, X_s) \, ds \right\} F(X_T) \mid \mathcal{F}_t \right].$$

However, since X_t is a Markov process, the only relevant information in $\mathcal{F}_t$ is given by the value X_t. Hence, by the definition of ϕ,

$$M_t = R_t^{-1} \, \phi(t, X_t). \qquad (4.6)$$

Our conclusion is that M_t as given in (4.6) must be a martingale. Itô's formula gives

$$d\phi(t, X_t) = \partial_t \phi(t, X_t) \, dt + \partial_x \phi(t, X_t) \, dX_t + \frac{1}{2} \partial_{xx} \phi(t, X_t) \, d\langle X \rangle_t.$$

In particular,

$$d\phi(t, X_t) = H_t\, dt + A_t\, dB_t,$$

with

$$H_t = \partial_t\phi(t, X_t) + m(t, X_t)\, \partial_x\phi(t, X_t) + \frac{1}{2}\, \sigma(t, X_t)^2\, \partial_{xx}\phi(t, X_t),$$

$$A_t = \sigma(t, X_t)\, \partial_x\phi(t, X_t).$$

Since $\langle R \rangle_t = 0$, the stochastic product rule implies that

$$d[R_t^{-1}\phi(t, X_t)] = R_t^{-1}\, d\phi(t, X_t) + \phi(t, X_t)\, d[R_t^{-1}],$$

and hence the dt term of $d[R_t^{-1}\phi(t, X_t)]$ is R_t^{-1} times

$$-r(t, X_t)\, \phi(t, X_t) + \partial_t\phi(t, X_t) + m(t, X_t)\, \partial_x\phi(t, X_t)$$

$$+\frac{1}{2}\, \sigma(t, X_t)^2\, \partial_{xx}\phi(t, X_t).$$

Since M_t is a martingale, the dt term must be identically zero. This happens only if for all t, x,

$$-r(t, x)\, \phi(t, x) + \partial_t\phi(t, x) + m(t, x)\, \partial_x\phi(t, x)$$

$$+\frac{1}{2}\, \sigma(t, x)^2\, \partial_{xx}\phi(t, x) = 0.$$

We have derived the following.[1]

Theorem 4.3.1 (Feynman-Kac formula)**.** *Suppose X_t satisfies*

$$dX_t = m(t, X_t)\, dt + \sigma(t, X_t)\, dB_t, \quad X_0 = x_0,$$

and $r(t, x) \geq 0$ is a discounting rate. Suppose a payoff F at time T is given with $\mathbb{E}[|F(X_T)|] < \infty$. If $\phi(t, x), 0 \leq t \leq T$ is defined as in (4.5), and $\phi(t, x)$ is C^1 in t and C^2 in x, then $\phi(t, x)$ satisfies the PDE

$$\partial_t\phi(t, x) = -m(t, x)\, \partial_x\phi(t, x)$$

$$-\frac{1}{2}\, \sigma(t, x)^2\, \partial_{xx}\phi(t, x) + r(t, x)\, \phi(t, x),$$

for $0 \leq t < T$, with terminal condition $\phi(T, x) = F(x)$.

---

[1]One of our assumptions in Theorem 4.3.1 is that ϕ is sufficiently differentiable. One can give conditions on the coefficients and payoff function F under which this holds, and they include the examples that we give, but the general theory is beyond what we will discuss in this book. See I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd. ed, Springer (1991), Section 5.7.B for a discussion of this.

Example 4.3.1. Suppose X_t satisfies (4.3). Then $m(t,x) = mx, \sigma(t,x) = \sigma x$ and ϕ is as given in (4.4). The Feynman-Kac formula gives

$$\partial_t \phi = r\,\phi - mx\,\partial_x \phi - \frac{\sigma^2 x^2}{2}\,\partial_{xx}\phi. \qquad (4.7)$$

This is a version of the *Black-Scholes PDE*. We will return to this later.

Let us give a somewhat different derivation of the Feynman-Kac formula using the generator. Again, suppose that X_t satisfies

$$dX_t = m(t, X_t)\,dt + \sigma(t, X_t)\,dB_t,$$

and that F is a function that does not grow too quickly. Let

$$f(t, x) = \mathbb{E}\left[F(X_T) \mid X_t = x\right].$$

Let $r(t) \geq 0$ be a discount rate which for ease we will assume is a deterministic function of time, and

$$R_t = R_0 \, \exp\left\{\int_0^t r(s)\,ds\right\}.$$

Let

$$\phi(t, x) = \mathbb{E}\left[(R_t/R_T)\,F(X_T) \mid X_t = x\right] = \exp\left\{-\int_t^T r(s)\,ds\right\}\,f(t, x)$$

Recall that

$$\partial_t f(t, x) = -L_x^t f(t, x),$$

where L^t is the generator

$$L^t h(x) = m(t, x)\,h'(x) + \frac{\sigma(t, x)}{2}\,h''(x).$$

Therefore,

$$
\begin{aligned}
\partial_t \phi(t, x) &= r(t)\,\exp\left\{-\int_t^T r(s)\,ds\right\}\,f(t, x) \\
&\quad - \exp\left\{-\int_t^T r(s)\,ds\right\}\,L_x^t f(t, x) \\
&= r(t)\,\phi(t, x) - L_x^t \phi(t, x) \\
&= r(t)\,\phi(t, x) - m(t, x)\,\partial_x \phi(t, x) - \frac{\sigma^2(t, x)}{2}\,\partial_{xx}\phi(t.x),
\end{aligned}
$$

which is the Feynman-Kac PDE.

4.4 Binomial approximations

In the next few sections we will be considering the SDE

$$dX_t = m(t, X_t)\, dt + \sigma(t, X_t)\, dB_t.$$

We will give some heuristic derivations based on discrete approximations. Recall that the Euler method to sample from this equation uses the rule

$$X(t + \Delta t) = X(t) + m(t, X(t))\, \Delta t + \sigma(t, X(t))\, \sqrt{\Delta t}\, Z$$

where $Z \sim N(0, 1)$. Here we wish to consider some *binomial* sampling schemes, that is, schemes such that given $X(t)$, $X(t + \Delta t)$ takes one of two values.

If X_t is Brownian motion with drift zero and variance σ^2,

$$dX_t = \sigma\, dB_t,$$

then the binomial scheme is approximation by random walk,

$$\mathbb{P}\{X(t + \Delta t) - X(t) = \sigma\, \sqrt{\Delta t} \mid X(t)\}$$

$$= \mathbb{P}\{X(t + \Delta t) - X(t) = -\sigma\, \sqrt{\Delta t} \mid X(t)\} = \frac{1}{2}.$$

In this case the values of $X(k\Delta t)$ lie in the lattice of points

$$\left\{ \ldots, -2\sigma\sqrt{\Delta t}, -\sigma\sqrt{\Delta t}, 0, \sigma\sqrt{\Delta t}, 2\sigma\sqrt{\Delta t}, \ldots \right\}.$$

If $m \equiv 0$ and $\sigma(t, x)$ is not constant, we can change this rule to

$$\mathbb{P}\{X(t + \Delta t) - X(t) = \pm\sigma(t, X(t))\, \sqrt{\Delta t} \mid X(t)\} = \frac{1}{2},$$

but then the values of $X(t)$ do not stay in a lattice.

Suppose σ is constant, but m varies,

$$dX_t = m(t, X_t)\, dt + \sigma\, dB_t.$$

There are two reasonable ways to incorporate drift in this binomial model. One is to use a version of the Euler's method and set

$$\mathbb{P}\{X(t + \Delta t) - X(t) = m(t, X_t)\, \Delta t \pm \sigma\, \sqrt{\Delta t} \mid X(t)\} = \frac{1}{2},$$

for which it is immediate that

$$E\left[X(t+\Delta t)-X(t)\mid \mathcal{F}_t\right]$$
$$= E\left[X(t+\Delta t)-X(t)\mid X(t)\right] = m(t,X_t)\,\Delta t. \qquad (4.8)$$

This method has the disadvantage that we do not stay on a lattice of points. Another approach is to always make the jumps $\pm\sigma\sqrt{\Delta t}$ but to change the probabilities based on the drift. In particular, we let

$$\mathbb{P}\{X(t+\Delta t)-X(t)=\sigma\sqrt{\Delta t}\mid X(t)\} = \frac{1}{2}\left[1+\frac{m(t,X_t)}{\sigma}\sqrt{\Delta t}\right],$$

$$\mathbb{P}\{X(t+\Delta t)-X(t)=-\sigma\sqrt{\Delta t}\mid X(t)\} = \frac{1}{2}\left[1-\frac{m(t,X_t)}{\sigma}\sqrt{\Delta t}\right].$$

The probabilities are chosen so that (4.8) holds for this scheme as well. These two binomial methods illustrate the two main ways to incorporate a drift term (in other words, to turn a fair game into an unfair game):

- Choose an increment with mean zero and then add a small amount to it.

- Change the probabilities of the increment so that it does not have mean zero.

Example 4.4.1. We will use the latter rule to sample from Brownian motion with constant drift $m\neq 0$ and $\sigma^2 = 1$. Our procedure for simulating uses

$$\mathbb{P}\{X(t+\Delta t)-X(t)=\pm\sqrt{\Delta t}\mid X(t)\} = \frac{1}{2}\left[1\pm m\sqrt{\Delta t}\right]. \qquad (4.9)$$

Suppose that $\Delta t = 1/N$ for some large integer N, and let us consider the distribution of $X(1)$. There are 2^N possible paths that the discrete process can take which can be represented by

$$\omega = (a_1, a_2, \ldots, a_N)$$

where $a_j = +1$ or -1 depending on whether that step goes up or down. Let $J = J(\omega)$ denote the number of $+1$'s, and define r by $J = (N/2) + r\sqrt{N}$. Note that the position at time 1 is

$$\begin{aligned}
X(1) &= \sqrt{\Delta t}\,[a_1 + \cdots + a_N] \\
&= J\sqrt{\Delta t} - (N-J)\sqrt{\Delta t} \\
&= 2r\sqrt{N}\sqrt{\Delta t} = 2r.
\end{aligned}$$

For each ω, the probability of its occurring is

$$q(\omega) = \left(\frac{1}{2}\right)^N \left[1 + m\sqrt{\Delta t}\right]^J \left[1 - m\sqrt{\Delta t}\right]^{N-J}.$$

If m were 0, then the probability would be be just $(1/2)^N$. Therefore, the *ratio* of the probabilities for $m \neq 0$ to $m = 0$ is

$$\left[1 + \frac{m}{\sqrt{N}}\right]^J \left[1 - \frac{m}{\sqrt{N}}\right]^{N-J}$$

$$= \left[1 - \frac{m^2}{N}\right]^{N/2} \left[1 + \frac{m}{\sqrt{N}}\right]^{r\sqrt{N}} \left[1 - \frac{m}{\sqrt{N}}\right]^{-r\sqrt{N}}.$$

Using the relation $\left(1 + \frac{a}{N}\right)^N \sim e^a$, we see that the limit of the right-hand side as $N \to \infty$ is

$$e^{-m^2/2} e^{2rm} = e^{mX(1)} e^{-m^2/2}.$$

Given this, we see that in order to sample from a Brownian motion with drift, we could first sample from a Brownian motion without drift and then weight the samples by the factor $e^{mX(1)} e^{-m^2/2}$. We will show how to do this directly in Section 5.2.

As one more application of binomial approximations, we will give a heuristic argument for the following theorem.

Theorem 4.4.1. *Suppose that*

$$dX_t = m(X_t)\,dt + \sigma\,dB_t,$$

where m is continuously differentiable, and let $p(t, x)$ denote the density of X_t at time t. Then

$$\partial_t p(t, x) = L_x^* p(t, x),$$

where

$$
\begin{aligned}
L^* f(x) &= -[m(x)\,f(x)]' + \frac{\sigma^2}{2}\,f''(x) \\
&= -m'(x)\,f(x) - m(x)\,f'(x) + \frac{\sigma^2}{2}\,f''(x).
\end{aligned}
$$

—

Recall that

$$Lf(x) = m(x) f'(x) + \frac{\sigma^2}{2} f''(x).$$

If m is constant, then as we saw before, one obtains L^* from L by just changing the sign of the drift. For varying m we get another term. We will derive the expression for L^* by using the binomial approximation

$$\mathbb{P}\{X(t + \Delta t) - X(t) = \pm\sigma \sqrt{\Delta t} \mid X(t)\} = \frac{1}{2} \left[1 \pm \frac{m(X_t)}{\sigma} \sqrt{\Delta t}\right].$$

Let $\epsilon = \sqrt{\Delta t}, \epsilon^2 = \Delta t$. To be at position $x = k\epsilon$ at time $t + \epsilon^2$, one must be at position $x \pm \sigma\epsilon$ at time t. This gives the relation

$$p(t + \epsilon^2, x) = p(t, x - \sigma\epsilon) \frac{1}{2} \left[1 + \frac{m(x - \sigma\epsilon)\,\epsilon}{\sigma}\right]$$

$$+ p(t, x + \sigma\epsilon) \frac{1}{2} \left[1 - \frac{m(x + \sigma\epsilon)}{\sigma}\,\epsilon\right]. \tag{4.10}$$

We know that

$$p(t + \epsilon^2, x) = p(t, x) + \partial_t p(t, x)\,\epsilon^2 + o(\epsilon^2). \tag{4.11}$$

The definition of the derivative (see Section 2.10.1), implies that

$$\frac{p(t, x + \sigma\epsilon) + p(t, x - \sigma\epsilon)}{2} = p(t, x) + \frac{\sigma^2 \epsilon^2}{2}\partial_{xx}p(t, x) + o(\epsilon^2).$$

$$p(t, x \pm \sigma\epsilon) = p(t, x) \pm \partial_x p(t, x)\,\sigma\epsilon + o(\epsilon),$$

$$m(x \pm \sigma\epsilon) = m(x) \pm m'(x)\,\sigma\epsilon + o(\epsilon),$$

Plugging in, we see that the right-hand side of (4.10) equals

$$p(t, x) + \epsilon^2 \left[\frac{\sigma^2}{2}\partial_{xx}p(t, x) - m'(x)\,p(t, x) - m(x)\,\partial_x p(t, x)\right] + o(\epsilon^2).$$

Comparing this with (4.11) gives us the result.

One can also derive this result by considering L^* as the adjoint of L. In other words, L^* is defined by

$$\int_{-\infty}^{\infty} L^* f(x)\, g(x)\, dx = \int_{-\infty}^{\infty} f(x)\, Lg(x)\, dx.$$

If

$$Lg(x) = m(x)\, g'(x) + \frac{\sigma^2}{2} g''(x),$$

then integration by parts gives

$$L^* f(x) = -[m(x)\, f(x)]' + \frac{\sigma^2}{2} f''(x).$$

4.5 Continuous martingales

Earlier we made the comment that Brownian motion is the only continuous martingale. We will make this explicit in this section.

Proposition 4.5.1. *Suppose that M_t is a continuous martingale with respect to a filtration $\{\mathcal{F}_t\}$ with $M_0 = 0$, and suppose that the quadratic variation of M_t is the same as that of standard Brownian motion,*

$$\langle M \rangle_t = \lim_{n \to \infty} \sum_{j < 2^n t} \left[M\left(\frac{j+1}{2^n} \right) - M\left(\frac{j}{2^n} \right) \right]^2 = t.$$

Then every $\lambda \in \mathbb{R}$,
$$\mathbb{E}\left[\exp\{i\lambda M_t\} \right] = e^{-\lambda^2 t/2}.$$

Sketch of proof. Fix λ and let $f(x) = e^{i\lambda x}$. Note that the derivatives of f are uniformly bounded in x. Following the proof of Itô's formula we can show that

$$f(M_t) - f(M_0) = N_t + \frac{1}{2} \int_0^t f''(M_s)\, ds = N_t - \frac{\lambda^2}{2} \int_0^t f(M_s)\, ds,$$

where N_t is a martingale. In particular, if $r < t$,

$$\mathbb{E}[f(M_t) - f(M_r)] = \frac{1}{2} \mathbb{E}\left[\int_r^t f''(M_s)\, ds \right] = -\frac{\lambda^2}{2} \int_r^t \mathbb{E}[f(M_s)]\, ds.$$

If we let $G(t) = \mathbb{E}[f(M_t)]$, we get the equation

$$G'(t) = -\frac{\lambda^2}{2} G(t), \quad G(0) = 1,$$

which has solution $G(t) = e^{-\lambda^2 t/2}$. □

Theorem 4.5.2. *Suppose that M_t is a continuous martingale with respect to a filtration $\{\mathcal{F}_t\}$ with $M_0 = 0$, and suppose that the quadratic variation of M_t is the same as that of standard Brownian motion*

$$\langle M \rangle_t = \lim_{n \to \infty} \sum_{j < 2^n t} \left[M \left(\frac{j+1}{2^n} \right) - M \left(\frac{j}{2^n} \right) \right]^2 = t.$$

Then M_t is a standard Brownian motion.

Proof. We need to show that M_t satisfies the conditions of a standard Brownian motion. We are given $M_0 = 0$ and that $t \mapsto M_t$ is continuous, so we only need to establish independent, normal increments. One way to express both of these conditions is to say that for every $s < t$, the conditional distribution of $M_t - M_s$ given $\mathcal{F}_s$ is $N(0, t - s)$. But the proposition above shows that for every λ,

$$E\left(e^{i\lambda(M_t - M_s)} \mid \mathcal{F}_s\right) = e^{-\lambda^2(t-s)/2}.$$

Since the characteristic function determines the distribution, we are finished. □

This theorem can be extended to say that any continuous martingale is a time change of a Brownian motion. To be precise, suppose that M_t is a continuous martingale. The quadratic variation can be defined as the process $\langle M \rangle_t$ such that $M_t^2 - \langle M \rangle_t$ is a local martingale. Let

$$\tau_s = \inf\{t : \langle M \rangle_t = s\}.$$

Then $Y_s = M_{\tau_s}$ is a continuous martingale whose quadratic variation is the same as standard Brownian motion. Hence Y_s is a standard Brownian motion.

4.6 Exercises

Exercise 4.1. A process X_t satisfies the Ornstein-Uhlenbeck SDE if

$$dX_t = mX_t\, dt + dB_t,$$

where $m \in \mathbb{R}$ and B_t is a standard Brownian motion. Suppose that $X_0 = 1$, $R > 1$ and $T = T_R$ is the first time that $X_t = R$ or $X_t = 0$.

1. Find a function F with $F(0) = 0$ and $F(x) > 0$ for $x > 0$ such that $F(X_{t \wedge T})$ is a martingale. (You may leave your answer in terms of a definite integral.)

2. Find the probability that $X_T = R$. You can write the answer in terms of the function F.

3. For what values of m is it true that

$$\lim_{R \to \infty} \mathbb{P}\{X_T = R\} = 0 \ ?$$

Exercise 4.2. Suppose B_t is a standard Brownian motion and X_t satisfies the SDE

$$dX_t = a \, \cot(X_t)\, dt + dB_t, \quad X_0 = x_0 \in (0, \pi/2),$$

where $a \in \mathbb{R}$. For $\epsilon > 0$, let

$$T_\epsilon = \min\{t : \sin(X_t) = \sin \epsilon\},$$

$$T = T_{0+} = \min\left\{t : X_t \in \{0, \pi\}\right\} = \min\{t : \sin(X_t) = 0\},$$

We will only consider the process X_t for $t \leq T$. Let

$$\tau = \min\{t : X_t = \pi/2\}.$$

1. Find a function $F(x)$ that is positive for $0 < x < \pi/2$ with $F(\pi/2) = 0$ and such that $M_t = F(X_{t \wedge T})$ is a local martingale for $t < T$. (It suffices to write F in the form

$$F(x) = \int_x^{\pi/2} g(y)\, dy,$$

where g is an explicit function.)

2. For which values of a is

$$F(0+) = \lim_{\epsilon \downarrow 0} F(\epsilon) < \infty \ ?$$

3. Assume that $0 < \epsilon < x_0 < \pi/2$ and let

$$q(x_0, \epsilon) = \mathbb{P}\{T_\epsilon < \tau\}.$$

Write $q(x_0, \epsilon)$ explicitly in terms of F.

4. For whch values of a is it the case that

$$\lim_{\epsilon \downarrow 0} q(x_0, \epsilon) = 0 \ ?$$

Exercise 4.3. Suppose B_t is a standard Brownian motion and let $X_t = e^{B_t} + 2$. Let

$$\phi(t, x) = \mathbb{E}\left[e^{-2(T-t)} \left(X_T^3 - 1 \right) e^{-X_T} \mid X_t = x \right], \quad 0 < t < T.$$

Use the Feynman-Kac theorem to find a second-order PDE that ϕ satisfies.

Exercise 4.4. Suppose $B_1^1, \ldots, B_t^d$ are independent standard Brownian motions and let

$$X_t = \sqrt{(B_t^1)^2 + \cdots + (B_t^d)^2}.$$

1. Use Itô's formula to show that

$$dX_t = \frac{a}{X_t}\, dt + dM_t,$$

where $a = (d-1)/2$ and M_t is a continuous martingale satisfying

$$dM_t = \sum_{j=1}^{d} A_t^j\, dB_t^j,$$

for suitable processes A_t^j.

2. Show that $\langle M \rangle_t = t$.

3. Explain why M_t is a standard Brownian motion. (Hint: use Proposition 4.5.1.)

Chapter 5

Change of measure and Girsanov theorem

5.1 Absolutely continuous measures

In order to state the Girsanov theorem in the next section, we will need to discuss absolute continuity and singularity of measures. For measures on discrete (countable) spaces, this is an elementary idea, but it is more subtle for uncountable spaces. We assume that we have a probability space $(\Omega, \mathcal{F})$ where Ω is the space of outcomes (sometimes called the *sample space*) and $\mathcal{F}$ is the collection of events. An event is a subset of Ω, and we assume that the set of events forms a σ-algebra (see Section 1.1). A *(positive) measure* is a function $\mu : \mathcal{F} \to [0, \infty]$ such that $\mu(\emptyset) = 0$ and if $A_1, A_2, \ldots$ are disjoint,

$$\mu \left(\bigcup_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} \mu(A_n).$$

The measure μ is a *probability (measure)* if $\mu(\Omega) = 1$.

Definition Suppose μ, ν are measures on $(\Omega, \mathcal{F})$.

- ν is *absolutely continuous* with respect to μ, written $\nu \ll \mu$, if for every $E \in \mathcal{F}$, if $\mu(E) = 0$, then $\nu(E) = 0$.

- μ and ν are *mutually absolutely continuous* or *equivalent* measures if $\nu \ll \mu$ and $\mu \ll \nu$.

- μ and ν are *singular* measures, written $\mu \perp \nu$, if there exists an event $E \in \mathcal{F}$ with $\mu(E) = 0$ and $\nu(\Omega \setminus E) = 0$.

Example 5.1.1. Suppose Ω is a countable set and $\mathcal{F}$ is the collection of all subsets of Ω. If $p : \Omega \to [0, \infty)$ is a function, then there is a corresponding measure μ defined by

$$\mu(E) = \sum_{\omega \in E} p(\omega).$$

Suppose ν is another measure given by the function q. Let

$$A_\mu = \{\omega : p(\omega) > 0\}, \quad A_\nu = \{\omega : q(\omega) > 0\}.$$

Then $\nu \ll \mu$ if and only if $A_\nu \subset A_\mu$, and

$$q(\omega) = \frac{d\nu}{d\mu}(\omega)\, p(\omega), \quad \omega \in \Omega,$$

where $d\nu/d\mu$ is defined on A_μ by

$$\frac{d\nu}{d\mu}(\omega) = \frac{q(\omega)}{p(\omega)}.$$

Note that for any event E,

$$\nu(E) = \sum_{\omega \in E} \frac{d\nu}{d\mu}(\omega)\, p(\omega).$$

In this case, ν and μ are equivalent if $A_\nu = A_\mu$, and $\nu \perp \mu$ if $A_\nu \cap A_\mu = \emptyset$.

Example 5.1.2. Suppose μ denotes Lebesgue measure on $\mathbb{R}$,

$$\mu(A) = \int_A d\mu = \int_A dx = \text{length}(A),$$

and X is a continuous random variable with density f. Let P_X denote the distribution of X, that is, the probability measure on $\mathbb{R}$ given by

$$P_X(A) = \mathbb{P}\{X \in A\} = \int_A f(x)\, dx = \int_A f\, d\mu.$$

Then $P_X \ll \mu$ and we can write

$$P_X(A) = \int_A \frac{dP_X}{d\mu}\, d\mu \quad \text{where} \quad \frac{dP_X}{d\mu} = f.$$

If $f(x) > 0$ for all x, then $P_X(A) > 0$ whenever $\mu(A) > 0$ and hence $\mu \ll P_X$. If Y is another continuous random variable with density g, let

$$A_X = \{x : f(x) > 0\}, \quad A_Y = \{y : g(y) > 0\}.$$

If $A_Y \subset A_X$, then $P_Y \ll P_X$ and we can write

$$\mathbb{P}\{Y \in A\} = \int_A g \, d\mu = \int_A \frac{g}{f} f \, d\mu = \int_A \frac{dP_Y}{dP_X} \, dP_X,$$

where

$$\frac{dP_Y}{dP_X} = \frac{g}{f}.$$

If $A_X \cap A_Y = \emptyset$, then $P_X \perp P_Y$.

Example 5.1.3. Suppose X is a discrete random variable taking values in the countable set A and Y is a continuous random variable with density g. If P_X, P_Y denote the distributions, then $P_X \perp P_Y$. Indeed,

$$P_X(\mathbb{R} \setminus A) = 0, \quad P_Y(A) = 0.$$

These examples demonstrate the following theorem whose proof can be found in any book on measure theory. A measure μ is σ-*finite* if we can write

$$\Omega = \bigcup_{n=1}^{\infty} A_n,$$

where $\mu(A_n) < \infty$ for each n.

Theorem 5.1.1 (Radon-Nikodym Theorem). *Suppose μ, ν are σ-finite measures on $(\Omega, \mathcal{F})$ with $\nu \ll \mu$. Then there exists a function f such that for every E,*

$$\nu(E) = \int_E f \, d\mu. \tag{5.1}$$

The function f is called the *Radon-Nikodym derivative* of ν with respect to μ and is denoted

$$f = \frac{d\nu}{d\mu}.$$

Roughly speaking, the ν measure of a point x is $(d\nu/d\mu)(x)$ times the μ measure of x. This interpretation is precise in the case of discrete measures,

such as in Example 5.1.1, but does not make precise sense when the point x gets zero measure in both measures. Example 5.1.2 shows that this subtlety is something we are already familiar with. If μ denotes length, then the density $f = dP_X/d\mu$ of a continuous random variable can be interpreted as saying that the probability of obtaining a value in $[x, x + dx]$ is $f(x)$ times dx, the length of $[x, x + dx]$.

If $(\Omega, \mathcal{F}, P)$ is a probability space, and Q is a probability measure with $Q \ll P$, then the Radon-Nikodym derivative

$$X = \frac{dQ}{dP}$$

is a nonnegative random variable with $\mathbb{E}[X] = 1$. If E is an event, then (5.1) can be written as

$$Q(E) = \mathbb{E}_P [X \, 1_E] . \tag{5.2}$$

Here $\mathbb{E}_P$ denotes expectation *using the probability measure P*. (Up to now, we have been considering a single probability measure $\mathbb{P}$ on a space and using $\mathbb{E}$ to denote expectation with respect to it. Since we are now going to consider different measures on the same space, we adopt the notation $\mathbb{E}_P$ to denote expectations using the measure P.) It is not hard to extend the relation (5.2) to give

$$\mathbb{E}_Q[Y] = \mathbb{E}_P \left[Y \, \frac{dQ}{dP} \right] .$$

Example 5.1.4. Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, and $\mathcal{G} \subset \mathcal{F}$ is a sub σ-algebra. As before, we think of $\mathcal{G}$ as "partial information". Let X be a nonnegative, integrable $\mathcal{F}$-measurable random variable. Then

$$Q(A) = \mathbb{E} [1_A X] , \quad A \in \mathcal{G},$$

defines a measure on $(\Omega, \mathcal{G})$ that satisfies $Q \ll \mathbb{P}$. Therefore, there exists a *$\mathcal{G}$-measurable* random variable Y such that for all $A \in \mathcal{G}$,

$$Q(A) = \mathbb{E} [1_A Y] .$$

This random variable Y is the conditional expectation $E(X \mid \mathcal{G})$ as defined in Section 1.1.

Example 5.1.5. Let the set of outcomes Ω be the set of continuous functions $f : [0, 1] \to \mathbb{R}$. If B_t denotes a Brownian motion with drift zero and variance

σ^2, there is a measure P_σ on Ω called *Wiener measure (with variance σ^2)*. We think of P_σ as the distribution of the "function-valued" random variable $t \mapsto B_t$. If V is a subset of Ω, then $P_\sigma(V)$ is the probability that the function $t \mapsto B_t$ lies in V. We claim that if $\sigma \neq \sigma'$, then $P_\sigma \perp P_{\sigma'}$. In order to show this, we need to find an event E such that $P_\sigma(E) = 1, P_{\sigma'}(E) = 0$. Let E_r denote the set of functions f such that

$$\lim_{n \to \infty} \sum_{j=1}^{2^n} \left[f\left(\frac{j}{2^n}\right) - f\left(\frac{j-1}{2^n}\right) \right]^2 = r^2.$$

Using what we know about the quadratic variation, we see that

$$P_\sigma(E_\sigma) = 1, \quad P_{\sigma'}(E_{\sigma'}) = 1.$$

Let Ω denote the set of continuous functions $f : [0,1] \to \mathbb{C}$ which is a Banach space (complete, normed, metric space) under the norm

$$\|f\| = \max\{|f(s)| : 0 \leq s \leq 1\}.$$

Let $\mathcal{F}$ denote the corresponding Borel σ-algebra, that is, the smallest σ-algebra under which all the open sets under $\| \cdot \|$ are measurable. The measures P_σ are defined on $(\Omega, \mathcal{F})$. It is easy to check that the functions $\Theta_s(f) = f(s)$ are measurable functions on this space, and hence so are the functions

$$\Psi_n(f) = \sum_{j=1}^{2^n} \left[f\left(\frac{j}{2^n}\right) - f\left(\frac{j-1}{2^n}\right) \right]^2,$$

and the sets

$$E_r = \left\{ f : \lim_{n \to \infty} \Psi_n(f) = r^2 \right\}.$$

We proved that $P_\sigma(E_\sigma) = 1$ in Theorem 2.8.2.

One rarely needs to deal with measures that are not σ-finite, but the Radon-Nikodym derivative does not hold for such measures. For example, if μ is counting measure and ν is Lebesgue measure, both on $\mathbb{R}$, then $\nu \ll \mu$, but there is no f such that (5.1) holds.

5.2 Giving drift to a Brownian motion

As we have already noted, there are two ways to take a fair game and make it unfair or vice versa:

- Play the game and then add a deterministic amount in one direction.

- Change the probabilities of the outcome.

We will consider these methods in the case of a standard Brownian motion B_t. For the first method, we can define a Brownian motion with drift m, by setting

$$W_t = mt + B_t.$$

We will now consider the second way by changing the probabilities. We have already motivated this with a binomial approximation in (4.9), but in this section we will transform the game directly using the Brownian motion.

Suppose that B_t is defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $\{\mathcal{F}_t\}$. To change the probability is to consider a different measure Q instead of $\mathbb{P}$. Let

$$M_t = e^{mB_t - \frac{m^2 t}{2}}. \tag{5.3}$$

We have seen that M_t is a martingale. In fact (see (3.20)), Itô's formula shows that M_t satisfies

$$dM_t = m\, M_t\, dB_t, \quad M_0 = 1.$$

For each event V that is $\mathcal{F}_t$ measurable, we define

$$Q_t(V) = \mathbb{E}\left[1_V\, M_t\right].$$

In other words, on the probability space $(\Omega, \mathcal{F}_t, \mathbb{P})$,

$$\frac{dQ_t}{d\mathbb{P}} = M_t.$$

If $s < t$ and V is $\mathcal{F}_s$-measurable, then it is also $\mathcal{F}_t$-measurable. In this case, $Q_s(V) = Q_t(V)$ as can be seen in the calculation

$$Q_t(V) = \mathbb{E}\left[1_V\, M_t\right] = \mathbb{E}\left[E(1_V\, M_t \mid \mathcal{F}_s)\right]$$

$$= \mathbb{E}\left[1_V\, E(M_t \mid \mathcal{F}_s)\right] = \mathbb{E}\left[1_V\, M_s\right] = Q_s(V). \tag{5.4}$$

Hence we can write just Q for the measure.

