Trading Volatility

Luke Marolda

June 14, 2025

1 Introduction

Welcome to Trading Volatility - all of my options knowledge, shared with you!

In my brief time exploring the world of options I have learned quite a lot. However, I've noticed that the lens in which options can be studied through is wide, ranging from deeply theoretical to surprisingly applicable. All of the courses I've taken, books or papers I've read, and videos I've watched, typically fall on either end of the spectrum. You either have a medium which emphasizes the use of options without the complex mathematical background, or focuses on the theory, ignoring the practical knowledge needed to trade these products. Why not have both at the same time?

As I write this book I plan to do my best at synthesizing these two lenses into one, pulling from both my academic and industry experiences in order to summarize and share my knowledge of options.

Enjoy!

Author's Note: My favorite part of this book is the visualizations. I had written a bunch of mathematical finance related code (check out my GitHub) at the end of my Junior year, and now I finally get to put it to good use.

2 Arbitrage Free Pricing

3 Basic Options Theory

Now that we have a basic understanding of arbitrage free pricing and the other foundational concepts of mathematical finance, we can begin to dive into the world of options. In this chapter, we will give a gentle introduction to options theory in a way that will prepare the reader for subsequent chapters.

3.1 Motivation

While options were first traded in London in 1690, their usage has exploded since their standardization in 1973 with the opening of CBOE, the Chicago Board Options Exchange. During the 1990s and onwards, the popularization of hedge funds and the birth of volatility as an asset class itself caused a rapid proliferation of derivative securities, ranging in complexity from simple calls and puts to more advanced contracts such as volatility swaps and variance swaps. [Ben14]

What is interesting about options is that they not only have directional exposure, but also volatility exposure. When you buy a stock, whether it rises 10 points in 1 day or 1 month has no impact on your position's payoff - regardless of the amount of time it took, you still only directly profit from the 10 point difference in what you bought the stock at and what it is at now. Options, however, are path-dependent securities. You have directional exposure, but also volatility exposure (and a number of other things). So in the same context of the 10 point stock move, whether that move is accomplished in 1 day or 1 month *does* actually matter for the value of an options position with the stock as the underlying.

With that being said, there are many benefits (and risks) to trading options. A core use case is for leverage - if you have a directional view on a stock you can earn a leveraged return by choosing to trade options versus taking a position in the stock itself. But this of course introduces other risks, as options prices move based on a variety of additional exposures. Another popular use of options is call overwriting: for a directional investor who owns a stock, this is accomplished by selling an out of the money call option to enhance yield. This has been a successful strategy over time as volatility is typically overpriced. But once again, engaging in a call overwriting strategy involves more decision making than simply holding a directional position in only the stock. At its core, both an equity and volatility view are required to trade options.

We will cover more strategies and common use cases of options later in this book. For now, we want to establish a base knowledge of options theory that we can build upon. A core advantage of options that I want you to remember is the following: options offer convexity - if markets move against the investor the only loss is the premium paid, versus a position in a stock itself or a forward, which has (virtually) unlimited loss. Think about this as we now introduce the definitions of calls and puts.

3.2 The Core

3.2.1 A Note on Exercise

Before we formally introduce the main types of options contracts, we must first understand exercise types. You will see that an option gives the holder the right to buy (or sell) the underlying security on a set of exercise dates. This set can range from a singular day at expiry, or any single time up until expiration. We split expiration types into three categories:

Definition 3.1. *Consider an option with expiration date T.*

1. *If the option can be exercised at any time $t \leq T$ then we say it is American*
2. *If the option can be exercised at a select number of times $t_1, ..., t_n$ then we say it is Bermudan*
3. *If the option can only be exercised at $t = T$ then we say it is European*

For the rest of this chapter, we will focus our discussion on European options. A complete theory of American/Bermudan options involves the notion of a stopping time, which we will cover later in a later chapter.

3.2.2 Calls

Definition 3.2. *A call option is the right, but not obligation, to buy or take a long position in a given asset at a fixed strike price K on or before a specified expiration date T.*

As previously mentioned, we are looking at the case where this position can only be taken at expiration, not before. With that being said, we can think of a call option as a security g whose payoff at expiry T is

$$g(x) = (x - K)^+$$

where x is the underlying security. For those unfamiliar with the positive part operator:

$$x^+ = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases}$$

This payoff function should make sense to the reader. Because a call gives you the right but not obligation to buy the underlying, the payoff you'll receive is simply the difference between the two if it's positive, and zero otherwise. This is because if the underlying is less than the strike at expiry, then you'd simply choose to not exercise it. We can see this payoff graphically in Figure 1. This figure shows the payoff at expiry for a call with $K = 100$. This image directly matches our formal definition, as the payoff is 0 when the underlying price is less than strike, and begins to grow linearly once the underlying surpasses the strike price. This is a classic diagram and is often referred to as the "hockey stick", due to the shape of a call's payoff.

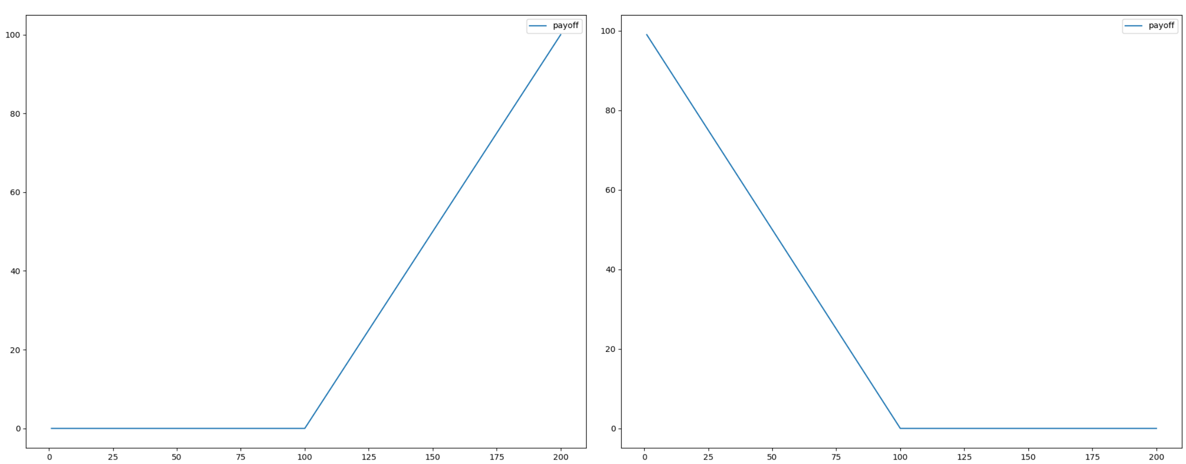
3.2.3 Puts

Definition 3.3. *A put option is the right, but not obligation, to sell or take a short position in a given asset at a fixed strike price K on or before a specified expiration date T.*

Similarly to calls, for the European case we can think of a put option as a security g whose payoff at expiry T is

$$g(x) = (K - x)^+$$

where x is the underlying security.



(a) Call (b) Put

Figure 1: Payoffs across price of the underlying.

Because a put gives you the right but not obligation to sell the underlying, the payoff you'll receive is simply the difference between the two if the underlying is less than strike, and zero otherwise. Once again, if the underlying is more than the strike at expiry, then you'd simply choose to not exercise it.

We see that calls and puts offer directional exposure to the underlying security. A call option profits when markets rise, giving the investor "long" exposure, and a put option profits when markets fall, giving "short" exposure. Therefore, options allow investors to employ long or short strategies. But as I mentioned earlier, the loss on an option position is limited to the premium paid for it. We visually see that the payoffs of calls and puts in the worst case are 0. Of course this means you could lose all of your upfront investment, but this characteristic is a key differentiation between purely directional securities such as stocks and forwards, which can have a negative payoff at expiration.

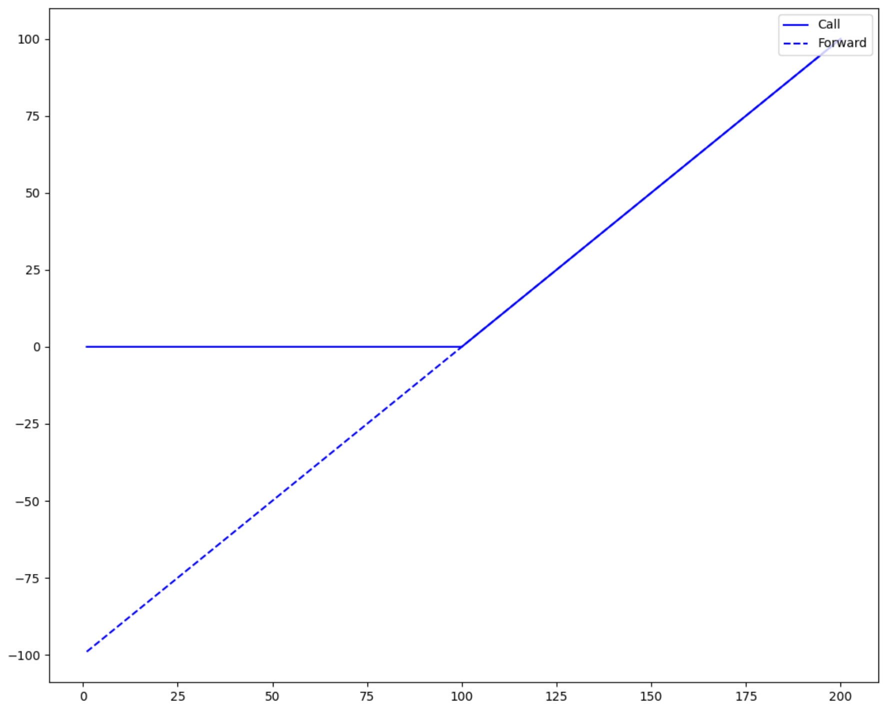


Figure 2: Payoff of a call versus a forward contract.

3.3 Pricing

Now that we've introduced calls and puts, how can we know how much they're worth? Options pricing has many layers of complexity, but for this introductory chapter we will mention some key ideas to think about.

3.3.1 Intrinsic Value and Time Value

When we introduced calls and puts in the above sections, we defined them by their payoff functions *at* maturity. Figure 1 displays what the options are worth at the time of expiration - so then how do we determine what the price of an option should look like prior to expiration? This is what is cool about options; we have this time dependence that vastly impacts the price of our security.
Consider a case where we have a call option on an underlying stock that is currently trading at \$100 and has 0 volatility. That is, the stock does not move at all, ever. Well then clearly the price of the call at any time prior to expiry, should be equal to the payoff at expiry, because there is no chance that the underlying will move in any direction. We call the payoff at expiration the *intrinsic value* of the option, and in this case the price of the call should be exactly equal to its intrinsic value. Note that the function g we defined for the definitions of puts and calls describe exactly the intrinsic values of the two securities.

