A MACHINE LEARNING APPROACH TO STOCHASTIC OPTIMAL CONTROL

PABLO EVER AVALOS ROBLES

Master's Program in Statistics

APPROVED:

Michael Pokojovy, Ph.D., Chair

Ori Rosen, Ph.D.

Thomas M. Fullerton, Jr., Ph.D.

Stephen L. Crites, Jr., Ph.D.
Dean of the Graduate School

A MACHINE LEARNING APPROACH TO STOCHASTIC OPTIMAL CONTROL

by

PABLO EVER AVALOS ROBLES

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

Department of Mathematical Sciences

THE UNIVERSITY OF TEXAS AT EL PASO

May 2022

Acknowledgements

Foremost, I would like to thank UTEP and the faculty of the mathematics department for giving me the opportunity to continue my education. I would like to express my gratitude to my thesis advisor, Dr. Michael Pokojovy, whom introduced me to this thesis topic and guided me throughout the process. I would also like to thank the rest of my thesis committee, Dr. Ori Rosen and Dr. Thomas M. Fullerton, Jr. for giving me the feedback needed to complete my thesis. I cannot see myself completing this thesis without the support from my advisor and the committee.

Lastly, I want to thank my family and friends that supported me throughout my graduate education. They were always there to reassure any doubts that I may have had. I will always be grateful of everyone that was there for me at any moment during my graduate degree.

Abstract

Merton's portfolio optimization problem is a well-renowned problem in financial mathematics which seeks to optimize the investment decision for an investor. In the simplest situation, the market consists of a risk-less asset (i.e. a bond) that pays back a relatively low interest rate, and a risky asset (i.e. a stock) that follows a geometric Brownian motion. The optimal allocation strategy of the investor's wealth is found by optimizing the expected utility along the stochastic evolution of the market. This thesis focuses on several different applications of this optimization problem. We look at pre-constructed analytical solutions and showcase the results. We formulate simulated allocation strategies and compare results. Lastly, we approach this optimization problem using machine learning, specifically, by training neural networks.

Table of Contents

List of Tables

List of Figures

Introduction

Merton's portfolio optimization problem is a well-renowned optimization problem in financial mathematics, named after Nobel laureate Robert Merton. In the simplest scenario, this continuous-time finance issue revolves around finding the optimal investment decision for the investor, who only has two possible options of investment. One is a risk-free asset (i.e. a savings account, or a bond) which usually pays back a relative low rate of interest. The other is a risky asset (i.e. a stock, or real estate) whose price is assumed to follow a geometric Brownian motion. Assuming the investor has a limited amount of time, the problem's goal is to maximize the personal expected utility gained from consuming the portfolio's wealth. The consumption pattern of the investor may vary. The process involves modeling a market and finding the optimal allocation of wealth between the risk-free asset (bond) and the risky asset (stock). We can then use a utility function to model the investor's attitude towards the confronted risk of investment, which when optimized, will allow us to, both, maximize the final/consumed wealth, and simultaneously control the risk of losing money.

The model can be considered a continuous-time market model, meaning the capital can be re-balanced at any moment before time has run out. In other words, wealth allocation can be switched between the different assets without an additional cost. It is also assumed that the assets can be sold or bought arbitrarily at any time. Lastly, the investor only gets information on current prices.

Merton formulated this problem in 1969 and solved it in 1971 using a stochastic optimal control approach. The value function of this optimization problem could be solved using dynamic programming, by deriving a nonlinear partial differential equation referred to as the Hamilton-Jacobi-Bellman (HJB) equation. This equation, however, is impractical to solve analytically

considering its nonlinearity. For special cases, however, a utility function can be considered in the constant relative risk aversion (CRRA) class, such as logarithmic or power utility, which is useful, because, for these cases, the optimal strategy is to keep a constant fraction of the current wealth in risky assets.

The goal of this thesis is to understand the applications of Merton's portfolio optimization problem and approach it using several different methods. We will look at pre-constructed analytical solutions and showcase the results. We will formulate simulated allocation strategies and compare results. Lastly, we approach this optimization problem using machine learning.

