# FM436 Stochastic Calculus Tutorial Notes

Robert Rogers, October 2023

## Contents

# 1. Why do we need stochastic calculus?

In the first section of this course you've seen a number of equations that describe current-period variables in terms of next-period variables, as a result of no-arbitrage restrictions or of agents optimising over time. For example, you saw how the fundamental theorem of asset pricing implies $P_t = E_t^*(P_{t+1})$.

What happens if you make these time intervals between *t* and *t+1* smaller and smaller? As the size of the steps approaches 0, you end up with a **differential equation** – a description of the rate of the change of a variable, given current values. So, in the coming section of the course when we work with continuous time, we will need to work with differential equations.

But you'll see that something strange happens when we introduce randomness into a continuous time model. The path of our continuous variables get very "squiggly." So "squiggly," in fact, that some of the normal calculus techniques you learned in school no longer work, and we need to invent new approaches. This section of the course introduces these new calculus techniques.

The material starts by introducing continuous time in a deterministic framework to show how simple differential equations can be used to price assets. We then introduce the main building block we will use for continuous time stochastic processes – Brownian motion.

Once we understand Brownian motion, we introduce the techniques we need to work with it – probability measures, Itô integrals, stochastic differential equations, and Itô's Lemma. Finally we show how these techniques can be employed to price an asset in a continuous time framework.

If you are interested in a more thorough and in-depth treatment of these topics you can read:

- Mikosch, Elementary Stochastic Calculus with Finance in View, World Scientific
- Shreve, Stochastic Calculus for Finance, Springer

Another source that some students have found useful is the series "Topics In Mathematics With Applications In Finance" from MIT OpenCourseWare. Lectures 17, 18, and 21 cover similar ground to these lecture notes.

# 2. Discrete and continuous time asset pricing without risk

First let's think about what happens when we move to discrete to continuous time in a totally deterministic model, with a simple no arbitrage constraint.

## a. Discrete time:

**Discounting**

Suppose we have some risk free bond B that returns the known risk free rate each period:

$$\frac{B_{t+1}}{B_t} \equiv R^f_{t+1}$$

If we want to look at the risk free rate over multiple periods we can just define:

$$\frac{B_{t+s}}{B_t} = \prod_{i=1}^{s} R^f_{t+i} \equiv R^f_{t\to t+s}$$

**Asset pricing**

If we want to know the price of a non-dividend paying asset with payoff of price $P_T$ at time T, there are two ways to approach this. These will seem very similar for now, but the distinction will make more sense later on.

1) Direct approach

As noted in the first section of this course the fundamental theorem of asset pricing implies:

$$P_t = \frac{1}{R^f_{t\to T}} E^*_t(P_T) = (since\ there\ is\ no\ risk) \frac{P_T}{R^f_{t\to T}}$$

In other words, we just need to discount the payoff using the risk free rate to find the current price.

2) Difference equation

We could also look at how the asset's price evolves in each period. Since there is no risk, a no arbitrage constraints tell us that for all t:

$$\frac{P_{t+1}}{P_t} = \frac{B_{t+1}}{B_t} = R^f_{t+1}$$

This is a "difference equation." It tells us how $P$ evolves. To find the actual level of $P_t$ at any given time, we need to pin down some value. Since we know $P_T$ (we call this a "terminal condition), we can plug in this equation repeatedly, working backwards, to find $P_t$

$$P_t = \frac{P_T}{\prod_{i=t}^{T} R^f_i} = \frac{P_T}{R^f_{t\to T}}$$

Note that if we had instead been given an "initial condition" $P_0$, would find $P_t = P_0 R^f_{t\to t+s}$.

## b. Continuous time:

**Discounting:**

Now let's stop assuming that there are discrete units of time, and start allowing for time to take any value on the real number line.

As before, we can define the risk free rate as the return on a risk free bond over any interval:

$$\frac{B_{t+s}}{B_t} \equiv R^f_{t\to t+s}$$

In practice, we often want to look at the risk-free rate over the smallest possible intervals. So we define the instantaneous risk free rate as the instantaneous returns on the risk-free bond:

$$\lim_{s\to 0}\frac{(B_{t+s}-B_t)/s}{B_t} = \frac{dB_t/dt}{B_t} \equiv r_t$$

How can we write $R^f_{t\to t+s}$ in terms of r? It might not be immediately obvious we can find an analytic expression. If we just plug in and integrate we don't an expression that looks easy to work with:

$$R^f_{t\to t+s} = \frac{B_{t+s}}{B_t} = \frac{1}{B_t}\left(B_t + \int_t^{t+s}\frac{dB_\tau}{d\tau}d\tau\right) = ??$$

But we can use a little trick to make this easier – take the log first. This will come in useful repeatedly. Note that:

$$\frac{dB_t/dt}{B_t} = \frac{d\log B_t}{dt} = r_t$$

Now we can just integrate both sides to find:

$$\log\frac{B_{t+s}}{B_t} = \int_t^{t+s}\frac{d\log B_\tau}{d\tau}d\tau = \int_t^{t+s} r_\tau d\tau$$

Hence:

$$\frac{B_{t+s}}{B_t} = e^{\int_t^{t+s} r_\tau d\tau} = R^f_{t\to t+s}$$

Or, if $r_t$ is constant (as we will often assume):

$$R^f_{t\to t+s} = e^{rs}$$

We will use exponential discounting with the continuously compounded risk free rate often, so you should get comfortable with this derivation and approach.

**Asset pricing:**

As in discrete time, there are two ways to think about finding the price of an asset with payoff $P_T$ at time T.

1) Direct approach

$$P_t = \frac{1}{R^f_{t\to T}}E^*_t(P_T) = (since\ there\ is\ no\ risk)\frac{P_T}{R^f_{t\to T}} = P_T e^{-\int_t^T r_s ds}$$

In other words, we just need to discount the payoff using the risk-free rate to find the current price.

2) Differential equation

As with our difference equations in discrete time, we can also price P by looking at what how the price evolves at all the smallest possible intervals of time between t and T:

$$\frac{dP_t/dt}{P_t} = \frac{dB_t/dt}{B_t} = r_t$$

This is a differential equation. Just as for the difference equation in discrete time, it tells us how $P_t$ evolves. But to pin down the level of $P_t$, we also need another condition. Differential equations like this, generally have unique solutions for the path of P given initial or terminal condition. However, when things get more complicated, those solutions are rarely available in closed form, except for a few lucky types of equation.

We can of course find a closed form expression for $P_t$ here using the same approach as in the subsection on discounting above:

$$\log \frac{P_T}{P_t} = \int_t^T \frac{d \log P_s}{ds} ds = \int_t^T r_s ds$$

$$P_t = P_T e^{-\int_t^T r_s ds}$$

And if we had worked with an initial condition of $P_0$ instead of a terminal condition of $P_T$, we would instead have: $P_t = P_0 e^{\int_0^t r_s ds}$

# 3. Stochastic processes

## a. Discrete

A stochastic process is a collection of random variables indexed by time. Or it may often be easier to think of it as a probability distribution over a space of paths. You have already been working with stochastic processes in this course.