Say that the strike of our option is also at \$100, the current price of the stock. By the way, if our option is struck at the current price of the underlying, we say that the option is "At-The-Money", or ATM for short. Well then the price of our call should be \$0 at any time before maturity, because our intrinsic value is

$$(100 - 100)^+ = 0$$

But now consider a call that has non-zero volatility. For simplicity, we'll say the stock has a 50% chance of moving up \$10 and a 50% chance of moving down \$10 before expiration. Based on our definition of a call, we know that this means the call has a 50% chance of profiting \$10, and a 50% chance of expiring worthless. Therefore, the fair value of the call prior to expiry should be \$5 (a simple expected value calculation).

Interesting, so as we increased the volatility of the underlying, not only did our call option increase in value, but it also became worth more than its intrinsic value at times prior to expiration. This shows us two things

1. As volatility increases, so does the price of options
2. The price of an option consists of more than its intrinsic value

We call this secondary component in price the *Time Value* of the option. Of course we saw that in the presence of no volatility, we had no additional time value. But as we introduced volatility, the time value of the option came into play. With this very simple example, we've illustrated that options have a clear exposure to volatility, and that an option's price is

$$\text{Price} = \text{Intrinsic Value} + \text{Time Value}$$

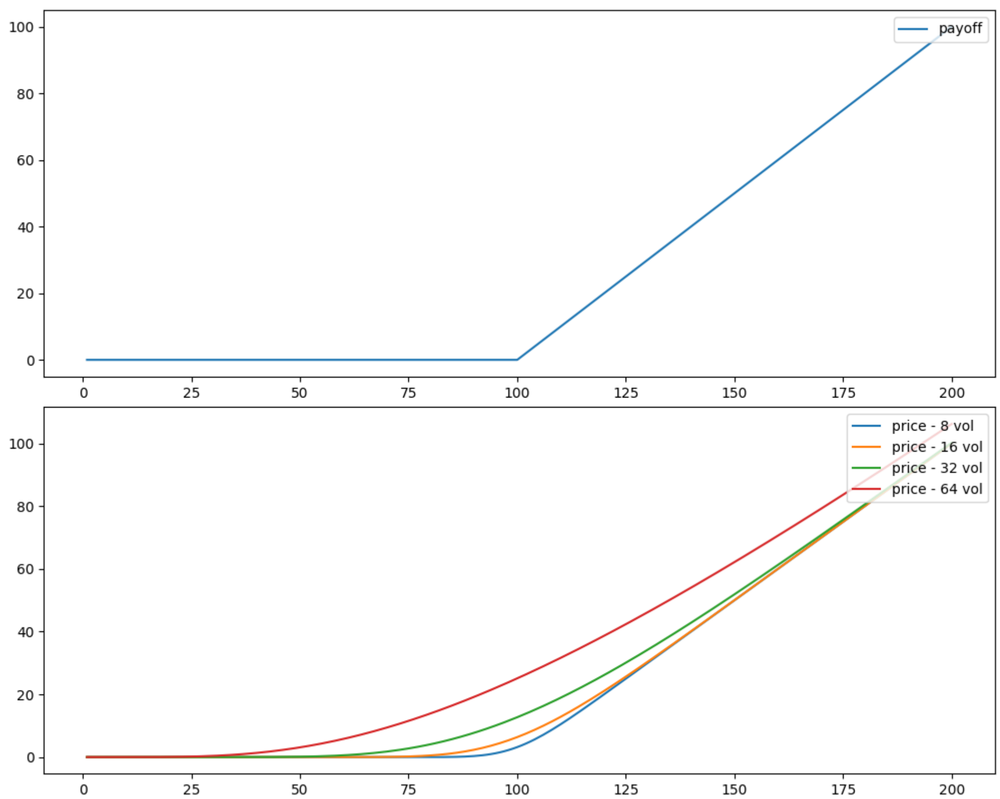


Figure 3: Price of a call across different volatilities.

3.3.2 Pricing Models

Now we know that an option's price should not only reflect its intrinsic payoff function, but should also incorporate the likelihood that it expires ITM: "in-the-money" (ie. profitable) or OTM: "out-the-money" (worthless). This should make sense - because we are dealing with expiration dates in the future, we have to be able to reason about the way the underlying asset is likely to move throughout the duration of the option's lifetime. Once we can reasonably model the underlying, we can then

utilize expectation and arbitrage-free pricing to tie everything together and settle on fair prices for these securities.
This process will be the subject of the next few chapters for this book. We need to decide how to accurately (and fairly) model the underlying (which will be equities for the majority of this book) and then can use these models to price options. Stocks, however, are complicated. Modeling randomness requires some high level mathematics, which we will develop over the next few chapters. For now, we briefly introduce some key factors that we'll expect to influence the price of our options.

Each of the following factors will play a role in the theoretical pricing models

1. K - Strike
2. T - Time to expiration
3. S - Price of the underlying
4. r - Interest rate
5. q - Dividend yield
6. σ - Volatility of the underlying

The first two should be trivial, as they are fixed within the terms of an options contract. The next three should also make sense, and are (mostly) known constants at the initiation of the contract. The last variable, the volatility, is where things get the most interesting. We saw above through a simple example that the price of options should increase with increasing volatility, but the way we define, observe, and model volatility will be the subject of its own chapter.

4 Probability Prep

Whether you like it or not, probability theory will be the foundation of many of our discussions about options. Remember, we want to work towards modeling a stock - which inherently involves randomness. Probability is how we make sense of randomness, so there are some essentials we must brush up on before we continue. Note that by no means is this meant to be a comprehensive coverage, rather we will focus on the essential concepts needed for the rest of of this book. For a full breakdown of probability theory, check out Dolev Artzi's "Probability and its Applications". Select definitions and proofs were based on the teachings of [Iye22].

4.1 The Basics

4.1.1 Random Variables

In essence, a random variable is a mathematical formalization of an object that depends on random events. The simplest example is the outcome of a coin flip - we can label this as a random variable. Although it sounds simple, there is actually a lot to the true definition of a random variable. But first, we must discuss sample spaces.

Continuing with the example of a coin flip, we can agree that there are two outcomes: heads or tails. Therefore, the *sample space* of outcomes consists of heads and tails, which we could possibly denote as $\Omega = (\omega_H, \omega_T)$. Of course this extends to all random variables - another simple example is the outcome of a dice roll, our sample space would then be the numbers we could roll: $\Omega = (\omega_1, \omega_2, \dots, \omega_6)$. Now I take that the reader knows this already, but I'm including it for clarity. With this in mind, we can strengthen our definition of a random variable by now calling it a *function.*

Definition 4.1. *A random variable X is a function $X : \Omega \to \mathbb{R}$ that maps the sample space outcomes to real values.*

To formalize our coin flip example, we'd define a random variable X as

$$X(\omega) = \begin{cases} 1 & \omega = \omega_H \\ 0 & \omega = \omega_T \end{cases}$$

where for the fair case, our respective probabilities would be $\mathbb{P}[\omega = \omega_H] = \mathbb{P}[\omega = \omega_T] = \frac{1}{2}$.

Even more formally, a random variable is a *measurable* function from a probability measure space to a measurable space. Now I am no measure theory specialist, so I won't bother to go into detail here (maybe another time though). But basically, when we talk about continuous time stochastic processes such as Brownian motion, our sample space gets more complex because it becomes all of its possible trajectories - which is the space of all continuous functions $\Omega = C([0, \infty))$. But this is uncountably infinite; we have no hope of defining any sort of probabilities $\mathbb{P}[A]$ for all $A \subset \Omega$. In this case we then restrict our attention to some σ-algebra $\mathcal{G}$, which is a subset of of some sets $A \subset \Omega$ on which we can actually define a probability measure. So then our most formal definition of a random variable incorporates measurability in regards to this σ-algebra $\mathcal{G}$. However, I'll leave it at that for now just to get you thinking.

The most important way to characterize a random variable is through the probabilities of the values it can take. For discrete random variables, we accomplish this through a *probability mass function (PMF)*, and for continuous random variables we use a *probability density function (PDF)*.

Definition 4.2. *A discrete random variable X must have a probability mass function p_X, where if x is any possible value of X, then $p_X(x)$ describes the probability of the event $\{X = x\}$:*

$$p_X(x) = \mathbb{P}[X = x]$$

So for our mentioned our coin flip example, we'd describe our random variable X with the following probability mass function:

$$p_X(x) = \begin{cases} \frac{1}{2} & x = 1 \\ \frac{1}{2} & x = 0 \end{cases}$$

Definition 4.3. *A continuous random variable X must have a probability density function (PDF) f_X such that*

$$\mathbb{P}[X \in B] = \int_B f_X(x)dx$$

for every subset B of the real line. In particular, the probability that X falls within an interval is

$$\mathbb{P}[a \leq X \leq b] = \int_a^b f_X(x)dx$$

We'll see an example of a probability density function when we discuss notable continuous distributions in a later section.

4.1.2 Expectation

Informally, the expected value of a random variable is the arithmetic mean of the possible values it can take, weighted by the probabilities of those outcomes. Formally, we don't necessarily have a simple formula. But, we will list a few common cases with a respective definition for expectation.

For a discrete random variable X with a finite list $x_1, \ldots, x_k$ of possible outcomes, each of which having respective probabilities $p_1, \ldots, p_k$, then expectation is defined as

$$\mathbf{E}[X] = \sum_{i=1}^{k} x_i p_i$$

For a random variable X with a probability density function f_X on the real number line, then the expectation of X is given by

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} x f_X(x)dx$$

Lastly, for an arbitrary real-valued random variable X defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, expectation is defined as the Lebesgue integral

$$\mathbf{E}[X] = \int_\Omega X(\omega) d\mathbb{P}(\omega)$$

With some definitions stated, now we can discuss some useful properties of expectation

1. (Linearity) Let X, Y be random variables and have $\alpha, \beta \in \mathbb{R}$. Then

$$\mathbf{E}[\alpha X + \beta Y] = \alpha \mathbf{E}[X] + \beta \mathbf{E}[Y]$$

2. (Positivity) Let X be a random variable. If $X \geq 0$ then $\mathbf{E}[X] \geq 0$. Further, if $X \geq 0$ and $\mathbf{E}[X] = 0$, then $X = 0$ almost surely.

3. (Layer Cake) Let X be a random variable. If $X \geq 0$, then

$$\mathbf{E}[X] = \int_0^\infty \mathbb{P}[X \geq t]dt$$

   More generally, if ϕ is an increasing function with $\phi(0) = 0$ then

$$\mathbf{E}[\phi(X)] = \int_0^\infty \phi'(t)\mathbb{P}[X \geq t]dt$$

4. (Unconscious Statistician) Let X be a random variable with PDF p_X. Then for any continuous function f we have

$$\mathbf{E}[f(X)] = \int_{-\infty}^{\infty} f(x)p_X(x)dx$$

4.1.3 Variance

Variance, being a measure of dispersion, is the expected value of the squared deviation from the mean of a random variable

$$\mathbf{Var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2]$$

Another essential measure of dispersion is a random variable's *standard deviation*, which is defined as the square root of the variance

$$\sigma_X = \sqrt{\mathbf{Var}[X]}$$

It is not a coincidence that both standard deviation and volatility are denoted by σ. In fact, the volatility of an underlying asset is essentially the standard deviation of the asset's prices over a defined period of time (we'll explain this in much more depth in a later chapter dedicated to volatility).

