Theory and Estimation of Drift Dynamics of Asset Price Process

A Thesis

submitted to
Indian Institute of Science Education and Research Pune
in partial fulfillment of the requirements for the
BS-MS Dual Degree Programme

by

Aakash Verma

Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

April, 2018

Supervisor: Dr. Anindya Goswami
© Aakash Verma 2018

All rights reserved

Certificate

This is to certify that this dissertation entitled Theory and Estimation of Drift Dynamics of Asset Price Process towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Aakash Verma at Indian Institute of Science Education and Research under the supervision of Dr. Anindya Goswami, Assistant Professor, Department of Mathematics , during the academic year 2017-2018.

Dr. Anindya Goswami

Committee:

Dr. Anindya Goswami

Prof. Uttara Naik-Nimbalkar

This thesis is dedicated to my beloved family.

Declaration

I hereby declare that the matter embodied in the report entitled Theory and Estimation of Drift Dynamics of Asset Price Process are the results of the work carried out by me at the Department of Mathematics, Indian Institute of Science Education and Research, Pune, under the supervision of Dr. Anindya Goswami and the same has not been submitted elsewhere for any other degree.

Aakash Verma

Acknowledgments

I would like to thank all the people who supported and helped me drive my way through to the completion of the thesis. First and foremost, I would like to immensely thank Dr. Anindya Goswami, for encouraging and showing me ways to articulate mathematical ideas at every step of my project with his expertise. He invested a significant amount of time in my project, and without him, this thesis would not have been possible. I would also like to extend my gratitude towards my committee member Prof. Uttara-Naik Nimbalkar for showing interest and helping me throughout the project and helping me in developing some of the techniques which helped me during the entire project. I would like to thank Milan Kumar Das who provided me endless support for my queries. Also, immense thank goes to Bhaskara Lab for providing excellent computational facilities without which I wouldn't have achieved conclusive results and also thanks to my institute IISER Pune for providing me excellent facilities and the opportunity to learn a great deal of science.

I would like to thank my parents who provided me much needed support; my elder brother who always has been my motivation, my friends for their guidance and support throughout my journey in IISER.

I will remain thankful to God for keeping me healthy and motivated to complete the project.

Abstract

This project aims at investigating the behaviour of drift dynamics in financial models. Earlier models used geometric Brownian motion (GBM) for capturing the behaviour of stock prices. In 1973, Black, Scholes and Merton used the GBM model [5] for pricing European style options and also gave a mathematical formula for the same. The flaw of this model was it took several market parameters as volatility, drift, interest rate as a constant which need not be the case with the actual market. Although this model was a flawed one, it provided the much-needed base required in modelling the market and motivated further research. To rectify these shortcomings, after a significant amount of research in the field of mathematical finance, several models were proposed and one such, the Markov modulated GBM model which takes the market parameters as a function of a Markov chain which evolves according to a specific transition rate. This opens a wide range of possibilities and ideas for research.

This project deals with surveying several of the key process involved in these models and simulating several of them. Also, several numerical experiments related to the theoretical results were carried out and analyzed. The problem of estimating the drift parameter in these models was investigated using theoretical calculations and experimenting them numerically.

Contents

Introduction

In the classical model of stock prices by Black-Scholes-Merton (BSM), which is assumed to follow Geometric Brownian Motion, the drift and the volatility of the prices are held constant. However, in reality, this is not the case, the empirical volatility varies over time. In regime switching model, it is assumed that the market has finitely many hypothetical observable economic states and those are realized for certain random intervals of time. In particular, the volatility is assumed to depend on those regimes or states and the state transitions are modelled by a pure jump process. The Market model with finite-state Markov regime is a very popular choice.

In comparison with Markov switching, the study of semi-Markov (SM) regime switching is relatively uncommon. In this type of models, one has an opportunity to incorporate some memory effect of the market. In particular, the knowledge of past stagnancy period can be fed into the option price formula to obtain the price value. Hence this type of models has greater appeal regarding applicability than the one with Markov switching.

It is also shown in the literature that by considering such kind of general model, many drawbacks of original Black-Scholes-Merton model can be fixed. While fixing the deficits, it also retains the mathematical tractability. More precisely, the price function still solves a parabolic PDE which can be computed numerically. Despite that, for a real-life application, one must know the values of the coefficients which appear in the equation. Even for Black-Scholes-Merton model, this is a challenge. For Markov and semi-Markov modulated GBM (MMGBM and SMGBM respectively), the PDE involves a few more parameters which should be estimated from the market data.

In this thesis, we put our concern on the parameter, called drift coefficient and its transition kernels of the related drift process, which is a Markov chain. Although these do not

appear in the option price equation for regime switching GBM (MMGBM) but do appear for regime switching Heston or Geometric Levy process. However, in this thesis, we do not discuss these models and restrict ourselves to regime switching MMGBM only. The main reason for this is that the exposition with restriction is within the scope of a time-bound masters project and the study or result might find immediate application for more realistic and complicated models.

Since in practice the regimes are not observed, the above estimation problem becomes rather involved. Therefore, before attacking this real-life problem, we revisit the standard theory of estimating transition parameter of an observed discrete-time Markov chain.

In above regards, we first look at being able to draw statistical inferences about the transition probabilities obtained from the unbroken observations of Markov chains. For this purpose, we look at Billingsley [3], and revisit the proof of important results from it and use it to draw inference on our numerical experiment done on the simulation of discrete-time Markov chain. More about the general aspects of statistical inferences from Markov chain can be found in Grenander [11], Bartlett [2]. Further, we investigate upon continuous-time Markov chain and semi-Markov processes and simulate them.

Now we return to the original problem. In the regime switching MMGBM one observes only the stock price which is driven by a Markov chain and Brownian motion. Thus in a sense, one can at most obtain a noisy observation of continuous-time Markov chain with Gaussian noise. To understand the mathematical properties of such noise, we look in detail at the chapter on the Brownian motion.

On the detailed study of Brownian motion, we revisit the definition and mathematical properties of it. Also, alongside we studied geometric Brownian motion (GBM) and simulated them using Box-Muller transform for generating normal random variables. Finally, we present the mathematical details of the regime-switching model alluded above. Further, we study several models which help in defining the main problem. We do not get into much details of those and just look at their option pricing equations. We then look at methods to simulate regime switching Markov modulated GBM. We use this simulation in further numerical experiments to analyze the main problem.

To investigate the original problem, we restrict ourselves to binary regime case, or in other words, we look at only 2-state Markov chain case. Henceforth, we look to validate

our model of regime switching MMGBM. Then we try to succinctly answer the problem of estimating the scale parameter coming from holding time distribution of continuous-time Markov chain. Similar accounts can be found in [9]. Further, the estimation of drift dynamics has 4 parameters namely, $\theta_1, \theta_2, \mu_1, \mu_2$ which are estimated using the knowledge of $\bar{\mu}$ which is the time-series of partial observations of μ and is available to us. Then we discuss different scenarios where the estimator has reasonably small Standard Error and also taking into the picture the other case of large Standard Error. Then we estimate the scale parameter on a live data. At last, we use the method of antithetic variates which is useful for reducing the variance by inducing negatively correlated random variables and also helpful in reducing the number of iterations used for simulation. We hope that these estimators and their variance reductions would be effective even for the larger class of models including Heston, Jump Diffusion models, and many more financial models.

Chapter 1

Preliminaries

In this chapter, we are going to look into theorems concerning generating random variables for the case when CDF might not be continuous and which will be relevant for the simulations to follow. However, before that, we will take a look at the definition of rcll and upper semicontinuous function useful for the theorems to follow. We have produced new proofs of these theorems using real analysis and study of the Royden [15].

Definition 1.0.1 (rcll Function). *Let (M,d) be a metric space and let $E \subseteq \mathbb{R}$. A function $f : E \to M$ is called a rcll function, if $\forall\ t \in E$,*

- *the left limit $f(t^-) = \lim_{x \uparrow t} f(x)$ exists and*
- *the right limit $f(t^+) = \lim_{x \downarrow t} f(x)$ exists and equals to $f(t)$.*

In other words, f is right-continuous with left limits.

Using above definition, it is easy to see that all continuous functions are rcll and so are all cumulative distribution functions. Now let us look forth to the definition of upper semicontinuous function.

Definition 1.0.2 (Upper Semicontinuous Function). *Suppose X is a topological space, then the function $G : X \to \mathbb{R} \cup \{-\infty, \infty\}$ is said to be upper semicontinuous if $\forall\ \alpha \in \mathbb{R}$, $G^{-1}([-\infty, \alpha)) = \{x \in X \mid G(x) < \alpha\}$ is an open set in X.*

Now using above definitions, we state and prove some theorems as under.

Theorem 1.0.1. *If G is a upper semicontinuous on $\mathbb{R}$ and a non-decreasing function then $G^{-1}([a,b)) = [c,d)$ for some real c, d.*

Proof. We have the following assertion as a consequence of G being upper semicontinuous and non-decreasing.

$$\begin{aligned} G^{-1}([a,b)) &= G^{-1}((-\infty,b) \cap (-\infty,a)^c) \\ &= G^{-1}((-\infty,b)) \cap (G^{-1}(-\infty,a))^c \\ &= (-\infty,d) \cap (-\infty,c)^c \\ &= [c,d) \end{aligned}$$

for some real c, d. Hence proved. □

Now we prove an essential theorem relating to semicontinuity. We elaborate more on this by stating and proving it as under.

Theorem 1.0.2. *If $F : \mathbb{R} \to \mathbb{R}$ is non-decreasing then $F^{\leftarrow}(u) : \mathbb{R} \to \mathbb{R} \cup \{-\infty, \infty\}$ defined by $F^{\leftarrow}(u) = \sup\{x \mid F(x) \leq u\}$ is non-decreasing and upper semicontinuous.*

Proof. By definition we have $F^{\leftarrow}(u) = \sup\{x \mid F(x) \leq u\}$ which implies $F^{\leftarrow}$ is non-decreasing. Showing the upper semicontinuity is essentially the same as showing the set $\mathbf{U} := \{u \mid F^{\leftarrow}(u) < \lambda\}$ is open. Now let $u_0 \in \mathbf{U}$, which means that it satisfies $\sup\{x \mid F(x) \leq u_0\} < \lambda^{'} < \lambda$ (for some $\lambda^{'}$ (by Archimedean property)). Thus if for some x, $F(x) \leq u_0$, then $x \leq F^{\leftarrow}(u_0) < \lambda^{'}$. Therefore, $F(\lambda^{'}) > u_0$. This has further implication that $\exists\ u_1$ such that $u_0 < u_1 < F(\lambda^{'})$ (again by Archimedean property). From the definition of $\mathbf{U}$, it is evident that $\mathbf{U}$ is either of the form $(-\infty, \beta)$ or $(-\infty, \beta]$ for some $\beta \leq \infty$ (if $F^{\leftarrow}$ is bounded and $\lambda \geq \sup F^{\leftarrow}(u)$) then $\beta = \infty$). Hence to show $\mathbf{U}$ is open, given a $u_0 \in \mathbf{U}$, it is enough to produce a $u_1 (> u_0)$ that lies in $\mathbf{U}$. We claim that above u_1 serves the purpose. Indeed, now only remaining bit is to check if this $u_1 \in \mathbf{U}$. We can prove this part as follows. Now assume, if possible $\lambda^{'} < F^{\leftarrow}(u_1)$ which implies $F(\lambda^{'}) \leq u_1$, but we already know from the earlier part of the proof that by our choice $u_1 < F(\lambda^{'})$. Therefore we have arrived at the contradiction. Hence, we must have $F^{\leftarrow}(u_1) \leq \lambda^{'}$, which will imply that $F^{\leftarrow}(u_1) \leq \lambda^{'} < \lambda$. Thus $u_1 \in \mathbf{U}$. Hence proved the function $F^{\leftarrow}(u)$ is non-decreasing and upper semicontinuous. □

Now we will state and prove the Theorem when CDF might not be continuous.

Theorem 1.0.3. *Let $X \sim F_X$, i.e. F_X is the CDF of a random variable X, and F_X need not be continuous or injection. Then*

$$Y := F^{\leftarrow}(U) \sim F_X,$$

where $F^{\leftarrow} : [0,1] \to \mathbb{R} \cup \{-\infty, \infty\}$ given by $F^{\leftarrow}(u) = \sup\{x \mid F_X(x) \leq u\} \in \mathbb{R} \cup \{-\infty, \infty\}$ and $U \sim U[0,1]$, the uniform distribution on the closed interval [0,1].

Proof. Now $F^{\leftarrow}(u) \in \mathbb{R} \ \forall \ u \in (0,1)$. So, $F^{\leftarrow}(U) \in \mathbb{R}$ (almost surely). Since F_X is upper semicontinuous and so $F^{\leftarrow}$ is rcll and non-decreasing function. Therefore we have $(F^{\leftarrow})^{-1}([a,b)) = [c,d)$ for some real c, d if $[a,b) \subset (0,1)$. Now if $[c,d) = (F^{\leftarrow})^{-1}([a,b))$, we have the following consequences which follow trivially from the above arguments,

i. $F^{\leftarrow}(c) \geq a$, $F^{\leftarrow}(c^-) \leq a$,

ii. $F^{\leftarrow}(d^-) \leq b, F^{\leftarrow}(d) \geq b$,

iii. $d = \sup\{u \mid \sup\{x \mid F_X(x) \leq u\} \leq b\}$,

iv. $c = \inf\{u \mid \sup\{x \mid F_X(x) \leq u\} \geq a\}$,

Now we claim that the following equalities hold,

$$F_X(b) = d, \tag{1.1}$$

$$F_X(a) = c. \tag{1.2}$$

We will prove only equation (1.1) as the proof of equation (1.2) is analogous. Let if possible, $F_X(b) < d$. Then using the right continuity of F_X, $\exists\ \epsilon > 0$ such that $F_X(b+\epsilon) < d$. Hence from (iii) $\exists\ u$ with $F_X(b+\epsilon) < u \leq d$. From (iii) and monotonicity of $F^{\leftarrow}$, $u \leq d$ implies $F^{\leftarrow}(u) \leq b$. However, $F^{\leftarrow}(u) \leq b$ as $F_X(b+\epsilon) < u$, $b + \epsilon \leq F^{\leftarrow}(u)$. So we have the following relation as a result $F^{\leftarrow}(u) \leq b < b + \epsilon \leq F^{\leftarrow}(u)$. Hence we arrive at a contradiction. Now let us consider the other case. If possible, assume $d < F_X(b)$. Then $\exists\ u^{'}$ such that $d < u^{'} < F_X(b)$. If $e := \sup\{x \mid F_X(x) \leq u^{'}\}$, then $F_X(e^-) \leq u^{'} < F_X(b)$. Therefore, from monotonicity of F_X, $F^{\leftarrow}(u^{'}) \leq e \leq b$ and $\sup\{x \mid F_X(x) \leq u^{'}\} \leq b$. Hence using (iii)

we have the following relation as result $d < u^{'} \leq \sup_{u}\{F^{\leftarrow}(u) \leq b\} = d$, a contradiction. Therefore, we must have $F_X(b) = d$. Hence (1.1) is true as claimed.

Thus using equation (1.1) and equation (1.2), we obtain the following relation.

$$\begin{aligned}
\mathbb{P}(Y \in [a,b)) &= \mathbb{P}(G(U) \in [a,b)) \\
&= \mathbb{P}(U \in (F^{\leftarrow})^{-1}([a,b))) \\
&= \mathbb{P}(U \in [F_X(a), F_X(b))) \\
&= F_X(b) - F_X(a).
\end{aligned}$$

Hence $Y \sim F_X$.

□

Chapter 2

Simulation and Statistical Inference on Markov Chains

In this chapter, we are going to discuss the essential concepts of the Markov chain and the methods to simulate them. We will discuss more upon the issue of simulating the Discrete-time, Continuous-time Markov chains, as well as look at the Semi-Markov processes. We will also look at the sequence generated from our algorithm of Discrete-time Markov chain and try to obtain an Empirical Probability Matrix (EPM) and look for the results related to its convergence to Transition Probability Matrix (TPM). We will see some critical theorems which will shape the course of this chapter and will motivate our study of it.

2.1 Historical Development of Markov Chain

To find the historical roots of Markov chain an extensive literature survey regarding the life of Andrey Markov and the origin of Markov chain was done. So starting with the development of these chains to their use in modern day world will motivate us in our research on this topic. Let us discuss the brief overview of this historical development.

To begin with, there were several historical events that shaped the development of Markov chain. The discovery of these chains was not accidental but the outcome of several years of

research and thoughts that Markov had put while reading through Russian literature. He found out and recorded the patterns in vowels and consonants as he read the literature. The methods used by Markov went beyond the realms of coin-flipping and dice-rolling (where each event is independent of the other) to chains of linked events (where what happens next depends on the current state of the system). If we consider the modern day scenario, the discovery made by Markov has not gone obsolete instead it has become more vital fuel for research. We discuss the importance of Markov chains with the help of many events. The methods and techniques developed by Markov help in identifying genes in DNA, powering algorithms for voice recognition, contextual computing, web search used by famous Tech-Companies such as Google (using Page-rank algorithm which incorporates the idea of Markov chains), Microsoft and many others to name. In other fields of Science, such as Physics it helps in simulating and analysing the collective behaviour of particles in a system. In Statistics, it is used for the process of withdrawing selected samples when we have a large set of possibilities. Even the current most popular field of Machine Learning, Data Science, Neuroscience use the Markov chains extensively. Hence it has become an exciting part of research more than ever before to understand how and where these chains work.

