Spring 2014

Approximating the wiener process

Branden James Wooten
James Madison University

Approximating the Wiener Process

A Project Presented to

The Faculty of the Undergraduate

College of Science and Mathematics along with

The College of Integrated Science and Engineering

James Madison University

In Partial Fullfillment of the Requirements

for the Degree of Bachlors of Science

Branden James Wooten

May 2014

Accepted by the Department of Mathematics and Computer Science, James Madison University, in partial fullfillment of the requirements for the Degree of Bachlors of Science.

FACULTY COMMITTEE:

____

Faculty Project Advisor: Stephen Lucas, Ph.D.,
Associate Professor, Mathematics

____

Reader: Ramon Mata-Toledo, Ph.D.,
Professor, Computer Science

____

Reader: Xunhua Wang, Ph.D.,
Associate Professor, Computer Science

HONORS PROGRAM APPROVAL:

____

Barry Falk, Ph.D.,
Director, Honors Program

For my mom, dad, and sister

Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank Dr. Lucas for his support, knowledge, and for the opportunity to work on this project. Dr. Lucas's insight and expertise helped to make this project possible. The time and information he devoted to my learning and the progression of this thesis cannot go unacknowledged.

I would like to express appreciation for my readers, Dr. Mata-Toledo and Dr. Wang for their influence, time, and assistance perfecting my thesis.

Additional thanks to my roommates, friends, and coworkers for their encouragement and support.

I appreciate the opportunity provided by the James Madison University Honors Program to complete this thesis.

Abstract

Wiener Processes, w_t, are random processes with mean zero, variance t. Wiener processes are difficult to work with as any realization is continuous and nowhere differentiable. Through the use of Karhunen-Loève expressions one can approximate the Wiener Process and run simulations to determine how long it takes before the truncated estimation is no longer a true Wiener Process. This project shows the necessary statistical tests needed to determine this information, along with many simulation examples and results. Furthermore, with the results of the approximated Wiener Process, one can solve stochastic differential equations that would ordinarily be extremely difficult to solve.

1 Introduction

Through the investigation of Karhunen-Loève (K-L) Expansions and Wiener Processes, the following question is proposed: Given a truncated K-L expansion, how many realizations or experiments are needed before it is clear that the process is not a Wiener process? A Wiener Process is a random process with mean zero, variance t.

The Wiener Process, an example represented by Figure 1 has many practical applications. "The process occurs frequently in pure and applied mathematics, economics, and physics" [1]. One specific example where Wiener Processes are used is in the financial industry. The process is also used wherever random noise is present. Wiener processes are used to represent the integral of Gaussian white noise processes and thus is useful in modeling noise in electronics, instrument errors in filtering theory, and unknown forces in control theory [1].

The main disadvantage of the Wiener Process is that it is a nowhere differentiable function which makes working with the function and calculus much more difficult.

1.1 Stochastic Differential Equations

The word "stochastic" is synonymous with the word "random." Stochastic differential equations are a type of differential equation, including random terms, which results in a solution that is itself a stochastic process. "Stochastic processes can be defined as a family of random variables from some probability space into a state space." [2]

Stochastic differential equations have many uses in the modern world, for example, fluctuating stock prices. In terms of this project, stochastic differential equations are related to Brownian motion or the Wiener Process. These types of equations are extremely hard to calculate accurately as $\frac{dw}{dt}$ does not exist. Equation 1 shows one realization of the analytical solution of a stochastic differential equation. One can compare the results from the

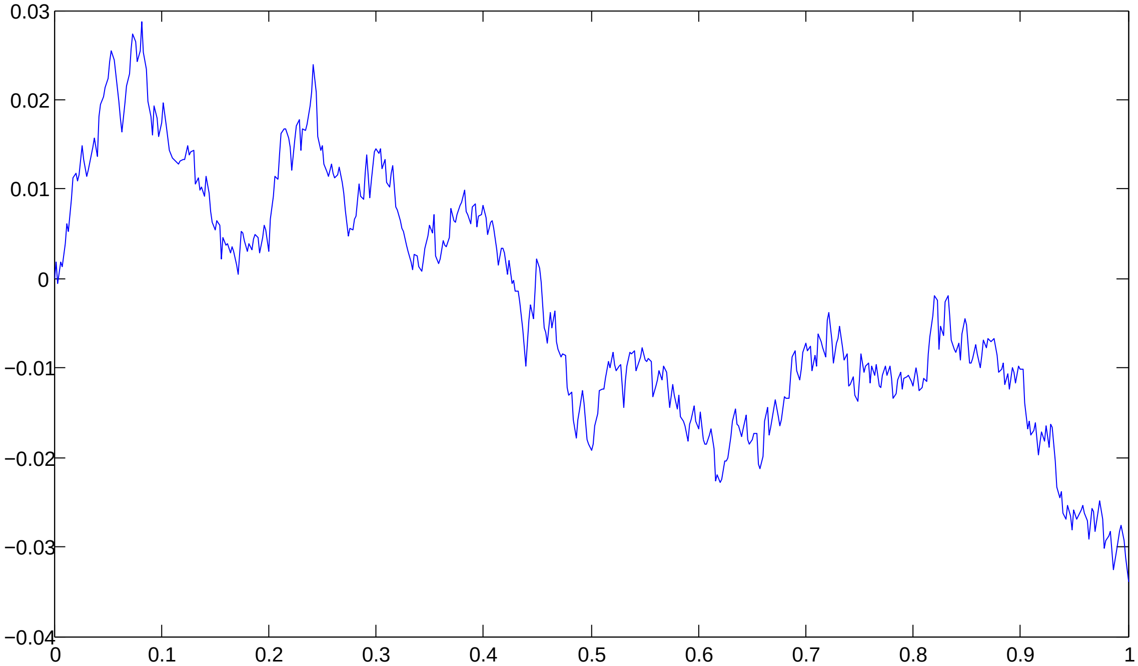

Figure 1: Wiener Process created by code 5.4

approximated version to the solution of

$$x_t = \mu x + \sigma x \frac{dw}{dt}, \tag{1}$$

where x_t is the derivative.

1.2 Wiener Process

The Wiener Process is a "continuous-time stochastic process named in honor of Norbert Wiener. It is often also called Brownian motion after Robert Brown" [1]. The Wiener process is used to represent the integral of the Gaussian white noise process in electronics engineering. At each time, the process $w(t)$ will be a Gaussian distribution with expected value $E[w(t)] = 0$ and variance $var[w(t)] = t$. Each time we examine the process it will follow a Gaussian Distribution.

1.3 Karhunen-Loève

The Karhunen-Loève or K-L expansion was originally discovered by Kosambi. The K-L expansion is a representation of stochastic processes similar to the expansion in Fourier series known as the Proper Orthogonal Decomposition. [4]

The Karhunen-Loève has a major disadvantage. The K-L expansion is a no-where differentiable function. This means that it is hard to use anywhere analytically since it cannot be differentiated at any given point.

Karhunen-Loève series expansion is based on the eigen-decomposition of the covariance function" [3]. The Karhunen-Loève Expansion for the Wiener Process is

$$W_t = \sqrt{2}\sum_{k=1}^{\infty} \frac{z_k \cdot \sin\left(k - \frac{1}{2}\pi t\right)}{k - \frac{1}{2}\pi} \tag{2}$$

where $0 \leq t \leq 1$ and each z_k is from a Gaussian Sample with $\mu = 0, \sigma^2 = 1$.

The purpose is to approximate a Wiener Process by truncating the infinite sum in equation 2. This replaces $w(t)$ with any analytic function.

1.3.1 Unfair Coin Example

Any standard statistics book will refer to the probability of a fair coin as 50/50. This shows the probability that the coin will land on heads 50% of the time. As a metaphoric example of what is planned with the Wiener Processes, one can examine flipping an unfair coin. Take for example, a coin that lands 45% of the time on heads and 55% of the time on tails. In this case, it is possible to find how many tosses before one discovers that the coin is in fact an unfair coin. In the beginning, 10 flips may not be enough to determine the coin's fairness, neither would 25 or 50. It takes about 1000 trials before it is clear that the coin is indeed an unfair coin.

Equation 3 shows the expansion of the Chi Square Test for the simple unfair coin example:

$$\begin{aligned}
\chi^2 = \sum_{i=1}^{n} \frac{(observed - expected)^2}{expected} &= \frac{(\frac{n}{2} - \frac{9n}{20})^2}{\frac{n}{2}} + \frac{(\frac{n}{2} - \frac{11n}{2})^2}{\frac{n}{2}} \\
&= \frac{2}{n}\left[\left(\frac{n}{20}\right)^2 + \left(\frac{-n}{20}\right)^2\right] \\
&= \frac{2}{n} \cdot \frac{n^2}{200} \\
&= \frac{n}{100}
\end{aligned} \tag{3}$$

This small example models the exact process that will take place with the K-L expansion. At first it is hard to tell whether the coin is actually fair or not, it all depends on the number of trials that are completed. The goal here is to use the Chi Square tests and figures below to determine how many trials it takes to determine that the coin that is being flipped is actually an unfair coin. For example, if the χ^2 value is larger than a given critical value, then reject the hypothesis that the coin is fair. It can be stated with 95% confidence that the coin is unfair based off of the Chi Square significance value of 3.84 using one degree of freedom.

The experiment is to truncate the Wiener Process, as represented by Equation 4, and through numerical experimentation (simulations), determine how many realizations before one can tell the truncated Wiener Process is not a true Wiener Process. Once this step is complete, this approximation can further be used to approximate Stochastic Differential Equations. Results from the approximated experiment can be compared (using the approximated Wiener Process values of the stochastic DEQs) to known results. The goal is to see if the results computed using the approximated value are close enough to the real values. If so, an easier way to solve stochastic DEQs and Wiener Processes has been found.

$$W_t = \sqrt{2} \sum_{k=1}^{n} \frac{z_k \cdot \sin\left(k - \frac{1}{2}\pi t\right)}{k - \frac{1}{2}\pi} \tag{4}$$

Figures 2 through 5 are examples of coin flips with n trials each. These plots indicate

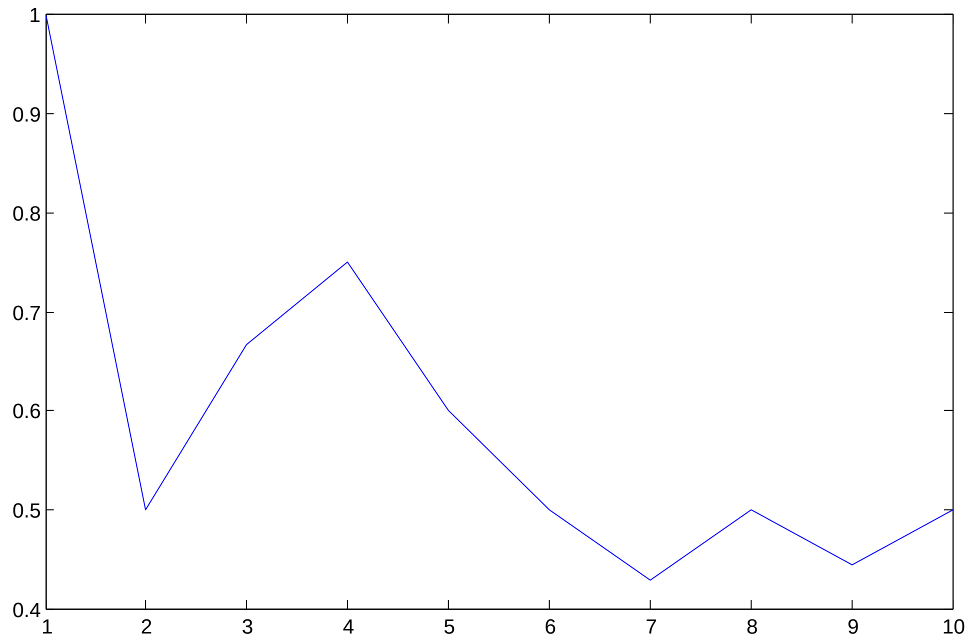

Figure 2: Cumulative Percentage of Heads - Fair Coin Toss, 10 Trials

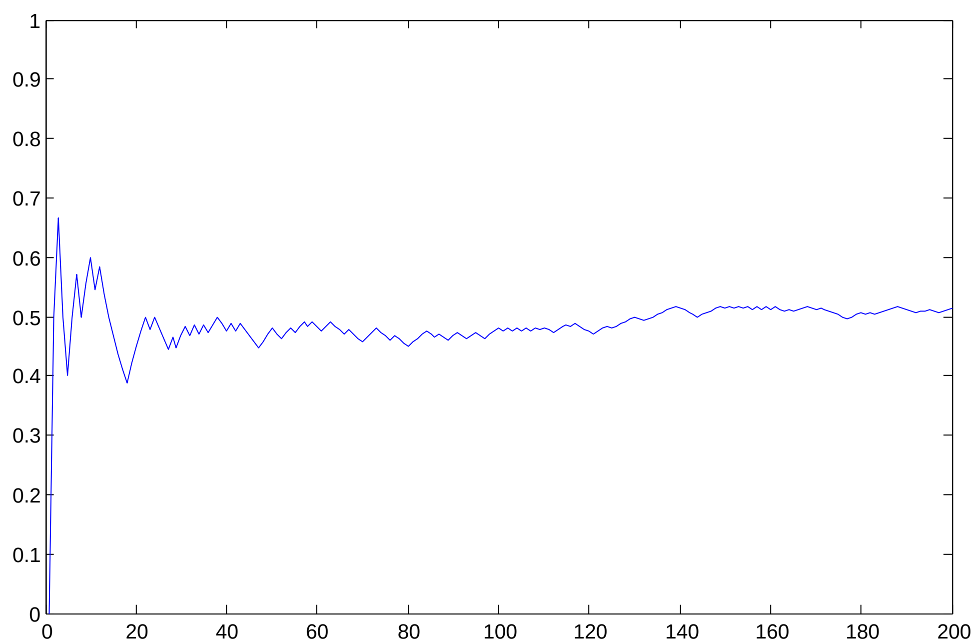

Figure 3: Cumulative Percentage of Heads - Fair Coin Toss, 200 Trials

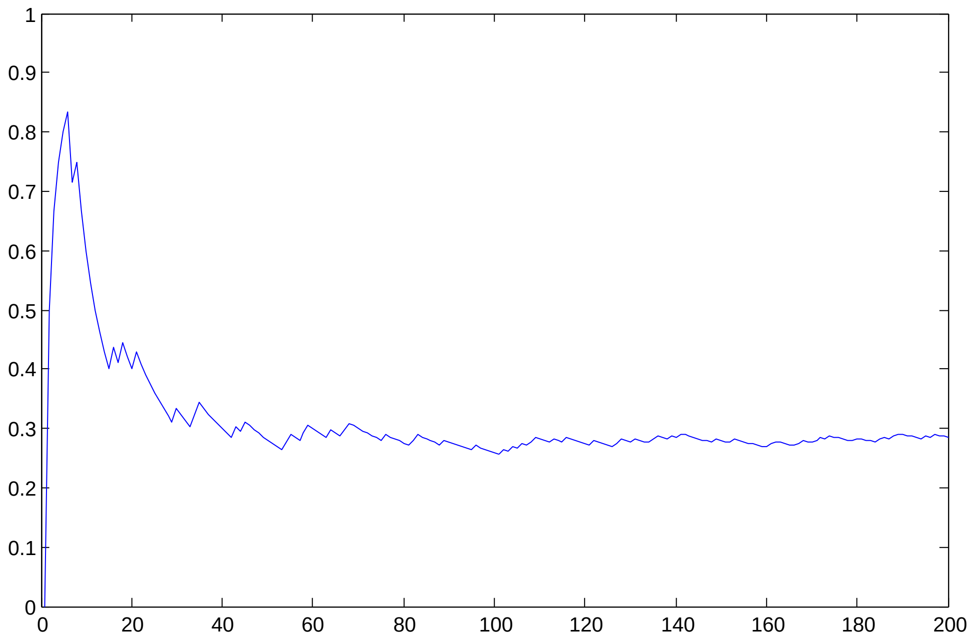

Figure 4: Cumulative Percentage of Heads - Un-Fair Coin Toss, 200 Trials, 30% Chance of Heads

