A GENERALIZED CORRELATED RANDOM WALK APPROXIMATION TO FRACTIONAL BROWNIAN MOTION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

BUKET COŞKUN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
STATISTICS

JULY 2018

Approval of the thesis:

A GENERALIZED CORRELATED RANDOM WALK APPROXIMATION TO FRACTIONAL BROWNIAN MOTION

submitted by **BUKET COŞKUN** in partial fulfillment of the requirements for the degree of **Master of Science in Statistics Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** ______________

Prof. Dr. Ayşen Dener Akkaya
Head of Department, **Statistics** ______________

Assoc. Prof. Dr. Ceren Vardar Acar
Supervisor, **Statistics Department, METU** ______________

Examining Committee Members:

Prof. Dr. İnci Batmaz
Department of Statistics, METU ______________

Assoc. Prof. Dr. Ceren Vardar Acar
Department of Statistics, METU ______________

Assoc. Prof. Dr. Seher Nur Sülkü
Department of Econometrics, Ankara Hacı Bayram Veli Uni. ______________

Date: ______________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: BUKET COŞKUN

Signature :

ABSTRACT

A GENERALIZED CORRELATED RANDOM WALK APPROXIMATION TO FRACTIONAL BROWNIAN MOTION

Coşkun, Buket
M.S., Department of Statistics
Supervisor : Assoc. Prof. Dr. Ceren Vardar Acar

July 2018, 70 pages

The application of fractional Brownian Motion (fBm) has drawn a lot of attention in a large number of areas, ranging from mathematical finance to engineering. The feature of long range dependency limited due to the value of Hurst parameter $H \in (1/2, 1)$ makes fBm the desired process for stochastic modelling. The simulation of fBm is also vital for the application in such fields. Hence, the development of an algorithm to simulate an fBm is required in both theoretical and practical aspects of fBm. In this study, we mainly propose a new fBm generation method by using the Hurst parameter and the correlation structure based on this parameter and suggest an algorithm to generate correlated random walk converging to fBm, with Hurst parameter, $H \in (1/2, 1)$. The increments of this random walk are simulated from Bernoulli distribution with proportion p, whose density is constructed using the link between correlation of multivariate Gaussian random variables and correlation of their dichotomized binary variables. We prove that the normalized sum of trajectories of this proposed random walk yields a Gaussian process whose scaling limit is the fBm.

Keywords: Fractional Brownian Motion, Simulation, Random Walks, Discretization, Convergence

ÖZ

GENELLEŞTİRİLMİŞ İLİŞKİLİ RASSAL YÜRÜYÜŞÜN KESİRLİ BROWN HAREKETİNE YAKINSAMASI

Coşkun, Buket
Yüksek Lisans, İstatistik Bölümü
Tez Yöneticisi : Doç. Dr. Ceren Vardar Acar

Temmuz 2018, 70 sayfa

Kesirli Brown hareketinin (kBh) uygulamaları matematiksel finanstan mühendisliğe kadar bir çok alanda dikkat çekmektedir. Hurst parametre değerinin $H \in (1/2, 1)$ olması nedeniyle sınırlanan uzun dönem bağımlılık özelliği kBh'yi stokastik modelleme için aranan süreç yapmaktadır. kBh 'nin simülasyonu bu alanlardaki uygulamaları açısından da önemlidir. Bu yüzden fBm üreten bir algoritmanın gelitirilmesi kBh için hem teorik hem de pratik açıdan gereklidir. Bu çalışmadaki temel amacımız $H \in (1/2, 1)$ olan, bu Hurst parametresini ve parametreye bağlı korelasyon yapısını kullanarak yeni bir kBh üretme yöntemi ve kBh'ye yakınsayan ilişkili rassal yürüyüş süreci üreten bir algoritma önermektir. Bu rassal yürüyüşlerdeki artışlar, çok değişkenli Gauss tipi rassal değişkenin korelasyonu ile onun ikili olarak kesikli halinin korelasyonu arasındaki ilişki kullanılarak ulaşılan dağılama sahip p oranından gelen Bernolli dalmndan üretilir. Bizde bu önerilen rassal yürüyüşten normalleştirilmiş toplamlarının ölçeklendirilmiş limiti kBh olan Gaussian sürecini ürettiğini kanıtladık.

Anahtar Kelimeler: Kesirli Brown Hareketi, Simulasyon, Rassal Yürüyüş, Kesikli hale Getirme, Yakınsama

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

a.s	Almost Surely
Bm	Brownian Motion
CLT	Central Limit Theorem
CRW	Correlated Random Walk
fBm	Fractional Brownian Motion
FFT	Fast Fourier Transform
fGn	Fractional Gaussian Noise
iid	Independent Identically Distributed
LRD	Long Range Dependence
MVN	Multivariate Normal
RMD	Random Midpoint Displacement
SRD	Short Range Dependence

CHAPTER 1

INTRODUCTION AND MOTIVATION

Most of the real data displaying long-range dependence can be modeled with self-similar processes. Fractional Brownian motion (fBm) is one of the simplest models demonstrating long-range dependence. In recent years, this phenomenon has become quite popular due to its applications in many areas. For instance, in the field of mathematical finance, Roger [31] has proposed an fBm model for the movement of share prices. In communication systems, Leland et al. [24] use the increments of fBm for modeling of Ethernet local area network (LAN) traffic. In biology, Lim and Muniandy [26] use the discrete-time version of fBm to model the non-coding sequence of human DNA by recognizing a DNA sequence as a fractal random walk.

Fractional Brownian motion has introduced by Kolmogorov [21] for the development of turbulence theory. It has named after Mandelbrot and Van Ness [28] since they define a fractal integral representation with respect to a standard Brownian motion, which can be generalized to fBm in the presence of dependent increments.

An fBm, denoted by $B^H(t)$, is a centred Gaussian process with stationary increments such that $E[B^H(t)B^H(u)] = \frac{C^2}{2}(t^{2H} + u^{2H} - |t-u|^{2H})$ for all $t, u \geq 0$, where C denotes the scale parameter, and H denotes the Hurst parameter or the parameter of self- similarity. It is already known that the increments of fBm can be either positively or negatively correlated depending on the Hurst parameter. In particular, an fBm with parameter $H = 1/2$ corresponds to a standard Brownian Motion. For $H \in (0, \frac{1}{2})$, its increments are negatively correlated and display short-range dependence. In contrast, for $H \in (\frac{1}{2}, 1)$, the auto-covariance of the fBm increments is positive. Thus two consecutive increments tend to have the same directions. Such fBm is called persistent. In addition, for such H, fBm has the long-range dependence property. Thanks to this property, fBM has been applied to many areas of science. Another nice characteristic of fBm is that it has stationary Gaussian increments, called fractional Gaussian noise (fGn), whose cumulative sum displays an fBm sample.

The simulation (or generation) of fBm has drawn a lot of attention due to its applications in diverse areas. In the literature, there is a large number of simulation methods. For instance, the method studied by Hosking [18] implicitly computes the fGn covariance matrix. The Cholesky method proposed by Asmussen [2] also uses the Cholesky decomposition. This method is applied to the same matrix, but the covariance matrix of fBm can also be used for Cholesky method. Another approach is the fast Fourier transform (FFT) method developed by Davies and Harte [9]. In this method, FFT algorithm is used in order to generate an fGn sample. Then, the covariance matrix of fGn is buried in a circulant covariance matrix, and this circulant matrix is diagonalized with FFT algorithm. In another study, the integral representation introduced by Mandelbrot and Van Ness [28] is used for a direct approximation of fBm. Another approach of generating an fBm process is the wavelet transform method. The method firstly generate the wavelet coefficients referring to an orthogonal basis. The fBm is reached via an inverse wavelet transformation. The fast implementation of this method is given in Abry and Sellan [1]. The other approximate and fast technique is the random midpoint displacement method proposed by Lau et al. [23]. He used an approach based on counting of the conditional distribution of fGn like the Hosking method. ts only difference is the generation based on the condition distribution given the last certain points instead of all past points.

Due to the content of this study, we also concentrate on several random walk approximations of fBm. Donsker's theorem expresses that standard Brownian motion can be constructed by random walks. As an analogue of this theorem, fBm can also be constructed by random walks. Taqqu [35] uses the normal random variables to show convergence. Dasgupta [8] shows this approximation by using the binary random variables and the stochastic integral representation of fBm. On the other hand, Sottinen [33] defines a random walk which convergences weakly to fBm by using a kernel function that converts the standard Bm to fBm for long-range dependence case. Szabados [34] uses moving average of an appropriate nested sequence of random walks uniformly converge to fBm when $H \in (\frac{1}{4}, 1)$. This approximation use the discrete form of moving average representation. Enriquez [16] proves that normalized correlated random walk converges weakly to fBm. The construction relies on correlated random walks including discrete time processes such that the distribution of each jump is a function of the preceding jump, which is defined as the parameter of persistence. Konstantopoulos and Sakhanenko [22] has introduced scaled random walks which use the weighted sum of iid random variables, converges to fBm under the sufficient condition for the weak convergence of normalized sums to fBm with $H > \frac{1}{2}$. Lindstrom [27] provides the same approximation with Konstantopoulos [22] for the case $H < \frac{1}{2}$.

The motivation of this study is to propose an fBm generation method by using the Hurst parameter and a correlation structure based on this parameter. For this purpose, we generate a correlated random walk which converges to fBm. In order to show this convergence, we write our theorem and prove it theoretically. In light of this theorem, we present a new simulation algorithm for fBm by using the correlated random walk. Our fBm construction is the generalization of the construction given by Enriquez [16]. His construction depends on the persistent random walks with a persistence parameter, which corresponds to the probability of producing the same jump as the last. As an enhancement to Enrique's study [16], our fBm construction use correlated random walks depending on a correlation structure counted by using the given Hurst parameter and the discretization proportion. Moreover, we use the relationship between the correlation of multivariate normal random variables and the correlation of their discretized version in order to establish a link between the persistence parameter and the discretization proportion.

This thesis is organized as follows. Chapter 2 provides the definitions and notations used throughout the paper. In addition, some important fBm simulation methods are expressed. The realization and discretization of fBm are presented in Chapter 3. Furthermore, After discretization, simulation results are given. Chapter 4 is devoted to the generation of fBm. The generation is performed in two steps. First, a relationship between the discretization proportion and persistence parameter is provided. Then the convergence of the correlated random walk to fBm is proved. In Chapter 5, a new algorithm is proposed to generate correlated random walks which converge to fBm. We give a brief review of this study and state our main results in the final chapter.

CHAPTER 2

DEFINITIONS, NOTATIONS AND SIMULATION METHODS

2.1 Random Walk

The random walk is the stochastic process that is one of the most fundamental topics in probability theory. The term "random walk" was originally proposed by Pearson [30] in a letter to Nature in 1905. In this paper, he asks his readers a question. Suppose a man starts its walk at a point and walks one step in a straight line, then he walks another step. After he repeats this process n times, he wants to know the distance he takes in n steps. Lord Rayleigh, physicist and 1904 Nobel physics winner, solved this question asymptotically. In earlier 1900s, the theory of random walk was also developed by economist Louis Bachelier in his thesis. He proposed the random walk as a financial time series model.

It is obviously seen that the random walk model is quite versatile and interdisciplinary. The question of Pearson stemmed from modelling of mosquito infestation in a forest, which is related to biology. Rayleigh's work related to physics and Bachelier's to economics. Today, the theory of random walk is useful in many disciplines like chemistry, ecology, computer sciences, psychology as well as biology, physics and economics.

After expressing the history of the random walk, now let us present the formal definition.

2.1.1 Definition and Properties

Assume $X_1, X_2, ...$ are independent, identically distributed (iid) binary random variables. Each takes the value 1 with the probability of p and -1 with the probability of $q = 1 - p$. The integer-time stochastic process $\{S_n, n \geq 1\}$ which is a sum of these iid random variables, is called a random walk.

$$S_n = \sum_{i=0}^{n} X_i. \tag{2.1}$$

If $p = q = \frac{1}{2}$, the random walk is called symmetric random walk. For instance, if a fair coin is flipped repeatedly, the coin will be either tail with the probability of p or will be head with the probability of $1 - p$. Suppose that the respective probabilities of tail and head are represented by p and $1 - p$, then S_n is represented by the number of heads minus the number of tails in the first n toss.

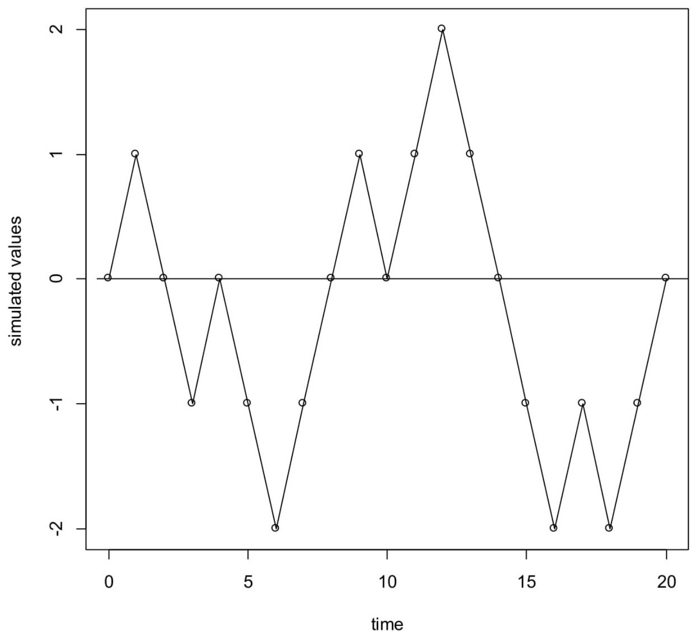

Figure 2.1: Random walk simulation for n=20 step size

Basic properties of a symmetric random walk are presented as follows:

i The random walk starts at time zero, $S_0 = 0$.

ii The expectation of a symmetric random walk process is equal to zero. The expectation of each increment can be calculated as $E(X_i) = \frac{1}{2} \times 1 + \frac{1}{2} \times -1 = 0$. Since the expected value of each increment takes the value of 0. The expectation of random walk is defined as $E(S_n) = E(X_1 + X_2 + ... + X_n) = E(X_1) + E(X_2) + ... + E(X_n) = 0$.

iii The variance of a symmetric random walk process equals the number of steps by independence. Similarly, the variance is obtained by $V(S_n) = V(X_1 + X_2 + ... + X_n) = V(X_1) + V(X_2) + ... + V(X_n) = n$ by independence of the increments. Note that each increment is distributed with variance $V(X_i) = E(X_i)^2 - [E(X_i)]^2 = 1$.

iv A symmetric random walk process has independent increments. For any finite set of discrete times n_i such that $i = 1, ..., k$ with $0 \leq n_0 \leq n_1 \leq ..... \leq n_k$, the increments $(S_{n_1} - S_{n_0}), (S_{n_2} - S_{n_1}), ..., (S_{n_k} - S_{n_{k-1}})$ are independent from each other. These increments are distributed with zero mean $E[S_{n_{i+1}} - S_{n_i}] = 0$ and finite variance $Var[S_{n_{i+1}} - S_{n_i}] = n_{i+1} - n_i$.

$$E[S_{n_{i+1}} - S_{n_i}] = E(\sum_{j=n_i}^{n_{i+1}} X_j) = \sum_{j=n_i}^{n_{i+1}} E(X_j) = 0 \tag{2.2}$$

$$Var[S_{n_{i+1}} - S_{n_i}] = Var(\sum_{j=n_i}^{n_{i+1}} X_j) = \sum_{j=n_i}^{n_{i+1}} Var(X_j) = \sum_{j=n_i}^{n_{i+1}} 1 = n_{i+1} - n_i \tag{2.3}$$

v A symmetric random walk has Markov property. This property states that the future state of the process does not depend on past only when the current state is given. $(S_n, n\epsilon Z_+)$ is Markov if and only if, for all $n\epsilon Z$, $E[S_{n+1}|S_n, ..., S_1] = E[S_{n+1}|S_n]$.

vi A symmetric random walk is a martingale. A martingale is defined as an integer-time stochastic process $\{Z_n; n \geq 1\}$ with the properties $E|Sn| < \infty$ for all $n \geq 1$. In Equation (2.1), X_i's are identically and independently distributed with zero mean. By applying the Markov property, the conditional expectation can be rewritten as

$$\begin{aligned} E[S_{n+1}|S_n] &= E[X_{n+1} + S_n|S_n, ..., S_1] \\ &= E[X_{n+1}] + E[S_n] \\ &= E[S_n]. \end{aligned} \tag{2.4}$$

2.1.2 Distribution of Random Walk

We can benefit from the Bernoulli process to find the distribution of S_n. Let us convert the random walk into Bernoulli distribution by using the transformation $Y_i = \frac{1}{2}(X_i + 1)$. Since X_i takes the value 1 or -1, Y_i takes the 1 or 0, respectively. D_n represents the sum of Y_i. From literature, it is well known that the sum of Bernoulli trials follows the Binomial distribution. Therefore, $D_n = \sum_{i=1}^{n} \frac{1}{2}(X_i + 1)$ follows the binomial distribution with parameters n and $\frac{1}{2}$. When $\sum_{i=1}^{n} X_i$ is substituted by S_n, D_n is $\frac{1}{2}(S_n + n)$. Then, the probability of S_n can be written as

$$P(S_n = y) = P(2D_n - n = y) = P(D_n = \frac{y+1}{2}) = \binom{n}{n+y}\frac{1}{2^n} \tag{2.5}$$

for $r = -n, -(n-2), ..., (n-2), n.$

2.2 Brownian Motion

In 1827, the Scottish Botanist Robert Brown discovered a motion of the pollen grains of some plants in liquids. While examining this movement, he realized that the place of the pollen particles had changed in a random motion. This phenomenon was described as Brownian motion in the paper of Brown [5]. However, the correct answer for the reason of pollen particles' motion could not be given until the discovery of the kinetic theory by Einstein [14]. He noticed that this motion stemmed from the collusion of the molecules in the liquid. After the physical construction of Brownian motion by Brown [5] and Einstein [14], the mathematical foundation was given by some mathematicians like Weiner [36], Donsker [13], Kolmogorov [21], Levy [25]. In recent years, this stochastic process has become a very significant one that is widely used in many disciplines such as economics, finance, biology, mathematical statistics, physics and management science. Especially, nowadays, it is used to model the stock prices and financial markets.

