# Semester Project (Fall 2019)
# Computational Techniques in Mathematical Finance

Arijit Paul

Reg. No. 20171089

November 17, 2019

## 1 Introduction

The subject of mathematical finance demands the handling and analysis of data in large sizes. Management of such large data is very difficult to do manually and various computational techniques are required for this purpose. In this project, we will look at some of these techniques in detail.

## 2 Highlights of the Project

Initially, we mostly focused on how to generate random numbers that follow various distributions. More specifically, we looked into some definitions and theorems that will enable us to generate random numbers from any distribution $F_X$, provided we have access to random numbers from Uniform[0,1] distribution.

Then we went on to describe how to generate Normal random numbers, where we discussed about the Box-Muller Method in some detail. We used this method to simulate Normal random numbers. From there on we moved to study Brownian Motion (BM) and some of its important properties. Using the Normal random numbers, we simulated and plotted a discrete approximation of a realization of a Brownian Motion. Once we were equipped with the ability to create realizations of BM, we were able to simulate a number of different stochastic processes.

In order to be certain of the fact that the numbers generated indeed follow the given distribution, we briefly discussed the theory behind the K-S Test.

Next, our discussions moved to pure-jump processes, where we studied the properties of Markov Pure-Jump Processes, and tried simulating them. We achieved this by first defining and simulating the embedded discrete-time Markov Chain, and then combining that with the transition times to form the Markov Pure-Jump Process.

# 3 Generation of Random Numbers

We begin this section by stating and proving a very powerful theorem.

**Theorem 3.1.** *Let $X$ have continuous CDF $F_X$ and define the random variable $Y$ as $Y = F_X(X)$. Them $Y$ is uniformly distributed on $(0, 1)$, that is, $P(Y \leq y) = y$ for $0 < y < 1$.*

There are a few things worth noting before we look into the proof of the theorem. If the CDF $F_X$ is strictly increasing, then its inverse, $F_X^{-1}$ is well-defined by

$$F_X^{-1}(y) = x \iff F_X(y) = y \tag{1}$$

However, when $F_X$ is not strictly increasing (a CDF cannot be decreasing), i.e., it is constant on some interval, then the above definition does not work. One way of resolving the issue is to define $F_X^{-1}$ as

$$F_X^{-1}(y) = sup\{x|F_X(x) \leq y\} \tag{2}$$

which agrees with our definition (1) for strictly increasing cases.

*Proof.* For $0 < y < 1$, we have

$$\begin{aligned} P(Y \leq y) &= P(F_X(X) \leq y) && (3)\\ &= P(F_X^{-1}[F_X(X)] \leq F_X^{-1}(y)) && (4)\\ &= P(X \leq F_X^{-1}(y)) && (5)\\ &= F_X(F_X^{-1}(y)) && (6)\\ &= y && (7) \end{aligned}$$

Now, Step (4) is true since by definition (2), $F_X^{-1}$ is non-decreasing, and that means

$$x \leq y \implies F_X^{-1}(x) \leq F_X^{-1}(y)$$

Step (5) is straightforward if $F_X$ is strictly increasing. However, if that is not the case, i.e., when it is constant for some interval $[x_1, x_2]$ in the domain of $F_X$, then $F_X^{-1}(x)$ might not be equal to $x$. Indeed, for any $x \in [x_1, x_2]$, we have $F_X^{-1}(x) = x_2$ from definition (2). However, the probability equality makes sense, since a flat region in the CDF curve implies that the region concerned has zero probability, i.e., $P(X \leq x) = P(X \leq x_2)$ , since $P(x < X < x_2) = F_X(x_2) - F_X(x) = 0$. □

Theorem 3.1 enables us to generate random numbers from various distributions by using Uniform random numbers. All one has to do is generate a Uniform random number $u$ between 0 and 1, and solve for $x$ in $F_X(x) = u$. However, this theorem only holds for continuous CDFs, not in general (Definition of CDF only implies right continuity). Even if the CDF is continuous, there is the task of inverting $F_X$, which might not be trivial. In the following sections we try to generalize the above notion.

## 3.1 Some Definitions

**Definition 3.1** (**rcll Function**). *Let (M,d) be a metric space and let $E \subseteq \mathbb{R}$. A function $f : E \to M$ is called a rcll function, if $\forall\ t \in E$,*

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

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

Using above definition, it is easy to see that all continuous functions are rcll and so are all cumulative distribution functions.

**Definition 3.2** (**Upper Semicontinuous Function**). *Suppose $X$ is a topological space, then the function $G : X \to \mathbb{R} \cup \{-\infty, \infty\}$ is said to be upper semicontinuous if $\forall\ \alpha \in \mathbb{R}$, $G^{-1}([-\infty, \alpha)) = \{x \in X \mid G(x) < \alpha\}$ is an open set in $X$.*

## 3.2 Relevant Theorems

**Theorem 3.2.** *If $G$ is a upper semicontinuous on $\mathbb{R}$ and a non-decreasing function then $G^{-1}([a, b)) = [c, d)$ for some real $c, d$.*

*Proof.* The following proof follows from basic set theory algebra and the fact that $G$ is upper semicontinuous and non-decreasing.

$$
\begin{aligned}
G^{-1}([a,b)) &= G^{-1}((-\infty, b) \cap (-\infty, a)^c) && ((-\infty, a)^c = [a, \infty)) && (8)\\
&= G^{-1}((-\infty, b)) \cap (G^{-1}(-\infty, a))^c && && (9)\\
&= (-\infty, d) \cap (-\infty, c)^c && && (10)\\
&= [c, d) && && (11)
\end{aligned}
$$

for some real $c, d$. Hence proved. □

**Theorem 3.3.** *If $F : \mathbb{R} \to \mathbb{R}$ is non-decreasing then $F^{\leftarrow}(u) : \mathbb{R} \to \mathbb{R} \cup \{-\infty, \infty\}$ defined by $F^{\leftarrow}(u) = \sup\{x \mid F(x) \leq u\}$ is non-decreasing and upper semicontinuous.*

Clearly, when $F$ is continuous and strictly increasing, the above definition implies $F^{\leftarrow} = F^{-1}$, where $F^{-1}(y) = x \iff F(x) = y$.

*Proof.* By definition we have $F^{\leftarrow}(u) = \sup\{x \mid F(x) \leq u\}$ which implies $F^{\leftarrow}$ is non-decreasing. Showing the upper semicontinuity is essentially the same as showing the set $\mathbf{U} := \{u \mid F^{\leftarrow}(u) < \lambda\}$ is open.

From the definition of $\mathbf{U}$, it is evident that $\mathbf{U}$ is either of the form $(-\infty, \beta)$ or $(-\infty, \beta]$ for some $\beta \leq \infty$ (if $F^{\leftarrow}$ is bounded and $\lambda \geq \sup F^{\leftarrow}(u)$) then $\beta = \infty$). Hence to show $\mathbf{U}$ is open, given a $u_0 \in \mathbf{U}$, it is enough to produce a $u_1 (> u_0)$ that lies in $\mathbf{U}$.

Now let $u_0 \in \mathbf{U}$, which means that it satisfies $\sup\{x \mid F(x) \leq u_0\} < \lambda$. By the Archimedian property, we can find a $\lambda^{'} < \lambda$ such that

$$\sup\{x \mid F(x) \leq u_0\} < \lambda^{'} < \lambda$$

Thus if for some $x$, $F(x) \leq u_0$, then $x \leq F^{\leftarrow}(u_0) < \lambda^{'}$. Therefore, $F(\lambda^{'}) > u_0$ . This has further implication that $\exists\ u_1$ such that $u_0 < u_1 < F(\lambda^{'})$ (again by Archimedean property). We now claim that this $u_1 \in \mathbf{U}$, and that will complete the proof. Indeed, if $u_1 \notin \mathbf{U}$ then by definition, $\lambda^{'} < F^{\leftarrow}(u_1)$ which implies $F(\lambda^{'}) \leq u_1$, giving us a contradiction. Hence, $u_1 \in \mathbf{U}$ and we are done.

□

After studying the above theorems, we are now well-equipped to tackle the case when the CDF is not continuous.