In the context of above discussion, we are motivated to do some significant amount of statistical analysis and research in the related field.

2.2 Simulation of Discrete-Time Markov Chain

In this section our main aim will be to discuss the concept of Discrete-time Markov chains in details and further detailed analysis will be done. In this section, we will deal with discrete-time, finite state Markov chain. In particular, we will state the definition of Stationary Markov Chain and the method to simulate it.

Definition 2.2.1. *A stochastic process* $\{ X_t\colon t \geq 0\}$ *is called a stationary Markov Chain if for all times* $n \geq 0$ *and all states* $i_0, i_1, \cdots, i_k, \cdots, i_{n-1}, i_n$, $i_k \in S$, *we have:*

$$\mathbb{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \cdots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j | X_n = i) = \mathbb{P}_{ij}. \tag{2.1}$$

Here $\mathbb{P}_{ij}$ gives the probability that the chain, whenever in state i, moves next (one unit of time later) into state j, and is referred to as transition probability. The square matrix

$P = (\mathbb{P}_{ij})$, where $i, j \in S$ is called the Transition Probability Matrix **(TPM)** . Since, when leaving the state i the chain must move to one of the state $j \in S$, each row of P sums to one (e.g., forms a probability distribution): For each i

$$\sum_{j \in S} \mathbb{P}_{ij} = 1.$$

Now moving on from the definition and dealing with the computational aspects of the DTMC (we will refer to Discrete-Time Markov Chain in short as DTMC wherever possible). We used the below-mentioned algorithm for simulating the DTMC, and we ran the code in python language to generate the samples. We describe the algorithm as follows:

Algorithm 1 To simulate Discrete-time Markov Chain up to first N steps given a TPM, initial state, state space $= \{1, \cdots, k\}$

1. Choose an initial state, $x_0 = i$, with probability $\mathbb{P}_i$, $i \in S$.
2. for($n = 1$ to N)
   generate $u \sim U[0, 1]$ independently.
   $x_n = F(u, x_{n-1})$.
   end.

The output for the above algorithm is $\{x_0, \cdots, x_N\}$, also $U[0, 1]$ denotes (Uniform distribution on $[0, 1]$) and the Function $F : [0, 1] \times S \to S$ is given by:

$$F(u, i) = \begin{cases} 0, & \text{if } u \leq \mathbb{P}_{j_0} \\ 1, & \text{if } \mathbb{P}_{j_0} < u \leq \mathbb{P}_{j_0} + \mathbb{P}_{j_1} \\ \vdots & \\ k-1, & \text{if } \sum_{m=0}^{k-2} \mathbb{P}_{j_m} < u \leq \sum_{m=0}^{k-1} \mathbb{P}_{j_m} (= 1). \end{cases}$$

The function $F(u, i)$ can also be viewed in terms of quantile function and significantly $F(u, i) = F^{\leftarrow}_{(X_n | X_{n-1})}(u) = i$ where $F^{\leftarrow}(u) = \sup\{x \mid F(x) \leq u\}$ or alternatively we have $F^{\leftarrow}(u) = \inf\{x \mid F(x) \geq u\}$ as they are essentially same except on measure zero set.

We can illustrate the behaviour of the above defined Function in the form of a figure 2.1 as below. The horizontal axis has consecutive intervals of lengths $\mathbb{P}_{j_0}$ to $\mathbb{P}_{j_{k-1}}$.

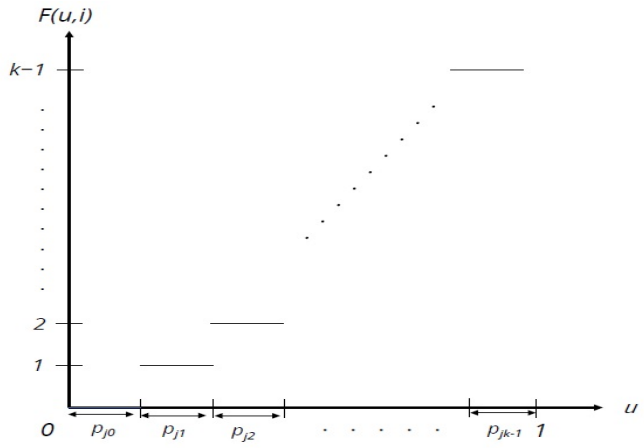


Figure 2.1: $F(u,i)$ vs u for fixed i.

An important mistake was observed in one of the books [6], the author has incorrectly used the quantile function in his simulation of random variable. He has defined quantile function in terms of *min* function which has a problem if we look at a step function.

Given a TPM Matrix 'P', the above algorithm produces a random sequence in S. It is easy to see from the definition of F, that

$$\mathbb{P}(x_n = i^{'} \mid x_{n-1} = i, x_{n-2} = i_{n-2}, \cdots, x_0 = i_0) = \mathbb{P}(F(u,i) = i^{'}) = \mathbb{P}_{ii'}.$$

Thus we say that Algorithm 1 produces a Markov chain with a given TPM Matrix. The empirical verification of the Markov chain is possible provided the chain is sufficiently large.

We define an Empirical Probability Matrix **(EPM)** as below.

Definition 2.2.2. *Given a sequence of Markov chain simulated up to n steps. An empirical probability matrix is*

$$\hat{\mathbb{P}}_{ij} = \frac{f_{ij}}{f_i},$$

where f_{ij} is the number of instantaneous transitions from i to j; f_i is the number of exits from state i. Hence $\left(\hat{\mathbb{P}}_{ij}\right)_{i,j\in S}$ is a probability matrix.

We will encounter the above terms in the upcoming sections as well where we will mean the same.

Now we state an important result regarding the convergence of EPM (obtained from a

sequence of DTMC) to TPM.

2.3 Theorem on the convergence of EPM:

We must discuss the convergence of EPM before talking about the rate of convergence. A decent amount of time was spent in understanding the Theorem from research paper [3] by Billingsley. Here in this part of the section, we will be stating that important theorem, and we will reproduce the proof in a detailed manner.

2.3.1 Preliminaries

First and foremost, let us look at the statement of the Central Limit Theorem for Multidimension which will be used in the proof to follow and then we will look at several definitions which help in understanding the Theorem and further the subsequent discussion.

We now state Central Limit Theorem for Multidimension.

Theorem 2.3.1 (**Central Limit Theorem for Multidimension).** *Consider X_i is a random vector in $\mathbb{R}^k$, with mean vector $\mu = E(X_i)$ and co-variance matrix $\sum$(amongst the components of the vector), and these random vectors are independent and identically distributed while summation is being done component-wise. The Central limit theorem states that when scaled, sums converge to a multivariate normal distribution. Let*

$$\boldsymbol{X_i} = \begin{bmatrix} X_{i(1)} \\ X_{i(2)} \\ \vdots \\ X_{i(k)} \end{bmatrix} \tag{2.2}$$

be the k-vector. The $\boldsymbol{X_i}$ means that it is a random vector not random variable. So now sum of these random vectors as mentioned before component-wise is

$$\begin{bmatrix} X_{1(1)} \\ X_{1(2)} \\ \vdots \\ X_{1(k)} \end{bmatrix} + \begin{bmatrix} X_{2(1)} \\ X_{2(2)} \\ \vdots \\ X_{2(k)} \end{bmatrix} + \cdots + \begin{bmatrix} X_{n(1)} \\ X_{n(2)} \\ \vdots \\ X_{n(k)} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n}[X_{i(1)}] \\ \sum_{i=1}^{n}[X_{i(2)}] \\ \vdots \\ \sum_{i=1}^{n}[X_{i(k)}] \end{bmatrix} = \sum_{i=1}^{n} \boldsymbol{X_i}$$

and the average is

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X_i} = \frac{1}{n}\begin{bmatrix} \sum_{i=1}^{n}[X_{i(1)}] \\ \sum_{i=1}^{n}[X_{i(2)}] \\ \vdots \\ \sum_{i=1}^{n}[X_{i(k)}] \end{bmatrix} = \begin{bmatrix} \bar{X}_{i(1)} \\ \bar{X}_{i(2)} \\ \vdots \\ \bar{X}_{i(k)} \end{bmatrix} = \bar{\boldsymbol{X}}_{\boldsymbol{n}}$$

and therefore

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\boldsymbol{X_i} - E(X_i)] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\boldsymbol{X_i} - \mu) = \sqrt{n}(\bar{\boldsymbol{X}}_{\boldsymbol{n}} - \mu).$$

The multidimensional central limit theorem states that

$$\sqrt{n}(\bar{\boldsymbol{X}}_{\boldsymbol{n}} - \mu) \xrightarrow{D} N_k(0, \Sigma),$$

where the co-variance matrix Σ equals to

$$\Sigma = \begin{bmatrix} Var(X_{1(1)}) & Cov(X_{1(1)}, X_{1(2)}) & Cov(X_{1(1)}, X_{1(3)}) & \cdots & Cov(X_{1(1)}, X_{1(k)}) \\ Cov(X_{1(2)}, X_{1(1)}) & Var(X_{1(2)}) & Cov(X_{1(2)}, X_{1(3)}) & \cdots & Cov(X_{1(2)}, X_{1(k)}) \\ Cov(X_{1(3)}, X_{1(1)}) & Cov(X_{1(3)}, X_{1(2)}) & Var(X_{1(3)}) & \cdots & Cov(X_{1(3)}, X_{1(k)}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Cov(X_{1(k)}, X_{1(1)}) & Cov(X_{1(k)}, X_{1(2)}) & Cov(X_{1(k)}, X_{1(3)}) & \cdots & Var(X_{1(k)}) \end{bmatrix}.$$

Now let us have a look at some of the definitions to enhance our understanding of the Theorem to follow.

Definition 2.3.1 (**Irreducible Markov Chain**). *A Markov Chain is said to be* ***irreducible*** *if it is possible to reach state j from every other state i.*

Definition 2.3.2 (**Periodicity**). *A state i is said to have period m if any return to the state i occurs in multiple of m steps. More formally,*

$$m = gcd\{n > 0 : \mathbb{P}(X_n = i \mid X_0 = i) > 0\}.$$

Here 'gcd' is the greatest common divisor provided the set is not empty. If $m = 1$ then the state is said to be aperiodic.

Definition 2.3.3 (**Transience**). *A state i said to be transient, if given that chain starts in state i, there is a non-zero probability that the chain will never return to i.*

Now we will take the discussion little further and define some terms. If we let the random variable T_i be the first return time to state i:

$$T_i = \inf\{n \geq 1 : X_n = i : X_0 = i\}.$$

$r_{ii}^{(n)} = \mathbb{P}(T_i = n)$ is the probability of returning to state i after n steps. Therefore, by the definition of transience we have a state i is transient if

$$\mathbb{P}(T_i < \infty) = \sum_{n=1}^{\infty} r_{ii}^{(n)} < 1.$$

Definition 2.3.4 (**Recurrence**). *The recurrence is defined as the negation of transience which means, a state i is recurrent if it is not transient.*

Now we will see the meaning of mean recurrence time. As by name, the mean recurrence time at the state i is the expected return time M_i defined as below.

$$M_i = E[T_i] = \sum_{n=1}^{\infty} n r_{ii}^{(n)}.$$

Definition 2.3.5 (**Positive Recurrence**). *A state i is positive recurrent if M_i is finite.*

Definition 2.3.6 (**Null-Recurrent**). *A state i is said to be null-recurrent if M_i is not finite.*

In other words, if a recurrent state i is not positive recurrent then it is null-recurrent.

Definition 2.3.7 (Ergodicity). *A state i is said to be ergodic if it is aperiodic and positive recurrent.*

Also an irreducible Markov chain is ergodic if it has an aperiodic state. Now we have covered all the definitions necessary to understand the theorem but still some of the terms may be unfamiliar so we will define them as under. We will also discuss the method of sampling used in the theorem briefly.

We consider $\{x_n\}_{n=0}^{N}$ to be the sample from the simulated Markov chain with transition probabilities $\mathbb{P}_{ij}$ and initial probabilities $\mathbb{P}_i$. If $\{a_0, a_1, \cdots, a_N\}$ is a sequence of $N+1$ states then the probability that $x_0, x_1, \cdots, x_N$ is this sequence is $\mathbb{P}_{a_0}\mathbb{P}_{a_0a_1}\cdots\mathbb{P}_{a_{N-1}a_N}$. For $i, j = 0, 1, \cdots, k-1$, let f_{ij} be the number of m, with $1 \leq m \leq N$ for which $a_m = i$ and $a_{m+1} = j$; meaning it is the number of instantaneous transition from i to j.

We define $f_{i.} = \sum_j f_{ij}$ and $f_{.j} = \sum_i f_{ij}$, where $\{f_{i.}\}$ and $\{f_{.j}\}$ are the frequency counts of $\{a_0, \cdots, a_{N-1}\}$ and $\{a_1, \cdots, a_N\}$ respectively from which it follows:

$$f_{i.} - f_{.i} = \delta_{ia_0} - \delta_{ia_N},$$

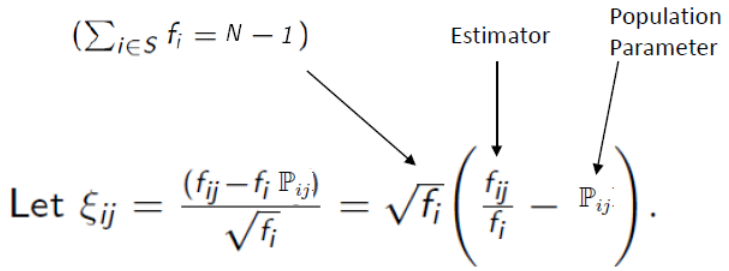
$$\sum_{ij} f_{ij} = \sum_i f_{i.} = \sum_j f_{.j} = N.$$

Here $f_{i.}$ denotes the number of exits from state i, and $f_{.i}$ denotes the number of entrances to state i.

For the purpose of simpler notation, in the next section we will denote '$f_{i.}$' by 'f_i'.

2.3.2 Main Result:

Now let us state the **Theorem 3.1** verbatim from research paper by Billingsley [3]. Here, S will denote state space and s will denote its cardinality.



Theorem 2.3.2 (Billingsley, 1961). *In the stationary, ergodic case, the distribution of the $s \times s$ dimensional matrix $\xi = \xi_{ij}$ converges as $N \to \infty$ to the normal distribution with mean 0 and the co-variance matrix $(\lambda_{ij,kl})$, where*

$$\lambda_{ij,kl} = \delta_{ik}(\delta_{jl}\mathbb{P}_{ij} - \mathbb{P}_{ij}\mathbb{P}_{il}).$$

Before proving the above Theorem we are going to discuss the following theorems which form the essential pillars of the proof to follow. While most of these theorems have been proved but some of them have only been stated to get the intended result.

We will now look forward to state a Lemma which provides us with tools in proving the Theorem 2.3.2.

