Elementary Stochastic Calculus

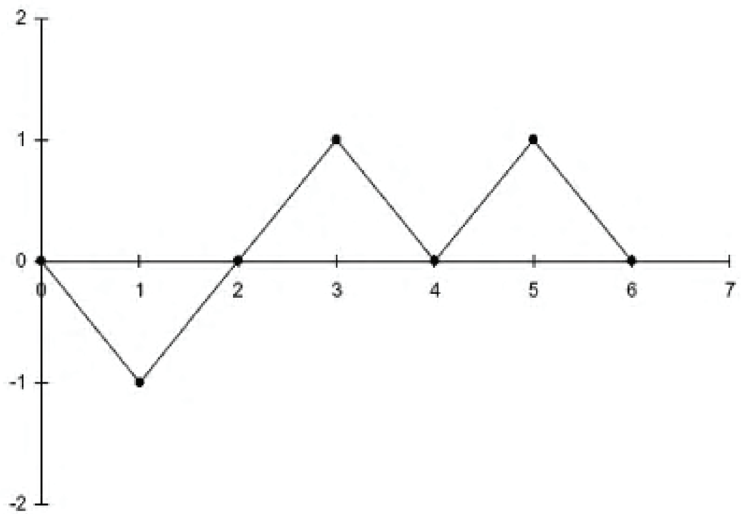
- Construction of Brownian Motion/Wiener Process
- Functions of a stochastic variable and Itô's Lemma
- Stochastic Integration
- The Itô Integral
- Examples of popular Stochastic Differential Equations

Introduction

Asset price random walks for stocks, indices, interest rates, etc. can be modelled using *Stochastic Differential Equations*. A stochastic differential equation is essentially a differential equation driven (or *perturbed*) by random noise. Most SDEs do not have analytical solutions and their simple looking form can be dangerously deceiving. Apart from a few special cases most equations have to be solved numerically (using simulations).

Construction of Brownian Motion

Brownian Motion can be constructed by careful scaling of a simple symmetric random walk. Consider the coin tossing experiment



where we define the random variable

$$R_i = \begin{cases} +1 & \text{if } H \\ -1 & \text{if } T \end{cases}$$

and examine the statistical properties of R_i.

Firstly the mean

$$\mathbb{E}\left[R_i\right] = (+1)\frac{1}{2} + (-1)\frac{1}{2} = 0$$

and secondly the variance

$$\begin{aligned} \mathbb{V}\left[R_i\right] &= \mathbb{E}\left[R_i^2\right] - \underbrace{\mathbb{E}^2\left[R_i\right]}_{=0} \\ &= \mathbb{E}\left[R_i^2\right] = 1 \end{aligned}$$

Suppose we now wish to keep a score of our winnings after the n^{th} toss - we introduce a new random variable

$$X_n = \sum_{i=1}^{n} R_i$$

This allows us to keep a track of our total winnings. This represents the position of a marker that starts off at the origin (no winnings).

So starting with no money means

$$X_0 = R_0$$

Now we can calculate expectations of X_n

$$\begin{aligned}
\mathbb{E}[X_n] &= \mathbb{E}\left[\sum_{i=1}^{n} R_i\right] \\
&= \sum_{i=1}^{n} \mathbb{E}[R_i] = 0
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}\left[X_n^2\right] = \\
&\qquad \mathbb{E}\left[R_1^2 + R_2^2 + \ldots . R_n^2 + 2R_1R_2 + \ldots\ldots + 2R_{n-1}R_n\right] \\
&\quad = \mathbb{E}\left[\sum_{i=1}^{n} R_i^2\right] + 2\mathbb{E}\left[\sum_{\substack{i=1 \\ j\neq i}}^{n} R_{ij}\right] = n \cdot 1 + 2 \cdot 0 \\
&\quad = n
\end{aligned}$$

Now look at the *quadratic variation* of the random walk.

After each toss, we have won or lost \$1. That is

$$X_n - X_{n-1} = \pm 1 \Longrightarrow |X_n - X_{n-1}| = 1$$

Hence

$$\sum_{i=1}^{n} \underbrace{(X_n - X_{n-1})^2}_{=1} = n$$

Let's now extend this by introducing time dependence. Perform six tosses of a coin in a time t. So each toss must be performed in time $t/6$, and a bet size of $\sqrt{t/6}$ (and not \$1), i.e. we win or lose $\sqrt{t/6}$ depending on the outcome.

Let's examine the quadratic variation for this experiment

$$
\begin{aligned}
& \sum_{i=1}^{6} (X_n - X_{n-1})^2 \\
= & \sum_{i=1}^{6} \left(\pm\sqrt{t/6} \right)^2 \\
= & \; 6 \times \frac{t}{6} = t
\end{aligned}
$$

Now speed up the game. So we perform n tosses within time t with each bet being $\sqrt{t/n}$. Time for each toss is t/n.