Definition 1 *A Brownian motion (Bm)* $\{W_t\}$ *at time* $t \geq 0$ *defined on the probability space* (Ω, F, P) *is a real-valued stochastic process such that*

i W_t *has almost surely continuous path.*

ii For all $0 \leq t_1 \leq t_2 \leq ..... \leq t_k$ *and* $k \geq 0$, $W_{t_1} - W_{t_0}, W_{t_2} - W_{t_1}, ..., W_{t_k} - W_{t_{k-1}}$ *are independent random variables, that is,* $W(t)$ *has independent increments.*

iii For all $0 \leq u \leq t$ *and* $h + u \geq 0$*, the law of* $W_t - W_u$ *is the same with the law of* $W_{t+h} - W_{u+h}$*. Besides, the law of* $W_t - W_u$ *normally distributed with expectation zero and variance* $t - u$*. Hence,* W_t *has stationary and normal increments.*

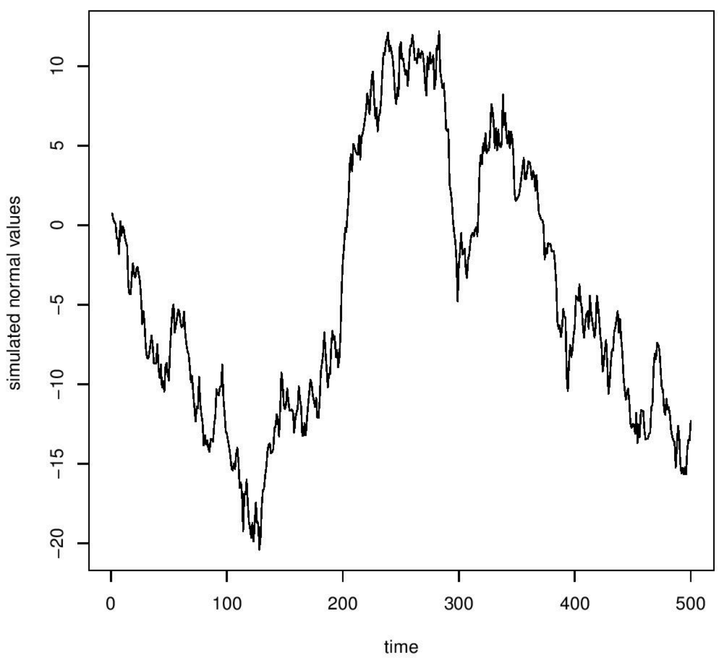

Figure 2.2: Brownian motion simulation for n=500 step size

Since the increments of a Brownian motion are distributed normally, the linear transformation of independent Gaussian random vector $(W_{t_1}, W_{t_2}, ..., W_{t_k})$ is jointly normally distributed for all $0 \leq t_1 < t_2..... \leq t_k$. Note that if all vectors are Gaussian, then the stochastic process W_t is also a Gaussian. Thus, the process $\{W_t : t \geq 0\}$ is a Gaussian process with zero mean and covariance function $Cov(W_t, W_u) = \sigma^2 min(u, t)$. The covariances, for all $u \leq t$, between increments is given by

$$Cov(W_t, W_u) = E(W_t W_u) - E(W_u)E(W_t) = E(W_t W_u).$$

Now substituting $(W_t - W_u) + W_u$ in W_t, we obtain

$$\begin{aligned} Cov(W_t, W_u)) &= E(((W_t - W_u) + W_u)W_u) \\ &= E((W_t - W_u)W_u) + E({W_u}^2). \end{aligned} \tag{2.6}$$

As the increments are independent, the covariance between increments is then given by

$$Cov(W_t, W_u) = E({W_u}^2) = Var(W_u) = \sigma^2 u \tag{2.7}$$

Similarly, for all $t \leq u$, the covariance equals to $\sigma^2 t$, or equivalently, $Cov(W_t, W_u) = \sigma^2 min(u, t)$. Moreover, Brownian motion is a Markov process. Assume now that $\{W_t : t \geq 0\}$ is a stochastic process. The Markov property says that if we know the process $\{W_t : t \geq 0\}$ on the interval $[0, s]$, for the prediction of the future $\{W_t : t \geq s\}$ this is as useful as just knowing the endpoint W_s.Suppose that W_t is a stochastic process on probability space with filtration $\mathcal{F}_t$, collection of σ algebra $(\Omega, \mathcal{F}, \mathcal{P})$, is called a Markov process if

$$\begin{aligned} P(W_{t+u}|\mathcal{F}_t) &= P(W_{t+u}|W_0, W_1, W_2, ..., W_t) \\ &= P(W_{t+u}|W_t) \end{aligned} \tag{2.8}$$

holds for every $u \geq t$. Note that if W_t is a Bm, the process $W_{t+u} - W_u$ is independent of the process W_u by independency of the increments.

Besides, Brownian motion is a martingale, that is, it owns the martingale property. For fix $0 \leq u \leq t$, a stochastic process W_t with filtration $\mathcal{F}_t$ is a martingale if the process W_t is integrable and the statement $E(W_t|\mathcal{F}_u) = W_u$ is almost surely true.

Recall that Bm has zero mean, $E(W_t) = 0$. Hence, $E(W_t)$ is obviously integrable since the condition $E(W_t) \leq \infty$ holds. Then

$$E[W_t|\mathcal{F}_u] = E[(W_t - W_u) + W_u|\mathcal{F}_u] = E[W_t - W_u|\mathcal{F}_u] + E[W_u|\mathcal{F}_u]. \qquad (2.9)$$

Since the process $W_t - W_u$ is independent of $\mathcal{F}_t$, we have $E(W_t - W_u|\mathcal{F}_u) = E(W_t - W_u) = 0$. We get

$$E[W_t|\mathcal{F}_u] = W_u + E[W_t - W_u] = W_u. \qquad (2.10)$$

Thus, W_t is a martingale.

Brownian motion has one basic property, which is called the invariance property. This property identifies a transformation on the space of functions, which changes the individual Brownian random functions but leaves their distribution unchanged. If a Bm is exposed to transformations, these transformed ones are also Bm. After transformations, their distributions remains stable but their Brownian functions are altered. Suppose that the process W_t is a standard Bm. The scaling version $\beta^{-1}W_{\beta^2 t}$, the time shifted version $W_{t+\alpha} - W_t$, the symmetric form $-W_t$, and the time inversion form $tW_{1/t}$ is also a standard Bm for any $\alpha, \beta > 0$ and $t > 0$.

2.3 Fractional Brownian Motion

Firstly, Kolmogorov [21] has introduced fractional Brownian motion with the name of "Wiener spiral". Then Mandelbrot and Vann Ness [28] have determined its name as fractional Brownian motion (fBm). They have also suggested a representation of fBm with respect to the standard Brownian motion. fBm is the abbreviation of fractional Brownian motion which is generally used in the literature. The dependence of fBm increments is qualified by a parameter, called Hurst, named after hydrologist Harold Edwin Hurst. In the study of Hurst [19], he has used the Hurst index to define the irregularity of Nile River in rate of flow. Moreover, he has used R/S analysis, which is one of the techniques in time series analysis, for modelling hydrological processes like flow, precipitation to estimate the Hurst parameter.

Definition 2 *A fractional Brownian motion is a centered Gausssian process* $\{B^H(t), t \geq 0\}$ *with the covariance function*

$$E[B^H(t)B^H(u)] = \frac{1}{2}(t^{2H} + u^{2H} - |t-u|^{2H}). \qquad (2.11)$$

This process has a parameter $H \in (0,1)$, called the Hurst parameter.

The fBm satisfies the following properties:

- $B^H(0) = 0$, the processes is beginning at point 0.
- $B^H(t)$ has stationary increments. This means the process $B^H(t+u) - B^H(u)$ has the same distribution as $B^H(t)$ for $u, t \geq 0$.

$$\begin{aligned}
& E[(B^H(t+k) - B^H(k))(B^H(u+k) - B^H(k))] \\
= \; & E[B^H(t+k)B^H(u+k)] - E[B^H(u+k)B^H(k)] - E[B^H(k)B^H(u+k)] \\
+ \; & E[B^H(k)B^H(k)] \\
= \; & \frac{1}{2}[t+k]^{2H} + [u+k]^{2H} - |t-u|^{2H} - [(t+k)^{2H} + k^{2H} - t^{2H}] \\
- \; & [(u-k)^{2H} + k^{2H} - u^{2H}] + 2k^{2H}]E[B^H(1)] \\
= \; & \frac{1}{2}(t^{2H} + u^{2H} - |t-u|^{2H})E[B^H(1)] = E[B^H(t)B^H(u)] \qquad (2.12)
\end{aligned}$$

- $B^H(t)$ is a Gaussian process with the variance of $B^H(t) = t^{2H}$ for all $t \geq 0$ and $H \in (0,1)$. We know

$$V[B^H(t)] = E[(B^H(t))^2] - E[B^H(t)]^2. \qquad (2.13)$$

The expectation of $B^H(t)$ is zero by the definition of Brownian motion, then

$$E[(B^H(t))^2] = \frac{1}{2}(t^{2H} + t^{2H} - |t-t|^{2H}) = t^{2H} \qquad (2.14)$$

- $B^H(t)$ has continuous trajectories.
- There is a correlation between two increments. The increments of fBm are also called fractional Gaussian noise (fGn), which can be written as $[B^H(t)] -$

$[B^H(t-1)]$. The distribution of fGn is Gaussian with zero mean and covariance function, for $u + k \leq t$ and $t - u = km$,

$$\begin{aligned} \rho^H(n) &= Cov(B^H(t+k) - B^H(t), B^H(u+k) - B^H(u)) \\ &= \frac{1}{2}k^{2H}[(m+1)^{2H} + (m-1)^{2H} - 2m^{2H}], \end{aligned} \tag{2.15}$$

where H represents the Hurst parameter.

Equation (2.15) will be zero for $H = 1/2$. Then this equation corresponding to a standard Brownian motion owns independent increments. However, for $H \neq 1/2$, the increments are not independent. If $1/2 < H < 1$, the increments are positively correlated. For this H, the fBm is said to be persistent. This means that the direction (up or down) of a jump is more likely to be followed by a jump with the same direction. In other words, when the fBm declines in the past, then declining in the future is more possible.

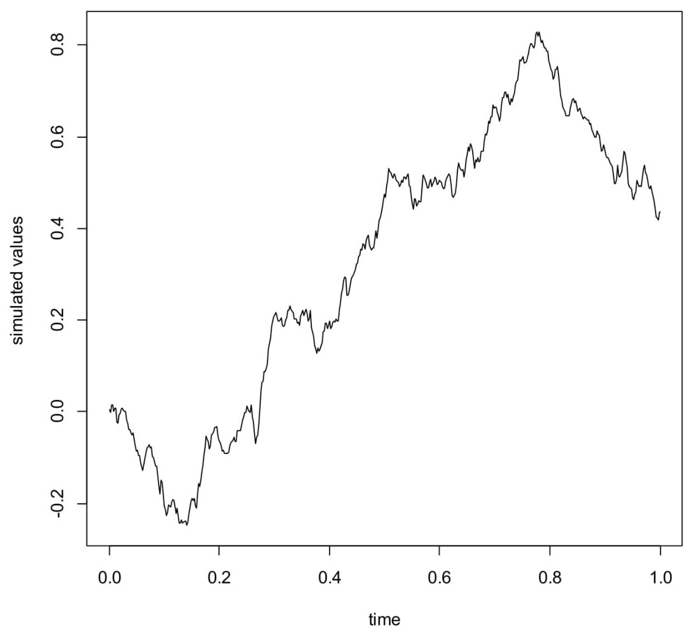

Figure 2.3: Fractional Brownian motion path with H=0.7

Contrarily, if $0 < H < 1/2$, fBm is said to have negatively correlated increments, a property of anti-persistence. This means that a jump up is more likely to be followed by a jump down. For $0 < H < 1/2$, they have negative correlation. This case refers to short range dependence. In other words, it is anti-persistent, that is, when the fBm declines in the past, then rising in the future is more possible, or conversely.

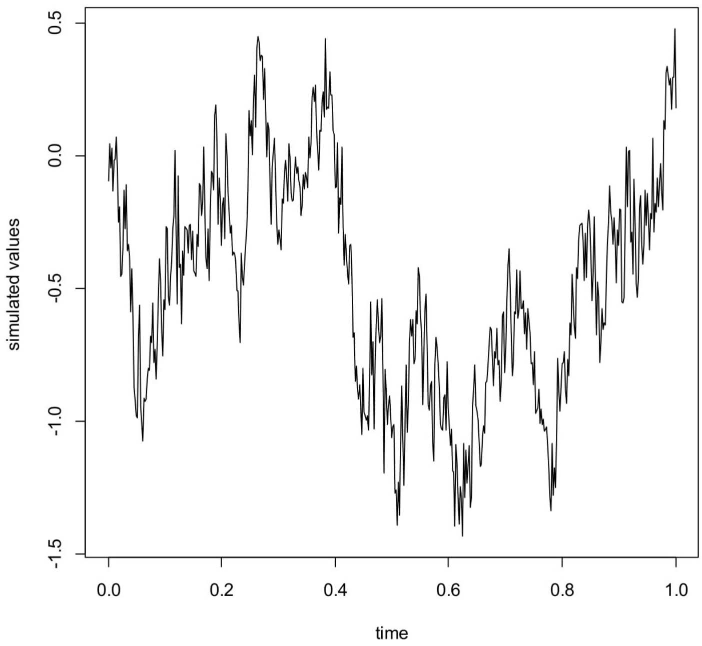

Figure 2.4: Fractional Brownian motion path with H=0.3

- The increments of fBm display long-range dependence when $H > \frac{1}{2}$. A stationary sequence $(Y_m)_{m \in N}$ has the characteristic of long-range dependence with the condition that $\rho(m) = Cov(Y_h, Y_{h+m})$ ensures $lim_{n \to \infty} \frac{\rho(m)}{bm^{-\beta}} = 1$, where b is any constant and $\beta \in (0, 1)$. Therefore, when m goes to infinity, Equation (2.15) converges to $H(2H-1)m^{2H-2}$. When $lim_{n \to \infty} \frac{\rho^H(m)}{H(2H-1)m^{2H-2}} = 1$, fBm increments have long-range dependence for the case $H > \frac{1}{2}$.

- fBm $B^H(t)$ is a self similar process as $B^H(\lambda t)$ and $\lambda^{-H} B^H(t)$ have the same laws.

2.4 Simulation Methods of Fractional Brownian Motion

Recently, fBm has been very popular thanks to its application in various fields such as economics, finance, engineering, hydrology, and communication. Hence, many people have shown a lot of interest in the simulation of fBm. Both exact and approximate techniques are proposed for the simulation of fBm because of the fractal structure which is hard to manage in the numerical calculation. In this part of the study, we present the most widely used methods to simulate a fractional Brownian motion.

2.4.1 Hosking Method

The Hosking method which is proposed by Hosking [18], is based on the generation of a stationary Gaussian process depending on a covariance function. This algorithm simulates a fractional Gaussian sequence Z_k. In other words, it calculates the conditional distribution Z_{r+1} given the past samples $Z_r, ..., Z_1, Z_0$. In fact, this conditional distribution is Gaussian with expectation and variance which are functions of r. Therefore, recursive simulation of Z_{r+1} can be done by generating standard normal random variable Z_0.

Assume that $\gamma(k) = E(X_r X_{r+k})$ represents the autocovariance function and let us take $\gamma(0) = 1$ for simplicity. $M(r) = \gamma(ij)_{i,j=0,...,r}$ denotes the covariance matrix, $d(r)$ is the $(r+1)$ column vector with elements $d(r)_k = \gamma(k+1), k = 0, ..., r$. Let us define the matrix $G(r) = (1(i = r - j))_{i,j=0,...,r}$ such that 1 stands for the indicator function. Hence, the matrix $M(r+1)$ can be decomposed as

$$M(r+1) = \begin{bmatrix} 1 & d(r)^T \\ d(r) & M(r) \end{bmatrix} = \begin{bmatrix} M(r) & G(r)d(r) \\ d(r)^T G(r) & 1 \end{bmatrix}.$$

The conditional distribution can be expressed with mean $\mu_{r+1} = E(Z_{r+1}|Z_r, ..., Z_0)$ and variance $\delta^2_{r+1} = Var(Z_{r+1}|Z_r, ..., Z_0)$. Then X_{r+1} can be generated from a standard normally distributed Z_0. Since we know μ_n, δ^2_r and $\iota_r = d(r)^T G(r) M(r)^{-1} d(r)$, next estimation of the mean and the variance can be generated recursively by using the recursion

$$\delta_{r+1}^2 = \delta_r^2 - \frac{(\lambda(r+2) - \iota_r)^2}{\delta_{r+1}^2}. \tag{2.16}$$

In order to generate an fBm sample, the cumulative sum can be computed from the initial recursion μ_1, δ_1^2 and ι_0.