We claim that

- The process $t \mapsto B_t$, *under the measure Q,* is a Brownian motion with drift m and $\sigma^2 = 1$.

We prove this by showing it satisfies the conditions to be a Brownian motion. The continuity of paths is immediate as well as $B_0 = 0$. To show the rest, it suffices to show that if $s, t \geq 0$, then $B_{t+s} - B_s$ is independent of $\mathcal{F}_s$ with a normal distribution with mean mt and variance t. We can establish this by showing it has the (conditional) moment generating function

$$E_Q \left(\exp\{\lambda(B_{t+s} - B_s)\} \mid \mathcal{F}_s \right) = e^{\lambda m t} e^{\lambda^2 t/2}. \tag{5.5}$$

Here we are writing E_Q to denote that the conditional expectation is taken using the measure Q. To establish (5.5), we need to show that if V is $\mathcal{F}_s$-measurable, then

$$
\begin{aligned}
\mathbb{E}_Q \left[1_V \, \exp\{\lambda(B_{t+s} - B_s)\} \right] &= \mathbb{E}_Q \left[1_V \, e^{\lambda m t} e^{\lambda^2 t/2} \right] \\
&= e^{\lambda m t} e^{\lambda^2 t/2} Q(V),
\end{aligned}
$$

or equivalently, by the definition of $\mathbb{E}_Q$,

$$\mathbb{E} \left[1_V \, \exp\{\lambda(B_{t+s} - B_s)\} M_{t+s} \right] = e^{\lambda m t} e^{\lambda^2 t/2} \mathbb{E}[1_V M_s].$$

Note that if $Y = B_{t+s} - B_s$, then Y is independent of $\mathcal{F}_s$ and

$$
\begin{aligned}
E \left(e^{\lambda Y} M_{t+s} \mid \mathcal{F}_s \right) &= M_s \, e^{-m^2 t/2} E \left(e^{\lambda Y} e^{mY} \mid \mathcal{F}_s \right) \\
&= M_s \, e^{-m^2 t/2} \mathbb{E} \left[e^{(\lambda + m)Y} \right] \\
&= M_s \, e^{-m^2 t/2} e^{(\lambda + m)^2 t/2} \\
&= M_s \, e^{\lambda^2 t/2} e^{\lambda m t}.
\end{aligned}
$$

Therefore, if V is $\mathcal{F}_s$-measurable,

$$
\begin{aligned}
\mathbb{E} \left[1_V \, \exp\{\lambda(B_{t+s} - B_s)\} M_{t+s} \right] & \\
&= \mathbb{E} \left[E(1_V \, \exp\{\lambda(B_{t+s} - B_s)\} M_{t+s} \mid \mathcal{F}_s) \right] \\
&= \mathbb{E} \left[1_V \, E(\exp\{\lambda(B_{t+s} - B_s)\} M_{t+s} \mid \mathcal{F}_s) \right] \\
&= e^{\lambda^2 t/2} e^{\lambda m t} \mathbb{E}[1_V M_s].
\end{aligned}
$$

Example 5.2.1. Suppose X_t is a geometric Brownian motion satisfying

$$dX_t = X_t \left[m \, dt + \sigma \, dB_t \right],$$

where B_t is a standard Brownian motion defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If $r \in \mathbb{R}$, then we can find a new probability measure Q such that

$$dB_t = r\,dt + dW_t,$$

where W_t is a Brownian motion with respect to Q. Then,

$$dX_t = X_t \left[(m + \sigma r)\,dt + \sigma\,dW_t\right].$$

Hence with respect to Q, X_t is a geometric Brownian motion with the same volatility but a new drift. From this we can see that the measures for geometric Brownian motions *with the same σ* are equivalent.

Example 5.2.2. We compute a quantity for Brownian motion with drift by using a standard Brownian motion. Suppose B_t is a standard Brownian motion, $a > 0$, $m \in \mathbb{R}$, and M_t is the martingale in (5.3). Let $T_a = \inf\{t : B_t = a\}$. Then under the new measure Q, B_t is a Brownian motion with drift m. Note that

$$Q\{T_a < \infty\} = \mathbb{E}\left[M_{T_a} \mathbf{1}\{T_a < \infty\}\right] = \mathbb{E}\left[M_{T_a}\right].$$

Here the expectation $\mathbb{E}$ is with respect to the original measure under which B is a standard Brownian motion; the second equality follows from the fact that $\mathbb{P}\{T_a < \infty\} = 1$. Therefore,

$$
\begin{aligned}
Q\{T_a < \infty\} &= \mathbb{E}\left[\exp\left\{mB_{T_a} - \frac{m^2 T_a}{2}\right\}\right] \\
&= e^{am}\,\mathbb{E}\left[\exp\left\{-\frac{m^2 T_a}{2}\right\}\right].
\end{aligned}
$$

In Example 2.7.1 we computed the density of T_a for Brownian motion. Given this we could compute the expectation on the right-hand side. However we will choose a different method. We know that driftless Brownian motion will hit the line $y = a$ with probability one, and hence so will Brownian motion with $m > 0$. Therefore, if $m > 0$, $Q\{T_a < \infty\} = 1$ and

$$\mathbb{E}\left[\exp\left\{-\frac{m^2 T_a}{2}\right\}\right] = e^{-am}.$$

If the Brownian motion has drift $-m$, then

$$Q\{T_a < \infty\} = e^{-am}\,\mathbb{E}\left[\exp\left\{-\frac{m^2 T_a}{2}\right\}\right] = e^{-2am}.$$

5.3 Girsanov theorem

The Girsanov theorem describes the probability measure one obtains by "weighting" or "tilting" the measure of a Brownian motion by a martingale. One example was given in the previous section where the martingale was

$$M_t = e^{mB_t - (m^2 t/2)}, \tag{5.6}$$

and in the new measure, B_t was a Brownian motion with drift. We will generalize this idea here.

Suppose M_t is a nonnegative martingale satisfying the exponential SDE

$$dM_t = A_t\, M_t\, dB_t, \quad M_0 = 1, \tag{5.7}$$

where B_t is a standard Brownian motion. The solution to this equation was given in (3.20),

$$M_t = e^{Y_t} \quad \text{where} \quad Y_t = \int_0^t A_s\, dB_s - \frac{1}{2} \int_0^t A_s^2\, ds. \tag{5.8}$$

For many applications it suffices to consider the equation (5.7) and not worry about the form of the solution (5.8). Solutions to (5.7) are local martingales, but as we have seen, they might not be martingales. For now we will *assume* that M_t is a martingale. In that case, we can define a probability measure $\mathbb{P}^*$ by saying that if V is an $\mathcal{F}_t$-measurable event, then

$$\mathbb{P}^*(V) = \mathbb{E}\left[1_V\, M_t\right]. \tag{5.9}$$

In other words, if we consider $\mathbb{P}, \mathbb{P}^*$ as being defined on $\mathcal{F}_t$-measurable events,

$$\frac{d\mathbb{P}^*}{d\mathbb{P}} = M_t.$$

If $s < t$ and V is $\mathcal{F}_s$-measurable, then V is also $\mathcal{F}_t$-measurable. Hence, in order for the above definition to be consistent, we need that for such V,

$$\mathbb{E}\left[1_V\, M_s\right] = \mathbb{E}\left[1_V\, M_t\right].$$

Indeed, this holds by the computation (5.4) which only uses the fact that M is a martingale and V is $\mathcal{F}_s$-measurable. We write $\mathbb{E}^*$ for expectations with respect to $\mathbb{P}^*$. If X is $\mathcal{F}_t$-measurable, then

$$\mathbb{E}^*[X] = \mathbb{E}\left[X\, M_t\right].$$

Theorem 5.3.1 (Girsanov Theorem). *Suppose M_t is a nonnegative martingale satisfying (5.7), and let $\mathbb{P}^*$ be the probability measure defined in (5.9). If*

$$W_t = B_t - \int_0^t A_s \, ds,$$

then with respect to the measure $\mathbb{P}^*$, W_t is a standard Brownian motion. In other words,

$$dB_t = A_t \, dt + dW_t,$$

where W is a $\mathbb{P}^*$-Brownian motion.

In other words, if we weight the probability measure $\mathbb{P}$ by the martingale, then in the new measure $\mathbb{P}^*$, the Brownian motion acquires a drift A_t. The example of the previous section is the martingale (5.6) with $A_t \equiv m$ and a constant drift m is obtained.

Let us give a heuristic derivation of Girsanov's theorem using a binomial approximation. Suppose Δt is given. In the binomial approximation to $\mathbb{P}$, we are equally likely to go up and down,

$$\mathbb{P}\left\{ B(t + \Delta t) - B(t) = \pm\sqrt{\Delta t} \mid B(t) \right\} = \frac{1}{2}.$$

The binomial approximation to (5.7) is

$$\mathbb{P}\left\{ M(t + \Delta t) = M(t) \left[1 \pm A(t) \sqrt{\Delta t} \right] \mid B(t) \right\} = \frac{1}{2}.$$

Therefore, in the weighted measure $\mathbb{P}^*$ the probability of a jump of $\pm\sqrt{\Delta}$ should be proportional to $1 \pm A(t) \sqrt{\Delta t}$. Since the sum of the two probabilities equals one, we see that

$$\mathbb{P}^*\left\{ B(t + \Delta t) - B(t) = \pm\sqrt{\Delta t} \mid B(t) \right\} = \frac{1}{2} \left[1 \pm A(t) \sqrt{\Delta t} \right].$$

As we saw in Section 4.4, this implies that

$$E^*[B(t + \Delta t) - B(t) \mid B(t)] = A(t) \, \Delta t,$$

that is, in the probability measure $\mathbb{P}^*$, the process obtains a drift of $A(t)$.

The condition that M_t be a martingale (and not just a local martingale) is necessary for Girsanov's theorem as we have stated it. Given only (5.7)

or (5.8), it may be hard to determine if M_t is a martingale, so it is useful to have a version that applies for local martingales. If we do not know M_t is a martingale, we can still use Theorem 5.3.1 if we are careful to stop the process before anything bad happens. To be more precise, suppose $M_t = e^{Y_t}$ satisfies (5.8), and note that

$$\langle Y \rangle_t = \int_0^t A_s^2 \, ds.$$

Let

$$T_n = \inf\{t : M_t + \langle Y \rangle_t = n\},$$

and let

$$A_t^{(n)} = \begin{cases} A_t, & t < T_n \\ 0, & t \geq T_n \end{cases}.$$

Then

$$dM_{t \wedge T_n} = A_t^{(n)} \, M_{t \wedge T_n} \, dB_t,$$

which is a square integrable martingale since

$$\begin{aligned}
\mathbb{E}\left[(M_{t \wedge T_n} - 1)^2\right] &= \int_0^t \mathbb{E}[A_s^{(n)} \, M_{s \wedge T_n}]^2 \, ds \\
&\leq n^2 \, \mathbb{E}\int_0^t [A_s^{(n)}]^2 ds \leq n^3.
\end{aligned}$$

There is the corresponding measure, which we might denote by $\mathbb{P}_n^*$, which gives a drift of A_t up to time T_n and then proceeds with drift 0. If $n < m$, then $\mathbb{P}_n^*$ and $\mathbb{P}_m^*$ are the same measure restricted to $t \leq T_n$. Hence we can write $\mathbb{P}^*$ for a measure on $B_t, 0 \leq t < T$, where

$$T = \lim_{n \to \infty} T_n.$$

This shows how to tilt the measure up to time T, There are examples such that $\mathbb{P}^*\{T < \infty\} > 0$. However, if for some fixed t_0, $\mathbb{P}^*\{T > t_0\} = 1$, then $M_t, 0 \leq t \leq t_0$ is a martingale. In other words, what prevents a solution to (5.7) from being a martingale is that *with respect to the new measure* $\mathbb{P}^*$, either M_t or $|A_t|$ goes to infinity in finite time. We summarize with a restatement of the Girsanov theorem.

Theorem 5.3.2 (Girsanov Theorem, local martingale form). *Suppose $M_t = e^{Y_t}$ satisfies (5.7)–(5.8), and let*

$$T_n = \inf\{t : M_t + |A_t| = n\}, \quad T = T_\infty = \lim_{n \to \infty} T_n.$$

Let $\mathbb{P}^*$ be the probability measure as above. If

$$W_t = B_t - \int_0^t A_s \, ds, \quad 0 \le t < T,$$

then with respect to the measure $\mathbb{P}^*$, $W_t, t < T$ is a standard Brownian motion. In other words,

$$dB_t = A_t \, dt + dW_t, \quad t < T,$$

where W is a $\mathbb{P}^*$-Brownian motion. If any one of the following three conditions hold, then $M_s, 0 \le s \le t$, is actually a martingale:

$$\mathbb{P}^*\{T > t\} = 1,$$

$$\mathbb{E}[M_t] = 1,$$

$$\mathbb{E}\left[\exp\left\{\frac{\langle Y \rangle_t}{2}\right\}\right] < \infty. \tag{5.10}$$

It is not always easy to see whether or not the local martingale in (5.7) will be a martingale. However, if any one of the three conditions at the end of the theorem hold, then it is a martingale. The first condition uses $\mathbb{P}^*$, the new measure, while the expectation $\mathbb{E}$ in the other two conditions is with respect to the original measure. The relation (5.10) is called the *Novikov condition*.

Even if M_t is not a martingale, since

$$M_t = \lim_{n \to \infty} M_{t \wedge T_n},$$

Fatou's lemma implies that

$$\mathbb{E}[M_t] = \mathbb{E}\left[\lim_{n \to \infty} M_{t \wedge T_n}\right] \le \lim_{n \to \infty} \mathbb{E}[M_{t \wedge T_n}] = \mathbb{E}[M_0] = 1.$$

An extension of this argument shows that nonnegative local martingales are supermartingales, that is, if $s < t$,

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

Example 5.3.1. Let B_t be a standard Brownian motion with $B_0 = 1$. Let $\tau = \inf\{t : B_t = 0\}$. Then $M_t = B_{t \wedge \tau}$ is a nonnegative martingale, satisfying

$$dM_t = dB_t = A_t\, M_t\, dB_t, \quad t < \tau,$$

where

$$A_t = \frac{1}{M_t} = \frac{1}{B_t}.$$

If we tilt the measure using M_t, then

$$dB_t = A_t\, dt + dW_t = \frac{dt}{B_t} + dW_t,$$

where W_t is a standard Brownian motion in the new measure $\mathbb{P}^*$. Note that this equation gives the Bessel process that we studied in Section 4.2 with $a = 1$. Using properties of the Bessel process, we see that

$$\mathbb{P}^*\{\tau < \infty\} = 0.$$

In other words, in the new measure the process avoids the origin.

Example 5.3.2. Let B_t be a standard Brownian motion with $B_0 = 1$ and $r \in \mathbb{R}$. Let $\tau = \inf\{t : B_t = 0\}$. For $t < \tau$, Itô's formula gives

$$
\begin{aligned}
dB_t^r &= r\, B_t^{r-1}\, dB_t + \frac{r(r-1)}{2} B_t^{r-2}\, dt \\
&= B_t^r \left[\frac{r}{B_t}\, dB_t + \frac{r(r-1)}{2B_t^2}\, dt \right].
\end{aligned}
$$

Let

$$M_t = \exp\left\{ -\int_0^t \frac{r(r-1)}{2B_s^2}\, ds \right\} B_t^r.$$

The product rule shows that M_t satisfies the exponential SDE

$$dM_t = \frac{r}{B_t}\, M_t\, dB_t, \quad t < \tau.$$

Therefore,

$$dB_t = \frac{r}{B_t}\, dt + dW_t, \quad t < \tau,$$

where W_t is a Brownian motion in the new measure. This equation is the Bessel equation. In particular, if $r \geq 1/2$, then $\mathbb{P}^*\{\tau = \infty\} = 1$, and using this we see that with $\mathbb{P}^*$-probability one

$$M_t + \int_0^t A_s^2 \, ds$$

is finite. Therefore, for $r \geq 1/2$, M_t is a martingale.

Example 5.3.3. Suppose that X_t satisfies

$$dX_t = X_t \left[m(t, X_t) \, dt + \sigma(t, X_t) \, dB_t \right],$$

where B_t is a standard Brownian motion with respect to the probability measure $\mathbb{P}$. Let us assume that $\sigma(t, x) > 0$ for all t, x. Suppose we want to find a probability measure $\mathbb{P}^*$ that is mutually absolutely continuous with respect to $\mathbb{P}$ and such that X_t is a martingale under $\mathbb{P}^*$. Then we would want

$$dB_t = -\frac{m(t, X_t)}{\sigma(t, X_t)} \, dt + dW_t,$$

where W_t is a standard Brownian motion with respect to $\mathbb{P}^*$. This would give

$$dX_t = X_t \, \sigma(t, X_t) \, dW_t.$$

The local martingale that we need to consider is M_t satisfying

$$dM_t = -\frac{m(t, X_t)}{\sigma(t, X_t)} M_t \, dB_t, \quad M_0 = 1.$$

In other words,

$$M_t = \exp\left\{ \int_0^t A_s \, dB_s - \frac{1}{2} \int_0^t A_s^2 \, ds \right\}, \quad A_t = -\frac{m(t, X_t)}{\sigma(t, X_t)}. \tag{5.11}$$

While M_t is a local martingale, we cannot say it is a martingale without verifying one of the conditions in Theorem 5.3.2.

Example 5.3.4. Suppose X_t is a Bessel process satisfying

$$dX_t = \frac{1}{X_t} \, dt + dB_t, \quad X_0 = 1,$$

where B_t is a standard Browian motion. In Section 4.2 it was shown that with probability one, the process X_t will never reach the origin and, in fact, $X_t \to \infty$. Let

$$M_t = \frac{1}{X_t}.$$

Using Itô's formula, we see that

$$dM_t = -\frac{1}{X_t^2} \, dX_t + \frac{1}{X_t^3} \, d\langle X \rangle_t = -\frac{1}{X_t} \, M_t \, dB_t.$$

Therefore, M_t is a local martingale with $M_0 = 1$ and we can apply the second form of the Girsanov theorem. We will ask if M_t is a martingale. Indeed, one can see that it is not by noting that $M_t \to 0$ and (with some more work that we omit) $\mathbb{E}[M_t] \to 0$. If M_t were a martingale, we would have $\mathbb{E}[M_t] = 1$ for all t. Another way to see that M_t is not a martingale is to consider the measure $\mathbb{P}^*$ given by tilting by the martingale. In this measure,

$$dB_t = -\frac{1}{X_t} \, dt + dW_t,$$

where W_t is a $\mathbb{P}^*$-Brownian motion. Therefore, $dX_t = dW_t$ which says that the distribution of X_t *in the measure* $\mathbb{P}^*$ is that of a standard Brownian motion. Since the Brownian motion reaches zero, we can see that with respect to $\mathbb{P}^*$, the martingale M_t reaches infinity in finite time.

Suppose M_t is a continuous, nonnegative local martingale. To show that M_t is a supermartingale, suppose that τ_n is an increasing sequence of stopping times such that $M_{t \wedge \tau_n}$ is a martingale. Let $s < t$. Suppose V is an $\mathcal{F}_s$-measurable event and let $V_k = V \cap \{\tau_k > s\}$. Since $M_{t \wedge \tau_n}$ is a martingale, if $n \geq k$,

$$\mathbb{E}\left[1_{V_k} M_{t \wedge \tau_n}\right] = \mathbb{E}\left[1_{V_k} M_{s \wedge \tau_n}\right] = \mathbb{E}\left[1_{V_k} M_s\right].$$

Hence Fatou's lemma implies that

$$\mathbb{E}\left[1_{V_k} M_t\right] \leq \liminf_{n \to \infty} \mathbb{E}\left[1_{V_k} M_{t \wedge \tau_n}\right] = \mathbb{E}\left[1_{V_k} M_s\right].$$

Since $\mathbb{E}\left[1_V E(M_t \mid \mathcal{F}_s)\right] = \mathbb{E}\left[1_V M_t\right]$, by letting $k \to \infty$ and using the monotone convergence theorem, we get

$$\mathbb{E}\left[1_V E(M_t \mid \mathcal{F}_s)\right] = \mathbb{E}\left[1_V M_t\right] \leq \mathbb{E}\left[1_V M_s\right].$$

Since this holds for every $\mathcal{F}_s$-measurable event V, we have $E(M_t \mid \mathcal{F}_s) \leq M_s$ with probability one. Also, if $V = \{E(M_t \mid \mathcal{F}_s) < M_s\}$ and $\mathbb{P}(V) > 0$, then

$$\mathbb{E}[M_t] = \mathbb{E}\left[1_V\, M_t\right] + \mathbb{E}\left[1_{V^c}\, M_t\right] < \mathbb{E}\left[1_V\, M_s\right] + \mathbb{E}\left[1_{V^c}\, M_s\right] = \mathbb{E}[M_s].$$

Conversely, if $\mathbb{E}[M_t] = \mathbb{E}[M_0]$ for all t, then M_t must be a martingale. (We emphasize that we are using the fact that $M_t \geq 0$.)

To prove the Girsanov theorem, let us first consider the case when there exists $K < \infty$ such that with probability one $|A_t| \leq K$ and $M_t \leq K$. By Theorem 4.5.2, we need only show that in the measure $\mathbb{P}^*$, W_t is a continuous martingale with quadratic variation $\langle W \rangle_t = t$. Continuity and the quadratic variation calculation are immediate (since they hold with probability one in the measure $\mathbb{P}$), so all we need to show is that W_t is a $\mathbb{P}^*$-martingale, that is, if $s < t$, then

$$E_{\mathbb{P}^*}(W_t \mid \mathcal{F}_s) = W_s.$$

In other words, if V is $\mathcal{F}_s$-measurable, we need to show that

$$\mathbb{E}^*\left[1_V\, W_s\right] = \mathbb{E}^*\left[1_V\, W_t\right],$$

which by definition means

$$\mathbb{E}\left[1_V\, W_s\, M_s\right] = \mathbb{E}\left[1_V\, W_t\, M_t\right].$$

In other words, we need to show that $Z_t = W_t\, M_t$ is a martingale with respect to $\mathbb{P}$. Since

$$dW_t = -A_t\, dt + dB_t, \quad dM_t = A_t\, M_t\, dB_t,$$

the product rule gives

$$dZ_t = W_t\, dM_t + M_t\, dW_t + d\langle W, M \rangle_t = (A_t\, W_t + 1)\, M_t\, dB_t,$$

which shows that Z_t is a local martingale. Also, since $|W_t| \leq |B_t| + tK$ we see that

$$\int_0^t \mathbb{E}[(A_s\, W_s + 1)^2\, M_s^2]\, ds < \infty,$$

and hence Z_t is a square-integrable martingale.

For more general A_t, M_t we use localization with the stopping times T_n as above. Note that

$$\mathbb{E}[M_t] \geq \lim_{n \to \infty} \mathbb{E}\left[M_t\, 1\{T_n > t\}\right] = \lim_{n \to \infty} \mathbb{P}^*\{T_n > t\} = \mathbb{P}^*\{T > t\},$$

and hence if $\mathbb{P}^*\{T > t\} = 1$, then $\mathbb{E}[M_t] = 1$ and $M_s, 0 \leq s \leq t$ is a martingale.

We will now show that the Novikov condition (5.10) implies that M is a martingale. For ease let $t = 1$. Let

$$\tau_r = \inf\{t : \langle Y \rangle_t = r\}.$$

The process

$$X_r = \int_0^{\tau_r} A_s \, dW_s$$

is a standard Brownian motion with respect to the measure $\mathbb{P}^*$. Also,

$$Y_r = X_r + \frac{1}{2} \int_0^{\tau_r} A_s \, ds = X_r + \frac{r}{2}.$$

In particular,

$$\max_{0 \leq s \leq \tau_r} M_s = \max_{0 \leq t \leq r} \exp\left\{X_t + \frac{t}{2}\right\}.$$

In other words, $T = \lim T_n$ can be defined as

$$T = \sup\{t : \langle Y \rangle_t < \infty\}.$$

Let V denote the event that $\tau_r \leq 1$ for all $r < \infty$. We need to show that $\mathbb{P}^*(V) = 0$.

Let

$$\rho_n = \min\left\{m \geq n : X_m \leq 0\right\}.$$

Since X_m is a $\mathbb{P}^*$-local martingale, for each n, $\mathbb{P}^*[V \cap \{\rho_n < \infty\}] = \mathbb{P}^*(V)$. Also,

$$M_{\rho_n} = \exp\left\{Y_{\rho_n}\right\} \leq e^{\rho_n/2}.$$

Let us fix n, let $\rho = \rho_n$, and note that

$$\mathbb{P}^*(V) \leq \sum_{m=n}^{\infty} \mathbb{E}\left[M_m \, \mathbf{1}\left\{\rho = m\right\}\right] \quad \leq \quad \sum_{m=n}^{\infty} e^{m/2} \, \mathbb{P}\{\rho = m\}$$

$$\leq \quad \mathbb{E}\left[e^{\langle Y \rangle_1/2} \, \mathbf{1}\{\rho = m\}\right].$$

We therefore, get

$$\mathbb{P}^*(V) \leq \mathbb{E}\left[e^{\langle Y \rangle_1/2} \, \mathbf{1}\{\tau_r < 1\}\right].$$

The events $\{\tau_r \leq 1\}$ shrink to a $\mathbb{P}$-null set as $r \to \infty$. The condition (5.10) implies that $e^{\langle Y \rangle_1/2}$ is integrable, and hence,

$$\lim_{r \to \infty} \mathbb{E}\left[e^{\langle Y \rangle_1/2} 1\{\tau_r \leq 1\}\right] = 0.$$

5.4 Black-Scholes formula

An *arbitrage* is a system that guarantees that a player (investor) will not lose money while also giving a positive probability of making money. If P and Q are equivalent probability measures, then an arbitrage under probability P is the same as an arbitrage under probability Q. This holds since for equivalent probability measures

$$P(V) = 0 \quad \text{if and only if} \quad Q(V) = 0,$$

$$P(V) > 0 \quad \text{if and only if} \quad Q(V) > 0.$$

This observation is the main tool for the pricing of options as we now show. We will consider a simple (European) call option for a stock whose price moves according to a geometric Brownian motion.

Suppose that the stock price S_t follows the geometric Brownian motion,

$$dS_t = S_t \left[m\, dt + \sigma\, dB_t\right], \tag{5.12}$$

and there also exists a risk-free bound R_t satisfying

$$dR_t = r\, R_t\, dt, \tag{5.13}$$

that is, $R_t = e^{rt} R_0$. Let T be a time in the future and suppose we have the option to buy a share of stock at time T for strike price K. The value of this option at time T is

$$F(S_T) = (S_T - K)_+ = \begin{cases} (S_T - K) & \text{if } S_T > K, \\ 0 & \text{if } S_T \leq K. \end{cases}$$

The goal is to find the price $f(t, x)$ of the option at time $t < T$ given $S_t = x$.

One possibility, which is *not* the Black-Scholes solution, is to price the option by the expected value (in time t dollars),

$$\hat{f}(t, x) = \mathbb{E}\left[e^{-r(T-t)} F(S_T) \mid S_t = x\right].$$

In (4.7) we showed that this function satisfies the PDE

$$\partial_t \hat{f}(t, x) = r \, \hat{f}(t, x) - m \, x \, \hat{f}'(t, x) - \frac{\sigma^2 \, x^2}{2} \, \hat{f}''(t, x). \tag{5.14}$$

Here and throughout this section we use primes for x-derivatives. If one sells an option at this price and uses the money to buy a bond at the current interest rate, then there is a positive probability of losing money.

The Black-Scholes approach to pricing is to let $f(t, x)$ be the value of a portfolio at time t, given that $S_t = x$, that can be *hedged* in order to guarantee a portfolio of value $F(S_T)$ at time T. By a portfolio, we mean an ordered pair (a_t, b_t) where a_t, b_t denote the number of units of stocks and bonds, respectively. Let V_t be the value of the portfolio at time t,

$$V_t = a_t \, S_t + b_t \, R_t. \tag{5.15}$$

We will manage the portfolio by switching between stocks and bonds so that, no matter how the price of the stock moves, the value at time T will be

$$V_T = (S_T - K)_+.$$

We assume that the portfolio is *self-financing*, that is, one does not add outside resources to the portfolio. The mathematical consequence of this assumption is that the change of the value of the portfolio is given only by the change of the price of the assets,

$$dV_t = a_t \, dS_t + b_t \, dR_t. \tag{5.16}$$

It may appear at first that (5.16) is a consequence of (5.15). However, the units of the assets a_t, b_t vary with time and hence we need to use the product rule. If V_t is defined by (5.15), the product rule implies that

$$d(a_t \, S_t) = a_t \, dS_t + S_t \, da_t + d\langle a, S \rangle_t,$$

$$d(b_t \, R_t) = b_t \, dR_t + R_t \, db_t.$$

Here we use the fact that R_t is differentiable in t to see that $\langle b, R \rangle_t = 0$. Hence, (5.16) is a strong *assumption* about the portfolio, and we will use it to determine the price and the hedging strategy.

If we assume (5.16) and plug in (5.12) and (5.13), we get

$$
\begin{aligned}
dV_t &= a_t\, S_t\, [m\, dt + \sigma\, dB_t] + b_t\, r\, R_t\, dt \\
&= a_t\, S_t\, [m\, dt + \sigma\, dB_t] + r\, [V_t - a_t\, S_t]\, dt \\
&= [m\, a_t\, S_t + r\, (V_t - a_t\, S_t)]\, dt + \sigma\, a_t\, S_t\, dB_t. \qquad (5.17)
\end{aligned}
$$

Under our definition, $V_t = f(t, S_t)$. Therefore, assuming f is sufficiently differentiable, Itô's formula shows that

$$
\begin{aligned}
dV_t &= df(t, S_t) \\
&= \partial_t f(t, S_t)\, dt + f'(t, S_t)\, dS_t + \frac{1}{2}\, f''(t, S_t)\, d\langle S \rangle_t \\
&= \left[\partial_t f(t, S_t) + m\, S_t\, f'(t, S_t) + \frac{\sigma^2 S_t^2}{2}\, f''(t, S_t) \right] dt \\
&\quad + \sigma\, S_t\, f'(t, S_t)\, dB_t. \qquad (5.18)
\end{aligned}
$$

By equating the dB_t terms in (5.17) and (5.18), we see that that the portfolio is given by

$$
a_t = f'(t, S_t), \qquad b_t = \frac{V_t - a_t\, S_t}{R_t}, \qquad (5.19)
$$

and then by equating the dt terms we get the Black-Scholes equation

$$
\partial_t f(t, x) = r\, f(t, x) - r\, x\, f'(t, x) - \frac{\sigma^2\, x^2}{2}\, f''(t, x).
$$

There are two things to note about this equation.

- The drift term m does not appear. If we think about this, we realize why our assumptions should give us an equation independent of m. Our price was based on being able to hedge our portfolio so that with probability one the value at time T is $(S_T - K)_+$. Geometric Brownian motions with the same σ but different m are mutually absolutely continuous and hence have the same events of probability one.

- The equation is exactly the same as (5.14) except that m has been replaced with r. Therefore, we can write

$$
f(t, x) = \mathbb{E}\left[e^{-r(T-t)}\, F(S_t) \mid S_t = x \right],
$$

where S satisfies

$$
dS_t = S_t\, [r\, dt + \sigma\, dB_t].
$$

Using this, one can compute $f(t,x)$ exactly; in the next section we do this and derive the *Black-Scholes formula*

$$f(T-t,x) = x\,\Phi\left(\frac{\log(x/K) + (r + \frac{\sigma^2}{2})t}{\sigma\sqrt{t}}\right)$$

$$-K\,e^{-rt}\,\Phi\left(\frac{\log(x/K) + (r - \frac{\sigma^2}{2})t}{\sigma\sqrt{t}}\right), \tag{5.20}$$

where Φ is the standard normal distribution function. If one knows r and σ (which from a practical perspective is a *big* assumption), one can just plug into this formula.

