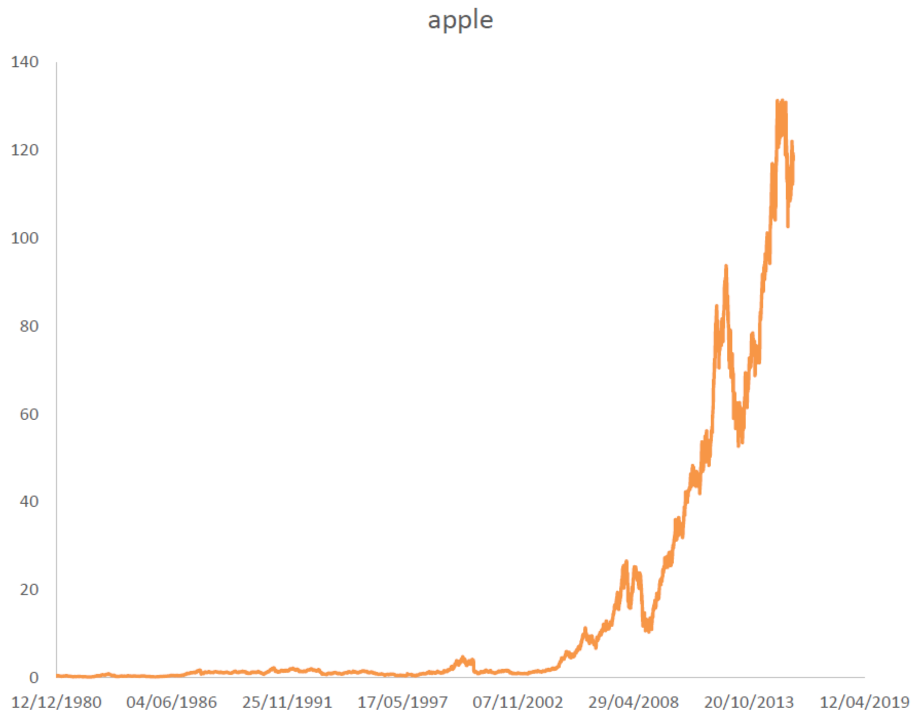
Stochastic Calculus

These are detailed notes for the stochastic calculus sessions. Any typos/edits/comments to

riaz.ahmad@fitchlearning.com



The diagram above shows the flucatuations in the price of apple stock over a 35 year period. This uncertainty that is quite conspicuous is the most important feature of financial modelling. Because there is so much randomness, the most successful mathematical models of financial assets have a probabilistic foundation.

The evolution of financial assets is random and depends on time. They are examples of *stochastic processes* which are random variables indexed (parameterized) with time. If the movement of an asset is discrete it is called a *random walk*. A continuous movement is called a *diffusion process*. We will consider the asset price dynamics to exhibit continuous behaviour and each random path traced out is called a *realization*. If we used a to denote the price of apple stock at time t, we could express this as a_t. This would be an example of a stochastic process.

Hence the need for defining a robust set of properties for the randomness observed in an asset price realization, which is **Brownian Motion.** This was named after the Scottish Botanist who in the summer of 1827, while examining grains of pollen of the plant Clarkia pulchella suspended in water under a microscope, observed minute particles, ejected from the pollen grains, executing a continuous highly irregular fidgety motion. Further study showed that finer particles moved more rapidly, and that the motion was stimulated by heat and by a decrease in the viscosity of the liquid. The findings were published in *A Brief Account of Microscopical Observations Made in the Months of June, July and August 1827* There is little doubt that as well as the most well known (and famous) stochastic process, Brownian motion is the most widely used.

The origins of quantitative finance can be traced back to the start of the twentieth century. Louis Jean-Baptiste Alphonse Bachelier (March 11, 1870 - April 28, 1946) is credited with being the first person to derive the price of an option where the share price movement was modelled by Brownian motion, as part of his PhD at the Sorbonne, entitled Théorie de la Spéculation (published 1900).

Thus, Bachelier may be considered a pioneer in the study of financial mathematics and one of the earliest exponents of Brownian Motion. Five years later Einstein used Brownian motion to study diffusions; which described the microscopic transport of material and heat. In 1920 Norbert Wiener, a mathematician at MIT provided a mathematical construction of Brownian motion together with numerous results about the properties of Brownian motion - in fact he was the first to show that Brownian motion exists and is a well-defined entity. Hence Wiener process is also used as a name for this, and denoted W, $W(t)$ or W_t.

The goal of any investment is to make a good return. Whether the investment is an equity, house, expensive painting or a case of fine wine, the most important thing is a return. The absolute value of the investment (i.e. in £, \$, €) is not an accurate indicator of a return. The **return** is the relative growth in the value of an asset, together with accumulated cashflows (such as dividends), over some period, based on the value that the asset started with.

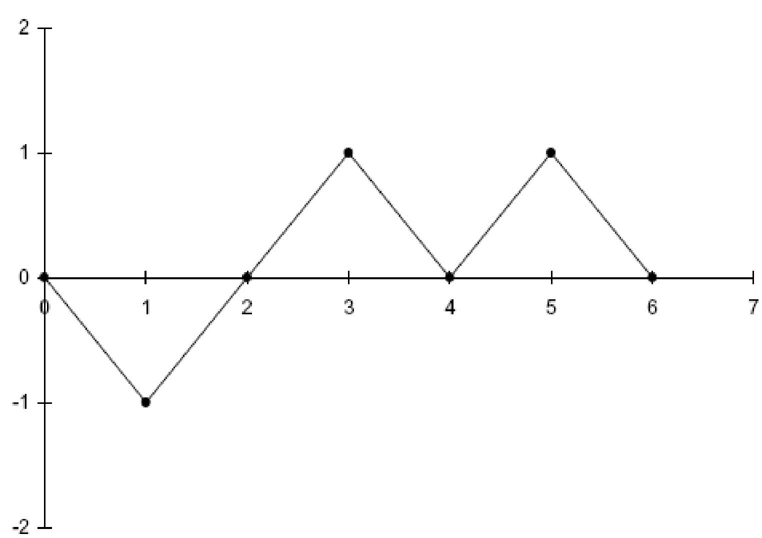
$$\text{Return} = \frac{\text{change in value of the asset} \;+\; \text{accumulated cashflows}}{\text{original value of the asset}}$$

So rather than considering stock prices (for example) directly, it is more useful to study returns over some period. Typically these are examined over a period of a day. To express this mathematically, if the asset value on the ith day is S_i, then the return R_i from day i to day $i+1$ is given by

$$R_i = \frac{S_{i+1} - S_i}{S_i}.$$

Construction of Brownian Motion

Brownian Motion can be constructed as a carefully scaled limit of a symmetric random walk, in the context of a simple gambling game. 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

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

and secondly the variance

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

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

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

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

$$W_0 = 0$$

Now we can calculate expectations of W_n

$$\mathbb{E}[W_n] = \mathbb{E}\left[\sum_{i=1}^{n} R_i\right] = \sum_{i=1}^{n} \mathbb{E}[R_i] = 0$$

$\mathbb{E}[X_n^2] =$

$$\begin{aligned}\mathbb{E}\left[X_n^2\right] &= \mathbb{E}\left[R_1^2 + R_2^2 + ...R_n^2 + 2R_1R_2 + ... + 2R_{n-1}R_n\right] \\ &= \mathbb{E}\left[\sum_{i=1}^{n} R_i^2\right] + \mathbb{E}\left[\sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n} R_iR_j\right] = \sum_{i=1}^{n}\mathbb{E}\left[R_i^2\right] + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n}\mathbb{E}[R_i]\,\mathbb{E}[R_j] \\ &= n \times 1 + 0 \times 0 = n\end{aligned}$$

A Note on Variations

Consider a function f_t, where $t_i = i\dfrac{t}{n}$, we can define different measures of how much f_t varies over time as

$$V^N = \sum_{i=1}^{n} \left|f_{t_i} - f_{t_{i-1}}\right|^N$$

The cases $N = 1, 2$ are important.

$$V = \sum_{i=1}^{n} \left|f_{t_i} - f_{t_{i-1}}\right| \ \textit{total variation} \text{ of trajectory - sum of absolute changes}$$

$$V^2 = \sum_{i=1}^{n} \left|f_{t_i} - f_{t_{i-1}}\right|^2 \ \textit{quadratic variation} \text{ - sum of squared changes}$$

Now look at the *quadratic variation* of the random walk. After each toss, we have won or lost \$1. That is

$$W_n - W_{n-1} = \pm 1 \Longrightarrow |W_n - W_{n-1}| = 1$$

Hence

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

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

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

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.

$$W_i - W_{i-1} = \pm\sqrt{t/n}$$

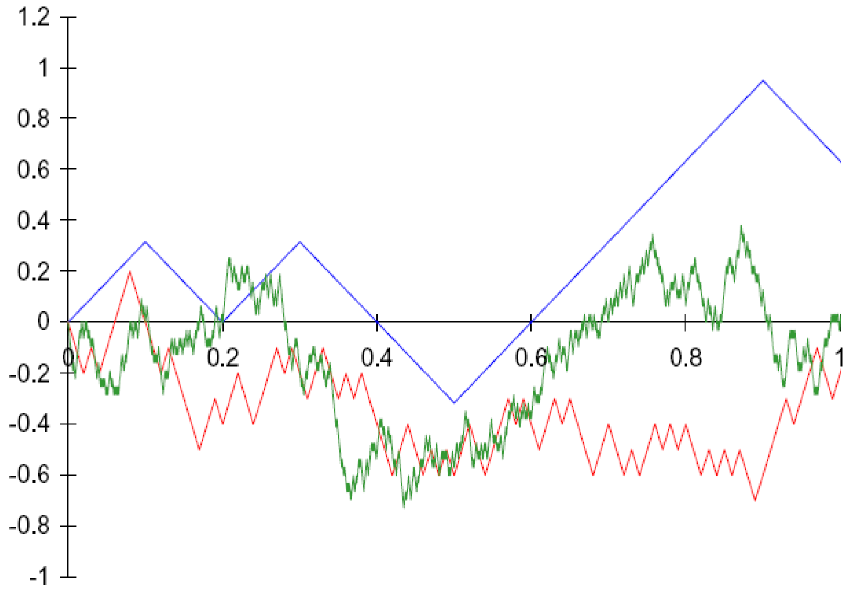
The quadratic variation is

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

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

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

$$\begin{aligned} \mathbb{E}\left[W_t^2\right] &= \mathbb{E}\left[\lim_{n\to\infty}\sum_{i=1}^{n} R_i^2\right] \\ &= \lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{E}\left[R_i^2\right] = \lim_{n\to\infty} n\cdot\left(\sqrt{t/n}\right)^2 \\ \mathbb{V}[W_t] &= \mathbb{E}\left[W_t^2\right] = t \end{aligned}$$

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

Alternative notation for Brownian motion/Wiener process is X_t or B_t.