Lemma 2.3.3. *Assume that the chain is stationary and ergodic and let $\zeta = (\zeta_1, \cdots, \zeta_s)$ be the random vector with components*

$$\zeta_i = \frac{f_i - N\mathbb{P}_i}{N^{\frac{1}{2}}}.$$

Then

$$\begin{cases} E\{\zeta_i\} = 0 \\ E\{\zeta_i \zeta j\} = \alpha_{ij} + o\left(\frac{1}{N}\right) \end{cases} \tag{2.3}$$

where

$$\alpha_{ij} = \delta_{ij}\mathbb{P}_i - \mathbb{P}_i\mathbb{P}_j + \mathbb{P}_i \sum_{m=1}^{\infty}(\mathbb{P}_{ij}^{(m)} - \mathbb{P}_j) + \mathbb{P}_j \sum_{m=1}^{\infty}(\mathbb{P}_{ji}^{(m)} - \mathbb{P}_i). \tag{2.4}$$

Moreover, the weak law of large number holds:

$$p \lim_{N\to\infty} \frac{f_i}{N} = \mathbb{P}_i. \tag{2.5}$$

Proof. To prove equation (2.3), we define the random variable $c_m(i)$ as:

$$c_m(i) = \begin{cases} 1, \textit{if } x_m = i, \\ 0, \textit{if } x_m \neq i. \end{cases}$$

Then $f_i = \sum_{m=1}^{N} c_m(i)$. From the stationarity of the chain we have,

$$E\{c_m(i)\} = 0(\mathbb{P}(x_m \neq i)) + 1(\mathbb{P}(x_m = i) = \mathbb{P}_i).$$

so ,

$$E\{\zeta_i\} = \frac{E(f_i) - N\mathbb{P}_i}{N^{\frac{1}{2}}} = \frac{N\mathbb{P}_i - N\mathbb{P}_i}{N^{\frac{1}{2}}} = 0.$$

Now, we will prove the other part of equation (2.3) as under,

$$\begin{aligned} E\{\zeta_i\zeta_j\} &= E\left[\left(\frac{f_j - N\mathbb{P}_j}{N^{\frac{1}{2}}}\right)\left(\frac{f_i - N\mathbb{P}_i}{N^{\frac{1}{2}}}\right)\right] \\ &= E\left[\frac{(f_j - N\mathbb{P}_j)(f_i - N\mathbb{P}_i)}{N}\right] \\ &= E\left[\frac{(\sum_{l=1}^{N} c_l(j) - N\mathbb{P}_j)(\sum_{m=1}^{N} c_m(i) - N\mathbb{P}_i)}{N}\right] \\ &= N^{-1}E\left[\sum_{l=1}^{N}(c_l(j) - \mathbb{P}_j)\sum_{m=1}^{N}(c_{m(i)} - \mathbb{P}_i)\right] \\ &= N^{-1}\sum_{l=1}^{N}\sum_{m=1}^{N} E\{(c_l(i) - \mathbb{P}_i)(c_m(j) - \mathbb{P}_j)\}. \end{aligned}$$

Hence,

$$E\{\zeta_i\zeta_j\} = N^{-1}\sum_{l=1}^{N}\sum_{m=1}^{N} E\{(c_l(i) - \mathbb{P}_i)(c_m(j) - \mathbb{P}_j)\}. \tag{2.6}$$

Again using stationarity we have,

$$E\{(c_l(i) - \mathbb{P}_i)(c_m(j) - \mathbb{P}_j)\} = \mathbb{P}\{x_l = i, x_m = j\} - \mathbb{P}_i\mathbb{P}_j = \begin{cases} \mathbb{P}_i\mathbb{P}_{ij}^{(m-l)} - \mathbb{P}_i\mathbb{P}_j, & \textit{if } m > l \\ \mathbb{P}_i\delta_{ij} - \mathbb{P}_i\mathbb{P}_j, & \textit{if } m = l \\ \mathbb{P}_i\mathbb{P}_{ji}^{(l-m)} - \mathbb{P}_j\mathbb{P}_i, & \textit{if } m < l. \end{cases}$$

So

$$E\{\zeta_i\zeta_j\} = N^{-1}\sum_{m=1}^{N}\left\{\sum_{l=1}^{m-1}(\mathbb{P}_i\mathbb{P}_{ij}^{(m-l)} - \mathbb{P}_i\mathbb{P}_j) + (\mathbb{P}_i\delta_{ij} - \mathbb{P}_i\mathbb{P}_j) + \sum_{l=1}^{m+1}(\mathbb{P}_i\mathbb{P}_{ij}^{(m-l)} - \mathbb{P}_i\mathbb{P}_j)\right\}. \quad (2.7)$$

Resolving the first term inside bracket of equation (2.7) we have:

$$\sum_{l=1}^{m-1}(\mathbb{P}_i\mathbb{P}_{ij}^{(m-l)} - \mathbb{P}_i\mathbb{P}_j) = \mathbb{P}_i\sum_{l=1}^{m-1}\mathbb{P}_{ij}^{(m-l)} - (m-1)\mathbb{P}_i\mathbb{P}_j.$$

Now we evaluate:

$$\sum_{l=1}^{m-1}\mathbb{P}_{ij}^{(m-l)} = [\mathbb{P}_{ij}^{(1)} + \mathbb{P}_{ij}^{(2)} + \cdots \mathbb{P}_{ij}^{(m-1)}].$$

So

$$\begin{aligned} N^{-1}\sum_{m=1}^{N}\{\textit{first Term}\} &= N^{-1}\left(\sum_{m=1}^{N}\mathbb{P}_i\sum_{l=1}^{m-1}\mathbb{P}_{ij}^{(m-l)} - \sum_{m=1}^{N}\sum_{l=1}^{m-1}\mathbb{P}_i\mathbb{P}_j\right) \\ &= N^{-1}\mathbb{P}_i\left(\sum_{m=1}^{N}[\mathbb{P}_{ij}^{(1)} + \cdots + \mathbb{P}_{ij}^{(m-1)}] - \sum_{m=1}^{N}(m-1)\mathbb{P}_j\right). \end{aligned}$$

We note that the right hand side of the equations has two summations. We first consider the first of these summations and rearrange the terms so as to get simpler expression in the following manner:

$$\begin{aligned} m = 1 \quad &: \quad \mathbb{P}_{ij}^{(0)}(=0) \\ m = 2 \quad &: \quad \mathbb{P}_{ij}^{(1)} \\ m = 3 \quad &: \quad \mathbb{P}_{ij}^{(1)} + \mathbb{P}_{ij}^{(2)} \\ \vdots \quad &: \quad \vdots \\ m = N \quad &: \quad \mathbb{P}_{ij}^{(1)} + \cdots + \mathbb{P}_{ij}^{(N-1)}. \end{aligned}$$

It is now evident that

$$\sum_{m=1}^{N}[\mathbb{P}_{ij}^{(1)} + \cdots + \mathbb{P}_{ij}^{(m-1)}] = (N-1)\mathbb{P}_{ij}^{(1)} + (N-2)\mathbb{P}_{ij}^{(2)} + \cdots + \mathbb{P}_{ij}^{(N-1)} = \sum_{m'=1}^{N-1}(N-m')\mathbb{P}_{ij}^{(m')}.$$

Similarly second summation inside the bracket is resolved.

$$\sum_{m=1}^{N}(m-1)\mathbb{P}_j = \mathbb{P}_j \sum_{m=1}^{N}(m-1)$$
$$= \mathbb{P}_j\Big[0+1+\cdots+(N-1)\Big] = \sum_{m'=1}^{N-1}(N-m')\mathbb{P}_j.$$

Combining above evaluated terms we get first term of equation (2.7) as under:

$$N^{-1}\sum_{m'=1}^{N-1}(N-m')\Big(\mathbb{P}_i\mathbb{P}_{ij}^{(m')} - \mathbb{P}_i\mathbb{P}_j\Big).$$

Resolving second term of equation (2.7) we have:

$$N^{-1}\sum_{m=1}^{N}(\mathbb{P}_i\delta_{ij} - \mathbb{P}_i\mathbb{P}_j) = N^{-1}N(\mathbb{P}_i\delta_{ij} - \mathbb{P}_i\mathbb{P}_j)$$
$$= (\mathbb{P}_i\delta_{ij} - \mathbb{P}_i\mathbb{P}_j).$$

Resolving the third term inside bracket of equation (2.7) we have:

$$\sum_{l=m+1}^{N}(\mathbb{P}_j\mathbb{P}_{ji}^{(l-m)} - \mathbb{P}_j\mathbb{P}_i) = \mathbb{P}_j\sum_{l=m+1}^{N}\mathbb{P}_{ji}^{(l-m)} - (N-m)\mathbb{P}_j\mathbb{P}_i.$$

Now we evaluate:

$$\sum_{l=m+1}^{N}\mathbb{P}_{ji}^{(l-m)} = \Big[\mathbb{P}_{ji}^{(1)} + \mathbb{P}_{ji}^{(2)} + \cdots \mathbb{P}_{ji}^{(N-m)}\Big].$$

So

$$N^{-1}\sum_{m=1}^{N}\{third\ \ Term\} = N^{-1}\left(\sum_{m=1}^{N}\mathbb{P}_j\sum_{l=m+1}^{N}\mathbb{P}_{ji}^{(l-m)} - \sum_{m=1}^{N}\sum_{l=m+1}^{N}\mathbb{P}_j\mathbb{P}_i\right)$$
$$= N^{-1}\mathbb{P}_j\left(\sum_{m=1}^{N}\Big[\mathbb{P}_{ji}^{(1)} + \cdots + \mathbb{P}_{ji}^{(N-m)}\Big] - \sum_{m=1}^{N}(N-m)\mathbb{P}_i\right).$$

We note that the right hand side of the equations has two summations. We first consider the first of these summations and rearrange the terms so as to get simpler expression in the

following manner:

$$\begin{aligned} m = N \quad &: \quad \mathbb{P}_{ji}^{(0)} (= 0) \\ m = N-1 \quad &: \quad \mathbb{P}_{ij}^{(1)} \\ m = N-2 \quad &: \quad \mathbb{P}_{ij}^{(1)} + \mathbb{P}_{ij}^{(2)} \\ \vdots \quad &: \quad \vdots \\ m = 1 \quad &: \quad \mathbb{P}_{ij}^{(1)} + \cdots + \mathbb{P}_{ij}^{(n-1)}. \end{aligned}$$

So we have,

$$\sum_{m=1}^{N} [\mathbb{P}_{ji}^{(1)} + \cdots + \mathbb{P}_{ji}^{(N-m)}] = (N-1)\mathbb{P}_{ji}^{(1)} + (N-2)\mathbb{P}_{ji}^{(2)} + \cdots + \mathbb{P}_{ji}^{(N-1)} = \sum_{m'=1}^{N-1} (N-m')\mathbb{P}_{ji}^{(m')}.$$

Similarly second summations inside the bracket simplifies as below:

$$\begin{aligned} \sum_{m=1}^{N} (N-m)\mathbb{P}_i &= \mathbb{P}_i \sum_{m=1}^{N} (N-m) \\ &= \mathbb{P}_i \Big[0 + 1 + \cdots + (N-1) \Big] = \sum_{m'=1}^{N-1} (N-m')\mathbb{P}_i. \end{aligned}$$

So third term of equation (2.7) becomes:

$$N^{-1} \sum_{m'=1}^{N-1} (N-m') \Big(\mathbb{P}_j \mathbb{P}_{ji}^{(m')} - \mathbb{P}_j \mathbb{P}_i \Big).$$

Combining all of the above we have:

$$\begin{aligned} E\{\zeta_i \zeta_j\} = N^{-1} \sum_{m'=1}^{N-1} (N-m') \Big(\mathbb{P}_i \mathbb{P}_{ij}^{(m')} - \mathbb{P}_i \mathbb{P}_j \Big) + (\mathbb{P}_i \delta_{ij} - \mathbb{P}_i \mathbb{P}_j) \\ + N^{-1} \sum_{m'=1}^{N-1} (N-m') \Big(\mathbb{P}_j \mathbb{P}_{ji}^{(m')} - \mathbb{P}_j \mathbb{P}_i \Big). \end{aligned} \tag{2.8}$$

Comparing the first term of the equation (2.8) with the first term of α_{ij} from the equation

(2.4) of the Lemma 2.3.3, the terms differs by the amount

$$N^{-1}\mathbb{P}_i \sum_{m=N}^{\infty} \Big(\mathbb{P}_{ij}^{(m)} - \mathbb{P}_j\Big) + N^{-1}\mathbb{P}_i \sum_{m=1}^{N-1} m\Big(\mathbb{P}_{ij}^{(m)} - \mathbb{P}_j\Big).$$

From the fact that in ergodic case of Markov chain $\exists$ constants γ and ρ, $\rho < 1$, such that

$$\Big|\mathbb{P}_{ij}^{(N)} - \mathbb{P}_j\Big| < \gamma\rho^N, \forall\ i,\ j,\ and\ N.$$

This implies that the series mentioned $\sum_{m=1}^{\infty}(\mathbb{P}_{ij}^{(m)} - \mathbb{P}_j)$ and $\sum_{m=1}^{\infty} m(\mathbb{P}_{ij}^{(m)} - \mathbb{P}_j)$ converge absolutely. Therefore the difference in above equation is of the order $o(1/N)$. Similarly the second sum also differs and the result of equation (2.3) follows henceforth.

Now, to prove the weak law of large numbers, equation (2.5), the part which will be needed for the proof of Theorem 2.3.2. We prove it as follows,

For weak law of large number to hold we must have:

$$\lim_{N\to\infty} \mathbb{P}\left[\left|\frac{f_i}{N} - \mathbb{P}_i\right| > \epsilon\right] = 0.$$

The variance of the random variable $f_i = \sum_{m=1}^{N} c_m(i)$ is derived using the relation that

$$Var\left(\sum_{i=1}^{N} X_i\right) = \sum_{i=1}^{N}\sum_{j=1}^{N} Cov\Big(X_i, X_j\Big).$$

So applying above relation to our use we have:

$$
\begin{aligned}
Var(f_i) = Var\sum_{m=1}^{N} c_m(i) &= \sum_{m=1}^{N}\sum_{l=1}^{N} Cov\Big(c_m(i), c_l(i)\Big) \\
&= \sum_{l=1}^{N}\sum_{m=1}^{N} E\Big\{\big(c_l(i)-\mathbb{P}_i\big)\big(c_m(i)-\mathbb{P}_i\big)\Big\} \\
&= NN^{-1}\sum_{l=1}^{N}\sum_{m=1}^{N} E\Big\{\big(c_l(i)-\mathbb{P}_i\big)\big(c_m(i)-\mathbb{P}_i\big)\Big\} \\
&= NE\big[\zeta_i\zeta_i\big] \text{ (from equation (2.6))} \\
&= N\Big[\alpha_{ij}+o(1/N)\Big]. \text{ (from equation (2.3))}
\end{aligned}
$$

So we have $Var\Big(\frac{f_i}{N}\Big) = \frac{1}{N^2}Var(f_i) = \frac{N[\alpha_{ij}+o(1/N)]}{N^2} = \frac{\alpha_{ij}}{N} + o\Big(\frac{1}{N^2}\Big)$.

The variance of random variable $\frac{f_i}{N}$ is $\Big[\frac{\alpha_{ij}}{N} + o\Big(\frac{1}{N^2}\Big)\Big]$. So the probability equivalent event is:

$$
\mathbb{P}_N\left[\left|\frac{f_i}{N}-\mathbb{P}_i\right| > \epsilon\right] = \mathbb{P}_N\left[\Big|f_i - N\mathbb{P}_i\Big| > N\epsilon\right].
$$

Now, by Chebyshev's inequality:

$$
\mathbb{P}_N\left[\left|\frac{f_i}{N}-\mathbb{P}_i\right| > \epsilon\right] \leq \frac{Var(f_i)}{(N\epsilon)^2} = \frac{1}{\epsilon^2}\frac{Varf_i}{N^2} = \frac{1}{\epsilon^2}\left(\frac{\alpha_{ij}}{N} + o\Big(\frac{1}{N^2}\Big)\right).
$$

So as N $\to \infty$ the above term $\to 0$ which gives us:

$$
\lim_{N\to\infty}\mathbb{P}_N\left[\left|\frac{f_i}{N}-\mathbb{P}_i\right| > \epsilon\right] = \lim_{N\to\infty}\left[\frac{1}{\epsilon^2}\left(\frac{\alpha_{ij}}{N} + o\Big(\frac{1}{N^2}\Big)\right)\right] \to 0.
$$

So we have

$$
\frac{f_i}{N} \xrightarrow{\mathbb{P}} \mathbb{P}_i.
$$

Hence proved Lemma 2.3.3. □

We generate the process x_n as mentioned in the algorithm for DTMC 1 as follows. Consider an independent collection of random variables x_1 and w_{in} $(i = 0, 1, \cdots, s;\ n = 0, 1, \cdots)$ such that

$$
\mathbb{P}\{x_1 = i\} = \mathbb{P}_i \text{ and } \mathbb{P}\{w_{in} = j\} = \mathbb{P}_{ij}.
$$

The way of sampling is same as we did in the Function 'F' of DTMC. Now by the way of

sampling we have intuitively:

$$\mathbb{P}\{x_k = a_k, 0 \leq k \leq N\} = \mathbb{P}_{a_0}\mathbb{P}_{a_0a_1}\cdots\mathbb{P}_{a_{N-1}a_N}. \tag{2.9}$$

Now to prove the Theorem 2.3.2, a convergence theorem needs to be used which is stated as below and proof of it can be referred to Section 20.6 of Cramér [8].

Theorem 2.3.4. *Let $\xi_1, \xi_2, \cdots$ be a sequence of random variables with distribution functions $F_1, F_2, \cdots$. Suppose that $F_n(x) \to F(x)$ as $n \to \infty$.*
Let $\eta_1, \eta_2, \cdots$ be another sequence of random variables, and suppose that the $\eta_n \xrightarrow{\mathbb{P}} c(constant)$. Put $X_n = \xi_n + \eta_n$; $Y_n = \xi_n\eta_n$; $Z_n = \frac{\xi_n}{\eta_n}$. Then the distribution functions of $X_n \to F(x-c)$. Further if $c > 0$, then distribution functions of $Y_n \to F\left(\frac{x}{c}\right)$ while the distribution of $Z_n \to F(cx)$.

