Lecture Notes for 5765/6895, Part I

1 A brief introduction to continuous-time models

We begin with our discrete-time model from the last semester

$$S_n = S(t_n) = S(0)e^{\sigma\sqrt{\Delta t}(X_1+X_2+\cdots+X_n)}, \qquad n = 0, 1, 2, \ldots \tag{1}$$

which is defined only for those discrete times t_n. It is natural to ask for ways to extend this model for continuous time t to fill the gaps. The obvious extension is

$$S(t) = S(0)e^{X(t)}, \qquad \text{for } t > 0 \tag{2}$$

where $X(t)$ extends the notion of the finite sum in the random walk. As we see from the discrete time model, one of the central questions to ask is about the correlations among stock prices at different times $t_1, t_2, \ldots$, and it is important to assume that price changes $S_1 - S_0, S_2 - S_1, \ldots$ are mutually independent. This suggests that we should concentrate on these changes as our main target, which motivates the study on returns $R_n = (S_{n+1} - S_n)/S_n$. If we let Δt to approach zero, naturally we would be working on the instantaneous return $dS(t)/S(t)$. This brings up the central question in stochastic process: how do we see the collection of random variables indexed by a continuous parameter t? If we think of S as a function of t that depends on a random factor, how should we define $dS(t)$ even as it may turn out that $S(t)$ may not be differentiable in t?

1.1 From random walk to Brownian motion

The extension procedure has many new elements, compared to the regular procedure we are familiar with in Newton calculus, and the ideas can be illustrated by considering the prototype process: Brownian motion. As we see from Eqs.(1) and (2), the question is to find the proper limiting procedure to link

$$\sigma\sqrt{\Delta t}\,(X_1 + X_2 + \cdots + X_n) \to X(t)$$

where $X_j, j = 1, \ldots$ are i.i.d. random variables with a binomial distribution.
To begin the discussion, we introduce the standard random walk

$$M_n = \sum_{j=1}^{n} X_j, \quad n = 1, \ldots \tag{3}$$

with

$$X_j = \begin{cases} 1 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{2} \end{cases} \tag{4}$$

Notice that this is different from the random walk used in the Ross text, where the probability of turning a head is

$$p = \frac{1}{2}\left(1 + \frac{\mu}{\sigma}\sqrt{\Delta t}\right) \tag{5}$$

which results in

$$E[X_i] = \frac{\mu}{\sigma}\sqrt{\Delta t} \tag{6}$$

The version of random walk we use here is called symmetric random walk and there is an obvious advantage in its simplicity. To address the nonzero mean issue, we can express

$$X(n\Delta t) = \mu n\Delta t + \sigma\sqrt{\Delta t}M_n \tag{7}$$

Then Eq.(1) can written as

$$S_n = S(t_n) = S(0)e^{X(t_n)} = S(0)e^{\mu t_n + \sigma\sqrt{\Delta t}M_n}, \qquad n = 0, 1, 2, \ldots \tag{8}$$

From now on, we will maintain the symmetric random walk notation for X_j and focus on the limit of

$$\sqrt{\Delta t}M_n = \sqrt{\frac{Tn}{N}} \cdot \frac{X_1 + X_2 + X_3 + \cdots + X_n}{\sqrt{n}}$$

as $N \to \infty$. The reason we take the trouble to write in this form is that for fixed $t = t_n = Tn/N$, as $N \to \infty$, $n \to \infty$ too, but the sum in equation divided by $\sqrt{n}$ will converge in distribution to a standard normal random variable, according to the cantral limit theorem, which implies that $\sqrt{\Delta t}M_n$ will converge to a normal random variable with mean zero and variance t_n. This turns out to be a starting point of Brownian motion. At this point, we can see that our goal is to establish

$$X(t_n) = \mu t_n + \sigma W(t_n) \tag{9}$$

which will allow us to extended from $t_n = n\Delta t$ to more general t

$$X(t) = \mu t + \sigma W(t) \tag{10}$$

At this point it is clear that we will need to establish a procedure

$$\sqrt{\Delta t}M_n = W(t_n) \longrightarrow W(t)$$

From the discussion above, it appears that the major requirements for the process $W(t)$ are:

1. $W(t) - W(s)$ and $W(u) - W(v)$ are independent as long as those two time intervals $[s, t]$ and $[v, u]$ do not overlap;
2. $E[W(t) - W(s)] = 0$;
3. $\mathrm{Var}(W(t) - W(s)) = |t - s|$.

It turns out that these conditions are the basic requirements for the Brownian motion. We can use the symmetric random walk introduced above to formally develop a procedure to arrive at that process. First we consider the properties of the discrete process $M_n, n = 1, 2, \dots$ in terms of the increments $M_{k_{i+1}} - M_{k_i}$:

1. Non-overlapping increments as random variables are independent;

2. Each increment has mean zero;

3. Variance of the increment is $k_{i+1} - k_i$;

4. Martingale property: for $k < l$,

$$E_k[M_l] = E_k[M_l - M_k + M_k] = E_k[M_l - M_k] + M_k = M_k \tag{11}$$

5. Quadratic variance

$$\sum_{j=0}^{k-1} (M_{j+1} - M_j)^2 = X_1^2 + X_2^2 + \cdots + X_k^2 = k \tag{12}$$

This starts to look like the set of requirements for our process, except that it does not have a time scale. Next, we consider the scaled random walk:

$$W^{(n)}(t) = \frac{1}{\sqrt{n}} M_{nt} \tag{13}$$

for $t > 0$ such that nt is an integer, as the random walk introduced above is defined only for nonnegative integers as its index. For $W(t)$ as a function of t, or a path, we would need to fill the gap for t in between two neighboring integer indices. A natural choice is a piecewise linear interpolation to connect all the dots. This scaled random walk has all the properties of the original random walk, namely

1. Independent increments: for $0 = t_0 < t_1 < t_2 \cdots < t_m$ such that nt_j is an integer, $W^{(n)}(t_1) - W^{(n)}(t_0)$, $W^{(n)}(t_2) - W^{(n)}(t_1), \cdots W^{(n)}(t_m) - W^{(n)}(t_{m-1})$ are mutually independent from each other;

2. Mean and variance, for $0 \le s \le t$, we have

$$E[W^{(n)}(t) - W^{(n)}(s)] = 0, \quad \mathrm{Var}\left(W^{(n)}(t) - W^{(n)}(s)\right) = t - s \tag{14}$$

3. Martingale property, for $t > s$

$$E_s[W^{(n)}(t)] = W^{(n)}(s) \tag{15}$$

4. Quadratic variation

$$\sum_{j=0}^{k-1} (W^{(n)}(t_{j+1}) - W^{(n)}(t_j))^2 = n(t_{j+1} - t_j) \cdot \left(\frac{1}{\sqrt{n}}\right)^2 = t_{j+1} - t_j \tag{16}$$

Our goal is to study the limit of $W^{(n)}(t)$ as $n \to \infty$, which will be our construction of the Brownian motion $W(t)$. The convergence in this matter is in the sense of distribution, namely, $W^{(n)}(t)$ converges to $W(t)$ in distribution. The Brownian motion $W(t)$ enjoys the aforementioned properties, with a notable exception, that is $W(t)$ as a function of t is nowhere differentiable.