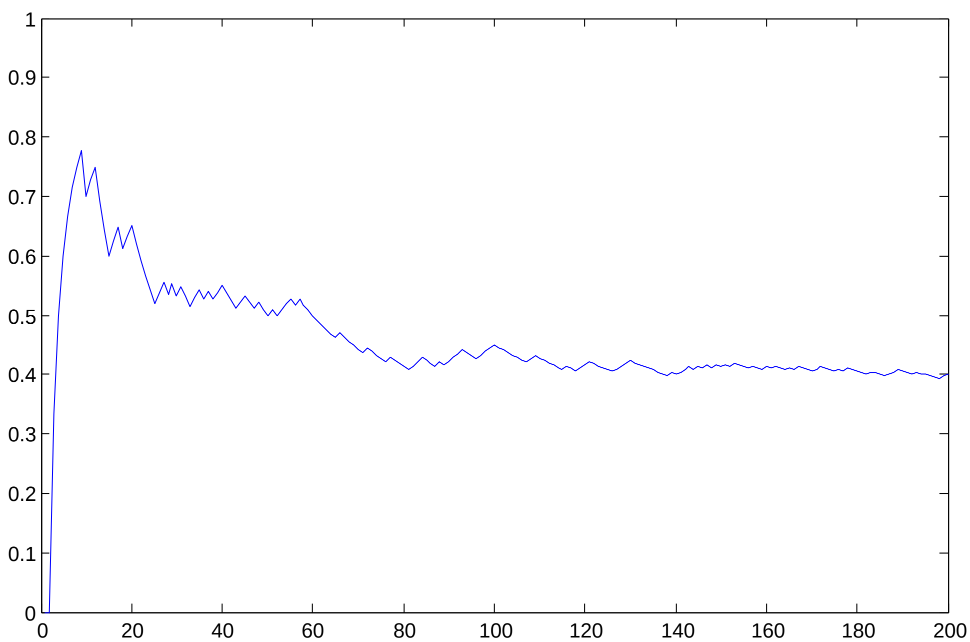

Figure 5: Cumulative Percentage of Heads - Un-Fair Coin Toss, 200 Trials, 40% Chance of Heads

how a coin is flipped and the importance of the number of trials.

Fair coins should have a distribution that is approximately normal, as shown by Figure 6. This can be compared to an unfair coin, where 70% of the time the coin lands on heads, such as Figure 7. The fair coin, as most likely suspected, is a symmetric curve centered around 50. Figure 7 shows a roughly symmetric curve centered around 70.

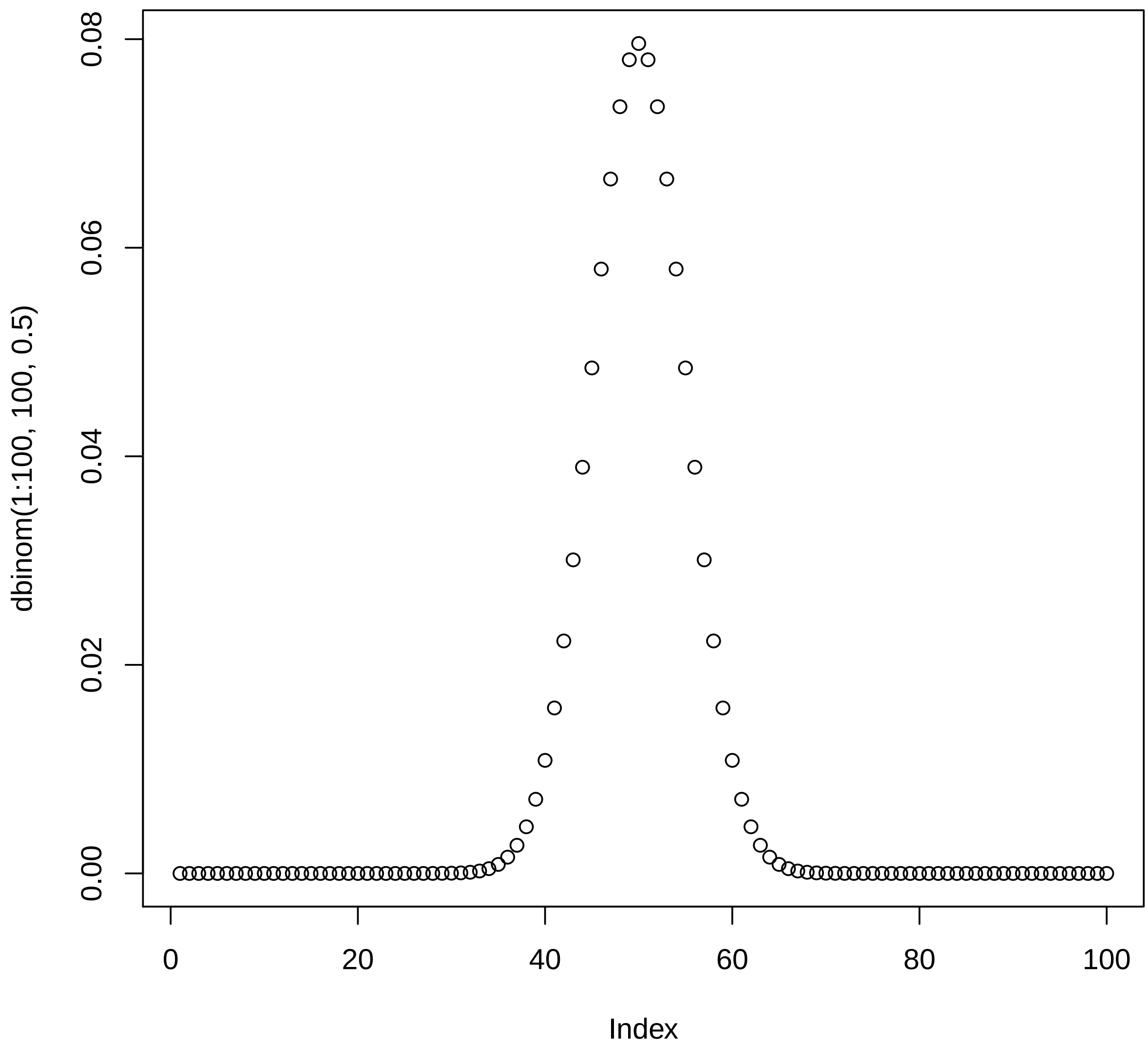

Figure 6: Fair Coin Distribution - Normal Curve

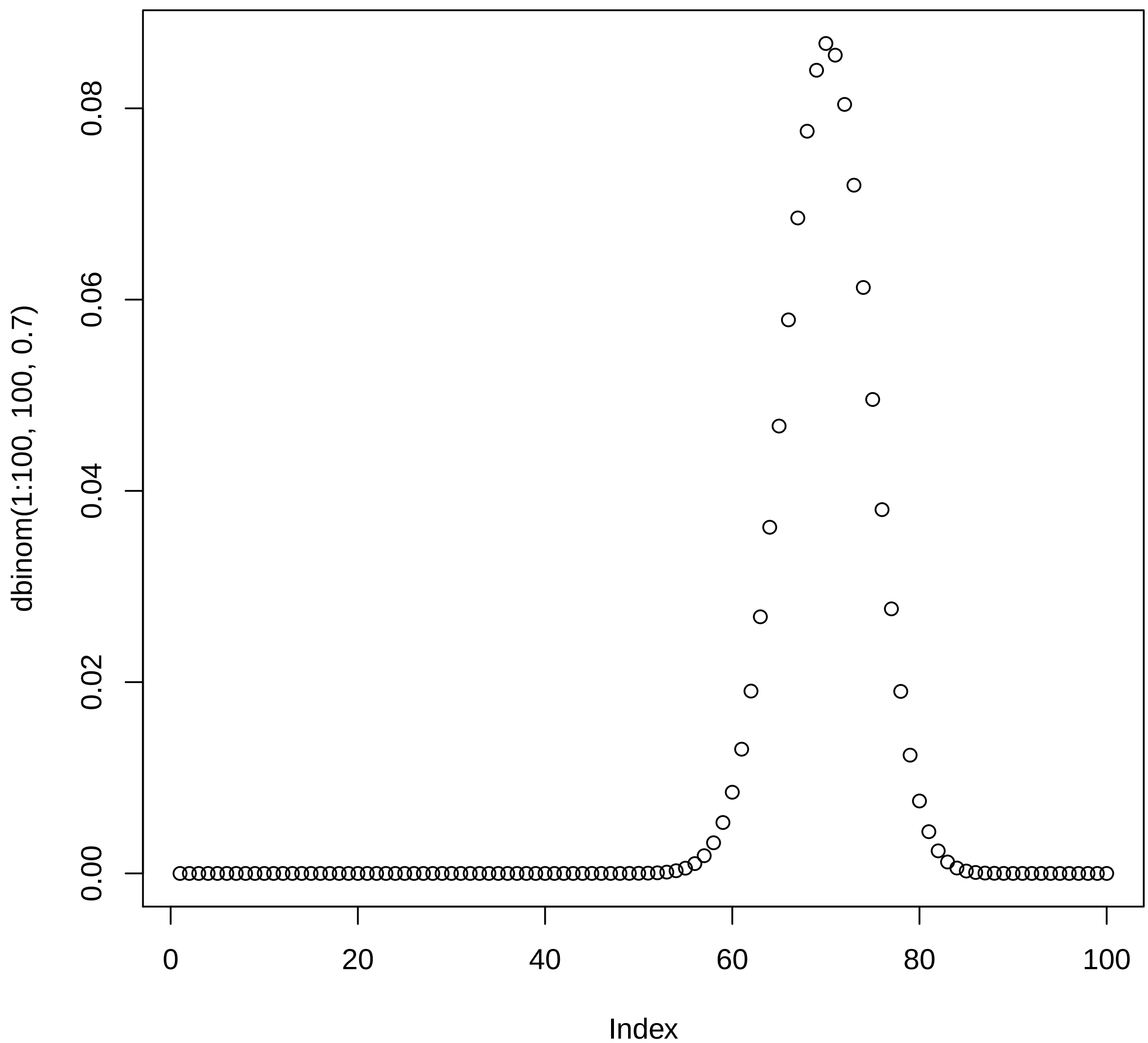

Figure 7: Un-Fair Coin Distribution

2 Statistical Tests

Throughout this experiment, different statistical tests were considered to determine which test would be the best for the analysis of the data being looked at. The goal was to determine how far one has to go before being able to tell that the truncated Wiener process is not normal. The statistical tests below can all be used to check whether a data sample comes from a normally distributed population. The Chi Square test cannot be used as it is only really good for discrete data.

2.1 Kolmogorov-Smirnov

The Kolmogorov-Smirnov test was first examined as a possibility for this project. The Kolmogorov-Smirnov (K-S) test is a test for the equality of continuous, one-dimensional probability distributions. It is a natural test for a uniform distribution with limits 0 and 1 [6]. The K-S test is just one example to measure goodness of fit for any continuous distribution. The K-S test only needs the data to be asymptotically normally distributed to work correctly. The two sample K-S test is known to be one of the best ways to compare two samples as it is sensitive to differences in both location and shape of cumulative distribution functions.

2.2 Anderson-Darling

The Anderson–Darling test is a statistical test based on the K-S test to determine if a given sample of data is drawn from a given probability distribution. The Anderson Darling test improves the Kolmogorov-Smirnov test in the tails of the distribution. The basic Anderson-Darling statistic is

$$A^2 = -n - S, \tag{5}$$

where

$$S = \sum_{i=1}^{n} \frac{2i-1}{n} [\ln(\phi(Y_i)) + \ln(1 - \phi(Y_{n+1-i}))] \tag{6}$$

The Anderson-Darling test can be used given any of the 5 cases below. Case 0 applies to this project; both the mean and variance are known: $\mu = 0, \sigma^2 = 1$.

Case 0:	$F(x)$ continuous, completely specified.
Case 1:	$F(x)$ is the normal distribution, σ^2 and μ estimated by $\hat{x}$.
Case 2:	$F(x)$ is the normal distribution, μ known and σ^2 estimated by $\sum_i (x_i - \mu)^2/n$
Case 3:	$F(x)$ is the normal distribution, both μ and σ^2 are unknown estimated by $\hat{x}$ and $s^2 = \sum_i (x_i - x)^2/(n-1)$
Case 4:	$F(x) = 1 - exp(-\theta x)$

Table 1: Anderson-Darling Cases

"If A^2 exceeds a given critical value, then the hypothesis of normality is rejected with some significance level. The critical values are given in Table 2 [5] below (valid for $n \geq 5$)" [6].

Significance	Case 0	Case 1	Case 2	Case 3
15%	1.610			0.576
10%	1.933	0.908	1.760	0.656
5%	2.492	1.105	2.323	0.787
2.5%	3.070	1.304	2.904	0.918
1%	3.857	1.573	3.690	1.092

Table 2: Anderson-Darling Significance Table

The test statistic significance levels can be calculated using

$$1.273^{-\pi^2 z/2} \tag{7}$$

To determine a significance level for use with a given statistic,

$$score = -\frac{3}{\pi^2} \ln\left(\frac{x}{1.273}\right) \tag{8}$$

where x is the percentage level.

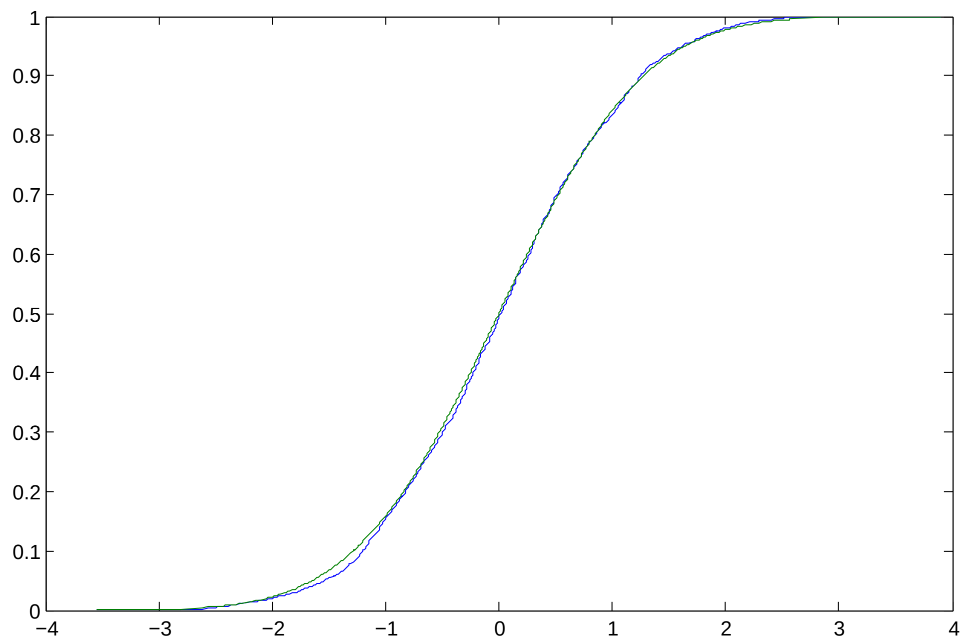

Figure 8: Anderson Darling Test to Verify Normality

For a given value z of the modified test statistic, the significance level α in the upper tail is given approximately by $\alpha(z)$.