**Theorem 3.4.** *Let $X \sim F_X$, i.e. $F_X$ is the CDF of a random variable X, and $F_X$ need not be continuous or injection. Then*

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

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

*Proof.* $F^{\leftarrow}(u) \in \mathbb{R}\ \forall\ u \in (0,1)$. So, $F^{\leftarrow}(U) \in \mathbb{R}$ (almost surely). Since $F_X$ is non-decreasing and so $F^{\leftarrow}$ is upper semicontinuous and non-decreasing. It is also evident from the definition that $F^{\leftarrow}$ is rcll, since the left limit exists and

$$\begin{aligned} \lim_{t \to u^+} F^{\leftarrow}(t) &= \lim_{t \to u^+} \sup\{x \mid F_X(x) \leq t\} && (12) \\ &= sup\{x \mid F_X(x) \leq u\} && (13) \\ &= F^{\leftarrow}(u) && (14) \\ & && (15) \end{aligned}$$

where (13) follows from the fact that the CDF is right continuous.

Also, by theorem 3.2, we have $(F^{\leftarrow})^{-1}([a,b)) = [c,d)$ for some real $c, d$ if $[a,b) \subset (0,1)$, and the following follow trivially:

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

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

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

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

Now we claim that the following two inequalities hold:

$$F_X(b) = d \tag{16}$$
$$F_X(a) = c \tag{17}$$

We will prove (16), as the proof of (17) is analogous. Let if possible, $F_X(b) < d$. Then using the right continuity of $F_X$, $\exists\ \epsilon > 0$ such that $F_X(b+\epsilon) < d$. Hence from (iii) $\exists\ u$ with $F_X(b+\epsilon) < u \leq d$. From (iii) and monotonicity of $F^{\leftarrow}$, $u \leq d$ implies $F^{\leftarrow}(u) \leq b$. However, $F^{\leftarrow}(u) \leq b$ as $F_X(b+\epsilon) < u$, $b + \epsilon \leq F^{\leftarrow}(u)$. So we have the following relation as a result:

$$F^{\leftarrow}(u) \leq b < b + \epsilon \leq F^{\leftarrow}(u).$$

Hence we arrive at a contradiction. Now let us consider the other case. If possible, assume $d < F_X(b)$. Then $\exists\ u^{'}$ such that $d < u^{'} < F_X(b)$. If $e := \sup\{x \mid F_X(x) \leq u^{'}\}$, then $F_X(e^{-}) \leq u^{'} < F_X(b)$. Therefore, from monotonicity of $F_X$, $F^{\leftarrow}(u^{'}) \leq e \leq b$ and $\sup\{x \mid F_X(x) \leq u^{'}\} \leq b$. Hence using (iii) we have the following relation as a result:

$$d < u^{'} \leq \sup_{u}\{F^{\leftarrow}(u) \leq b\} = d$$

which is a contradiction. Therefore, we must have $F_X(b) = d$. Hence (16) is true as claimed.

Hence, we arrive at the following conclusion:

$$\begin{aligned}
\mathbb{P}(Y \in [a,b)) &= \mathbb{P}(F^{\leftarrow}(U) \in [a,b)) && (18)\\
&= \mathbb{P}(U \in (F^{\leftarrow})^{-1}([a,b))) && (19)\\
&= \mathbb{P}(U \in [F_X(a), F_X(b))) && (20)\\
&= F_X(b) - F_X(a). && (21)\\
\text{Hence } Y \sim F_X. && (22)\\
& && (23)
\end{aligned}$$

□

### 3.3 Normal Random Numbers

In this section, we will discuss methods for generating random numbers from Normal (Gaussian) distribution. We assume that we have access to $Uniform[0,1]$ random numbers, and discuss ways of generating $N(0,1)$ numbers from them. Numbers from $N(\mu, \sigma^2)$ can then be obtained using a simple location-scale transformation.

### 3.3.1 Numerical Approximation Method

If $X$ is a random variable such that $X \sim N(0,1)$, then the pdf of $X$ is given by (for $x \in \mathbb{R}$)

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{24}$$

We will, however, require the CDF, which is obtained by integrating the pdf as follows

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{25}$$

However, this integral does not have a closed form and we have to resort to numerical methods (trapezoidal method, for example). But, the lower limit of $-\infty$ does not make sense to a computer. Here we notice that the probability that $X$ lies below $-5$ is practically negligible ($P(X \leq -5) = 2.86651571879 \times 10^{-7}$) and we can safely ignore it. So, for this calculation we will use the CDF

$$F_X(x) = \int_{-5}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad\qquad x \in [-5, 5] \tag{26}$$

with $F_X(x) = 0$ for $x < -5$ and $F_X(x) = 1$ for $x > 5$.

Now all that remains is inverting $F_X$. For this purpose we can use various computing tricks. This method, although less efficient that the algorithm that follows, has the added advantage that it can be used for generating random numbers from almost all distributions.

### 3.3.2 Box-Muller Transformation

The procedure described above is not very efficient, since it requires numerical approximations in a number of steps and errors may creep in. Therefore, we shall look at another method that generates two $N(0,1)$ numbers from two $Uniform(0,1)$ numbers. This method is known as the **Box-Muller Transformation**.

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

$$h(x,y) = \frac{1}{2\pi} e^{\left(-\frac{x^2+y^2}{2}\right)} \tag{27}$$

where $x, y \in \mathbb{R}$

Next we consider the polar coordinates of the point $(X,Y)$ given by the relation:

$$(R, \Theta) = \left(\sqrt{X^2+Y^2}, arctan(\frac{Y}{X})\right) \tag{28}$$

with $R \geq 0$ and $\Theta \in [0, 2\pi)$ The above map is very clearly bijective, and in order to obtain the joint density $g$ of $R$ and $\Theta$, we look at the inverse of the above transform:

$$(X, Y) = (Rcos\Theta, Rsin\Theta) \tag{29}$$

and find the Jacobian, $J$ as:

$$|J| = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} \tag{30}$$

$$= \begin{vmatrix} cos\theta & -rsin\theta \\ sin\theta & rcos\theta \end{vmatrix} \tag{31}$$

$$= r(cos^2\theta + sin^2\theta) \tag{32}$$

$$= r \tag{33}$$

The joint density is therefore: $g(r, \theta) = |J|h(x, y)$ , i.e.,

$$g(r, \theta) = \frac{1}{2\pi} r e^{-\frac{r^2}{2}} \tag{34}$$

for $r \geq 0, 0 \leq \theta < \pi$. This is the product of two pdf's, namely $Uniform[0, 2\pi]$ distribution and $Rayleigh(1)$ distribution. The $Rayleigh(\sigma)$ distribution is defined by:

$$f(r; \sigma) = \frac{r}{\sigma^2} e^{\frac{-r^2}{2\sigma^2}}, r \geq 0 \tag{35}$$

with the CDF

$$F_R(r; \sigma) = 1 - e^{\frac{-r^2}{2\sigma^2}}, r \geq 0 \tag{36}$$

where $\sigma$ is the scale parameter (1 in our case). Now, clearly the CDF is continuous and a strictly increasing function. Therefore, by theorem 3.1 $F_R(R)$ follows $Uniform[0, 1]$ distribution, and $F_R^{-1}(U) = \sqrt{-2\sigma^2 \ln(1 - U)}$ follows $Rayleigh(\sigma)$ distribution, where $U \sim Uniform[0, 1]$. Since we need $Rayleigh(1)$ distribution, we use the relation

$$R = \sqrt{-2 \ln(1 - U)} \tag{37}$$

On the other hand, if $\Theta \sim Uniform[0, 2\pi]$, we have the CDF

$$F_\Theta(\theta) = \begin{cases} \frac{\theta}{(2\pi)}, 0 \leq \theta \leq 2\pi \\ 0, \text{ otherwise .} \end{cases} \tag{38}$$

It follows that $F_\Theta(\Theta)$ follows $Uniform[0, 1]$ distribution and the inverse function

$$\Theta = 2\pi V \tag{39}$$

follows $Uniform[0, 2\pi]$ distribution (where $V \sim Uniform[0, 1]$). Note that in (26), the point $2\pi$ is included as well, which is the same as $\theta = 0$. But, since $P(\Theta = 2\pi) = 0$ this does not pose much problem.

