Anton LeKang

ECE 3522: Stochastic Processes in Signals and Systems

Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA 1912

CA8

# Problem Statement

The objective of this computer assignment is to look at a uniform sum of independent variables, and to observe what happens with the central limit theorem. From increasing the amount of data points, the fundamental principles of the Central Limit Theorem can be observed. The whole objective of the Central Limit Theorem is to show that eventually the data will come out to a normal distribution. Even though these variable are two independent one, by increasing the size and summing them help reiterate the theorem. This computer assignment will use the sine wave, and Gaussian white noise function in MATLAB to observe the effects on the total signal.

# Approach and Results

Part 1: Create a function to generate a sum of uniformly distributed mutually independent random variables

Mainly this part of the lab can be observed in the code section. But to create this function the min and max must be set to the given parameters of -1, and 1. The function also has two other inputs, n which is the number of random variables, and N which is the total sample generated.

Here is what the function looks like:

Part 2: Looks at the RMS error between Gaussian, and the actual distribution

Figure 1 – Number of random variables is 10

Figure 2 – Number of random variables is 100

These two figures show the first two images where varying number of random variables and looking at the Gaussian and the sum. Figure 1 represents when there are 10 random variables, and figure 2 represents when there are 100. The blue line represents the histogram of the summation function from part 1, and the red line shows the error between the actual distribution and Gaussian.

Again these two figures above pertain to the second part of this computer assignment where the RMS of the Gaussian and the summation are plotted. The figure on the left represents when there are 1,000 random variables. These plots go higher than the desired random variable of 100 to represent a better Gaussian fit. The figure on the right shows the data when there are 10,000 random variables. Again the blue line represents the uniformly distributed sum, and the red line represents the MSE between the actual data and the Gaussian.

Figure 3 – Number of random variables is 1000

Figure 4 – Number of random variables is 10000

Part 3: Plot RMS and gausian overlay for n=1, 10, and 10000

Figure 5 – Sum and Gaussian 10 data points

Figure 6 - Sum and Gaussian 100 data points

These two figures show the first of three that were plotted representing the actual distribution with the overlay of the Gaussian fit. The blue line represents the distribution of the sum, and the red line is supposed to be the Gaussian of the sum. Unfortunately the Gaussian should not just be a straight line and is not coming outright. The Gaussian normal pdf function is used with the signal, but the output is not coming correct.

Figure 7 – Sum and Gaussian 1000 data points

This is the last figure for part three of this computer assignment. Again the number of data points was increased to show a better Gaussian fit. The blue line shows a nice interpretation of the summing function with the two independent variables. The function is starting to look more and more Gaussian as the number of plots increased. Again the red plot which is supposed to be the Gaussian is not coming out correcting. The plot should also get better as the number of data points increases..

# MATLAB Code

Part 1:

function CA\_8\_2

 %number of samples generated

 N = 10e3;

 %generate a sum of random variables vector for variying

 %values of n

 for n = 10000

 %generate sum of RV's

 S = SRV\_gen(n, N, -1, 1);

 z = length(S);

 %calculate mean and variance

 mn(n) = mean(S);

 vr(n) = var(S);

 sq(n) = sqrt(n);

 %find gaussian distribution

 gaus = normpdf(S,mn(n),sq(n))

 %find actual distribution

 dist = hist(S, z);

 dist = dist/100;

 %find mean squarred error and store

 MSE = mse(gaus, dist)

 figure(1)

 plot(gaus, 'r');

 hold on

 plot(dist, 'b');

% %plot distributions for certain values of n

% if n = [1, 10, 100]

% figure(i);

% plot(gaus);

% hold on

% plot(dist, 'b');

% i = i+1;

% end

 end

%

% figure(4)

% plot(MSE);

end

%generates sum of RV

function S = SRV\_gen(n, N, min, max)

 x = min + (max-min)\*rand(n, N);

 [z,q] = size(x);

 %sum the collumns

 for i = 1:z

 S(i) = sum(x(i, :));

 end

end

Part 2: Similar to part 1

function CA\_8\_2

 %number of samples generated

 N = 10e3;

 %generate a sum of random variables vector for variying

 %values of n

 for n = 100

 %generate sum of RV's

 S = SRV\_gen(n, N, -1, 1);

 z = length(S);

 %calculate mean and variance

 mn(n) = mean(S);

 vr(n) = var(S);

 %find gaussian distribution

 gaus = normpdf(S,mn(n),vr(n))

 %find actual distribution

 dist = hist(S, z);

 dist = dist/100000

 %find mean squarred error and store

 figure(1)

 plot(gaus, 'r');

 hold on

 plot(dist, 'b');

 end

% figure(4)

% plot(MSE);

end

%generates sum of RV

function S = SRV\_gen(n, N, min, max)

 %create random matrix of dimension n-by-N

 % n-> # of RV's (collumns)

 % N-> # of samples (rows)

 % range of 'min' to 'max'

 x = min + (max-min)\*randn(n, N);

 [z,q] = size(x);

 %sum the collumns

 for i = 1:z

 S(i) = sum(x(i, :));

 end

end

These two scripts of code show the code that was used to produce the plots and figures from above. It is important to note that the main function the produces the sum of the uniformly distributed variables is the SRV\_gen function. This function takes the number of data points, the numbers of samples generated, and the min and max to create the sum. This sum is then use multiple times, including when trying to produce the Gaussian. Lastly MSE function is used to create the mean squared error between the data.

# Conclusions

The main objective of this assignment was to observe the Central Limit Theorem, and to get a basic knowledge on how it is computed. In the first part of this assignment a function was created to sum mutually independent random variables. It turns out that as this is plotted with an increase in the number of random variables the plot got more and more Gaussian. This is seen in the first couple of figure above. The other objective that is plotted is the root mean squared between the Gaussian Fit of the distribution and the actual distribution. Similarly as the number of variables increased, that line became more and more uniform. That is the main purpose of the Central Limit Theorem. As the distribution of random variables increase, the curve becomes more like a normal distribution. It makes sense because the sum has multiple variables in it, including the mean, variance of the data. The last part of this computer assignment looks at the actual distribution over the Gaussian fit. Now this Gaussian fit function was not working and it kept producing a straight line. It should have looked more Gaussian like the actual data as the number of random variables increased. This was a good lab at looking the Central Limit Theorem. It activity increased the knowledge we have been looking at all year with the Gaussian distribution and the increase in variables/sizes. I look forward to applying it to future computer assignments.