Lecture Notes for Chapter 6

This is the chapter that brings together the mathematical tools (Brownian motion, Itô calculus) and the financial justifications (no-arbitrage pricing) to produce the derivative pricing methodology that dominated derivative trading in the last 40 years. We have seen that the no-arbitrage condition is the ultimate check on a pricing model and we have seen how this condition is verified on a binomial tree. Extending this to a model that is based on a stochastic process that can be made into a martingale under certain probability measure, a formal connection is made with so-called risk-neutral pricing, and the probability measure involved leads to a world that is called risk-neutral world. We will examine three aspects involved in this theory: no-arbitrage condition, martingale, and risk-neutral world, and show how these concepts lead to the pricing methodology as we know today.

1 Prototype model: a one-step binomial tree

First we take another look at the one-step binomial model and inspect the implications of no-arbitrage condition imposed on this model, and the emergence of the so-called risk-neutral probability measure.

To focus on the main issue, let us assume that the risk-free interest rate $r = 0$. We have this simplest model for a stock price over one period of time: it starts with a price S_0, and can move to one of two states after this time period: S_+ or S_- (assuming $S_+ > S_-$). In order that there is no-arbitrage opportunity for any investor, we must have $S_- < S_0 < S_+$. Imagine if $S_0 < S_- < S_+$, then an investor can buy the stock and he/she is guaranteed to gain a profit since in either scenario the stock price will be up. On the other hand, if $S_- < S_+ < S_0$, then an investor can short the stock and he/she will also be locked in to a net profit as the stock price will drop in either scenario. These are arbitrage opportunities and the only way to deny these opportunities is to make sure that S_0 is sandwiched between S_- and S_+.

What is the implication of this arbitrage-free (or no-arbitrage) condition $S_- < S_0 < S_+$? There exists a value $p : 0 < p < 1$ such that

$$S_0 = pS_+ + (1-p)S_- \tag{1}$$

This seemingly innocent representation is actually quite profound: it guarantees the existence of of probability measure $\tilde{\mathbb{P}}$ under which

$$S_0 = \tilde{\mathbb{E}}[S_1] \tag{2}$$

where the random variable S_1 denotes the price after the time period, and takes one of the values S_- and S_+. The probability measure equipped with this expectation is that the up move S_+ occurs with a probability p, and S_- occurs with a probability $1-p$. Remember that we did not say anything about actual probabilities of the up and down moves when we set up the model. These values p and $1-p$ are determined from the possible stock prices S_- and S_+. The expectation formula suggests that we can *imagine a world* where the expectation of the future stock price is actually the current stock price. In that world, no investor would ask for any compensation for the risk taken, and this is why we call this probability measure the risk-neutral measure.

Thus we find an equivalence between the arbitrage-free condition and the existence of the risk-neutral probability measure. One more consequence of this argument is that we also have a process for S where S is expected to have fluctuations averaged to be zero, or that the drift is zero, which is what behind the concept of martingale in stochastic processes.

How is this model related to our goal to price stock derivatives? We will price the derivative based on the arbitrage-free principle again: if we can form a portfolio that consists of a share of the derivative and some shares of the underlying stock, and make the portfolio riskless, then the portfolio value should be easily determined as it has no fluctuation, which leads to

$$P_0 = C_0 + \Delta \cdot S_0 = C_1 + \Delta \cdot S_1 = \tilde{\mathbb{E}}\left[C_1 + \Delta \cdot S_1\right] = \tilde{\mathbb{E}}\left[C_1\right] + \Delta \cdot \tilde{\mathbb{E}}\left[S_1\right]$$

The previous expectation for S_1 implies that

$$C_0 = \tilde{\mathbb{E}}[C_1] \tag{3}$$

and this is the pricing formula for the derivative and we see it is specifically derived for the risk-neutral measure.

2 Stochastic preliminaries

In order to introduce the more general model extended from the above ideas, we need some stochastic process preliminaries.

2.1 Probability space

A probability space refers to a triple $(\Omega, \mathcal{F}, \mathbb{P})$, which specifies the sample space Ω, a σ-algebra $\mathcal{F}$, and a probability measure $\mathbb{P}$:

1. Ω is called the sample space, and it contains all the possible outcomes;
2. $\mathcal{F}$ is a σ-algebra that contains all the *events*, namely all subsets of Ω. The collection of events (subsets) is more general than a collection of outcomes, as it includes all *possible combinations* of the outcomes;
3. $\mathbb{P}$ is probability measure that returns a value between 0 and 1 for each event in $\mathcal{F}$. In particular, it returns 0 for the empty set, and 1 for Ω, and $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ for nonintersecting events A and B.

