Draft version January 15, 2020
Typeset using LaTeX **twocolumn** style in AASTeX63

Modeling Lab 1: Getting My Toolbox Ready

Kathlyn M. Purcell[1]

[1] *Villanova University*

1. QUESTION 1

In question 1, I used a model for a Gaussian function to make a string for mean 1, standard deviation 0.1 and length 1 million. This was done using one line of code and generated 1 million random numbers following those constraints. This string was used in every subsequent solution.

2. QUESTION 2

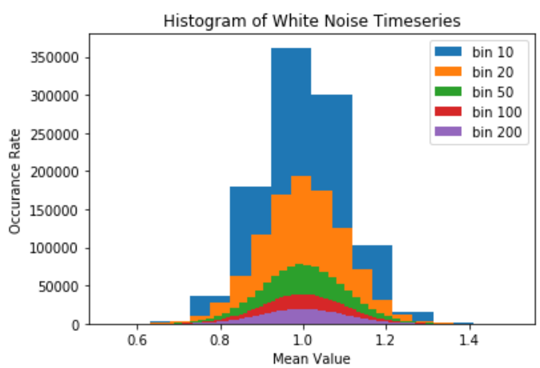

Figure 1. Overlayed Histogram of a White Noise Timeseries with Varied Binning, *Made through PYPLOT*

For this question I plotted a histogram with 5 overplotted graphs depending on their bin number shown below in Figure 1. 10 bin is the 'tallest' due to the fact that it was not distributed over a large bin length; this makes it have only 10 sections to store 1 million numbers, so all the bin- 10 sections were relatively 'tall'. 200 binning was the 'shortest'; this was due to the fact that the bulk of the 1 million variables could spread itself out over 200 sections resulting in smaller, but more, bins... hence the 'shorter' histogram.

3. QUESTION 3

The autocorrelation resulted in a strong peak at the 0 lag value. Figure 2 shows this effect. This was concurrent with our Gaussian predictions where the directly correlated values would result in a very strong peak when the maximum values were multiplied. As the lag shifted further from perfect correlation there is a drop in autocorrelation values on both sides of the 0 value

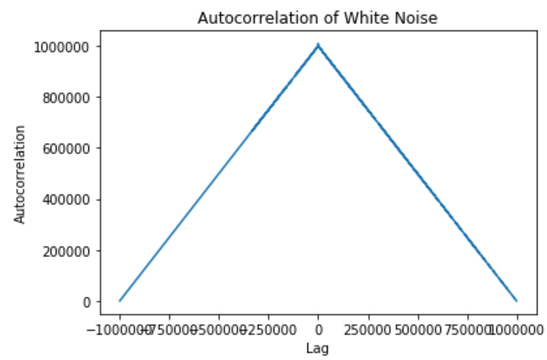

Figure 2. Autocorrelatted White Noise Time Scale, *Made through PYPLOT*

signifying the diminishing multiplication values as the maximum Gaussian values were only able to multiply with the smaller, less probable values as they fell off the further they strayed from the mean of 1. Even though our values were randomly selected, over the entire 1 million point range of the White Noise sample they were able to even out to produce relatively constant (looking) autocorrelation lines.

4. QUESTION 4

With 1-, 2-, 3-, 4-, 5-σ one would expect probabilities of 68.2%, 95.4, 99.7, 99.90, and 100, respectivly, for the data points to fall into the standard deviations (NDT Resource Center). This was modeled over 10 iterations of our randomly generated Gaussian function to obtain values of the following in Table 2 at the end of the paper.

This table came about by running the program 10 times to determine the range of random-ness in the separate 1 through 5 sigma deviation sections. It is clear that the distribution is relatively accurate. The discrepancies arise because the series is randomly generated. The largest deviation is in the 5σ distribution due to its small size. When any number is randomly selected from this region the relative change is greater due to the smaller sample size.

5. QUESTION 5

The cumulative function was compiled and resulted in a trendline that was increasing at a linear rate. This concurs with our logical predictions because the sample size was so great, and the standard deviation was so small that the overall trendline would experience little shifting either up or down. See Figure 3 below.