We will try to use above theorems to leverage. Hence, consider the random vector $\eta_{ij} := \frac{f_{ij} - f_i\mathbb{P}_{ij}}{(N\mathbb{P}_i)^{\frac{1}{2}}}$. This random vector will have the same distribution by Theorem 2.3.4 if it can be proved that for each fixed i and j, the difference of

$$\frac{g_{ij} - \lfloor N\mathbb{P}_i \rfloor \mathbb{P}_{ij}}{(N)^{\frac{1}{2}}} - \frac{f_{ij} - f_i\mathbb{P}_{ij}}{(N\mathbb{P}_i)^{\frac{1}{2}}} \tag{2.10}$$

goes to 0 in probability. Since the ratio of ξ_{ij} and η_{ij} goes to 1 in probability by the equation (2.5), it will then follow by above Theorem 2.3.4 that ξ has a limiting distribution. Now we need to show that equation (2.10) goes to 0 in probability. So we show it as below.
Let us define e_m as follows,

$$e_m = \begin{cases} 1 - \mathbb{P}_{ij}, & \text{if } w_{im} = j \\ -\mathbb{P}_{ij}, & \text{if } w_{im} \neq j. \end{cases}$$

and put $S_m = e_1 + \cdots + e_m$.

$$\begin{aligned} E[e_m] &= (1-\mathbb{P}_{ij})\mathbb{P}_{ij} + -\mathbb{P}_{ij}(1-\mathbb{P}_{ij}) \\ &= \mathbb{P}_{ij} - \mathbb{P}_{ij}^2 - \mathbb{P}_{ij} + \mathbb{P}_{ij}^2 \\ &= 0. \\ \sigma^2 = Var[e_m] &= (1-\mathbb{P}_{ij})^2\mathbb{P}_{ij} + (-\mathbb{P}_{ij})^2(1-\mathbb{P}_{ij}) \\ &= (1+\mathbb{P}_{ij}^2 - 2\mathbb{P}_{ij})\mathbb{P}_{ij} + \mathbb{P}_{ij}^2(1-\mathbb{P}_{ij}) \\ &= \mathbb{P}_{ij} + \mathbb{P}_{ij}^3 - 2\mathbb{P}_{ij}^2 + \mathbb{P}_{ij}^2 - \mathbb{P}_{ij}^3 \\ &= \mathbb{P}_{ij} - \mathbb{P}_{ij}^2 = \mathbb{P}_{ij}(1-\mathbb{P}_{ij}). \end{aligned}$$

Then the $e_m's$ are iid with mean 0 and variance $\sigma^2 = \mathbb{P}_{ij}(1-\mathbb{P}_{ij})$, and the difference will now become

$$\frac{(S_{\lfloor N\mathbb{P}_i \rfloor} - S_{f_i})}{N^{\frac{1}{2}}}. \tag{2.11}$$

Now we'll require the Kolmogorov's inequality for the remaining bit of the argument and we will use it directly without proving. So we state it as below.

Kolmogorov's Inequality: Let $X_1, \cdots, X_n : \Omega \to \mathbb{R}$ be independent random variable defined on common probability space $(\Omega, F, \mathbf{P})$ with $E[X_k] = 0$ and $\mathrm{Var}X_k < +\infty$, where $k = 1, \cdots, n$. Then for each $c > 0$ we have,

$$\mathbb{P}\Big(\underset{1\le k\le n}{max}|S_k| \ge c\Big) \le \frac{1}{c^2}Var[S_n] \equiv \frac{1}{c^2}\sum_{k=1}^{n} VarX_k,$$

where $S_k = X_1 + \cdots + X_n$.

Now to show that the equation (2.10) goes to 0 in probability is same as proving it for the equation (2.11), for which we prove it as under.

Proof. Given $\epsilon > 0$, choose n_0 such that $N \ge n_0$, then

$$\mathbb{P}\{|f_i - \lfloor N\mathbb{P}_i \rfloor| > N\epsilon^3\} < \epsilon,$$

which holds by equation (2.5). Now if $N \geq n_0$, then

$$\begin{aligned}
&\mathbb{P}\left\{\frac{|S_{\lfloor N\mathbb{P}_i\rfloor} - S_{f_i}|}{N^{\frac{1}{2}}} > \epsilon\right\} \\
&\qquad \leq \mathbb{P}\{|f_i - \lfloor N\mathbb{P}_i\rfloor| > N\epsilon^3\} + \mathbb{P}\left\{\underset{m-[N\mathbb{P}_i]\leq N\epsilon^3}{max}|S_{\lfloor N\mathbb{P}_i\rfloor} - S_m| > \epsilon N^{\frac{1}{2}}\right\} \\
&\qquad \leq \epsilon + 2\mathbb{P}\left\{\underset{1\leq m\leq N\epsilon^3}{max}|S_m| > \frac{\epsilon N^{\frac{1}{2}}}{2}\right\} \\
&\qquad \leq \epsilon + 2\Big(\frac{4}{\epsilon^2 N}(N\epsilon^3\sigma^2\Big) = \epsilon + 2(4\epsilon\sigma^2) = (1 + 8\sigma^2)\epsilon,
\end{aligned}$$

where the last inequality follows from the earlier stated Kolmogorov's inequality and since ϵ was arbitrary, equation (2.11) goes to zero in probability.

Now using the acquired knowledge from previous Theorems and Lemma we will prove the Theorem 2.3.2 as below.

Proof of Theorem *2.3.2.* Since the process produced by discussed algorithm will have proper joint distribution by equation (2.9), so using it to compute the distributions of f_{ij}. We already know $(f_{i1}, \cdots, f_{is})$ is the frequency count of $(w_{i1}, \cdots, w_{if_i})$. By weak law of large numbers (equation (2.5)), $f_i \to N\mathbb{P}_i$ with high probability. It is better to compare $(f_{i1}, \cdots, f_{is})$ with the frequency count $(g_{i1}, \cdots, g_{is})$ of $\Big(w_{i1}, \cdots, w_{\lfloor N\mathbb{P}_i\rfloor}\Big)$.
Now we have from the independence of array w_{iN} and Central Limit Theorem for MultiDimension 2.3.2 that the s^2 random variables

$$\frac{(g_{ij} - \lfloor N\mathbb{P}_i\rfloor\mathbb{P}_{ij})}{(N\mathbb{P}_i)^{\frac{1}{2}}}$$

are asymptotically jointly normally distributed with co-variance matrix given$(\lambda_{ij,kl})$ by $\lambda_{ij,kl} = \delta_{ik}(\delta_{jl}\mathbb{P}_{ij} - \mathbb{P}_{ij}\mathbb{P}_{il})$. Using Convergence Theorem 2.3.4 proved earlier we have that the $s^2 -$ *dimensional* random vector with components $\eta_{ij} = \frac{f_{ij} - f_i\mathbb{P}_{ij}}{(N\mathbb{P}_i)^{\frac{1}{2}}}$ will have the same mentioned distribution. Hence, proved Theorem 2.3.2. □

2.4 Rate of Convergence (Irreducible Discrete-Time Markov Chain)

We have already built upon the Theorem 2.3.2 which is very strong, so we will look at a special version of this and look at some experimental results and their numerical significance.

Now given a sequence of irreducible Markov chain obtained using the algorithm for DTMC, we recall the way we defined an Empirical Probability Matrix in terms of empirical probabilities as below. Following the notations introduced in earlier proved Theorem. We have,
Notations used:

$f_{i.}$ = number of exits from state i.
f_{ij} = number of instantaneous transitions from state i to state j.

We discuss the empirical probability once again before talking about the convergence. We have by definition $\hat{\mathbb{P}}_{ij} = \frac{f_{ij}}{f_{i.}}$ (gives the Empirical Transition Probability).

Now we will discuss about the theoretical way of defining rate of convergence of Empirical Probability Matrix (obtained from given Markov chain)(EPM) to Transition Probability Matrix (TPM) but before that we discuss more on the error estimate of this convergence.

$$\begin{aligned}\mathbb{P}\left(\hat{\mathbb{P}}_{ij} = \frac{c}{m}\middle| f_{i.} = m\right) &= \mathbb{P}\left(f_{ij} = c\middle| f_{i.} = m\right)\left[\ as\ \hat{\mathbb{P}}_{ij} = \frac{f_{ij}}{f_{i.}} \implies f_{ij} = f_{i.}\hat{\mathbb{P}}_{ij}\right] \\ &= \binom{m}{c}\mathbb{P}_{ij}^{c}(1-\mathbb{P}_{ij})^{c}\ (\text{as } f_{ij} \sim Bin(m, \mathbb{P}_{ij})\ .\end{aligned}$$

Hence

$$\begin{aligned}E\left(\hat{\mathbb{P}}_{ij}\middle| f_{i.} = m\right) &= \sum_{c=0}^{m}\frac{c}{m}\binom{m}{c}\mathbb{P}_{ij}^{c}(1-\mathbb{P}_{ij})^{m-c} \\ &= \frac{1}{m}\sum_{c=0}^{m}c\binom{m}{c}\mathbb{P}_{ij}^{c}(1-\mathbb{P}_{ij})^{m-c} \\ &= \frac{1}{m}m\mathbb{P}_{ij} = \mathbb{P}_{ij}.\end{aligned}$$

Therefore $\hat{\mathbb{P}}_{ij}$ is an unbiased estimator of $\mathbb{P}_{ij}$.

For Variance using the equation $V(X) = E(X^2) - (E(X))^2$ we have :

$$\begin{aligned} V\Big(\hat{\mathbb{P}}_{ij}\Big|f_{i.} = m\Big) &= \sum_{c=0}^{m} \Big(\frac{c}{m}\Big)^2 \binom{m}{c} \mathbb{P}_{ij}^{c}(1-\mathbb{P}_{ij})^{m-c} - \mathbb{P}_{ij}^2 \\ &= \left(\frac{1}{m^2}\sum_{c=0}^{m} c^2 \binom{m}{c} \mathbb{P}_{ij}^{c}(1-\mathbb{P}_{ij})^{m-c}\right) - \mathbb{P}_{ij}^2 \\ &= \frac{1}{m^2}\Big(m\mathbb{P}_{ij}(1-\mathbb{P}_{ij}) + (m\mathbb{P}_{ij})^2\Big) - \mathbb{P}_{ij}^2 \\ &= \frac{\mathbb{P}_{ij}(1-\mathbb{P}_{ij})}{m}. \end{aligned}$$

Hence $V\Big(\hat{\mathbb{P}}_{ij}\Big|f_{i.} = m\Big) \to 0$ as m $\to \infty$.
Since $\frac{f_{i.}}{N}$ = stationary probability of i^{th} state > 0.

$$\lim_{N\to\infty} f_{i.} = \infty \textbf{ (almost surely)}.$$

Hence $\hat{\mathbb{P}}_{ij}$ is a consistent estimator of $V\Big(\hat{\mathbb{P}}_{ij}\Big|f_{i.} = m\Big) = E\Big(|\hat{\mathbb{P}}_{ij} - \mathbb{P}_{ij}|^2\Big|f_{i.} = m\Big)$ (squared L^2 norm of error)

$$= \frac{\mathbb{P}_{ij}(1-\mathbb{P}_{ij})}{m} \leq \frac{1}{4m}.$$

Thus $\Big|\Big|\hat{\mathbb{P}}_{ij} - \mathbb{P}_{ij}\Big|\Big|_{L^2} = \sqrt{\mathbb{P}_{ij}(1-\mathbb{P}_{ij})}m^{-\frac{1}{2}}$, we notice as $N \to \infty, m \to \infty$ linearly as $\frac{(f_{i.})}{N} = \pi_i, (\pi_i \in \mathbb{R})$. Thus L^2 error decays in order $o(N^{-\frac{1}{2}})$.

We have already obtained the entry-wise error estimate above but for the experimental purpose we have used directly the norm of EPM(n) and TPM matrices.

Now we give a numerical analysis which shows the rate of convergence of the EPM to TPM. We use the methods used in previously described DTMC algorithm 1 to generate the EPM and perform the series of numerical operations to find out the rate of convergence. We define the terminology to be used in the rate plot as below:

$$r_N := \frac{|TPM - EPM^{(N)}|}{|TPM - EPM^{(2N)}|},\ \alpha_N := \log_2(r_N).$$

In the figure We denote the sum of the sequence $\{\alpha_N\}_N$ by $\overline{\alpha_N}$. This particular sum is more formally known as *Cesàro* sum. So in our use the *Cesàro* sum is $\overline{\alpha_N} = \frac{1}{n}\sum_{p=1}^{n}\alpha_p$. To illustrate the reason for this to be *Cesàro* sum. Consider $\{\alpha_p\}_p$ to be the sequence of α's. Then,

$$\text{for } p \geq 2, \text{ define } a_p := \alpha_p - \alpha_{p-1}, a_1 = \alpha_1$$

So, $\sum_{p=1}^{k} a_p$ is a telescopic sum and hence we end up with α_k.

Now, the *Cesàro* sum of $\sum a_n = \frac{1}{n}\sum_{k=1}^{N}\alpha_k \to \overline{\alpha_N}$.

Now, for large N, if α is the rate of convergence of $\{x_N\}$ to x, which are entries of $EPM^{(N)}$ and TPM matrices respectively, then $|x - x_N| \sim cN^{-\alpha}$.

$$\frac{|x - x_N|}{|x - x_{2N}|} \approx \frac{cN^{-\alpha}}{c(2N)^{-\alpha}} = 2^{\alpha}.$$

or

$$\alpha \approx \log_2\left(\frac{|x - x_N|}{|x - x_{2N}|}\right).$$

Hence, we obtain α.
In general:

$$|x - x_N| = cN^{-\alpha} + o(N^{-\alpha}).$$

then

$$\begin{aligned}\frac{|x - x_N|}{|x - x_{2N}|} &= \frac{cN^{-\alpha} + o(N^{-\alpha})}{c(2N)^{-\alpha} + o((2N)^{-\alpha})} \\ &= \frac{1}{2^{-\alpha} + \frac{o((2N)^{-\alpha})}{cN^{-\alpha}}} + \frac{\frac{o(N^{-\alpha})}{cN^{-\alpha}}}{\frac{c(2N)^{-\alpha}}{cN^{-\alpha}} + \frac{o((2N)^{-\alpha})}{cN^{-\alpha}}}.\end{aligned}$$

Now above equation on right hand side has 2 terms. So considering first of these terms and trying to resolve it we have:

$$\frac{o((2N)^{-\alpha})}{cN^{-\alpha}} \to 0 \text{ as } N \to \infty.$$

Now considering the second term. We first resolve numerator as below:

$$\frac{o(N^{-\alpha})}{cN^{-\alpha}} \to 0 \text{ as } N \to \infty.$$

Now looking at denominator we have:

$$\frac{c(2N)^{-\alpha}}{cN^{-\alpha}} \to (\text{fixed } +ve \text{ quantity (bounded)}) + \frac{o((2N)^{-\alpha})}{cN^{-\alpha}} \to 0 \text{ as } N \to \infty).$$

Hence the second term is 0.
So we have:

$$\lim_{N\to\infty} \frac{|x - x_N|}{|x - x_{2N}|} = \frac{1}{2^{-\alpha}}.$$

so,

$$\alpha = \log_2 \lim_{N\to\infty} \left(\frac{|x - x_N|}{|x - x_{2N}|} \right).$$

Since log function is continuous, so

$$\alpha = \lim_{N\to\infty} \log_2 \left(\frac{|x - x_N|}{|x - x_{2N}|} \right).$$

or

$$\alpha = \lim_{N\to\infty} \alpha_N, \text{ where } \alpha_N = \log_2 \left(\frac{|x - x_N|}{|x - x_{2N}|} \right).$$

2.4.1 Numerical Experiment:

Now we have the following plot as a result of computation done by us on previously simulated Discrete-time Markov chain sequence to obtain the EPM. We are considering the state space to be $S = \{0, 1, 2\}$, the Transition Probability Matrix from [10] as follows,

$$TPM = \begin{bmatrix} 0 & 0.67 & 0.33 \\ 0.5 & 0 & 0.5 \\ 0.33 & 0.67 & 0 \end{bmatrix}$$

to generate a chain of length $N = 10^4$.

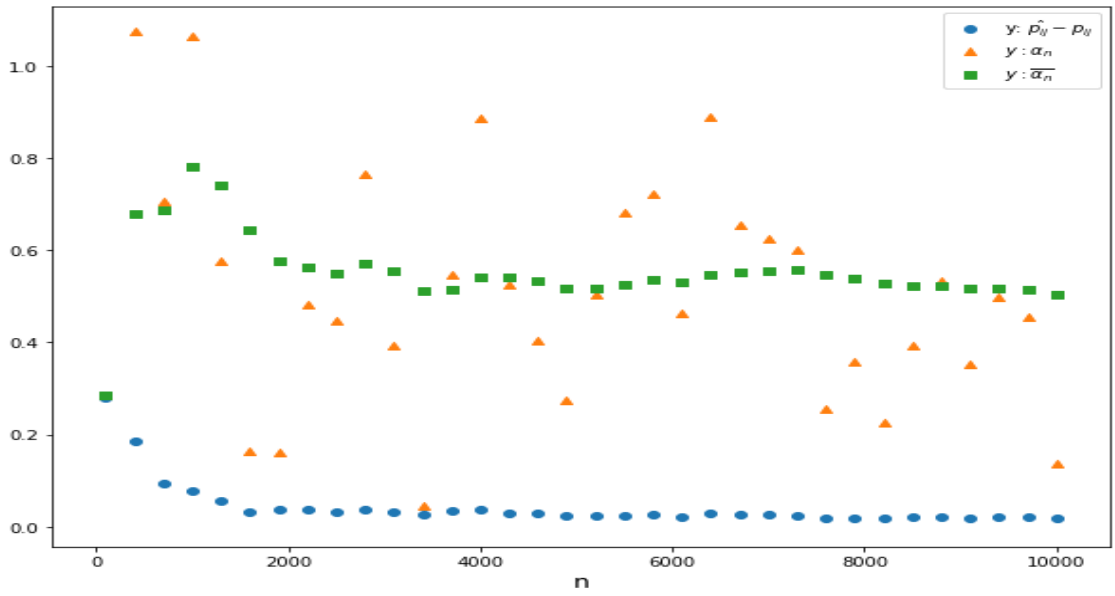


Figure 2.2: Estimation error & rate of convergence with increasing chain length.

