Chapter 3
Continuous processes

Stock prices are not trees. The discrete trees of the previous chapter are only an approximation to the way that prices actually move. In practice, a price can change at any instant, rather than just at some fixed tick-times when a portfolio can be calmly rebalanced. The binary choice of a single jump 'up' or 'down' only becomes subtle as the ticks get closer and closer, giving the tree more and ever-shorter branches. But such trees grow too complex and we stop being able to see the wood.

We shall have to start from scratch in the continuous world. The discrete models will guide us – the intuitions gained there will be more than useful – but limiting arguments based on letting δt tend to zero are too dangerous to be used rigorously. We will encounter a representation theorem which establishes the basis of risk-free construction and again it will be martingale measures that prime the expectation operator correctly. But processes and measures will be harder to separate intuitively – we will need a calculus to help us. And changes in measure will affect processes in surprising ways. We will no longer be able to proceed in full generality – we will concentrate on Brownian motion and its relatives. If there is one overarching principle to this chapter it is that Brownian motion is sophisticated enough to produce interesting models and simple enough to be tractable. Given the subtleties of working with continuous processes, a simple calculus based on Brownian motion will be more than enough for us.

3.1 Continuous processes

We want randomness. With our discrete stock price model we didn't have any old random process. We forcibly limited ourselves to a binomial tree. We started simply and hoped (with some justification) that complex enough market models could be built from such humble materials. The single binomial branching was the building block for our 'realistic' market. For the continuous world we need an analogous basis – something simple and yet a reasonable starting point for realism.

What is a continuous process? Three small-scale principles guide us. Firstly, the value can change at any time and from moment to moment. Secondly, the actual values taken can be expressed in arbitrarily fine fractions – any real number can be taken as a value. And lastly the process changes continuously – the value cannot make instantaneous jumps. In other words, if the value changes from 1 to 1.05 it must have passed through, albeit quickly, all the values in between.

At least as a starting point, we can insist that stock market indices or prices of individual securities behave this way. Even though they move in a 'sharp-edged' way, it isn't too unrealistic to claim that they nonetheless display continuous process behaviour.

And as far back as Bachelier in 1900, who analysed the motion of the Paris stock exchange, people have gone further and compared the prices to one particular continuous process – the process followed by a randomly moving gas particle, or *Brownian motion* (figure 3.2).

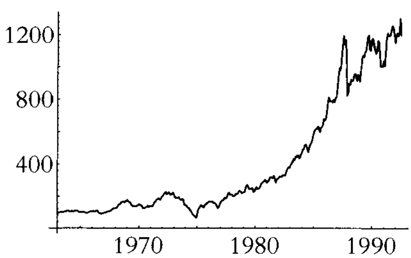


Figure 3.1 UK FTA index, 1963–92

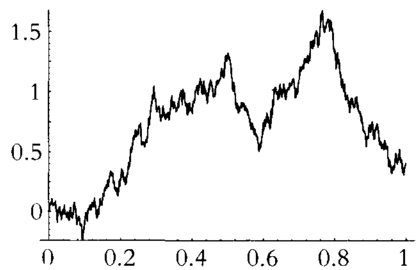


Figure 3.2 Brownian motion

Locally the likeness can be striking – both display the same jaggedness, and the same similarity under scale changes – the jaggedness never smooths out as the magnification increases. But globally, the similarity fades – figure 3.1

doesn't *look like* figure 3.2. At an intuitive level, the global structure of the stock index is different. It grows, gets 'noisier' as time passes, and doesn't go negative. Brownian motion can't be the whole story.

But we only want a basis – the single binomial branching didn't look promising right away. We shouldn't run ahead of ourselves. Brownian motion will prove a remarkably effective component to build continuous processes with – *locally* Brownian motion looks realistic.

Brownian motion

It was nearly a century after botanist Robert Brown first observed microscopic particles zigzagging under the continuous buffeting of a gas that the mathematical model for their movements was properly developed. The first step to the analysis of Brownian motion is to construct a special family of discrete binomial processes.

The random walk $W_n(t)$

For n a positive integer, define the binomial process $W_n(t)$ to have:

(i) $W_n(0) = 0$,

(ii) layer spacing $1/n$,

(iii) up and down jumps equal and of size $1/\sqrt{n}$,

(iv) measure $\mathbb{P}$, given by up and down probabilities everywhere equal to $\frac{1}{2}$.

In other words, if $X_1, X_2, \ldots$ is a sequence of independent binomial random variables taking values $+1$ or -1 with equal probability, then the value of W_n at the ith step is defined by:

$$W_n\left(\tfrac{i}{n}\right) = W_n\left(\tfrac{i-1}{n}\right) + \frac{X_i}{\sqrt{n}}, \qquad \text{for all } i \geqslant 1.$$

The first two steps are shown in figure 3.3. What does W_n look like as n gets large?

Instead of blowing out of control, the family portraits (figure 3.4) appear to be settling down towards something as n increases. The moves of size

$1/\sqrt{n}$ seem to force some kind of convergence. Can we make a formal statement? Consider for example, the distribution of W_n at time 1: for a particular W_n, there are $n + 1$ possible values that it can take, ranging from $-\sqrt{n}$ to $\sqrt{n}$. But the distribution always has zero mean and unit variance. (Because $W_n(1)$ is the sum of n IID random variables, each with zero mean and variance $1/n$.)

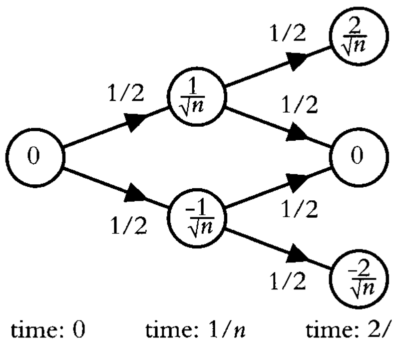


Figure 3.3 The first two steps of the random walk W_n

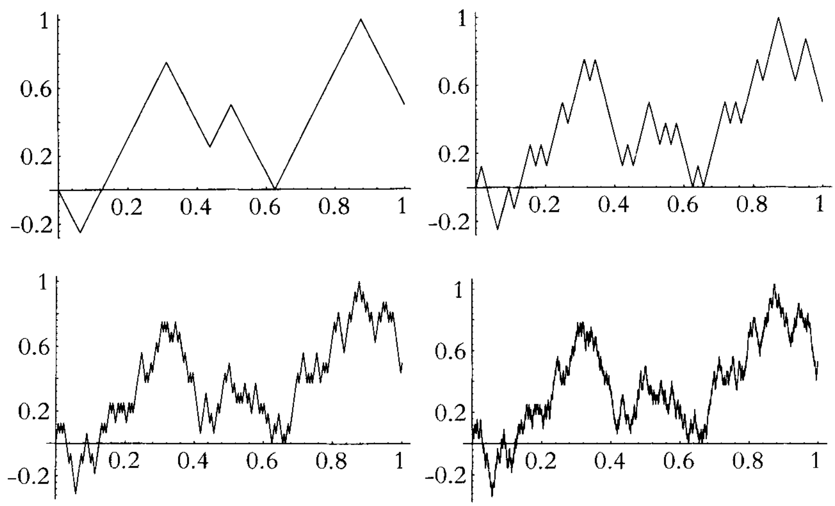


Figure 3.4 Random walks of 16, 64, 256 and 1024 steps respectively

Moreover the central limit theorem gives us a limit for these binomial distributions – as n gets large, the distribution of $W_n(1)$ tends towards the unit normal $N(0,1)$. In fact, the value of $W_n(t)$ is the same as

$$W_n(t) = \sqrt{t}\left(\frac{\sum_{i=1}^{nt} X_i}{\sqrt{nt}}\right).$$

The distribution of the ratio in brackets tends, by the central limit theorem, to a normal $N(0,1)$ random variable. And so the distribution of $W_n(t)$ tends to a normal $N(0,t)$.

There is a formal unity underlying the family – all the marginal distributions tend towards the same underlying normal structure.

And not just all the marginal distributions, but all the *conditional* marginal distributions as well. Each random walk W_n has the property that its future movements away from a particular position are independent of where that position is (and indeed independent of its entire history of movements up to that time). Additionally such a future displacement $W_n(s+t) - W_n(s)$ is binomially distributed with zero mean and variance t. Thus again, the central limit theorem gives us a constant limiting structure, and all conditional marginals tend towards a normal distribution of the same mean and variance.

The marginals converge, the conditional marginals converge, and the temptation is irresistible to say that the distributions of the processes converge too. And indeed they do, though this isn't the place to set up the careful formal framework to make sense of that statement. The distribution of W_n converges, and it converges towards *Brownian motion.*

Formally:

Brownian motion

The process $W = (W_t : t \geqslant 0)$ is a $\mathbb{P}$-Brownian motion if and only if

(i) W_t is continuous, and $W_0 = 0$,

(ii) the value of W_t is distributed, under $\mathbb{P}$, as a normal random variable $N(0,t)$,

(iii) the increment $W_{s+t} - W_s$ is distributed as a normal $N(0,t)$, under $\mathbb{P}$, and is independent of $\mathcal{F}_s$, the history of what the process did up to time s.

These are both necessary and sufficient conditions for the process W to be Brownian motion. The last condition, though an exact echo of the behaviour of the discrete precursors $W_n(t)$, is subtle. Many processes that have marginals $N(0,t)$ are not Brownian motion. In the continuous world, just as it was in the discrete, it is not just the marginals (conditional on the process' value at time zero) that count, but *all* the marginals conditional on *all* the histories $\mathcal{F}_s$. It will in fact be the daunting task of specifying all these that drives us to a Brownian calculus.

Exercise 3.1 If Z is a normal $N(0,1)$, then the process $X_t = \sqrt{t}Z$ is continuous and is marginally distributed as a normal $N(0,t)$. Is X a Brownian motion?

Exercise 3.2 If W_t and $\tilde{W}_t$ are two independent Brownian motions and ρ is a constant between -1 and 1, then the process $X_t = \rho W_t + \sqrt{1-\rho^2}\tilde{W}_t$ is continuous and has marginal distributions $N(0,t)$. Is this X a Brownian motion?

It is also worth noting just how *odd* Brownian motion really is. We won't stop to prove them, but here is a brief peek into the bestiary:

- Although W is continuous everywhere, it is (with probability one) differentiable nowhere.

- Brownian motion will eventually hit any and every real value no matter how large, or how negative. It may be a million units above the axis, but it will (with probability one) be back down again to zero, by some later time.

- Once Brownian motion hits a value, it immediately hits it again *infinitely* often, and then again from time to time in the future.

- It doesn't matter what scale you examine Brownian motion on – it looks just the same. Brownian motion is a fractal.

Brownian motion is often also called a Wiener process, and is a (one-dimensional) Gaussian process.

Brownian motion as stock model

We had our misgivings about Brownian motion as a global model for stock behaviour, but we don't have to use it on its own. Brownian motion wanders. It has mean zero, whereas the stock of a company normally grows at some rate – and historically we expect prices to rise if only because of inflation. But we can add in a drift artificially. For example the process $S_t = W_t + \mu t$, for some constant μ reflecting nominal growth, is called Brownian motion with drift.

And if it looks too noisy, or not noisy enough, we can scale the Brownian motion by some factor: for example, $S_t = \sigma W_t + \mu t$, for a constant noise factor σ.

How are we doing? Consider the stock market data shown in figure 3.1. We could estimate σ and μ for the best fit [in this case, $\sigma = 91.3$ and $\mu = 37.8$] and simulate a sample path.

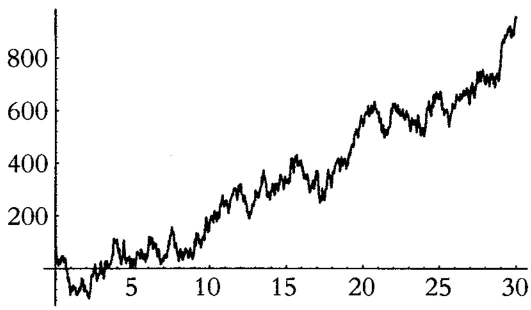


Figure 3.5 Brownian motion plus drift

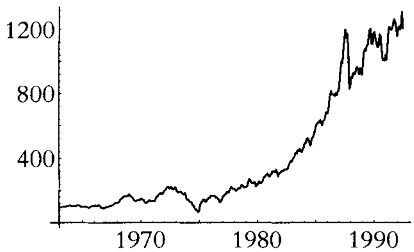
Not bad – the process has long-term upwards growth, as we want. But in this particular case, we have a glitch right away. The process went negative, which we may not want for the price of a stock of a limited liability company.

Exercise 3.3 Show that, for all values of σ ($\sigma \neq 0$), μ, and $T > 0$ there is always a positive probability that S_T is negative. (Hint: consider the marginal distribution of S_T.)

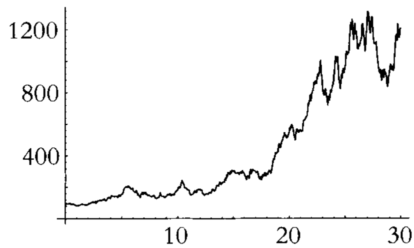
We can though be more adventurous in shaping Brownian motion to our ends. Consider for example, taking the exponential of our process:

$$X_t = \exp(\sigma W_t + \mu t).$$

Now we mirror the stock market's long-term exponential growth (and for good measure we start off quietly and get noisier). Again finding a best fit for σ and μ [$\sigma = 0.178$ and $\mu = 0.087$, a 'noisiness' of 17.8% and an annual drift of 8.7%] we can simulate a sample path (figure 3.6).



UK FTA index, 1963–92
Figure 3.1



Exponential Brownian motion
Figure 3.6

This process is, not surprisingly, well known and it is usually called exponential Brownian motion with drift, or sometimes geometric Brownian motion with drift. It is not the only model for stocks – and indeed we will look at others later on – but it is simple and not that bad. (Could you tell which picture was which without the captions?) Brownian motion can prove an effective building block.

