Foundations of Reinforcement Learning with Applications in Finance

Ashwin Rao, Tikhon Jelvis

1 Derivatives Pricing and Hedging

In this chapter, we cover two applications of MDP Control regarding financial derivatives' pricing and hedging (the word *hedging* refers to reducing or eliminating market risks associated with a derivative). The first application is to identify the optimal time/state to exercise an American Option (a type of financial derivative) in an idealized market setting (akin to the "frictionless" market setting of Merton's Portfolio problem from Chapter **??**). Optimal exercise of an American Option is the key to determining it's fair price. The second application is to identify the optimal hedging strategy for derivatives in real-world situations (technically refered to as *incomplete markets*, a term we will define shortly). The optimal hedging strategy of a derivative is the key to determining it's fair price in the real-world (incomplete market) setting. Both of these applications can be cast as Markov Decision Processes where the Optimal Policy gives the Optimal Hedging/Optimal Exercise in the respective applications, leading to the fair price of the derivatives under consideration. Casting these derivatives applications as MDPs means that we can tackle them with Dynamic Programming or Reinforcement Learning algorithms, providing an interesting and valuable alternative to the traditional methods of pricing derivatives.

In order to understand and appreciate the modeling of these derivatives applications as MDPs, one requires some background in the classical theory of derivatives pricing. Unfortunately, thorough coverage of this theory is beyond the scope of this book and we refer you to Tomas Bjork's book on Arbitrage Theory in Continuous Time (Björk 2005) for a thorough understanding of this theory. We shall spend much of this chapter covering the very basics of this theory, and in particular explaining the key technical concepts (such as arbitrage, replication, risk-neutral measure, market-completeness etc.) in a simple and intuitive manner. In fact, we shall cover the theory for the very simple case of discrete-time with a single-period. While that is nowhere near enough to do justice to the rich continuous-time theory of derivatives pricing and hedging, this is the best we can do in a single chapter. The good news is that MDP-modeling of the two problems we want to solve - optimal exercise of american options and optimal hedging of derivatives in a real-world (incomplete market) setting - doesn't require one to have a thorough understanding of the classical theory. Rather, an intuitive understanding of the key technical and economic concepts should suffice, which we bring to life in the simple setting of discrete-time with a single-period. We start this chapter with a quick introduction to derivatives, next we describe the simple setting of a single-period with formal mathematical notation, covering the key concepts (arbitrage, replication, risk-neutral measure, market-completeness etc.), state and prove the all-important fundamental theorems of asset pricing (only for the single-period setting), and finally show how these two derivatives applications can be cast as MDPs, along with the appropriate algorithms to solve the MDPs.

1.1 A Brief Introduction to Derivatives

If you are reading this book, you likely already have some familiarity with Financial Derivatives (or at least have heard of them, given that derivatives were at the center of the 2008 financial crisis). In this section, we sketch an overview of financial derivatives and refer

you to the book by John Hull (Hull 2010) for a thorough coverage of Derivatives. The term "Derivative" is based on the word "derived" - it refers to the fact that a derivative is a financial instrument whose structure and hence, value is derived from the *performance* of an underlying entity or entities (which we shall simply refer to as "underlying"). The underlying can be pretty much any financial entity - it could be a stock, currency, bond, basket of stocks, or something more exotic like another derivative. The term *performance* also refers to something fairly generic - it could be the price of a stock or commodity, it could be the interest rate a bond yields, it could be the average price of a stock over a time interval, it could be a market-index, or it could be something more exotic like the implied volatility of an option (which itself is a type of derivative). Technically, a derivative is a legal contract between the derivative buyer and seller that either:

- Entitles the derivative buyer to cashflow (which we'll refer to as derivative *payoff*) at future point(s) in time, with the payoff being contingent on the underlying's performance (i.e., the payoff is a precise mathematical function of the underlying's performance, eg: a function of the underlying's price at a future point in time). This type of derivative is known as a "lock-type" derivative.
- Provides the derivative buyer with choices at future points in time, upon making which, the derivative buyer can avail of cashflow (i.e., *payoff*) that is contingent on the underlying's performance. This type of derivative is known as an "option-type" derivative (the word "option" refering to the choice or choices the buyer can make to trigger the contingent payoff).

Although both "lock-type" and "option-type" derivatives can both get very complex (with contracts running over several pages of legal descriptions), we now illustrate both these types of derivatives by going over the most basic derivative structures. In the following descriptions, current time (when the derivative is bought/sold) is denoted as time $t = 0$.

1.1.1 Forwards

The most basic form of Forward Contract involves specification of:

- A future point in time $t = T$ (we refer to T as expiry of the forward contract).
- The fixed payment K to be made by the forward contract buyer to the seller at time $t = T$.

In addition, the contract establishes that at time $t = T$, the forward contract seller needs to deliver the underlying (say a stock with price S_t at time t) to the forward contract buyer. This means at time $t = T$, effectively the payoff for the buyer is $S_T - K$ (likewise, the payoff for the seller is $K - S_T$). This is because the buyer, upon receiving the underlying from the seller, can immediately sell the underlying in the market for the price of S_T and so, would have made a gain of $S_T - K$ (note $S_T - K$ can be negative, in which case the payoff for the buyer is negative).

The problem of forward contract "pricing" is to determine the fair value of K so that the price of this forward contract derivative at the time of contract creation is 0. As time t progresses, the underlying price might fluctuate, which would cause a movement away from the initial price of 0. If the underlying price increases, the price of the forward would naturally increase (and if the underlying price decreases, the price of the forward would naturally decrease). This is an example of a "lock-type" derivative since neither the buyer

nor the seller of the forward contract need to make any choices at time $t = T$. Rather, the payoff for the buyer is determined directly by the formula $S_T - K$ and the payoff for the seller is determined directly by the formula $K - S_T$.

1.1.2 European Options

The most basic forms of European Options are European Call and Put Options. The most basic European Call Option contract involves specification of:

- A future point in time $t = T$ (we refer to T as the expiry of the Call Option).
- Underlying Price K known as *Strike Price*.

The contract gives the buyer (owner) of the European Call Option the right, but not the obligation, to buy the underlying at time $t = T$ at the price of K. Since the option owner doesn't have the obligation to buy, if the price S_T of the underlying at time $t = T$ ends up being equal to or below K, the rational decision for the option owner would be to not buy (at price K), which would result in a payoff of 0 (in this outcome, we say that the call option is *out-of-the-money*). However, if $S_T > K$, the option owner would make an instant profit of $S_T - K$ by *exercising* her right to buy the underlying at the price of K. Hence, the payoff in this case is $S_T - K$ (in this outcome, we say that the call option is *in-the-money*). We can combine the two cases and say that the payoff is $f(S_T) = \max(S_T - K, 0)$. Since the payoff is always non-negative, the call option owner would need to pay for this privilege. The amount the option owner would need to pay to own this call option is known as the fair price of the call option. Identifying the value of this fair price is the highly celebrated problem of *Option Pricing* (which you will learn more about as this chapter progresses).

A European Put Option is very similar to a European Call Option with the only difference being that the owner of the European Put Option has the right (but not the obligation) to *sell* the underlying at time $t = T$ at the price of K. This means that the payoff is $f(S_T) = \max(K - S_T, 0)$. Payoffs for these Call and Put Options are known as "hockey-stick" payoffs because if you plot the $f(\cdot)$ function, it is a flat line on the *out-of-the-money* side and a sloped line on the *in-the-money* side. Such European Call and Put Options are "Option-Type" (and not "Lock-Type") derivatives since they involve a choice to be made by the option owner (the choice of exercising the right to buy/sell at the Strike Price K). However, it is possible to construct derivatives with the same payoff as these European Call/Put Options by simply writing in the contract that the option owner will get paid $\max(S_T - K, 0)$ (in case of Call Option) or will get paid $\max(K - S_T, 0)$ (in case of Put Option) at time $t = T$. Such derivatives contracts do away with the option owner's exercise choice and hence, they are "Lock-Type" contracts. There is a subtle difference - setting these derivatives up as "Option-Type" means the option owner might act "irrationally" - the call option owner might mistakenly buy even if $S_T < K$, or the call option owner might for some reason forget/neglect to exercise her option even when $S_T > K$. Setting up such contracts as "Lock-Type" takes away the possibilities of these types of irrationalities from the option owner.

A more general European Derivative involves an arbitrary function $f(\cdot)$ (generalizing from the hockey-stick payoffs) and could be set up as "Option-Type" or "Lock-Type."

1.1.3 American Options

The term "European" above refers to the fact that the option to exercise is available only at a fixed point in time $t = T$. Even if it is set up as "Lock-Type," the term "European" typically

means that the payoff can happen only at a fixed point in time $t = T$. This is in contrast to American Options. The most basic forms of American Options are American Call and Put Options. American Call and Put Options are essentially extensions of the corresponding European Call and Put Options by allowing the buyer (owner) of the American Option to exercise the option to buy (in the case of Call) or sell (in the case of Put) at any time $t \leq T$. The allowance of exercise at any time at or before the expiry time T can often be a tricky financial decision for the option owner. At each point in time when the American Option is *in-the-money* (i.e., positive payoff upon exercise), the option owner might be tempted to exercise and collect the payoff but might as well be thinking that if she waits, the option might become more *in-the-money* (i.e., prospect of a bigger payoff if she waits for a while). Hence, it's clear that an American Option is always of the "Option-Type" (and not "Lock-Type") since the timing of the decision (option) to exercise is very important in the case of an American Option. This also means that the problem of pricing an American Option (the fair price the buyer would need to pay to own an American Option) is much harder than the problem of pricing an European Option.

So what purpose do derivatives serve? There are actually many motivations for different market participants, but we'll just list two key motivations. The first reason is to protect against adverse market movements that might damage the value of one's portfolio (this is known as *hedging*). As an example, buying a put option can reduce or eliminate the risk associated with ownership of the underlying. The second reason is operational or financial convenience in trading to express a speculative view of market movements. For instance, if one thinks a stock will increase in value by 50% over the next two years, instead of paying say $100,000 to buy the stock (hoping to make $50,000 after two years), one can simply buy a call option on $100,000 of the stock (paying the option price of say $5,000). If the stock price indeed appreciates by 50% after 2 years, one makes $50,000 - $5,000 = $45,000. Although one made $5000 less than the alternative of simply buying the stock, the fact that one needs to pay $5000 (versus $50,000) to enter into the trade means the potential *return on investment* is much higher.

Next, we embark on the journey of learning how to value derivatives, i.e., how to figure out the fair price that one would be willing to buy or sell the derivative for at any point in time. As mentioned earlier, the general theory of derivatives pricing is quite rich and elaborate (based on continuous-time stochastic processes), and we don't cover it in this book. Instead, we provide intuition for the core concepts underlying derivatives pricing theory in the context of a simple, special case - that of discrete-time with a single-period. We formalize this simple setting in the next section.

1.2 Notation for the Single-Period Simple Setting

Our simple setting involves discrete time with a single-period from $t = 0$ to $t = 1$. Time $t = 0$ has a single state which we shall refer to as the "Spot" state. Time $t = 1$ has n random outcomes formalized by the sample space $\Omega = \{\omega_1, \ldots, \omega_n\}$. The probability distribution of this finite sample space is given by the probability mass function

$$\mu : \Omega \to [0, 1]$$

such that

$$\sum_{i=1}^{n} \mu(\omega_i) = 1$$

This simple single-period setting involves $m+1$ fundamental assets $A_0, A_1, \ldots, A_m$ where A_0 is a riskless asset (i.e., it's price will evolve deterministically from $t=0$ to $t=1$) and $A_1, \ldots, A_m$ are risky assets. We denote the Spot Price (at $t=0$) of A_j as $S_j^{(0)}$ for all $j = 0, 1, \ldots, m$. We denote the Price of A_j in ω_i as $S_j^{(i)}$ for all $j = 0, \ldots, m, i = 1, \ldots, n$. Assume that all asset prices are real numbers, i.e., in $\mathbb{R}$ (negative prices are typically unrealistic, but we still assume it for simplicity of exposition). For convenience, we normalize the Spot Price (at $t=0$) of the riskless asset A_O to be 1. Therefore,

$$S_0^{(0)} = 1 \text{ and } S_0^{(i)} = 1 + r \text{ for all } i = 1, \ldots, n$$

where r represents the constant riskless rate of growth. We should interpret this riskless rate of growth as the "time value of money" and $\frac{1}{1+r}$ as the riskless discount factor corresponding to the "time value of money."

1.3 Portfolios, Arbitrage and Risk-Neutral Probability Measure

We define a portfolio as a vector $\theta = (\theta_0, \theta_1, \ldots, \theta_m) \in \mathbb{R}^{m+1}$, representing the number of units held in the assets $A_j, j = 0, 1, \ldots, m$. The Spot Value (at $t=0$) of portfolio θ, denoted by $V_\theta^{(0)}$, is:

$$V_\theta^{(0)} = \sum_{j=0}^{m} \theta_j \cdot S_j^{(0)} \tag{1.1}$$

The Value of portfolio θ in random outcome ω_i (at $t=1$), denoted by $V_\theta^{(i)}$, is:

$$V_\theta^{(i)} = \sum_{j=0}^{m} \theta_j \cdot S_j^{(i)} \text{ for all } i = 1, \ldots, n \tag{1.2}$$

Next, we cover an extremely important concept in Mathematical Economics/Finance - the concept of *Arbitrage*. An Arbitrage Portfolio θ is one that "makes money from nothing." Formally, an arbitrage portfolio is a portfolio θ such that:

- $V_\theta^{(0)} \leq 0$
- $V_\theta^{(i)} \geq 0$ for all $i = 1, \ldots, n$
- There exists an $i \in \{1, \ldots, n\}$ such that $\mu(\omega_i) > 0$ and $V_\theta^{(i)} > 0$

Thus, with an Arbitrage Portfolio, we never end up (at $t=0$) with less value than what we start with (at $t=1$) and we end up with expected value strictly greater than what we start with. This is the formalism of the notion of *arbitrage*, i.e., "making money from nothing." Arbitrage allows market participants to make infinite returns. In an efficient market, arbitrage would disappear as soon as it appears since market participants would immediately exploit it and through the process of exploiting the arbitrage, immediately eliminate the arbitrage. Hence, Finance Theory typically assumes "arbitrage-free" markets (i.e., financial markets with no arbitrage opportunities).