Hence, we arrive at the required formulae for obtaining the standard normal variables$X$ and $Y$:

$$X = R\cos(\Theta) = \sqrt{-2\ln(1-U)}\cos(2\pi V) \tag{40}$$

$$Y = R\sin(\Theta) = \sqrt{-2\ln(1-U)}\sin(2\pi V) \tag{41}$$

Once we obtain $X$ and$Y$, it is a simple matter to location-scale transformation to obtain $N(\mu, \sigma^2)$ random variables. However, the pair of random variables obtained in this method are independent, and hence, uncorrelated. One can obtain a pair of correlated normal random variables using uncorrelated standard normal variables in the following method:

Suppose we have random variables $Z_1$ and $Z_2$ such that $Z_i \sim N(0,1)$ for $i = 1, 2$. Assume the correlation, $\rho(Z_1, Z_2) = 0$. We wish to generate $P, Q \sim N(0,1)$ with correlation, $\rho(P, Q) = \rho_0$. For this purpose, we would require the following lemmas:

**Lemma 3.5.** *Suppose, $X$ be a random variable such that $X \sim N(\mu, \sigma^2)$, then for $a \in \mathbb{R}$, $aX \sim N(a\mu, a^2\sigma^2)$.*

*Proof.* Let, $Y = aX$. Then we have the CDF of $Y$ as:

$$\begin{aligned} F(Y) = P(Y \leq y) &= P(aX \leq y) && (42)\\ &= P(X \leq \frac{y}{a}) && (43)\\ &= \int_{-\infty}^{\frac{y}{a}} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx && (44) \end{aligned}$$

Therefore, the pdf of $Y$,

$$\begin{aligned} f(y) &= \frac{dF(y)}{dy} && (45)\\ &= \frac{d}{dy}\Big(\int_{-\infty}^{\frac{y}{a}} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx\Big) && (46)\\ &= \Big(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\frac{y}{a}-\mu)^2}{2\sigma^2}}\Big)\Big(\frac{1}{a}\Big) - 0 && (47)\\ &= \frac{1}{(a\sigma)\sqrt{2\pi}} e^{-\frac{(x-a\mu)^2}{2(a\sigma)^2}} && (48) \end{aligned}$$

Which is the pdf of $N(a\mu, a^2\sigma^2)$. □

**Lemma 3.6.** *Suppose, $X$ and $Y$ are independent random variables such that, $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$. Then*

$$Z = X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2). \tag{49}$$

*Proof.* $X$ and $Y$ are independent continuous random variables. Define $W = X$ and $Z = X + Y$. Hence, $X = W$ and $Y = Z - W$.

The Jacobian for the density transformation corresponding to the transformation $(X, Y) \mapsto (W, Z)$ is clearly 1.

Now, we have

$$f_{W,Z}(w, z) = f_{X,Y}(w, z - w) \tag{50}$$
$$= f_X(w) f_Y(z - w) \tag{51}$$

where (51) follows from the independence of $X$ and $Y$. Therefore, the marginal distribution of $Z$ is given by

$$f_Z(z) = \int_{-\infty}^{\infty} f_{W,Z}(w, z) dw \tag{52}$$
$$= \int_{-\infty}^{\infty} f_X(w) f_Y(z - w) dw \tag{53}$$
$$= \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dw \tag{54}$$

(54) is called the convolution of $f_X$ and $f_Y$.

Now,

$$f_X(x) = \frac{1}{\sigma_X \sqrt{2\pi}} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}}$$

and

$$f_Y(y) = \frac{1}{\sigma_Y \sqrt{2\pi}} e^{-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}}$$

Substitution of these values in (54) and subsequent algebraic manipulations yield the desired result. □

Now we are well-equipped to handle our original problem. Say, we have $P = aZ_1 + bZ_2$ and $Q = cZ_1 + dZ_2$, with $a^2 + b^2 = 1, c^2 + d^2 = 1$. Then we have,

$$E[P] = aE[Z_1] + bE[Z_2] \tag{55}$$
$$= a.0 + b.0 \tag{56}$$
$$= 0 \tag{57}$$

Similarly, $E[Q] = 0$. And,

$$Var[P] = a^2 Var[Z_1] + b^2 Var[Z_2] \tag{58}$$
$$= a^2.1 + b^2.2 \tag{59}$$
$$= a^2 + b^2 = 1 \tag{60}$$

Similarly, $Var[Q] = c^2 + d^2 = 1$. Next, we look at the covariance of $P$ and $Q$.

$$\begin{aligned} Cov(P,Q) &= Cov(aZ_1 + bZ_2, cZ_1 + dZ_2) && (61)\\ &= ac.Cov(Z_1,Z_1) + ad.Cov(Z_1,Z_2) + bc.Cov(Z_2,Z_1) + bd.Cov(Z_2,Z_2) && (62)\\ &= ac.1 + ad.0 + bc.0 + ad.1 && (63)\\ &= ac + bd && (64)\\ &= ac + \sqrt{(1-a^2)(1-c^2)} && (65) \end{aligned}$$

Now, $\rho(P,Q) = Cov(P,Q)$, since $Var[P] = Var[Q] = 1$. If we want $\rho(P,Q) = \rho_0$, we have to solve $Cov(P,Q) = \rho_0$ for $a$ and $c$. One such choice can be $a = 1$ and $c = \rho_0$, i.e., $P = Z_1$ and $Q = \rho_0 Z_1 + \sqrt{(1-\rho_0^2)}Z_2$.

# 4 Simulation of Normal Random Numbers

## 4.1 K-S Test

Before going into the algorithms for generating random numbers, we make a slight digression, and look at how we can test if a sequence of real numbers follow a certain distribution. One way of verifying this would be to use the Kolmogorov–Smirnov test (or K-S test), which is a non-parametric test of the equality of continuous (or discontinuous), one-dimensional probability distributions that can be used to compare a sample with a reference probability distribution (one-sample K–S test), or to compare two samples (two-sample K–S test).

We concern ourselves with the one-sample K-S test only. The Kolmogorov–Smirnov statistic quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution. The null distribution of this statistic is calculated under the null hypothesis that the sample is drawn from the reference distribution. In this case, the distribution considered under the null hypothesis may be continuous, purely discrete or mixed.

The empirical distribution, $F_n$, for $n$ iid ordered observations$X_i$ is defined by

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{[-\infty,x]}(X_i)$$

where $I_{[-\infty,x]}$ denotes the indicator function. The K-S statistic for a given CDF $F$ is

$$D_n = \sup_x |F_n(x) - F(x)| \tag{66}$$

It can be shown that if the sample comes from the distribution $F$, then $D_n$ converges almost surely as $n$ goes to infinity. The critical values, $D_{n,\alpha}$ do not depend on $F$ (here $\alpha$ is the level of significance) and for $n \geq 50$, the critical value for $\alpha = 0.05$ is approximately given by

$$D_{n,0.05} = \frac{1.358}{\sqrt{n}}$$

If $D_n \geq D_{n,0.05}$, we reject the null hypothesis that the sample comes from a population following the distribution $F$.

### 4.2 Generation of Uniform Random Numbers

The CDf of $Uniform[0,1]$ distribution is given by

$$F_U(u) = \begin{cases} 0, \text{if } u < 0 \\ u, \text{if } u \in [0,1) \\ 1, \text{if } u \geq 1. \end{cases} \tag{67}$$

However, it is not possible to produce truly random numbers by means of any algorithm. The best we can do is produce numbers that "appear" random. There are many sophisticated algorithms to generate these "random-looking" numbers (also called Pseudo-random numbers). The tools we would be using are relatively simple (and consequently, don't perform that well in tests for randomness) and these class of random number generators are called "Linear Congruential Generators (LCG)".

In this method, one generates a sequence of numbers based on some modular operations. The sequence is completely deterministic, and depends only on the initial seed (apart from the pre-determined parameters). Now, one might ask how a deterministic sequence of numbers are being passed off as "random". That is why we are conducting a K-S test, to check if our sequence can be considered "random enough".

The fact that the sequence is deterministic is actually advantageous as it grants reproducibility to a numerical experiment. Therefore, even though there are ways to randomize the numbers more effectively (by invoking the last few digits of the system, for example), we chose this method as it grants us the ability to reproduce any interesting results we find.

The LCG we are using uses the recursion relation

$$X_{n+1} = (aX_n + c) \bmod m \tag{68}$$

With $X_0$ is taken as the seed. $X_n/m$ is taken as the $(n+1)^{th}$ uniform random number. Since there are only finitely many values (namely, $m$) that $X_n$ can take, it is inevitable that there will be repetitions, and once one number is repeated, the sequence falls into a cycle, which is especially undesirable. However, suitable choices of the parameters $a, m$ and $c$ can go a long way in maximizing the period of a loop.

According to the Hull-Dobell Theorem, for $c \neq 0$, correctly chosen parameters allow the period to be equal to $m$, and that happens if and only if

1. $m$ and $c$ are relatively prime
2. $(a-1)$ is divisible by all prime factors of m

3. $(a-1)$ is divisible by 4 if $m$ is divisable by 4.

For our LCG we have used the values:

$$m = 2^{48}$$
$$a = 25214903917$$
$$c = 11$$

For each run of 50,000 numbers, we have tabulated the value of the seed, a histogram of the uniform random numbers generated, the "difference function" (which calculates the absolute difference between the empirical and the theoretical CDFs for the purpose of the K-S test, i.e., $|F_n(i) - F(i)|$ for each $i \in \{1, 2, ..., 50000\}$), and the value of the K-S statistic denoting whether we can take the sequence to be coming from uniform random numbers. The results are tabulated in $Table - 1$.