In practice, standard deviation is easier to interpret because it has the same units of the random variable itself - however, the variance is more commonly used as it grants us nicer properties. For example, the variance of a sum of independent random variables is equal to the sum of their variances.

Below are some properties of variance that we will use frequently

1. Let X be a random variable and have $\alpha, \beta \in \mathbb{R}$. Then

$$\mathbf{Var}[\alpha X + \beta] = \alpha^2 \mathbf{Var}[X]$$

2. (Variance in Terms of Moments) Let X be a random variable. Then we have

$$\mathbf{Var}[X] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$$

3. Let X, Y be two random variables. If X and Y are independent, then

$$\mathbf{Var}[X + Y] = \mathbf{Var}[X] + \mathbf{Var}[Y]$$

4.1.4 Conditional Expectation

Before discussing conditional expectation, we must first understand the concept of a filtration. In the context of stochastic processes (sequences of random variables), a filtration can be interpreted as an indexed family of objects, each representing all historical but not future information about the stochastic process at a given point in time. Each of these sub-objects gain in complexity as we move forward in time, because more information becomes available.

This must sound quite abstract, so I'll explain with an example. Consider a sequence of n coin tosses. Then our filtration $\mathcal{F}$ is a sequence of objects $(\mathcal{F}_i)_{i \leq n}$ where each $\mathcal{F}_i$ represents all events described using the first i coin tosses. Of course, as more coin tosses are observed, more information can be deduced, so we have an ordering $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_n$. Filtrations are often referred to as *information sets* due to this intuitive explanation. In simplest terms, $\mathcal{F}_n$ represents all information available up to time n.

In continuous time, the coin toss example loses its meaning, so we instead think of filtrations in terms of Brownian motion. A filtration $\mathcal{F}$ is still a sequence of objects $(\mathcal{F}_i)_{i \leq t}$, but now each $\mathcal{F}_i$ represents all events described using the trajectory of a Brownian motion up to point i. We still have the same ordering property and intuitive understanding, it is just now backed by a continuous stochastic process. If we say a random variable is "$\mathcal{F}_t$-measurable", this means that it is observed from the information available up to time t.

With this in mind, we can define conditional expectation, as the expected value conditioned on all information up to a point in time. Unlike traditional expectation, there is no simple formula here. However, we can think of $\mathbf{E}_t[X] = \mathbf{E}[X|\mathcal{F}_t]$ as the expected value of X given $\mathcal{F}_t$.

Definition 4.4. *The conditional expectation $\mathbf{E}_t[X]$ is the unique random variable such that*

1. *$\mathbf{E}_t[X]$ is $\mathcal{F}_t$-measurable*
2. *For every $A \in \mathcal{F}_t$, $\int_A \mathbf{E}_t[X] d\mathbb{P} = \int_A X d\mathbb{P}$*

In all honesty, this definition is not very applicable. Therefore, we will now provide some more useful properties that can be used when evaluating conditional expectation.

1. (Linearity) Let X, Y be random variables and have $\alpha, \beta \in \mathbb{R}$. Then
$$\mathbf{E}_t[\alpha X + \beta Y] = \alpha \mathbf{E}_t[X] + \beta \mathbf{E}_t[Y]$$
2. (Positivity) Let X be a random variable. If $X \geq 0$ then $\mathbf{E}_t[X] \geq 0$. Further, if $X \geq 0$ and $\mathbf{E}_t[X] = 0$, then $X = 0$ almost surely.
3. (Tower Property) Let X be a random variable. If $0 \leq s \leq t$, then
$$\mathbf{E}_s[\mathbf{E}_t[X]] = \mathbf{E}_s[X]$$
4. Let X, Y be random variables. If X is $\mathcal{F}_t$-measurable then
$$\mathbf{E}_t[XY] = X\mathbf{E}_t[Y]$$
A special case of this is choosing $Y = 1$. In this case, $\mathbf{E}_t[X] = X$
5. Let Y be a random variable. If Y is independent of $\mathcal{F}_t$, then
$$\mathbf{E}_t[Y] = \mathbf{E}[Y]$$

Combining the last two properties gives us a special lemma that we will detail below, called the independence lemma.

Theorem 4.1. *(Independence Lemma) Let X, Y be random variables. If X is $\mathcal{F}_t$-measurable and Y is independent of $\mathcal{F}_t$, then for any function $f(x, y) : \mathbb{R}^2 \to \mathbb{R}$ we have*
$$\mathbf{E}_t[f(X, Y)] = g(X) \ \ \textit{where} \ \ g(x) = \mathbf{E}[f(x, Y)]$$

4.1.5 Moment Generating Functions

Next, we'll discuss moment generating functions, which provide an alternative specification for the random variable's probability distribution. It provides an additional route to analytical results compared to working directly with a random variable's probability density function or cumulative density function.

Definition 4.5. *The moment generating function (MGF) of X is defined by*
$$M_X(\lambda) = \mathbf{E}[e^{\lambda X}]$$

Theorem 4.2. *A sequence of random variables $(X_n) \to X$ (in distribution) if and only if $M_{X_n} \to M_X$ pointwise.*

The proof of the above is out of the scope of this book, but a rigorous proof can be found in Billingsley's "Probability and Measure".

4.1.6 Characteristic Functions

Similar to moment generating functions, we have another defining function for random variables, the characteristic function. There is a significant advantage to characteristic functions, as every random variable must have one. The moment generating function, however, has limitations - they do not always exist for all λ. For some distributions, it only exists for a range of λ values around zero. Further, for heavy-tailed distributions, such as the Cauchy distribution, it may not exist at all. This makes characteristic functions a great addition to our probability toolkit.

Definition 4.6. *The characteristic function (CF) of X is defined by*

$$\phi_X(\lambda) = \mathbf{E}[e^{i\lambda X}]$$

Now we will list some important properties of characteristic functions. These will be important for some of the proofs to follow, but are also just good to know in general.

Theorem 4.3. *Let X, Y be random variables, where $Y = \alpha X$ with $x \in \mathbb{R}$. Then $\phi_Y(\lambda) = \phi_X(\alpha\lambda)$.*

Theorem 4.4. *Let X, Y be two independent random variables. Then $\phi_{X+Y}(\lambda) = \phi_X(\lambda)\phi_Y(\lambda)$*

Theorem 4.5. *Let X, Y be two random variables. The joint characteristic function of X, Y is defined by $\phi_{X,Y}(\lambda, \mu) = \mathbf{E}[e^{i(\lambda X + \mu Y)}]$. If X, Y are independent, then $\phi_{X,Y}(\lambda, \mu) = \phi_X(\lambda)\phi_Y(\mu)$. Conversely, if $\phi_{X,Y}(\lambda, \mu) = \phi_X(\lambda)\phi_Y(\mu)$ for every $\mu, \lambda \in \mathbb{R}$, then X and Y must be independent.*

I will leave the proofs for the above properties as an exercise for the reader. They can be accomplished using the definitions of independence and the properties of expectation and exponentiation.

Theorem 4.6. *Let X be a random variable. Then $\mathbf{E}[X^n] = (-i)^n \phi_X^{(n)}(0) = M_X^{(n)}(0)$.*

This can simply be checked by the reader. I find this property particularly useful, as it can be helpful for quickly computing higher moments of random variables.

Theorem 4.7. *A sequence of random variables $(X_n) \to X$ (in distribution) if and only if $\phi_{X_n} \to \phi_X$ pointwise.*

Once again, the proof for the above is out of the scope of this book. We will use it without proof, but an interested reader can find what they're looking for by checking out Levy's Continuity Theorem.

4.2 Notable Distributions

4.2.1 Normal Distribution

Definition 4.7. *We say a random variable X is normally distributed (or Gaussian) with mean μ and variance σ^2 if the PDF of X is*

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

We denote this as $X \sim N(\mu, \sigma^2)$

Below, we list the moment generating function for a standard normal random variable. These can be computed by definition. If $X \sim N(0, 1)$ then $\phi_X(\lambda) = e^{-\lambda^2/2}$ and $M_X(\lambda) = e^{\lambda^2/2}$.

4.3 Stochastic Processes

Now that we have a basic understanding of random variables, we can turn our attention to analysis of *sequences* of random variables, ie. stochastic processes.

Definition 4.8. *Let $X_1, X_2, \ldots, X_n$ be a sequence of random variables indexed by time. We call this sequence a stochastic process.*

Typically, we will be working with stochastic processes that involve a sequence of "i.i.d." variables, which means that each of the X_i are independent and identically distributed.

Definition 4.9. *We say that a stochastic process $X_1, X_2, \ldots, X_n$ is adapted if and only if X_n is $\mathcal{F}_n$-measurable $\forall n \geq 0$*

4.3.1 Central Limit Theorem

Consider the following stochastic process: Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. random variables with the following

1. $\mathbf{E}[X_n] = 0$
2. $\mathbf{Var}[X_n] = 1$

Now define $S_n = \sum_{k=1}^{n} X_k$ with $S_0 = 0$. This is simply the sum of the sequence up to time n. We are interested in the behavior of S_n as $n \to \infty$. To answer this question, we will briefly look at the law of large numbers and central limit theorem in the context of this sequence.

Theorem 4.8. *(Law of Large Numbers)* $\lim_{n\to\infty} \frac{1}{n} S_n = 0$

Now this is not exactly the formal definition of the true weak or strong law of large numbers, but we can still prove this claim heuristically to help our understanding.

Proof. This is a heuristic proof - it is not intended to be formal but is to demonstrate the theorem stated above. To proceed, we will show that both the expectation and the variance of S_n approach 0 as n approaches infinity. With this, we can then claim convergence in probability to 0. First, we look at expectation:

$$\mathbf{E}[\frac{1}{n} S_n] = \frac{1}{n}\mathbf{E}[\sum_{k=1}^{n} X_k] = \frac{1}{n}\sum_{k=1}^{n}\mathbf{E}[X_k] = 0$$

And of course the limit of this is also 0. Now the order in which I did this might make some measure theorists cringe, as I took expectation and then the limit. This process of passing expectation through the limit is a complicated one, and in general is not always possible. The true rigorous results involving when we can or can't do this, are stated in Lebesgue's Convergence Theorems. But for now, lets hold off on the measure theory.

