Anton LeKang

ECE 3522: Stochastic Processes in Signals and Systems

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

CA5

# Problem Statement

The objective of this computer assignment is to build on the concept of error, and to observe what happens to different types of error from increasing the number of data point. The first part of this lab looks at the error of the estimate and the error of the variance. These errors look at the overall actual value minus the current mean and variance. If there are random data points, the mean and variance will change from variance the number of points. The second part of this computer assignment looks at the error between a true normal distribution, with the estimate of that distribution from random values. This again will create different results based on the number of data values. This computer assignment will use the random number generator on MATLAB to observe what happens with the error of a data set.

# Approach and Results

Part 1: Mean and Variance Estimate

The first part of the computer assignment loos at the error between the mean value, and the variance value. Figures 1 and 3 represent when there were 1,000 random variables. Figures 2 and 4 represent when there was 100,000 random variables. Although it was asked to look at 10,000,000 variables, this was not plotted because it took to long for the graph to come up, and it sounded like my computer was going to blow up. From looking at the two graphs, it is evident that the error decreases as the number of samples increases. When looking at figure 1 and 2, the error was far off from zero until more samples were added into the data. Similarly figures 3 and 4 look at the error between the variance of the data points. As the number of data points increased, the error go closer and closer to zero. This mean the values were getting close to the actual mean and variance values.

Figure 1 – Mean Estimate with 1,000 data values

Figure 2 – Mean Estimate With 100,000 data values

Figure 3 – Variance Estimate with 1,000 data values

Figure 4 – Variance Estimate with 100,000 data values

Part 2: Covariance

This next part of the computer assignment looked at the mean-squared error between a normal distribution, and an estimate of the distribution. From looking at this, an understanding can be seen from looking at data with certain bin values. The first three figures shown the true distribution, estimate, and the error for 100 data values.

Figure 6 – Estimated Distribution of 100 data values

Figure 5 –True Distribution of 100 data values

Figure 7 – Mean squared error between the true and estimated

As seen in the three figures above. The uniform distribution and the estimate distribution are fairly close to each other, but not exactly because the error ends up around 2.5. The next three figures show the true, estimated, and error for 10,000 data values.

Figure 9 – True Distribution of 10,000 data values

Figure 8 –True Distribution of 10,000 data values

 Figure 10 – Mean squared error between the true and estimated

As seen in the three figures above, as the number of samples is increased, the mean squared error before them decreases. Now the value is around 1.4. Also as the data values increase, the pdf of the estimate as seen in figure 8 start to get normal or flatter. This coincides well with the fact that the error is decreasing, since the plots are getting closer and closer to each other. The next three figures show the true, estimate, and error for when there are 10,000,000 data values.

 Figure 12 – Estimate Distribution with 10,000,000 data values

 Figure 11 – True Distribution with 10,000,000 data values

 Figure 13 – Mean squared error between the true and estimated

These last three figures shown the distribution when there are 10,000,000 data points. Again the uniform distribution gives a good uniform plot, but also the estimate distribution looks very similar to it. As the samples increase, the pdf of the estimate gets closer and closer to 1. Just like the last figures, the error also gets smaller. This time the error is averaged around .8, so as the samples increased, the error decreased.

# MATLAB Code

Part 1:

clc

clear all

 D = rand(1, 10e4);

 mn = mean(D);

 vr = var(D);

 z = length(D);

 for i = 1:z

 p = D(1:i);

 x = length(p);

 u = mean(p);

 s = var(p);

 %store errors in matricies

 Em(i) = (u-mn);

 Ev(i) = (s-vr);

end

figure(1)

%plot

semilogx(Em);

xlabel('Number of Points (n)');

ylabel('Error of mean estimate');

title('Mean stimate with 10e4 data points')

figure(2)

semilogx(Ev);

xlabel('Number of Points (n)');

ylabel('Error of varience estimate');

title('Variance stimate with 10e4 data points')

Part 2:

function CA\_6\_2

%generate array of random

F = rand(1, 10e6);

z = length(F);

%actual uniform distribution

xbins = (0:1/z:1);

y = length(xbins);

h = hist(F, xbins);

Preal = h;

%estimated uniform distribution

xbins2 = (0:.01:1);

y = length(xbins2);

h = hist(F, xbins2);

Pest = h/100000;

sum = 0;

sum2 = 0;

for j = 1:z

 sum = sum + Preal(j);

end

for j = 1:y

 sum2 = sum2 + Pest(j);

end

sum

sum2

q = length(Preal);

r = length(Pest);

%calculate mean squared error

figure(1)

plot(xbins2, Pest)

xlabel('Data Bins');

ylabel('Index');

title('Estimated Uniform Distribution')

figure(2)

plot(xbins, Preal)

xlabel('Data Bins');

ylabel('Index');

title('Actual Uniform Distribution')

for i = 1:101

 seg\_real = Pest(1:i);

 seg\_est = Preal(1:i);

 M(i) = mse(seg\_real, seg\_est);

end

M

figure(3)

semilogx(M);

xlabel('Number of Points (n)');

ylabel('Mean Squared Error');

title('Error between actual and estimated')

end

These are the scripts for each part of the computer assignment. In the first part of the assignment the mean error and variance error of a set of data points are looked at in comparison with the current mean and variance. The error is computed by taking the original mean and variance, and then subtracting the current value from it. Rand creates the sample of random variables which is the data used in this computer assignment. The second script shows the calculations for creating the mean-squared error from looking at the true, and estimated distributions. Again the distributions were created by using random values, and then a histogram is created to make a PDF. The estimated value is normalized by 100,000 because the largest data value going into it was 10,000,000. The true distribution has 100 bins, from 0 to 1 in increments of .01. The Sum each of the values is the computes, and then that values is then used to create the mean-squared error. This data is plotted, and displayed.

# Conclusions

Looking at the error between variables can be a very useful tool when observing different samples. There are many characteristics that play into the error, but a large determining factor is the size of the sample. As seen in this CA, the size of the data sample increases, the error between those data values and the actual data values decreases. As seen in the first part of the computer assignment, the error got closer to zero as the data points increased. When there was only 1,000 data points, the error was very choppy, this happened until more data points were inputted. In the second part of the lab, the mean-squared error was looked at by comparing the true normal distribution with an estimated distribution. From looking at figures 5-13, the error decreased as the sampled sized increase. At first the error was very uneven and the estimated distribution was not that normal. But then when there was 10,000,00 data values, the estimated distribution looked very normalized, and the actually distribution had a ton of data values within it, that’s why it was dark blue. This computer was very valuable in understand the importance of the sample size, and connecting it with the error of the data. I look forward to applying this understanding to future computer assignments.