Another construction of a Brownian motion

Brownian Motion can be constructed as a carefully scaled limit of a symmetric random walk, in the context of a simple gambling game. Consider the coin tossing experiment with payoff

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

Let

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

this is the position of a random walker starting at the origin and after each toss moves one unit up or down with equal probability of a half. We know

$$\mathbb{E}\left[X_n\right] = \sum_{i=1}^{n} \mathbb{E}\left[Z_i\right] = 0$$

$$\mathbb{V}\left[X_n\right] = \sum_{i=1}^{n} \mathbb{V}\left[Z_i\right] = n$$

Now introduce time t; steps N. That is, perform N tosses in time t. Can we find a continuous time limit? Each step takes $\delta t = t/N$. If we let $N \to \infty$, steps of size ± 1 would become infinite (or may possibly collapse to zero). We require a finite random walk. So look for a suitable scaling of the time-step (away from ± 1), keeping in mind the Central Limit Theorem. Let

$$Y = \alpha_N Z$$

for some suitable α_N and let

$$\{X_n : n = 0, 1, ..., N\}$$

be the path of the random walk with steps of size α_N.

Thus

$$\mathbb{E}\left[X_N\right] = 0 \ \forall N$$

$$\begin{aligned} \mathbb{V}\left[X_N\right] &= \mathbb{E}\left[(X_N)^2\right] \\ &= \mathbb{E}\left[\left(\sum_{i=1}^{N} Y_i\right)^2\right] = \mathbb{E}\left[\alpha_N^2 \sum_{i=1}^{N} Z_i^2\right] \\ &= \alpha_N^2 N = \left(\frac{t}{\delta t}\right) \alpha_N^2 \end{aligned}$$

We need $\alpha_N^2/\delta t \sim O(1)$. Choose $\alpha_N^2/\delta t = 1$. Therefore

$$\mathbb{E}\left[(X_N)^2\right] = \mathbb{V}\left[X_N\right] = t.$$

As $N \to \infty$, the random walk

$$\{X_t : t \in [0, \infty)\}$$

converges to Brownian Motion.

Quadratic Variation

Consider a function f_t, which has at most a finite number of jumps or discontinuities. Define the Quadratic Variation

$$Q\left[f\right] = \lim_{N\to\infty}\sum_{i=0}^{N-1}\left|f_{t_{i+1}} - f_{t_i}\right|^2$$

Take the time period $[0, T]$ with N partitions so $dt = T/N$; $t_i = idt$. For Brownian Motion W_t on the interval $[0, T]$ we have

$$\begin{aligned} Q\left[W_t\right] &= \lim_{N\to\infty}\sum_{i=0}^{N-1}\left|W_{t_{i+1}} - W_{t_i}\right|^2 \\ &= \lim_{N\to\infty}\sum_{i=0}^{N-1}\left|\pm\sqrt{dt}\right|^2 = \lim_{N\to\infty} Ndt \\ &= T \end{aligned}$$

So the Quadratic Variation $Q\left[W_t\right] = T$.

The Variation given

$$\begin{aligned} V\left[W_t\right] &= \lim_{N\to\infty}\sum_{i=0}^{N-1}\left|W_{t_{i+1}} - W_{t_i}\right| \\ &= \lim_{N\to\infty}\sum_{i=0}^{N-1}\left|\pm\sqrt{dt}\right| \\ &= \lim_{N\to\infty} N\sqrt{dt} \to \infty \end{aligned}$$

Properties of a Wiener Process/Brownian motion

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

- Brownian motion starts at zero, i.e. $W_0 = 0$ (with probability one).
- Continuity - paths of W_t are continuous (no jumps) a.s. (differentiable nowhere)
- Brownian motion has independent Gaussian increments, with zero mean and variance equal to the temporal extension of the increment. That is 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 Central Limit Theorem we have a distribution that is normal. $W_t - W_s$ has a pdf given by

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

- More specifically $W_{t+s} - W_t$ is independent of W_t. This means if

  $$0 \le t_0 \le t_1 \le t_2 \le .....$$

  $$\begin{aligned} dW_1 &= W_1 - W_0 \text{ is independent of } dW_2 = W_2 - W_1 \\ dW_3 &= W_3 - W_2 \text{ is independent of } dW_4 = W_4 - W_3 \\ &\text{and so on} \end{aligned}$$

Also called *standard Brownian motion* if the above properties hold. More importantly is the result (in stochastic differential equations)

$$dW = W_{t+dt} - W_t \sim N(0, dt)$$

- Brownian motion has *stationary increments.* A stochastic process $(X_t)_{t\ge0}$ is said to be *stationary* if X_t has the same distribution as X_{t+h} for any $h > 0$. This can be checked by defining the increment process $I = (I_t)_{t\ge0}$ by

  $$I_t := W_{t+h} - W_t.$$

  Then $I_t \sim N(0, h)$, and $I_{t+h} = W_{t+2h} - W_{t+h} \sim N(0, h)$ have the same distribution. This is equivalent to saying that the process $(W_{t+h} - W_t)_{h\ge0}$ has the same distribution $\forall t$.

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 at different times. This 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.* Martingales are very important in finance.

Think back to the way the betting game has been constructed. Martingales are essentially stochastic processes that are meant to capture the concept of a fair game in the setting of a gambling environment and thus there exists a rich history in the modelling of gambling games. Although this is a key example area for us, they nevertheless are present in numerous application areas of stochastic processes.

Before discussing the Martingale property of Brownian motion formally, some general background information.

A stochastic process $\{X_n : 0 \le n < \infty\}$ is called a $\mathbb{P}-$ **martingale** with respect to the information **filtration** $\mathcal{F}_n$, and probability distribution $\mathbb{P}$, if the following two properties are satisfied

P1 $\mathbb{E}_n^{\mathbb{P}}\left[|X_n|\right] < \infty \ \forall \ n \ge 0$

P2 $\mathbb{E}_n^{\mathbb{P}}\left[X_{n+m} | \mathcal{F}_n\right] = X_n, \ \forall \ n, m \ge 0$

The first property is simply a technical integrability condition (fine print), i.e. the expected value of the absolute value of X_n must be finite for all n. Such a finiteness condition appears whenever integrals defined over $\mathbb{R}$ are used (think back to the properties of the Fourier Transform for example).

The second property is the one of key importance. This is another expectation result and states that the expected value of X_{n+m} given $\mathcal{F}_n$ is equal to X_n for all non-negative n and m.

The symbol $\mathcal{F}_n$ denotes the information set called a filtration and is the flow of information associated with a stochastic process. This is simply the information we have in our model at time n. It is recognising that at time n we have already observed all the information $\mathcal{F}_n = (X_0, X_1, ..., X_n)$.

So the expected value at any time in the future is equal to its current value - the information held at this point it is the best forecast. Hence the importance of Martingales in modelling fair games. This property is modelling a fair game, our future payoff is equal to the current wealth.

It is also common to use t to depict time

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

Taking expectations of both sides gives

$$\mathbb{E}_t\left[M_T\right] = \mathbb{E}_t\left[M_t\right];\ t < T$$

so martingales have constant mean.

Now replacing the equality in **P2** with an inequality, two further important results are obtained. A process M_t which has

$$\mathbb{E}_t^{\mathbb{P}}\left[M_T \middle| \mathcal{F}_t\right] \geq M_t$$

is called a *submartingale* and if it has

$$\mathbb{E}_t^{\mathbb{P}}\left[M_T \middle| \mathcal{F}_t\right] \leq M_t$$

is called a *supermartingale*.

Using the earlier betting game as an example (where probability of a win or a loss was $\frac{1}{2}$)

$$\begin{array}{rcl} \text{submartingale - gambler wins money on average } \mathbb{P}(H) & > & \frac{1}{2} \\ \text{supermartingale- gambler loses money on average } \mathbb{P}(H) & < & \frac{1}{2} \end{array}$$

The above defintions tell us that every martingale is also a submartingale and a supermartingale. The converse is also true.

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

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

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

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

$$\mathbb{E}^{\mathbb{P}}\left[W_t \middle| \mathcal{F}_s\right] = W_s.$$

Not a Martingale: Brownian motion with drift

$$X_t = W_t + at,$$

with $X_s = W_s + as$ for $s < t$.

$$\begin{aligned} \mathbb{E}\left[W_t + at \middle| \mathcal{F}_s\right] &= \mathbb{E}\left[W_t - W_s + W_s + at \middle| \mathcal{F}_s\right] \\ &= \mathbb{E}\left[W_t - W_s \middle| \mathcal{F}_s\right] + \mathbb{E}\left[W_s + at \middle| \mathcal{F}_s\right] \\ &= W_s + at \\ &\neq X_s. \end{aligned}$$

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).

Markov is also called memoryless as it is a stochastic process in which the distribution of future states depends only on the present state and not on how it arrived there. "It doesn't matter how you arrived at your destination".

Now write the **Markov Property** more formally. Consider time t_0 such that $t_0 \in [0, T]$ together with a given function $h = h(y)$. Define

$$\mathbb{E}^{t_0,x}[h(W_T)]$$

as the expectation of $h(W_T)$ given $W_{t_0} = x$. Now let $\xi \in \mathbb{R}$ be given and start with with initial condition $W_0 = \xi$. The Markov property states

$$\mathbb{E}^{0,\xi}[h(W_T)|\mathcal{F}_{t_0}] = \mathbb{E}^{t_0,W_{t_0}}[h(W_T)].$$

So the only relevant information is W_{t_0}. Imagine starting the process (or SDE) at time t_0 with value W_{t_0}. The past is irrelevant to future dynamics.