Similarly, we can compute the variance:

$$\mathbf{Var}[\frac{1}{n} S_n] = \frac{1}{n^2}\mathbf{Var}[S_n] = \frac{1}{n^2}\mathbf{Var}[\sum_{k=1}^{n} X_k] = \frac{1}{n^2}\sum_{k=1}^{n}\mathbf{Var}[X_k] = \frac{1}{n^2} n = \frac{1}{n}$$

We were able to pass the variance through the summation because each of the X_k are independent. Now, taking the limit of this gives us our desired conclusion

$$\lim_{n\to\infty} \frac{1}{n} = 0$$

□

Now what happens if we consider $\frac{1}{\sqrt{n}} S_n$ instead?

Theorem 4.9. *(Central Limit Theorem) As $n \to \infty$, then $\frac{1}{\sqrt{n}} S_n$ converges to a standard normal distribution $N(0,1)$. That is, for any bounded continuous function f,*

$$\mathbf{E}[f(\frac{S_n}{\sqrt{n}})] = \mathbf{E}[f(N(0,1)]$$

Proof. We will proceed using Theorem 4.6. That is, if we can show that the characteristic function of $\frac{1}{\sqrt{n}} S_n$ converges to the that of a standard normal, then we are done. Formally, we need to show:

$$\lim_{n\to\infty} \phi_{\frac{S_n}{\sqrt{n}}}(\lambda) = e^{-\lambda^2/2}, \forall \lambda \in \mathbb{R}$$

First, we will use some of the properties we introduced for characteristic functions

$$\phi_{\frac{S_n}{\sqrt{n}}}(\lambda) = \phi_{S_n}(\frac{\lambda}{\sqrt{n}})$$

then using the definition of $S_n = \sum_{k=1}^n X_k$ where each of the X_k are independent

$$\phi_{S_n}(\frac{\lambda}{\sqrt{n}}) = \phi_{X_1}(\frac{\lambda}{\sqrt{n}})\phi_{X_2}(\frac{\lambda}{\sqrt{n}})\cdots\phi_{X_n}(\frac{\lambda}{\sqrt{n}})$$

by definition we also know that each of the X_k have the same distribution, so $\phi_{X_k}(\lambda) = \phi_X(\lambda), \forall\lambda \in \mathbb{R}$. Combining this gives us

$$\phi_{\frac{S_n}{\sqrt{n}}}(\lambda) = \phi_X(\frac{\lambda}{\sqrt{n}})^n$$

To proceed, we can use Taylor's Theorem to approximate $\phi_X(\frac{\lambda}{\sqrt{n}})$ centered at 0

$$\phi_{\frac{S_n}{\sqrt{n}}}(\lambda) = (\phi_X(0) + \frac{\lambda}{\sqrt{n}}\phi'_X(0) + \frac{\lambda^2}{2n}\phi''_X(0) + O(\frac{1}{n^{3/2}}))^n$$

using the definition of the characteristic function, we know $\phi_X(0) = 1$, $\phi'_X(0) = 0$, and $\phi''_X(0) = -1$ therefore

$$\phi_{\frac{S_n}{\sqrt{n}}}(\lambda) = (1 + 0 - \frac{\lambda^2}{2n} + O(\frac{1}{n^{3/2}}))^n$$

Now we can take the limit. We know that the error term $O(\frac{1}{n^{3/2}})$ will approach 0 as $n \to \infty$, so let's ignore that:

$$\lim_{n\to\infty} \phi_{\frac{S_n}{\sqrt{n}}}(\lambda) = \lim_{n\to\infty}(1 - \frac{\lambda^2}{2n})^n$$

The reader should now recognize this expression, it is the limit definition of e! This brings us to the desired result:

$$\lim_{n\to\infty} \phi_{\frac{S_n}{\sqrt{n}}}(\lambda) = e^{-\lambda^2/2}$$

To show the more general version for any bounded continuous function f, one can utilize the same arguments as above.

□

4.3.2 Martingales

There is a special type of stochastic process that will come up very frequently in this book - a martingale. It will be useful to think of a martingale as a fair game, ie. stopping the game based on the information available today will not change your expected return. In other words, the expected value of the process at a future time t, conditioned on all of the information available up to prior time s, is simply equal to the process at time s.

Definition 4.10. *An adapted process M is a martingale if for every $0 \le s \le t$, we have $\mathbf{E}_s[M_t] = M_s$.*

Start to think about how this could be relevant for "fair pricing" of securities. If we can have an asset that is modeled as a martingale (accounting for the time value of money), then this will give us a baseline for being able to fairly price other securities with this asset as the underlying (ex. options).

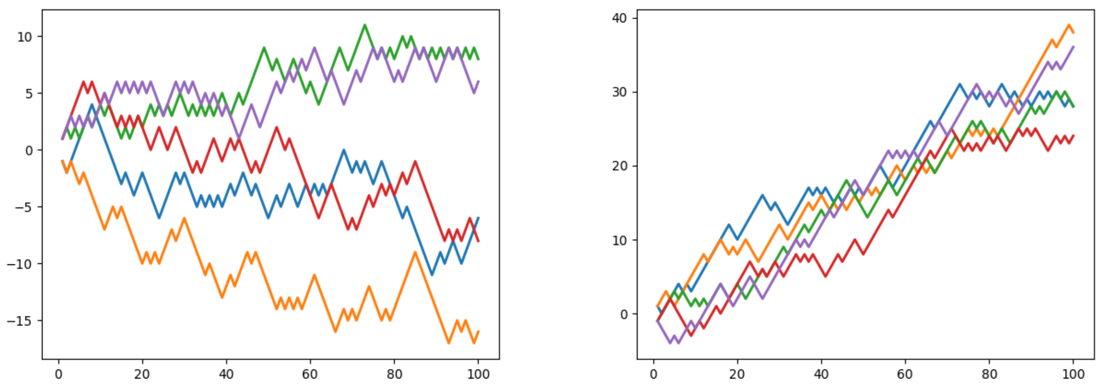
4.3.3 Random Walk

Now we will look at a stochastic process that will play a very important role in how we model stocks. We'll take a look at a simple random walk, which we will then take the limit of to obtain brownian motion.

Definition 4.11. *(Simple symmetric random walk) Let $\epsilon_1, \ldots, \epsilon_n$ be a sequence of independent random variables with $\mathbb{P}[\epsilon_n = 1] = \mathbb{P}[\epsilon_n = -1] = \frac{1}{2}$ for all n. We then define the process*

$$X_0 = 0,\ X_n = \sum_{k=1}^n \epsilon_k$$

as a simple symmetric random walk.



(a) Symmetric

(b) Non-symmetric with $\mathbb{P}[\epsilon_n = 1] = 0.70$

Figure 4: Random walks, 5 iterations.

Note that this is just an extension of a Bernoulli process. You can think of a random walk as just a sequence of coin flips in time, where every increment if we flip heads we move *up* by 1, and if we flip tails we move *down* by 1. Also notice that $\mathbf{E}[\epsilon_n] = 0$ and $\mathbf{Var}[\epsilon_n] = 1$ for all n.
The reader should be able to see that a simple symmetric random walk is in fact a martingale. We can see this below with some quick manipulation using the properties of conditional expectation

$$\mathbf{E}_n[X_{n+1}] = \mathbf{E}_n[X_n + \epsilon_{n+1}] = X_n + \mathbf{E}[\epsilon_{n+1}] = X_n$$

However, it is important to realize that this was only *symmetric* random walks. In these cases, we have the property that $\mathbf{E}[\epsilon_n] = 0$ for all n as mentioned earlier. We can also look at non-symmetric random walks, which are not martingales but are still interesting nonetheless.

At this point, the reader might start to see how random walks can aid us with modeling stocks. I mean, don't these graphs sort of look like how stocks behave? Well, not exactly. We defined random walks in a discrete sense, where movements up or down occur every iteration of n. But what exactly is an iteration of n? A year, a month, a day, a second? In the real world, stocks (basically) move instantaneously, so the next natural step would be to ask ourselves what happens to a random walk when the time in between coin flips approaches 0. This is the basic idea of brownian motion.

4.3.4 Brownian Motion

Now we will work towards our formal definition of brownian motion. To start, we will do exactly what we mentioned above: examine the simple random walk as the time in between coin flips goes to 0. To do this, we'll first divide the half-line $[0, \infty)$ into small intervals of size δ. So instead of our previous intervals of $1, 2, ..., n$, we now have $\delta, 2\delta, 3\delta, ...n\delta$. We can then define our coin tosses to meet this new setup - assume that at each time slot, we toss a fair coin. So between point 0 and any point $t = n\delta$, we have tossed n coins. But now, instead of moving up and down by 1, we will move up and down by $\sqrt{\delta}$.

Doing this brings us a new process. Let W_t be defined on $[0, \infty)$ where at each value t the value of W_t is given by

$$W_0 = 0,\ W_t = W_{n\delta} = \sqrt{\delta} \sum_{i=1}^{n} \epsilon_i$$

Note that the reason we chose step sizes of $\sqrt{\delta}$ is so that the variance of W_t remains equal to t. We can see now that

$$\mathbf{E}[W_t] = \sqrt{\delta} \sum_{i=1}^{n} \mathbf{E}[\epsilon_i] = 0$$

$$\mathbf{Var}[W_t] = \mathbf{Var}[\sqrt{\delta}\sum_{i=1}^{n}\epsilon_i] = \delta\sum_{i=1}^{n}\mathbf{Var}[\epsilon_i] = n\delta = t$$

Well what does the distribution of W_t look like? To answer this question, we will use the Central Limit Theorem which we defined earlier. Using some slight manipulation, we have the following

$$W_t = \sqrt{\delta}\sum_{i=1}^{n}\epsilon_i = \sqrt{\delta}\sum_{i=1}^{t/\delta}\epsilon_i = \sqrt{t}\frac{1}{\sqrt{t/\delta}}\sum_{i=1}^{t/\delta}\epsilon_i = \sqrt{t}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_i$$

Now let's take a step back and visualize what will happen as we take $\delta \to 0$. The behavior of W_t depends on what happens between point 0 and point $t = n\delta$. Between these two points, there are n flips. So as we take $\delta \to 0$, this is the same as taking $n \to \infty$. Essentially, we are flipping faster and faster (less time in between flips) until eventually between any two points, we've had an infinite amount of coin flips. Below we can see what happens to the trajectory of our random walk as we flip faster. As you can tell, things are starting to look more like the movement of stocks.