Figure 8, generated using code 5.2, is the graphical representation of a randomly created data set with a normal distribution, $n = 2000$. This figure uses the Anderson Darling test to determine if the data set is actually normal. The green curve represents the error function for any cumulative distribution function:

$$F_x(x;\mu,\sigma) = \frac{1}{2}\left[1 + erf\left(\frac{\ln x - \mu}{\sigma\sqrt{2}}\right)\right] \tag{9}$$

Since both the green and blue curves are nearly identical, it can be said that this data sample is normally distributed as the error is so small, that the data must be considered normal.

Figure 9 is an example with an Anderson Darling test where $n = 50$. Clearly 50 iterations is not enough to tell if the distribution fits the normal curve or not. It's not until about

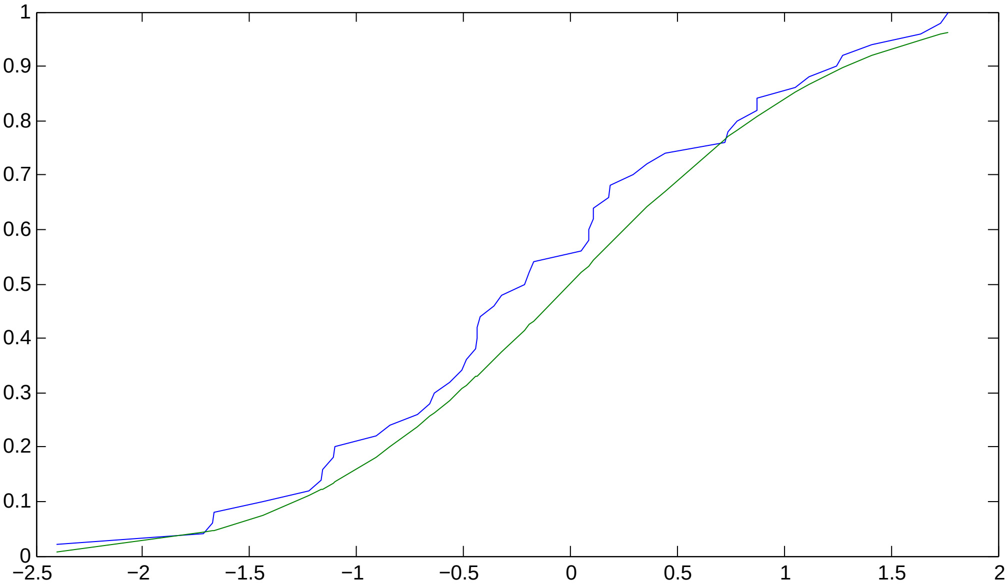

Figure 9: Anderson Darling Test to Verify Normality - Not Normal

$n = 1000$ that one can be fairly confident that the sample fits the curve. Figure 10 runs the Anderson Darling test for normality on a sample with $n = 1000$. Here it is clear the sample is roughly normal.

2.3 Shapiro-Wilk

The Shapiro-Wilk test is very similar to the Anderson-Darling however it takes more care with the end points. However, it has been known to not work well when numbers are repeated frequently. Anderson-Darling was chosen over the Shapiro-Wilk test because it was more applicable to the goal with concern to the tails of the distribution as well as Shapiro-Wilk only tests normality.

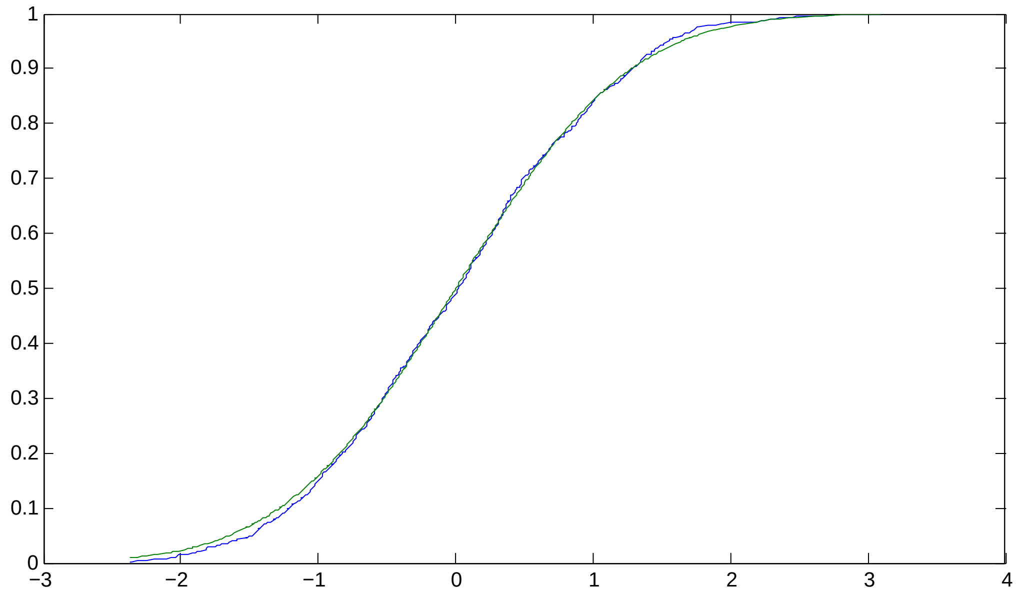

Figure 10: Anderson Darling Test to Verify Normality

2.4 Why Anderson-Darling

The Anderson-Darling Test was chosen in place of the Chi-Square test, the Shapiro-Wilk test, and the K-S test because the statistic is more sensitive to the tails of a distribution, rather than the center. The goal of the Anderson-Darling test is to determine when one can no longer say that the truncated K-L expansion is truly Gaussian, meaning that the truncated Wiener Process expression is no longer working. The Anderson-Darling test will provide the stopping point when a truncation of the Wiener Process is no longer valid.

3 Results

3.1 Unfair Coin Example

The unfair coin example is a trivial model for what was expected when the K-L expansion is run with similar simulations. A simulation of flipping a coin was run with various numbers of realizations. When looking at a probability of 50/50 for a coin, Figure 11 shows the cumulative chi square test with 50000 realizations. By calculating the Chi Square value for each group $1 - n$, $n = 1$ to 5000 it can be seen how the value varies as the number of realizations increases and how the value remains below the critical value of 3.84 given from the 5% significance value of the Chi Square test with 1 degree of freedom. Comparing this to an unfair coin with probability of 70% heads and 30% tails a much different plot emerges. Figures 12 - 15 show this result. Unlike the fair coin, Figures 12 - 15 show how quickly the Chi Square test can determine that the coin is unfair when the simulation is run with 70% heads 30% tails. It is obvious almost immediately.

Table 3 shows the progression of how easy it is to determine that a coin is unfair based on the number of realizations. When the coin is closer to fair it takes longer to determine that it is truly an unfair coin. As the unfairness level increases, it becomes easier and easier to determine that the coin is truly an unfair coin. For example when a coin is 51% fair, Figures 17 through 19, it takes over 4000 tries to determine that the coin is truly an unfair coin. But when the unfairness level is 70% it quickly falls into the unfair coin category, taking roughly 10 trials to determine.

For the example when a coin is flipped with 45% probability of heads, equation 3 shows the theoretical value for the Chi Square test at a given n value. Figure 16 shows the theoretical diagonal line and the critical value. This theoretical value can be calculated for any given probability and Chi Square test.