2.4.2 Cholesky Method

The Cholesky method introduced by Asmussen [2] uses the Cholesky decomposition of the covariance matrix. It implies that the covariance matrix $\Pi(r)$ can be separated as $L(r)L(r)^T$, where $L(r)$ denotes an $(r+1)\times(r+1)$ lower triangular matrix and "T" stands for the transpose. Assume that l_{jk} represents the element (j,k) of the matrix $L(r)$ for $j,k = 0,1,...,r$. Then $L(r)$ corresponds to the lower triangular matrix if $l_{jk} = 0$ for $k > j$. Note that this separation is possible only when $L(r)$ is a symmetric positive definite matrix.

The matrix scheme can be written as

$$\begin{bmatrix} \Upsilon(0) & \Upsilon(1) & \dots & \Upsilon(r) \\ \Upsilon(1) & \Upsilon(0) & \dots & \Upsilon(r-1) \\ \vdots & \vdots & \ddots & \vdots \\ \Upsilon(r) & \Upsilon(r-1) & \dots & \Upsilon(0) \end{bmatrix} = \begin{bmatrix} l_{00} & 0 & 0 & 0 \\ l_{10} & l_{11} & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{00} & l_{01} & \dots & l_{00} \end{bmatrix} \times \begin{bmatrix} l_{00} & l_{10} & \dots & l_{r0} \\ 0 & l_{11} & \dots & l_{r1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & l_{rr} \end{bmatrix}.$$

It can easily be seen that $l_{00}^2 = \Upsilon(0)$, $l_{10}l_{00} = \Upsilon(1)$ and $l_{10}^2 + l_{11}^2 = \Upsilon(0)$ for $j = 1$. If $j \geq 1$, the entries of the matrix $L(r)$ can be identified by

$$\begin{aligned} l_{j0} &= \frac{\Upsilon(j)}{l_{00}}, \\ l_{j,k} &= \frac{1}{l_{kk}}\Big(\Upsilon(j-k) - \sum_{i=0}^{k-1} l_{ji}l_{ki}\Big) \quad 0 < k \leq n, \\ l_{jj}^2 &= \Upsilon(0) - \sum_{i=0}^{j-1} l_{ji}^2. \end{aligned}$$

Then, $Y_{r+1} = \sum_{i=0}^{r+1} l_{r+1,i}V_i$ generates the fractional Gaussian noise sequence for given iid standard normal variables $(V_j)_{j=0,...,r+1}$. If we consider this in a matrix form, the main idea is to simulate $Y(r) = L(r)V(r)$ recursively. When $\Upsilon(r)$ is a positive definite matrix, this guarantees the non-negativity of l_{jj}^2. Then, the covariance structure of $Y(r)$ can be obtained as

$$Cov(Y(r)) = Cov(L(r)V(r)) = L(r)Cov(V(r))L(r)^T = L(r)L(r)^T = \Upsilon(r). \tag{2.17}$$

Although the numerical implementation of the Cholesky method is simple, it has drawbacks because of the storage in memory.

2.4.3 Fast Fourier Transform Method

This approach has firstly proposed by Davies et al. [9]. In later times, both Dietrich and Newsam [12] and Wood and Chan [37] have also developed this method. Similar to Hosking and Cholesky techniques, the aim of this method is to derive the square root of covariance matrix. Suppose that sample size is M and for $a \in M$, the size of covariance matrix is $M = 2^a$. For obtaining the square root of covariance matrix, the key idea is to embed covariance matrix in a circulant covariance matrix with size $2M = 2^{a+1}$. Then, the circulant covariane matrix G can be defined as

$$\begin{bmatrix} \upsilon(0) & \upsilon(1) & \dots & \upsilon(M-1) & 0 & \upsilon(M-1) & \dots & \upsilon(2) & \upsilon(1) \\ \upsilon(1) & \upsilon(0) & \dots & \upsilon(M-2) & \upsilon(M-1) & 0 & \dots & \upsilon(3) & \upsilon(2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \upsilon(M-1) & \upsilon(M-2) & \dots & \upsilon(0) & \upsilon(1) & \upsilon(2) & \dots & \upsilon(M-1) & \upsilon(0) \\ 0 & \upsilon(M-1) & \dots & \upsilon(1) & \upsilon(0) & \upsilon(1) & \dots & \upsilon(M-2) & \upsilon(M-1) \\ \upsilon(M-1) & 0 & \dots & \upsilon(2) & \upsilon(1) & \upsilon(0) & \dots & \upsilon(M-3) & \upsilon(M-2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \upsilon(1) & \upsilon(2) & \dots & 0 & \upsilon(M-1) & \upsilon(M-2) & \dots & \upsilon(1) & \upsilon(0) \end{bmatrix}$$

Here, $\upsilon(.)$ represents the covariance function of fGn. The algorithm can be conducted by the following theorem. If every circulant matrix G is split into $G = CQC^T$, where Q is the dioganal matrix of eigenvalues of G, and C stands for the unitary matrix identified as

$$(C)_{ij} = \frac{1}{\sqrt{2M}} exp(-2\pi k \frac{ij}{2M}), \quad \text{for } i, j = 0, ..., 2M-1 \tag{2.18}$$

where $i = \sqrt{-1}$. The matrix Q can be formed from eigenvalues given by

$$\lambda_k = \sum_{i=0}^{2M-1} w_i exp(2\pi k \frac{ij}{2M}), \quad \text{for } i,j = 0,...,2M-1 \tag{2.19}$$

where w_i refers to $i+1$ th element of the first row of G. In order to obtain an fBm sample, below steps can be followed.

Step 1) Find the eigenvalues by using Equation (2.19).

Step 2) Compute $V = C^T Q$. For the derivation of the matrix V, apply the following simulation scheme;

- Derive two standard normal random variables V_0 and V_M
- Derive two independent standard normal random variables $Q_i^{(1)}$ and $Q_i^{(2)}$ and obtain

$$\begin{aligned} V_i &= \frac{1}{\sqrt{2}} Q_i^{(1)} + k Q_i^{(2)} \\ V_{2M-i} &= \frac{1}{\sqrt{2}} Q_i^{(1)} - k Q_i^{(2)} \end{aligned}$$

Step 3) Calculate $X = CQ^{1/2}V$ such that

$$X_k = \frac{1}{\sqrt{2M}} \sum_{i=0}^{2M-1} \sqrt{\lambda_i} V_i exp(2\pi k \frac{ij}{2M}) \tag{2.20}$$

More efficiently, this computation is applied by Fast Fourier Transformation. A sample of fGn can be generated by taking the first M elements of X. The advantages of this method its speed.

Note that the function "fbm(H,n)" allows the generation of fBm with hurst parameter H and lenght n by fast Fourier transform method in R. This function can be reach from "somebm" package in R.

2.4.4 Stochastic Representation Method

Recall that fBm is defined by Mandelbroth and Van Ness [28] through stochastic integral representation in terms of Brownian Motion. The main point is to approximate this integral through Riemann sums in order to simulate the process.

By approximating the integral representation

$$B^H(t) = \frac{1}{\Gamma(\frac{1}{2}+H)}\Big(\int_{-\infty}^{0}(t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}dB(s) + \int_0^t (t-s)^{H-\frac{1}{2}}dB(s)\Big), \tag{2.21}$$

where $B(s)$ represents standard Bm and Γ. denotes gamma function, if the first integral part of Equation (2.21) is truncated, say at $-b$, then the approximation given by

$$\tilde{B}^H(t) = C_H\Big(\sum_{k=-b}^{0}[(n-k)^{H-1/2} - (-k)^{H-1/2}B_1(k)] + \sum_{k=0}^{n}(n-k)^{H-1/2}B_2(k)\Big), \tag{2.22}$$

for $n = 1, 2, ..., N$ where C_H is the constant depending on H, B_1 and B_2 represents the iid standard normal vectors. If the truncation parameter increases, the approximation is more efficient. However, this method is not best way to simulate the fBm despite its simplicity.

2.4.5 The Wavelet-Based Synthesis Method

The wavelet-based simulation procedure is proposed by Arby and Sellan [1]. They suggest a wavelet representation which decorrelates the high-frequencies, that is,

$$B_H(t) = \sum_{k=-\infty}^{\infty}\phi_H(t-k)S_h^k + \sum_{i=0}^{\infty}\sum_{k=-\infty}^{\infty}\lambda_i(k)2^{iH}\psi_H(2^i t - k) - b_0, \tag{2.23}$$

where b_0 represents an arbitrary constant, S_h^k is a partial sum of fractional ARIMA $(0, H-\frac{1}{2}, 0)$, $\lambda_i(k)$ is iid Gaussian random variables for $i \geq 0$, $k \in \mathcal{Z}$, ϕ_H is a properly selected fractional scaling function, and ψ_H is a wavelet. Especially, the functions ϕ_H and ψ_H are identified via a connected orthogonal scaling function and a wavelet associated with a multiresolution analysis.

Decorrelation of the high frequencies of Equation (2.23) which means the independency of the Gaussian coefficients $\lambda_i(k)$, enables a fast simulation of fBm. Let

$$S_h^k(\iota) = 2^{\iota}(H+1)\int_{\mathcal{R}}(B_H(t)+b_0)f(2^{\iota t-k})dt \tag{2.24}$$

be the conveniently normalized approximation coefficients in the wavelet extention of fBm at the scale $2^{-\iota}$, where the function $f:\mathcal{R}\to\mathcal{R}$ is biorthogonal to the scaling function ϕ_H. The algorithm includes high and low pass filters, symbolized by $v^{(r)}$ and $u^{(r)}$ with $r=\frac{1}{2}+H$, respectively. As indicated by Arby and Sellan [1], the fractional filters $v^{(r)}$ and $u^{(r)}$ enable the link

$v^{(r)} = g^{(r)} * v, \quad u^{(r)} = f^{(r)} * v,$

where v and u are the high and low-pass filters associated with the initial Multiresolution analysis, * represents a convolution and the filters $g^{(r)} = \{g_n^{(r)}\}$ and $f^{(r)} = \{f_n^{(r)}\}$ are identified by z-transformations as

$g^{(r)} = (1+z^{-1})^r = \sum_{n=-\infty}^{\infty} g_n^{(r)} z^{-n},$

$f^{(r)} = (1-z^{-1})^{-r} = \sum_{n=-\infty}^{\infty} f_n^{(r)} z^{-n}.$

In practice, as the filters $v^{(r)}$ and $u^{(r)}$ are infinite and $g_n^{(r)}$ may diverge as $n\to\infty$, Arby and Sellan [1] propose to set

$$v^{(r)} = g^{(1)} * tg^{(d)} * v, \quad u^{(r)} = f^{(1)} * tf^{(d)} * u \tag{2.25}$$

where $d = H-\frac{1}{2}$, $tg^{(d)}$ and $tf^{(d)}$ denote $g^{(d)}$ and $f^{(d)}$ truncated at some a priori defined cutoff level.

Here, the opinion is to generate a fractional ARIMA $(0, H+\frac{1}{2}, 0)$ sequence of a finite length and use Equation (2.24) with truncated filters given by Equation (2.25) to generate a much longer process $S_h^k(\iota)$ at required approximation level ι. The properly normalized sequence $S_h^k(\iota)$ is taken for the approximation of fBm at the scale $2^{-\iota}$.

Note that the simulation can easily be done by MATLAB with "wfbm(H,L)" code which returns fBm signals of the Hurst parameter $0<H<1$ and length L. This code follows the algorithm proposed by Arby and Sellan [1].

2.4.6 Random Midpoint Displacement

Lau et al.[23] put forward to the random midpoint displacement (RMD) approach based on the counting of the conditional distribution of fGn like Hosking method. Its only difference is the generation of the conditional distribution Z_{n+1} given last certain generated points rather than all past sample points. The general form of RMD, called Conditionalized RMD, is also proposed by Narros [29]. The only difference from RMD is to use more sample points in conditional part.

Let $W(t)$ denote the fBm and $X_{j,k}$ is the fGn of a certain interval k in a given level j. Initially, the conditional distribution of $\{Z(\frac{1}{2})|Z(0)Z(1)\}$ is computed. Then, fGn becomes $X_{j,k} = Z(k2^{-j}) - Z((k-1)2^{-j})$ for $j = 1, ..., k,\ k = 1, ..., 2^i$. For a different j, a scaled sample of fGn with size 2^i has a relation given by

$$X_{j,2k-1} + X_{j,2k} = X_{j-1,k} \tag{2.26}$$

An fBm sample can be obtained by taking sum of these scaled fGn sample. Due to Equation (2.26), it is sufficient to generate $X_{j,k}$ for odd k. Suppose that the sample points $X_{j,1}, ..., X_{j,2m}$ are generated for $m \in 0, 1, ..., 2^{i-1} - 1$. Then, by condition at the past values, the point $X_{j,2m+1}$ can be computed as

$$\begin{aligned} X_{j,2m-1} = {} & e(j,m)(X_{j,max(2m-s+1,1)}, ..., X_{j,2m}, X_{j-1,m+1}, .., X_{j-1,min(m+r,2^{j-1})})' \\ & + \sqrt{\varphi(j,m)}V_{j,m}, \end{aligned} \tag{2.27}$$

where $e(j,m)$ is a row vector such that

$$\begin{aligned} & e(j,m)(X_{j,max(2m-s+1,1)}, ..., X_{j,2m}, X_{j-1,m+1}, .., X_{j-1,min(m+r,2^{j-1})})' \\ & = E[X_{j,2m-1}|(X_{j,max(2m-s+1,1)}, ..., X_{j,2m}, X_{j-1,m+1}, .., X_{j-1,min(m+r,2^{j-1})})'] \end{aligned}$$

and $V_{j,m}$ represents an independent set of Gaussian variables, $\varphi(j,m)$ stands for the scalar given by

$$Var[X_{j,2m-1}|(X_{j,max(2m-s+1,1)}, ..., X_{j,2m}, X_{j-1,m+1}, .., X_{j-1,min(m+r,2^{j-1})})']$$

Here, the expectation and the variance are conditioned on the number of intervals which is shown by integers s and m. For more detailed information, see Dieker [11].

2.4.7 Simulation by Random Walks

In our study, we propose random walks that converge to fBm. Therefore, we focus specially on random walk approximation to fBm which are available in the literature. Donsker theorem expresses that a standard Bm can be constructed by the random walk. As analogue of this theorem, fBm can also be constructed by random walks.

Sottinen [33] defines a random walk which converges weakly to fBm by using following kernel representation of fBm in respect of a standard Brownian Motion for the case $H > \frac{1}{2}$ as follows:

$$B(t) = \int_0^t k(t,s)dW_s, \tag{2.28}$$

where W is a standard Bm and $k(t,s)$ is the kernel function. He proposes a theorem that a random walk $B^{(n)}(t)$ can be constructed by using the transformation given in Equation (2.28). Let

$$B^{(n)}(t) = \int_0^t k^{(n)}(t,s)dW_s^{(n)}, \tag{2.29}$$

where k represents the kernel, $W_s^{(n)}$ denotes the scaled random walk such that $W_s^{(n)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} \epsilon_i^{(n)}$ where $\epsilon_i^{(n)}$ denotes the iid random variable with zero mean and variance 1. Then the random walk $B^{(n)}(t)$ converges weakly to fBm.
He shows the convergence in two steps. First is shpwing that the following finite-dimensional distribution of random walk $Z^{(n)}$,

$$X^{(n)} = \sum_{j=1}^{a} c_j Z^{(n)}(t_j) \tag{2.30}$$

for random $c_1, c_2, \cdots c_a$ and $t_1, t_2 \cdots, t_a \in [0,T]$, converge to the finite-dimensional distribution of fBm. Second part includes the tightness of the random walk $Z^{(n)}$.

Konstantopoulos [22] introduces the random walks which are obtained by the weighted sums of iid random variables under the condition given in Equation (2.33). Initially, the following random variable $\{Y_i, i \in Z\}$ can be defined by the weighted sum of random walks $\{X_j, j \in Z\}$.

$$Y_i = \sum_{j=-\infty}^{\infty} X_j w_{i-j} \tag{2.31}$$

where $\{w_i\}$ represents the weights.In order to generalize it, they define a scaled process for $f(n) > 0$ as

$$Z_{n,t} = (f(n))^{-1} \sum_{j=1}^{[nt]} X_i. \tag{2.32}$$

Konstantopoulos [22] also asserts that if the condition

$$V_n = \sum (w_{j+1} + \cdots + w_{j+n}) \sim Ln^{2H} \qquad as \qquad n \to \infty \tag{2.33}$$

is satisfied, then the process $Z_{n,t}$ converges weakly to fBm.

Lindstrom [27] provides the same approximation with Konstantopoulos [22]. Conversely, he is interested in the case $H < 1/2$. Eventhough, for $H > 1/2$, it is adequate to show Equation (2.33) to prove the convergence. For $H < 1/2$, due to the delicate calculation of the integral, Lindstrom [27] uses the representation given in Equation (2.21).