Let us look at an example. Consider the earlier random walk S_n given by

$$S_n = \sum_{i=1}^{n} X_i$$

which defined the winnings after n flips of the coin. The X_i's are IID with mean μ. Now define

$$M_n = S_n - n\mu.$$

We will demonstrate that M_n is a Martingale.

Start by writing

$$\mathbb{E}_n[M_{n+m}|\mathcal{F}_n] = \mathbb{E}_n[S_{n+m} - (n+m)\mu].$$

So this is an expectation conditional on information at time n. Now work on the right hand side.

$$\begin{aligned}
&= \mathbb{E}_n\left[\sum_{i=1}^{n+m} X_i - (n+m)\mu\right] \\
&= \mathbb{E}_n\left[\sum_{i=1}^{n} X_i + \sum_{i=n+1}^{n+m} X_i\right] - (n+m)\mu \\
&= \sum_{i=1}^{n} X_i + \mathbb{E}_n\left[\sum_{i=n+1}^{n+m} X_i\right] - (n+m)\mu \\
&= \sum_{i=1}^{n} X_i + \sum_{i=n+1}^{m} \mathbb{E}_n[X_i] - (n+m)\mu = \sum_{i=1}^{n} X_i + m\mu - (n+m)\mu \\
&= \sum_{i=1}^{n} X_i - n\mu = S_n - n\mu \\
\mathbb{E}_n[M_{n+m}] &= M_n.
\end{aligned}$$

Returning to the point about non-differentiability, using the classical sense of differentiation

$$\begin{aligned}
\lim_{\delta t \to 0} \frac{W_{t+\delta t} - W_t}{\delta t} &= \lim_{\delta t \to 0} \frac{O\left(\sqrt{\delta t}\right)}{\delta t} \\
&= \lim_{\delta t \to 0} \frac{1}{\sqrt{\delta t}}
\end{aligned}$$

which does not exist.

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:

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

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

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

$$dW^2 = dt.$$

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 (right most) squaring. Let $Y(t_j) = (W(t_j) - W(t_{j-1}))^2$, 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 $Y(t_j) = (W(t_j) - W(t_{j-1}))^2$ and taking the expectation, we arrive at:

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

Computing the expectation

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

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

Now, since $Z(t_j) = W(t_j) - W(t_{j-1})$ 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 $\psi \sim N(\mu, \sigma^2)$ 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

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

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

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

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

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

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 $3\frac{t^2}{n^2}$.

and

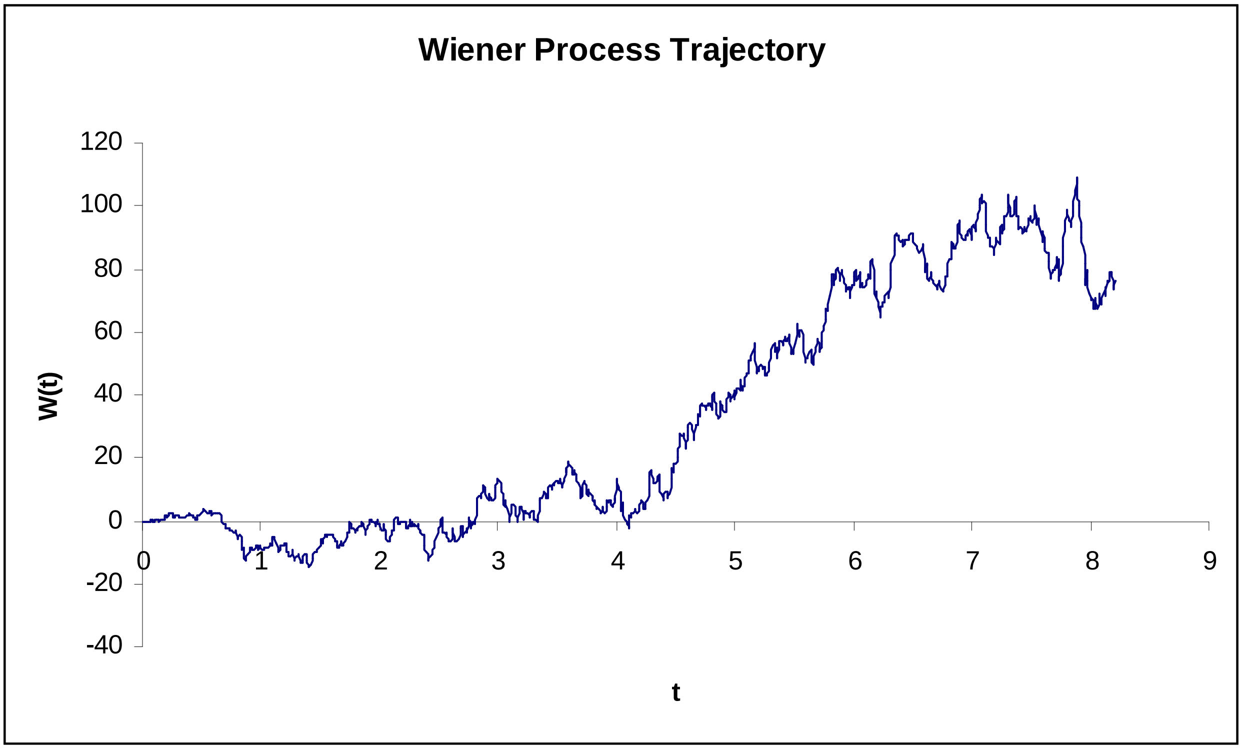
$$\mathbb{E}\left[(W(t_j) - W(t_{j-1}))^2\right] = \frac{t}{n} \qquad \text{for } j = 1, \ldots, n$$

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:

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

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:

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

Taylor Series and Itô

If we were to do a naive Taylor series expansion of $F(W_t)$, completely disregarding the nature of W_t, and treating dW_t as a small increment in W_t, we would get

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

ignoring higher-order terms. We could argue that $F(W_t + dW_t) - F(W_t)$ was just the 'change in' F and so

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

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 dW_t^2 term isn't really random at all. The dW_t^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 dW_t^2.

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

$$dW_t^2 = dt.$$

and in practice you will get the right result.

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

$$\frac{dF}{dW_t} = 2W_t \text{ and } \frac{d^2F}{dW_t^2} = 2.$$

Therefore Itô's lemma tells us that

$$dF = dt + 2W_t dW_t.$$

This is an example of a **stochastic differential equation (SDE),** written more generally as

$$dF = A(W_t)\,dt + B(W_t)\,dW_t.$$

There are two parts. The part before the plus, $A(W_t)\,dt$, is the deterministic bit. The random component follows $B(W_t)\,dW_t$. More importantly $A(W_t)$ is the drift; $B(W_t)$ is the diffusion.

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

$$\begin{aligned} t &\longrightarrow t + dt \\ W_t &\longrightarrow W_t + dW_t. \end{aligned}$$

Using Taylor as before

$$\begin{aligned} f(t+dt, W_t + dW_t) &= f(t, W_t) + \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial W_t} dW_t + \tfrac{1}{2}\frac{\partial^2 f}{\partial W_t^2} dW_t^2 + .. \\ df &= f(t+dt, W_t + dW_t) - f(t, W_t) \end{aligned}$$

This gives another form of Itô:

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

This is also a SDE.

Examples:

1. Obtain a SDE for $f = te^{W_t}$. We need $\frac{\partial f}{\partial t} = e^{W_t}$; $\frac{\partial f}{\partial W_t} = te^{W_t} = \frac{\partial^2 f}{\partial W_t^2}$, then substituting in $(*)$

$$df = \left(e^{W_t} + \tfrac{1}{2}te^{W_t}\right) dt + te^{W_t} dW_t.$$

We can factor out te^{W_t} and rewrite the above as

$$\frac{df}{f} = \left(\tfrac{1}{t} + \tfrac{1}{2}\right) dt + dW_t.$$

2. Consider the function of a stochastic variable $f = t^2 W_t^n$

$$\frac{\partial f}{\partial t} = 2tW_t^n; \ \frac{\partial f}{\partial W_t} = nt^2 W_t^{n-1}; \ \frac{\partial^2 f}{\partial W_t^2} = n\left(n-1\right) t^2 W_t^{n-2},$$

in $(*)$ gives

$$df = \left(2tW_t^n + \tfrac{1}{2}n\left(n-1\right) t^2 W_t^{n-2}\right) dt + nt^2 W_t^{n-1} dW_t.$$

Itô multiplication table:

$\times$	dt	dW_t
dt	$dt^2 = 0$	$dtdW_t = 0$
dW_t	$dW_t dt = 0$	$dW_t^2 = dt$

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 $f\left(t, W\left(t\right)\right)$:

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

Example: Show that

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

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

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

Substituting all these terms in to the formula and noting that $f\left(0, W\left(0\right)\right) = 1$ verifies the result.

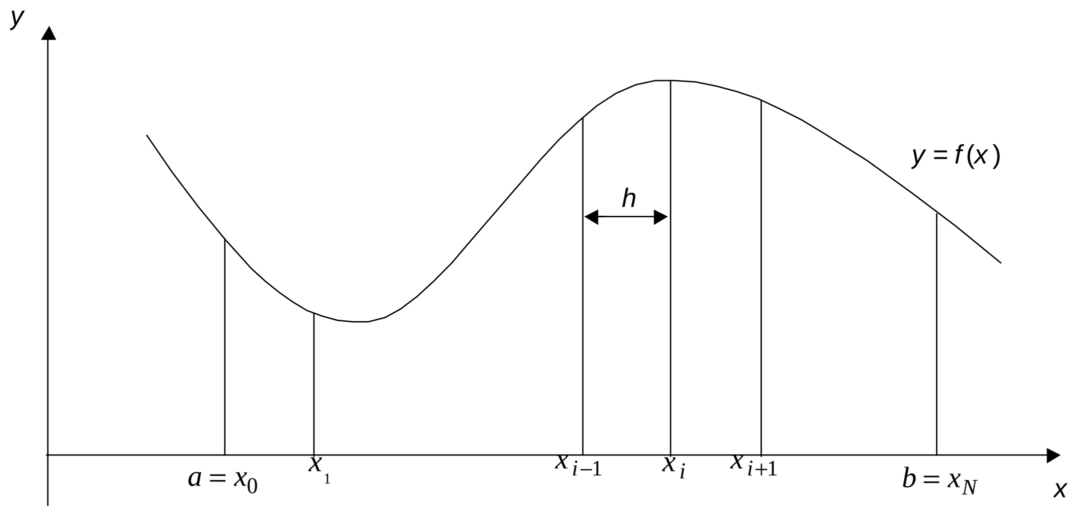
Naturally if $f = f\left(W\left(t\right)\right)$ then the integral formula simply collapses to

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

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 $x_0 = a < x_1 < x_2 <$ ...... $< x_{N-1} < x_N = b$, where the length of an interval $dx = x_i - x_{i+1}$ tends to zero as $N \to \infty$. 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;

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

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 $N\to\infty$, $f(t)$ 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

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

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 $W_i = W(t_i)$ and therefore we know $f(t_i,W_i)$. The only uncertainty is in the $W_{i+1}-W_i$ term.

