A Project Report

On

Reinforcement Learning for pricing American Options

Hrithik Bandaru	Rakshith S.	Ritvik Raj Padige
17XJ1A0508	17XJ1A0539	17XJ1A0542

Under the supervision of

Dr. Manoj Kumar Yadav

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF

SE-308: THIRD YEAR PROJECT

MAHINDRA ECOLE CENTRALE

COLLEGE OF ENGINEERING

HYDERABAD (May 2020)

ACKNOWLEDGMENTS

In performing our assignment, we had to take the help and guideline of some respected persons, who deserve our greatest gratitude. The completion of this assignment gives us much Pleasure. We would like to show our gratitude **Mr Manoj Kumar Yadav, Dept. of Mathematics, Mahindra Ecole Centrale** for giving us a good guideline for the paper throughout numerous consultations. We would also like to expand our deepest gratitude to all those who have directly and indirectly guided us in writing this assignment.

In addition, a thank you to Professor Manoj Kumar Yadav, who introduced us to the Methodology of work, and whose passion for the "underlying structures" had lasting effect. We also thank the institution for consent to include copyrighted pictures as a part of our paper.

Many people, especially our classmates and team members itself, have made valuable comments and suggestions on this proposal which gave us an inspiration to improve our paper. We thank all the people for their help directly and indirectly to complete our paper.

Mahindra
École Centrale
COLLEGE OF ENGINEERING

Hyderabad

Certificate

This is to certify that the project report entitled "**Reinforcement Learning for Pricing American Options**" submitted by Mr. Hrithik Bandaru (ID No. 17XJ1A0508), Mr. Rakshith.S (ID No. 17XJ1A0539) and Mr. Ritvik Raj Padige (ID No. 17XJ1A0542) in partial fulfillment of the requirements of the course PR 301, Project Course, embodies the work done by him/her under my supervision and guidance.

(SUPERVISOR NAME & Signature)

Mahindra Ecole Centrale, Hyderabad

Date:

ABSTRACT

Options are significant apparatuses for modern finance. This venture talks about the techniques of Reinforcement Learning (RL)— in particular, dynamic programming — to handle the topic of learning exercise policies for American choices. We likewise analyzed the outcomes acquired from the Dynamic Programming method with the Monte Carlo Method.

Our outcomes show that the exercise policies found by Monte Carlo method are more fulfilling than those found by Dynamic Programming, for the information considered. Likewise, the arrangements got from Reinforcement Learning, a worldview of Machine Learning will propel the best in class in a significant and testing field of use. This likewise uncovers computational finance stays a productive field for the potential execution of ML methods.

CONTENTS

1.1 Options

What is an option? An option is a contract between two parties: **the buyer** and **the seller**. The price at which the transaction occurs is referred to as the **premium**. That premium is quoted on a per-share basis. Because an option contract represents the right to buy or sell 100 shares of stock, the premium (paid by the buyer to the seller) is that per-share premium multiplied by 100. Thus, if you sell one option at $1.20, you receive $120. The following description of options is not complicated. **There are two types of options**:

A **call** option gives its owner the right to buy and a put option gives its owner the right to sell:

A SPECIFIED ITEM (CALLED THE UNDERLYING ASSET)

AT A SPECIFIED PRICE (CALLED THE STRIKE PRICE)

FOR A SPECIFIED PERIOD OF TIME (ANY TIME BEFORE THE OPTION'S EXPIRATION DATE)

Thus, in return for paying a premium to purchase an option: The owner of a call option has the right (but not the obligation) to buy the underlying asset at the strike price any time before the expiration date.

The owner of a **put** option has the right (but not the obligation) to sell the underlying asset at the strike price any time before the expiration date. The underlying asset is almost always 100 shares of a specified stock or ETF.

1.5 Advantages of Options

There are four key advantages options may give an investor:

- They may provide increased cost-efficiency
- They may be less risky than equities
- They have the potential to deliver higher percentage returns
- They offer a number of strategic alternatives.