1.2 Some other properties of Brownian motion

Some other notations are used for Brownian motion, such as W_t, B_t, or $B(t)$. We can also use a two-point joint distribution to describe the process. For example, the covariance of $W(t)$ and $W(s)$ can be calculated as

$$E[W(s)W(t)] = E\left[W(s)\left(W(t) - W(s)\right)\right] + E[W^2(s)] = \mathrm{Var}(W(s)) = s, \quad \text{if } s < t \tag{17}$$

Therefore in general we have $E[W(s)W(t)] = \min(s,t) = s \wedge t$.
Suppose we have a sequence $t_1 < t_2 < \cdots < t_m$, the covariance matrix for $W(t_1), W(t_2), \dots, W(t_m)$ is

$$\begin{bmatrix} E[W^2(t_1)] & E[W(t_1)W(t_2)] & \dots & E[W(t_1)W(t_m)] \\ E[W(t_2)W(t_1)] & E[W^2(t_2)] & \dots & E[W(t_2)W(t_m)] \\ & \cdots & & \\ & & \dots & \\ E[W(t_m)W(t_1)] & E[W(t_m)W(t_2)] & \dots & E[W^2(t_m)] \end{bmatrix} = \begin{bmatrix} t_1 & t_1 & \dots & t_1 \\ t_1 & t_2 & \dots & t_2 \\ & \cdots & & \\ & & \dots & \\ t_1 & t_2 & \dots & t_m \end{bmatrix} \tag{18}$$

We can also describe Brownian motion from the moment generating function, which is

$$\begin{aligned} \phi(u_1, u_2, \dots, u_m) &= E\left[e^{i(W(t_1)u_1 + \cdots + W(t_m)u_m)}\right] \\ &= \exp\{\frac{t_1}{2}(u_1 + \cdots + u_m)^2 + \frac{t_2 - t_1}{2}(u_2 + \cdots + u_m)^2 + \\ &\quad \cdots + \frac{t_m - t_{m-1}}{2}u_m^2\} \end{aligned} \tag{19}$$

1.3 Quadratic variation and the dW notation

Now we have discussed Brownian motion and we can view stock price as some process based on Brownian motion. It is natural to consider processes that have a functional dependence on $W(t)$. Let $f(x)$ be a smooth function with all the derivatives we ask for, we now explore the following two aspects:

1. $f(W(T)) - f(W(0))$ expressed as an integral from 0 to T;

2. The differential $df(W(t))$ in terms of dt and $dW(t)$ terms.

Notice that the integrals and differentials will have to be reintroduced simply because $W(t)$ is not differentiable in t.

The integral and differential questions are closely related, as we know from Newton calculus, for example

$$f(W(T)) - f(W(0)) = \sum_{j=0}^{n-1} [f(W(t_{j+1}) - W(t_j)] = \sum (\Delta f) \tag{20}$$

How do we express Δf? A Taylor's expansion gives

$$(\Delta f)_j = f'_j \Delta W_j + \frac{1}{2} f''_j (\Delta W_j)^2 + \cdots \tag{21}$$

Here we include the second-order term just in case. But how do we interpret the following sums?

$$\sum f'_j \Delta W_j, \quad \text{and} \quad \sum f''_j (\Delta W_j)^2$$

If W is differentiable in t, we can easily relate this to $\int f' W' dt$. But here we would need to treat ΔW_j as a random variable. For this purpose, we should consider the total variations

$$\sum_{j=0}^{n-1} |f(t_{j+1}) - f(t_j)|^\alpha$$

where $\alpha = 1$ refers to the first-order variation $(FV_T(f))$ and $\alpha = 2$ refers to the quadratic (second-order) variation $([f, f](T))$, in the limit as $n \to \infty$.

(A). $\alpha = 1$

If $f \in C^1$, we have the first-order variation $FV_T(f) = \int_0^T |f'(t)|\, dt$. On the other hand, if $f(t) = W(t)$, $FV_T(W) = \infty$.

(B). $\alpha = 2$

The quadratic variation of a function $f(t)$ from 0 to T is

$$[f, f](T) = \lim_{||\Pi|| \to 0} \sum_{j=0}^{n-1} [f(t_{j+1}) - f(t_j)]^2 \tag{22}$$

where $\Pi = \{t_0, t_1, \ldots, t_n\}$ denotes a partition of $[0, T]$ and $||\Pi||$ refers to the largest subinterval of the partition. For smooth function f, we have $[f, f](T) = 0$. But then if we denote $Q_\Pi = \sum_{j=0}^{n-1} [W(t_{j+1}) - W(t_j)]^2$,

$$[W, W](T) = \lim_{||\Pi|| \to 0} Q_\Pi = T \tag{23}$$

To show this, we just need to verify $E[Q_\Pi] = T$ and $\mathrm{Var}(Q_\Pi) = 0$. The first part is quite straightforward

$$E[Q_\Pi] = \sum \mathrm{Var}\left(W(t_{j+1} - W(t_j)\right) = \sum \Delta t = T \tag{24}$$

For the second part, we need to use some properties of independent normal random variables,

$$
\begin{aligned}
\mathrm{Var}(Q_\Pi) &= \sum \mathrm{Var}(\Delta W_j)^2 \\
&= \sum \left[E[(\Delta W_j)^4] - \left(E[(\Delta W_j)^2] \right)^2 \right] \\
&= \sum \left[3(\Delta t)^2 - (\Delta t)^2 \right] \\
&= 2 \sum (\Delta t)^2 \\
&= 2T\Delta t
\end{aligned} \tag{25}
$$

We see that as $\Delta t \to 0$, $\mathrm{Var}(Q_\Pi) \to 0$.

In the limit $\Delta t \to 0$, it is customary to replace Δt with dt. But how about ΔW? The common notation gives

$$\Delta W \quad \text{“}\to\text{”} \quad dW(t)$$

in the sense that

$$\sum \Delta W \longrightarrow \int dW$$

and we often write

$$dW(t) \cdot dW(t) \quad \text{“}=\text{”} \quad dt$$

with an understanding that

$$E[dW(t) \cdot dW(t)] = dt, \quad \text{and} \quad \mathrm{Var}\left(dW(t) \cdot dW(t) \right) = 0$$

1.4 Itô's integral and Itô's formula

The Black-Scholes model is motived by the decomposition of the stock return over a very short period of time into a drift component (deterministic) and a fluctuation component driven by Brownian motion. As $\Delta t \to 0$, we expect the continuous time model

$$\frac{dS(t)}{S(t)} = \alpha\, dt + \sigma\, dW(t) \tag{26}$$

to be solved via integration. A temptation is to express the left-hand-side as $d(\log S(t))$ and then the integration step takes us to a solution for $S(t)$. Once we realize that $S(t)$ is driven by a Brownian motion, we will see that

$$d \log S(t) \neq \frac{dS(t)}{S(t)}$$

Our new challenge is to develop a calculus tool to address

$$df(W(t))$$

and

$$\int_0^t \Delta(s, W(s))\, dW(s)$$

for so-called adapted process $\Delta(t)$. Contrary to what we are used to in calculus, we will start with the integral. The reasons are: the integral can be interpreted as the limit of a sum of random variables; and it's possible to apply these limit theorems (such as the Law of Large Numbers and the Central Limit Theorem) in probability to obtain powerful results. A more subtle consideration is in the choice of the time to evaluate the integrand, as we will see in the following modeling of profit and loss (P and L) in a trading setting.