3.2 Stochastic calculus

Shaping Brownian motion with functions may be powerful, but it brings a dangerous complexity. Consider any smooth (differentiable) curve. Globally it can have almost any behaviour it likes, because the condition that it is differentiable does nothing to affect it at a large scale. Suppose we zoom in though, pinning down a small section under a microscope. In figure 3.7, we focus in on the point of a particular differentiable curve with x-co-ordinate of 1.7, increasing the magnification by a factor of about ten each time.

Reading the graphs from left to right and line by line, each small box is expanded to form the frame for the next graph. As the process continues, the graph section becomes smoother and straighter, until eventually it *is* straight – it is a small straight line.

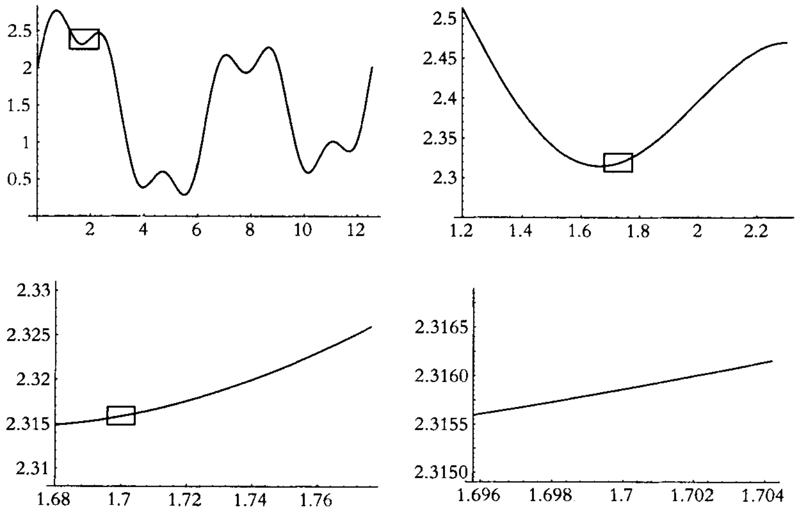


Figure 3.7 Progressive magnification around the point 1.7

Differentiable functions, however strange their global behaviour, are at heart built from straight line segments. Newtonian calculus is the formal acknowledgement of this.

With a Newtonian construction, we could decide to build up a family of nice functions by specifying how they are locally built up out of our building block, the straight line. We would write the change in value of a *Newtonian function* f over a time interval at t of infinitesimal length dt as

$$df_t = \mu_t \, dt,$$

where μ_t is our scaling function, the slope or drift of the magnified straight line at t.

Then we could explore our universe of Newtonian functions. Consider, for example:

(1) The equation $df_t = \mu \, dt$, for some constant μ. What is f? That is, what does it look like? How does it behave globally? Could we draw it? If

we stick together straight line segments of slope μ, then intuitively we just produce a straight line of slope μ. If f_0, for example, was equal to zero then we might guess (correctly) that f_t could be written in more familiar notation as $f_t = \mu t$.

(2) The equation $df_t = t\,dt$. Here we have a slope at time t of value t – what does this look like? Simple integration comes to the rescue. If $f_0 = 0$, then we could again pin down f_t as $f_t = \frac{1}{2}t^2$. The going was a bit harder here, but we managed it, and we can check it ourselves by differentiation: $f_t' = t$ as we require.

What about uniqueness though? In the first example, our intuition dismissed the possibility of another solution, but what about here? The construction metaphor ($df_t = t\,dt$ tells us how to build f_t, and thus given a starting place and a deterministic building plan we ought to produce just one possible f_t) suggests that $f_t = \frac{1}{2}t^2$ is the unique solution and indeed we can formalise this.

Uniqueness of Newtonian differentials

Two complementary forms of uniqueness operate here.

- If f_t and $\tilde{f}_t$ are two differentiable functions agreeing at 0 ($f_0 = \tilde{f}_0$) and they have identical drifts ($df_t = d\tilde{f}_t$), then the processes are equal: $f_t = \tilde{f}_t$ for all t. In other words, f is unique given the drift μ_t (and f_0).
- Secondly, given a differentiable function f_t, there is only one drift function μ_t which satisfies $f_t = f_0 + \int_0^t \mu_s\,ds$ (for all t). So μ is unique given f.

Instead of just giving the drift μ_t directly, we might have a problem where the drift itself depends on the current value of the function. Specifically, if the drift μ_t equals $\mu(f_t, t)$, where $\mu(x, t)$ is a known function, then

$$df_t = \mu(f_t, t)\,dt$$

is called an ordinary differential equation (ODE). If there is a differentiable function f which satisfies it (with given f_0), it forms a *solution*. There are plenty of ODEs which have no solutions, and plenty more which do not have unique solutions. (The uniqueness of the solution to an ODE cannot be deduced just from the uniqueness of Newtonian differentials in the box.)

(3) The equation $df_t = f_t\,dt$. Now things are harder, as direct integration is not a route to the solution. We could guess – say $f_t = e^t$ – and then check by differentiation. This solution happens to be unique for $f_0 = 1$.

(4) The equation $df_t = f_t t^{-2}\,dt$. This is an example of a bad case, where solutions need neither exist nor be unique. Given $f_0 = 0$, there are an infinite number of solutions, namely $f_t = a\exp(-1/t)$ for every possible value of an arbitrary constant a. However, for $f_0 \neq 0$, there are no solutions at all.

Perhaps our universe of Newtonian functions isn't so benign. It is clear that though ODEs are powerful construction tools, they are also dangerous ones. There are plenty of 'bad' ODEs which we haven't a clue how to explore.

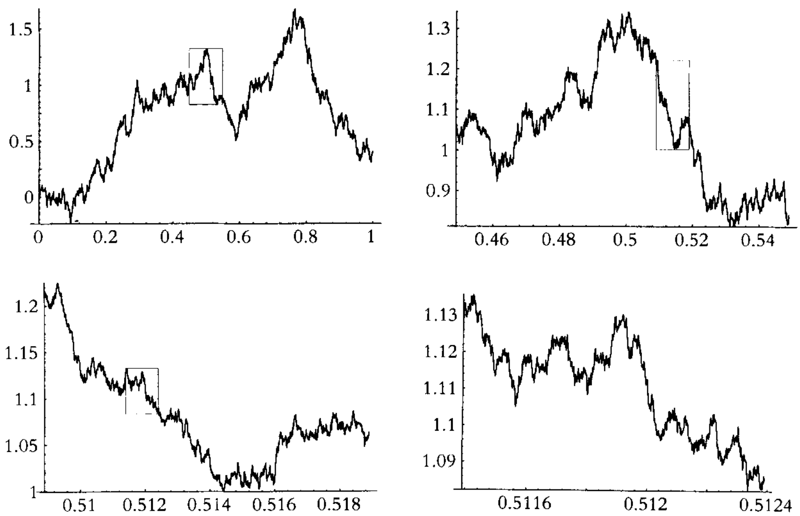


Figure 3.8 'Zooming in' on Brownian motion

Stochastic differentials

And if it was bad for Newtonian differentials, consider a construction procedure based on Brownian motion. Zooming in on Brownian motion doesn't produce a straight line (figure 3.8)

As before, each box is expanded by suitable horizontal and vertical scaling to frame the next graph. The self-similarity of Brownian motion means that each new graph is also a Brownian motion, and just as noisy.

But of course this self-similarity is ideal for a building block – we could build global Brownian motion out of lots of local Brownian motion segments. And we could build general random processes from small segments of Brownian motion (suitably scaled). If we built using straight line segments (suitably scaled) too, we could include Newtonian functions as well.

A stochastic process X will have both a Newtonian term based on dt and a Brownian term, based on the infinitesimal increment of W which we will call dW_t. The Brownian term of X can have a noise factor σ_t, and so the infinitesimal change of X_t is

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

As in the Newtonian case, the drift μ_t can depend on the time t. But it can also be random and depend on values that X (or indeed W) took up until t itself. And of course, so can the noisiness σ_t. Such processes, like X and σ, whose value at time t can depend on the history $\mathcal{F}_t$, but not the future, are called *adapted* to the filtration $\mathcal{F}$ of the Brownian motion W.

We call σ_t the *volatility* of the process X at time t and μ_t the *drift* of X at t.

Stochastic processes

What does our universe look like? As with Newtonian differentials, finding this out entails 'integrating' stochastic differentials in some way. We can, though, formally define what it is to be a (continuous) stochastic process.

This definition of stochastic process (see box) is not universal, and in particular it excludes discontinuous cases such as Poisson processes. Nevertheless it will be quite adequate for all the models we will meet.

The technical condition that σ and μ must be $\mathcal{F}$-previsible processes means that they are adapted to the filtration $\mathcal{F}$, and that they may have some jump discontinuities. In terms of stochastic analysis, this defines stochastic processes to be semimartingales whose drift term is absolutely continuous. This class is closed under all the operations used later, and all the models considered will lie within it.

And as it happens, we can provide a uniqueness result to mirror the classical setup.

Stochastic processes

A stochastic process X is a continuous process $(X_t : t \geqslant 0)$ such that X_t can be written as

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

where σ and μ are random $\mathcal{F}$-previsible processes such that $\int_0^t (\sigma_s^2 + |\mu_s|) \, ds$ is finite for all times t (with probability 1). The differential form of this equation can be written

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

Uniqueness of volatility and drift

Two complementary forms of uniqueness operate here.

- Firstly, if two processes X_t and $\tilde{X}_t$ agree at time zero ($X_0 = \tilde{X}_0$) and they have identical volatility σ_t and drift μ_t, then the processes are equal: $X_t = \tilde{X}_t$ for all t. In other words, X is unique given σ_t and μ_t (and X_0).
- Secondly, given the process X, there is only one pair of volatility σ_t and drift μ_t which satisfies $X_t = X_0 + \int_0^t \sigma_s \, dW_s + \int_0^t \mu_s \, ds$ (for all t). This uniqueness of σ_t and μ_t given X comes from the *Doob–Meyer decomposition* of semimartingales.

In the special case when σ and μ depend on W only through X_t, such as $\sigma_t = \sigma(X_t, t)$, where $\sigma(x, t)$ is some deterministic function, the equation

$$dX_t = \sigma(X_t, t) \, dW_t + \mu(X_t, t) \, dt$$

is called a stochastic differential equation (SDE) for X. And it will generally be easier to write down the SDE (if it exists) for a particular X then it is to provide an explicit solution for the SDE. As in the Newtonian case (ODEs), an SDE need not have a solution, and if it does it might not be unique. Usage of the term SDE does tend to spread out from this strict definition to include the stochastic differentials of processes whose volatility and drift depends not only on X_t and t, but also on other events in the history $\mathcal{F}_t$.

But can we recognise the world we have created, perhaps in terms of W_t, the Brownian motion we have some handle on?

Partially. In the simple case, where σ and μ are both constants, meaning that X has constant volatility and drift, the SDE for X is

$$dX_t = \sigma\, dW_t + \mu\, dt.$$

It isn't too hard to guess what the solution to this is:

$$X_t = \sigma W_t + \mu t,$$

(assuming that $X_0 = 0$). And our meagre understanding of W_t and dW_t at least gives us some confidence that the differential form of σW_t is $\sigma\, dW_t$. As σ and μ are independent of X, the uniqueness result could form a part of a proof that this is the only solution.

But consider the only slightly more complex SDE (echoing the Newtonian ODE of example (3) above),

$$dX_t = X_t\big(\sigma\, dW_t + \mu\, dt\big).$$

We're at sea.

3.3 Itô calculus

Intuitive integration doesn't carry us very far. We need tools to manipulate the differential equations, just as Newtonian calculus has the chain rule, product rule, integration by parts, and so on.

How far could Newton carry us? Suppose we had some function f of Brownian motion, say $f(W_t) = W_t^2$. Could we use a simple chain rule to produce the stochastic differential df_t? Under Newtonian rules, $d(W_t^2)$ would be $2W_t\, dW_t$, which doesn't look too implausible. But we should check via integration, because

$$\text{if} \quad \int_0^t d(W_s^2) = 2\int_0^t W_s\, dW_s, \quad \text{then} \quad W_t^2 = 2\int_0^t W_s\, dW_s.$$

How can we tackle $\int_0^t W_s\,dW_s$? Consider dividing up the time interval $[0,t]$ into a partition $\{0, t/n, 2t/n, \ldots, (n-1)t/n, t\}$ for some n. Then we could approximate the integral with a summation over this partition, that is

$$2\int_0^t W_s\,dW_s \approx 2\sum_{i=0}^{n-1} W\left(\tfrac{it}{n}\right)\left(W\left(\tfrac{(i+1)t}{n}\right) - W\left(\tfrac{it}{n}\right)\right).$$

Now something begins to worry us. The difference term inside the brackets is just the increment of Brownian motion from one particular partition point to the next. By property (iii) of Brownian motion, that increment is independent of the Brownian motion up to that point, and in particular it is independent of the Brownian motion term $W(it/n)$. Also the increment has zero mean, which means that so too must the product of the increment and $W(it/n)$. So the summation consists of terms with zero mean, forcing it to have zero mean itself.

But W_t^2 has mean t, because of the variance structure of Brownian motion, so $2W_t\,dW_t$ *cannot* be the differential of W_t^2, because its integral doesn't even have the right expectation.

What went wrong? Consider a Taylor expansion of $f(W_t)$ for some smooth f:

$$df(W_t) = f'(W_t)\,dW_t + \tfrac{1}{2}f''(W_t)\,(dW_t)^2 + \tfrac{1}{3!}f'''(W_t)\,(dW_t)^3 + \ldots$$

Over-familiar with Newtonian differentials, we assumed that $(dW_t)^2$ and higher terms were zero. But as we have observed before, Brownian motion is odd. Take $(dW_t)^2$, given the same partitioning of $[0,t]$ we just used: $\{0, t/n, 2t/n, \ldots, t\}$. We can model the integral of $(dW_t)^2$ by the (hopefully convergent) approximation

$$\int_0^t (dW_t)^2 = \sum_{i=1}^{n}\left(W\left(\tfrac{ti}{n}\right) - W\left(\tfrac{t(i-1)}{n}\right)\right)^2.$$

But if we let $Z_{n,i}$ be

$$Z_{n,i} = \frac{W\left(\frac{ti}{n}\right) - W\left(\frac{t(i-1)}{n}\right)}{\sqrt{t/n}},$$

then for each n, the sequence $Z_{n,1}, Z_{n,2}, \ldots$ is a set of IID normals $N(0,1)$. (Because each increment $W\left(\frac{ti}{n}\right) - W\left(\frac{t(i-1)}{n}\right)$ is a normal $N(0,t/n)$, independent of the ones before it, by Brownian motion fact (iii).)