Comment: The result in the theory asserts the rate to be $\frac{1}{2}$ for large N. However, the empirical verification of this is tricky as seen in the plot of $\overline{\alpha_N}$ in Figure 2.2. In Figure 2.2, $\overline{\alpha_N}$ varies from 0 to 1. However, for large N (≥ 5000), the $\overline{\alpha_N}$ remains close to $\frac{1}{2}$. Also note that due to truncation error we have problems related to round-off which has been tried to fix by using the fixed point precision method (to be precise, the precision of 8 digits) but still we may have these errors due to level of calculations performed on data. Also the computation has been done by taking norm of $EPM^{(N)}$ and TPM matrices but in theory we have done it component wise for these matrices. Now, the chaotic behaviour as seen in the Figure 2.2 can be explained via following empirical computation. For each $i, j \in S$ we have:

$$\begin{aligned}
\alpha_{ij}(N) = \log_2\left(\frac{\hat{\mathbb{P}_{ij}}^{(N)} - \mathbb{P}_{ij}}{\hat{\mathbb{P}_{ij}}^{(2N)} - \mathbb{P}_{ij}}\right) &\overset{CLT}{\approx} \frac{1}{2}\log_2\left(\frac{|\mathbf{N}(0, \sqrt{\mathbb{P}_{ij}(1-\mathbb{P}_{ij}}(\pi_i N)^{-\frac{1}{2}})|^2}{|\mathbf{N}(0, \sqrt{\mathbb{P}_{ij}(1-\mathbb{P}_{ij}}(2\pi_i N)^{-\frac{1}{2}})|^2}\right) \\
&\approx \frac{1}{2}\log_2\left(\frac{\mathbb{P}_{ij}(1-\mathbb{P}_{ij})(\pi_i N)^{-1}\chi_1}{\mathbb{P}_{ij}(1-\mathbb{P}_{ij})(2\pi_i N)^{-1}\chi_2}\right). \\
&\approx \frac{1}{2}\Big(\log_2(2F(1,1))\Big),
\end{aligned}$$

where χ_1 and χ_2 in the above expression are denoting two independent χ_1^2 random vari-

able. Since $\frac{\chi_1}{\chi_2} \sim F(1,1)$, therefore, we obtain the last term of above expression for $p-value = 5\%$, we have $\mathbb{P}\left(F(1,1) \in \left(\frac{1}{161}, 161\right)\right) < 95\%$ and so $\mathbb{P}(\log_2(F(1,1) \in (-7.33, 7.34)) < 95\%$. So we have,

$$\begin{aligned}\alpha_{ij}(N) &\approx \frac{1}{2}\log_2\left(2F(1,1)\right) \\ &= \frac{1}{2} + \frac{1}{2}\log_2(F(1,1)) \\ &\in \left(\frac{1}{2} - 3.67, \frac{1}{2} + 3.67\right).\end{aligned}$$

Thus the asymptotic distribution of $\alpha_{ij}(N)$ has a large standard error.

2.5 Continuous-time Markov Chain

In this part we will go through the definition of Continuous-time Markov chain and the method to simulate it. Before simulating we will also define the Q-matrix.

Definition 2.5.1. *A stochastic process $\{X(t): t \geq 0\}$ is called a Continuous-time Markov Chain if $\forall\ t \geq 0$, $s \geq 0$, $i \in S$, $j \in S$,*

$$\mathbb{P}(X(s+t) = j | X(s) = i, \{X(u) : 0 \leq u < s\}) = \mathbb{P}(X(s+t) = j | X(s) = i) = \mathbb{P}_{ij}(t),$$

where $\mathbb{P}_{ij}$ represents the probability that the chain will be in state j after t units of time given it is in state i now.

Definition 2.5.2 (**Q-matrix**). *A transition rate-matrix or Q-matrix is a matrix describing the rate a continuous-time Markov chain moves between the states. The element $q_{ij}(i \neq j)$ denotes the rate of departing the state i and arriving at a state j whereas diagonal elements $q_{ii} = -\sum_{j \neq i} q_{ij}$. A Q-matrix satisfies the following conditions:*

1. $0 \leq -q_{ii} < \infty$.
2. $0 \leq q_{ij}$: *for* $i \neq j$.

3. $\sum_j q_{ij} = 0 : \ \forall \ i.$

Now let us look at the way of simulating the Continuous-time Markov chain.

Algorithm 2 To simulate a Continuous-time Markov Chain, given a Q-matrix, Time-Horizon, initial state:

---

1. Choose an initial state, say $X_0 = i_0$, set $n = 0, t_0 = 0$.
2. Generate $H_{X_0} \sim \exp(-q_{X_0X_0})$, set $t_1 = H_{X_0}$.
3. while $t_n < T$:
   $X_{n+1} = F(x, X_n), x \sim U[0, 1]$.
4. set $n = n + 1$.
5. Generate $H_{X_n} \sim exp(-q_{X_nX_n})$, set $t_{n+1} = t_n + H_{X_n}$.
6. if $t_{n+1} > T$ stop
   else repeat from (3).

2.6 Semi-Markov Processes

We can now deduce what if the holding times H do not have an exponential distribution, so then the resulting process $\{X_t : t \geq 0\}$ will not, in general, possess the Markov property; it will not be a CTMC. Instead, it is called a Semi-Markov process having holding time distribution, say Z_{X_n}. Now using Tanmay [13], we have the following definition.

Definition 2.6.1. *A Semi-Markov process is a process $\{X_t\}_{t\geq 0}$ that satisfies the following properties:*

1. *X_t is a piece-wise constant rcll process with discontinuities at a discrete set $\{T_n\}_{n\geq 1}$.*
2. *The transition probabilities satisfy:*
   $\mathbb{P}[X_{T_{n+1}} = j, T_{n+1} - T_n \leq y|(X_0, T_0), (X_1, T_1), \cdots, (X_{T_n}, T_n)] =$
   $\mathbb{P}[X_{T_{n+1}} = j, T_{n+1} - T_n \leq y|X_{T_n}]$.

Algorithm 3 To simulate a Semi-Markov process, given a Instantaneous Transition-rate Matrix, Time-Horizon, initial state, general holding Time distribution, say Z_{X_n}:

Now let us look at the method we used for simulating the Semi-Markov process.

1. Choose an initial state, say $X_0 = i_0$, set $n = 0$, $t_0 = 0$.

2. while $t_n < T$:
   $X_{n+1} = F(x, X_n)$, $x \sim U[0, 1]$.

3. Generate $H \sim Z_{X_n, X_{n+1}}$, set $t_{n+1} = t_n + H$.

4. if $t_{n+1} > T$ stop
   else $n = n + 1$
   repeat from (2).

Chapter 3

Brownian Motion

3.1 Development of Brownian Motion

Robert Brown was the first person to study the continuous irregular movement of pollen grains suspended in fluid. Henceforth, the phenomenon was named after him as Brownian Motion. In mathematics Brownian motion is described by the Wiener process, a continuous-time stochastic process named after Norbert Wiener who studied this process in great details.

Let us now discuss more on the Brownian motion.

Definition 3.1.1. *The random variables $\{B_t\}_{t\geq 0}$, is said to be Standard Brownian Motion if it satisfies the following conditions:*

1. $B_0 = 0$.
2. $t \to B_t$ *is almost surely continuous, i.e* $\mathbb{P}\{\omega \in \Omega \mid$ *the path* $t \to B_t(\omega)$ *is a continuous function*$\} = 1$.
3. B_t *has independent increments, i.e.* $\forall\ t_1, t_2, \cdots, t_n$, *the* $B_{t_2} - B_{t_1}$, $B_{t_3} - B_{t_2}, \cdots, B_{t_n} - B_{t_{n-1}}$ *are independent random variables. More precisely, for* $0 \leq s < t$, $B_t - B_s$ *is independent of* $B_u, u \in [0, s]$.
4. $B_t - B_s \sim N(0, t-s)$ *(for* $0 \leq s \leq t$*), where* $N(\mu, \sigma^2)$ *denotes the normal distribution with mean* μ *and variance* σ^2.

The Generalized Brownian Motion is $\mathbf{W} = \{W_s : s \in [0, \infty)\}$ can be constructed from Standard Brownian motion as $W_s = \mu s + \sigma B_s$ where $B_s = \{B_s : s \in [0, \infty)\}$ is the standard Brownian motion and μ is the drift parameter and σ is the volatility parameter.

We have noticed that since Brownian Motion can take negative values, so using it to model stock price dynamics is not a good idea. So the idea is to look for a non-negative random variable. We introduce here the Geometric Brownian Motion, the non-negative variant of Brownian motion. We recall it using [14] as below.

Definition 3.1.2 (Geometric Brownian Motion)**.** *Let $\{B_t\}_{t\geq 0}$ be the standard Brownian Motion, then*

$$S_t = S_0 e^{W_t}, \tag{3.1}$$

where $W_t = \mu t + \sigma B_t$ is the generalized Brownian motion with drift and S_0 is the initial value.

Going back to equation (3.1), if we take logarithm both sides, it yields us the Brownian Motion; $W_t = \ln\left(\frac{S_t}{S_0}\right) = \ln(S_t) - \ln(S_0)$. $\ln(S_t) = \ln(S_0) + W_t$ is normally distributed with mean $(\mu t + \ln(S_0))$ and variance $\sigma^2 t$. Hence $\ln(S_t) - \ln(S_u) = W_t - W_u$ is independent of $\ln(S_v)$ for $v \in [0, u]$, i.e. $\frac{S_t}{S_u}$ is independent of S_v for $v \in [0, u]$.

Definition 3.1.3. *A discrete-time stochastic process $\{X_n\}_{n\geq 0}$ that satisfies for any time n,*

$$E(|X_n|) < \infty,$$

$$E(X_{n+1}|X_1, \cdots, X_n) = X_n,$$

is called a discrete-time martingale.

In other words, conditional expected value of future state, given all past observations, is equal to the most recent observation. Let us look at the definition of continuous-time martingale.

Definition 3.1.4. *A continuous-time stochastic process $\{X_t\}_{t\geq 0}$ is a continuous-time martingale if $E(|X_t|) < \infty$ and $E(X_{t+h}|X_s, 0 \leq s \leq t) = X_t$ for $h \geq 0, t \geq 0$.*

Theorem 3.1.1. *Brownian motion is a continuous-time martingale.*

Proof. To show that Brownian motion is a continuous-time martingale is same as showing $E(B_t|\{B_u, u \in [0,s]\}) = B_s$. So, using $B_t - B_s \sim N(0, t-s)$, and independent to $\{B_u, u \in [0,s]\}$.

$$\begin{aligned} E(B_t|\{B_u, u \in [0,s]\}) &= E(B_t - B_s + B_s|\{B_u, u \in [0,s]\}) \\ &= E(B_t - B_s|\{B_u, u \in [0,s]\}) + E(B_s|\{B_u, u \in [0,s]\}) \\ &= 0 + E(B_s|\{B_u, u \in [0,s]\}) \\ &= 0 + B_s \\ &= B_s. \end{aligned}$$

Hence Proved. □

Before we begin with methods for simulation we will discuss about methods which can be used to help us in performing the simulation efficiently.

3.2 Box-Muller Transform

As we saw in the first chapter, the method of simulation is dependent on the behaviour and knowledge of function $F(x)$ and also whether it is analytic and invertible. For the purpose of simulating Brownian motion and other stochastic process, we will need to generate random numbers from Normal Distribution, but the normal cumulative distribution function cannot be inverted analytically. Hence our earlier used method will fail. So this particular problem can be resolved by using a relatively simple method that allows us simulating two normal random distribution from two uniform random numbers. This particular method is termed as Box-Muller Transform.

Consider two random variables X,Y distributed as Standard Normal, and independent of each another. The joint pdf then can be defined as:

$$h(x,y) = \frac{1}{2\pi} e^{-\frac{(x^2+y^2)}{2}}$$

Consider transformation of variables from Cartesian(x, y) into polar coordinates(r, θ) by:

$$x = r\cos(\theta)$$

$$y = r\sin(\theta).$$

The joint pdf $g(r, \theta) := h(x, y)|J|$, J is the Jacobian. Hence,

$$g(r, \theta) = \frac{1}{2\pi} r e^{-\frac{r^2}{2}}$$

for $r \geq 0, 0 \leq \theta \leq 2\pi$. This is the product of two pdf's, namely Uniform distribution and Rayleigh distribution. We know Rayleigh distribution is defined by:

$$f(x; \sigma) = \frac{x}{\sigma^2} e^{\frac{-x^2}{2\sigma^2}}, x \geq 0$$

$$F(x; \sigma) = 1 - e^{\frac{-x^2}{2\sigma^2}}, x \in [0, \infty)$$

where σ is the scale parameter of the distribution. We know a quantile function [7] is defined by $Q(p) = \inf\{x \in \mathbb{R} : p \leq F(x)\}$, and if F(x) is continuous and strictly monotonic function then $Q = F^{-1}$. In our case, we know the CDF, $F_X(x)$ is continuous and strictly monotonic function, so we have $F_X(x) = P(X \leq x) = p$. So here we have $p = 1 - e^{\frac{r^2}{2}}$ which resolves to $r = \sqrt{-2\ln(1-p)} = G^{-1}(p)$ and so $R = \sqrt{-2\ln(1-U)}$ simulates the Rayleigh distribution, given $U \sim U(0, 1)$. Also,

$$I(\theta) = \begin{cases} \frac{\theta}{(2\pi)}, 0 \leq \theta \leq 2\pi \\ 0, \text{ otherwise .} \end{cases}$$

So $\theta = 2\pi V$ simulates a uniform distribution between 0 and 2π, where $V \sim V(0, 1)$. Hence, we arrive at the formulas to simulate a pair of normal random variables X and Y:

$$X = R\cos(\theta) = \sqrt{-2\ln(1-U)}\cos(2\pi V)$$

$$Y = R\sin(\theta) = \sqrt{-2\ln(1-U)}\sin(2\pi V)$$

This method provides us with pair of Normal random numbers $X, Y \sim N(0, 1)$, we can use it to our benefit by using both of them and thus we can reduce the number of times for generating normal numbers by half the amount. Let us see the algorithm for generating

normal random numbers using Box-Muller Transform.

Algorithm 4 Generating Normal random numbers with given mean (m) and variance (s)

1. Generate 2 random numbers U, V, where $U, V \sim U[0, 1]$.
2. To generate pair Standard normal number, set $X_1 = \sqrt{-2\ln(1-U)}\cos(2\pi V)$ and $X_2 = \sqrt{-2\ln(1-U)}\sin(2\pi V)$.
3. To generate Normal random number with given mean and variance, we set $Z_1 = m + \sqrt{s}X_1$ and $Z_2 = m + \sqrt{s}X_2$.

3.3 Simulation of Brownian Motion

Now let us look at the method of simulating Brownian motion.

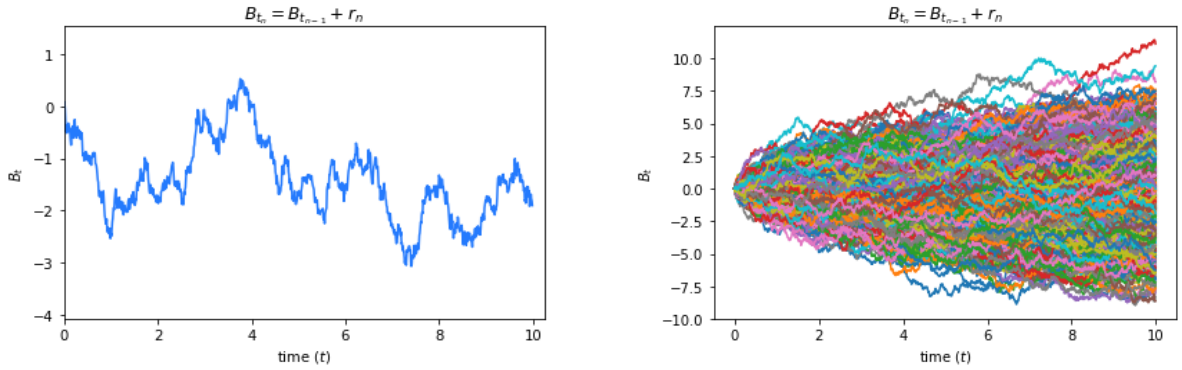
Algorithm 5 Method for simulating standard Brownian motion at times $0 = t_0 < t_1 < t_2 < \cdots < t_k$:

1. Set a definite time increment $t_i - t_{i-1} = dt$.
2. Sequentially generate normal random variable (using Box-Muller transform Algorithm [4]) $r_1, r_2, \cdots, r_k$, with mean 0 and variance dt.
3. We have already set the time-increments dt, then by definition we can recursively define realizations of Brownian motion as,

$$\begin{aligned} B_{t_1} &= r_1, \\ B_{t_2} &= B_{t_1} + r_2 = r_1 + r_2, \\ &\vdots \\ B_{t_k} &= \sum_{i=1}^{k} r_i. \end{aligned}$$

We observe that to generate Brownian motion, we only need to generate normal random variables using Box-Muller transform. Following plots are generated using above algorithm,

where we have used discrete points for generating such points and to approximate with continuous-time we have used small increments. Also we can check the Gaussian behaviour of Brownian motion by plotting several realizations of Brownian motion as shown below.