Next, we describe another very important concept in Mathematical Economics/Finance - the concept of a *Risk-Neutral Probability Measure*. Consider a Probability Distribution $\pi : \Omega \to [0, 1]$ such that

$$\pi(\omega_i) = 0 \text{ if and only if } \mu(\omega_i) = 0 \text{ for all } i = 1, \ldots, n$$

Then, π is said to be a Risk-Neutral Probability Measure if:

$$S_j^{(0)} = \frac{1}{1+r} \cdot \sum_{i=1}^{n} \pi(\omega_i) \cdot S_j^{(i)} \text{ for all } j = 0, 1, \ldots, m \tag{1.3}$$

So for each of the $m+1$ assets, the asset spot price (at $t = 0$) is the riskless rate-discounted expectation (under π) of the asset price at $t = 1$. The term "risk-neutral" here is the same as the term "risk-neutral" we used in Chapter **??**, meaning it's a situation where one doesn't need to be compensated for taking risk (the situation of a linear utility function). However, we are not saying that the market is risk-neutral - if that were the case, the market probability measure μ would be a risk-neutral probability measure. We are simply defining π as a *hypothetical construct* under which each asset's spot price is equal to the riskless rate-discounted expectation (under π) of the asset's price at $t = 1$. This means that under the hypothetical π, there's no return in excess of r for taking on the risk of probabilistic outcomes at $t = 1$ (note: outcome probabilities are governed by the hypothetical π). Hence, we refer to π as a risk-neutral probability measure. The purpose of this hypothetical construct π is that it helps in the development of Derivatives Pricing and Hedging Theory, as we shall soon see. The actual probabilities of outcomes in Ω are governed by μ, and not π.

Before we cover the two fundamental theorems of asset pricing, we need to cover an important lemma that we will utilize in the proofs of the two fundamental theorems of asset pricing.

Lemma 1.3.1. *For any portfolio $\theta = (\theta_0, \theta_1, \ldots, \theta_m) \in \mathbb{R}^{m+1}$ and any risk-neutral probability measure $\pi : \Omega \to [0, 1]$,*

$$V_\theta^{(0)} = \frac{1}{1+r} \cdot \sum_{i=1}^{n} \pi(\omega_i) \cdot V_\theta^{(i)}$$

Proof. Using Equations (1.1), (1.3) and (1.2), the proof is straightforward:

$$\begin{aligned} V_\theta^{(0)} &= \sum_{j=0}^{m} \theta_j \cdot S_j^{(0)} = \sum_{j=0}^{m} \theta_j \cdot \frac{1}{1+r} \cdot \sum_{i=1}^{n} \pi(\omega_i) \cdot S_j^{(i)} \\ &= \frac{1}{1+r} \cdot \sum_{i=1}^{n} \pi(\omega_i) \cdot \sum_{j=0}^{m} \theta_j \cdot S_j^{(i)} = \frac{1}{1+r} \cdot \sum_{i=1}^{n} \pi(\omega_i) \cdot V_\theta^{(i)} \end{aligned}$$

□

Now we are ready to cover the two fundamental theorems of asset pricing (sometimes, also refered to as the fundamental theorems of arbitrage and the fundamental theorems of finance!). We start with the first fundamental theorem of asset pricing, which associates absence of arbitrage with existence of a risk-neutral probability measure.

1.4 First Fundamental Theorem of Asset Pricing (1st FTAP)

Theorem 1.4.1 (First Fundamental Theorem of Asset Pricing (1st FTAP))**.** *Our simple setting of discrete time with single-period will not admit arbitrage portfolios if and only if there exists a Risk-Neutral Probability Measure.*

Proof. First we prove the easy implication - if there exists a Risk-Neutral Probability Measure π, then we cannot have any arbitrage portfolios. Let's review what it takes to have an arbitrage portfolio $\theta = (\theta_0, \theta_1, \ldots, \theta_m)$. The following are two of the three conditions to be satisfied to qualify as an arbitrage portfolio θ (according to the definition of arbitrage portfolio we gave above):

- $V_\theta^{(i)} \geq 0$ for all $i = 1, \ldots, n$
- There exists an $i \in \{1, \ldots, n\}$ such that $\mu(\omega_i) > 0$ ($\Rightarrow \pi(\omega_i) > 0$) and $V_\theta^{(i)} > 0$

But if these two conditions are satisfied, the third condition $V_\theta^{(0)} \leq 0$ cannot be satisfied because from Lemma (1.3.1), we know that:

$$V_\theta^{(0)} = \frac{1}{1+r} \cdot \sum_{i=1}^{n} \pi(\omega_i) \cdot V_\theta^{(i)}$$

which is strictly greater than 0, given the two conditions stated above. Hence, all three conditions cannot be simultaneously satisfied which eliminates the possibility of arbitrage for any portfolio θ.

Next, we prove the reverse (harder to prove) implication - if a risk-neutral probability measure doesn't exist, then there exists an arbitrage portfolio θ. We define $\mathbb{V} \subset \mathbb{R}^m$ as the set of vectors $v = (v_1, \ldots, v_m)$ such that

$$v_j = \frac{1}{1+r} \cdot \sum_{i=1}^{n} \mu(\omega_i) \cdot S_j^{(i)} \text{ for all } j = 1, \ldots, m$$

with $\mathbb{V}$ defined as spanning over all possible probability distributions $\mu : \Omega \to [0, 1]$. $\mathbb{V}$ is a bounded, closed, convex polytope in $\mathbb{R}^m$. By the definition of a risk-neutral probability measure, we can say that if a risk-neutral probability measure doesn't exist, the vector $(S_1^{(0)}, \ldots, S_m^{(0)}) \notin \mathbb{V}$. The Hyperplane Separation Theorem implies that there exists a non-zero vector $(\theta_1, \ldots, \theta_m)$ such that for any $v = (v_1, \ldots, v_m) \in \mathbb{V}$,

$$\sum_{j=1}^{m} \theta_j \cdot v_j > \sum_{j=1}^{m} \theta_j \cdot S_j^{(0)}$$

In particular, consider vectors v corresponding to the corners of $\mathbb{V}$, those for which the full probability mass is on a particular $\omega_i \in \Omega$, i.e.,

$$\sum_{j=1}^{m} \theta_j \cdot (\frac{1}{1+r} \cdot S_j^{(i)}) > \sum_{j=1}^{m} \theta_j \cdot S_j^{(0)} \text{ for all } i = 1, \ldots, n$$

Since this is a strict inequality, we will be able to choose a $\theta_0 \in \mathbb{R}$ such that:

$$\sum_{j=1}^{m} \theta_j \cdot (\frac{1}{1+r} \cdot S_j^{(i)}) > -\theta_0 > \sum_{j=1}^{m} \theta_j \cdot S_j^{(0)} \text{ for all } i = 1, \ldots, n$$

Therefore,

$$\frac{1}{1+r} \cdot \sum_{j=0}^{m} \theta_j \cdot S_j^{(i)} > 0 > \sum_{j=0}^{m} \theta_j \cdot S_j^{(0)} \text{ for all } i = 1, \ldots, n$$

This can be rewritten in terms of the Values of portfolio $\theta = (\theta_0, \theta_1, \ldots, \theta)$ at $t = 0$ and $t = 1$, as follows:

$$\frac{1}{1+r} \cdot V_\theta^{(i)} > 0 > V_\theta^{(0)} \text{ for all } i = 1, \ldots, n$$

Thus, we can see that all three conditions in the definition of arbitrage portfolio are satisfied and hence, $\theta = (\theta_0, \theta_1, \ldots, \theta_m)$ is an arbitrage portfolio.

□

Now we are ready to move on to the second fundamental theorem of asset pricing, which associates replication of derivatives with a unique risk-neutral probability measure.

1.5 Second Fundamental Theorem of Asset Pricing (2nd FTAP)

Before we state and prove the 2nd FTAP, we need some definitions.

Definition 1.5.1. A Derivative D (in our simple setting of discrete-time with a single-period) is specified as a vector payoff at time $t = 1$, denoted as:

$$(V_D^{(1)}, V_D^{(2)}, \ldots, V_D^{(n)})$$

where $V_D^{(i)}$ is the payoff of the derivative in random outcome ω_i for all $i = 1, \ldots, n$

Definition 1.5.2. A Portfolio $\theta = (\theta_0, \theta_1, \ldots, \theta_m) \in \mathbb{R}^{m+1}$ is a *Replicating Portfolio* for derivative D if:

$$V_D^{(i)} = V_\theta^{(i)} = \sum_{j=0}^{m} \theta_j \cdot S_j^{(i)} \text{ for all } i = 1, \ldots, n \tag{1.4}$$

The negatives of the components $(\theta_0, \theta_1, \ldots, \theta_m)$ are known as the *hedges* for D since they can be used to offset the risk in the payoff of D at $t = 1$.

Definition 1.5.3. An arbitrage-free market (i.e., a market devoid of arbitrage) is said to be *Complete* if every derivative in the market has a replicating portfolio.

Theorem 1.5.1 (Second Fundamental Theorem of Asset Pricing (2nd FTAP))**.** *A market (in our simple setting of discrete-time with a single-period) is Complete if and only if there is a unique Risk-Neutral Probability Measure.*

Proof. We will first prove that in an arbitrage-free market, if every derivative has a replicating portfolio (i.e., the market is complete), then there is a unique risk-neutral probability measure. We define n special derivatives (known as *Arrow-Debreu securities*), one for each random outcome in Ω at $t = 1$. We define the time $t = 1$ payoff of *Arrow-Debreu security* D_k (for each of $k = 1, \ldots, n$) as follows:

$$V_{D_k}^{(i)} = \mathbb{I}_{i=k} \text{ for all } i = 1, \ldots, n$$

where $\mathbb{I}$ represents the indicator function. This means the payoff of derivative D_k is 1 for random outcome ω_k and 0 for all other random outcomes.

Since each derivative has a replicating portfolio, denote $\theta^{(k)} = (\theta_0^{(k)}, \theta_1^{(k)}, \ldots, \theta_m^{(k)})$ as the replicating portfolio for D_k for each $k = 1, \ldots, m$. Therefore, for each $k = 1, \ldots, m$:

$$V_{\theta^{(k)}}^{(i)} = \sum_{j=0}^{m} \theta_j^{(k)} \cdot S_j^{(i)} = V_{D_k}^{(i)} = \mathbb{I}_{i=k} \text{ for all } i = 1, \ldots, n$$

Using Lemma (1.3.1), we can write the following equation for any risk-neutral probability measure π, for each $k = 1, \ldots, m$:

$$\sum_{j=0}^{m} \theta_j^{(k)} \cdot S_j^{(0)} = V_{\theta^{(k)}}^{(0)} = \frac{1}{1+r} \cdot \sum_{i=1}^{n} \pi(\omega_i) \cdot V_{\theta^{(k)}}^{(i)} = \frac{1}{1+r} \cdot \sum_{i=1}^{n} \pi(\omega_i) \cdot \mathbb{I}_{i=k} = \frac{1}{1+r} \cdot \pi(\omega_k)$$

We note that the above equation is satisfied for a unique $\pi : \Omega \to [0, 1]$, defined as:

$$\pi(\omega_k) = (1+r) \cdot \sum_{j=0}^{m} \theta_j^{(k)} \cdot S_j^{(0)} \text{ for all } k = 1, \ldots, n$$

which implies that we have a unique risk-neutral probability measure.

Next, we prove the other direction of the 2nd FTAP. We need to prove that if there exists a risk-neutral probability measure π and if there exists a derivative D with no replicating portfolio, then we can construct a risk-neutral probability measure different than π.

Consider the following vectors in the vector space $\mathbb{R}^n$

$$v = (V_D^{(1)}, \ldots, V_D^{(n)}) \text{ and } v_j = (S_j^{(1)}, \ldots, S_j^{(n)}) \text{ for all } j = 0, 1, \ldots, m$$

Since D does not have a replicating portfolio, v is not in the span of $\{v_0, v_1, \ldots, v_m\}$, which means $\{v_0, v_1, \ldots, v_m\}$ do not span $\mathbb{R}^n$. Hence, there exists a non-zero vector $u = (u_1, \ldots, u_n) \in \mathbb{R}^n$ orthogonal to each of $v_0, v_1, \ldots, v_m$, i.e.,

$$\sum_{i=1}^{n} u_i \cdot S_j^{(i)} = 0 \text{ for all } j = 0, 1, \ldots, n \tag{1.5}$$

Note that $S_0^{(i)} = 1 + r$ for all $i = 1, \ldots, n$ and so,

$$\sum_{i=1}^{n} u_i = 0 \tag{1.6}$$

Define $\pi' : \Omega \to \mathbb{R}$ as follows (for some $\epsilon \in \mathbb{R}^+$):

$$\pi'(\omega_i) = \pi(\omega_i) + \epsilon \cdot u_i \text{ for all } i = 1, \ldots, n \tag{1.7}$$

To establish π' as a risk-neutral probability measure different than π, note:

- Since $\sum_{i=1}^{n} \pi(\omega_i) = 1$ and since $\sum_{i=1}^{n} u_i = 0$, $\sum_{i=1}^{n} \pi'(\omega_i) = 1$
- Construct $\pi'(\omega_i) > 0$ for each i where $\pi(\omega_i) > 0$ by making $\epsilon > 0$ sufficiently small, and set $\pi'(\omega_i) = 0$ for each i where $\pi(\omega_i) = 0$
- From Equations (1.7), (1.3) and (1.5), we have for each $j = 0, 1, \ldots, m$:

$$\frac{1}{1+r} \cdot \sum_{i=1}^{n} \pi'(\omega_i) \cdot S_j^{(i)} = \frac{1}{1+r} \cdot \sum_{i=1}^{n} \pi(\omega_i) \cdot S_j^{(i)} + \frac{\epsilon}{1+r} \cdot \sum_{i=1}^{n} u_i \cdot S_j^{(i)} = S_j^{(0)}$$

□

Together, the two FTAPs classify markets into:

- Market with arbitrage $\Leftrightarrow$ No risk-neutral probability measure
- Complete (arbitrage-free) market $\Leftrightarrow$ Unique risk-neutral probability measure
- Incomplete (arbitrage-free) market $\Leftrightarrow$ Multiple risk-neutral probability measures

The next topic is derivatives pricing that is based on the concepts of *replication of derivatives* and *risk-neutral probability measures*, and so is tied to the concepts of *arbitrage* and *completeness*.

1.6 Derivatives Pricing in Single-Period Setting

In this section, we cover the theory of derivatives pricing for our simple setting of discrete-time with a single-period. To develop the theory of how to price a derivative, first we need to define the notion of a *Position*.

Definition 1.6.1. A *Position* involving a derivative D is the combination of holding some units in D and some units in the fundamental assets $A_0, A_1, \ldots, A_m$, which can be formally represented as a vector $\gamma_D = (\alpha, \theta_0, \theta_1, \ldots, \theta_m) \in \mathbb{R}^{m+2}$ where α denotes the units held in derivative D and α_j denotes the units held in A_j for all $j = 0, 1 \ldots, m$.