### 4.3 Generation of Normal Random Numbers

Once we obtain a sequence of uniform random numbers, we used that list and the Box-Muller Method to generate normal random numbers. For each seed of LCG, we have only proceeded to the next step (i.e., generation of the random numbers), if the K-S test came out to be positive.

We have computed 5 runs, corresponding to each of the 5 runs in $Table - 1$. Again, each run has 50,000 numbers, and we have tabulated the seed corresponding to the LCG, the histogram, the "difference function", and the results of the K-S Tests. The test fails to reject any of our runs as a sequence of normal random numbers, and hence, we consider the our algorithm to be feasible. Results are tabulated in $Table - 2$.

Figure 1: Table 1

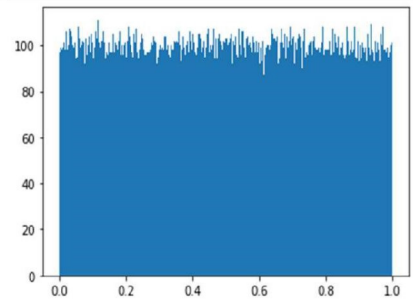
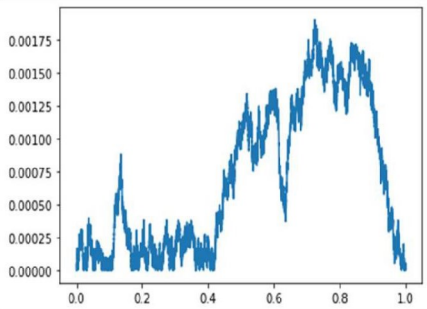
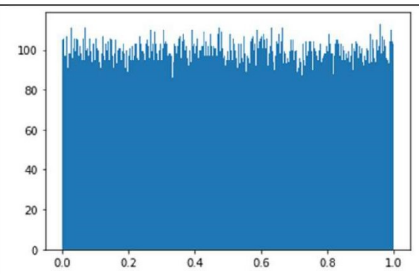
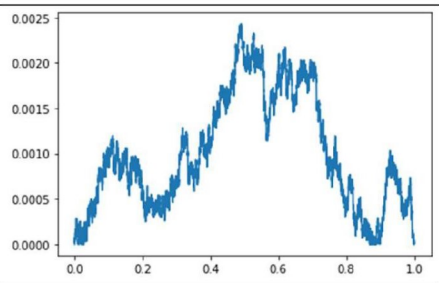
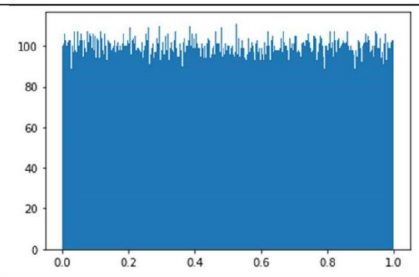
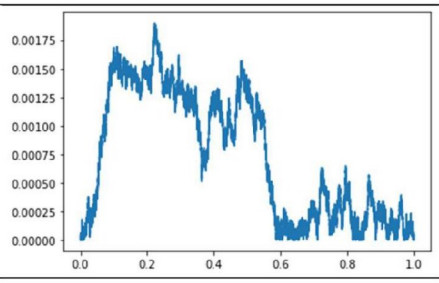
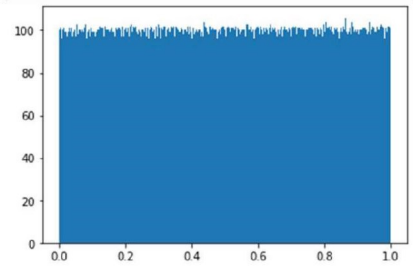
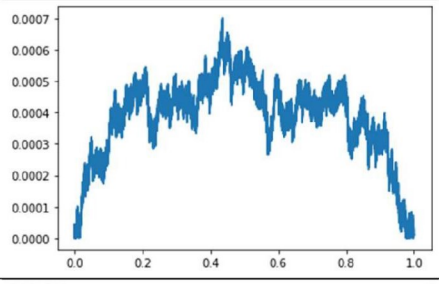
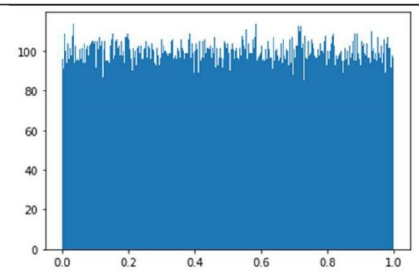
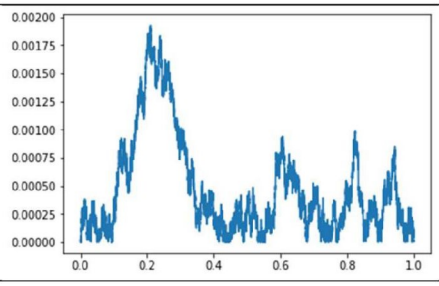
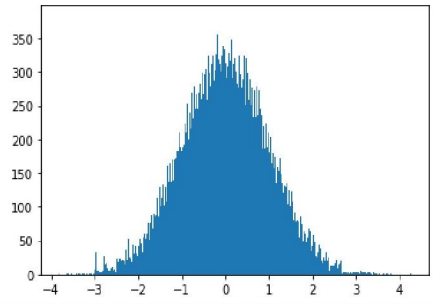
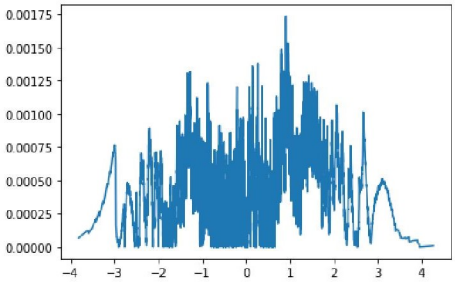
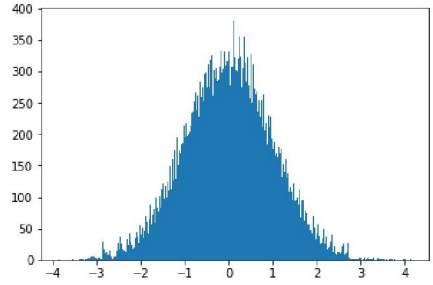
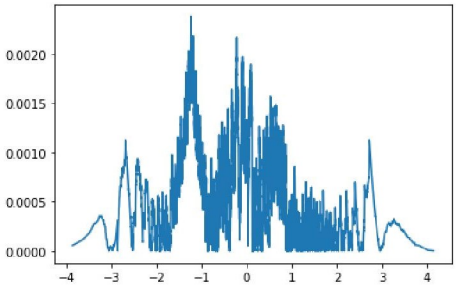
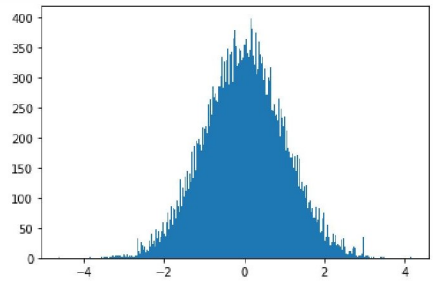
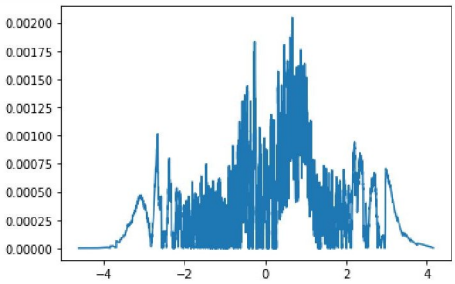
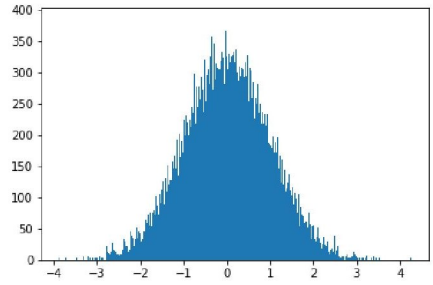
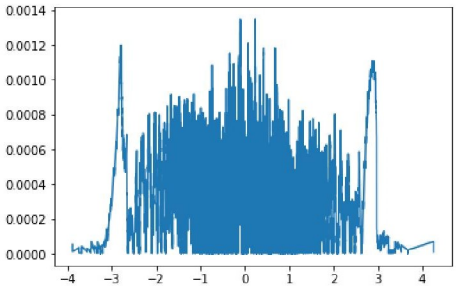
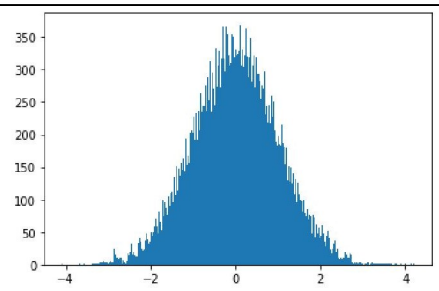
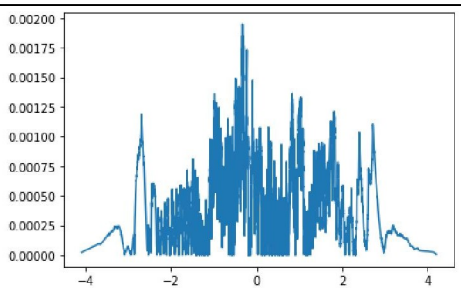
| Seed | Histogram (500 bins) | Difference Function | K-S Test Result |
|---|---|---|---|
| 1234567 |  |  | $D_{cr}$=0.0060<br>$D_{ob}$=0.0019<br>Hence, we accept the sequence as uniform random numbers. |
| 9876543 |  |  | $D_{cr}$=0.0060<br>$D_{ob}$=0.0024<br>Hence, we accept the sequence as uniform random numbers. |
| 2345678 |  |  | $D_{cr}$=0.0060<br>$D_{ob}$=0.0018<br>Hence, we accept the sequence as uniform random numbers. |
| 7777777 |  |  | $D_{cr}$=0.0060<br>$D_{ob}$=0.0007<br>Hence, we accept the sequence as uniform random numbers. |
| 7 |  |  | $D_{cr}$=0.0060<br>$D_{ob}$=0.0019<br>Hence, we accept the sequence as uniform random numbers. |