$$X_n - X_{n-1} = \pm\sqrt{t/n}$$

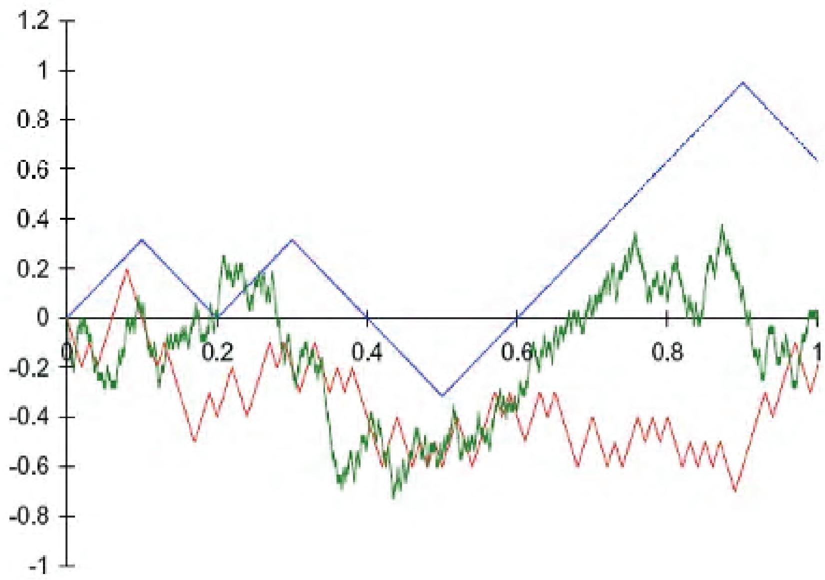
The quadratic variation is

$$
\begin{aligned}
\sum_{i=1}^{n} (X_n - X_{n-1})^2 &= n \times \left(\pm\sqrt{t/n} \right)^2 \\
&= t
\end{aligned}
$$

As n becomes larger and larger, time between subsequent tosses decreases and the bet sizes become smaller. The time and bet size decrease in turn like

$$\begin{aligned}\text{time decrease} &\sim O\left(\frac{1}{n}\right) \\ \text{bet size} &\sim O\left(\frac{1}{\sqrt{n}}\right)\end{aligned}$$

The diagram below shows a series of coin tossing experiments.



The scaling we have used has been chosen carefully to both keep the random walk finite and also not becoming zero. i.e. In the limit $n \longrightarrow \infty$, the random walk stays finite. It has an expectation conditional on a starting value of zero, of

$$\begin{aligned}
\mathbb{E}\left[X_t\right] &= \mathbb{E}\left[\lim_{n\to\infty}\sum_{i=1}^{n} R_i\right] \\
&= \lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{E}\left[R_i\right] \\
&= 0
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[X_t^2\right] &= \mathbb{E}\left[\lim_{n\to\infty}\sum_{i=1}^{n} R_i^2\right] \\
&= \lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{E}\left[R_i^2\right] \\
&= t
\end{aligned}$$

This limiting process as dt tends to zero is called Brownian Motion.

Properties of a Wiener Process

A stochastic process $\{W(t) : t \in \mathbb{R}_+\}$ is defined to be Brownian motion (or a Wiener process) if

- $W(0) = 0$ (with probability one)

- Continuity - paths of $W(t)$ are continuous (no jumps). Differentiable nowhere.

- for each $t > 0$ and $s > 0$, $W(t) - W(s)$ is normal with mean 0 and variance $|t - s|$,

  i.e. $(W(t) - W(s)) \sim N(0, |t - s|)$. Coin tosses are Binomial, but due to a large number and the C.L.T we have a distribution that is normal. That is $W(t) - W(s)$ has a pdf given by

$$p(x) = \frac{1}{\sqrt{2\pi(t-s)}} \exp\left(-\frac{x^2}{2(t-s)}\right)$$

So Brownian motion has independent Gaussian increments.

- $W(t+s) - W(t)$ is independent of $W(t)$.
  This means $dW_1 = W(t_1) - W(t_0)$ is independent of $dW_2 = W(t_2) - W(t_1)$, is independent of $dW_3 = W(t_3) - W(t_2)$, ..., $dW_n = W(t_n) - W(t_{n-1})$.

Also called *standard Brownian motion*.

If we want to be a little more pedantic then we can write some of the properties above as

$$W_t \sim N^{\mathbb{P}}(0, t)$$

i.e. W_t is normally distributed under the probability measure $\mathbb{P}$.

The *covariance function* for a Brownian motion can be calculated as follows. If $t > s$,

$$\begin{aligned}
\mathbb{E}\left[W_t W_s\right] &= \mathbb{E}\left[\left(W_t - W_s\right) W_s + W_s^2\right] \\
&= \underbrace{\mathbb{E}\left[W_t - W_s\right]}_{N(0,|t-s|)} \mathbb{E}\left[W_s\right] + \mathbb{E}\left[W_s^2\right] \\
&= (0) \cdot 0 + \mathbb{E}\left[W_s^2\right] \\
&= s
\end{aligned}$$

The first term on the second line follows from independence of increments. Similarly, if $s > t$; then $\mathbb{E}\left[W_t W_s\right] = t$ and it follows that

$$\mathbb{E}\left[W_t W_s\right] = \min\left\{t, s\right\}.$$

Brownian motion is a *martingale*.

A stochastic process M_t is called a $\mathbb{P}-$martingale if for $t < T$

$$\mathbb{E}_t^{\mathbb{P}}\left[M_T\right] = M_t$$

That is, it's a conditional expectation and we write formally as

$$\mathbb{E}_t^{\mathbb{P}}\left[M_T \middle| \mathcal{F}_t\right] = M_t;\ t < T$$

$\mathcal{F}_t$ here is an <u>information set</u> called (using probability terminology) a *filtration*.