Therefore, a *Position* is an extension of the Portfolio concept that includes a derivative. Hence, we can naturally extend the definition of *Portfolio Value* to *Position Value* and we can also extend the definition of *Arbitrage Portfolio* to *Arbitrage Position*.

We need to consider derivatives pricing in three market situations:

- When the market is complete
- When the market is incomplete
- When the market has arbitrage

1.6.1 Derivatives Pricing when Market is Complete

Theorem 1.6.1. *For our simple setting of discrete-time with a single-period, if the market is complete, then any derivative D with replicating portfolio $\theta = (\theta_0, \theta_1, \ldots, \theta_m)$ has price at time $t = 0$ (denoted as value $V_D^{(0)}$):*

$$V_D^{(0)} = V_\theta^{(0)} = \sum_{j=0}^{n} \theta_j \cdot S_j^{(i)} \tag{1.8}$$

Furthermore, if the unique risk-neutral probability measure is $\pi : \Omega \to [0, 1]$, then:

$$V_D^{(0)} = \frac{1}{1+r} \cdot \sum_{i=1}^{n} \pi(\omega_i) \cdot V_D^{(i)} \tag{1.9}$$

Proof. It seems quite reasonable that since θ is the replicating portfolio for D, the value of the replicating portfolio at time $t = 0$ (equal to $V_\theta^{(0)} = \sum_{j=0}^{n} \theta_j \cdot S_j^{(i)}$) should be the price (at $t = 0$) of derivative D. However, we will formalize the proof by first arguing that any candidate derivative price for D other than $V_\theta^{(0)}$ leads to arbitrage, thus dismissing those other candidate derivative prices, and then argue that with $V_\theta^{(0)}$ as the price of derivative D, we eliminate the possibility of an arbitrage position involving D.

Consider candidate derivative prices $V_\theta^{(0)} - x$ for any positive real number x. Position $(1, -\theta_0 + x, -\theta_1, \ldots, -\theta_m)$ has value $x \cdot (1+r) > 0$ in each of the random outcomes at $t = 1$. But this position has spot $(t = 0)$ value of 0, which means this is an Arbitrage Position, rendering these candidate derivative prices invalid. Next consider candidate derivative prices $V_\theta^{(0)} + x$ for any positive real number x. Position $(-1, \theta_0 + x, \theta_1, \ldots, \theta_m)$ has value $x \cdot (1 + r) > 0$ in each of the random outcomes at $t = 1$. But this position has spot $(t = 0)$ value of 0, which means this is an Arbitrage Position, rendering these candidate derivative prices invalid as well. So every candidate derivative price other than $V_\theta^{(0)}$ is invalid. Now our goal is to *establish* $V_\theta^{(0)}$ as the derivative price of D by showing that we eliminate the possibility of an arbitrage position in the market involving D if $V_\theta^{(0)}$ is indeed the derivative price.

Firstly, note that $V_\theta^{(0)}$ can be expressed as the riskless rate-discounted expectation (under π) of the payoff of D at $t = 1$, i.e.,

$$V_\theta^{(0)} = \sum_{j=0}^{m} \theta_j \cdot S_j^{(0)} = \sum_{j=0}^{m} \theta_j \cdot \frac{1}{1+r} \cdot \sum_{i=1}^{n} \pi(\omega_i) \cdot S_j^{(i)} = \frac{1}{1+r} \cdot \sum_{i=1}^{n} \pi(\omega_i) \cdot \sum_{j=0}^{m} \theta_j \cdot S_j^{(i)}$$
$$= \frac{1}{1+r} \cdot \sum_{i=1}^{n} \pi(\omega_i) \cdot V_D^{(i)} \quad (1.10)$$

Now consider an *arbitrary portfolio* $\beta = (\beta_0, \beta_1, \ldots, \beta_m)$. Define a position $\gamma_D = (\alpha, \beta_0, \beta_1, \ldots, \beta_m)$. Assuming the derivative price $V_D^{(0)}$ is equal to $V_\theta^{(0)}$, the Spot Value (at $t = 0$) of position γ_D, denoted $V_{\gamma_D}^{(0)}$, is:

$$V_{\gamma_D}^{(0)} = \alpha \cdot V_\theta^{(0)} + \sum_{j=0}^{m} \beta_j \cdot S_j^{(0)} \quad (1.11)$$

Value of position γ_D in random outcome ω_i (at $t = 1$), denoted $V_{\gamma_D}^{(i)}$, is:

$$V_{\gamma_D}^{(i)} = \alpha \cdot V_D^{(i)} + \sum_{j=0}^{m} \beta_j \cdot S_j^{(i)} \text{ for all } i = 1, \ldots, n \quad (1.12)$$

Combining the linearity in Equations (1.3), (1.10), (1.11) and (1.12), we get:

$$V_{\gamma_D}^{(0)} = \frac{1}{1+r} \cdot \sum_{i=1}^{n} \pi(\omega_i) \cdot V_{\gamma_D}^{(i)} \quad (1.13)$$

So the position spot value (at $t = 0$) is the riskless rate-discounted expectation (under π) of the position value at $t = 1$. For any γ_D (containing any arbitrary portfolio β), with derivative price $V_D^{(0)}$ equal to $V_\theta^{(0)}$, if the following two conditions are satisfied:

- $V_{\gamma_D}^{(i)} \geq 0$ for all $i = 1, \ldots, n$
- There exists an $i \in \{1, \ldots, n\}$ such that $\mu(\omega_i) > 0$ ($\Rightarrow \pi(\omega_i) > 0$) and $V_{\gamma_D}^{(i)} > 0$

then:

$$V_{\gamma_D}^{(0)} = \frac{1}{1+r} \cdot \sum_{i=1}^{n} \pi(\omega_i) \cdot V_{\gamma_D}^{(i)} > 0$$

This eliminates any arbitrage possibility if D is priced at $V_\theta^{(0)}$.

To summarize, we have eliminated all candidate derivative prices other than $V_\theta^{(0)}$, and we have established the price $V_\theta^{(0)}$ as the correct price of D in the sense that we eliminate the possibility of an arbitrage position involving D if the price of D is $V_\theta^{(0)}$.

Finally, we note that with the derivative price $V_D^{(0)} = V_\theta^{(0)}$, from Equation (1.10), we have:

$$V_D^{(0)} = \frac{1}{1+r} \cdot \sum_{i=1}^{n} \pi(\omega_i) \cdot V_D^{(i)}$$

□

Now let us consider the special case of 1 risky asset ($m = 1$) and 2 random outcomes ($n = 2$), which we will show is a Complete Market. To lighten notation, we drop the subscript 1 on the risky asset price. Without loss of generality, we assume $S^{(1)} < S^{(2)}$. No-arbitrage requires:

$$S^{(1)} \leq (1+r) \cdot S^{(0)} \leq S^{(2)}$$

Assuming absence of arbitrage and invoking 1st FTAP, there exists a risk-neutral probability measure π such that:

$$S^{(0)} = \frac{1}{1+r} \cdot (\pi(\omega_1) \cdot S^{(1)} + \pi(\omega_2) \cdot S^{(2)})$$

$$\pi(\omega_1) + \pi(\omega_2) = 1$$

With 2 linear equations and 2 variables, this has a straightforward solution, as follows:

$$\pi(\omega_1) = \frac{S^{(2)} - (1+r) \cdot S^{(0)}}{S^{(2)} - S^{(1)}}$$

$$\pi(\omega_2) = \frac{(1+r) \cdot S^{(0)} - S^{(1)}}{S^{(2)} - S^{(1)}}$$

Conditions $S^{(1)} < S^{(2)}$ and $S^{(1)} \leq (1+r) \cdot S^{(0)} \leq S^{(2)}$ ensure that $0 \leq \pi(\omega_1), \pi(\omega_2) \leq 1$. Also note that this is a unique solution for $\pi(\omega_1), \pi(\omega_2)$, which means that the risk-neutral probability measure is unique, implying that this is a complete market.

We can use these probabilities to price a derivative D as:

$$V_D^{(0)} = \frac{1}{1+r} \cdot (\pi(\omega_1) \cdot V_D^{(1)} + \pi(\omega_2) \cdot V_D^{(2)})$$

Now let us try to form a replicating portfolio (θ_0, θ_1) for D

$$V_D^{(1)} = \theta_0 \cdot (1+r) + \theta_1 \cdot S^{(1)}$$

$$V_D^{(2)} = \theta_0 \cdot (1+r) + \theta_1 \cdot S^{(2)}$$

Solving this yields Replicating Portfolio (θ_0, θ_1) as follows:

$$\theta_0 = \frac{1}{1+r} \cdot \frac{V_D^{(1)} \cdot S^{(2)} - V_D^{(2)} \cdot S^{(1)}}{S^{(2)} - S^{(1)}} \text{ and } \theta_1 = \frac{V_D^{(2)} - V_D^{(1)}}{S^{(2)} - S^{(1)}} \tag{1.14}$$

Note that the derivative price can also be expressed as:

$$V_D^{(0)} = \theta_0 + \theta_1 \cdot S^{(0)}$$

1.6.2 Derivatives Pricing when Market is Incomplete

Theorem (1.6.1) assumed a complete market, but what about an incomplete market? Recall that an incomplete market means some derivatives can't be replicated. Absence of a replicating portfolio for a derivative precludes usual no-arbitrage arguments. The 2nd FTAP says that in an incomplete market, there are multiple risk-neutral probability measures which means there are multiple derivative prices (each consistent with no-arbitrage).

To develop intuition for derivatives pricing when the market is incomplete, let us consider the special case of 1 risky asset ($m = 1$) and 3 random outcomes ($n = 3$), which we will show is an Incomplete Market. To lighten notation, we drop the subscript 1 on the risky asset price. Without loss of generality, we assume $S^{(1)} < S^{(2)} < S^{(3)}$. No-arbitrage requires:

$$S^{(1)} \leq S^{(0)} \cdot (1 + r) \leq S^{(3)}$$

Assuming absence of arbitrage and invoking the 1st FTAP, there exists a risk-neutral probability measure π such that:

$$S^{(0)} = \frac{1}{1+r} \cdot (\pi(\omega_1) \cdot S^{(1)} + \pi(\omega_2) \cdot S^{(2)} + \pi(\omega_3) \cdot S^{(3)})$$

$$\pi(\omega_1) + \pi(\omega_2) + \pi(\omega_3) = 1$$

So we have 2 equations and 3 variables, which implies there are multiple solutions for π. Each of these solutions for π provides a valid price for a derivative D.

$$V_D^{(0)} = \frac{1}{1+r} \cdot (\pi(\omega_1) \cdot V_D^{(1)} + \pi(\omega_2) \cdot V_D^{(2)} + \pi(\omega_3) \cdot V_D^{(3)})$$

Now let us try to form a replicating portfolio (θ_0, θ_1) for D

$$V_D^{(1)} = \theta_0 \cdot (1 + r) + \theta_1 \cdot S^{(1)}$$

$$V_D^{(2)} = \theta_0 \cdot (1 + r) + \theta_1 \cdot S^{(2)}$$

$$V_D^{(3)} = \theta_0 \cdot (1 + r) + \theta_1 \cdot S^{(3)}$$

3 equations & 2 variables implies there is no replicating portfolio for *some* D. This means this is an Incomplete Market.

So with multiple risk-neutral probability measures (and consequent, multiple derivative prices), how do we go about determining how much to buy/sell derivatives for? One approach to handle derivative pricing in an incomplete market is the technique called *Superhedging*, which provides upper and lower bounds for the derivative price. The idea of Superhedging is to create a portfolio of fundamental assets whose Value *dominates* the derivative payoff in *all* random outcomes at $t = 1$. Superhedging Price is the smallest possible Portfolio Spot ($t = 0$) Value among all such Derivative-Payoff-Dominating portfolios. Without getting into too many details of the Superhedging technique (out of scope for this book), we shall simply sketch the outline of this technique for our simple setting.

We note that for our simple setting of discrete-time with a single-period, this is a constrained linear optimization problem:

$$\min_{\theta} \sum_{j=0}^{m} \theta_j \cdot S_j^{(0)} \text{ such that } \sum_{j=0}^{m} \theta_j \cdot S_j^{(i)} \geq V_D^{(i)} \text{ for all } i = 1, \ldots, n \tag{1.15}$$

Let $\theta^* = (\theta_0^*, \theta_1^*, \ldots, \theta_m^*)$ be the solution to Equation (1.15). Let SP be the Superhedging Price $\sum_{j=0}^{m} \theta_j^* \cdot S_j^{(0)}$.

After establishing feasibility, we define the Lagrangian $J(\theta, \lambda)$ as follows:

$$J(\theta, \lambda) = \sum_{j=0}^{m} \theta_j \cdot S_j^{(0)} + \sum_{i=1}^{n} \lambda_i \cdot (V_D^{(i)} - \sum_{j=0}^{m} \theta_j \cdot S_j^{(i)})$$

So there exists $\lambda = (\lambda_1, \ldots, \lambda_n)$ that satisfy the following KKT conditions:

$$\lambda_i \geq 0 \text{ for all } i = 1, \ldots, n$$

$$\lambda_i \cdot (V_D^{(i)} - \sum_{j=0}^{m} \theta_j^* \cdot S_j^{(i)}) = 0 \text{ for all } i = 1, \ldots, n \text{ (Complementary Slackness)}$$

$$\nabla_\theta J(\theta^*, \lambda) = 0 \Rightarrow S_j^{(0)} = \sum_{i=1}^{n} \lambda_i \cdot S_j^{(i)} \text{ for all } j = 0, 1, \ldots, m$$

This implies $\lambda_i = \frac{\pi(\omega_i)}{1+r}$ for all $i = 1, \ldots, n$ for a risk-neutral probability measure $\pi : \Omega \to [0, 1]$ (λ can be thought of as "discounted probabilities").

Define Lagrangian Dual

$$L(\lambda) = \inf_\theta J(\theta, \lambda)$$

Then, Superhedging Price

$$SP = \sum_{j=0}^{m} \theta_j^* \cdot S_j^{(0)} = \sup_\lambda L(\lambda) = \sup_\lambda \inf_\theta J(\theta, \lambda)$$

Complementary Slackness and some linear algebra over the space of risk-neutral probability measures $\pi : \Omega \to [0, 1]$ enables us to argue that:

$$SP = \sup_\pi \sum_{i=1}^{n} \frac{\pi(\omega_i)}{1+r} \cdot V_D^{(i)}$$

This means the Superhedging Price is the least upper-bound of the riskless rate-discounted expectation of derivative payoff across each of the risk-neutral probability measures in the incomplete market, which is quite an intuitive thing to do amidst multiple risk-neutral probability measures.