The three objects are integral parts of a system, in particular the concept of $\mathcal{F}$ should be appreciated, as it specifies the kind of events for which $\mathbb{P}$ is supposed to be applied.
As an example, look at a system where n states are possible, each with a probability of occurring $0 < p_i < 1, i = 1, \ldots, n$, and $\sum_{i=1}^{n} p_i = 1$. $\mathcal{F}$ is the set that contains all possible combinations of states, such as the event that either state 1 or state 2 occurs, which has probability $p_1 + p_2$. This would be a legitimate probability space. With this probability space, we can answer questions like

$$\mathbb{P}\,(\text{one of the states } 1 \leq j \leq k \text{ occurs}) = p_1 + \ldots + p_k$$

$$\mathbb{E}[V] = p_1 v_1 + p_2 v_2 + \ldots + p_n v_n$$

2.2 Random variable

A random variable X can be introduced only after a specification of the probability space. The lack of rigorous definition can lead to confusion if the probability space is not carefully constructed. The connection between a random variable and the probability space is that we want to refer the event that X is in a Borel subset of $\mathbb{R}$ to an event in $\mathcal{F}$ as part of the probability space, so we can assign a probability to the event.
The most important information about a random variable X is its distribution function, either in cumulative form

$$F(x) = \mathbb{P}\{X \leq x\},$$

or in mass function (discrete case) $p(x) = \mathbb{P}\{X = x\}$, or density function (continuous case) $f(x) = F'(x)$, respectively.

2.3 Filtration

This concept is crucial for a proper description of a stochastic process, and it addresses particularly the issue of flow of information in time. As we know, a process is an ordered set of random variables, or a collection of random variables indexed by time. As time passes, we can imagine that the collection expands and we can describe events in more and more details. The relation to our discussions of trading strategies is that we would like to base our trading decisions on the information received so far, or partial information of the outcome. A mathematical description of the partial information is therefore critical to a rigorous setup for our models.
Here is an example involving coin tosses modeled as a discrete time process $\{\omega_1, \omega_2, \ldots\}$, where each ω_k is either a head (H) or a tail (T). Suppose we observe up to $T = 3$ and the sample space will contain all the possible outcomes, such as HHH (three consecutive heads), THT (first and third heads, second tail), and so on. Suppose we observe only the first toss, then the event

$$\{\omega_1 = H\} = \{HHH, HHT, HTH, HTT\} = A_H$$

refers to the a collection of outcomes which we would not be able to distinguish at time 1. In another word, at time 1 we cannot tell the difference between HHH and HHT. The details are only resolved as we move step by step into the future. We can similarly define another collection $A_T = \{\omega_1 = T\}$. At time 1, all we can observe are contained in a set

$$\mathcal{F}_1 = \{\emptyset, \Omega, A_H, A_T\}$$

which is a σ-algebra. Move along further in time, at time 2, we can observe outcomes in more detail, namely we can tell the difference between A_{HH} (the first two tosses being heads), and A_{HT} (the first head, second tail). Listing all the combinations involving unions, intersections, and complementary sets, we have another σ-algebra $\mathcal{F}_2$, which contains $\mathcal{F}_1$.
Extending this argument, we have a filtration $\mathcal{F}_n$, which satisfies

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \subset \cdots \subset \mathcal{F}$$

Roughly speaking, a filtration is a procedure to resolve the details of all the events according to time, and it involves an increasing collection of events indexed by the time variable . In continuous time, a filtration is denoted by $\{\mathcal{F}_t, t \geq 0\}$
Here is a simple example: the value X_t is going to be revealed at time t, we will know for sure if the event $\{X_t > \alpha\}$ (a Borel set) occurred or not by time t. In this case we say

$$\{X_t > \alpha\} \in \mathcal{F}_t,$$

but the same event is not in any of $\mathcal{F}_s$ for $s < t$.
We place so much emphasis on the information at time t because we want to construct trading strategies that can be easily followed. This means that we need a precise definition of what information that is available at time t and how we give orders based on this information.

