Review of Module 1

The Binomial Model

The model has made option pricing accessible to MBA students and finance practitioners preparing for the CFA®. It is a very useful tool for conveying the ideas of delta hedging and no arbitrage, in addition to the subtle concept of risk neutrality and option pricing. Here the model is considered in a slightly more mathematical way.

The basic assumptions in option pricing theory consist of two forms, key:

- Short selling allowed
- No arbitrage opportunities

and relaxable

- Frictionless markets
- Perfect liquidity
- Known volatility and interest rates
- No dividends on the underlying

The key assumptions underlying the binomial model are:

- an asset value changes only at discrete time intervals
- an asset's worth can change to one of only two possible new values at each time step.

The one period model - Replication

Another way of looking at the Binomial model is in terms of replication: we can replicate the option using only cash (or bonds) and the asset. That is, mathematically, simply a rearrangement of the earlier equations. It is, nevertheless a very important interpretation.
In one time step:

1. The asset moves from $S_0 = s$ to $S_1 = s_u$ or $S_1 = s_d$.
2. An option X pays off x_u if the asset price is s_u and x_d if the price is s_d.
3. There is a bond market in which a pound invested today is continuously compounded at a constant (risk-free) rate r and becomes e^r, one time-step later.

Now consider a portfolio of ψ bonds and ϕ assets which at time $t = 0$, will have an initial value of

$$V_0 = \phi S_0 + \psi$$

Now with this money we can buy or sell bonds or stocks in order to obtain a new portfolio at time-step 1.

Can we construct a hedging strategy which will guarantee to pay off the option, whatever happens to the asset price?

The Hedging Strategy

We arrange the portfolio so that its value is exactly that of the required option pay-out at the terminal time regardless of whether the stock moves up or down.

This is because having two unknowns $\phi,\ \psi$, the amount of stock and bond, and we wish to match the two possible terminal values, $x_u,\ x_d$, the option payoffs. Thus we need to have

$$\begin{aligned} x_u &= \phi s_u + \psi e^r, \\ x_d &= \phi s_d + \psi e^r. \end{aligned}$$

Solving for $\phi,\ \psi$ we have

$$\begin{aligned} \phi &= \frac{x_u - x_d}{s_u - s_d} \\ \psi &= e^{-r}\frac{x_d s_u - x_u s_d}{s_u - s_d} \end{aligned}$$

This is a *hedging strategy.*

At time step 1, the value of the portfolio is

$$V_1 = \left\{ \begin{array}{ll} x_u & \text{if } S_1 = s_u \\ x_d & \text{if } S_1 = s_d \end{array} \right\}$$

This is the option payoff. Thus, given $V_0 = \phi S_0 + \psi$ we can construct the above portfolio which has the same payoff as the option. Hence the price for the option must be V_0. Any other price would allow arbitrage as you could play this hedging strategy, either buying or selling the option, and make a guaranteed profit.

Thus the fair, arbitrage-free price for the option is given by

$$\begin{aligned} V_0 &= (\phi S_0 + \psi) \\ &= \frac{x_u - x_d}{s_u - s_d} s + e^{-r}\frac{x_d s_u - x_u s_d}{s_u - s_d} \\ &= e^{-r}\left(\frac{e^r s - s_d}{s_u - s_d} x_u + \frac{s_u - e^r s}{s_u - s_d} x_d\right). \end{aligned}$$

Define

$$q = \frac{e^r s - s_d}{s_u - s_d},$$

then we conclude that

$$V_0 = e^{-r}\left(q x_u + (1-q)\, x_d\right)$$

where

$$0 \le q \le 1.$$

We can think of q as a probability induced by insistence on no-arbitrage, i.e. the so-called *risk-neutral probability.* It has nothing to do with the real probabilities of s_u and s_d occurring; these are p and $1-p$, in turn.

The option price can be viewed as the discounted expected value of the option pay-off with respect to the probabilities q,

$$\begin{aligned} V_0 &= e^{-r}\left(q x_u + (1-q)\, x_d\right) \\ &= \mathbb{E}_q\left[e^{-r} X\right]. \end{aligned}$$

The fact that the risk neutral/fair value (or $q-$value) of a call is less than the expected value of the call (under the real probability p), is not a puzzle.

Pricing a call using the real probability, p, you will probably make a profit, but you might also might make a loss. Pricing an option using the risk-neutral probability, q, you will certainly make neither a profit nor a loss.

Assume an asset which has value S and during a time step δt can either rise to uS or fall to vS with $0 < v < 1 < u$. So as earlier probabilities of a rise and fall in turn are p and $1-p$.

$$\underbrace{\begin{array}{ccc} & & uS \\ & \nearrow & \\ S & & \\ & \searrow & \\ & & vS \end{array}}_{\delta t}$$

$$\underbrace{\begin{array}{ccc} & & V^{+} - \Delta uS \\ & \nearrow & \\ V - \Delta S & & \\ & \searrow & \\ & & V^{-} - \Delta vS \end{array}}_{\delta t}$$

Also set $uv = 1$ so that after an up and down move, the asset returns to S. Hence a *recombining* tree.

To implement the Binomial model we need a model for asset price evolution to predict future possible spot prices. So use

$$\delta S = \mu S \delta t + \sigma S \phi \sqrt{\delta t},$$

i.e. discrete version of GBM. The 3 constants u, v, p are chosen to give the binomial model the same drift and diffusion as the SDE. For the correct drift, choose

$$pu + (1-p)\, v = e^{\mu \delta t} \tag{a}$$

and for the correct standard deviation set

$$pu^2 + (1-p)\, v^2 = e^{\left(2\mu + \sigma^2\right)\delta t} \tag{b}$$

$u\,(a) + v\,(a)$ gives

$$(u+v)\, e^{\mu\ \delta t} = pu^2 + uv - puv + pvu + v^2 - pv^2.$$

Rearrange to get

$$(u+v)\, e^{\mu \delta t} = pu^2 + + (1-p)\, v^2 + uv$$

and we know from (b) that $pu^2 + (1-p)\, v^2 = e^{\left(2\mu+\sigma^2\right)\delta t}$ and $uv = 1$. Hence we have

$$\begin{aligned} (u+v)\, e^{\mu\delta t} &= e^{\left(2\mu+\sigma^2\right)\delta t} + 1 \Rightarrow \\ (u+v) &= e^{-\mu\delta t} + e^{\left(\mu+\sigma^2\right)\delta t}. \end{aligned}$$

Now recall that the quadratic equation $ax^2 + bx + c = 0$ with roots α and β has

$$\alpha + \beta = -\frac{b}{a};\ \alpha\beta = \frac{c}{a}.$$

We have

$$\begin{aligned}(u+v) &= e^{-\mu\delta t}+e^{\left(\mu+\sigma^2\right)\delta t}\equiv-\frac{b}{a}\\ uv &= 1\equiv\frac{c}{a}\end{aligned}$$

hence u and v satisfy

$$(x-u)(x-v)=0$$

to give the quadratic

$$x^2-(u+v)\,x+uv=0\Rightarrow$$

$$x=\frac{(u+v)\pm\sqrt{(u+v)^2-4uv}}{2}$$

so with $u>1$

$$u=\frac{1}{2}\left(e^{-\mu\delta t}+e^{\left(\mu+\sigma^2\right)\delta t}\right)+\frac{1}{2}\sqrt{\left(e^{-\mu\delta t}+e^{(\mu+\sigma^2)\delta t}\right)^2-4}$$

In this model, the hedging argument gives

$$V^+-\Delta uS=V^--\Delta vS$$

which leads to $\Delta=\dfrac{V^+-V^-}{(u-v)\,S}$. Because all other terms are known choose Δ to eliminate risk.

We know tomorrow's option value therefore price today is tomorrow's value discounted for interest rates

$$V-\Delta S=\frac{1}{1+r\delta t}\left(V^+-\Delta uS\right)$$

so $(1+r\delta t)(V-\Delta S)=V^+-\Delta uS$ and replace using the definition of Δ above

$$(1+r\delta t)\,V=V^+\left(\frac{-v+1+r\delta t}{(u-v)}\right)+V^-\left(\frac{u-1-r\delta t}{(u-v)}\right)$$

where the risk-neutral probabilities are

$$\begin{aligned}q &= \frac{-v+1+r\delta t}{(u-v)}\\ 1-q &= \frac{u-1-r\delta t}{(u-v)}.\end{aligned}$$

So $(1+r\delta t)\,V=V^+q+V^-(1-q)\,.$

Finally we have

$$V=\frac{V^+-V^-}{(u-v)}+\frac{uV^--vV^+}{(1+r\delta t)(u-v)}$$

$$q=\frac{e^{r\delta t}-v}{(u-v)}$$

The Continuous Time Limit

Performing a Taylor expansion around $\delta t = 0$ we have

$$\begin{aligned}
u \sim \frac{1}{2}\left(\left(1-\mu\delta t+\cdots\right)+\left(1+\left(\mu+\sigma^2\right)\delta t+\cdots\right)\right)+\frac{1}{2}\left(e^{-2\mu\delta t}+2e^{\sigma^2\delta t}+e^{2\left(\mu+\sigma^2\right)\delta t}-4\right)^{\frac{1}{2}} \\
=\left(1+\frac{1}{2}\sigma^2\delta t+\cdots\right)+\frac{1}{2}\left(1-2\mu\delta t+2+2\sigma^2\delta t+1+2\mu\delta t+2\sigma^2\delta t-4+\cdots\right) \\
=\left(1+\frac{1}{2}\sigma^2\delta t+\cdots\right)+\frac{1}{2}\left(4\sigma^2\delta t+\cdots\right)^{\frac{1}{2}}
\end{aligned}$$

Ignoring the terms of order $\delta t^{\frac{3}{2}}$and higher we get the result

$$=\left(1+\sigma\delta t^{\frac{1}{2}}+\frac{1}{2}\sigma^2\delta t\right)+\cdots$$