Compare this to a definition such as

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

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

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

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

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

so that

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

$$\begin{aligned} &\sum_{i=0}^{N-1} W_{i+1}^3 - \sum_{i=0}^{N-1} W_i^3 - \sum_{i=0}^{N-1} 3W_i\left(W_{i+1} - W_i\right)^2 \\ &-\sum_{i=0}^{N-1}\left(W_{i+1} - W_i\right)^3 \end{aligned}$$

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.

Exercise: Consider the Itô integral of the form

$$\int_0^T f(t, W(t))\,dW(t) = \lim_{N\to\infty} \sum_{i=0}^{N-1} f(t_i, W_i)(W_{i+1} - W_i).$$

The interval $[0, T]$ is divided into N partitions with end points

$$t_0 = 0 < t_1 < t_2 < \ \ldots\ldots \ < t_{N-1} < t_N = T,$$

where the length of an interval $t_{i+1} - t_i$ tends to zero as $N \to \infty$.

We know from Itô's lemma that

$$2\int_0^T W(t)\,dW(t) = W(T)^2 - W(0)^2 - T.$$

Show from the definition of the Itô integral that the result can also be found by initially writing the integral

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

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

Consider the Itô integral of the form

$$\int_0^T f(t, W(t))\,dW(t) = \lim_{N\to\infty} \sum_{i=0}^{N-1} f(t_i, W_i)(W_{i+1} - W_i).$$

The interval $[0, T]$ is divided into N partitions with end points

$$t_0 = 0 < t_1 < t_2 < \ \ldots\ldots \ < t_{N-1} < t_N = T,$$

where the length of an interval $t_{i+1} - t_i$ tends to zero as $N \to \infty$.

We know from Itô's lemma that

$$4\int_0^T W^3(t)\,dW(t) = W^4(T) - W^4(0) - 6\int_0^T W^2(t)\,dt$$

Show from the definition of the Itô integral that the result can also be found by initially writing the integral

$$4\int_0^T W^3 dX = \lim_{N\to\infty} 4\sum_{i=0}^{N-1} W_i^3 \left(W_{i+1} - W_i\right)$$

Hint: use $4b^3(a-b) = a^4 - b^4 - 4b(a-b)^3 - 6b^2(a-b)^2 - (a-b)^4$**.**

The other important property that the Itô integral has is that it is a martingale. We know that

$$W_{i+1} - W_i$$

is a martingale; i.e. in the context

$$\mathbb{E}\left[W_{i+1} - W_i\right] = 0.$$

Since

$$\mathbb{E}\left[\sum_{i=0}^{N-1} f\left(t_i, W_i\right)\left(W_{i+1} - W_i\right)\right] = \sum_{i=0}^{N-1} f\left(t_i, W_i\right)\mathbb{E}\left[W_{i+1} - W_i\right] = 0$$

Thus

$$\mathbb{E}\left[\int_0^T f\left(t, W\left(t\right)\right) dW\left(t\right)\right] = 0.$$

This is, essentially a consequence of the Itô integral being non-anticipatory, as discussed earlier. No other stochastic integral has this property.

Diffusion Process

A stochastic process, X_t is called a *diffusion process* if it satisfies

$$dX_t = A(X_t, t)\,dt + B(X_t, t)\,dW_t \tag{1}$$

This is also an example of a Stochastic Differential Equation (SDE) for the process X_t and consists of two components:

1. $A(X_t,t)\,dt$ is deterministic – coefficient of dt is known as the *drift* of the process.
2. $B(X_t, t)\,dW_t$ is random – coefficient of dW_t is known as the *diffusion* or *volatility* of the process.

We say X_t evolves according to (or follows) this process.

For example

$$dG_t(t) = (G_t + G_{t-1})\,dt + dW_t$$

is not a diffusion (although it is a SDE)

- $A \equiv 0$ and $B \equiv 1$ reverts the process back to Brownian motion.
- Called time-homogeneous if A and B are not dependent on t.
- $dX_t{}^2 = B^2 dt$.

We say (1) is a SDE for the process X_t or a *Random Walk* for dX_t. If $A,\ B$ are not dependent on t the diffusion is said to be time homogeneous (i.e. time dependent).

Suppose for (1) the initial condition (t_0, x) are given, i.e. $X_{t_0} = x$. The diffusion (1) can be written in integral form as

$$X_t = x + \int_{t_0}^{t} A(X_s, s)\,ds + \int_{t_0}^{t} B(X_s, s)\,dW_s;\ t \geq t_0.$$

Knowing the path of the Wiener Process up to time t, we can evaluate X_t.
Diffusion processes have two main characteristics. They are:

(i) continuous

(ii) Markov processes

Thus the Markov property is one of the most important properties for the study of stochastic differential equations.

Remark: A diffusion X_t is a *Markov* process if - once the present state $X_t = g$ is given, the past $\{X_\tau,\ \tau < t\}$ is irrelevant to the future dynamics.

We have seen that Brownian motion can take on negative values so its direct use for modelling stock prices is unsuitable. Instead a non-negative variation of Brownian motion called geometric Brownian motion (GBM) is used

If for example we have a diffusion X_t

$$dX_t = \mu X_t dt + \sigma X_t dW_t \tag{2}$$

then the drift is $A(X_t, t) = \mu X_t$ and diffusion is $B(X_t, t) = \sigma X_t$.

The process (2) is also called Geometric Brownian Motion (GBM).

Brownian motion $W(t)$ is used as a basis for a wide variety of models. Consider a pricing process $\{S(t) : t \in \mathbb{R}_+\}$: we can model its instantaneous change dS by a SDE

$$dS = a(S,t)\,dt + b(S,t)\,dW_t \tag{3}$$

By choosing different coefficients a and b we can have various properties for the diffusion process.

A very popular finance model for generating asset prices is the GBM model given by (2). The instantaneous return on a stock $S(t)$ is a constant coefficient SDE

$$\frac{dS}{S} = \mu dt + \sigma dW_t \tag{4}$$

where μ and σ are the return's drift and volatility, respectively.

$$\underbrace{\frac{dS}{S}}_{\substack{\text{relative change in stock}\\ \text{between } t \text{ and } t+dt}} = \underbrace{\mu}_{\substack{\text{mean return of stock}\\ \text{between } t \text{ and } t+dt}} dt + \underbrace{\sigma}_{\substack{\text{standard deviation of}\\ \text{stock}}} \underbrace{dW_t}_{\substack{\text{new random}\\ \text{term}}}$$

Now suppose we have a function $V = V(S_t, t)$ where S_t is a process which evolves according to (4). If $S_t \longrightarrow S_t + dS_t$, $t \longrightarrow t + dt$ then a natural question to ask is "what is the jump in V?" To answer this we return to Taylor, which gives

$$\begin{aligned} & V(S_t + dS_t, t + dt) \\ = \; & V(S_t, t) + \frac{\partial V}{\partial t} dt + \frac{\partial V}{\partial S_t} dS_t + \frac{1}{2} \frac{\partial^2 V}{\partial S^2} dS_t^2 + O\left(dS_t^3, dt^2\right) \end{aligned}$$

So S_t follows

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

Remember that

$$\mathbb{E}(dW_t) = 0, \quad dW_t^2 = dt$$

we only work to $O(dt)$ - anything smaller we ignore and we also know that

$$dS_t^2 = \sigma^2 S_t^2 dt$$

So the change dV when $V(S_t, t) \rightarrow V(S_t + dS_t, t + dt)$ is given by

$$dV = \frac{\partial V}{\partial t} dt + \frac{\partial V}{\partial S_t} (S_t \mu dt + S_t \sigma dW_t) + \frac{1}{2} \sigma^2 S_t^2 \frac{\partial^2 V}{\partial S_t^2} dt$$

Re-arranging to have the standard form of a SDE $dG = a(G, t)\, dt + b(G, t)\, dW_t$ gives

$$dV = \left(\frac{\partial V}{\partial t} + \mu S_t \frac{\partial V}{\partial S_t} + \frac{1}{2} \sigma^2 S_t^2 \frac{\partial^2 V}{\partial S_t^2} \right) dt + \sigma S_t \frac{\partial V}{\partial S_t} dW_t. \tag{5}$$

This is Itô's Formula in two dimensions.

Naturally if $V = V(S)$ then (5) simplifies to the shorter version

$$dV = \left(\mu S_t \frac{dV}{dS_t} + \frac{1}{2} \sigma^2 S_t^2 \frac{d^2 V}{dS_t^2} \right) dt + \sigma S_t \frac{dV}{dS_t} dW_t. \tag{6}$$

Further Examples

In the following cases S evolves according to GBM.

Given $V = t^2 S^3$ obtain the SDE for V, i.e. dV. So we calculate the following terms

$$\frac{\partial V}{\partial t} = 2tS^3, \quad \frac{\partial V}{\partial S} = 3t^2 S^2 \rightarrow \frac{\partial^2 V}{\partial S^2} = 6t^2 S.$$

We now substitute these into (5) to obtain

$$dV = \left(2tS^3 + 3\mu t^2 S^3 + 3\sigma^2 S^3 t^2\right) dt + 3\sigma t^2 S^3 dW.$$

Now consider the example $V = \exp(tS)$

Again, function of 2 variables. So

$$\begin{aligned} \frac{\partial V}{\partial t} &= S \exp(tS) = SV \\ \frac{\partial V}{\partial S} &= t \exp(tS) = tV \\ \frac{\partial^2 V}{\partial S^2} &= t^2 V \end{aligned}$$

Substitute into (5) to get

$$dV = V\left(S + \mu tS + \frac{1}{2}\sigma^2 S^2 t^2\right) dt + (\sigma StV)\, dW.$$

Not usually possible to write the SDE in terms of V – but if you can do so - do not struggle to find a relation if it does not exist. Always works for exponentials.