Figure 2: Table 2

| Seed | Histogram (500 bins) | Difference Function | K-S Test Result |
|---|---|---|---|
| 1234567 |  |  | $D_{cr}$=0.00608<br>$D_{ob}$=0.00173<br>Hence, we accept the sequence as normal random numbers. |
| 9876543 |  |  | $D_{cr}$=0.00608<br>$D_{ob}$=0.00237<br>Hence, we accept the sequence as normal random numbers. |
| 2345678 |  |  | $D_{cr}$=0.00608<br>$D_{ob}$=0.00205<br>Hence, we accept the sequence as normal random numbers. |
| 7777777 |  |  | $D_{cr}$=0.00608<br>$D_{ob}$=0.00135<br>Hence, we accept the sequence as normal random numbers. |
| 7 |  |  | $D_{cr}$=0.00608<br>$D_{ob}$=0.00195<br>Hence, we accept the sequence as normal random numbers. |

# 5 Brownian Motion

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

Let us now discuss more on the Brownian motion.

**Definition 5.1.** *The family of random variables $\{B_t\}_{t\geq 0}$, is said to be Standard Brownian Motion if it satisfies the following conditions:*

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

We will not try and prove that such a family of random variables exist but assume their existence (this can be proved by the application of Kolmogorov's Extension Theorem). Next we try to simulate a discrete-time approximation of Brownian Motion. Following that, we would look at some of the important properties related to this stochastic process.

## 5.1 Simulating an Approximation to Brownian Motion

Once we have a list of standard normal numbers, we can use the following algorithm to generate a discrete time approximation to Brownian Motion.

We choose a small value $h > 0$ which would correspond to the time increment. We also set $B_0 = 0$. Next, we generate a sequence of random numbers from a $N(0, h)$ distribution which can be achieved by a simple scale transformation of standard normal numbers. We call this sequence $\{r_n\}_{n\in\mathbb{N}}$. Now, we define

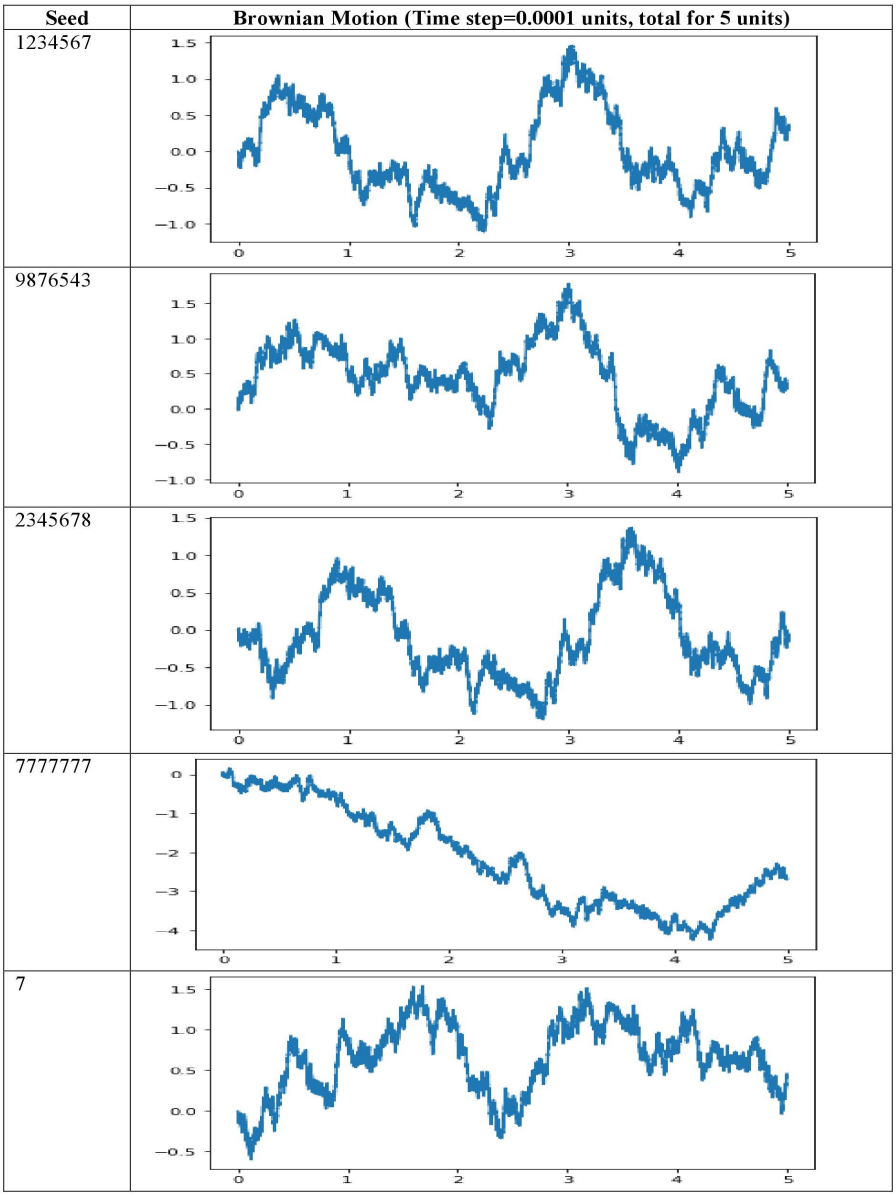
$$B_{t+h} = B_t + r_n \quad \forall t \in \{0, h, 2h, 3h, ...\}. \tag{69}$$

Here $r_n$ is a random number from $N(0, h)$ distribution. Obviously, $B_{t+h} - B_t = r_n \sim N(0, h)$, and these increments are independent if the terms of the sequence $\{r_n\}_{n\in\mathbb{N}}$ are independent. However, for property 4 to hold, we must show that $B_{t+mh} - B_t \sim N(0, mh)$. Now, $B_{t+mh} - B_t$ is a

sum of $m$ i.i.d random variables from $N(0, h)$ distribution. By recursively applying Lemma 3.6, we see that the sum is again a normal random variable, with mean and variance equal to the sum of means and variances of the given random variables. Hence, $B_{t+mh} - B_t \sim N(0, mh)$.

But the above algorithm gives us discrete points. In order to satisfy property 1, we connected successive points by straight lines to get a continuous graph. The following table, $(Table - 3)$ tabulates 5 runs of this algorithm, each corresponding to one of the normal random number sequences in $Table - 2$.

Figure 3: Table 3



## 5.2 Some Properties of Brownian Motion

Now we look at some of the important properties of Brownian Motion. First, we would require the following definitions.