2.4 Conditional expectation

Having introduced the concept of filtration, we are ready to discuss the concept of conditional expectation. The intuitive idea of conditional expectation is not difficult to acquire: given what happened up to certain time t, we want to find the best informed "prediction". That approach was naive: to properly introduce conditional expectation, we need a filtration. If we think of expectation as some sort of averaging, then conditional expectation refers to averaging over certain subset, a sub-σ-algebra. Of course this "average" depends on which particular sub-σ-algebra we specified, therefore it is a "function" of the sub-σ-algebra. In the language of probability theory, a conditional expectation is a random variable, as we don't know which event from the sub-σ-algebra is going to be observed.
There are two special situations that deserve to be mentioned: one is that the information contained in this subset does not provide any help in determining the value of X, we say that X is independent of this particular sub-σ-algebra, the

other is that the information in this subset determines completely the value of X (we say that X is measurable with respect to this sub-σ-algebra). In between these two situations we have the more interesting case where a conditional expectation gives some estimate of X, but not completely.
The standard notation for conditional expectation is

$$\mathbb{E}_n[X] = \mathbb{E}\left[X|\mathcal{F}_n\right] \tag{4}$$

for the discrete time case, and

$$\mathbb{E}_t[X] = \mathbb{E}\left[X|\mathcal{F}_t\right] \tag{5}$$

for the continuous time case.

2.5 Stopping time

A stopping time is a random variable and we often use τ to denote. The intuitive meaning is that a random variable is a stopping time if you can definitely say yes or no to the following question at time t: has this time τ arrived? In terms of the filtration, we have the definition for a stopping time: τ is a stopping time if

$$\{\omega : \tau \leq t\} \in \mathcal{F}_t$$

Examples are easy to construct: the first time the stock price reaches 100 is a stopping time, but the last time the stock price reaches 100 is *not* a stopping time, as we will never know if it is the last time that the stock hits that level.

2.6 Markov process

A Markov process is a special situation where

$$\mathbb{E}\left[X|\mathcal{F}_t\right] = \mathbb{E}\left[X|X_t = x\right] = f(x)$$

which implies that the conditional expectation does not require the information of the path from 0 to t. Instead, the information at a single point t is sufficient to determine the conditional expectation.

2.7 Martingale

A martingale is a special Markov process where

$$\mathbb{E}\left[X_t|\mathcal{F}_s\right] = \mathbb{E}\left[X_t|X_s = x\right] = X_s, \quad \text{for } t > s \tag{6}$$

The interpretation is that if we pause at any time s and try to predict a future value $X_t (t > s)$, the best prediction is the current value X_s. After some pondering, we realize that this process has no tendency (bias) to either go up or go down. If we picture a gambling game, at any time, no matter how much you have won or lost so far, the game always starts afresh and the chances to win or lose are **equal**.

To give an example, we consider

$$X_k = Z_1 + Z_2 + \cdots + Z_k,$$

where the i.i.d.'s

$$Z_j = \begin{cases} 1 \\ -1 \end{cases}$$

with equal probabilities 0.5 and 0.5. We can calculate

$$\mathbb{E}[X_k] = 0$$

but also

$$\mathbb{E}[X_k|\mathcal{F}_j] = X_j$$

as this expectation is taken assuming $\mathcal{F}_j$ is specified, implying $Z_1, \ldots, Z_j$ already observed, therefore no expectation taken with respect to these variables.

2.8 Equivalence of measures and change of measure

Two measures on the same sample space, with the same event spaces, are said to be equivalent if for all events E, probabilities under these two measures are either both zero, or both positive. As long as the probabilities are not zero, the magnitudes do not matter. Intuitively, when we move between two equivalent measures, we should agree on the same possible events, we are free to redistribute probabilities, as long as we agree on what's possible (positive probabilities), and what's impossible (zero probability).

For two equivalent measures P and Q, we can express the expectation of some random variable under P in terms of another expectation of the same random variable, modified by a random factor called the Radon-Nikodým derivative, under the other measure Q. This change of measure is fundamental in the Black-Scholes model that turns a Brownian motion with a drift to the standard Brownian motion (without drift).