One more example: That is $S(t)$ evolves according to GBM and $V = V(S) = S^n$. So use

$$dV = \left[\mu S\frac{dV}{dS} + \frac{1}{2}\sigma^2 S^2 \frac{d^2V}{dS^2}\right] dt + \left[\sigma S\frac{dV}{dS}\right] dW.$$

$$V'(S) = nS^{n-1} \rightarrow V''(S) = n(n-1)S^{n-2}$$

Therefore Itô gives us $dV =$

$$\left[\mu SnS^{n-1} + \frac{1}{2}\sigma^2 S^2 n(n-1)S^{n-2}\right] dt + \left[\sigma SnS^{n-1}\right] dW$$

$$dV = \left[\mu nS^n + \frac{1}{2}\sigma^2 n(n-1)S^n\right] dt + [\sigma nS^n]\, dW$$

Now we know $V(S) = S^n$, which allows us to write

$$dV = V\left[\mu n + \frac{1}{2}\sigma^2 n(n-1)\right] dt + [\sigma n]\, V dW$$

with drift $= V\left[\mu n + \frac{1}{2}\sigma^2 n(n-1)\right]$ and diffusion $= \sigma nV$.

Example

1. (a) Itô's lemma can be used to deduce the following formula for stochastic differential equations and stochastic integrals

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

for a function $F(W(\tau), \tau)$ where $dW(\tau)$ is an increment of a Brownian motion.
If $W(0) = 0$ evaluate

$$\int_0^t \tau^2 \sin W\, dW(\tau).$$

$$\downarrow \frac{\partial F}{\partial W} = t^2 \sin W \longrightarrow \quad F = -t^2 \cos W \downarrow$$

$$\frac{\partial^2 F}{\partial W^2} = t^2 \cos W \qquad \frac{\partial F}{\partial t} = -2t \cos W$$

and substitute into the integral formula

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

(b) Suppose the stochastic process $S(t)$ evolves according to Geometric Brownian Motion (GBM), where

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

Obtain a SDE $df(S, t)$ for each of the following functions

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

i $f(S, t) = \alpha^t + \beta t S^n \qquad \alpha,\ \beta$ are constants

$$\begin{aligned} \frac{\partial f}{\partial t} &= \alpha^t \log a + \beta S^n; \ \frac{\partial f}{\partial S} = n\beta t S^{n-1}; \ \frac{\partial^2 f}{\partial S^2} = n(n-1)\beta t S^{n-2} \\ df &= \left(\alpha^t \log a + \beta S^n + n\mu\beta t S^n + \frac{1}{2} n(n-1)\beta t \sigma^2 S^n\right) dt + \sigma n \beta t S^n dW \end{aligned}$$

ii $f(S, t) = \log tS + \cos tS$

$$\begin{aligned} \frac{\partial f}{\partial t} &= \frac{1}{t} - S \sin tS; \ \frac{\partial f}{\partial S} = \frac{1}{S} - t \sin tS; \ \frac{\partial^2 f}{\partial S^2} = -\frac{1}{S^2} - t^2 \cos tS \\ df &= \left(\frac{1}{t} - S \sin tS + \mu S\left(\frac{1}{S} - t \sin tS\right) + \frac{1}{2}\sigma^2 S^2 \left(-\frac{1}{S^2} - t^2 \cos tS\right)\right) dt + \\ & \quad \sigma S \left(\frac{1}{S} - t \sin tS\right) dW \end{aligned}$$

Important Cases - Equities and Interest Rates

If we now consider S which follows a lognormal random walk, i.e. $V = \log(S)$ then substituting into (6) gives

$$d\left((\log S)\right) = \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma dW$$

Integrating both sides over a given time horizon (between t_0 and T)

$$\int_{t_0}^{T} d\left((\log S)\right) = \int_{t_0}^{T} \left(\mu - \frac{1}{2}\sigma^2\right) dt + \int_{t_0}^{T} \sigma dW \quad (T > t_0)$$

we obtain

$$\log \frac{S(T)}{S(t_0)} = \left(\mu - \frac{1}{2}\sigma^2\right)(T - t_0) + \sigma\left(W(T) - W(t_0)\right)$$

Assuming at $t_0 = 0$, $W(0) = 0$ and $S(0) = S_0$ the exact solution becomes

$$S_T = S_0 \exp\left\{\left(\mu - \frac{1}{2}\sigma^2\right) T + \sigma\phi\sqrt{T}\right\}. \tag{7}$$

(7) is of particular interest when considering the pricing of a simple European option due to its non path dependence. Stock prices cannot become negative, so we allow S, a non-dividend paying stock to evolve according to the lognormal process given above - and acts as the starting point for the Black-Scholes framework.

However μ is replaced by the risk-free interest rate r in (7) and the introduction of the risk-neutral measure - in particular the Monte Carlo method for option pricing.

Interest rates exhibit a variety of dynamics that are distinct from stock prices, requiring the development of specific models to include behaviour such as return to equilibrium, boundedness and positivity. The earliest interest rate models took as their starting point a stochastic model for the short rate, or instantaneous interest rate denoted r_t. These were one factor models meaning the single source of randomnessThe interest rate for the shortest possible deposit is commonly called the spot interest rate, or simply the **spot rate**.This is defined as the rate of interest for the infinitessimal interval $[t, t+dt]$, with

$$r_t dt = \text{total interest gained in } [t, t+dt] .$$

The spot rate was introduced specifically for the purpose of efficient interest rate modelling and is not traded in the markets. In practice one takes yield on a liquid finite maturity bond e.g. a one month US Treasury bill. Unlike equities, there are numerous models for capturing the dynamics of interest rates. These short rate models for r_t are expressed as a SDE

$$dr_t = u(r_t, t)\, dt + w(r_t, t)\, dW_t$$

for given coefficients $u(r_t, t)$ and $w(r_t, t)$.

Here we consider another important example of a stochastic differential equation, put forward by Vasicek in 1977. This model has a mean reverting Ornstein-Uhlenbeck process for the short rate and is used for generating interest rates, given by

$$dr_t = (\eta - \gamma r_t)\, dt + \sigma dW_t. \tag{8}$$

So drift is $(\eta - \gamma r_t)$ and volatility given by σ.

γ refers to the *speed of reversion* or simply the *speed.* $\frac{\eta}{\gamma} (= \overline{r})$ denotes the mean (equlibrium) rate, and we can rewrite this random walk (7) for dr_t in a more popular form as

$$dr_t = -\gamma (r_t - \overline{r})\, dt + \sigma dW_t.$$

The dimensions of γ are 1/time, hence $1/\gamma$ has the dimensions of time (years). For example a rate that has speed $\gamma = 3$ takes one third of a year to revert back to the mean, i.e. 4 months. $\gamma = 52$ means $1/\gamma = 1/52$ years i.e. 1 week to mean revert (hence very rapid). The mean reverting behaviour is consistent with economics theory supporting the idea that interest rates should fluctuate along a long term mean equlibrium rate, determined by economic equlilibrium between demand and supply - a comon feature of all short rate models. The one disadvantage of the Vasicek model is that interest rates can become negative, which is undesirable in economics and highly disliked by economists - they below in a zero lower bound. This bizarre concept in its simplest form means a lender pays another party to borrow its money. That is, the bank will charge its customers for having accounts. However, it is not as odd as it seems given the situation of negative interest rates in Switzerland in the 60s and more recently in Japan. That said, you and I do not need to worry as this is not a concept about to hit the high street.
Returning to the mathematics! By setting $X_t = r_t - \overline{r}$, X_t is a solution of

$$dX_t = -\gamma X_t dt + \sigma dW_t;\; X_0 = \alpha, \tag{9}$$

hence it follows that X_t is an Ornstein-Uhlenbeck process and an analytic solution for this equation exists. (9) can be written as $dX_t + \gamma X_t dt = \sigma dW_t$.

Multiply both sides by an integrating factor $e^{\gamma t}$

$$\begin{aligned} e^{\gamma t}(dX_t + \gamma X_t dt) &= \sigma e^{\gamma t} dW_t \\ d\left(e^{\gamma t} X_t\right) &= \sigma e^{\gamma t} dW_t \end{aligned}$$

Integrating over $[0,t]$ gives

$$\begin{aligned}\int_0^t d\left(e^{\gamma s}X_s\right) &= \int_0^t \sigma e^{\gamma s} dW_s \\ e^{\gamma s}X_s\big|_0^t &= \int_0^t \sigma e^{\gamma s} dW_s \rightarrow e^{\gamma t}X_t - X_0 = \int_0^t \sigma e^{\gamma s} dW_s\end{aligned}$$

$$X_t = \alpha e^{-\gamma t} + \sigma \int_0^t e^{\gamma(s-t)} dW_s. \tag{10}$$

By using integration by parts, i.e. $\int v\ du = uv - \int u\ dv$ we can simplify (10).

$$\begin{aligned}&u = W_s \\ &v = e^{\gamma(s-t)} \rightarrow dv = \gamma e^{\gamma(s-t)} ds\end{aligned}$$

Therefore

$$\int_0^t e^{\gamma(s-t)} dW_s = W_t - \gamma \int_0^t e^{\gamma(s-t)} W_s\ ds$$

and we can write (10) as

$$X_t = \alpha e^{-\gamma t} + \sigma \left(W_t - \gamma \int_0^t e^{\gamma(s-t)} W_s\ ds\right)$$

allowing numerical treatment for the integral term.