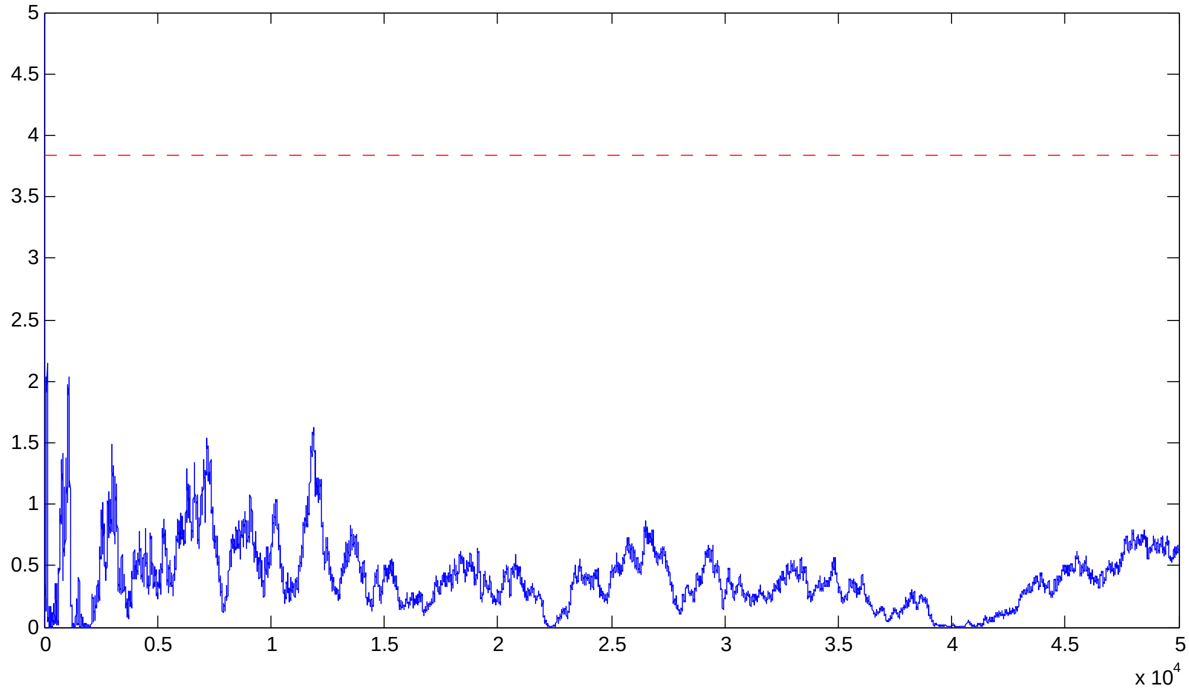

Figure 11: Chi Square Test Results - Fair Coin, 50000 Trials

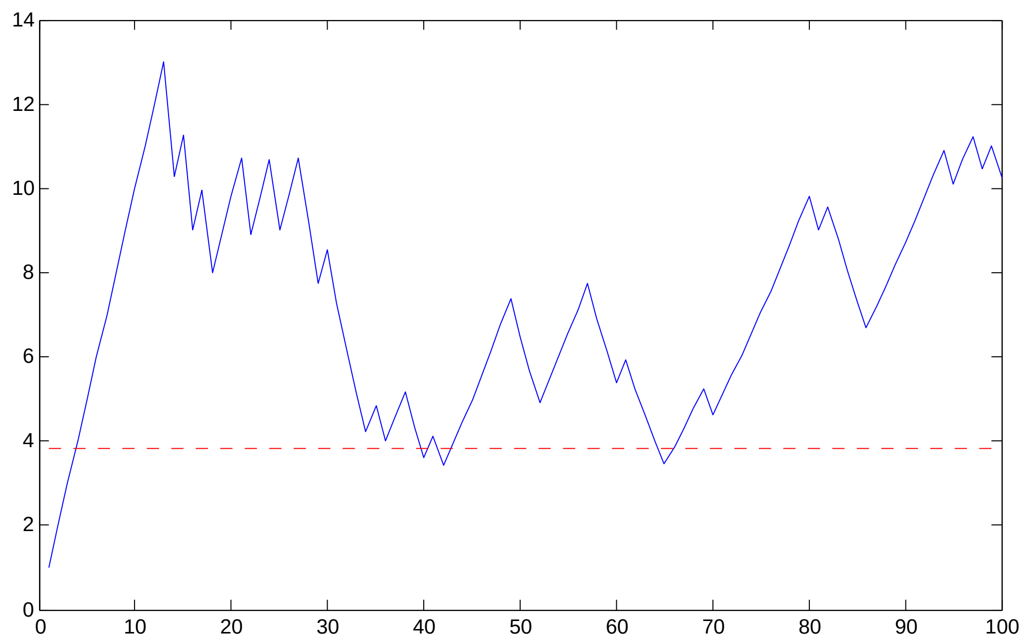

Figure 12: Chi Square Test Results - Unfair Coin - 70%, 100 Trials

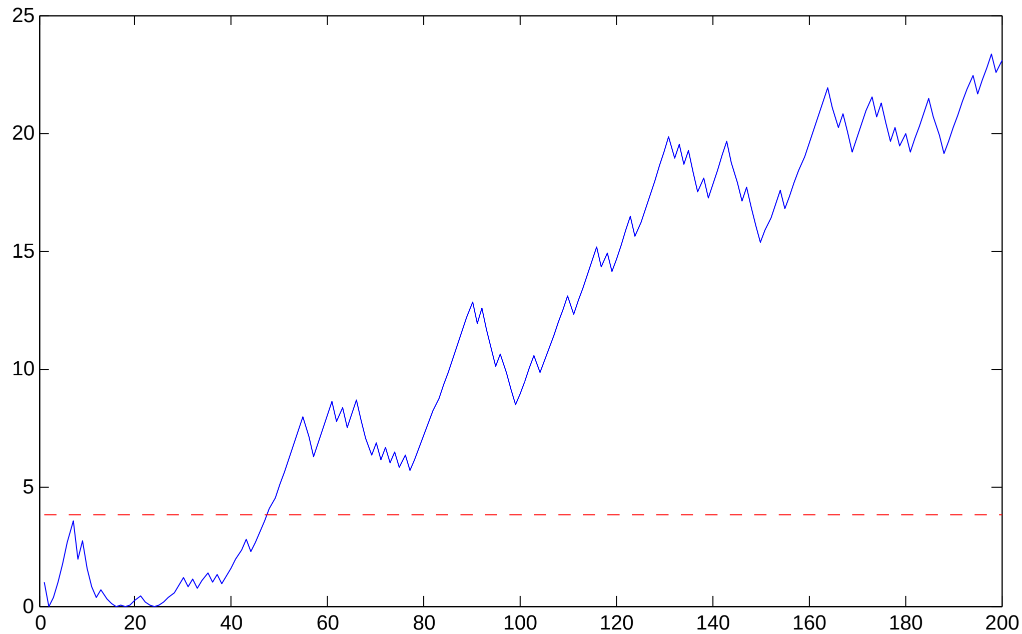

Figure 13: Chi Square Test Results - Unfair Coin - 70%, 200 Trials

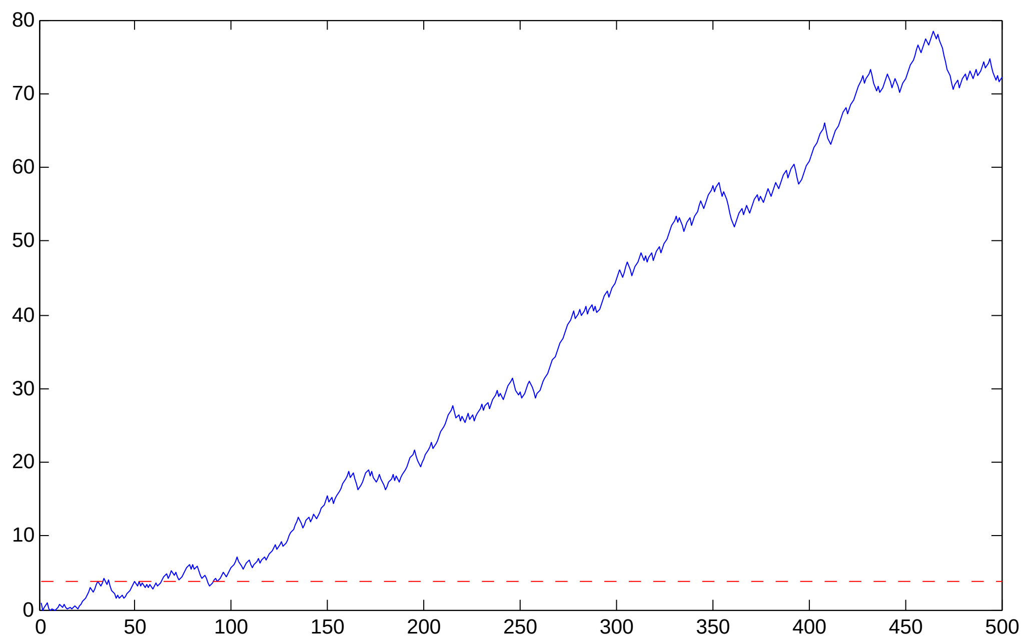

Figure 14: Chi Square Test Results - Unfair Coin - 70%, 500 Trials

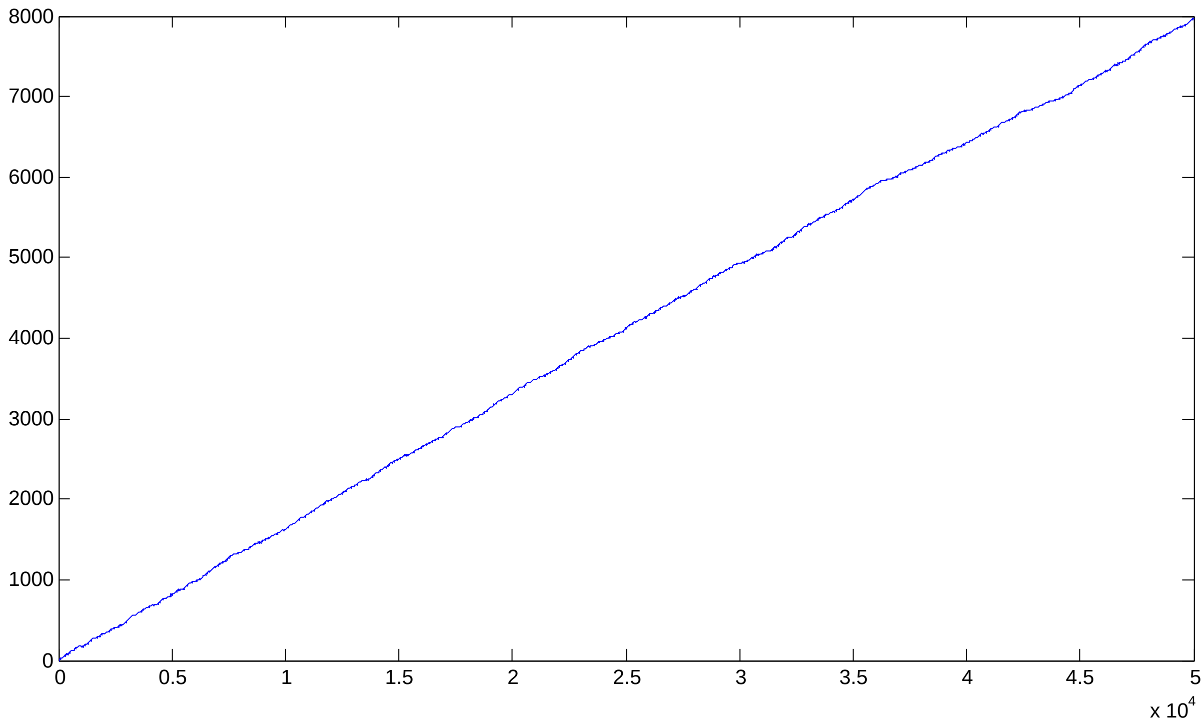

Figure 15: Chi Square Test Results - Unfair Coin - 70%, 50000 Trials

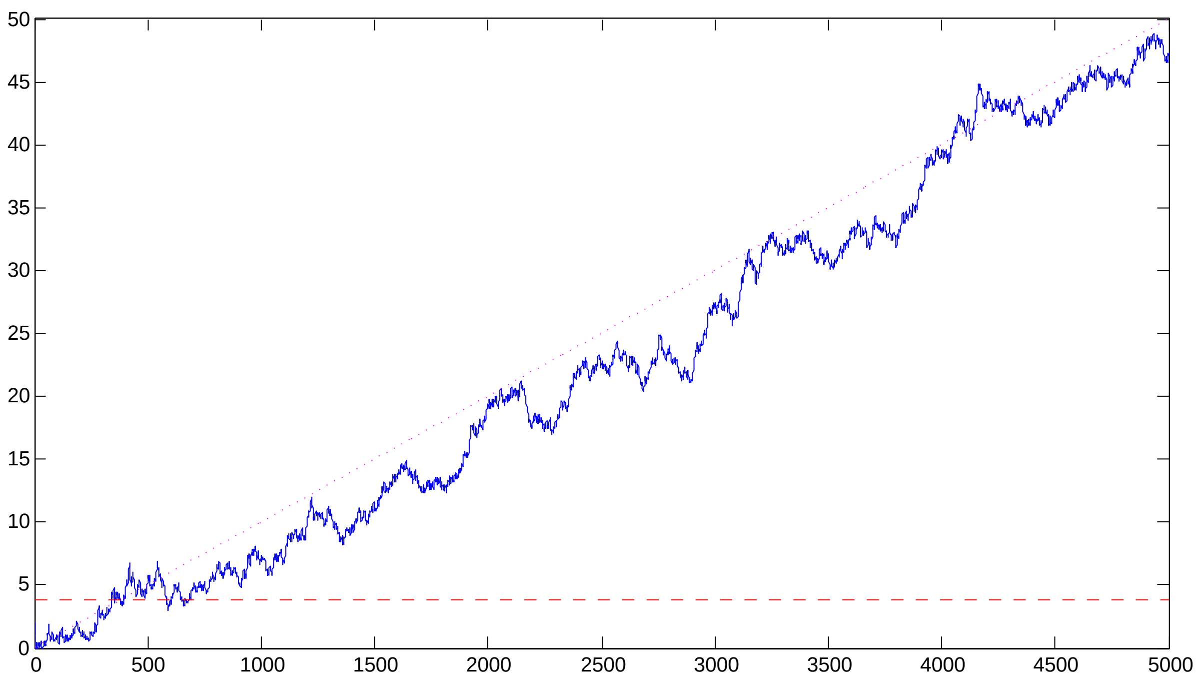

Figure 16: Chi Square Test Results - Unfair Coin - 45%, 5000 Trials - With Theoretical

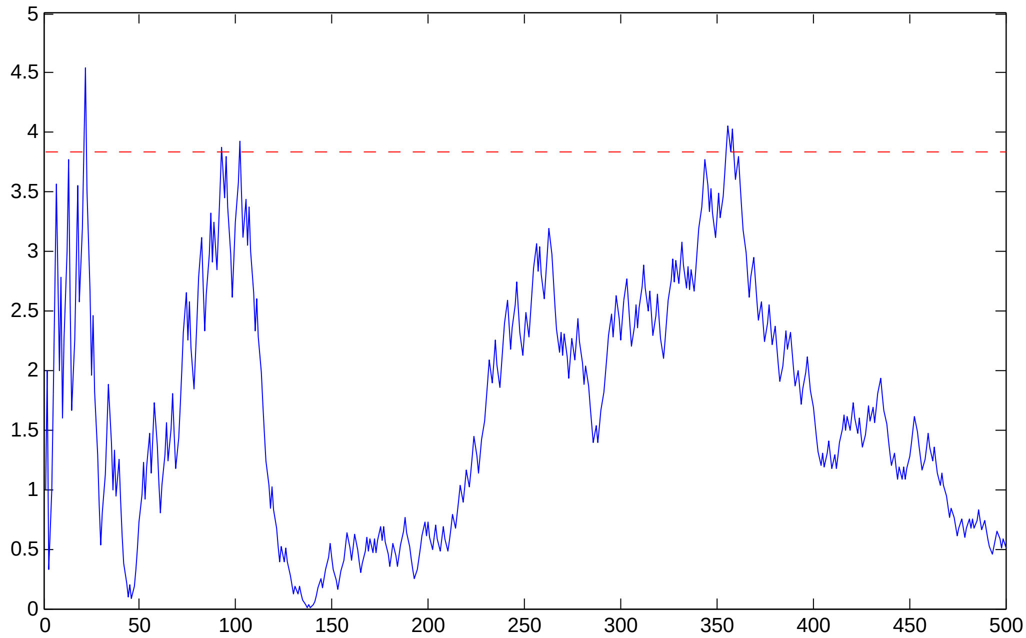

Figure 17: Chi Square Test Results - Unfair Coin - 51%, 500 Trials

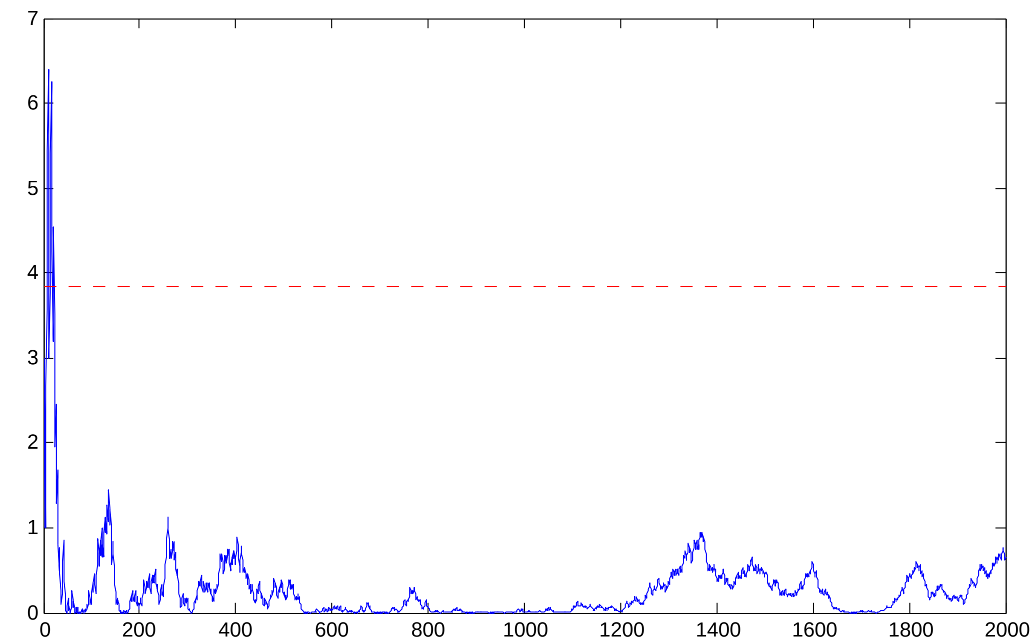

Figure 18: Chi Square Test Results - Unfair Coin - 51%, 2000 Trials

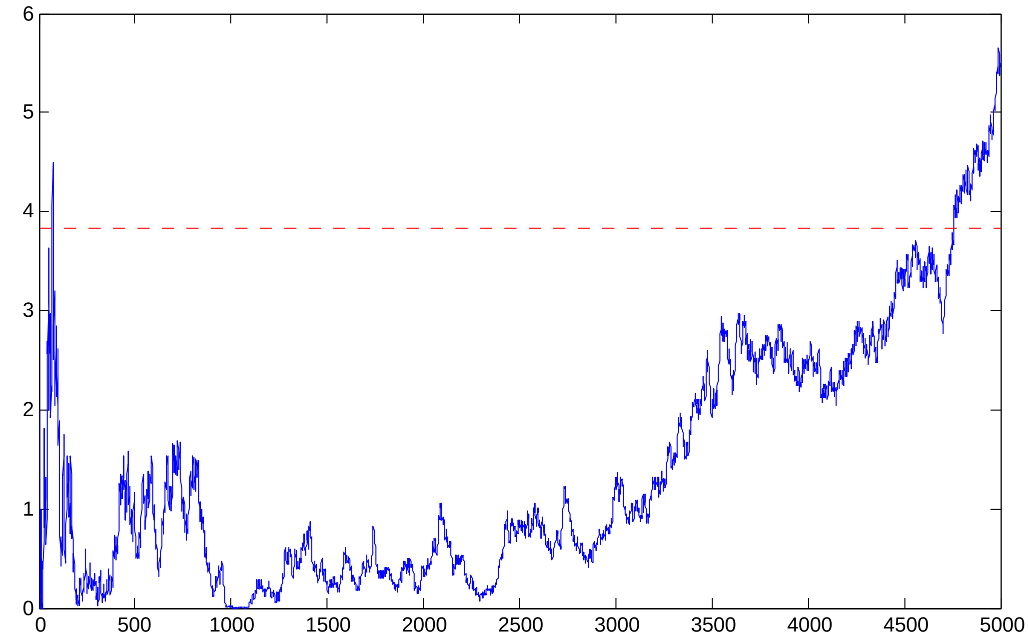

Figure 19: Chi Square Test Results - Unfair Coin - 51%, 5000 Trials

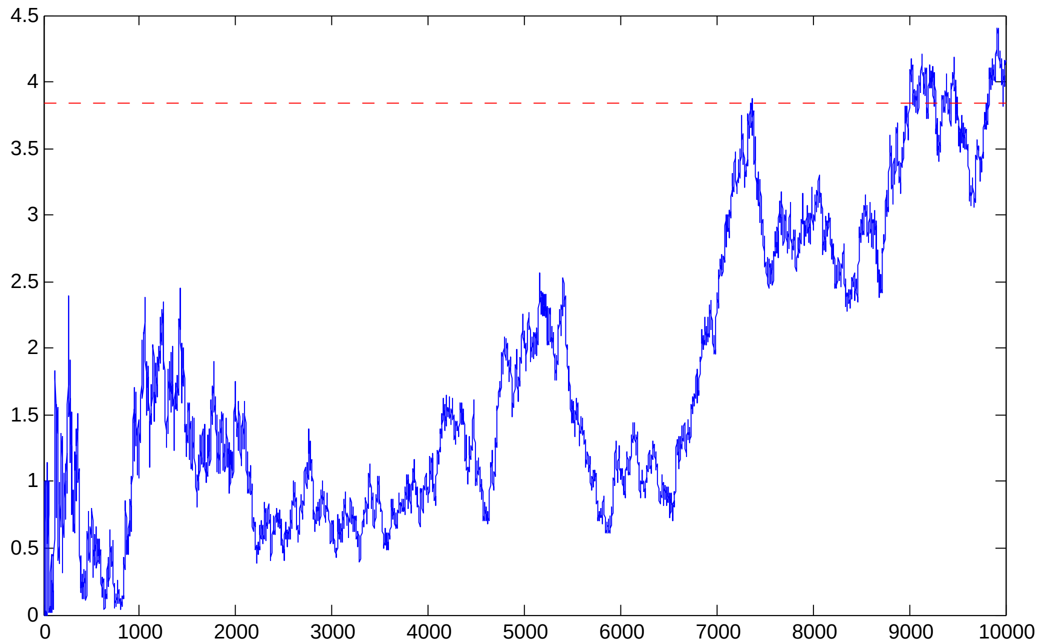

Figure 20: Chi Square Test Results - Unfair Coin - 51%, 10000 Trials

Fairness Percentage	Realizations Till Determined Unfair
50%	-
51%	4500
60%	50
70%	10

Table 3: Number of realizations to determine a coin is unfair

3.2 Karhunen-Loève & Wiener Process

For a data set of 3000 randomly generated points, the Anderson Darling statistic of 3.0122, indicates that the data is normal and that rejection of the null hypothesis is not needed.