On the other hand, Szabados [34] utilizes moving average of an appropriate nested sequence of simple random walks uniformly converge to fBm for the case $H \in (\frac{1}{4}, 1)$. This approximation use the discrete form of moving average representation in given Equation (2.21) by Mandelbrot and Vann Ness [28].

2.4.8 Construction by Correlated Random Walk

Taqqu [35] uses the normal random variables to show the convergence. In his study, he proves the weak convergence of normalized sum of stationary random variables that display long range dependence. He states the theorem as

Theorem 1 *The sequence B_t^N weakly converges to $\bar{B}_t$ with the properties;*

i $\bar{B}_t$ *is almost surely continuous.*

ii $\bar{B}_{(0)} = 0$

iii $\bar{B}_t$ *is self similar with order* H *i.e;*

$$P(\bar{B}_{kt_1} \leq y_1, \bar{B}_{kt_2} \leq y_2, \cdots, \bar{B}_{kt_q} \leq y_q)$$
$$= P(k^H \bar{B}_{t_1} \leq y_1, k^H \bar{B}_{t_2} \leq y_2, \cdots, k^H \bar{B}_{t_q} \leq y_q)$$

iv For $\lambda \leq \frac{1}{H}$, $E(\bar{B}_t) = 0$ *and* $E|\bar{B}_t|^\lambda < \infty$.

if the sequence B_t^N *on Skorokhod space* $D[0,1]$ *presented by Sokorokhod [32] as an option of uniform topology to work on weak convergence of a stochastic process with jumps and satisfies the conditions;*

i $B_t^N = \frac{M_{[Nt]}}{L(N)N^{2H}}$ *where* $M_N = \sum_{j=1}^{N} X_j$, L *is slowly varying function and* X_i *represents the stationary sequence with zero mean.*

ii As $N \to \infty$, $E(M^N)^2 = O(N^{2H}L(N))$.

iii As $N \to \infty$, $E|(M^N)|^{2d} = O((E(M^N)^2))^d)$.

iv When $N \to \infty$, *finite dimensional distribution of* B_t^N *converge.*

Notably, the proof of this theorem is provided in his paper based on Bilingsley [4].

Lemma 1 *For a stationary Gaussian Sequence* Y_i *with* $E(Y_i) = 0$, $E(Y_i^2) = 1$ *and correlation* $r(k) = E(Y_i Y_{i+k})$*;*

$$Z_N(t) = \frac{1}{a_N} \sum_{i=1}^{Nt} Y_i \tag{2.34}$$

under the condition that ${a_N}^2$ *asymptotically equals to* $N^{2H}L(N)$, *that is* $Var(\sum_{i=1}^{Nt} Y_i)$ *asymptotically proportional to* $N^{2H}L(N)$ *as* $N \to \infty$, *weakly converges to* $\sqrt{C}B^H(t)$ *as* $N \to \infty$ *if the condition*

$$\sum_{j=1}^{N} \sum_{i=1}^{N} r(j-i) \sim CN^{2H}L(N) \qquad \textit{as} \quad n \to \infty \tag{2.35}$$

,

where C *is a positive constant and* L *represents a slowly varying function for* $r(k) \sim$

$k^{2H-2}L(k)$ *in the case* $H > 1/2$ *and* $r(k) \sim -k^{2H-2}L(k)$ *in the case* $H < 1/2$ *with* $r(0) + 2\sum_{k=1}^{\infty} r(k) = 0$.

Here, the notation "$\sim$" corresponds to asymptotic equivalence. He also states that the same consequences are satisfied for ${a_N}^2 * \sim \frac{2}{2H(2H-1)}N^{2H}L(N)$ which we have used by showing our convergence theorem. Moreover, Enriquez [16] proves that normalized CRW converges weakly to fBm. This paper is more crucial since we propose a correlated random walk which generalizes the construction of persistent random walk mentioned in Enriquez [16]. Thus, we underline this method in this section. He separates fBm into two cases as $H \geq \frac{1}{2}$ and $H < \frac{1}{2}$. We focus on the construction for the first case throughout this study. He defines the correlated random walk with a parameter, called the persistence parameter ρ, referring to the probability of the next jump being the same as the previous jump with jump sizes of -1 and $+1$.

Definition 3 *A discrete process* X^{ρ} *is called the correlated random walk having persistence parameter* ρ *for all* $\rho \in [0, 1]$ *if the process satisfy following conditions;*

i *The process is starting at 0,* $X_0^{\rho} = 0$.

ii *All* X_j' *s are identically distributed with probability* $1/2$, $P(X_j = -1) = 1/2$ *and* $P(X_j = +1) = 1/2$.

iii *For all* $n \geq 1$*, the jump or increment* $\epsilon_n^{\rho} = X_n^{\rho} - X_{n-1}^{\rho}$ *is almost surely equal to* -1 *and* $+1$.

iv *For all* $n \geq 1$, $\rho = P(\epsilon_{n+1}^{\rho} = \epsilon_n^{\rho} | \sigma(X_k^{\rho}, 0 \leq k \leq\ n)$

Furthermore, the correlation between two increments ϵ^{ρ} with n step distance is calculated as follows.

Proposition 1 *For all* $n \geq 0$ *and* $m \geq 1$*, we have*

$$E[\epsilon_m^{\rho}\epsilon_{m+n}^{\rho}] = (2\rho - 1)^n$$

Then, P^ρ is defined, which is the law of correlated random walk X^ρ. After adding randomness to persistence, instead of persistence ρ, probability measure μ on $[0, 1]$ is assigned and used. Hence, the law of correlated random walk is related to μ. Therefore, the law of correlated random walk is based on measure μ and X^ρ is P^μ : $\int_0^1 P^\rho d\mu(\rho)$.
Moreover, the representation of the increments ϵ^ρ are converted to $\epsilon^\mu = X_n^\mu - X_{n-1}^\mu$. Thus, Proposition 1 is modified as below:

Proposition 2 *For all $n \geq 0$ and $m \geq 1$ we have*

$$E[\epsilon_m^\mu \epsilon_{m+n}^\mu] = \int_0^1 (2\rho - 1)^n d\mu(\rho)$$

Then, the proof of convergence theorem depending on the construction of the CRW, is given in Enriquez [16].

Theorem 2 *For $H \in (\frac{1}{2}, 1)$, let μ^H with density $2^{3-2H}(1-H)(1-\rho)^{1-2H}$ denote the probability on $[\frac{1}{2}, 1]$ and $(X^{\mu^H,i})_{i\geq 1}$ be a sequence of independent process with law P^{μ^H}. Then*

$$L^D \lim_{N\to\infty} L \lim_{M\to\infty} a_H \frac{X_{[Nt]}^{\mu^H,1} + X_{[Nt]}^{\mu^H,2} ... + X_{[Nt]}^{\mu^H,M}}{N^H\sqrt{M}} = B_H(t)$$

where $c_H = \frac{H(2H-1)}{\Gamma(3-2H)}$, L and L^D denotes the convergence in distribution and the convergence in the sense of weak convergence in the Skolorohod topology on $D[0, 1]$, the space of cadlag functions on $[0, 1]$.

It is obviously seen that CRW X^μ with the law of P^μ weakly converges to fBm $B_H(t)$ by Lemma1. This convergence is shown in two steps. First one is that the summation of great number of paths converges to a discrete Gaussian process by CLT. Secondly, when this Gaussian process, which satisfies the correlation condition stated in Taqqu [35], is scaled, it converges to fBm.

In Enriquez [16], it is also emphasized that the theorem is conducted for probability measure having moments of $\frac{1}{n^{2-2H}L(n)}$ given in Taqqu [35]. However, choosing μ^H is appropriate for simulation purposes.

CHAPTER 3

REALIZATION AND DISCRETIZATION OF FRACTIONAL BROWNION MOTION

Just as a random walk converges to a standard Brownian motion, it is expected that the correlated random walk converges to fBm. Enriquez [16] proves that the correlated random walk with persistence parameter convergences to fBm. We consider whether a correlated random walk with discretization parameter also converges to fBm. In this chapter, it is initially mentioned how a random walk converges to Brownian motion. Then, theoretical variance-covariance matrix of fBm increments is created and MVN data is generated. And then, this generated data is discretized with a proportion p. Finally, simulation results using dicretized variables are presented.

3.1 Convergence of Random Walks to Brownian Motion

Let $S_n = \sum_{j=0}^{n} X_j$ be one dimensional simple random walk. If the X_j's are merged by linear interpolation, then S_t can be defined for non-integer, $t \in [0, n]$, satisfying $n \leq t \leq n+1$.

$$\begin{aligned} S_t &= S_n + (t-n)X_{n+1} \\ &= \sum_{j=0}^{n} X_j + (t-n)X_{n+1} \end{aligned} \tag{3.1}$$

Now, let us rescale the random walk by taking $1/\sqrt{n}$ as step size for each step instead of step size 1 in simple random walk and compress time to the interval $[0, 1]$. Then, the scaled random walk is obtained as

$$W_t^n = \frac{1}{\sqrt{n}} \sum_{j=0}^{nt} X_j. \tag{3.2}$$

In this rescaled process, there are number of nt steps and each has $\pm\frac{1}{\sqrt{n}}$ step size. This $\pm\frac{1}{\sqrt{n}}$ step size is taking in time unit $\frac{1}{n}$. Recall that the expectation of each X_j is zero and the variance of it equals 1. Hence, W_1^n can be written as

$$W_1^n = \frac{\sum_{j=0}^{n} X_j - E(\sum_{j=0}^{n} X_j)}{\sqrt{V(\sum_{j=0}^{n} X_j)}}. \tag{3.3}$$

By Central Limit Theorem, as $n \to \infty$, W_1^n converges in distribution to a standard Normal distribution. Similarly, by CLT, ${W_t}^n$ converges in distribution to normal distribution $N(0, t)$.

$$W_t^n = \frac{1}{\sqrt{n}} \sum_{j=0}^{nt} X_j = \sqrt{t} \frac{\sum_{j=0}^{nt} X_j - ntE(X_1)}{\sqrt{ntV(X_1)}} \tag{3.4}$$

Assume $0 \le s \le t \le 1$, the sum of $n(t-s)$ iid random variables $W_t^n - W_s^n$ converges in distribution to a normal distribution $N(0, t-s)$ by Central Limit Theorem.

$$\begin{aligned} W_t^n - W_s^n &= \frac{1}{\sqrt{n}} \sum_{j=0}^{nt} X_j - \frac{1}{\sqrt{n}} \sum_{j=0}^{ns} X_j = \frac{1}{\sqrt{n}} \sum_{j=ns}^{nt} X_j \\ &= \sqrt{t-s} \frac{\sum_{j=ns}^{nt} X_j - \sqrt{n(t-s)}E(X_1)}{\sqrt{n(t-s)V(X_1)}} \end{aligned} \tag{3.5}$$

When $n \to \infty$, the independency of $(W_s^n - W_0^n, W_t^n - W_s^n)$ can be seen as below. Assume $m \le ns \le m+1$, then

$$W_s^n - W_0^n = \frac{1}{\sqrt{n}} \sum_{j=0}^{ns} X_j = \frac{1}{\sqrt{n}} \sum_{j=0}^{m} X_j + \frac{(ns-m)X_{m+1}}{\sqrt{n}} \tag{3.6}$$

$$W_t^n - W_s^n = \frac{1}{\sqrt{n}} \sum_{j=ns}^{nt} X_j = \frac{(m+1-ns)X_{m+1}}{\sqrt{n}} + \frac{1}{\sqrt{n}} \sum_{j=m+2}^{nt} X_j \tag{3.7}$$

Note that the only common term is X_{m+1}. If $n \to \infty$, this term will be negligible.

Then, any pair of $W_s^n - W_0^n$ and $W_t^n - W_s^n$ become independent Normal random variables with $W_s^n - W_0^n \sim N(0, s)$ and $W_t^n - W_s^n \sim N(0, t-s)$. Therefore, for all $0 \leq t_1 \leq \cdots \leq t_k \leq 1$, the processes $W_{t_1}, W_{t_2} - W_{t_1}, \cdots, W_{t_k} - W_{t_{k-1}}$ converge to independent normal random variables with expectations 0 and variances $t_1, t_2 - t_1, \cdots t_k - t_{k-1}$ as $n \to \infty$.

Briefly, the random walk converges to;

i As $n \to \infty$, $W_1^n \to W_1 \sim N(0, 1)$ and $W_t^n \to W_t \sim N(0, t)$.

ii As $n \to \infty$, $W_t^n - W_s^n \to W_t - W_s \sim N(0, t-s)$.

iii $W_{t_1}, W_{t_2} - W_{t_1}, \cdots, W_{t_k} - W_{t_{k-1}}$ are independent normal variables. Each has zero expectations and variances of $t_1, t_2 - t_1, \cdots t_k - t_{k-1}$, respectively.

3.2 Variance-Covariance Matrix of Fractional Brownian Motion

It is known that fBm $\{B_t^H, t \geq 0\}$ is a Gaussian process with zero mean and covariance function

$$Cov[B^H(j)B^H(k)] = \frac{1}{2}[j^{2H} + k^{2H} - 2|j-k|^{2H}],$$

where $H \in (0, 1)$ is the parameter of fBm, called Hurst index. fBm has also stationary Gaussian increments, called fractional Gaussian noise (fGn). The covariance of fGns can be calculated by using Equation (3.8).

$$\begin{aligned} & Cov[B^H(j) - B^H(j-1))(B^H(k) - B^H(k-1))] \\ =& E[B^H(j) - B^H(j-1))(B^H(k) - B^H(k-1))] - E[B^H(j) - B^H(j-1)] \\ & E\left[(B^H(k) - B^H(k-1))\right] \end{aligned}$$

Since the second part of the equation is zero, we have

$$
\begin{aligned}
&= E[B^H(j)B^H(k) - B^H(j)B^H(k-1) - B^H(j-1)B^H(k) \\
&+ B^H(j-1)B^H(k-1)] \\
&= E[B^H(j)B^H(k)] - E[B^H(j)B^H(k-1)] - E[B^H(j-1)B^H(k)] \\
&+ E[B^H(j-1)B^H(k-1)]
\end{aligned}
$$

Let us substitute Equation (3.8) into Equation (3.8).

$$
\begin{aligned}
&= \frac{1}{2}[j^{2H} + k^{2H} - 2|j-k|^{2H}] - \frac{1}{2}[j^{2H} + (k-1)^{2H} - 2|j-k+1|^{2H}] \qquad (3.8) \\
&- \frac{1}{2}[(j-1)^{2H} + k^{2H} - 2|j-k-1|^{2H}] + \frac{1}{2}[(j-1)^{2H} + (k-1)^{2H} \\
&- 2|j-k|^{2H}] \\
&= \frac{1}{2}[j^{2H} + k^{2H} - 2|j-k|^{2H} - j^{2H} - (k-1)^{2H} + 2|j-k+1|^{2H} \\
&- (j-1)^{2H} - k^{2H} + 2|j-k-1|^{2H} + (j-1)^{2H} + (k-1)^{2H} - 2|j-k|^{2H}] \\
&= \frac{1}{2}[|j-k+1|^{2H} + |j-k-1|^{2H} - 2|j-k|^{2H}]
\end{aligned}
$$

Suppose that $j-k=m$ for $k<j$, the autocovariance function of the increments for m th lag corresponds to

$$
\begin{aligned}
\gamma(m) &= E[B^H(k+m) - B^H(k+m-1))(B^H(k) - B^H(k-1))] \\
&= \frac{1}{2}[|m+1|^{2H} + |m-1|^{2H} - 2m^{2H}] \qquad (3.9)
\end{aligned}
$$

Let us obtain the variance of the increments.

$$
V(B^H(j)) = E[(B^H(j))^2] = \gamma(0) = \frac{1}{2}[1^{2H} + |-1|^{2H} - 0^{2H}] = 1 \qquad (3.10)
$$

Since the variance of the increments is equal to 1 , the autocorrelation function of the increments for m th lag is equal to Equation (3.9).

Then, in order to make discretization in time, fBm is divided into the number of n time intervals in the interval $[0,t]$. Suppose the interval length is equal to $\lambda = \frac{t}{n}$. Then, the covariance structure of the increments can be written as

$$
\gamma(m) = \frac{1}{2}[|\lambda(m+1)|^{2H} + |\lambda(m-1)|^{2H} - 2|\lambda m|^{2H}]. \qquad (3.11)
$$

Let us take $t = 1$ to squeeze the fBm points to the interval [0,1]. Thus, for $m = j - k$, $j = 0, 1, 2, ..., n$ and $k = 0, 1, 2..., j$, the variance-covariance matrix of fBm increments can be produced as follows:

$$\begin{bmatrix} V(B_{\frac{1}{n}} - B_0) & Cov(B_{\frac{1}{n}} - B_0, B_{\frac{2}{n}} - B_{\frac{1}{n}}) & \dots\dots\dots\dots & Cov(B_1 - B_{\frac{n-1}{n}}, B_{\frac{1}{n}} - B_0) \\ & V(B_{\frac{2}{n}} - B_{\frac{1}{n}}) & \dots\dots\dots\dots & Cov(B_1 - B_{\frac{n-1}{n}}, B_{\frac{2}{n}} - B_{\frac{1}{n}}) \\ \vdots & \ddots & \vdots & \vdots \\ & \dots & V(B_{\frac{n-1}{n}} - B_{\frac{n-2}{n}}) & Cov(B_1 - B_{\frac{n-1}{n}}, B_{\frac{n-1}{n}} - B_{\frac{n-2}{n}}) \\ \dots & \dots & \dots & V(B_1 - B_{\frac{n-1}{n}}) \end{bmatrix} \tag{3.12}$$

3.3 Multivariate Normal Data Generation

It is known that one way to generate fBm is to simulate Multivariate Gaussian random variables with the covariance matrix. Therefore, we generate MVN values with the theoretical correlation matrix of fBm increments to obtain the matrix (3.12). Recall that the covariance and correlation of increments are the same. As we use the covariance matrix of increments, after conducting simulations, we sum them to obtain a realization of fBm.