We can easily generalize this to the case where the stock price satisfies

$$dS_t = S_t\,[m(t, S_t)\,dt + \sigma(t, S_t)\,dB_t],$$

$$dR_t = r(t, S_t)\,R_t\,dt.$$

Under the assumption of a self-financing portfolio, we again get (5.17) and (5.18), and by equating coefficients we get the Black-Scholes equation

$$\partial_t f(t,x) = r(t,x)\,f(t,x) - r(t,x)\,x\,f'(t,x) - \frac{\sigma(t,x)^2\,x^2}{2}\,f''(t,x). \tag{5.21}$$

As before, the drift term $m(t,x)$ does not appear in the equation. The function f can be given by

$$f(t,x) = \mathbb{E}\left[(R_t/R_T)\,F(S_T) \mid S_t = x\right],$$

where S_t, R_t satisfy

$$dS_t = S_t\,[r(t, S_t)\,dt + \sigma(t, S_t)\,dB_t],$$

$$dR_t = r(t, S_t)\,R_t\,dt.$$

Mathematical justification of this arguments requires sufficient assumptions on m and σ such that the equation (5.21) has a solution. Finding explicit solutions to such equations is often impossible, but one can either solve the PDE numerically or do Monte Carlo simulations of the associated diffusion process S_t.

5.5   Martingale approach to Black-Scholes equation

We will give a different approach to deriving the Black-Scholes formula. Suppose for the moment that the risk-free bond has rate $r(t, x)$ and that the volatility is given by $\sigma(t, x)$. If R_t denotes the value of the bond at time t, then

$$dR_t = r(t, S_t)\, R_t\, dt, \quad R_t = R_0 \, \exp\left\{ \int_0^t r(s, S_s)\, ds \right\}.$$

As explained in the previous section, if we want a strategy to hedge a portfolio so that its value is determined at time T, then the strategy must be independent of the drift coefficient $m(t, x)$. For this reason, we may assume that the stock price satisfies

$$dS_t = S_t \left[r(t, S_t)\, dt + \sigma(t, S_t)\, dB_t \right], \tag{5.22}$$

and then the value of the portfolio at time t satisfies

$$V_t = f(t, S_t) = E_Q \left[(R_t/R_T)\, F(S_T) \mid S_t \right] = E_Q \left[(R_t/R_T)\, F(S_T) \mid \mathcal{F}_t \right].$$

We write $\mathbb{E}_Q, E_Q$ to emphasize that the expectations are taken with respect to the measure under which S_t satisfies (5.22).

Let $\tilde{S}_t = S_t/R_t, \tilde{V}_t = V_t/R_t$ be the stock price and portfolio value, respectively, discounted by the bond rate. The product rule shows that $\tilde{S}_t$ satisfies

$$d\tilde{S}_t = \sigma(t, R_t)\, \tilde{S}_t\, dB_t.$$

In other words (under some growth restrictions on σ, for example, if σ is uniformly bounded), $\tilde{S}_t$ is a martingale. Also,

$$\tilde{V}_t = V_t/R_t = R_t^{-1}\, E_Q \left((R_t/R_T)\, F(S_T) \mid \mathcal{F}_t \right)$$

$$= E_Q \left(R_T^{-1}\, F(S_T) \mid \mathcal{F}_t \right) = E_Q[\tilde{V}_T \mid \mathcal{F}_t].$$

We have just demonstrated the following principle.

Theorem 5.5.1. *Suppose S_t satisfies*

$$dS_t = S_t \left[m(t, S_t)\, dt + \sigma(t, S_t)\, dB_t \right],$$

and a risk-free bond R_t is available at rate $r(t, S_t)$,

$$dR_t = r(t, S_t) R_t dt.$$

Suppose that the Brownian motion is defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and that there exists a probability measure Q that is mutually absolutely continuous with respect to $\mathbb{P}$ such that under Q, the discounted stock price $\tilde{S}_t = S_t/R_t$ is a martingale. Suppose there is an option at time T with value $F(S_T)$ such that $\mathbb{E}_Q[R_T^{-1} |F(S_T)|] < \infty$. Then the arbitrage-free price of the option at time t is

$$V_t = R_t E_Q \left(R_T^{-1} F(S_T) \mid \mathcal{F}_t \right). \tag{5.23}$$

A nice thing about this formulation is that $V_T = F(S_T)$ follows directly from the formula. However, unlike the previous approach, we do not immediately get the expression for the portfolio needed to hedge the option.

Example 5.5.1. We will derive the Black-Scholes formula leaving some of the calculus calculations to Exercise 5.11. Suppose that r, σ are constants and $F(S_T) = (S_T - K)_+$. The discounted values are $\tilde{S}_t = e^{-rt} S_t$, $\tilde{V}_t = e^{-rt} V_t$ and

$$\tilde{V}_T = e^{-rT} F(S_T) = e^{-rT} (S_T - K)_+ = (\tilde{S}_T - \tilde{K})_+,$$

where $\tilde{K} = e^{-rT} K$. Under the measure Q, $\tilde{S}_t$ satisfies

$$d\tilde{S}_t = \sigma \tilde{S}_t dB_t,$$

which implies that

$$
\begin{aligned}
\tilde{S}_T &= \tilde{S}_t \exp \left\{ \int_t^T \sigma dB_s - \frac{1}{2} \int_t^T \sigma^2 ds \right\} \\
&= \tilde{S}_t \exp \left\{ \sigma(B_T - B_t) - \frac{\sigma^2(T - t)}{2} \right\}.
\end{aligned}
$$

In other words, the conditional distribution of $\tilde{S}_T$ given $\tilde{S}_t$ is that of

$$Z = \exp \left\{ aN + y \right\},$$

where $a = \sigma\sqrt{T - t}$, N is a standard normal random variable, and

$$y = \log \tilde{S}_t - \frac{a^2}{2}.$$

Straightforward calculus shows that Z has a density

$$g(z) = \frac{1}{az} \phi \left(\frac{-y + \log z}{a} \right),$$

where ϕ is the standard normal density, and hence

$$\tilde{V}_t = \int_{\tilde{K}}^{\infty} (z - \tilde{K}) \, g(z) \, dz.$$

Another computation gives

$$\tilde{V}_t = \tilde{S}_t \, \Phi \left(\frac{\log(\tilde{S}_t/\tilde{K}) + \frac{a^2}{2}}{a} \right) - \tilde{K} \, \Phi \left(\frac{\log(\tilde{S}_t/\tilde{K}) - \frac{a^2}{2}}{a} \right),$$

which implies that

$$V_t = e^{rt} \, \tilde{V}_t = S_t \, \Phi \left(\frac{\log(S_t/K) + rs + \frac{a^2}{2}}{a} \right)$$

$$-e^{-rs} \, K \, \Phi \left(\frac{\log(S_t/K) + rs - \frac{a^2}{2}}{a} \right),$$

where $s = T - t$. Plugging in $a = \sigma\sqrt{s}$ gives (5.20).

One of the hypotheses in Theorem 5.5.1 is that if S_t satisfies

$$dS_t = S_t \left[m(t, S_t) \, dt + \sigma(t, S_t) \, dB_t \right],$$

then there exists a probability measure Q under which

$$dS_t = S_t \left[r(t, S_t) \, dt + \sigma(t, S_t) \, dW_t \right], \tag{5.24}$$

where W_t is a Q-Brownian motion. Indeed, if S_t satisfies (5.24), then the discounted price $\tilde{S}_t = S_t/R_t$ satisfies

$$d\tilde{S}_t = \tilde{S}_t \, \sigma(t, S_t) \, dW_t = \tilde{S}_t \, \sigma(t, R_t \tilde{S}_t) \, dW_t. \tag{5.25}$$

The Girsanov theorem tells us that the way to obtain Q is to tilt by the local martingale M_t where

$$dM_t = M_t \, \frac{r(t, S_t) - m(t, S_t)}{\sigma(t, S_t)} \, dB_t,$$

for then in the measure Q,

$$dB_t = \frac{r(t, S_t) - m(t, S_t)}{\sigma(t, S_t)} \, dt + dW_t.$$

In order for the theorem to hold, we need that M_t is actually a martingale and that $\tilde{S}_t$ as given in (5.25) is a martingale. While one can give general conditions when this is true, it is often just as easy to check this in each case. If $|r - m|/\sigma$ is uniformly bounded, these conditions are satisfied.

5.6 Martingale approach to pricing

We generalize the discussion in the previous section to the pricing of claims that are functions of the entire history of the prices of an asset. Suppose S_t denotes the price of an asset satisfying

$$dS_t = S_t \, [m_t \, dt + \sigma_t \, dB_t] \tag{5.26}$$

where B_t is a standard Brownian motion. Let $\mathcal{F}_t$ denote the information in $\{B_s : 0 \le s \le t\}$, and as usual we assume that m_t, σ_t are processes adapted to the filtration $\{\mathcal{F}_t\}$. We also assume that there is a risk-free bond R_t satisfying

$$dR_t = r_t \, R_t \, dt, \quad R_t = R_0 \, \exp\left\{ \int_0^t r_s \, ds \right\},$$

where r_t is also adapted.

Let T be a fixed future time and assume that V is an $\mathcal{F}_T$-measurable random variable. We call V a *claim (at time T)*. The examples we have seen are of the form $V = F(S_T)$, but other examples are

$$V = \max_{0 \le t \le T} S_t, \quad V = \frac{1}{T} \int_0^T S_t \, dt.$$

We will start with the following definition.

Definition, first try If V is a claim at time T, then the (*arbitrage-free*) *price $V_t, 0 \le t \le T$*, of a claim V_T is the minimum value of a self-financing portfolio that can be hedged to guarantee that its value at time T is V.

Our goal is to determine the price V_t and the corresponding portfolio (a_t, b_t), where a_t denotes the number of units of S and b_t the number of units of R. This will require some mathematical assumptions that we will make as we need them. Recall that

$$V_t = a_t \, S_t + b_t \, R_t,$$

and (a_t, b_t) is self-financing if

$$dV_t = a_t \, dS_t + b_t \, dR_t.$$

We will start by giving two bad, but unrealistic, examples that show that we need to take some care. The two examples are similar. In the first example, we allow the stock price to fluctuate too much. In the second, we choose a very risky portfolio similar to the martingale betting strategy.

Example 5.6.1. Suppose for ease that $r_t \equiv 0$ and $R_0 = 1$ so that $R_t \equiv 1$. Let $S_t = e^{Z_t}$ where Z_t is the "martingale betting strategy revisited" in Example 4.1.1. We recall that Z_t satisfies

$$dZ_t = A_t \, dB_t, \quad Z_0 = 0,$$

and the "bets" A_t are chosen so that with probability one $Z_1 \geq 1$. Itô's formula shows that S_t satisfies

$$dS_t = S_t \left[\frac{A_t^2}{2} \, dt + A_t \, dB_t \right].$$

We cannot find an equivalent measure such that S_t is a martingale. Indeed, we know that with probability one $S_1 > S_0$, and hence this fact must be true in any equivalent measure.

Example 5.6.2. Suppose that $R_t \equiv 1$ and $S_t = e^{B_t - (t/2)}$ which satisfies

$$dS_t = S_t \, dB_t, \quad S_0 = 1.$$

Let A_t be as in Example 5.6.1, and choose a portfolio with $a_t = A_t/S_t$. The value of the portfolio is $V_t = a_t \, S_t + b_t \, R_t = A_t + b_t$. If the portfolio is to be self-financing, we need that

$$dV_t = a_t \, dS_t + b_t \, dR_t = \frac{A_t}{S_t} \, S_t \, dB_t = A_t \, dB_t.$$

We choose b_t so that this holds (this may require choosing b_t to be negative). Since $V_0 = 1$, we see that $V_t = 1 + Z_t$ where Z_t is as in Example 5.6.1. In particular, $V_1 = 2$, and this portfolio hedges the claim $V \equiv 2$ at time $T = 1$. Similarly for any constant J, we can find A_t such that with probability one $Z_t \geq J$. Hence given any V_0 (even negative values), we can find a portfolio that hedges the claim $V \equiv 2$. One disadvantage of these portfolios, however, is that they allow the value at times $0 < t < 1$ to be negative. It will turn out that we can eliminate examples like this by restricting to portfolios whose value at all times is nonnegative.

With the bad examples in mind, we now proceed to develop the theory. To start, we will consider discounted prices. Let

$$\tilde{S}_t = (R_0/R_t)\, S_t, \qquad \tilde{V}_t = (R_0/R_t)\, V_t$$

denote the discounted stock price and discounted value, respectively. The portfolio (a_t, b_t) is the same whether or not we are using discounted units and the discounted value $\tilde{V}_t$ is given by

$$\tilde{V}_t = a_t\, \tilde{S}_t + b_t\, R_0.$$

(Note that the "discounted bond value" is R_0.) Using the product formula, we see that

$$d\tilde{S}_t = \tilde{S}_t\, [(m_t - r_t)\, dt + \sigma_t\, dB_t]\,.$$

Our goal is to find a self-financing portfolio (a_t, b_t) such that with probability one

$$\tilde{V}_T = a_T\, \tilde{S}_T + b_T\, R_0 = \tilde{V}.$$

Since this must happen with probability one, we may consider a mutually absolutely continuous measure. We let Q be the probability measure (if it exists) that is mutually absolutely continuous with respect to $\mathbb{P}$ such that under Q the discounted stock price is a martingale. Recalling (5.11), we can see that the Girsanov theorem tells us to choose

$$dQ = M_t\, d\mathbb{P},$$

where M_t satisfies

$$dM_t = \frac{r_t - m_t}{\sigma_t}\, M_t\, dB_t, \qquad M_0 = 1. \tag{5.27}$$

The solution to this last equation is a local martingale, but it not necessarily a martingale. If it is not a martingale, then some undesirable conclusions may result as in our examples above. Our first assumption will be that it is a martingale.

- **Assumption 1.** The local martingale defined in (5.27) is actually a martingale.

This assumption implies that Q is mutually absolutely continuous with respect to $\mathbb{P}$. Theorem 5.3.2 gives a number of ways to establish $Q \ll \mathbb{P}$. If $Q \ll \mathbb{P}$, then we also get $\mathbb{P} \ll Q$ if $\mathbb{P}\{M_t > 0\} = 1$. Let

$$W_t = B_t - \int_0^t \frac{r_s - m_s}{\sigma_s} \, ds,$$

which is a Brownian motion with respect to Q. Plugging in we see that

$$d\tilde{S}_t = \sigma_t \, \tilde{S}_t \, dW_t. \tag{5.28}$$

This shows that $\tilde{S}_t$ is a *local* martingale with respect to Q. We will want this to be a martingale, and we make this assumption.

- **Assumption 2.** The Q-local martingale $\tilde{S}_t$ satisfying (5.28) is actually a Q-martingale.

Again, Theorem 5.3.2 gives some sufficient conditions for establishing that the solution to (5.28) is a Q-martingale. We write $\mathbb{E}_Q$ and E_Q for expectations (regular and conditional) with respect to Q.

Definition A claim V at time T is called a *contingent claim* if $V \geq 0$ and

$$\mathbb{E}_Q \left[\tilde{V}^2 \right] < \infty.$$

The *(arbitrage-free) price* $V_t, 0 \leq t \leq T$, of a contingent claim V_T is the minimum value of a self-financing portfolio that can be hedged to guarantee that its value never drops below zero and at time T equals V.

Given a contingent claim, we can set

$$\tilde{V}_t = E_Q \left[\tilde{V} \mid \mathcal{F}_t \right].$$

This is a square integrable martingale and $\tilde{V}_T = \tilde{V}$. We would like to find a portfolio (a_t, b_t) satisfying

$$V_t = a_t\, S_t + b_t\, R_t,$$

that is self-financing,

$$dV_t = a_t\, dS_t + b_t\, dR_t.$$

Recall that if A_t is an adapted process with

$$\int_0^t \mathbb{E}_Q[A_s^2]\, ds < \infty,$$

then

$$Z_t = \int_0^t A_s\, dW_s$$

is a square integrable martingale. Let us assume for the moment that there exists such a process A_s such that

$$\tilde{V}_t = \tilde{V}_0 + \int_0^t A_s\, dW_s,$$

that is

$$d\tilde{V}_t = A_t\, dW_t. \tag{5.29}$$

We compute,

$$\begin{aligned}
dV_t &= R_t\, d\tilde{V}_t + \tilde{V}_t\, dR_t \\
&= R_t\, A_t\, dW_t + \tilde{V}_t\, dR_t \\
&= \frac{A_t}{\sigma_t\, \tilde{S}_t}\, R_t\, d\tilde{S}_t + \tilde{V}_t\, dR_t \\
&= \frac{A_t}{\sigma_t\, \tilde{S}_t}\, [dS_t - \tilde{S}_t\, dR_t] + \tilde{V}_t\, dR_t \\
&= \frac{A_t}{\sigma_t\, \tilde{S}_t}\, dS_t + \left[\tilde{V}_t - \frac{A_t}{\sigma_t}\right] dR_t.
\end{aligned}$$

Therefore, if

$$a_t = \frac{A_t}{\sigma_t\, \tilde{S}_t}, \quad b_t = \tilde{V}_t - \frac{A_t}{\sigma_t}, \tag{5.30}$$

the portfolio is self-financing and

$$a_t\, S_t + b_t\, R_t = \frac{A_t}{\sigma_t\, \tilde{S}_t}\, \tilde{S}_t\, R_t + \left[\tilde{V}_t - \frac{A_t}{\sigma_t}\right] R_t = R_t\, \tilde{V}_t = V_t.$$

Along the way we made the assumption that we could write V_t as (5.29). It turns out, as we discuss in the next section, that this can always be done although we cannot guarantee that the process A_t is continuous or piecewise continuous. Knowing existence of the process is not very useful if one cannot find A_t. For now we just write as an assumption that the computations work out.

- **Assumption 3**. We can write $\tilde{V}_t$ as (5.29), and if we define a_t, b_t as in (5.30), then the stochastic integral

$$V_t = \int_0^t a_s\, dS_s + \int_0^t b_s\, dR_s,$$

  is well defined.

Theorem 5.6.1. *If V is a contingent claim and assumptions 1-3 hold, then the arbitrage-free price is*

$$V_t = R_t\, E_Q(\tilde{V}_T \mid \mathcal{F}_t).$$

We have done most of the work in proving this theorem. What remains is to show that if (a_t^*, b_t^*) is a self-financing strategy with value

$$V_t^* = a_t^*\, S_t + b_t^*\, R_t,$$

such that with probability one, $V_t^* \geq 0$ for all t and $V_T^* \geq V$, then for all t, with probability one $V_t^* \geq V_t$. When we say "with probability one" this can be with respect to either $\mathbb{P}$ or Q since one of our assumptions is that the two measures are mutually absolutely continuous. Let $\tilde{V}_t^* = V_t^*/R_t$ be the discounted values. The product rule gives

$$dV_t^* = d(R_t\tilde{V}_t^*) = R_t\, d\tilde{V}_t^* + \tilde{V}_t^*\, dR_t = R_t\, d\tilde{V}_t^* + \left[a_t^*\, \tilde{S}_t + b_t^*\right] dR_t,$$

and the self-financing assumptions implies that

$$dV_t^* = a_t^*\, dS_t + b_t^*\, dR_t = a_t^* \left[R_t\, d\tilde{S}_t + \tilde{S}_t\, dR_t\right] + b_t^*\, dR_t.$$

By equating coefficients, we see that

$$dV_t^* = a_t^* \, d\tilde{S}_t = a_t^* \, \sigma_t \, \tilde{S}_t \, dW_t.$$

In particular, $\tilde{V}_t^*$ is a nonnegative local martingale. We have seen that this implies that $\tilde{V}_t^*$ is a supermartingale and

$$E_Q \left[\tilde{V}_T^* \mid \mathcal{F}_t \right] \leq \tilde{V}_t^*.$$

If $V \leq \tilde{V}_T^*$, then

$$E_Q \left[\tilde{V}_T^* \mid \mathcal{F}_t \right] \geq E_Q \left[V \mid \mathcal{F}_t \right] = V_t.$$

Example 5.6.3. We will give another derivation of the Black-Scholes equation. Assume that the stock price is a diffusion satisfying

$$dS_t = S_t \left[m(t, S_t) \, dt + \sigma(t, S_t) \, dB_t \right],$$

and the bond rate satisfies

$$dR_t = r(t, S_t) \, R_t \, dt.$$

The product rule implies that the discounted stock price satisfies

$$d\tilde{S}_t = \tilde{S}_t \left[(m(t, S_t) - r(t, S_t)) \, dt + \sigma(t, S_t) \, dB_t \right].$$

If V is a claim of the form $V = F(S_T)$, let ϕ be the function

$$\phi(t, x) = E_Q \left[(R_t / R_T) \, V \mid S_t = x \right],$$

and note that

$$V_t = R_t \, \tilde{V}_t = R_t \, E_Q \left[R_T^{-1} V \mid \mathcal{F}_t \right] = \phi(t, S_t).$$

Assuming sufficient smoothness, Itô's formula gives

$$d\phi(t, S_t) = \partial_t \phi(t, S_t) \, dt + \phi'(t, S_t) \, dS_t + \frac{1}{2} \phi''(t, S_t) \, d\langle S \rangle_t.$$

Recalling that

$$dS_t = S_t \left[r(t, S_t) \, dt + \sigma(t, S_t) \, dW_t \right],$$

we see that

$$d\tilde{V}_t = d[R_t^{-1}\,\phi(t, S_t)] = J_t\,dt + A_t\,dW_t,$$

where

$$J_t = R_t^{-1}\left[\partial_t\phi(t, S_t) + \frac{\sigma(t, S_t)^2\,S_t^2}{2}\,\phi''(t, S_t)\right.$$

$$\left. +r(t, S_t)\,S_t\,\phi'(t, S_t) - r(t, S_t)\phi(t, S_t)\right],$$

$$A_t = R_t^{-1}\,S_t\,\sigma(t, S_t)\,\phi'(t, S_t) = \tilde{S}_t\,\sigma(t, S_t)\,\phi'(t, S_t).$$

Since $\tilde{V}_t$ is a Q-martingale, $J_t = 0$, giving the Black-Scholes PDE again, and $d\tilde{V}_t = A_t\,dW_t$. Plugging into (5.30), we recover the hedging portfolio,

$$a_t = \frac{A_t}{\sigma(t, S_t)\,\tilde{S}_t} = \phi'(t, S_t),$$

$$b_t = \tilde{V}_t - \frac{A_t}{\sigma(t, S_t)} = R_t^{-1}\left[V_t - S_t\,\phi'(t, S_t)\right].$$

This can be compared to (5.19).

Example 5.6.4. Suppose S_t is a geometric Brownian motion

$$dS_t = S_t\left[m\,dt + \sigma\,dB_t\right],$$

and the bond rate is constant r. Suppose that the claim is the average stock price over the interval $[0, T]$,

$$V = \frac{1}{T}\int_0^T S_t\,dt.$$

In the new measure Q, the discounted stock price $\tilde{S}_t = e^{-rt}\,S_t$ satisfies

$$d\tilde{S}_t = \sigma\,\tilde{S}_t\,dW_t,$$

where W_t is a Q-Brownian motion. The discounted value is

$$\tilde{V}_t = E_Q\left[\frac{e^{-rT}}{T}\int_0^T S_s\,ds \mid \mathcal{F}_t\right].$$

Since $\int_0^t S_s \, ds$ is $\mathcal{F}_t$-measurable, we get

$$
\begin{aligned}
T \, e^{rT} \, \tilde{V}_t &= \int_0^t S_s \, ds + E_Q \left[\int_t^T S_s \, ds \mid \mathcal{F}_t \right] \\
&= \int_0^t S_s \, ds + \int_t^T E_Q \left[S_s \mid \mathcal{F}_t \right] \, ds.
\end{aligned}
$$

The second equality uses a form of Fubini's theorem that follows from the linearity of conditional expectation. Since $\tilde{S}_s$ is a Q-martingale, if $s > t$,

$$
E_Q \left[S_s \mid \mathcal{F}_t \right] = e^{rs} \, E_Q \left[\tilde{S}_s \mid \mathcal{F}_t \right] = e^{rs} \, \tilde{S}_t = e^{r(s-t)} \, S_t.
$$

Therefore,

$$
\int_t^T E_Q \left[S_s \mid \mathcal{F}_t \right] \, ds = S_t \int_t^T e^{r(s-t)} \, ds = \frac{e^{r(T-t)} - 1}{r} \, S_t,
$$

and

$$
\begin{aligned}
\tilde{V}_t &= \frac{e^{-rT}}{T} \int_0^t S_s \, ds + \frac{e^{-rt} - e^{-rT}}{rT} \, S_t \\
&= \frac{e^{-rT}}{T} \int_0^t S_s \, ds + \frac{1 - e^{-r(T-t)}}{rT} \, \tilde{S}_t, \quad\quad (5.31)
\end{aligned}
$$

$$
V_t = e^{rt} \, \tilde{V}_t = \frac{e^{-r(T-t)}}{T} \int_0^t S_s \, ds + \frac{1 - e^{-r(T-t)}}{rT} \, S_t.
$$

Note that $V_T = V$ which we needed, and the price at time 0 is

$$
V_0 = \frac{1 - e^{-rT}}{rT} \, S_0.
$$

The hedging portfolio can be worked out with a little thought. We will start with all the money in stocks and as time progresses we move money into bonds. Suppose that during time interval $[t, t+\Delta t]$ we convert $u \Delta t$ units of stock into bonds. Then the value of these units of bonds at time T will be about $u \, e^{r(T-t)} \, S_t \, \Delta t$. If we choose $u = e^{r(t-T)}/T$, then the value will be about $S_t \, \Delta t / T$ and hence the value of all our bonds will be about $\frac{1}{T} \int_0^T S_s \, ds$. This gives us

$$
\frac{da_t}{dt} = -\frac{e^{r(t-T)}}{T}
$$

and using the terminal condition $a_T = 0$, we get

$$a_t = \frac{1 - e^{r(t-T)}}{rT}. \tag{5.32}$$

This is a special case where the hedging strategy does not depend on the current stock price S_t. If we want to use the formula we derived, we use (5.31) to give

$$d\tilde{V}_t = \frac{1 - e^{r(t-T)}}{rT} d\tilde{S}_t = \frac{1 - e^{r(t-T)}}{rT} \sigma \tilde{S}_t \, dW_t.$$

Plugging this into (5.30) gives (5.32).

5.7 Martingale representation theorem

In the last section we assumed in (5.29) that a continuous martingale could be represented as a stochastic integral. In fact, this is always that case. Suppose B_t is a standard Brownian motion and $\{\mathcal{F}_t\}$ is the filtration given by the Brownian motion. Suppose there is a claim V at time T and let

$$V_t = \mathbb{E}[V \mid \mathcal{F}_t].$$

Then there exists an adapted process A_t such that

$$\mathbb{E}[V \mid \mathcal{F}_t] = \mathbb{E}[V] + \int_0^t A_s \, dB_s.$$

In order to be precise, we need to allow a wider class of processes than those with continuous and piecewise continuous paths. We will not go into detalis here, but any allowable process can be approximated by continuous and piecewise continuous processes. From practical perspective, knowing that such an A_s exists is not so useful unless one can give it explicitly. In the examples we give the SDE for V_s explicitly.

This theorem states roughly that all of the randomness in the system is given by the Brownian motion B_t and the only way to get martingales with only this information is to vary the "betting strategy" on the Brownian motion. To illustrate this idea, we will derive the representation theorem in the easier case of random walk.

Suppose we have a binomial model with $\Delta t = 1/N$ and $\Delta x = \sqrt{\Delta t} = 1/\sqrt{N}$. We have independent random variables

$$X_{\Delta t}, X_{2\Delta t}, \ldots$$

each with

$$\mathbb{P}\left\{X_{k\Delta t} = \Delta x\right\} = \mathbb{P}\left\{X_{k\Delta t} = -\Delta x\right\} = \frac{1}{2}.$$

Let $\mathcal{F}_{k\Delta t}$ denote the information in $\{X_{\Delta t}, X_{2\Delta t}, , \ldots, X_{k\Delta t}\}$, and assume that $M_{k\Delta t}$ is a martingale with respect to $\mathcal{F}_{k\Delta t}$. The martingale property implies that

$$E\left[M_{(k+1)\Delta t} \mid \mathcal{F}_{k\Delta t}\right] = M_{k\Delta t}. \tag{5.33}$$

Since $\{M_{(k+1)\Delta t}\}$ is $\mathcal{F}_{(k+1)\Delta t}$-measurable, its value depends on the values of

$$X_{\Delta t}, X_{2\Delta t}, \ldots, X_{(k+1)\Delta t}.$$

When we take a conditional expectation with respect to $\mathcal{F}_{k\Delta t}$ we are given the value of the vector

$$\mathbf{X}_{k\Delta} = \left(X_{\Delta t}, X_{2\Delta t}, \ldots, X_{k\Delta t}\right).$$

Given a particular value of $\mathbf{X}_{k\Delta t}$, there are only two possible values for $M_{(k+1)\Delta t}$ corresponding to the values when $X_{(k+1)\Delta t} = \Delta x$ and $X_{(k+1)\Delta t} = -\Delta x$, respectively. Let us denote the two values by

$$[b + a]\,\Delta x, \quad [b - a]\,\Delta x,$$

where $b\,\Delta x$ is the average of the two values. The two numbers a, b depend on $\mathbf{X}_{k\Delta t}$ and hence are $\mathcal{F}_{k\Delta t}$-measurable. The martingale property (5.33) tells us that

$$b\,\Delta x = M_{k\Delta t},$$

and hence

$$M_{(k+1)\Delta t} - M_{k\Delta t} = \pm a\,\Delta x = a\,X_{(k+1)\Delta t}.$$

If we write $J_{(k+1)\Delta t}$ for the number a, then $J_{(k+1)\Delta t}$ is $\mathcal{F}_{k\Delta t}$-measurable, and

$$M_{k\Delta t} = M_0 + \sum_{j=1}^{k} J_{j\Delta t}\,X_{j\Delta t}.$$

This is the form of the stochastic integral with respect to random walk as in Section 1.6.

5.8   Exercises

Exercise 5.1. For each of the following random variables X_j on $\mathbb{R}$, let μ_j be the distribution considered as a probability measure on $\mathbb{R}$. For each pair state whether or not $\mu_j \ll \mu_k$. Since there are six random variables, there should be 30 answers.

$$X_1 \text{ normal mean 2, variance 7}$$

$$X_2 \text{ binomial with parameters } n = 7, p = .3$$

$$X_3 \text{ Poisson with parameter 2}$$

$$X_4 = e^{X_1}$$

$$X_5 \text{ uniform on } [0, 1]$$

$$X_6 = X_1 X_3 + X_2 \text{ assuming } X_1, X_2, X_3 \text{ independent }.$$

Exercise 5.2. Consider the martingale betting strategy from Chapter 1. Let W_n denote the winnings at time n and let

$$M_n = 1 - W_n,$$

so that at each time n, M_n equals 0 or 2^n. Note that $M_0 = 1$. Let $T = \min\{n : M_n = 0\}$. Let $\mathcal{F}_n$ be the information in $M_0, M_1, \ldots, M_n$.

1. Explain why M_n is a nonnegative martingale.

2. Define a new probability measure Q_n by saying that for any event V that is $\mathcal{F}_n$-measurable,

$$Q_n(V) = \mathbb{E}\left[M_n \, 1_V\right].$$

   Show that if $m < n$ and V is $\mathcal{F}_m$-measurable, then $Q_m(V) = Q_n(V)$.

3. Given the last part, we can write Q rather than just Q_n. Find the transition probability

$$Q\left\{M_{n+1} = 2^{n+1} \mid M_n = 2^n\right\}.$$

4. Find the Q-probability that $T < \infty$.

5. Is M_n a martingale with respect to the measure Q?

Exercise 5.3. Suppose B_t is a standard Brownian motion on $(\Omega, \mathbb{P})$. For each of the following choices of $X_t, 0 \leq t \leq 1$, state whether there is an equivalent probability measure Q such that the X_t is a standard Brownian motion in the new measure. If the answer is yes, give dQ/dP at $t = 1$. In all cases assume that $B_0 = 0, X_0 = 0$.

$$dX_t = 2\,dt + dB_t,$$

$$dX_t = 2\,dt + 6\,dB_t,$$

$$dX_t = 2B_t\,dt + dB_t.$$

Exercise 5.4. Let B_t be a standard Brownian motion with $B_0 = 0$. Let $m > 0$ and let $X_t = e^{-mB_t^2}$.