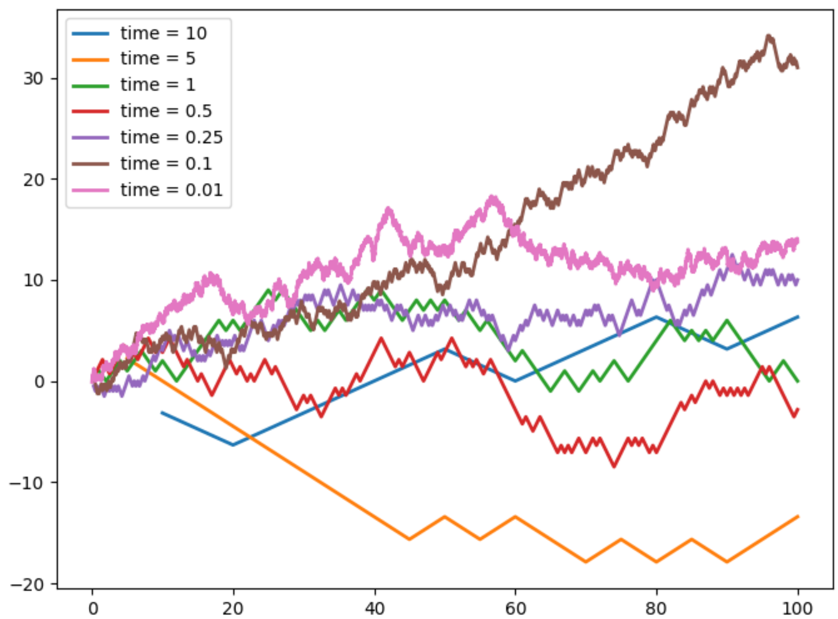


Figure 5: Random walk with various times in between coin flips.

Now this puts us in a position to use our Central Limit Theorem definition from earlier:

$$W_t = \sqrt{t}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_i \xrightarrow{n\to\infty} \sqrt{t}N(0,1) = N(0,t)$$

We can do the same process to see that the distribution between two points $s < t$ is

$$W_t - W_s \sim N(0, t-s)$$

Furthermore, we can see that $W_t - W_s$ is independent of W_s. In simple terms, the value of W_s only depends on the coin tosses up until time s, whereas the value of $W_t - W_s$ only depends on the coin tosses after s. By definition, we have independent tosses so therefore these values are independent. Another thing to note is the continuity and differentiability of the process we just uncovered. Informally, as $\delta \to 0$, we exhibit no jumps in the trajectory of W_t which grants us continuity. However, we absolutely cannot claim differentiability. Let's think about the process before we take $\delta \to 0$: at an individual coin flip, we move up or down 1 over the time period of δ, which is then scaled by $\sqrt{\delta}$. In other words

$$\frac{d}{dt}W_t \approx \sqrt{\delta}\frac{\pm 1}{\delta} = \pm\frac{1}{\sqrt{\delta}}$$

As we take $\delta \to 0$, we have no hope of this existing. The key here is that we have no guarantee on its *direction*, which we certainly need to claim differentiability. In summary, brownian motion is continuous everywhere, but differentiable nowhere.

Moving forward, we won't actually use this construction to define brownian motion. Rather, we present what we just discovered in the more commonly found definition:

Definition 4.12. *A standard Brownian motion is a continuous process such that*

1. $W_0 = 0$, $W_t - W_s \sim N(0, t-s)$
2. $W_t - W_s$ *is independent of* $\mathcal{F}_s$

Fantastic! Now that we have defined Brownian motion, we can finally explore how we'll actually use it to model assets.

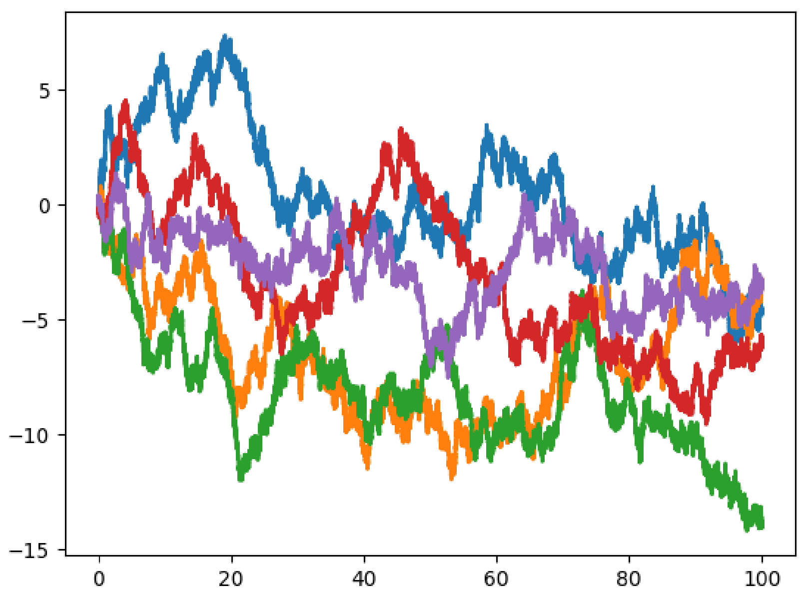


Figure 6: Brownian motion, 5 iterations.

5 Stochastic Calculus

6 Everything Black Scholes

In this chapter, we will derive the trillion dollar equation, the Black-Scholes PDE. The majority of the math from this chapter was taught to me by [Iye22]

6.1 Assumptions & Definitions

We make a few core assumptions in this derivation, namely:

1. Assets are liquid, ie. no slippage
2. Money market with constant continuously compounded interest rate r. Borrowing interest rate is equal to the lending rate
3. No transaction costs

Now the most important assumption, is that our underlying asset follows a Geometric Brownian Motion:

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

where

$$\alpha = \text{Mean Return}$$

$$\sigma = \text{Volatility}$$

and W_t is a standard Brownian Motion

6.1.1 Geometric Brownian Motion

For unfamiliar readers the above assumption might seem unusual. What is a Geometric Brownian motion and why do we model stocks as such? Consider the following:

We know that stocks generally exhibit positive drift over time. That is, stocks go up. This would incline us to naively model a stock as follows

$$dS_t = \alpha S_t dt$$

which of course we could write in the more familiar form

$$S_t = Ce^{\alpha t}$$

This is purely exponential growth. But stocks don't behave like this, if they did we would all be millionaires. Yes, stocks tend to exhibit a similar upward drift that we see with exponential growth, but also with random fluctuations. To model this mathematically, we introduce the dW_t term, which gives us the random behavior we are looking for, but still maintains the positive drift over time. Hence the form:

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

where $\alpha, \sigma \in \mathbb{R}$ are constants representing the mean return and mean volatility of the stock, respectively. The fact that these are constants (especially the volatility) is important, and is an assumption we will revisit in a later chapter.

Using some of our stochastic calculus knowledge from the previous chapter, we can use this to find a closed form expression for S_t.

Proposition 6.1. *For the geometric Brownian motion $dS_t = \alpha S_t dt + \sigma S_t dW_t$ we have*

$$S_t = S_0 \exp\left((\alpha - \frac{\sigma^2}{2})t + \sigma W_t\right)$$

Proof. To arrive at the above expression we will apply Ito's lemma to $\ln(S_t)$:

$$d\ln(S_t) = 0dt + \frac{1}{S_t}dS_t - \frac{1}{2}\frac{1}{S_t^2}d[S,S]_t$$

Recall from the last chapter what we learned about the quadratic variation of stochastic process defined by an Ito integral. Consider our geometric Brownian motion but written in integral form:

$$dS_t = \alpha S_t dt + \sigma S_t dW_t \implies S_t = S_0 + \int_0^t \alpha S_r dr + \int_0^t \sigma S_r dW_r$$

The Riemann term has finite first variation, and hence has 0 quadratic variation. Thus the quadratic variation of this stochastic process only depends on the Ito term, which we know is

$$[S,S]_t = \int_0^t \sigma^2 S_r^2 dr \implies d[S,S]_t = \sigma^2 S_t^2 dt$$

Using this and our expression for dS_t brings us to

$$d\ln(S_t) = \frac{1}{S_t}(\alpha S_t dt + \sigma S_t dW_t) - \frac{1}{2}\frac{1}{S_t^2}\sigma^2 S_t^2 dt$$

$$d\ln(S_t) = \alpha dt + \sigma dW_t - \frac{\sigma^2}{2}dt = (\alpha - \frac{\sigma^2}{2})dt + \sigma dW_t$$

Finally, taking an integral from 0 to t and using the properties of log gives

$$\ln(\frac{S_t}{S_0}) = (\alpha - \frac{\sigma^2}{2})t + \sigma W_t \implies S_t = S_0 \exp\left((\alpha - \frac{\sigma^2}{2})t + \sigma W_t\right)$$

□

Using this closed form expression we can now plot some iterations of geometric Brownian motion. Starting to look more like stocks, right?

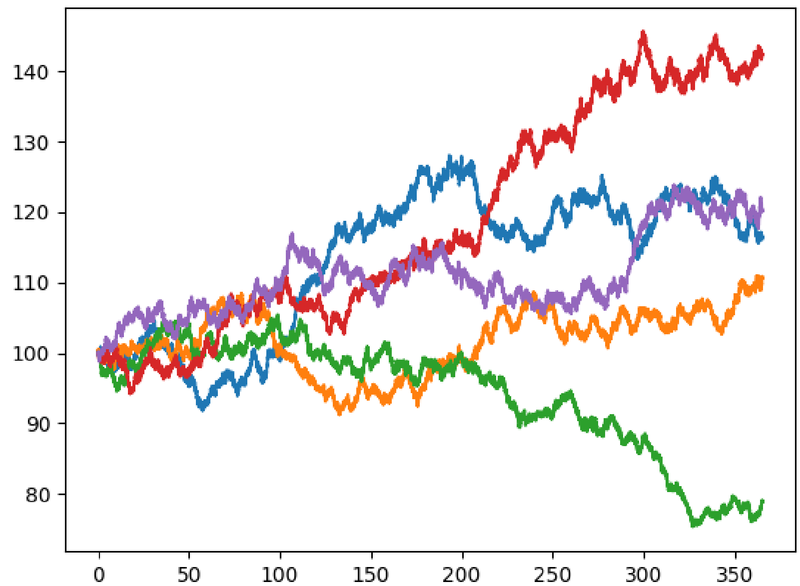


Figure 7: Geometric Brownian motion with $S_0 = 100$, $\alpha = 0.05$ and $\sigma = 0.15$, 5 iterations.

6.2 Setup

Now what question are we even trying to solve with the Black-Scholes PDE? What is our goal?
Consider a security that matures at time T and makes a single payment $V_T = g(S_T)$ at maturity T, where g is some deterministic function. In plain english, we are looking at some tradeable instrument that makes a single payment at a future time, where that payment is purely a function of the underlying security price at that time.
What we are looking to find is the arbitrage-free price of this security at some time $t < T$. That is, what can we say about V_t?
In our case, the underlying security is a stock, which we know behaves randomly. The problem statement becomes: can we find the price of a security whose payoff is based on the stock price at some fixed time in the future, even though the value of the stock at that time is unknown?
Now of course our common examples fall into this category, ie. European Call/Put. We'll list these cases below to demonstrate the usefullness of the presented problem:

$$\text{Call: } g(x) = (x - K)^+$$

$$\text{Put: } g(x) = (K - x)^+$$

where K is the fixed strike price of the Call or Put.
To tackle the evaluation of V_t we will use a technique that is fundamental to all of mathematical finance:

$$\text{PRICING} = \text{REPLICATION}$$

That is, we can construct a "self-financing" portfolio of traded assets that exactly replicates the payoff of the instrument we are trying to price. This means that the replicating portfolio's payoffs should exactly match those of the instrument we are pricing, making it a synthetic replica if you will.
This technique matches the role in industry of pricing such a security. If I as a bank sell you a call, then how can I use the proceeds of the sale to synthetically construct a 'hedging' portfolio that ensures my obligations are always covered. Hence the statement that pricing is exactly the same thing as replicating.
With that being said, let $X_t = f(t, S_t)$ be the wealth of a replicating portfolio of V_t, where f is some deterministic function of time and the underlying asset, which in our case is a stock.