(a) Brownian motion with t=10 units

(b) 1000 realizations of Brownian Motion

Figure 3.1: Standard Brownian Motion Simulation with dt = 0.01

Next, we look at few more simulations which deal with General Brownian motion with drift parameter μ, and volatility term σ and also we look at the method to simulate GBM.

Algorithm 6 Simulating the Brownian motion with drift μ and volatility term σ at times $0 = t_0 < t_1 < t_2 < \cdots < t_k$:

1. Set a definite time increment $t_i - t_{i-1} = dt$.
2. Sequentially generate normal random variable (using Box-Muller transform Algorithm [4]) $r_1, r_2, \cdots, r_k$, with mean 0 and variance dt.
3. We can then by definition recursively define,

$$\begin{aligned} X_{t_1} &= \sigma r_1 + \mu dt, \\ X_{t_2} &= X_{t_1} + \sigma r_2 + \mu dt = \sigma(r_1 + r_2) + 2\mu dt, \\ &\vdots \\ X_{t_k} &= \sum_{i=1}^{k} \Big(\sigma r_i + \mu dt\Big) = \sigma \sum_{i=1}^{k} (r_i) + k\mu dt. \end{aligned}$$

Algorithm 7 Simulating Geometric Brownian motion (GBM) (with drift μ and variance term σ) at times $0 = t_0 < t_1 < t_2 < \cdots < t_k$:

1. Set a definite time increment $t_i - t_{i-1} = dt$.
2. Sequentially generate normal random variable (using Box-Muller transform Algorithm [4]) $r_1, r_2, \cdots, r_k$, with mean 0 and variance dt.
3. Next we generate log-normal random variables Y_i by setting $Y_i = e^{\sigma r_i + \left(\mu - \frac{1}{2}\sigma^2\right)dt}$, $i \in 1, 2, \cdots, k$.
4. We can then recursively define,

$$\begin{aligned} S_{t_1} &= S_0 Y_1 \\ S_{t_2} &= S_{t_1} Y_2 = S_0 Y_1 Y_2 \\ &\vdots \\ S_{t_k} &= S_{t_{k-1}} Y_k = S_0 \prod_{i=1}^{k} Y_i. \end{aligned}$$

Chapter 4

Stochastic Models of Asset Price Dynamics

Black-Scholes model was one of the early models to study and replicate the stock price dynamics in great details, and it also gave the formula for pricing the European call and put options. The model was first formulated through the hard work of Fischer Black and Myron Scholes. They went on derive the equation, to say, partial differential equation, known as Black-Scholes equation, governing the prices of options over time. Subsequently, the work by Robert Merton during his study of Black-Scholes model by employing mathematics, he formulated the Black-Scholes option pricing formula. For this work, Merton and Scholes went on to win Nobel Prize in Economics in 1997. Let us first define few basic terms of option pricing theory.

4.1 Basics of option pricing

In this section, we talk about some basic terms of option pricing theory which will be helpful in the forthcoming sections.

Definition 4.1.1. *An option is an agreement (contract) which gives owner the right, but not obligation to buy (call) or sell (put) an underlying asset or financial instrument at a specified strike price* (K) *at maturity time of the contract* (T)*, which is a specified date.*

The details mentioned in the definition have to be mentioned in the clause of the contract, where by underlying asset can be a company stock, currency, commodity as silver, etc. A simple example can be of a contract that binds two parties for a sale of a commodity such as petrol or crude oil, four months from now. Now let us discuss the problem of option pricing without getting into much detail.

Since while signing a contract the buyer of the option is transferring risk to the writer of the option. He must pay him fairly to agree to the contract. Let us consider a stock, whose value at maturity time T is S_T. Now, if the writer has to buy the stock at maturity time T to sell that to buyer as promised, he would lose $C_T = \max(S_T - K, 0) = (S_T - K)^+$ as mostly the option will be exercised only if the buyer gains some profit out of it. Hence, the price of the option at maturity T is known to be C_T. But the problem of pricing involves that how much the buyer of the option should pay the writer of the option at time $t = 0$ for an asset worth $(S_T - K)^+$ at T, the answer depends on the statistical properties of S_T. There were several models which spanned out to solve this particular problem and we will discuss few of them in detail.

4.2 Black-Scholes model

The stock prices, $\{S_t\}_{t\geq 0}$ in BSM model follow geometric Brownian motion whose dynamics is given by

$$dS_t = \mu S_t dt + \sigma S_t dW_t, S_0 \geq 0,$$

where W_t is a Brownian motion, and μ is percentage drift, and σ is percentage volatility. The solution of above SDE is given by

$$S_t = S_0 e^{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t}.$$

The Black-Scholes equation is a PDE, which describes the price of the option over time and also Black-Scholes model only talks about the European style of option pricing. If at time t the stock price $S_t = S$ and $\phi(S, t)$ is the price of derivative as a function of stock price and time, r is the annualized risk-free interest rate, continuously compounded, μ is the drift parameter, volatility σ is the standard deviation of stock returns per unit time, then ϕ

satisfies the following PDE:

$$\frac{\partial \phi}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 \phi}{\partial S^2} + rS\frac{\partial \phi}{\partial S} - r\phi = 0.$$

The corresponding solution, known as Black-Scholes formula for call option price with strike price K is,

$$C(S_t, t) = N(d_1)S_t - N(d_2)Ke^{-r(T-t)},$$

$$d_1 = \frac{1}{\sigma\sqrt{T-t}}\left[\ln\left(\frac{S_t}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)\right],$$
$$d_2 = d_1 - \sigma\sqrt{T-t},$$

where $N(.)$ is the CDF of Standard Normal distribution, $T-t$ is the time to maturity, S_t is the spontaneous stock price used for immediate settlement, and r and σ are as mentioned above.

4.3 Drawbacks and many generalization

The Black-Scholes model has the following assumptions laid on it.

Assumptions:

1. There is no arbitrage opportunity (i.e. No opportunity of making risk-less profit exists.)
2. It is possible to borrow and lend cash at a known constant risk-free interest rate.
3. It is possible to buy and sell any amount, even fractional, of stock (which includes short selling).
4. The above transactions do not incur any fees or cost (i.e., frictionless market).
5. The stock price follows a GBM with constant drift and volatility.
6. The underlying security does not pay a dividend.

With above assumptions, we can figure out the drawbacks of this model.

Drawbacks of BSM model:

1. To begin with the model assumes the underlying factors to be known and also assumes them to be constant throughout the life of option. But none of this is true in the real world scenario, the underlying riskfree rate, volatility, and dividends are unknown and they may change in short span of time with high variance and thus leading to high fluctuation of option prices.

2. It assumes the stock prices have continuous sample paths, implying it fails to model the opening gaps and jumps.

3. Also it is not suitable for pricing style other than European call options.

But even with these many drawbacks, the Black-Scholes model is widely popular as it is simple and provides a ready-made value for the option. But for more accurate modelling of option pricing, there have been many generalization over the years such as Heston model, Jump Diffusion model, MMGBM model. We are going to discuss them briefly below.

4.3.1 Heston Model

Heston model, named as such after Steven Heston, is one of the key generalization to overcome the drawback of BSM model. It is a model which assumes the volatility of the asset is not constant, not deterministic, but rather it follows a random process, to be more precise, a CIR process.

Model Description

The Heston model assumes the stock price, $\{S_t\}_{t\geq 0}$ follows the dynamics given by equation

$$dS_t = \mu S_t dt + \sqrt{\nu_t} S_t dW_t, S_0 > 0,$$

where ν_t, the square of the instantaneous volatility, is a CIR process:

$$d\nu_t = \kappa(\theta - \nu_t)dt + \sigma\sqrt{\nu_t} dW_t^{'}, \nu_0 > 0,$$

where $dW_t dW_t^{'} = \rho dt$. The parameters in the above mentioned equation represent the following:

1. θ is the long run mean of ν_t.
2. κ is the speed of mean reversion.
3. σ is the volatility of volatility.

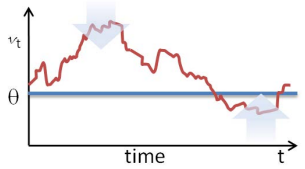


Figure 4.1: θ, ν_t vs time (t)

Also, the following conditions need to be satisfied,

1. Feller condition: $\sigma^2 < 2\kappa\theta$ to assure non-negativity of ν.
2. Square integrability of S: $\sigma \leq \frac{\kappa}{(2\rho+\sqrt{2})^+}$.

Now if at time t, the stock price S_t is equal to S and the volatility $\sqrt{\nu_t} = \sqrt{\nu}$, then the locally risk minimizing price C_t of a call option with a maturity time T and strike price K is given by,

$$C_t = \phi(t, S, \nu),$$

where ϕ satisfies the following PDE:

$$\frac{\partial \phi}{\partial t} + rS\frac{\partial \phi}{\partial S} + (\kappa(\theta - \nu) - \rho\sigma(\mu - r))\frac{\partial \phi}{\partial \nu} + \frac{1}{2}\nu S^2 \frac{\partial^2 \phi}{\partial S^2} + \rho\sigma\nu S\frac{\partial^2 \phi}{\partial S \partial \nu} + \frac{1}{2}\sigma^2 \nu \frac{\partial^2 \phi}{\partial \nu^2} - r\phi = 0.$$

$$\phi(t, S, \nu) = (S - K)^+ \quad \forall S, \nu.$$

For more extensive research on the related topic, one can refer [4]

4.3.2 Jump Diffusion Model

As previously discussed, BSM model assumed the stock prices to follow GBM and henceforth, it does not account for jumps which might be there in stock prices. Although for

longer time-horizon one may see continuous behaviour but for day scale jumps may be observed. Therefore, we look more upon this using jump diffusion model [12]. The equation for dynamics of stock prices is given by:

$$dS_t = S_t\Big(\mu dt + \sigma dW_t + \int_{\mathbb{R}} \eta(z)N(dz, dt)\Big),$$

where $N(dz, dt)$ is a Poisson random measure on $\mathbb{R} \times [0, \infty)$ with intensity measure $\nu(dz)dt$, where ν is a finite Borel measure. W and N are assumed to be independent. The bounded continuous function $\eta : \mathbb{R} \to (-1, \infty)$ is the jump size coefficient. Now, at time t, if the stock price S_t is equal to S then the locally risk minimizing price C_t of a call option with a maturity time T and strike price K is given by, $C_t = \phi(t, S)$, where

$$\frac{\partial \phi}{\partial t} + (r + \beta_1)S\frac{\partial \phi}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 \phi}{\partial S^2} + \int_{\mathbb{R}} (\beta_2(z)[\phi(t, S(1 + \eta(z)) - \phi(t, S)]\nu(dz) = r\phi(t, S),$$

where,

$$\beta_1 = \frac{(\mu - r)\int \eta^2(z)\nu(dz) - \sigma^2 \int \eta(z)\nu(dz)}{\sigma^2 + \int \eta^2(z)\nu(dz)},$$
$$\beta_2(z) = 1 - \frac{(\mu - r + \int(\eta(z)\nu(dz))\eta(z)}{\sigma^2 + \int \eta^2(z)\nu(dz)},$$

and $\phi(t, S) = (S - K)^+$.

More work on option pricing theory in regime-switching Jump Diffusion model can be found in [1]

4.3.3 Regime Switching Model

By definition of Geometric Brownian motion we know, it follows the SDE $dS_t = \mu S_t dt + \sigma S_t dB_t$, where B_t is the Brownian motion, and μ (the drift), and σ (volatility) are constants. The solution of this SDE and stock price comes out to be $S_t = S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t}$, but our current discussion revolves around the Markov-modulated GBM under which we don't assume μ (drift) and σ (volatility) are constants instead we assume them to be function of

Continuous-time Markov chain and so the SDE of Markov-modulated GBM becomes,

$$dS_t = \mu(X(t))S_t dt + \sigma(X(t))S_t dB_t,$$

where X(t) is a Markov chain as discussed above. For example, say Market is in state 1 if the instantaneous drift is less than 0.15 and in state 2 if that is greater or equal to 0.15, so our chain will be a two state process and we can model our market using GBM as, either

$$dS_t = \mu(1)S_t dt + \sigma S_t dB_t,$$

or,

$$dS_t = \mu(2)S_t dt + \sigma S_t dB_t,$$

for a special case when $\sigma(1) = \sigma(2) = \sigma$ and combining it together would result in equation formally discussed as,

$$dS_t = \mu(X(t))S_t dt + \sigma S_t dB_t,$$

where X(t) is a Markov chain whose state space $S = \{1, 2\}$. The solution of this SDE is given by

$$S_t = S_0 e^{\int_0^t \left(\mu(X_u) - \frac{\sigma^2}{2}\right) du + \sigma B_t}.$$

We can rewrite the above equation for S_t as under.

$$S_{t_{i+1}} = S_{t_i} \ e^{\left(\left(\mu_i - \frac{\sigma^2}{2}\right) dt \ + \ \sigma(B_{t_i+dt} - B_{t_i})\right)}, \tag{4.1}$$

where $i \in S$, and S is the state space.

4.4 Simulation of MMGBM

We will now look upon an algorithm to simulate MMGBM model as it is simpler than other alternatives, using 2-state Continuous-time Markov chain to observe the dynamics of stock prices. For the purpose of simulating MMGBM, the underlying rate transition matrix, Q should be known to generate holding times from exponential distribution of continuous-time Markov chain, and also to mention that the chain under our consideration is irreducible as

it was in Chapter 2.

Algorithm 8 Simulating Markov-Modulated GBM (with drift μ and volatility σ) upto Time-horizon T, given $Q-$matrix.

1. Choose an initial state, say $X_0 = i_0$, set $n = 0$.
2. Simulate a continuous-time Markov chain with a given Q-matrix, and the time-horizon T as in Algorithm [2] and obtain transition times, say t_n.
3. Simulate $\{B_{u_i}\}_i$ as in Algorithm [5] $\forall\ u_i \leq T$, set $u_i = idt\ \forall\ i$.
4. Set a definite time-increment, say dt, and also $u_0 = 0, i = 0$.
5. Define $\mu_i = \mu(X_t)$, X_t is the underlying embedded markov chain for $t_n \leq u_i < t_{n+1}$, $\forall$ i such that $u_i \leq T$.
6. Compute $S_{u_{i+1}} = S_{u_i} \exp\left(\left(\mu_i - \frac{\sigma^2}{2}\right)dt + \sigma(B_{u_{i+1}} - B_{u_i})\right)$, as in equation 4.1 until $u_{i+1} > T$.
7. Set $i = i + 1$.
8. $u_i = idt$, go and iterate from step 4.
9. return Output (u_i, S_{u_i}), $i \geq 1$.

We obtain the following plot after simulating MMGBM. Here we have considered our Q-matrix to be,

$$Q = \begin{bmatrix} -0.33 & 0.33 \\ 0.5 & -0.5 \end{bmatrix}$$

.

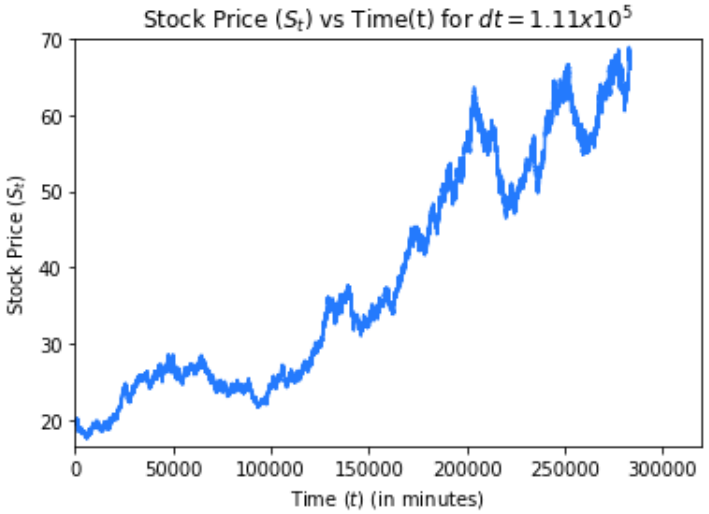


Figure 4.2: MMGBM with 2 states: $\{0, 1\}$, and $\mu = \{0.1, 0.2\}$, $\sigma = \{0.01\}$, $T = 3$ years, $S_0 = 20$ units, time increments $dt = \frac{1}{(250 \times 360)} = 1.1 \times 10^{-5}$ units.