3 Risk-Neutral Pricing

The objective for this chapter is to establish a pricing formula for any derivatives with a known payoff function. The key to success is whether we can find such a probability measure for the market model so the price is expressed as the conditional expectation under this so-called risk-neutral probability measure. The ultimate justification for a price is whether it will allow arbitrage opportunities on the market. We have several things at hand: no-arbitrage condition that puts a confidence tag on the price, the existence of a risk-neutral probability measure, and the martingale representation that guarantees a replicating strategy. One mathematical theorem that makes the connection possible is Girsanov's theorem, and unlike some theorems where the multi-dimensional case simply extends, Girsanov's theorem for multi-dimension reveals something not easily appreciated in

one-dimension. The intricacy and interplay of these objects deserve to be examined from several angles. First we want to move to a world where everyone is risk-neutral, so the expected return on every security is just the risk-free interest rate. The feature in this world is that the discounted portfolio value should be a martingale under this probability measure. Girsanov's theorem then tells us what's needed to be able to move to that world. These conditions become the conditions for the existence of the risk-neutral probability measure. The magic of this shift in Brownian motion is that any portfolio characterized through an adapted process $\Delta(t)$ that involves self-financing, after proper discount, is a martingale under the risk-neutral probability measure. Once martingale, we can express the value at an earlier time as the conditional expectation of the value at a later time. Finally, what connects the portfolio with the derivative is the martingale representation, which says any two martingales can be represented by each other. The discounted conditional expectation of the payoff function is one, and the discounted value process for a portfolio is another one, and we can relate them using the martingale representation theorem. The existence of the conditional expectation process guarantees the existence of the portfolio, which eventually matches the payoff function of the derivative. With this full circle of arguments, we establish the pricing methodology for any derivatives.
Finally, two fundamental theorems of asset pricing are in place:

1. Existence of a risk-neutral probability measure for the market implies no-arbitrage opportunities;

2. For a market model that has risk-neutral probability measure, the model is complete if and only if the risk-neutral probability measure is unique.

The first one is just what we discussed at the beginning of this paragraph. The second introduces the concept of complete market which in this incarnation just says that every derivative security can be hedged using the underlying and other derivatives available on the market. The description of complete market makes us more aware of the limitation of the Black-Scholes model, and also points to more realistic models that describe incomplete markets.

3.1 No-arbitrage and risk-neutral measure in multi steps

The extension of the one-step model in section 1 to multi steps is easy to implement, but not so trivial to justify the connection between no-arbitrage and the existence of a risk-neutral measure. Instead of just requiring

$$\mathbb{E}[P(t)] = P(0)$$

for a security with price P_t, we will need

$$\mathbb{E}_s[P(t)] = P(s)$$

for any $t > s$. This suggests the use of martingale in the model. The example on page 129 of the text illustrates this point quite well, where the model has

$$E[S(t)] = S(0), \quad \text{for } t = 1, 2$$

but

$$E_1[S(2)] \neq S(1).$$

The failure of the last condition leads to an arbitrage opportunity. The purpose of this example is to show that the martingale condition is essential to rule out arbitrage opportunities in the model.

3.2 Implying of risk-neutral probability measure from market

The risk-neutral idea sounds unnatural and we wonder if there is such a thing that can be connected with the market. If we assume the validity of the pricing formula for a derivative

$$V(S_0, 0) = \mathbb{E}[F(S_T)] = \int F(x)p(x)dx,$$

with a probability density function $p(x)$ for the random variable S_T, we will attempt to recover $p(x)$ if prices $V(S_0, 0)$ can all be observed. Suppose that we have a collection of call options with the same expiration, but different strike prices K. For convenience, we assume that the function $C_K(S_0, 0)$ is available from the market for all values of $K > 0$. Assuming $K_1 < K_2$, we consider the portfolio

- long a call with strike $K_1 - \epsilon$;
- short a call with strike K_1;
- short a call with strike K_2;
- long a call with strike $K_2 + \epsilon$.