Likewise, the *Subhedging* price SB is defined as:

$$\max_\theta \sum_{j=0}^{m} \theta_j \cdot S_j^{(0)} \text{ such that } \sum_{j=0}^{m} \theta_j \cdot S_j^{(i)} \leq V_D^{(i)} \text{ for all } i = 1, \ldots, n$$

Likewise arguments enable us to establish:

$$SB = \inf_\pi \sum_{i=1}^{n} \frac{\pi(\omega_i)}{1+r} \cdot V_D^{(i)}$$

This means the Subhedging Price is the highest lower-bound of the riskless rate-discounted expectation of derivative payoff across each of the risk-neutral probability measures in the incomplete market, which is quite an intuitive thing to do amidst multiple risk-neutral probability measures.

So this technique provides an lower bound (SB) and an upper bound (SP) for the derivative price, meaning:

- A price outside these bounds leads to an arbitrage
- Valid prices must be established within these bounds

But often these bounds are not tight and so, not useful in practice.

The alternative approach is to identify hedges that maximize Expected Utility of the combination of the derivative along with it's hedges, for an appropriately chosen market/trader Utility Function (as covered in Chapter **??**). The Utility function is a specification of reward-versus-risk preference that effectively chooses the risk-neutral probability measure (and hence, Price).

Consider a concave Utility function $U : \mathbb{R} \to \mathbb{R}$ applied to the Value in each random outcome $\omega_i, i = 1, \ldots n$, at $t = 1$ (eg: $U(x) = \frac{1-e^{-ax}}{a}$ where $a \in \mathbb{R}$ is the degree of risk-aversion). Let the real-world probabilities be given by $\mu : \Omega \to [0, 1]$. Denote $V_D = (V_D^{(1)}, \ldots, V_D^{(n)})$ as the payoff of Derivative D at $t = 1$. Let us say that you buy the derivative D at $t = 0$ and will receive the random outcome-contingent payoff V_D at $t = 1$. Let x be the candidate derivative price for D, which means you will pay a cash quantity of x at $t = 0$ for the privilege of receiving the payoff V_D at $t = 1$. We refer to the candidate hedge as Portfolio $\theta = (\theta_0, \theta_1, \ldots, \theta_m)$, representing the units held in the fundamental assets.

Note that at $t = 0$, the cash quantity x you'd be paying to buy the derivative and the cash quantity you'd be paying to buy the Portfolio θ should sum to 0 (note: either of these cash quantities can be positive or negative, but they need to sum to 0 since "money can't just appear or disappear"). Formally,

$$x + \sum_{j=0}^{m} \theta_j \cdot S_j^{(0)} = 0 \tag{1.16}$$

Our goal is to solve for the appropriate values of x and θ based on an *Expected Utility* consideration (that we are about to explain). Consider the Utility of the position consisting of derivative D together with portfolio θ in random outcome ω_i at $t = 1$:

$$U(V_D^{(i)} + \sum_{j=0}^{m} \theta_j \cdot S_j^{(i)})$$

So, the Expected Utility of this position at $t = 1$ is given by:

$$\sum_{i=1}^{n} \mu(\omega_i) \cdot U(V_D^{(i)} + \sum_{j=0}^{m} \theta_j \cdot S_j^{(i)}) \tag{1.17}$$

Noting that $S_0^{(0)} = 1, S_0^{(i)} = 1 + r$ for all $i = 1, \ldots, n$, we can substitute for the value of $\theta_0 = -(x + \sum_{j=1}^{m} \theta_j \cdot S_j^{(0)})$ (obtained from Equation (1.16)) in the above Expected Utility expression (1.17), so as to rewrite this Expected Utility expression in terms of just $(\theta_1, \ldots, \theta_m)$ (call it $\theta_{1:m}$) as:

$$g(V_D, x, \theta_{1:m}) = \sum_{i=1}^{n} \mu(\omega_i) \cdot U(V_D^{(i)} - (1 + r) \cdot x + \sum_{j=1}^{m} \theta_j \cdot (S_j^{(i)} - (1 + r) \cdot S_j^{(0)}))$$

We define the *Price* of D as the "breakeven value" x^* such that:

$$\max_{\theta_{1:m}} g(V_D, x^*, \theta_{1:m}) = \max_{\theta_{1:m}} g(0, 0, \theta_{1:m})$$

The core principle here (known as *Expected-Utility-Indifference Pricing*) is that introducing a $t = 1$ payoff of V_D together with a derivative price payment of x^* at $t = 0$ keeps the Maximum Expected Utility unchanged.

The $(\theta_1^*, \ldots, \theta_m^*)$ that achieve $\max_{\theta_{1:m}} g(V_D, x^*, \theta_{1:m})$ and $\theta_0^* = -(x^* + \sum_{j=1}^m \theta_j^* \cdot S_j^{(0)})$ are the requisite hedges associated with the derivative price x^*. Note that the Price of V_D will NOT be the negative of the Price of $-V_D$, hence these prices simply serve as bid prices or ask prices, depending on whether one pays or receives the random outcomes-contingent payoff V_D.

To develop some intuition for what this solution looks like, let us now write some code for the case of 1 risky asset (i.e., $m = 1$). To make things interesting, we will write code for the case where the risky asset price at $t = 1$ (denoted S) follows a normal distribution $S \sim \mathcal{N}(\mu, \sigma^2)$. This means we have a continuous (rather than discrete) set of values for the risky asset price at $t = 1$. Since there are more than 2 random outcomes at time $t = 1$, this is the case of an Incomplete Market. Moreover, we assume the CARA utility function:

$$U(y) = \frac{1 - e^{-a \cdot y}}{a}$$

where a is the CARA coefficient of risk-aversion.

We refer to the units of investment in the risky asset as α and the units of investment in the riskless asset as β. Let S_0 be the spot ($t = 0$) value of the risky asset (riskless asset value at $t = 0$ is 1). Let $f(S)$ be the payoff of the derivative D at $t = 1$. So, the price of derivative D is the breakeven value x^* such that:

$$\max_{\alpha} \mathbb{E}_{S \sim \mathcal{N}(\mu,\sigma^2)}[\frac{1 - e^{-a \cdot (f(S) - (1+r) \cdot x^* + \alpha \cdot (S - (1+r) \cdot S_0))}}{a}]$$
$$= \max_{\alpha} \mathbb{E}_{S \sim \mathcal{N}(\mu,\sigma^2)}[\frac{1 - e^{-a \cdot (\alpha \cdot (S - (1+r) \cdot S_0))}}{a}] \quad (1.18)$$

The maximizing value of α (call it α^*) on the left-hand-side of Equation (1.18) along with $\beta^* = -(x^* + \alpha^* \cdot S_0)$ are the requisite hedges associated with the derivative price x^*.

We set up a `@dataclass MaxExpUtility` with attributes to represent the risky asset spot price S_0 (`risky_spot`), the riskless rate r (`riskless_rate`), mean μ of S (`risky_mean`), standard deviation σ of S (`risky_stdev`), and the payoff function $f(\cdot)$ of the derivative (`payoff_func`).

```
@dataclass(frozen=True)
class MaxExpUtility:
    risky_spot: float  # risky asset price at t=0
    riskless_rate: float  # riskless asset price grows from 1 to 1+r
    risky_mean: float  # mean of risky asset price at t=1
    risky_stdev: float  # std dev of risky asset price at t=1
    payoff_func: Callable[[float], float]  # derivative payoff at t=1
```

Before we write code to solve the derivatives pricing and hedging problem for an incomplete market, let us write code to solve the problem for a complete market (as this will serve as a good comparison against the incomplete market solution). For a complete market, the risky asset has two random prices at $t = 1$: prices $\mu + \sigma$ and $\mu - \sigma$, with probabilities of 0.5 each. As we've seen in Section 1.6.1, we can perfectly replicate a derivative payoff in this complete market situation as it amounts to solving 2 linear equations in 2 unknowns (solution shown in Equation (1.14)). The number of units of the requisite hedges are simply the negatives of the replicating portfolio units. The method `complete_mkt_price_and_hedges` (of the `MaxExpUtility` class) shown below implements this solution, producing a dictionary comprising of the derivative price (`price`) and the hedge units α (`alpha`) and β (`beta`).

```
def complete_mkt_price_and_hedges(self) -> Mapping[str, float]:
    x = self.risky_mean + self.risky_stdev
    z = self.risky_mean - self.risky_stdev
    v1 = self.payoff_func(x)
    v2 = self.payoff_func(z)
    alpha = (v1 - v2) / (z - x)
    beta = - 1 / (1 + self.riskless_rate) * (v1 + alpha * x)
    price = - (beta + alpha * self.risky_spot)
    return {"price": price, "alpha": alpha, "beta": beta}
```

Next we write a helper method `max_exp_util_for_zero` (to handle the right-hand-side of Equation (1.18)) that calculates the maximum expected utility for the special case of a derivative with payoff equal to 0 in all random outcomes at $t = 1$, i.e., it calculates:

$$\max_{\alpha} \mathbb{E}_{S \sim \mathcal{N}(\mu,\sigma^2)}[\frac{1 - e^{-a \cdot (-(1+r) \cdot c + \alpha \cdot (S - (1+r) \cdot S_0))}}{a}]$$

where c is cash paid at $t = 0$ (so, $c = -(\alpha \cdot S_0 + \beta)$).

The method `max_exp_util_for_zero` accepts as input `c: float` (representing the cash c paid at $t = 0$) and `risk_aversion_param: float` (representing the CARA coefficient of risk aversion a). Refering to Section **??** in Appendix **??**, we have a closed-form solution to this maximization problem:

$$\alpha^* = \frac{\mu - (1+r) \cdot S_0}{a \cdot \sigma^2}$$
$$\beta^* = -(c + \alpha^* \cdot S_0)$$

Substituting α^* in the Expected Utility expression above gives the following maximum value for the Expected Utility for this special case:

$$\frac{1 - e^{-a \cdot (-(1+r) \cdot c + \alpha^* \cdot (\mu - (1+r) \cdot S_0)) + \frac{(a \cdot \alpha^* \cdot \sigma)^2}{2}}}{a} = \frac{1 - e^{a \cdot (1+r) \cdot c - \frac{(\mu - (1+r) \cdot S_0)^2}{2\sigma^2}}}{a}$$

```
def max_exp_util_for_zero(
    self,
    c: float,
    risk_aversion_param: float
) -> Mapping[str, float]:
    ra = risk_aversion_param
    er = 1 + self.riskless_rate
    mu = self.risky_mean
    sigma = self.risky_stdev
    s0 = self.risky_spot
    alpha = (mu - s0 * er) / (ra * sigma * sigma)
    beta = - (c + alpha * self.risky_spot)
    max_val = (1 - np.exp(-ra * (-er * c + alpha * (mu - s0 * er))
                          + (ra * alpha * sigma) ** 2 / 2)) / ra
    return {"alpha": alpha, "beta": beta, "max_val": max_val}
```

Next we write a method `max_exp_util` that calculates the maximum expected utility for the general case of a derivative with an arbitrary payoff $f(\cdot)$ at $t = 1$ (provided as input `pf: Callable[[float, float]]` below), i.e., it calculates:

$$\max_{\alpha} \mathbb{E}_{S \sim \mathcal{N}(\mu,\sigma^2)}[\frac{1 - e^{-a \cdot (f(S) - (1+r) \cdot c + \alpha \cdot (S - (1+r) \cdot S_0))}}{a}]$$

Clearly, this has no closed-form solution since $f(\cdot)$ is an arbitary payoff. The method `max_exp_util` uses the `scipy.integrate.quad` function to calculate the expectation as an

integral of the CARA utility function of $f(S) - (1+r) \cdot c + \alpha \cdot (S - (1+r) \cdot S_0)$ multiplied by the probability density of $\mathcal{N}(\mu, \sigma^2)$, and then uses the `scipy.optimize.minimize_scalar` function to perform the maximization over values of α.

```
from scipy.integrate import quad
from scipy.optimize import minimize_scalar

def max_exp_util(
    self,
    c: float,
    pf: Callable[[float], float],
    risk_aversion_param: float
) -> Mapping[str, float]:
    sigma2 = self.risky_stdev * self.risky_stdev
    mu = self.risky_mean
    s0 = self.risky_spot
    er = 1 + self.riskless_rate
    factor = 1 / np.sqrt(2 * np.pi * sigma2)

    integral_lb = self.risky_mean - self.risky_stdev * 6
    integral_ub = self.risky_mean + self.risky_stdev * 6

    def eval_expectation(alpha: float, c=c) -> float:

        def integrand(rand: float, alpha=alpha, c=c) -> float:
            payoff = pf(rand) - er * c\
                     + alpha * (rand - er * s0)
            exponent = -(0.5 * (rand - mu) * (rand - mu) / sigma2
                         + risk_aversion_param * payoff)
            return (1 - factor * np.exp(exponent)) / risk_aversion_param

        return -quad(integrand, integral_lb, integral_ub)[0]

    res = minimize_scalar(eval_expectation)
    alpha_star = res["x"]
    max_val = - res["fun"]
    beta_star = - (c + alpha_star * s0)
    return {"alpha": alpha_star, "beta": beta_star, "max_val": max_val}
```

Finally, it's time to put it all together - the method `max_exp_util_price_and_hedge` below calculates the maximizing x^* in Equation (1.18). First, we call `max_exp_util_for_zero` (with c set to 0) to calculate the right-hand-side of Equation (1.18). Next, we create a wrapper function `prep_func` around `max_exp_util`, which is provided as input to `scipt.optimize.root_scalar` to solve for x^* in the right-hand-side of Equation (1.18). Plugging x^* (`opt_price` in the code below) in `max_exp_util` provides the hedges α^* and β^* (`alpha` and `beta` in the code below).

```
from scipy.optimize import root_scalar

def max_exp_util_price_and_hedge(
    self,
    risk_aversion_param: float
) -> Mapping[str, float]:
    meu_for_zero = self.max_exp_util_for_zero(
        0.,
        risk_aversion_param
    )["max_val"]

    def prep_func(pr: float) -> float:
        return self.max_exp_util(
            pr,
            self.payoff_func,
            risk_aversion_param
        )["max_val"] - meu_for_zero

    lb = self.risky_mean - self.risky_stdev * 10
    ub = self.risky_mean + self.risky_stdev * 10
    payoff_vals = [self.payoff_func(x) for x in np.linspace(lb, ub, 1001)]
```

```
        lb_payoff = min(payoff_vals)
        ub_payoff = max(payoff_vals)
        opt_price = root_scalar(
            prep_func,
            bracket=[lb_payoff, ub_payoff],
            method="brentq"
        ).root
        hedges = self.max_exp_util(
            opt_price,
            self.payoff_func,
            risk_aversion_param
        )
        alpha = hedges["alpha"]
        beta = hedges["beta"]
        return {"price": opt_price, "alpha": alpha, "beta": beta}
```

The above code for the class `MaxExpUtility` is in the file rl/chapter8/max_exp_utility.py. As ever, we encourage you to play with various choices of S_0, r, μ, σ, f to create instances of `MaxExpUtility`, analyze the obtained prices/hedges, and plot some graphs to develop intuition on how the results change as a function of the various inputs.