We can rewrite our approximation for $\int (dW_s)^2$ as

$$\int_0^t (dW_s)^2 \approx t \sum_{i=1}^{n} \frac{Z_{n,i}^2}{n}.$$

By the weak law of large numbers (just like the strong law but only talking about the distribution of random variables), the distribution of the right-hand side summation converges towards the constant expectation of each $Z_{n,i}^2$, namely 1. Thus $\int_0^t (dW_s)^2 = t$, or in differential form $(dW_t)^2 = dt$.

We can't ignore $(dW_t)^2$; it only looks second order because of the notation. What about $(dW_t)^3$ and so on? It turns out that they *are* zero. (For example, $\mathbb{E}(|dW_t|^3)$ has size $(dt)^{3/2}$, which is negligible compared with dt.) So Taylor gives us:

$$df(W_t) = f'(W_t)\, dW_t + \tfrac{1}{2} f''(W_t)\, dt + 0.$$

The formal version of this surprising departure from Newtonian differentials is the deservedly famous *Itô's formula* (sometimes seen modestly as Itô's lemma).

Itô's formula

If X is a stochastic process, satisfying $dX_t = \sigma_t\, dW_t + \mu_t\, dt$, and f is a deterministic twice continuously differentiable function, then $Y_t := f(X_t)$ is also a stochastic process and is given by

$$dY_t = \Big(\sigma_t f'(X_t)\Big)\, dW_t + \Big(\mu_t f'(X_t) + \tfrac{1}{2}\sigma_t^2 f''(X_t)\Big)\, dt.$$

Returning to our W_t^2, we can apply Itô with $X = W$ and $f(x) = x^2$ and we have

$$d(W_t^2) = 2W_t\, dW_t + dt, \quad \text{or} \quad W_t^2 = 2\int_0^t W_s\, dW_s + t,$$

which at least has the right expectation.

More generally, if X is still just the Brownian motion W, then $f(X)$ has differential

$$df(W_t) = f'(W_t)\, dW_t + \tfrac{1}{2} f''(W_t)\, dt,$$

as hinted above.

Exercise 3.4 If $X_t = \exp(W_t)$, then what is dX_t?

SDEs from processes

Itô's most immediate use is to generate SDEs from a functional expression for a process. Consider the exponential Brownian motion we set up in section 3.1:

$$X_t = \exp(\sigma W_t + \mu t).$$

What SDE does X follow? We know we can handle the term inside the brackets but we have to take a stochastic differential of the exponential function as well. With the right formulation though, we can use Itô's formula.

Suppose we took Y_t to be the process $\sigma W_t + \mu t$, and f to be the exponential function $f(x) = e^x$. Then Y_t *is* simple enough that we can write down its differential immediately: $dY_t = \sigma\, dW_t + \mu\, dt$. But of course the X_t we want can be written as $X_t = f(Y_t)$, so one application of Itô's formula gives us

$$dX_t = \sigma f'(Y_t)\, dW_t + \left(\mu f'(Y_t) + \tfrac{1}{2}\sigma^2 f''(Y_t)\right) dt.$$

The exponential function is particularly pleasant, as $f'(Y_t) = f''(Y_t) = f(Y_t) = X_t$, so we can rewrite the differential as

$$dX_t = X_t\left(\sigma\, dW_t + (\mu + \tfrac{1}{2}\sigma^2)\, dt\right).$$

Here, the variable σ is sometimes called the *log-volatility* of the process, because it is the volatility of the process $\log X_t$, and which is often abbreviated just to volatility notwithstanding that term's existing definition. We will also use the name *log-drift* for the drift μ of $\log X_t$, which is different from the drift of dX_t/X_t above.

Processes from SDEs

Much like differentiation (easy, but its inverse can be impossible), using Itô to convert processes to SDEs is relatively straightforward. And if that were all we ever wanted to do there would be few problems. But it isn't – one of the key needs we have is to go in the opposite direction and convert SDEs to processes. Or in other words, to solve them.

In general we can't. Most stochastic differential equations are just too difficult to solve. But a few, rare examples can be, and just like some ODEs they depend on an inspired guess and then a proof that the proposed solution is an actual solution via Itô. Such a solution to an SDE is called a *diffusion*.

Suppose we are asked to solve the SDE

$$dX_t = \sigma X_t \, dW_t.$$

We need an inspired guess – so we notice that the stochastic term $(\sigma X_t \, dW_t)$ from this SDE is the same as the SDE we generated via Itô in the section above. Moreover, *if* we choose μ to be $-\frac{1}{2}\sigma^2$, then the drift term in the SDE would match our SDE as well. We guess then that

$$X_t = \exp(\sigma W_t - \tfrac{1}{2}\sigma^2 t).$$

What does Itô tell us? That dX_t is indeed $\sigma X_t \, dW_t$, which is what we wanted. So we have found *one* solution, and as it turns out, the only solution (up to constant multiples). Soluble SDEs are scarce, and this one is special enough to have a name: the Doléans exponential of Brownian motion.

Let's go back then to the SDE we tripped over earlier:

$$dX_t = X_t\big(\sigma \, dW_t + \mu \, dt\big).$$

We could match both drift and volatility terms for this SDE and the SDE of $\exp(\sigma W_t + \nu t)$ if and only if we take ν to be $\mu - \frac{1}{2}\sigma^2$. So that is our guess, that

$$X_t = \exp\big(\sigma W_t + (\mu - \tfrac{1}{2}\sigma^2)t\big).$$

And again Itô confirms our intuition.

Exercise 3.5 What is the solution of $dX_t = X_t(\sigma \, dW_t + \mu_t \, dt)$, for μ_t a general bounded integrable function of time?

The product rule

Another Newtonian law was the product rule, that $d(f_t g_t) = f_t\,dg_t + g_t\,df_t$. In the stochastic world, there are two (seemingly) separate cases.

In the more significant case, X_t and Y_t are adapted to the same Brownian motion W, in that

$$dX_t = \sigma_t\,dW_t + \mu_t\,dt,$$
$$dY_t = \rho_t\,dW_t + \nu_t\,dt.$$

By applying Itô's formula to $\frac{1}{2}\left((X_t + Y_t)^2 - X_t^2 - Y_t^2\right) = X_t Y_t$, we can see that

$$d(X_t Y_t) = X_t\,dY_t + Y_t\,dX_t + \sigma_t \rho_t\,dt.$$

The final term above is actually $dX_t\,dY_t$ (following from $(dW_t)^2 = dt$) and again marks the difference between Newtonian and Itô calculus.

In the other case, X_t and Y_t are two stochastic processes adapted to two different and independent Brownian motions, such as

$$dX_t = \sigma_t\,dW_t + \mu_t\,dt,$$
$$dY_t = \rho_t\,d\tilde{W}_t + \nu_t\,dt,$$

where σ_t and ρ_t are the respective volatilities of X and Y, μ_t and ν_t are their drifts, and W and $\tilde{W}$ are two independent Brownian motions. Here

$$d(X_t Y_t) = X_t\,dY_t + Y_t\,dX_t,$$

just as in the Newtonian case.

At a deeper level these two stochastic cases can be reconciled by viewing X_t and Y_t as both adapted to the two-dimensional Brownian motion $(W_t, \tilde{W}_t)$, as will be explained in section 6.3.

Exercise 3.6 Show that if B_t is a zero-volatility process and X_t is any stochastic process, then

$$d(B_t X_t) = B_t\,dX_t + X_t\,dB_t.$$

3.4 Change of measure – the C-M-G theorem

Something remains hidden from us. One of the central themes of the previous chapter was the importance of separating process and measure. Yet we don't seem to mention measures in our stochastic differentials. We may have our basic tools for manipulating stochastic processes, but they are a manipulation of differentials of Brownian motion, not a manipulation of measure. We haven't actually ignored the importance of measure – W_t is not strictly a Brownian motion *per se*, but a Brownian motion with respect to some measure $\mathbb{P}$, a $\mathbb{P}$-Brownian motion. And thus our stochastic differential formulation describes the behaviour of the process X with respect to the measure $\mathbb{P}$ that makes the W_t (or of course the dW_t) a Brownian motion. But the only tool we have seen so far gives us no clue how W_t let alone X_t changes as the measure changes.

As it happens, Brownian motions change in easy and pleasant ways under changes in measure. And thus by extension through their differentials, so do stochastic processes.

Change of measure – the Radon–Nikodym derivative

To get some intuitive feel for the effects of a change of measure, we should go back for a while to discrete processes. Consider a simple two-step random walk:

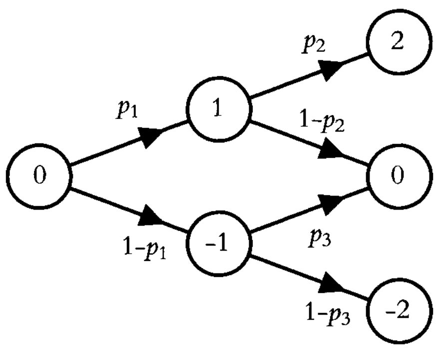


Figure 3.9 Two-step recombinant tree

To get from time 0 to time 2, we can follow four possible paths $\{0, 1, 2\}$,

$\{0, 1, 0\}$, $\{0, -1, 0\}$ and $\{0, -1, -2\}$. Suppose we specified the probability of taking these paths:

Table 3.1 Path probabilities

Path	Probability	
$\{0, 1, 2\}$	$p_1 p_2$	$=: \pi_1$
$\{0, 1, 0\}$	$p_1(1 - p_2)$	$=: \pi_2$
$\{0, -1, 0\}$	$(1 - p_1)p_3$	$=: \pi_3$
$\{0, -1, -2\}$	$(1 - p_1)(1 - p_3)$	$=: \pi_4$

We could view this mapping of paths to path probabilities as encoding the measure $\mathbb{P}$. If we knew π_1, π_2, π_3 and π_4, then (as long as all of them are strictly between 0 and 1) we know p_1, p_2 and p_3. Thus if we represent our process with a non-recombining tree, we can label each of the paths at the end with the π-information encoding the measure.

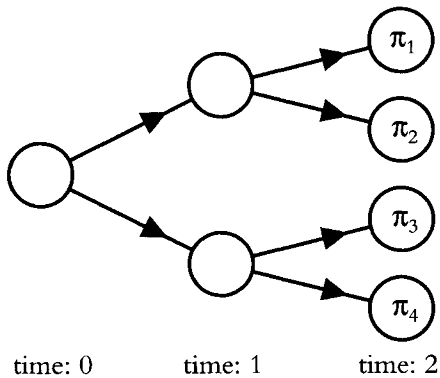


Figure 3.10 Tree with path probabilities marked

Now suppose we had a different measure $\mathbb{Q}$ with probabilities q_1, q_2 and q_3. Again we can code this up with path probabilities, say π'_1, π'_2, π'_3 and π'_4. And again if each π' is strictly between 0 and 1, π'_1, π'_2, π'_3 and π'_4 uniquely decides $\mathbb{Q}$.

And with this encoding, there is a very natural way of encoding the differences between $\mathbb{P}$ and $\mathbb{Q}$, giving some idea of how to distort $\mathbb{P}$ so as to produce $\mathbb{Q}$. If we form the ratio π'_i/π_i for each path i, we write the mapping of paths to this ratio as $\frac{d\mathbb{Q}}{d\mathbb{P}}$. This random variable (random because it depends on the path) is called the *Radon–Nikodym derivative* of $\mathbb{Q}$ with respect to $\mathbb{P}$ up to time 2.

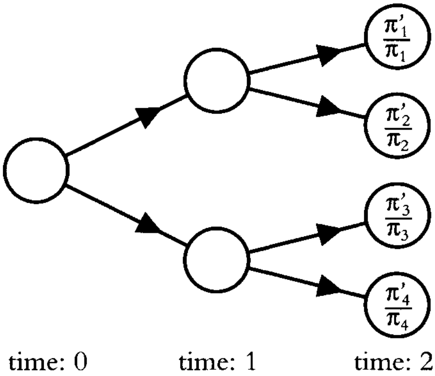


Figure 3.11 Tree with Radon–Nikodym derivative marked

From $\frac{d\mathbb{Q}}{d\mathbb{P}}$ we can derive $\mathbb{Q}$ from $\mathbb{P}$. How? If we have $\mathbb{P}$, then we have $\pi_1, \pi_2, \ldots, \pi_4$, and $\frac{d\mathbb{Q}}{d\mathbb{P}}$ gives us the ratios π'_i/π_i, so we have $\pi'_1, \pi'_2, \ldots, \pi'_4$. And thus $\mathbb{Q}$.

What about p_i or q_i being zero or one? Two things happen – firstly it can become impossible to back out the p_i from the π_i. Consider if p_1 is zero then both π_1 and π_2 are zero and so information about p_2 is lost. But then of course, the paths corresponding to π_1 and π_2 are both impossible (probability zero), so in some sense p_2 really isn't relevant. If we restrict ourselves to only providing π_i for possible paths, then we *can* recover the corresponding p's.

The second problem has a similar flavour but is more serious. Suppose one of the p's is zero, but none of the q's are. Then at least one π_i will be zero when none of the π'_i are. Not all the ratios π'_i/π_i will be well defined, and thus $\frac{d\mathbb{Q}}{d\mathbb{P}}$ can't exist. We could suppress those paths which had path probability zero, but now we *have* lost something. Those paths may have been $\mathbb{P}$-impossible but they are $\mathbb{Q}$-possible. If we throw them away, then we have lost information about $\mathbb{Q}$ just where it is relevant – paths which are $\mathbb{Q}$-possible. Somehow we can't define $\frac{d\mathbb{Q}}{d\mathbb{P}}$ if $\mathbb{Q}$ allows something which $\mathbb{P}$ doesn't. And of course *vice versa*.