The value of the portfolio, assuming $K_2 - K_1$ small, is

$$\begin{aligned} & C_{K_1-\epsilon} - C_{K_1} - C_{K_2} + C_{K_2+\epsilon} \\ \approx & -\frac{\partial C_K}{\partial K}_{|K_1} \cdot \epsilon + \frac{\partial C_K}{\partial K}_{|K_2} \cdot \epsilon \\ \approx & (K_2 - K_1)\frac{\partial^2 C_K}{\partial K^2} \cdot \epsilon \end{aligned}$$

On the other hand, the payoff of the portfolio will converge to

$$F(S) = \begin{cases} 1, & K_1 < S < K_2 \\ 0, & \text{otherwise} \end{cases}$$

so

$$
\begin{aligned}
&C_{K_1-\epsilon} - C_{K_1} - C_{K_2} + C_{K_2+\epsilon} \\
\approx&\epsilon \int_{K_1}^{K_2} p(x)dx \\
\approx&\epsilon(K_2 - K_1)p(\xi)
\end{aligned}
$$

We therefore conclude

$$
p(\xi) \approx \frac{\partial^2 C_K}{\partial K^2}
$$

for $K_1 < \xi < K_2$. Finally we let $K_1 \to K_2$ so we have

$$
p(S) = \frac{\partial^2 C_K}{\partial K^2}(S) \tag{7}
$$

This shows that the risk-neutral density for the stock price can be inferred from the option price, given that the option price is a smooth function of its strike K.

3.3 Accommodating interest rate

The interest effect (time value of money) in the case of deterministic rate can be summarized in the discount factor

$$
Z(t,T) = \text{price at } t \text{ of a zero-coupon bound maturing at } T.
$$

The option price with interest rate taken into account is

$$
C(0) = Z(0,T) \int p(S)F(S)\,dS \tag{8}
$$

or

$$
\frac{C(0)}{Z(0,T)} = \mathbb{E}\left[\frac{C(T)}{Z(T,T)}\right] \tag{9}
$$

since $Z(T,T) = 1$.
This suggests that we can work with the notion of discounted asset price $\tilde{X}(t) = Z^{-1}(t,T)X(t)$ in the following martingale pricing.

3.4 Discrete and continuous time martingale pricing

As we emphasized, the price for a derivative security that we are looking for is one such that it causes no arbitrage. In the binomial model, which is our prototype for discrete time models, the fact that there is no arbitrage implies that there will be some outcomes outperforming bank deposits (at risk free rates), and some outcomes underperforming. With discount factor properly factored in, we will have the current security price is **between** the good (outperform) and the bad (underperform) future prices. Therefore, it is possible to choose a probability measure such that

$$\frac{S_0}{B_0} = \tilde{\mathbb{E}}\left[\frac{S_1}{B_1}\right] \tag{10}$$

Notice that the expectation is taken with respect to the risk-neutral measure, whose existence is guaranteed by the no-arbitrage condition. In general we can say (from time n to time $n+1$)

$$\frac{S_n}{B_n} = \tilde{\mathbb{E}}_n\left[\frac{S_{n+1}}{B_{n+1}}\right] \tag{11}$$

Using the repeated conditioning property $E_n[X] = E_n\left[E_{n+1}[X]\right]$, we can generalize the above to

$$\frac{S_n}{B_n} = \tilde{\mathbb{E}}_n\left[\frac{S_m}{B_m}\right], \quad m > n \tag{12}$$

How does this help us to price derivatives such as an option? We need the no-arbitrage idea again: consider a portfolio with one share of the call, and Δ_n shares of the stock. In the one-step model, we know that by choosing

$$\Delta_n = \frac{C_{n+1}(H) - C_{n+1}(T)}{S_{n+1}(H) - S_{n+1}(T)} \tag{13}$$

we will have the option position hedged. Here H and T denote the up and move cases respectively. In doing so we can have the risk eliminated, meaning P_{n+1} will be the same value in both H and T outcomes. No-arbitrage argument requires

$$\frac{P_n}{B_n} = \tilde{\mathbb{E}}_n\left[\frac{P_{n+1}}{B_{n+1}}\right] \tag{14}$$

for any probability measure, as P_{n+1}/B_{n+1} is now the same in both outcomes. In particular, the above formula is valid for the risk-neutral measure,

$$\frac{C_n}{B_n} + \Delta_n \frac{S_n}{B_n} = \tilde{\mathbb{E}}_n\left[\frac{C_{n+1}}{B_{n+1}}\right] + \Delta_n \tilde{\mathbb{E}}_n\left[\frac{S_{n+1}}{B_{n+1}}\right]$$

Using Eq.(**??**), we therefore have

$$\frac{C_n}{B_n} = \tilde{\mathbb{E}}_n\left[\frac{C_{n+1}}{B_{n+1}}\right] \tag{15}$$

and

$$\frac{C_n}{B_n} = \tilde{\mathbb{E}}_n\left[\frac{C_m}{B_m}\right], \quad m > n \tag{16}$$

In the continuous time model, it will be more subtle. As we assumed in the discrete model that Δ_n is determined at time n and it remains the same until time $n+1$.