Based off of the Anderson-Darling test results for a data set with 3000 points with 3 terms in the K-L expansion, it can be determined that the best stopping point is roughly 2500 realizations. At roughly this point, the process crosses the critical value, 2.492, which gives the 95% confidence that process is normal. It is at this point that the truncated Wiener Process is no longer valid and too far from the actual Wiener Process. Using the truncated process for values larger than roughly 2000 would result in inaccurate conclusions. Table 4 shows this result.

When the K-L has 7 terms, the results are slightly different. Table 5 shows these results. It takes about 50,000 trials before one can notice the data is not normal. At roughly 2500 realizations, the data crosses the critical value of 2.492. For the 7 term K-L, it is at this

n Trials	Realizations Till Determined Not Normal	Anderson Darling Statistic
100	-	0.3536
500	-	1.1959
700	-	0.3913
800	-	1.0178
1000	-	1.1959
3000	2750	3.0122
5000	3500	2.2677
10000	3000	2.4400
50000	5000	21.2969

Table 4: Number of realizations to determine a 3 Term K-L is not a true K-L

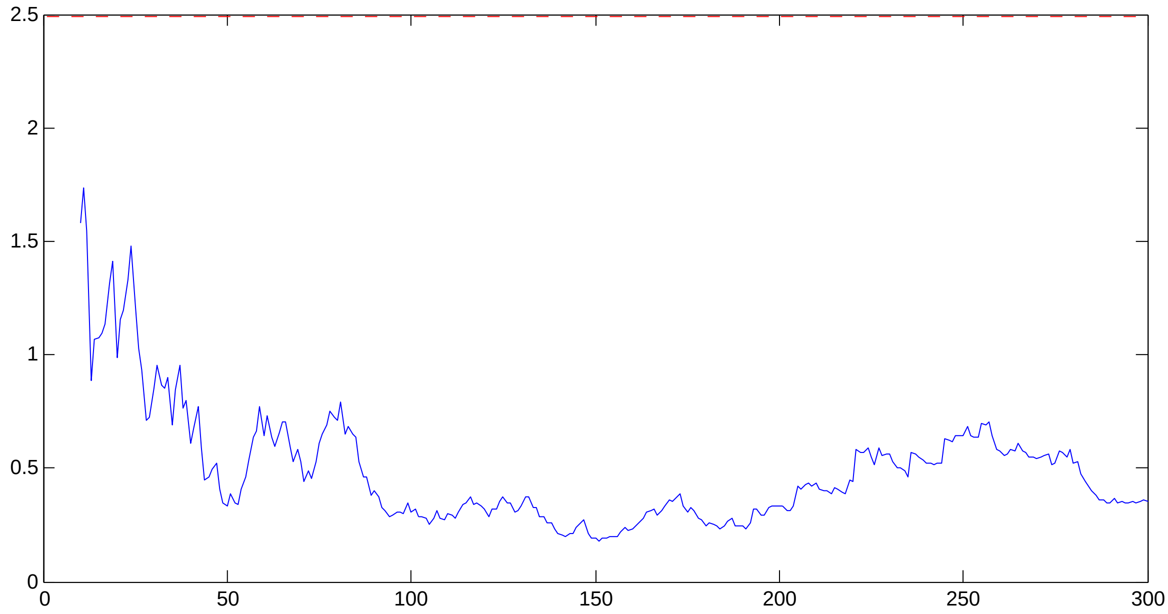

Figure 21: Anderson Darling Test - 3 Term K-L - 100 Trials

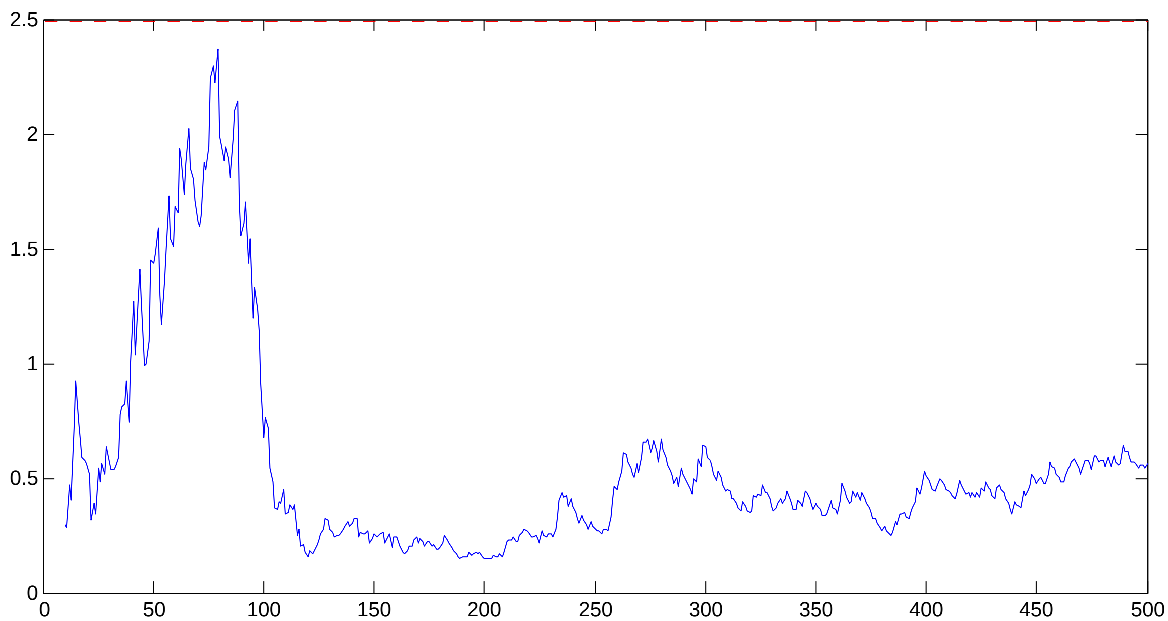

Figure 22: Anderson Darling Test - 3 Term K-L - 500 Trials

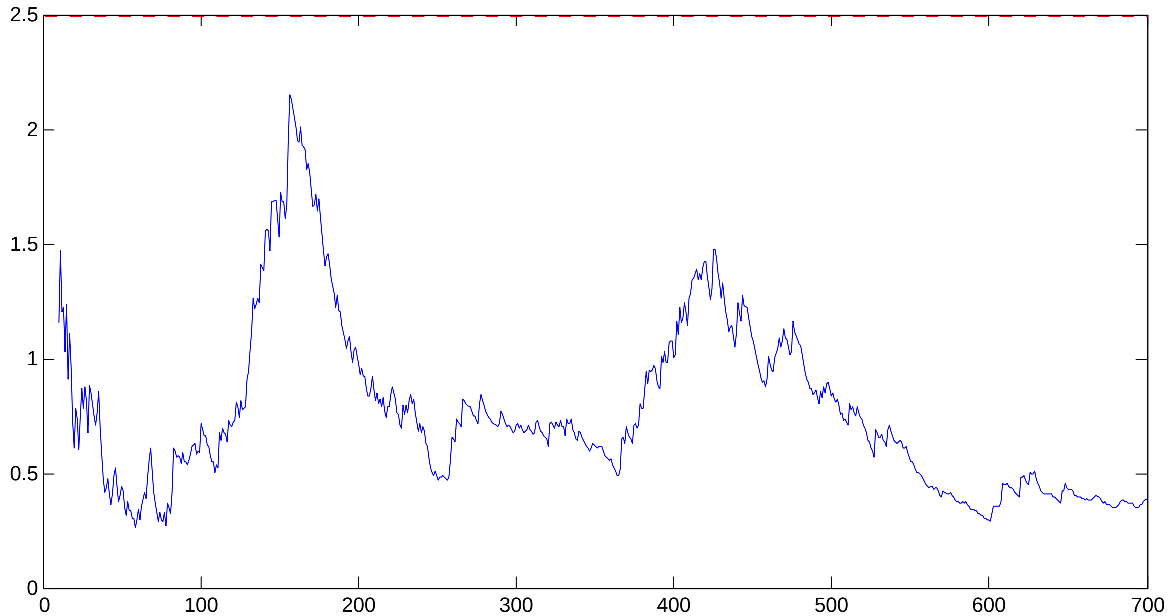

Figure 23: Anderson Darling Test - 3 Term K-L - 700 Trials

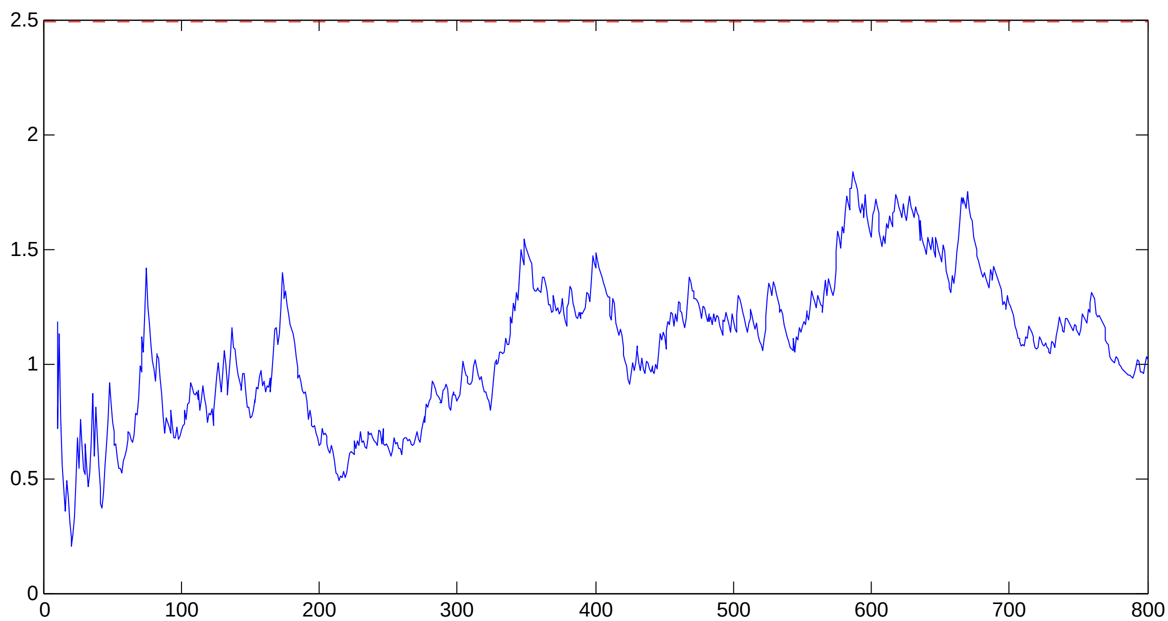

Figure 24: Anderson Darling Test - 3 Term K-L - 800 Trials

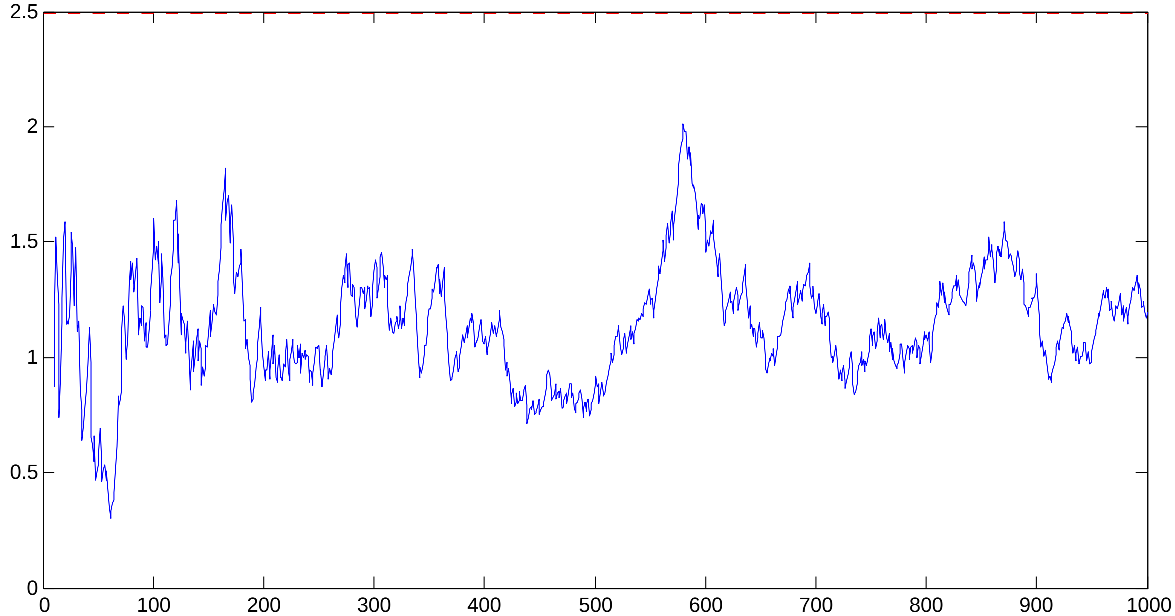

Figure 25: Anderson Darling Test - 3 Term K-L - 1,000 Trials

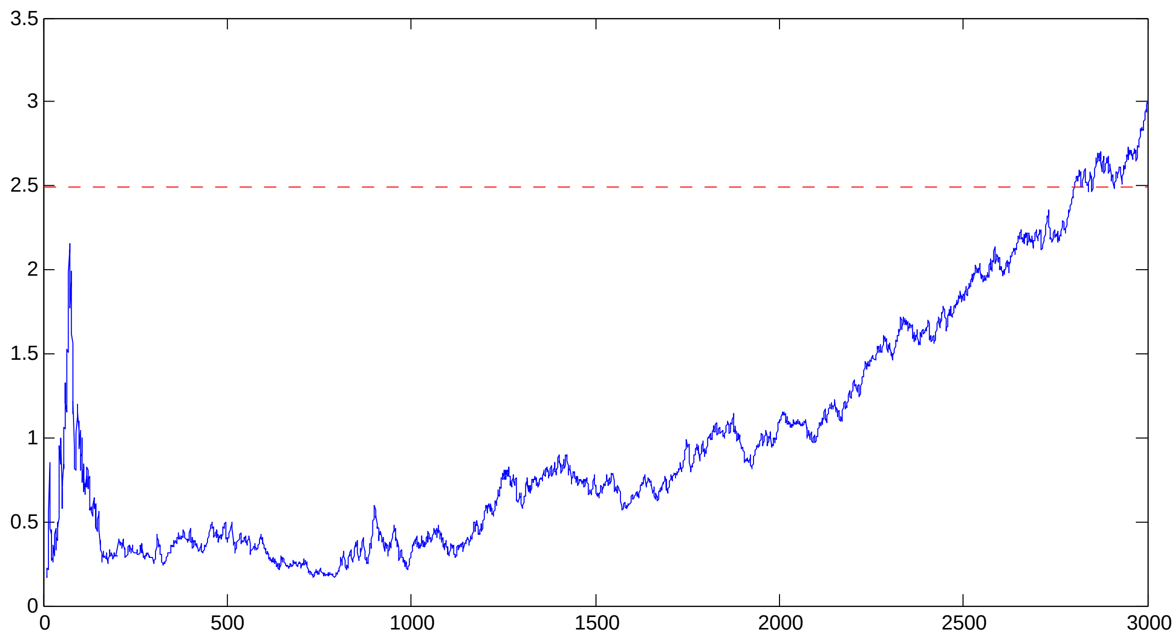