This is important enough to formalise.

Equivalence

Two measures $\mathbb{P}$ and $\mathbb{Q}$ are *equivalent* if they operate on the same sample space and agree on what is possible. Formally, if A is any event in the sample space,

$$\mathbb{P}(A) > 0 \iff \mathbb{Q}(A) > 0.$$

In other words, if A is possible under $\mathbb{P}$ then it is possible under $\mathbb{Q}$, and if A is impossible under $\mathbb{P}$ then it is also impossible under $\mathbb{Q}$. And *vice versa.*

We can only meaningfully define $\frac{d\mathbb{Q}}{d\mathbb{P}}$ and $\frac{d\mathbb{P}}{d\mathbb{Q}}$ if $\mathbb{P}$ and $\mathbb{Q}$ are equivalent, and then only where paths are $\mathbb{P}$-possible. But of course if paths are $\mathbb{P}$-impossible then we know how $\mathbb{Q}$ acts on those paths – if $\mathbb{Q}$ is equivalent to $\mathbb{P}$ then they are $\mathbb{Q}$-impossible as well.

Thus two measures $\mathbb{P}$ and $\mathbb{Q}$ must be equivalent before they will have Radon–Nikodym derivatives $\frac{d\mathbb{Q}}{d\mathbb{P}}$ and $\frac{d\mathbb{P}}{d\mathbb{Q}}$.

Expectation and $\frac{d\mathbb{Q}}{d\mathbb{P}}$

While we are still working with discrete processes, we should stock up on some facts about expectation and the Radon–Nikodym derivative. One of the reasons for defining it was the efficient coding it represented. Everything we needed to know about $\mathbb{Q}$ could be extracted from $\mathbb{P}$ and $\frac{d\mathbb{Q}}{d\mathbb{P}}$.

Consider then a claim X known by time 2 on our discrete two-step process. The claim X is a random variable, or in other words a mapping from paths to values – we can let x_i denote the value the claim takes if path i is followed. So the expectation of X with respect to $\mathbb{P}$ is given by

$$\mathbb{E}_{\mathbb{P}}(X) = \sum_i \pi_i x_i,$$

where i ranges over all four possible paths. And the expectation of X with respect to $\mathbb{Q}$ is

$$\mathbb{E}_{\mathbb{Q}}(X) = \sum_i \pi'_i x_i = \sum_i \pi_i \left(\frac{\pi'_i}{\pi_i} x_i\right) = \mathbb{E}_{\mathbb{P}}\left(\frac{d\mathbb{Q}}{d\mathbb{P}} X\right).$$

Just like X, $\frac{d\mathbb{Q}}{d\mathbb{P}}$ is a random variable which we can take the expectation of. And the conversion from $\mathbb{Q}$ to $\mathbb{P}$ is pleasingly simple: $\mathbb{E}_{\mathbb{Q}}(X) = \mathbb{E}_{\mathbb{P}}(\frac{d\mathbb{Q}}{d\mathbb{P}} X)$.

Attractive though this is, it represents just one simple case: $\frac{d\mathbb{Q}}{d\mathbb{P}}$ is defined with a particular time horizon in mind – the ends of the paths, in this case $T = 2$. We specified X at this time and we only wanted an unconditioned expectation. In formal terms, the result we derived was

$$\mathbb{E}_{\mathbb{Q}}(X_T \mid \mathcal{F}_0) = \mathbb{E}_{\mathbb{P}}\Big(\frac{d\mathbb{Q}}{d\mathbb{P}} X_T \;\Big|\; \mathcal{F}_0\Big),$$

where T is the time horizon for $\frac{d\mathbb{Q}}{d\mathbb{P}}$ and X_T is known at time T. What about $\mathbb{E}_{\mathbb{Q}}(X_t|\mathcal{F}_s)$ for t not equal to T and s not equal to zero? We need somehow to know $\frac{d\mathbb{Q}}{d\mathbb{P}}$ not just for the ends of paths but everywhere – $\frac{d\mathbb{Q}}{d\mathbb{P}}$ is a random variable, but we would like a process.

Radon–Nikodym process

We can do this by letting the time horizon vary, and setting ζ_t to be the Radon–Nikodym derivative taken up to the horizon t. That is, ζ_t is the Radon–Nikodym derivative $\frac{d\mathbb{Q}}{d\mathbb{P}}$ but only following paths up to time t, and only looking at the ratio of probabilities up to that time. For instance, at time 1, the possible paths are $\{0, 1\}$ and $\{0, -1\}$ and the derivative ζ_1 has values on them of q_1/p_1 and $(1 - q_1)/(1 - p_1)$ respectively. At time zero, the derivative process ζ_0 is just 1, as the only 'path' is the point $\{0\}$ which has probability 1 under both $\mathbb{P}$ and $\mathbb{Q}$. Concretely, we can fill in ζ_t on our tree in terms of the p's and q's (figure 3.12).

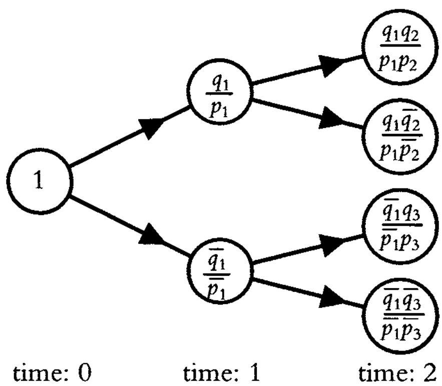


Figure 3.12 Tree with ζ_t process marked ($\bar{p}_i = 1 - p_i$, $\bar{q}_i = 1 - q_i$)

In fact there is another expression for ζ_t as the conditional expectation of the T-horizon Radon–Nikodym derivative,

$$\zeta_t = \mathbb{E}_{\mathbb{P}}\left(\frac{d\mathbb{Q}}{d\mathbb{P}} \,\middle|\, \mathcal{F}_t\right),$$

for every t less than or equal to the horizon T.

Exercise 3.7 Prove that this equation holds for $t = 0, 1, 2$.

We can see that the expectation with respect to $\mathbb{P}$ unpicks the $\frac{d\mathbb{Q}}{d\mathbb{P}}$ in just the right way. The process ζ_t represents just what we wanted – an idea of the amount of change of measure so far up to time t along the current path. If we wanted to know $\mathbb{E}_{\mathbb{Q}}(X_t)$ it would be $\mathbb{E}_{\mathbb{P}}(\zeta_t X_t)$, where X_t is a claim known at time t. If we want to know $\mathbb{E}_{\mathbb{Q}}(X_t|\mathcal{F}_s)$ then we need the amount of change of measure from time s to time t – which is just ζ_t/ζ_s. That is, the change up to time t with the change up to time s removed. In other words

$$\mathbb{E}_{\mathbb{Q}}(X_t \mid \mathcal{F}_s) = \zeta_s^{-1}\mathbb{E}_{\mathbb{P}}(\zeta_t X_t \mid \mathcal{F}_s).$$

Exercise 3.8 Prove this on the tree.

Radon-Nikodym summary

Given $\mathbb{P}$ and $\mathbb{Q}$ equivalent measures and a time horizon T, we can define a random variable $\frac{d\mathbb{Q}}{d\mathbb{P}}$ defined on $\mathbb{P}$-possible paths, taking positive real values, such that

(i) $\mathbb{E}_{\mathbb{Q}}(X_T) = \mathbb{E}_{\mathbb{P}}\left(\frac{d\mathbb{Q}}{d\mathbb{P}} X_T\right)$, for all claims X_T knowable by time T.

(ii) $\mathbb{E}_{\mathbb{Q}}(X_t \mid \mathcal{F}_s) = \zeta_s^{-1}\mathbb{E}_{\mathbb{P}}(\zeta_t X_t \mid \mathcal{F}_s)$, $s \leqslant t \leqslant T$,

where ζ_t is the process $\mathbb{E}_{\mathbb{P}}(\frac{d\mathbb{Q}}{d\mathbb{P}}|\mathcal{F}_t)$.

Change of measure – the continuous Radon–Nikodym derivative

What now? To define a measure for Brownian motion it seems we have to be able to write down the likelihood of every possible path the process can take, ranging across not only a continuous-valued state space but also a continuous-valued time line. Standard probability theory gives some clue to the technology required, if we were content merely to represent the marginal distributions for the process at each time. Despite the continuous nature of the state space, we know that we can express likelihoods in terms of a probability density function.

For example, the measure $\mathbb{P}$ on the real numbers, corresponding to a normal $N(0,1)$ random variable X, can be represented via the density $f_{\mathbb{P}}(x)$, where

$$f_{\mathbb{P}}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

In some loose sense, $f_{\mathbb{P}}(x)$ represents the relative likelihood of the event $\{X = x\}$ occurring. Or, less informally the probability that X lies between x and $x + dx$ is approximately $f_{\mathbb{P}}(x)dx$. In exact terms, the probability that X takes a value in some subset A of the reals is

$$\mathbb{P}(X \in A) = \int_A \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx.$$

For example, the chance of X being in the interval $[0,1]$ is the integral of the density over the interval, $\int_0^1 f_{\mathbb{P}}(x)\, dx$, which has value 0.3413.

But marginal distributions aren't enough – a single marginal distribution won't capture the nature of the process (we can see that clearly even on a discrete tree). Nor will all the marginal distributions for each time t. We need nothing less than all the marginal distributions at each time t *conditional on every history* $\mathcal{F}_s$ for all times $s < t$. We need to capture the idea of a likelihood of a path in the continuous case, by means of some conceptual handle on a particular path specified for all times $t < T$.

One approach is to specify a path if not for all times before the horizon T, then at least for some arbitrarily large yet still finite set of times $\{t_0 = 0, t_1, \ldots, t_{n-1}, t_n = T\}$. Consider then, the set of paths which go through the points $\{x_1, \ldots, x_n\}$ at times $\{t_1, \ldots, t_n\}$. If there were just one time t_1 and one point x_1, then we could write down the likelihood of such a path. We could use the probability density function of W_{t_1}, $f^1_{\mathbb{P}}(x)$, which is the density function of a normal $N(0, t_1)$, or

$$f_{\mathbb{P}}^1(x) = \frac{1}{\sqrt{2\pi t_1}} \exp\left(-\frac{x^2}{2t_1}\right).$$

And if we can do this for one time t_1, then we can for finitely many t_i. All we require is the joint likelihood function $f_{\mathbb{P}}^n(x_1, \ldots, x_n)$ for the process taking values $\{x_1, \ldots, x_n\}$ at times $\{t_1, \ldots, t_n\}$.

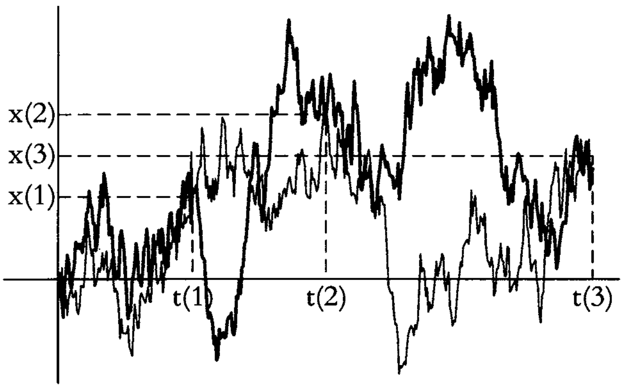


Figure 3.13 Two Brownian motions agreeing on the set $\{t_1, t_2, t_3\}$

Joint likelihood function for Brownian motion
If we take t_0 and x_0 to be zero, and write Δx_i for $x_i - x_{i-1}$ and $\Delta t_i = t_i - t_{i-1}$, then given the third condition of Brownian motion that increments $\Delta W_i = W(t_i) - W(t_{i-1})$ are mutually independent, we can write down

$$f_{\mathbb{P}}^n(x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi \Delta t_i}} \exp\left(-\frac{(\Delta x_i)^2}{2\Delta t_i}\right).$$

So we can write down a likelihood function corresponding to the measure $\mathbb{P}$ for a process on a finite set of times. And in the continuous limit, we have a handle on the measure $\mathbb{P}$ for a continuous process. If A is some subset of $\mathbb{R}^n$, then the $\mathbb{P}$-probability that the random n-vector $(W_{t_1}, \ldots, W_{t_n})$ is in A is exactly the integral over A of the likelihood function $f_{\mathbb{P}}^n$.

Radon–Nikodym derivative – continuous version

Suppose $\mathbb{P}$ and $\mathbb{Q}$ are equivalent measures. Given a path ω, for every ordered time mesh $\{t_1, \ldots, t_n\}$ (with $t_n = T$), we define x_i to be $W_{t_i}(\omega)$, and then the derivative $\frac{d\mathbb{Q}}{d\mathbb{P}}$ up to time T is defined to be the limit of the likelihood ratios

$$\frac{d\mathbb{Q}}{d\mathbb{P}}(\omega) = \lim_{n\to\infty} \frac{f^n_{\mathbb{Q}}(x_1, \ldots, x_n)}{f^n_{\mathbb{P}}(x_1, \ldots, x_n)},$$

as the mesh becomes dense in the interval $[0, T]$.

This continuous-time derivative $\frac{d\mathbb{Q}}{d\mathbb{P}}$ still satisfies the results that

(i) $$\mathbb{E}_{\mathbb{Q}}(X_T) = \mathbb{E}_{\mathbb{P}}\left(\frac{d\mathbb{Q}}{d\mathbb{P}} X_T\right),$$

(ii) $$\mathbb{E}_{\mathbb{Q}}(X_t \mid \mathcal{F}_s) = \zeta_s^{-1}\mathbb{E}_{\mathbb{P}}(\zeta_t X_t \mid \mathcal{F}_s), \qquad s \leqslant t \leqslant T,$$

where ζ_t is the process $\mathbb{E}_{\mathbb{P}}(\frac{d\mathbb{Q}}{d\mathbb{P}} | \mathcal{F}_t)$, and X_t is any process adapted to the history $\mathcal{F}_t$.