Examples:

- Binary white noise: $x_{t+1} = \begin{cases} 1 \; with \; prob \; ½ \\ -1 \; with \; prob \; ½ \end{cases}$
- Simple random walk: $x_{t+1} = \begin{cases} x_t + 1 \; with \; prob \; ½ \\ x_t - 1 \; with \; prob \; ½ \end{cases}$
- Price with lognormal return $P_{t+1} = P_t e^{z_{t+1}}$ where $z_{t+1}$ has a normal distribution

**Pricing a derivative:**

Suppose we there is a risky stock whose price is in each period is $S_t$.

We want to find the price at time t of a derivative that pays some function $f(S_T)$ at time T. We could again take two approaches:

1) Direct approach:

$$P_t = \frac{1}{R^f_{t\to T}} E^*_t(f(S_T)) = E_t(M_{t\to T} f(S_T))$$

The price of $P_t$ therefore depends on f function, but also on the distribution and variance of $S_T$, due to Jensen's inequality. With specific assumptions on the distribution of $M_{t\to T} f(S_T)$ (e.g. lognormality), we could calculate or numerically estimate the price at time t.

2) Difference equation:

In each period the derivatives price must follow:

$$P_t = \frac{1}{R^f_{t+1}} E^*_t(P_{t+1}) = E_t(M_{t+1} P_{t+1})$$

And in the final period: $P_T = f(S_T)$

This gives us a difference equation with a terminal condition. Therefore, similar to the approach in the deterministic section, we could work backwards step by step to find $P_t$, if we make some assumption on the distribution of $S_t$ in each period. This is, for example, how you learned to approach pricing in the binomial tree model.

## b. Continuous:

How can we define a "reasonable" continuous stochastic process non-degenerate processes? It's not straightforward.

Think about what it would "look like" if you tried to define binary random noise or simple random walk the exact same way as we did for discrete time:

$$x_{t+1} = \begin{cases} x_t + 1 \; with \; prob \; ½ \\ x_t - 1 \; with \; prob \; ½ \end{cases}$$

Or would it be possible to define the following?

$$\forall s > 0: \;\; x_{t+s} = \begin{cases} x_t + 1 \; with \; prob \; ½ \\ x_t - 1 \; with \; prob \; ½ \end{cases}$$

It's tricky to define a continuous stochastic process! In practice, we rely a lot on one well-understood process which we will describe in the next section: Brownian motion.

# 4. Brownian motion (aka Wiener process)

## a. Existence

It turns out that you can define a process that is continuous, and normally distributed over every possible interval. This is Brownian motion (sometimes also called a Wiener process).

I will call the following the "Brownian motion existence theorem" (but won't prove it):

There exists a probability distribution over the set of continuous functions $B: \mathbb{R} \to \mathbb{R}$ satisfying:

i. $B(0) = 0$
ii. (Normally distributed): For all $0 \le s < t$, the distribution of $B(t) - B(s)$ is the normal distribution with mean 0 and variance $t - s$
iii. (Independent increments): For any set of non-overlapping time intervals $[s_i, t_i]$ where $i = 1,2, \dots N$, the random variables $B(t_i) - B(s_i)$ are mutually independent.

We refer to a particular instance of a path chosen according to the Brownian motion as a sample Brownian path.

## b. Facts about Brownian motion

1. The values attained by a sample Brownian path (these I won't prove):
   a. Have a standard deviation of $\sqrt{t}$ – i.e. a sample path is more likely than not to be somewhere in the range $[-\sqrt{t}, \sqrt{t}]$
   b. Cross the t-axis infinitely often
   c. Eventually hit every real number
2. Brownian motion is the limit of a random walk as the number of steps approaches infinity
3. Brownian motion is continuous, but nowhere differentiable

## c. Brownian motion as the limit of a random walk

How do we get this marvellous process? One way to think about it is as the limit of a random walk as the size of each steps approaches 0.

Proof:

Consider a random walk that lasts for $n$ steps along a time axis of from 0,1, where each increment is either $+\frac{1}{\sqrt{n}}$ or $-\frac{1}{\sqrt{n}}$ with probability ½ . I.e:

$$Y_0 = 0$$

$$Y_{t+\frac{1}{n}} = \begin{cases} Y_t + \frac{1}{\sqrt{n}} & with\ prob\ ½ \\ Y_t - \frac{1}{\sqrt{n}} & with\ prob\ ½ \end{cases} \quad for\ t \in \left\{0, \frac{1}{n}, \frac{2}{n}, \dots \frac{n-1}{n}\right\}$$

Note that:

i. We have chosen each time increment to be $\frac{1}{n}$ so that the $n$ steps go from 0 to 1
ii. We have chosen the increments to be $\frac{1}{\sqrt{n}}$ so that the variance of each step is $\frac{1}{n}$, so the variance of $Y_t$ is constant no matter how many steps there are. $var(Y_t) = tn * \frac{1}{n} = t$. And $var(Y_t - Y_s) = t - s$ for $t, s \in \left\{0, \frac{1}{n}, \frac{2}{n}, \dots 1\right\}$

If we take the limit as $n \to \infty$, then we can apply the central limit theorem (CLT) to show this is normally distributed over all intervals (i.e. a Brownian motion from 0 to 1).

The CLT states for a random variable X with variance 1

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i \to^d N(0,1)$$

If we take any interval $\{t, s\}$ such that $t, s \in \left\{0, \frac{1}{n}, \frac{2}{n}, \dots 1\right\}$, then we have:

$$Y_t - Y_s = \sum_{i=1}^{n(t-s)} \left( Y_{s+\frac{i}{n}} - Y_{s+\frac{i-1}{n}} \right)$$

If we let $X_i = \sqrt{n} \left( Y_{s+\frac{i}{n}} - Y_{s+\frac{i-1}{n}} \right)$ and $N = n(t-s)$, then we can rewrite this:

$$Y_t - Y_s = \frac{1}{\sqrt{n}} \sum_{i=1}^{N} X_i = \sqrt{t-s} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i$$

Since $var(X_i) = 1$ (by construction), the CLT tells us:

$$Y_t - Y_s \to^d N(0, t-s)$$

Hence as $n \to \infty$, the random walk series $Y$ approaches a Brownian motion from 0 to 1.

## d. Non-differentiability

One striking fact about Brownian motion is that it is everywhere continuous but nowhere differentiable, with probability 1. Here I present an argument why this is the case (although not exactly a formal proof). This isn't a course in probability theory, so don't sweat the details of this too much – just try to get the intuition.

Think about the fact that the standard deviation of $B(t)$ is $\sqrt{t}$. For any probability p the "p percentile confidence interval" of where $B(t)$ lies will be proportional to $\sqrt{t}$ – i.e. will be $[-k\sqrt{t}, k\sqrt{t}]$ for some value of $k$. Draw what this looks like on a graph. What happens to its values as $t \to 0$? What happens to the slope of its boundaries as $t \to 0$?