For a Brownian motion, again where $t < T$

$$\begin{aligned}\mathbb{E}_t^{\mathbb{P}}\left[W_T\right] &= \mathbb{E}_t^{\mathbb{P}}\left[W_T - W_t + W_t\right] \\ &= \underbrace{\mathbb{E}_t^{\mathbb{P}}\left[W_T - W_t\right]}_{N(0,|T-t|)} + \mathbb{E}_t^{\mathbb{P}}\left[W_t\right]\end{aligned}$$

The next step is important one - and requires a little subtlety

The first term is zero. We are taking expectations at time $t-$ hence W_t is known, i.e. $\mathbb{E}_t^{\mathbb{P}}\left[W_t\right] = W_t$. So

$$\mathbb{E}_t^{\mathbb{P}}\left[W_T\right] = W_t.$$

Another important property of Brownian motion is that of a *Markov process*. That is if you observe the path of the B.M from 0 to t and want to estimate W_T where $T > t$ then the only relevant information for predicting future dynamics is the value of W_t. That is, the past history is fully reflected in the present value. So the conditional distribution of W_t given up to $t < T$ depends only on what we know at t (latest information).

Mean Square Convergence

Consider a function $F(X)$. If

$$\mathbb{E}\left[(F(X) - l)^2\right] \longrightarrow 0$$

then we say that $F(X) = l$ in the *mean square limit*, also called *mean square convergence*. We present a full derivation of the mean square limit. Starting with the quantity:

$$\mathbb{E}\left[\left(\sum_{j=1}^{n}\big(X(t_j) - X(t_{j-1})\big)^2 - t\right)^2\right]$$

where $t_j = \dfrac{jt}{n} = j\Delta t$.

Hence we are saying that *up to mean square convergence,*

$$dX^2 = dt.$$

This is the symbolic way of writing this property of a Wiener process, as the partitions Δt become smaller and smaller.

Developing the terms inside the expectation

First, we will simplify the notation in order to deal more easily with the outer (rightmost) squaring. Let $Y(t_j) = \left(X(t_j) - X(t_{j-1})\right)^2$, then we can rewrite the expectation as:

$$\mathbb{E}\left[\left(\sum_{j=1}^{n} Y(t_j) - t\right)^2\right]$$

Expanding we have:

$$\mathbb{E}\left[(Y(t_1) + Y(t_2) + \ldots + Y(t_n) - t) \times (Y(t_1) + Y(t_2) + \ldots + Y(t_n) - t)\right]$$

The term inside the Expectation is equal to

$$
\begin{aligned}
&Y(t_1)^2 + Y(t_1)Y(t_2) + \ldots + Y(t_1)Y(t_n) - Y(t_1)t \\
&+Y(t_2)^2 + Y(t_2)Y(t_1) + \ldots + Y(t_2)Y(t_n) - Y(t_2)t \\
&\vdots \\
&+Y(t_n)^2 + Y(t_n)Y(t_1) + \ldots + Y(t_n)Y(t_{n-1}) - Y(t_n)t \\
&-tY(t_1) - tY(t_2) - \ldots - tY(t_n) + t^2
\end{aligned}
$$

Rearranging

$$
\begin{aligned}
&Y(t_1)^2 + Y(t_2)^2 + \ldots + Y(t_n)^2 \\
&2Y(t_1)Y(t_2) + 2Y(t_1)Y(t_3) + \ldots + 2Y(t_{n-1})Y(t_n) \\
&-2Y(t_1)t - 2Y(t_2)t - \ldots - 2Y(t_n)t \\
&+t^2
\end{aligned}
$$

We can now factorize to get

$$
\sum_{j=1}^{n} Y(t_j)^2 + 2\sum_{i=1}^{n}\sum_{j<i} Y(t_i)Y(t_j) - 2t\sum_{j=1}^{n} Y(t_j) + t^2
$$

Substituting back $Y(t_j) = \Big(X(t_j) - X(t_{j-1})\Big)^2$ and taking the expectation, we arrive at:

$$\begin{aligned}
&\mathbb{E}\,[\,\sum_{j=1}^{n} \Big(X(t_j) - X(t_{j-1})\Big)^4 \\
&+2\sum_{i=1}^{n}\sum_{j<i} (X(t_i) - X(t_{i-1}))^2 \Big(X(t_j) - X(t_{j-1})\Big)^2 \\
&-2t\sum_{j=1}^{n} \Big(X(t_j) - X(t_{j-1})\Big)^2 \\
&+t^2\,]
\end{aligned}$$

Computing the expectation

By linearity of the expectation operator, we can write the previous expression as:

$$
\begin{aligned}
&\sum_{j=1}^{n} \mathbb{E}\left[\left(X(t_j) - X(t_{j-1})\right)^4\right] \\
&+2\sum_{i=1}^{n}\sum_{j<i} \mathbb{E}\left[(X(t_i) - X(t_{i-1}))^2 \left(X(t_j) - X(t_{j-1})\right)^2\right] \\
&-2t\sum_{j=1}^{n} \mathbb{E}\left[\left(X(t_j) - X(t_{j-1})\right)^2\right] \\
&+t^2
\end{aligned}
$$

Now, since $Z(t_j) = X(t_j) - X(t_{j-1})$ follows a Normal distribution with mean 0 and variance $\frac{t}{n}$ $(= dt)$, it follows (standard result) that its fourth moment is equal to $3\frac{t^2}{n^2}$. We will show this shortly.