Just as the measure $\mathbb{P}$ can be approached through a limiting time mesh, so can the Radon-Nikodym derivative $\frac{d\mathbb{Q}}{d\mathbb{P}}$. The event of paths agreeing with ω on the mesh, $A = \{\omega' : W_{t_i}(\omega') = W_{t_i}(\omega),\ i = 1, \ldots, n\}$, gets smaller and smaller till it is just the single point-set $\{\omega\}$. The Radon–Nikodym derivative can be thought of as the limit

$$\frac{d\mathbb{Q}}{d\mathbb{P}}(\omega) = \lim_{A\to\{\omega\}} \frac{\mathbb{Q}(A)}{\mathbb{P}(A)}.$$

Simple changes of measure – Brownian motion plus constant drift

We have the mechanics of change of measure but still no clue about what change of measure does in the continuous world. Suppose, for example, we had a $\mathbb{P}$-Brownian motion W_t. What does W_t *look* like under an equivalent measure $\mathbb{Q}$ – is it still recognisably Brownian motion or something quite different?

Foresight can provide one simple example. Consider W_t a $\mathbb{P}$-Brownian motion, then (out of nowhere) define $\mathbb{Q}$ to be a measure equivalent to $\mathbb{P}$ via

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left(-\gamma W_T - \tfrac{1}{2}\gamma^2 T\right),$$

for some time horizon T. What does W_t look like with respect to $\mathbb{Q}$?

One place to start, and it is just a start, is to look at the marginal of W_T under $\mathbb{Q}$. We need to find the likelihood function of W_T with respect to $\mathbb{Q}$, or something equivalent. One useful trick is to look at moment-generating functions:

Identifying normals
A random variable X is a normal $N(\mu, \sigma^2)$ under a measure $\mathbb{P}$ if and only if

$$\mathbb{E}_{\mathbb{P}}\big(\exp(\theta X)\big) = \exp\big(\theta\mu + \tfrac{1}{2}\theta^2\sigma^2\big), \qquad \text{for all real } \theta.$$

To calculate $\mathbb{E}_{\mathbb{Q}}\big(\exp(\theta W_T)\big)$, we can use fact (i) of the Radon–Nikodym derivative summary, which tells us that it is the same as the $\mathbb{P}$-expectation $\mathbb{E}_{\mathbb{P}}\big(\frac{d\mathbb{Q}}{d\mathbb{P}} \exp(\theta W_T)\big)$. This equals

$$\mathbb{E}_{\mathbb{P}}\big(\exp(-\gamma W_T - \tfrac{1}{2}\gamma^2 T + \theta W_T)\big) = \exp\big(-\tfrac{1}{2}\gamma^2 T + \tfrac{1}{2}(\theta - \gamma)^2 T\big),$$

because W_T is a normal $N(0, T)$ with respect to $\mathbb{P}$.

Simplifying the algebra, we have

$$\mathbb{E}_{\mathbb{Q}}\big(\exp(\theta W_T)\big) = \exp(-\theta\gamma T + \tfrac{1}{2}\theta^2 T),$$

which is the moment-generating function of a normal $N(-\gamma T, T)$. Thus the marginal distribution of W_T, under $\mathbb{Q}$, is also a normal with variance T but with mean $-\gamma T$.

What about W_t for t less than T? The marginal distribution of W_T is what we would expect if W_t under $\mathbb{Q}$ were a Brownian motion plus a constant drift $-\gamma$. Of course, a lot of other process also have a marginal normal $N(-\gamma T, T)$ distribution at time T, but it would be an elegant result if the sole effect of changing from $\mathbb{P}$ to $\mathbb{Q}$ via $\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp(-\gamma W_T - \frac{1}{2}\gamma^2 T)$ were just to punch in a drift of $-\gamma$.

And so it is. The process W_t is a Brownian motion with respect to $\mathbb{P}$ and Brownian motion with constant drift $-\gamma$ under $\mathbb{Q}$. Using our two results about $\frac{d\mathbb{Q}}{d\mathbb{P}}$, we can prove the three conditions for $\tilde{W}_t = W_t + \gamma t$ to be $\mathbb{Q}$-Brownian motion:

(i) $\tilde{W}_t$ is continuous and $\tilde{W}_0 = 0$;

(ii) $\tilde{W}_t$ is a normal $N(0,t)$ under $\mathbb{Q}$;

(iii) $\tilde{W}_{t+s} - \tilde{W}_s$ is a normal $N(0,t)$ independent of $\mathcal{F}_s$.

The first of these is true and (ii) and (iii) can be re-expressed as

(ii)′ $\mathbb{E}_{\mathbb{Q}}\big(\exp(\theta\tilde{W}_t)\big) = \exp(\frac{1}{2}\theta^2 t)$;

(iii)′ $\mathbb{E}_{\mathbb{Q}}\big(\exp\big(\theta(\tilde{W}_{t+s} - \tilde{W}_s)\big) \mid \mathcal{F}_s\big) = \exp(\frac{1}{2}\theta^2 t)$.

Exercise 3.9 Show that (ii)′ and (iii)′ are equivalent to (ii) and (iii) respectively, and prove them using the change of measure process $\zeta_t = \mathbb{E}_{\mathbb{P}}(\frac{d\mathbb{Q}}{d\mathbb{P}}|\mathcal{F}_t)$.

That both W_t and $\tilde{W}_t$ are Brownian motion, albeit with respect to different measures, seems paradoxical. But switching from $\mathbb{P}$ to $\mathbb{Q}$ just changes the relative likelihood of a particular path being chosen. For example, W *might* follow a path which drifts downwards for a time at a rate of about $-\gamma$. Although that path is $\mathbb{P}$-unlikely, it is $\mathbb{P}$-possible. Under $\mathbb{Q}$, on the other hand, such a path is much more likely, and the chances are that is what we see. But it still could be just improbable Brownian motion behaviour.

We can see this in the Radon–Nikodym derivative $\frac{d\mathbb{Q}}{d\mathbb{P}}$ which is large when W_T is very negative, and small when W_T is closer to zero or positive. This is just the consequence of the common sense thought that paths which end up negative are more likely under $\mathbb{Q}$ (Brownian motion plus downward drift) than they are under $\mathbb{P}$ (driftless Brownian motion). Correspondingly, paths which finish near or above zero are less likely under $\mathbb{Q}$ than $\mathbb{P}$.

Cameron–Martin–Girsanov

So this one change of measure just changed a vanilla Brownian motion into one with drift – nothing else. And of course, drift is one of the elements of our stochastic differential form of processes. In fact *all* that measure changes on Brownian motion can do is to change the drift. All the processes that we are interested in are representable as instantaneous differentials made up of some amount of Brownian motion and some amount of drift. The mapping of stochastic differentials under $\mathbb{P}$ to stochastic differentials under $\mathbb{Q}$ is both natural and pleasing.

This is what our theorem provides.

Cameron–Martin–Girsanov theorem
If W_t is a $\mathbb{P}$-Brownian motion and γ_t is an $\mathcal{F}$-previsible process satisfying the boundedness condition $\mathbb{E}_{\mathbb{P}} \exp\left(\frac{1}{2}\int_0^T \gamma_t^2\,dt\right) < \infty$, then there exists a measure $\mathbb{Q}$ such that

(i) $\mathbb{Q}$ is equivalent to $\mathbb{P}$

(ii) $$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left(-\int_0^T \gamma_t\,dW_t - \frac{1}{2}\int_0^T \gamma_t^2\,dt\right)$$

(iii) $\tilde{W}_t = W_t + \int_0^t \gamma_s\,ds$ is a $\mathbb{Q}$-Brownian motion.

In other words, W_t is a drifting $\mathbb{Q}$-Brownian motion with drift $-\gamma_t$ at time t.

Within constraints, if we want to turn a $\mathbb{P}$-Brownian motion W_t into a Brownian motion with some specified drift $-\gamma_t$, then there's a $\mathbb{Q}$ which does it.

Within limits, drift is measure and measure drift.

Conversely to the theorem,

Cameron–Martin–Girsanov converse
If W_t is a $\mathbb{P}$-Brownian motion, and $\mathbb{Q}$ is a measure equivalent to $\mathbb{P}$, then there exists an $\mathcal{F}$-previsible process γ_t such that

$$\tilde{W}_t = W_t + \int_0^t \gamma_s\,ds$$

is a $\mathbb{Q}$-Brownian motion. That is, W_t plus drift γ_t is $\mathbb{Q}$-Brownian motion. Additionally the Radon–Nikodym derivative of $\mathbb{Q}$ with respect to $\mathbb{P}$ (at time T) is $\exp(-\int_0^T \gamma_t\,dW_t - \frac{1}{2}\int_0^T \gamma_t^2\,dt)$.

C-M-G and stochastic differentials

The C-M-G theorem applies to Brownian motion, but all our processes are disguised Brownian motions at heart. Now we can see the rewards of our Brownian calculus instantly – C-M-G becomes a powerful tool for controlling the drift of *any* process.

3.4 Change of measure – the C-M-G theorem

Suppose that X is a stochastic process with increment

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

where W is a $\mathbb{P}$-Brownian motion. Suppose we want to find if there is a measure $\mathbb{Q}$ such that the drift of process X under $\mathbb{Q}$ is $\nu_t \, dt$ instead of $\mu_t \, dt$. As a first step, dX can be rewritten as

$$dX_t = \sigma_t \left(dW_t + \left(\frac{\mu_t - \nu_t}{\sigma_t} \right) dt \right) + \nu_t \, dt.$$

If we set γ_t to be $(\mu_t - \nu_t)/\sigma_t$, and if γ then satisfies the C-M-G growth condition ($\mathbb{E}_{\mathbb{P}} \exp(\frac{1}{2} \int_0^T \gamma_t^2 \, dt) < \infty$) then indeed there is a new measure $\mathbb{Q}$ such that $\tilde{W}_t := W_t + \int_0^t (\mu_s - \nu_s)/\sigma_s \, ds$ is a $\mathbb{Q}$-Brownian motion.

But this means that the differential of X under $\mathbb{Q}$ is

$$dX_t = \sigma_t \, d\tilde{W}_t + \nu_t \, dt,$$

where $\tilde{W}$ is a $\mathbb{Q}$-Brownian motion – which gives X the drift ν_t we wanted.

We can also set limits on the changes that changing to an equivalent measure can wreak on a process. Since the change of measure can only change the Brownian motion to a Brownian motion plus drift, the volatility of the process must remain the same.

Examples – changes of measure

1. Let X_t be the drifting Brownian process $\sigma W_t + \mu t$, where W is a $\mathbb{P}$-Brownian motion and σ and μ are both constant. Then using C-M-G with $\gamma_t = \mu/\sigma$, there exists an equivalent measure $\mathbb{Q}$ under which $\tilde{W}_t = W_t + (\mu/\sigma)t$ and $\tilde{W}$ is a $\mathbb{Q}$-Brownian motion up to time T. Then $X_t = \sigma \tilde{W}_t$, which is (scaled) $\mathbb{Q}$-Brownian motion.
   The measures also give rise to different expectations. For example, $\mathbb{E}_{\mathbb{P}}(X_t^2)$ equals $\mu^2 t^2 + \sigma^2 t$, but $\mathbb{E}_{\mathbb{Q}}(X_t^2) = \sigma^2 t$.

2. Let X_t be the exponential Brownian motion with SDE

   $$dX_t = X_t(\sigma \, dW_t + \mu \, dt),$$

   where W is $\mathbb{P}$-Brownian motion. Can we change measure so that X has the new SDE

   $$dX_t = X_t(\sigma \, dW_t + \nu \, dt),$$

for some arbitrary constant drift ν?
Using C-M-G with $\gamma_t = (\mu - \nu)/\sigma$, there is indeed a measure $\mathbb{Q}$ under which $\tilde{W}_t = W_t + (\mu - \nu)t/\sigma$ is a $\mathbb{Q}$-Brownian motion. Then X does have the SDE

$$dX_t = X_t(\sigma\, d\tilde{W}_t + \nu\, dt),$$

where $\tilde{W}$ is a $\mathbb{Q}$-Brownian motion.

3.5 Martingale representation theorem

We can solve some SDEs with Itô; we can see how SDEs change as measure changes. But central to answering our pricing question in chapter two was the concept of a measure with respect to which the process was expected to stay the same, the *martingale measure* for our discrete trees. The price of derivatives turned out to be an expectation under this measure, and the construction of this expectation even showed us the trading strategy required to justify this price. And so it is here.

First the description again:

Martingales
A stochastic process M_t is a *martingale* with respect to a measure $\mathbb{P}$ if and only if

(i) $\mathbb{E}_{\mathbb{P}}(|M_t|) < \infty$, for all t

(ii) $\mathbb{E}_{\mathbb{P}}(M_t \mid \mathcal{F}_s) = M_s$, for all $s \leqslant t$.

The first condition is merely a technical sweetener, it is the second that carries the weight. A martingale measure is one which makes the expected future value conditional on its present value and past history merely its present value. It isn't *expected* to drift upwards or downwards.

Some examples:

(1) Trivially, the constant process $S_t = c$ (for all t) is a martingale with respect to any measure: $\mathbb{E}_{\mathbb{P}}(S_t|\mathcal{F}_s) = c = S_s$, for all $s \leqslant t$, and for any

measure $\mathbb{P}$.

(2) Less trivially, $\mathbb{P}$-Brownian motion is a $\mathbb{P}$-martingale. Intuitively this makes sense – Brownian motion doesn't move consistently up or down, it's as likely to do either. But we should get into the habit of checking this formally: we need $\mathbb{E}_{\mathbb{P}}(W_t|\mathcal{F}_s) = W_s$. Of course we have that the increment $W_t - W_s$ is independent of $\mathcal{F}_s$ and distributed as a normal $N(0, t-s)$, so that $\mathbb{E}_{\mathbb{P}}(W_t - W_s|\mathcal{F}_s) = 0$. This yields the result, as

$$\mathbb{E}_{\mathbb{P}}(W_t|\mathcal{F}_s) = \mathbb{E}_{\mathbb{P}}(W_s|\mathcal{F}_s) + \mathbb{E}_{\mathbb{P}}(W_t - W_s|\mathcal{F}_s) = W_s + 0.$$

(3) For any claim X depending only on events up to time T, the process $N_t = \mathbb{E}_{\mathbb{P}}(X|\mathcal{F}_t)$ is a $\mathbb{P}$-martingale (assuming only the technical constraint $\mathbb{E}_{\mathbb{P}}(|X|) < \infty$).