**Definition 5.2.** *Let $f : [a, b] \to \mathbb{R}$ be a continuous function. If $\Pi$ denotes the set of all partitions of the interval $[a, b]$, then the total variation of $f$ (denoted by $TV(f)$) is defined as:*

$$TV(f) = \sup_{\pi \in \Pi} \Big( \sum_{i=0}^{n-1} |f(x_{i+1}^{(n)}) - f(x_i^{(n)})| \Big) \tag{70}$$

*where $\pi$ is a partition of $[a, b]$ of size $n$.*

For example, consider $f : [0, 1] \to \mathbb{R}$ defined by $f(x) = x$. We have,

$$\sum_{i=0}^{n-1} |f(x_{i+1}^{(n)}) - f(x_i^{(n)})| = \sum_{i=0}^{n-1} |x_{i+1}^{(n)} - x_i^{(n)}| = \sum_{i=0}^{n-1} (x_{i+1}^{(n)} - x_i^{(n)}) = 1 - 0 = 1$$

Hence, $TV(f) = 1$.

**Definition 5.3.** *Let $f : [a, b] \to \mathbb{R}$ be a continuous function. If $\Pi$ denotes the set of all partitions of the interval $[a, b]$, then the $p^{th}$ variation of $f$ (denoted by $V^p(f)$) is defined as:*

$$V^p(f) = \sup_{\pi \in \Pi} \Big( \sum_{i=0}^{n-1} |f(x_{i+1}^{(n)}) - f(x_i^{(n)})|^p \Big) \tag{71}$$

*where $\pi$ is a partition of $[a, b]$ of size $n$.*

**Definition 5.4.** *(Quadratic Variation of a Stochastic Process) Let $X : [0, \infty) \to \mathbb{R}$ be a continuous function. Consider a sequence of partitions*

$$\pi^{(n)} = \{t_i | 0 \leq i \leq n \text{ and } 0 = t_0 < t_1 < ... < t_n\}$$

*with the following properties:*

1. $\lim_{n\to\infty} t_n = \infty$
2. $\lim_{n\to\infty} \Delta t_i = \lim_{n\to\infty} (t_{i+1} - t_i) = 0$

*Then the quadratic variance of $X$ is given by*

$$\langle X \rangle = \lim_{n\to\infty} \sum_{i=0}^{n-1} |X(t_{i+1}) - X(t_i)|^2 \tag{72}$$

If we consider a stochastic process $X(t)$ upto a time $T$, then the quadratic variance can be represented as $\langle X \rangle_T$ and $\{\langle X \rangle_T\}_{T \geq 0}$ is called the quadratic variation process of $X_n$. After discussing all these definitions, we are now well equipped to state a rather strong theorem.

**Theorem 5.1.** *(Levy's Theorem) The quadratic variation process of Brownian Motion is the identity function of time (almost surely), i.e.,*

$$\langle B\rangle_T = T \quad a.s.\ \forall T \geq 0$$

*where $B : [0,\infty) \to \mathbb{R}$ is a Brownian Motion.*

*Proof.* Let $\pi^{(n)} = \{t_i^{(n)} | 0 = t_0^{(n)} < t_1^{(n)} < ... < t_{m_n}^{(n)} = T\}$ be a partition of the interval $[0,T]$which satisfies the conditions mentioned above. This describes a sequence of partitions for which the $n^{th}$ partition has $m_n$ elements. Now consider

$$\begin{aligned}
E\Big(\sum_{i=0}^{m_n-1} |B(t_{i+1}^{(n)}) - B(t_i^{(n)})|^2\Big) &= \sum_{i=0}^{m_n-1} E\Big(|B(t_{i+1}^{(n)}) - B(t_i^{(n)})|^2\Big) && (73)\\
&= \sum_{i=0}^{m_n-1} Var\Big(B(t_{i+1}^{(n)}) - B(t_i^{(n)})\Big) && (74)\\
&= \sum_{i=0}^{m_n-1} \Big(t_{i+1}^{(n)} - t_i^{(n)}\Big) && (75)\\
&= T && (76)\\
& && (77)
\end{aligned}$$

Hence we have obtained, $E(\langle B\rangle_T) = T$. Here step (74) follows because the expectation is zero. Now, if we can show that $Var(\langle B\rangle_T) = 0$ then we will be done. We have,

$$\begin{aligned}
Var\Big(\sum_{i=0}^{m_n-1} |B(t_{i+1}^{(n)}) - B(t_i^{(n)})|^2\Big) &= \sum_{i=0}^{m_n-1} Var\Big(|B(t_{i+1}^{(n)}) - B(t_i^{(n)})|^2\Big) && (78)\\
&= \sum_{i=0}^{m_n-1} \Big(E\big(|B(t_{i+1}^{(n)}) - B(t_i^{(n)})|^4\big) - \big(E(|B(t_{i+1}^{(n)}) - B(t_i^{(n)})|^2)\big)^2\Big) && (79)\\
&= \sum_{i=0}^{m_n-1} \Big(3(t_{i+1}^{(n)} - t_i^{(n)})^2 - (t_{i+1}^{(n)} - t_i^{(n)})^2\Big) && (80)\\
&= 2\sum_{i=0}^{m_n-1} (t_{i+1}^{(n)} - t_i^{(n)})^2 && (81)\\
&\leq 2\Big(\sup_i (t_{i+1}^{(n)} - t_i^{(n)})\Big) \sum_{i=0}^{m_n-1} (t_{i+1}^{(n)} - t_i^{(n)}) && (82)\\
&= 2\delta_n T && (83)
\end{aligned}$$

where $\delta_n = \sup_i (t_{i+1}^{(n)} - t_i^{(n)})$. The step (78) follows from the independent increment property of Brownian Motion, and consequently the Covariance term is zero. Step (80) is obtained by using the

fact that the Kurtosis of a normal random variable is 3, and that leads to the relation $E(X^4) = 3\sigma^2$ in our case, $\sigma^2$ being the variance.

But, we know that $(t_{i+1}^{(n)} - t_i^{(n)}) \to 0$ as $n \to \infty$. Hence, $\delta_n \to 0$ as $n \to \infty$. Therefore we have,

$$Var\Big( \sum_{i=0}^{m_n - 1} |B(t_{i+1}^{(n)}) - B(t_i^{(n)})|^2 \Big) \to 0 \text{ as } n \to \infty$$

which implies $Var(\langle B \rangle_T) = 0$. This implies

$$P(\{\omega | \langle B \rangle_T(\omega) = T\}) = 1 \ a.s. \ \forall \ T > 0 \tag{84}$$

It means that if $T > 0$ then almost surely(except possibly on a set with measure zero) for all Brownian Motion on $[0, T]$ the quadratic variation is $T$. However, this statement is only made for a point $T$. We would like to prove the more general statement

$$P(\{\omega | \langle B \rangle_T(\omega) = T\}) = 1 \ a.s. \ \forall \ T > 0 \implies P(\{\omega | \langle B \rangle_t(\omega) = t \ \ \forall \ t > 0\}) = 1 \tag{85}$$

(85) says that if $N_t = \{\omega | \langle B \rangle_t(\omega) \neq t\}$ then

$$P(N_T) = 0 \ \forall \ T > 0 \implies P(\cup_{0 < t \leq T} N_t) = 0$$

From set theory, we have, $P(\cup_{i \in I} S_i) \leq \sum_{i \in I} P(S_i)$, if the set $I$ is countable. But in the above statement, the union is over the index set $(0, T]$, which is uncountable, and we cannot use the above inequality. This can be resolved as follows:

Let $\{r_i\}_{i \in \mathbb{Q}}$ be a denumeration of the countable set $Q = \mathbb{Q} \cap [0, T]$. Then,

$$P(\cup_{i \in \mathbb{N}} N_{r_i}) \leq \sum_{i \in \mathbb{N}} 0 = 0 \tag{86}$$

But, probability cannot be less than 0 by Kolmogorov axioms. Hence,

$$P(\cup_{i \in \mathbb{N}} N_{r_i}) = 0 \implies P(\{\omega | \langle B \rangle_t(\omega) = t \ \ \forall \ t \in Q\}) = 1 \tag{87}$$

We observe that $\langle B \rangle_t : [0, \infty) \to \mathbb{R}$ is continuous in $t$, since $B$ is a continuous function of $t$. Consider $t_0 \in (0, T] \setminus Q$, and let $\{s_n\}_{n \in \mathbb{N}}$ be a rational sequence for which $s_n \in Q \ \forall \ n \in \mathbb{N}$ and $\lim_{n \to \infty} s_n = t_0$.