To make this a bit more specific:

> Let $p$ be any probability in $(0,1)$
>
> Let $k$ be the inverse standard normal CDF of $\frac{p}{2}$: $k = \Phi^{-1}(\frac{p}{2})$
>
> Since $B(t) \sim N(0, t)$, we know that $B(t) \in (-k\sqrt{t}, k\sqrt{t})$ with probability $p$
>
> As $t \to 0$, our interval $(-k\sqrt{t}, k\sqrt{t})$ comes arbitrarily close to 0 for any probability p. So it makes sense that it converges
>
> But the *slope* of the boundaries of interval approaches infinity, i.e. $\lim_{t \to 0} \frac{d(k\sqrt{t})}{dt} = \infty$. So it makes sense that we can't differentiate the function – the possible realisations don't converge towards any slope as $t \to 0$.

Below I provide an informal proof of continuity:

> Given distance $\delta$, the probability $B(t) \in (-\delta, \delta)$ is $p_\delta(t) = 2\Phi\left(\frac{\delta}{\sqrt{t}}\right) - 1$

$\forall\delta, \lim_{t\to 0} p_\delta(t) = 2 - 1 = 1$

So for any tiny distance $\delta$, if I pick a small enough $t$, I can make the probability that $B(t)$ is less than $\delta$ away from $B(0)$ arbitrarily high. So $B$ is continuous at 0 with probability 1.

And an informal proof of non-differentiability:

Denote the slope of $B$ between 0 and t by $b(t)$: $b(t) \equiv \frac{B(t)}{t}$

Suppose $B$ is differentiable at $t = 0$, and $\frac{dB(0)}{dt} = G \geq 0$. This means $\lim_{t\to 0} b(t) = G$

For any $t > 0$, the probability that $b(t) < 0$ is $\frac{1}{2}$

For any $\epsilon > 0$ The probability $b(t) > 2G$ is $p_{2G}(t) = 1 - \Phi(2G\sqrt{t})$

$\forall G > 0$: $\lim_{t\to 0} p_{2G}(t) = 1 - 0.5 = 0.5$

Thus the probability $P\big(b(t) \in (0,2G)\big) \to 0$ as $t \to 0$

So G cannot be the derivative B at 0, because for any unit time $\epsilon > 0$, we can find a point $b(t)$ with within that time (i.e. $t < \epsilon$), such that $b(t)$ lies outside of $(0,2G)$ with arbitrarily high probability.

## e. Brownian motion with drift

We'll often want to consider processes that are functions of a Brownian motion, rather than just the Brownian motion itself. E.g. a Brownian motion with a drift is a Brownian motion, plus some cumulative function of time:

$$X_t = X_0 + \int_0^t \mu_s\, ds + \sigma B_t$$

Where $\mu_s$ is the drift for time s. If the drift is constant, then we have the simpler expression:

$$X_t = X_0 + \mu t + \sigma B_t$$

$\sigma$ is a constant volatility (aka "diffusion") term.

## f. Geometric Brownian motion

In finance we often work with a process called "geometric Brownian motion" (as opposed to "arithmetic" Brownian motion), which has instantaneous *returns* that follow a Brownian motion. A geometric Brownian motion process $S$, with a constant drift can be simply defined as:

$$S_t = S_0 e^{Mt + \sigma B_t}$$

where $M, \sigma$ are constants. A GBM is lognormally distributed over all intervals, and is always greater than 0, as long as $S_0 > 0$.

# 5. Probability measures in continuous time

## a. The risk neutral measure

Earlier in the course, you saw that no arbitrage implied that you could set up a special "probability measure" under which all assets return the risk-free rate. You rewrote the SDF times the physical probability as a new "risk neutral probability."

It turns out you can do the exact same thing in continuous time. In general, if any two mappings of states to probabilities are "equivalent" – meaning they share an assessment of what has 0 and 1 probability – then you can always change measure between them.

Since we've defined Brownian motion as having 0 mean, we write that a Brownian motion is associated with a specific "probability measure." E.g. if we say $B_t$ is Brownian motion under the physical probability measure and $B_t^Q$ is Brownian motion under the risk neutral measure Q, that means that $E_t(B_s) = 0, E_t^Q(B_s) \neq 0, E_t\left(B_s^Q\right) \neq 0, E_t^Q\left(B_s^Q\right) = 0$.[1]

## b. Key facts about changing measure with Brownian motions

Just as in discrete time, it can be useful to "change measure" from the physical probabilities to the risk neutral probabilities when dealing with Brownian motions. There are three important things for you to know about changing probability measure with Brownian motions:

1. When you change measure, you add a drift term.
    a. If you compare two Brownian motions under measures Q and P: $B_t^Q = B_t^P + \int_0^t \eta_s ds$, where $\eta_s$ is the new drift applied at time s. If $\eta_s$ is constant we can write this in a simpler way: $B_t^Q = B_t + \eta t$
    b. This is similar to what you learned earlier in discrete time: $E\left(R_{i,t+1}\right) \neq E^*\left(R_{i,t+1}\right)$
2. When you change measure, the variance does not change
    a. $var_s\left(B_t^Q\right) = var_s(B_t^P) = t - s$
    b. So if we consider a process $X_t = \mu t + \sigma B_t$, under a different measure $\mu$ will be different but $\sigma$ (the "diffusion" term) will be the same
    c. The variance of a Brownian motion is known with certainty from observing it for any time period, so it does not change under different probability measures.
    d. The continuous, "fractal" nature of Brownian motion means even if I gave you a microsecond of data, you could chop it up into an infinite number of periods and observe the variance. You will see this in more detail in subsection on "quadratic variation" later.
    e. This is different from what you learned in the discrete time section. It was not the case in general that $var_t\left(R_{i,t+1}\right) = var_t^*(R_{i,t+1})$
3. Any two Brownian motions with drift are equivalent ("Girsanov's Theorem")
    a. This means that for any process we can always change to a risk neutral measure under which the expected instantaneous returns are the risk free rate.
    b. More generally, for any drift $\eta_s$ that we want to generate, there must exist a probability measure Q such that : $B_t^Q = B_t^P + \int_0^t \eta_s ds$ (as long as $\eta_s$ is not a function of future realisations of $B_t^Q$)

---

[1] The risk neutral measure is typically denoted by a Q or a * superscript

c. For example, if we have a process $X_t = X_0 + \mu t + \sigma B_t$, and we want to change to a measure in which X has 0 drift, we could simply state that there must be a measure Q such that $B_t^Q = B_t - \frac{\mu}{\sigma} t$. Then $X_t = X_0 + \sigma B_t^Q$

For this course you will mainly just use the physical measure and the risk neutral measure. And what you need to know is that you can switch between them. When you switch to the risk neutral measure, the Brownian motion picks up a drift that's equal to whatever is needed to make the process you are interested have an expected return of the instantaneous risk free rate: $r_t$. The diffusion term does not change.

There can also be times where it can be useful to use to other probability measures, but you will probably not need to do so in this course. For example, if you a pricing an option on the relative price of two assets, it may be useful to use a measure under which the drift of their relative prices is 0.

# 6. Itô integrals and stochastic differential equations

We have shown that Brownian motion is not differentiable. This seems like a problem for us, because in section 2, we also saw that continuous time asset pricing uses differential equations (e.g. we used the statement $\frac{dP_t/dt}{P_t} = r_t$ for our deterministic model). We'd like to be able to get a differential equation for $f(B_t)$, but we can't write $\frac{df(B_t)}{dt} = f'(B_t)\frac{dB_t}{dt}$.