Example (3) is an elegant little trick for producing martingales – and as we shall see (and have already seen in chapter two) central to pricing derivatives. First why? Convince yourself that $N_t = \mathbb{E}_{\mathbb{P}}(X|\mathcal{F}_t)$ is a well-defined *process* – the first stage of the alchemy is the introduction of a time line into the random variable X. Now for N_t to be a $\mathbb{P}$-martingale, we require $\mathbb{E}_{\mathbb{P}}(N_t|\mathcal{F}_s) = N_s$, but for this we merely need to be satisfied that

$$\mathbb{E}_{\mathbb{P}}\Big(\mathbb{E}_{\mathbb{P}}(X|\mathcal{F}_t) \;\Big|\; \mathcal{F}_s\Big) = \mathbb{E}_{\mathbb{P}}(X \mid \mathcal{F}_s).$$

That is, that conditioning firstly on information up to time t and then on information up to time s is just the same as conditioning up to time s to begin with. This property of conditional expectation is the *tower law.*

Exercise 3.10 Show that the process $X_t = W_t + \gamma t$, where W_t is a $\mathbb{P}$-Brownian motion, is a $\mathbb{P}$-martingale if and only if $\gamma = 0$.

Representation

In chapter two, we had a binomial representation theorem – if M_t and N_t are both $\mathbb{P}$-martingales then they share more than just the name – locally they can only differ by a scaling, by the size of the opening of each particular branching. We could represent *changes* in N_t by scaled *changes* in the other non-trivial $\mathbb{P}$-martingale. Thus N_t itself can be represented by the scaled *sum* of these changes.

In the continuous world:

Martingale representation theorem

Suppose that M_t is a $\mathbb{Q}$-martingale process, whose volatility σ_t satisfies the additional condition that it is (with probability one) always non-zero. Then if N_t is any other $\mathbb{Q}$-martingale, there exists an $\mathcal{F}$-previsible process ϕ such that $\int_0^T \phi_t^2 \sigma_t^2 \, dt < \infty$ with probability one, and N can be written as

$$N_t = N_0 + \int_0^t \phi_s \, dM_s.$$

Further ϕ is (essentially) unique.

This is virtually identical to the earlier result, with summation replaced by an integral. As we are getting used to, the move to a continuous process extracts a formal technical penalty. In this case, the $\mathbb{Q}$-martingale's volatility must be positive with probability 1 – but otherwise our chapter two result has carried across unchanged. If there is a measure $\mathbb{Q}$ under which M_t is a $\mathbb{Q}$-martingale, then any other $\mathbb{Q}$-martingale can be represented in terms of M_t. The process ϕ_t is simply the ratio of their respective volatilities.

Driftlessness

We need just one more tool. Thrown into the discussion of martingales was the intuitive description of a martingale as neither drifting up or drifting down. We have, though, a technical definition of drift via our stochastic differential formulation. An obvious question springs to mind: are stochastic processes with no drift term always martingales, and *vice versa* can martingales always be represented as just $\sigma_t \, dW_t$ for some $\mathcal{F}$-previsible volatility process σ_t?

Nearly.

One way round we can do for ourselves with the martingale representation theorem. If a process X_t is a $\mathbb{P}$-martingale then with W_t a $\mathbb{P}$-Brownian motion, we have an $\mathcal{F}$-previsible process ϕ_t such that

$$X_t = X_0 + \int_0^t \phi_s \, dW_s.$$

This is just the integral form of the increment $dX_t = \phi_t \, dW_t$, which has no drift term.

The other way round is true (up to a technical constraint), but harder. For reference:

A collector's guide to martingales
If X is a stochastic process with volatility σ_t (that is $dX_t = \sigma_t \, dW_t + \mu_t \, dt$) which satisfies the technical condition $\mathbb{E}\big[(\int_0^T \sigma_s^2 \, ds)^{\frac{1}{2}}\big] < \infty$, then

$$X \text{ is a martingale} \iff X \text{ is driftless } (\mu_t \equiv 0).$$

If the technical condition fails, a driftless process may not be a martingale. Such processes are called *local martingales*.

Exponential martingales

The technical constraint can be tiresome. For example, take the (driftless) SDE for an exponential process $dX_t = \sigma_t X_t \, dW_t$. The condition (in this case, $\mathbb{E}\big[(\int_0^T \sigma_s^2 X_s^2 \, ds)^{\frac{1}{2}}\big] < \infty$) is difficult to check, but for these specific exponential examples, a better (more practical) test is:

A collector's guide to exponential martingales
If $dX_t = \sigma_t X_t \, dW_t$, for some $\mathcal{F}$-previsible process σ_t, then

$$\mathbb{E}\Big(\exp\big(\tfrac{1}{2}\textstyle\int_0^T \sigma_s^2 \, ds\big)\Big) < \infty \Rightarrow X \text{ is a martingale.}$$

We also note that the solution to the SDE is $X_t = X_0 \exp\big(\int_0^t \sigma_s \, dW_s - \frac{1}{2}\int_0^t \sigma_s^2 \, ds\big)$.

Exercise 3.11 If σ_t is a bounded function of both time and sample path, show that $dX_t = \sigma_t X_t \, dW_t$ is a $\mathbb{P}$-martingale.

3.6 Construction strategies

We have the mathematical tools – Itô, Cameron–Martin–Girsanov, and the martingale representation theorem – now we need some idea of how to hook them into a financial model. In the simplest models, Black–Scholes for example, we'll have a market consisting of one random security and a riskless cash account bond; and with this comes the idea of a portfolio.

The portfolio (ϕ, ψ)
A *portfolio* is a pair of processes ϕ_t and ψ_t which describe respectively the number of units of security and of the bond which we hold at time t. The processes can take positive or negative values (we'll allow unlimited short-selling of the stock or bond). The security component of the portfolio ϕ should be *$\mathcal{F}$-previsible*: depending only on information up to time t but not t itself.

There is an intuitive way to think about previsibility. If ϕ were left-continuous (that is, ϕ_s tends to ϕ_t as s tends upwards to t from below) then ϕ would be previsible. If ϕ were only right-continuous (that is, ϕ_s tends to ϕ_t only as s tends downwards to t from above), then ϕ need not be.

Self-financing strategies

With the idea of a portfolio comes the idea of a strategy. The description (ϕ_t, ψ_t) is a dynamic strategy detailing the amount of each component to be held at each instant. And one particularly interesting set of strategies or portfolios are those that are financially self-contained or *self financing*.

A portfolio is self-financing if and only if the change in its value only depends on the change of the asset prices. In the discrete framework this was captured via a difference equation, and in the continuous case it is equivalent to an SDE.

What SDE?

With stock price S_t and bond price B_t, the value, V_t, of a portfolio (ϕ_t, ψ_t) at time t is given by $V_t = \phi_t S_t + \psi_t B_t$. At the next time instant, two things happen: the old portfolio changes value because S_t and B_t have changed price; and the old portfolio has to be adjusted to give a new portfolio as instructed by the trading strategy (ϕ, ψ). If the cost of the adjustment is perfectly matched by the profits or losses made by the portfolio then no extra money is required from outside – the portfolio is self-financing.

In our discrete language, we had the difference equation

$$\Delta V_i = \phi_i \, \Delta S_i + \psi_i \, \Delta B_i.$$

In continuous time, we get a stochastic differential equation:

Self-financing property
If (ϕ_t, ψ_t) is a portfolio with stock price S_t and bond price B_t, then

$$(\phi_t, \psi_t) \quad \text{is self-financing} \iff dV_t = \phi_t \, dS_t + \psi_t \, dB_t.$$

Suppose the stock price S_t is given by a simple Brownian motion W_t (so $S_t = W_t$ for all t), and the bond price B_t is constant ($B_t = 1$ for all t). What kind of portfolios are self-financing?

(1) Suppose $\phi_t = \psi_t = 1$ for all t. If we hold a unit of stock and a unit of bond for all time without change, then the value of the portfolio ($V_t = W_t + 1$) may fluctuate, but it will all be due to fluctuation of the stock. Intuitively, no extra money is needed to come in to uphold the (ϕ_t, ψ_t) strategy and none comes out – this (ϕ_t, ψ_t) portfolio ought to be self-financing.

Checking this formally, $V_t = W_t + 1$ implies that $dV_t = dW_t$ which is the same as $\phi_t \, dS_t + \psi_t \, dB_t$, as we required (remembering that $dB_t = 0$).

(2) Suppose $\phi_t = 2W_t$ and $\psi_t = -t - W_t^2$. Here (ϕ_t, ψ_t) is a portfolio, ϕ_t is previsible, and the value $V_t = \phi_t S_t + \psi_t B_t = W_t^2 - t$. By Itô's formula, $dV_t = 2W_t \, dW_t$ which is identical to $\phi_t \, dS_t + \psi_t \, dB_t$ as required.

Exercise 3.12 Verify the Itô claim in (2) above (which also shows that $W_t^2 - t$ is a martingale).

Surprising though it seems: holding as many units of stock as twice its current price, though a rollercoaster strategy, is exactly offset by the stock profits and the changing bond holding of $-(t + W_t^2)$. The (ϕ_t, ψ_t) strategy could (in a perfect market) be followed to our heart's content without further funding.

The second example should convince us that being self-financing is not an automatic property of a portfolio. The Itô check worked, but it could easily have failed if ψ_t had been different – the (ϕ_t, ψ_t) strategy would have required injections or forced outflows of cash. Every time we claim a portfolio is self-financing we have to turn the handle on Itô's formula to check the SDE.

Trading strategies

Now we can define a replicating strategy for a claim:

Replicating strategy

Suppose we are in a market of a riskless bond B and a risky security S with volatility σ_t, and a claim X on events up to time T.

A *replicating strategy* for X is a self-financing portfolio (ϕ, ψ) such that $\int_0^T \sigma_t^2 \phi_t^2 \, dt < \infty$ and $V_T = \phi_T S_T + \psi_T B_T = X$.

Why should we care about replicating strategies? For the same reason as we wanted them in the discrete market models. The claim X gives the value of some derivative which we need to pay off at time T. We want a price if there is one, as of now, given a model for S and B.

If there is a replicating strategy (ϕ_t, ψ_t), then the price of X at time t *must be* $V_t = \phi_t S_t + \psi_t B_t$. (And specifically, the price at time zero is $V_0 = \phi_0 S_0 + \psi_0 B_0$.) If it were lower, a market player could buy one unit of the derivative at time t and sell ϕ_t units of S and ψ_t units of B against it, continuing to be short (ϕ, ψ) until time T. Because (ϕ, ψ) is self-financing and the portfolio is worth X at time T guaranteed, the bought derivative and sold portfolio would safely cancel at time T, and no extra money is required between times t and T. The profit created by the mismatch at time t can be banked there and then without risk. And, as usual with arbitrage, one unit could have been many; no risk means no fear.

And of course if the derivative price had been higher than V_t, then we could have *sold* the derivative and *bought* the self-financing (ϕ, ψ) to the same effect. Replicating strategies, *if* they exist, tie down the price of the claim X not just at payoff but everywhere.

We can lay out a battle plan. We define a market model with a stock price process complex enough to satisfy our need for realism. Then, using whatever tools we have to hand we find replicating strategies for all useful claims X. And if we can, we can price derivatives in the model. The rest of the book consists of upping the stakes in complexity of models and of claims.

3.7 Black–Scholes model

We need a model to cut our teeth on. We have the tools and we've seen the overall approach at the end of chapter two. So taking the stock model of section 3.1, we will use the Cameron–Martin–Girsanov theorem (section 3.4) to change it into a martingale, and then use the martingale representation theorem (section 3.5) to create a replicating strategy for each claim. Itô will oil the works.

The model

Our first model – basic Black–Scholes
We will posit the existence of a deterministic r, μ and σ such that the bond price B_t and the stock price follow

$$\begin{aligned} B_t &= \exp(rt), \\ S_t &= S_0 \exp(\sigma W_t + \mu t), \end{aligned}$$

where r is the *riskless interest rate*, σ is the *stock volatility* and μ is the *stock drift*. There are no transaction costs and both instruments are freely and instantaneously tradable either long or short at the price quoted.

We need a model for the behaviour of the stock – simple enough that we actually can find replicating strategies but not so simple that we can't bring ourselves to believe in it as a model of the real world.

Following in Black and Scholes' footsteps, our market will consist of a riskless constant-interest rate cash bond and a risky tradable stock following an exponential Brownian motion.

As we've seen in section 3.1, it is at least a plausible match to the real world. And as we shall see here, it is quite hard enough to start with.

Zero interest rates

If there's one parameter that throws up a smokescreen around a first run at an analysis of the Black–Scholes model, it's the interest rate r. The problems it causes are more tedious than fatal – as we'll see soon, the tools we have are powerful enough to cope. But we'll temporarily simplify things, and set r to be zero.

So now we begin. For an arbitrary claim X, knowable by some horizon time T, we want to see if we can find a replicating strategy (ϕ_t, ψ_t).

Finding a replicating strategy

We shall follow a three-step process outlined in this box here.

Three steps to replication

(1) Find a measure $\mathbb{Q}$ under which S_t is a martingale.

(2) Form the process $E_t = \mathbb{E}_{\mathbb{Q}}(X|\mathcal{F}_t)$.

(3) Find a previsible process ϕ_t, such that $dE_t = \phi_t \, dS_t$.

The tools described earlier on will be essential to do this. We shall use the Cameron–Martin–Girsanov theorem (section 3.4) for the first step and the martingale representation theorem (section 3.5) for the third one.