By continuity of $\langle B \rangle_t$, $\langle B \rangle_{s_n} \to \langle B \rangle_{t_0}$. Again, since $s_n \in Q \ \forall \ n \in \mathbb{N}$, we have

$$\langle B \rangle_{s_n} = s_n \ \forall \ n \in \mathbb{N} \implies \langle B \rangle_{s_n} \to t_0 \tag{88}$$

$$\implies \langle B \rangle_{t_0} = t_0 \tag{89}$$

But $t_0$ was an arbitrary element of $(0, T] \setminus Q$, so

$$\langle B \rangle_t = t \ \forall \ t \in (0, T] \setminus Q. \tag{90}$$

By combining (87) and (90), we conclude

$$\langle B \rangle_t = t \ \forall \ t \in (0,T] \implies P(\cup_{0<t\leq T} N_t) = 0 \tag{91}$$

This concludes the proof of Levy's theorem.

□

# 6 Pure-Jump Processes

**Definition 6.1.** *Let $\{X_t\}_{t\geq 0}$ be a stochastic process such that*

1. *$t \mapsto X_t$ is right-continuous with left limits (i.e., it is an rcll function)*
2. *On any interval $[a.b]$, $P(\#\{t \in [a,b] | X_t \neq X_{t^-}\} < \infty) = 1$, where $\#$ denotes the cardinality of the concerned set.*
3. *$\forall \ t \in [0,T], \ \ t \mapsto X_t$ is a step function with probability 1.*

*Then, $X = \{X_t\}_{t\geq 0}$ is a pure-jump process. Here, by a step function, we mean a real-valued function that can be written as a finite linear combination of indicator function on intervals.*

Having looked at the definition of pure-jump processes, we turn our attention to the special case of Markov Pure-Jump Processes.

## 6.1 Markov Pure-Jump Processes

**Definition 6.2.** *Suppose $X = \{X_t\}_{t\geq 0}$ is a pure jump process. Let*

$$\{T_1 < T_2 < ...\} := \{t \in [0,T] | X_t \neq X_{t^-}\}, \ T_0 = 0$$

*be the consecutive transition times. $X$ is a Markov Pure-Jump Process (MPP) if*

$$P(X_t \in A | X_s : s \in [0, t^{'}]) = P(X_t \in A | X_{t^{'}}) \ \ \forall \ t^{'} \leq t.$$

The property alluded to in the above definition is known as the Markov Property. Needless to say that there are Markov Processes (stochastic processes with the above mentioned property) that are not Pure-Jump Processes, but we restrict our attention to MPP only.

## 6.2 Some Properties of MPP

**Definition 6.3.** *It in the above definition of MPP, $P(X_t \in A|X_{t'}) = P(X_{t-t'} \in A|X_0) \ \ \forall\ t^{'} \leq t$, i.e., the time information is lost in the process, then that MPP is lnown as Time-Homogeneous MPP.*

**Definition 6.4.** *The quantity $n(t) := \sup\{n|T_n \leq t\}$ is called the class transition index. This implies that $T_{n(t)} \leq t < T_{n(t)+1} \ \ \forall\ t \geq 0$. Where there is no scope of confusion, we might drop the $'t'$, and denote successive class transition index as $T_0, T_1$ and so on.*

**Definition 6.5.** *A discrete-time Markov chain is a sequence of random variables $X_1, X_2, X_3, ...$ with the Markov property, namely that the probability of moving to the next state depends only on the present state and not on the previous states. We will get back to discrete-time Markov chains again when we try and simulate such a process.*

**Definition 6.6.** *Let $S$ be the state space of a MPP (i.e., the set of all possible values that $X_t$ can attain). For our purposes, we assume $S$ to be denumerable. We define $\bar{X}_n := X_{T_n} \ \ \in S$ and $\bar{X} := \{\bar{X}_n\}_{n\in\mathbb{N}}$. Then $\bar{X}_n$ is a discrete-time Markov Chain and is called the Embedded Discrete Time Markov Chain for the given MPP.*

Now, if $X = \{X_t\}_{t\geq 0}$ is any pure-jump process, it is easy to see that the probability

$$P\Big(X_{T_{n+1}} \in A,\ (T_{n+1} - T_n) \in (a,b)|X_0, T_0, X_{T_1}, T_1, ..., X_{T_n}, T_n\Big) \tag{92}$$

that dictates the time of next transition, along with the destination of the next transition completely determines the law of the aforementioned process. Now, if $X$ is a time-homogeneous MPP, then (92) reduces to

$$\begin{aligned} P\Big(X_{T_{n+1}} \in A,\ (T_{n+1} - T_n) \in (a,b)|X_{T_n}, T_n\Big) &= P\Big(X_{T_1} \in A,\ (T_1 - T_0) \in (a,b)|X_0, T_0,\Big) && (93)\\ &= P\Big(X_{T_1} \in A,\ T_1 \in (a,b)|X_0,\Big) && (94)\\ &= P\Big(X_{T_1} \in A|X_0\Big)P\Big(T_1|X_0\Big) && (95)\\ &= P\Big(\bar{X}_1 \in A|\bar{X}_0\Big)P\Big(T_1|\bar{X}_0\Big) && (96) \end{aligned}$$

where in (93) we have used the Markov property as well as the time-homogeneous property. (94) follows from the definition that $T_0 = 0$. Step (95) is again a consequence of the Markov property, and it says that the transition time given the current state is independent of future states.

Thus, if we have a time-homogeneous MPP, all we need to know are the above two quantities in order to specify the process completely. Next, we discuss one very important property of the random variable $T_n$, and state and prove a very important theorem.

**Lemma 6.1.** *Let $T_n$ denote the transition time in a time-homogeneous MPP. Then*

$$P(T_n > t + \delta|T_n > t) = P(T_n > \delta) \tag{97}$$

Property (97) is known as memorylessness, or the memoryless property.

*Proof.*

$$P(T_n > t+\delta | T_n > t) = P(T_n > \delta) = P(T_0 > t+\delta | T_0 > t) = P(T_0 > \delta) \tag{98}$$
$$= P(X_{t'} = x \ \forall t^{'} \in [0, t+\delta] \ | X_s = x \ \forall s \in [o,t]) \tag{99}$$
$$= P(X_{t'} = x \ \forall t^{'} \in [0, t] \cup (t, t+\delta] \ | X_s = x \ \forall s \in [o,t]) \tag{100}$$
$$= P(X_{t'} = x \ \forall t^{'} \in (t, t+\delta] \ | X_s = x \ \forall s \in [o,t]) \tag{101}$$
$$= P(X_{t'} = x \ \forall t^{'} \in (t, t+\delta] \ | X_t = x) \tag{102}$$
$$= P(X_{t'} = x \ \forall t^{'} \in (0, \delta] \ | X_0 = x) \tag{103}$$
$$= P(T_0 > \delta | T_0 > 0) \tag{104}$$
$$= P(T_0 > \delta) \tag{105}$$
$$= P(T_n > \delta) \tag{106}$$

where in steps (98), (103) and (106) we have use the time homogeneity and in step (102) we have used the Markov property. □

The following theorem specifies the CDF of a continuous random variable with the memory-less property.

**Theorem 6.2.** *Let $T$ be a continuous random variable with the memory-less property. Then $T \sim exp(\lambda)$ for some positive constant $\lambda$.*

*Proof.* Given that $T$ is a continuous random variable, i.e., the CDF of $T$, say $F$, is continuously differentiable, and has a pdf .(say $f$) Now, from the memory-less property, we have

$$P(T > t+\delta | T > t) = P(T > \delta) \tag{107}$$
$$\implies \frac{P(T > t+\delta, T > t)}{P(T > t)} = P(T > \delta) \tag{108}$$
$$\implies \frac{P(T > t+\delta)}{P(T > t)} = P(T > \delta) \tag{109}$$
$$\implies 1 - F(t+\delta) = (1 - F(t))(1 - F(\delta)) \tag{110}$$
$$\implies 1 - F(t+\delta) = 1 - [F(t) + F(\delta) - F(t)F(\delta)] \tag{111}$$
$$\implies F(t+\delta) - F(t) = F(\delta)(1 - F(t)) \tag{112}$$
$$\implies \lim_{\delta \to 0} \frac{F(t+\delta) - F(t)}{\delta} = \lim_{\delta \to 0} \frac{F(\delta)}{\delta}(1 - F(t)) \tag{113}$$
$$\implies \frac{\frac{d}{dt}F(t)}{1 - F(t)} = -f(0) \tag{114}$$
$$\implies 1 - F(t) = ce^{-f(0)t} \tag{115}$$
$$\implies F(t) = 1 - c^{-f(0)t} \tag{116}$$

Putting $t = 0$ yields $F(0) = 0 = 1 - c \implies c = 1$. Hence, setting $f(0) = \lambda$, we get

$$F(t) = 1 - e^{-\lambda t}$$

which is the CDF of $exp(\lambda)$ distribution.