The Itô product rule

Let X_t, Y_t be two one-dimensional Itô processes, where

$$\begin{aligned} dX_t &= a(t, X_t)\, dt + b(t, X_t)\, dW_t^{(1)}, \\ dY_t &= c(t, Y_t)\, dt + d(t, Y_t)\, dW_t^{(2)} \end{aligned}$$

By applying the two-dimensional form of Itô's lemma with $f(t, x, y) = xy$

$$df = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial y}\delta y + \tfrac{1}{2}\frac{\partial^2 f}{\partial x^2}\delta x^2 + \tfrac{1}{2}\frac{\partial^2 f}{\partial y^2}\delta y^2 + \frac{\partial^2 f}{\partial x \partial y}\delta x \delta y$$

$$\begin{array}{lll} \frac{\partial f}{\partial t} = 0 & \frac{\partial f}{\partial x} = y & \frac{\partial f}{\partial y} = x \\ \frac{\partial^2 f}{\partial x^2} = 0 & \frac{\partial^2 f}{\partial y^2} = 0 & \frac{\partial^2 f}{\partial x \partial y} = 1 \end{array}$$

which gives

$$df = y\delta x + x\delta y + \delta x \delta y$$

to give

$$d(X_t Y_t) = X_t dY_t + Y_t dX_t + dX_t dY_t.$$

Now consider a pair of stochastic processes that are independent standard Brownian motions, i.e. $W_t^{(1)}$, $W_t^{(2)}$ such that $Z_t = W_t^{(1)} W_t^{(2)}$, then

$$d(Z_t) = W_t^{(1)} dW_t^{(2)} + W_t^{(2)} dW_t^{(1)} + \rho dt.$$

The Itô rule for ratios

X_t, Y_t be two one-dimensional Itô processes, where

$$\begin{aligned} dX_t &= \mu_X(t, X_t)\, dt + \sigma_X(t, X_t)\, dW_t^{(1)}, \\ dY_t &= \mu_Y(t, Y_t)\, dt + \sigma_Y(t, Y_t)\, dW_t^{(2)}. \end{aligned}$$

And suppose

$$dW_t^{(1)} dW_t^{(2)} = \rho dt.$$

By applying the two-dimensional form of Itô's lemma with $f(X, Y) = X/Y$.

We already know that for $f(t, X, Y)$

$$\begin{aligned} df &= \frac{\partial f}{\partial X} dX + \frac{\partial f}{\partial Y} dY + \tfrac{1}{2}\frac{\partial^2 f}{\partial X^2} dX^2 + \tfrac{1}{2}\frac{\partial^2 f}{\partial Y^2} dY^2 + \frac{\partial^2 f}{\partial X \partial Y} dX\, dY \\ &= \left(\mu_X \frac{\partial f}{\partial X} + \mu_Y \frac{\partial f}{\partial Y} + \tfrac{1}{2}\sigma_X^2 \frac{\partial^2 f}{\partial X^2} + \tfrac{1}{2}\sigma_Y^2 \frac{\partial^2 f}{\partial Y^2} + \rho \sigma_X \sigma_Y \frac{\partial^2 f}{\partial X \partial Y}\right) dt \\ &\quad + \sigma_X \frac{\partial f}{\partial X} dW_t^{(1)} + \sigma_Y \frac{\partial f}{\partial Y} dW_t^{(2)} \end{aligned}$$

$$\begin{array}{lll} \frac{\partial f}{\partial t} = 0 & \frac{\partial f}{\partial X} = 1/Y & \frac{\partial f}{\partial Y} = -X/Y^2 \\ \frac{\partial^2 f}{\partial x^2} = 0 & \frac{\partial^2 f}{\partial Y^2} = 2X/Y^3 & \frac{\partial^2 f}{\partial X \partial Y} = -1/Y^2 \end{array}$$

which gives

$$\begin{aligned} df &= \left(\mu_X \frac{1}{Y} - \mu_Y \frac{X}{Y^2} + \sigma_Y^2 \frac{X}{Y^3} - \rho \sigma_X \sigma_Y \frac{1}{Y^2}\right) dt + \sigma_X \frac{1}{Y} dW_t^{(1)} - \sigma_Y \frac{X}{Y^2} dW_t^{(2)} \\ \frac{df}{f} &= \left(\frac{\mu_X}{X} - \frac{\mu_Y}{Y} + \frac{\sigma_Y^2}{Y^2} - \frac{\rho \sigma_X \sigma_Y}{XY}\right) dt + \frac{\sigma_X}{X} dW_t^{(1)} - \frac{\sigma_Y}{Y} dW_t^{(2)} \end{aligned}$$

Another common form is

$$d\left(\frac{X}{Y}\right) = \frac{X}{Y}\left(\frac{dX}{X} - \frac{dY}{Y} - \frac{dX\,dY}{XY} + \left(\frac{dY}{Y}\right)^2\right)$$

As an example suppose we have

$$\begin{aligned} dS_1 &= 0.1dt + 0.2dW_t^{(1)}, \\ dS_2 &= 0.05dt + 0.1dW_t^{(2)}, \end{aligned}$$

$\rho = 0.4$

$$d\left(\frac{S_1}{S_2}\right) = \left(\mu_X \frac{1}{Y} - \mu_Y \frac{X}{Y^2} + \sigma_Y^2 \frac{X}{Y^3} - \rho\sigma_X\sigma_Y \frac{1}{Y^2}\right) dt + \sigma_X \frac{1}{Y} dW_t^{(1)} - \sigma_Y \frac{X}{Y^2} dW_t^{(2)}$$

where

$$\begin{aligned} \mu_X &= 0.1;\ \mu_Y = 0.05 \\ \sigma_X &= 0.2;\ \sigma_Y = 0.1 \end{aligned}$$

$$d\left(\frac{S_1}{S_2}\right) = \left(\frac{0.1}{S_2} - 0.05\frac{S_1}{S_2^2} + 0.01\frac{S_1}{S_2^3} - 0.008\frac{1}{S_2^2}\right) dt + 0.2\frac{1}{S_2} dW_t^{(1)} - 0.1\frac{S_1}{S_2^2} dW_t^{(2)}$$

Discrete Time Random Walks

When simulating a random walk we write the SDE given by (6) in discrete form

$$\delta S = S_{i+1} - S_i = rS_i\delta t + \sigma S_i\phi\sqrt{\delta t}$$

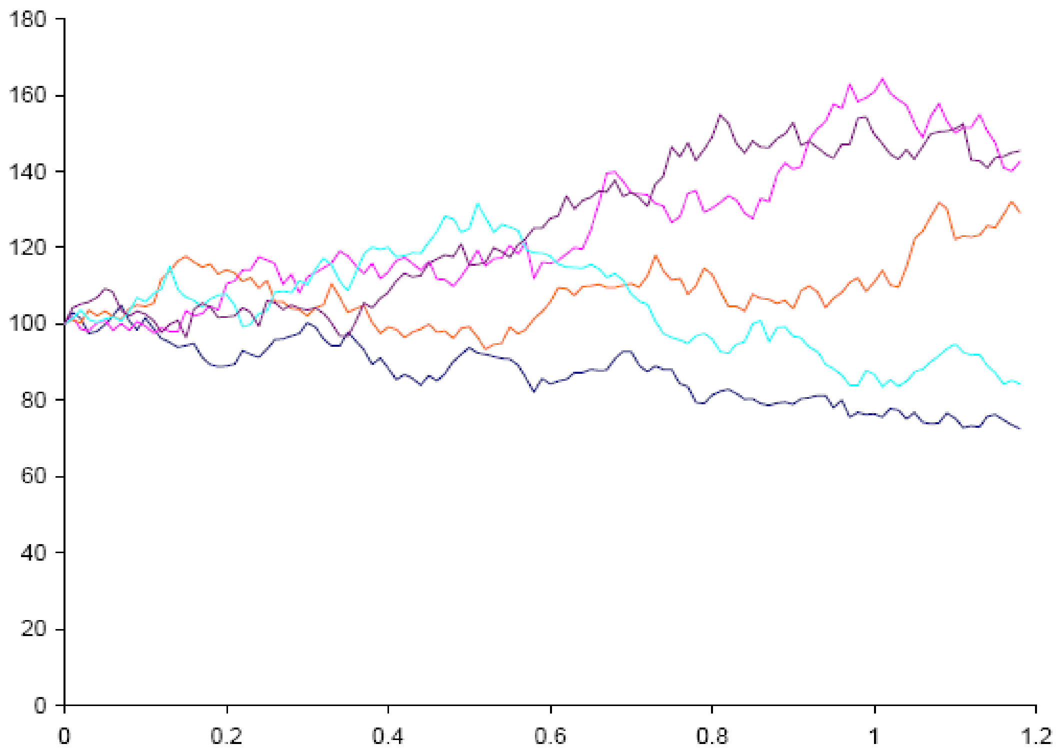
which becomes

$$S_{i+1} = S_i\left(1 + r\delta t + \sigma\phi\sqrt{\delta t}\right). \tag{11}$$

This gives us a time-stepping scheme for generating an asset price realization if we know S_0, i.e. $S(t)$ at $t = 0$. $\phi \sim N(0,1)$ is a random variable with a standard Normal distribution.

Alternatively we can use discrete form of the analytical expression (7)

$$S_{i+1} = S_i \exp\left\{\left(r - \frac{1}{2}\sigma^2\right)\delta t + \sigma\phi\sqrt{\delta t}\right\}.$$



So we now start generating random numbers. In C++ we produce uniformly distributed random variables and then use the Box Muller transformation (Polar Marsaglia method) to convert them to Gaussians.

This can also be generated on an Excel spreadsheet using the in-built random generator function RAND(). A crude (but useful) approximation for ϕ can be obtained from

$$\sum_{i=1}^{12} \mathrm{RAND}() - 6$$

where $\mathrm{RAND}() \sim U[0,1]$.

A more accurate (but slower) ϕ can be computed using $\mathrm{NORMSINV}(\mathrm{RAND}())$.

Dynamics of Vasicek Model

The Vasicek model

$$dr_t = \gamma\left(\overline{r} - r_t\right)dt + \sigma dW_t$$

is an example of a *Mean Reverting Process* - an important property of interest rates. γ refers to the *reversion rate* (also called the speed of reversion) and $\overline{r}$ denotes the *mean rate*.

γ acts like a "spring". Mean reversion means that a process which increases has a negative trend (γ pulls it down to a mean level $\overline{r}$), and when r_t decreases on average γ pulls it back up to $\overline{r}$.