Step one

For two different reasons – firstly we need to apply the Cameron–Martin–Girsanov theorem, and secondly we need to be able to tell if S_t is a $\mathbb{Q}$-martingale for a given $\mathbb{Q}$ – we want to find an SDE for S_t.

The stock follows an exponential Brownian motion, $S_t = \exp(\sigma W_t + \mu t)$, so the logarithm of the stock price, $Y_t = \log(S_t)$, follows a simple drifting Brownian motion $Y_t = \sigma W_t + \mu t$. Thus the SDE for Y_t is easy to write down: $dY_t = \sigma\, dW_t + \mu\, dt$. But, of course, Itô makes it possible to write down the SDE for $S_t = \exp(Y_t)$ as

$$dS_t = \sigma S_t\, dW_t + (\mu + \tfrac{1}{2}\sigma^2) S_t\, dt.$$

In order for S_t to be a martingale, the first thing to do is to kill the drift in this SDE. If we let γ_t be a process with constant value $\gamma = (\mu + \frac{1}{2}\sigma^2)/\sigma$, then the C-M-G theorem says that there is a measure $\mathbb{Q}$ such that $\tilde{W}_t = W_t + \gamma t$ is $\mathbb{Q}$-Brownian motion. (The technical boundedness condition is satisfied because γ_t is constant.) Substituting in, the SDE is now

$$dS_t = \sigma S_t\, d\tilde{W}_t.$$

No drift term, thus S_t could be a $\mathbb{Q}$-martingale. The exponential martingales box (section 3.5) contains a condition in terms of σ for S_t to be a martingale under $\mathbb{Q}$. As σ is constant, the condition holds which means that S_t must be a $\mathbb{Q}$-martingale. Consequently, $\mathbb{Q}$ is the *martingale measure* for S_t.

Step two

Given $\mathbb{Q}$, we can convert X into a process by forming $E_t = \mathbb{E}_{\mathbb{Q}}(X|\mathcal{F}_t)$. This is, as we have already discussed in example (3) of section 3.5, a $\mathbb{Q}$-martingale.

Step three

Since there is a $\mathbb{Q}$, under which both E_t and S_t are $\mathbb{Q}$-martingales, we can invoke the martingale representation theorem. There exists a previsible process ϕ_t which constructs $E_t = \mathbb{E}_{\mathbb{Q}}(X|\mathcal{F}_t)$ out of S_t. (To use the theorem, we need to check that the volatility of S_t is always positive, but this is true because the volatility is just σS_t, and both σ and S_t are always positive.) Formally:

$$E_t = \mathbb{E}_{\mathbb{Q}}(X|\mathcal{F}_t) = \mathbb{E}_{\mathbb{Q}}(X) + \int_0^t \phi_s\, dS_s,$$

or, of course, $dE_t = \phi_t\, dS_t$. So the martingale representation theorem tells us an important fact: given a $\mathbb{Q}$ that makes S_t a $\mathbb{Q}$-martingale with positive volatility, $dE_t = \phi_t\, dS_t$ for some ϕ_t.

We need a replicating strategy (ϕ_t, ψ_t), and it's tempting to believe that we have got one half of it. So we should try it, setting ψ_t to be the only thing it can be, given that we want the portfolio to be worth E_t for all t.

Replicating strategy

Our strategy is to:

- hold ϕ_t units of stock at time t and
- hold $\psi_t = E_t - \phi_t S_t$ units of the bond at time t.

Is it self-financing? The value of the portfolio at time t is

$$V_t = \phi_t S_t + \psi_t B_t = E_t,$$

because the bond B_t is constantly equal to 1. Thus $dV_t = dE_t$, but of course $dE_t = \phi_t \, dS_t$, from the martingale representation theorem.

Since dB_t is zero, we have the self-financing condition we want, namely $dV_t = \phi_t \, dS_t + \psi_t \, dB_t$.

Since the terminal value of the strategy V_T is $E_T = X$, we have a replicating strategy for X – which means there is an arbitrage price for X at all times. Specifically there is an arbitrage price for X at time zero – the value of the (ϕ_t, ψ_t) portfolio at time zero, which makes the price E_0, or $\mathbb{E}_{\mathbb{Q}}(X)$. In other words, the price of the claim X is its expected value under the measure that makes the stock process S_t a martingale.

It is worth pausing to let a few surprises sink in. The first is just the fact that there are replicating strategies for arbitrary claims. The model that we have chosen isn't too unrealistic – it has the right kind of behaviour and a healthy degree of randomness. So we might expect to fail in our search for replicating strategies. It is after all particularly odd that despite the lack of knowledge about the claim's eventual value, we can nevertheless trade in the market in such a way that we always produce it.

The second surprise, and just as important, is that the price of the derivative has such a simple expression – the expected value of the claim. It is the easiest thing to forget that this is *not* the expectation of the claim with respect to the real measure of S_t, which is the measure that makes it an exponential Brownian motion with drift μ and volatility σ. All *that* expectation could give us would be a long-term average of the claim's payout. And though that could be a useful thing to know in order to judge whether punting with the derivative is worthwhile in the long run, it doesn't give a price. There is a replicating strategy and thus an arbitrage price for the claim. And arbitrage always wins out.

The price happens to be an expectation, but not *the* expectation in a traditional statistics sense. It could only be *the* expectation if quite by chance the drift μ we believe in for the stock were exactly and precisely right to make S_t a martingale in the first place ($\mu = -\frac{1}{2}\sigma^2$).

The third surprise is the simplicity of the process S_t under its martingale measure. If we actually want to crank the handle and calculate derivative prices for a particular claim, we have to be able to calculate the expected value of the claim under the martingale measure $\mathbb{Q}$. Since the claim depends on S_t, this normally involves calculating the expected value, under $\mathbb{Q}$, of some function of the values of S_t up to $t = T$. If S_t were an unpleasant process under $\mathbb{Q}$, then this task could be unpleasant too. But S_t is *also* an exponential Brownian motion under $\mathbb{Q}$. If we solve the SDE, then

$$S_t = \exp(\sigma \tilde{W}_t - \tfrac{1}{2}\sigma^2 t),$$

and we find that S_t has the same constant volatility σ and a new but also constant drift of $-\frac{1}{2}\sigma^2$. So if we felt that S_t was tractable under its original measure, it is also tractable under the martingale measure.

Non-zero interest rates

Now we can bring the interest rate r back in again. What happens if r is non-zero? We can't just ignore it. Suppose we did, and considered a forward contract with claim $S_T - k$ for some price k. We already know that the k which gives the forward contract a zero value at time zero is $k = S_0 e^{rT}$. The arbitrage to produce this is easy to figure out. But our rule, when r was zero, of simply taking the expected value of the claim under the martingale measure for S_t cannot work. In fact,

$$\mathbb{E}_{\mathbb{Q}}\big(S_T - S_0 e^{rT}\big) = S_0\big(1 - e^{rT}\big) \neq 0.$$

Even discounting the claim won't help in this case. So our rule of finding a measure which makes S_t into a martingale only holds true when r is zero. When r is not zero, the inexorable growth of cash gets in the way.

So we take a guess. If the growth of cash is annoying, simply remove it by discounting everything. We call B_t^{-1} the discount process, and form a discounted stock $Z_t = B_t^{-1} S_t$ and a discounted claim $B_T^{-1} X$.

In this discounted world, we could be forgiven for thinking that r was zero again. So maybe our analysis will work again. Of course, this is all just heuristic justification, and the proof is only in the doing. If we can't find a replicating strategy then, attractive as our guess is, it is also wrong.

Fortunately, we can. Focusing on our discounted stock process Z_t, it is not too hard to write down an SDE

$$dZ_t = Z_t\big(\sigma\, dW_t + (\mu - r + \tfrac{1}{2}\sigma^2)\, dt\big).$$

Exercise 3.13 Prove it.

Step one

To make Z_t into a martingale, we can invoke C-M-G just as before, only now to introduce a drift of $(\mu - r + \frac{1}{2}\sigma^2)/\sigma$ to the underlying Brownian motion. So there exists (another) $\mathbb{Q}$ equivalent to the original measure $\mathbb{P}$ and a $\mathbb{Q}$-Brownian motion $\tilde{W}_t$ such that

$$dZ_t = \sigma Z_t\, d\tilde{W}_t.$$

So Z_t, under $\mathbb{Q}$, is driftless and a martingale.

Step two

We need a process which hits the discounted claim and is also a $\mathbb{Q}$-martingale. And, as before, conditional expectation provides it, namely by forming the process $E_t = \mathbb{E}_{\mathbb{Q}}\big(B_T^{-1}X \mid \mathcal{F}_t\big)$.

Step three

The discounted stock price Z_t is a $\mathbb{Q}$-martingale; and so is the conditional expectation process of the discounted claim E_t. Thus the martingale representation theorem gives us a previsible ϕ_t such that $dE_t = \phi_t\, dZ_t$.

We want to hit the real claim with amounts of the real stock, but in our shadow discounted world we can hit the discounted claim by holding ϕ_t units of the discounted stock. So just as a guess, let us try ϕ_t out in the real world as well.

What about the bond holding? The bond holding in the discounted world is $\psi_t = E_t - \phi_t Z_t$, so we can try that in the real world too. Some reassurance comes from the fact that at time T we will be holding ϕ_T units of the stock and ψ_T units of the bond which will be worth $\phi_T S_T + \psi_T B_T = B_T E_T = X$.

So our replicating strategy is to

- hold ϕ_t units of the stock at time t, and
- hold $\psi_t = E_t - \phi_t Z_t$ units of the bond.

Are we right? The value V_t of the portfolio (ϕ_t, ψ_t) is given by $V_t = \phi_t S_t + \psi_t B_t = B_t E_t$. Thus following exercise 3.6, we can write dV_t as

$$dV_t = B_t\, dE_t + E_t\, dB_t.$$

But dE_t is $\phi_t\, dZ_t$ (our fact from the martingale representation theorem), and so $dV_t = \phi_t B_t\, dZ_t + E_t\, dB_t$. A bit of rearrangement tells us that $E_t = \phi_t Z_t + \psi_t$, and thus

$$dV_t = \phi_t B_t\, dZ_t + (\phi_t Z_t + \psi_t)\, dB_t = \phi_t (B_t\, dZ_t + Z_t\, dB_t) + \psi_t\, dB_t.$$

But, from exercise 3.6 again, $d(B_t Z_t) = B_t\, dZ_t + Z_t\, dB_t$, and since $S_t = B_t Z_t$, we have

$$dV_t = \phi_t\, dS_t + \psi_t\, dB_t.$$

That is, (ϕ_t, ψ_t) is self-financing.

Self-financing strategies

A portfolio strategy (ϕ_t, ψ_t) of holdings in a stock S_t and a non-volatile cash bond B_t has value $V_t = \phi_t S_t + \psi_t B_t$ and discounted value $E_t = \phi_t Z_t + \psi_t$, where Z is the discounted stock process $Z_t = B_t^{-1} S_t$. Then the strategy is self-financing if either

$$dV_t = \phi_t\, dS_t + \psi_t\, dB_t,$$

or equivalently $$dE_t = \phi_t\, dZ_t.$$

A strategy is self-financing if changes in its value are due only to changes in the assets' values, or equivalently if changes in its discounted value are due only to changes in the discounted values of the assets.

Since we know that $V_T = X$, then we have proved that (ϕ_t, ψ_t) is a replicating strategy for X. Our guesses came good.

Summary

Suppose we have a Black–Scholes model for a continuously tradable stock and bond, that is assuming the existence of a constant r, μ and σ such that their respective prices can be represented as $S_t = S_0 \exp(\sigma W_t + \mu t)$ and $B_t = \exp(rt)$. Then all integrable claims X, knowable by some time horizon T, have associated replicating strategies (ϕ_t, ψ_t). In addition, the arbitrage price of such a claim X is given by

$$V_t = B_t \mathbb{E}_{\mathbb{Q}}(B_T^{-1} X \mid \mathcal{F}_t) = e^{-r(T-t)} \mathbb{E}_{\mathbb{Q}}(X | \mathcal{F}_t),$$

where $\mathbb{Q}$ is the martingale measure for the discounted stock $B_t^{-1} S_t$.

The important measure $\mathbb{Q}$ is not the measure which makes the stock a martingale, but the measure that makes the *discounted* stock a martingale. And the arbitrage price of the claim is the expectation under $\mathbb{Q}$ of the *discounted* claim.

So when interest rates are non-zero, what are the new rules? They are just discounted versions of the old rules:

Three steps to replication (discounted case)

(1) Find a measure $\mathbb{Q}$ under which the discounted stock price Z_t is a martingale.

(2) Form the process $E_t = \mathbb{E}_{\mathbb{Q}}(B_T^{-1} X \mid \mathcal{F}_t)$.

(3) Find a previsible process ϕ_t, such that $dE_t = \phi_t \, dZ_t$.

Call options

We should price something. Following Black and Scholes, we'll price a call option – the right but not the obligation to buy a unit of stock for a predetermined amount at a particular exercise date, say T. If we let this predetermined amount be k (in financial terms, the *strike* of the option), then in formal notation, our claim is $\max(S_T - k, 0)$. Or in more convenient notation, $(S_T - k)^+$.

First we should find V_0, the value of the replicating strategy (and thus the option) at time zero. Our formula tells us that this is given by

$$e^{-rT}\mathbb{E}_{\mathbb{Q}}\big((S_T - k)^+\big),$$

where $\mathbb{Q}$ is the martingale measure for $B_t^{-1}S_t$.

But how do we find this? The first thing to notice is the simplicity of the claim. The value $(S_T - k)^+$ only depends on the stock price at one point in time – namely the expiry time, T. So to find the expectation of this claim we need only find the marginal distribution of S_T under $\mathbb{Q}$.

And to do that, we can look at the process for S_t written in terms of the $\mathbb{Q}$-Brownian motion $\tilde{W}_t$. Since $d(\log S_t) = \sigma\, d\tilde{W}_t + (r - \frac{1}{2}\sigma^2)\, dt$, if we denote the stock price at time zero, S_0, by s, we have that $\log S_t = \log s + \sigma\tilde{W}_t + (r - \frac{1}{2}\sigma^2)t$, and thus $S_t = s\exp\big(\sigma\tilde{W}_t + (r - \frac{1}{2}\sigma^2)t\big)$.