1.3 Basics of Options and Key terms

1.3.1 Strike Price – the strike price (or exercise price) of an option is the fixed price at which the owner of the option can buy (in the case of a call), or sell (in the case of a put), the underlying security or commodity. The strike price may be set by reference to the spot price (market price) of the underlying security or commodity on the day an option is taken out, or it may be fixed at a discount or at a premium. The strike price is a key variable in a derivatives contract between two parties. Where the contract requires delivery of the underlying instrument, the trade will be at the strike price, regardless of the market price of the underlying instrument at that time.

1.3.2 Premium - The absolute expense of a choice is known as the premium. The premium normally increments as the option turns out to be further in-the-cash (where the strike cost turns out to be increasingly ideal comparable to the current underlying price). The premium diminishes as the option turns out to be progressively out-of-the-money (when the strike cost is less favorable according to the underlying security).

1.3.3 Out of the Money (OTM)

“Out of the money” (OTM) is an expression used to describe an option contract that only contains intrinsic value. These options will have a delta of less than 50.0.

An OTM call option will have a strike price that is higher than the market price of the underlying asset. Alternatively, an OTM put option has a strike price that is lower than the market price of the underlying asset.

1.3.4 Intrinsic value

Intrinsic value is also used in options pricing to determine how in-the-money an option is or how much profit currently exists.

For review, an options contract grants the buyer the right, but not the commitment, to buy or sell the underlying security at a preset price called the strike price. Options have expiration dates at which they can be exercised or converted to the shares of the underlying security. A call option allows an investor to buy assets such as a stock while a put option allows an investor to sell the asset. If the market price at expiration is above the strike price, the call option is profitable or in-the-money. If the market price is below the strike of the put option, the put is profitable. If either option is not profitable at expiry, the options expire worthlessly, and the buyer loses the upfront fee or premium paid at the onset.

The intrinsic value of both call and put options is the difference between the underlying stock's price and the strike price. In the case of both call and put options, if the calculated value is negative, the intrinsic value is zero. In other words, intrinsic value only measures the profit as determined by the difference between the option's strike price and market price.

It's important to note the intrinsic value does not include the premium meaning it's not the true profit of the trade since it doesn't include the initial cost. Intrinsic value only shows how in-the-money an option is considering its strike price and the market price of the underlying asset.

1.3.5 Time Value – The value of an option based on the amount of time before the contract expires. Time is valuable to investors because of the possibility that an option's intrinsic value will increase during the contract's time frame. As the expiration date approaches, time value decreases. This is known as time decay or "theta," after the options pricing model used to calculate it.

1.3.6 Interest Rates – Interest rates have little, however quantifiable, impacts on option prices. By and large, as interest rates rise, call premiums increment and put premiums decline. This is a direct result of the expenses related to owning the underlying: The purchase acquires either interest expense (if the cash is borrowed) or lost interest income (if existing assets are utilized to buy the shares). In either case, the purchaser will have interest costs.

1.4 American and European Options

American Options - American options outline the timeframe when the option holder can exercise their option contract rights. These rights allow the holder to buy or sell—depending on if the option is a call or put—the underlying asset, at the set strike price on or before the predetermined expiration date. Since investors have the freedom to exercise their options at any point during the life of the contract, American style options are more valuable than the limited European options. However, the ability to exercise early carries an added premium or cost. The last day to exercise a weekly American option is normally on the Friday of the week in which the option contract expires. Conversely, the last day to exercise a monthly American option is normally the third Friday of the month.

European Options - European option is a version of an options contract that limits execution to its expiration date. In other words, if the underlying security such as a stock has moved in price an investor would not be able to exercise the option early and take delivery of or sell the shares. Instead, the call or put action will only take place on the date of option maturity.