In discrete time we can approximate this by writing (as earlier)

$$r_{i+1} = r_i + \gamma\left(\overline{r} - r_i\right)\delta t + \sigma\phi\sqrt{\delta t}$$

To gain an understanding of the properties of this model, look at dr in the absence of randomness

$$\begin{aligned} dr &= -\gamma\left(r - \overline{r}\right)dt \\ \int \frac{dr}{\left(r - \overline{r}\right)} &= -\gamma \int dt \\ r(t) &= \overline{r} + k\exp\left(-\gamma t\right) \end{aligned}$$

So γ controls the rate of exponential decay.

Producing Standardized Normal Random Variables

Consider the RAND() function in Excel that produces a uniformly distributed random number over 0 and 1, written $\mathbf{Unif}_{[0,1]}$. We can show that for a large number N,

$$\lim_{N\to\infty}\sqrt{\frac{12}{N}}\left(\sum_{1}^{N}\mathbf{Unif}_{[0,1]}-\tfrac{N}{2}\right)\sim N(0,1).$$

Introduce $\mathbf{U}_i$ to denote a uniformly distributed random variable over $[0,1]$ and sum up. Recall that

$$\begin{aligned}\mathbb{E}[\mathbf{U}_i] &= \tfrac{1}{2}\\ \mathbb{V}[\mathbf{U}_i] &= \tfrac{1}{12}\end{aligned}$$

The mean is then

$$\mathbb{E}\left[\sum_{i=1}^{N}\mathbf{U}_i\right]=N/2$$

so subtract off $N/2$, so we examine the variance of $\left(\sum_{1}^{N}\mathbf{U}_i-\tfrac{N}{2}\right)$

$$\begin{aligned}\mathbb{V}\left[\sum_{1}^{N}\mathbf{U}_i-\tfrac{N}{2}\right] &= \sum_{1}^{N}\mathbb{V}[\mathbf{U}_i]\\ &= N/12\end{aligned}$$

As the variance is not 1, write

$$\mathbb{V}\left[\alpha\left(\sum_{1}^{N}\mathbf{U}_i-\tfrac{N}{2}\right)\right]$$

for some $\alpha\in\mathbb{R}$. Hence $\alpha^2\frac{N}{12}=1$ which gives $\alpha=\sqrt{12/N}$ which normalises the variance. Then we achieve the result

$$\sqrt{\frac{12}{N}}\left(\sum_{1}^{N}\mathbf{U}_i-\tfrac{N}{2}\right).$$

Rewrite as

$$\frac{\left(\sum_{1}^{N}\mathbf{U}_i-N\times\tfrac{1}{2}\right)}{\sqrt{\tfrac{1}{12}}\sqrt{N}}.$$

and for $N\to\infty$ by the Central Limit Theorem we get $N(0,1)$

Generating Correlated Normal Variables

Consider two uncorrelated standard Normal variables ε_1 and ε_2 from which we wish to form a correlated pair ϕ_1, & ϕ_2 $(\sim N(0,\ 1))$, such that $\mathbb{E}[\phi_1\phi_2] = \rho$. The following scheme can be used

1. $\mathbb{E}[\varepsilon_1] = \mathbb{E}[\varepsilon_2] = 0$; $\mathbb{E}[\varepsilon_1^2] = \mathbb{E}[\varepsilon_2^2] = 1$ and $\mathbb{E}[\varepsilon_1\varepsilon_2] = 0$ $(\because\ \varepsilon_1,\ \varepsilon_2$ are uncorrelated$)$.
2. Set $\phi_1 = \varepsilon_1$ and $\phi_2 = \alpha\varepsilon_1 + \beta\varepsilon_2$ (i.e. a linear combination).
3. Now

$$\begin{aligned}
\mathbb{E}[\phi_1\phi_2] &= \rho = \mathbb{E}[\varepsilon_1(\alpha\varepsilon_1 + \beta\varepsilon_2)] \\
\mathbb{E}[\varepsilon_1(\alpha\varepsilon_1 + \beta\varepsilon_2)] &= \rho \\
\alpha\mathbb{E}\left[\varepsilon_1^2\right] + \beta\mathbb{E}[\varepsilon_1\varepsilon_2] &= \rho \to \alpha = \rho
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[\phi_2^2\right] &= 1 = \mathbb{E}\left[(\alpha\varepsilon_1 + \beta\varepsilon_2)^2\right] \\
&= \mathbb{E}\left[\alpha^2\varepsilon_1^2 + \beta^2\varepsilon_2^2 + 2\alpha\beta\varepsilon_1\varepsilon_2\right] \\
&= \alpha^2\mathbb{E}\left[\varepsilon_1^2\right] + \beta^2\mathbb{E}\left[\varepsilon_2^2\right] + 2\alpha\beta\mathbb{E}[\varepsilon_1\varepsilon_2] = 1 \\
\rho^2 + \beta^2 &= 1 \to \beta = \sqrt{1-\rho^2}
\end{aligned}$$

4. This gives $\phi_1 = \varepsilon_1$ and $\phi_2 = \rho\varepsilon_1 + \left(\sqrt{1-\rho^2}\right)\varepsilon_2$ which are correlated standardized Normal variables.

Transition Probability Density Functions for Stochastic Differential Equations

To match the mean and standard deviation of the trinomial model with the continuous-time random walk we choose the following definitions for the probabilities

$$\begin{aligned}\phi^{+}(y,t) &= \frac{1}{2}\frac{\delta t}{\delta y^{2}}\left(B^{2}(y,t)+A(y,t)\,\delta y\right),\\ \phi^{-}(y,t) &= \frac{1}{2}\frac{\delta t}{\delta y^{2}}\left(B^{2}(y,t)-A(y,t)\,\delta y\right)\end{aligned}$$

We first note that the expected value is

$$\begin{aligned}&\phi^{+}(\delta y)+\phi^{-}(-\delta y)+\left(1-\phi^{+}-\phi^{-}\right)(0)\\ =\;&\left(\phi^{+}-\phi^{-}\right)\delta y\end{aligned}$$

We already know that the mean and variance of the continuous time random walk given by

$$dy = A(y,t)\,dt + b(y,t)\,dW$$

is, in turn,

$$\begin{aligned}\mathbb{E}[dy] &= Adt\\ \mathbb{V}[dy] &= B^{2}dt.\end{aligned}$$

So to match the mean requires

$$\left(\phi^{+}-\phi^{-}\right)\delta y = A\delta t$$

The variance of the trinomial model is $\mathbb{E}[u^{2}]-\mathbb{E}^{2}[u]$ and hence becomes

$$\begin{aligned}&(\delta y)^{2}\left(\phi^{+}+\phi^{-}\right)-\left(\phi^{+}-\phi^{-}\right)^{2}(\delta y)^{2}\\ =\;&(\delta y)^{2}\left(\phi^{+}+\phi^{-}-\left(\phi^{+}-\phi^{-}\right)^{2}\right).\end{aligned}$$

We now match the variances to get

$$(\delta y)^{2}\left(\phi^{+}+\phi^{-}-\left(\phi^{+}-\phi^{-}\right)^{2}\right)=B^{2}\delta t$$

First equation gives

$$\phi^{+}=\phi^{-}+A\tfrac{\delta t}{\delta y}$$

which upon substituting into the second equation gives

$$(\delta y)^{2}\left(\phi^{-}+\alpha+\phi^{-}-\left(\phi^{-}+\alpha-\phi^{-}\right)^{2}\right)=B^{2}\delta t$$

where $\alpha = A\frac{\delta t}{\delta y}$. This simplifies to

$$2\phi^{-}+\alpha-\alpha^{2}=B^{2}\tfrac{\delta t}{(\delta y)^{2}}$$

which rearranges to give

$$\begin{aligned}\phi^{-} &= \frac{1}{2}\left(B^{2}\tfrac{\delta t}{(\delta y)^{2}}+\alpha^{2}-\alpha\right)\\ &= \frac{1}{2}\left(B^{2}\tfrac{\delta t}{(\delta y)^{2}}+\left(A\tfrac{\delta t}{\delta y}\right)^{2}-A\tfrac{\delta t}{\delta y}\right)\\ &= \frac{1}{2}\tfrac{\delta t}{(\delta y)^{2}}\left(B^{2}+A^{2}\delta t-A\delta y\right)\end{aligned}$$

δt is small compared with δy and so

$$\phi^- = \frac{1}{2}\frac{\delta t}{(\delta y)^2}\left(B^2 - A\delta y\right).$$

Then

$$\phi^+ = \phi^- + A\frac{\delta t}{\delta y} = \frac{1}{2}\frac{\delta t}{(\delta y)^2}\left(B^2 + A\delta y\right).$$

Note

$$\left(\phi^+ + \phi^-\right)(\delta y)^2 = B^2\delta t$$

Derivation of the Fokker-Planck/Forward Kolmogorov Equation

Recall that $y',\ t'$ are futures states.