6.2.1 Self Financing Condition

Note that we used the term 'self-financing' when describing the process of replication, but what exactly does this mean? And how can we define this mathematically?
In the simplest terms, this means that the purchase of a new asset in this portfolio must be financed by the sale of an old one. It is a portfolio that only uses tradeable assets and has no external cash flows.
In our case, the portfolio can consist of tradeable assets, which would be shares of stock and cash in the money market. At any time t we'll hold Δ_t shares of stock and hold the rest of our portfolio in cash.
It follows that the value of our portfolio at any time would be:

$$X_t = \Delta_t S_t + (X_t - \Delta_t S_t)$$

This should make sense to the reader. Our portfolio wealth at time t is given by the number of shares we hold at that time (Δ_t) multiplied by the value of the stock itself (S_t), plus the remaining amount ($X_t - \Delta_t S_t$) in cash.
Therefore, we can describe an infinitesimally small change in our portfolio by:

$$dX_t = \Delta_t dS_t + r(X_t - \Delta_t S_t)dt$$

Once again this should make sense as our change in portfolio wealth due to the stock holdings is the number of shares we hold multiplied by the change in stock price, and the change due to the cash holding is the amount of cash we hold multiplied by the continuously compounded rate multiplied by the change in time.
This expression is what we will use mathematically as the 'self-financing condition'.

6.3 Deriving the PDE

With that out of the way, now we can actually proceed with the core derivation. Recall that we are considering a security that makes a single payment of $V_T = g(S_T)$ at maturity time T.
Note the shorthand notation: $\partial_t f = \frac{\partial f}{\partial t}$ etc.

Theorem 6.1. *If the security can be replicated, and $f = f(t,x)$ is a function such that the wealth of the replicating portfolio is given by $X_t = f(t, S_t)$, then:*

$$\partial_t f + rx\partial_x f + \frac{\sigma^2 x^2}{2}\partial_x^2 f - rf = 0 \qquad x > 0, t < T,$$
$$f(t,0) = g(0)e^{-r(T-t)} \qquad t \leq T,$$
$$f(T,x) = g(x) \qquad x \geq 0.$$

What we see above is the infamous Black-Scholes partial differential equation, with boundary conditions.

Proof. Assume that the security is replicable and $f = f(t,x)$ is a function such that the wealth of the replicating portfolio is given by $X_t = f(t, S_t)$.
By definition, we then know that the replicating portfolio must be self-financing, and thus we have

$$dX_t = \Delta_t dS_t + r(X_t - \Delta_t S_t)dt$$

We also know by assumption that S_t follows a geometric brownian motion and so

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

Combining these two expressions gives us

$$dX_t = \Delta_t(\alpha S_t dt + \sigma S_t dW_t) + r(X_t - \Delta_t S_t)dt$$

which then yields

$$dX_t = ((\alpha - r)\Delta_t S_t + rX_t)dt + \sigma\Delta_t S_t dW_t \qquad (*)$$

Next, we will apply Ito's Lemma on $X_t = f(t, S_t)$. Note the shorthand notation $f = f(t, S_t)$.

$$dX_t = df = \partial_t f dt + \partial_x f dS_t + \frac{1}{2}\partial_x^2 f d[S,S]_t$$

Once again, we know that S_t follows a geometric brownian motion, so we know that the quadratic variation of this stochastic process is $[S,S]_t = \sigma^2 S_t^2$ and hence $d[S,S]_t = \sigma^2 S_t^2 dt$. This brings us to

$$dX_t = df = \partial_t f dt + \partial_x f(\alpha S_t dt + \sigma S_t dW_t) + \frac{1}{2}\partial_x^2 f \sigma^2 S_t^2 dt$$

which then yields

$$dX_t = (\partial_t f + \alpha S_t \partial_x f + \frac{1}{2}\sigma^2 S_t^2 \partial_x^2 f)dt + \sigma S_t \partial_x f dW_t \qquad (**)$$

Now equate the dW_t and dt terms in $(*)$ and $(**)$

$$dW_t : \qquad \sigma\Delta_t S_t = \sigma S_t \partial_x f$$

this yields what is known as the delta hedging rule, which has vast applications:

$$\Delta_t = \partial_x f(t, S_t)$$

Recalling that $X_t = f(t, S_t)$ we get

$$dt : \qquad (\alpha - r)\Delta_t S_t + rf = \partial_t f + \alpha S_t \partial_x f + \frac{1}{2}\sigma^2 S_t^2 \partial_x^2 f$$

Using the delta hedging rule we just uncovered, we have

$$\alpha S_t \partial_x f - r\partial_x f S_t + rf = \partial_t f + \alpha S_t \partial_x f + \frac{1}{2}\sigma^2 S_t^2 \partial_x^2 f$$

Cancelling like terms, we finally arrive at

$$\partial_t f + rS_t \partial_x f + \frac{1}{2}\sigma^2 S_t^2 \partial_x^2 f - rf = 0$$

□

6.4 Using the PDE

Now we'd also like to show the converse, so that we can actually use solutions to the Black Scholes PDE to give us the arbitrage free price of securities.

Theorem 6.2. *If f solves the Black-Scholes PDE, then the security can be replicated and $X_t = f(t, S_t)$ is the wealth of the replicating portfolio at any time $t \leq T$.*

Note that the wealth of the replicating portfolio is equal to the arbitrage-free price of the security, so proving the above statement will unlock the blueprint of computing the price of this security, which was our original goal.

Proof. We assume that f solves the equations presented in Theorem 3.1. We choose the following

1. $X_0 = f(0, S_0)$
2. $\Delta_t = \partial_x f(t, S_t)$

The second choice makes use of the delta hedging rule that we uncovered in the previous theorem.
To prove our claim we will first show that $X_t = f(t, S_t), \forall t \in (0, T)$
It suffices to show

$$e^{-rt}X_t = e^{-rt}f(t, S_t), \forall t \in (0, T)$$

To continue, we will apply Ito's Lemma on the LHS of this expression

$$d(e^{-rt}X_t) = -re^{-rt}X_t dt + e^{-rt}dX_t + 0$$

Recall the self financing condition on X_t as well as the fact that S_t is a geometric brownian motion

$$d(e^{-rt}X_t) = -re^{-rt}X_t dt + e^{-rt}(\Delta_t dS_t + r(X_t - \Delta_t S_t)dt)$$

$$d(e^{-rt}X_t) = -re^{-rt}X_t dt + e^{-rt}(\Delta_t(\alpha S_t dt + \sigma S_t dW_t) + r(X_t - \Delta_t S_t)dt)$$

Cleaning this up...

$$d(e^{-rt}X_t) = (\alpha - r)e^{-rt}\Delta_t S_t dt + e^{-rt}\Delta_t \sigma S_t dW_t$$

Now we can also apply Ito's Lemma in combination with the product rule to the RHS of the desired expression

$$d(e^{-rt}f(t, S_t)) = (-re^{-rt}f + e^{-rt}\partial_t f)dt + e^{-rt}\partial_x f dS_t + \frac{1}{2}e^{-rt}\partial_x^2 f d[S, S]_t$$

Recalling that $dS_t = \alpha S_t dt + \sigma S_t dW_t$ and $d[S, S]_t = \sigma^2 S_t^2 dt$

$$d(e^{-rt}f(t, S_t)) = e^{-rt}(-rf + \partial_t f + \partial_x f \alpha S_t + \frac{1}{2}\partial_x^2 f \sigma^2 S_t^2)dt + e^{-rt}\partial_x f \sigma S_t dW_t$$

Now add and subtract $rS_t\partial_x f$ from the dt term

$$d(e^{-rt}f(t, S_t)) = e^{-rt}(-rf + \partial_t f + \partial_x f \alpha S_t + \frac{1}{2}\partial_x^2 f \sigma^2 S_t^2 + rS_t\partial_x f - rS_t\partial_x f)dt + e^{-rt}\partial_x f \sigma S_t dW_t$$

Since we assumed that f solves the PDE, this simplifies to

$$d(e^{-rt}f(t, S_t)) = e^{-rt}(\alpha - r)\partial_x f S_t dt + e^{-rt}\partial_x f \sigma S_t dW_t$$

bringing us to

$$d(e^{-rt}f(t, S_t)) = e^{-rt}(\alpha - r)\Delta_t S_t dt + e^{-rt}\Delta_t \sigma S_t dW_t$$

which is precisely equal to what we derived for the LHS of our desired expression
Therefore we have the following

$$d(e^{-rt}(X_t - f(t, S_t))) = 0$$

thus integrating from 0 to t gives

$$e^{-rt}(X_t - f(t, S_t)) = e^{-r0}(X_0 - f(0, S_0))$$

but we have $X_0 - f(0, S_0) = 0$ by choice so this yields our desired result

$$X_t = f(t, S_t), \forall t \in (0, T)$$

Now that we have this, we can then take the limit

$$X_T = \lim_{t \to T} X_t = \lim_{t \to T} f(t, S_t) = f(T, S_T) = g(T)$$

which implies that the wealth of the replicating portfolio at time T is equal to the payoff of the security. Combining this gives us that $X_t = f(t, S_t), \forall t \in [0, T]$, completing the argument that the security is replicable and the wealth of the replicating portfolio X_t is equal to the arbitrage free price of the security.

□

Remark. The choice to show $X_t = f(t, S_t), \forall t \in (0, T)$ and then utilize a limit argument to extend to the closed interval $[0, T]$ was not arbitrary. To be able to apply Ito's Lemma, we need a function $f(t, x)$ continuous in t and twice continuous in x. Consider the case of a call option, where $g(x) = (x - K)^+$. This is discontinuous in x at at $t = T$, which is why we used the closed interval $(0, T)$ when applying Ito's Lemma.

Remark. The above proof is still a work in progress - need to explain the choice of discounting in the proof

At this point we can now find explicit solutions f to the Black Scholes PDE in order to find the arbitrage free price of a security.

6.5 Finding Solutions to the PDE

Great, so now we have a PDE that we can find solutions to in order to price securities, but how do we actually find solutions? You might have noticed that the Black Scholes PDE looks suspiciously similar to the Heat Equation. It is in fact a variant - with some clever substitutions we can transform the Black Scholes PDE into the Heat Equation, whose solution is well documented and can be found with a standard change of variables. However, we aren't going to do this. Instead, we will revisit and prove the following proposition in a later chapter using some new tools we will uncover.