Running this code for $S_0 = 100, r = 5\%, \mu = 110, \sigma = 25$ when buying a call option (European since we have only one time period) with strike price $= 105$, the method `complete_mkt_price_and hedges` gives an option price of 11.43, risky asset hedge units of -0.6 (i.e., we hedge the risk of owning the call option by short-selling 60% of the risky asset) and riskless asset hedge units of 48.57 (i.e., we take the \$60 proceeds of short-sale less the \$11.43 option price payment = \$48.57 of cash and invest in a riskless bank account earning 5% interest). As mentioned earlier, this is the perfect hedge if we had a complete market (i.e., two random outcomes). Running this code for the same inputs for an incomplete market (calling the method `max_exp_util_price_and_hedge` for risk-aversion parameter values of $a = 0.3, 0.6, 0.9$ gives us the following results:

```
--- Risk Aversion Param = 0.30 ---
{'price': 23.279, 'alpha': -0.473, 'beta': 24.055}
--- Risk Aversion Param = 0.60 ---
{'price': 12.669, 'alpha': -0.487, 'beta': 35.998}
--- Risk Aversion Param = 0.90 ---
{'price': 8.865, 'alpha': -0.491, 'beta': 40.246}
```

We note that the call option price is quite high (23.28) when the risk-aversion is low at $a = 0.3$ (relative to the complete market price of 11.43) but the call option price drops to 12.67 and 8.87 for $a = 0.6$ and $a = 0.9$ respectively. This makes sense since if you are more risk-averse (high a), then you'd be less willing to take the risk of buying a call option and hence, would want to pay less to buy the call option. Note how the risky asset short-sale is significantly less (~47% - ~49%) compared the to the risky asset short-sale of 60% in the case of a complete market. The varying investments in the riskless asset (as a function of the risk-aversion a) essentially account for the variation in option prices (as a function of a). Figure 1.1 provides tremendous intuition on how the hedges work for the case of a complete market and for the cases of an incomplete market with the 3 choices of risk-aversion parameters. Note that we have plotted the negatives of the hedge portfolio values at $t = 1$ so as to visualize them appropriately relative to the payoff of the call option. Note that the hedge portfolio value is a linear function of the risky asset price at $t = 1$. Notice how the slope and intercept of the hedge portfolio value changes for the 3 risk-aversion scenarios and how they compare against the complete market hedge portfolio value.

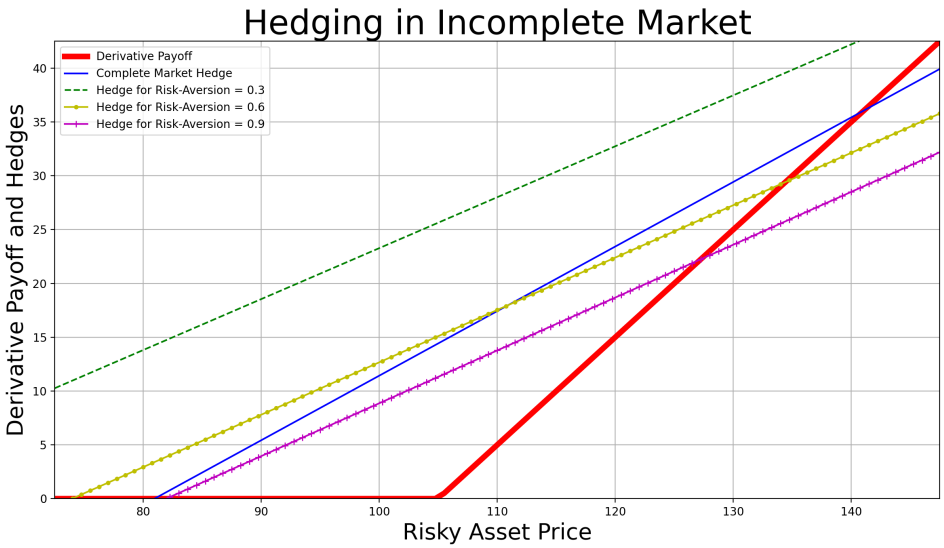


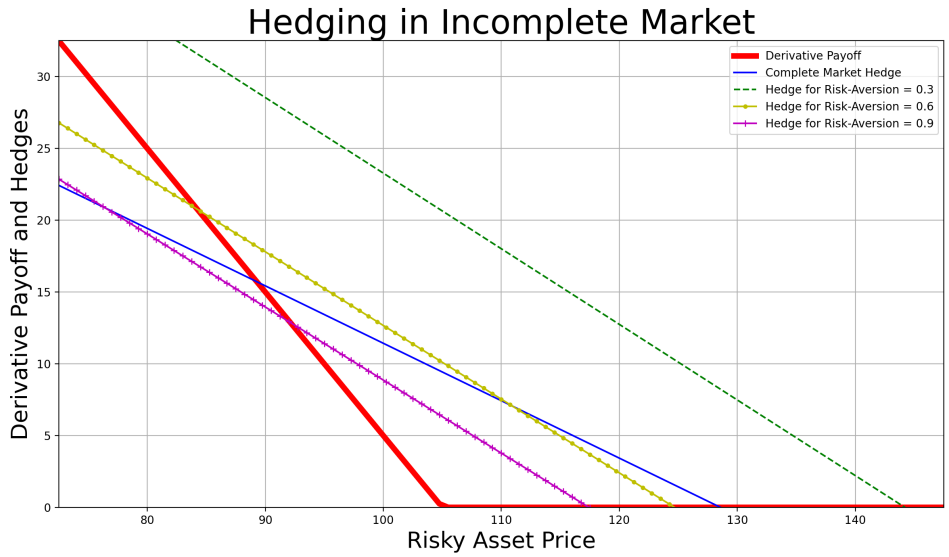
Figure 1.1: Hedges when buying a Call Option

Now let us consider the case of selling the same call option. In our code, the only change we make is to make the payoff function `lambda x: - max(x - 105.0, 0)` instead of `lambda x: max(x - 105.0, 0)` to reflect the fact that we are now selling the call option and so, our payoff will be the negative of that of an owner of the call option.

With the same inputs of $S_0 = 100, r = 5\%, \mu = 110, \sigma = 25$, and for the same risk-aversion parameter values of $a = 0.3, 0.6, 0.9$, we get the following results:

```
--- Risk Aversion Param = 0.30 ---
{'price': -6.307, 'alpha': 0.527, 'beta': -46.395}
--- Risk Aversion Param = 0.60 ---
{'price': -32.317, 'alpha': 0.518, 'beta': -19.516}
--- Risk Aversion Param = 0.90 ---
{'price': -44.236, 'alpha': 0.517, 'beta': -7.506}
```

We note that the sale price demand for the call option is quite low (6.31) when the risk-aversion is low at $a = 0.3$ (relative to the complete market price of 11.43) but the sale price demand for the call option rises sharply to 32.32 and 44.24 for $a = 0.6$ and $a = 0.9$ respectively. This makes sense since if you are more risk-averse (high a), then you'd be less willing to take the risk of selling a call option and hence, would want to charge more for the sale of the call option. Note how the risky asset hedge units are less (~52% - 53%) compared to the risky asset hedge units (60%) in the case of a complete market. The varying riskless borrowing amounts (as a function of the risk-aversion a) essentially account for the variation in option prices (as a function of a). Figure **??** provides the visual intuition on how the hedges work for the 3 choices of risk-aversion parameters (along with the hedges for the complete market, for reference).



Note that each buyer and each seller might have a different level of risk-aversion, meaning each of them would have a different buy price bid/different sale price ask. A transaction can occur between a buyer and a seller (with potentially different risk-aversion levels) if the buyer's bid matches the seller's ask.

1.6.3 Derivatives Pricing when Market has Arbitrage

Finally, we arrive at the case where the market has arbitrage. This is the case where there is no risk-neutral probability measure and there can be multiple replicating portfolios (which can lead to arbitrage). So this is the case where we are unable to price derivatives. To provide intuition for the case of a market with arbitrage, we consider the special case of 2 risky assets ($m = 2$) and 2 random outcomes ($n = 2$), which we will show is a Market with Arbitrage. Without loss of generality, we assume $S_1^{(1)} < S_1^{(2)}$ and $S_2^{(1)} < S_2^{(2)}$. Let us try to determine a risk-neutral probability measure π:

$$S_1^{(0)} = e^{-r} \cdot (\pi(\omega_1) \cdot S_1^{(1)} + \pi(\omega_2) \cdot S_1^{(2)})$$

$$S_2^{(0)} = e^{-r} \cdot (\pi(\omega_1) \cdot S_2^{(1)} + \pi(\omega_2) \cdot S_2^{(2)})$$

$$\pi(\omega_1) + \pi(\omega_2) = 1$$

3 equations and 2 variables implies that there is no risk-neutral probability measure π for various sets of values of $S_1^{(1)}, S_1^{(2)}, S_2^{(1)}, S_2^{(2)}$. Let's try to form a replicating portfolio $(\theta_0, \theta_1, \theta_2)$ for a derivative D:

$$V_D^{(1)} = \theta_0 \cdot e^r + \theta_1 \cdot S_1^{(1)} + \theta_2 \cdot S_2^{(1)}$$

$$V_D^{(2)} = \theta_0 \cdot e^r + \theta_1 \cdot S_1^{(2)} + \theta_2 \cdot S_2^{(2)}$$

2 equations and 3 variables implies that there are multiple replicating portfolios. Each such replicating portfolio yields a price for D as:

$$V_D^{(0)} = \theta_0 + \theta_1 \cdot S_1^{(0)} + \theta_2 \cdot S_2^{(0)}$$

Select two such replicating portfolios with different $V_D^{(0)}$. The combination of one of these replicating portfolios with the negative of the other replicating portfolio is an Arbitrage Portfolio because:

- They cancel off each other's portfolio value in each $t = 1$ states
- The combined portfolio value can be made to be negative at $t = 0$ (by appropriately choosing the replicating portfolio to negate)

So this is a market that admits arbitrage (no risk-neutral probability measure).

1.7 Derivatives Pricing in Multi-Period/Continuous-Time Settings

Now that we have understood the key concepts of derivatives pricing/hedging for the simple setting of discrete-time with a single-period, it's time to do an overview of derivatives pricing/hedging theory in the full-blown setting of multiple time-periods and in continuous-time. While an adequate coverage of this theory is beyond the scope of this book, we will sketch an overview in this section. Along the way, we will cover two derivatives pricing applications that can be modeled as MDPs (and hence, tackled with Dynamic Programming or Reinforcement Learning Algorithms).

The good news is that much of the concepts we learnt for the single-period setting carry over to multi-period and continuous-time settings. The key difference in going over from single-period to multi-period is that we need to adjust the replicating portfolio (i.e., adjust θ) at each time step. Other than this difference, the concepts of arbitrage, risk-neutral probability measures, complete market etc. carry over. In fact, the two fundamental theorems of asset pricing also carry over. It is indeed true that in the multi-period setting, no-arbitrage is equivalent to the existence of a risk-neutral probability measure and market completeness (i.e., replication of derivatives) is equivalent to having a unique risk-neutral probability measure.

1.7.1 Multi-Period Complete-Market Setting

We learnt in the single-period setting that if the market is complete, there are two equivalent ways to conceptualize derivatives pricing:

- Solve for the replicating portfolio (i.e., solve for the units in the fundamental assets that would replicate the derivative payoff), and then calculate the derivative price as the value of this replicating portfolio at $t = 0$.
- Calculate the probabilities of random-outcomes for the unique risk-neutral probability measure, and then calculate the derivative price as the riskless rate-discounted expectation (under this risk-neutral probability measure) of the derivative payoff.

It turns out that even in the multi-period setting, when the market is complete, we can calculate the derivative price (not just at $t = 0$, but at any random outcome at any future time) with either of the above two (equivalent) methods, as long as we appropriately adjust the fundamental assets' units in the replicating portfolio (depending on the random outcome) as we move from one time step to the next. It is important to note that when we alter the fundamental assets' units in the replicating portfolio at each time step, we need to respect the constraint that money cannot enter or leave the replicating portfolio (i.e., it is a

self-financing replicating portfolio with the replicating portfolio value remaining unchanged in the process of altering the units in the fundamental assets). It is also important to note that the alteration in units in the fundamental assets is dependent on the prices of the fundamental assets (which are random outcomes as we move forward from one time step to the next). Hence, the fundamental assets' units in the replicating portfolio evolve as random variables, while respecting the self-financing constraint. Therefore, the replicating portfolio in a multi-period setting in often refered to as a *Dynamic Self-Financing Replicating Portfolio* to reflect the fact that the replicating portfolio is adapting to the changing prices of the fundamental assets. The negatives of the fundamental assets' units in the replicating portfolio form the hedges for the derivative.

To ensure that the market is complete in a multi-period setting, we need to assume that the market is "frictionless" - that we can trade in real-number quantities in any fundamental asset and that there are no transaction costs for any trades at any time step. From a computational perspective, we walk back in time from the final time step (call it $t = T$) to $t = 0$, and calculate the fundamental assets' units in the replicating portfolio in a "backward recursive manner." As in the case of the single-period setting, each backward-recursive step from outcomes at time $t + 1$ to a specific outcome at time t simply involves solving a linear system of equations where each unknown is the replicating portfolio units in a specific fundamental asset and each equation corresponds to the value of the replicating portfolio at a specific outcome at time $t + 1$ (which is established recursively). The market is complete if there is a unique solution to each linear system of equations (for each time t and for each outcome at time t) in this backward-recursive computation. This gives us not just the replicating portfolio (and consequently, hedges) at each outcome at each time step, but also the price at each outcome at each time step (the price is equal to the value of the calculated replicating portfolio at that outcome at that time step).