Firstly we know that $Z(t_j) \sim N\left(0, \frac{t}{n}\right)$, i.e.

$$\mathbb{E}\left[Z(t_j)\right] = 0, \quad \mathbb{V}\left[Z(t_j)\right] = \frac{t}{n}$$

therefore we can construct its PDF. For any random variable $\psi \sim N\left(\mu, \sigma^2\right)$ its probability density is given by

$$p(\psi) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(\psi-\mu)^2}{\sigma^2}\right)$$

hence for $Z(t_j)$ the PDF is

$$p(z) = \frac{1}{\sqrt{t/n}\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{z^2}{t/n}\right)$$

$$\begin{aligned} \mathbb{E}\left[\left(X(t_j) - X(t_{j-1})\right)^4\right] &= \mathbb{E}\left[Z^4\right] \\ &= 3\frac{t^2}{n^2} \qquad \text{for } j = 1, \ldots, n \end{aligned}$$

So

$$\begin{aligned}\mathbb{E}\left[Z^4\right] &= \int_{\mathbb{R}} Z^4 p(z)\,dz \\ &= \sqrt{\frac{n}{2t\pi}} \int_{\mathbb{R}} Z^4 \exp\left(-\frac{1}{2}\frac{z^2}{t/n}\right) dz\end{aligned}$$

now put

$$u = \frac{z}{\sqrt{t/n}} \longrightarrow du = \sqrt{n/t}dz$$

Our integral becomes

$$\sqrt{\frac{n}{2t\pi}} \int_{\mathbb{R}} \left(\sqrt{\frac{t}{n}}u\right)^4 \exp\left(-\tfrac{1}{2}u^2\right) \sqrt{\frac{t}{n}}du$$

$$
\begin{aligned}
&= \sqrt{\frac{1}{2\pi}}\frac{t^2}{n^2}\int_{\mathbb{R}} u^4 \exp\left(-\tfrac{1}{2}u^2\right) du \\
&= \frac{t^2}{n^2}.\sqrt{\frac{1}{2\pi}}\int_{\mathbb{R}} u^4 \exp\left(-\tfrac{1}{2}u^2\right) du \\
&= \frac{t^2}{n^2}.\mathbb{E}\left[u^4\right].
\end{aligned}
$$

So the problem reduces to finding the fourth moment of a standard normal random variable. Here we do not have to explicitly calculate any integral. Two ways to do this.

Either use the MGF as we did earlier and obtained the fourth moment to be three.

Or the other method is to make use of the fact that the kurtosis of the standardised normal distribution is 3.

That is

$$\mathbb{E}\left[\frac{(\phi-\mu)^4}{\sigma^4}\right] = \mathbb{E}\left[\frac{(\phi-0)^4}{1^4}\right] = 3.$$

Hence $\mathbb{E}\left[u^4\right] = 3$ and we can finally write $3\dfrac{t^2}{n^2}$.

and

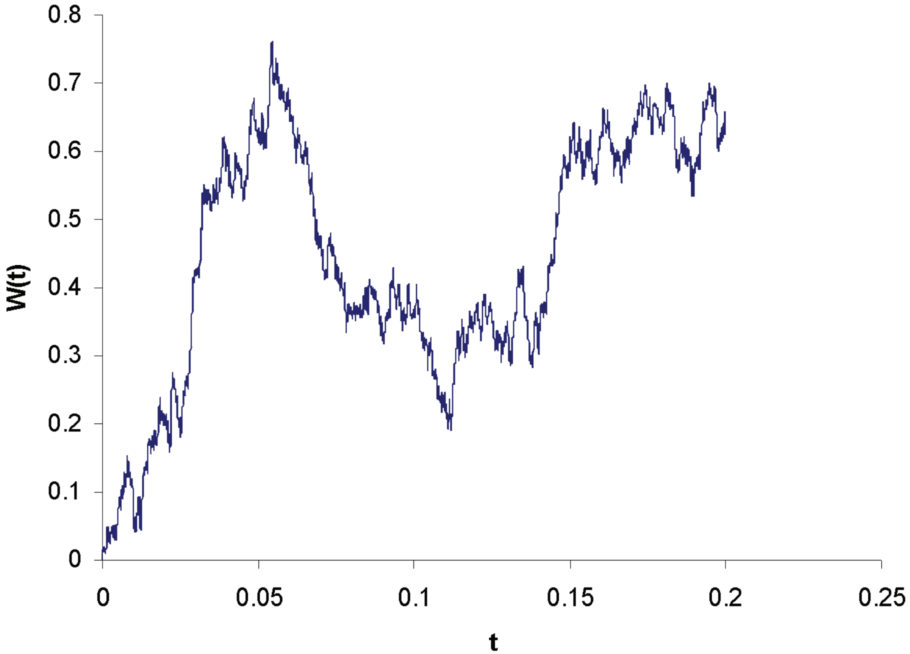
$$\mathbb{E}\left[\Big(X(t_j)-X(t_{j-1})\Big)^2\right] = \frac{t}{n} \qquad \text{for } j = 1, \ldots, n$$

Because of the single summation, the fourth moment and the variance multiplied by t actually recur n times. Because of the double summation, the product of variances occurs $\frac{n(n-1)}{2}$ times.