Profit and Loss: Suppose you trade a stock at times $t_0, t_1, \ldots$, and the stock prices at these times are $S_0, S_1, \ldots$, with number of shares at t_i equal to Δ_i, the Profit and Loss over one time period is $\Delta_i(S_{i+1} - S_i)$, and the total Profit and Loss (**PnL**) is

$$\sum_{i=0}^{n-1} \Delta_i(S_{i+1} - S_i)$$

The practice in trading is that you decide on your position (the number of shares to hold) $\Delta(t)$ at time t, at the price $S(t)$, given all the information available at that time, then *hold* it until the next time to trade $(t + \Delta t)$. At that time the price has changed to $S(t + \Delta t)$ so the change in the value is $\Delta(t)(S(t + \Delta t) - S(t))$. Once the new price is revealed and the PnL for this period is realized, you can choose a new position $\Delta(t + \Delta t)$, which may or may not be the same as $\Delta(t)$. You would need to use amount from other part of the portfolio to take the new position for the next time period which would cost you $\Delta(t + \Delta t)S(t + \Delta t)$.
If $\Delta(t)$ and $S(t)$ were deterministic smooth functions of t, the above sum would converge to the Stejes integral $\int \Delta(t)dS(t)$. In fact, the limit would be the same even if you evaluate Δ somewhere else in the time interval $[t, t + \Delta t]$. Obviously in the trading setting this does not make any sense, as you cannot go back in time to adjust your holdings for each period of time. This sits well with Itô's version of stochastic integral, namely, whenever we encounter ΔdS, it is assumed to be a limit of

$$\Delta(t)(S(t + \Delta t) - S(t))$$

This is in strong contrast with the notion of a Riemann sum that involves

$$\Delta(\xi)\,(S(t + \Delta t) - S(t))$$

where an arbitrary $\xi \in [t, t + \Delta t]$ can be used.
To appreciate this subtle point, we just need to remind ourselves that both Δ and S are random and we will need the expectation (most likely conditional) of the PnL. It is there that the independent increment property will come into play. Assuming independent increments, as intervals $[0, t)$ and $[t, t + \Delta t)$ do not overlap, we will have

$$E[\Delta(t)(S(t + \Delta t) - S(t))] = E[\Delta(t)] \cdot E[S(t + \Delta t) - S(t)]$$

At this stage we limit our processes to be driven by a Brownian motion $W(t)$ and address three specific cases of integrals that involve $W(t)$;

A. $\int_0^T W(t)dg(t)$, where $g(t)$ is a smooth non-random function

As $W(t)$ is continuous in t, this integral can still be defined through a Riemann sum

$$\lim_{\Delta t\to 0}\sum W(\xi_j)(g(t_{j+1})-g(t_j))$$

and we can justify it, for example, by the fact that its expectation and variance both make proper sense in the limit.

B. $\int_0^T f(t)\,dW(t)$, where $f(t)$ is a continuous non-random function of t

We can also look at the Riemann sum to see if it makes sense

$$\lim_{\Delta t\to 0}\sum f(\xi_j)(W(t_{j+1})-W(t_j))$$

This can be viewed as a linear combination of independent normal random variables. As f is continuous in t, the choices of ξ_j will not matter as we pass the limit. We can claim that this integral is also properly defined in the Riemann sense.

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

Now this is going to be quite different in that it involves products of random variables, once we discretize it into a sum. In Itô's convention, for each term in the sum $f(W)$ is taken at the beginning of the time period, while dW is taken to be the difference over the time period, so the expectation of the product can be represented by the product of expectations. Still, we will see surprising result from the following example.

Consider $f(x)=x$ so the integral in question is $\int_0^T W(t)\,dW(t)$ and we would expect the answer to be $W^2(T)/2$. However

$$E\left[\sum W_j(W_{j+1}-W_j)\right]=\sum E[W_j]\cdot E[W_{j+1}-W_j]=0$$

Here we use the notation $W_j=W(t_j)$ and the mean zero property of Brownian motion increments. The crucial step in the above is the independence of $W_{j+1}-W_j$ and $f(W_j)$ as they cover two non-overlapping time intervals. Imagine what will happen if you replace $W_j(W_{j+1}-W_j)$ with $W_{j+1}(W_{j+1}-W_j)$. Because of this observation, we know that our guess for the integral is wrong and the next guess is

$$\int_0^T W(t)\,dW(t)=\frac{1}{2}W^2(T)-\frac{T}{2} \tag{27}$$

because the expectation of $W^2(T)$ is the variance of $W(T)$ which is just T. It turns out that this time the answer is correct. To see how it is derived, a formal elementary derivation can be obtained by manipulating the finite sum $\sum W_j(W_{j+1}-W_j)$.

From now on we will focus on the Itô's integral $\int_0^T \Delta(t)\, dW(t)$, for any adapted $\Delta(t)$ (which may be random, but completely determined by time t), with the understanding that the integrand is always evaluated at the beginning of each time interval in the discrete version. We can also consider

$$I(t) = \int_0^t \Delta(s)\, dW(s)$$

as a process, once we let the upper limit in the integral to vary. This corresponds to the accumulated PnL up to time t in the trading setting. As a process, it enjoys many properties:

1. $I(t)$ is a martingale;
2. $E[I^2(t)] = \int_0^t E[\Delta^2(s)]\, ds$;
3. The quadratic variation $[I(t), I(t)] = \int_0^t \Delta^2(s)\, ds$.

Once we introduced the integral and process properly, we can start the discussion of Itô's formula. Let us begin with the Taylor expansion for $f(t,x)$ where f is continuously differentiable in t and twice continuously differentiable in x:

$$f(t+\Delta t, x+\Delta x) = f(t,x)+f_t\,\Delta t+f_x\,\Delta x+\frac{1}{2}f_{tt}(\Delta t)^2+\frac{1}{2}f_{xx}(\Delta x)^2+f_{tx}\,\Delta x\cdot\Delta t+\cdots \tag{28}$$

Suppose $x = X(t)$ is a smooth function of t, then the above expansion leads to

$$df = f_t\, dt + f_x X'(t)\, dt \tag{29}$$

when we take the limit as $\Delta t \to 0$. Suppose we take $x = W(t)$ which is not smooth at all and $W'(t)$ is not defined, we should be more careful with the expansion. In particular, we shall consider each term as a random variable and look at the mean and variance to see what order (in Δt) it turns up. The most interesting term is the one involving f_{xx} as the expectation of $(\Delta X)^2$ is just Δt that is on par with the order of the first derivative terms.
We therefore have the following Itô's formula for Brownian motion, assuming the differentiability of f as above.

(A). Integral form

$$f(T, W(T)) = f(0, W(0)) + \int_0^T f_t\, dt + \int_0^T f_x\, dW(t) + \frac{1}{2}\int_0^T f_{xx}\, dt \tag{30}$$

(B). Differential form

$$df = f_t\, dt + f_x\, dW(t) + \frac{1}{2}f_{xx}\, dt \tag{31}$$

For example, if $f(x) = \frac{1}{2}x^2,\ f_t = 0,\ f_x = x,\ f_{xx} = 1$, therefore

$$\frac{1}{2}W^2(T) = \int_0^T W(t)\, dW(t) + \frac{1}{2}\int_0^T dt$$

which gives us the result in (27).
Next we consider $X(t)$ to be driven by $W(t)$ via a stochastic differential equation

$$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))\, dW(t) \tag{32}$$

with the Black-Scholes model as a special example where μ and σ are constants. Itô's formula in this more general case can be written as follows.