Equivalently, we can do a backward-recursive calculation in terms of the risk-neutral probability measures, with each risk-neutral probability measure giving us the transition probabilities from an outcome at time step t to outcomes at time step $t+1$. Again, in a complete market, it amounts to a unique solution of each of these linear system of equations. For each of these linear system of equations, an unknown is a transition probability to a time $t + 1$ outcome and an equation corresponds to a specific fundamental asset's prices at the time $t + 1$ outcomes. This calculation is popularized (and easily understood) in the simple context of a Binomial Options Pricing Model. We devote Section 1.8 to coverage of the original Binomial Options Pricing Model and model it as a Finite-State Finite-Horizon MDP (and utilize the Finite-Horizon DP code developed in Chapter **??** to solve the MDP).

1.7.2 Continuous-Time Complete-Market Setting

To move on from multi-period to continuous-time, we simply make the time-periods smaller and smaller, and take the limit of the time-period tending to zero. We need to preserve the complete-market property as we do this, which means that we can trade in real-number units without transaction costs in continuous-time. As we've seen before, operating in continuous-time allows us to tap into stochastic calculus, which forms the foundation of much of the rich theory of continuous-time derivatives pricing/hedging. With this very rough and high-level overview, we refer you to Tomas Bjork's book on Arbitrage Theory in Continuous Time (Björk 2005) for a thorough understanding of this theory.

To provide a sneak-peek into this rich continuous-time theory, we've sketched in Appendix **??** the derivation of the famous Black-Scholes equation and it's solution for the case of European Call and Put Options.

So to summarize, we are in good shape to price/hedge in a multi-period and continuous-time setting if the market is complete. But what if the market is incomplete (which is typical in a real-world situation)? Founded on the Fundamental Theorems of Asset Pricing (which applies to multi-period and continuous-time settings as well), there is indeed considerable literature on how to price in incomplete markets for multi-period/continuous-time, which includes the superhedging approach as well as the *Expected-Utility-Indifference* approach, that we had covered in Subsection 1.6.2 for the simple setting of discrete-time with single-period. However, in practice, these approaches are not adopted as they fail to capture real-world nuances adequately. Besides, most of these approaches lead to fairly wide price bounds that are not particularly useful in practice. In Section 1.10, we extend the *Expected-Utility-Indifference* approach that we had covered for the single-period setting to the multi-period setting. It turns out that this approach can be modeled as an MDP, with the adjustments to the hedge quantities at each time step as the actions of the MDP - calculating the optimal policy gives us the optimal derivative hedging strategy and the associated optimal value function gives us the derivative price. This approach is applicable to real-world situations and one can even incorporate all the real-world frictions in one's MDP to build a practical solution for derivatives trading (covered in Section 1.10).

1.8 Optimal Exercise of American Options cast as a Finite MDP

In this section, we tackle the pricing of American Options in a discrete-time, multi-period setting, assuming the market is complete. To satisfy market completeness, we need to assume that the market is "frictionless" - that we can trade in real-number quantities in any fundamental asset and that there are no transaction costs for any trades at any time step. In particular, we employ the Binomial Options Pricing Model to solve for the price (and hedges) of American Options. The original Binomial Options Pricing Model was developed to price (and hedge) options (including American Options) on an underlying whose price evolves according to a lognormal stochastic process, with the stochastic process approximated in the form of a simple discrete-time, finite-horizon, finite-states process that enables enormous computational tractability. The lognormal stochastic process is basically of the same form as the stochastic process of the underlying price in the Black-Scholes model (covered in Appendix **??**). However, the underlying price process in the Black-Scholes model is specified in the real-world probability measure whereas here we specify the underlying price process in the risk-neutral probability measure. This is because here we will employ the pricing method of riskless rate-discounted expectation (under the risk-neutral probability measure) of the option payoff. Recall that in the single-period setting, the underlying asset price's expected rate of growth is calibrated to be equal to the riskless rate r, under the risk-neutral probability measure. This calibration applies even in the multi-period and continuous-time settings. For a continuous-time lognormal stochastic process, the lognormal drift will hence be equal to r in the risk-neutral probability measure (rather than μ in the real-world probability measure, as per the Black-Scholes model). Precisely, the stochastic process S for the underlying price in the risk-neutral probability measure is:

$$dS_t = r \cdot S_t \cdot dt + \sigma \cdot S_t \cdot dz_t \tag{1.19}$$

where σ is the lognormal dispersion (often refered to as "lognormal volatility" - we will simply call it volatility for the rest of this section). If you want to develop a thorough understanding of the broader topic of change of probability measures and how it affects the

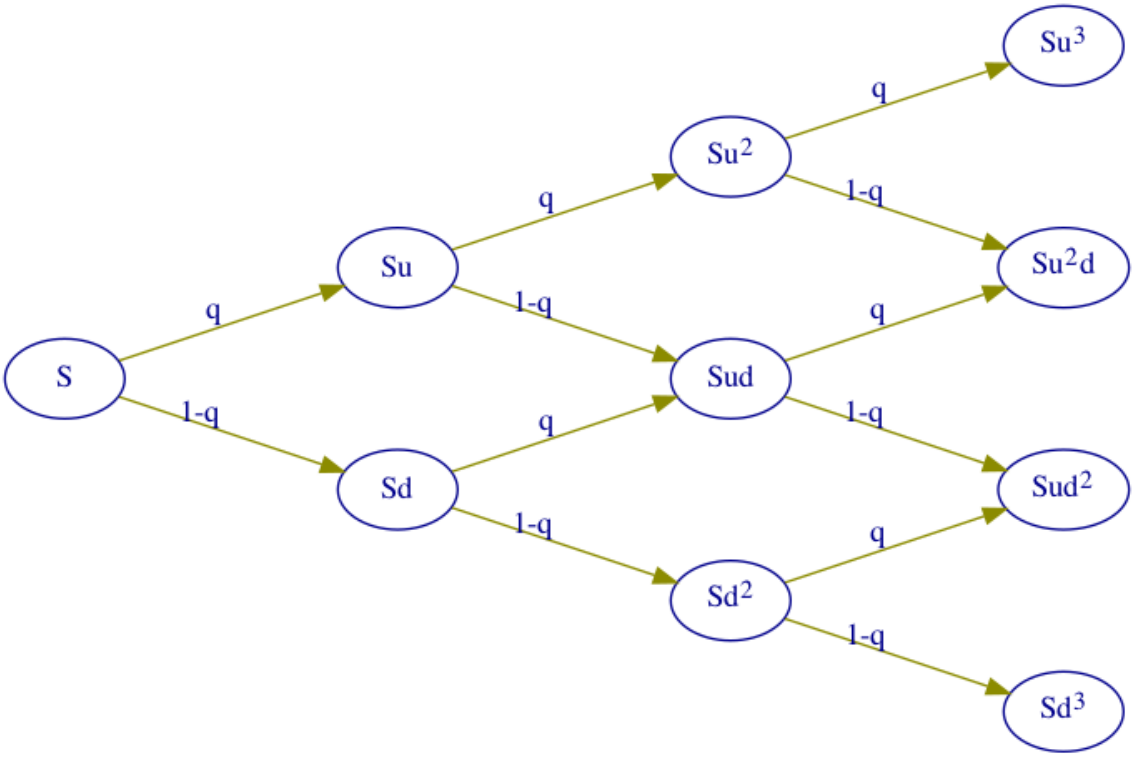


Figure 1.2: Binomial Option Pricing Model (Binomial Tree)

drift term (beyond the scope of this book, but an important topic in continuous-time financial pricing theory), we refer you to the technical material on Radon-Nikodym Derivative and Girsanov Theorem.

The Binomial Options Pricing Model serves as a discrete-time, finite-horizon, finite-states approximation to this continuous-time process, and is essentially an extension of the single-period model we had covered earlier for the case of a single fundamental risky asset. We've learnt previously that in the single-period case for a single fundamental risky asset, in order to be a complete market, we need to have exactly two random outcomes. We basically extend this "two random outcomes" pattern to each outcome at each time step, by essentially growing out a "binary tree." But there is a caveat - with a binary tree, we end up with an exponential (2^i) number of outcomes after i time steps. To contain the exponential growth, we construct a "recombining tree," meaning an "up move" followed by a "down move" ends up in the same underlying price outcome as a "down move" followed by an "up move" (as illustrated in Figure 1.2). Thus, we have $i+1$ price outcomes after i time steps in this "recombining tree." We conceptualize the ascending-sorted sequence of $i+1$ price outcomes as the (time step $= i$) states $\mathcal{S}_i = \{0, 1, \ldots, i\}$ (since the price movements form a discrete-time, finite-states Markov Process). Since we are modeling a lognormal process, we model the discrete-time price moves as multiplicative to the price. We denote $S_{i,j}$ as the price after i time steps in state j (for any $i \in \mathbb{Z}_{\geq 0}$ and for any $0 \leq j \leq i$). So the two random prices resulting from $S_{i,j}$ are $S_{i+1,j+1} = S_{i,j} \cdot u$ and $S_{i+1,j} = S_{i,j} \cdot d$ for some constants u and d (that are calibrated). The important point is that u and d remain constant across time steps i and across states j at each time step i (as seen in Figure 1.2).

Let q be the probability of the "up move" (typically, we use p to denote real-world prob-

ability and q to denote the risk-neutral probability) so that $1-q$ is the probability of the "down move." Just like u and d, the value of q is kept constant across time steps i and across states j at each time step i (as seen in Figure 1.2). q, u and d need to be calibrated so that the probability distribution of log-price-ratios $\{\log\left(\frac{S_{i,0}}{S_{0,0}}\right), \log\left(\frac{S_{i,1}}{S_{0,0}}\right), \ldots, \log\left(\frac{S_{i,i}}{S_{0,0}}\right)\}$ after i time steps (with each time step of interval $\frac{T}{n}$ for a given expiry time $T \in \mathbb{R}^+$ and a fixed number of time steps $n \in \mathbb{Z}^+$) serves as a good approximation to $\mathcal{N}((r-\frac{\sigma^2}{2})\frac{iT}{n}, \frac{\sigma^2 iT}{n})$ (that we know to be the risk-neutral probability distribution of $\log\left(\frac{S_{\frac{iT}{n}}}{S_0}\right)$ in the continuous-time process defined by Equation (1.19), as derived in Section **??** in Appendix **??**), for all $i = 0, 1, \ldots, n$. Note that the starting price $S_{0,0}$ of this discrete-time approximation process is equal to the starting price S_0 of the continuous-time process.

This calibration of q, u and d can be done in a variety of ways and there are indeed several variants of Binomial Options Pricing Models with different choices of how q, u and d are calibrated. We shall implement the choice made in the original Binomial Options Pricing Model that was proposed in a seminal paper by Cox, Ross, Rubinstein (Cox, Ross, and Rubinstein 1979). Their choice is best understood in two steps:

- As a first step, ignore the drift term $r \cdot S_t \cdot dt$ of the lognormal process, and assume the underlying price follows the martingale process $dS_t = \sigma \cdot S_t \cdot dz_t$. They chose d to be equal to $\frac{1}{u}$ and calibrated u such that for any $i \in \mathbb{Z}_{\geq 0}$, for any $0 \leq j \leq i$, the variance of two equal-probability random outcomes $\log\left(\frac{S_{i+1,j+1}}{S_{i,j}}\right) = \log(u)$ and $\log\left(\frac{S_{i+1,j}}{S_{i,j}}\right) = \log(d) = -\log(u)$ is equal to the variance $\frac{\sigma^2 T}{n}$ of the normally-distributed random variable $\log\left(\frac{S_{t+\frac{T}{n}}}{S_t}\right)$ for any $t \geq 0$ (assuming the process $dS_t = \sigma \cdot S_t \cdot dz_t$). This yields:
$$\log^2(u) = \frac{\sigma^2 T}{n} \Rightarrow u = e^{\sigma\sqrt{\frac{T}{n}}}$$
- As a second step, q needs to be calibrated to account for the drift term $r \cdot S_t \cdot dt$ in the lognormal process under the risk-neutral probability measure. Specifically, q is adjusted so that for any $i \in \mathbb{Z}_{\geq 0}$, for any $0 \leq j \leq i$, the mean of the two random outcomes $\frac{S_{i+1,j+1}}{S_{i,j}} = u$ and $\frac{S_{i+1,j}}{S_{i,j}} = \frac{1}{u}$ is equal to the mean $e^{\frac{rT}{n}}$ of the lognormally-distributed random variable $\frac{S_{t+\frac{T}{n}}}{S_t}$ for any $t \geq 0$ (assuming the process $dS_t = r \cdot S_t \cdot dt + \sigma \cdot S_t \cdot dz_t$). This yields:
$$qu + \frac{1-q}{u} = e^{\frac{rT}{n}} \Rightarrow q = \frac{u \cdot e^{\frac{rT}{n}} - 1}{u^2 - 1} = \frac{e^{\frac{rT}{n}+\sigma\sqrt{\frac{T}{n}}} - 1}{e^{2\sigma\sqrt{\frac{T}{n}}} - 1}$$

This calibration for u and q ensures that as $n \to \infty$ (i.e., time step interval $\frac{T}{n} \to 0$), the mean and variance of the binomial distribution after i time steps matches the mean $(r-\frac{\sigma^2}{2})\frac{iT}{n}$ and variance $\frac{\sigma^2 iT}{n}$ of the normally-distributed random variable $\log\left(\frac{S_{\frac{iT}{n}}}{S_0}\right)$ in the continuous-time process defined by Equation (1.19), for all $i = 0, 1, \ldots n$. Note that $\log\left(\frac{S_{i,j}}{S_{0,0}}\right)$ follows a random walk Markov Process (reminiscent of the random walk examples in Chapter **??**) with each movement in state space scaled by a factor of $\log(u)$.

Thus, we have the parameters u and q that fully specify the Binomial Options Pricing Model. Now we get to the application of this model. We are interested in using this model

for optimal exercise (and hence, pricing) of American Options. This is in contrast to the Black-Scholes Partial Differential Equation which only enabled us to price options with a fixed payoff at a fixed point in time (eg: European Call and Put Options). Of course, a special case of American Options is indeed European Options. It's important to note that here we are tackling the much harder problem of the ideal timing of exercise of an American Option - the Binomial Options Pricing Model is well suited for this.