Chapter 5

Study of Drift Coefficient

In this chapter we will be looking to answer the question, "Why the calibration of μ is important?" and to answer this question we are going to use the knowledge from the previous chapter, from [16] about several models discussed there and see if we can investigate more upon the question. Let us begin by reviewing those models in order again in regards to our question.

BSM Model: For BSM model calibration of μ is irrelevant as Black-Scholes formula for option pricing doesn't involve μ.

Heston & Jump-Diffusion Model: In Heston and Jump Diffusion model, option-price equation depends on the drift coefficient (μ). The knowledge of this coefficient is must to find out the value of the option.

Last of all the option price equation discussed for Regime-Switching (generalizations to Heston or Jump Diffusion) Model also involves the value of μ.

5.1 An approach to drift estimation

So as seen before that drift (μ) is involved in each of the modelling equations (Heston, Jump-Diffusion), but the problem of estimating it was not discussed. In this section, we are going to see the difficulty involved in estimating it through these models, although any of

the mentioned models could be used, but for our purpose, we have used Regime-Switching generalization of BSM Model. Several simulations were done in these regards alongside the case study of empirically deriving the conditions for estimating μ. The first thing is to validate the model and then look for the squeezed duration, by which we mean we will compute the moving drift (same way as we calculate moving average) for suitable window size and set a particular cutoff percentage. Henceforth, we look down how many times this moving drift falls below the specified cutoff percentage. We then collect the duration data for varying values of scale parameter θ of the exponential distribution, assuming the underlying stochastic process is continuous-time Markov chain (so Holding times $\sim \exp(\theta)$). Let us discuss the process in more details with the help of following steps:

1. First of all we run our algorithm for simulation of MMGBM(as in 8) for a fixed scale parameter θ and then use it to produce simulated stock prices data for a given Time-Horizon. This data would be minute-wise data, we do so by considering ($250 \times 6 \times 60$ minutes i.e., trading for 250 days, 6 hours $= 6 \times 60$ minutes.)

2. From this Stock prices data, we compute the relative return $r_i = \frac{S_i - S_{i-1}}{S_{i-1}}$, we do so as to see the relative change of stock prices over the day.

3. Now we choose a suitable windows size n, and then find out the moving drift over this windows (or to say drift estimators, say $\hat{\mu}_j$, at j^{th} time) using $\hat{\mu}_j = \frac{1}{dt}\frac{1}{n}\sum_{i=0}^{n-1} r_{j-i}$.

4. For computing the duration data we define,

   (a) Given $\hat{\mu}_j$, let $a_0 = 0$ and define,

   $$b_{i-1} = \min\{k > a_{i-1} \mid \hat{\mu}_k > 0\},$$

   $$a_i = \min\{k > b_{i-1} \mid \hat{\mu}_k < 0\},$$

   then squeezed duration is given by, $d_i = b_i - a_i$.

5. We will end up with several values of d_i's for a particular scale parameter θ, say $\{d_1, d_2, \cdots, d_L\}$, so one can find $\bar{d}$ by taking mean of these values, that is, $\bar{d} := \frac{1}{L}\sum_{i=1}^{L} d_i$.

More generally, we can define the squeezed duration as,

Definition 5.1.1. *An asset is said to be in p-squeeze at i-th time step if the empirical drift $\hat{\mu}_j$ defined above is below c_p, the p-th percentile of $\hat{\mu}$. We define for $i = 1, 2, \cdots$, let $a_0 = 0$ and define,*

$$b_{i-1} = \min\{k > a_{i-1} \mid \hat{\mu}_k > c_p\}$$
$$a_i = \min\{k > b_{i-1} \mid \hat{\mu}_k < c_p\}.$$

Therefore, the time durations for p-squeezes are $\{d_1, d_2, \cdots, d_L\}$ and $L = \max\{i | b_i < \infty\}$, provided $L \geq 1$ and squeezed durations are defined in similar manner as above. But, for our purpose, we will use the formerly defined squeezed duration.

5.2 An estimator of drift and its demerits

In this section we are going to talk about complexity involved in discussing above steps. To begin with we first simulate MMGBM, and then look at the standard error of drift estimator. Now, for standard we will try to look at BSM model and check if the variance is small enough to use for MMGBM model. We have the following for BSM model,

$$dS_t = S_t[\mu dt + \sigma B_t],$$

or

$$dS_t = S_t[\mu dt + \sigma N(0, dt)],$$

by considering increments of size dt. Now stock prices return can be computed as,

$$r_i = \frac{S_i - S_{i-1}}{S_{i-1}} = \mu dt + \sigma N(0, dt), r_i \sim N(\mu dt, \sigma^2 dt).$$

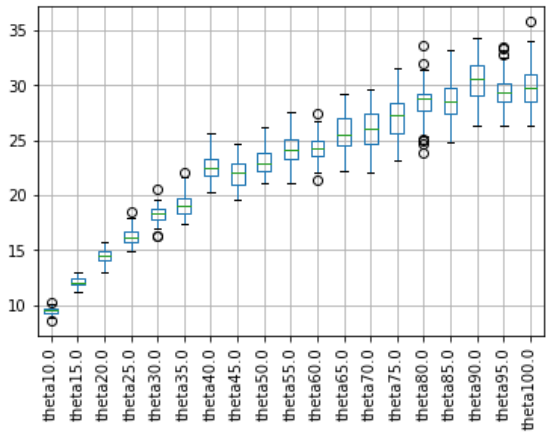
Now consider the random variable $\hat{\mu}_j^{(n)}$, which is moving drift over a windows size n. Then we have,

$$\hat{\mu}_j^{(n)} = \frac{1}{dt}\frac{1}{n}\sum_{i=0}^{n-1} r_{j-i}, \tag{5.1}$$

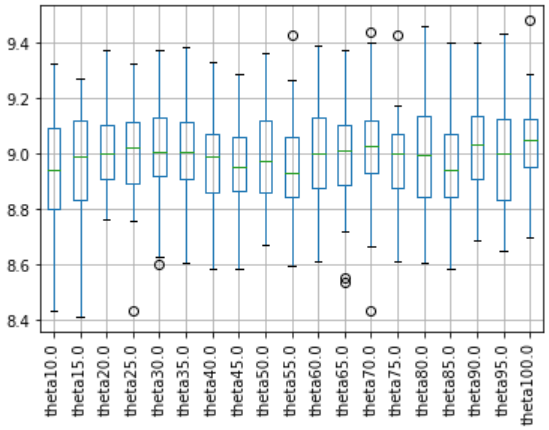
where each of these r_{j-i} are iid random variables and $Var(\hat{\mu}_j^{(n)}) = \frac{\sigma^2}{ndt}$. Now, if we consider realistic values as per market model, we have $\sigma = 0.2, dt = \frac{1}{250\times360} = \frac{1}{90000}$, and to find suitable windows size n, we have, say we want $Var(\hat{\mu}) < 0.1$, then we have,

$$\begin{aligned}\frac{(4\times10^{-2})(9\times10^{4})}{n} &< 0.1,\\ \frac{3600}{0.1} &< n\\ n &> 36000.\end{aligned}$$

But, the problem with such a large windows size is that it will average out the expected behaviour of changing drift parameter and hence this will not be healthy approach. This shows that for large value σ, our method of estimating drift will fail, while if we consider $\sigma = 0.01$, and then choose a small windows size, say $n = 20$, we get $Var(\hat{\mu}) = \frac{(1\times10^{-4})\times(9\times10^{4})}{20} = \frac{9}{20} = 0.45$. Now, comparing it with $\sigma = 0.2$, we have, $Var(\hat{\mu}) = \frac{(4\times10^{-2})\times(9\times10^{4})}{20} = \frac{3600}{20} = 180$, we get a very large variance, method fails. So, we can apply this method of estimation only for small σ. This project investigated more upon the same that for large values of σ, it is not useful to estimate drift in regime switching setting by this method. We obtain the following Box plots for MMGBM model when we plot mean squeezed duration (Sqd) of drift estimator against varying scale parameter θ of exponential distribution of 2-state Continuous-time Markov chain.



(a) $\sigma = 0.01, N = 50, n = 20, \mu_1 = -1.0, \mu_2 = 2.0$.



(b) $\sigma = 0.2, N = 50, n = 20, \mu_1 = -1.0, \mu_2 = 2.0$.

Figure 5.1: mean Squeezed duration (Sqd) vs $\theta \in [10, 100]$.

We can clearly, see for fixed μ values and window size $n = 20$, $\sigma = 0.01$, we observe

monotonic increasing behaviour as expected, while a washed out behaviour for large σ as 0.2. So, from now on we will only look at the results where values of $\sigma \leq 0.01$ and $n = 20$ as it would be beneficial for keeping the behaviour of mean squeezed duration from fading away.

5.3 Estimation of scale parameter in MMGBM model

Now, given a data the problem of estimating scale parameter θ can be done by employing the simulation technique which we will discuss in further details. For our purpose we have restricted ourselves to Binary regime case and can use probability estimate p, also we have considered 2-state, say $\{1, 2\}$ Markov chain case only. Hence the drift dynamics has 4 unknown parameters, namely, $\mu_1, \mu_2, \theta_1, \theta_2$, but we know $\bar{\mu}$ which is the observable in form time-series and is defined by $\bar{\mu} = \frac{\theta_1\mu_1+\theta_2\mu_2}{\theta_1+\theta_2}$. We can determine the value of μ from probability estimate p using $\bar{\mu} = p\mu_1 + (1-p)\mu_2$. For our purpose if we consider $\mu_1 = -1.0$, then we can obtain the value of μ_2 for a particular value of $\bar{\mu}$ and probability estimate p. Now, we proceed further to estimate our scale parameter. We discuss more on the method of simulating and estimating the scale parameter. Our original estimator of θ is the mean of N different mean squeezed durations obtained by simulation, that is running the simulation N times to obtain as many values of $\bar{d}$ for each θ. More precisely, if $\bar{d} = \frac{1}{L}\sum_{i=1}^{L} d_i$, where d_i are squeezed durations obtained from single simulation,then after N different independent simulations we will obtain $\bar{d_1}, \bar{d_2}, \cdots, \bar{d_N}$, then we set

$$m := \frac{1}{N}\sum_{i=1}^{N} \bar{d}^{(i)},$$

as the key value, to be used in deriving estimator Let us look at the algorithm for estimating scale parameter theta. From now on, we will denote mean squeezed duration by $mean(Sqd)$ wherever required.

Algorithm 9 Estimating θ for mean(Sqd)

1. Simulate MMGBM model, say for $\theta \in [\theta_c, \theta_d]$.
2. Use output from previous step to calculate mean(Sqd), say $\bar{d} \; \forall \; \theta$.
3. Repeat step 1 and step 2 for N times.
4. After obtaining N different values for each θ, obtain the mean value for each θ corresponding to N measurements, to say obtain m value discussed for each corresponding θ.
5. Fit linear regression line and obtain the slope, say a and intercept value, say b. The equation of regression line passing through these points is of the form $m = a\theta + b$.
6. Obtain the mean(Sqd), say m^* from single simulation as in model validation.
7. Estimated θ value, say θ^* is obtained by solving the equation, $\theta^* = \frac{m^* - b}{a}$

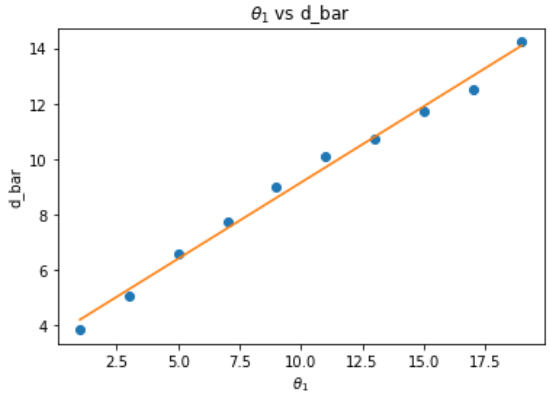
Let us look at the numerical experiment performed where we obtained 3 years data and found mean of mean(Sqd), that is m value as discussed above, then we obtain regression fit by performing above algorithm [9] for $\theta \in [1.0, 3.0, \cdots, 19]$, $\sigma = 0.01$ and the time-horizon $= 3$ years and $N = 50$ iterations.

The following are the values for linear regression fit performed on the simulated data.

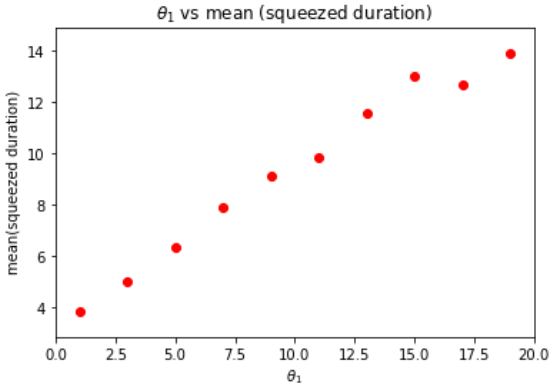
Slope (Linear Fit)	Intercept (Linear Fit)
0.55	3.66

We have the following plots to compare and infer from.

Now to empirically verify our estimation of scale parameter, we choose a particular θ value, say $\theta = 7.0$ from a single run of simulation and try to match it. The value of mean(Sqd), $m^* = 7.86$ was obtained from a single run of simulation corresponding estimated $\theta^* = 7.63$ We can match it with the following plot of a single simulation. We observe that the original scale parameter θ is very close to the estimate scale parameter but actually little less than it.



(a) Linear Regression fit on 50 iterations of mean(Sqd)



(b) Single realization of above plot with same parameters

Figure 5.2: Comparing estimated scale parameter with original.

Chapter 6

Variance Reduction of Estimator

In previous chapter we saw the method of obtaining mean(Sqd) for different θ and corresponding box plots were obtained but we didn't talk much about the variance between the normal random variables used in MMGBM simulation, while using the classical variates can lead to results having larger variance which can be reduced by employing the method on Antithetic-Variates which will be discussed in next section. Further, we will look at the difference in the estimation of both these methods using simulation algorithms from previous chapter.

6.1 Variance reduction method

In descriptive statistics, if we want to estimate, say an unknown mean $\mu = E(X)$ of a distribution, we collect n iid random variables from the given distribution, $X_1, \cdots, X_n$ and use the formula for sample mean

$$\overline{X}(n) = \frac{1}{n} \sum_{j=1}^{n} X_j. \tag{6.1}$$

Also, $\sigma^2 = Var(X)$ denote the variance of the distribution, we can make a conclusion that

(6.2)

Before we begin discussion about the method, we need to know the applicability and reason of using this particular method. In our simulation, instead of collecting data, as was the case with in classical statistics, instead we simulate the random variables and generate iid $X_1, \cdots, X_n$ r.v., so for applicability purpose, for sufficiently large n (say $n = 10^5$) and we can then use central limit theorem and therefore, our method of constructing confidence interval can also be used. Thus we can easily obtain confidence interval using this technique.

Now using more clever technique and using the simulation more smartly, we can get better estimate of the confidence intervals. For this very purpose we will try to induce and generate negatively correlated random variables $X_1, \cdots, X_n$, with same mean but smaller variance, so variance of estimator in 6.1 will become smaller and thus leading to smaller confidence interval and resulting in better estimates.

6.1.1 Method of Antithetic Variates

Let X_i denote the identically distributed random variables having the same mean μ and variance σ^2 but we are not assuming they are independent (since we want to have correlation between our random variables). Let $n = 2k$, for some $k \geq 1$, that is, n will be even. Now

$$\overline{X}(n) = \frac{1}{2k} \sum_{i=1}^{2k} X_i = \frac{1}{k} \sum_{i=1}^{k} Y_i = \overline{Y}(k), \tag{6.3}$$

where each of Y_i are defined by,

$$\begin{aligned} Y_1 &= \frac{X_1 + X_2}{2}, \\ Y_2 &= \frac{X_3 + X_4}{2}, \\ &\vdots \\ Y_k &= \frac{X_{n-1} + X_n}{2}. \end{aligned}$$

Also from equation (6.3), we can conclude that the two estimators $\overline{Y}(m)$ and $\overline{X}(n)$ for $E(X)$ are identical.