(A). Integral form

$$\begin{aligned} & f(T, W(T)) \\ = \;& f(0, W(0)) + \int_0^T f_t\, dt + \int_0^T f_x\, dX(t) + \frac{1}{2}\int_0^T f_{xx}\,(dX)^2 \\ = \;& f(0, W(0)) + \int_0^T \left(f_t + \mu f_x + \frac{1}{2}\sigma^2 f_{xx}\right) dt + \int_0^T \sigma f_x\, dW(t) \end{aligned} \tag{33}$$

(B). Differential form

$$\begin{aligned} df &= f_t\, dt + f_x\, dX(t) + \frac{1}{2} f_{xx}\,(dX)^2 \\ &= \left(f_t + \mu f_x + \frac{1}{2}\sigma^2 f_{xx}\right) dt + \sigma f_x\, dW(t) \end{aligned} \tag{34}$$

Here we use a symbolic notation

$$(dX)^2 = (\mu dt + \sigma dW)^2 = \mu^2 (dt)^2 + 2\mu\sigma\, dt \cdot dW(t) + \sigma^2 (dW(t))^2 = \sigma^2 dt$$

Example: let $f(t, x) = \log x,\ f_t = 0, f_x = 1/x, f_{xx} = -1/x^2$, and we choose the process X:

$$dX = aX\, dt + bX\, dW(t)$$

Then

$$\begin{aligned} d\log X(t) &= \left(ax\frac{1}{x} + \frac{1}{2}b^2x^2\left(-\frac{1}{x^2}\right)\right) dt + bx\frac{1}{x}\, dW(t) \\ &= \left(a - \frac{b^2}{2}\right) dt + b\, dW(t) \end{aligned} \tag{35}$$

If we substitute the variable X by S, a, b by α and σ, we have

$$d\log S = \left(\alpha - \frac{\sigma^2}{2}\right) dt + \sigma\, dW(t) \tag{36}$$

Now we see what we have missed at the beginning of this section. In fact, we have

$$d \log S(t) = \frac{dS(t)}{S(t)} - \frac{\sigma^2}{2} dt \tag{37}$$

Actually Eq.(36) can now be solved by direct integration:

$$\log S(t) - \log S(0) = \left(\alpha - \frac{\sigma^2}{2} \right) t + \sigma W(t)$$

or

$$S(t) = S(0) \exp \left\{ \left(\alpha - \frac{\sigma^2}{2} \right) t + \sigma W(t) \right\} \tag{38}$$

2 Black-Scholes-Merton PDE and dynamic hedging

2.1 Derivation of Black-Scholes-Merton PDE

With a brief preparation in Itô's calculus, we are now ready to derive the Black-Scholes-Merton PDE for a European call price $C(t, S(t))$. The ansatz is that the call price at time t depends on the stock price at time t only, and this is justified by an no-arbitrage argument. Our major assumptions are

1. The underlying stock price follows the process

$$\frac{dS}{S} = \alpha dt + \sigma dW(t), \tag{39}$$

2. The risk-free interest r is a constant;
3. Supplies are unlimited and short selling (you sell something you don't own by borrowing) is allowed;
4. No transaction cost.

The approach is to set up a portfolio and a strategy so that the portfolio tracks the call price no matter whether the stock price itself moves up or down. This can be done in the following steps:

1. set up the portfolio by a proper combination of stock and money market deposit, calculate its change in time;
2. calculate the change in the call price;
3. set these changes equal to each.

At time t, suppose we have a total of $X(t)$ to invest. We invest in $\Delta(t)$ shares of the underlying stock, which comes to $\Delta(t)S(t)$ in value, then put the rest in a money market which earns a risk-free interest at rate r. If the stock value is over $X(t)$ then we need to borrow and we assume the rate is also r. We write as

$$X(t) = \Delta(t)S(t) + (X(t) - \Delta(t)S(t))$$

As time passes by Δt, we now have

$$X(t + \Delta t) = \Delta(t)S(t + \Delta t) + (X(t) - \Delta(t)S(t))(1 + r\Delta t)$$

In differential terms, we have

$$\begin{aligned} dX(t) &= \Delta dS + r(X - \Delta S)\,dt \\ &= \Delta(\alpha S\,dt + \sigma S\,dW) + r(X - \Delta S)\,dt \\ &= (\alpha - r)\Delta S\,dt + rX\,dt + \sigma\Delta S\,dW \end{aligned} \tag{40}$$

We can also use Itô's formula to differentiate the discounted stock price

$$\begin{aligned} d(e^{-rt}S(t)) &= -re^{-rt}S\,dt + e^{-rt}dS \\ &= (\alpha - r)e^{-rt}S\,dt + e^{-rt}\sigma S\,dW \end{aligned} \tag{41}$$

These two combine to give

$$\begin{aligned} d(e^{-rt}X(t)) &= \Delta \cdot d(e^{-rt}S(t)) \\ &= (\alpha - r)e^{-rt}\Delta \cdot S\,dt + e^{-rt}\sigma\Delta \cdot S\,dW \end{aligned} \tag{42}$$

Next we look at the change in the option price

$$dC(t, S(t)) = \left(C_t + \alpha S C_S + \frac{1}{2}\sigma^2 S^2 C_{SS}\right) dt + \sigma S C_S\,dW \tag{43}$$

and

$$d(e^{-rt}C(t, S(t))) = e^{-rt}\left(C_t + \alpha S C_S + \frac{1}{2}\sigma^2 S^2 C_{SS} - rC\right) dt + e^{-rt}\sigma S C_S\,dW \tag{44}$$

Comparing Eq.(42) with Eq.(44), in order for them to match, we first need to set

$$\Delta = \frac{\partial C}{\partial S} \tag{45}$$

to eliminate the dW terms, and then we arrive at the celebrated Black-Scholes-Merton PDE

$$C_t + rSC_S + \frac{1}{2}\sigma^2 S^2 C_{SS} - rC = 0. \tag{46}$$

As we know for any PDE problem, it comes with some initial and/or boundary conditions. For a European call, we have a condition at $t = T$:

$$C(T, S) = \max(S - K, 0) = (S - K)^+ \tag{47}$$

The region in $t - S$ plane to solve this equation is $0 < t < T$ and $S > 0$. If you know something about heat equation, you would notice that this is backward in time, namely you set the terminal condition at $t = T$ and solve backward in time to obtain the solution $C(t, S)$ for $0 < t < T$. In practice, $C(0, S)$ gives us the call price at $t = 0$ when the observed underlying stock price is S. To solve this PDE problem, we note that it is still a linear equation and the equation can be converted to the standard heat equation after several changes of variables. First we introduce $x = \log S$ and the equation becomes a constant coefficient problem

$$C_t + (r - \frac{1}{2}\sigma^2)C_x + \frac{1}{2}\sigma^2 C_{xx} = rC \tag{48}$$

Then another change of variable will take it to

$$C_\tau = \frac{1}{2}\sigma^2 C_{uu} \tag{49}$$

For the European call, the solution for any $0 < t < T$ is

$$C(t, S) = SN(d_1) - Ke^{-r(T-t)}N(d_2) \tag{50}$$

where $N(x)$ is the normal cumulative distribution function, and

$$d_1 = \frac{\log\left(\frac{S}{K}\right) + (r + \frac{1}{2}\sigma^2)(T - t)}{\sigma\sqrt{T - t}}, \quad d_2 = d_1 - \sigma\sqrt{T - t} \tag{51}$$

2.2 Dynamic hedging

The idea behind this approach is quite profound: it says that an option can be **replicated** by a dynamically adjusted portfolio that consists of some shares of the underlying stock and a money market account. The number of shares Δ is calculated from the partial derivative of $C(t, S)$ with respect to S, and that function $C(t, S)$ is the solution of the Black-Scholes-Merton PDE supplied with the call terminal condition. If we have all the questions answered and follow this so-called dynamic hedging scheme, the portfolio will match *exactly* the option value, no matter what the stock price later becomes. The final punch line is that this current portfolio value must be the same as the current option value, therefore the cost of the portfolio is the original price of the option.