We can now conclude that the expectation is equal to:

$$
\begin{aligned}
& 3n\frac{t^2}{n^2} + n(n-1)\frac{t^2}{n^2} - 2tn\frac{t}{n} + t^2 \\
= \; & 3\frac{t^2}{n} + t^2 - \frac{t^2}{n} - 2t^2 + t^2 = 2\frac{t^2}{n} \\
= \; & O(\frac{1}{n})
\end{aligned}
$$

So, as our partition becomes finer and finer and n tends to infinity, the quadratic variation will tend to t *in the mean square limit*.



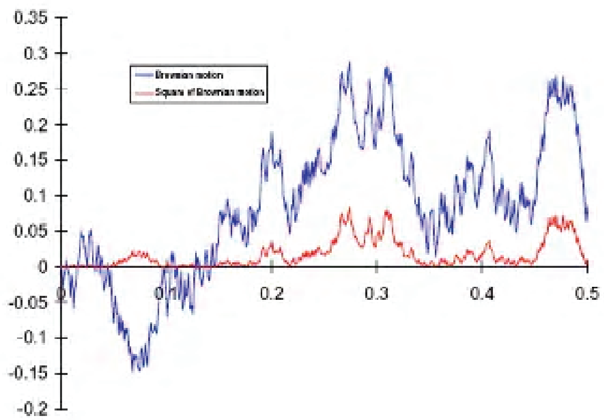
The diagram above represents a realisation of a Wiener process, with $\Delta t = 0.0001$.

Numerical Scheme:

$$
\begin{aligned}
\text{Start} &: \quad t_0,\ W_0 = 0;\ \text{define } \Delta t = T/n \\
\text{loop } i &= 1, 2, \cdots, n : \\
t_i &= t_{i-1} + \Delta t \\
\text{draw } \phi &\sim N(0,1) \\
W_i &= W_{i-1} + \phi\sqrt{\Delta t}
\end{aligned}
$$

Functions of stochastic variables and Itô's lemma

Now we will see the idea of a function of a stochastic variable. Below is shown a realization of a Brownian motion $X(t)$ and the function $F(X) = X^2$.



Whenever we have functions of a variable it is natural to want to know how to differentiate and manipulate these functions.

What are the rules of calculus when variables are stochastic?

The first point to note is that in the stochastic world we really have two 'variables.'

These are time t and the Brownian motion X.

We are used to writing ordinary and partial differential equations in the form

$$\frac{dF}{d\cdot}$$

or

$$\frac{\partial F}{\partial \cdot}$$

where the quantities on the bottom are the independent variables.

So might expect something similar in the stochastic world.

We immediately hit a problem, however.

Because dX is of size $\sqrt{dt}$ it is much bigger than dt.

This means that we have to be careful whenever we think about gradients/slopes/derivatives/sensitivities, since these are limits as dt goes to zero.

For this reason, in the stochastic world we instead work with *stochastic differential equations*.

These take the form

$$dF = \cdots dt + \cdots dX.$$

So, what are the rules of calculus?

Since X is stochastic, so is F, and we can ask 'what is the stochastic differential equation for F?'

If $F(X) = X^2$ what is the equation for dF?

If $F = X^2$ is it true that $dF = 2X\ dX$?

No.

- The ordinary rules of calculus do not generally hold in a stochastic environment.

Then what are the rules of calculus?

We are going to throw caution to the wind, pretend that there are no problems or subtleties, use Taylor series. . . and see what happens!

Taylor Series . . . and Itô

If we were to do a naive Taylor series expansion of F, completely disregarding the nature of X, and treating dX as a small increment in X, we would get

$$F(X+dX) = F(X) + \frac{dF}{dX}dX + \frac{1}{2}\frac{d^2F}{dX^2}dX^2,$$

ignoring higher-order terms.

We could argue that $F(X+dX) - F(X)$ was just the 'change in' F and so

$$dF = \frac{dF}{dX}dX + \frac{1}{2}\frac{d^2F}{dX^2}dX^2.$$

This is *almost* correct.

Because of the way that we have defined Brownian motion, and have seen how the quadratic variation behaves, it turns out that the dX^2 term isn't really random at all.

The dX^2 term becomes (as all time steps become smaller and smaller) the same as its average value, dt.

Taylor series and the 'proper' Itô are very similar. The only difference being that the correct Itô's lemma has a dt instead of a dX^2.

You can, with little risk of error, use Taylor series with the 'rule of thumb'

$$dX^2 = dt.$$

- and in practice you will get the right result.

We can now answer the question, "If $F = X^2$ what is dF?" In this example

$$\frac{dF}{dX} = 2X \quad \text{and} \quad \frac{d^2F}{dX^2} = 2.$$

Therefore Itô's lemma tells us that

$$dF = dt + 2X\,dX.$$

This is an example of a **stochastic differential equation (SDE)**.

Now consider a slight extension. A function of a stochastic variable $f = f(t, W(t))$, so we can allow both t and $W(t)$ to change, i.e.

$$\begin{aligned} t &\longrightarrow t + dt \\ W &\longrightarrow W + dW. \end{aligned}$$

Using Taylor as before

$$\begin{aligned} f(t + dt, W + dW) &= f(t, W) + \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial W}dW + \frac{1}{2}\frac{\partial^2 f}{\partial W^2}dW^2 + .. \\ df &= f(t + dt, W + dW) - f(t, W) = \\ &\quad \left(\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial W^2}\right)dt + \frac{\partial f}{\partial W}dW \end{aligned}$$

This gives another form of Itô:

$$df = \left(\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial W^2}\right)dt + \frac{\partial f}{\partial W}dW. \qquad (*)$$

This is also a SDE.