The situation is quite different in the continuous case. First we introduce our continuous time model for the stock price and bond price:

$$dS_t = \mu S_t \, dt + \sigma S_t \, dW_t \tag{17}$$
$$dB_t = r B_t \, dt \tag{18}$$

Using Itô's formula, we can compute for the discounted stock price

$$\begin{aligned} d\left[\frac{S_t}{B_t}\right] &= \frac{1}{B_t} dS_t + S_t\left(-\frac{1}{B_t}\right) dB_t \\ &= \mu\frac{S_t}{B_t} dt + \sigma\frac{S_t}{B_t} dW_t - r\frac{S_t}{B_t} dt \\ &= (\mu - r)\frac{S_t}{B_t} dt + \sigma\frac{S_t}{B_t} dW_t \end{aligned} \tag{19}$$

If $\mu = r$, then we will have a martingale for S_t/B_t since

$$d\left[\frac{S_t}{B_t}\right] = \sigma\frac{S_t}{B_t} dW_t \tag{20}$$

as $\sigma\frac{S_t}{B_t}$ is determined at time t (so-called adapted), and

$$E\left[d\left[\frac{S_t}{B_t}\right]\right] = 0, \tag{21}$$

so we can write literally

$$\frac{S_t}{B_t} = E_t\left[\frac{S_T}{B_T}\right]$$

This is the continuous time analogy of Eq.(**??**). However, in reality we do not necessarily have $\mu = r$ where μ is the expected growth rate of the stock in the real world. This crucial step would require the following section to explain.

3.5 Obtaining the risk-neutral measure: Girsanov's theorem

The previous section suggests that our Black-Scholes model shows that the discounted stock price would have no drift if the expected growth rate $\mu = r$, the risk free interest rate. If indeed we have $\mu = r$ in a world, what would be the implications? Here are some of the features:

- Investors seek no compensation for the risk taken, as long as the expected growth matches the risk free interest rate;
- Discounted stock price would be a martingale;
- Arbitrage opportunities would not exist.

They all sound very nice but we can see that they are impractical in our world. Can such probability exist in another world? Girsanov's theorem answers this question.
Theorem (Girsanov) Let W_t be a Brownian motion with sample space Ω and probability measure $\mathbb{P}$. Then for a reasonable ν, there exists an equivalent measure $\mathbb{Q}$ on the same Ω such that

$$\tilde{W}_t = W_t - \nu t$$

is a Brownian motion under $\mathbb{Q}$. The change of measure formula is also available explicitly but we do not need it here.
The idea behind the theorem is that we can manipulate the probabilities so that the process with a tendency to move upward (or downward) can be made to "correct" itself so the bias is removed. Intuitively, suppose the process has a negative drift, then we can downplay those downward path to an extend, and overplay those upward so that the the negative drift is corrected. The downplay/overplay effects are achieved by modifying the probability measure.
Going back to Eq.(**??**), we can introduce $\tilde{W}_t$ with $\nu = -(\mu - r)/\sigma$ such that

$$d\left[\frac{S_t}{B_t}\right] = \sigma \frac{S_t}{B_t}\, d\tilde{W}_t \tag{22}$$

Then the discounted stock price will be a martingale under the new probability measure determined by $\tilde{W}_t$. We denote this probability measure by a tilde or star and called it the risk-neutral probability measure. Now we can use it to price any derivative with the price

$$V_0 = B_0 \tilde{\mathbb{E}}\left[\frac{V_T}{B_T}\right] \tag{23}$$

In particular, when the payoff $V_T = F(S_T) = (S_T - K)^+$, we have the Black-Scholes formula reproduced:

$$\begin{aligned}
C_0 &= e^{-rT}\tilde{\mathbb{E}}\left[(S_T - K)^+\right] \\
&= e^{-rT}\tilde{\mathbb{E}}\left[\left(S_0 \exp\left((r - \frac{1}{2}\sigma^2)T + \sigma\sqrt{T}Z\right) - K\right)^+\right] \\
&= S_0 N(d_1) - Ke^{-rT}N(d_2)
\end{aligned}$$