1.5 Difference between European and American Options

1.5.1 Difference One: Settlement price – The first major difference between American- and European-style options is the manner in which the settlement price of the underlying index is determined. The settlement price is the official closing price for the index and is the basis for determining which options are in the money and the intrinsic value of those options

When a stock closes for trading on the third Friday of the month, the last trade determines the settlement price. It is obvious which options finish in- or out-of-the money and which options will be exercised by their owners. As anyone would reasonably anticipate, the last price is the final price. However, the rules are completely different with European-style index options.

1.5.2 Difference Two: Cash-Settled – European options are cash-settled. If an option is ITM when the settlement price is determined, the exercise/assignment process involves transferring the intrinsic value of the option, in cash, from the account of the person who is short the option to the option owner's account. When using American options, shares of the underlying asset (not cash) exchange hands via exercise and assignment.

Example

> **YOU BOUGHT FOUR NDX JUN 1950 CALLS @ $450 (OPTIONS ON THE NASDAQ 100 INDEX) ON A DAY WHEN YOU FELT BULLISH. AT EXPIRATION, THE SETTLEMENT PRICE OF NDX IS 1950.97. YOUR OPTIONS ARE ITM, BUT THIS IS NOT A VICTORY. THE INTRINSIC VALUE OF EACH CALL IS $97. $388 (4 X $97) IS TRANSFERRED INTO YOUR ACCOUNT AND THE LONG CALL OPTIONS ARE REMOVED. IT IS AS IF YOU SOLD THE CALLS @ $97. NO SHARES WERE BOUGHT OR SOLD. THERE IS NO CONCERN WITH BUYING OR SELLING A BUNCH OF DIFFERENT STOCKS TO RID YOUR PORTFOLIO OF ANY UNWANTED POSITIONS— BECAUSE THERE IS ONLY CASH.**

1.5.3 Difference Three: Exercise Rights – The owner of an American-style option has the right to exercise any time before it expires. European style options cannot be exercised before expiration. Thus, owners of European-style options who do not want to hold a position have no choice but to sell their options. This represents a small inconvenience for the option buyer and a nice bonus for the option seller who never has to be concerned with being assigned an exercise notice early.

1.6 Reinforcement Learning

Reinforcement learning is a computational approach to understanding and automating goal-directed learning and decision making. It is distinguished from other computational approaches by its emphasis on learning by an agent from direct interaction with its environment, without requiring exemplary supervision or complete models of the environment. In our opinion, reinforcement learning is the first field to seriously address the computational issues that arise when learning from interaction with an environment in order to achieve long-term goals. Reinforcement learning uses the formal framework of Markov decision processes to define the interaction between a learning agent and its environment in terms of states, actions, and rewards. This framework is intended to be a simple way of representing essential features of the artificial intelligence problem. These features include a sense of cause and effect, a sense of uncertainty and nondeterminism, and the existence of explicit goals. The concepts of value and value function are key to most of the reinforcement learning methods that we consider in this book. We take the position that value functions are important for efficient search in the space of policies. The use of value functions distinguishes reinforcement learning methods from evolutionary methods that search directly in policy space guided by evaluations of entire policies.

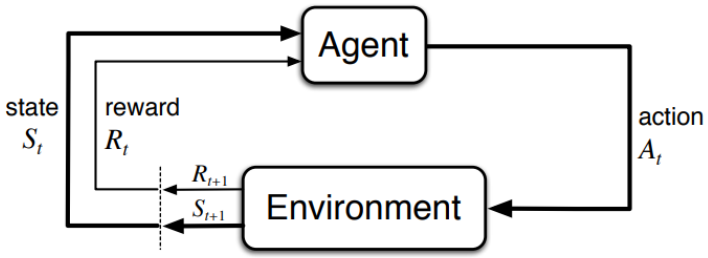

Figure 1: Inserted from Ref(2)

1.7 Monte Carlo Simulation

Monte Carlo Simulation is a mathematical technique that generates random variables for modelling risk or uncertainty of a certain system. The random variables or inputs are modelled on the basis of probability distributions such as normal, log normal, etc. Different iterations or simulations are run for generating paths and the outcome is arrived at by using suitable numerical computations. Monte Carlo Simulation is the most tenable method used when a model has uncertain parameters or a dynamic complex system needs to be analyzed. It is a probabilistic method for modelling risk in a system.