Proposition 6.2. *An explicit solution to the Black Scholes PDE presented in Theorem 3.1 is the following*

$$f(t, x) = \int_{-\infty}^{\infty} e^{-r\tau} g(x \exp((r - \frac{\sigma^2}{2})\tau + \sigma\sqrt{\tau} y)) \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} dy, \;\; \tau = T - t$$

Essentially, for a security that makes a single payment of $V_T = g(S_T)$ at maturity time T, we can use this explicit solution to find an analytical expression for the arbitrage free price of such a security. To demonstrate this, we will use Proposition 3.1 in the case of a European Call option

Theorem 6.3. *For European Calls, $g(x) = (x - K)^+$, and*

$$f(t, x) = c(t, x) = xN(d_+(T - t, x)) - Ke^{-r(T-t)} N(d_-(T - t, x))$$

where

$$d_\pm(\tau, x) = \frac{1}{\sigma\sqrt{\tau}} (\ln(\frac{x}{K}) + (r \pm \frac{\sigma^2}{2})\tau)$$

and

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy$$

Proof. Using Proposition 3.1 and $g(x) = (x - K)^+$, we have the following

$$f(t, x) = c(t, x) = \int_{-\infty}^{\infty} e^{-r\tau} [x \exp((r - \frac{\sigma^2}{2})\tau + \sigma\sqrt{\tau} y) - K]^+ \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} dy, \;\; \tau = T - t$$

To get rid of the positive part operator, consider where the integrand is non-zero:

$$x\exp((r-\frac{\sigma^2}{2})\tau+\sigma\sqrt{\tau}y)-K>0$$

$$\exp((r-\frac{\sigma^2}{2})\tau+\sigma\sqrt{\tau}y)>\frac{K}{x}$$

Note that this is valid because the stock price x is non-negative. After further manipulation we then have

$$y>\frac{1}{\sigma\sqrt{\tau}}(\ln(\frac{K}{x})-(r-\frac{\sigma^2}{2})\tau)$$

If we let

$$d_\pm(\tau,x)=\frac{1}{\sigma\sqrt{\tau}}(\ln(\frac{x}{K})+(r\pm\frac{\sigma^2}{2})\tau)$$

then we see that the integrand is non-negative for

$$y>-d_-(\tau,x)$$

Therefore

$$c(t,x)=\int_{-d_-}^{\infty}e^{-r\tau}(x\exp((r-\frac{\sigma^2}{2})\tau+\sigma\sqrt{\tau}y)-K)\frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}dy,\ \ \tau=T-t$$

which we will split into two terms and simplify separately

$$c(t,x)=\int_{-d_-}^{\infty}e^{-r\tau}(x\exp((r-\frac{\sigma^2}{2})\tau+\sigma\sqrt{\tau}y)\frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}dy-\int_{-d_-}^{\infty}e^{-r\tau}K\frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}dy$$

Starting with the second term:

$$Ke^{-r\tau}\frac{1}{\sqrt{2\pi}}\int_{-d_-}^{\infty}e^{-\frac{y^2}{2}}dy=Ke^{-r\tau}(1-N(-d_-))=Ke^{-r\tau}N(d_-)$$

Now for the second term. Note that we can combine the exponent terms and complete the square:

$$-\frac{1}{2}y^2+\sigma\sqrt{\tau}y+r\tau-\tau\frac{\sigma^2}{2}-r\tau=-\frac{1}{2}(y^2-2\sigma\sqrt{\tau}y+\tau\sigma^2)=-\frac{1}{2}(y-\sigma\sqrt{\tau})^2$$

with this in mind the second term becomes

$$x\frac{1}{\sqrt{2\pi}}\int_{-d_-}^{\infty}e^{-\frac{1}{2}(y-\sigma\sqrt{\tau})^2}dy$$

let $u=y-\sigma\sqrt{\tau},du=dy$

$$x\frac{1}{\sqrt{2\pi}}\int_{-d_--\sigma\sqrt{\tau}}^{\infty}e^{-\frac{1}{2}u^2}du=x(1-N(-d_--\sigma\sqrt{\tau}))=xN(d_-+\sigma\sqrt{\tau})$$

Note that

$$d_-+\sigma\sqrt{\tau}=\frac{1}{\sigma\sqrt{\tau}}(\ln(\frac{x}{K})+\tau r-\tau\frac{\sigma^2}{2}+\sigma^2\tau)=\frac{1}{\sigma\sqrt{\tau}}(\ln(\frac{x}{K})+(r+\frac{\sigma^2}{2})\tau)=d_+(\tau,x)$$

Therefore, combining these simplified terms we arrive at

$$c(t,x)=xN(d_+(T-t,x))-Ke^{-r(T-t)}N(d_-(T-t,x))$$

□

6.6 Put-Call Parity

Now we could go through the same process to find an explicit solution for European Puts, but is there a simpler way? Consider the following identity:

$$(x-K)^+ - (K-x)^+ = x - K$$

This is immediately clear from the definition of the positive part operator. But how does this help us? Well remember that the payoff of a European Call at maturity is $g(S_T) = (S_T - K)^+$ and for a European Put is $g(S_T) = (K - S_T)^+$. So, if we consider a portfolio that is long a call and short a put, then our payoff at maturity is

$$(S_T - K)^+ - (K - S_T)^+ = S_T - K$$

which is precisely the payoff of a forward contract with the same strike as the options. Pretty cool, so we basically constructed a "synthetic forward" by combining two options positions. Using what we know about arbitrage-free pricing and what we learned in the previous section, we can price this synthetic forward. We know the value of this portfolio at maturity, so if we want to price it at a time $t < T$ then we can simply discount strike which gives us the following proposition, what we call Put-Call parity.

Proposition 6.3. *Let $c(t,x)$ and $p(t,x)$ be the arbitrage-free prices of a Call and Put respectively, both with maturity T and strike K. Then we have*

$$c(t,x) - p(t,x) = x - Ke^{-r(T-t)}$$

and hence

$$p(t,x) = c(t,x) - x + Ke^{-r(T-t)}$$

Therefore, as stated we have an easier way to determine an explicit solution for the price of a put option, rather than going through the same tedious process we did for calls.

Theorem 6.4. *For European Puts, $g(x) = (K-x)^+$, and*

$$f(t,x) = p(t,x) = Ke^{-r(T-t)}N(-d_-(T-t,x)) - xN(-d_+(T-t,x))$$

where

$$d_\pm(\tau,x) = \frac{1}{\sigma\sqrt{\tau}}(\ln(\frac{x}{K}) + (r \pm \frac{\sigma^2}{2})\tau)$$

and

$$N(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{y^2}{2}}dy$$

7 Risk Management I

In this chapter, we will begin to shift our attention to the more practical aspects of options trading. We have established a strong theoretical backing in the pricing of these securities, so now we can begin to appreciate their dynamic risk management. Of course, we will not shy away from mathematical explanations of an option's risk profile, however, we will also aim to establish a strong intuition when it comes to understanding how an option's price reacts to different factors.

To start, lets review what we already know. In Chapter 1 we observed that an option's price consists of two components: intrinsic value and time value. The following risks we will discuss are concerned with both the option's time value and intrinsic value. Recall that we also discovered that an option's price increases with volatility, hence the idea that options provide such volatility exposure.

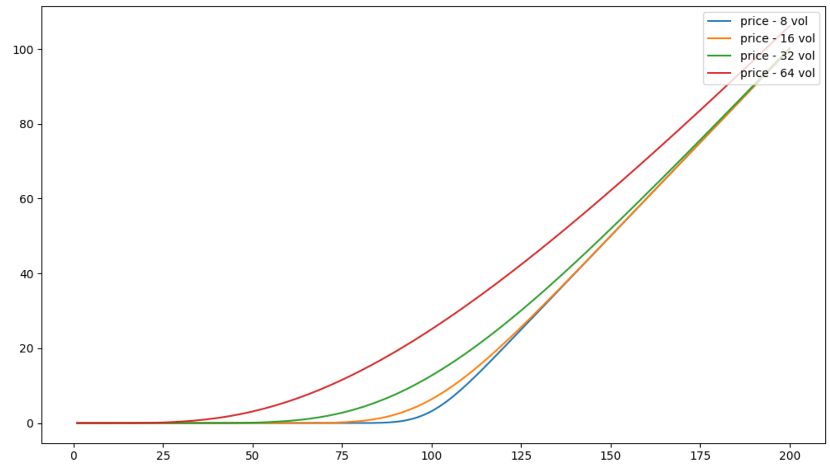


Figure 8: Price of a call across different volatilities, maturity fixed at $T = 1$ year.

But observe in particular how the price changes more dramatically around the strike price, ie. at the money. How an option's price changes at different points in the dimension of underlying (spot) price will be a critical dynamic for understanding risk. Similarly, we can also observe how price changes with respect to another dimension: time to maturity. As we increase the time to maturity of an option, we are essentially allowing more time for our underlying asset to move. In general, this will increase the time value of our option and hence will increase the overall price. However, depending on where the underlying price is located relative to our strike price, the impact will also be different.

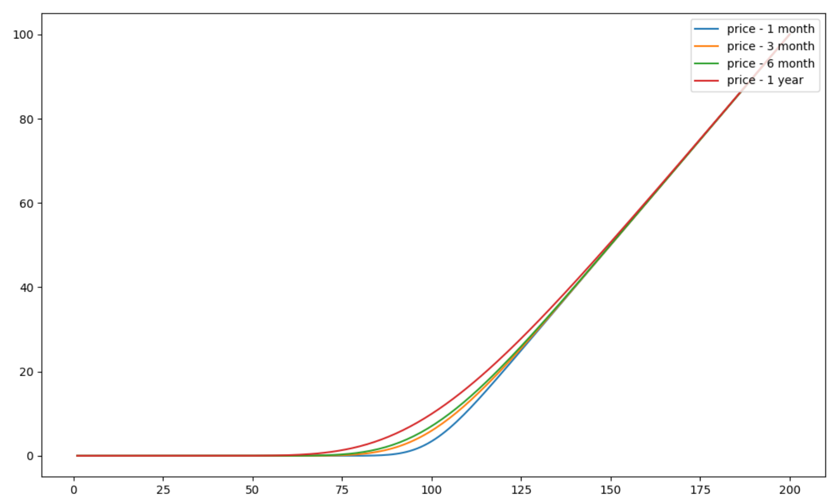


Figure 9: Price of a call across different maturities, volatility fixed at $\sigma = 0.25$.

How option's prices and risks change over the dimension of time will also be essential for this chapter.

In fact, understanding how price and risk changes with respect to volatility and maturity across spot space will allow us to describe the majority of an option's core risk characteristics.