1. Find a function g such that

$$M_t := X_t \exp\left\{\int_0^t g(X_s)\,ds\right\}$$

   is a local martingale.

2. What SDE does M_t satisfy?

3. Let Q be the probability measure obtained by tilting by M_t. Find the SDE for B_t in terms of a Q-Brownian motion.

4. Explain why M_t is actually a martingale.

Exercise 5.5. Let B_t be a standard Brownian motion with $B_0 = 1$. Let $T = \min\{t : B_t = 0\}$ Let $r > 0$ and let $X_t = B_t^r$.

1. Find a function g such that

$$M_t := X_t \exp\left\{\int_0^t g(X_s)\,ds\right\}$$

   is a local martingale for $t < T$. (Do not worry about what happens after time T.)

2. What SDE does M_t satisfy?

3. Let Q be the probability measure obtained by tilting by M_t. Find the SDE for B_t in terms of a Q-Brownian motion.

Exercise 5.6. Suppose $a, m > 0$ and W_t is a Brownian motion with drift m and variance 1. Let
$$T_a = \inf\{t : W_t = a\}.$$
Find the density of T_a. Hint: look at Examples 2.7.1 and 5.2.2.

Exercise 5.7. Suppose the price of a stock S_t follows a geometric Brownian motion
$$dS_t = S_t \left[3\, dt + dB_t\right],$$
whee B_t is a standard Brownian motion with respect to a measure $\mathbb{P}$. Suppose there is a risk-free bond with rate $r = .05$. Let Q be the measure, mutually absolutely continuous with respect to $\mathbb{P}$, under which the discounted stock price $\tilde{S}_t$ is a martingale.

1. Give the SDE for $\tilde{S}_t$ in terms of W_t, a standard Brownian motion with respect to Q.

2. Suppose there is a claim at time $T = 2$ of $V = S_2^2$. Is this a contingent claim?

3. Find the discounted value $\tilde{V}_t$ of the claim for $0 \leq t \leq 2$.

4. Give the SDE for $\tilde{V}_t$ in terms of the Brownian motion W.

5. Find the portfolio (a_t, b_t) that hedges the claim.

6. Find the value V_t.

Exercise 5.8. Repeat Exercise 5.7 with $V = S_2^3$.

Exercise 5.9. Repeat Exercise 5.7 with claim
$$V = \int_0^2 s\, S_s \, ds.$$

Exercise 5.10. Repeat Exercise 5.7 with claim
$$V = \int_0^2 S_s^2 \, ds.$$

Exercise 5.11. Here we do the calculus needed to finish the derivation of the Black-Scholes formula in Example 5.5.1.

1. Show that if N has a standard normal distribution and $a \neq 0, y$ are constants, then the density of $Z = \exp\{aN + y\}$ is

$$g(z) = \frac{1}{az} \phi \left(\frac{-y + \log z}{a} \right),$$

   where ϕ denotes the density of N.

2. Show that for all x,

$$\int_x^\infty (z - x) \, g(z) \, dz = e^{y + (a^2/2)} \, \Phi \left(\frac{y - \log x + a^2}{a} \right) - x \, \Phi \left(\frac{y - \log x}{a} \right).$$

3. Use these calculations to check the details of Example 5.5.1.

Chapter 6

Jump processes

6.1 Lévy processes

The assumptions that led to Brownian motion were:

- Independent increments

- Stationary increments

- Continuous paths.

If we want to consider more general processes, we need to give up at least one of these assumptions. In this chapter we will not assume continuity of paths.

Definition A stochastic process X_t is called a *Lévy process* if it has independent, stationary increments. That is to say, for every $s, t > 0$, the random variable $X_{s+t} - X_s$ is independent of $\{X_r : r \leq s\}$ and has the same distribution as $X_t - X_0$.

Brownian motion is the only Lévy process with continuous paths. We have already discussed one example of a Lévy process with discontinuous paths, the Poisson process, and we will describe it in more detail in Section 6.2. Let us give two examples that can be derived from Brownian motion.

Example 6.1.1. Suppose B_t is a standard Brownian motion and $T_s = \inf\{t : B_t = s\}$. Using the strong Markov property, we can see that $T_s, s \geq 0$ has i.i.d. increments. Therefore it is a Lévy process. Note that it is increasing,

that is, with probability one if $r < s$, then $T_r < T_s$. We calculated the density of T_s in Example 2.7.1,

$$f_s(t) = \frac{s}{t^{3/2}\sqrt{2\pi}}\, e^{-\frac{s^2}{2t}}, \quad 0 < t < \infty.$$

In particular, the distribution of T_s is not normal. We claim that (with probability one) the function $s \mapsto T_s$ is not continuous. To see this, let $M = \max_{0 \le t \le 1} B_t$ and let $s_0 \in (0,1)$ be a time t with $B_t = M$. Then by definition of M, T_s we see that $T_M \le s_0$, but $T_s > 1$ for $s > M$, showing that T is not continuous at s_0. The scaling property of Brownian motion implies that T_s has the same distribution as $s^2\, T_1$. A tricky calculation which we omit computes the characteristic function

$$\mathbb{E}[e^{i(r/s^2)\,T_s}] = \mathbb{E}[e^{irT_1}] = \int_{-\infty}^{\infty} e^{irt}\, f_1(t)\, dt = e^{\Psi(r)},$$

where

$$\Psi(r) = \begin{cases} |2r|^{1/2}(1-i) & \text{if } r \ge 0 \\ |2r|^{1/2}(1+i) & \text{if } r \le 0 \end{cases}.$$

Example 6.1.2. Let $B_t = (B_t^1, B_t^2)$ be a standard two-dimensional Brownian motion. Let

$$T_s = \inf\{t : B_t^1 = s\},$$

and

$$X_s = B^2(T_s).$$

Using the strong Markov property, one can show that the increments are i.i.d. and similarly to Example 6.1.1 that the paths as discontinuous. The scaling property of Brownian motion implies that X_s has the same distribution as $s\, X_1$. The density of X_1 turns out to be Cauchy,

$$f(x) = \frac{1}{\pi\,(x^2 + 1)}, \quad -\infty < x < \infty,$$

with characteristic function

$$\mathbb{E}\left[e^{irX_t}\right] = \int_{-\infty}^{\infty} e^{irx}\, f(x)\, dx = e^{-|r|}.$$

Not every distribution can arise as the distribution of the increments of a Lévy process. Indeed, suppose that X_t is a Lévy process. Then we can write

$$X_1 = \sum_{j=1}^{n} Y_{j,n}, \quad \text{where } Y_{j,n} = X\left(\frac{j}{n}\right) - X\left(\frac{j-1}{n}\right).$$

Definition A random variable X has an *infinitely divisible distribution* if for each n we can find independent, identically distributed random variables $Y_{1,n}, \dots, Y_{n,n}$ such that X has the same distribution as

$$Y_{1,n} + \cdots + Y_{n,n}.$$

We just noted that if X_t is a Lévy process, then X_1 (and, in fact, X_t for each t) has an infinitely divisible distribution. The converse is true and it not too difficult to show. If F is the distribution function of an infinitely divisible distribution, then there is a Lévy process X_t such that X_1 has distribution F. Normal and Poisson random variables are infinitely divisible:

- If $X \sim N(m, \sigma^2)$, then X has the same distribution as the sum of n independent $N(m/n, \sigma^2/n)$ random variables.

- If X has a Poisson distribution with mean λ, then X has the same distribution as the sum of n independent Poisson random variables with mean λ/n.

The goal of the next few sections is to show that every infinitely divisible random variable is the sum of a normal random variable and a "generalized Poisson" or "jump" random variable. The category "generalized Poisson" is rather large and will include the distributions in Examples 6.1.1 and 6.1.2.

A precise definition of a Lévy process needs another condition. For Brownian motion, we put continuity of the paths into the assumptions. To construct Brownian motion, we first defined Brownian motion on dyadic rationals times, showed that the corresponding process was uniformly continuous, and then extended to other times by continuity. In the case of Lévy processes, we similarly can first define the process at dyadic times to satisfy independent, stationary increments. As we discuss later, we then prove that with probability one, for all times t, the limits

$$X_{t-} = \lim_{q \uparrow t} X_q, \quad X_{t+} = \lim_{q \downarrow t} X_q,$$

exist where the limits are taken over the dyadic rationals. We then define X_t to be X_{t+}.

6.2   Poisson process

The Poisson process is the basic jump Lévy process. It is obtained by taking jumps of size one at a particular rate λ. Suppose we have such a process, and let $p(s)$ denote the probability that there is at least one jump in the interval $[t, t + s]$. If the increments are to be stationary, this quantity must not depend on t. Rate λ means that the probability that there is a jump some time during the time interval $[t, t + \Delta t]$ is about $\lambda \Delta t$; more precisely,

$$p(\Delta t) = \lambda \, \Delta t + o(\Delta t), \quad \Delta t \downarrow 0.$$

We can use this observation to construct the Poisson process.

We consider the *waiting times* between jumps. Let X_t denote the number of jumps that have occurred by time t and let

$$T = \inf\{t : X_t = 1\}$$

denote the amount of time until the first jump. Using i.i.d. increments, we see that

$$
\begin{aligned}
\mathbb{P}\{T > t\} &= \prod_{j=1}^{n} \mathbb{P}\left\{\text{no jump during } \left[\frac{(j-1)t}{n}, \frac{jt}{n}\right]\right\} \\
&= \left[1 - p\left(\frac{t}{n}\right)\right]^n.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathbb{P}\{T > t\} &= \lim_{n \to \infty} \left[1 - p\left(\frac{t}{n}\right)\right]^n \\
&= \lim_{n \to \infty} \left[1 - \frac{\lambda t}{n} + o\left(\frac{\lambda t}{n}\right)\right]^n = e^{-\lambda t}.
\end{aligned}
$$

Recall that a random variable T has an *exponential distribution* with rate λ if it has density

$$f(t) = \lambda \, e^{-\lambda t}, \quad 0 < t < \infty,$$

and hence
$$\mathbb{P}\{T > t\} = \int_t^\infty f(s)\,ds = e^{-\lambda t}.$$

Our assumptions imply that the waiting times of a Poisson distribution with parameter λ must be exponential with rate λ. Note that

$$\mathbb{E}[T] = \int_0^\infty t\,f(t)\,dt = \frac{1}{\lambda},$$

so that the mean waiting time is (quite reasonably) the reciprocal of the rate. This observation gives a way to construct a Poisson process. This construction also gives a good way to simulate Poisson processes (see Exercise 6.1).

- Let $T_1, T_2, T_3, \ldots$ be independent random variables each exponential with rate λ.

- Let $\tau_0 = 0$ and for positive integer n,

$$\tau_n = T_1 + \cdots + T_n.$$

   In other words, τ_n is the time at which the nth jump occurs.

- Set
$$X_t = n \quad \text{for} \quad \tau_n \le t < \tau_{n+1}.$$

Note that we have defined the process so that the paths are right-continuous,

$$X_t = X_{t+} := \lim_{s \downarrow t} X_s.$$

The paths also have limits from the left, that is, for every t the limit

$$X_{t-} = \lim_{s \uparrow t} X_s$$

exists. If t is a time that the process jumps, that is, if $t = \tau_n$ for some $n > 0$, then
$$X_t = X_{t+} = X_{t-} + 1.$$

At all other times the path is continuous, $X_{t-} = X_{t+}$.

The i.i.d. increments follow from the construction. We can use the assumptions to show the following.

- The random variable $X_{t+s} - X_s$ has a Poisson distribution with mean λt, that is,

$$\mathbb{P}\{X_{t+s} - X_s = k\} = e^{-\lambda t}\frac{(\lambda t)^k}{k!}.$$

One way to derive this is to write a system of differential equations for the functions

$$q_k(t) = \mathbb{P}\{X_t = k\}.$$

In the small time interval $[t, t + \Delta t]$, the chance that there is more than one jump is $o(\Delta t)$ and the chance there is exactly one jump is $\lambda \Delta t + o(\Delta t)$. Therefore, up to errors that are $o(\Delta t)$,

$$\mathbb{P}\{X_{t+\Delta t} = k\} = \mathbb{P}\{X_t = k - 1\}(\lambda \Delta t) + \mathbb{P}\{X_t = k\}[1 - \lambda \Delta t].$$

This gives

$$q_k(t + \Delta t) - q_k(t) = \lambda \Delta t \ [q_{k-1}(t) - q_k(t)] + o(\Delta t),$$

or

$$\frac{dq_k(t)}{dt} = \lambda [q_{k-1}(t) - q_k(t)].$$

If we assume $X_0 = 0$, we also have the initial conditions $q_0(0) = 1$ and $q_k(0) = 0$ for $k > 0$. We can solve this system of equations recursively, and this yields the solutions

$$q_k(t) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}.$$

(Although it takes some good guesswork to start with the equations and find $q_k(t)$, it is easy to verify that $q_k(t)$ as given above satisfies the equations.)

When studying infinitely divisible distributions, it will be useful to consider characteristic functions, and for notational ease, we will take logarithms. Since the characteristic function is complex-valued, we take a little care in defining the logarithm.

Definition If X is a random variable, then its *characteristic exponent* $\Psi(s) = \Psi_X(s)$ is defined to be the continuous function satisfying

$$\mathbb{E}[e^{isX}] = e^{\Psi(s)}, \quad \Psi(0) = 0.$$

If X_t is a Lévy process, then the characteristic exponent of the process is the characteristic exponent of $X_1 - X_0$.

Note that if X, Y are independent, then

$$\Psi_{X+Y} = \Psi_X + \Psi_Y. \tag{6.1}$$

- If $X \sim N(m, \sigma^2)$, then

$$\Psi(s) = ims - \frac{\sigma^2}{2} s^2.$$

- If X is Poisson with mean λ, then

$$\Psi(s) = \lambda\left[e^{is} - 1\right].$$

This can be seen from the computation

$$\mathbb{E}[e^{isX}] = \sum_{n=0}^{\infty} e^{isn}\, \mathbb{P}\{X = n\} = \sum_{n=0}^{\infty} \frac{(\lambda e^{is})^n}{n!}\, e^{-\lambda} = e^{\lambda e^{is}}\, e^{-\lambda}.$$

If X_t is a Lévy process starting at the origin with characteristic exponent Ψ, then i.i.d. increments and the relation (6.1), imply that

$$\Psi_{X_t} = t\,\Psi_{X_1} = t\,\Psi.$$

An important property of Lévy processes is the following.

- Suppose X_t^1, X_t^2 are independent Lévy processes with $\Psi_{X^1} = \Psi_1, \Psi_{X^2} = \Psi_2$. Then $X_t = X_t^1 + X_t^2$ is a Lévy process with $\Psi_{X_t} = \Psi_1 + \Psi_2$.

For example if X_t^1 is a Brownian motion with drift m and variance σ^2 and X_t^2 is an independent Poisson process with rate λ, then $X_t = X_t^1 + X_t^2$ is a Lévy process with

$$\Psi(s) = ims - \frac{\sigma^2 s^2}{2} + \lambda\left[e^{is} - 1\right]. \tag{6.2}$$

A Lévy process is a Markov process so we can talk about its generator,

$$Lf(x) = \lim_{t \downarrow 0} \frac{\mathbb{E}[f(X_t) \mid X_0 = x] - f(x)}{t}.$$

Recall that for a Brownian motion B_t with drift m and variance σ^2,

$$Lf(x) = m\,f'(x) + \frac{\sigma^2}{2}\, f''(x).$$

Moreover, if

$$f(t, x) = \mathbb{E}\left[F(B_t) \mid B_0 = x\right],$$

then f satisfies the heat equation

$$\partial_t f(t, x) = L_x f(t, x). \tag{6.3}$$

We now compute the generator for the Poisson process. Although we think of the Poisson process as taking integer values, there is no problem extending the definition so that $X_0 = x$. In this case the values taken by the process are $x, x + 1, x + 2, \ldots$ Up to terms that are $o(t)$, $\mathbb{P}\{X_t = X_0 + 1\} = 1 - \mathbb{P}\{X_t = X_0\} = \lambda t$. Therefore as $t \downarrow 0$,

$$\mathbb{E}[f(X_t) \mid X_0 = x] = \lambda t\, f(x + 1) + [1 - \lambda t]\, f(x) + o(t),$$

and

$$Lf(x) = \lambda\left[f(x + 1) - f(x)\right].$$

The same argument shows that if $f(t, x)$ is defined by

$$f(t, x) = \mathbb{E}\left[F(X_t) \mid X_0 = x\right],$$

then f satisfies the heat equation (6.3) with the generator L. We can view the generator as the operator on functions f that makes (6.3) hold.

The generators satisfy the following linearity property: if X_t^1, X_t^2 are independent Lévy processes with generators L_1, L_2, respectively, then $X_t = X_t^1 + X_t^2$ has generator $L_1 + L_2$. For example, if X is the Lévy process with Ψ as in (6.2),

$$Lf(x) = m\, f'(x) + \frac{\sigma^2}{2}\, f''(x) + \lambda\left[f(x + 1) - f(x)\right].$$

6.3   Compound Poisson process

In the Poisson process, the process jumps from x to $x + 1$ at rate λ. The compound Poisson process generalizes this by allowing the jump size to be random. For the moment we think of the process as having two parameters, a jump rate λ and a distribution function F. We construct the process as follows.

- Let $T_1, T_2, T_3, \ldots$ be independent random variables each exponential with rate λ.

- Let N_t denote the corresponding Poisson process as in the previous section,

$$N_t = n \quad \text{if} \quad T_1 + \cdots + T_n \leq t < T_1 + \cdots + T_{n+1}.$$

- Let $Y_1, Y_2, Y_3, \ldots$ be independent random variables, independent of $T_1, T_2, \ldots$, with distribution function F, and let

$$S_n = Y_1 + \cdots + Y_n, \quad S_0 = 0.$$

- Set $X_t = S_{N_t}$.

We call the process X_t a *compound Poisson process (CPP)* starting at the origin.

Let $\mu^{\#}$ denote the distribution of the random variables Y_j, that is, if $V \subset \mathbb{R}$, then $\mu^{\#}(V) = \mathbb{P}\{Y_1 \in V\}$. We set

$$\mu = \lambda \mu^{\#}$$

which is a measure on $\mathbb{R}$ with $\mu(\mathbb{R}) = \lambda$. For the usual Poisson process $\mu^{\#}, \mu$ are just "point masses" on the point 1,

$$\mu^{\#}(\{1\}) = 1, \quad \mu(\{1\}) = \lambda, \quad \mu^{\#}(\mathbb{R} \setminus \{1\}) = \mu(\mathbb{R} \setminus \{1\}) = 0.$$

The measure μ encodes both λ and $\mu^{\#}$ so we can consider μ as the parameter for the compound Poisson process. For any measure μ with $\lambda = \mu(\mathbb{R}) < \infty$ there is such a process. The measure μ is called the *Lévy measure* for the process. Without loss of generality we assume that $\mu(\{0\}) = 0$ since "jumps of size zero" do not affect the process X_t. The construction shows that the increments of X_t are i.i.d., and hence X_t is a Lévy process.

Proposition 6.3.1. *Suppose X_t is a CPP with Lévy measure μ with $X_0 = 0$. Then*

$$\Psi(s) = \int_{-\infty}^{\infty} [e^{isx} - 1] \, d\mu(x),$$

$$Lf(x) = \int_{-\infty}^{\infty} [f(x+y) - f(x)] \, d\mu(y). \tag{6.4}$$

Moreover, if

$$\sigma^2 := \int x^2 \, d\mu(x) < \infty, \tag{6.5}$$

and

$$m = \int x \, d\mu(x),$$

then $M_t = X_t - tm$ is a square integrable martingale with $\mathrm{Var}[M_t] = t\sigma^2$.

Proof. Let

$$\phi(s) = \mathbb{E}[e^{isY_j}] = \int_{-\infty}^{\infty} e^{isx} \, d\mu^{\#}(x),$$

denote the characteristic function of a random variable with distribution function F. Then,

$$\mathbb{E}[e^{isX_t}] = \sum_{n=0}^{\infty} \mathbb{P}\{N_t = n\} \, \mathbb{E}[e^{isX_t} \mid N_t = n].$$

Conditioned on $N_t = n$, the distribution of X_t is that of $Y_1 + \cdots + Y_n$, and hence

$$\mathbb{E}[e^{isX_t} \mid N_t = n] = \phi(s)^n.$$

Since N_t is Poisson with mean $t\lambda$,

$$
\begin{aligned}
\mathbb{E}[e^{isX_t}] &= \sum_{n=0}^{\infty} e^{-t\lambda} \frac{(t\lambda)^n}{n!} \phi(s)^n \\
&= e^{-t\lambda} \sum_{n=0}^{\infty} \frac{(t\lambda\phi(s))^n}{n!} \\
&= \exp\{t\lambda[\phi(s) - 1]\} \\
&= \exp\left\{t\lambda \int_{-\infty}^{\infty} [e^{isx} - 1] \, d\mu^{\#}(x)\right\} \\
&= \exp\left\{t \int_{-\infty}^{\infty} [e^{isx} - 1] \, d\mu(x)\right\}.
\end{aligned}
$$

The second-to-last equality uses the fact that $\mu^{\#}$ is a probability measure.

The computation for the generator L follows from the definition of the CPP. In a small time Δt, the probability that there is a jump is $\lambda \Delta t + o(\Delta t)$, and given that there is a jump, the amount jumped is given by the distribution $\mu^{\#}$. In other words, the probability that there is a jump in time Δt whose size is in (a, b) is given up to an error of size $o(\Delta t)$ by

$$\lambda \, \Delta t \, \mu^{\#}(a, b) = \Delta t \, \mu(a, b).$$

Therefore,

$$\mathbb{E}\left[f(X_{\Delta t}) \mid X_0 = x\right]$$

$$= \left[1 - \lambda \Delta t\right] f(x) + \Delta t \int_{-\infty}^{\infty} f(x+y)\, d\mu(y) + o(\Delta t)$$

$$= f(x) + \Delta t \int_{-\infty}^{\infty} \left[f(x+y) - f(x)\right] d\mu(y) + o(\Delta t),$$

which is a restatement of (6.4).

The moments of X_t can be computed by differentiating its characteristic function,

$$\phi'_{X_t}(0) = i\,\mathbb{E}[X_t], \quad \phi''_{X_t}(0) = -\,\mathbb{E}[X_t^2],$$

assuming that the moments exist. For any positive integer n, if

$$\int_{-\infty}^{\infty} |x|^n\, d\mu(x) < \infty,$$

then we can differentiate under the integral n times to see that

$$\Psi^{(n)}(s) = i^n \int_{-\infty}^{\infty} x^n\, e^{isx}\, d\mu(x).$$

In particular, if (6.5) holds,

$$\Psi'(0) = i \int_{-\infty}^{\infty} x\, d\mu(x), \quad \Psi''(0) = -\int_{-\infty}^{\infty} x^2\, d\mu(x),$$

and since $\phi_{X_t}(s) = \exp\{t\Psi(s)\}$ and $\Psi(0) = 0$, we get

$$\phi'_{X_t}(0) = t\Psi'(0) = it \int_{-\infty}^{\infty} x\, d\mu(x) = itm,$$

$$\phi''_{X_t}(0) = t\Psi''(0) + [t\Psi'(0)]^2 = -[t\sigma^2 + (tm)^2].$$

Therefore,

$$\mathbb{E}[X_t] = tm, \quad \mathrm{Var}[X_t] = \mathbb{E}\left[(X_t - tm)^2\right] = \mathbb{E}\left[X_t^2\right] - (\mathbb{E}[X_t])^2 = t\sigma^2.$$

If $\mathcal{F}_s$ denote the information contained in $\{X_r : r \leq s\}$ and $s < t$, then $X_t - X_s$ is independent of $\mathcal{F}_s$ and

$$E[X_t \mid \mathcal{F}_s] = X_s + E[X_t - X_s \mid \mathcal{F}_s] = X_s + \mathbb{E}[X_t - X_s] = X_s + (t-s)m,$$

and hence

$$\mathbb{E}[X_t - tm \mid \mathcal{F}_s] = X_s - sm.$$

This shows that M_t is a martingale and

$$\mathbb{E}[M_t^2] = \mathbb{E}[(X_t - tm)^2] = \mathrm{Var}[X_t] = t\,\sigma^2.$$

$\square$

We call $M_t = X_t - mt$ the *compensated compound Poisson process (compensated CPP)* associated to X_t. It is a Lévy process. If L denotes the generator for X_t, then M_t has generator

$$
\begin{aligned}
L_M f(x) &= Lf(x) - mf'(x) \\
&= \int_{-\infty}^{\infty} [f(x+y) - f(x) - y\,f'(x)]\ d\mu(y).
\end{aligned}
\tag{6.6}
$$

The quadratic variation of the CPP is defined as before

$$\langle X \rangle_t = \lim_{n\to\infty} \sum_{j \le tn} \left[X\left(\frac{j}{n}\right) - X\left(\frac{j-1}{n}\right) \right]^2.$$

Note that the terms in the sum are zero unless there is a jump in the time interval $[(j-1)/n, j/n]$. Hence we see that

$$\langle X \rangle_t = \sum_{s \le t} [X_s - X_{s-}]^2.$$

Unlike the case of Brownian motion, for fixed t the random variable $\langle X \rangle_t$ is not constant. We can similarly find the quadratic variation of the martingale $M_t = X_t - mt$. By expanding the square, we see that this it is the limit as $n \to \infty$ of three sums

$$\sum_{j \le tn} \left[X\left(\frac{j}{n}\right) - X\left(\frac{j-1}{n}\right) \right]^2$$

$$+ \frac{2m}{n} \sum_{j \le tn} \left[X\left(\frac{j}{n}\right) - X\left(\frac{j-1}{n}\right) \right] + \sum_{j \le tn} \frac{m^2}{n^2}.$$

Since there are only finitely many jumps, we can see that the second and third limits are zero and hence

$$\langle M \rangle_t = \langle X \rangle_t = \sum_{s \le t} [X_s - X_{s-}]^2.$$

This next proposition generalizes the last assertion of the previous proposition and sheds light on the meaning of the generator L. In some sense, this is an analogue of Itô's formula for CPP. Recall that if X_t is a diffusion satisfying

$$dX_t = m(X_t)\,dt + \sigma(X_t)\,dB_t,$$

then the Markov process X_t has generator

$$Lf(x) = m(x)\,f'(x) + \frac{\sigma^2(x)}{2}\,f''(x),$$

and Itô's formula gives

$$df(X_t) = Lf(X_t)\,dt + f'(X_t)\,\sigma(X_t)\,dB_t.$$

In other words, if

$$M_t = f(X_t) - \int_0^t Lf(X_s)\,ds,$$

then M_t a (local) martingale satisfying

$$dM_t = f'(X_t)\,\sigma(X_t)\,dB_t.$$

Proposition 6.3.2. *Suppose X_t is a compound Poisson process with Lévy measure μ, and suppose that f is a continuous function such that for all x, t,*

$$\mathbb{E}\left[f(X_t)^2 \mid X_0 = x\right] < \infty.$$

Then

$$M_t = f(X_t) - \int_0^t Lf(X_s)\,ds$$

is a square integrable martingale with

$$\langle M \rangle_t = \langle f(X) \rangle_t = \sum_{s \le t} \left[f(X_s) - f(X_{s-})\right]^2.$$

The proof of this proposition is similar to the derivation of Itô's formula. We will first show that M_t is a martingale, that is, $E[M_t \mid \mathcal{F}_s] = M_s$. The argument is essentially the same for all s so we will assume that $s = 0$ and $X_0 = x$. We need to show that

$$\mathbb{E}\left[f(X_t) - f(X_0) - \int_0^t Lf(X_s)\,ds\right] = 0.$$

This argument is the same for all t, so let us assume $t = 1$. Then, as in the proof of Itô's formula, we write

$$f(X_1) - f(X_0) - \int_0^1 Lf(X_s)\,ds$$

$$= \sum_{j=1}^n \left[f\left(X_{\frac{j}{n}}\right) - f\left(X_{\frac{(j-1)}{n}}\right) - \int_{(j-1)/n}^{j/n} Lf(X_s)\,ds\right].$$

We write the expectation of the right-hand side as the sum of two terms

$$\sum_{j=1}^n \mathbb{E}\left[f\left(X_{\frac{j}{n}}\right) - f\left(X_{\frac{(j-1)}{n}}\right) - \frac{1}{n}Lf\left(X_{\frac{(j-1)}{n}}\right)\right] \qquad (6.7)$$

$$\sum_{j=1}^n \mathbb{E}\left[\frac{1}{n}Lf\left(X_{\frac{(j-1)}{n}}\right) - \int_{(j-1)/n}^{j/n} Lf(X_s)\,ds\right], \qquad (6.8)$$

The definition of L implies that

$$E\left[f(X_{t+\Delta t}) \mid \mathcal{F}_t\right] = f(X_t) + Lf(X_t)\,\Delta t + o(\Delta t),$$

and hence

$$\mathbb{E}\left[f(X_{t+\Delta t}) - f(X_t) - Lf(X_t)\,\Delta t\right] = o(\Delta t).$$

This shows that the sum in (6.7) has n terms that are $o(1/n)$ and hence the limit is zero. The terms inside the expectation in (6.8) equal zero unless there is a jump between time $(j-1)/n$ and j/n. This occurs with probability $O(1/n)$ and in this case the value of the random variable is $O(1/n)$. Hence the expectations are $O(1/n^2)$ and the sum of n of them has limit zero.

The computation of the quadratic variation is essentially the same as in Proposition 6.3.1.

Suppose that $X_t = S_{N_t}$ where N_t is a Poisson process of rate λ. Let E_n be the event that $N_{t+(1/n)} - N_t \geq 2$ for some $t \leq 1$. We know that $\cap_n E_n$ has probability zero, and hence since $\mathbb{E}[|f(X_1)|] < \infty$,

$$\lim_{n \to \infty} \mathbb{E}\left[|f(X_1)| \, 1_{E_n}\right] = 0.$$

Therefore,

$$\mathbb{E}[f(X_1)] = \lim_{n \to \infty} \mathbb{E}[f(X_1) 1_{E_n^c}].$$

However,

$$\mathbb{E}\left[[f(X_{t+s}) - f(X_t)] \, 1_{E_n^c}\right] = s \, \mathbb{E}[Lf(X_t)][1 + O(s)].$$

We therefore get

$$\mathbb{E}\left[f(X_1) - f(X_0)\right] = \lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} \mathbb{E}[Lf(X_{j/n})].$$

Right-continuity implies that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} Lf(X_{j/n}) = \int_0^1 Lf(X_s) \, ds,$$

and the bound $\mathbb{E}[|f(X_1)|] < \infty$ can be used to justify the interchange of limit and expectation.

For a CPP the paths $t \mapsto X_t$ are piecewise constant and are discontinuous at the jumps. As for the usual Poisson process, we have defined the path so that is it right-continuous and has left-limits. We call a function *cadlag* (also written *càdlàg*), short for *continue à droite, limite à gauche*, if the paths are right-continuous everywhere and have left-limits. That is, for every t, the limits

$$X_{t+} = \lim_{s \downarrow t} X_s, \quad X_{t-} = \lim_{s \uparrow t} X_s,$$

exist and $X_t = X_{t+}$. The paths of a CPP are cadlag. We can write

$$X_t = X_0 + \sum_{0 \leq s \leq t} [X_s - X_{s-}].$$

Although as written this sum is over an uncountable number of times s, the term $X_s - X_{s-}$ is only nonzero at those s at which the path is discontinuous and this is a finite set.

In the next section, we will use the following maximal lemma.