Since $uv = 1$ this implies that $v = u^{-1}$. Using the expansion for u obtained earlier we have

$$\begin{aligned}
v = \left(1+\sigma\delta t^{\frac{1}{2}}+\frac{1}{2}\sigma^2\delta t+\cdots\right)^{-1} \\
=\left(1+\sigma\delta t^{\frac{1}{2}}\left(1+\frac{1}{2}\sigma\delta t^{\frac{1}{2}}\right)\cdots\right)^{-1} \\
=\left(1-\sigma\delta t^{\frac{1}{2}}\left(1+\frac{1}{2}\sigma\delta t^{\frac{1}{2}}\right)+\left(\sigma\delta t^{\frac{1}{2}}\left(1+\frac{1}{2}\sigma\delta t^{\frac{1}{2}}\right)^2+\cdots\right)\right) \\
=1-\sigma\delta t^{\frac{1}{2}}-\frac{1}{2}\sigma^2\delta t+\sigma^2\delta t+\cdots \\
=1-\sigma\delta t^{\frac{1}{2}}+\frac{1}{2}\sigma^2\delta t
\end{aligned}$$

So we have

$$\begin{aligned}
u &\sim 1+\sigma\sqrt{\delta t}+\frac{1}{2}\sigma^2\delta t \\
v &\sim 1-\sigma\sqrt{\delta t}+\frac{1}{2}\sigma^2\delta t
\end{aligned}$$

So to summarise we can write

$$\begin{aligned}
u &= e^{\sigma\sqrt{\delta t}} \\
v &= e^{-\sigma\sqrt{\delta t}} \\
q &= \frac{e^{r\delta t}-v}{(u-v)}
\end{aligned}$$

and use these to build the asset price tree using u and v, and then value the option backwards from T using

$$e^{r\delta t}V\left(S,t\right)=qV\left(uS,t+\delta t\right)+\left(1-q\right)V\left(vS,t+\delta t\right)$$

and at each stage the headge ratio Δ is obtained using

$$\Delta=\frac{V^+-V^-}{(u-v)\,S}=\frac{V\left(uS,t+\delta t\right)-V\left(vS,t+\delta t\right)}{(u-v)\,S}$$

Note that

$$\Delta = \frac{V^+ - V^-}{(u-v)\,S} \sim \frac{2\sigma\sqrt{\delta t}S\frac{\partial V}{\partial S}}{2\sigma\sqrt{\delta t}S} = \frac{\partial V}{\partial S}$$

Now expand

$$\begin{aligned} V^+ &= V\left(uS, t+\delta t\right) \sim V + \delta t\frac{\partial V}{\partial t} + \sigma\sqrt{\delta t}S\frac{\partial V}{\partial S} + \tfrac{1}{2}\sigma^2\delta t S^2\frac{\partial^2 V}{\partial S^2}, \\ V^- &= V\left(vS, t+\delta t\right) \sim V + \delta t\frac{\partial V}{\partial t} - \sigma\sqrt{\delta t}S\frac{\partial V}{\partial S} + \tfrac{1}{2}\sigma^2\delta t S^2\frac{\partial^2 V}{\partial S^2}. \end{aligned}$$

Then

$$\begin{aligned} V &= \frac{V^+ - V^-}{(u-v)} + \frac{uV^- - vV^+}{(1+r\delta t)\,(u-v)} \\ &= \frac{2\sigma\sqrt{\delta t}S}{2\sigma\sqrt{\delta t}}\frac{\partial V}{\partial S} + \frac{\left(1+\sigma\sqrt{\delta t}\right)V^- - \left(1-\sigma\sqrt{\delta t}\right)V^+}{(1+r\delta t)\,2\sigma\sqrt{\delta t}} \end{aligned}$$

Rearranging to give

$$(1+r\delta t)\,2\sigma\sqrt{\delta t}V = 2\sigma\sqrt{\delta t}S\,(1+r\delta t)\,\frac{\partial V}{\partial S} + \left(V^- - V^+\right) + \sigma\sqrt{\delta t}\left(V^- + V^+\right),$$

and so

$$\begin{aligned} (1+r\delta t)\,2\sigma\sqrt{\delta t}V &= 2\sigma\sqrt{\delta t}S\,(1+r\delta t)\,\frac{\partial V}{\partial S} - 2\sigma\sqrt{\delta t}S\frac{\partial V}{\partial S} + \\ & \quad 2\sigma\sqrt{\delta t}\left(V + \tfrac{1}{2}\sigma^2\delta t S^2\frac{\partial^2 V}{\partial S^2} + \delta t\frac{\partial V}{\partial t}\right), \end{aligned}$$

$$(1+r\delta t)\,V = S\,(1+r\delta t)\,\frac{\partial V}{\partial S} - S\frac{\partial V}{\partial S} + \left(V + \tfrac{1}{2}\sigma^2\delta t S^2\frac{\partial^2 V}{\partial S^2} + \delta t\frac{\partial V}{\partial t}\right),$$

divide through by δt and allow $\delta t \longrightarrow 0$

$$rV = rS\frac{\partial V}{\partial S} + \tfrac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2} + \frac{\partial V}{\partial t}$$

and hence the Black-Scholes Equation.

Probability

Probability theory provides the necessary structure to model the uncertainty that is central to finance and is the chief reason for its powerful influence in mathematical finance. Any formal discussion of random variables requires defining the triple $(\Omega, \mathcal{F}, \mathbb{P})$; as it forms the foundation of the *probabilistic universe.* This three-tuple is called a *probability space* and comprises of

1. the sample space Ω
2. the filtration $\mathcal{F}$
3. the probability measure $\mathbb{P}$

Basic set theoretic notions have special interpretations in probability theory. Here are some

- The **complement** in Ω of the event A, written A^c is interpreted as "not A" and occurs iff A does not occur.
- The **union** $A \cup B$ of two events A and B is the event "at least one of A or B occurs".
- The **intersection** $A \cap B$ of two events A and B is the event "both A and B occur". Events A and B are said to be **mutually exclusive** if they are disjoint, $A \cap B = \emptyset$, and so both cannot occur together.
- The **inclusion** relation $A \subseteq B$ means the "occurrence of A implies the occurrence of B".

Example The daily closing price of a risky asset, e.g. share price on the FTSE100. Over the course of a year (252 business days)

$$\Omega = \{S_1, S_2, S_3, \ldots, S_{252}\}$$

We could define an event e.g. $\Psi = \{S_i : S_i \geq 110\}$

Outcomes of experiments are not always numbers, e.g. 2 heads appearing; picking an ace from a deck of cards, or the coin flipping example above. We need some way of assigning real numbers to each random event. Random variables assign numbers to events. Thus a *random variable* (RV) X is a function which maps from the sample space Ω to the set of real numbers

$$X : \omega \in \Omega \to \mathbb{R},$$

i.e. it associates a number $X(\omega)$ with each outcome ω. A more robust definition will follow.

Consider the example of tossing a coin and suppose we are paid £1 for each head and we lose £1 each time a tail appears. We know that $\mathbb{P}(\mathrm{H}) = \mathbb{P}(\mathrm{T}) = \frac{1}{2}$. So now we can assign the following outcomes

$$\mathbb{P}(1) = \frac{1}{2};\ \mathbb{P}(-1) = \frac{1}{2}$$

Mathematically, if our random variable is X, then

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

or using the notation above $X : \omega \in \{\text{H,T}\} \rightarrow \{-1, 1\}$.

Returning to the coin tossing game we see the sample space Ω has two events: $\omega_1 =$ Head; $\omega_2 =$ Tail.

So now

$$\Omega = \{\omega_1, \omega_2\}$$

And the P&L from this game is a RV X defined by

$$\begin{aligned} X(\omega_1) &= +1 \\ X(\omega_2) &= -1 \end{aligned}$$

$\Omega = \{\omega_1, \omega_2\} \Rightarrow \Omega \subset 2^{\Omega} = \{\emptyset, \{-1\}, \{+1\}, \{-1, +1\}\}$.

In a multi-period market, information about the market is revealed in stages. The n period Binomial model demonstrates the way this information becomes available.

Some events may be completely determined by the end of the first trading period, others by the end of the second or third, and others will only be available at the termination of all trading. These events can be classified in the following way; consider time $t \leq T$, define

$$\mathcal{F}_t = \{\text{all events determined in the first } t \text{ trading periods}\}$$

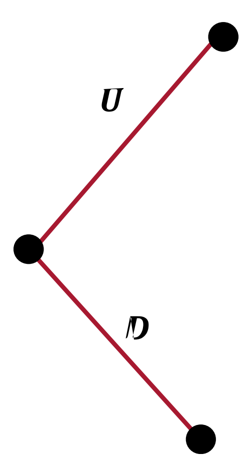
The binomial stock price model is a discrete time stochastic model of a stock price process in which a fictitious coin is tossed and the stock price dynamics depend on the outcome of the coin tosses e.g. a Head means the stock rises by one unit, a tail means the stock falls by that same amount. Start by introducing some new probabilistic terminology and concepts.

Suppose $\mathbb{T} := \{0, 1, 2, ..., n\}$ represents a discrete time set.

The sample space $\Omega = \Omega_n$, the set of all outcomes of n coin tosses; each sample point $\omega \in \Omega$ is of length n, written as $\omega = \omega_1\omega_2....\omega_n$, where each $\{\omega_t : t \in \mathbb{T}\}$ is either U (due to a Head) or D (due to a Tail), representing the outcome of the t^{th} coin toss. So e.g. three coin tosses would give a sample path $\omega = \omega_1\omega_2\omega_3$ of length 3.

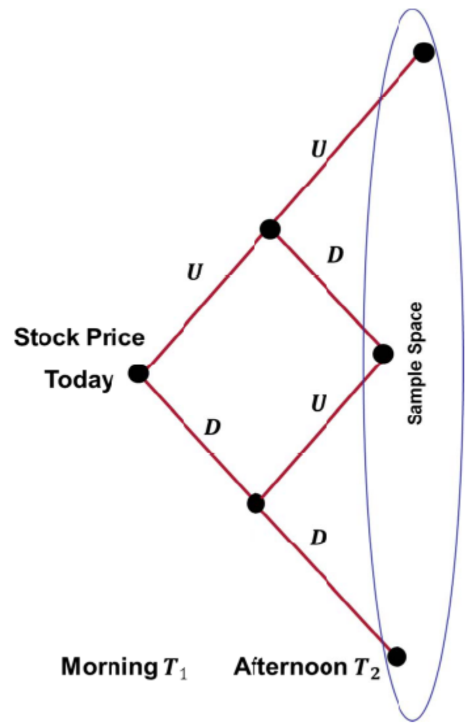
We are interested in a *stochastic process* due to the dynamic nature of asset prices.