Monte Carlo Simulation uses probability distribution for modelling a stochastic or a random variable. Different probability distributions are used for modelling input variables such as normal, lognormal, uniform, and triangular. From probability distribution of input variable, different paths of outcome are generated.

Compared to deterministic analysis, the Monte Carlo method provides a superior simulation of risk. It gives an idea of not only what outcome to expect but also the probability of occurrence of that outcome. It is also possible to model correlated input variables.For instance, Monte Carlo Simulation can be used to compute the value at risk of a portfolio. This method tries to predict the worst return expected from a portfolio, given a certain confidence interval for a specified time period.

1.8 Markov Decision Processes

A mathematical representation of a complex decision-making process is "Markov Decision Processes" (MDP). MDP is defined by:

- A state S, which represents every state that one could be in, within a defined world.
- A model or transition function T; which is a function of the current state, the action taken and the state where we end up. This transition produces a certain probability of ending up in state S', starting from the state S and taking the action A.
- Actions are things I can do in a particular state.
- A reward is a scaler value for being in a state. It tells us the usefulness of entering the state.

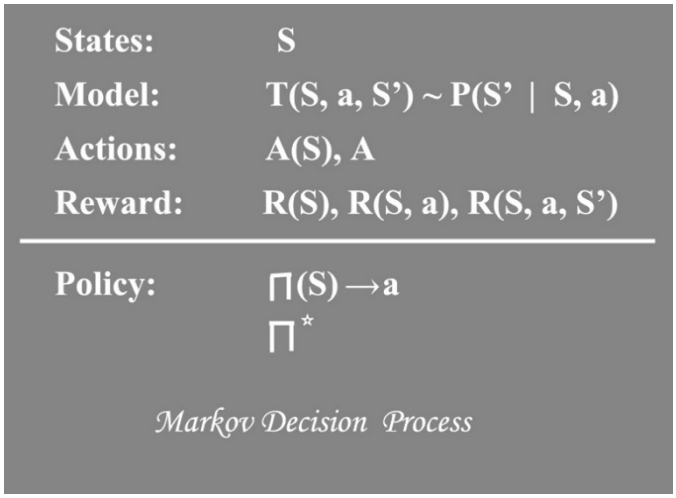

Figure 2: Inserted from Ref (8)

1.9 Algorithms for American Option Pricing

1.9.1 Dynamic Programming

Basically, we want to solve a big problem that is hard. We can first solve a few smaller but similar problems, if those can be solved, then the solution to the big problem will be easy to get. To solve each of those smaller problems, we use the same idea, we first solve a few even smaller problems. Continue doing it, we will eventually encounter a problem we know how to solve Dynamic programming has the same feature, the difference is that at each step, there might be some optimization involved.

1.9.2 Monte Carlo Method

While the previous approach assumes, we have a complete knowledge of the environment, many times this is not the case. Monte Carlo (MC) methods are able to learn directly from experience or episodes rather than relying on the prior knowledge of the environment dynamics. The term "Monte Carlo" is often used broadly for any estimation method whose operation involves a significant random component. Interestingly, in many cases is possible to generate experiences sampled according to the desired probability distributions but infeasible to obtain the distributions in explicit form.

PROBLEM DEFINITION

The aim of this project was to evaluate and improve American Options using Reinforcement Algorithm i.e. Monte Carlo methods and compare the results obtained with Dynamic Programming Method. At present, we have worked policy evaluation using dynamic programming and monte Carlo methods. The policy improvement is left as future work.