Examples:

1. Obtain a SDE for $f = te^{W(t)}$. We need $\frac{\partial f}{\partial t} = \quad W(t)$; $\frac{\partial f}{\partial W} = te^{W(t)} = \frac{\partial^2 f}{\partial W^2}$, then substituting in $(*)$
$$df = \left(\quad W(t) + \tfrac{1}{2} te^{W(t)} \right) dt + te^{W(t)} dW.$$
We can factor out $te^{W(t)}$ and rewrite the above as
$$\frac{df}{f} = \left(\tfrac{1}{t} + \tfrac{1}{2} \right) dt + dW.$$

2. Consider the function of a stochastic variable $f = t^2 W^n (t)$
$$\frac{\partial f}{\partial t} = 2tW^n; \ \frac{\partial f}{\partial W} = nt^2 W^{n-1}; \frac{\partial^2 f}{\partial W^2} = n(n-1)\, t^2 W^{n-2},$$
in $(*)$ gives
$$df = \left(2tW^n + \tfrac{1}{2} n (n-1)\, t^2 W^{n-2} \right) dt + nt^2 W^{n-1} dW.$$

A Formula for Stochastic Integration

If we take the 2D form of Itô given by $(*)$, rearrange and integrate over $[0, t]$, we obtain a very nice formula for integrating functions of the form $f(t, W(t))$:

$$\int_0^t \frac{\partial f}{\partial W} dW = f(t, W(t)) - f(0, W(0)) - \int_0^t \left(\frac{\partial f}{\partial \tau} + \frac{1}{2} \frac{\partial^2 f}{\partial W^2} \right) d\tau$$

Example: Show that

$$\int_0^t \left(t + {}^{W} \right) dW = tW + {}^{W} - 1 - \int_0^t \left(W_\tau + \frac{1}{2} {}^{W_\tau} \right) d\tau.$$

Comparing this to the stochastic integral formula above, we see that $\frac{\partial f}{dW} \equiv t + {}^{W} \Longrightarrow f = tW + {}^{W}$. Also

$$\frac{\partial^2 f}{\partial W^2} = {}^{W_t}, \quad \frac{\partial f}{\partial t} = W_t.$$

Substituting all these terms in to the formula and noting that $f(0, W(0)) =$

1 verifies the result.

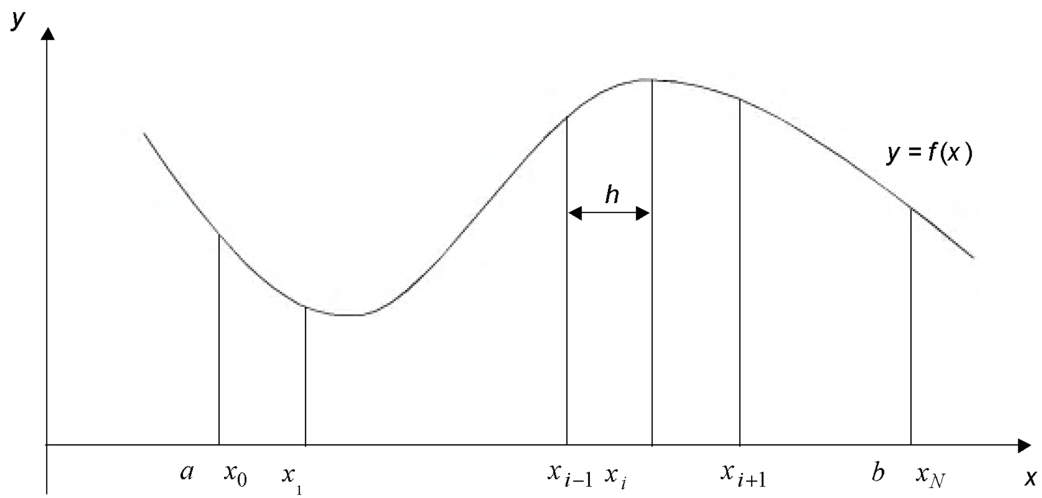
Naturally if $f = f(W(t))$ then the integral formula simply collapses to

$$\int_0^t \frac{df}{dW} dW = f(W(t)) - f(W(0)) - \frac{1}{2}\int_0^t \frac{d^2 f}{dW^2} d\tau$$

Itô Integral

Recall the usual Riemann definition of a definite integral

$$\int_a^b f(x)\, dx$$



which represents the area under the curve between $x = a$ and $x = b$, where the curve is the graph of $f(x)$ plotted against x.

Assuming f is a "well behaved" function on $[a, b]$, there are many different ways (which all lead to the same value for the definite integral).

Start by partitioning $[a, b]$ into N intervals with end points $x_0 = a < x_1 < x_2 < \ \ldots\ldots \ < x_{N-1} < x_N = b$, where the length of an interval $dx = x_i - x_{i+1}$ tends to zero as $N \longrightarrow \infty$. So there are N intervals and $N + 1$ points x_i.