Suppose before the markets open we guess the possible outcomes of the stock price, this will give us our sample space. The sample path will tell us what just happened. Consider a stock price which over the next time step can go up U or go down D.



$$
\begin{aligned}
\Omega_1 &= \{U, D\}. \\
& \quad 2^1 \text{ outcomes} \\
\omega &= \omega_1 \text{ length } 1
\end{aligned}
$$

Then a two time period model looks like



So the sample space at the end of two time periods is

$$\Omega_2 = \{UU, UD, DU, DD\}.$$

$$
\begin{aligned}
& \quad 2^2 \text{ outcomes} \\
\omega &= \omega_1\omega_2 \text{ length } 2
\end{aligned}
$$

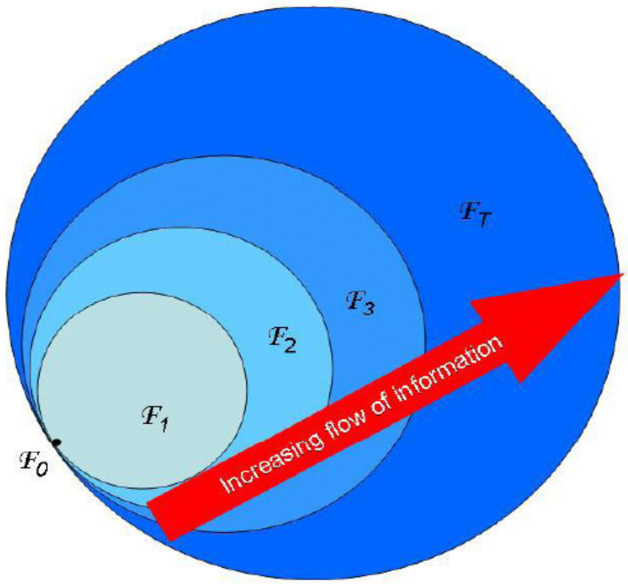
For this experiment a **sample path** or trajectory would be one realisation e.g. DU or DD. Generally in probability theory, the sample space is of greater interest. As the number of time periods becomes larger and larger it becomes increasingly difficult to track all of the possible outcomes and corresponding sample space generated through time, i.e. $\Omega_1, \Omega_2, \Omega_3, ..., \Omega_t, \Omega_{t+1,}....$

The filtration, F, is an indication of how an increasing family of events builds up over time as more results become available, it is much more than just a family of events. The filtration, F is a set formed of all possible combinations of events $A \subset \Omega$, their unions and complements. So for example if we want to know what events can occur we are also interested in what cannot happen. The filtration F is an object in Measure Theory called a $\sigma-$algebra (also called a $\sigma-$field). $\sigma-$algebras can be interpreted as records of information. Measure theory was brought to probability by Kolmogorov.

Now let F be the non-empty set of all subsets of Ω; then F $\left(\subset 2^{\Omega}\right)$ is a $\sigma-$algebra (also called a $\sigma-$field), that is, a collection of subsets of Ω with the properties:

1. $\emptyset \in$ F
2. If $A \subset$ F then $A^c \subset$ F (closed under complements)
3. If the sequence $A_i \subset$ F $\forall i \in \mathbb{N} \Longrightarrow \bigcup_{i=1}^{\infty} A_i \subset$ F (closed under countable unions)

The second property also implies that $\Omega \subset$ F. In addition $\bigcap_{i=1}^{\infty} A_i \subset$ F. The pair $(\Omega,\ \mathcal{F})$ is called a *measurable space.*



Key Fact: For $0 \leq t_1 \leq t_2 \leq\ \ldots\ \leq T$,

$$\mathcal{F}_{t_1} \subseteq \mathcal{F}_{t_2} \subseteq \ldots \subseteq \mathcal{F}_T \equiv \mathcal{F}$$

Since we consider that information gets constantly recorded and accumulates up until the end of the experiment T without ever getting lost or forgotten, it is only logical that with the passage of time the filtration increases.

In general it is very difficult to describe explicitly the filtration. In the case of (say) the binomial model, this can be done. Consider a 3-period binomial model. At the end of each time period new information becomes available, allowing us to predict the stock price trajectory.

Example Consider a 3-period binomial model. At the end of each period, new information becomes available to help us predict the actual stock trajectory. So take $n = 3$; $\Omega = \Omega_3$, given by the finite set

$$\Omega_3 = \{UUU, UUD, UDU, UDD, DUU, DUD, DDU, DDD\},$$

the set of all possible outcomes of three coin tosses. At time $t = 0$, before the start of trading we only have the **trivial filtration**

$$\mathcal{F}_0 = \{\Omega, \emptyset\}$$

since we do not have any information regarding the trajectory of the stock. The trivial $\sigma-$ algebra $\mathcal{F}_0$ contains no information: knowing whether the outcome ω of the two tosses is in $\emptyset$; (it is not) and whether it is in Ω (it is) tells you nothing about ω, in accordance with the idea that at time zero one knows nothing about the eventual outcome ω of the three coin tosses. All one can say is that $\omega \notin \emptyset$ and $\omega \in \Omega$ and so $\mathcal{F}_0 = \{\Omega, \emptyset\}$.

Now define the following two subsets of Ω :

$$A_U = \{UUU, UUD, UDU, UDD\}, \ A_D = \{DUU, DUD, DDU, DDD\}.$$

We see A_U is the subset of outcomes where a Head appears on the first throw, A_D is the subset of outcomes where a Tail lands on the first throw. After the first trading period $t = 1$,(11am) we know whether the initial move was an up move or down move. Hence

$$\mathcal{F}_1 = \{\Omega, \emptyset, A_U, A_D\}$$

Define also

$$\begin{aligned} A_{UU} &= \{UUU, UUD\}, \ A_{UD} = \{UDU, UDD\}, \\ A_{DU} &= \{DUU, DUD\}, \ A_{DD} = \{DDU, DDD\} \end{aligned}$$

corresponding to the events that the first two coin tosses result in HH, HT, TH, TT respectively. This is the information we have at the end of the 2nd trading period $t = 2$,(1 pm). This means at the end of the second trading period we have accumulated increasing information. Hence

$$\mathcal{F}_2 = \{\Omega, \emptyset, A_U, A_D, A_{UU}, A_{UD}, A_{DU}, A_{DD} + \text{all unions of these}\},$$

which can be written as follows

$$\begin{aligned} \mathcal{F}_2 = \ & \{\Omega, \emptyset, A_U, A_D, A_{UU}, A_{UD}, A_{DU}, A_{DD} \\ & A_{UU} \cup A_{DU}, A_{UU} \cup A_{DD}, A_{UD} \cup A_{DU}, A_{UD} \cup A_{DD} \\ & A_{UU}^c, A_{UD}^c, A_{DU}^c, A_{DD}^c\} \end{aligned}$$

We see

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2$$

Then $\mathcal{F}_2$ is a $\sigma-$algebra which contains the "information of the first two tosses" of "the information up to time 2". This is because, if you know the outcome of the first two tosses, you can say whether the outcome $\omega \in \Omega$ of all three tosses satisfies $\omega \in A$ or $\omega \notin A$ for each $A \in \mathcal{F}_2$.

Similarly, $\mathcal{F}_3 \equiv \mathcal{F}$, the set of all subsets of Ω, contains full information about the outcome of all three tosses. The sequence of increasing $\sigma-$ algebras $\mathbb{F} = \{\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3\}$ is a filtration.

Adapted Process A stochastic process S_t is said to be *adapted* to the filtration $\mathcal{F}_t$ (or $\mathcal{F}_t$ measurable or $\mathcal{F}_t -$ adapted) if the value of S at time t is known given the information set $\mathcal{F}_t$.

We place a *probability measure* $\mathbb{P}$ on $\{\Omega, \mathcal{F}\}$. $\mathbb{P}$ is a special type of "function", called a measure which assigns probabilities to subsets (i.e. the outcomes); the theory also comes from Measure Theory. Whereas cumulative density functions (CDF) are defined on intervals such as $\mathbb{R}$; probability measures are defined on general sets, giving greater power, generalisation and flexibility. A probability measure $\mathbb{P}$ is a function mapping $\mathbb{P} : \mathcal{F} \longrightarrow [0,1]$ with the properties

(i) $\mathbb{P}(\Omega) = 1$,

(ii) if $A_1, A_2, ...$ is a sequence of disjoint sets in $\mathcal{F}$, then $\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mathbb{P}(A_k)$.

Example Recall the usual coin toss game with the earlier defined results. As the outcomes are equiprobable the probability measure defined as $\mathbb{P}(\omega_1) = \frac{1}{2} = \mathbb{P}(\omega_2)$.

The interpretation is that for a set $A \in$F there is a probability in $[0,1]$ that the outcome of a random experiment will lie in the set A. We think of $\mathbb{P}(A)$ as this probability. The $A \in$F is called an *event*. For $A \in$F we can define

$$\mathbb{P}(A) := \sum_{\omega \in A} \mathbb{P}(\omega), \qquad (*)$$

as A has finitely many elements. Letting the probability of H on each coin toss be $p \in (0,1)$, so that the probability of T is $q = 1 - p$. For each $\omega = (\omega_1, \omega_2, \ldots \omega_n) \in \Omega$ we define

$$\mathbb{P}(\omega) := p^{\text{Number of H in } \omega} q^{\text{Number of T in } \omega}.$$

Then for each $A \in$F we define $\mathbb{P}(A)$ according to $(*)$.

In the finite coin toss space, for each $t \in \mathbb{T}$ let F$_t$ be the $\sigma-$ algebra generated by the first t coin tosses. This is a $\sigma-$ algebra which encapsulates the information one has if one observes the outcome of the first t coin tosses (but not the full outcome ω of all n coin tosses). Then F$_t$ is composed of all the sets A such that F$_t$ is indeed a $\sigma-$ algebra, and such that if the outcome of the first t coin tosses is known, then we can say whether $\omega \in A$ or $\omega \notin A$, for each $A \in$ F$_t$. The increasing sequence of $\sigma-$ algebras $(\mathcal{F}_t)_{t \in \mathbb{T}}$ is an example of a *filtration*. We use this notation when working in continuous time.