When we introduced basic options theory in Chapter 1, we mentioned that their theoretical pricing models are concerned with six core inputs:

1. K - Strike
2. T - Time to expiration
3. S - Price of the underlying
4. r - Interest rate
5. q - Dividend yield
6. σ - Volatility of the underlying

Throughout this chapter, as we describe different risks associated with options we will essentially be answering the following questions: *how does the price of an option change with respect to one of these inputs, holding all others constant?* To the mathematically curious, this sounds all too familiar - this is the concept of a partial derivative. Therefore, the following risks are simply just the partial derivatives of an option's price. For me personally, thinking of them as such is helpful and intuitive, and so I urge you to remember the following idea: trading an option is just trading a mathematical object that can be represented by a function. When we want to know how this object's price will change with respect to changes in the market, we are merely asking how our function will behave with respect to changes in one of its parameters.

Lastly, note that we will be describing risk in the context of a Black-Scholes world. What do I mean by this? While this model gives us massive insight into the dynamics of options trading, it is not perfect. There are some important flawed assumptions in this model that we will revisit in a later chapter, which will have an impact on risk dynamics. Real world dynamics have some key differences from this mathematical landscape. However, for the purpose of an initial introduction, working in a Black-Scholes world will suffice. Now with that being said, lets get started.

7.1 Delta

The first and most important risk we will discuss is an option's *Delta.* In simplest terms, delta describes our exposure to directional movements in the underlying security.

Definition 7.1. *Delta is the rate of change in the theoretical option value with respect to changes in the underlying asset's price. More formally, delta is the first partial derivative of the value V of the option with respect to the underlying's price S.*

$$\Delta = \frac{\partial V}{\partial S}$$

This should look familiar from the previous chapter, as we utilized the delta hedging rule to derive our Black-Scholes equation. That is why it is so powerful - it tells us how many shares in the underlying we must own to exactly replicate an option's exposure to directional price movements in that asset. For a vanilla call, delta will be a number between 0 and 1, and for a vanilla put it will be between 0 and -1. Depending on price, a call option behaves as if one owns 1 share of the underlying stock (ie. deep in the money), or owns nothing (ie. far out of the money), or something in between, and conversely for a put option.

Proposition 7.1. *The delta for European calls and puts at any time $\tau = T - t$ are given by the following, respectively:*

$$\Delta_C = e^{-r\tau}N(d_1), \;\; \Delta_P = -e^{-r\tau}N(-d_1)$$

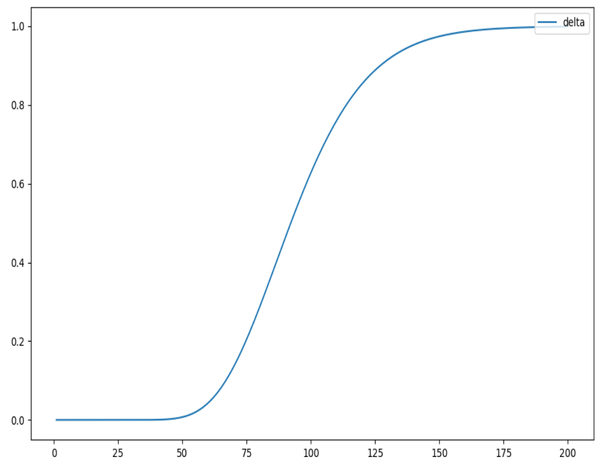
where

$$d_1 = \frac{1}{\sigma\sqrt{\tau}}(\ln(\frac{S}{K}) + (r - q + \frac{\sigma^2}{2})\tau), \;\; N(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{y^2}{2}}\,dy$$

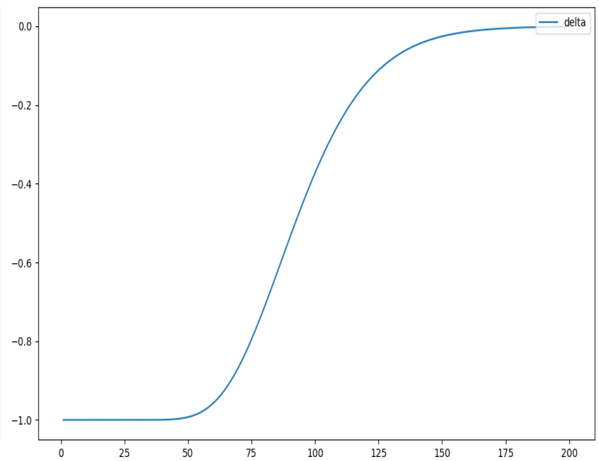
This of course matches our intuition that call delta is positive and put delta is negative. Also recalling put-call parity, we can relate the delta for calls and puts. For a call and put of the same strike we have the following

$$\Delta_C - \Delta_P = 1 \implies \Delta_C = \Delta_P + 1, \Delta_P = \Delta_C - 1$$

Using the closed form expression from above, we can see how the delta of a call or put option changes across spot space.



(a) Call Delta



(b) Put Delta

Figure 10: Delta for a call and put with $K = 100, T = 1, \sigma = 0.25$.

We can see that the dynamics of delta for calls and puts are nearly identical except for the opposing signs. Based on these figures, it might seem like delta can serve as a proxy for the probability that an option expires in the money. Well, the absolute value of delta is in fact close to, but not identical to, the implied probability that an option will expire in-the-money under the risk-neutral measure. It can serve as an estimate for traders, but the true implied probability of finishing in the money is closer related to the price of a digital option (we will revisit this later).

7.1.1 Gamma

The second most important risk of an option is its *Gamma*. At its essence, the gamma is what represents an option's *convexity* - increased profits in upward movements of the underlying and decreased losses in downward movements.

Definition 7.2. *Gamma is the rate of change in an option's delta with respect to changes in the underlying asset's price. More formally, gamma is the second partial derivative of the value V of the option with respect to the underlying's price S.*

$$\Gamma = \frac{\partial \Delta}{\partial S} = \frac{\partial^2 V}{\partial S^2}$$

In contrast to delta, the gamma of puts and calls are the same sign: positive. This is because both puts and calls offer the desirable convexity of being long an option. In general, long options positions will be long gamma whereas short positions will be short gamma.

The convexity of an option can be interpreted as the value of optionality. Recall the critical Jensen's inequality: the expected value of a convex function is greater than or equal to the function of the expected value

$$\mathbf{E}[f(X)] \geq f(\mathbf{E}[X])$$

In our case, the value of an option is due to the convexity of the ultimate payout: one has the option to buy an asset or not (for a call), and the ultimate payout function is convex. Thus, if one purchases a call option, the expected value of the option is higher than simply taking the expected future value of the underlying and inputting it into the option payout function. The price of the option – the value of the optionality – thus reflects the convexity of the payoff. Generally, being long convexity (gamma) means one benefits from the realized volatility of the underlying.

Of course, there is no free lunch. The benefits convexity will come at a cost - that cost being time. We will soon discuss the theta of an option, its intimate relationship with gamma, and how both relate to implied and realized volatility.

Proposition 7.2. *The gamma for European calls and puts at any time $\tau = T - t$ is given by the following:*

$$\Gamma = e^{-q\tau}\frac{\phi(d_1)}{S\sigma\sqrt{\tau}} = Ke^{-r\tau}\frac{\phi(d_2)}{S^2\sigma\sqrt{\tau}}$$

where

$$d_2 = d_1 - \sigma\sqrt{\tau}, \;\; \phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

Knowing that gamma is the second derivative of an option's price with respect to spot price, we can see that its profile should just be the derivative of the delta graph we previously viewed.

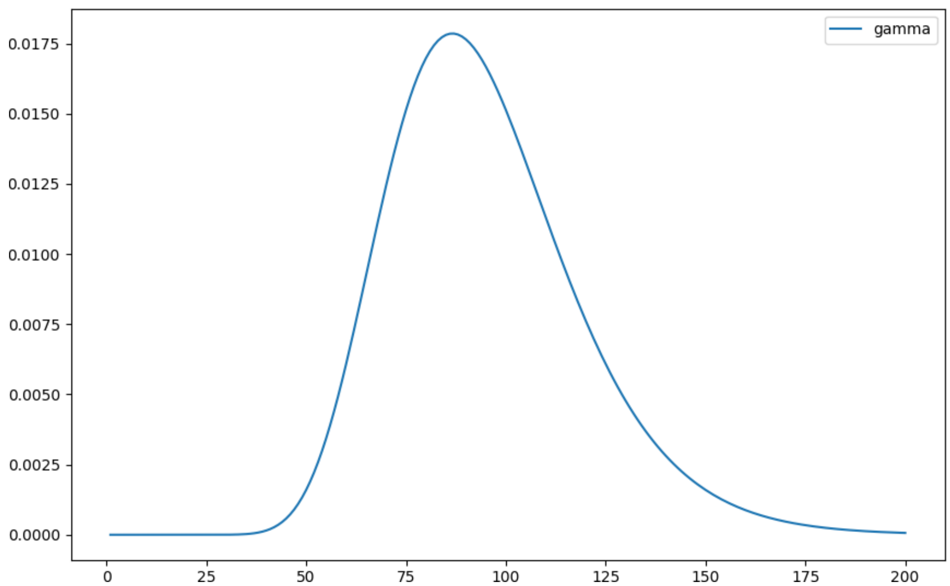


Figure 11: Gamma for a call and put with $K = 100, T = 1, \sigma = 0.25$

As expected, the gamma of an option is highest when underlying spot is (roughly) at-the-money. This should make sense as our optionality is the most desirable at this point. When spot trades at the strike price of the option, then the holder will benefit the most from the added optionality because whether the option will expire in or out of the money is the most uncertain. When we are deep ITM or OTM the optionality means less - we have more certainty in which state our option's position will expire in. This intuitive explanation also matches our understanding of delta: when we are deep OTM, the underlying price must move up by a large amount (an improbable event) to end in the money, hence why our delta is near-zero and has a rate of change of near-zero. As we approach strike price, the delta increases at a higher rate towards ~ 0.50. Then once we pass ATM and move deep ITM, the underlying price must move now down by a large amount (another improbable event) in order to end out of the money, hence why the delta is practically 1 and behaves the same as the underlying from a directional view. This matches the behavior of the derivative of the delta graph with respect to spot.

This is what I love about options theory. For every intuitive explanation of the security, there exists a corresponding backing in mathematical theory. The ultimate ability is to understand options from both perspectives, which is exactly what we are trying to establish in this book.

7.2 Theta

7.3 Vega

7.4 Rho

7.5 Higher Order Greeks

7.5.1 Vanna

7.5.2 Veta

7.5.3 Volga

7.5.4 Theta vs. Time-to-expiry

7.5.5 Charm

8 Everything Volatility

8.1 The Basics

8.2 Limitations of Black Scholes

9 Risk Management II

References

[Ben14] Colin Bennett. *Trading Volatility.* Colin Bennett, 2014.

[Iye22] Gautam Iyer. Lecture notes on continuous time finance fall 2022. pages 1–21, 2022.