It is easy to obtain a sample data from Multivariate Gaussian distribution. Assume $Z \sim N_q(\mu, \Sigma)$ where Σ is $q \times q$, positive definite, that is $|\Sigma^{-1}| > 0$, symmetric variance-covariance matrix, and μ denotes the vector of mean. The sample data from MVN distribution is derived by the Cholesky decomposition of the covariance matrix Σ and univariate standard normal vector. As long as $z = (z_1, z_2, \cdots, z_q)$ are q independent standard normal random variables, $Z' = \mu + Vz$ is the random sample from MVN such that Σ is decomposed into multiplication of lower and upper triangular matrix, $\Sigma = VV^T$.

We apply the Cholesky method for partitioning the covariance matrix of fGn to find the square root of it which is given in (3.12). Let δ_n denote a positive definite and $n \times n$

variance covariance matrix of fBm increments. By the Cholesky method, δ_n can be split into the multiplication of lower and upper triangular matrix, $\Lambda_n \Lambda_n^T$. Then, a sample $X = \mu + \Lambda_n z$ from the MVN distribution is generated. Here, $z = (z_1, z_2, ..., z_n)$ is the standard normal vector, μ the mean vector. Here, we take the mean as zero vector since the expectation of fGn is equal to zero.

In our case, for each $j = 1, 2, \cdot\cdot\cdot, k$, we firstly generate the normal random vector $X_j = \Lambda_n Z$, where Z is a standard normal random vector. Λ_n is $n \times n$ lower triangular matrix decomposing from the variance covariance matrix of fBm increments through Cholesky method.

Indeed, we generate the number of k replications of X_j with $N(0, \Lambda_n)$. The vector X_j for $j = 1, ..., k$ is the j th fBm trajectory. The entry X_{ji} refers to i th increment of j th fBm trajectory. Then, we obtain the n dimensional multivariate normal random variables with $n \times n$ covariance matrix as shown below.

$$X = \begin{bmatrix} X_{11} & X_{12} & X_{13} & \dots & \dots & \dots & X_{1n} \\ X_{21} & X_{22} & X_{23} & \dots & \dots & \dots & X_{2n} \\ X_{31} & X_{32} & X_{33} & \dots & \dots & \dots & X_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ X_{j,1} & X_{j,2} & X_{j,3} & \dots & X_{j,i} & \dots & X_{j,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{k1} & X_{k2} & X_{k3} & \dots & \dots & \dots & X_{kn} \end{bmatrix} \tag{3.13}$$

Here, as seen from matrix (3.12), n corresponds to the number of time intervals for a trajectory of fBm. Let us take the sum of each row and $n \to \infty$, each converges to fBm. In order to understand whether the generation is correct, we again compute the covariance matrix and see that it matches up with the theoretical covariance matrix of fBm.

3.4 Discretization of Generated Multivariate Normal Data

We focus on the discertization of MVN generated in the previous section. Hence, in this section, we introduce the discretization procedure and dichotomous variables

In literature, different from innate binary variables such as male or female, yes or no, success or failure, some dichotomous variables are created via discretization of continuous ones. These binary variables such as body length of a person with statues short or tall, high or low amount of money, can be produced by threshold value. Dichotomous variables are more crucial for many scientific fields, despite their drawbacks like information loss, low reliability. Nonetheless, these variables are easy to implement to be used in many areas such as psychology, criminology, biology, and sociology.

It is well known that a correlation among two continuous variables is generally calculated with the Pearson correlation. On the other hand, provided that these two are dichotomized via a threshold term, the correlation name will change. Tetrachoric correlation coefficient is assigned for correlation between two dichotomies variable before discretization. After discretization, the correlation between these dichotomized variables is called phi correlation coefficient.

Demirtas and Vardar-Acar [10] emphasize that when both variables are discretized, the magnitude of these correlations can easily be transformed to the binary case under the normality assumption. Suppose that the distribution of two continuous variables is bivariate normal. After dichotomization of bivariate normal variables, the connection between the tetrachoric correlation and the phi coefficient is known.

Let Z_i's denote the normal variables. These are dichotomized to produce Y_i's which represent binary variables. Assume that $Y_1, Y_2, ..., Y_j$ is a J binary random variables, such that $E[Y_j] = p_j$ for $j = 1, 2, ..., J$ and $Corr[Y_j, Y_k] = \sigma_{jk}$ for $j = 1, 2, ..., J-1; k = 2, 3..., J$. Let $Z = (Z_1, Z_2, ..., Z_J)^T$ represents the J-dimensional multivariate normal random variables with zero mean and $Corr[Z_j, Z_k] = \delta_{jk}$ for $j = 1, 2, ..., J-1; k = 2, 3..., J$.

Then, there is a link between tetrachoric correlation (δ_{jk}) and phi coefficient (σ_{jk}) as below:

$$\sigma_{jk} = \frac{[\Phi(z(p_j), z(p_k), \delta_{jk}) - p_j p_k]}{\sqrt{p_j(1-p_j)p_k(1-p_k)}}, \tag{3.14}$$

where $z(p_i)$ represents p_i th quantile of standard normal distribution for i=1,2. $\Phi[z_1, z_2, \delta]$ is the cumulative distribution function of standard bivariate normal with correlation coefficient δ. Explicitly, $\Phi[z_1, z_2, \delta] = \int_{-\infty}^{z_1}\int_{-\infty}^{z_2} f(z_1, z_2, \delta)dz_1dz_2$ where $f(z_1, z_2, \delta) = [2\pi^{-1}(1-\delta)^{\frac{-1}{2}}] \times exp[-({z_1}^2 - 2z_1z_2\delta + {z_2}^2]/2(1-\delta^2)$.

Note that the link in Equation (3.14) has introduced by Emrich [15]. This relation is used in simulation study to generate binary outcomes. Once J dimentional MVN random variables $Z = (Z_1, Z_2, ...Z_J)^T$ are generated, binary variables is determined by setting $Y_j = 1$ if $Z_j \leq z(p_j)$ and $Y_j = 0$ if $Z_j > z(p_j)$. Then, the expectation and correlation structure for binary variables are obtained as follows:

$$E[Y_j] = P[Y_j = 1] = P[Z_j \leq z(p_j)] = p_j$$

and

$$\begin{aligned} corr(Y_k, Y_j) &= cov(Y_k, Y_j)/(p_k q_k p_j q_j)^{1/2} \\ &= [P\{Y_k = 1, Y_j = 1\} - p_k p_j]/(p_k q_k p_j q_j)^{1/2} \\ &= [P\{Z_k \leq z(p_k), Z_j \leq z(p_j)\} - p_k p_k]/(p_k q_k p_j q_j)^{1/2} \\ &= [\Phi(z(p_k), z(p_j), \delta_{jk}] - p_k p_j]/(p_k q_k p_j q_j)^{1/2}. \end{aligned}$$

After generating the normal outcomes (Z_j), binary variables (Y_j) can be defined by setting $Y_j = 1$ for $Z_j \leq z(p_j)$ and 0 if otherwise. Equivalently, binary random variables can be created by setting $Y_j = 1$ if $Z_j \geq z(1-p_j)$, without causing any change in the correlation.

The relation for non-normality case is also available. In this situation, we can employ power polynomials that established upon the assertion that the first four moments of the distribution ordinarily reflect its characteristics. The polynomial transformation, $Y = eZ^3 + fZ^2 + gZ + h$ where e, f, g and h are some constants, Z represents the standard normal distribution and Y is standardized. Hence, these constants, which can be calculated by moments, are included in the distribution of Y. Equations obtained by profiting from moments of standard normal distribution can be settled via some optimization techniques. Then, correlations among non-normal variables can be written with respect to correlation among normal variables. When this equation is combined with the Equation (3.14), the relationship for non-normal case can be reached.

Since fBm is a Gaussian process, we focus on the relationship for the normal case. When setting of binary variables is $Y_j = 0$ if $Z_j > z(p_j)$ and $Y_j = 1$ if $Z_j \leq z(p_j)$, Y can take the value 1 or 0. However, random walk process $S_n = \sum_{i=0}^{n} Y_i$ is the sum of identically distributed binary random variables $Y_1, Y_2, ..., Y_n$. Thus, Y can take the value of 1 or -1 with the probability of p or $(1-p)$, respectively.

And so, let us take $Y_j = 1$ if $Z_j > z(p)$ and take $Y_j = -1$ if $Z_j \leq z(p)$. First of all, in order to compute the correlation, we find

$$\begin{aligned} E[Y_j] &= 1P(Y_j = 1) + (-1)P(Y_j = -1) \\ &= 1P[Z_j > z(p)] + (-1)P[Z_j \leq z(p)] \\ &= (-1)p + 1(1-p) \\ &= 1 - 2p \end{aligned} \tag{3.15}$$

and

$$\begin{aligned} E[{Y_j}^2] &= 1^2P[Y_j = 1] + (-1)^2P[Y_j = -1] \\ &= 1P[Z_j > z(p)] + 1P[Z_j \leq z(p)] \\ &= p + (1-p) = 1. \end{aligned} \tag{3.16}$$

Then, the variance is obtained as

$$
\begin{aligned}
Var[Y_j] &= E[{Y_j}^2] - E[Y_j]^2 = 1 - (1-2p)^2 \\
&= 4p(1-p).
\end{aligned}
\tag{3.17}
$$

We also calculate

$$
\begin{aligned}
E[Y_jY_k] = & \, 1 \cdot 1 \cdot P[Y_j = 1, Y_k = 1] \; + \; (-1)\cdot(-1)\cdot P[Y_j = -1, Y_k = -1] \\
& + (-1)\cdot 1 \cdot P[Y_j = -1, Y_k = 1] + 1 \cdot (-1) \cdot P[Y_j = 1, Y_k = -1] \\
& = P[Z_j > z(p), Z_k > z(p)] \; + \; P[Z_j \le z(p), Z_k \le z(p)] \\
& - P[Z_j \le z(p), Z_k > z(p)] \; - \; P[Z_j > z(p), Z_k \le z(p)].
\end{aligned}
\tag{3.18}
$$

In order to calculate this expectation, note that

$$
P[Z_j \le z(p), Z_k \le z(p)] = \Phi[z(p), z(p), \delta_{jk}]
\tag{3.19}
$$

and $P[Z_j \le z(p)] = p$, $P[Z_j > z(p)] = 1 - p$ and Equation (3.19) is known from the relation between the bivariate and dichotomous random variables.Moreover,

$$
\begin{aligned}
P[Z_j > z(p), Z_k \le z(p)] &= P[Z_j \le z(p)] \; - \; P[Z_j \le z(p), Z_k \le z(p)] \\
&= p - \Phi[z(p), z(p), \delta_{jk}]
\end{aligned}
\tag{3.20}
$$

and

$$
\begin{aligned}
P[Z_j > z(p), Z_k > z(p)] &= P[Z_j > z(p)] \; - \; P[Z_j > z(p), Z_k \le z(p)] \\
&= (1-p) - (p - \Phi[z(p), z(p), \delta_{jk}]) \\
&= \Phi[z(p), z(p), \delta_{jk}] - 2p + 1.
\end{aligned}
\tag{3.21}
$$

Therefore,

$$
\begin{aligned}
E[Y_jY_k] &= [\Phi(z(p), z(p), \delta_{jk}) - 2p + 1] + [\Phi(z(p), z(p), \delta_{jk})] \\
& \quad -[p - \Phi(z(p), z(p), \delta_{jk})] - [p - \Phi(z(p), z(p), \delta_{jk})] \\
&= 4\Phi(z(p), z(p), \delta_{jk}) - 4p + 1
\end{aligned}
\tag{3.22}
$$

Now, let us calculate the correlation.

$$Corr(Y_jY_k) = \frac{E[Y_jY_k] - E[Y_j][Y_k]}{\sqrt{Var(Y_j)Var(Y_k)}} \tag{3.23}$$

By using Equations(3.16), (3.17) and (3.22), we get

$$\begin{aligned} Corr(Y_jY_k) &= \frac{4\Phi(z(p), z(p), \delta_{jk}) - 4p + 1 - [1 - 2p]^2}{4p(1-p)} \\ &= \frac{\Phi(z(p), z(p), \delta_{jk}) - p^2}{p(1-p)} \end{aligned} \tag{3.24}$$

As a result, a correlated binary data consisting of 1 and -1 which will form our newly suggested dependent random walk possesses a certain covariance structure $\sigma_{jk} = [\Phi(z(p), z(p), \delta_{jk}) - p^2]/p(1-p)$ for $p = p_j = p_k$.

3.5 Simulation Studies

In this section, we carry out a simulation study to foresee whether the correlated random walk with a different parameter, called discretization proportion, can also convergences to fBm. As mentioned earlier, a random sample from the MVN distribution is discretized depending on corresponding normal quantiles given the proportions. After discretization of MVN data, we simultaneously simulate binary and normal variables. Whereas the normal variables correspond to fBm, the binary variables correspond to the correlated random walk.

We simultaneously draw a simulated fBm path and a correlated random walk path. Firstly, we simulate various fBm paths by using some well-known methods such as the Cholesky method, the wavelet based simulation method, and the fast Fourier transform method. In addition, we generate our fBm paths by using different values of the Hurst parameter H.

Several examples of these paths are displayed in the following figures. In Figure 3.1, the fBm path highlighted with black color is simulated via Cholesky method.

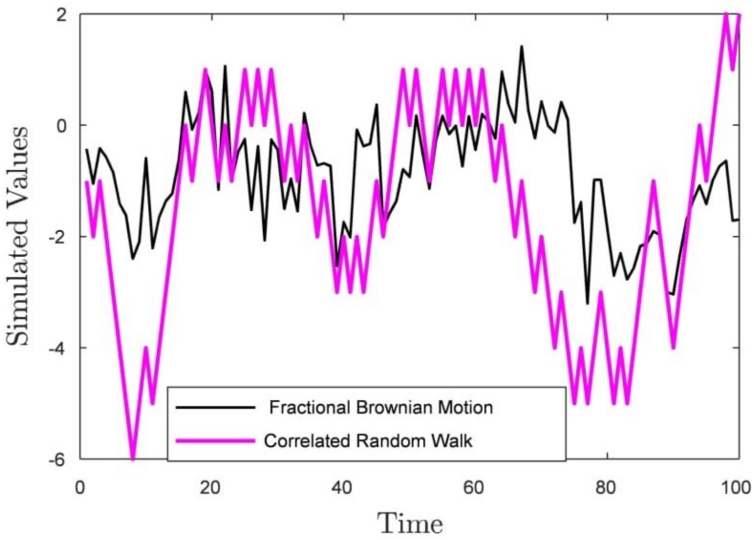

Figure 3.1: Simulated path with H=0.1 by Cholesky Method

Firstly, fBm with Hurst parameter 0.1 is generated. When $H < 0.5$, the time series has an anti-persistent behaviour. Thus, $H = 0.1$ generates an fBm series with the characteristic of mean-reverting, which means that the movement with a direction (up or down) is more likely followed by a movement with opposite direction. After discretization of fBm increments (fBm jumps), we obtain a sample path from the binary distribution. The path highlighted with pink color represents the cumulative sum of these binary values. The comparison of two paths indicates that they are liable to show similar behaviours. It can be seen that two paths are quite similar in terms of the directions of the path and the peaks along the path. However, in some of the time points, small gaps exist between the two paths.

In Figure 3.2, an fBm path is represented by the black line. This path is generated by Cholesky method. Here, different from Figure 3.1, fBm is simulated with the Hurst parameter of 0.3. Here, fBm is a persistent series. The comparison of two paths indicates that they move in a similar manner. However, in some of the time points, gaps are obviously seen.The pink line shows the correlated random walk path which is obtained from discretized fBm increments. When fBm and CRW are drawn in Figure 3.2, it can be seen that the directions of the jumps resemble each other.