So the marginal distribution for S_T is given by s times the exponential of a normal with mean $(r - \frac{1}{2}\sigma^2)T$ and variance $\sigma^2 T$. Thus if we let Z be a normal $N(-\frac{1}{2}\sigma^2 T, \sigma^2 T)$, we can write S_T as $se^{(Z+rT)}$ and thus the claim as the expectation $e^{-rT}\mathbb{E}\big((se^{(Z+rT)} - k)^+\big)$, which equals

$$\frac{1}{\sqrt{2\pi\sigma^2 T}}\int_{\log(k/s)-rT}^{\infty}(se^x - ke^{-rT})\exp\left(-\frac{(x+\frac{1}{2}\sigma^2 T)^2}{2\sigma^2 T}\right)dx.$$

This integral can be decomposed by a change of variables into a couple of standard cumulative normal integrals. If we use the notation $\Phi(x)$ to denote $(2\pi)^{-\frac{1}{2}}\int_{-\infty}^{x}\exp(-y^2/2)\,dy$, the probability that a normal $N(0,1)$ has value less than x, then we can calculate that $V_0 = V(s,T)$, where

Black–Scholes formula

$$V(s,T) = s\Phi\left(\frac{\log\frac{s}{k} + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right) - ke^{-rT}\Phi\left(\frac{\log\frac{s}{k} + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right).$$

This is the Black–Scholes formula for pricing European call options. (Put options, the right to sell a unit of stock for k, can be priced as a call less a forward – put-call parity.)

Exercise 3.14 Find the change of variable and thus prove the Black–Scholes formula.

3.8 Black–Scholes in action

If a stock has a constant volatility of 18% and constant drift of 8%, with continuously compounded interest rates constant at 6%, what is the value of an option to buy the stock for \$25 in two years time, given a current stock price of \$20?

The description fits the Black–Scholes conditions. Thus using $s = 20$, $k = 25$, $\sigma = 0.18$, $r = 0.06$, and $t = 2$, we can calculate V_0 as \$1.221.

Exercise 3.15 What information about the drift was required?

Price dependence

For values of the current stock price s much smaller than the exercise price k, the value of the formula itself gets small, signifying that the option is out of the money and unlikely to recover in time. Conversely, for values of s much greater than k, the option loses most of its optionality, and becomes a forward. Correspondingly the option price is approximately $s - ke^{-rT}$, which is the current value of a stock forward struck at price k for time T.

Time dependence

As the time to maturity T gets smaller, the chances of the price moving much more decreases and the option value gets closer and closer to the claim value taken at the current price, $(s - k)^+$.

For larger times, however, the option value gets larger. An option with almost infinite time to maturity would have value approaching s, as the cost now of price k is almost zero. It can be seen in figure 3.14 that as the time to expiration gets closer to zero, the curve gets closer to the option shape $(s - k)^+$.

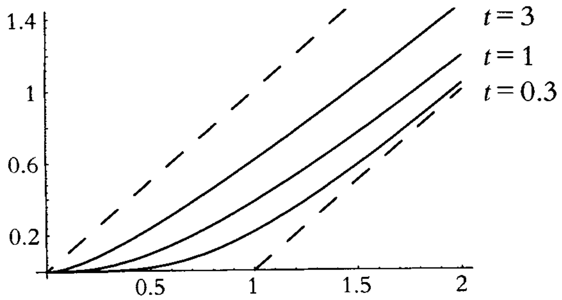


Figure 3.14 Option price against stock price for times 3, 1, and 0.3. Exercise price $k = \$1$, interest rate $r = 0$, volatility $\sigma = 1$.

Volatility dependence

All else being equal, the option is worth more the more volatile the stock is. At one extreme, if σ is very small, the option resembles a riskless bond and is just worth $(s - ke^{-rT})^+$, which is the value of the corresponding forward if the option will be in the money and is zero otherwise. At the other extreme, if σ is very large, the option is worth s.

American options

Sometimes an option has more optionality about it than just choosing between two alternatives at the maturity date. American options are the most well-known examples of such derivatives, giving the right to, say, purchase a unit of stock for a strike price k at any time up to and including the expiration date T, rather than only at that date. The buyer of the option then has to make decisions from moment to moment to decide when and if to call the option.

The buyer of an American call has the choice when to stop, and that choice can only use price information up to the present moment. Such a (random) time is called a *stopping time*. Following a strategy which will result in exercising the option at the stopping time τ, the corresponding payoff is

$$(S_\tau - k)^+ \qquad \text{at time } \tau.$$

If the option issuer knew in advance which stopping time the investor will use, the cost at time zero of hedging that payoff is

$$\mathbb{E}_{\mathbb{Q}}\big(e^{-r\tau}(S_\tau - k)^+\big).$$

As we do not know which τ will be used, we have to prepare for the worst possible case, and charge the maximum value (maximised over *all* possible stopping strategies),

$$V_0 = \sup_\tau \mathbb{E}_{\mathbb{Q}}\big(e^{-r\tau}(S_\tau - k)^+\big).$$

Pricing derivatives with optionality
In general, if the option purchaser has a set of options A, and receives a payoff X_a at time T, after choosing a in A, then the option issuer should charge

$$V_0 = \sup_{a \in A} \mathbb{E}_{\mathbb{Q}}\big(e^{-rT} X_a\big)$$

for it. If the purchaser does not exercise the option optimally, then the issuer's hedge will produce a surplus by date T.

That hedge in full

Returning to the original European option, one thing that would be useful to know would be the actual replicating strategy required, that is, to actually find out how much stock would be required at each point of time to artificially construct the derivative.

The amount of stock, ϕ_t, comes from the martingale representation theorem, but unfortunately, the theorem merely states that ϕ_t exists. However the martingale representation theorem, at heart, tells us that the reason that the discounted claim can be built from the discounted stock is that, being martingales under the same measure, one is locally just a scaled version of the other. The process ϕ_t is merely the ratio of volatilities. Thus, intuitively, if we looked at the ratio of the change in the value of the option caused by a move in the stock price and the change in the stock price used, this ought to be something like ϕ_t. And if we have a restricted enough claim where the only input required from the filtration for pricing the claim is the stock price at the current moment, and moreover that the functional relation implied by this between the value of the claim and the current stock price is smooth, then we could guess that the partial derivative of the option value with respect to the stock price is the ϕ_t we want.

And so it is. For the often-encountered case where the claim depends only on the terminal value, the option value *is* a well-behaved function of the current stock price. Suppose the derivative X is a function of the terminal value of the stock price, so that $X = f(S_T)$ for some function $f(s)$. Then the following is true.

Terminal value pricing

If the derivative X equals $f(S_T)$, for some f, then in the value of the derivative at time t is equal to $V_t = V(S_t, t)$, where $V(s,t)$ is given by the formula

$$V(s,t) = \exp\big(-r(T-t)\big)\mathbb{E}_{\mathbb{Q}}\big(f(S_T) \mid S_t = s\big)$$

And then the trading strategy is given by $\phi_t = \frac{\partial V}{\partial s}(S_t, t)$.

Why? Consider dV_t, the infinitesimal change in the value of the option. Remembering that $dS_t = \sigma S_t\, d\tilde{W}_t + rS_t\, dt$, then Itô gives us

$$dV_t = d(V(S_t,t)) = \left(\sigma S_t \frac{\partial V}{\partial s}\right) d\tilde{W}_t + \left(rS_t \frac{\partial V}{\partial s} + \tfrac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 V}{\partial s^2} + \frac{\partial V}{\partial t}\right) dt.$$

But we also know that $dV_t = \phi_t\, dS_t + \psi_t\, dB_t$, from the self-financing condition. And since $dB_t = rB_t\, dt$ we have

$$dV_t = (\sigma S_t \phi_t)\, d\tilde{W}_t + (rS_t\phi_t + r\psi_t B_t)\, dt.$$

But SDE representations are unique – so the volatility terms must match, giving $\phi_t = \frac{\partial V}{\partial s}$. The amount of stock in the replicating portfolio at any stage is the derivative of the option price with respect to the stock price.

Using this substitution for ϕ_t and the fact that $V_t = S_t\phi_t + \psi_t B_t$, we can also match the drift terms of the two SDEs to get a partial differential equation for V as

$$\tfrac{1}{2}\sigma^2 s^2 \frac{\partial^2 V}{\partial s^2} + rs\frac{\partial V}{\partial s} - rV + \frac{\partial V}{\partial t} = 0.$$

Notoriously, this PDE, coupled with the boundary condition that $V(s,T)$ must equal $f(s)$, gives another way of solving the pricing equation.

Explicit Black–Scholes hedge

The call option is a terminal value claim, as described earlier, and so we can find an expression for the hedge itself. The amount of stock held is the derivative of the value function with respect to stock price. In symbols

$$\phi_t = \frac{\partial V}{\partial s}(S_t, T-t) = \Phi\left(\frac{\log\frac{S_t}{k} + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}\right).$$

Because ϕ is always between zero and one, we need only ever have a bounded long position in the stock. Also the value of the bond holding at any time is

$$B_t\psi_t = -ke^{-r(T-t)}\Phi\left(\frac{\log\frac{S_t}{k} + (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}\right),$$

which, although always a borrowing, is bounded by the exercise price k.

There are two possibilities as the time approaches maturity. If the option is out of the money, that is the stock price is less than the exercise price, then both the bond and the stock holding go to zero, reflecting the increasing worthlessness of the option. Alternatively, if the price stays above the exercise value, then the stock holding grows to one unit and the value of the bond to $-k$. This combination exactly balances the now certain demand for a unit of stock in return for cash amount k.

Example – hedging in continuous time

This can be seen operating in practice. Below are two possible realisations of a stock price which starts at \$10. Both are exponential Brownian motions with volatility 20% and growth drift of 15%.

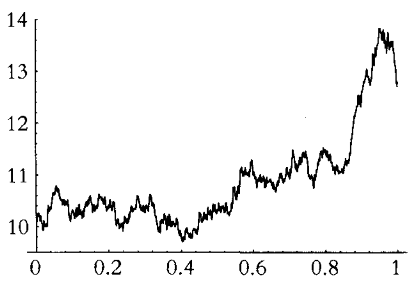


Figure 3.15a Stock price (A)

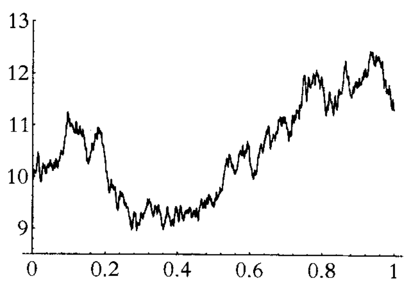


Figure 3.15b Stock price (B)

Let us price an option on this stock, to buy it at time $T = 1$ for the strike price of $k = \$12$, assuming interest rates are 5%. We can calculate both the evolving worth of the option V_t and the amount of stock to be held, ϕ_t, to hedge the contract.

In the case (A), these processes are shown in figure 3.16.

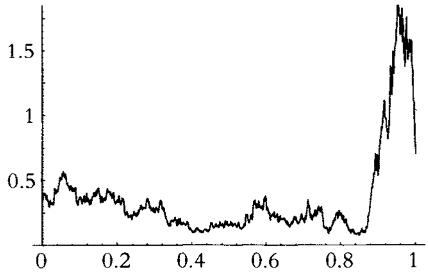


Figure 3.16a Option value (A)

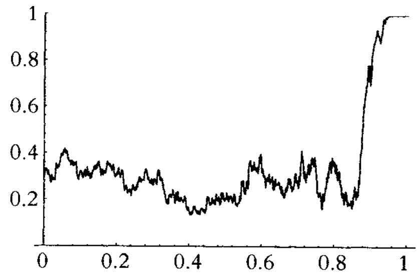


Figure 3.16b Stock hedge (A)

As time progresses, the option becomes in the money and the option value moves like the stock price. Also the hedge gets closer and closer to one, signifying that the option will be exercised.

In the case (B), these processes are shown in Figure 3.17.

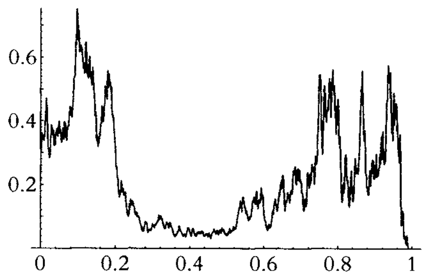


Figure 3.17a Option value (B)

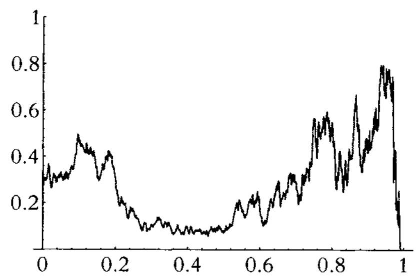


Figure 3.17b Stock hedge (B)

This time the option is not exercised and both the value of it and the hedge go to zero over time.

Exercise 3.16 A stock has current price \$10 and moves as an exponential Brownian motion with upward drift of 15% a year (continuously compounded) and volatility of 20% a year. Current interest rates are constant at 5%. What is the value of an option on the stock for \$12 in a year's time?

Exercise 3.17 For the same stock, what is the value of a derivative which pays off \$1 if the stock price is more than \$10 in a year's time?

Conclusions

Even with a respectable stochastic model for the stock, we can replicate any claim. Not something we had any right to expect. The replicating portfolio has a value given by the expected discounted claim, with respect to a measure which makes the discounted stock a martingale. Moreover, changing to the martingale measure has a remarkably simple effect on the process S_t – only the drift changes, to another constant value. The stock remains an exponential Brownian motion; even the volatility σ stays the same.

These three surprises conspire to make the result look easier to get at than perhaps it really is. Something subtle and beautiful really is going on under all the formalism and the result only serves to obscure it. Before we push on, stop and admire the view.