Discretising x gives

$$x_i = a + idx$$

Now consider the definite integral

$$\int_0^T f(t)\, dt.$$

With Riemann integration there are a number of ways we can approximate this:

1. left hand rectangle rule;

$$\int_0^T f(t)\,dt = \lim_{N\to\infty} \sum_{i=0}^{N-1} f(t_i)(t_{i+1} - t_i)$$

or

2. right hand rectangle rule;

$$\int_0^T f(t)\,dt = \lim_{N\to\infty} \sum_{i=0}^{N-1} f(t_{i+1})(t_{i+1} - t_i)$$

or

3. trapezium rule;

$$\int_0^T f(t)\,dt = \lim_{N\to\infty} \sum_{i=0}^{N-1} \frac{1}{2}\left(f(t_i) + f(t_{i+1})\right)(t_{i+1} - t_i)$$

or

4. midpoint rule

$$\int_0^T f(t)\,dt = \lim_{N\to\infty} \sum_{i=0}^{N-1} f\left(\tfrac{1}{2}(t_i + t_{i+1})\right)(t_{i+1} - t_i)$$

In the limit $N \longrightarrow \infty$, $f(t)$ we get the same value for each definition of the definite integral, provided the function is integrable.

Now consider the stochastic integral of the form

$$\int_0^T f(t, W)\,dW = \int_0^T f(t, W(t))\,dW(t)$$

where $W(t)$ is a Brownian motion. We can define this integral as

$$\lim_{N\to\infty} \sum_{i=0}^{N-1} f(t_i, W_i)(W_{i+1} - W_i),$$

where $W_i = W(t_i)$, or as

$$\lim_{N\to\infty} \sum_{i=0}^{N-1} f(t_{i+1}, W_{i+1}) (W_{i+1} - W_i),$$

or as

$$\lim_{N\to\infty} \sum_{i=0}^{N-1} f\left(t_{i+\frac{1}{2}}, W_{i+\frac{1}{2}}\right) (W_{i+1} - W_i),$$

where $t_{i+\frac{1}{2}} = \frac{1}{2}(t_i + t_{i+1})$ and $W_{i+\frac{1}{2}} = W\left(t_{i+\frac{1}{2}}\right)$ or in many other ways. So clearly drawing parallels with the above Riemann form.

Very Important: In the case of a stochastic variable $dW(t)$ the value of the stochastic integral **does** depend on which definition we choose.

In the case of a stochastic integral, the definition

$$I = \lim_{N \to \infty} \sum_{i=0}^{N-1} f\left(t_i, W_i\right)\left(W_{i+1} - W_i\right),$$

is special. This definition results in the **Itô Integral**.

It is special because it is **non-anticipatory**; given that we are at time t_i we know $W_i = W\left(t_i\right)$ and therefore we know $f\left(t_i, W_i\right)$. The only uncertainty is in the $W_{i+1} - W_i$ term.

Compare this to a definition such as

$$\lim_{N \to \infty} \sum_{i=0}^{N-1} f\left(t_{i+1}, W_{i+1}\right)\left(W_{i+1} - W_i\right),$$

which is **anticipatory**; given that at time t_i we know W_i but are uncertain about the future value of W_{i+1}. Thus we are uncertain about *both* the value of

$$f\left(t_{i+1}, W_{i+1}\right)$$

and the value of $\left(W_{i+1} - W_i\right)$ – there exists uncertainty in both of these quantities. That is, evaluation of this integral requires us to anticipate the future value of W_{i+1} so that we may evaluate $f\left(t_{i+1}, W_{i+1}\right)$.

The main thing to note about Itô integrals is that I is a random variable (unlike the deterministic case). Additionally, since I is essentially the limit of a sum of normal random variables, then by the CLT I is also normally distributed, and can be characterized by its mean and variance.

Example: Show that Itô's lemma implies that

$$3\int_0^T W^2 dW = W\left(T\right)^3 - W\left(0\right)^3 - 3\int_0^T W\left(t\right) dt.$$

Show that the result also can be found by writing the integral

$$3\int_0^T W^2 dW = \lim_{N\to\infty} 3\sum_{i=0}^{N-1} {W_i}^2 \left(W_{i+1} - W_i\right)$$

Hint: use $3b^2(a-b) = a^3 - b^3 - 3b(a-b)^2 - (a-b)^3$.

The Itô integral here is defined as

$$\int_0^T 3W^2(t)\,dW(t) = \lim_{N\to\infty} \sum_{i=0}^{N-1} 3W_i^2 \left(W_{i+1} - W_i\right)$$

Now note the hint:

$$3b^2(a-b) = a^3 - b^3 - 3b(a-b)^2 - (a-b)^3$$

hence

$$
\begin{aligned}
&\equiv 3W_i^2\,(W_{i+1} - W_i) \\
&= W_{i+1}^3 - W_i^3 - 3W_i\,(W_{i+1} - W_i)^2 - (W_{i+1} - W_i)^3\,,
\end{aligned}
$$

so that

$$
\sum_{i=0}^{N-1} 3W_i^2\,(W_{i+1} - W_i) =
$$

$$
\sum_{i=0}^{N-1} W_{i+1}^3 - \sum_{i=0}^{N-1} W_i^3 - \sum_{i=0}^{N-1} 3W_i\,(W_{i+1} - W_i)^2
$$
$$
- \sum_{i=0}^{N-1} (W_{i+1} - W_i)^3
$$

Now the first two expressions above give

$$\begin{aligned}\sum_{i=0}^{N-1} W_{i+1}^3 - \sum_{i=0}^{N-1} W_i^3 &= W_N^3 - W_0^3 \\ &= W(T)^3 - W(0)^3 .\end{aligned}$$

In the limit $N \to \infty$, i.e. $dt \to 0$, $(W_{i+1} - W_i)^2 \to dt$, so

$$\lim_{N\to\infty} \sum_{i=0}^{N-1} 3W_i (W_{i+1} - W_i)^2 = \int_0^T 3W(t)\, dt$$

Finally $(W_{i+1} - W_i)^3 = (W_{i+1} - W_i)^2 \cdot (W_{i+1} - W_i)$ which when $N \to \infty$ behaves like $dW^2 dW \sim O\left(dt^{3/2}\right) \longrightarrow 0$.