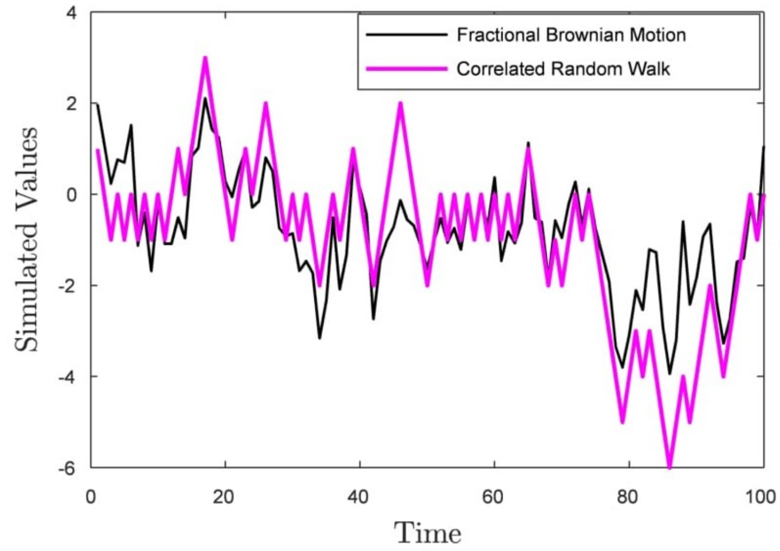

Figure 3.2: Simulated path with H=0.3 by Cholesky Method

In Figure 3.3, an fBm path highlighted with black color is also derived via Cholesky method. Here, an fBm with Hurst index 0.6 is generated. Note that the Hurst pa-

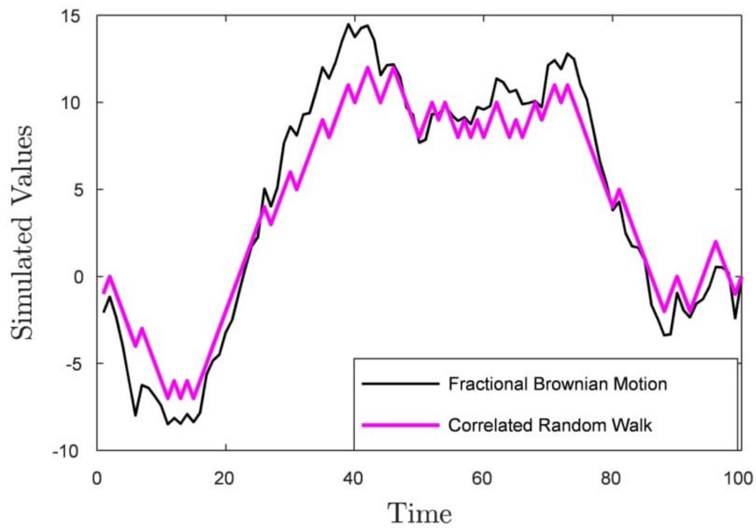

Figure 3.3: Simulated path with H=0.6 by Cholesky Method

rameter for the case $H > 1/2$ indicates a persistent series. A persistent series means that the direction (up or down) of the next value is more likely to be the same as the direction of the current value. Firstly, an fBm sample is generated by Cholesky method.Then, the differences between the sample points (increments) are discretized via the relationship given in Equation 3.14. Hence, we obtain a sample data from the binary distribution. Then, the cumulative sum of these binary values represents the correlated random walk path.This random walk path is drawn in Figure 3.3 with the pink line. When these two paths (fBm and correlated random walk) are are drawn in 3.3, it can be said that they look like each other with respect to the direction of the jumps.

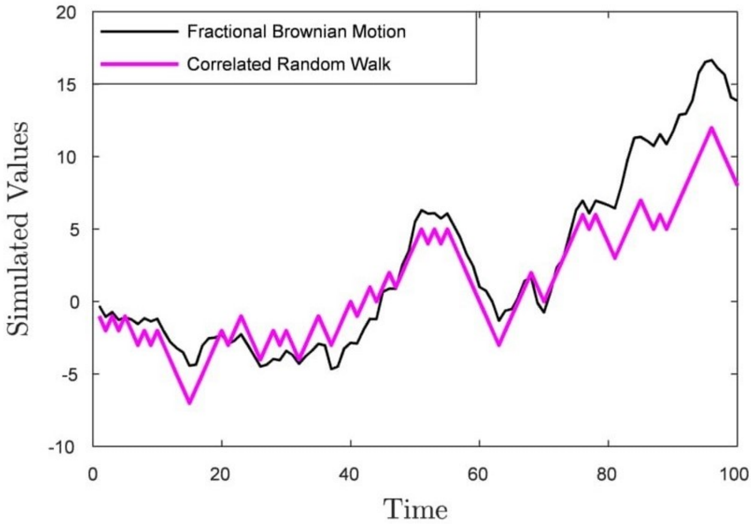

Figure 3.4: Simulated path with H=0.8 by Cholesky Method

The black line in Figure 3.4 represents the fBm path with Hurst parameter $H = 0.8$. This path is simulated by Cholesky method. Here, fBm has a parameter of $H > 1/2$. This shows the presence of a long-range dependence. A long-range dependent process is the same as a long-memory process. Initially, binary data is derived by discretization of fBm path. Then, the pink line indicates the correlated random walk that refers to the cumulative sum of these binary values. If we compare the fBm and its random walk version, it is observed that two move in similar directions at each time point. Furthermore, it can easily be seen that gaps in Figure 3.4 are narrower than gaps in Figures 3.1, 3.2 and 3.3 which are obtained by fBm with all other Hurst parameters.

In order to compare different methods, fBm's with the same Hurst parameters are also generated by using fast Fourier transform method proposed by Davies et al.[9].

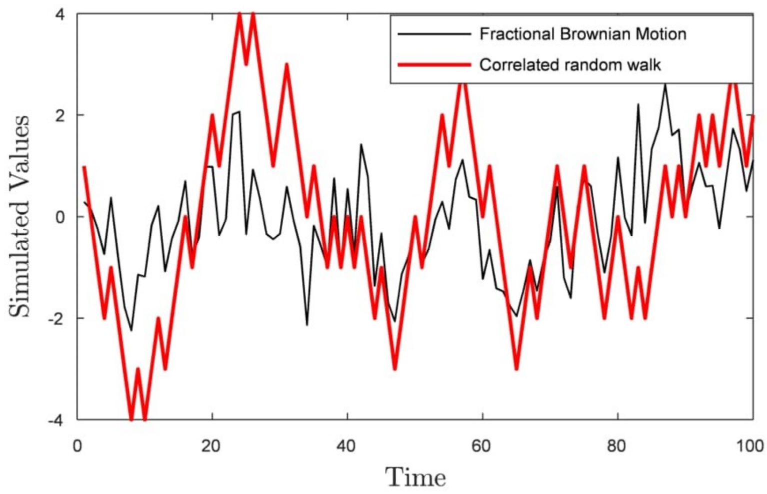

Figure 3.5: Simulated path with H=0.1 by Fast Fourier Transform Method

Figure 3.5 shows the generated fBm path highlighted with black by using fast Fourier transform method. Here, we again generate an fBm path with Hurst dimension 0.1. This dimension corresponds to an anti-persistence series which exhibits a time series with many jagged lines. Then, the red line stands for the correlated random walk corresponding to discretized fBm increments. The random walk path highlighted with red color is the cumulative sum of binary values. The comparison of fBm and correlated random walk shows that the directions of consecutive jumps are the same. Yet, there are some gaps between them. Let us now look at the same case for $H = 0.3$.

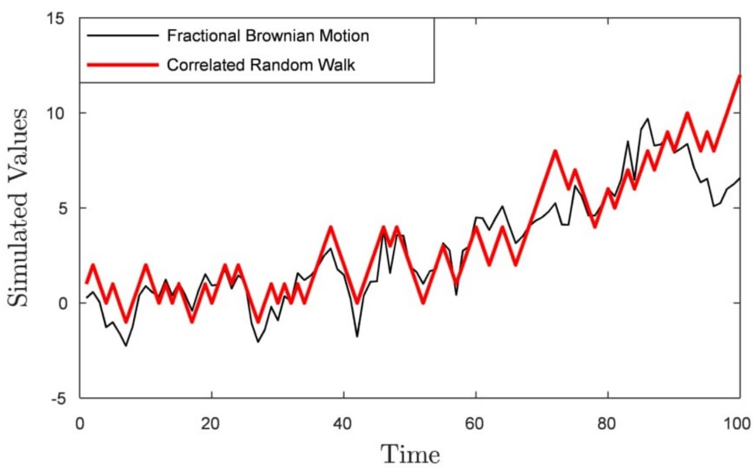

Figure 3.6: Simulated path with H=0.3 by Fast Fourier Transform Method

If we look at Figure 3.6, the line drawn with black represents the fBm path which is generated via FFT method. This simulation is done for the Hurst parameter 0.3 which means that if the series is an up movement then the next movement is more likely to be down. Let us discretize these fBm increments. Then, we obtain a sample path from binary distribution. Then, the red line represents the correlated random walk which is generated by taking the cumulative sum of these binary values. For the FFT method, we obtain the closer results for this Hurst index by the comparison of Cholesky method stated in Figure 3.2.

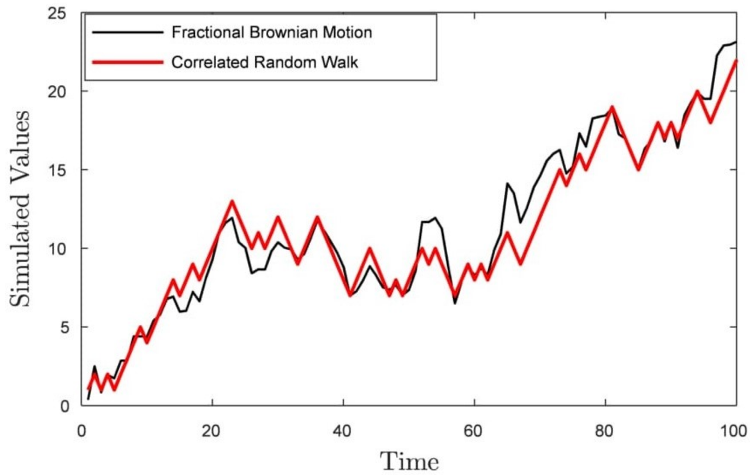

Figure 3.7: Simulated path with H=0.6 by Fast Fourier Transform Method

When Figures 3.7 and 3.8 are analyzed, it can be seen that fBm paths highlighted with black color are generated via FFT method for Hurst indexes $H = 0.6$ and $H = 0.8$, respectively. Each shows a persistent series. For the value of $H = 0.6$, if the last move is up, then there is a 60% chance that the next move again is up. For the value of 0.8, the series is less jagged than for the value of 0.6. When an fBm path is discretized, the cumulative sum of these values is drawn with the red line. Then, we can say that the red and black lines show similar behavior in terms of the direction of the jumps. Moreover, the gaps between them are small. This means that when the Hurst parameter increases, the closer results are obtained.

As illustrated in Figures 3.9 and 3.10, an fBm path is simulated by using the Wavelet based simulation procedure proposed by Abry and Sellan [1]. Firstly, the fBm path

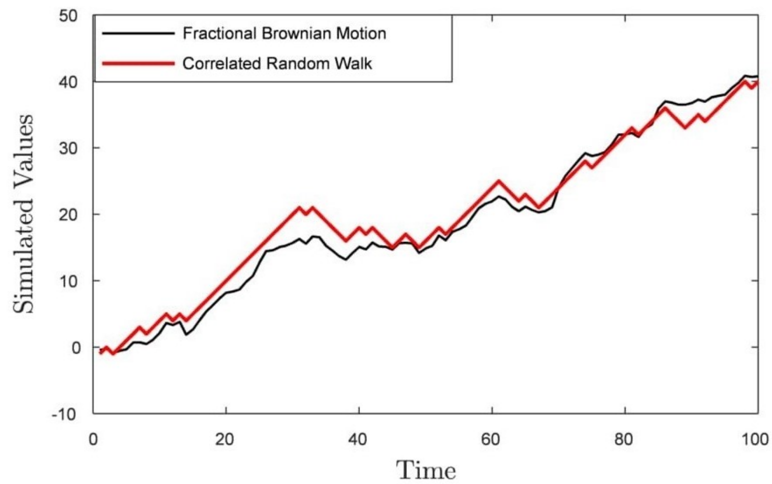

Figure 3.8: Simulated path with H=0.8 by Fast Fourier Transform Method

highlighted with black color is drawn for Hurst dimensions $H = 0.1$ and $H = 0.3$, respectively. After fBm increments are discretized to create binary values, these values are summed and drawn with orange line as can be seen from Figures 3.9 and 3.10.

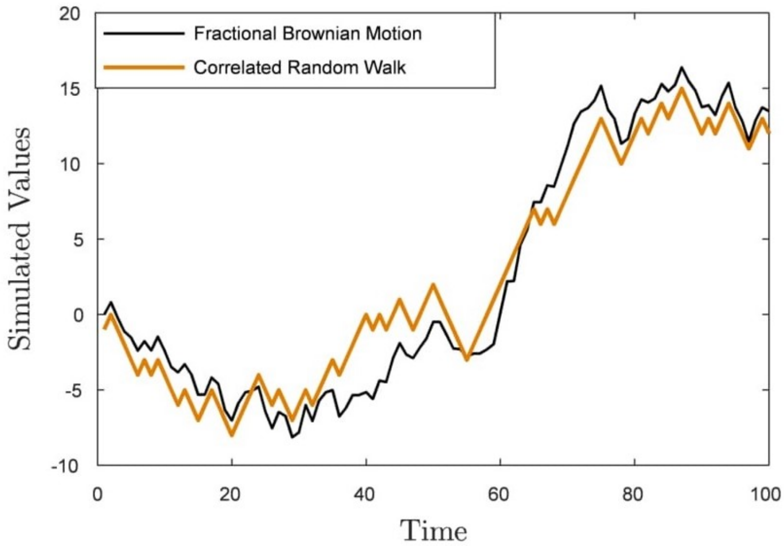

Figure 3.9: Simulated path with H=0.1 by Wavelet Based Simulation Method

Note that this sum corresponds to correlated random walk. Then, it is obviously seen that both fBm and CRW move in the same direction. For the Hurst parameters of 0.1 and 0.3, this method gives better results than the Cholesky and FFT method. If we

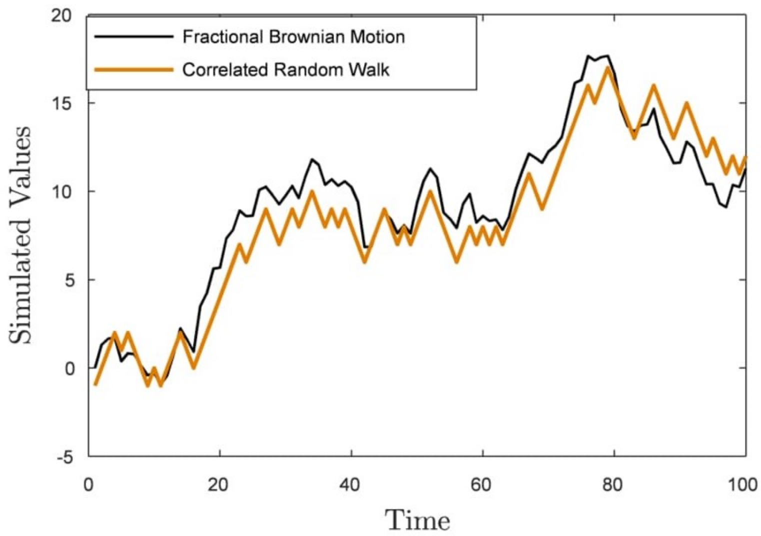

Figure 3.10: Simulated path with H=0.3 by Wavelet Based Simulation Method

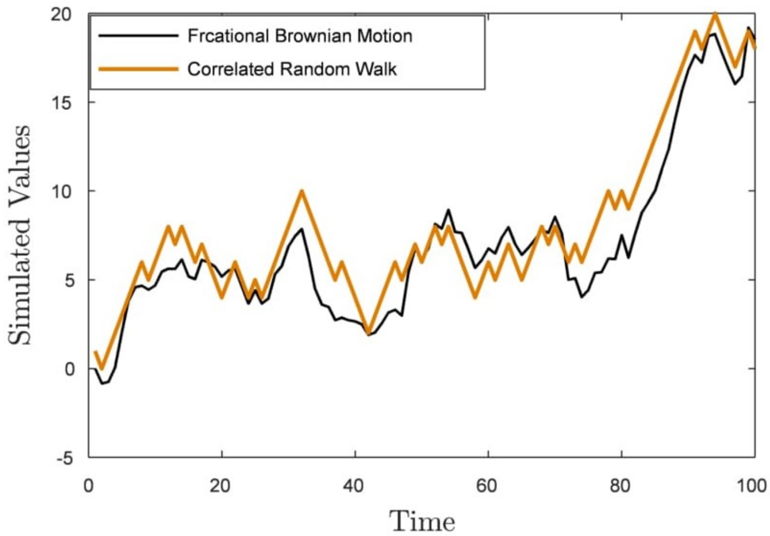

Figure 3.11: Simulated path with H=0.6 by Wavelet Based Simulation Method

compare Figures 3.9 and 3.10, when the Hurst parameter increases, the behavior of the direction get better and gaps decrease. In Figures 3.11 and 3.12, simulation of fBm with Hurst values $H = 0.6$ and $H = 0.8$ is carried out by using the wavelet based transformation method. $H > 1/2$ shows the presence of a long-range dependency in time series. Note that long-range dependence is the same as a long-memory process.