Lemma 6.3.3. *If X_t is a CPP satisfying* (6.5), *$M_t = X_t - tm$, and*

$$K = \max_{0 \leq t \leq 1} |M_t|,$$

then

$$\mathbb{P}\{|K| \geq a\} \leq \frac{\sigma^2}{a^2}.$$

Proof. Let

$$K_n = \max\{|M_{j/n}| : j = 1, 2, \ldots, n\}.$$

Since X_t has piecewise constant paths, $K = \lim_{n \to \infty} K_n$ and it suffices to show for each n that

$$\mathbb{P}\{|K_n| \geq a\} \leq \frac{\sigma^2}{a^2}.$$

Fix n and let $Z_j = M_{j/n}, \overline{Z}_n = \max\{|Z_j| : j = 1, \ldots, n\}$. Then Z_j is a discrete-time martingale, and Corollary 1.7.2 gives

$$\mathbb{P}\{\overline{Z}_n \geq a\} \leq \frac{\mathbb{E}[Z_n^2]}{a^2} = \frac{\mathbb{E}[M_1^2]}{a^2} = \frac{\sigma^2}{a^2}.$$

$\square$

6.4   Integration with respect to compound Poisson processes

Defining the integral with respect to a compound Poisson process is easy. If X_t is a CPP and A_t is another process, we could let

$$\int_0^t A_s \, dX_s = \sum_{0 \leq s \leq t} A_s \left[X_s - X_{s-} \right]. \tag{6.9}$$

As noted before, there are only a finite number of nonzero terms in the sum so the sum is well defined. This definition requires no assumptions on the process A_s. However, if we want the integral to satisfy some of the properties of the Itô integral, we will need to assume more.

Suppose that $\mathbb{E}[X_1] = m, \text{Var}[X_1] = \sigma^2$, and let M_t be the square integrable martingale $M_t = X_t - mt$. Then if the paths of A_t are Riemann integrable, the integral

$$Z_t = \int_0^t A_s \, dM_s = \int_0^t A_s \, dX_s - m \int_0^t A_s \, ds$$

is well defined. Let $\mathcal{F}_t$ denote the information contained in $\{M_s : s \leq t\}$ which is the same as the information contained in $\{X_s : s \leq t\}$. In analogy with the Itô integral, we might hope that if A_t is square integrable, piecewise continuous, and adapted to $\{\mathcal{F}_t\}$, then Z_t would be a martingale. However, this is not always the case as we now show.

Example 6.4.1. Suppose X_t is the CPP that takes jumps of rate 1 and when it jumps it chooses ± 1 each with the same probability. In other words, the Lévy measure of X is the probability measure with $\mu(\{1\}) = \mu(\{-1\}) = 1/2$. Then $m = 0, \sigma^2 = 1$, and $A_t = X_t - X_{t-}$ is adapted to $\{\mathcal{F}_t\}$. It is discontinuous only at the jumps of X_t and hence is piecewise continuous. However, if we define the integral as in (6.9), then

$$\int_0^t A_s \, dX_s = \sum_{0 \leq s \leq t} [X_s - X_{s-}]^2,$$

which has positive expectation (since it is nonnegative and has a positive probability to be strictly positive).

The problem in our setup is that we allow a betting strategy that sees a jump at time s and immediately changes the bet to take advantage of this. In our frameweork, we will not allow these instantaneous changes by restricting to strategies that are *left*-continuous.

Proposition 6.4.1. *Suppose X_t is a CPP with $\sigma^2 < \infty$ and $M_t = X_t - mt$. Suppose that A_t is a process satisfying:*

- *each A_t is measurable with respect to $\{X_s : 0 \leq s \leq t\}$;*

- *with probability one, the paths of A_t are left-continuous, that is, $A_t = A_{t-}$;*

- *for every $t < \infty$,*

$$\int_0^t \mathbb{E}[A_s^2] \, ds < \infty.$$

Then

$$Z_t = \int_0^t A_s \, dM_s$$

is a square integrable martingale with

$$\mathbb{E}\left[Z_t^2\right] = \sigma^2 \int_0^t \mathbb{E}\left[A_s^2\right] \, ds. \qquad (6.10)$$

Rather than assuming that A_s is left-continuous, we can assume that A_s has cadlag paths, but then we change the definition in (6.9) to

$$\int_0^t A_s \, dX_s = \sum_{0 \le s \le t} A_{s-} \left[X_s - X_{s-}\right].$$

Proof of Proposition 6.4.1. If $s < t$, then arguing as before,

$$E[A_s \left(M_t - M_s\right) \mid \mathcal{F}_s] = A_s \, E[M_t - M_s \mid \mathcal{F}_s] = 0, \qquad (6.11)$$

$$
\begin{aligned}
\mathbb{E}[A_s^2 \left(M_t - M_s\right)^2] &= \mathbb{E}\left[E(A_s^2 \left(M_t - M_s\right)^2 \mid \mathcal{F}_s)\right] \\
&= \mathbb{E}\left[A_s^2 \, \mathbb{E}[(M_t - M_s)^2]\right] \\
&= \sigma^2 \left(t - s\right) \mathbb{E}[A_s^2]. \qquad (6.12)
\end{aligned}
$$

For each t, the probability that $M_{t-} \ne M_t = 0$. Therefore, with probability one, for all rational times t, $M_t = M_{t-}$. Suppose that A_s is a simple process that changes values only at times in $\{0, 1/n, 2/n, 3/n, \ldots\}$. Then with probability one, for all t, if $j/n \le t < (j+1)/n$,

$$Z_t = \int_0^t A_s \, dM_s = \sum_{i=0}^{j-1} A\left(\frac{i}{n}\right) \left[M\left(\frac{i+1}{n}\right) - M\left(\frac{i}{n}\right)\right].$$

This expression uses the fact $M_t = M_{t-}$ for all rational t. Using this expression and (6.11) and (6.12) we see that Z_t is a square integrable martingale satisfying (6.10).

We now assume that A_t is adapted, left-continuous, and uniformly bounded, that is, with probability one, $|A_t| \le C$ for all t. Define A_t^n by

$$A_t^n = A_{j/n} \quad \text{if} \quad \frac{j}{n} < t \le \frac{j+1}{n}.$$

Left-continuity implies that with probability one for all t

$$A_t = \lim_{n\to\infty} A_t^n,$$

and hence

$$\int_0^t A_s \, dM_s = \lim_{n\to\infty} \int_0^t A_s^n \, dM_s.$$

Using uniform boundedness, one can interchange limits and expectations to get (6.11) and (6.12).

Finally to remove the boundedness assumption, we let $T_n = \inf\{t : |A_t| = n\}$ and let $A_{t,n} = A_{t\wedge T_n}$. As $n \to \infty$, $A_{t,n} \to A_t$, and we can argue as before. $\qquad\square$

Example 6.4.2. Suppose X_t is a CPP with Lévy measure μ satisfying

$$\int_{-\infty}^{\infty} e^{2y} \, d\mu(y) < \infty.$$

Let $f(x) = e^x$ and $S_t = f(X_t) = e^{X_t}$ which we can consider as a simple model of an asset price with jumps. This is an analogue of geometric Brownian motion for jump processes. Note that

$$f(x + y) - f(x) = e^x \, h(y) \quad \text{where} \quad h(y) = e^y - 1,$$

and hence

$$Lf(x) = \int_{-\infty}^{\infty} [f(x + y) - f(x)] \, d\mu(y) = re^x,$$

where

$$r = \int_{-\infty}^{\infty} h(y) \, d\mu(y) < \infty.$$

Note that

$$S_t - S_{t-} = S_{t-} \, h(X_t - X_{t-}). \tag{6.13}$$

In particular, the jump times for S are the same as the jump times for X. Let

$$\hat{X}_t = \sum_{s\le t} h(X_t - X_{t-})$$

be the CPP with $\hat{X}_0 = 0$ that takes a jump of size $h(y)$ whenever X_t takes a jump of size y. This process has Lévy measure $\hat{\mu}$ where

$$\hat{\mu}(h(V)) = \mu(V), \tag{6.14}$$

with mean $\hat{m} = r$. If $\hat{Y}_t = \hat{X}_t - r\,t$ is the compensated CPP, we can write (6.13) as

$$dS_t = S_{t-}\,d\hat{X}_t = S_{t-}\,d\hat{Y}_t + r\,S_{t-}\,dt = S_{t-}\left[d\hat{Y}_t + r\,dt\right].$$

Since S_t is piecewise continuous, it does not matter in the dt integral whether we write S_t or S_{t-}. However, we must write S_{t-} in the $d\hat{X}_t$ and $d\hat{Y}_t$ integrals.

If $r = 0$, that is, if

$$\int_{-\infty}^{\infty} [e^y - 1]\,d\mu(y) = 0,$$

then $\hat{X}_t$ and hence S_t are actually martingales. If $r \neq 0$, we let $\tilde{S}_t = e^{-rt}\,S_t = e^{X_t - rt}$. Since $t \mapsto e^{-rt}$ is differentiable, we can use the product rule

$$d\tilde{S}_t = S_t\,d[e^{-rt}] + e^{-rt}\,dS_t = \tilde{S}_{t-}\,d\hat{Y}_t.$$

Example 6.4.3. We will proceed backwards through the last example to solve the exponential differential equation for compound Poisson processes. Suppose $\hat{X}_t$ is a CPP whose Lévy measure $\hat{\mu}$ satisfies $\hat{\mu}\{x : x \leq -1\} = 0$ and

$$\int_{-1}^{\infty} x^2\,d\hat{\mu}(x) < \infty.$$

Let

$$\hat{m} = \int_{-1}^{\infty} x\,d\hat{\mu}(x) < \infty,$$

and let $\hat{Y}_t = \hat{X}_t - \hat{m}\,t$ be the compensated CPP which is a square integrable martingale. Let μ be the measure as defined in (6.14) and let X_t denote the corresponding process as in the previous example. In other words, if $\hat{X}_t$ has a jump of size y, then X_t has a jump of size $h^{-1}(y) = \log[1 + y]$. Then the solutions to the exponential differential equation

$$dZ_t = Z_{t-}\,d\hat{X}_t, \quad dM_t = M_{t-}\,d\hat{Y}_t,$$

are

$$Z_t = Z_0\,e^{X_t}, \quad M_t = Z_0\,e^{X_t - \hat{m}t}.$$

Note that we can write

$$Z_t = Z_0 \exp \left\{ \sum_{0 < s \le t} \log[1 + (\hat{X}_s - \hat{X}_{s-})] \right\}.$$

Example 6.4.4. More generally, the equations

$$dZ_t = A_{t-} \, Z_{t-} \, dX_t, \quad dM_t = A_{t-} \, M_{t-} \, dY_t$$

where X is a CPP with compensated process $Y_t = X_t - mt$, have solutions

$$Z_t = Z_0 \exp \left\{ \sum_{0 < s \le t} \log[1 + A_{s-} \, (X_s - X_{s-})] \right\},$$

$$M_t = Z_t \exp \left\{ -m \int_0^t A_{s-} \, ds \right\} = Z_t \exp \left\{ -m \int_0^t A_s \, ds \right\}.$$

6.5 Change of measure

Suppose X_t is a CPP defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with Lévy measure μ and corresponding filtration $\{\mathcal{F}_t\}$. We would like to do the analogue of the Girsanov theorem to change the measure.

Suppose that strictly positive martingale M_t is a strictly positive martingale with respect to $\{\mathcal{F}_t\}$ with $M_0 = 1$. As before, we can define a new probability measure Q by saying that if V is $\mathcal{F}_t$-measurable, then

$$Q[V] = \mathbb{E} \left[M_t \, 1_V \right].$$

If we let $\mathbb{P}_t, Q_t$ denote $\mathbb{P}, Q$ restricted to $\mathcal{F}_t$, then $\mathbb{P}_t$ and Q_t are mutually absolutely continuous with

$$\frac{dQ_t}{d\mathbb{P}_t} = M_t, \quad \frac{d\mathbb{P}_t}{dQ_t} = \frac{1}{M_t}.$$

If ν is a another measure on $\mathbb{R}$, we can ask if we can find a martingale so that under the measure Q, X_t is a CPP with Lévy measure ν. A necessary condition for this to hold is that μ and ν are mutually absolutely continuous measures on $\mathbb{R}$. Indeed, if V is a subset of $\mathbb{R}$ with $\mu(V) > 0, \nu(V) = 0$, then X_t under $\mathbb{P}$ has positive probability of having $X_s - X_{s-} \in V$ for some $s \le 1$, while this has zero probability under Q.

Maybe a bit surprising is the fact that the converse holds. Suppose ν, μ are mutually absolutely continuous and let $f = d\nu/d\mu$ which we assume is strictly positive. With respect to $\mathbb{P}$, X_t satisfies

$$\mathbb{P}\{X_{t+\Delta t} - X_t \in V\} = \mu(V)\,\Delta t + o(\Delta t),$$

and we would like

$$Q\{X_{t+\Delta t} - X_t \in V\} = \nu(V)\,\Delta t + o(\Delta t),$$

which would give

$$\frac{Q\{X_{t+\Delta t} - X_t \in V\}}{\mathbb{P}\{X_{t+\Delta t} - X_t \in V\}} = \frac{\nu(V)}{\mu(V)} + o(\Delta t).$$

To get this we choose

$$M_t = \exp\left\{\sum_{s \leq t} f(X_s - X_{s-}) - rt\right\}, \quad r = \nu(\mathbb{R}) - \mu(\mathbb{R}).$$

Suppose we model an asset price by $S_t = e^{X_t}$ where X_t is a CPP, and suppose that ν is any measure equivalent to μ with

$$\int_{-\infty}^{\infty} [e^y - 1]\,d\nu(y) = 0.$$

Then Q is equivalent to $\mathbb{P}$, and (see Example 6.4.2) S is a martingale with respect to Q. In particular, there are many different measures Q that are equivalent to $\mathbb{P}$ that make S_t a martingale. This is in contrast to the case when X_t is a Brownian motion where such a measure is unique.

6.6   Generalized Poisson processes I

The generalized Poisson process is like the CPP except that the jump times form a dense subset of the real line. Suppose that μ is a measure on $\mathbb{R}$. We want to consider the process X_t that, roughly speaking, satisfies the condition that the probability of a jump whose size lies in $[a, b]$ occurring in time Δt is about $\mu[a, b]\,\Delta t$. The CPP assumes that

$$\mu(\mathbb{R}) = \int_{-\infty}^{\infty} d\mu(x) < \infty,$$

and this implies that there are only a finite number of jumps in a bounded time interval. In this section we allow an infinite number of jumps ($\mu(\mathbb{R}) = \infty$), but require the expected sum of the absolute values of the jumps in an interval to be finite. This translates to the condition

$$\int_{-\infty}^{\infty} |x| \, d\mu(x) < \infty. \tag{6.15}$$

This implies that

$$\mu\{x : |x| > \epsilon\} < \infty \qquad \text{for every} \qquad \epsilon > 0, \tag{6.16}$$

but it is possible for $\mu\{x : |x| > 0\}$ to be infinite.

To be more specific, let μ_ϵ denote the measure μ restricted to jumps of absolute value strictly greater than ϵ. For each ϵ, we have a compound process with Lévy measure μ_ϵ which can be written as

$$X_t^\epsilon = \sum_{0 \le s \le t} (X_s^\epsilon - X_{s-}^\epsilon).$$

Let us write

$$V_t^\epsilon = \sum_{0 \le s \le t} \left| X_s^\epsilon - X_{s-}^\epsilon \right|,$$

which is a CPP with measure $\bar{\mu}_\epsilon$ defined by

$$\bar{\mu}_\epsilon(a, b) = \mu_\epsilon(a, b) + \mu_\epsilon(-b, -a), \quad \epsilon \le a < b < \infty.$$

In other words V_t^ϵ has the same jumps as X_t^ϵ except that they all go in the positive direction. As $\epsilon \downarrow 0$, V_t^ϵ increases (since we are adding jumps and they are all in the same direction) and using (6.15), we see that

$$\lim_{\epsilon \downarrow 0} \mathbb{E}\left[V_t^\epsilon\right] = t \int_{-\infty}^{\infty} |x| \, d\mu(x) < \infty.$$

Hence we can define

$$V_t = \lim_{\epsilon \downarrow 0} V_t^\epsilon$$

and $\mathbb{E}[V_t] < \infty$. Similarly, we can define

$$X_t = \lim_{\epsilon \downarrow 0} X_t^\epsilon.$$

Since $|X_t^\epsilon| \leq V_t$, the dominated convergence theorem is used to justify interchanges of limits.

Let $\mathcal{T}_n$ be the set of times such that the process $X_t^{1/n}$ is discontinuous, and let

$$\mathcal{T} = \bigcup_{n=1}^{\infty} \mathcal{T}_n.$$

Since each $\mathcal{T}_n$ is finite, the set $\mathcal{T}$ is countable, but if $\mu(\mathbb{R}) = \infty$, $\mathcal{T}$ will be infinite and, in fact, dense in every interval. Even though there are jumps in every interval, the paths are cadlag, that is, they are right continuous and the limits

$$X_{t-} = \lim_{s \uparrow t} X_s$$

exist for every s.

Calculating as in the previous section we can see the following.

Proposition 6.6.1. *Suppose X_t is a generalized Poisson process with Lévy measure μ. Then*

$$\Psi(s) = \int_{-\infty}^{\infty} [e^{ixs} - 1] \, d\mu(x),$$

$$\mathbb{E}[X_t] = tm, \quad m = \int_{-\infty}^{\infty} x \, d\mu(x),$$

$$Lf(x) = \int_{-\infty}^{\infty} [f(x+y) - f(x)] \, d\mu(y),$$

$$\mathrm{Var}[X_t] = \mathbb{E}[(X_t - mt)^2] = t \int_{-\infty}^{\infty} x^2 \, d\mu(x).$$

One may worry about the existence of the integrals above. Taylor's theorem implies that

$$e^{ixs} - 1 = ixs - \frac{x^2 s^2}{2} + O(|xs|^3).$$

Using this and (6.15) we can see that

$$\int_{|x| \leq 1} |e^{ixs} - 1| \, d\mu(x) < \infty.$$

Also,

$$\int_{|x|>1} |e^{ixs} - 1| \, d\mu(x) \le 2\,\mu\{x : |x| > 1\} < \infty.$$

For the generator L we assume that f is C^1 and hence for y near the origin

$$f(x + y) - f(x) = O(|y|).$$

Then,

$$\int_{|x| \le 1} |f(x + y) - f(x)| \, d\mu(y) < \infty.$$

Finiteness of the integral

$$\int_{|x|>1} |f(x + y) - f(x)| \, d\mu(y)$$

requires some assumptions on the growth of f at infinity.

We can relax the conditions somewhat by requiring (6.16) to hold but replacing (6.15) with the weaker condition

$$\int_{-1}^{1} |x| \, d\mu(x) < \infty. \tag{6.17}$$

Indeed, if (6.16) and (6.17) hold we can write

$$\mu = \mu_1 + \mu_2,$$

where μ_1 is μ restricted to $\{|x| \le 1\}$ and μ_2 is μ restricted to $\{|x| > 1\}$. To μ_1 we associate the generalized Poisson process X_t^1 and to μ_2 we associate an independent CPP, and then set $X_t = X_t^1 + X_t^2$. In this case it is possible that $\mathbb{E}[|X_t|] = \infty$. However, the formulas for Ψ and L are the same as above.

Example 6.6.1 (Positive Stable Processes). Suppose that $0 < \beta < 1$ and μ is defined by

$$d\mu(x) = c\,x^{-(1+\beta)} \, dx, \quad 0 < x < \infty,$$

where $c > 0$. In other words, the probability of a jump of size between x and $x + \Delta x$ in time Δt, is approximately

$$c\,x^{-(1+\beta)} \, (\Delta t) \, (\Delta x).$$

Note that

$$\int_\epsilon^\infty c\, x^{-(1+\beta)}\, dx = \frac{c}{\beta}\, \epsilon^{-\beta},$$

$$\int_0^1 x\, d\mu(x) = c \int_0^1 x^{-\beta}\, dx = \frac{c}{1-\beta} < \infty,$$

$$\int_0^\infty x\, d\mu(x) = c \int_0^\infty x^{-\beta}\, dx = \infty.$$

Therefore, μ satisfies (6.16) and (6.17), but not (6.15). The corresponding generalized Poisson process has

$$\Psi(s) = c \int_0^\infty [e^{isx} - 1]\, x^{-(1+\beta)}\, dx.$$

With careful integration, this integral can be computed but we omit it. Using the change of variables $y = rx$, we can see that if $r > 0$, then

$$\Psi(rs) = c \int_0^\infty [e^{irsx} - 1]\, x^{-(1+\beta)}\, dx = cr^\beta \int_0^\infty [e^{isy} - 1]\, y^{-(1+\beta)}\, dy.$$

This implies that the distribution of rX_1 is the same as that of X_{r^β}. In particular, $X_{1/n}$ has the same distribution as $n^{-1/\beta}\, X_1$ and hence we can write

$$X_1 = \frac{Z_1 + \cdots + Z_n}{n^{1/\beta}}, \quad Z_j = n^{1/\beta}\, [X_{j/n} - X_{(j-1)/n}],$$

where $Z_1, \ldots, Z_n$ are independent with the same distribution as X_1. The process X_t is called the *positive stable process* with exponent β.

Example 6.6.2 (Gamma process). Suppose $\lambda > 0$ and μ is defined by

$$d\mu(x) = \frac{e^{-\lambda x}}{x}\, dx, \quad 0 < x < \infty.$$

Note that

$$\int_0^\infty d\mu(x) = \infty, \quad \int_0^\infty x\, d\mu(x) < \infty.$$

Using a table of integrals, we can see that the characteristic exponent is

$$\Psi(s) = \int_0^\infty [e^{isx} - 1]\, d\mu(x) = \int_0^\infty \frac{e^{(is-\lambda)x} - e^{-\lambda x}}{x}\, dx = \log \frac{\lambda}{\lambda - is}.$$

In other words, the characteristic function of X_t is $\phi_{X_t}(s) = [\lambda/(\lambda-is)]^t$. This is the characteristic function of a Gamma random variable with parameters λ and t, that is, the density of X_t is

$$f_t(x) = \frac{\lambda}{\Gamma(t)}\,(\lambda x)^{t-1}\,e^{-\lambda x}, \quad 0 < x < 1,$$

where

$$\Gamma(t) = \int_0^\infty x^{t-1}\,e^{-x}\,dx,$$

is the Gamma function. Important values of t are the following.

- X_1 has the distribution of an exponential random variable with rate λ.

- If n is a positive integer, then X_n has the distribution of the sum of n independent random variables each exponential with rate λ. If N_s is a Poisson process with rate λ with $N_0 = 0$, then X_n has the same distribution as $\inf\{s : N_s = n\}$.

- If $\lambda = 1/2$ and n is a positive integer, then $X_{n/2}$ has a χ^2 distribution with n degrees of freedom, that is, it has the same distribution as

$$Z_1^2 + \cdots + Z_n^2,$$

 where $Z_1, \ldots, Z_n$ are independent, standard normal random variables.

The process X_t is called the *Gamma process with parameter* λ.

6.7 Generalized Poisson processes II

In the previous section, we assumed that the expected sum of the absolute values of the jump was finite. There are jump Lévy processes that do not satisfy this condition. These processes have many "small" jumps. Although the absolute values of the jumps are not summable, there is enough cancellation between the positive jumps and negative jumps to give a nontrivial process.

We will construct such a process associated to a Lévy measure μ satisying the following assumptions.

- For every $\epsilon > 0$, the number of jumps of absolute value greater than ϵ in a finite time interval is finite. More precisely, if μ_ϵ denotes μ restricted to $\{|x| > \epsilon\}$, then

$$\mu_\epsilon(\mathbb{R}) = \mu\{x : |x| > \epsilon\} < \infty \quad \text{for every} \quad \epsilon > 0. \tag{6.18}$$

- All jumps are bounded in absolute value by 1, that is,

$$\mu\{x : |x| > 1\} = 0. \tag{6.19}$$

- The second moment of the measure is finite,

$$\sigma^2 := \int_{-\infty}^{\infty} x^2 \, d\mu(x) < \infty. \tag{6.20}$$

Note that (6.20) can hold even if (6.15) does not hold.

Let

$$m_\epsilon = \int_{-\infty}^{\infty} x \, d\mu_\epsilon(x) = \int_{\epsilon < |x| \leq 1} x \, d\mu(x),$$

which is finite and well defined for all $\epsilon > 0$ by (6.18)–(6.20). For each $\epsilon > 0$, there is a CPP X_t^ϵ associated to the measure μ_ϵ and a correponding compensated Poisson process $M_t^\epsilon = X_t^\epsilon - m_\epsilon t$.

Definition Suppose μ is a measure on $\mathbb{R}$ satisfying (6.18)–(6.20). Then the *compensated generalized Poisson process (CGPP) with Lévy measure* μ is given by

$$Y_t = \lim_{\epsilon \downarrow 0} M_t^\epsilon = \lim_{\epsilon \downarrow 0}[X_t^\epsilon - m_\epsilon t]. \tag{6.21}$$

Two important cases are the following.

- If

$$\int_{-\infty}^{\infty} |x| \, d\mu(x) < \infty,$$

then the (uncompensated) generalized Poisson process

$$X_t = \lim_{\epsilon \downarrow 0} X_t^\epsilon,$$

as in the previous section exists. In this case

$$Y_t = X_t - mt,$$

where

$$m = \lim_{\epsilon \downarrow 0} m_\epsilon = \int_{-\infty}^{\infty} x \, d\mu(x).$$

- If μ is symmetric about the origin, that is, if for every $0 < a < b$ the rate of jumps with increments in $[a, b]$ is the same as that of jumps in $[-b, -a]$, then $m_\epsilon = 0$ for every ϵ and

$$Y_t = \lim_{\epsilon \downarrow 0} X_t^\epsilon.$$

Proposition 6.7.1. *If Y_t is a CGPP with Lévy measure μ, then*

$$\Psi(s) = \int_{-1}^1 [e^{isx} - 1 - isx] \, d\mu(x),$$

$$Lf(s) = \int_{-\infty}^\infty [f(x + y) - f(x) - yf'(x)] \, d\mu(y).$$

The latter quantity is defined provided that f is a C^2 function. If also $\mathbb{E}[f(Y_t)^2] < \infty$, then

$$M_t = f(Y_t) - \int_0^t Lf(Y_s) \, ds$$

is a square integrable martingale with

$$\langle M \rangle_t = \sum_{s \leq t} [f(Y_s) - f(Y_{s-})]^2.$$

Proof. Let $X^\epsilon = X_1^\epsilon$ and $M^\epsilon = M_1^\epsilon = X^\epsilon - m_\epsilon$. Let $\Psi^\epsilon = \Psi_{M_\epsilon}$ and note that

$$\Psi^\epsilon(s) = \Psi_{X_\epsilon}(s) - im_\epsilon s = \int_{|x| > \epsilon} [e^{isx} - 1] d\mu(x) - im_\epsilon s.$$

Since

$$m_\epsilon = \int_{|x| > \epsilon} x \, d\mu(x),$$

we can write

$$\Psi^\epsilon(s) = \int_{|x| > \epsilon} [e^{isx} - 1 - isx] \, d\mu(x).$$

Using Taylor series, we see that $|e^{isx} - 1 - isx| = O(s^2 x^2)$, and hence (6.20) implies that for each s,

$$\int |e^{isx} - 1 - isx| \, d\mu(x) < \infty.$$

This allows us to use the dominated convergence theorem to conclude that

$$\lim_{\epsilon \downarrow 0} \Psi^\epsilon(s) = \int_{-1}^{1} [e^{isx} - 1 - isx] \, d\mu(x).$$

Since $Y = \lim_{\epsilon \downarrow 0} M^\epsilon$, we can conclude that this limit equals $\Psi(s)$.

The second equation is obtained by taking the limit of (6.6).

$\square$

A particular case of the proposition is $f(x) = x$ in which we conclude that Y_t is a square integrable martingale with

$$\langle Y \rangle_t = \sum_{s \leq t} [Y_s - Y_{s-}]^2.$$

The same argument shows that the moment generating function also exists and

$$\mathbb{E}[e^{sY_t}] = \exp \left\{ t \int_{-1}^{1} [e^{sx} - 1 - sx] \, d\mu(x) \right\}, \quad s \in \mathbb{R}.$$

In particular, $\mathbb{E}[e^{sY_t}] < \infty$ for all s, t. (This is in contrast to CCPs for which this expectation could be infinite.)

We will establish existence of the limit in (6.21) both in L^2 and with probability one. Let us first consider the case $t = 1$. Let

$$Y^\epsilon = Y_1^\epsilon = X_1^\epsilon - m_\epsilon.$$

It suffices to prove the existence of

$$Y = \lim_{\epsilon \downarrow 0} Y^\epsilon.$$

Let $Z_n = Y^{2^{-n}} - Y^{2^{-(n+1)}}$, so that

$$Y = \sum_{n=0}^{\infty} Z_n.$$

The random variables Z_n are independent, mean zero, and

$$\sum_{n=0}^{\infty} \mathrm{Var}[Z_n] < \infty.$$

Therefore $\sum Z_n$ converges to a random variable Y in both L^2 and with probability one (see Exercise 1.14). If $2^{-n} \leq \epsilon < 2^{-n+1}$, then

$$\mathbb{E}\left[(Y^\epsilon - \sum_{k=0}^{n} Z_k)^2\right] \leq \int_{2^{-n}}^{2^{-n+1}} x^2 \, d\mu(x),$$

which goes to zero as $n \to \infty$. Hence

$$\lim_{\epsilon \downarrow 0} \mathbb{E}\left[|Y^\epsilon - Y|^2\right] = 0.$$

Convergence with probability one can be deduced from Lemma 6.3.3.

We now define Y_q as in (6.21) for $q \in \mathcal{D}$ as the limit with probability one. For other t we define

$$Y_{t+} = \lim_{q \downarrow t} Y_q, \qquad Y_{t-} = \lim_{q \uparrow t} Y_q. \tag{6.22}$$

Here we use the shorthand that when q is the index, the limit is taken over $\mathcal{D}$ only. Define the jump time $\mathcal{T}_\epsilon$ to be the set of times s such that X^ϵ is discontinuous and $\mathcal{T} = \cup_{j=1}^{\infty} \mathcal{T}_{1/j}$. With probability one each $\mathcal{T}_\epsilon$ is finite and hence $\mathcal{T}$ is countable.

Proposition 6.7.2. *With probability one, the limits in (6.22) exists for all t. Moreover, they are equal unless $t \in \mathcal{T}$.*

Proof. Let
$$S_\epsilon^2 = \mathbb{E}\left[(Y - Y_\epsilon)^2\right] = \int_{|x| \leq \epsilon} x^2 \, d\mu(x).$$

Let
$$M^\epsilon = \sup\left\{|Y_q - Y_q^\epsilon| : q \in \mathcal{D}, q \leq 1\right\}.$$

Using Lemma 6.3.3 (first for $Y_t^\delta - Y_t^\epsilon$ and then letting $\delta \downarrow 0$), we can see that

$$\mathbb{P}\{M \geq a\} \leq \frac{S_\epsilon^2}{a^2}.$$

We can find $\epsilon_n \downarrow 0$ such that

$$\sum_{n=1}^{\infty} n^2 \, S_{\epsilon_n}^2 < \infty.$$

Then if $M_n = M^{\epsilon_n}$, by the Borel-Cantelli lemma, with probability one, for all n sufficiently large $M_n \leq 1/n$ and this inequality implies that for all t,

$$\left| \limsup_{q \downarrow t} Y_{q+} - \lim_{q \downarrow t} Y^{\epsilon_n}_{q+} \right| \leq \frac{1}{n}.$$

Similar estimates hold for the liminf and for limits from the left.   □

We now define Y_t for all t by $Y_t = Y_{t+}$, and then the paths of Y_t are cadlag.

6.8   The Lévy-Khinchin characterization

The next theorem tells us that all Lévy processes can be given as independent sums of the processes we have described.

Theorem 6.8.1. *Suppose X_t is a Lévy process with $X_0 = 0$. Then X_t can be written as*

$$X_t = m\,t + \sigma\,B_t + C_t + Y_t, \tag{6.23}$$

where $m \in \mathbb{R}, \sigma \geq 0$, and B_t, C_t, Y_t are independent of the following types:

- *B_t is a standard Brownian motion,*

- *C_t is a compound Poisson process with Lévy measure μ_C satisfying*

$$\mu_C\{|x| \leq 1\} = 0,$$

$$\mu_C(\mathbb{R}) = \mu_C\{|x| > 1\} < \infty,$$

- *Y_t is a compensated generalized Poisson process with Lévy measure μ_Y satisfying*

$$\mu_Y(\{0\}) = 0, \quad \mu_Y\{|x| > 1\} = 0,$$

$$\int_{-\infty}^{\infty} x^2 \, d\mu_Y(x) < \infty.$$

The process $C_t + Y_t$ is called a *pure jump Lévy process*. We can summarize the theorem by saying that every Lévy process is the independent sum of a Brownian motion (with mean m and variance σ^2) and a pure jump process (with Lévy measure $\mu = \mu_C + \mu_Y$.)