2.0 This problem answers the following questions:

- How Monte Carlo methods provides better results than Dynamic programming Method in policy evaluation ?

The idea and methods we used to solve the problem were influenced by the book Reinforcement Learning: An Introduction Second edition, in progress Richard S. Sutton and Andrew G. Barto c 2014, 2015.

3.1 DYNAMIC PROGRAMMING

Convergence of iterative policy evaluation on a small grid-world. The left column is the sequence of approximations of the state-value function for the random policy (all actions equal). The right column is the sequence of greedy policies corresponding to the value function estimates (arrows are shown for all actions achieving the maximum). The last policy is guaranteed only to be an improvement over the random policy, but in this case it, and all policies after the third iteration, are optimal.

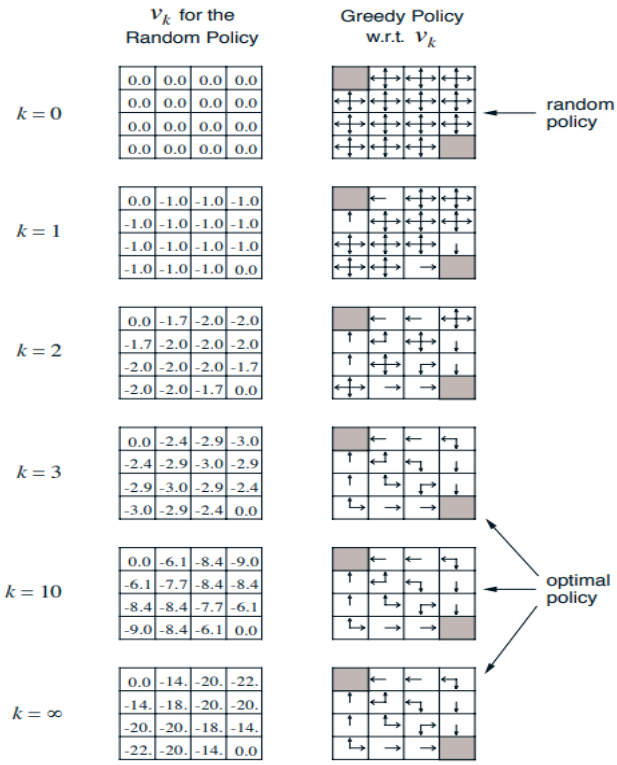

Figure 3: Inserted from Ref (2)

Notice in the right column that as we update the values of the states, we can now generate more and more efficient policies until we reach the optimal "rules" a robot must follow to end up in the termination states as fast as possible.

$$\begin{aligned} v_\pi(s) &= \mathbb{E}_\pi\left[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \cdots \mid S_t = s\right] \\ &= \mathbb{E}_\pi[R_{t+1} + \gamma v_\pi(S_{t+1}) \mid S_t = s] \\ &= \sum_a \pi(a|s) \sum_{s',r} p(s', r|s, a)\Big[r + \gamma v_\pi(s')\Big], \end{aligned}$$

Understanding the algorithms implemented in solving the problems.

4.1 IMPORTANT TERMINOLOGIES AND SYMBOLS

π ← policy to be evaluated.

V ← An arbitrary state-value function.

s ← An empty list.

G← Temporary Variable.

θ← A small positive number.

γ← Gamma (discounting rate)

a← Action

r← Reward

4.2 Dynamic programming algorithm

$$V(s) \leftarrow \sum_a \pi(a|s) \sum_{s',r} p(s', r|s, a)\left[r + \gamma V(s')\right]$$

The *V(s')* is multiplied by a *gamma*, which is the discounting factor. In our case we use *gamma*=1 but the idea of the discounting factor is that immediate rewards (the *r* in our equation) are more important than the future rewards (reflected by the value of *s'*) and we can adjust the *gamma* to reflect this fact. The idea of discounting factor is that immediate rewards (the *r* in our equation) are more important than the future rewards (reflected by the value of *s'*) and we can adjust the *gamma* to reflect this fact.