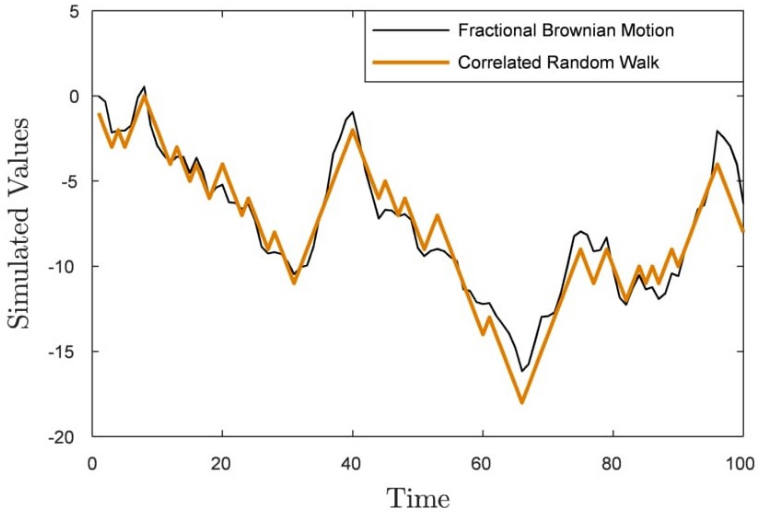

Figure 3.12: Simulated path with H=0.8 by Wavelet Based Simulation Method

A long-memory process implies that past increment (jump) is auto-correlated with an increment in the future. In all figures of Section 3.5, fBm is represented by black line. If the increments of this fBm are discretized in order to build up binary values, then the sum of these values is drawn with orange line as can be seen from Figures 3.11 and 3.12. If we compare these two plots, the jumps of both paths move in the same direction. Conversely, the approximation in Figure 3.12 seems more efficient.

As a summary, all simulation plots are obtained using various values of H and different kinds of fBm generation methods. The differences between fBm points (increments) are discretized through the relationship given in Equation 3.14. Thus, we get binary values. If we take the cumulative sum of these binary values, we get a random walk path. The correlated random walk paths are shown in all plots with colourful lines according to simulation methods. After these two paths (fBm and random walk) are drawn in the same plots, it can be said that they look like each other with respect to the direction of the jumps for all Hurst parameters. However, when the Hurst parameter is approximate to one, simulation paths are close to each other. Moreover, the best result is given by wavelet based transform method since gaps between two paths are small for the whole Hursth values H.

As a conjecture from these simulations, the question of "Do these generated correlated random walks converges to fBm for all discretization proportions ?" arises. Even if there are some gaps in the figures, this gives us an idea about how the convergence depends on some conditions for the value of p. After encountering the paper of Enriquez [16], we find the density of discretization proportion p which allows that the correlated random walk converges to fBm. Therefore, we suggest a new algorithm of generating fBm for a given Hurst parameter, $H > 1/2$. Next chapter presents the theoretical background and the proof.

CHAPTER 4

THEORETICAL CONSTRUCTION OF CORRELATED RANDOM WALK

As mentioned in the previous chapter, We have noted that the main questions, which arise from our simulation study, are i) Do the correlated random walk with explicit discretization proportion converge to fBm ? ii) Can we write an algorithm to generate an fBm through such correlated random walk? For examining the first question, we conduct a relation between persistence parameter and sicretization proportion in Section 4.1. For considering the second question, we prove the convergence of a correlated random walk with discretization proportion to fBm in Section 4.2. The study of Enriquez [16] guides us in finding the relationship, showing the convergence, and proposing the simulation algorithm to generate an fBm..

4.1 Relation between Persistence ρ and Discretization Proportion p of Fractional Brownian Motion

Recall that the definition of CRW with persistence parameter ρ is given in Definition 3. In order to establish a connection between p and ρ, we utilize from the fourth item (iv) given in Definition 3. According to this item,

$$P(\epsilon_{n+1}^{\rho} = \epsilon_n^{\rho}) = \rho, \tag{4.1}$$

where $\epsilon_n^{\rho} = X_n^{\rho} - X_{n-1}^{\rho}$ corresponds to the increments or jumps in each time step, takes values of $+1$ or -1.

In order to prevent any confusion, the notation of Y is defined as the binary variable instead of ϵ. Hence, the Equation (4.1) turns into

$$P(Y_{n+1}^{\rho} = Y_n^{\rho}) = \rho \tag{4.2}$$

Since the value of Y can only take the value of $+1$ or -1, both Y_{n+1} and Y_n can also take the value $+1$ or -1. Thus,

$$P(Y_{n+1}^{\rho} = Y_n^{\rho}) = P(Y_{n+1}^{\rho} = 1, Y_n^{\rho} = 1) + P(Y_{n+1}^{\rho} = -1, Y_n^{\rho} = -1)$$

$$= P[Z_{n+1}^{\rho} > z(p), Z_n^{\rho} > z(p)] + P[Z_{n+1}^{\rho} \leq z(p), Z_n^{\rho} \leq z(p)],$$

Substitute Equation (3.19) and (3.21) into Equation (4.3), then,

$$\begin{aligned} \rho &= \Phi(z(p), z(p), \delta_{n,n+1}) - [2p-1] + \Phi(z(p), z(p), \delta_{n+1,n}) \\ &= 2\Phi(z(p), z(p), \delta_{n,n+1}) - 2p + 1 \end{aligned} \tag{4.3}$$

where $\delta_{n,n+1} = \begin{bmatrix} 1 & \gamma(1) \\ \gamma(1) & 1 \end{bmatrix}$ is the correlation matrix, where $\gamma(1)$ denotes one step correlation of fBm increments and $P(Z \leq z(p)) = p$ is the proportion of discretization, Φ is the cumulative distribution function of the standard bivariate normal variables and $z(p)$ represents p th quantile of the standard normal distribution.

After observing this relationship, we define our correlated random walk as shown below.

Definition 4 *The correlated random walk X^p with discretization probability p is a discrete-time process, such that*

i *Walk is starting at 0, $X_0 = 0$.*

ii *All X_j' s are identically distributed with probability p, $P(X_j = -1) = p$ and $P(X_j = +1) = 1 - p$.*

iii *For all $n \geq 1$, $Y_n = X_n - X_{n-1}$ is equal to -1 and $+1$ almost surely.*

iv *For all $n \geq 1$, $P(Y_{n+1} = Y_n | \sigma(X_k^p, 0 \leq k \leq n)) = 2\Phi(z(p), z(p), \delta_{n,n+1}) - 2p + 1$*

The density of persistence is known from Theorem 2 in Chapter 2. To reach the density of the discretization proportion, we apply the transformation of persistence by using the relationship given in Equation (4.3).

On $[\frac{1}{2}, 1]$, the density function of the persistence ρ is

$$f_\rho(\rho) = (1 - H)2^{3-2H}(1 - \rho)^{1-2H}. \tag{4.4}$$

Let us change the variable by using the relationship $\rho = 2\Phi[z(p), z(p), \delta_{n+1,n}] - 2p + 1$. Then, the density function of p is obtained as

$$f_p(p) = (1 - H)2^{3-2H}(1 - (2\Phi[z(p), z(p), \delta_{n+1,n}] - 2p + 1))^{1-2H}|J|, \tag{4.5}$$

where $|J|$ is the Jacobian term with equality

$$|J| = |\frac{d}{dp}2[\Phi(z(p), z(p), \delta_{n+1,n}] - 2p + 1|.$$

Hence,

$$\begin{aligned} f_p(p) = &(1 - H)2^{3-2H}(2p - 2\Phi(z(p), z(p), \delta_{n+1,n})^{1-2H} \\ &\times|\frac{d}{dp}2\Phi(z(p), z(p), \delta_{n+1,n} - 2p + 1| \end{aligned}$$

4.2 Theoretical Convergence of Correlated Random Walk to Fractional Brownian Motion

After conducting our simulation study, we come up with the question that "Does correlated random walks generated by discretization with any probability p converge to fBm? ". This seemed to be a natural question because the gaps are observed between the correlated random walk and fBm. Through the work of Enriquez [16], we prove the convergence of the correlated random walk to fBm by using Equation (4.6) corresponding to the density of the discretization proportion.

The idea behind the convergence theorem consists of two stages. Firstly, let us generate different correlated random walk paths made up of -1 and $+1$. When we take the sum of N th steps of each random walk for $k \geq 1$, then the summation is normally distributed as the number of paths gets larger. Then, the result of the limit generates a discrete centered Gaussian sequence.

Second, by using the Lemma 5.1 in Taqqu [35], a discrete Gaussian sequence is summed and multiplied with a constant a and divided into N^H, then it converges weakly to fBm $B_H(t)$. This is feasible if the sum of all correlations between stationary Gaussian sequences satisfies the conditions in Taqqu [35].

In the following theorem , we prove how a correlated random walk with discretization proportion, which satisfies explicit density, converges to fBm.

Theorem 3 *Let μ^H be the distribution function of discretization proportion with density $f_p(p) = (1-H)2^{3-2H}(2p-2\Phi(z(p), z(p), \delta_{n+1,n})^{1-2H}|\frac{d}{dp}2\Phi(z(p), z(p), \delta_{n+1,n}-2p+1$ and $(X^{\mu^H,i})_{i\geq 1}$ be a sequence of independent process,*

$$\lim_{N\to\infty}\lim_{M\to\infty} a_H \frac{X^{\mu^H,1}_{[Nt]} + X^{\mu^H,2}_{[Nt]} ... + X^{\mu^H,M}_{[Nt]}}{N^H\sqrt{M}} = B^H(t)$$

where $a_H = \sqrt{\frac{H(2H-1)}{\Gamma(3-2H)}}$, N is the number of time steps and M be the number of trajectories, $B^H(t)$ is the fBm with Hurst index $H \in [1/2, 1]$.

Proof *Let $X^{\mu^H,i}_k$ be k th step of i th binary variables -1 or $+1$ trajectory for $1 \leq i \leq M$ where M represents the number of trajectories.*

It is known that CLT states that for any independent random sequence X_n, sample mean or sum of this sequence is distributed normally for large sample size n.

Thus, CLT proposes that $\lim_{M\to\infty}\frac{X^{\mu^H,1}_k + X^{\mu^H,2}_k ... + X^{\mu^H,M}_k}{\sqrt{M}}$ is the discrete Gaussian process because the expectation $E[\sum_{i=1}^M (X^{\mu^H,i})] = 0$ and variance $Var[\sum_{i=1}^M (X^{\mu^H,i})] = M$ with the variance of binary variables $Var(X_i) = 1$.

Suppose that this discrete Gaussian Process is $(Z^H_k)_{k\geq 1}$. It has stationary increments $\varepsilon^H_k = (Z^H_{k+1} - Z^H_k)$ with $E[\varepsilon^H_k] = 0$ and $E[(\varepsilon^H_k)^2] = 1$ and $E[\varepsilon^H_k \varepsilon^H_{k+n}]$ is found as *follows:*

By Proposition 2, n step correlation is written as:

$$r(n) = E[\varepsilon_k^H \varepsilon_{k+n}^H] = \int_{p_{lower}}^{p_{upper}} [2 \cdot (2\Phi(z(p), z(p), \delta_{n,n+1}) - 2p + 1) - 1]^n d_p(p), \tag{4.6}$$

where p_{lower} and p_{upper} denotes the range of p values satisfying the inequality

$$\frac{1}{2} \leq 2\Phi[z(p), z(p), \delta_{n,n+1}] - 2p + 1 < 1.$$

Here, $d_p(p) = \frac{d}{dp}(P_p) = f_p(p)dp$, where P_p denotes the law of correlated random walk X^p. Then substitute $f_p(p)dp$ given in Equation (4.6) into Equation (4.6), we observe

$$\begin{aligned} &= \int_{p_{lower}}^{p_{upper}} [4\Phi(z(p), z(p), \delta_{n,n+1}) - 4p + 1]^n (1 - H) 2^{3-2H} \\ &\times (2p - 2\Phi(z(p), z(p), \delta_{n,n+1}))^{1-2H} \frac{d}{dp} (2\Phi(z(p), z(p), \delta_{n,n+1}) - 2p + 1) dp. \end{aligned}$$

In order to solve the integral, let us change the variable.

$$\begin{aligned} \nu &= 4\Phi(z(p), z(p), \delta_{n,n+1}) - 4p + 1 \\ \nu - 1 &= 4\Phi(z(p), z(p), \delta_{n,n+1}) - 4p \\ \frac{\nu - 1}{2} &= 2\Phi(z(p), z(p), \delta_{n,n+1}) - 2p \\ \frac{\nu - 1}{2} + 1 &= 2\Phi(z(p), z(p), \delta_{n,n+1}) - 2p + 1 \\ \frac{\nu + 1}{2} &= 2\Phi(z(p), z(p), \delta_{n,n+1}) - 2p + 1 \end{aligned}$$

By taking the derivative of both side with respect to p,

$$\begin{aligned} \frac{d(\frac{\nu+1}{2})}{dp} &= \tfrac{d}{dp}[2\Phi(z(p), z(p), \delta_{n,n+1}) - 2p + 1] \\ d\Big(\frac{\nu + 1}{2}\Big) &= \tfrac{d}{dp}[2\Phi(z(p), z(p), \delta_{n,n+1}) - 2p + 1]dp \end{aligned}$$

Now, the range of p *becomes*

$$\frac{1}{2} \leq \rho \leq 1$$

$$\frac{1}{2} \leq 2\Phi(z(p), z(p), \delta_{n,n+1}) - 2p + 1 \leq 1. \tag{4.7}$$

Equivalently,

$$0 \leq 4\Phi(z(p), z(p), \delta_{n,n+1}) - 4p + 1 \leq 1$$

$$0 \leq \nu \leq 1 \tag{4.8}$$

Thus,

$$r(n) = (1-H)2^{3-2H} \int_0^1 \nu^n \Big(\frac{1-\nu}{2}\Big)^{1-2H} d\Big(\frac{\nu+1}{2}\Big)$$

$$= (1-H)2^{3-2H} \int_0^1 \nu^n \frac{(1-\nu)^{1-2H}}{2^{1-2H}} \frac{1}{2} d\nu$$

$$= \frac{(1-H)2^{3-2H}}{2^{2-2H}} \int_0^1 \nu^n (1-\nu)^{1-2H} d\nu. \tag{4.9}$$

Note that $\int_0^1 \nu^n (1-\nu)^{1-2H} d\nu$ *corresponds to the Beta function with parameter* $(n + 1, 2-2H)$*. Hence* $r(n)$ *becomes*

$$r(n) = (2-2H)\frac{\Gamma(n+1), \Gamma(2-2H)}{\Gamma(n+3-2H)}$$

$$= (2-2H)\frac{n!(1-2H)!}{(n+2-2H)!}$$

$$\sim_{n\to\infty} = \Gamma(3-2H)\frac{1}{n^{2-2H}}. \tag{4.10}$$

The Equation (4.10) *can be rearranged as:*

$$= \frac{1}{a_H^2}\frac{H(2H-1)}{n^{2-2H}} \qquad \text{where} \qquad a_H = \sqrt{\frac{H(2H-1)}{\Gamma 3 - 2H}} \tag{4.11}$$

such that

$$a_H^2 \sum_{i=1}^{N} \sum_{j=1}^{N} r(|i-j|) = a_H^2 (r(0) + \sum_{i=1}^{N-1} [r(0) + 2 \sum_{k=1}^{i} r(k)]) \tag{4.12}$$

$$\sim_{n \to \infty} N^{2H}$$

In Taqqu [35], Theorem 2.1 *and Lemma* 5.1 *bring this conclusion directly.*

CHAPTER 5

PROPOSING AN ALGORITHM FOR GENERATING FRACTIONAL BROWNIAN MOTION THROUGH CORRELATED RANDOM WALK

By considering the relationship between the persistence parameter and the discretization proportion, we are able to construct a new correlated random walk which converges to fBm. Initially, we obtain the values of p which satisfy the density given in Equation (4.6). This is done via the density of persistence and the relationship given in Equation (4.3).

Now let us consider

$$F(a) = P(\rho \leq a) = u, \qquad \text{where u is from Uniform } (0,1).$$

$$= \int_{\frac{1}{2}}^{a} f_\rho(\rho) d\rho$$

$$= \int_{\frac{1}{2}}^{a} 2^{3-2H}(1-H)(1-\rho)^{1-2H} d\rho$$

Let $\vartheta = 1 - \rho$, then

$$u = \int_{\frac{1}{2}}^{1-a} (1-H)2^{3-2H}\vartheta^{1-2H}(-d\vartheta)$$

$$= (H-1)2^{3-2H} \int_{\frac{1}{2}}^{1-a} \vartheta^{1-2H} d\vartheta$$

$$= (H-1)2^{3-2H} \int_{\frac{1}{2}}^{1-a} \vartheta^{1-2H} d\vartheta$$

$$= (H-1)2^{3-2H} \frac{\vartheta^{1-2H}}{2-2H} \Big|_{\frac{1}{2}}^{1-a}$$

$$= 2^{2-2H}[2^{2-2H} - (1-a)^{2-2H}]$$

and

$$1 - (2-2a)^{2-2H} = u$$

$$(2-2a)^{2-2H} = 1 - u$$

$$2(1-a) = (1-u)^{\frac{1}{2-2H}}$$

$$a = 1 - \frac{(1-u)^{\frac{1}{2-2H}}}{2} \tag{5.1}$$

Obviously, the proportion ρ has the density of $1 - \frac{(1-u)^{\frac{1}{2-2H}}}{2}$ for $H > 1/2$. The values obtained from the Equation (5.1) are replaced by ρ given in the relationship between the persistence and the discretization proportion. Thus, the proportion ρ takes the values which satisfy the equation

$$-\frac{(1-u)^{\frac{1}{2-2H}}}{4} = \Phi(z(p), z(p), \delta_{n,n+1}) - p. \tag{5.2}$$

When the dichotomous values are generated using the value of p which satisfies the density given in Equation (5.2) and these values are generated for a large number of trajectories, their sum corresponds to the values of the correlated random walk. And so, by Theorem 3 in Chapter 4, if the sums of these CRW points for each time step are scaled by $N^H\sqrt{M}$, this process converges weakly to fBm for the large number of trajectories M, and the large number of time steps N.