When moving to continuous time we will write $(\mathcal{F}_t)_{t \in [0,T]}$.

If we are concerned with developing a more measure theory based rigorous approach then working structures such as $\sigma-$algebras becomes more important - so we do not need to worry too much about this in our financial mathematics setting.

We can compute the probability of any event. For instance,

$$\begin{aligned} \mathbb{P}(A_U) &= \mathbb{P}(\text{H on first toss}) = \mathbb{P}\{UUU, UUD, UDU, UDD\} \\ &= p^3 + 2p^2 q + pq^2 \\ &= p, \end{aligned}$$

and similarly $\mathbb{P}(A_T) = q$. This agrees with the mathematics and our intuition.

Explanation of probability measure: If the number of basic events is very large we may prefer to think of a continuous probability distribution. As the number of discrete events tends to infinity, the probability of any individual event usually tends to zero. In terms of random variables, the probability that the random variable X takes a given value tends to zero.

So, the individual probabilities p_i are no longer useful. Instead we have a probability density function $p(x)$ with the property that

$$\Pr(x \leq X \leq x + dx) = p(x)\, dx$$

for any infinitesimal interval of length dx (think of this as a limiting process starting with a small interval whose l
It is also called a density because it is the probability of finding X on an interval of length dx divided by the length of the interval. Recall that the following are analogous

$$\begin{aligned} \int_{-\infty}^{\infty} p(x)\, dx &= 1 \\ \sum_i p_i &= 1. \end{aligned}$$

The (cumulative) distribution function of a random variable is defined by

$$P(x) = \Pr(X \leq x).$$

It is an increasing function of x with $P(-\infty) = 0$ and $P(\infty) = 1$; note that $0 \leq P(x) \leq 1$. It is related to the density function by

$$p(x) = \frac{dP(x)}{dx}$$

provided that $P(x)$ is differentiable. Unlike $P(x)$, $p(x)$ may be unbounded or have singularities such as delta functions.

$\mathbb{P}$ is the probability measure, a special type of "function", called a measure, assigning probabilities to subsets (i.e. the outcomes); the mathematics emanates from Measure Theory. Probability measures are similar to cumulative density functions (CDF); the chief difference is that where PDFs are defined on intervals (e.g. $\mathbb{R}$), probability measures are defined on general sets. We are now concerned with mapping subsets on to $[0,1]$. The following definition of the expectation has been used

$$\begin{aligned}
\mathbb{E}[h(X)] &= \int_{\mathbb{R}} h(x)\, p(x)\, dx \\
&= \int_{\mathbb{R}} h(x)\, dP(x)
\end{aligned}$$

We now write as a Lebesgue integral with respect to the measure $\mathbb{P}$

$$\mathbb{E}^{\mathbb{P}}[h(X(\omega))] = \int_{\Omega} h(\omega)\, \mathbb{P}(d\omega).$$

So integration is now done over the sample space (and not intervals).

If $\{W_t : t \in [0,T]\}$ is a Brownian motion or any general stochastic process $\{S_n : n = 0, ..., N\}$. The probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is the set of all paths (continuous functions) and $\mathbb{P}$ is the probability of each path.

Example Recall the usual coin toss game with the earlier defined results. As the outcomes are equiprobable the probability measure defined as $\mathbb{P}(\omega_1) = \frac{1}{2} = \mathbb{P}(\omega_2)$.

There is a very powerful relation between expectations and probabilities. In our formula for the expectation, choose $f(X)$ to be the *indicator function* $\mathbf{1}_{x\in A}$ for a subset $A \subset \Omega$ defined

$$\mathbf{1}_{x\in A} = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

i.e. when we are in A, the indicator function returns 1.

The expectation of the indicator function of an event is the probability associated with this event:

$$\begin{aligned}
\mathbb{E}[\mathbf{1}_{X\in A}] &= \int_{\Omega} \mathbf{1}_{x\in A} d\mathbb{P} \\
&= \int_A d\mathbb{P} + \int_{\Omega\backslash A} d\mathbb{P} \\
&= \int_A d\mathbb{P} \\
&= \mathbb{P}(A)
\end{aligned}$$

which is simply the probability that the outcome $X \in A$.

Conditional Expectations

What makes a conditional expectation different (from an unconditional one) is information (just as in the case of conditional probability). In our probability space, $(\Omega, \mathcal{F}, \mathbb{P})$ information is represented by the filtration $\mathcal{F}$; hence a conditional expectation with respect to the (usual information) filtration seems a natural choice.

$$Y = \mathbb{E}\left[X \middle| \mathcal{F} \right]$$

is the expected value of the random variable conditional upon the filtration set $\mathcal{F}$. In general

- In general Y will be a random variable
- Y will be adapted to the filtration $\mathcal{F}$.

Conditional expectations have the following useful properties: If X, Y are integrable random variables and a, b are constants then

1. **Linearity**:
$$\mathbb{E}\left[aX + bY \middle| \mathcal{F} \right] = a\mathbb{E}\left[X \middle| \mathcal{F} \right] + b\mathbb{E}\left[Y \middle| \mathcal{F} \right]$$

2. **Tower Property (i.e. Iterated Expectations)**: if $\mathcal{G} \subset \mathcal{F}$
$$\mathbb{E}\left[\mathbb{E}\left[X \middle| \mathcal{F} \right] \middle| \mathcal{G} \right] = \mathbb{E}\left[X \middle| \mathcal{G} \right].$$
This property states that if taking iterated expectations with respect to several levels of information, we may as well take a single expectation subject to the smallest set of available information. The special case is
$$\mathbb{E}\left[\mathbb{E}\left[X \middle| \mathcal{F} \right] \right] = \mathbb{E}\left[X \right].$$

3. **Taking out what is known**: X is $\mathcal{F}-$ adapted, then the value of X is know once we know $\mathcal{F}$. Therefore
$$\mathbb{E}\left[X \middle| \mathcal{F} \right] = X.$$
and hence by extension if X is $\mathcal{F}-$ measurable, but not Y then
$$\mathbb{E}\left[XY \middle| \mathcal{F} \right] = X\mathbb{E}\left[Y \middle| \mathcal{F} \right].$$

4. **Independence**: X is independent of $\mathcal{F}$, then knowing $\mathcal{F}$ is of no use in predicting X
$$\mathbb{E}\left[X \middle| \mathcal{F} \right] = \mathbb{E}\left[X \right].$$

5. **Positivity**: If $X \geq 0$ then $\mathbb{E}\left[X \middle| \mathcal{F} \right] \geq 0$.

6. **Jensen's inequality**: Let f be a convex function, then
$$f\left(\mathbb{E}\left[X \middle| \mathcal{F} \right] \right) \leq f\mathbb{E}\left[f\left(X \right) \middle| \mathcal{F} \right]$$

Solving the Diffusion Equation

The Heat/Diffusion equation

Consider the equation

$$\frac{\partial u}{\partial t}=c^2\frac{\partial^2 u}{\partial x^2}$$

for the unknown function $u=u\left(x,t\right)$, c^2 is a positive constant. The idea is to obtain a solution in terms of Gaussian curves. So $u\left(x,t\right)$ represents a probability density.

We assume a solution of the following form exists:

$$u\left(x,t\right)=t^{-1/2}f\left(\frac{x}{t^{1/2}}\right)$$

and the non-dimensional variable

$$\xi=\frac{x}{t^{1/2}}$$

which allows us to obtain the following derivatives

$$\frac{\partial \xi}{\partial x}=t^{-1/2};\quad \frac{\partial \xi}{\partial t}=-\frac{1}{2}xt^{-3/2}$$

we can now say

$$u\left(x,t\right)=t^{-1/2}f\left(\xi\right)$$

therefore

$$\frac{\partial u}{\partial x}=\frac{\partial u}{\partial \xi}\frac{\partial \xi}{\partial x}=t^{-1/2}f'\left(\xi\right)\times\frac{1}{t^{1/2}}=t^{-1}f'\left(\xi\right)$$

$$\frac{\partial^2 u}{\partial x^2}=\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}\right)=\frac{\partial}{\partial x}\left(t^{-1}f'\left(\xi\right)\right)=t^{-3/2}f''\left(\xi\right)$$

$$\begin{aligned}\frac{\partial u}{\partial t} &= t^{-1/2}\frac{\partial}{\partial t}f\left(\xi\right)-\frac{1}{2}t^{-3/2}f\left(\xi\right)\\ &= t^{-1/2}\left(-\frac{1}{2}xt^{-3/2}\right)f'\left(\xi\right)-\frac{1}{2}t^{-3/2}f\left(\xi\right)\\ &= -\frac{1}{2}\xi t^{-3/2}f'\left(\xi\right)-\frac{1}{2}t^{-3/2}f\left(\xi\right)\end{aligned}$$

and then substituting

$$\begin{aligned}\frac{\partial u}{\partial t} &= -\frac{1}{2}t^{-3/2}\left(\xi f'\left(\xi\right)+f\left(\xi\right)\right)\\ \frac{\partial^2 u}{\partial x^2} &= t^{-3/2}f''\left(\xi\right)\end{aligned}$$

gives

$$-\frac{1}{2}t^{-3/2}\left(\xi f'\left(\xi\right)+f\left(\xi\right)\right)=c^2t^{-3/2}f''\left(\xi\right)$$

simplifying to the ODE

$$-\frac{1}{2}\left(f+\xi f'\right)=c^2f''.$$

We have an exact derivative on the left hand side, i.e. $\frac{d}{d\xi}\left(\xi f\right)=f+\xi f'$, hence

$$-\frac{1}{2}\frac{d}{d\xi}\left(\xi f\right)=c^2f''$$

and we can integrate once to get

$$-\frac{1}{2}\left(\xi f\right)=c^2 f'+K.$$

We set $K=0$ in order to get the correct solution, i.e.

$$-\frac{1}{2}\left(\xi f\right)=c^2 f'$$

which can be solved as a simple first order variable separable equation:

$$f\left(\xi\right)=A\exp\left(-\tfrac{1}{4c^2}\xi^2\right)$$

A is a normalizing constant, so write

$$A\int_{\mathbb{R}}\exp\left(-\tfrac{1}{4c^2}\xi^2\right)d\xi=1.$$

Now substitute $s=\xi/2c$, so $2cds=d\xi$

$$2cA\underbrace{\int_{\mathbb{R}}\exp\left(-s^2\right)ds}_{=\sqrt{\pi}}=1,$$

which gives $A=\dfrac{1}{2c\sqrt{\pi}}$. Returning to

$$u\left(x,t\right)=t^{-1/2}f\left(\xi\right)$$

becomes

$$u\left(x,t\right)=\frac{1}{2c\sqrt{\pi t}}\exp\left(-\frac{x^2}{4tc^2}\right).$$

Hence the random variable x is Normally distributed with mean zero and standard deviation $c\sqrt{2t}$.