Input π, the policy to be evaluated
Initialize an array $V(s) = 0$, for all $s \in \mathcal{S}^+$
Repeat
  $\Delta \leftarrow 0$
  For each $s \in \mathcal{S}$:
    $v \leftarrow V(s)$
    $V(s) \leftarrow \sum_a \pi(a|s) \sum_{s',r} p(s', r|s, a)\left[r + \gamma V(s')\right]$
    $\Delta \leftarrow \max(\Delta, |v - V(s)|)$
until $\Delta < \theta$ (a small positive number)
Output $V \approx v_\pi$

4.3 Monte Carlo Method algorithm

For each simulation we save the 4 values: (1) the initial state, (2) the action taken, (3) the reward received and (4) the final state. In the end, a simulation is just an array containing *x* arrays of these values.Now, from these simulations, we iterate from the end of the "experience" array, and compute G as the previous state value in the same experience (weighed by gamma, the discount factor) plus the received reward in that state. We then store *G* in an array of *Returns(St)*. Finally, for each state we compute the average of the Returns(St) and we set this as the state value at a particular iteration.

Input: a policy π to be evaluated

Initialize:

$V(s) \in \mathbb{R}$, arbitrarily, for all $s \in \mathcal{S}$

$Returns(s) \leftarrow$ an empty list, for all $s \in \mathcal{S}$

Loop forever (for each episode):

Generate an episode following π: $S_0, A_0, R_1, S_1, A_1, R_2, \ldots, S_{T-1}, A_{T-1}, R_T$

$G \leftarrow 0$

Loop for each step of episode, $t = T-1, T-2, \ldots, 0$:

$G \leftarrow \gamma G + R_{t+1}$

Unless S_t appears in $S_0, S_1, \ldots, S_{t-1}$:

Append G to $Returns(S_t)$

$V(S_t) \leftarrow$ average$(Returns(S_t))$

Results and Conclusion

- Technically, we don’t have to compute all the state-values for all the states if we don’t want. We could just focus on a particular grid point and start all the simulations from that initial state to sample episodes that include that state, ignoring all others. This can radically decrease the computational expense. By this we can clearly see that using Monte Carlo methods is clearly beneficial than using Dynamic Programming.

- Dynamic Programming approaches assume complete knowledge of the environment (the MDP). In practice, we often don't have full knowledge of how the world works.

- Monte Carlo (MC) methods can learn directly from experience collected by interacting with the environment. An episode of experience is a series of (State, Action, Reward, Next State) tuples.

- MC methods work based on episodes. We sample episodes of experience and make updates to our estimates at the end of each episode. MC methods have high variance (due to lots of random decisions within an episode) but are unbiased.

- As we said before, this approach does not require a full understanding of the environment dynamics and we can learn directly from experience or simulation.

References

1. Prof. Mengdi Wang's and Prof. Dimitri Bertsekas' lecture notes.
2. Reinforcement Learning: An Introduction Second edition, in progress Richard S. Sutton and Andrew G. Barto.
3. Policy Iteration Adaptive Dynamic Programming Algorithm for Discrete-Time Nonlinear Systems Derong Liu, Fellow, IEEE, and Qinglai Wei, Member, IEEE.
4. https://artint.info/html/ArtInt_228.html
5. https://www.investopedia.com/articles/optioninvestor/06/options4advantages.asp
6. https://corporatefinanceinstitute.com/resources/knowledge/trading-investing/options-calls-and-puts/
7. https://www.nerdwallet.com/blog/investing/options-trading-definitions/
8. https://www.geeksforgeeks.org/markov-decision-process/
9. Reinforcement Learning: An Introduction Author(s): Richard Sutton; Andrew G Barto Publisher: Bradford Book, Year: 2018.
10. The Rookie's Guide to Options Author: Mark D. Wolfinger 2nd Edition.