3.6 Hedging and self-financing strategies

One important step we have not yet justified: in constructing a portfolio to be risk less, we assume that some Δ can be obtained to achieve the perfect hedge. This is easy to do in the binomial model, but not so obvious in the Black-Scholes model. The purpose is to justify the pricing formula

$$\frac{V_t}{B_t} = \tilde{\mathbb{E}}_t\left[\frac{V_T}{B_T}\right].$$

One way to do this is to show that the left-hand-side is a martingale under the risk-neutral measure. Using Itô's formula

$$\begin{aligned} d\left[\frac{V_t}{B_t}\right] &= \frac{1}{B_t} dV_t - \frac{V_t}{B_t^2}\, dB_t \\ &= \frac{1}{B_t}\left[\left(\frac{\partial V}{\partial t} + rS\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} - rV\right) dt + \sigma S \frac{\partial V}{\partial S}\, d\tilde{W}_t\right] \end{aligned}$$

To make sure there is no drift in this expansion, we must have $V(S,t)$ to satisfy the Black-Scholes PDE

$$\frac{\partial V}{\partial t} + rS\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} = rV \tag{24}$$

Still, why should we have the discounted derivative price to be a martingale under the risk-neutral measure?

The justification is based on the hedging of the derivative. Suppose the derivative in question is a call (with strike K, expiration T) having price $C_t = C(S_t, t)$. We start with the amount $C(S_0, 0)$ we received when we sold this call, and wish to do something with the stock to have cover this call position. Namely, we want to use this amount $C(S_0, 0)$ to replicate the payoff $(S_T - K)^+$ at time T. How should we do it? More specifically, how many shares of the stock should we purchase to begin with, and how should be adjust?

The answer comes from the martingale representation theorem:

Theorem (martingale representation) If W_t is a Brownian motion with filtration $\mathcal{F}_t$, suppose M_t is another martingale with the same filtration, then there exists an adapted process ϕ_t such that

$$dM_t = \phi_t\, dW_t$$

here ϕ_t is adapted, meaning that its value is revealed to us by time t. The secret of hedging ratio is in this ϕ_t, which is guaranteed by the representation theorem.

Now assume that $M_t = C_t/B_t$ is a martingale under the risk-neutral measure, according to the representation theorem, we must have a ϕ_t such that $dM_t = \phi_t\, d\tilde{W}_t$. Again, using Itô's formula

$$d\left(\frac{C_t}{B_t}\right) = \sigma \frac{S_t}{B_t}\frac{\partial C}{\partial S}\, d\tilde{W}_t = \frac{\partial C}{\partial S} d\left(\frac{S_t}{B_t}\right)$$

This connects the change in C_t to the change in S_t, and the factor tells us how many shares of the stock we should get so the price changes are cancelled out.

Now we can answer the question near the end of section 3.6: given $C(S_0, 0)$, how should we replicate the call? We first write

$$\begin{aligned} C(S_0, 0) &= \left(C(S_0, 0) - \frac{\partial C}{\partial S}\cdot S_0\right) + \frac{\partial C}{\partial S}\cdot S_0 \\ &= \alpha_0 S_0 + \beta_0 S_0 \end{aligned}$$

Let $\alpha_t = C(S_t, t) - \frac{\partial C}{\partial S} \cdot S_t$, and $\beta_t = \frac{\partial C}{\partial S}(S_t, t)$, so we consider the portfolio consisting of α_t units of the bond, β_t shares of the stock, with value

$$P_t = \alpha_t B_t + \beta_t S_t.$$

By the time we get to T, if $S_T > K$, $\frac{\partial C}{\partial S} = 1$, so $P_T = (S_T - K - S_T)B_T + S_T = S_T - K$; and if $S_T \leq K$, $\frac{\partial C}{\partial S} = 0$, so $P_T = C(S_T, T) = 0$. Combining these two, we have the correct payoff $(S_T - K)^+$ for the portfolio. Our conclusion is that we have successfully replicated the call by this portfolio.
How do we construct this portfolio? We should buy and sell according to the latest α and β, but we also need to check if the change caused by changes in α and β will cancel out. We can verify that

$$dC_t = \alpha_t \, dB_t + \beta \, dS_t$$

So this portfolio is a so-called self-financing portfolio.