As mentioned earlier, we want to model the problem of Optimal Exercise of American Options as a discrete-time, finite-horizon, finite-states MDP. We set the terminal time to be $t = T + 1$, meaning all the states at time $T + 1$ are terminal states. Here we will utilize the states and state transitions (probabilistic price movements of the underlying) given by the Binomial Options Pricing Model as the states and state transitions in the MDP. The MDP actions in each state will be binary - either exercise the option (and immediately move to a terminal state) or don't exercise the option (i.e., continue on to the next time step's random state, as given by the Binomial Options Pricing Model). If the exercise action is chosen, the MDP reward is the option payoff. If the continue action is chosen, the reward is 0. The discount factor γ is $e^{-\frac{rT}{n}}$ since (as we've learnt in the single-period case), the price (which translates here to the Optimal Value Function) is defined as the riskless rate-discounted expectation (under the risk-neutral probability measure) of the option payoff. In the multi-period setting, the overall discounting amounts to composition (multiplication) of each time step's discounting (which is equal to γ) and the overall risk-neutral probability measure amounts to the composition of each time step's risk-neutral probability measure (which is specified by the calibrated value q).

Now let's write some code to determine the Optimal Exercise of American Options (and hence, the price of American Options) by modeling this problem as a discrete-time, finite-horizon, finite-states MDP. We create a `dataclass OptimalExerciseBinTree` whose attributes are `spot_price` (specifying the current, i.e., time=0 price of the underlying), `payoff` (specifying the option payoff, when exercised), `expiry` (specifying the time T to expiration of the American Option), `rate` (specifying the riskless rate r), `vol` (specifying the lognormal volatility σ), and `num_steps` (specifying the number n of time steps in the binomial tree). Note that each time step is of interval $\frac{T}{n}$ (which is implemented below in the method `dt`). Note also that the `payoff` function is fairly generic taking two arguments - the first argument is the time at which the option is exercised, and the second argument is the underlying price at the time the option is exercised. Note that for a typical American Call or Put Option, the payoff does not depend on time and the dependency on the underlying price is the standard "hockey-stick" payoff that we are now fairly familiar with (however, we designed the interface to allow for more general option payoff functions).

The set of states $\mathcal{S}_i$ at time step i (for all $0 \leq i \leq T + 1$) is: $\{0, 1, \ldots, i\}$ and the method `state_price` below calculates the price in state j at time step i as:

$$S_{i,j} = S_{0,0} \cdot e^{(2j-i)\sigma\sqrt{\frac{T}{n}}}$$

Finally, the method `get_opt_vf_and_policy` calculates u (`up_factor`) and q (`up_prob`), prepares the requisite state-reward transitions (conditional on current state and action) to move from one time step to the next, and passes along the constructed time-sequenced transitions to `rl.finite_horizon.get_opt_vf_and_policy` (which we had written in Chapter **??**) to perform the requisite backward induction and return an `Iterator` on pairs of `V[int]` and `FiniteDeterministicPolicy[int, bool]`. Note that the states at any time-step i are the integers from 0 to i and hence, represented as `int`, and the actions are represented as `bool` (`True` for exercise and `False` for continue). Note that we represent an early terminal

state (in case of option exercise before expiration of the option) as -1.

```
from rl.distribution import Constant, Categorical
from rl.finite_horizon import optimal_vf_and_policy
from rl.dynamic_programming import V
from rl.policy import FiniteDeterministicPolicy

@dataclass(frozen=True)
class OptimalExerciseBinTree:

    spot_price: float
    payoff: Callable[[float, float], float]
    expiry: float
    rate: float
    vol: float
    num_steps: int

    def dt(self) -> float:
        return self.expiry / self.num_steps

    def state_price(self, i: int, j: int) -> float:
        return self.spot_price * np.exp((2 * j - i) * self.vol *
                                        np.sqrt(self.dt()))

    def get_opt_vf_and_policy(self) -> \
            Iterator[Tuple[V[int], FiniteDeterministicPolicy[int, bool]]]:
        dt: float = self.dt()
        up_factor: float = np.exp(self.vol * np.sqrt(dt))
        up_prob: float = (np.exp(self.rate * dt) * up_factor - 1) / \
            (up_factor * up_factor - 1)
        return optimal_vf_and_policy(
            steps=[
                {NonTerminal(j): {
                    True: Constant(
                        (
                            Terminal(-1),
                            self.payoff(i * dt, self.state_price(i, j))
                        )
                    ),
                    False: Categorical(
                        {
                            (NonTerminal(j + 1), 0.): up_prob,
                            (NonTerminal(j), 0.): 1 - up_prob
                        }
                    )
                } for j in range(i + 1)}
                for i in range(self.num_steps + 1)
            ],
            gamma=np.exp(-self.rate * dt)
        )
```

Now we want to try out this code on an American Call Option and American Put Option. We know that it is never optimal to exercise an American Call Option before the option expiration. The reason for this is as follows: Upon early exercise (say at time $\tau < T$), we borrow cash K (to pay for the purchase of the underlying) and own the underlying (valued at S_τ). So, at option expiration T, we owe cash $K \cdot e^{r(T-\tau)}$ and own the underlying valued at S_T, which is an overall value at time T of $S_T - K \cdot e^{r(T-\tau)}$. We argue that this value is always less than the value $\max(S_T - K, 0)$ we'd obtain at option expiration T if we'd made the choice to not exercise early. If the call option ends up in-the-money at option expiration T (i.e., $S_T > K$), then $S_T - K \cdot e^{r(T-\tau)}$ is less than the value $S_T - K$ we'd get by exercising at option expiration T. If the call option ends up not being in-the-money at option expiration T (i.e., $S_T \leq K$), then $S_T - K \cdot e^{r(T-\tau)} < 0$ which is less than the 0 payoff we'd obtain at option expiration T. Hence, we are always better off waiting until option expiration (i.e. it

is never optimal to exercise a call option early, no matter how much in-the-money we get before option expiration). Hence, the price of an American Call Option should be equal to the price of an European Call Option with the same strike price and expiration time. However, for an American Put Option, it is indeed sometimes optimal to exercise early and hence, the price of an American Put Option is greater then the price of an European Put Option with the same strike price and expiration time. Thus, it is interesting to ask the question: For each time $t < T$, what is the threshold of underlying price S_t below which it is optimal to exercise an American Put Option? It is interesting to view this threshold as a function of time (we call this function as the optimal exercise boundary of an American Put Option). One would expect that this optimal exercise boundary rises as one gets closer to the option expiration T. But exactly what shape does this optimal exercise boundary have? We can answer this question by analyzing the optimal policy at each time step - we just need to find the state k at each time step i such that the Optimal Policy $\pi_i^*(\cdot)$ evaluates to True for all states $j \leq k$ (and evaluates to False for all states $j > k$). We write the following method to calculate the Optimal Exercise Boundary:

```
    def option_exercise_boundary(
        self,
        policy_seq: Sequence[FiniteDeterministicPolicy[int, bool]],
        is_call: bool
    ) -> Sequence[Tuple[float, float]]:
        dt: float = self.dt()
        ex_boundary: List[Tuple[float, float]] = []
        for i in range(self.num_steps + 1):
            ex_points = [j for j in range(i + 1)
                         if policy_seq[i].action_for[j] and
                         self.payoff(i * dt, self.state_price(i, j)) > 0]
            if len(ex_points) > 0:
                boundary_pt = min(ex_points) if is_call else max(ex_points)
                ex_boundary.append(
                    (i * dt, opt_ex_bin_tree.state_price(i, boundary_pt))
                )
        return ex_boundary
```

`option_exercise_boundary` takes as input `policy_seq` which represents the sequence of optimal policies π_i^* for each time step $0 \leq i \leq T$, and produces as output the sequence of pairs $(\frac{iT}{n}, B_i)$ where

$$B_i = \max_{j:\pi_i^*(j)=True} S_{i,j}$$

with the little detail that we only consider those states j for which the option payoff is positive. For some time steps i, none of the states j qualify as $\pi_i^*(j) = True$, in which case we don't include that time step i in the output sequence.

To compare the results of American Call and Put Option Pricing on this Binomial Options Pricing Model against the corresponding European Options prices, we write the following method to implement the Black-Scholes closed-form solution (derived as Equations **??** and **??** in Appendix **??**):

```
from scipy.stats import norm
    def european_price(self, is_call: bool, strike: float) -> float:
        sigma_sqrt: float = self.vol * np.sqrt(self.expiry)
        d1: float = (np.log(self.spot_price / strike) +
                     (self.rate + self.vol ** 2 / 2.) * self.expiry) \
            / sigma_sqrt
        d2: float = d1 - sigma_sqrt
```

```
        if is_call:
            ret = self.spot_price * norm.cdf(d1) - \
                strike * np.exp(-self.rate * self.expiry) * norm.cdf(d2)
        else:
            ret = strike * np.exp(-self.rate * self.expiry) * norm.cdf(-d2) - \
                self.spot_price * norm.cdf(-d1)
        return ret
```

Here's some code to price an American Put Option (changing `is_call` to `True` will price American Call Options):

```
from rl.gen_utils.plot_funcs import plot_list_of_curves

spot_price_val: float = 100.0
strike: float = 100.0
is_call: bool = False
expiry_val: float = 1.0
rate_val: float = 0.05
vol_val: float = 0.25
num_steps_val: int = 300

if is_call:
    opt_payoff = lambda _, x: max(x - strike, 0)
else:
    opt_payoff = lambda _, x: max(strike - x, 0)

opt_ex_bin_tree: OptimalExerciseBinTree = OptimalExerciseBinTree(
    spot_price=spot_price_val,
    payoff=opt_payoff,
    expiry=expiry_val,
    rate=rate_val,
    vol=vol_val,
    num_steps=num_steps_val
)

vf_seq, policy_seq = zip(*opt_ex_bin_tree.get_opt_vf_and_policy())
ex_boundary: Sequence[Tuple[float, float]] = \
    opt_ex_bin_tree.option_exercise_boundary(policy_seq, is_call)
time_pts, ex_bound_pts = zip(*ex_boundary)
label = ("Call" if is_call else "Put") + " Option Exercise Boundary"
plot_list_of_curves(
    list_of_x_vals=[time_pts],
    list_of_y_vals=[ex_bound_pts],
    list_of_colors=["b"],
    list_of_curve_labels=[label],
    x_label="Time",
    y_label="Underlying Price",
    title=label
)

european: float = opt_ex_bin_tree.european_price(is_call, strike)
print(f"European Price = {european:.3f}")

am_price: float = vf_seq[0][NonTerminal(0)]
print(f"American Price = {am_price:.3f}")
```

This prints as output:

```
European Price = 7.459
American Price = 7.971
```

So we can see that the price of this American Put Option is significantly higher than the price of the corresponding European Put Option. The exercise boundary produced by this code is shown in Figure 1.3. The locally-jagged nature of the exercise boundary curve is because of the "diamond-like" local-structure of the underlying prices at the nodes in the

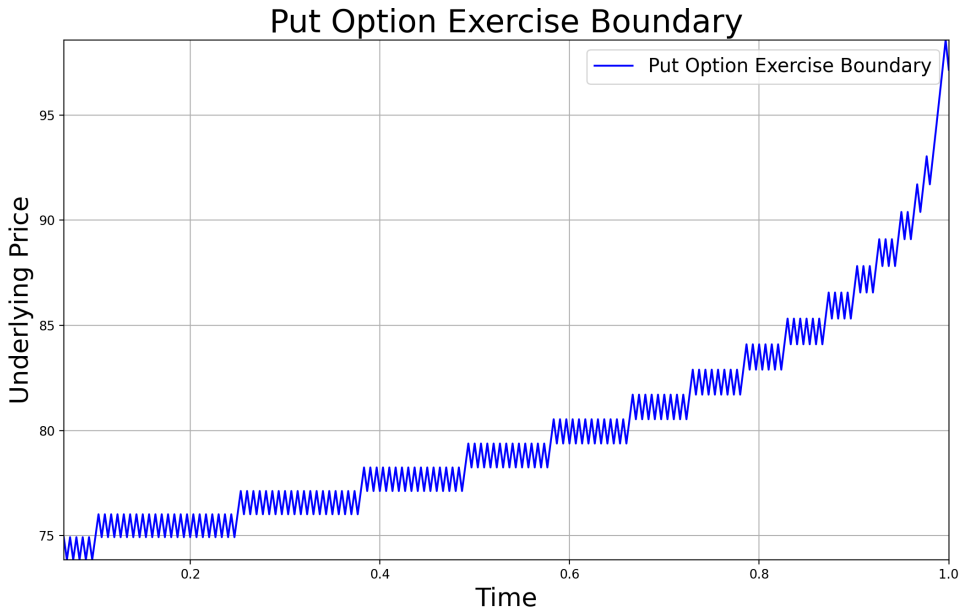


Figure 1.3: Put Option Exercise Boundary

binomial tree. We can see that when the time to expiry is large, it is not optimal to exercise unless the underlying price drops significantly. It is only when the time to expiry becomes quite small that the optimal exercise boundary rises sharply towards the strike price value.

Changing `is_call` to `True` (and not changing any of the other inputs) prints as output:

```
European Price = 12.336
American Price = 12.328
```

This is a numerical validation of our proof above that it is never optimal to exercise an American Call Option before option expiration.

The above code is in the file rl/chapter8/optimal_exercise_bin_tree.py. As ever, we encourage you to play with various choices of inputs to develop intuition for how American Option Pricing changes as a function of the inputs (and how American Put Option Exercise Boundary changes). Note that you can specify the option payoff as any arbitrary function of time and the underlying price.

1.9 Generalizing to Optimal-Stopping Problems

In this section, we generalize the problem of Optimal Exercise of American Options to the problem of Optimal Stopping in Stochastic Calculus, which has several applications in Mathematical Finance, including pricing of exotic derivatives. After defining the Optimal Stopping problem, we show how this problem can be modeled as an MDP (generalizing the MDP modeling of Optimal Exercise of American Options), which affords us the ability to solve them with Dynamic Programming or Reinforcement Learning algorithms.

First we define the concept of *Stopping Time*. Informally, Stopping Time τ is a random time (time as a random variable) at which a given stochastic process exhibits certain behavior. Stopping time is defined by a *stopping policy* to decide whether to continue or stop

a stochastic process based on the stochastic process' current and past values. Formally, it is a random variable τ such that the event $\{\tau \leq t\}$ is in the σ-algebra $\mathcal{F}_t$ of the stochastic process, for all t. This means the stopping decision (i.e., *stopping policy*) of whether $\tau \leq t$ only depends on information up to time t, i.e., we have all the information required to make the stopping decision at any time t.

A simple example of Stopping Time is *Hitting Time* of a set A for a process X. Informally, it is the first time when X takes a value within the set A. Formally, Hitting Time T_{X_A} is defined as:

$$T_{X,A} = \min\{t \in \mathbb{R} | X_t \in A\}$$

A simple and common example of Hitting Time is the first time a process exceeds a certain fixed threshold level. As an example, we might say we want to sell a stock when the stock price exceeds \$100. This \$100 threshold constitutes our stopping policy, which determines the stopping time (hitting time) in terms of when we want to sell the stock (i.e., exit owning the stock). Different people may have different criterion for exiting owning the stock (your friend's threshold might be \$90), and each person's criterion defines their own stopping policy and hence, their own stopping time random variable.