The machine learning algorithm that is considered is a simple neural network system. Neural network systems are trained algorithms that run input values through a series of hidden neuron layers which eventually lead to a predicted output value. This process will require stochastic gradient descent implemented in the backpropagation of the system training since we do not work with pre-labeled training data. The goal will be to predict the best allocation strategy from running current wealth through a trained neural network.

Theoretical Background on Bellman's Dynamic Programming Principle

Before we can tackle the optimization problem, we must establish a framework where stochastic control is used. The primary contribution of this thesis will be in the application of the model, but we will take the time to construct the mathematical theory as similarly seen by [Tikosi, 2016] and [Aboagye, 2018]. This presentation does not contain any new developments but solely serves the purpose of presenting a concise review.

Brownian Motion

Let us start building the basic ideas of stochastic finance, by following the work of [Karatzas and Scheve, 1998] and [Steele, 2001].

Let us consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with filtration $(\mathcal{F}_t)_{t\in[0,T]}$. We assume the filtration is right-continuous and $\mathcal{F}_0$ contains all the measure 0 sets. Let $T > 0$ be a non-random terminal time, and $(W_t)_{t\in[0,T]}$ denote a Brownian motion, which is a stochastic process.

Definition: A process W = $(W_t)_{t\epsilon[0,T]}$ is a $\mathbb{P}$**-Brownian motion** if it is $\mathcal{F}_t$-adapted and it satisfies

1. W is continuous with $W_0 = 0$,
2. $W_t - W_s$ is independent of $\mathcal{F}_s, 0 \leq s < t \leq T$,
3. For any finite $0 \leq s < t \leq T, W_t - W_s \sim N(0, t-s)$ under the probability measure $\mathbb{P}$.

We can also denote a vector of higher dimensional Brownian motion:

$$W_t = (W_t^1, \dots, W_t^n)^\top$$

where the W^i are independent Brownian motions, all adapted to the same filtration $\mathcal{F}$. This is essential in modeling volatility, or in our case, the fluctuation of how a stock would act in a market setting.

Remark: The expected value of Brownian motion W_t at any time t is zero, that is, $\mathbb{E}[W_t] = 0$, and variance is t, $Var(W_t) = t$, since $W_t \sim N(0, t)$ when $s = 0$. Although not practically relevant for this thesis, we still note that if running time would be left indefinite, then the variance would also run to infinity.

STOCHASTIC CONTROL

We must then introduce basic stochastic control theory based on work by [Saß, 2006]. We will consider Itô processes, which satisfy stochastic differential equations driven by Brownian motion. We have the stochastic differential equation:

$$\mathrm{d}X_t = b(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t.$$

By stochastic Picard & Lindelöf theorem, this equation has a strong solution when drift $b: [0,\infty) \times \mathbb{R}^n \to \mathbb{R}^n$ and diffusion coefficient $\sigma: [0,\infty) \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$ satisfy the following for all $0 \le s, t$ and $x, y \in \mathbb{R}^n$

$$\| b(s,x) - b(t,y) \| + \| \sigma(s,x) - \sigma(t,y) \| \le K(\| y - x \| + |t - s|)$$

$$\| b(t,x) \|^2 + \| \sigma(t,x) \|^2 \le K^2(1 + \| x \|^2)$$

for some positive constant K.

Definition: An $\mathcal{F}$-progressively measurable process $(u_t)_{t\epsilon[0,T]}$ with values in some set $\mathcal{U} \subseteq \mathbb{R}^p$ is called a **control process**. An n-dimensional process $(Y_t)_{t\epsilon[0,T]}$ controlled by u_t if it is defined by

$$\mathrm{d}Y_t = b(t, Y_t, u_t)\mathrm{d}t + \sigma(t, Y_t, u_t)\mathrm{d}W_t$$

where $Y_0 = y_0$,

$$b: [0,T] \times \mathbb{R}^n \times \mathcal{U} \to \mathbb{R}^n$$

$$\sigma: [0,T] \times \mathbb{R}^n \times \mathcal{U} \to \mathbb{R}^{n\times m}.$$

and $(W_t)_{t\epsilon[0,T]}$ denotes the m-dimensional Brownian motion.