Figure 26: Anderson Darling Test - 3 Term K-L - 3,000 Trials

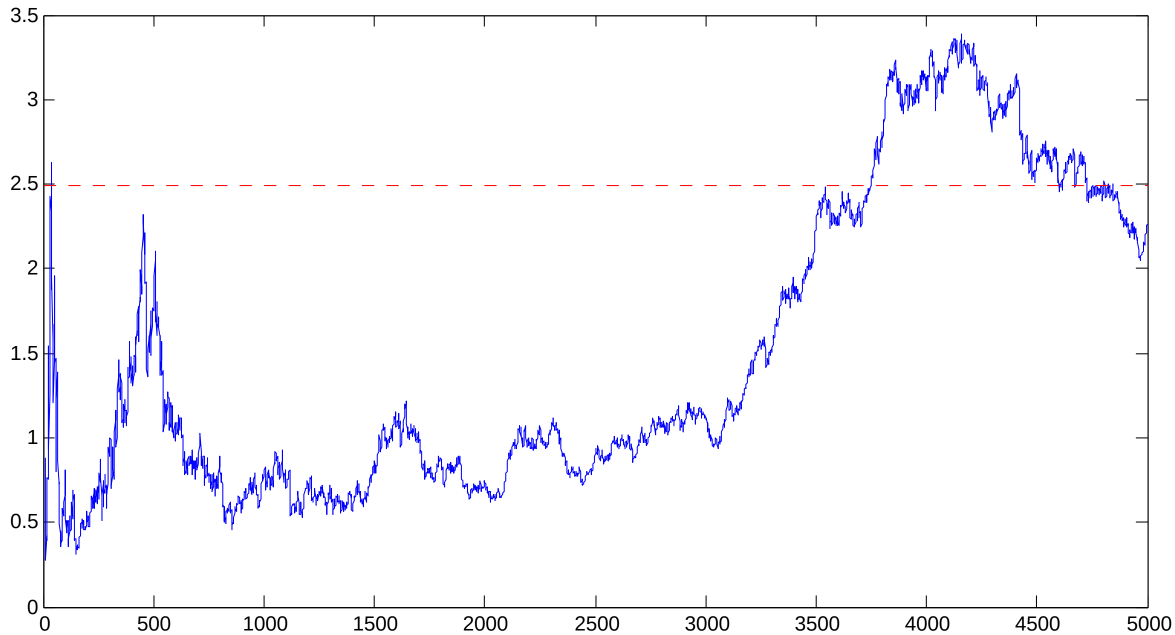

Figure 27: Anderson Darling Test - 3 Term K-L - 5,000 Trials

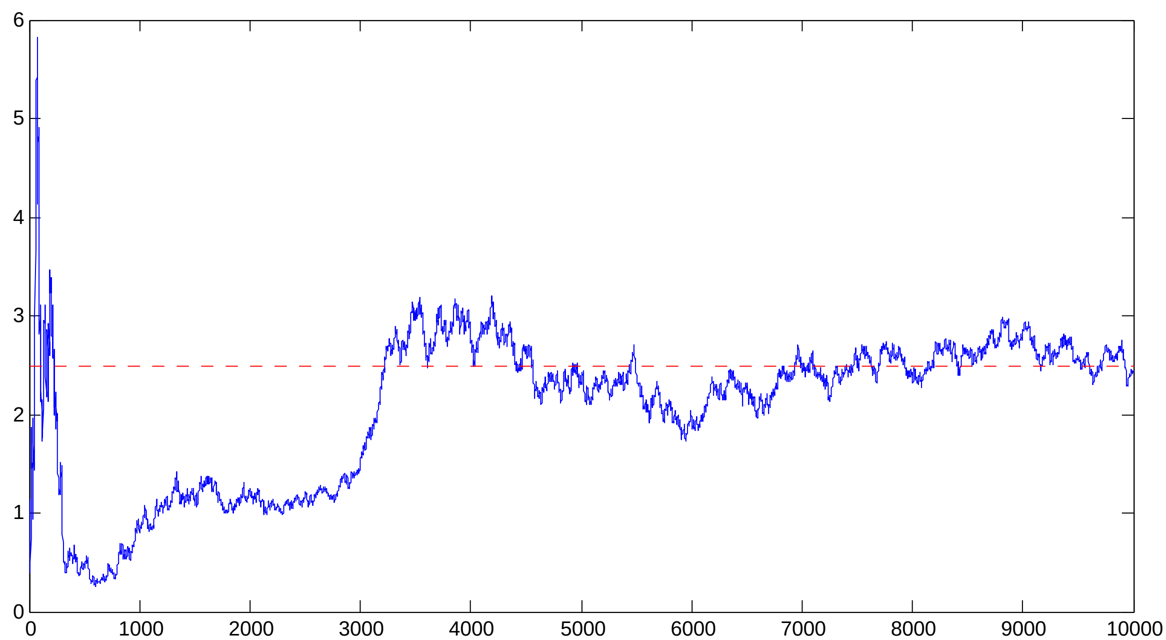

Figure 28: Anderson Darling Test - 3 Term K-L - 10,000 Trials

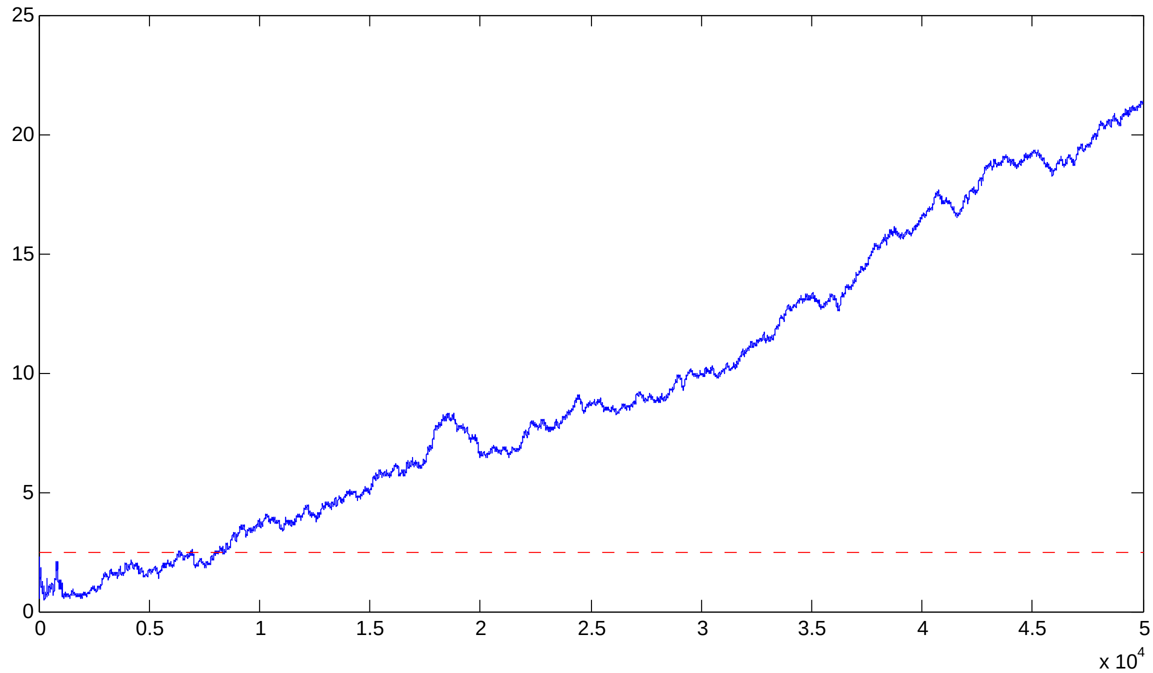

Figure 29: Anderson Darling Test - 3 Term K-L - 50,000 Trials

point that the truncated process should no longer be used, as inaccurate results may be concluded since the truncated K-L is no longer accurately portraying the non-truncated version. Figures 30 - 39 graphically displays these results.

Now that the number of realizations to determine a K-L is no longer a true Wiener Process has been determined, we can use this to solve stochastic differential equations as long as the number of realizations is kept below this limit.

n Trials	Realizations Till Determined Not Normal	Anderson Darling Statistic
100	-	0.6320
500	-	0.4203
700	-	0.8085
800	-	1.0742
1000	-	0.5924
3000	-	0.5327
5000	-	0.8282
10000	-	0.6350
50000	2750	3.0849
100000	2000	8.1503

Table 5: Number of realizations to determine a 7 Term K-L is not a true K-L

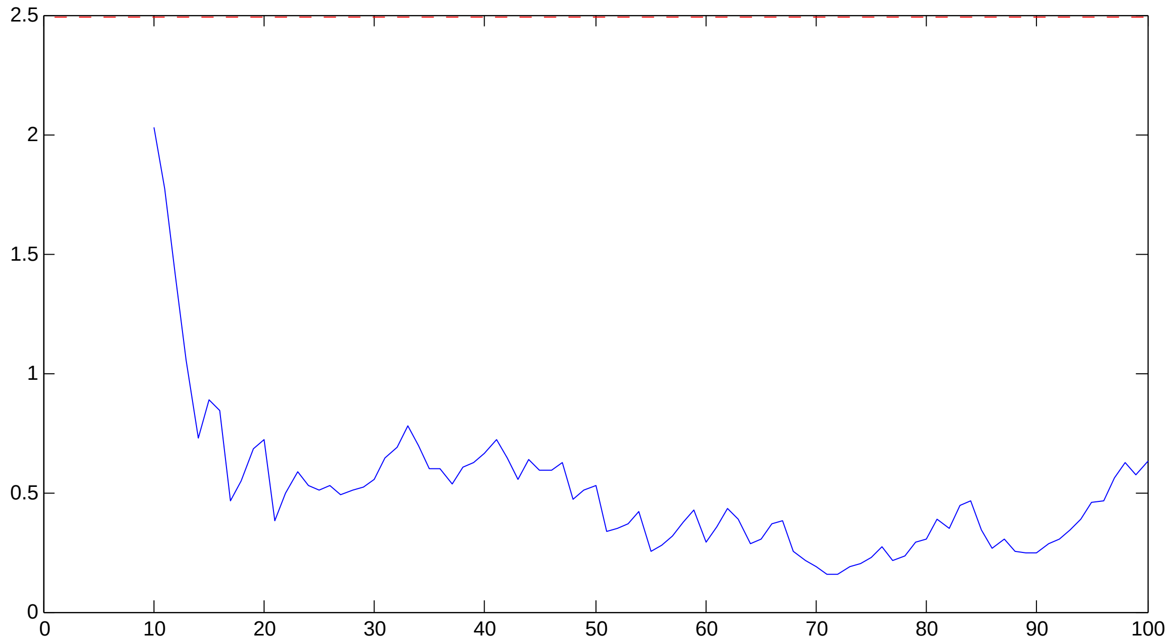

Figure 30: Anderson Darling Test - 7 Term K-L - 100 Trials

3.3 A Stochastic ODE

Equation 10 is a trivial example of solving differential equations using the separation of variables method.

$$\begin{aligned} \frac{dy}{dx} &= 6y^2x \\ \mathrm{d}y &= 6y^2x\ dx \\ \frac{dy}{y^2} &= 6x\ dx \\ \int y^{-2}dy &= \int x\ dx \\ -\frac{1}{y} &= 3x^2 + C \end{aligned} \tag{10}$$

For the purpose of this project, equation 11 represents what can be solved using the