Now that we have defined Stopping Time, we are ready to define the Optimal Stopping problem. *Optimal Stopping* for a stochastic process X is a function $W(\cdot)$ whose domain is the set of potential initial values of the stochastic process and co-domain is the length of time for which the stochastic process runs, defined as:

$$W(x) = \max_{\tau} \mathbb{E}[H(X_\tau)|X_0 = x]$$

where τ is a set of stopping times of X and $H(\cdot)$ is a function from the domain of the stochastic process values to the set of real numbers.

Intuitively, you should think of Optimal Stopping as searching through many Stopping Times (i.e., many Stopping Policies), and picking out the best Stopping Policy - the one that maximizes the expected value of a function $H(\cdot)$ applied on the stochastic process at the stopping time.

Unsurprisingly (noting the connection to Optimal Control in an MDP), $W(\cdot)$ is called the Value function, and H is called the Reward function. Note that sometimes we can have several stopping times that maximize $\mathbb{E}[H(X_\tau)]$ and we say that the optimal stopping time is the smallest stopping time achieving the maximum value. We mentioned above that Optimal Exercise of American Options is a special case of Optimal Stopping. Let's understand this specialization better:

- X is the stochastic process for the underlying's price in the risk-neutral probability measure.
- x is the underlying security's current price.
- τ is a set of exercise times, each exercise time corresponding to a specific policy of option exercise (i.e., specific stopping policy).
- $W(\cdot)$ is the American Option price as a function of the underlying's current price x.
- $H(\cdot)$ is the option payoff function (with riskless-rate discounting built into $H(\cdot)$).

Now let us define Optimal Stopping problems as control problems in Markov Decision Processes (MDPs).

- The MDP *State* at time t is X_t.
- The MDP *Action* is Boolean: Stop the Process or Continue the Process.

- The MDP *Reward* is always 0, except upon Stopping, when it is equal to $H(X_\tau)$.
- The MDP *Discount Factor* γ is equal to 1.
- The MDP probabilistic-transitions are governed by the Stochastic Process X.

A specific policy corresponds to a specific stopping-time random variable τ, the Optimal Policy π^* corresponds to the stopping-time τ^* that yields the maximum (over τ) of $\mathbb{E}[H(X_\tau)|X_0 = x]$, and the Optimal Value Function V^* corresponds to the maximum value of $\mathbb{E}[H(X_\tau)|X_0 = x]$.

For discrete time steps, the Bellman Optimality Equation is:

$$V^*(X_t) = \max(H(X_t), \mathbb{E}[V^*(X_{t+1})|X_t])$$

Thus, we see that Optimal Stopping is the solution to the above Bellman Optimality Equation (solving the Control problem of the MDP described above). For a finite number of time steps, we can run a backward induction algorithm from the final time step back to time step 0 (essentially a generalization of the backward induction we did with the Binomial Options Pricing Model to determine Optimal Exercise of American Options).

Many derivatives pricing problems (and indeed many problems in the broader space of Mathematical Finance) can be cast as Optimal Stopping and hence can be modeled as MDPs (as described above). The important point here is that this enables us to employ Dynamic Programming or Reinforcement Learning algorithms to identify optimal stopping policy for exotic derivatives (which typically yields a pricing algorithm for exotic derivatives). When the state space is large (eg: when the payoff depends on several underlying assets or when the payoff depends on the history of underlying's prices, such as Asian Options-payoff with American exercise feature), the classical algorithms used in the finance industry for exotic derivatives pricing are not computationally tractable. This points to the use of Reinforcement Learning algorithms which tend to be good at handling large state spaces by effectively leveraging sampling and function approximation methodologies in the context of solving the Bellman Optimality Equation. Hence, we propose Reinforcement Learning as a promising alternative technique to pricing of certain exotic derivatives that can be cast as Optimal Stopping problems. We will discuss this more after having covered Reinforcement Learning algorithms.

1.10 Pricing/Hedging in an Incomplete Market cast as an MDP

In Subsection 1.6.2, we developed a pricing/hedging approach based on *Expected-Utility-Indifference* for the simple setting of discrete-time with single-period, when the market is incomplete. In this section, we extend this approach to the case of discrete-time with multi-period. In the single-period setting, the solution is rather straightforward as it amounts to an unconstrained multi-variate optimization together with a single-variable root-solver. Now when we extend this solution approach to the multi-period setting, it amounts to a sequential/dynamic optimal control problem. Although this is far more complex than the single-period setting, the good news is that we can model this solution approach for the multi-period setting as a Markov Decision Process. This section will be dedicated to modeling this solution approach as an MDP, which gives us enormous flexibility in capturing the real-world nuances. Besides, modeling this approach as an MDP permits us to tap into some of the recent advances in Deep Learning and Reinforcement Learning (i.e. Deep Reinforcement Learning). Since we haven't yet learnt about Reinforcement Learning algorithms, this section won't cover the algorithmic aspects (i.e., how to solve the MDP) - it

will simply cover how to model the MDP for the *Expected-Utility-Indifference* approach to pricing/hedging derivatives in an incomplete market.

Before we get into the MDP modeling details, it pays to remind that in an incomplete market, we have multiple risk-neutral probability measures and hence, multiple valid derivative prices (each consistent with no-arbitrage). This means the market/traders need to "choose" a suitable risk-neutral probability measure (which amounts to choosing one out of the many valid derivative prices). In practice, this "choice" is typically made in ad-hoc and inconsistent ways. Hence, our proposal of making this "choice" in a mathematically-disciplined manner by noting that ultimately a trader is interested in maximizing the "risk-adjusted return" of a derivative together with it's hedges (by sequential/dynamic adjustment of the hedge quantities). Once we take this view, it is reminiscent of the *Asset Allocation* problem we covered in Chapter **??** and the maximization objective is based on the specification of preference for trading risk versus return (which in turn, amounts to specification of a Utility function). Therefore, similar to the Asset Allocation problem, the decision at each time step is the set of adjustments one needs to make to the hedge quantities. With this rough overview, we are now ready to formalize the MDP model for this approach to multi-period pricing/hedging in an incomplete market. For ease of exposition, we simplify the problem setup a bit, although the approach and model we describe below essentially applies to more complex, more frictionful markets as well. Our exposition below is an adaptation of the treatment in the Deep Hedging paper by Buehler, Gonon, Teichmann, Wood, Mohan, Kochems (Bühler et al. 2018).

Assume we have a portfolio of m derivatives and we refer to our collective position across the portfolio of m derivatives as D. Assume each of these m derivatives expires by time T (i.e., all of their contingent cashflows will transpire by time T). We model the problem as a discrete-time finite-horizon MDP with the terminal time at $t = T+1$ (i.e., all states at time $t = T + 1$ are terminal states). We require the following notation to model the MDP:

- Denote the derivatives portfolio-aggregated *Contingent Cashflows* at time t as $X_t \in \mathbb{R}$.
- Assume we have n assets trading in the market that would serve as potential hedges for our derivatives position D.
- Denote the number of units held in the hedge positions at time t as $\boldsymbol{\alpha}_t \in \mathbb{R}^n$.
- Denote the cashflows per unit of hedges at time t as $\boldsymbol{Y}_t \in \mathbb{R}^n$.
- Denote the prices per unit of hedges at time t as $\boldsymbol{P}_t \in \mathbb{R}^n$.
- Denote the trading account value at time t as $\beta_t \in \mathbb{R}$.

We will use the notation that we have previously used for discrete-time finite-horizon MDPs, i.e., we will use time-subscripts in our notation.

We denote the State Space at time t (for all $0 \leq t \leq T+1$) as $\mathcal{S}_t$ and a specific state at time t as $s_t \in \mathcal{S}_t$. Among other things, the key ingredients of s_t include: $\boldsymbol{\alpha}_t, \boldsymbol{P}_t, \beta_t, D$. In practice, s_t will include many other components (in general, any market information relevant to hedge trading decisions). However, for simplicity (motivated by ease of articulation), we assume s_t is simply the 4-tuple:

$$s_t := (\boldsymbol{\alpha}_t, \boldsymbol{P}_t, \beta_t, D)$$

We denote the Action Space at time t (for all $0 \leq t \leq T$) as $\mathcal{A}_t$ and a specific action at time t as $\boldsymbol{a}_t \in \mathcal{A}_t$. $\boldsymbol{a}_t$ represents the number of units of hedges traded at time t (i.e., adjustments to be made to the hedges at each time step). Since there are n hedge positions (n assets to be traded), $\boldsymbol{a}_t \in \mathbb{R}^n$, i.e., $\mathcal{A}_t \subseteq \mathbb{R}^n$. Note that for each of the n assets, it's corresponding

component in $\boldsymbol{a}_t$ is positive if we buy the asset at time t and negative if we sell the asset at time t. Any trading restrictions (eg: constraints on short-selling) will essentially manifest themselves in terms of the exact definition of $\mathcal{A}_t$ as a function of s_t.

State transitions are essentially defined by the random movements of prices of the assets that make up the potential hedges, i.e., $\mathbb{P}[\boldsymbol{P}_{t+1}|\boldsymbol{P}_t]$. In practice, this is available either as an explicit transition-probabilities model, or more likely available in the form of a *simulator*, that produces an on-demand sample of the next time step's prices, given the current time step's prices. Either way, the internals of $\mathbb{P}[\boldsymbol{P}_{t+1}|\boldsymbol{P}_t]$ are estimated from actual market data and realistic trading/market assumptions. The practical details of how to estimate these internals are beyond the scope of this book - it suffices to say here that this estimation is a form of supervised learning, albeit fairly nuanced due to the requirement of capturing the complexities of market-price behavior. For the following description of the MDP, simply assume that we have access to $\mathbb{P}[\boldsymbol{P}_{t+1}|\boldsymbol{P}_t]$ in *some form*.

It is important to pay careful attention to the sequence of events at each time step $t = 0, \ldots, T$, described below:

1. Observe the state $s_t := (\boldsymbol{\alpha}_t, \boldsymbol{P}_t, \beta_t, D)$.
2. Perform action (trades) $\boldsymbol{a}_t$, which produces trading account value change $= -\boldsymbol{a}_t^T \cdot \boldsymbol{P}_t$ (note: this is an inner-product in $\mathbb{R}^n$).
3. These trades incur transaction costs, for example equal to $\gamma \cdot abs(\boldsymbol{a}_t^T) \cdot \boldsymbol{P}_t$ for some $\gamma \in \mathbb{R}^+$ (note: abs, denoting absolute value, applies point-wise on $\boldsymbol{a}_t^T \in \mathbb{R}^n$, and then we take it's inner-product with $\boldsymbol{P}_t \in \mathbb{R}^n$).
4. Update α_t as:
$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{\alpha}_t + \boldsymbol{a}_t$$
At termination, we need to force-liquidate, which establishes the constraint: $\boldsymbol{a}_T = -\boldsymbol{\alpha}_T$.
5. Realize end-of-time-step cashflows from the derivatives position D as well as from the (updated) hedge positions. This is equal to $X_{t+1} + \boldsymbol{\alpha}_{t+1}^T \cdot \boldsymbol{Y}_{t+1}$ (note: $\boldsymbol{\alpha}_{t+1}^T \cdot \boldsymbol{Y}_{t+1}$ is an inner-product in $\mathbb{R}^n$).
6. Update trading account value β_t as:
$$\beta_{t+1} = \beta_t - \boldsymbol{a}_t^T \cdot \boldsymbol{P}_t - \gamma \cdot abs(\boldsymbol{a}_t^T) \cdot \boldsymbol{P}_t + X_{t+1} + \boldsymbol{\alpha}_{t+1}^T \cdot \boldsymbol{Y}_{t+1}$$
7. MDP Reward $r_{t+1} = 0$ for all $t = 0, \ldots, T-1$ and $r_{T+1} = U(\beta_{T+1})$ for an appropriate concave Utility function (based on the extent of risk-aversion).
8. Hedge prices evolve from $\boldsymbol{P}_t$ to $\boldsymbol{P}_{t+1}$, based on price-transition model of $\mathbb{P}[\boldsymbol{P}_{t+1}|\boldsymbol{P}_t]$.

Assume we now want to enter into an incremental position of derivatives-portfolio D' in m' derivatives. We denote the combined position as $D \cup D'$. We want to determine the *Price* of the incremental position D', as well as the hedging strategy for $D \cup D'$.

Denote the Optimal Value Function at time t (for all $0 \leq t \leq T$) as $V_t^* : \mathcal{S}_t \to \mathbb{R}$. Pricing of D' is based on the principle that introducing the incremental position of D' together with a calibrated cash payment/receipt (Price of D') at $t = 0$ should leave the Optimal Value (at $t = 0$) unchanged. Precisely, the Price of D' is the value x^* such that

$$V_0^*((\boldsymbol{\alpha}_0, \boldsymbol{P}_0, \beta_0 - x^*, D \cup D')) = V_0^*((\boldsymbol{\alpha}_0, \boldsymbol{P}_0, \beta_0, D))$$

This Pricing principle is known as the principle of *Indifference Pricing*. The hedging strategy for $D \cup D'$ at time t (for all $0 \leq t < T$) is given by the associated Optimal Deterministic Policy $\pi_t^* : \mathcal{S}_t \to \mathcal{A}_t$

1.11 Key Takeaways from this Chapter

- The concepts of Arbitrage, Completeness and Risk-Neutral Probability Measure.
- The two fundamental theorems of Asset Pricing.
- Pricing of derivatives in a complete market in two equivalent ways: A) Based on construction of a replicating portfolio, and B) Based on riskless rate-discounted expectation in the risk-neutral probability measure.
- Optimal Exercise of American Options (and it's generalization to Optimal Stopping problems) cast as an MDP Control problem.
- Pricing and Hedging of Derivatives in an Incomplete (real-world) Market cast as an MDP Control problem.

Bibliography

Björk, Tomas. 2005. *Arbitrage Theory in Continuous Time*. 2. ed., reprint. Oxford [u.a.]: Oxford Univ. Press. http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+505893878&sourceid=fbw_bibsonomy.

Bühler, Hans, Lukas Gonon, Josef Teichmann, and Ben Wood. 2018. "Deep Hedging." http://arxiv.org/abs/1802.03042.

Cox, J., S. Ross, and M. Rubinstein. 1979. "Option Pricing: A Simplified Approach." *Journal of Financial Economics* 7: 229–63.

Hull, John C. 2010. *Options, Futures, and Other Derivatives*. Seventh. Pearson.