We have not included the generalized Poisson process in our decomposition (6.23). If X_t is such a process we can write it as

$$X_t = C_t + \hat{X}_t$$

where C_t denotes the sum of jumps of absolute value greater than one. Then we can write $\hat{X}_t = Y_t + \hat{m}t$ where Y_t is a compensated generalized Poisson process and $\hat{m} = \mathbb{E}[\hat{X}_1]$, which by the assumptions is finite.

The conditions on μ_C, μ_Y can be summarized as follows:

$$\mu(\{0\}) = 0, \quad \int_{-\infty}^{\infty} [x^2 \wedge 1] \, d\mu(x) < \infty. \tag{6.24}$$

To specify the distribution of a Lévy processes, one needs to give $m \in \mathbb{R}, \sigma > 0$ and a measure μ satisfying (6.24). We then set μ_C to be μ restricted to $\{|x| > 1\}$ and μ_Y to be μ restricted to $\{|x| \leq 1\}$. We have already shown that (m, σ, μ) determine a Lévy process. To prove the theorem, therefore, we need to show that any Lévy process can be written like this.

We will not give the entire proof of this theorem, but let us sketch how one derives it. Suppose X_t is a Lévy process. Suppose that the process has jumps. For every $a > 0$, we can write

$$X_t = \hat{X}_{t,a} + \hat{C}_{t,a},$$

where $\hat{C}_{t,a}$ denotes the movement by jumps of absolute value greater than a and $\hat{X}_{t,a}$ denotes a Lévy process with all jumps bounded by a. For each a one can show that $\hat{C}_{t,a}$ is a CPP independent of $\hat{X}_{t,a}$. We let a go to zero, and after careful analysis we see that

$$X_t = Z_t + C_{t,1} + Y_t,$$

where Y_t is compensated generalized Poisson process and Z_t is a Lévy process with continuous paths. We then show that Z_t must be a Brownian motion. The following theorem is related and proved similarly.

Theorem 6.8.2. *(Lévy-Khinchin) If X has an infinitely divisible distribution, then there exist m, σ, μ_C, μ_Y satisfying the conditions of Theorem 6.8.1 such that*

$$\Psi_X(s) = ims - \frac{\sigma^2}{2} s^2 + \int_{|x|>1} [e^{isx} - 1] \, d\mu_C(x)$$

$$+ \int_{|x|\leq 1} [e^{isx} - 1 - isx] \, d\mu_Y(x).$$

We show how to prove this theorem and then prove another theorem which is a key step in the proof of Theorem 6.8.1.

Here we will show that if X has an infinitely divisible distribution, then the characteristic function of X can be written as $\phi(s) = e^{\Psi(s)}$ where

$$\Psi(s) = ims - \frac{\sigma^2 s}{2} + \int [e^{isx} - 1 - isx 1\{|x| \leq 1\}] \, \mu(dx),$$

where $m \in \mathbb{R}, \sigma^2 \geq 0$, and μ is a measure on $\mathbb{R}$ with $\mu\{0\} = 0$ and

$$\int \frac{x^2}{1+x^2} \, \mu(dx) < \infty.$$

We fix X and allow all constants to depend on the distribution of X. Let s_0 be such that $|\phi(s) - 1| \leq 1/2, \quad s \leq s_0$.

Since X is infinitely divisible we can write

$$X = Y_{1,n} + \cdots + Y_{n,n},$$

where $Y_{1,n}, \ldots, Y_{n,n}$ are i.i.d. We write μ_n for the distribution of $Y_{1.n}$ and $\nu_n = n\mu_n$. We let $\tilde{Y}_{j,n} = Y_{j,n} 1\{|Y_{j,n}| \leq 1\}$ and let

$$\tilde{m}_n = \mathbb{E}[\tilde{Y}_{j,n}], \quad \tilde{\sigma}_n^2 = \text{Var}[\tilde{Y}_{j,n}].$$

Let $\phi_n(s) = e^{\Psi_n(s)}$ denote the characteristic function of $Y_{1,n}$ and note that $\phi(s) = \phi_n(s)^n$. It suffices to find m, σ^2, μ and a subsequence n_j such that for $|s| \leq s_0$,

$$\lim_{j\to\infty} n_j \Psi_{n_j}(s) = ims - \frac{\sigma^2 s}{2} + \int [e^{isx} - 1 - isx 1\{|x| \leq 1\}] \, \mu(dx). \quad (6.25)$$

Lemma 6.8.3.

- *For all $\epsilon > 0$, there exists $M_\epsilon < \infty$ such that for all n, $\nu_n\{|x| \geq M_\epsilon\} \leq \epsilon$.*

- *For every $y > 0$, there exists $r < \infty$ such that for all n, $\nu_n\{|x| \geq y\} \leq r$.*

-
$$\sup_n n[\tilde{\mu}_n + \tilde{\sigma}_n^2] < \infty.$$

Proof. By symmetry it suffices to estimate $\nu_n\{x \geq r\}$. We first make the easy observation that for every $r < \infty$, for all n sufficiently large $\mu_n(-\infty, 1) \geq r/n$. Indeed, if this were not true, then there exists a subsequence n_j such that with probability at least

$$\left(1 - \frac{r}{n_j}\right)^{n_j} \sim e^{-r}$$

we would have $Y_{j,n_j} \geq 1$ for $j = 1, \ldots, n_j$ which implies that with positive probability $X = \infty$.

Let $M, r < \infty$ and let $q_n = \min\{r/n, \mu_n[M, \infty)\}$ and for $n > 2/r$ define $\delta_n^{\pm}$ to be the infimum (supremum) of all numbers such that

$$\mathbb{P}\{Y_{1,n} \geq \delta_n^+\} \leq q_n, \quad \mathbb{P}\{Y_{1,n} \leq \delta_n^-\} \leq q_n.$$

Using this we can find a random variable with the same distribution as $Y_{1,n}$ by the following.

- With probability $1 - 2q_n$ choose from distribution ρ_1 and with probability $2q_n$ choose from distribution ρ_2.

- Distribution ρ_2 satisfies $\rho_2(-\infty, 1] = \frac{1}{2}, \rho_2[M, \infty) = \frac{1}{2}$.

By focusing on the first time that one chooses from ρ_2 we see that the conditional distribution on a set of probability $p_n := 1 - (1 - q_n)^n$ has a "spread" of at least $M - 1$. Another way of saying this is that if x is the median of the distribution of the sum of the remaining terms, there is probability at last $p_n/4$ of being less than $x+1$ and probability at last $p_n/4$ of being greater than $x + M$.

If $a > 0$, let $r_n = n\mathbb{P}\{a \leq Y_{1,n}\}$ and assume $r_n \geq 1$. Using the first part we can find b (independent of n, a) with $\mathbb{P}\{a \leq Y_{1,n} \leq b\} \geq r_n/(2n)$. If $0 < s < 1/2b$, then

$$|1 - e^{iys}| \geq c\, y^2\, s^2, \quad 0 \leq |y| \leq b.$$

Find $0 \le s \le 1/2b$ such that $|\phi(s)| \ge 1/e$. We have for n sufficiently large

$$|\phi_n(s)| \le 1 - \frac{cs^2 a^2 r_n}{2n}.$$

and hence

$$\liminf_{n \to \infty} 2ncs^2 a^2 r_n \le 1.$$

Let $Y_n = \sum_j \tilde{Y}_{1,n}$. Using the second part, we see that $\mathbb{P}\{\tilde{Y}_{j,n} = Y_{j,n}, j = 1, \ldots, n\} \ge c$, and hence the $\{Y_n\}$ are tight, that is, for every $\epsilon > 0$, there exists K_ϵ such that for all n, $\mathbb{P}\{|Y_{j,n}| > K_\epsilon\} < \epsilon$. Chebyshev's inequality gives

$$\mathbb{P}\left\{|Y_n| \le \frac{n|\tilde{m}_n|}{2}\right\} \le \mathbb{P}\left\{|Y_n - n\tilde{m}_n| \ge \frac{n|\tilde{m}_n|}{2}\right\} \le \frac{4n\tilde{\sigma}_n^2}{n^2 \tilde{m}_n^2}.$$

If $n\tilde{\sigma}_n^2$ is bounded and n_j is a subsequence with $n_j|\tilde{m}_{n_j}| \to \infty$, then we would have

$$\limsup_{y \to \infty} \mathbb{P}\left\{|Y_{n_j}| \ge \frac{n_j|\tilde{m}_{n_j}|}{2}\right\} = 1,$$

which contradicts the tightness of Y_n. Suppose $n_j\tilde{\sigma}_{n_j}^2 \to \infty$. Then tightness gives that $|\tilde{m}_{n_j}|^2 = o(\tilde{\sigma}_{n_j}^2)$ and hence $\mathbb{E}[|\tilde{Y}_{1,n_j}|^3] \le \mathbb{E}[\tilde{Y}_{1,n_j}^2] \sim \tilde{\sigma}_{n_j}^2$. To be a bit more precise, we would assume that

$$\frac{Y_n - n\,\tilde{m}_n}{\sqrt{n\tilde{\sigma}_n^2}}$$

would approach a normal distribution. If $n\tilde{\sigma}_n^2 \to \infty$ this would indicate that Y_n is not tight and one can make this precise with the Berry-Esseen theorem (using the upper bound on the thrid moment). Hence $n\tilde{\sigma}_n^2$ is bounded and hence $n|\tilde{m}_n|$ is bounded.

$\square$

Using the lemma we can find a measure μ with $\mu\{0\} = 0$ and a subsequence n_j such that for all bounded continuous functions f that vanish in a neighborhood of the origin,

$$\lim_{j \to \infty} \int f(x)\nu_{n_j}(dx) = \int f(x)\mu(dx).$$

By taking a subsubsequence if necessary we can also conclude that there exists $\tilde{m}, \tilde{\sigma}^2$ such that

$$\lim_{j \to \infty} n_j \, \tilde{m}_{n_j} = \tilde{m}, \quad \lim_{j \to \infty} n_j \, \sigma^2_{n_j} = \tilde{\sigma}^2.$$

We will assume that $\mu\{-1, 1\} = 0$. If this is not the case, find b with $\mu\{-b, b\} = 0$ and do the proof similarly defining $\tilde{Y}_{1,n} = Y_{1,n} \, 1\{|\tilde{Y}_{1,n}| \leq b\}$. We will establish (6.25) with this μ, $m = \tilde{m}$ and

$$\sigma^2 = \tilde{\sigma}^2 - \int_{|x| \leq 1} x^2 \, \mu(dx).$$

For the remainder we fix s with $|s| \leq s_0$ and allow constants to depend on s. We note that

$$|\phi_n(s) - 1| \leq c/n,$$

and hence

$$\phi(s) = [1 + \phi_n(s) - 1]^n = \exp\left\{n(\phi_n(s) - 1)\right\} [1 + O(n^{-1})],$$

$$
\begin{aligned}
\Psi(s) + O(n^{-1}) &= n\left[\phi_n(s) - 1\right] \\
&= \int [e^{isx} - 1] \, \nu_n(dx) \\
&= isn\tilde{m}_n + \int [e^{isx} - 1 - isx 1\{|x| \leq 1\}] \, \nu_n(dx).
\end{aligned}
$$

$$\lim_{j \to \infty} \int_{|x| > a} [e^{isx} - 1 - isxg_\epsilon(x)] \, \nu_{n_j}(dx) = \int_{|x| > a} [e^{isx} - 1 - isxg_\epsilon] \, \mu(dx).$$

Using $\mu\{-a, a-1, 1\} = 0$ (to handle the discontinuity of the integrand at $x = \pm a, \pm 1$), we can see that

$$\lim_{j \to \infty} \int_{|x| > a} [e^{isx} - 1 - isx 1\{|x| \leq 1\}] \, \nu_{n_j}(dx)$$

$$= \int_{|x| > a} [e^{isx} - 1 - isx 1\{|x| \leq 1\}] \, \mu(dx).$$

Also,

$$\left| \int_{|x| \le a} [e^{isx} - 1 - isx 1\{|x| \le 1\}] \, \nu_n(dx) + \frac{s^2}{2} \int_{|x| \le a} x^2 \, \nu_n(x) \right|$$

$$\le \quad c \int_{|x| \le a} |x|^3 \nu_n(dx)$$

$$\le \quad c \, a \int_{|x| \le a} x^2 \nu_n(dx)$$

$$\le \quad ca.$$

Also, as $j \to \infty$, again using $\mu\{\pm a, \pm 1\} = 0$,

$$\int_{|x| \le a} x^2 \, \nu_{n_j}(x) \quad = \quad \int_{|x| \le 1} x^2 \, \nu_{n_j}(x) - \int_{a < |x| \le 1} x^2 \, \nu_{n_j}(x)$$

$$= \quad o(1) + \tilde{\sigma}^2 - \int_{a < |x| \le 1} x^2 \mu(dx).$$

By taking $a \to 0$ at an appropriate rate, we see that

$$\lim_{j \to \infty} \int [e^{isx} - 1 - isx 1\{|x| \le 1\}] \, \nu_{n_j}(dx)$$

$$= -\frac{s^2}{2} \left[\tilde{\sigma}^2 - \int x^2 \, 1\{|x| \le 1\} \, \mu(dx) \right]$$

$$+ \int [e^{isx} - 1 - isx 1\{|x| \le 1\}] \, \nu(dx).$$

Theorem 6.8.4. *Suppose X_t is a Lévy process with continuous paths. Then X_t is a Brownian motion.*

Proof. All we need to show is that X_1 has a normal distribution. Let

$$X_{j,n} = X_{j/n} - X_{(j-1)/n},$$

$$M_n = \max \{|X_{j,n}| : j = 1, \ldots, n\}.$$

Continuity of the paths implies that $M_n \to 0$ and hence for every $a > 0$, $\mathbb{P}\{M_n < a\} \to 1$. Independence of the increments implies that

$$\mathbb{P}\{M_n < a\} \quad = \quad \left[1 - \mathbb{P}\{|X_{1/n}| \ge a\}\right]^n$$

$$\le \quad \exp\left\{-n\mathbb{P}\{|X_{1/n}| \ge a\}\right\}.$$

Therefore, for every $a > 0$,

$$\lim_{n\to\infty} n\,\mathbb{P}\{|X_{1/n}| \geq a\} = 0. \tag{6.26}$$

This implies that there exists a sequence $a_n \downarrow 0$ with

$$\lim_{n\to\infty} n\,\mathbb{P}\{X_{1/n} \geq a_n\} = 0. \tag{6.27}$$

We claim that all the moments of X_1 are finite. To see this let $J = \max_{0\leq t\leq 1} |X_t|$ and find k such that $\mathbb{P}\{J \geq k\} \leq 1/2$. Then using continuity of the paths, by stopping at the first time t that $|X_t| = nk$, we can see that

$$\mathbb{P}\{J \geq (n+1)k \mid J \geq nk\} \leq 1/2,$$

and hence

$$\mathbb{P}\{J \geq nk\} \leq (1/2)^n,$$

from which finiteness of the moments follows. Let $m = \mathbb{E}[X_1], \sigma^2 = \mathrm{Var}[X_1]$, and note that $\mathbb{E}[X_1^2] = m^2 + \sigma^2$. Our goal is to show that $X_1 \sim N(m, \sigma^2)$.

Let

$$\tilde{X}_{j,n} = X_{j,n}\,1\{|X_{j,n}| \leq a_n\}, \quad Z_n = \sum_{j=1}^{n} \tilde{X}_{j,n}.$$

Let ϕ_n denote the characteristic function of $\tilde{X}_{j,n}$ and hence $[\phi_n]^n$ is the characteristic function of Z_n. It follows from (6.27) that $Z_n \to X_1$ in distribution, so it suffices to prove that for every s

$$\lim_{n\to\infty} [\phi_n(s)]^n = \exp\left\{ims - \frac{\sigma^2 s^2}{2}\right\}.$$

Also, it is easy to check that $\mathbb{E}[Z_n] \to m$, $\mathrm{Var}[Z_n] \to \sigma^2$ and hence

$$\mathbb{E}[\tilde{X}_{1,n}] = \frac{m\,[1 + o(1)]}{n}, \quad \mathbb{E}[\tilde{X}_{1,n}^2] = \left[\frac{m^2}{n^2} + \frac{\sigma^2}{n}\right](1 + o(1)).$$

Also, since $|\tilde{X}_{1,n}| \leq a_n$,

$$\mathbb{E}[|\tilde{X}_{1,n}|^3] \leq a_n\,\mathbb{E}[\tilde{X}_{1,n}^2] = a_n \left[\frac{m^2}{n^2} + \frac{\sigma^2}{n}\right][1 + o(1)].$$

(This estimate uses (6.27) which in turn uses the fact that the paths are continuous.) Using these estimates, we see that for fixed s,

$$\phi_n(s) = 1 + \frac{ims}{n} - \frac{\sigma^2 s^2}{2n} + o\left(\frac{1}{n}\right),$$

where the $o(\cdot)$ term depend on s. This implies that

$$
\begin{aligned}
\lim_{n \to \infty} \phi_n(s)^n &= \lim_{n \to \infty}\left[1 + \frac{ims}{n} - \frac{\sigma^2 s^2}{2n} + o\left(\frac{1}{n}\right)\right]^n \\
&= \exp\left\{ims - \frac{\sigma^2 s^2}{2}\right\}.
\end{aligned}
$$

$\square$

6.9   Integration with respect to Lévy processes

Suppose X_t is a Lévy process which by Theorem 6.8.1 we can write as

$$X_t = m\,t + \sigma\,B_t + C_t + Y_t.$$

We will define the integral

$$\int_0^t A_s\,X_s$$

by

$$m\int_0^t A_s\,ds + \sigma\int_0^t A_s\,dB_s + \int_0^t A_s\,dC_s + \int_0^t A_s\,dY_s.$$

The first integral is a Riemann integral, the second is an Itô integral as in Chapter 3, the third integral was defined in Section 6.4, and this leaves us only the fourth integral to define. As we saw in Section (6.4), the third integral does not have the properties we want unless we assume that A_t is left continuous. The same problems exists for the fourth integral so we will restrict to such processes.

Assume that Y_t is a process with Lévy measure μ satisfying (6.18)–(6.20). Then Y_t is a square integrable martingale with $\mathbb{E}[Y_t] = \sigma^2\,t$ (this is not the same σ as in the previous paragraph). This will allow us to define

$$\int_0^t A_s\,dY_s$$

as an Itô integral as in Chapter 3.

We start with simple processes. Suppose A_t is a simple process as in Section 3.2.2. To be specific, suppose that times $0 \le t_0 < t_1 < \cdots < t_n < \infty$ are given, and $A_s = A_{t_j}$ t for $t_j \le s < t_{j+1}$. Then, we define

$$\int_0^t A_s \, dY_s = \sum_{i=0}^{j-1} A_{t_i} \left[Y_{t_{i+1}} - Y_{t_i} \right] + A_{t_j} \left[Y_t - Y_{t_j} \right].$$

With probability one, the paths of Y_t are continuous at the times $t_1, \ldots, t_n$ so the definition above does not care whether we choose the simple process A_s to be right continuous or left continuous. The following proposition is proved in the same way as Proposition 3.2.1.

Proposition 6.9.1. *Suppose $A_t, \hat{A}_t$ are bounded simple processes adapted to the filtration of the process Y_t.*

- **Linearity.** *If a, b are constants, then $aA_t + b\hat{A}_t$ is also a simple process and*

$$\int_0^t (aA_s + b\hat{A}_s) \, dY_s = a \int_0^t A_s \, dY_s + b \int_0^t \hat{A}_s \, dY_s.$$

  Moreover, if $0 < r < t$,

$$\int_0^t A_s \, ds = \int_0^r A_s \, dY_s + \int_r^t A_s \, dY_s.$$

- **Martingale property.** *The process*

$$Z_t = \int_0^t A_s \, dY_s$$

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

- **Variance rule.** *Z_t is square integrable and*

$$\mathrm{Var}[Z_t] = \sigma^2 \mathbb{E}\left[Z_t^2 \right] = \int_0^t \mathbb{E}[A_s^2] \, ds.$$

- **Cadlag paths.** *With probability one, the function $t \mapsto Z_t$ is cadlag.*

We now suppose that A_s is an adapted process with left continuous paths. If $\tilde{A}_s$ is an adapted process with cadlag paths, we could take $A_s = \tilde{A}_{s-}$. We approximate A_s by a simple process $A_s^{(n)}$ by choosing

$$A_s^{(n)} = 2^n \int_{(j-1)2^{-n}}^{j2^{-n}} A_r \, dr \qquad \text{if} \quad \frac{j}{2^n} < r \le \frac{j+1}{2^n}.$$

The integral is defined by

$$\int_0^t A_s \, dY_s = \lim_{n \to \infty} \int_0^t A_s^{(n)} \, dY_s.$$

Let us give be more precise about this definition. Suppose that A_t is a process with left continuous paths with right limits, that is, the left continuous version of a processes with cadlag paths. Assume that exists $K < \infty$ such that with probability one. $|A_t| \le K$ for all t. Since with probability one, the paths are left continuous, we know that

$$\lim_{n \to \infty} A_t^{(n)} = A_t.$$

SInce $|A_t^{(n)}| \le K$, we can used dominated convergence (for Lebesgue measure) to show that with probability one,

$$\lim_{n \to \infty} \int_0^t |A_s^{(n)} - A_s|^2 \, ds = 0,$$

and dominated convegence for $\mathbb{P}$ to see that

$$\lim_{n \to \infty} \mathbb{E}\left[\int_0^t |A_s^{(n)} - A_s|^2 \, ds \right] = \lim_{n \to \infty} \int_0^t \mathbb{E}\left[|A_s^{(n)} - A_s|^2 \right] \, ds = 0.$$

As in the case for Brownian motion, we use the proposition to see that for fixed t, the sequence of random variables

$$Z_t^{(n)} := \int_0^t A_s^{(n)} \, dA_s,$$

is a Cauchy sequence in L^2 and has a limit Z_t in L^2.

We use this method to define the integral at dyadic times t, an we define it at other times by

$$Z_t = \lim_{t_n \downarrow t} Z_{t_n}$$

where t_n are dyadic rationals.

6.10　Symmetric stable process

An important example of a Lévy process is the symmetric stable process. A random variable X has a symmetric stable distribution with parameter $\alpha \in (0, 2]$ if $\Psi_X(s) = -|cs|^\alpha$ for some $c > 0$. If $\alpha = 2$, this means that X has a normal distribution with mean zero and variance $2c^2$. For $0 < \alpha < 2$ these distributions arise in Lévy processes that we call symmetric α-stable processes. The term stable comes from the following property.

Proposition 6.10.1. *If X has a symmetric α-stable distribution and $X_1, \ldots, X_n$ are independent random variables with the same distribution as X, then X has the same distribution as*

$$Y = \frac{X_1 + \cdots + X_n}{n^{1/\alpha}}.$$

Proof. Using properties of characteristic functions, we get

$$\Psi_Y(s) = \sum_{j=1}^n \Psi_{X_j}(s/n^{1/\alpha}) = n\Psi_X(s/n^{1/\alpha}) = \Psi_X(s).$$

$\square$

Definition If $0 < \alpha \leq 2$, then a Lévy process X_t is a *symmetric stable process* if

$$\Psi(s) = -|cs|^\alpha$$

for some $c > 0$.

Proposition 6.10.2. *Suppose $0 < \alpha < 2$, $C > 0$ and X_t is a Lévy process with Lévy measure*

$$d\mu(x) = \frac{C}{|x|^{1+\alpha}}\, dx. \qquad (6.28)$$

Then X_t is a symmetric α-stable process with $\Psi(s) = -Cb_\alpha|s|^\alpha$ where

$$b_\alpha = 2\int_0^\infty \frac{1 - \cos y}{y^{1+\alpha}}\, dy < \infty.$$

The condition $0 < \alpha < 2$ is needed to guarantee that μ as defined in (6.28) satisfies (6.24). Since $1 - \cos(y) = O(y^2)$ as $y \to 0$, we can see that b_α is finite. In fact, it can be calculated

$$b_\alpha = \frac{\pi}{\Gamma(\alpha + 1)\sin(\alpha\pi/2)},$$

where Γ denotes the Gamma function. The proof of the proposition is a simple computation. We write

$$\Psi(s) = C \int_{|x|\leq 1} [e^{isx} - 1 - isx] \frac{dx}{|x|^{1+\alpha}} + C \int_{|x|>1} [e^{isx} - 1] \frac{dx}{|x|^{1+\alpha}}.$$

Using $e^{isx} = \cos(sx) + i\sin(sx)$ and the fact that μ is symmetric about the origin, we see that

$$\begin{aligned}\Psi(s) &= 2C \int_0^\infty \frac{\cos(sx) - 1}{x^{1+\alpha}}\,dx \\ &= 2C|s|^\alpha \int_0^\infty \frac{\cos(y) - 1}{y^{1+\alpha}}\,dy = -Cb_\alpha|s|^\alpha.\end{aligned}$$

Suppose X_t has a symmetric α-stable process with Lévy measure as in (6.28) with $C = 1/b_\alpha$. It can be shown that X_1 has a symmetric density g_α,

$$\mathbb{P}\{a \leq X \leq b\} = \int_a^b g_\alpha(x)\,dx.$$

This density is bounded. Unfortunately, except for the case $\alpha = 1$, there is no explicit form expression for the density. Despite this fact, it is known that

$$g_\alpha(x) \sim \frac{1}{|x|^{1+\alpha}}, \qquad |x| \to \infty,$$

and hence

$$\mathbb{P}\{|X_1| \geq K\} \sim \int_{|x|\geq K} g_\alpha(x)\,dx = \frac{2}{\alpha K^\alpha}.$$

This comes from the fact that the easiest way for $|X_1|$ to be unusually large is for there to be a single very large jump, and the probability of a jump of size at least K by time one is asymptotic to $\mu\{|x| \geq K\}$.

If $\alpha = 1$, then $b_\alpha = \pi$ and the corresponding Lévy measure is

$$d\mu(x) = \frac{1}{\pi\,x^2}\,.$$

The density of X_1 is the *Cauchy density*

$$g_1(x) = \frac{1}{\pi\,(x^2 + 1)}.$$

The easiest way to check this is to compute the characteristic function of g_t and obtain $e^{-|s|}$.

Symmetric stable distribution arise as limit distributions for sums of symmetric random variables with "heavy tails". The next proposition gives one version.

Proposition 6.10.3. *Suppose $\alpha, c > 0$ and $X_1, X_2, \ldots$ are independent, identically distributed random variables with a bounded density f satisfying $f(x) = f(-x)$ and*

$$f(x) \sim \frac{c}{x^{1+\alpha}}, \quad x \to \infty, \tag{6.29}$$

where $\sim$ means the ratio of the two sides converges to one.

- *If $0 < \alpha < 2$, and*

$$Z_n = \frac{X_1 + \cdots + X_n}{n^{1/\alpha}}, \tag{6.30}$$

then Z_n converges in distribution to an α-symmetric random variable.

- *If $\alpha > 2$, and*

$$Z_n = \frac{X_1 + \cdots + X_n}{\sqrt{n}},$$

then Z_n converges in distribution to a centered normal random variable.

- *If $\alpha = 2$, and*

$$Z_n = \frac{X_1 + \cdots + X_n}{\sqrt{n \log n}},$$

then Z_n converges in distribution to a centered normal random variable.

If $\alpha > 2$, then $\mathrm{Var}[X_j] < \infty$, and hence the result is a restatement of the central limit theorem.

We will prove the proposition for $0 < \alpha < 2$. Our first observation is that if $X_1, X_2, \ldots$ are independent, identically distributed random variables whose characteristic exponent Ψ satisfies

$$\Psi(s) = -r\,|s|^\alpha + o(|s|^\alpha), \quad s \to 0, \tag{6.31}$$

then Z_n as defined in (6.30) converges to an α-stable distribution. This follows since for each s,

$$\lim_{n \to \infty} \Psi_{Z_n}(s) = \lim_{n \to \infty} n\,\Psi(s/n^{1/\alpha}) = -r\,|s|^\alpha.$$

Therefore, we only need to show that (6.29) implies (6.31).

Let ϕ denote the characteristic function of X_1, and note that $-\Psi(s) = -\log \phi(s) \sim \phi(s) - 1$ as $s \to 0$. By changing variables, we see that if $s > 0$,

$$
\begin{aligned}
\phi(s) - 1 &= \int_{-\infty}^{\infty} [e^{isx} - 1] f(x) \, dx \\
&= 2 \int_0^{\infty} [\cos(sx) - 1] f(x) \, dx \\
&= 2s^{-1} \int_0^{\infty} [\cos y - 1] f(y/s) \, dy \\
&= 2s^{\alpha} \int_0^{\infty} [\cos y - 1] s^{-(1+\alpha)} f(y/s) \, dy
\end{aligned}
$$

We claim that

$$
\lim_{s \downarrow 0} \int_0^{\infty} [\cos y - 1] s^{-(1+\alpha)} f(y/s) \, dy = cI,
$$

where

$$
I = \int_0^{\infty} \frac{\cos y - 1}{y^{1+\alpha}} \, dy,
$$

from which (6.31) follows with $r = 2cI$. To see this, let $\epsilon = (2-\alpha)/6 > 0$ and note that

$$
\left| \int_0^{s^{1-\epsilon}} [\cos y - 1] s^{-(1+\alpha)} f(y/s) \, dy \right| \leq C s^{-(1+\alpha)} \int_0^{s^{1-\epsilon}} y^2 \, dy
$$

$$
\leq C s^{(2-\alpha)/2} \longrightarrow 0.
$$

Also, for $y \geq s^{1-\epsilon}$,

$$
f(y/s) = c \, (y/s)^{-(1+\alpha)} [1 + o(1)],
$$

and hence

$$
\lim_{s \downarrow 0} \int_{s^{1-\epsilon}}^{\infty} [\cos y - 1] s^{-(1+\alpha)} f(y/s) \, dy
$$

$$
= c \lim_{s \downarrow 0} \int_{s^{1-\epsilon}}^{\infty} [\cos y - 1] y^{-(1+\alpha)} \, dy = cI.
$$

Proposition 6.10.3 only uses stable processes for $\alpha \leq 2$. The next proposition that shows that there are no nontrivial stable process for $\alpha > 2$.

Proposition 6.10.4. *Suppose $0 < \beta < 1/2$, and $X_1, X_2, \ldots$ are independent, identically distributed each having the same distribution as*

$$\frac{X_1 + \cdots + X_n}{n^\beta}.$$

Then $\mathbb{P}\{X_j = 0\} = 1$.

We use the following lemma to prove Proposition 6.10.4.

Lemma 6.10.5. *Suppose X is a nonconstant random variable with characteristic function ϕ. Then there exists $\epsilon > 0$ such that if $|s| \leq \epsilon$, then*

$$|\phi(s)| \leq 1 - \epsilon\, s^2.$$

Proof. Let us first assume that X is bounded with mean m and variance $\sigma^2 > 0$. Then, the characteristic function satisfies

$$\phi(s) = e^{ims} \left[1 - \frac{\sigma^2}{2} s^2 + O(|s|^3)\right], \quad s \downarrow 0,$$

and hence there exists $\epsilon > 0$ such that

$$|\phi(s)| \leq \left|1 - \frac{\sigma^2}{4} s^2\right|, \quad |s| \leq \epsilon.$$

For general nonconstant X, we can find a bounded interval I such that $\mathbb{P}\{X \in I\} = \rho > 0$ and such the variance σ^2 of the conditional distribution given $X \in I$ is positive. Then,

$$\phi(s) = \rho\, \mathbb{E}[e^{isX} \mid X \in I] + (1 - \rho)\, \mathbb{E}[e^{isX} \mid X \notin I].$$

There exists $\epsilon > 0$ such that if $|s| \leq \epsilon$, $|\mathbb{E}[e^{isX} \mid X \in I]| \leq 1 - (\sigma^2 s^2/4)$ and hence

$$|\phi(s)| \leq \rho \left(1 - \frac{\sigma^2}{4} s^2\right) + 1 - \rho = 1 - \frac{\rho\sigma^2}{4} s^2.$$

$\square$

To prove Proposition 6.10.4, let ϕ denote the characteristic function of X_j. Then

$$\phi(s) = \phi(n^{-\beta} s)^n.$$

If the distribution of X_j is nontrivial, then there exists $\epsilon > 0$ such that for $|s| \leq \epsilon$,

$$|\phi(s)| \leq 1 - \epsilon\, s^2,$$

and hence for every $s \neq 0$,

$$\begin{aligned}
|\phi(s)| = \lim_{n\to\infty} |\phi(n^{-\beta} s)|^n &\leq \lim_{n\to\infty} \left(1 - \epsilon\, n^{-2\beta} s^2\right)^n \\
&\leq \lim_{n\to\infty} \exp\left\{-\epsilon\, s^2\, n^{1-2\beta}\right\} = 0.
\end{aligned}$$

Since ϕ is continuous at 0 with $\phi(0) = 1$, this is a contradiction. Therefore, $\mathbb{P}\{X_j = c\} = 1$ for some c and since $c = (c + \cdots + c)/n^\beta$, $c = 0$.