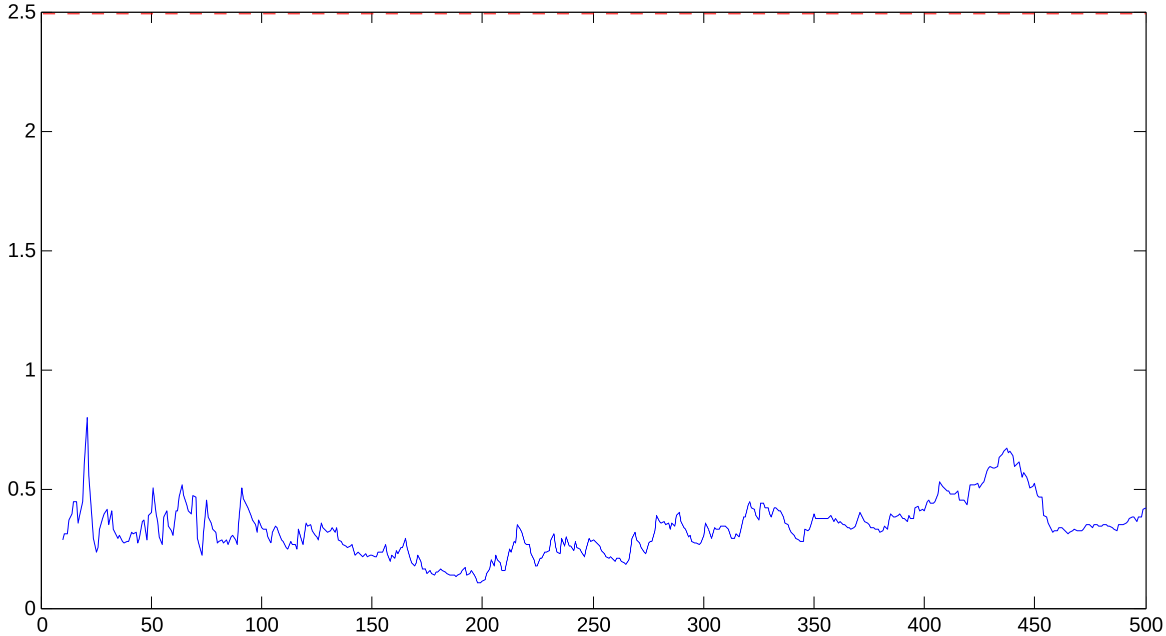

Figure 31: Anderson Darling Test - 7 Term K-L - 500 Trials

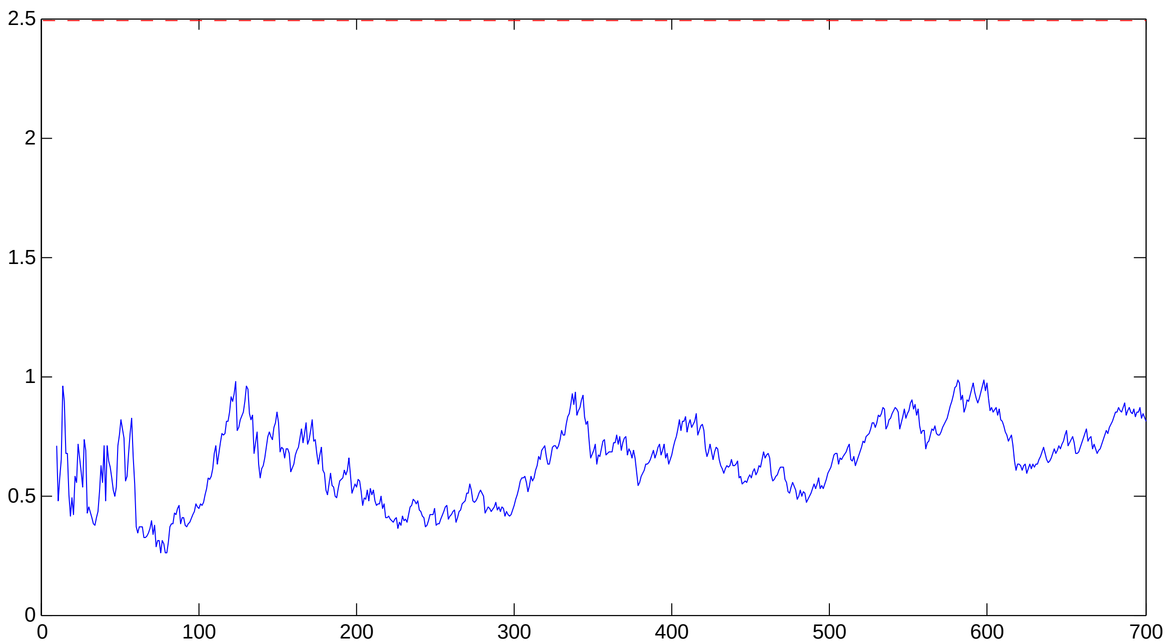

Figure 32: Anderson Darling Test - 7 Term K-L - 700 Trials

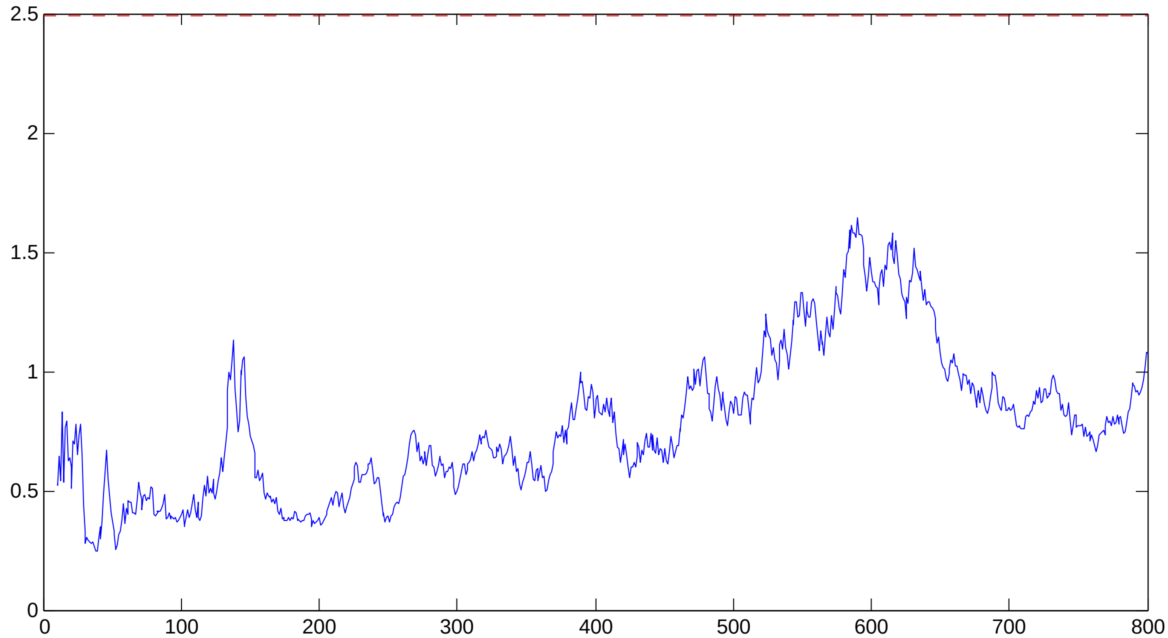

Figure 33: Anderson Darling Test - 7 Term K-L - 800 Trials

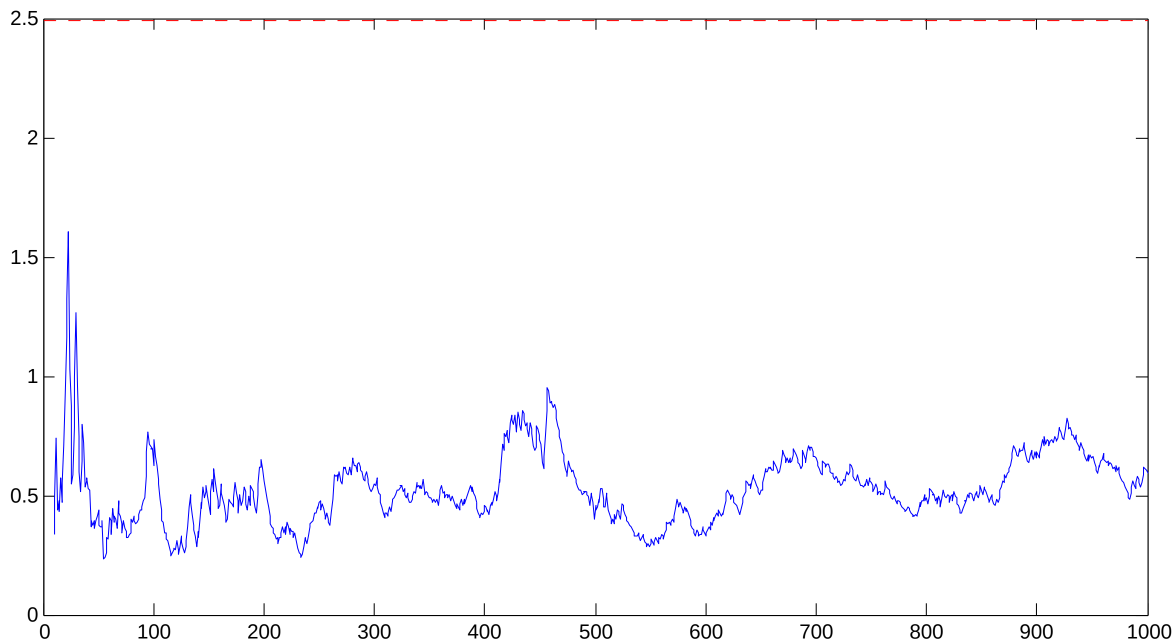

Figure 34: Anderson Darling Test - 7 Term K-L - 1,000 Trials

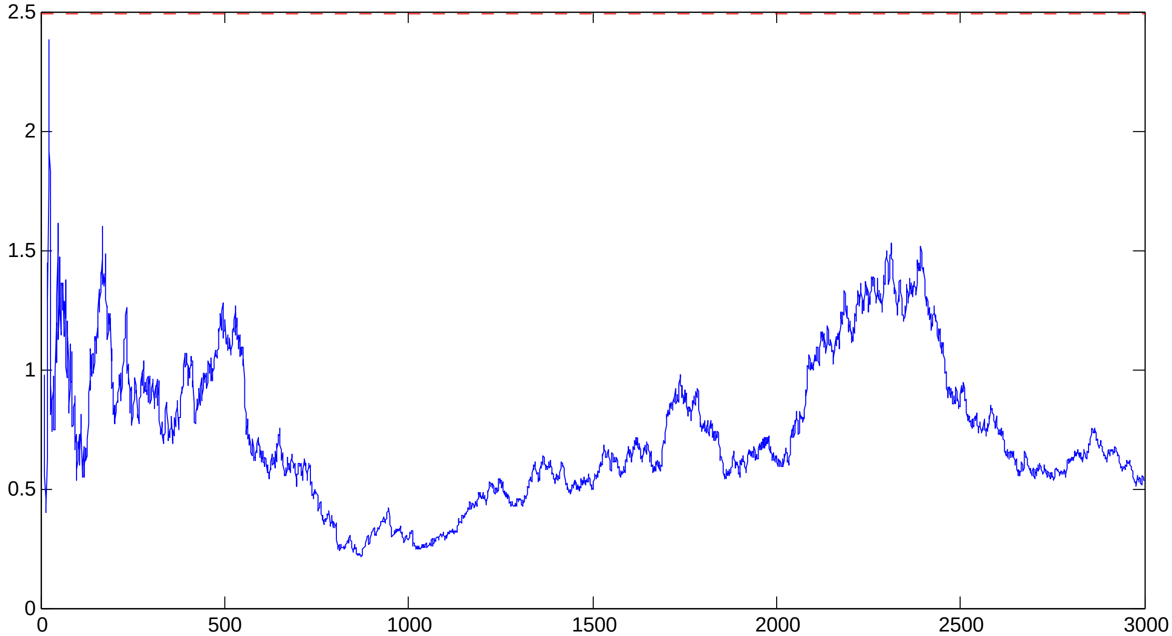

Figure 35: Anderson Darling Test - 7 Term K-L - 3,000 Trials

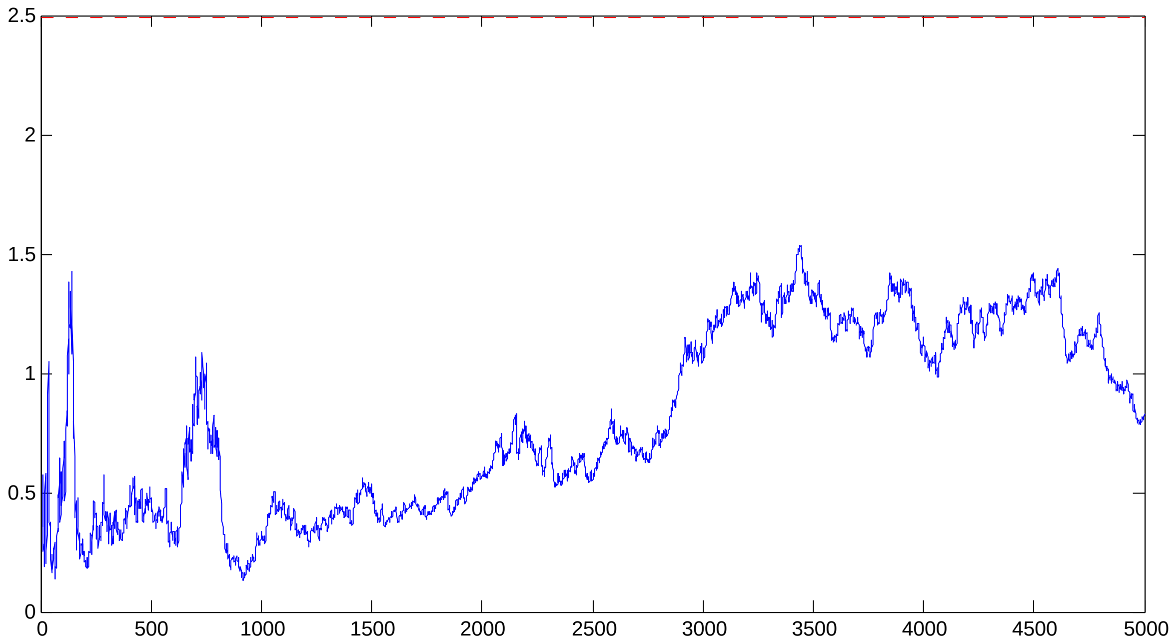

Figure 36: Anderson Darling Test - 7 Term K-L - 5,000 Trials

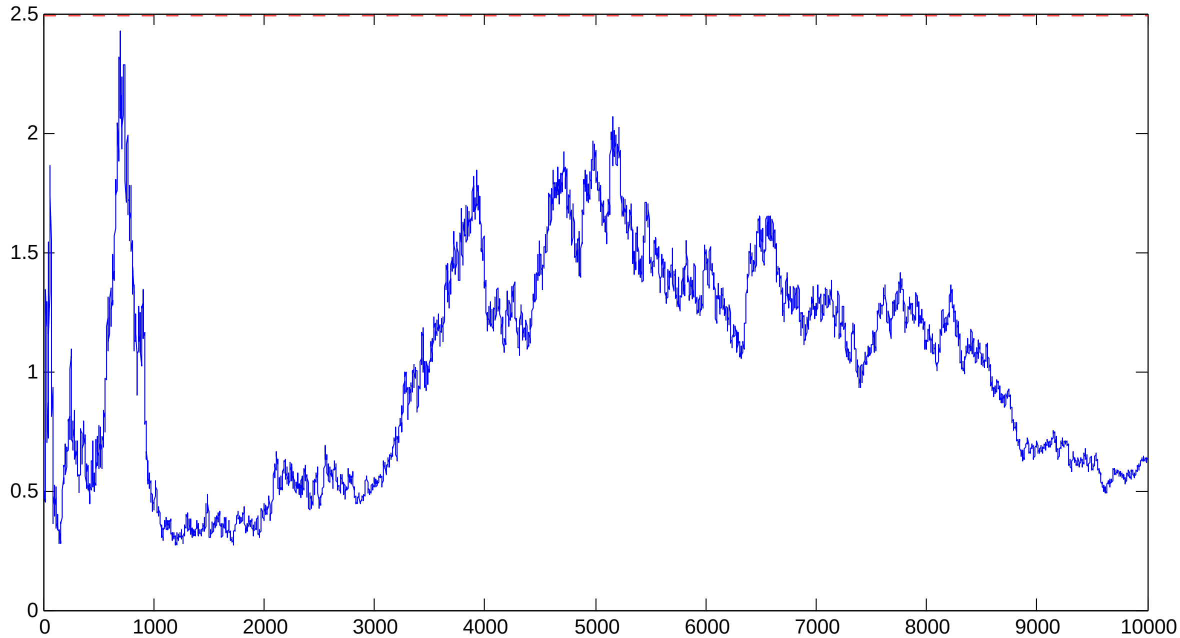

Figure 37: Anderson Darling Test - 7 Term K-L - 10,000 Trials

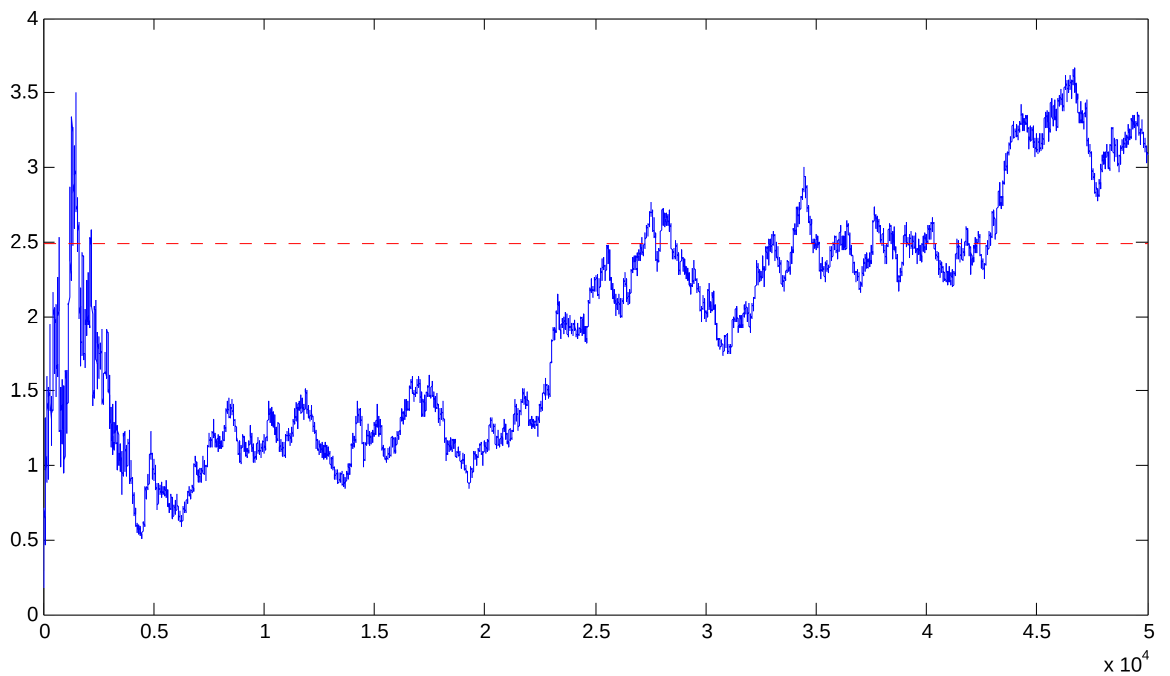

Figure 38: Anderson Darling Test - 7 Term K-L - 50,000 Trials

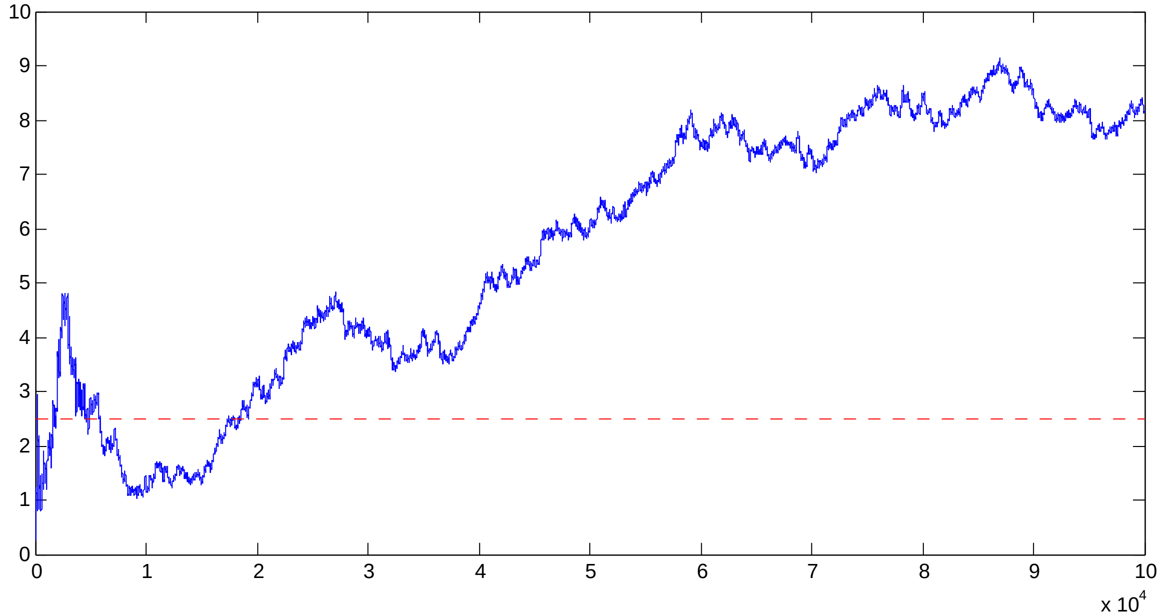

Figure 39: Anderson Darling Test - 7 Term K-L - 100,000 Trials

truncated Wiener Process information from above.

$$\begin{aligned}
\frac{dx}{dt} &= \mu x + \sigma x(\alpha_1 \sin \beta_1 t + \alpha_2 \sin \beta_2 t) \\
\frac{dx}{dt} &= (\mu + \sigma\alpha \cos \beta_1 t + \sigma\alpha \cos \beta_2 t)x \\
\int \frac{dx}{dt} &= \int (\mu + \sigma x_1 \cos \beta_1 t + \sigma\alpha_2 \cos \beta_2 t) \\
&= \mu t + \frac{\sigma\alpha_1}{\beta_1} \sin \beta_1 t + \frac{\sigma\alpha_2}{\beta_2} \sin \beta_2 t + C
\end{aligned} \tag{11}$$

Since $x(t) = 1 \Rightarrow A = 1$ it can then be said that

$$x = Ae^{\mu t} \frac{\sigma\alpha_1}{\beta_1} \sin \beta_1 t + \frac{\sigma\alpha_2}{\beta_1} \sin \beta_2 t \tag{12}$$

We can now look at the the stochastic differential equation:

$$x_t = \mu x + \sigma x \frac{dw}{dt} \tag{13}$$

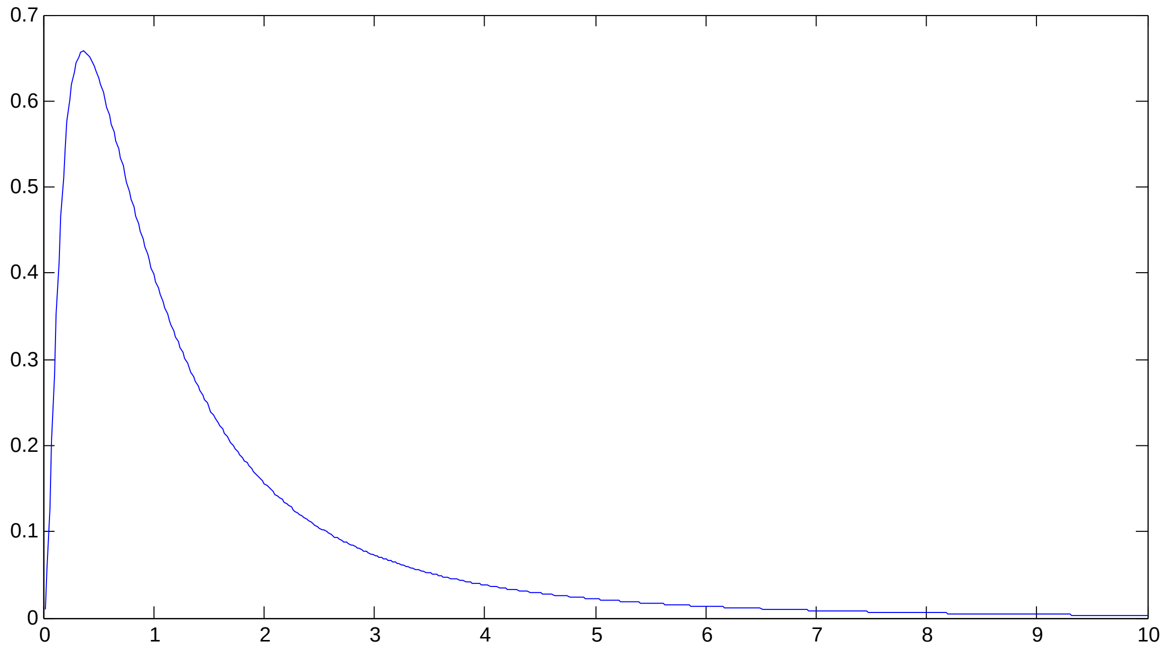

Figure 40: Log Normal Distribution

This has the analytical solution:

$$x = \int \mu x + \int \sigma x \frac{dw}{dt}\ dt \tag{14}$$

The error for these equations comes into play based off the log normal distribution, Figure 40. The error will all depend on σ as that would be a larger approximation of the random piece of the equation.

4 Conclusion

Through the use of statistical tests to determine normality, it can be determined that the truncated K-L expansion can be used to help solve problems including the Wiener Process. Since the process cannot be differentiated up to a given number of realizations, using the truncated version gives practically the same results. This means that this problem that was once super difficult to do, has now been extremely simplified. Given the number of trails, one can tell when the truncated version no longer will be a good choice. If the number of realizations needed is less than that given value, then one can resort to the truncated process so the equation can be solved quicker and easier. Through these results, stochastic differential equations, such as the example above, can be solved much more easily.

4.1 Further Investigation

This project has shown how to approximate a Wiener Process using a trucnated K-L expansion. It has further shown that not many terms are needed before one can notice that it is no longer working. Further research can show how to replace a complicated differential equation with a simple deterministic differential equation to assist in analyzing different equations inorder to make solving them easier.

5 Code

5.1 Unfair Coin

```
clear all;
close all;

coinFairness = .7;
numberOfTrials = 5000;
numHeads = 0;
trials = rand(numberOfTrials, 1);

heads = zeros(numberOfTrials,1);
chi = zeros(numberOfTrials,1);

for i = 1:numberOfTrials