Hence putting together gives

$$W(T)^3 - W(0)^3 - \int_0^T 3W(t)\, dt$$

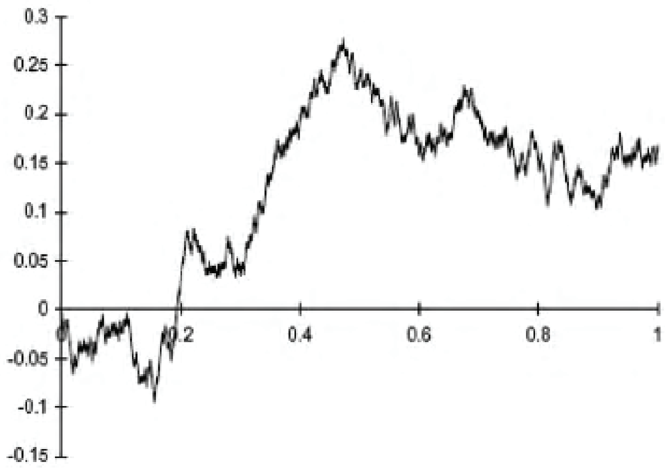
which is consistent with Itô's lemma.

Some pertinent examples

The first example simple Brownian motion but with a drift:

$$dS = \mu dt + \sigma dW.$$

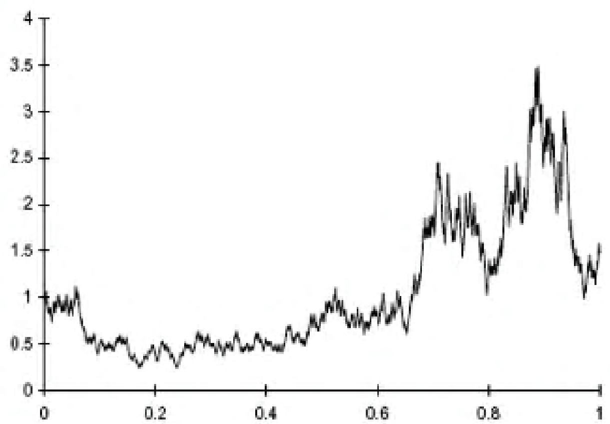
In this realization S has gone negative.



Our second example is similar to the above but the drift and randomness scale with S:

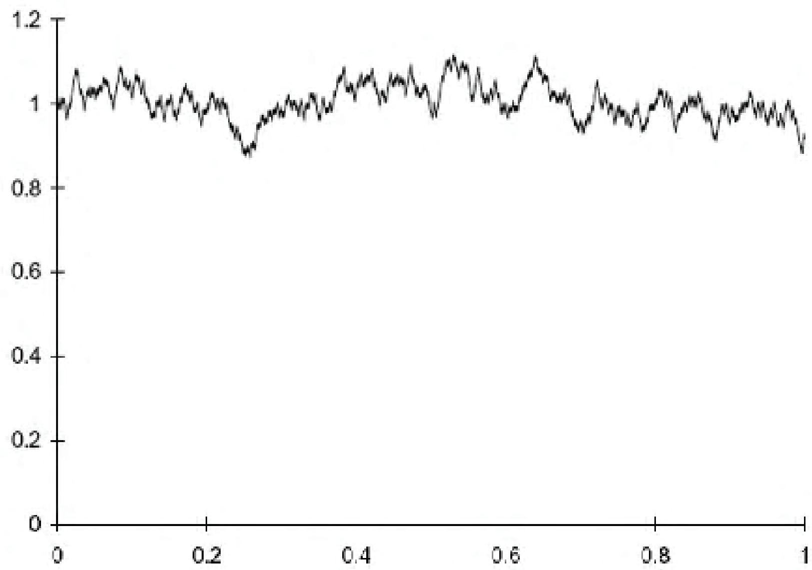
$$dS = \mu S dt + \sigma S dW.$$

If S starts out positive it can never go negative; the closer that S gets to zero the smaller the increments dS.



The third example is

$$dS = (\nu - \mu S)dt + \sigma dW.$$



This random walk is an example of a **mean-reverting** random walk.

If S is large, the negative coefficient in front of dt means that S will move down on average, if S is small it rises on average. There is still no incentive for S to stay positive in this random walk.

With r instead of S this random walk is the Vasicek model for the short-term interest rate.

The final example is similar to the third but we are going to adjust the random term slightly:

$$dS = (\nu - \mu S)dt + \sigma S^{1/2} dW.$$

Now if S ever gets close to zero the randomness decreases, perhaps this will stop S from going negative?

This particular stochastic differential equation for S will be important later on, it is the Cox, Ingersoll & Ross model for the short-term interest rate.