Applied Stochastic Calculus

Stochastic Process

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*.

We need a definition and set of properties for the randomness observed in an asset price realization, which will be *Brownian Motion*.

This is named after the Scottish Botanist who in 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 fidgety motion. In 1900 Louis Bachelier was the first person to model the share price movement using Brownian motion as part of his PhD. Five years later Einstein used Brownian motion to study diffusions. 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.

Construction of Brownian Motion and properties

We construct Brownian motion using a simple symmetric random walk. Define a random variable

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

Let

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

which defines the marker's position after the n^{th} toss of the game. This is conditional upon the marker starting at position $X = 0$, so at each time-step moves one unit either to the left or right with equal probability. Hence the distribution is binomial with

$$\begin{aligned} \text{mean} &= \frac{1}{2}(+1) + \frac{1}{2}(-1) = 0 \\ \text{variance} &= \frac{1}{2}(+1)^2 + \frac{1}{2}(-1)^2 = 1 \end{aligned}$$

This can be approximated to a Normal distribution due to the Central Limit Theorem.

Is there a continuous time limit to this discrete random walk? Let's introduce time dependency. Take a time period for our walk, say $[0, t]$ and perform N steps. So we partition $[0, t]$ into N time intervals, so each step takes

$$\delta t = t/N.$$

Speed up this random walk so let $N \to \infty$. The problem with the original step sizes of ± 1 gives the variance that is infinite. We rescale the space step, keeping in mind the central limit theorem. Let

$$Y = \alpha_N Z$$

for some α_N to be found, and let $\left\{X_n^N; n = 0, ..., N\right\}$ such that $X_0^N = 0$; be the path/trajectory of the random walk with steps of size α_N.

Thus we now have

$$\mathbb{E}\left[X_n^N\right] = 0, \ \forall n$$

and

$$\begin{aligned}\mathbb{V}\left[X_n^N\right] &= \mathbb{E}\left[\left(X_n^N\right)^2\right] = N\mathbb{E}\left[Y^2\right] \\ &= N\alpha_N^2 \mathbb{E}\left[Z^2\right] = N\alpha_N^2\left(\frac{1}{2} + \frac{1}{2}\right) \\ &= \left(\frac{t}{\delta t}\right)\alpha_N^2.\end{aligned}$$

Obviously we must have $\alpha_N^2/\delta t = O(1)$. Choosing $\alpha_N^2/\delta t = 1$ gives

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

As $N \to \infty$, the symmetric random walk $\left\{X_{[tN]}^N; t \in [0, \infty)\right\}$ converges to a **standard Brownian motion** $\{W_t; t \in [0, \infty)\}$. So $W_t \sim N(0, dt)$.

With $t = n\delta t$ we have

$$\frac{dW_t}{dt} = \lim_{\delta t \to 0} \frac{W_{t+\delta t} - W_t}{\delta t} \to \infty.$$

Quadratic Variation

Consider a function $f(t)$ on the interval $[0, T]$. Discretising by writing $t = idt$ and $dt = T/N$ we can define the variation $\mathbf{V}^n$ of f for $n = 1, 2, ..$ as

$$\mathbf{V}^n[f] = \lim_{N \to \infty} \sum_{i=0}^{N-1} \left|f_{t_i+1} - f_{t_i}\right|^n.$$

Of interest is the quadratic variation

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

If $f(t)$ has more than a finite number of jumps or a singularity then $Q[f] = \infty$.

For a Brownian motion on $[0, T]$ we have

$$\begin{aligned}Q[W_t] &= \lim_{N \to \infty} \sum_{i=0}^{N-1} \left(W_{t_i+1} - W_{t_i}\right)^2 \\ &= \lim_{N \to \infty} \sum_{i=0}^{N-1} \left(W((i+1)dt) - W(idt)\right)^2 \\ &= \lim_{N \to \infty} \sum_{i=0}^{N-1} dt = \lim_{N \to \infty} \sum_{i=0}^{N-1} \frac{T}{N} = T.\end{aligned}$$

Suppose that $f(t)$ is a differentiable function on $[0,T]$. Then to leading order, we have

$$f_{t_i+1} - f_{t_i} = f((i+1)\,dt) - f(idt) \sim f'(t_i)\,dt$$

so,

$$\begin{aligned} Q[f] &\sim \lim_{N\to\infty}\sum_{i=0}^{N-1}(f'(t_i)\,dt)^2 \\ &\sim \lim_{N\to\infty} dt\sum_{i=0}^{N-1}(f'(t_i))^2\,dt \\ &\sim \left(\lim_{N\to\infty}\frac{T}{N}\right)\int_0^T (f'(t_i))^2\,dt \\ &= 0. \end{aligned}$$

The quadratic variation of $f(t)$ is zero. This argument remains valid even if $f'(t)$ has a finite number of jump discontinuities. Thus a Brownian motion, W_t, has at worst a finite number of discontinuities, but an infinite number of discontinuities in its derivative, W_t'. It is continuous but not differentiable, almost everywhere. For us the important result is

$$dW_t^2 = dt$$

or more importantly we can write (up to mean square limit)

$$\mathbb{E}\left[dW_t^2\right] = dt.$$

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), i.e. $\mathbb{P}(W_0 = 0) = 1$.
- Continuity - paths of W_t are continuous (no jumps) with probability 1. 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
  $$0 \le t_0 \le t_1 \le t_2 \le .....$$
  $$\begin{aligned} dW_1 &= W_1 - W_0 \text{ is independent of } dW_2 = W_2 - W_1 \\ dW_3 &= W_3 - W_2 \text{ is independent of } dW_4 = W_4 - W_3 \\ &\quad\text{and so on} \end{aligned}$$

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 $(X_t)_{t\geq 0}$ 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 $I = (I_t)_{t\geq 0}$ 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 $(W_{t+h} - W_t)_{h\geq 0}$ 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. Let can be calculated as follows. If $t > s$,

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

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

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 $\{X_n : 0 \leq n < \infty\}$ 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 $\mathbb{E}_n^{\mathbb{P}}[|X_n|] < \infty \ \forall\ n \geq 0$

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

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

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

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

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 - 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).

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".

Let us look at an example. Consider the <u>earlier</u> 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

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

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

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

Functions of a stochastic variable and Stochastic Differential Equations

In continuous time models, changes are (infinitesimally) small. Calculus is used to analyse small changes, hence an extension of 'ordinary' deterministic calculus to variables governed by a diffusion process.

Start by recalling a Taylor series expansion, i.e. Taylor's theorem: let $f(x)$ be a sufficiently differentiable function of x, for small δx,

$$f(x+\delta x) = f(x) + f'(x)\,\delta x + \frac{1}{2} f''(x)\,\delta x^2 + O\left(\delta x^3\right).$$

So we are approximating using the tangent or quadratic. The infinitesimal version is

$$df = f'(x)\,\delta x$$

where we have defined

$$df = f(x+\delta x) = f(x)$$

where $\delta x \ll 1$. Hence $\delta x^2 \ll \delta x$, and

$$df \sim \frac{df}{dx}\delta x + \ldots.$$

How does this work for functions of a stochastic variable?

Suppose that $x = W(t)$ is Brownian motion, so $f = f(W)$

$$\begin{aligned} df &\sim \frac{df}{dW}dW + \frac{1}{2}\frac{d^2 f}{dW^2}(dW)^2 + \ldots. \\ &\sim \frac{df}{dW}dW + \frac{1}{2}\frac{d^2 f}{dW^2}dt + \ldots. \end{aligned}$$

This is the most basic version of Itô's lemma; for a function of a Wiener process (or Brownian motion) $W(t)$ or W_t, given by

$$df = \frac{df}{dW}dW + \frac{1}{2}\frac{d^2 f}{dW^2}dt.$$

Now consider a simple example $f = W^2$ then

$$\begin{aligned} d\left(W^2\right) &= 2W\,dW + \frac{1}{2}(2)\,dt \\ df &= 2W\,dW + dt, \end{aligned}$$

which is a consequence of Brownian motion and stochastic calculus. In normal calculus the $+dt$ term would not be present.