This strategy to use the underlying stock, with the number of shares dynamically adjusted according to the partial derivative of the option price with respect to the underlying, is called dynamic hedging, and it is the most common strategy employed in practice to hedge derivative positions. It is simple to follow but the cost can be substantial as the position needs to be frequently adjusted - in practice we can only do that much and there will be hedging errors incurred.

Another angle to look at dynamic hedging is to look for a portfolio that has the stock risk factor eliminated. Consider the following positions at time t:

1. One share of the call;

2. $-\Delta$ shares of the underlying stock;

3. A loan in the amount $M(t) = Ke^{-r(T-t)}N(d_2)$ at interest rate r.

The total value of the portfolio at t is

$$P = C(t, S(t)) - \Delta \cdot S(t) + M(t)$$

Now suppose that everything else being held, and just the stock price moves by a small amount, the change in P is the change in stock price multiplied by

$$\frac{\partial P}{\partial S} = \frac{\partial C}{\partial S} - \Delta = 0$$

if we choose Δ to be $\partial C/\partial S$. When this partial derivative is zero, we say that this portfolio is **delta-neutral**.

2.3 Self-financing portfolio

Suppose you are putting together a portfolio that consists of two kinds of asset with prices U and V, each with number of shares/units ϕ and ψ respectively, the total value of the portfolio at time t is

$$X(t) = \phi(t)U(t) + \psi(t)V(t) \tag{52}$$

If we compute the differential in time, with Itô's calculus in mind, we should have

$$dX(t) = \phi dU + U d\phi + d\phi \cdot dU + \psi dV + V d\psi + d\psi \cdot dV. \tag{53}$$

However, if the change is caused only by the price changes in U and V, then just two terms are needed

$$dX(t) = \phi dU + \psi dV \tag{54}$$

In a dynamic trading setting, the numbers of shares ϕ and ψ are also variables so the above is in general not valid. However, if we are restricted to a certain trading practice we can still ensure that the above is valid. For example, we can imagine that for each trading period the numbers of shares are held, only to be changed after Δt, and when we change the numbers of shares we make sure the total value remains the same. The implication of this statement is that whatever gain/loss from one asset is reinvested/compensated in the other, and no other resources are called in to make up the difference. Portfolios with this property (54) are called *self-financing portfolios*. To see how this practice works, notice that by the time you readjust the positions (changing ϕ and ψ), the updated portfolio value is

$$X(t+\Delta t) = \phi(t)U(t+\Delta t) + \psi(t)V(t+\Delta t), \tag{55}$$

but then we will change ϕ from $\phi(t)$ to $\phi(t+\Delta t)$, and ψ from $\psi(t)$ to $\psi(t+\Delta t)$, with the restriction that no fund is added or taken away from the portfolio, so you have the same total value

$$X(t+\Delta t) = \phi(t+\Delta t)U(t+\Delta t) + \psi(t+\Delta t)V(t+\Delta t) \tag{56}$$

This requires us to balance

$$(\phi(t+\Delta t)-\phi(t))U(t+\Delta t)+(\psi(t+\Delta t)-\psi(t))V(t+\Delta t)=0. \tag{57}$$

This suggests that as $\Delta t \to 0$, we should have

$$Ud\phi + Vd\psi = 0. \tag{58}$$

You may be wondering what happen to those products of differentials. Here we should point out that (58) corresponds to

$$(\phi(t+\Delta t)-\phi(t))U(t)+(\psi(t+\Delta t)-\psi(t))V(t)=0, \tag{59}$$

and the difference between (57) and (59) actually explains the missing terms $d\phi \cdot dU$ and $d\psi \cdot dV$.

Here is a numerical example with self-financing. Suppose the stock price at time t_0 is $S(t_0) = \$10$, and the money market unit value is $M(t_0) = \$1$. We hold $\Delta(t_0) = 2$ shares of the stock, $\Gamma(t_0) = 15$ units of the money market fund at that time. The total portfolio value at t_0 is

$$X(t_0) = 2 \times 10 + 15 \times 1 = \$35$$

Next we are at time t_1 and the stock price has moved to \$11 and the money market unit price is now $M(t_1) = 1.01$. Suppose that our trading strategy tells us that we should now hold $\Delta(t_1) = 3$ shares of the stock. The question is how much money we would have to move from the money market account. First, there is no change in the holdings between t_0 and t_1 so the portfolio value at t_1 is

$$X(t_1) = 2 \times 11 + 15 \times 1.01 = \$37.15$$

At t_1, we are required to adjust our stock positions because the new delta is 3, but there is no money taken away from the portfolio, nor any injected. So we must have

$$X(t_1) = 3 \times 11 + \Gamma(t_1) \times 1.01 = \$37.15$$

This is an equation for Γ and we have $\Gamma(t_1) \approx 4.11$ units of the money market, which is to say that we will need to withdraw 10.89 units from the money market (with an amount of \$11 to buy one more share of the stock). As we see that self-financing poses some restrictions on how you can trade.

2.4 Black-Scholes-Merton PDE derived from a riskless portfolio

There is another angle to derive the BSM PDE, which is again based on the no-arbitrage principle with a point of view that a riskless portfolio should earn just the riskless interest rate. We will proceed to construct such a portfolio based on the stock and a call option on the stock. As the stock and its call option move in the same direction but different proportions, it is natural to construct a portfolio with one share of the call, and short Δ shares of the underlying stock, so the value at t is

$$N(t) = C(t, S(t)) - \Delta(t) \cdot S(t) \tag{60}$$

as we see from section 2.2, this portfolio will be delta neutral for small changes in the stock price, if we choose the right Δ (in this case we can guess that it is $\partial C/\partial S$), we should be able to offset the gain/loss from the stock and call options. The trouble is that we will need to rebalance the positions after the price change. However, the rebalance should be made based on the requirement that no fund is taken out, or injected into the portfolio, which comes to the self-financing concept discussed in section 2.3. First let us compute the instantaneous changes:

$$dN(t) = dC - \Delta \cdot dS - S \cdot d\Delta - dS \cdot d\Delta \tag{61}$$

and

$$dC = C_t dt + C_S dS + \frac{1}{2} C_{SS} (dS)^2$$

If we require self-financing so $S \cdot d\Delta + dS \cdot d\Delta = 0$, then

$$dN(t) = \left(C_t + \alpha S C_S - \alpha S \Delta + \frac{\sigma^2}{2} S^2 C_{SS} \right) dt + (\sigma S C_S - \sigma S \Delta) dW(t)$$

If we want to eliminate the risk in S which appears in W, we should make sure that there is no dW term present in $dN(t)$, this suggests that we should choose

$$\Delta = C_S = \frac{\partial C}{\partial S}. \tag{62}$$

On the other hand, now that dN has no random component, it should just earn the riskfree interest rate. That is

$$dN(t) = rN(t)dt.$$

Using our calculated dN, we have

$$C_t + \alpha S C_S - \alpha S \Delta + \frac{\sigma^2}{2} S^2 C_{SS} = r(C - S\Delta) = r(C - SC_S)$$

or

$$C_t + rSC_S + \frac{1}{2} \sigma^2 S^2 C_{SS} = rC \tag{63}$$

This is the same equation as (46) derived in section 2.1.
We can summarize these two approaches that yield the same partial differential equation. One is to form a portfolio to replicate the call option ($X(t) \to C(t, S(t))$), where we want to make sure that X has the same value as C no matter what happens to the stock price in time, and in the end $X(T) = (S(T) - K)^+$ as promised. The no-arbitrage argument shows that $X(0)$ should be the same as $C(0, S(0))$, therefore the price of call option at any time before the expiration can be computed based on the number of shares of the stock, the stock price at

that time, and the money market balance at that time. The other approach is to mix the stock with the call so that the risk in each offsets the other. The no-arbitrage argument in this case says that a riskless portfolio is nothing but a money market account that earns the riskfree interest rate. We know that the answer is $N(t) = C(t) - \Delta(t) \cdot S(t)$ where Δ is the rate of change in the option to the change in stock price. Both arguments will lead to the same PDE (63) for the option price.