So how can we rescue something that looks like differential equations, but with Brownian motion?

Consider that when we use differential equations, we're often not really interested in the point in time values of $\frac{dP_t}{dt}$ or $\frac{df(P_t)}{dt}$. We're really interested in finding some level (e.g. $P_t$), and we use integrals over the differential equations to do so. E.g. we used the fact that $\int_0^T \frac{d\log P_t}{dt}dt = \log P_T - \log P_t$ to find $P_t$ in part 2.

So maybe even if the derivative $\frac{dB_t}{dt}$ doesn't exist, we can rescue something that does the job that we would want the integral $\int_0^T \frac{dB_t}{dt}dt$ or $\int_0^T \frac{df(B_t)}{dt}dt$ to do, and still use some sort of differential equation.

## a. The Itô integral

The usual Reimann integral that we learn for calculus can be defined as:

$$\int_0^T f(t)dt = \lim_{n\to\infty}\sum_{i=1}^{n} f\left(\frac{i-1}{n}T\right)\left(\frac{T}{n}\right)$$

I.e. split up $T$ into $n$ different increments, and evaluate $f\left(\frac{i-1}{n}T\right)$ times the increment length.

The fundamental theorem of calculus tells us that for a differentiable function F:

$$F(X) - F(0) = \int_0^X F'(x)dx$$

We can create a similar statement using something that we'll call the "Itô integral", which allows us to work with a non-differentiable function $B(t)$.

**Definition: Itô integral**

$$\int_0^T f(t)dB(t) = \lim_{n\to\infty}\sum_{i=1}^{n} f\left(\frac{i-1}{n}T\right)\left(B\left(\frac{i}{n}T\right) - B\left(\frac{i-1}{n}T\right)\right)$$

I.e. split up *B(T)* into $n$ different increments, and evaluate $f\left(\frac{i-1}{n}T\right)$ times each increment length.

Clearly, a similar statement to the fundamental theorem of calculus holds:

$$B(T) - B(0) = \int_0^T dB(t)$$

Because all we've done is divided $B$ into little increments and then summed those up together again.

## b. Stochastic differential equations

Now we can write things that look like differential equations, but use non-differentiable functions.

As a simple example, let's try to write a differential equation that describes the price of a stock $S(t)$ whose instantaneous returns follow a Brownian motion at time $t$ (ignoring any drift term for now).[2] I.e. changes in $S(t)$ are equal to $S(t)$ times changes in the Brownian motion times some volatility term $\sigma$. I would be tempted to write:

$$\frac{dS(t)}{dt} = S(t)\sigma\frac{dB(t)}{dt}$$

But I cannot use this differential equation because, because $\frac{dB(t)}{dt}$ does not exist.

However, I can represent this relationship using Itô integrals. When I write that the instantaneous returns follow a Brownian motion, I'm saying that for very small increments (i.e. very large number of increments $n$):

$$\frac{S\left(\frac{i}{n}T\right) - S\left(\frac{i-1}{n}T\right)}{S\left(\frac{i-1}{n}T\right)} \approx \sigma\left(B\left(\frac{i}{n}T\right) - B\left(\frac{i-1}{n}T\right)\right)$$

If we add a summation around this expression, and take the limit as $n \to \infty$ we have:

$$\lim_{n\to\infty}\sum_{i=1}^{n} S\left(\frac{i}{n}T\right) - S\left(\frac{i-1}{n}T\right) = \lim_{n\to\infty}\sum_{i=1}^{n} S\left(\frac{i-1}{n}T\right)\sigma\left(B\left(\frac{i}{n}T\right) - B\left(\frac{i-1}{n}T\right)\right)$$

This matches our definition of the Itô integral. We can therefore write:

$$S(\mathrm{T}) - S(0) = \int_0^{\mathrm{T}} dS(t) = \int_0^{\mathrm{T}} S(t)\sigma dB(t)$$

If this integral relationship holds over any arbitrary interval, then it is often convenient to just drop the integrals from the notation. So we just write:

$$dS(t) = S(t)\sigma dB(t)$$

This is a **stochastic differential equation (SDE)**. It is defined to mean that the Itô integral of the left and right hand sides are equal over all possible intervals.

## c. Working with stochastic differential equations:

Two important and simple facts will help you work with stochastic differential equations

1) Itô integrals can be combined with Reimann integrals, and thus SDEs can include $dt$

E.g. let's add a drift to our Geometric Brownian Motion and say that the stock returns follow a Brownian motion, plus a constant mean that we'll call a "drift".

Now we're saying that each little increment of $S(t)$ is equal to $S(t)$ times a little increment of $B(t)$ (scaled by $\sigma$) plus a little increment of time (scaled by $\mu$). In terms of Itô integrals we can write this as:

$$S(T_2) - S(T_1) = \int_{T_1}^{T_2} dS(t) = \int_{T_1}^{T_2} S(t)\sigma dB(t) + \int_{T_1}^{T_2} S(t)\mu dt$$

---

[2] This is Geometric Brownian motion

over any arbitrary interval $[T_1, T_2]$. The last integral is an ordinary Riemann integral. We can then write this as an SDE:

$$dS(t) = S(t)\sigma dB(t) + S(t)\mu dt$$

2) SDEs can be manipulated stochastic differential equations as normal, although you cannot divide or multiply by the infinitesimal $d$ parts.

E.g. we could write the equation above as:

$$\frac{dS(t)}{S(t)} = \sigma dB(t) + \mu dt$$

$$\frac{dS(t)}{S(t)} - \sigma dB(t) - \mu dt = 0$$

Etc…

But you cannot write:

$$\frac{dS(t)}{dt} - \sigma \frac{dB(t)}{dt} = \mu$$

## d. SDEs and martingales

A process $X$ is said to be a "martingale" if $E_t(X_s) = X_t$ for all $s \geq t$. I.e. the best predictor of future value is current value.

When is an Itô integral a martingale? If you look at the summation notation you'll see we're just summing up a lot of independent normal variables. Which means that unless $f$ is a function of future occurrences of $B_t$, then you'll have a martingale.

We call a function $f$ that is only a function of past and contemporaneous events (not future events) an "adapted process" in SDE jargon. In general, most processes we deal with in finance will be "adapted processes" because e.g. a trading strategy or a derivative price today can't be based on where stocks will go in the future!

**Fact:** If $f$ is an adapted process, then $\int_0^T f(s, B_s)dB_s$ is a martingale. I.e.:

$$E_t\left(\int_t^T f(s, B_s)dB_s\right) = 0$$

Or in SDE notation we can write: $E_t(f(s, B_s)dB_s) = 0 \quad for\ any\ s \geq t$

# 7. Itô's Lemma

## a. Why do we need Itô's Lemma?

For normal calculus, one very important tool we use a lot is the chain rule:

$$\frac{df(g(x))}{dx} = f'(g(x))g'(x)$$

We will need a version of this for SDEs:

$$df(B(t)) = ?$$

For example, in section 2 we used a trick to find the bond price implied given by the differential equation: $\frac{dP(t)/dt}{P(t)} = r_t$. We differentiated $\log P(t)$ and used $\frac{d \log P(t)}{dt} = \frac{dP(t)/dt}{P(t)}$.