By using Theorem 3, we propose the following algorithm to simulate the fBm.

i Obtain p values from the equation

$$-\frac{(1-u)^{\frac{1}{2-2H}}}{4} = \Phi(z(p), z(p), \delta_{n,n+1}) - p$$

,where u is from Uniform$(0,1)$ distribution.

ii Simulate M independent $\{p_j\}$ $j = 1, 2, ..., M$ for M trajectories.

iii Simulate M replications according to

If $t_i = 1$, $\epsilon_1^j = 2 \times Bernoulli(\frac{1}{2}) - 1$ for $X_1^j = \epsilon_1^j$.

If $t_i > 1$, $\epsilon_{t_i}^j = \epsilon_{t_{i-1}}^j \times (2 \times Bernoulli(2\Phi[z(p_j), z(p_j), \delta_{n,n+1}] - 2p_j)$

,where $X_{t_i}^j = X_{t_{i-1}}^j + \epsilon_{t_i}^j$

iv For each t_i, compute

$$B_{t_i}^H = a_H \frac{X_{[Nt_i]}^{\mu^H,1} + X_{[Nt_i]}^{\mu^H,2} ... + X_{[Nt_i]}^{\mu^H,M}}{N^H \sqrt{M}}.$$

In this algorithm, the construction of the fBm is similar to the construction of the standard BM from random walks. However, in reality, this algorithm may not sufficiently give fast results due to the simulation for the large number of paths.

By using this algorithm, the samples of fBm are obtained as the following figures. Figures 5.1 and 5.2 are obtained using the number of 300 time steps and the number of 100.000 trajectories. These plots refer approximately to fBm paths. However, as the number of time steps is increased, CRW converges faster. Here, we take less number of time steps in order to facilitate the visibility.

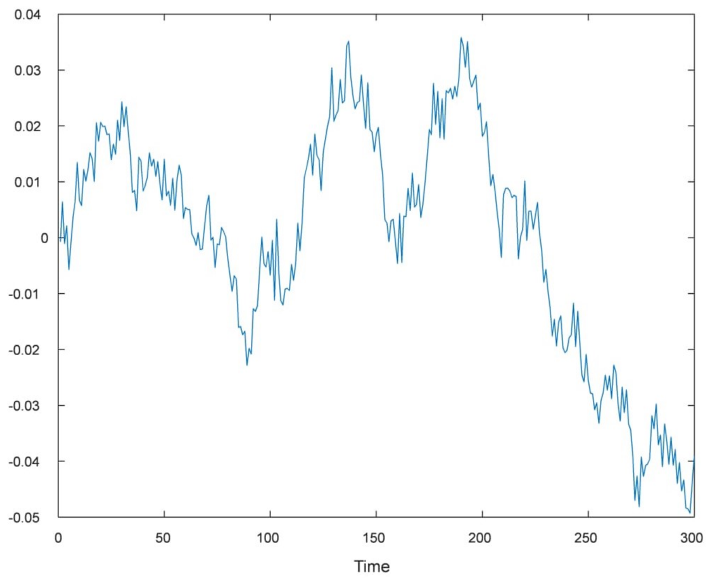

Figure 5.1: Fractional Brownian Motion Path from Correlated Random Walk with H=0.6

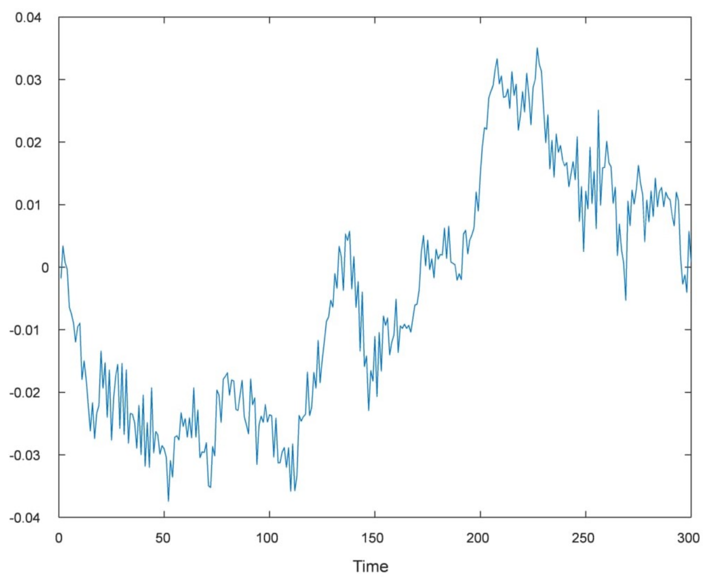

Figure 5.2: Fractional Brownian Motion Path from Correlated Random Walk with H=0.8

CHAPTER 6

CONCLUSION

In this thesis, our aim is to propose a new fBm generation method by using given Hurst parameter and using the correlation structure based on this parameter. Our generation is the generalized version of Enriquez [16]. He has suggested an fBm construction through a persistent random walk with a parameter of persistence. This parameter refers to the probability of making the same jump as the previous one. Besides, the literature [10,14] consists of the connection between the Pearson correlation and the phi coefficient. We know that a correlation between two continuous variables is calculated with the Pearson correlation. On the other hand, the phi coefficients represents the correlation between two dichotomous variables is. In this study, the Pearson correlation corresponds to the correlation between the bivariate normal random variables. The phi coefficient corresponds to the correlation between the correlated binary outcomes which are created from the discretization of the generated normal outcomes. We propose an optional method for the construction of an fBm by using the relationship between the normally distributed continuous variables and their discretized version. Enriquez [16] generates a persistent random walk with persistence parameter in order to construct an fBm. As an improvement to the study of Enriquez [16], we generate a correlated random walk with a correlation structure calculated using the given Hurst parameter and the discretization proportion, to construct an fBm. Furthermore, we write our theorem to show that the correlated random walk with discretization proportion converges to fBm. This theorem enables us to develop a new algorithm for the simulation of fBm.

Diverse properties of the fBm such as self similarith and long-range dependence are characterized by the Hurst parameter H. One reason for the application of fBm is that it has a long memory when $H > 1/2$. Thus, we generate our fBm for the case $H > 1/2$.

In the introduction chapter, we describe the development of our motivation in this study. In Chapter 2, we continue with introducing the main definitions, notations and properties that we need for presenting our study. We also review the major simulation methods for generating an fBm. In Chapter 3, we mention the convergence of a random walk to a standard Brownian motion. Then the variance-covariance matrix of fBm increments are created and multivariate normal vectors are generated by using the Cholesky decomposition. Due to the definition of the CRW, we then discretize the increments with a proportion p, and take the cumulative sum of the discretized increments to obtain a CRW sample. We perform the discretization for generating fBm samples by using different simulation methods and Hurst parameters. When a theoretical fBm sample and its discretized form are compared, we realize that these two lines move in the same direction for all values of p. Hence, we consider that we can generate a CRW converging to an fBm by using the discretization procedure. Here, we profit from the paper of Enriquez [16] in order to prove the convergence. Afterwards, in chapter 4, we establish a link between the persistence parameter ρ and the discretization proportion p. Then, we describe our correlated random walk which is generated from discretized Bernoulli variables with proportion p. An explicit density is assigned to this proportion thanks to the link given in Equation (4.3). It is easily seen that the correlated random walk with a discretization proportion which has an explicit density converges to fBm. We show the convergence based on the fact that the great number of such random walks produce discrete Gaussian process after sum of the whole random walks is normalized. When this process owns a correlation that satisfies the terms specified in Taqqu [35], it converges to fBm. Based on this convergence, we propose a new algorithm for simulating an fBm in Chapter 5. We also provide our proposed MATLAB codes.

Recall that we mention the construction of fBm for the long-range dependence throughout this study. As a future study, we may also consider the construction of fBm for the short-range dependence. The short-range dependence corresponds to the presence of negatively correlated increments. In the presence of short-range dependence, the definition of the persistence must be changed. Then, the persistence means the probability of having the opposite jump to the last one. In addition, as in this study, using the lemma in Taqqu [35] is not enough to get the convergence. We need also to show the tightness of the family of the processes by controlling the criterion in Billingsley [4]. Therefore, we are leaving it for the future study.

REFERENCES

[1] ABRY, P., AND SELLAN, F. The wavelet-based synthesis for fractional brownian motion proposed by f. sellan and y. meyer: Remarks and fast implementation, 1996.

[2] ASMUSSEN, S. *Stochastic simulation with a view towards stochastic processes*. University of Aarhus. Centre for Mathematical Physics and Stochastics (MaPhySto)[MPS], 1998.

[3] BIAGINI, F., HU, Y., ØKSENDAL, B., AND ZHANG, T. *Stochastic calculus for fractional Brownian motion and applications*. Springer Science & Business Media, 2008.

[4] BILLINGSLEY, P. *Convergence of probability measures*. John Wiley & Sons, 1968.

[5] BROWN, R. *A Brief Account of Microscopical Observations Made... on the Particles Contained in the Pollen of Plants, and on the General Existence of Active Molecules in Organic and Inorganic Bodies*. 1828.

[6] CAGLAR, M. Simulation of fractional brownian motion with micropulses. *Advances in Performance Analysis 3* (2000), 43–69.

[7] COLLINS, J., AND DE LUCA, C. Upright, correlated random walks: A statistical-biomechanics approach to the human postural control system. *Chaos: An Interdisciplinary Journal of Nonlinear Science 5*, 1 (1995), 57–63.

[8] DASGUPTA, A. *Fractional Brownian motion: Its properties and applications to stochastic integration*. PhD thesis, University of North Caroline in Chapel Hill, 1997.

[9] Davies, R. B., and Harte, D. Tests for hurst effect. *Biometrika 74*, 1 (1987), 95–101.

[10] Demirtas, H., and Vardar-Acar, C. Anatomy of correlational magnitude transformations in latency and discretization contexts in monte-carlo studies. In *Monte-Carlo Simulation-Based Statistical Modeling*. Springer, 2017, pp. 59–84.

[11] Dieker, T. Simulation of fractional brownian motion. *MSc theses, University of Twente, Amsterdam, The Netherlands* (2004).

[12] Dietrich, C. R., and Newsam, G. N. Fast and exact simulation of stationary gaussian processes through circulant embedding of the covariance matrix. *SIAM Journal on Scientific Computing 18*, 4 (1997), 1088–1107.

[13] Donsker, M., and Varadhan, S. Asymptotics for the wiener sausage. *Communications on Pure and Applied Mathematics 28*, 4 (1975), 525–565.

[14] Einstein, A. *Investigations on the Theory of the Brownian Movement*. Courier Corporation, 1956.

[15] Emrich, L. J., and Piedmonte, M. R. A method for generating high-dimensional multivariate binary variates. *The American Statistician 45*, 4 (1991), 302–304.

[16] Enriquez, N. A simple construction of the fractional brownian motion. *Stochastic Processes and their Applications 109*, 2 (2004), 203–223.

[17] Fagan, W. F., and Calabrese, J. M. The correlated random walk and the rise of movement ecology. *The Bulletin of the Ecological Society of America 95*, 3 (2014), 204–206.

[18] Hosking, J. R. Modeling persistence in hydrological time series using fractional differencing. *Water resources research 20*, 12 (1984), 1898–1908.

[19] Hurst, H. E. Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng. 116* (1951), 770–808.

[20] KIJIMA, M., AND TAM, C. M. Fractional brownian motions in financial models and their monte carlo simulation. In *Theory and Applications of Monte Carlo Simulations*. InTech, 2013.

[21] KOLMOGOROV, A. N. Wienersche spiralen und einige andere interessante kurven in hilbertscen raum, cr (doklady). *Acad. Sci. URSS (NS) 26* (1940), 115–118.

[22] KONSTANTOPOULOS, T., AND SAKHANENKO, A. I. Convergence and convergence rate to fractional brownian motion for weighted random sums. *1*, 0 (2004), 47–63.

[23] LAU, W.-C., ERRAMILLI, A., WANG, J. L., AND WILLINGER, W. Self-similar traffic generation: The random midpoint displacement algorithm and its properties. In *Communications, 1995. ICC'95 Seattle,'Gateway to Globalization', 1995 IEEE International Conference on* (1995), vol. 1, IEEE, pp. 466–472.

[24] LELAND, W. E., TAQQU, M. S., WILLINGER, W., AND WILSON, D. V. On the self-similar nature of ethernet traffic (extended version). *IEEE/ACM Transactions on networking 2*, 1 (1994), 1–15.

[25] LÉVY, P., AND LOÈVE, M. *Processus stochastiques et mouvement brownien*. Gauthier-Villars Paris, 1965.

[26] LIM, S., AND MUNIANDY, S. Fractional brownian motion: Theory and application to dna walk. In *Biological Physics 2000*. World Scientific, 2001, pp. 214–233.

[27] LINDSTRØM, T. A random walk approximation to fractional brownian motion. *arXiv preprint arXiv:0708.1905* (2007).

[28] MANDELBROT, B. B., AND VAN NESS, J. W. Fractional brownian motions, fractional noises and applications. *SIAM review 10*, 4 (1968), 422–437.

[29] NORROS, I., MANNERSALO, P., AND WANG, J. L. Simulation of fractional brownian motion with conditionalized random midpoint displacement. *Advances in Performance Analysis 2*, 1 (1999), 77–101.

[30] PEARSON, K. The problem of the random walk. *Nature 72*, 1865 (1905), 294.

[31] ROGERS, L. C. G. Arbitrage with fractional brownian motion. *Mathematical Finance 7*, 1 (1997), 95–105.

[32] SKOROKHOD, A. V. Limit theorems for stochastic processes. *Theory of Probability & Its Applications 1*, 3 (1956), 261–290.

[33] SOTTINEN, T. Fractional brownian motion, random walks and binary market models. *Finance and Stochastics 5*, 3 (2001), 343–355.

[34] SZABADOS, T. Strong approximation of fractional brownian motion by moving averages of simple random walks. *Stochastic processes and their applications 92*, 1 (2001), 31–60.

[35] TAQQU, M. S. Weak convergence to fractional brownian motion and to the rosenblatt process. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 31*, 4 (1975), 287–302.

[36] WIENER, N. The average of an analytic functional and the brownian movement. *Proceedings of the National Academy of Sciences 7*, 10 (1921), 294–298.

[37] WOOD, A. T., AND CHAN, G. Simulation of stationary gaussian processes in [0, 1] d. *Journal of computational and graphical statistics 3*, 4 (1994), 409–432.

APPENDIX A

APPENDIX A

Here the Matlab codes for given algorithm in Chapter 4 is offered.

```
function [fBMpath] = fBmalgorithm (H,n,M)
%
% This function help to generate fractional Brownian motion by correlated random
walks with Hurst index H.
%
% Input:
% H <- Hurst parameter
% n <- number of time steps
% M <- number of trajectories
%
% Output:
% fBmpath <- An fBm path constructed from correlated random walk
%
% Step 1: Find suitable p values
%
% One step correlation of fBm increments

fbmcorr = 0.5*((abs(2))^(2*H)+(abs(2-1-1))^(2*H)-2*(abs(2-1))^(2*H));

for i=1:M;
        u=rand;
```

```
            value(1,i)=((-1)*(1-u) ^ (1/(2-2*H)))/4;
end
p= 0:0.01:0.5;

for j=1:51
            pr = p(1,j);
            mu = [0 0];
            bivarcorr=[1 fbmcorr; fbmcorr 1];
            cdfprob = mvncdf([norminv(pr,0,1),norminv(pr,0,1)],mu,bivarcorr);
            cal(1,j)=cdfprob-pr;
end

for i=1:M;
      for j=1:50;
            if (value(1,i) <=cal(1,51))
                pm(1,i)= 0.5;
            elseif (value(1,i)<= cal(1,j)) && (value(1,i) >= cal(1,j+1))
          pm(1,i)= p(1,j+1);
          end
      end
end

% Step 2:Obtain M independent p values for M trajectories

for i=1:M;
            pe = pm(1,i);
            cdfprob = mvncdf([norminv(pe,0,1),norminv(pe,0,1)],mu,bivarcorr);
            trp(1,i) = 2*cdfprob-2*pe+1;
end
```

% Step 3:Construct M correlated random walk trajectories

```
for i=1:M;
        pe=pm(1,i);
        cdfprob = mvncdf([norminv(pe,0,1),norminv(pe,0,1)],mu,bivarcorr);
        trp(1,i)=2*cdfprob-2*pe+1;
end

for k=1:M
        u=rand;
        if (u <= 0.5 );
     ber=0;
        else
     ber=1;
        end

E(k,1)=2*ber-1;
mj= trp(1,k);

    for w=2:n
        u=rand;
              if (u <= mj);
        berpj=0;
              else
        berpj=1;
              end
    E(k,w)=E(k,w-1)*(2*berpj - 1);
    end
end
```

```
for j=1:M;
    X(j,1)=E(j,1);
    for i=2:n;
        X(j,i)=X(j,i-1)+E(j,i);
    end
end
```

Step 4: Generate a fBm path by using Theorem 3.

```
for k=1:n;
    btjpath(1,k) = (sqrt(H*(2*H-1)/gamma(2*H-1))*sum(X(:,k))/((sqrt(M))*n^H));
end
```