The optimization objective is

$$J(t,x,u) = E[\int_t^T \psi(t, X_t^u, u_t)\mathrm{d}t + \Psi(\mathrm{T}, \mathrm{X}_\mathrm{T}^\mathrm{u}) | \mathrm{X}_\mathrm{t}^u = x]$$

We can denote an admissible set of controls by $\mathcal{A}(t,x)$ which contains all the controls that fulfill the following:

1. The control process $u = (u_s)_{s\epsilon[t,T]}$ is progressively measurable with values in $\mathcal{U}$ and $E\left[\int_t^T ||u_s||^2 \mathrm{d}s\right] < \infty$.
2. The stochastic differential equation describing the controlled process has a unique strong solution $(X_s)_{s\in[t,T]}$ with $X_t = x$ and

$$E_{t,x}\left[\sup_{t\leq s\leq T} \| X_s \|^2\right] < \infty.$$

3. The optimization criterion $J(t,x,u)$ is well defined.

Our goal is to maximize the value function of the control problem defined by

$$V(t,x) = \sup_{u\in\mathcal{A}(t,x,)} J(t,x,u).$$

We must find the optimal value $V(0,x_0)$ and the optimal control strategy u^*, from which this value is obtained.

DYNAMIC PROGRAMMING

We will be using what we call dynamic programming to break down the optimization problem into smaller sub-problems to be able to achieve the best overall optimum. We will briefly go over how this method works and will apply it to the portfolio optimization problem later in the thesis. We construct the theory similarly done by [Tikosi, 2016]. To use dynamic programming, we must define a specific optimal substructure. We use the **Bellman Principle**. To this end, we introduce the value function:

$$V(t,x) = \sup_{u\in\mathcal{A}(t,x)} E_{t,x}\left[\int_t^{t_1} \psi(s, X_s^u, u_s)ds + V\left(t_1, X_{t_1}^u\right)\right]$$

The Bellman principle is used to solve optimal control problems by isolating part of the whole optimization problem. An optimal control on an interval ($[t, t_1]$ in our case) will remain optimal if we continue optimally at t_1. We can then continue by applying Itô's formula to $V(t_1, X_{t_1}^u)$ if the wealth process has sufficient smoothness properties. We then end up with:

$$
\begin{aligned}
V(t,x) = \sup_{\mathcal{u}\epsilon\mathcal{A}(t,x)} E_{t,x}\Bigg[\int_t^{t_1} \psi(s, X_s^u, u_s)\mathrm{d}s + V(t, X_t) \\
+ \int_t^{t_1} V_t(s, X_s)\big(D_x V(s, X_s)\big)^\top b(s, X_s, u_s)\mathrm{d}s + \int_t^{t_1} \big(D_x V(s, X_s)\big)^\top \sigma(s, X_s, u_s)\mathrm{d}W_s \\
+ \frac{1}{2}\int_t^{t_1} tr\big(D_{xx} V(s, V_s)\big)^\top \sigma(s, X_s, u_s)\sigma(s, X_s, u_s)^\top \mathrm{d}W_s\Bigg]
\end{aligned}
$$

The expectation of $\int_t^{t_1}\big(D_x V(s, X_s)\big)^\top \sigma(s, X_s, u_s)\mathrm{d}W_s$ is 0, considering it is a martingale. We can also use notation $a(s, X_s, u_s) = \sigma(s, X_s, u_s)\sigma(s, X_s, u_s)^\top$ for the diffusion matrix. Applying this, we obtain

$$
\begin{aligned}
V(t,x) = \sup_{\mathcal{u}\epsilon\mathcal{A}(t,x)} E_{t,x}\Bigg[\int_t^{t_1} \psi(s, X_s^u, u_s)\mathrm{d}s + V(t, X_t) \\
+ \int_t^{t_1} V_t(s, X_s)\big(D_x V(s, X_s)\big)^\top b(s, X_s, u_s)\mathrm{d}s \\
+ \frac{1}{2}\int_t^{t_1} tr\big(D_{xx} V(s, V_s)\big)^\top a(s, X_s, u_s)\mathrm{d}W_s\Bigg]
\end{aligned}
$$