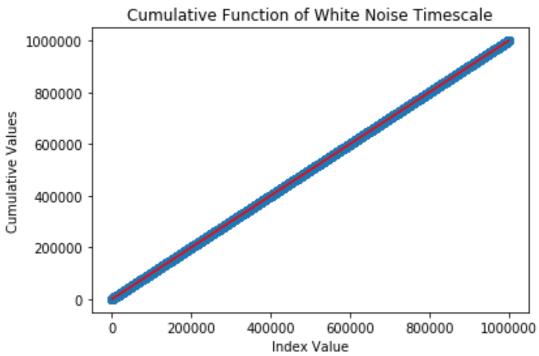

Figure 3 . Cumulative Scatterplot of a White Noise Timeseries with Least Squares Trendline Fit to the Data, *Made through PYPLOT*

This table represents that the program was run 10 times resulting in relatively constant but varying slope and r squares values and a wildly varying intercept value when the program generated different random values every time it was run.

5.1. *Slope*

This concurs further when the line is fit to a trendline of least squares and results in a near constant slope of 1. This is given in Table 1. Our mean value for this Gaussian White Noise timeseries was 1 so it follows logically that the slope results in a linear value of 1.

Slope	Intercept	R squared value
1.0000728986705358	4.373829519725405	0.9999999948131219
1.0000541588386727	33.09480732353404	0.9999999968549568
0.9999349155941792	-47.02312075352529	0.9999999968658031
1.0001517951371557	-88.10609549528453	0.9999999914730278
1.0001442325450842	-33.0111585764098	0.999999995598952
1.0000588919272002	-41.829512981232256	0.9999999928545146
1.0000371099738339	24.52080549352104	0.999999994231322
1.0000913339630342	-5.217409530130681	0.9999999931828204
1.000172370727833	66.64866257092217	0.9999999901532981
1.0000814518251664	35.71174247196177	0.9999999948205299
0.9999012490260744	-8.83918115781853	0.9999999976298882

Table 1. Slope and Intercept Values from Several Iterations of a Randomly Generated White Noise Timescale That Has Been Interpreted Through a Cumlative Function

5.2. *Intercept*

Again Table 1 should be referenced, this time the intercept values are of note. The intercept on the other hand resulted in a widely varying value from around -60 to +60. This can be explained by the large sample size. If the intercept is a sizable negative number, it can be assumed that there were low values dominating the Gaussian fit. If the intercept is a large positive number, then the domination can be seen as higher values than the mean of 1. This is to say that if the intercept was determined to be close to 1 then the high and low trend of the random-ness of the string from Question 1 had canceled itself out and produced a prefect Gaussian curve. Since the intercepts in this section varied to a great degree we can retroactively affirm that Question 1 was completed accurately and a truly random string was formed.

Sigma Number	Distribution Percentage	Number out of 1 Million
1	0.681585	681585
2	0.954227	272642
3	0.997223	42996
4	0.99994	2717
5	1	60
1	0.68284	682840
2	0.954541	271701
3	0.997313	42772
4	0.999937	2624
5	1	63
1	0.682553	682553
2	0.954377	271824
3	0.997282	42905
4	0.999931	2649
5	1	69
1	0.683173	683173
2	0.954785	271612
3	0.997278	42493
4	0.999928	2650
5	0.999999	71
1	0.682658	682658
2	0.95492	272262
3	0.997383	42463
4	0.999948	2565
5	1	52
1	0.682874	682874
2	0.9542	271326
3	0.99726	43060
4	0.999911	2651
5	1	89
1	0.682609	682609
2	0.954439	271830
3	0.997215	42776
4	0.999934	2719
5	0.999999	65

Table 2. pt. 1: Understanding the Variations in the Levels of Sigma 1-5 when the Program Randomly Generates a White Noise Time Function

Sigma Number	Distribution Percentage	Number out of 1 Million
1	0.682628	682628
2	0.954443	271815
3	0.997301	42858
4	0.999928	2627
5	1	72
1	0.683329	683329
2	0.954448	271119
3	0.997271	42823
4	0.999946	2675
5	1	54
1	0.682681	682681
2	0.954504	271823
3	0.997294	42790
4	0.999939	2645
5	1	61
1	0.683025	683025
2	0.954561	271536
3	0.997372	42811
4	0.999942	2570
5	1	58

Table 3. pt. 2: Understanding the Variations in the Levels of Sigma 1-5 when the Program Randomly Generates a White Noise Time Function