3 Martingale and risk-neutral pricing

Martingales are a class of stochastic processes that are first used to describe certain gambling games where your future gain/loss standing is not expected to improve or deteriorate based on your current standing. In our notation, a process $M(t)$ is a martingale if it satisfies

$$E_t\left[M(T)\right] = E\left[M(T)|\mathcal{F}_t\right] = M(t) \tag{64}$$

for any $t < T$. The most famous example is no doubt the Brownian motion, because

$$E_t[W(T)] = E_t[W(T) - W(t) + W(t)] = E_t[W(T) - W(t)] + W(t) = W(t),$$

as we used the increment property of Brownian motion. Of course martingales are not limited to Brownian motions, but they are derived from Brownian motion in the so-called martingale representation theorem, which says that any martingale $M(t)$ can be related to $W(t)$ either in the integral form

$$M(t) = \int_0^t \Delta(s)\, dW(s), \tag{65}$$

or in the differential form

$$dM(t) = \Delta(t)\, dW(t). \tag{66}$$

Here $\Delta(t)$ can be stochastic, but must be so-called adapted, namely its value must be determined by time t. In the language of measure theory, $\Delta(t)$ is adapted if it is measurable with respect to the filtration $\mathcal{F}_t$.

3.1 Advantage of being a martingale

The advantage of being a martingale is precisely in the benefit of (64). Imagine if we can claim that the stock price process is a martingale, we can just use

$$S(t) = E_t[S(T)]$$

to determine the price at t based on what you expect from the company at later time T. This is actually the basis for one major approach to price stocks.

How do we verify that some process is a martingale? At least in theory this question can be answered easily by the equation (66): we just need to look at the differential of the process and see if it is just some adapted process multiplied by $dW(t)$. The other way to answer the question is to see if the dt term is zero.

3.2 Stock price as a martingale and pricing derivatives

Suppose that the stock price follows the geometric Brownian motion process

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

If $\alpha = 0$, we see that $S(t)$ is indeed a martingale. If we construct a portfolio that consists of Δ shares of the stock and some amount in the money market, we write

$$X(t) = \Delta(t)S(t) + (X(t) - \Delta(t)S(t))$$

and we trade in the fashion that

$$dX(t) = \Delta(t)\, dS(t) + r(X(t) - \Delta(t)S(t))\, dt.$$

If we have $r = 0$, then for whatever adapted $\Delta(t)$, $X(t)$ will be a martingale since

$$dX(t) = \alpha\Delta(t) \cdot S(t)\, dW(t). \tag{67}$$

If we can choose a particular $\Delta(t)$ to lead to a portfolio with value $X(t)$ that ends up with the payoff function $F(S(T))$ when it reaches T (this remains to be verified), then we have $X(0)$ as the price of the derivative with that payoff. The martingale property just allows us to price

$$V(0) = X(0) = E[X(T)] = E[F(S(T))] \tag{68}$$

Several questions remain: how to make sure that we can have a portfolio that gives whatever the derivative ends up with, no matter what happens to the underlying stock? More importantly, even if we can answer that question, how do we extend this pricing methodology to more general situations where $\alpha \neq 0$ and $r \neq 0$?

3.3 Change of measure - Girsanov theorem

We begin with a discrete example: suppose we have a sample space $\Omega = \{H, T\}$, and a probability measure

$$P(H) = P(T) = \frac{1}{2}. \tag{69}$$

Now we introduce another probability measure $\tilde{P}$ where the probabilities assigned to the events are modified:

$$\tilde{P}(H) = \frac{1}{3}, \quad \tilde{P}(T) = \frac{2}{3}. \tag{70}$$

These two probabilities measures are considered equivalent in probability theory as

$$0 < Z(\omega) = \frac{\tilde{P}(\omega)}{P(\omega)} < \infty, \quad \omega = H, T \tag{71}$$

This condition guarantees that these two measures agree on what are possible and what are not possible, even though they could differ in the actual value. It is impossible to have an event that has zero probability while a finite (between 0 and 1) probability in another measure. With this new measure $\tilde{P}$, the expected value of a random variable $X(\omega)$ is no longer its usual average, instead

$$\begin{aligned}
\tilde{E}[X] &= X(H)\tilde{P}(H) + X(T)\tilde{P}(T) \\
&= X(H) \cdot \frac{\tilde{P}(H)}{P(H)} \cdot P(H) + X(T) \cdot \frac{\tilde{P}(T)}{P(T)} \cdot P(T) \\
&= E[X \cdot Z]
\end{aligned}$$

where Z as defined in (71) is called Radon-Nikodým derivative of $\tilde{P}$ with respect to P, or more intuitively, the relative weights placed on a particular event:

$$Z(H) = \frac{2}{3}, \quad Z(T) = \frac{4}{3}.$$

Here we see that H has relatively less weight under $\tilde{P}$ and T has more weight under $\tilde{P}$. As both probability measures are constrained by the requirement that the total probability is one, we must have

$$E[Z] = 1. \tag{72}$$

Next we consider the example that is our focus, the one related to the normal distribution. Suppose $\Omega = R$, $X \sim N(0,1)$ is a standard normal random variable under P. If we consider the new random variable

$$\tilde{X} = X + \mu \sim N(\mu, 1)$$

which is certainly not a standard normal random variable under P. Actually the mean is μ so those with positive values come with higher probabilities if $\mu > 0$. What if we want to readjust the probabilities so that the expected value of $\tilde{X}$ is zero again, under a new probability measure? An obvious choice is to decrease the probabilities for larger values and increase the probabilities for smaller values. Namely, we should choose a new probability measure $\tilde{P}$ such that the resulting Radon-Nikodým derivative Z takes on values smaller than one for large x values and larger values for small x values. In particular, we have $\tilde{X}$ under the original measure

$$dP = P\left\{\tilde{X} \in (x, x+dx)\right\} = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}}\, dx$$

and hope that under the new measure

$$d\tilde{P} = \tilde{P}\left\{\tilde{X} \in (x, x+dx)\right\} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx$$

so

$$Z(X) = \frac{d\tilde{P}(X)}{dP(X)} = e^{-\mu X - \frac{1}{2}\mu^2} \tag{73}$$

Again, this $Z(X)$ is called the Radon-Nikodým derivative, and we can verify that the ratio is always between 0 and ∞ for $-\infty < X < \infty$. We can also check

$$\begin{aligned} E\left[\tilde{X}Z\right] &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(x+\mu)e^{-\mu x-\frac{\mu^2}{2}}\cdot e^{-\frac{x^2}{2}}\,dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(x+\mu)e^{-\frac{(x+\mu)^2}{2}}\,dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}x'e^{-\frac{x'^2}{2}}\,dx' \\ &= \tilde{E}[\tilde{X}] \end{aligned} \tag{74}$$

In case we take the conditional expectation of Z, we will get a new process, as conditioning for different t leads to a different random variable, so we define

$$Z(t) = E_t[Z] \tag{75}$$

as the Radon-Nikodým derivative process.
The following result is more subtle: when Y depends on the uncertain information only after t, or in the language of measure theory, Y is measurable with respect to $\mathcal{F}_t$, then

$$\tilde{E}[Y] = E[YZ(t)] \tag{76}$$

The interpretation of this result is that the ratio correction is only accounted for those after time t, because what happened before t is considered known already.
Given the necessary background in a very brief way discussed above, we can state Girsanov theorem in a simple and intuitive form. Suppose we have $W(t)$ a Brownian motion under the measure P, and let us assume that $\Theta(t)$ is a given function of t, then in order to make

$$\tilde{W}(t) = W(t) + \int_0^t \Theta(u)\,du \tag{77}$$

a Brownian motion under $\tilde{P}$, we just need to introduce a RN derivative process

$$Z(t) = \exp\left\{-\int_0^t \Theta(u)\,dW(u) - \frac{1}{2}\int_0^t \Theta(u)^2\,du\right\} \tag{78}$$

that defines the ratio of the density functions. We should be able to spot the analogy between this ratio and that in (73).