We then subtract $V(t,x)$, divide by $(t - t_1)$, and let t_1 tend to *t*. We also take the supremum and limit after checking if taking the limit can be interchanged with the expectation. Taking the conditional expectation when $X_t = x$, we know $V(t, X_t) = V(t, x)$. Consequently, we end up with

$$
0 = \sup_{\mathcal{u}\epsilon\mathcal{U}}\left[\psi(t,x,u) + V_t(t,x) + \big(D_x V(t,x)\big)^\top b(t,x,u) + \frac{1}{2} tr\big(\big(D_{xx} V(t,x)\big) a(t,x,u)\big)\right]
$$

From here, we can define a differential operator that depends on *u*:

$$
\mathcal{L}^u f(t,x) = V_t(t,x) + \big(D_x f(t,x)\big)^\top b(t,x,u) + \frac{1}{2} tr\big(\big(D_{xx} V(t,x)\big) a(t,x,u)\big)
$$

Meaning we ultimately end up with:

$$
0 = \sup_{\mathcal{u}\epsilon\mathcal{U}}[\psi(t,x,u) + \mathcal{L}^u V(t,x)]
$$

We refer to this equation as the **Hamilton-Jacobi-Bellman equation** (HJB). With this, we derived a necessary condition for the value function V.

MODEL

We continue the work of [Karatzas and Scheve, 1998] and [Steele, 2001] to model an initial market. The initial market model will consider a simple portfolio, where the investor has a base initial wealth which can be partitioned and invested into a market with two separate types of investment assets. A model like the one being constructed is called the **Black-Scholes model**. One type of investment will be a single low-risk bond with evolving prices which can be denoted by

$$\mathrm{d}B_t = B_t r \mathrm{d}t$$

$$\text{where } B_0 = 1,$$

$$\text{meaning } B_t = e^{rt}.$$

The other type of investment will be a stock with fluctuating prices

$$\mathrm{d}S_t = Diag(S_t)(\mu \mathrm{d}t + \sigma \mathrm{d}W_t)$$

$$\text{where } S_0 = s_0 > 0,$$

$$\text{with drift } \mu \in \mathbb{R}, \text{volatility } \sigma > 0,$$

and W is an m-dimensional Brownian motion. We can then use Itô's formula to derive the explicit solution to the stochastic differential equation:

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

We can also point out that this type of process is also said to follow a geometric Brownian motion and follows constant drift.

Remark: Considering the definition, $S_t = S_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t\right)$, which is dependent on W_t, with the strength of the stochastic Picard-Lindelöf theorem, S_t is considered an Itô process. This

means that with a probability of 1, the path is continuous with finite variance at time t provided S_0 has a finite variance. Like with the definition of Brownian motion, although not practically relevant in this thesis, we note that the variance of S_t goes to infinity as t goes to infinity, unless $\mu < \frac{\sigma^2}{2}$.

If the investor begins with initial wealth x_0, then the wealth at time t can be said to be modeled by:

$$\mathrm{d}X_t = N_t^B \mathrm{d}B_t + N_t^S \mathrm{d}S_t$$

$$\text{where } X_0 = x_0 > 0.$$

Here, N_t^B *and* N_t^S represent the possibly fractional number of assets (bonds and stocks respectively) that are held by the investor at time *t*. N_t^B *and* N_t^S are non-negative and we assume that the wealth is also always non-negative, meaning

$$X(t) \geq 0, \text{ for } 0 < t < T.$$

Now that we have a general procedure as to how to build a market model, we introduce utility which allows us to find the optimal market allocation that will maximize the investor's satisfaction.

Utility

Utility is an essential concept to define in Merton's portfolio analysis. General economic utility is defined as the satisfaction gained from taking an action or consuming a product. In our case, utility represents the personal satisfaction of an investor from the outcome of their investment. This can be represented by utility functions that describes the pattern of utility gained from consuming wealth.

To begin constructing utility functions, we need define several assumptions. First, we shall assume that the investor is risk-averse, meaning the investor will act on investments which are in his favor, choosing a more predictable low-return investment over a risky high-return investment. This implies that we are looking at a strictly concave utility function, since the utility of an action will outweigh the demand of the action. Secondly, we assume that the investor will always benefit