More generally, suppose $F = F(t, W)$, is a function of time and Brownian motion, then Taylor's theorem is

$$dF(t,W) = \frac{\partial F}{\partial t}dt + \frac{\partial F}{\partial W}dW + \frac{1}{2}\frac{\partial^2 F}{\partial W^2}(dW)^2 + O(dW)^3$$

where we know $dW^2 = dt$, so Itô's lemma becomes

$$dF(t,W) = \left(\frac{\partial F}{\partial t} + \frac{1}{2}\frac{\partial^2 F}{\partial W^2}\right)dt + \frac{\partial F}{\partial W}dW.$$

Two important examples of Itô's lemma are

- $f(W(t)) = \log W(t)$ for which Itô gives

$$d\log W(t) = \frac{dW}{W} - \frac{dt}{2W^2}$$

- $g(W(t)) = e^{W(t)}$ for which Itô implies

$$de^{W(t)} = e^{W(t)}dW + \frac{1}{2}e^{W(t)}dt$$

If we write $S = e^{W(t)}$ then this becomes

$$dS = SdW + \frac{1}{2}Sdt$$

or

$$\frac{dS}{S} = \frac{1}{2}dt + dW$$

Geometric Brownian motion

In the Black-Scholes model for option prices, we denote the (risky) underlying (equity) asset price by $S(t)$ or S_t. Typical to also suppress the t and simply write the stock price as S. We model the instantaneous return during time dt,

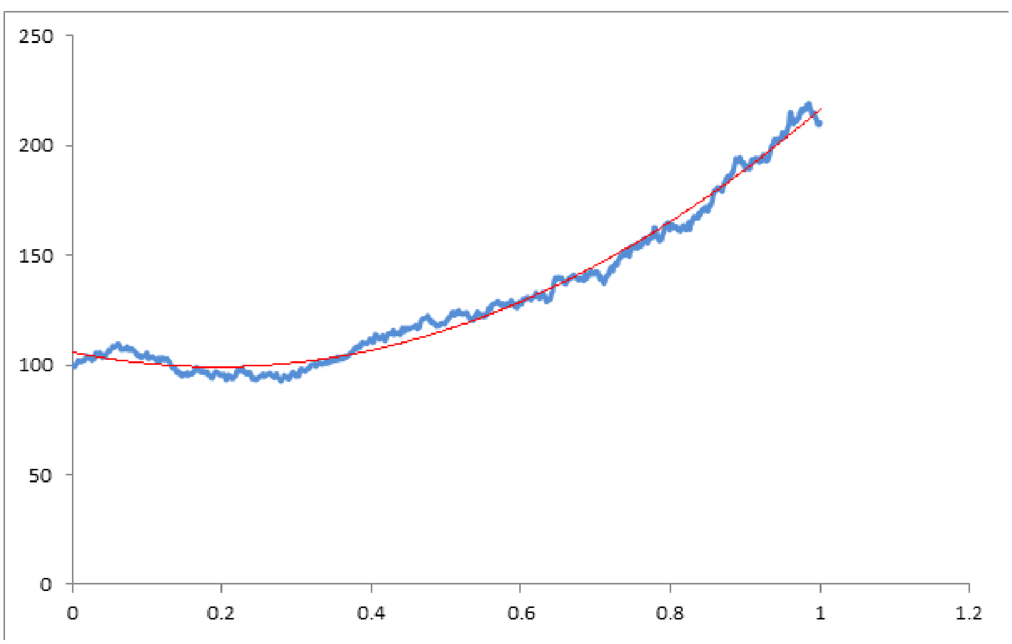
$$\frac{dS}{S} = \frac{dS(t)}{S(t)} = \frac{S(t+dt) - S(t)}{S(t)},$$

as a Normally distributed random variable,

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

where μdt is the expected return over dt and $\sigma^2 dt$ is the variance of returns (about the expected return).

We can think of μ as being a measure of the exponential growth of the expected asset price in time and σ is a measure of size of the random fluctuations about that exponential trend or a measure of the risk.



If we have

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

or more conveniently

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

then as $dW^2 = dt$,

$$\begin{aligned} dS^2 &= (\mu S dt + \sigma S dW)^2 \\ &= \sigma^2 S^2 dW^2 + 2\mu\sigma S^2 dt dW + \mu^2 S^2 dt^2 \end{aligned}$$

$$dS^2 = \sigma^2 S^2 dt + ...$$

In the limit $dt \to 0$,

$$dS^2 = \sigma^2 S^2 dt$$

This leads to Itô lemma for Geometric Brownian motion (GBM).

If $V = V(t, S)$, is a function of S and t, then Taylor's theorem states

$$dV = \frac{\partial V}{\partial t} dt + \frac{\partial V}{\partial S} dS + \frac{1}{2} \frac{\partial^2 V}{\partial S^2} dS^2$$

so if S follows GBM,

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

then $dS^2 = \sigma^2 S^2 dt$ and we obtain

Itô lemma for Geometric Brownian motion;

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

where the partial derivatives are evaluated at S and t.

If $V = V(S)$ then we obtain the shortened version of Itô;

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

Following on from the earlier example, $S(t) = e^{W(t)}$, for which

$$dS = \frac{1}{2} S dt + S dW$$

we find that we can solve the SDE $\frac{dS}{S} = \mu dt + \sigma dW$.

If we put $S(t) = Ae^{at + bW(t)}$ then from the earlier form of Itô's lemma we have

$$\begin{aligned} dS &= \left(aS + \frac{1}{2} b^2 S \right) dt + bS dW \\ & \quad \text{or} \\ \frac{dS}{S} &= \left(a + \frac{1}{2} b^2 \right) dt + b dW \\ & \quad \text{comparing with} \\ \frac{dS}{S} &= \mu dt + \sigma dW \\ & \quad \text{gives} \\ b &= \sigma; \; a = \mu - \frac{1}{2} \sigma^2 \end{aligned}$$

Another way to arrive at the same result is to use Itô for GBM. Using $f(S) = \log S(t)$ with

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

gives

$$\begin{aligned} d(\log S) &= \left(\mu S \frac{\partial}{\partial S}(\log S) + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2}{\partial S^2}(\log S)\right) dt + \sigma S \frac{\partial}{\partial S}(\log S)\, dW \\ &= \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma dW \end{aligned}$$

and hence

$$\begin{aligned} \int_0^t d(\log S(\tau)) &= \int_0^t \left(\mu - \frac{1}{2}\sigma^2\right) d\tau + \sigma \int_0^t dW \\ \log \frac{S(t)}{S(0)} &= \left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W(t) \end{aligned}$$

Taking exponentials and rearranging gives the earlier result. We have also used $W(0) = 0$.

Itô multiplication table:

$\times$	dt	dW
dt	$dt^2 = 0$	$dtdW = 0$
dW	$dWdt = 0$	$dW^2 = dt$

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

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

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

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

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

Diffusion Process

G is called a diffusion process if

$$dG(t) = A(G,t)\,dt + B(G,t)\,dW(t) \tag{1}$$

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

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

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

For example

$$dG(t) = (G(t) + G(t-1))\,dt + dW(t)$$

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.
- $dG^{\,2} = B^2 dt$.

We say (1) is a SDE for the process G or a *Random Walk* for dG.

The diffusion (1) can be written in integral form as

$$G(t) = G(0) + \int_0^t A(G,\tau)\,d\tau + \int_0^t B(G,\tau)\,dW(\tau)$$

Remark: A diffusion G is a *Markov* process if - once the present state $G(t) = g$ is given, the past $\{G(\tau),\ \tau < t\}$ 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 $G(t)$

$$dG = \mu G dt + \sigma G dW \tag{2}$$

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

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

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

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

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

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

An Extension of Itô's Lemma (2D)

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

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

So S follows

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

Remember that

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

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

$$dS^2 = \sigma^2 S^2 dt$$

So the change dV when $V(S, t) \to V(S + dS, t + dt)$ is given by

$$dV = \frac{\partial V}{\partial t} dt + \frac{\partial V}{\partial S} [S \mu dt + S \sigma dW] + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} dt$$

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

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

This is Itô's Formula in two dimensions.

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

$$dV = \left(\mu S \frac{dV}{dS} + \frac{1}{2} \sigma^2 S^2 \frac{d^2 V}{dS^2} \right) dt + \sigma S \frac{dV}{dS} dW. \tag{6}$$

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

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

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

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

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 $dV =$

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

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

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 $= \sigma n V$.

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. Here we consider another important example of a SDE, 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 rate, and we can rewrite this random walk (7) for dr_t as

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

The dimensions of γ are $1/\text{time}$, 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).
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 $e^{\gamma t}$

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

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.

Leaving the result in the form of (10) allows the calculation of the mean, variance and other moments. Start with the expected value $\mathbb{E}[X_t] =$

$$\begin{aligned} \mathbb{E}[X_t] &= \mathbb{E}\left[\alpha e^{-\gamma t} + \sigma \int_0^t e^{\gamma(s-t)} dW_s\right] \\ &= \mathbb{E}\left[\alpha e^{-\gamma t}\right] + \sigma \mathbb{E}\left[\int_0^t e^{\gamma(s-t)} dW_s\right] \\ &= \alpha e^{-\gamma t} + \sigma \int_0^t e^{\gamma(s-t)} \mathbb{E}[dW_s] \end{aligned}$$

Recall that Brownian motion is a Martingale; the Itô integral is a Martingale, hence

$$\mathbb{E}[X_t] = \alpha e^{-\gamma t}$$

To calculate the variance we have $\mathbb{V}[X_t] = \mathbb{E}[X_t^2] - \mathbb{E}^2[X_t]$

$$\begin{aligned} &= \mathbb{E}\left[\left(\alpha e^{-\gamma t} + \sigma \int_0^t e^{\gamma(s-t)} dW_s\right)^2\right] - \alpha^2 e^{-2\gamma t} \\ &= \mathbb{E}\left[\alpha^2 e^{-2\gamma t}\right] + \sigma^2 \mathbb{E}\left[\left(\int_0^t e^{\gamma(s-t)} dW_s\right)^2\right] + 2\sigma\alpha e^{-\gamma t} \underbrace{\mathbb{E}\left[\int_0^t e^{\gamma(s-t)} dW_s\right]}_{\text{Itô integral}} - \alpha^2 e^{-2\gamma t} \\ &= \sigma^2 \mathbb{E}\left[\left(\int_0^t e^{\gamma(s-t)} dW_s\right)^2\right] \end{aligned}$$

Now use Itô's Isometry

$$\mathbb{E}\left[\left(\int_0^t Y_s dW_s\right)^2\right] = \mathbb{E}\left[\int_0^t Y_s^2 ds\right],$$

So

$$\begin{aligned}\mathbb{V}[X_t] &= \sigma^2 \mathbb{E}\left[\int_0^t e^{2\gamma(s-t)} ds\right] \\ &= \frac{\sigma^2}{2\gamma} e^{2\gamma(s-t)}\Big|_0^t = \frac{\sigma^2}{2\gamma}\left(1 - e^{-2\gamma t}\right)\end{aligned}$$