3.4 Change of measure in derivative pricing

Let us start with our stock price model

$$\frac{dS}{S} = \alpha\,dt + \sigma\,dW(t)$$

where α is the expected growth rate which we hope to be able to go without it. It is desirable if we can just call the right-hand-side to be $\sigma\, d\tilde{W}(t)$, where $\tilde{W}$ is a Brownian motion under another probability measure $\tilde{P}$. Using Girsanov theorem, we identify here that we should choose $\Theta = \alpha/\sigma$ so

$$Z(t) = \exp\left\{-\frac{\alpha}{\sigma}W(t) - \frac{1}{2}\frac{\alpha^2}{\sigma^2}t\right\} \tag{79}$$

The advantage is that now $S(t)$ will be a martingale under $\tilde{P}$, and then any

$$V(t) = \tilde{E}_t\left[V(T)\right] = \tilde{E}_t\left[F(S(T))\right] \tag{80}$$

as a process will also be a martingale under $\tilde{P}$, because it is a conditional expectation. We just need to show that V defined as such will be the price of the derivative with payoff $F(S(T))$. The argument goes like this: first we construct a class of portfolios with total value $X(t)$ at time t, assuming the risk free interest rate $r = 0$. In this class of portfolios, there are two components, one with $\Delta(t)$ shares of the underlying stock, and the other consisting of an amount of $X(t) - \Delta(t)S(t)$ in the money market (with interest rate $r = 0$). The change over an infinitesimal period of time is

$$dX(t) = \Delta(t)dS(t) + r(X(t) - \Delta(t)S(t)) = \Delta(t)dS(t) = \Delta(t)\cdot\sigma\cdot S(t)\, d\tilde{W}(t)$$

ensures that $X(t)$ is a martingale under $\tilde{P}$. Here we also assumed self-financing in the trading. Our arguments would like to show that $X(t)$ can be used to replicate a derivative, guaranteed by the use of martingale representation theorem, so $X(t) = V(t)$ is the price of the derivative since it gets us $F(S(T))$ no matter what happens to $S(T)$.
Here are the detailed steps:

1. We just showed that $X(t)$ traded as described (two types of assets, self-financing) is a martingale under $\tilde{P}$;

2. $V(t) = \tilde{E}_t[V(T)]$ as a conditional expectation is always a martingale under $\tilde{P}$;

3. Apply the martingale representation theorem so we know that V can be related to $\tilde{W}$ through

   $$dV(t) = \Gamma(t)\, d\tilde{W}(t)$$

   for some adapted $\Gamma(t)$;

4. Let $\Delta = \Gamma/(\sigma S(t))$, which tells us how to construct the X, so we have

   $$dV(t) = dX(t)$$

   If we start with $X(0) = V(0)$, then $dV = dX$ leads to

   $$X(T) = V(T)$$

so it does replicate the derivative, therefore by the no-arbitrage argument that two assets doing the same thing must have the same price,

$$V(0) = X(0)$$

as defined in (80) is the price of the derivative in question.

As we see in the pricing formula, the price is given as a conditional expectation of the payoff under $\tilde{P}$. Once we have the distribution of S under this measure, we don't have to go back to the expectation under the original measure P. The question is, what is the new measure? In the case $r = 0$ we discussed above, the instantaneous return is a Brownian motion multiplied by σ, in another word, the expected grown rate is exactly zero and α is no where to be seen. Once we take care of the complication with $\alpha \neq 0$, we should be able to treat the case $r > 0$. This is done using the notion of discount. Let us define the discounted stock price and the discounted derivative price (everything in the future is discounted to the time $t = 0$ using the interest rate r).

$$\tilde{S}(t) = e^{-rt}S(t), \quad \tilde{V}(t) = e^{-rt}V(t). \tag{81}$$

Then

$$\begin{aligned} d\tilde{S} &= e^{-rt}dS - re^{-rt}Sdt \\ &= e^{-rt}\left(dS - rSdt\right) \\ &= \tilde{S}\left((\alpha - r)\,dt + \sigma\,dW(t)\right) \end{aligned} \tag{82}$$

If we introduce $\tilde{W}$ such that

$$d\tilde{W} = \frac{\alpha - r}{\sigma}dt + dW(t) \tag{83}$$

then $\tilde{S}$ will be a martingale under $\tilde{P}$. This parameter $\Theta = (\alpha - r)/\sigma$ is called the market price of risk. As $\alpha - r$ is the excess return generated by the stock and σ measure the level of risk, the ratio gives the expected excess return for a unit level of risk taken, so this ratio is called market price of risk. The RN process involved is

$$Z(t) = \exp\left\{-\frac{\alpha - r}{\sigma}W(t) - \frac{1}{2}\left(\frac{\alpha - r}{\sigma}\right)^2 t\right\}. \tag{84}$$

With this notation,

$$\tilde{S}(0) = \tilde{E}\left[\tilde{S}(t)\right] = \tilde{E}\left[e^{-rt}S(t)\right] \tag{85}$$

and we can express the option pricing as

$$V(0) = \tilde{V}(0) = \tilde{E}\left[e^{-rt}V(t)\right] \tag{86}$$

In general we have the pricing formula in terms of the risk-neutral probability measure

$$V(t) = \tilde{E}_t\left[e^{-r(T-t)}V(T)\right] = \tilde{E}_t\left[e^{-r(T-t)}F(S(T))\right] \tag{87}$$

This probability measure $\tilde{P}$ is called the risk-neutral probability measure, as it describes a particular world in which the probabilities are assigned such that every stock has the same expected growth rate r, or in another word, every discounted stock price is a martingale. Different stocks are different only in that their volatilities σ are different. This is quite counter-intuitive to most investors: typically higher returns are expected for higher risk levels, meaning that the expected growth rate does depend on σ. This tendency can be explained by the attitudes of the those so-called risk-seeking investors who would love to take risks to generate higher returns. If the expected growth is the same r for every investor in that world, it only means that all investors are indifferent to the risks involved, as long as they generate the same expected growth rate which is the same riskfree rate r. This is why that world is called a risk-neutral world as investors in that world just don't care about the risk and they do not demand extra returns for compensation of the risks they are required to take.

3.5 Change of measure in pricing of exotic options

Here we give one example to show another application of the change of measure technique: deriving a formal pricing formula for some exotic options. Exotic means that these options are path-dependent and the pricing of such derivatives usually would require Monte Carlo simulations, except for some special cases where some clever tricks can be found. The example we consider here, the down-and-out call barrier option, is one of such lucky cases. The payoff function for this derivative can be written as

$$F = (S(T) - K)^+ \cdot I_{\underline{S}(T) \geq H} \tag{88}$$

for some barrier $H < K$. Here I_A is the indicator function of event A (taking value 1 if A occurs and 0 otherwise), and the process

$$\underline{S}(t) = \min_{u \leq t} S(u) \tag{89}$$

gives the minimum price S achieved so far up until t.