To find the stock price implied by the geometric Brownian motion SDE: $\frac{dS(t)}{S_t} = \mu t + \sigma dB(t)$, we might like to try the same trick. But what is $d \log S(t)$?

Your first guess might be to apply to apply the chain rule as normal and state:

$$df(B(t)) = f'(B(t))dB(t)$$

$$d \log B(t) = \frac{dB(t)}{B(t)}$$

But it turns out this is incorrect. It is incorrect because of special property of Brownian motion: "quadratic variation of t".

## b. Quadratic variation

**Definition**

Let's define an object called "Quadratic variation".

For any function $f(t)$, the quadratic variation from 0 to T is:

$$\lim_{n\to\infty} \sum_{i=1}^{n} \left( f\left(\frac{i}{n}T\right) - f\left(\frac{i-1}{n}T\right)\right)^2$$

In other words: find the changes in the function over a large number of tiny increments between 0 and T. The quadratic variation is the sum of the squares of these changes.

In Itô integral notation we could write this object as:

$$\int_0^T (df(t))^2$$

**Quadratic variation of differentiable functions:**

It should not be surprising that for any differentiable function, the quadratic variation is 0. The square of a tiny number is much smaller than the number itself. And simply adding up all the changes in f gets you $f(T) - f(0)$, which is finite. So it makes sense that the sum of the squared increments approaches 0.

In in Itô integral notation, this means: $\int_0^T (df(t))^2 = 0$ for any differentiable function f

In SDE notation, this can be written: $(df(t))^2 = 0$

**Proof:**

We can rewrite the changes $f\left(\frac{i}{n}T\right) - f\left(\frac{i-1}{n}T\right)$ as the time increment $\frac{T}{N}$ multiplied by the average slope of $f$ from $\frac{i-1}{n}T$ to $\frac{i}{n}T$:

$$f\left(\frac{i}{n}T\right) - f\left(\frac{i-1}{n}T\right) = f^{slope}(i)\frac{T}{N} \ where \ f^{slope}(i) = \left(f\left(\frac{i}{n}T\right) - f\left(\frac{i-1}{n}T\right)\right)\frac{N}{T}$$

Let $f'_{max}(T)$ bet the highest absolute value of derivative attained by f along $[0, T]$.
$f'_{max}(T) \equiv \sup|f'(t)|$ for $t \in [0, T]$

For all $i \in \{1,2, \dots n\}$, the$f^{slope}(i)$ must be less in absolute value than $f'_{max}(T)$ by the mean value theorem. I.e. $f^{slope}(i)^2 \leq f'_{max}(T)^2$

Hence $\left(f\left(\frac{i}{n}T\right) - f\left(\frac{i-1}{n}T\right)\right)^2 \leq f'_{max}(T)^2\left(\frac{T}{n}\right)^2$

$\lim_{n\to\infty} \sum_{i=1}^{n}\left(\frac{T}{n}\right)^2 = 0$, so:

$$\lim_{n\to\infty} \sum_{i=1}^{n}\left(f\left(\frac{i}{n}T\right) - f\left(\frac{i-1}{n}T\right)\right)^2 \leq f'_{max}(T)^2 \lim_{n\to\infty}\sum_{i=1}^{n}\left(\frac{T}{n}\right)^2 = 0$$

**Quadratic variation of Brownian motion:**

An important result is that the quadratic variation of Brownian motion is T:

$$\forall T > 0, \lim_{n\to\infty}\sum_{i=1}^{n}\left(B\left(\frac{i}{n}T\right) - B\left(\frac{i-1}{n}T\right)\right)^2 = T$$

As an Itô integral this says: $\int dB_t^2 = T$

And as an SDE: $dB_t^2 = dt$

And in words: if you sum the squares of all the infinitesimal changes in $B_t$, you get the changes in t.

**Proof:**

Simply apply the (strong) law of large numbers.

$n\left(B\left(\frac{i}{n}T\right) - B\left(\frac{i-1}{n}T\right)\right)^2$ is a random variable with mean $T$, because $\left(\frac{i}{n}T\right) - B\left(\frac{i-1}{n}T\right) \sim N\left(0, \frac{T}{n}\right)$

So by LLN:

$$\frac{1}{n}\sum_{i=1}^{n} n\left(B\left(\frac{i}{n}T\right) - B\left(\frac{i-1}{n}T\right)\right)^2 \to^p T$$

If you go through the same logic with the “covariation” of two correlated Brownian motions B and Z, you will find that their quadratic covariation is $\rho T$. As an SDE, we can write this fact as: $dB_t dZ_t = \rho dt$.

If we calculate the quadratic covariation of a Brownian motion and a differentiable function, we will find it is 0. We can write this fact as: $dB_t dt = 0$.

Finally, if you calculate the cubic, quartic, etc variation of a Brownian motion, you will also find it is 0 (i.e. $dB_t^k = 0 \ for \ k > 2$). The appendix contains more detail on these calculations.

So, if we collect all the facts we learned about quadratic variation (in SDE notation), we have:

$$dB_t^2 = dt$$
$$dB_t dt = 0$$
$$dt^2 = 0$$
$$dB_t dZ_t = \rho dt$$
$$dB_t^k = 0 \; for \; k > 2$$

## c. Demonstrating Itô's lemma

Consider a differentiable function $f(t)$.

If we want to know the change in $f(t)$ over some small increment $\Delta f(t)$, we can use a Taylor expansion:

$$\Delta f(t) = f'(t)\Delta t + \frac{1}{2} f''(t)\Delta t^2 + \frac{1}{6} f'''(t)\Delta t^3 + \cdots$$

So if we want to evaluate the sum of an infinite number of small changes in $f(t)$:

$$\int_0^T df(t)$$

We can just plug in the Taylor expansion into the definition of the integral to get:

$$\int_0^T df\big(g(t)\big) = \int_0^T f'(t)dt + \int_0^T \frac{1}{2} f''(t)dt^2 + \int_0^T \frac{1}{6} f'''(t)dt^3 \ldots$$

Because the quadratic variation of t is 0 (and higher power variations are also 0), then the terms associated with $dt^2, dt^3$, etc… are 0 we are left (unsurprisingly) with:

$$\int_0^T f'(t)dt$$

If you're not sure what's going on here, it may help to write out the steps above in summation form:

$$\lim_{n\to\infty} \sum_{i=1}^{n} \left( f\left(\frac{i}{n}T\right) - f\left(\frac{i-1}{n}T\right) \right) = \lim_{n\to\infty} \sum_{i=1}^{n} \left( f'\left(\frac{i-1}{n}T\right)\frac{T}{n} + \frac{1}{2} f''\left(\frac{i-1}{n}T\right)\left(\frac{T}{n}\right)^2 + \cdots \right)$$
$$= \lim_{n\to\infty} \sum_{i=1}^{n} \left( f'\left(\frac{i-1}{n}T\right)\frac{T}{n} \right)$$

Where the last step is because $t$ has 0 quadratic or higher order variation.