    if (trials(i) < coinFairness)
        numHeads = numHeads + 1;
    end

    heads(i) = numHeads/i;

    theoretical(i) = i/100;
    chi(i) = i*(numHeads-0.5*i)^2/(0.5*i*(i-0.5*i));

end
plot(1:numberOfTrials,heads);
plot(1:numberOfTrials, chi);
hold on
plot([1, numberOfTrials], [3.84, 3.84], 'r--')
plot(1:numberOfTrials,theoretical, 'm:')
disp(sum(chi));
disp(chiTest);
```

5.2 Determination of Normality

```
% Given data, use the Anderson Darling test to verify normal
n=2000;
x=zeros(n,3);
x(:,1)=sort(randn(n,1));
x(:,2)=linspace(1/n,1,n);
x(:,3)=0.5*(1+erf(x(:,1)/sqrt(2)));

plot(x(:,1),x(:,2),x(:,1),x(:,3))

s=0;

for k=1:n,
  s=s+(2*k-1)*(log(x(k,3))+log(1-x(n+1-k,3)));
end;


asqr=-n-s/n;
disp(asqr)
```

5.3 Anderson-Darling Example Code

5.3.1 Successive Sets of Data

```
% Given data, use the Anderson Darling test on successive sets of data.
n=2000; y=randn(n,1)*sqrt(0.9331); asqr=zeros(n,1);

for k=10:n
  x=zeros(k,3);
  x(:,1)=sort(y(1:k,1));
  x(:,2)=linspace(1/k,1,k);
  x(:,3)=0.5*(1+erf(x(:,1)/sqrt(2)));
  s=0; for j=1:k, s=s+(2*j-1)*(log(x(j,3))+log(1-x(k+1-j,3))); end;
  asqr(k)=-k-s/k;
end
plot(10:n,asqr(10:n))
disp(asqr(2000))
```

```
% Given data from KL, use the Anderson Darling test to verify normal
n=50000;
x=zeros(n,3);
for i=1:n
  for k=1:3
    y=(k-1/2)*pi;
    x(i,1)=x(i,1)+randn*sin(y)/(y);
  end
end
x(:,1)=sort(x(:,1))*sqrt(2);
x(:,2)=linspace(1/n,1,n);
x(:,3)=0.5*(1+erf(x(:,1)/sqrt(2)));

plot(x(:,1),x(:,2),x(:,1),x(:,3))

s=0; for k=1:n, s=s+(2*k-1)*(log(x(k,3))+log(1-x(n+1-k,3))); end;
asqr=-n-s/n;
disp(asqr)
```

```
% Given data, use the Anderson Darling test on successive sets of data.
n=100000; y=zeros(n,1);
for i=1:n
  for k=1:7
    t=(k-1/2)*pi;
    y(i)=y(i)+randn*sin(t)/(t);
  end
end
y=y*sqrt(2);
asqr=zeros(n,1);

x=zeros(9,3); x(:,1)=sort(y(1:9));

for k=10:n
  temp=x(:,1);
  x=zeros(k,3);
  x(:,1)=sort([temp;y(k)]);
  x(:,2)=linspace(1/k,1,k);
  x(:,3)=0.5*(1+erf(x(:,1)/sqrt(2)));
  s=0; for j=1:k, s=s+(2*j-1)*(log(x(j,3))+log(1-x(k+1-j,3))); end;
  asqr(k)=-k-s/k;
end
plot(10:n,asqr(10:n))
disp(asqr(2000))
```

5.4 Sample Wiener Process

```
clear all

x = linspace(0,1,501);
y = zeros(1,501);

for i = 1:500
    y(i+1) = y(i) + randn / 500;
end

plot(x,y);
```

5.5 Log Normal Distribution

```
clear all

x = linspace(0,10,500);
y = zeros(500,1);

mu = 0;
sigma = 1;

for i = 1:500
    y(i) = 1/(x(i)*sigma*sqrt(2*pi)) * e^((-(log(x(i))-mu)^2)/(2*sigma^2));
end

plot(x,y);
```

References

[1] Allison June Barlow Chaney. *Wiener Process*. URL: `http://www.princeton.edu/~achaney/tmve/wiki100k/docs/Wiener_process.html`.

[2] J. L. Doob. *The Development of Rigor in Mathematical Probability*. Amer. Math. Monthly. 1996. URL: `http://www.princeton.edu/~achaney/tmve/wiki100k/docs/Wiener_process.html`.

[3] S. P. Huang, S. T. Quek, and K. K. Phoon. "Convergence study of the truncated Karhunen–Loeve expansion for simulation of stochastic processes". In: *Wiley Online Library* (2001). DOI: `10.1002/nme.255`. URL: `http://onlinelibrary.wiley.com/doi/10.1002/nme.255/abstract`.

[4] C.K. Raju, ed. *Kosambi the Mathematician* 44.20 (2009): *Economic and Political Weekly*. 33-45. URL: `http://www.jstor.org/stable/40279011`.

[5] M. A. Stephens. "EDF Statistics for Goodness of Fit and Some Comparisons". In: *Journal of the American Statistical Association* 69.347 (September 1974), 730–737.

[6] M. A. Stephens. "Use of the Kolmogorov-Smirnov, Cramet-Con Mises and Related Statistics without Extensive Tables". In: *Journal of the Royal Stasticial Society* 32.1 (1970), pp. 115–122.