Now, we will maximize this relation to our purpose as we will show in further steps, but for following discussion we will use $Y(k)$. Also, we have $E(Y_i) = \frac{E(X)+E(X)}{2} = E(X) = \mu$ (since X_i all have the same distribution and hence the same mean). So letting $Y = \frac{X_1+X_2}{2}$ denote a generic Y_i. Now we can re-frame our problem as to estimate $\mu = E(Y)$. Now we will compute variances,

$$\begin{aligned} Var(Y) &= \frac{1}{4}\Big(Var(X_{1)} + Var(X_2) + 2Cov(X_1, X_2)\Big) \\ &= \frac{1}{4}\Big(\sigma^2 + \sigma^2 + 2Cov(X_1, X_2)\Big) \\ &= \frac{1}{2}\Big(\sigma^2 + Cov(X_1, X_2)\Big). \end{aligned}$$

Now, if we $Cov(X_1, X_2) = 0$, i.e. if X_i were iid, and thus the $Var(Y) = \frac{\sigma^2}{2}$ (as previously), and $Var(\overline{Y}(k)) = \frac{\sigma^2}{n}$. But our are assuming them to be dependent and correlated, so this case is not possible.

But, now comes the other case of having $Cov(X_1, X_2) < 0, Var(Y) < \frac{\sigma^2}{2}, Var(\overline{Y}(k)) < \frac{\sigma^2}{n}$ and so the variance will be reduced. So, our task will be: if somehow we can create some negative correlation within each pair $(X_1, X_2), (X_3, X_4), \cdots$, but the pairs themselves should be iid (for the applicability of CLT) then $Var(\overline{Y}(m))$ will be reduced than the trivial way of using iid random variables X_i's.

We can generate a negatively correlated random variables for any distribution, say $F_X(x) = \mathbb{P}(X \leq x)$ with inverse $F^{-1}(y), y \in [0, 1]$, so we could generate a negatively correlated pair as $X_1 = F^{-1}(U), X_2 = F^{-1}(1 - U)$ and since we know $F^{-1}(y)$ is a monotonic increasing function of y. The random variables U and (1-U) have a correlation with coefficient $\rho = 1$, and thus they are negatively correlated to a largest extent and thus monotonicity preserves the property of negative correlation: $\rho_{X_1,X_2} < 0$ (not -1 always though).

Generalizing the above fact, we can say that for a monotone function (increasing or decreasing) g in each variable, it can be indeed shown that the pair of random variables $X_1 = g(U_1, \cdots, U_m)$ and $X_2 = g(1-U_1, \cdots, 1-U_m)$ are negatively correlated and are called as antithetic variates. One important thing to know is $U_1, \cdots, U_m$ and $1 - U_1, \cdots, 1 - U_m$ have the same distribution, then so will X_1 and X_2 have; they will have the same mean $E(X)$, but the only difference being they have induced negative correlation (whenever g is monotone), henceforth the two are negatively related random variables ($Cov(X_1, X_2) < 0$),

where each $U_i \sim U(0,1)$. Now, let us look forth to algorithm for simulating antithetic variates to estimate mean μ.

Algorithm 10 Algorithm for simulating antithetic variates to estimate mean $\mu = E(X)$, where $X = g(U_1, \cdots, U_m)$ is a monotonic function in U_i

1. Generate uniform random variables $U_1, \cdots, U_k$, where $U \sim U(0,1)$. To construct first pair of antithetic variates, set $X_1 = g(U_1, \cdots, U_m)$ and $X_2 = g(1 - U_1, \cdots, 1 - U_m)$.
2. Now, generate independently m new iid uniform random variables and construct another pair X_3, X_4 as before, and so on for further pair as we reach m (suitably large) desired number of pairs.
3. Now use the estimate which we found previously,

$$\overline{Y}(k) = \sum_{j=1}^{k} Y_j,$$

where,

$$\begin{aligned} Y_1 &= \frac{X_1 + X_2}{2}, \\ Y_2 &= \frac{X_3 + X_4}{2}, \\ &\vdots \\ Y_k &= \frac{X_{2k-1} + X_{2k}}{2}. \end{aligned}$$

6.2 Implementation of antithetic method

In this section we will look at implementing antithetic methods in the context of MMGBM simulation and compare it with method of using classical variates. First using we will simulate MMGBM model using algorithm [8] for varying values of θ and then we will obtain mean(Sqd) and repeat this for N_1 iterations to obtain the N_1 values of mean(Sqd) $\forall\ \theta$. But since we want to observe the variance reduction in these mean(Sqd) values, we will find

out mean of the N_1 value of mean(Sqd) obtained and repeat this procedure, say N_2 times and by this method we will end up with N_2 values. We will repeat the same procedure for antithetic method but this time instead of taking B_t in MMGBM simulation we will use both B_t and $-B_t$ and repeat the simulation for $\frac{N1}{2}$ iteration rather than $N1$ iterations and carry on as previously mentioned. We will again obtain N_2 values for mean(Sqd) for antithetic method. Now, next step will be to compare the Standard deviation of the values obtained from classical variates with that of antithetic variates. Now, let us look at the numerical experiment performed for the same. So in antithetic case, we will have $\tilde{m}$, i.e. the mean of mean(Sqd) for N observations, precisely

$$\tilde{m} = \frac{1}{N}\sum_{i=1}^{N} \bar{d}^{(i)} \text{ (N is even) },$$

as mentioned in previous section to apply antithetic property on simulation, where $\bar{d}(2i-1)$, $\bar{d}(2i)$ are from antithetic pair respectively.

Numerical Experiment

As was mention previously, we are going to consider a value of σ which is less than 0.01. For the purpose of numerical experiment we have considered θ from $[5, 10, \cdots, 100]$ and $\sigma = 10^{-2.5} \approx 0.0032$, $\mu_1 = -1.0$, $\mu_2 = 2.0$, $N_1 = 50$ iterations, $N_2 = 20$ and window size as mentioned in previous chapter will be fixed at $n = 20$. After this we will find out the values of 20 different values of standard deviation for both classical variates and antithetic variates and the difference of the corresponding values should be positive as classical variates should have larger standard deviation than antithetic variates. First of all us try to envisage why and how antithetic approach in this case will work.

Now for the considered experiment we have MMGBM model with 2-state Continuous-time Markov chain governing the behaviour of it and σ is constant. So let us see the model behaviour using following equations for antithetic behaviour considering equations where we will have W_t and $-W_t$, negatively correlated random variables as the Brownian motion process:

$$dS_t^1 = S_t^1(\mu_t dt + \sigma dW_t),$$

so solution of above SDE will be,

$$S_t^1 = S_0 e^{\int_0^t (\mu_u - \sigma^2)du + \sigma dW_u}.$$

and second equation is,

$$dS_t^2 = S_t^2(\mu_t dt - \sigma dW_t),$$

and for this we have the solution,

$$S_t^2 = S_0 e^{\int_0^t (\mu_u - \sigma^2)du - \sigma dW_u}.$$

Now for the drift estimator, we consider return from the two equations is,

$$r_t^1 = \mu_t dt + \sigma dW_t,$$

$$r_t^2 = \mu_t dt - \sigma dW_t.$$

Assume $ndt = \Delta$, where ndt is as in equation (5.1) and define $\frac{1}{\Delta}\sigma(W_{t+\Delta} - W_t) = Z_t \sim N\left(0, \frac{\sigma^2}{\Delta}\right)$, $\forall t$. Now, writing drift estimator for both equations used above in terms of integration we have,

$$\hat{\mu_t}^1 = \frac{1}{\Delta}\int_t^{t+\Delta} \mu_t dt + Z_t,$$

where $Z_t \sim N\left(0, \frac{\sigma^2}{\Delta}\right)$ as before. Similarly, we have:

$$\hat{\mu_t}^2 = \frac{1}{\Delta}\int_t^{t+\Delta} \mu_t dt - Z_t.$$

The moving average of actual drift (which is not observed) can then be defined as:

$$\hat{\mu_t}^0 = \frac{1}{\Delta}\int_t^{t+\Delta} \mu_u du.$$

Hence $\hat{\mu_t}^1$ and $\hat{\mu_t}^2$ are negatively correlated noisy observation of $\hat{\mu_t}^0$.

Here we show one illustrative plot of $\hat{\mu_t}^0$ and $\hat{\mu_t}^1$ against t, where $\mu_1 = -1.0, \mu_2 = 2.0, \sigma = 0.01$. Here the smooth line corresponds to $\hat{\mu_t}^0$, whereas the zig-zag line corresponds to $\hat{\mu_t}^1$. Interestingly, $\hat{\mu_t}^2$ (which is not drawn) would just be the reflection of $\hat{\mu_t}^1$ with respect to $\hat{\mu_t}^0$.

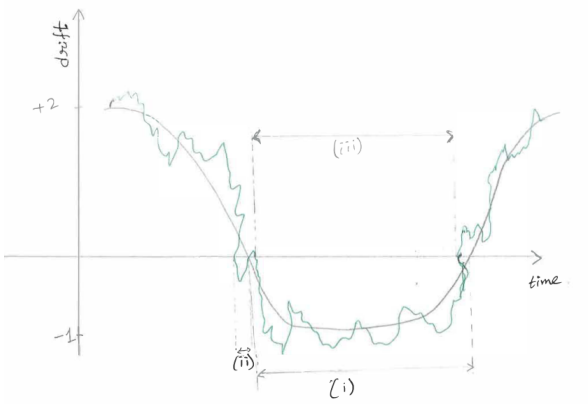


Figure 6.1: Estimated behaviour of MMGBM model

Observation

We can observe the curve in which the duration is defined as the interval during which the drift estimator falls below 0 and then moves up 0. To be more precise (i) shows the duration of $\hat{\mu_t}^0$, whereas $(ii$ and (iii) gives that of $\hat{\mu_t}^1$. The interval (i) is further denoted by $[a_i^0, b_i^0]$, whereas (ii) and (iii) by $[a_i^+, b_i^+]$. Now, we can visualize it even further by assuming the incoming points to arrive at time-axis to be a_i^+, a_i^- of $\hat{\mu_t}^1$ and $\hat{\mu_t}^2$ and b_i^+, b_i^- to be the outgoing points coming from $\hat{\mu_t}^1$ and $\hat{\mu_t}^2$ respectively. Then, we have the following cases for a_i^+, b_i^+ and a_i^-, b_i^-.

$$a_i^+ > a_i^0 \iff a_i^- < a_i^0.$$

$$b_i^+ > b_i^0 \iff b_i^- < b_i^0.$$

There are only four cases possible and if the above two implications are true then the following are also true,

$$a_i^+ < a_i^0, \text{ and } b_i^+ > b_i^0,$$

$$d_i^+ > d_i^0,$$

and also,

$$a_i^- > a_i^0 \text{ and } b_i^- < b_i^0,$$

$$d_i^- < d_i^0.$$

Now, we can see all possible four cases are covered, also both d_i^+ and d_i^- can't be of the same sign with above implications. We can then successively define $\bar{d}$ and m, mean of mean(Sqd) for our purpose.

Now, moving onto the numerical experiment we can graphically see the behaviour of the Standard deviation of m, mean of mean(Sqd) with different values of θ as below.

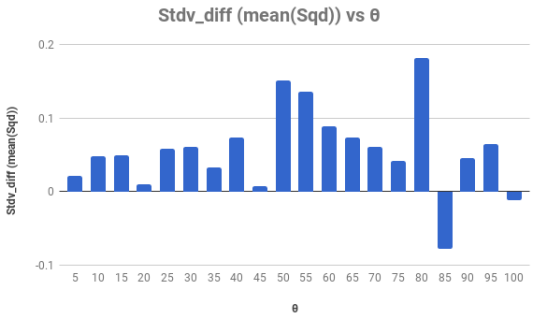


Figure 6.2: Plot for Stdev_diff (mean(Sqd)) vs θ

The following is the table for the values obtained doing numerical experiment with above-mentioned values for parameters.

Table 6.1: Numerical values for above experiment.

θ	**Stdev. (Classical)**	**Stdev. (Antithetic)**	**Stdev. (diff.)**
5	0.0577367611	0.0359386832	0.0217980779
10	0.079494534	0.0313407427	0.0481537914
15	0.1009686248	0.0516524004	0.0493162245
20	0.0795935088	0.0701220665	0.0094714423
25	0.1318180055	0.0735437574	0.058274248
30	0.1451433792	0.084037411	0.0611059682
35	0.1124458149	0.079468698	0.0329771169
40	0.1685399351	0.0947249054	0.0738150297
45	0.1144421889	0.1073623293	0.0070798595
50	0.2462864286	0.0958376658	0.1504487627
55	0.2463423197	0.110849276	0.1354930438
60	0.1878360133	0.0991509208	0.0886850925
65	0.2027822357	0.1299753555	0.0728068802
70	0.1866740955	0.1254993606	0.0611747349
75	0.2079093798	0.1663525028	0.041556877
80	0.403347321	0.2215420894	0.1818052316
85	0.1409312091	0.2192182348	-0.0782870257
90	0.3229277453	0.2777608255	0.0451669198
95	0.2701569909	0.2052659026	0.0648910883
100	0.2290936326	0.2412089429	-0.0121153103

We can see most of the times difference between Standard Deviation of Classical variate and antithetic is positive except few times. One explanation is that this negative Std. Deviation will tend to change with different simulation each time and will be marginal. Alternatively, it can be explained by assuming the Sqd mean for 50 iterations to be scattered around a point, so when we use antithetic method then we end up closer to the actual estimated mean value while in classical variates the chances of being far away from actual mean are high, but there can be times when antithetic mean is little more far away from actual mean than obtained. But even then this difference will be small, and we see from the above plot that is indeed the case.

Chapter 7

Conclusion

The problem of estimating drift requires the knowledge of several key asset pricing processes as Markov chain, where for Discrete-time Markov chain we obtained empirical results and studied the relevance about the theoretical results using [3], Brownian motion which were studied in sufficient details in their respective chapters. Further, the way these processes are beneficial in modelling the stock price dynamics were studied via several financial models, and further we focused and restricted our approach to regime switching GBM and used this model for the extensive study of drift dynamics and scale parameter estimation. The knowledge of all these processes and models would have been incomplete without empirical verification via computation codes. So, we had simulated most of the asset price processes using several useful techniques for our analysis and study. Now, with many simulations involved we also looked into much important methods of reducing variance using antithetic approach and verified it empirically.

The content discussed, and numerical experiments performed in this thesis are a subset of the broader pool, where we can study the semi-Markov modulated GBM with parametric holding time distribution from a broader family of distribution functions, such as Gamma, Weibull, Truncated exponential, Beta, etc. Now, this project has helped in the cause of studying the well known MMGBM model which will help in studying the semi-Markov modulated GBM. The course of the project helped in assimilating the ideas from the different field of mathematics and finance which will be helpful for the further study of financial models.

Bibliography

[1] Omkar Manjarekar Anindya Goswami and Anjana Ramachandran. "Option Pricing in a Regime-Switching Jump Diffusion Model". In: *working paper* ().

[2] Maurice Stevenson Bartlett. *An introduction to stochastic processes: with special reference to methods and applications*. CUP Archive, 1978.

[3] Patrick Billingsley. "Statistical Methods in Markov Chains". In: *The Annals of Mathematical Statistics* 32.1 (1961), pp. 12–40. ISSN: 00034851. URL: http://www.jstor.org/stable/2237603.

[4] Arunangshu Biswas and Ludger Overbeck. "Option pricing in a regime-switching stochastic volatility model". In: *to appear in Statistics & Probability Letters* ().

[5] Fischer Black and Myron Scholes. "The pricing of options and corporate liabilities". In: *Journal of political economy* 81.3 (1973), pp. 637–654.

[6] Massimiliano Bonamente. *Statistics and Analysis of Scientific Data. Graduate Texts in Physics*. Springer, 2017.

[7] George Casella and Roger L Berger. *Statistical inference*. Vol. 2. Duxbury Pacific Grove, CA, 2002.

[8] H. Cramér. *Mathematical Methods of Statistics*. Princeton Mathematical Series. Princeton University Press, 1999. ISBN: 9780691005478. URL: https://books.google.co.in/books?id=CRTKKaJO0DYC.

[9] Milan Kumar Das and Anindya Goswami. "Testing of binary regime switching models using squeeze duration". In: *working paper* ().

[10] Anindya Goswami and Ravi Kant Saini. "Volterra equation for pricing and hedging in a regime switching market". In: *Cogent Economics & Finance* 2.1 (2014). DOI: 10.1080/23322039.2014.939769. URL: http://doi.org/10.1080/23322039.2014.939769.

[11] Ulf Grenander. "Stochastic processes and statistical inference". In: *Arkiv för matematik* 1.3 (1950), pp. 195–277.

[12] Steven G Kou. "A jump-diffusion model for option pricing". In: *Management science* 48.8 (2002), pp. 1086–1101.

[13] Tanmay S. Patnakar. "Asset Pricing in a Semi-Markov Modulated Market with Time-dependent Volatility". MS Thesis. IISER Pune, Apr. 2016. URL: https://arxiv.org/pdf/1609.04907v1.pdf.

[14] Sheldon M Ross. *Introduction to probability models*. Academic press, 2014.

[15] Halsey Lawrence Royden and Patrick Fitzpatrick. *Real analysis*. Vol. 2. Macmillan New York, 1968.

[16] Steven E Shreve. *Stochastic calculus for finance II: Continuous-time models*. Vol. 11. Springer Science & Business Media, 2004.