6.11   Exercises

Exercise 6.1. Suppose $F(t) = \mathbb{P}\{T \leq t\}$ is the distribution function of a continuous random variable T. Define the "inverse" function G by

$$G(r) = \sup\{t : F(t) \leq r\}.$$

1. Suppose U is a random variable distributed uniformly on $[0, 1]$. Show that $G(U)$ has distribution function F.

2. Find G in the case that T is exponential with rate λ.

3. Use the waiting time method as described in Section 6.2 to sample from a Poisson process X_t with rate $\lambda = 3/2$. Run at least 3000 trials to estimate the distributon of X_3.

4. Compare the simulations with the actual distribution for X_3.

Exercise 6.2. Suppose Y_t is a Poisson process with parameter 2.

1. Find $\mathbb{P}\{Y_3 \geq 2\}$.

2. Find $\mathbb{P}\{Y_4 \geq Y_1 + 2 \mid Y_1 = 4\}$.

3. Find $\mathbb{P}\{Y_1 = 1 \mid Y_3 = 4\}$.

4. Find a nonrandom function $a(t)$ such that

$$X_t = Y_t - a(t)$$

is a martingale, that is, if $s \leq t$,

$$E[X_t \mid \mathcal{F}_s] = E[X_t \mid Y_s] = X_s.$$

Here $\mathcal{F}_t$ denotes the information in $\{Y_s : s \leq t\}$.

Exercise 6.3. Suppose X_t is a compound Poisson process with $\lambda = 2$, and measure $\mu^\#$ is given by a $N(0, 1)$ distribution. In other words, the process jumps when a Poisson process of rate 2 jumps and when it jumps it chooses a jump size from a $N(0, 1)$ distribution.

- Find the function f so that the Lévy measure μ can be written as $d\mu = f(x)\, dx$.

- Find $\mathbb{E}[X_t]$.

- Find $\mathbb{E}[X_t^2]$

- What is the generator L of the process X_t?

- Does there exist a process S_t such that for each t, S_t is measurable with respect to $\mathcal{F}_t$, the information in $\{X_s : 0 \leq s \leq t\}$, and such that

    - With probability one, S_t is differentiable with respect to t,
    - $M_t = X_t^2 - S_t$ is a martingale

  If so find it.

Exercise 6.4. Let X_t be as in Exercise 6.3 and let

$$Z_t = \exp\{X_t\}.$$

1. Find $\mathbb{E}[Z_t]$.

2. Find a continuous process S_t, adapted to the information in $\{\mathcal{F}_t\}$, with $S_t > 0$ such that S_t is differentiable with respect to t (except at the t at which X_t jumps) and

$$M_t = Z_t - S_t$$

is a martingale? Explain why S_t is not differentiable at the jump times of X_t.

3. Does there exist a process A_t, adapted to the information in $\{\mathcal{F}_t\}$, with $A_t > 0$ such that A_t is differentiable with respect to t (except at the t at which X_t jumps) and

$$M_t = Z_t \, A_t$$

is a martingale?

Exercise 6.5. Suppose Y_t is a Cauchy process, that is, a Lévy process such that Y_1 has a Cauchy distribution. Show why the following statement holds: for every $r > 0$ and $t > 0$,

$$\mathbb{P}\left\{\max_{0 \le s \le t} Y_s > r\right\} \quad (*) \quad 2\,\mathbb{P}\{Y_t > r\}.$$

Here $(*)$ is one of the following: $>, =, <$. Your task is to figure out which of $>, =, <$ should go into the statement and explain why the relation holds. (Hint: go back to the derivation of the reflection principle for Brownian motion. The only things about the Cauchy process that you should need to use are that it is symmetric about the origin and has jumps. Indeed, the same answer should be true for any symmetric Lévy process with jumps.)

Exercise 6.6. Suppose X_t is a Poisson process with rate 1 and let $r > 0$ with filtration $\{\mathcal{F}_t\}$, and

$$\tilde{S}_t = e^{X_t - rt}.$$

1. Find a strictly positive martingale M_t with $M_0 = 1$ such that $\tilde{S}_t$ is a martingale with respect to the tilted measure Q given by

$$Q(V) = \mathbb{E}\left[1_V \, M_t\right], \quad V \in \mathcal{F}_t.$$

2. Is your choice of M_t and Q unique?

Exercise 6.7. Suppose $f : \mathbb{R} \to \mathbb{R}$ is a function that is right continuous at every point. Let

$$K_f = \sup\{|f(x)| : 0 \le x \le 1\}.$$

1. Give an example to show that it is possible that $K_f = \infty$.

2. Show that if f is a cadlag function, then $K_f < \infty$.

Exercise 6.8. Suppose X_t is a generalized Poisson process with Lévy measure μ given by

$$d\mu(x) = \frac{dx}{x}, \quad 0 < x < 1.$$

1. Is this a generalized Poisson process of Type I?

2. Find $\mathbb{E}[X_t], \mathrm{Var}[X_t]$.

3. What is the probability that there are no jumps of size greater than $1/2$ by time $t = 2$?

4. Let
$$f(t, x) = \mathbb{E}\left[X_t^4 \mid X_0 = x\right].$$
Find the function $g(x)$ such that

$$\frac{d}{dt} f(t, x) \mid_{t=0} = g(x).$$

Exercise 6.9. Suppose X_t is as in Exercise 6.8 and B_t is an independent standard Brownian motion. Let $Y_t = X_t + t + 2B_t$ and answer the same questions for Y_t.

Exercise 6.10. Do Exercise 6.8 where

$$d\mu(x) = \frac{dx}{|x|^{5/2}}, \quad 0 < |x| < 1.$$

Chapter 7

Fractional Brownian motion

7.1 Definition

The assumptions of independent, stationary increments along with continuous paths give Brownian motion. In the last chapter we dropped the continuity assumption and obtained Lévy processes. In this chapter we will retain the assumptions of stationary increments and continuous paths, but will allow the increments to be dependent. The process X_t we construct is called *fractional Brownian motion* and depends on a parameter $H \in (0, 1)$ called the *Hurst index*. It measures the correlation of the increments.

- When $H > 1/2$, the increments are positively correlated, that is, if the process has been increasing, then it is more likely to continue increasing.

- When $H < 1/2$, the increments are negatively correlated, that is, if the process has been increasing, then it is more likely to decrease in the future.

- If $H = 1/2$, the increments are uncorrelated and the process is the same as Brownian motion.

We will also assume that the process is self-similar.

- If $a > 0$, then the distribution of

$$Y_t := a^{-H} X_{at}$$

  is the same as that of X_t. In particular, X_t has the same distribution as $t^H X_1$.

Our final assumption is that the process is a centered Gaussian process.

- For every $t_1, \ldots, t_k$, the random vector $(X_{t_1}, \ldots, X_{t_k})$ has a joint normal distribution with mean zero.

We normalize so that $\mathrm{Var}(X_1) = \mathbb{E}[X_1^2] = 1$ and hence

$$\mathrm{Var}(X_t) = \mathrm{Var}(t^H X_1) = t^{2H} \mathrm{Var}(X_1) = t^{2H}.$$

Definition If $H \in (0,1)$, the *fractional Brownian motion with Hurst parameter H (fBm_H)* is the centered (mean zero) Gaussian process X_t with continuous paths such that for all s, t,

$$\mathbb{E}\left[(X_t - X_s)^2\right] = |t - s|^{2H}.$$

Since
$$\mathbb{E}\left[(X_t - X_s)^2\right] = \mathbb{E}\left[X_t^2\right] + \mathbb{E}\left[X_s^2\right] - 2\mathbb{E}\left[X_s X_t\right],$$

it follows that

$$\mathrm{Cov}(X_s, X_t) = \mathbb{E}\left[X_s X_t\right] = \frac{1}{2}\left[s^{2H} + t^{2H} - |s - t|^{2H}\right]. \qquad (7.1)$$

If $H = 1/2$, then fractional Brownian motion is the same as usual Brownian motion.

As in the case of Brownian motion, we must show that such a process exists. We will discuss this in the next section, but for now we assume it does exist. If $s < t$, note that

$$
\begin{aligned}
\mathbb{E}\left[X_s\left(X_t - X_s\right)\right] &= \mathbb{E}\left[X_s X_t\right] - \mathbb{E}\left[X_s^2\right] \\
&= \frac{1}{2}\left[t^{2H} - s^{2H} - (t - s)^{2H}\right] \\
&\begin{cases} > 0, & H > 1/2 \\ = 0, & H = 1/2 \\ < 0, & H < 1/2 \end{cases}.
\end{aligned}
$$

Since $X_{t+\delta} - X_t \sim N(0, \delta^{2H})$, we can write roughly

$$|X_{t+\delta} - X_t| \approx \delta^H.$$

In other words, the Hölder exponent of fBm_H is given by the Hurst index H. If $H > 1/2$, the paths are "smoother" than Brownian paths and if $H < 1/2$, the paths are "rougher".

To determine correlations for large t, suppose t is large. Then

$$\mathbb{E}[(X_{s+1} - X_s)(X_{s+t+1} - X_{s+t}]$$
$$= \mathbb{E}[X_1 (X_{t+1} - X_t)]$$
$$= \mathbb{E}[X_1 X_{t+1}] - \mathbb{E}[X_1 X_t]$$
$$= \frac{1}{2} \left[(t+1)^{2H} + (t-1)^{2H} - 2t^{2H} \right]$$
$$\sim H (2H - 1) t^{2H-2}.$$

The coefficient is positive for $H > 1/2$ and negative for $H < 1/2$. As t goes to infinity the size of the correlation goes to zero like a power law.

7.2 Stochastic integral representation

We will give an expression for the fractional Brownian motion in terms of a stochastic integral. It will be useful to consider time going from $-\infty$ to ∞. Suppose B_t^1, B_t^2 are independent standard Brownian motions starting at the origin. If

$$B_t = \left\{ \begin{array}{ll} B_t^1, & t \geq 0 \\ B_{-t}^2, & t \leq 0 \end{array} \right. ,$$

then B_t is standard Brownian motion from time $-\infty$ to ∞. The centering $B_0 = 0$ is rather arbitrary; we really think of this process as a collection of increments $\{B_t - B_s\}$ for $s < t$. Viewed this way, this is sometimes called *white noise*. We let $\mathcal{F}_t$ denote the information contained in the random variables $\{B_s - B_r : r < s \leq t\}$.

Suppose $f(r, t)$ is a continuous (nonrandom) function, and let

$$Y_t = \int_{-\infty}^t f(r, t) \, dB_r = \lim_{n \to \infty} \sum f \left(\frac{k}{n}, t \right) \Delta B(k, n),$$

where

$$\Delta B(k, n) = B_{(k+1)/n} - B_{k/n},$$

and the sum is over all integers (positive and negative) with $(k+1)/n \leq t$. Then Y_t is a limit of centered normal random variables and hence is mean zero with variance

$$\mathbb{E}[Y_t^2] = \int_{-\infty}^t f(r, t)^2 \, dr.$$

We assume this is finite for each t. Moreover if $s < t$, we can write

$$Y_t - Y_s = \int_{-\infty}^{s} [f(r,t) - f(r,s)]\, dB_r + \int_{s}^{t} f(r,t)\, dB_r.$$

The right-hand side is the sum of two independent normal random variables: the first is $\mathcal{F}_s$-measurable and the second is independent of $\mathcal{F}_s$. Hence $Y_t - Y_s$ has a normal distribution. More generally, one can check that Y_t is a Gaussian process whose covariance is given for $s < t$ by

$$\mathbb{E}\left[Y_s Y_t\right] = \int_{-\infty}^{s} f(r,s)\, f(r,t)\, dr.$$

If we make the further assumption that $f(r,t) = \phi(t-r)$ for a one-variable function ϕ, then the process Y_t is stationary and has the form

$$Y_t = \int_{-\infty}^{t} \phi(t-r)\, dB_r. \tag{7.2}$$

Proposition 7.2.1. *If*

$$\phi(s) = c\, s^{H-\frac{1}{2}},$$

and Y_t is defined as in (7.2), then $X_t = Y_t - Y_0$ is fBm_H. Here

$$c = c_H = \left[\frac{1}{2H} + \int_{0}^{\infty} [(1+r)^{H-\frac{1}{2}} - r^{H-\frac{1}{2}}]^2\, dr\right]^{-1/2}.$$

Proof. By construction, X_t is a Gaussian process with continuous paths. Since Y_t is stationary, the distribution of $X_t - X_s = Y_t - Y_s$ is the same as that of $X_{t-s} = Y_{t-s} - Y_0$. In particular, $\mathbb{E}[(X_t - X_s)^2] = \mathbb{E}[X_{t-s}^2]$. Hence we only need to show that $\mathbb{E}[X_t^2] = t^{2H}$. Note that

$$X_t = \int_{-\infty}^{0} [\phi(t-r) - \phi(-r)]\, dB_r + \int_{0}^{t} \phi(t-r)\, dB_r.$$

The random variables on the right-hand side are independent and

$$c^{-2} \operatorname{Var} \left[\int_{-\infty}^{0} [\phi(t-r) - \phi(-r)] \, dB_r \right]$$

$$= \int_{-\infty}^{0} [(t-r)^{H-\frac{1}{2}} - (-r)^{H-\frac{1}{2}}]^2 \, dr$$

$$= \int_{0}^{\infty} [(t+r)^{H-\frac{1}{2}} - r^{H-\frac{1}{2}}]^2 \, dr$$

$$= \int_{0}^{\infty} [(t+st)^{H-\frac{1}{2}} - (st)^{H-\frac{1}{2}}]^2 \, t \, ds$$

$$= t^{2H} \int_{0}^{\infty} [(1+s)^{H-\frac{1}{2}} - s^{H-\frac{1}{2}}]^2 \, ds,$$

$$c^{-2} \operatorname{Var} \left[\int_{0}^{t} \phi(t-r) \, dB_r \right] = \int_{0}^{t} (t-r)^{2H-1} \, dr = \frac{t^{2H}}{2H}.$$

If we choose c as in (7.3), we get $\mathbb{E}[X_t^2] = t^{2H}$. □

7.3 Simulation

Because the fractional Brownian motion has long range dependence it is not obvious how to do simulations. The stochastic integral represetation (7.2) is difficult to use because it uses the value of the Brownian motion for all negative times. However, there is a way to do simulations that uses only the fact that fBm_H is a Gaussian process with continuous paths. Let us choose a step size $\Delta t = 1/N$; continuity tells us that it should suffice to sample

$$Y_1, Y_2, \ldots$$

where $Y_j = X_{j/N}$. For each n, the random vector $(Y_1, \ldots, Y_n)$ has a centered Gaussian distribution with covariance matrix $\Gamma = [\Gamma_{jk}]$. Given Γ we claim that we can find numbers a_{jk} with $a_{jk} = 0$ if $k > j$, and independent standard normal random variables $Z_1, Z_2, \ldots$ such that

$$Y_n = a_{n1} Z_1 + \cdots + a_{nn} Z_n. \tag{7.3}$$

In matrix notation, $A = [a_{jk}]$ is a lower triangular matrix such that $\Gamma = AA^T$.

This decomposition $\Gamma = AA^T$ is sometimes called the *Cholesky decomposition*. We will now show that it exists by giving an algorithm for finding the matrix. We start by setting

$$a_{11} = \sqrt{\Gamma_{11}}.$$

Suppose we have found the first $k - 1$ rows of A. This is a lower triangular $(k-1) \times (k-1)$ matrix. Suppose $j < k$. Then,

$$\Gamma_{jk} = \mathbb{E}\left[Y_j Y_k\right] = \sum_{i=1}^{j} a_{ji}\, a_{ki}.$$

The coefficients a_{jk} for $j = 1, \ldots, k - 1$ can be found by solving these $k - 1$ linear equation in $k - 1$ unknowns (one uses the fact that Γ is nonnegatve definite to show there is a unique solution), and then one sets

$$a_{kk} = \sqrt{\Gamma_{kk} - \sum_{i=1}^{k-1} a_{ki}^2}.$$

We see that this decomposition is unique if we impose the condition $a_{kk} \geq 0$.

¿From a practical perspective, these computations are done with computer packages. Note that one only needs to compute the coefficients a_{jk} once and store them. Then multiple simulations can be done using (7.3).

The Cholesky decomposition can be derived from the Gram-Schmidt process. Consider the Hilbert space of mean-zero L^2 random variables and let H_n denote the finite dimensonal subspace spanned by $Y_1, \ldots, Y_n$. Let P_n denote the projection onto H_n which is the same as the conditional expectation,

$$P_n X = E[X \mid Y_1, \ldots, Y_n].$$

Then we define
$$Z_n = \frac{Y_n - P_{n-1}Y_n}{\|Y_n - P_{n-1}Y_n\|_2}.$$

Since H_n is also the subspace spanned by $Z_1, \ldots, Z_n$, we can write

$$P_{n-1}Y_n = \sum_{j=1}^{n-1} a_{jn}\, Z_j,$$

for unique scalars a_{jn}, and we set $a_{nn} = \|Y_n - P_{n-1}Y_n\|_2$.

Chapter 8

Harmonic functions

8.1 Dirichlet problem

Recall that the Laplacian of a function f on $\mathbb{R}^d$ is defined by

$$\nabla^2 f(x) = \sum_{j=1}^{d} \partial_{jj} f(x).$$

The Laplacian is closely related to mean values of functions. The *(spherical) mean value of f on the sphere of radius ϵ about x* is defined by

$$MV(x; \epsilon, f) = \frac{\int_{|y-x|=\epsilon} f(y)\, d\sigma(y)}{\int_{|y-x|=\epsilon} 1\, d\sigma(y)},$$

where σ denotes $(d-1)$-dimensional surface measure. The mean value can also be described in terms of Brownian motion. Suppose B_t is a d-dimensional Brownian motion starting at x and let

$$\tau_\epsilon = \inf\{t : |B_t - B_0| = \epsilon\}.$$

Since Brownian motion is rotationally invariant, the distribution of B_{τ_ϵ} is uniform over the sphere $\{y : |y - x| = \epsilon\}$. Hence we can write

$$MV(x; \epsilon, f) = \mathbb{E}^x\left[f(B_{\tau_\epsilon})\right]. \tag{8.1}$$

Proposition 8.1.1. *Suppose f is a C^2 function in a neighborhood of $x \in \mathbb{R}^d$. Then*

$$\nabla^2 f(x) = 2d \lim_{\epsilon \downarrow 0} \frac{MV(x; \epsilon, f) - f(x)}{\epsilon^2}.$$

Proof. Without loss of generality, assume $x = 0$ and $f(x) = 0$, and write $MV(\epsilon)$ for $MV(0; f, \epsilon)$. Let $b_j = \partial_j f(0), a_{jk} = \partial_{jk} f(0)$. Taylor's theorem implies that

$$f(y) = \sum_{j=1}^{d} b_j\, y_j + \frac{1}{2} \sum_{j=1}^{d} a_{jk} y_j y_k + o(|y|^2), \quad y \to 0,$$

where we write $y = (y_1, \ldots, y_d)$. Therefore,

$$MV(\epsilon) = \sum_{j=1}^{d} b_j\, MV(0; y_j, \epsilon) + \frac{1}{2} \sum_{j=1}^{d} a_{jk}\, MV(0; y_j y_k, \epsilon) + o(\epsilon^2).$$

Note that $MV(0; y_j, \epsilon) = 0$ since y_j is an odd function. Similarly, $MV(0; y_j y_k, \epsilon) = 0$ if $j \neq k$. Linearity of the integral implies that

$$\sum_{j=1}^{d} MV(0; y_j^2, \epsilon) = MV(0; \sum_{j=1}^{d} y_j^2, \epsilon) = MV(0; \epsilon^2, \epsilon) = \epsilon^2.$$

Symmetriy implies that $MV(0; y_j^2, \epsilon) = MV(0; y_k^2, \epsilon)$, and hence $MV(y_j^2, \epsilon) = \epsilon/d$. We therefore have

$$MV(\epsilon) = \frac{1}{2} \sum_{j=1}^{d} a_{jj} \left(\epsilon^2/d \right) + o(\epsilon^2) = \frac{\epsilon^2}{2d} \nabla^2 f(0) + o(\epsilon^2).$$

$\square$

Definition Suppose f is a function defined on an open set $D \subset \mathbb{R}^d$.

- f is *harmonic in* D if f is C^2 and $\nabla^2 f(x) = 0$ for all $x \in D$. f is harmonic at x if it is harmonic in a neighborhood of x.

- A function f satisfies the *(spherical) mean value property* on D if for every $x \in D$ and every $\epsilon > 0$ with $\{y : |x - y| \leq \epsilon\} \subset D$,

$$f(x) = MV(f; x, \epsilon).$$

Proposition 8.1.2.

- *Suppose f is a harmonic function defined on an open set $D \subset \mathbb{R}^d$. Let B_t be a standard d-dimensional Brownian motion starting at $x \in D$ and let*

$$\tau = \inf\{t : B_t \notin D\}.$$

  Then $M_t := f(B_{t \wedge \tau})$ is a local martingale for $t < \tau$ satisfying

$$dM_t = \nabla f(B_t) \cdot dB_t, \quad t < \tau.$$

- *Suppose also that f is defined on ∂D so that $f : \overline{D} \to \mathbb{R}$ is a bounded, continuous function. Then M_t is a bounded continuous martingale. If $\mathbb{P}\{\tau < \infty\} = 1$, then*

$$f(x) = \mathbb{E}^x\left[f(B_\tau)\right]. \tag{8.2}$$

Proof. The first statement follows from Itô's formula. In fact, we already did this calculation in Theorem 3.7.2. Since bounded local martingales are martingales, the second statement follows from the optional sampling theorem. $\square$

The expression (8.2) is a generalization of the mean value property (8.1). It states that we can take mean values over sets other than spheres as long as integrate with respect to the correct the measure. This measure is often called *harmonic measure (in D starting at x)* and is defined by

$$\mathrm{hm}_D(V; x) = \mathbb{P}^x\{B_\tau \in V\}, \quad V \subset \partial D.$$

Then (8.2) can be written

$$f(x) = \int_{\partial D} f(y) \, \mathrm{hm}_D(dy; x).$$

The next proposition shows that we could use the mean value propery to define harmonic functions. In fact, this is the more natural definition.

Proposition 8.1.3. *Suppose f is a continuous function on an open set $D \subset \mathbb{R}^d$. Then f is harmonic in D if and only f satisfies the mean value property on D.*

Proof. If f is harmonic, then f restricted to a closed ball of radius ϵ contained in D is bounded. Therefore, the mean value property is a particular case of (8.2).

If f is C^2 and satisfies the mean value property, then $\nabla^2 f(x) = 0$ by Proposition 8.1.1. Hence we need only show that f is C^2. We will, in fact, show that f is C^∞.

It is a standard exercise in advanced calculus to show that for every $\delta > 0$ we can find a nonnegative radially symmetric C^∞ function $\phi = \phi_\delta$ such that $\phi(y) = 0$ for $|y| \geq \delta$ and
$$\int \phi(y)\,dy = 1.$$

Here $dy = dy_1 \cdots dy_d$. If $x \in D$ with $\{y : |x - y| \leq \epsilon\} \subset D$, and $\phi = \phi_{\epsilon/2}$, then the mean value property implies that

$$f(x) = \int f(x + y)\,\phi(y)\,dy = \int f(z)\,\phi(z - x)\,dz. \qquad (8.3)$$

Since f is locally bounded and ϕ is bounded and C^∞, we can differentiate with respect to x by bringing the derivatives inside the integral. $\square$

The proof showed that we did not need to assume that f is continuous. It suffices for f to be measurable and locally bounded so that derivatives can be taken on the right-hand side of (8.3).

We will solve the classic Dirichlet problem for harmonic functions. Suppose D is a bounded open set and $F : \partial D \to \mathbb{R}$ is a continuous function. The goal is to find a continuous function $f : \overline{D} \to \mathbb{R}$ that is harmonic in D with $f(x) = F(x)$ for $x \in \partial D$. Suppose that such a function f existed. Let $\tau = \tau_D = \inf\{t \geq 0 : B_t \in \partial D\}$. Since $\overline{D}$ is compact, f must be bounded, and hence
$$M_t = f(B_{t \wedge \tau})$$
is a continuous, bounded martingale. Arguing as in (8.2) we see that

$$f(x) = \mathbb{E}^x\left[f(B_\tau)\right] = \mathbb{E}^x\left[F(B_\tau)\right]. \qquad (8.4)$$

The right-hand side gives the only candidate for the solution. The strong Markov property can be used to see that this candidate satisfies the mean value property and the last proposition gives that f is harmonic in D.

It is a little more subtle to check if f is continuous on $\overline{D}$. This requires further assumptions on D which can be described most easily in terms of Brownian motion. Suppose $z \in \partial D$ and $x \in D$ with x near z. Can we say that $f(x)$ is near $F(z)$? Since F is continuous, this will be true if B_τ is near

z. To make this precise, one defines a point $z \in \partial D$ to be *regular* if for every $\epsilon > 0$ there exists $\delta > 0$ such that if $x \in D$ with $|x - z| < \delta$, then

$$\mathbb{P}^x \{|B_\tau - z| \geq \epsilon\} \leq \epsilon.$$

Theorem 8.1.4. *Suppose D is a bounded open set such that each $z \in \partial D$ is regular. Suppose F is a continuous function on ∂D. Then there exists a unique continuous function $f : \overline{D} \to \mathbb{R}$ that is harmonic on D and agrees with F on ∂D given by* (8.4).

Example 8.1.1. Let D be the annular region

$$D = \{x \in \mathbb{R}^d : r < |x| < R\},$$

and let $F(x) = 1$ if $|x| = R$, $F(x) = 0$ if $|x| = r$. Then for $x \in D$,

$$f(x; r, R) = \mathbb{E}^x [F(B_\tau)] = \mathbb{P}^x \{|B_\tau| = R\}.$$

We claim that

$$f(x; r, R) = \frac{x - r}{R - r}, \quad d = 1,$$

$$f(x; r, R) = \frac{\log |x| - \log r}{\log R - \log r}, \quad d = 2,$$

$$f(x; r, R) = \frac{r^{d-2} - |x|^{d-2}}{r^{d-2} - R^{d-2}}, \quad d \geq 3.$$

One can check this by noting that $\nabla^2 f(x) = 0$ and f has the correct boundary condition. The theorem implies that there is only one such function. Note that

$$\lim_{R \to \infty} f(x; r, R) = \begin{cases} 0, & d \leq 2 \\ 1 - (r/|x|)^{2-d}, & d \geq 3 \end{cases},$$

$$\lim_{r \downarrow 0} f(x; r, R) = \begin{cases} x/R & d = 1 \\ 1, & d \geq 2 \end{cases}. \tag{8.5}$$

We have already seen the $d = 1$ case as the gambler's ruin estimate for Brownian motion.

Example 8.1.2. Let $d \geq 2$. It follows from (8.5) that if $x \neq 0$, then the probability that a Brownian motion starting at x ever hits zero is zero. Let $D = \{x \in \mathbb{R}^d : 0 < |x| < 1\}$. Then 0 is *not* a regular point of ∂D, since if we start near the origin the Brownian motion will not exit D near 0. Let

$F(0) = 0$, $F(y) = 1$ if $|y| = 1$. Then, the only candidate for the solution of the Dirichlet problem is

$$f(x) = \mathbb{E}^x \left[F(B_\tau) \right] = 1, \quad x \in D.$$

If $f(0) = 0$, this function is not continuous at the origin.

If $D = U_r = \{x : |x| < r\}$ is the ball of radius r about the origin, then the harmonic measure $\mathrm{hm}_{U_r}(\cdot; x)$ is known explicitly. It is absolutely continuous with respect to $(d-1)$-dimensional surface measure s on ∂U_r. Its density is called the *Poisson kernel*,

$$H_r(x, y) = \frac{r^2 - |x|^2}{r\, \omega_{d-1}\, |x - y|^d},$$

where

$$\omega_{d-1} = \int_{|y|=1} ds(y)$$

denotes the surface area of the unit ball. If $F : \partial U_r \to \mathbb{R}$ is continuous, then the unique solution to the Dirichlet problem is

$$f(x) = \mathbb{E}^x \left[F(B_{\tau_{U_r}}) \right] = \int_{|y|=r} F(y)\, H_r(x, y)\, ds(y). \tag{8.6}$$

To verify this, one checks by direct computation the following facts:

- If $y \in \partial U_r$, the function $h(x) = H(x, y)$ is harmonic on U_r;

- If $x \in U_r$,
$$\int_{|y|=r} H_r(x, y)\, ds(y) = 1;$$

- For every $\epsilon > 0$, there exists $\delta > 0$ such that if $x \in U_r, y \in \partial U_r$ and $|x - y| < \delta$, then
$$\int_{|z|=r, |z-y| \geq \epsilon} H_r(x, z)\, ds(z) \leq \epsilon.$$

¿From these one concludes that f as defined by the right-hand side of (8.6) is harmonic in U_r and continuous on $\overline{U}_r$.

The reader may note that we did not need the probabilistic interpretation of the solution in order to verify that (8.6) solves the Dirichlet problem. Indeed, the solution using the Poisson kernel was discovered before the relationship with Brownian motion. An important corollary of this explicit solution is the following theorem; we leave the verification as Exercise 8.1. The key part of the theorem is the fact that the same constant C works for all harmonic functions.

Theorem 8.1.5.

1. *For every positive integer n, there exists $C = C(d, n) < \infty$ such that if f is a harmonic function on an open set $D \subset \mathbb{R}^d$, $x \in D$, $\{y : |x - y| < r\} \subset D$, and $j_1, \ldots, j_n \in \{1, \ldots, d\}$ then*

$$\left| \partial_{x_{j_1}} \cdots \partial_{x_{j_n}} f(x) \right| \le C \, r^{-n} \sup_{|y-x|<r} |f(y)|.$$

2. *(Harnack inequality) For every $0 < u < 1$, there exists $C = C(d, u) < \infty$ such that if f is a positive harmonic function on an open set $D \subset \mathbb{R}^d$, $x \in D$, $\{y : |x - y| < r\} \subset D$, then if $|x - z| \le ur$,*

$$C^{-1} f(x) \le f(z) \le C f(x).$$

8.2 h-processes

Suppose h is a positive harmonic function on an open set $D \subset \mathbb{R}^d$, and let B_t be a standard Brownian motion starting at $x \in D$. Let $\tau = \tau_D$ be the first time t with $B_t \notin D$. Then $M_t = h(B_t)$ is a positive local martingale for $t < \tau$ satisfying

$$dM_t = \frac{\nabla h(B_t)}{h(B_t)} M_t \, dB_t, \quad t < \tau.$$

Let τ_n be the minimum of τ and the first time t with $h(B_t) \ge n$. Then $M_{t \wedge \tau_n}$ is a continuous bounded martingale.

We can use the Girsanov theorem to consider the measure obtained by weighting by the local martingale M_t. To be more precise, if V is an event that depends only on $B_t, 0 \le t \le \tau_n$, then

$$\mathbb{P}^*(V) = h(x)^{-1} \, \mathbb{E}^* \left[M_{\tau_n} \, 1_V \right].$$

One can use the Girsanov theorem (more precisely, a simple generalization of the theorem to d-dimensional Brownian motion), to see that

$$dB_t = \frac{\nabla h(B_t)}{h(B_t)} dt + dW_t,$$

where W_t is a standard Brownian motion with respect to $\mathbb{P}^*$. The process B_t in the new measure $\mathbb{P}^*$ is often called the *(Doob) h-process* associated to the positive harmonic function h. It is defined for $t < \tau$.