□

## 6.3 Simulation of MPP

### 6.3.1 Discrete-Time Markov Chain

We first look at the definition of discrete-time Markov Chain once more. This holds a special significance, since for any time-homogeneous MPP, we can define its embedded Markov chain, and that, along with the transition times will determine the MPP.

**Definition 6.7.** *A stochastic process* $\{ X_t\colon t \geq 0\}$ *is called a discrete-time Markov Chain if for all times* $n \geq 0$ *and all states* $i_0, i_1, \cdots, i_k, \cdots, i_{n-1}, i_n$, $(i_k \in S)$, *we have*

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, ..., X_0 = i_0) = P(X_{n+1} = j | X_n = i) = p_{ij}. \tag{117}$$

Here $p_{ij}$ gives the probability that the chain, whenever in state $i$, moves next (one unit of time later) into state $j$, and is referred to as transition probability. The square matrix $\mathbb{P} = (p_{ij})$, where $i, j \in S$ is called the Transition Probability Matrix **(TPM)** . Since, when leaving the state $i$ the chain must move to one of the state $j \in S$, each row of $\mathbb{P}$ sums to one (e.g., forms a probability distribution): For each $i$

$$\sum_{j \in S} p_{ij} = 1.$$

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

1. Let $P(X_0 = i) = q_i$. Now, we can define a CDF for the random variable $X_0$ by setting

$$F_{X_0}(x) = \sum_{i=1}^{\lfloor x \rfloor} q_i \tag{118}$$

which defines a step function which is right-continuous. Now we can use the uniform random number generator we made before and generate a $U[0, 1]$ random number $u$ and find $X_0$ by using the percentile function,

$$X_0 = F^{\leftarrow}_{X_0}(u) = i \;\text{ if } u \in \Big( \sum_{j=1}^{i-1} q_j, \sum_{j=1}^{i} q_j \Big] \tag{119}$$

2. By definition, $P(X_1 = j|X_0 = x_0) = p_{x_0 j}$ and so on. Now, generating independent $U[0,1]$ random numbers $u_1, u_2, ..., u_n$ and using the TPM, we can easily simulate

$$X_1 = F^{\leftarrow}_{X_1|X_0}(u_1) = j \;\; \text{if} \;\; u_1 \in \Big( \sum_{i=1}^{j-1} p_{x_0 i}, \sum_{i=1}^{j} p_{x_0 i} \Big] \tag{120}$$

$$\vdots \tag{121}$$

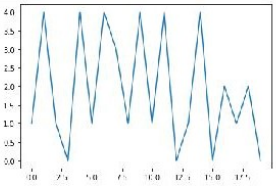
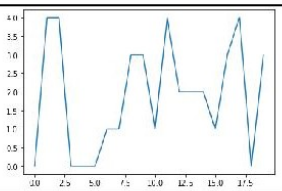
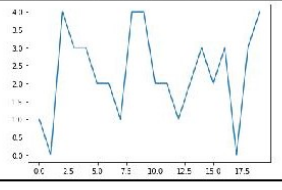
$$X_n = F^{\leftarrow}_{X_n|X_{n-1}}(u_n) = j \;\; \text{if} \;\; u_n \in \Big( \sum_{i=1}^{j-1} p_{x_{n-1} i}, \sum_{i=1}^{j} p_{x_{n-1} i} \Big] \tag{122}$$

3. In order to check if the sequence generated is indeed a Markov chain with the prescribed transition matrix, we can verify the ratios

$$\frac{\text{no. of transitions from i to j}}{\text{no. of occurrences of i}} \rightarrow p_{ij} \;\; as \;\; n \rightarrow \infty \tag{123}$$

The above algorithm is implemented, and the results are tabulated in $Table - 4$.

Figure 4: Table 4

| Initial Dist. | Transition Matrix | | | | | Plot of First 20 Timesteps | Empirical Transition Matrix (1000 timesteps) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.10 | 0.20 | 0.40 | 0.10 | 0.10 | 0.20 |  | 0.19 | 0.41 | 0.09 | 0.09 | 0.22 |
| 0.20 | 0.20 | 0.20 | 0.10 | 0.30 | 0.20 | | 0.20 | 0.21 | 0.11 | 0.29 | 0.19 |
| 0.20 | 0.10 | 0.20 | 0.30 | 0.20 | 0.20 | | 0.09 | 0.20 | 0.31 | 0.21 | 0.19 |
| 0.20 | 0.10 | 0.30 | 0.10 | 0.30 | 0.20 | | 0.12 | 0.29 | 0.08 | 0.31 | 0.20 |
| 0.30 | 0.30 | 0.20 | 0.10 | 0.10 | 0.30 | | 0.29 | 0.20 | 0.10 | 0.10 | 0.31 |
| 0.05 | 0.20 | 0.40 | 0.10 | 0.10 | 0.20 |  | 0.19 | 0.38 | 0.12 | 0.10 | 0.21 |
| 0.25 | 0.20 | 0.20 | 0.10 | 0.30 | 0.20 | | 0.21 | 0.19 | 0.11 | 0.29 | 0.20 |
| 0.25 | 0.10 | 0.20 | 0.30 | 0.20 | 0.20 | | 0.09 | 0.21 | 0.30 | 0.19 | 0.21 |
| 0.15 | 0.10 | 0.30 | 0.10 | 0.30 | 0.20 | | 0.09 | 0.31 | 0.11 | 0.29 | 0.20 |
| 0.30 | 0.30 | 0.20 | 0.10 | 0.10 | 0.30 | | 0.30 | 0.21 | 0.10 | 0.10 | 0.29 |
| 0.05 | 0.19 | 0.38 | 0.12 | 0.10 | 0.21 |  | 0.20 | 0.37 | 0.11 | 0.11 | 0.21 |
| 0.25 | 0.21 | 0.19 | 0.11 | 0.29 | 0.20 | | 0.22 | 0.19 | 0.11 | 0.28 | 0.20 |
| 0.25 | 0.09 | 0.21 | 0.30 | 0.19 | 0.21 | | 0.07 | 0.21 | 0.32 | 0.20 | 0.20 |
| 0.15 | 0.09 | 0.31 | 0.11 | 0.29 | 0.20 | | 0.08 | 0.32 | 0.11 | 0.30 | 0.19 |
| 0.30 | 0.30 | 0.21 | 0.10 | 0.10 | 0.29 | | 0.30 | 0.21 | 0.10 | 0.11 | 0.30 |

As we can see, the elements of the so-called "Empirical Transition Matrix" obtained by applying

the formula (123) are quite close to the elements of the TPM for 1000 time-steps, verifying that our Markov Chain indeed follows the Probability laws prescribed by the Transition Matrix.

### 6.3.2 Transition Times and Simulation of MPP

Since we have already proved that the transition times for a time-homogeneous MPP follows exponential distribution, and since we can obtain random numbers from any distribution now, it is straightforward to obtain the transition times.

After that, all we have to do is just consider the transitions at these times, and we will get a realization of an MPP.

## 7 Conclusions

This project was instrumental in building our understanding of various stochastic processes. Concepts such as Random Variables, Methods of obtaining various random numbers, Brownian Motion and Markov Pure-Jump Processes were discussed in some detail, with emphasis on some of their properties. A sizable part of this project was related to computation, which not only helped me in honing my programming skills in python, but also helped me in gaining better insight into the mathematical properties of these processes. I also learnt the use of LATEX, which I believe would be of great use in the future.

## 8 References


1. Principles of Mathematical Analysis by W. Rudin
2. Statistical Inference by Casella and Berger
3. Introduction to Stochastic Processes by Hoel, Port and Stone
4. Computational Techniques in Mathematical Finance (Project Report) by Visakh Narayanan
5. Theory and Estimation of Drift Dynamics of Asset Price Process (MS Thesis) by Aakash Verma
6. www.statweb.stanford.edu
7. www.onlinelibrary.wiley.com
8. www.wikipedia.org

# 9 Acknowledgements

I am extremely grateful to Prof. Anindya Goswami, who gave us the opportunity to undertake this project. His constant guidance and encouragement during the course of this semester showed me the way and urged me to learn more about Stochastic Processes. I would also like to extend my thanks to Mr. Akhil Babu, who also undertook this project with me, and have helped me on multiple occasions. Lastly, I would like to extend my gratitude to my family and friends who continued to support me throughout teh course of this project.