Returning to the integral in (10)

$$\int_0^t e^{\gamma(s-t)} dW_s$$

let's use the stochatsic integral formula to verify the result. Recall

$$\int_0^t \frac{\partial f}{\partial W} dW = f(t, W_t) - f(0, W_0) - \int_0^t \left(\frac{\partial f}{\partial s} + \frac{1}{2}\frac{\partial^2 f}{\partial W^2}\right) ds$$

so $\frac{\partial f}{\partial W} \equiv e^{\gamma(s-t)} \Longrightarrow f = e^{\gamma(s-t)} W_s,\ \frac{\partial f}{\partial s} = \gamma e^{\gamma(s-t)} W_s,\ \frac{\partial^2 f}{\partial W^2} = 0$

$$\begin{aligned}\int_0^t e^{\gamma(s-t)} dW_s &= W_t - 0 - \int_0^t \left(\gamma e^{\gamma(s-t)} W_s + \tfrac{1}{2} \times 0\right) ds \\ &= W_t - \gamma \int_0^t e^{\gamma(s-t)} W_s ds.\end{aligned}$$

We have used an integrating factor to obtain a solution of the Ornstein Uhlenbeck process. Let's look at $d\left(e^{\gamma t} U_t\right)$ by using Itô. Consider a function $V(t, U_t)$ where $dU_t = -\gamma U_t dt + \sigma dW_t$, then

$$\begin{aligned}dV &= \left(\frac{\partial V}{\partial t} - \gamma U \frac{\partial V}{\partial U} + \frac{1}{2}\sigma^2 \frac{\partial^2 V}{\partial U^2}\right) dt + \sigma \frac{\partial V}{\partial U} dW \\ d\left(e^{\gamma t} U\right) &= \left(\frac{\partial}{\partial t}\left(e^{\gamma t} U\right) - \gamma U \frac{\partial}{\partial U}\left(e^{\gamma t} U\right) + \frac{1}{2}\sigma^2 \frac{\partial^2}{\partial U^2}\left(e^{\gamma t} U\right)\right) dt + \\ &\quad \sigma \frac{\partial}{\partial U}\left(e^{\gamma t} U\right) dW \\ &= \left(\gamma e^{\gamma t} U - \gamma U e^{\gamma t}\right) dt + \sigma e^{\gamma t} dW \\ &= \sigma e^{\gamma t} dW\end{aligned}$$

Example: The Ornstein-Uhlenbeck process satisfies the spot rate SDE given by

$$dX_t = \kappa(\theta - X_t)\,dt + \sigma dW_t,\ \ X_0 = x,$$

where κ, θ and σ are constants. Solve this SDE by setting $Y_t = e^{\kappa t}X_t$ and using Itô's lemma to show that

$$X_t = \theta + (x-\theta)\,e^{-\kappa t} + \sigma\int_0^t e^{-\kappa(t-s)}dW_s.$$

First write Itô for Y_t given $dX_t = A(X_t,t)\,dt + B(X_t,t)\,dW_t$

$$\begin{aligned} dY_t &= \left(\frac{\partial Y_t}{\partial t} + A(X_t,t)\frac{\partial Y_t}{\partial X_t} + \tfrac{1}{2}B^2(X_t,t)\frac{\partial^2 Y_t}{\partial X_t^2}\right)dt + B(X_t,t)\frac{\partial Y_t}{\partial X_t}dW_t \\ &= \left(\frac{\partial Y_t}{\partial t} + \kappa(\theta - X_t)\frac{\partial Y_t}{\partial X_t} + \tfrac{1}{2}\sigma^2\frac{\partial^2 Y_t}{\partial X_t^2}\right)dt + \sigma\frac{\partial Y_t}{\partial X_t}dW_t \end{aligned}$$

$$\frac{\partial Y_t}{\partial t} = \kappa e^{\kappa t}X_t;\ \frac{\partial Y_t}{\partial X_t} = e^{\kappa t};\ \frac{\partial^2 Y_t}{\partial X_t^2} = 0.$$

$$\begin{aligned} d\left(e^{\kappa t}X_t\right) &= \left(\kappa e^{\kappa t}X_t + \kappa(\theta - X_t)\,e^{\kappa t}\right)dt + \sigma e^{\kappa t}dW_t \\ &= \kappa\theta e^{\kappa t}dt + \sigma e^{\kappa t}dW_t \end{aligned}$$

$$\begin{aligned} \int_0^t d\left(e^{\kappa s}X_s\right) &= \kappa\theta\int_0^t e^{\kappa s}ds + \sigma\int_0^t e^{\kappa s}dW_s \\ e^{\kappa t}X_t - x &= \theta e^{\kappa t} - \theta + \sigma\int_0^t e^{\kappa s}dW_s \\ X_t &= xe^{-\kappa t} + \theta - \theta e^{-\kappa t} + \sigma e^{-\kappa t}\int_0^t e^{\kappa s}dW_s \\ X_t &= \theta + (x-\theta)\,e^{-\kappa t} + \sigma\int_0^t e^{-\kappa(t-s)}dW_s. \end{aligned}$$

Consider

$$dr_t = \kappa(\theta - r_t)\,dt + \sigma dW_t,$$

and show by suitable integration for $s < t$

$$r_t = r_s e^{-\kappa(t-s)} + \theta\left(1 - e^{-\kappa(t-s)}\right) + \sigma\int_s^t e^{-\kappa(t-u)}dW_u.$$

The lower limit gives us an initial condition at time $s < t$. Expand $d\left(e^{\kappa t}r_t\right)$

$$\begin{aligned} d\left(e^{\kappa t}r_t\right) &= \left(\kappa e^{\kappa t}r_t dt + e^{\kappa t}dr_t\right) \\ &= e^{\kappa t}\left(\kappa\theta dt + \sigma dW_t\right) \end{aligned}$$

Now integrate both sides over $[s,t]$ to give for each $s < t$

$$\begin{aligned} \int_s^t d\left(e^{\kappa u}r_u\right) &= \kappa\theta\int_s^t e^{\kappa u}du + \sigma\int_s^t e^{\kappa u}dW_u \\ e^{\kappa t}r_t - e^{\kappa s}r_s &= \theta e^{\kappa t} - \theta e^{\kappa s} + \sigma\int_s^t e^{\kappa u}dW_u \end{aligned}$$

rearranging and dividing through by $e^{\kappa t}$

$$\begin{aligned} r_t &= e^{-\kappa(t-s)}r_s + \theta - \theta e^{-\kappa(t-s)} + \sigma e^{-\kappa t}\int_s^t e^{\kappa s}dW_u \\ r_t &= e^{-\kappa(t-s)}r_s + \theta\left(1 - e^{-\kappa(t-s)}\right) + \sigma\int_s^t e^{-\kappa(t-u)}dW_u \end{aligned}$$

so that r_t conditional upon r_s is normally distributed with mean and variance given by

$$\begin{aligned} \mathbb{E}\left[r_t|\, r_s\right] &= e^{-\kappa(t-s)}r_s + \theta\left(1 - e^{-\kappa(t-s)}\right) \\ \mathbb{V}\left[r_t|\, r_s\right] &= \frac{\sigma^2}{2\kappa}\left(1 - e^{-2\kappa(t-s)}\right) \end{aligned}$$

We note that as $t \to \infty$, the mean and variance become in turn

$$\begin{aligned} \mathbb{E}\left[r_t|\, r_s\right] &= \theta \\ \mathbb{V}\left[r_t|\, r_s\right] &= \frac{\sigma^2}{2\kappa} \end{aligned}$$

Example: Given $U = \log Y$, where Y satisfies the diffusion process

$$\begin{aligned} dY &= \frac{1}{2Y}dt + dW \\ Y(0) &= Y_0 \end{aligned}$$

use Itô's lemma to find the SDE satisfied by U.

Since $U = U(Y)$ with $dY = a(Y,t)\,dt + b(Y,t)\,dW$, we can write

$$dU = \left(a(Y,t)\frac{dU}{dY} + \frac{1}{2}b^2(Y,t)\frac{d^2U}{dY^2}\right)dt + b(Y,t)\frac{dU}{dY}dW$$

Now $U = \log(Y)$ so $\frac{dU}{dY} = \frac{1}{Y}, \frac{d^2U}{dY^2} = -\frac{1}{Y^2}$ and substituting in

$$dU = \left(\frac{1}{2Y}\left(\frac{1}{Y}\right) + \frac{1}{2}(1)^2\left(-\frac{1}{Y^2}\right)\right)dt + \frac{1}{Y}dW$$

$$dU = e^{-U}dW$$

Example: Consider the stochastic volatility model

$$d\sqrt{v} = \left(\alpha - \beta\sqrt{v}\right)dt + \delta dW$$

where v is the variance. Show that

$$dv = \left(\delta^2 + 2\alpha\sqrt{v} - 2\beta v\right)dt + 2\delta\sqrt{v}dW$$

Setting the variable $X = \sqrt{v}$ giving $dX = \underbrace{(\alpha - \beta X)}_{A}dt + \underbrace{\delta}_{B}dW$. We now require a SDE for dY, where $Y = X^2$. So $dv =$

$$\begin{aligned} dY &= \left(A\frac{dY}{dX} + \frac{1}{2}B^2\frac{d^2Y}{dX^2}\right)dt + B\frac{dY}{dX}dW \\ &= \left((\alpha - \beta X)(2X) + \frac{1}{2}\delta^2 \times 2\right)dt + \delta \times 2XdW \\ &= \left(2\alpha X - 2\beta X^2 + \delta^2\right)dt + 2\delta\sqrt{v}dW \\ &= \left(\delta^2 + 2\alpha\sqrt{v} - 2\beta v\right)dt + 2\delta\sqrt{v}dW \end{aligned}$$

(Harder) Example: Consider the dynamics of a non-traded asset S_t given by

$$\frac{dS_t}{S_t} = \alpha\left(\theta - \log S_t\right)dt + \sigma dW_t$$

where the constants $\sigma, \alpha > 0$. If $T > t$, show that

$$\log S_T = e^{-\alpha(T-t)}\log S_t + \left(\theta - \frac{1}{2\alpha}\sigma^2\right)\left(1 - e^{-\alpha(T-t)}\right) + \sigma\int_t^T e^{-\alpha(T-s)}dW_s.$$

Hence show that

$$\log S_T \sim N\left(e^{-\alpha(T-t)}\log S_t + \left(\theta - \frac{1}{2\alpha}\sigma^2\right)\left(1 - e^{-\alpha(T-t)}\right), \sigma^2\left(\frac{1 - e^{-2\alpha(T-t)}}{2\alpha}\right)\right)$$