As an example, suppose that D is the unit ball, $y = (1, 0, \ldots, 0) \in \partial D$ and

$$h(x) = \omega_{d-1} H_1(x, y) = \frac{1 - |x|^2}{|x - y|^d},$$

is the Poisson kernel. Then the h-process can be viewed as Brownian motion "conditioned so that $B_\tau = y$", where $\tau = \tau_D$. This is not precise because the conditioning is with respect to a event of probability zero. We claim that the $\mathbb{P}^*$-probability that $B_\tau = y$ equals one. To see this, assume that $B_0 = 0$ and let

$$T_n = \inf\{t : h(B_t) = n^3\}, \quad T'_n = \inf\{t > T_n : h(B_t) = n\}.$$

We claim that $\mathbb{P}^*\{T_n < \tau\} = 1$. Indeed, if we let $\tau_r = \inf\{t : |B_t| = r\}$, then we can check directly that

$$\lim_{r \uparrow 1} \mathbb{P}^*\{h(B_{\tau_r}) \geq n^3\} = \lim_{r \uparrow 1} \int_{|x|=1, h(rx) \geq n^3} h(rx) \, d\sigma(x) = 1,$$

and hence

$$\lim_{r \uparrow 1} \mathbb{P}^*\{T_n < \tau_r\} = 1.$$

Also, since $h(B_{T'_n}) = n^{-2} h(B_{T_n})$,

$$\mathbb{P}^*\{T'_n < \tau\} \leq n^{-2},$$

$$\sum_{n=1}^{\infty} \mathbb{P}^*\{T'_n < \tau\} < \infty.$$

Therefore, by the Borel-Canelli lemma, we see that with $\mathbb{P}^*$-probability one for all n sufficiently large $h(B_t) \geq n$ for $t \geq T_n$.

8.3 Time changes

To prepare for the next section we will consider time changes of Brownian motion. If B_t is a standard Brownian motion and $Y_t = B_{at}$, then Y_t is a Brownian motion with variance parameter a. We can write

$$Y_t = \int_0^t \sqrt{a}\, dW_s,$$

where $W_t = a^{-1/2} B_{at}$ is a standard Brownian motion. We generalize this by considering $Y_t = B_{a(t)}$ where a is a strictly increasing continuous function. We allow $a(t)$ to be random, but it must be adapted to the Brownian motion. In other words, for each t, the random time $a(t)$ must be a stopping time for the Brownian motion B_t. We will make the further assumption that a is a C^1 function, that is, we can write

$$a(t) = \int_0^t \dot{a}(s)\, ds,$$

where $s \mapsto \dot{a}(s)$ is a continuous nonnegative function. The assumption that $a(t)$ is strictly increasing implies that there are no open intervals on which $\dot{a}(s)$ is identically zero. If $Y(t) = B_{a(t)}$, then Y_t is a continuous local martingale and we can write

$$Y_t = \int_0^t \sqrt{\dot{a}(s)}\, dW_s,$$

where W_t is the standard Brownian motion

$$W_t = \int_0^t [\dot{a}(s)]^{-1/2}\, dB_{a(s)}. \tag{8.7}$$

One can verify this is a standard Brownian motion by checking that is a martingale with $\langle W \rangle_t = t$. More generally, if X_t satisfies the SDE

$$dX_t = R_t\, dt + A_t\, dB_t,$$

and $Y_t = X_{a(t)}$, then Y_t satisfies the SDE

$$dY_t = \dot{a}(t)\, R_{a(t)}\, dt + \sqrt{\dot{a}(t)}\, A_{a(t)}\, dW_t,$$

where W_t is the standard Brownian motion (8.7).

Example 8.3.1. Suppose that X_t satisfies the Bessel equation

$$dX_t = \frac{r}{X_t}\, dt + dB_t, \quad X_0 = 1,$$

where $r \in \mathbb{R}$. This is defined for $t < T = \inf\{s : X_s = 0\}$. For $t < T$, let $L_t = \log X_t$. Then Itô's formula shows that

$$
\begin{aligned}
dL_t &= \frac{1}{X_t}\, dX_t - \frac{1}{2\, X_t^2}\, d\langle X\rangle_t \\
&= \left(r - \frac{1}{2}\right) e^{-2L_t}\, dt + e^{-L_t}\, dB_t.
\end{aligned}
$$

Consider the time change with $\dot{a}(t) = e^{2L_t}$, and let $Y_t = L_{a(t)}$. Then Y_t satisfies

$$dY_t = \left(r - \frac{1}{2}\right) dt + dW_t,$$

where W_t is a standard Brownian motion. In other words, Y_t is a Brownian motion with variance parameter 1 and drift $m = r - \frac{1}{2}$. From this we can see that $Y_t \to \infty$ if $r > 1/2$, $Y_t \to -\infty$ if $r < 1/2$, and

$$-\infty = \liminf_{t\to\infty} Y_t < \limsup_{t\to\infty} Y_t = \infty,$$

if $r = 1/2$. One can use this to give another proof of Proposition 4.2.1. Note that if $r < 1/2$, then in the new time, it take an infinite amount of time for $L = \log X$ to reach $-\infty$. However, in the original time, this happens at time $T < \infty$. In other words,

$$a(T) = \int_0^T \frac{ds}{\dot{a}(s)} = \infty.$$

8.4   Complex Brownian motion

If B_t^1, B_t^2 are independent standard Brownian motions then

$$B_t = B_t^1 + i\, B_t^2$$

is called a *(standard) complex Brownian motion*. Note that this is the same as a two-dimensional Brownian motion except that the point in the plane is viewed as a complex number.

Suppose $f : \mathbb{C} \to \mathbb{C}$ is a nonconstant holomorphic (complex differentiable) function. We will consider $X_t = f(B_t)$. Near any point z with $f'(z) \neq 0$, the function looks like $f(w) = f(z) + f'(z)(w - z)$. Multiplication by $f'(z)$ is the same as a dilation by $|f'(z)|$ and a rotation by $\arg f'(z)$. A rotation of a standard two-dimensional Brownian motion gives a standard two-dimensional Brownian motion and a dilation gives a Brownian motion with a different variance. This leads us to believe that X_t is a time-change of standard Brownian motion.

Let us make this more precise. Let

$$a(t) = \int_0^t |f'(B_s)|^2 \, ds.$$

We consider the time-change in which it takes time $a(t)$ to traverse $f(B[0, t])$. Equivalently, in time t one has traversed $f(B[0, a^{-1}(t)])$.

Theorem 8.4.1. *Let $Y_t = X_{a^{-1}(t)}$. Then Y_t is a standard complex Brownian motion.*

Proof. Write $f = u + iv$, $X_t = U_t + iV_t$, $Y_t = \hat{U}_t + i\hat{V}_t$. The Cauchy-Riemann equations imply that $\partial_x u = \partial_y v$, $\partial_y u = -\partial_x v$ and

$$\nabla^2 u = \partial_x \partial_x u + \partial_y \partial_y u = \partial_x \partial_y v + \partial_y(-\partial_x v) = 0.$$

Similarly $\nabla^2 v = 0$. Since $U_t = u(B_t^1, B_t^2)$, $V_t = v(B_t^1, B_t^2)$, then Itô's formula (see Theorem 3.7.2) gives

$$dU_t = \nabla u(B_t) \cdot dB_t = \partial_x u(B_t) \, dB_t^1 + \partial_y u(B_t) \, dB_t^2.$$

$$\begin{aligned} dV_t = \nabla v(B_t) \cdot dB_t &= \partial_x v(B_t) \, dB_t^1 + \partial_y v(B_t) \, dB_t^2 \\ &= -\partial_y u(B_t) \, dB_t^1 + \partial_x u(B_t) \, dB_t^2. \end{aligned}$$

Note that

$$\begin{aligned} d\langle U \rangle_t &= \left([\partial_x u(B_t)]^2 + [\partial_y u(B_t)]^2\right) dt \\ &= \left([\partial_x u(B_t)]^2 + [-\partial_x v(B_t)]^2\right) dt = |f'(B_t)|^2 \, dt, \\ d\langle V \rangle_t &= |f'(B_t)|^2 \, dt, \\ d\langle U, V \rangle_t &= 0. \end{aligned}$$

Therefore, $\hat{U}_t, \hat{V}_t$ are martingales with quadratic variation t and hence are standard Brownian motions. Since $\langle U, V \rangle_t = 0$, we also have $\langle \hat{U}, \hat{V} \rangle_t = 0$, and hence $\hat{U}_t, \hat{V}_t$ are independent. $\qquad \square$

The last theorem assumes that the function f is entire, that is, it is defined and holomorphic on all of $\mathbb{C}$. Suppose D is *domain*, that is, an open, connected subset of $\mathbb{C}$, and suppose that f is a nonconstant holomorphic function on D. Let B_t be a complex Brownian motion, and let

$$\tau_D = \inf\{t \geq 0 : B_t \notin D\}.$$

Then the process X_t as above is defined for $t < \tau_D$, and the process Y_t is defined for $t < a(\tau_D-)$ with the processes taking values in $f(D)$.

Theorem 8.4.2. *If f is a holomorphic function on a domain D, and X_t, Y_t are defined as above, then $Y_t, 0 \leq t < a(\tau_D-)$ is a standard complex Brownian motion.*

One very important case is when f maps D one-to-one onto $f(D)$. Such functions are often called *conformal transformations*. In this case f^{-1} is a conformal transformation from $f(D)$ to D, and

$$a(\tau_D-) = \inf\{t : Y_t \notin f(D)\}.$$

In particular, this shows that harmonic measure is a conformal invariant,

$$\mathrm{hm}_D(V; z) = \mathrm{hm}_{f(D)}(f(V); f(z)).$$

8.5   Exercises

Exercise 8.1. Use (8.6) to prove Theorem 8.1.5.

Exercise 8.2. Suppose $D \subset \mathbb{C}$ is a domain and ϕ is a harmonic function on D (here we are also viewing D as a subset of $\mathbb{R}^2$). Let $f : D \to f(D)$ be a conformal transformation and let

$$\psi(w) = \phi(f^{-1}(w)).$$

Show that ψ is a harmonic function on $f(D)$ in two different ways:

- Do a direct computation to show that $\nabla^2\psi = 0$.

- Use the conformal invariance of harmonic measure and (8.2).

Chapter 9

Additional topics

Here we include some additional topics that are covered in the advanced version of this class. Although we use the regular font, this material should be considered at the level of the alternate font, that is, for readers who have a background in measure theory and have been reading both fonts.

9.1   Zero-One Laws

Suppose B_t is a Brownian motion which for convenience we will assume has continuous paths everywhere (not just on an event of probability one). Let

$$\mathcal{F}_t = \sigma\{B_s : s \le t\}, \quad \mathcal{G}_t = \sigma\{B_{s+t} - B_t : s > 0\}.$$

Let $\mathcal{F}_\infty$ denote the smallest σ-algebra containing each $\mathcal{F}_t$. By continuity we see that $\mathcal{F}$ is also the σ-algebra generated by $\cup_{t>0}\mathcal{G}_t$. The assumptions of Brownian motion imply that for each t, $\mathcal{F}_t$ and $\mathcal{G}_t$ are independent σ-algebras. We define the tail σ-algebras at the beginning and the end, respectively by

$$\mathcal{F}_{0+} = \bigcap_{t>0} \mathcal{F}_t, \quad \mathcal{G}_\infty = \bigcap_{t>0} \mathcal{G}_t.$$

An example of an event in $\mathcal{F}_{0+}$ is

$$\{\text{for all } \epsilon > 0 \text{ there exists } 0 < t < \epsilon \text{ with } B_t = 0\}.$$

An example in $\mathcal{G}_\infty$ is

$$\left\{\limsup_{t\to\infty} \frac{B_t}{\sqrt{2t \log\log t}} = 1\right\}.$$

Theorem 9.1.1. *[Blumenthal 0-1 Law] If $A \in \mathcal{F}_{0+}$ or $A \in \mathcal{G}_\infty$, then $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.*

This theorem is a straightforward generalization of the Kolmogorov 0-1 Law which concerns sequences of independent random variables. We will discuss the proof for $\mathcal{F}_{0+}$. Let $\mathcal{F}^0 = \bigcup_{t>0} \mathcal{F}_t$ and note that $\mathcal{F}^0$ is an algebra of sets, that is, closed under complements and finite unions, with $\sigma(\mathcal{F}^0) = \mathcal{F}_\infty$. The main lemma used is the following.

- If $\mathcal{F}^0$ is an algebra of subsets and $\mathcal{F} = \sigma(\mathcal{F}^0)$, then for each $A \in \mathcal{F}$ and $\epsilon > 0$, there exists $A_0 \in \mathcal{F}^0$ with $\mathbb{P}(A \Delta A_0) < \epsilon$. Here Δ denotes symmetric difference, $A \Delta A_0 = (A \setminus A_0) \cup (A_0 \setminus A)$.

To see this, one needs only show that the set of $A \in \mathcal{F}$ that can be approximated by sets in $\mathcal{F}^0$ as above is a σ-algebra containing $\mathcal{F}^0$.

Using the lemma, if $A \in \mathcal{F}_{0+}$ and $\epsilon > 0$, we can find $A_0 \in \mathcal{F}^0$ with $\mathbb{P}(A \Delta A_0) < \epsilon$. But A, A_0 are independent, so

$$\mathbb{P}(A) - \epsilon \leq \mathbb{P}(A \cap A_0) = \mathbb{P}(A)\, \mathbb{P}(A_0) \leq \mathbb{P}(A)\, [\mathbb{P}(A) + \epsilon].$$

Letting $\epsilon \downarrow 0$, we get $\mathbb{P}(A) \leq \mathbb{P}(A)^2$ which can only hold if $\mathbb{P}(A) \in \{0, 1\}$.

9.2 Law of the Iterated Logarithm

In this section we will prove the following theorem. Throughout this section, B_t will denote a standard one-dimensional Brownian motion and let

$$Y_t = \frac{B_t}{\sqrt{t}}.$$

Note that Y_t has a standard normal distribution for each t.

Theorem 9.2.1 (Law of the Iterated Logarithm). *With probability one,*

$$\limsup_{t \to \infty} \frac{Y_t}{\sqrt{2 \log \log t}} = 1.$$

Before giving the proof, we will show where the denominator comes from. Recall that as $x \to \infty$,

$$\mathbb{P}\{Y_t > x\} \sim \frac{c}{x}\, e^{-x^2/2}. \tag{9.1}$$

Let $\rho > 1, t_n = \rho^n$ and for each $r > 0$, let

$$V_n = \{Y_{t_n} > \sqrt{2r \log \log t_n}\}.$$

Using (9.1), we see that

$$\mathbb{P}(V_n) \sim \frac{c^*}{\sqrt{\log n}} n^{-r},$$

and hence

$$\sum_{n=1}^{\infty} \mathbb{P}(V_n) = \infty \qquad \text{if and only if} \quad r \geq 1.$$

While this suggests why the theorem might be true, it does not give a complete proof for two reasons: the events $\{V_n\}$ are not independent and we only considered times t_n of the form ρ^n.

Let A_r, A_r^-, denote the events

$$A_r = \left\{ \limsup_{t \to \infty} \frac{Y_t}{\sqrt{\log \log t}} < \sqrt{2r} \right\}, \quad A_r^- = \left\{ \liminf_{t \to \infty} \frac{Y_t}{\sqrt{\log \log t}} > -\sqrt{2r} \right\},$$

and note that

$$A_r \cup A_r^- = \left\{ \limsup_{t \to \infty} \frac{|Y_t|}{\sqrt{\log \log t}} < \sqrt{2r} \right\}.$$

These are tail events, that is, for each t they are independent of $\mathcal{F}_t := \sigma\{B_s : s \leq t\}$. Hence, using symmetry and Theorem 9.1.1,

$$\mathbb{P}(A_r) = \mathbb{P}(A_r^-) = \mathbb{P}(A_r \cup A_r^-) \in \{0, 1\}.$$

To prove Theorem 9.1, it suffices to show that $\mathbb{P}(A_r) = 0$ for $r < 1$ and $\mathbb{P}(A_r \cup A_r^-) = 1$ for $r > 1$. In both cases, we write $t_n = \rho^n$ with $\rho > 1$ and we will choose a value of ρ depending on r. If $r < 1$ we will choose a very large value while for $r > 1$ we choose a value very close to 1.

Suppose that $r < 1$ and let V_n denote the event

$$V_n = \{Y_{t_n} \geq \sqrt{2r \log \log t_n}\}.$$

We will find a $\rho > 1$ such that with probability one, V_n occurs infinitely often. We will need a slight generalization of the Borel-Cantelli lemma.

- **Claim.** Suppose $U_1, U_2, \ldots$ are events and let $\mathcal{G}_n$ denote the σ-algebra generated by $\{U_1, \ldots, U_n\}$. Suppose that there exist $q_n > 0$ with $\sum q_n = \infty$ such that for all n, $\mathbb{P}(U_n \mid \mathcal{F}_{n-1}) \geq q_k$. Then with probability one, U_n occurs infinitely often.

Indeed, under these assumptions, for each $n > 0$,

$$\mathbb{P}(U_n^c \cap U_{n+1}^c \cap U_{n+2}^c \cap \cdots) \leq \prod_{j=n}^{\infty}(1 - q_j) = 0.$$

Using the fact that

$$\mathbb{P}(|B_{t_n}| \geq x \mid \mathcal{F}_{t_{n-1}}) \geq \begin{cases} \mathbb{P}\{B_{t_n} - B_{t_{n-1}} \geq x\} & \text{if } B_{t_{n-1}} \geq 0 \\ \mathbb{P}\{B_{t_n} - B_{t_{n-1}} \leq -x\} & \text{if } B_{t_{n-1}} < 0 \end{cases},$$

we get

$$\begin{aligned}
\mathbb{P}(V_n \mid V_1, \ldots, V_{n-1}) &\geq \mathbb{P}\left\{ \frac{B_{t_n} - B_{t_{n-1}}}{\sqrt{t_n}} \geq \sqrt{2r \log\log t_n} \right\} \\
&= \mathbb{P}\left\{ \frac{B_{t_n} - B_{t_{n-1}}}{\sqrt{t_n - t_{n-1}}} \geq \sqrt{2r'[\log n + \log \rho]} \right\} \\
&\geq c_\rho [\log n]^{-1/2} \, n^{-r'}.
\end{aligned}$$

where

$$r' = r \frac{t_n}{t_n - t_{n-1}} = \frac{r}{1 - \rho^{-1}}.$$

For fixed $r < 1$, we can find $\rho > 1$ such that $r' < 1$, and hence we have with probability one, V_n occurs infinitely often.

Now suppose $r > 1$ and let U_n denote the event

$$U_n = \{Y_t \geq \sqrt{2r \log\log t} \text{ for some } t_{n-1} \leq t \leq t_n\}.$$

Note that

$$U_n \subset \left\{ \max_{0 \leq t \leq t_n} B_t \geq \sqrt{2rt_{n-1} \log\log t_{n-1}} \right\}.$$

Hence by the reflection principle,

$$\begin{aligned}
\mathbb{P}(U_n) &\leq 2\,\mathbb{P}\left\{ B_{t_n} \geq \sqrt{2rt_{n-1} \log\log t_{n-1}} \right\} \\
&= 2\,\mathbb{P}\{Y_{t_n} \geq \sqrt{2r'[\log(n-1) + \log \rho]}\},
\end{aligned}$$

where $r' = r/\rho$. If $r > 1$, we can find $\rho > 1$ with $r/\rho > 1$, and for this choice we have

$$\sum_{n=1}^{\infty} \mathbb{P}(U_n) < \infty.$$

Hence, by the Borel-Cantelli Lemma, with probability one, U_n occurs only finitely often and hence for all t sufficiently large, $Y_t \leq \sqrt{2r \log \log t}$.

9.3 Zero set and local time

Let B_t be a standard one-dimensional Brownian motion and define the *zero set* to be the collection of times at which the Brownian motion is at origin.

$$Z = \{s : B_s = 0\}, \quad Z_t = Z \cap [0, t] = \{s \leq t : B_s = 0\}.$$

We will call a time $t \in Z$ *left-isolated* or *right-isolated*, respectively, if there exists $\epsilon > 0$ such that $(t - \epsilon, t) \cap Z = \emptyset$ or $(t, t + \epsilon) \cap Z = \emptyset$, respectively. We say that t is *isolated* if it is both left-isolated and right-isolated.

Theorem 9.3.1. *With probability one the following holds.*

1. *Z is a closed set containing no positive rational numbers with Lebesgue measure zero.*

2. *The set of left-isolated points and the set of right-isolated points are both countable.*

3. *0 is not a right-isolated point on Z.*

4. *There are no isolated points.*

5. *Z is an uncountable set and for each t, Z_t is homeomorphic to the usual Cantor set.*

Proof.

Let $\mathbb{Q}_+$ denote the set of positive rationals.

1. Note that Z is the inverse image of the closed set $\{0\}$ under the continuous function B and hence is closed. For given $t > 0$, $\mathbb{P}\{t \in Z\} = \mathbb{P}\{B_t = 0\} = 0$. Since $\mathbb{Q}_+$ is countable, this implies that

$\mathbb{P}\{t \in Z \text{ for some } t \in \mathbb{Q}_+\} = 0$. If ℓ denotes Lebesgue measure, we have

$$\mathbb{E}\left[\ell(Z_t)\right] = \mathbb{E}\left[\int_0^t 1\{B_s = 0\}\, ds\right] = \int_0^t \mathbb{P}\{B_s = 0\}\, ds = 0.$$

Since $\ell(Z_t) \geq 0$, this implies that with probability one $\ell(Z_t) = 0$.

2. For each $q \in \mathbb{Q}_+$, let $\tau_q = \min\{t > q : B_t = 0\}$ With probability one, for all $q \in \mathbb{Q}_+$, $0 < \tau_q < \infty$. Also, if $t \in Z$ is left-isolated, there exists $q \in \mathbb{Q}_+$ such that $t = \tau_q$. Hence we see that the set of left-isolated points is the same as $\{\tau_q : q \in \mathbb{Q}_+\}$ and hence is countable. Similarly we can define $\sigma_q = \max\{t < q : B_t = 0\}$ to see that the set of right-isolated points is the same as $\{\sigma_q : q \in \mathbb{Q}_+\}$. Note that the random times τ_q are stopping times for the Brownian motion but the times σ_q are not.

3. There are various ways to do this. One is to note that the following events are in $\mathcal{F}_{0+}$:

$$A^+ = \{\text{for all } \epsilon > 0 \text{ there exists } 0 < t < \epsilon \text{ with } B_t \geq 0\}.$$

$$A^- = \{\text{for all } \epsilon > 0 \text{ there exists } 0 < t < \epsilon \text{ with } B_t \leq 0\}.$$

By symmetry, $\mathbb{P}(A^+) = \mathbb{P}(A^-)$ and clearly $\mathbb{P}(A^+ \cup A^-) = 1$. By the Blumenthal 0-1 law we see that $\mathbb{P}(A^+) = \mathbb{P}(A^-) = 1$, and 0 is not right-isolated on the event $A^+ \cap A^-$.

4. Since the random times $\tau_q, q \in \mathbb{Q}_+$ are stopping times, the strong Markov property implies that $\mathbb{P}\{\tau_q \text{ is right-isolated}\} = 0$. Therefore,

$$\mathbb{P}\{\tau_q \text{ is right-isolated for some } q \in \mathbb{Q}_+\} = 0.$$

Since every left-isolated point equals τ_q for some $q \in \mathbb{Q}_+$, this shows that no points are isolated.

5. For every $q \in \mathbb{Q}_+$, let I_q be the open interval (σ_q, τ_q). Then, we have $\sigma_q < q < \tau_q$; $B_{\sigma_q} = B_{\tau_q} = 0$; and $B_t \neq 0$ for $t \in I_q$. Hence we can write

$$Z = [0, \infty) \setminus \bigcup_{q \in \mathbb{Q}_+} I_q.$$

The remainder is an exercise in pointwise topology that we omit.

□

For the usual Cantor function, there is a corresponding Cantor function that is a continuous, nondecreasing function that has derivative zero off the Cantor set. There is a corresponding (random) function L_t associated to the zero set called the *local time (at 0)*. Let us state some properties first. Roughly speaking, L_t denotes the (normalized) amount of time that the Brownian motion has spent at zero. Since the Lebesgue of the zero set is zero, we must measure "amount of time" in a way other than the Lebesgue measure.

Similarly to the way we discussed Brownian motion, we will first state the properties of local time and they we will discuss how to construct the object.

Definition The local time associated to a standard Brownian motion at 0 is a continuous, nondecreasing process $L_t, t \geq 0$, adapted to the filtration of the Brownian motion satisfying the following.

1. For each t,
$$\mathbb{E}\left[L_t\right] = \int_0^t \frac{ds}{\sqrt{2\pi s}} = \sqrt{\frac{2t}{\pi}}. \tag{9.2}$$

2. If $a < b$, then $L_b > L_a$ if and only if $Z \cap (a, b) \neq \emptyset$.

3. If $a > 0$, and $\hat{L}_t = a^{-1/2} L_t$, then $\hat{L}_t, t \geq 0$ has the same distribution as $L_t, t \geq 0$. In particular L_t has the same distribution as $\sqrt{t}\, L_1$.

4. For each t, the distribution of L_t given $\mathcal{F}_t$ depends only on (L_t, B_t) and is given as follows. Let T be a random variable giving the amount of time for an independent Brownian motion to reach $|B_t|$ and let $\tilde{L}_s$ be a process with the same distribution as L_t independent of the Brownian motion. Then for $s > t$, the distribution of $L_{s+t} - L_t, s \geq t$ given $\mathcal{F}_t$ is the same as $Y_s, s \geq 0$ where
$$Y_s = \begin{cases} 0, & t \leq s \leq t + T \\ \tilde{L}_{s-T}, & s \geq T \end{cases}.$$

Note that the integrand in (9.2) is the density of B_s evaluated at the origin and can be viewed heuristically as "the probability that $B_s = 0$ ". There is some arbitrariness in the normalization and the definition should be considered up to multiplicative constant. Indeed, some authors choose to use

a different normalization so one should check before using exact formulas. The local time can be constructed using an approximation of the event $\{B_t = 0\}$. Let

$$L_{t,\epsilon} = \frac{1}{2\epsilon} \lim_{\epsilon \downarrow 0} \int_0^t 1\{|B_t| \leq \epsilon\}\, ds.$$

Theorem 9.3.2. *For each $t > 0$, the limit*

$$L_t = \lim_{\epsilon \downarrow 0} L_{t,\epsilon}$$

exists in L^2 and with probability one.

Proof. We first show that for fixed $b > 0$ the limit with $t = b$ exists. Fix $b > 0$ and $0 < \rho < 1$; we allow constants to depend on ρ and b. Let $J_n = L_{b,\rho^n}$. We will show that there exists c such that

$$\mathbb{E}\left[(J_{n+1} - J_n)^2\right] \leq c\, \rho^n. \tag{9.3}$$

Before doing this, let us show how this gives us the result. Using (9.3), we see that $\{J_n\}$ is a Cauchy sequence in L^2 and hence has a limit in L^2 with

$$\mathbb{E}\left[(J - J_n)^2\right] \leq c \sum_{j=n}^{\infty} \rho^j \leq c'\, \rho^n.$$

Markov's inequality states that

$$\mathbb{P}\{|J - J_n| \leq \rho^{n/4}\} \leq \frac{\mathbb{E}\left[(J - J_n)^2)\right]}{\rho^{n/2}} \leq c'\, \rho^{n/2}.$$

Hence using the Borel-Cantelli lemma, we see that with probability one $J_n \to J$. If $\rho^{n+1} \leq \epsilon \leq \rho^n$, we have

$$\rho\, J_{n+1} \leq L_{t,\epsilon} \leq \rho J_n,$$

and hence with probability one,

$$\rho\, J \leq \liminf_{\epsilon \downarrow 0} L_{t,\epsilon} \leq \limsup_{\epsilon \downarrow 0} L_{t,\epsilon} \leq \frac{1}{\rho}\, J.$$

By choosing a sequence $\rho_m \uparrow 1$, we get the result.

To prove (9.3), let $\Delta_n(t) = I_{n+1}(t) - I_n(t)$, and note that for $t \geq \rho^{2n}$.

$$\mathbb{E}\left[\frac{1}{2\rho^n} 1\{|B_t| \leq \rho^n\}\right] = \frac{1}{2\rho^n}\mathbb{P}\{|B_t| \leq \rho^n\} = \frac{1}{\sqrt{2\pi t}}\left[1 + O\left(\frac{\rho^{2n}}{t}\right)\right],$$

Therefore, for all $t \leq b$,

$$|\mathbb{E}[\Delta_n(t)]| \leq c \min\left\{\frac{\rho^{2n}}{t^{3/2}}, \frac{1}{t^{1/2}}\right\}.$$

Smilarly , if $s < t$, on the event that $\{|B_s| \leq \rho^n\}$ we have

$$|E[\Delta_n(t) \mid \mathcal{F}_s]| \leq c \min\left\{\frac{\rho^{2n}}{(t-s)^{3/2}}, \frac{1}{(t-s)^{1/2}}\right\}.$$

Therefore,

$$|\mathbb{E}[\Delta_n(s)\,\Delta_n(t)]| = |\mathbb{E}[\Delta_n(s)\,E(\Delta_n(t) \mid \mathcal{F}_s)]|$$
$$\leq c \min\left\{\frac{\rho^{2n}}{(t-s)^{3/2}}, \frac{1}{(t-s)^{1/2}}\right\}.$$

For $|t - s| \leq \rho^{2n}$, we will use

$$|\mathbb{E}[\Delta_n(s)\,\Delta_n(t)]| \leq \rho^{-n}\mathbb{E}\left[|\Delta_n(s)|\right] \leq \rho^{-n}\,\mathbb{E}[I_n(s) + I_{n+1}(s)] \leq c\,\rho^{-n}$$

$$\mathbb{E}[(J_{n+1} - J_n)^2] = \mathbb{E}\left[\left(\int_0^b \Delta_n(t)\,dt\right)^2\right]$$
$$= \mathbb{E}\left[\int_0^b \int_0^b \Delta_n(t)\,\Delta_n(s)\,ds\,dt\right]$$
$$= 2\int_0^b \int_s^b |\mathbb{E}[\Delta_n(s)\,\Delta_n(t)]|\,dt\,ds$$
$$\leq c\int_0^b \int_s^b \min\left\{\frac{\rho^{2n}}{(t-s)^{3/2}}, \frac{1}{(t-s)^{1/2}}\right\}\,dt\,ds$$
$$\leq c\,\rho^n.$$

$\square$

Since the limit also exists in L^2, we have

$$\mathbb{E}[L_t] = \lim_{\epsilon \downarrow 0} \mathbb{E}[L_{t,\epsilon}] = \int_0^t \frac{ds}{\sqrt{2\pi s}} = \sqrt{\frac{2t}{\pi}},$$

$$\mathbb{E}[L_t^2] = \lim_{\epsilon \downarrow 0} \mathbb{E}[L_{t,\epsilon}^2] = 2 \int_0^t \int_r^t \frac{ds}{\sqrt{2\pi r}\,\sqrt{2\pi\,(s-r)}}\, dr = t.$$

Similarly, using scaling and Fatou's lemma, we can show that for all integers k

$$\mathbb{E}[L_t^k] = t^{k/2}\, \mathbb{E}[L_1^k] \leq \lim_{\epsilon \downarrow 0} \mathbb{E}[L_{1,\epsilon}^k] = t^{k/2}\, C_k$$

for a C_k that can be computed exactly. The inequality is actually an equality, but we do not need this fact. If $s < t$, then the distribution of $L_t - L_s$ is stochastically dominated by that of $L_t - L_s$, and hence

$$\mathbb{E}[(L_t - L_s)^k)] \leq C_k\,(t-k)^{k/2}.$$

Using this and the Kolomogorov criterion (see Exercise 2.13), we can see that with probability one $t \mapsto L_t$ is a continuous function.