This is a call option with a twist: it gives the holder the right to purchase the stock for the price K at time T, *provided* that the price at no time before t falls below a specified level H. With the risk-neutral pricing methodology, we claim the price at time $t < T$ is

$$\begin{aligned} V(t) &= e^{-r(T-t)} \tilde{E}_t \left[(S(T) - K)^+ \cdot I_{\underline{S}(T) \geq H} \right] \\ &= e^{-r(T-t)} \tilde{E}_t \left[(S(T) - K) \cdot I_U \right] \\ &= e^{-r(T-t)} \tilde{E}_t \left[S(T) I_U \right] - e^{-r(T-t)} K \tilde{E}_t \left[I_U \right] \\ &= e^{-r(T-t)} \tilde{E}_t \left[S(T) I_U \right] - e^{-r(T-t)} K \tilde{P}_t \left[U \right] \end{aligned} \tag{90}$$

where $U = \{S(T) \geq K, \underline{S}(T) \geq H\}$ is the event where a positive payoff is realized.

Without loss of generality, we consider the case $t = 0$ for the initial price of the option, and there are two expected values to be computed. To illustrate the methodology, we consider only the second one (the more obvious one)

$$\tilde{E}\left[I_U\right] = \tilde{P}\left\{U\right\} = \tilde{P}\left\{S(T) \geq K, \underline{S}(T) \geq H\right\} \tag{91}$$

and S is assumed to be given by the solution from the Black-Scholes model

$$S(t) = S(0)e^{Y(t)}, \quad Y(t) = \left(r - \frac{\sigma^2}{2}\right)t + \sigma\tilde{W}(t).$$

There is an advantage in describing in terms of Y, rather than the original S, as the event U can now be described as

$$Y(T) \geq \log\left(\frac{K}{S(0)}\right), \quad m^Y(T) \geq \log\left(\frac{H}{S(0)}\right), \tag{92}$$

which is the same as

$$-Y(T) \leq \log\left(\frac{S(0)}{K}\right), \quad M^{-Y}(T) \leq \log\left(\frac{S(0)}{H}\right). \tag{93}$$

Here we use the notation

$$m^Y(t) = \min_{0\leq u\leq t} Y(u), \quad M^Y(t) = \max_{0\leq u\leq t} Y(u). \tag{94}$$

To simplify our notation, in the following we omit the tilde sign in expectation and probability calculations. For the probability (91), consider first the special case where $r - \sigma^2/2 = 0$, since here we just have $-Y(t) = \sigma W(t)$ so the event in question is

$$U = \left\{W : W(T) \leq b, \max_{t\leq T} W(t) \leq c\right\} \tag{95}$$

for $b = \frac{1}{\sigma}\log\left(\frac{S(0)}{K}\right) < c = \frac{1}{\sigma}\log\left(\frac{S(0)}{H}\right)$.

This probability can be immediately evaluated with the help of the **reflection principle** for Brownian motions, which says that for each branch of a Brownian path, there is an equally likely path that mirrors it. Suppose that we have a path that starts at $W(0) = 0$, and gets to $W(t) = c$ at time some $t > 0$, and eventually lands below $b < c$ at time $T > t$, by a distance $c - b$. The reflection principle suggests that there is equally likely to be a path that agrees with the previous one up to time t, but continues to go up at T to land above, also by a distance $c - b$, or $W(T) > c + (c - b) = 2c - b$, which is the same as $W(T) > 2c - b$ without the condition that W crosses c at some time t, as the condition is automatically satisfied. To be more specific,

$$\begin{aligned} & P\left\{W(T) \leq b, M^W(T) > c\right\} \\ = \; & P\left\{W(T) \geq 2c - b, M^W(T) > c\right\} \\ = \; & P\left\{W(T) \geq 2c - b\right\} \end{aligned} \tag{96}$$

This can be used to calculate

$$\begin{aligned}
& P\left\{W(T) \le b, M^W(T) \le c\right\} \\
= \;& P\left\{W(T) \le b\right\} - P\left\{W(T) \le b, M^W(T) > c\right\} \\
= \;& P\left\{W(T) \le b\right\} - P\left\{W(T) > 2c - b\right\} \\
= \;& \Phi\left(\frac{b}{\sqrt{T}}\right) - \left(1 - \Phi\left(\frac{2c-b}{\sqrt{T}}\right)\right) \\
= \;& \Phi\left(\frac{b}{\sqrt{T}}\right) - \Phi\left(\frac{b-2c}{\sqrt{T}}\right) \qquad (97)
\end{aligned}$$

Using this result, for the probability we want to evaluate,

$$P(U) = \Phi\left(\frac{\log\frac{S(0)}{K}}{\sigma\sqrt{T}}\right) - \Phi\left(\frac{\log\frac{S(0)}{K} - 2\log\frac{S(0)}{H}}{\sigma\sqrt{T}}\right). \qquad (98)$$

The main step is to extend this to the case $r - \sigma^2/2 \neq 0$. For this we will need a change of measure so that

$$\frac{1}{\sigma}Y(t) = \left(\frac{r}{\sigma} - \frac{\sigma}{2}\right)t + W(t)$$

is a Brownian motion under another probability measure Q, in which we can apply the reflection principle. This calculation is quite long and we will only sketch the procedure. First we notice that if we introduce $W_1 = W + \mu t$, the event

$$U = \left\{W_1 : W_1(T) \le b, M^{W_1}(T) \le c\right\}$$

is not the same as

$$\left\{W(T) \le b - \mu T, M^W(T) \le c - \mu T\right\}$$

because

$$\max_{t \le T} W_1(t) = \max_{t \le T}\left(W(t) + \mu t\right) \neq \max_{t \le T} W(t) + \mu T.$$

We should be reminded that we need to calculate $P\{U\} = E[I_U]$.
Let us introduce the joint distribution function for random variables $X_1 = X(T)$ and $X_2 = M^X(T)$

$$F^X(T, b, c) = P\left\{X_1 \le b, X_2 \le c\right\}, \qquad (99)$$

with the joint density function

$$f^X(T, b, c) = \frac{\partial^2}{\partial b\,\partial c}F^X(T, b, c). \qquad (100)$$

For $W_1 = W + \mu t$, we can introduce the RN process

$$Z(t) = \exp\left\{-\mu W(t) - \frac{1}{2}\mu^2 t\right\} = \frac{dQ}{dP},$$

so that W_1 will be a Brownian motion under the new measure Q, which means

$$P^Q\{U\} = F^{W_1}(T,b,c) = \Phi\left(\frac{b}{\sqrt{T}}\right) - \Phi\left(\frac{b-2c}{\sqrt{T}}\right) \tag{101}$$

under Q and we can calculate

$$P\{U\} = E[I_U] = E^Q\left[I_U Z^{-1}\right]. \tag{102}$$

This can be written as an integral once the joint density function for W_1 and M^{W_1} under Q is known, which is calculated by differentiating (101). After representing the integrand in terms of W_1 and using the joint density function, we arrive at

$$P\left\{W_1(T) \le b, M^{W_1}(T) \le c\right\} = \Phi\left(\frac{b-\mu T}{\sqrt{T}}\right) - e^{2\mu c}\Phi\left(\frac{b-2c-\mu T}{\sqrt{T}}\right) \tag{103}$$

Finally we can plug in the values for b and c to obtain the probability as we set forth for.