However, if we replace $t$ with a Brownian motion $B(t)$, we will get a different result, because of its non-0 quadratic variation.

The Taylor expansion is now:

$$\int_0^T df\big(B(t)\big) = \int_0^T f'\big(B(t)\big)dB(t) + \int_0^T \frac{1}{2} f''\big(B(t)\big)dB(t)^2 + \cdots$$

Since $B(t)$ has quadratic variation of t, we have:

$$\int_0^T \frac{1}{2} f''\big(B(t)\big) dB(t)^2 = \frac{1}{2} f''\big(B(t)\big) t$$

And so the Taylor expansion is equal to :

$$\int_0^T df\big(B(t)\big) = \int_0^T f'\big(B(t)\big) dB(t) + \frac{1}{2} f''\big(B(t)\big) t$$

Or in SDE terms:

$$df\big(B(t)\big) = f'\big(B(t)\big) dB(t) + \frac{1}{2} f''\big(B(t)\big) dt$$

This equation is Itô's Lemma – the key result in stochastic calculus.

## d. Itô's Lemma

To make this statement slightly more general and useful, we will consider:

- A process $X$ that has some drift, rather than just a pure Brownian motion (e.g. a stock price)
- A function $f$ that is a function of time and the value of $X$ (e.g. an option price)

**Itô's lemma:**

Let $f(t, X_t)$ be a twice differentiable function of two variables and $dX_t = \mu_t dt + \sigma_t dB_t$ be a stochastic process with Brownian motion $B$. Then:

$$df(t, X_t) = \left( \frac{\partial f}{\partial t} + \mu_t \frac{\partial f}{\partial X} + \frac{1}{2} \frac{\partial^2 f}{\partial X^2} \sigma_t^2 \right) dt + \frac{\partial f}{\partial X} \sigma_t dB_t$$

Note: I have switched derivative notation from $f'$ to $\frac{\partial f}{\partial t}$ for clarity, since I now have two partial derivatives of $f$.

To prove derive this version, simply differentiate $f$ as in the previous section:

$$df(t, X_t) = \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial X} dX_t + \frac{1}{2} \frac{\partial^2 f}{\partial X^2} dX_t^2$$

And then plug in for $dX_t^2$ using our quadratic variation results:

$$dX_t^2 = \mu_t^2 dt^2 + 2\mu_t \sigma_t dt dB_t + \sigma_t^2 dB_t^2 = \sigma_t^2 dt$$

## e. Itô's Lemma with two Brownian motions

We can use our quadratic variation result:

$$dB_t dZ_t = \rho dt$$

to find Itô's Lemma with two correlated Brownian motions.

If:

$$dX_t = \mu_t dt + \sigma_t dB_{1,t}$$

$$dY_t = \alpha_t dt + \beta_t dB_{2,t}$$

and $f(t, X_t, Y_t)$ is a twice differentiable function. Then:

$$df(t, X_t, Y_t) = \left( \frac{\partial f}{\partial t} + \frac{\partial f}{\partial X}\mu_t + \frac{\partial f}{\partial Y}\alpha_t + \frac{1}{2}\frac{\partial^2 f}{\partial X^2}\sigma_t^2 + \frac{1}{2}\frac{\partial^2 f}{\partial Y^2}\beta_t^2 + \frac{\partial^2 f}{\partial Y \partial X}\sigma_t \beta_t \rho \right) dt + \frac{\partial f}{\partial X}\sigma dB_{1,t} + \frac{\partial f}{\partial Y}\beta dB_{2,t}$$

## f. Itô Isometry

Another useful implication quadratic variation is a property called "Itô Isometry", which states that for any "adapted" function $f$ (see definition in section 6.d):

$$var_0\left(\int_0^T f(t, B_t) dB_t\right) = E_0\left(\int_0^T f(t, B_t)^2 dt\right)$$

This is useful for calculating the variance of stochastic processes.

To see why this is, first note that because $E_0\left(\int_0^T f(t, B_t) dB_t\right) = 0$ (again, see section 6.d):

$$var_0\left(\int_0^T f(t, B_t) dB_t\right) = E_0\left(\left(\int_0^T f(t, B_t) dB_t\right)^2\right)$$

Then note that all of the increments of $B_t$ are independent, by the definition of Brownian motion. So when we square the integral, all of the products of the $B_t$ increments from different time periods will have 0 expectation. So we'll just be left with:

$$E_0\left(\int_0^T f(t, B_t)^2 dB_t^2\right) = E_0\left(\int_0^T f(t, B_t)^2 dt\right)$$

If that doesn't make sense to you, you can think of what's going on using summation notation. Independent increments of Brownian motion implies:

$$E_0\left(\left(B\left(\frac{i}{n}T\right) - B\left(\frac{i-1}{n}T\right)\right)\left(B\left(\frac{j}{n}T\right) - B\left(\frac{j-1}{n}T\right)\right)\right) = 0 \quad for\ i \neq j$$

So:

$$E_0\left(\left(\int_0^T f(t, B_t) dB_t\right)^2\right) = E_0\left(\left(\lim_{n\to\infty}\sum_{i=1}^n f\left(\frac{i-1}{n}T\right)\left(B\left(\frac{i}{n}T\right) - B\left(\frac{i-1}{n}T\right)\right)\right)^2\right)$$

$$= E_0\left(\lim_{n\to\infty}\sum_{i=1}^n f\left(\frac{i-1}{n}T\right)^2\left(B\left(\frac{i}{n}T\right) - B\left(\frac{i-1}{n}T\right)\right)^2\right) = \int_0^T f(t, B_t)^2 dt$$

# 8. Pricing an asset

Now we have all the tools we need to tackle pricing problems using the two different methods.

Suppose we have some derivative that we know at time T will pay off some function of a stock price $F(S_T)$. If the stock price follows some Itô process:

$$dS_t = \mu(t, S_t)dt + \sigma(t, S_t)dB_t$$

(e.g. if $\sigma(t, S_t) = \sigma S_t$ and $\mu(t, S_t) = \mu S_t$ then we have geometric Brownian motion with drift)

How does the price of the derivative evolve over time? Let's assume the derivative price at time $t$ is a function of $t$ and the stock price and call it $f(t, S_t)$.

We have two possible approaches here – directly calculate the risk neutral expectation of $F(S_T)$, or solve for the price using differential equations. There is a theorem (which we will not cover) called the "Feynman-Kac theorem" that guarantees that these two approaches will always give the same answer.

## a. Direct approach (aka "Martingale method")

Just as in discrete time, one path to the solution is to use the Fundamental Theorem of Asset Pricing, which tells us that the price is be the risk-neutral expectation of its discounted payoff.

So under the risk neutral measure Q:

$$f(t, S_t) = E_t^Q\left(\frac{1}{R_{f,t\to T}} F(S_T)\right)$$

If the risk free rate is constant (as we will often assume), then we can write this:

$$f(t, S_t) = e^{-r(T-t)} E_t^Q\big(F(S_T)\big)$$

We can then find $f(t, S_t)$ by just computing this expectation using the risk neutral distribution of $S_T$, numerically or analytically. E.g. if $S_t$ is a geometric Brownian motion, then $S_T$ is lognormally distributed over discrete intervals.