We have $p(y,t;y',t') =$

$$\begin{aligned}
&\phi^-(y'+\delta y,t'-\delta t)\,p(y,t;y'+\delta y,t'-\delta t)\\
&+\big(1-\phi^-(y',t'-\delta t)-\phi^+(y',t'-\delta t)\big)\,p(y,t;y',t'-\delta t)\\
&+\phi^+(y'-\delta y,t'-\delta t)\,p(y,t;y'-\delta y,t'-\delta t)
\end{aligned}$$

Expand each of the terms in Taylor series about the point y',t' to find

$$p(y,t;y'+\delta y,t'-\delta t) = p(y,t;y',t') + \delta y\frac{\partial p}{\partial y'} + \frac{1}{2}\delta y^2\frac{\partial^2 p}{\partial y'^2} - \delta t\frac{\partial p}{\partial t'} + ...,$$

$$p(y,t;y',t'-\delta t) = p(y,t;y',t') - \delta t\frac{\partial p}{\partial t'} + ...,$$

$$p(y,t;y'-\delta y,t'-\delta t) = p(y,t;y',t') - \delta y\frac{\partial p}{\partial y'} + \frac{1}{2}\delta y^2\frac{\partial^2 p}{\partial y'^2} - \delta t\frac{\partial p}{\partial t'} + ...,$$

$$\phi^+(y'-\delta y,t'-\delta t) = \phi^+(y',t') - \delta y\frac{\partial \phi^+}{\partial y'} + \frac{1}{2}\delta y^2\frac{\partial^2 \phi^+}{\partial y'^2} - \delta t\frac{\partial \phi^+}{\partial t'} + ...,$$

$$\phi^+(y',t'-\delta t) = \phi^+(y',t') - \delta t\frac{\partial \phi^+}{\partial t'} + ...,$$

$$\phi^-(y'+\delta y,t'-\delta t) = \phi^-(y',t') + \delta y\frac{\partial \phi^-}{\partial y'} + \frac{1}{2}\delta y^2\frac{\partial^2 \phi^-}{\partial y'^2} - \delta t\frac{\partial \phi^-}{\partial t'} + ...,$$

$$\phi^-(y',t'-\delta t) = \phi^-(y',t') - \delta t\frac{\partial \phi^-}{\partial t'} + ...,$$

Substituting in our equation for $p(y,t;y',t')$, ignoring terms smaller than δt, noting that $\delta y \sim O\left(\sqrt{\delta t}\right)$, gives

$$\frac{\partial p}{\partial t'} = -\frac{\partial}{\partial y'}\left(\frac{1}{\delta y}\left(\phi^+-\phi^-\right)p\right) + \frac{1}{2}\frac{\partial^2}{\partial y'^2}\left(\left(\phi^+-\phi^-\right)p\right).$$

Noting the earlier results

$$\begin{aligned}
A &= \frac{(\delta y)^2}{\delta t}\left(\frac{1}{\delta y}\left(\phi^+-\phi^-\right)\right),\\
B^2 &= \frac{(\delta y)^2}{\delta t}\left(\phi^++\phi^-\right)
\end{aligned}$$

gives the *forward equation*

$$\frac{\partial p}{\partial t'} = \frac{1}{2}\frac{\partial^2}{\partial y'^2}\left(B^2(y',t')\,p\right) - \frac{\partial}{\partial y'}\left(A(y',t')\,p\right)$$

The initial condition used is

$$p(y,t;y',t') = \delta(y'-y)$$

As an example consider the important case of the distribution of stock prices. Given the random walk for equities, i.e. Geometric Brownian Motion

$$\frac{dS}{S} = \mu dt + \sigma dW.$$

So $A(S', t') = \mu S'$ and $B(S', t') = \sigma S'$. Hence the forward equation becomes

$$\frac{\partial p}{\partial t'} = \frac{1}{2}\frac{\partial^2}{\partial S'^2}\left(\sigma^2 {S'}^2 p\right) - \frac{\partial}{\partial S'}\left(\mu S' p\right).$$

More on this and solution technique later, but note that a transformation reduces this to the one dimensional heat equation and the *similarity reduction method* which follows is used.

The Steady-State Distribution

As the name suggests 'steady state' refers to time independent. Random walks for interest rates and volatility can be modelled with stochastic differential equations which have steady-state distributions. So in the long run, i.e. as $t' \longrightarrow \infty$ the distribution $p(y, t; y', t')$ settles down and becomes independent of the starting state y and t. The partial derivatives in the forward equation now become ordinary ones and the unsteady term $\frac{\partial p}{\partial t'}$ vanishes.

The resulting forward equation for the steady-state distribution $p_\infty(y')$ is governed by the ordinary differential equation

$$\frac{1}{2}\frac{d^2}{dy'^2}\left(B^2 p_\infty\right) - \frac{d}{dy'}\left(A p_\infty\right) = 0.$$

Example: The Vasicek model for the spot rate r evolves according to the stochastic differential equation

$$dr = \gamma(\overline{r} - r)\, dt + \sigma dW$$

Write down the Fokker-Planck equation for the transition probability density function for the interest rate r in this model.

Now using the steady-state version for the forward equation, solve this to find the **steady state** probability distribution $p_\infty(r')$, given by

$$p_\infty = \tfrac{1}{\sigma}\sqrt{\frac{\gamma}{\pi}}\exp\left(-\tfrac{\gamma}{\sigma^2}\left(r' - \overline{r}\right)^2\right).$$

Solution:

For the SDE $dr = \gamma(\overline{r} - r)\, dt + \sigma dW$ where drift $= \gamma(\overline{r} - r)$ and diffusion is σ the Fokker Planck equation becomes

$$\frac{\partial p}{\partial t'} = \frac{1}{2}\sigma^2\frac{\partial^2 p}{\partial {r'}^2} - \gamma\frac{\partial}{\partial r'}\left(\left(\overline{r} - r'\right)p\right)$$

where $p = p(r',\ t')$ is the transition PDF and the variables refer to future states. In the steady state case, there is no time dependency, hence the Fokker Planck PDE becomes an ODE with

$$\frac{1}{2}\sigma^2\frac{d^2 p_\infty}{dr^2} - \gamma\frac{d}{dr}\left(\left(\overline{r} - r\right)p_\infty\right) = 0$$

$p_\infty = p_\infty(r)$. The prime notation and subscript have been dropped simply for convenience at this stage. To solve the steady-state equation:

Integrate wrt r

$$\frac{1}{2}\sigma^2\frac{dp}{dr} - \gamma\left(\left(\overline{r} - r\right)p\right) = k$$

where k is a constant of integration and can be calculated from the conditions, that as $r \to \infty$

$$\left\{\begin{array}{c} \dfrac{dp}{dr} \to 0 \\ p \to 0 \end{array}\right. \Rightarrow k = 0$$

which gives

$$\frac{1}{2}\sigma^2\frac{dp}{dr} = -\gamma\left(\left(r - \overline{r}\right)p\right),$$

a first order variable separable equation. So

$$\begin{aligned} \frac{1}{2}\sigma^2\int\frac{dp}{p} &= -\gamma\int\left(\left(r - \overline{r}\right)\right)dr \to \\ \frac{1}{2}\sigma^2\ln p &= -\gamma\left(\frac{r^2}{2} - \overline{r}r\right) + C, \qquad C \text{ is arbitrary.} \end{aligned}$$

Rearranging and taking exponentials of both sides to give

$$p = \exp\left(-\tfrac{2\gamma}{\sigma^2}\left(\frac{r^2}{2} - \overline{r}r\right) + D\right) = E\exp\left(-\tfrac{2\gamma}{\sigma^2}\left(\frac{r^2}{2} - \overline{r}r\right)\right)$$

Complete the square to get

$$\begin{aligned} p &= E\exp\left(-\tfrac{\gamma}{\sigma^2}\left[\left(r - \overline{r}\right)^2 - \overline{r}^2\right]\right) \\ p_\infty &= A\exp\left(-\tfrac{\gamma}{\sigma^2}\left(r' - \overline{r}\right)^2\right). \end{aligned}$$

There is another way of performing the integration on the rhs. If we go back to $-\gamma\int\left(r - \overline{r}\right)dr$ and write as

$$-\gamma\int\tfrac{1}{2}\tfrac{d}{dr}\left(r - \overline{r}\right)^2 dr = \frac{-\gamma}{2}\left(r - \overline{r}\right)^2$$

to give

$$\frac{1}{2}\sigma^2\ln p = \frac{-\gamma}{2}\left(r - \overline{r}\right)^2 + C.$$

Now we know as p_∞ is a PDF

$$\begin{aligned} \int_{-\infty}^{\infty} p_\infty\, dr' &= 1 \to \\ A\int_{-\infty}^{\infty}\exp -\left(\tfrac{\gamma}{\sigma^2}\left(r' - \overline{r}\right)^2\right) dr' &= 1 \end{aligned}$$

A few (related) ways to calculate A. Now use the error function, i.e.

$$\int_{-\infty}^{\infty} e^{-x^2}dx = \sqrt{\pi}$$

So put

$$x = \sqrt{\tfrac{\gamma}{\sigma^2}}\left(r' - \overline{r}\right) \to dx = \sqrt{\tfrac{\gamma}{\sigma^2}}dr'$$

which transforms the integral above

$$\frac{A\sigma}{\sqrt{\gamma}}\int_{-\infty}^{\infty} e^{-x^2}\ dx = 1 \rightarrow A\sigma\sqrt{\frac{\pi}{\gamma}} = 1$$

therefore

$$A = \tfrac{1}{\sigma}\sqrt{\frac{\gamma}{\pi}}.$$

This allows us to finally write the steady-state transition PDF as

$$p_\infty = \tfrac{1}{\sigma}\sqrt{\frac{\gamma}{\pi}}\exp\left(-\tfrac{\gamma}{\sigma^2}\left(r' - \overline{r}\right)^2\right).$$

The *backward equation* is obtained in a similar way to the forward

$p\left(y, t; y', t'\right) =$

$$\begin{aligned} &\phi^+\left(y,t\right)p\left(y+\delta y, t+\delta t; y', t'\right) \\ &+\left(1-\phi^-\left(y,t\right)-\phi^+\left(y,t\right)\right)p\left(y, t+\delta t; y', t'\right) \\ &+\phi^-\left(y,t\right)p\left(y-\delta y, t+\delta t; y', t'\right) \end{aligned}$$

and expand using Taylor. The resulting PDE is

$$\frac{\partial p}{\partial t} + \frac{1}{2}B^2\left(y,t\right)\frac{\partial^2 p}{\partial y^2} + A\left(y,t\right)\frac{\partial p}{\partial y} = 0.$$

So the forward equation can be obtained from the backward equation using the transformation $t' = T - t$,

$$\frac{\partial p}{\partial t} = \frac{1}{2}\sigma^2 S^2\frac{\partial^2 p}{\partial S^2} + \mu S\frac{\partial p}{\partial S}.$$

Write $p = p\left(S', t\right)$ as $p = p\left(\xi, t\right)$ where $\xi = \log S'$

$$\frac{\partial p}{\partial S'} = \frac{1}{S'}\frac{\partial p}{\partial \xi};\ \frac{\partial^2 p}{\partial S'^2} = \frac{1}{S'}\frac{\partial p}{\partial \xi}$$

To solve, reduce to a 1D heat equation initially.

This can be solved with a starting condition of $S' = S$ at $t' = t$ to give the transition pdf

$$p\left(S, t; S', T\right) = \frac{1}{\sigma S'\sqrt{2\pi(t'-t)}}e^{-\left(\log(S/S') + \left(\mu - \frac{1}{2}\sigma^2\right)(t'-t)\right)^2/2\sigma^2(t'-t)}.$$