Writing Itô for the SDE where $f = f\left(S_t\right)$ gives

$$df = \left(\alpha\left(\theta - \log S_t\right)S_t\frac{df}{dS} + \tfrac{1}{2}\sigma^2 S_t^2\frac{d^2 f}{dS^2}\right)dt + \sigma S_t\frac{df}{dS}dW_t.$$

Hence if $f\left(S_t\right) = \log S_t$ then

$$\begin{aligned} d\left(\log S_t\right) &= \left(\alpha\left(\theta - \log S_t\right) - \tfrac{1}{2}\sigma^2\right)dt + \sigma dW_t \\ &= \alpha\left(\theta - \frac{1}{2\alpha}\sigma^2 - \log S_t\right)dt + \sigma dW_t \\ &= -\alpha\left(\log S_t - \mu\right)dt + \sigma dW_t \end{aligned}$$

where $\mu = \theta - \frac{1}{2\alpha}\sigma^2$. Going back to

$$df = -\alpha\left(f - \mu\right)dt + \sigma dW_t$$

and now write $x_t = f - \mu$ which gives $dx_t = df$ and we are left with an Ornstein-Uhlenbeck process

$$dx_t = -\alpha x_t dt + \sigma dW_t.$$

Following the earlier integrating factor method gives

$$\begin{aligned} d\left(e^{\alpha t}x_t\right) &= \sigma e^{\alpha t}dW_t \\ \int_t^T d\left(e^{\alpha s}x_s\right) &= \sigma\int_t^T e^{\alpha s}dW_s \\ x_T &= e^{-\alpha(T-t)}x_t + \sigma\int_t^T e^{-\alpha(T-s)}dW_s. \end{aligned}$$

Now replace these terms with the original variables and parameters

$$\log S_T - \left(\theta - \frac{1}{2\alpha}\sigma^2\right) = e^{-\alpha(T-t)}\left(\log S_T - \left(\theta - \frac{1}{2\alpha}\sigma^2\right)\right) + \sigma\int_t^T e^{-\alpha(T-s)}dW_s,$$

which upon rearranging and factorising gives

$$\log S_T = e^{-\alpha(T-t)}\log S_T + \left(\theta - \frac{1}{2\alpha}\sigma^2\right)\left(1 - e^{-\alpha(T-t)}\right) + \sigma\int_t^T e^{-\alpha(T-s)}dW_s.$$

Now consider $\mathbb{E}[\log S_T] =$

$$\begin{aligned} & e^{-\alpha(T-t)}\log S_T + \left(\theta - \frac{1}{2\alpha}\sigma^2\right)\left(1 - e^{-\alpha(T-t)}\right) + \sigma\mathbb{E}\left[\int_t^T e^{-\alpha(T-s)}dW_s\right] \\ = \; & e^{-\alpha(T-t)}\log S_T + \left(\theta - \frac{1}{2\alpha}\sigma^2\right)\left(1 - e^{-\alpha(T-t)}\right) \end{aligned}$$

Recall $\mathbb{V}[aX+b] = a^2\mathbb{V}[X]$. So write $\mathbb{V}[\log S_T] =$

$$\begin{aligned} & \mathbb{V}\left[e^{-\alpha(T-t)}\log S_T + \left(\theta - \frac{1}{2\alpha}\sigma^2\right)\left(1 - e^{-\alpha(T-t)}\right) + \sigma\int_t^T e^{-\alpha(T-s)}dW_s\right] \\ = \; & \underbrace{\mathbb{V}\left[e^{-\alpha(T-t)}\log S_T + \left(\theta - \frac{1}{2\alpha}\sigma^2\right)\left(1 - e^{-\alpha(T-t)}\right)\right]}_{=0} + \mathbb{V}\left[\sigma\int_t^T e^{-\alpha(T-s)}dW_s\right] \\ = \; & \sigma^2\mathbb{V}\left[\int_t^T e^{-\alpha(T-s)}dW_s\right] = \sigma^2\mathbb{E}\left[\left(\int_t^T e^{-\alpha(T-s)}dW_s\right)^2\right] \end{aligned}$$

because we have already obtained from the expectation that $\mathbb{E}\left[\int_t^T e^{-\alpha(T-s)}dW_s\right] = 0.$

Now use Itô's Isometry, i.e.

$$\mathbb{E}\left[\left(\int_0^t Y_s dX_s\right)^2\right] = \mathbb{E}\left[\int_0^t Y_s^2 ds\right],$$

$$\begin{aligned} \mathbb{V}[\log S_T] &= \sigma^2\mathbb{E}\left[\left(\int_t^T e^{-\alpha(T-s)}dW_s\right)^2\right] \\ &= \sigma^2\mathbb{E}\left[\int_t^T e^{-2\alpha(T-s)}ds\right] \\ &= \sigma^2\mathbb{E}\left[\frac{1}{2\alpha}e^{-2\alpha(T-s)}\bigg|_t^T\right] \\ &= \frac{\sigma^2}{2\alpha}\left(1 - e^{-2\alpha(T-t)}\right) \end{aligned}$$

Hence verified.

Example: Consider the SDE for the variance process v

$$dv = \varepsilon(m - \sigma)\,dt + \xi\sigma dW_t,$$

where $v = \sigma^2$. ε, ξ, m are constants. Using Itô's lemma, show that the volatility σ satisfies the SDE

$$d\sigma = a(\sigma, t)\,dt + b(\sigma, t)\,dW_t,$$

where the precise form of $a(\sigma, t)$ and $b(\sigma, t)$ should be given.

Consider the stochastic volatility model

$$dv = \varepsilon\left(m - \sqrt{v}\right)dt + \xi\sqrt{v}dW_t$$

If $F = F(v)$ then Itô gives

$$dF = \left(\varepsilon (m - \sigma) \frac{dF}{dv} + \frac{1}{2} \xi^2 v \frac{d^2 F}{dv^2} \right) dt + \xi \sqrt{v} \frac{dF}{dv} dW_t.$$

For $F(v) = v^{1/2}$; $\frac{dF}{dv} = \frac{1}{2} v^{-1/2}$, $\frac{d^2 F}{dv^2} = -\frac{1}{4} v^{-3/2}$

$$\begin{aligned} dF &= d\sigma = \left(\frac{\varepsilon}{2} (m - \sigma) v^{-1/2} - \frac{1}{8} \xi^2 v^{-1} \right) dt + \frac{\xi}{2} dW_t \\ &= \left(\frac{\varepsilon}{2\sigma} (m - \sigma) - \frac{1}{8\sigma} \xi^2 \right) dt + \frac{\xi}{2} dW_t \end{aligned}$$

$$a(\sigma, t) = \left(\frac{\varepsilon}{2\sigma} (m - \sigma) - \frac{1}{8\sigma} \xi^2 \right); \quad b(\sigma, t) = \frac{\xi}{2}$$

Higher Dimensional Itô

There is a multi-dimensional form of Itô's lemma. Let us consider the two-dimensional version initially, as this can be generalised nicely to the $N-$dimensional case, driven by a Brownian motion of any number (not necessarily the same number) of dimensions. Let

$$W_t := \left(W_t^{(1)}, W_t^{(2)} \right)$$

be a two-dimensional Brownian motion, where $W_t^{(1)}, W_t^{(2)}$ are independent Brownian motions, and define the two-dimensional Itô process

$$X_t := \left(X_t^{(1)}, X_t^{(2)} \right)$$

such that

$$dS_i = \mu_i S_i dt + \sigma_i S_i dW_i$$

Consider the case where N shares follow the usual Geometric Brownian Motions, i.e.

$$dS_i = \mu_i S_i dt + \sigma_i S_i dW_i,$$

for $1 \le i \le N$. The share price changes are correlated with correlation coefficient ρ_{ij}. By starting with a Taylor series expansion

$$\begin{gathered} V(t + \delta t, S_1 + \delta S_1, S_2 + \delta S_2, \ldots.., S_N + \delta S_N) = \\ V(t, S_1, S_2, \ldots.., S_N) + \tfrac{\partial V}{\partial t} + \sum_{i=1}^{N} \tfrac{\partial V}{\partial S_i} dS_i + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=i}^{N} \tfrac{\partial^2 V}{\partial S_i \partial S_j} + \ldots. \end{gathered}$$

which becomes, using $dW_i dW_j = \rho_{ij} dt$

$$dV = \left(\frac{\partial V}{\partial t} + \sum_{i=1}^{N} \mu_i S_i \frac{\partial V}{\partial S_i} + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sigma_i \sigma_j \rho_{ij} S_i S_j \frac{\partial^2 V}{\partial S_i \partial S_j} \right) dt + \sum_{i=1}^{N} \sigma_i S_i \frac{\partial V}{\partial S_i} dW_i.$$

We can integrate both sides over 0 and t to give

$$\begin{aligned} V(t, S_1, S_2, \ldots.., S_N) = {} & V(0, S_1, S_2, \ldots.., S_N) + \\ & \int_0^t \left(\tfrac{\partial V}{\partial \tau} + \sum_{i=1}^{N} \mu_i S_i \tfrac{\partial V}{\partial S_i} + \tfrac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sigma_i \sigma_j \rho_{ij} S_i S_j \tfrac{\partial^2 V}{\partial S_i \partial S_j} \right) d\tau \\ & + \int_0^t \sum_{i=1}^{N} \sigma_i S_i \tfrac{\partial V}{\partial S_i} dW_i. \end{aligned}$$

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

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

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

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,

$$\lim_{N\to\infty}\sqrt{\frac{12}{N}}\left(\sum_1^N U(0,1)-\tfrac{N}{2}\right)\sim N(0,1).$$

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 $\left(\sum_1^N \mathbf{U}_i-\frac{N}{2}\right)$

$$\begin{aligned} \mathbb{V}\left[\sum_1^N \mathbf{U}_i-\tfrac{N}{2}\right] &= \sum_1^N \mathbb{V}[\mathbf{U}_i] \\ &= N/12 \end{aligned}$$

As the variance is not 1, write

$$\mathbb{V}\left[\lambda\left(\sum_1^N \mathbf{U}_i-\tfrac{N}{2}\right)\right]$$

for some $\lambda\in\mathbb{R}$. Hence $\lambda^2\frac{N}{12}=1$ which gives $\lambda=\sqrt{12/N}$ 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

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

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.