This sounds simple, but it can be challenging to find $E_t^Q\big(F(S_T)\big)$ if $S_T$ does not have a simple closed form solution or $F(S_T)$ is not a standard distribution like lognormal.

Note: In general, this is called the "martingale method". It is so called because it uses the fact that the asset price divided by the price of the risk free asset (i.e. multiplied by the risk free rate) is a martingale under the risk neutral measure:

$$E_t\left(R_{f,v\to T} f(v, S_v)\right) = E_t\left(E_v\big(f(v, S_v)\big)\right) = R_{f,t\to T} f(t, S_t)$$

## b. Differential equations

**Deriving the SDE for the asset**

Just as in the deterministic continuous time section, another approach is to find a differential equation that defines the path of prices.

Suppose again we have a derivative that pays off $F(S_T)$ at time T, and

$$dS_t = \mu(t, S_t)dt + \sigma(t, S_t)dB_t$$

And we want to find the derivative price $f(t, S_t)$

By applying Itô's lemma, we can find the SDE for describing the evolution of $f(t, S_t)$:

$$df(t, S_t) = \left(\frac{\partial f}{\partial t} + \mu(t, S_t)\frac{\partial f}{\partial S} + \frac{1}{2}\sigma(t, S_t)^2\frac{\partial^2 f}{\partial S^2}\right)dt + \sigma(t, S_t)\frac{\partial f}{\partial S}dB_t$$

**Determining the pricing PDE**

In the deterministic model, all assets must have a return of $r_t$ at all times, so we could find the price by just imposing $df(t, S_t) = f(t, S_t)r_t dt$. With stochastic returns things are not so simple because there can be a risk premium. So how do we come up with an equivalent equation?

One way is to use the risk neutral measure. Under the risk neutral measure, all assets have an expected return of $r_t$. So we rewrite the SDE for the stock under the risk free measure as:

$$dS_t = S_t r_t dt + \sigma(t, S_t)dB_t^Q$$

And, since the asset must also have risk neutral expected returns of $r_t$, we have:

$$E_t^Q\big(df(t, S_t)\big) = \left(\frac{\partial f}{\partial t} + S_t r_t\frac{\partial f}{\partial S} + \frac{1}{2}\sigma(t, S_t)^2\frac{\partial^2 f}{\partial S^2}\right)dt = r_t f(t, S_t)dt$$

Or, dropping the $dt$:

$$r_t f(t, S_t) = \frac{\partial f}{\partial t} + S_t r_t\frac{\partial f}{\partial S} + \frac{1}{2}\sigma(t, S_t)^2\frac{\partial^2 f}{\partial S^2}$$

We call this the "pricing partial differential equation (PDE)". It is called a PDE, not an SDE because we have removed the stochastic element ($dB_t^Q$). This equation can be solved numerically or in some cases analytically to find the price.

Note that this is the continuous time equivalent of the binomial option pricing model you learned. If you reread that section of the lecture notes you will clearly see the parallels.

**Determining the pricing PDE by replicating portfolios**

A second way to arrive at the same pricing PDE, is through a no arbitrage logic.

If we can construct a portfolio composed of the derivative and the stock that gives us a risk-free return, then the return on the portfolio must be the risk free rate.

From the SDE for *f*, we can see that the derivative's diffusion term – i.e. its exposure to the stock's Brownian motion – is $\sigma(t, S_t)\frac{\partial f}{\partial S}$, whereas the stock's is $\sigma(t, S_t)$. So if we hold 1 unit of the derivative and go short $\frac{\partial f}{\partial S}$ units of the stock we will have no exposure to the Brownian motion and returns will be entirely deterministic and risk free.

Let's call this hedged portfolio $\theta_t = f(t, S_t) - \frac{\partial f}{\partial S}S_t$. We can easily find the differential equation that defines the path of its values:

$$d\theta_t = df(t, S_t) - \frac{\partial f}{\partial S}dS_t = \left(\frac{\partial f}{\partial t} + \frac{1}{2}\sigma(t, S_t)^2\frac{\partial^2 f}{\partial S^2}\right)dt$$

Since the hedged portfolio has a risk free return, by no arbitrage it must return the risk free rate!

$$\frac{d\theta_t}{\theta_t} = r_t dt$$

So plugging in and rearranging gives us the same pricing PDE that we derived using the risk neutral measure:

$$\left(\frac{\partial f}{\partial t}+\frac{1}{2}\sigma(t,S_t)^2\frac{\partial^2 f}{\partial S^2}\right)dt=\left(f(t,S_t)-\frac{\partial f}{\partial S}S_t\right)r_t dt$$

$$r_t f(t,S_t)=\frac{\partial f}{\partial t}+S_t r_t\frac{\partial f}{\partial S}+\frac{1}{2}\sigma(t,S_t)^2\frac{\partial^2 f}{\partial S^2}$$

And a terminal condition $f(T,S_T)=F(S_t)$

**Black Scholes PDE**

Often, we will not allow $r_t$ and $\sigma$ to both be time varying in this course, because it makes finding solutions very difficult. If $\sigma(t,S_t)=S_t\sigma$, and $r_t=r$, then this is known as the Black Scholes PDE:

$$rf(t,S_t)=\frac{\partial f}{\partial t}+S_t r\frac{\partial f}{\partial S}+\frac{1}{2}S_t^2\sigma^2\frac{\partial^2 f}{\partial S^2}$$

## c. Solving PDEs and SDEs

Differential equations generally have unique solutions for a given initial condition, as long as the drift and diffusions are "reasonable."[3] This is why they turn out to be quite useful! However, these solutions are rarely available in closed form.

If a closed form solution exists, there is no specific process you can always follow to find it. You may need to conjecture a form for the solution, and then differentiate it to find the value of the coefficients.

At times you may also be able to simplify the problem by adding additional conditions based on the nature of the problem: e.g. maybe you can figure out what the price should be as the underlying asset price approaches 0 or infinity, or if the object is an infinitely lived option, perhaps you could impose that its price should not be a function of time (so $\frac{\partial f}{\partial t}=0$).

You have already solved one SDE in the exercises: the Brownian motion with constant drift ($\frac{dS_t}{St}=\mu dt+\sigma dBt$). A more difficult example employing the method of conjecturing a solution is provided in the appendix.

In the large majority of cases in actual practice there are no closed form solutions. Instead, numerical methods are used. Typically these are:

- Finite differences: Take your starting point and take small steps using first order Taylor approximation (this can only be applied to PDEs with no stochastic component)
- Monte Carlo: As above, but draw small independent random numbers at each step to simulate your $B_t$ increments
- Tree methods: Using the fact that a Brownian motion is the limit of a random walk, split the space from 0 to T into a number of small steps and find all the possible values at time T of this random walk and the probability of each.

If you're interested in more detail on how these work you can check, for example, "Options, futures, and other derivatives" by John Hull.

[3] Formally, SDEs must satisfy some regularity conditions, including a space-variable Lipshitz condition and a spatial growth condition, but these are not very interesting for us.

# 9. Wrapping up

We've covered a lot of content this week. You will not need to know every detail and proof in these notes for the exam. If you can understand and remember the following 10 key pieces of information, you should do well in the last few weeks of FM436:

1. Brownian motion is normally distributed over all intervals, and each increment is independent.
2. Brownian motion is the limit of a random walk
3. Brownian motion is not differentiable
4. You can switch between the physical and risk neutral measure, and the only thing that changes is the drift term
5. What stochastic differential equations look like and mean
6. If a SDE has no drift, then the process is a martingale (and vice versa)
7. $dB_t^2 = dt$ and $dB_t dt = 0$ and $dB_t dZ_t = \rho dt$
8. Itô's lemma
9. To find the price a European derivative you can either take the expected discounted payoffs under the risk free measure (the "martingale method") or derive a PDE for the prices
10. PDEs and SDEs generally have solutions but usually not analytical ones.

# Appendix: Probability space vocabulary terminology

If you read any textbooks or articles that cover this material, you are likely to encounter some of the basic probability theory terminology described here.

A proof or example often starts with the sentence: “Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration $\mathbb{F}$”

$\omega$ is a **state** – e.g. heads for a coin flip, or a sample Brownian path for a Brownian motion.

$\Omega$, the **sample space**, denotes the set of all possible states, e.g. $\{\omega_1, \omega_2, \omega_3, \omega_4\}$

An **event** is a subset of states: $A \subseteq \Omega$ – e.g. you have at least two heads, or the sample path crosses above 2 before time T.

$\mathcal{F}$, the **sigma-algebra**, is a collection of events such that: 1. It includes $\Omega$ itself, and 2. It is closed to unions and complements. Intuitively $\mathcal{F}$, describes how “refined” our information is. E.g. $\{\{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}, \Omega, \emptyset\}$. For example,f we play a game where someone flips two coins and tells me how many were heads or tails, the sigma algebra would be {{HH}, {TT}, {HT, TH}}

$\mathbb{F}$, the **filtration**, is a sequence of sigma-algebras $(\mathcal{F})_{t \geq 0}$, that describes the evolution of our information over time. E.g. if I repeat the coin flip game described in the sigma-algebra section twice, the filtration is first: {{HH}, {TT}, {HT, TH}}, and second: all the permutations of possible results from the first round and from the second round.

$\mathbb{P}$, the **probability measure**, assigns probabilities to events in $\mathcal{F}$ in a consistent way. It is a mapping $\mathbb{P}: \mathcal{F} \to [0,1]$ such that:

- $\mathbb{P}(\Omega) = 1$
- $\mathbb{P}(\emptyset) = 1$, and for $A \in \mathcal{F}, \mathbb{P}(A) \geq 0$
- For disjoint events $A_i \in \mathcal{F}$:

$$\mathbb{P}(\cup A_i) = \sum_i \mathbb{P}(A_i)$$

# Appendix: Quadratic covariation

## a. Quadratic covariation of two Brownian motions:

Suppose we have two correlated Brownian motions B and Z. By the same logic as quadratic variation we can define quadratic covariation as

$$\lim_{n\to\infty}\sum_{i=1}^{n}\left(B\left(\frac{i}{n}T\right)-B\left(\frac{i-1}{n}T\right)\right)\left(Z\left(\frac{i}{n}T\right)-Z\left(\frac{i-1}{n}T\right)\right)$$

It's straightforward to show by the law of large numbers this is equal to $\rho T$

We could write this fact as: $dB_t dZ_t = \rho dt$. I.e. if you sum up all the little changes in $B_t$ multiplied by the little changes in $Z_t$, you get the correlation times the changes in t.

## b. Quadratic covariation of Brownian motion with a differentiable function:

Note also that for any differentiable function of time $f(t)$, the quadratic covariation with a Brownian motion will be 0 – you can verify this using the same logic we used to show the quadratic variation of a differentiable function is 0. If we consider the function $f(t) = t$, this gives us the important fact: $dB_t dt = 0$ and $dt^2 = 0$. I.e. if you sum up all the little changes in $B_t$ multiplied by t, you get 0.

## c. Higher order variation of a Brownian motion:

In general, for a standard normal $X$ with mean 0 and variance $\sigma^2$:

$$E(X^k) = \begin{cases} 0 \; if \; k \; is \; odd \\ (k-1)!!\,\sigma^k \; if \; k \; is \; even \end{cases}$$

So

$$E\left(\left(B\left(\frac{i}{n}T\right)-B\left(\frac{i-1}{n}T\right)\right)^k\right) = \begin{cases} 0 \; if \; k \; is \; odd \\ (k-1)!!\left(\frac{T}{n}\right)^k \; if \; k \; is \; even \end{cases}$$

$$\lim_{n\to\infty}\sum_{i=1}^{n}\left(B\left(\frac{i}{n}T\right)-B\left(\frac{i-1}{n}T\right)\right)^k \to^p \lim_{n\to\infty} nE\left(\left(B\left(\frac{i}{n}T\right)-B\left(\frac{i-1}{n}T\right)\right)^k\right) = 0 \; for \; any \; k > 2$$

# Appendix: Example of solving a SDE

The "Ornstein-Uhlenbeck" process is defined by the SDE:

$$dX_t = -\alpha X_t dt + \sigma B_t; \quad X_0 = x_0$$

This process is frequently used to describe random processes that revert to a mean ($\alpha$). E.g. the risk free short term interest rate is often modelled with this process. Can you see based on the SDE why it would show mean reversion?

If we want to solve for $X_t$ we'll need to conjecture a form. Because of the mean-reverting property, we might consider that the solution involves some decaying weighted average of past increments of $B_t$. I.e. we want old increments of $B_t$ to "matter less" as time passes and the process reverts to the mean. So we could conjecture:

$X_t = f(t, B_t) = a(t)\left(x_0 + \int_0^t b(s)dB_s\right)$ for some functions $a$ and $b$ of $t$, with $a(0)=1$

Don't worry: you aren't expected to be able to come up with this conjecture on your own! Now to see if this conjecture works for some $a$ and $b$, we differentiate $X$:

$$\frac{df(t, B_t)}{dt} = \frac{a'(t)}{a(t)} f(t, B_t)$$

$$\frac{df(t, B_t)}{dB_t} = a(t)b(t)dB_t$$

$$\frac{df^2(t, B_t)}{dB_t^2} = 0$$

So applying Itô's lemma:

$$dX_t = \frac{a'(t)}{a(t)} X_t dt + a(t)b(t)dB_t$$

Now we can match our coefficients. The process was defined by:

$$dX_t = -\alpha X_t dt + \sigma B_t$$

And our conjectured form gives us:

$$dX_t = \frac{a'(t)}{a(t)} X_t dt + a(t)b(t)dB_t$$

So if the conjecture is right, then we would need:

$$\frac{a'(t)}{a(t)} = -\alpha$$

$$a(t)b(t) = \sigma$$

This is satisfied if:

$$a(t) = e^{-\alpha t}$$

$$b(t) = \sigma e^{\alpha t}$$

So our conjecture works. Thus the solution to the SDE is given by:

$$X_t = f(t, B_t) = e^{-\alpha t}\left(x_0 + \sigma \int_0^t e^{\alpha s} dB_s\right)$$