Edward Moore

ECE 3512: Stochastic Processes in Signals and Systems

Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA 19121

# Problem Statement

In real life, often times mean and variance need to be calculated before you have access to the entire set of data. Here, we will estimate the mean and variance of a set of random numbers and compute the error of the estimates compared to the expected value if the data is completely uniform. The concept of mean-squared error (MSE) will also be explored. The MSE between the measured and expected uniform distribution will be calculated with the use of a histogram of randomly generated numbers.

# Approach and Results

Mean & Variance

First, the mean and variance were estimated, according to Equations 1 and 2, for an array of randomly generated numbers between 0 and 1. The Matlab representation of these equations is provided in the appendix. Once the estimates of the mean and variance have been calculated the error between the estimated values and the expected values can be calculated according to Equations 3 and 4, shown below. This requires no more than 1 additional line of code in the loop for each equation that calculated the previous values. Results of these calculations are shown below in Figure 1.

|  |  |
| --- | --- |
| Equation 1 | Equation 2 |
| Equation 3 | Equation 4 |

Since the plots of figure 1 represent calculations that accrue data, overtime the mean and error estimates all converge. Similarly, because the data is uniformly distributed the variance will see little to no change.

Mean-squared Error

Once these calculations have been completed, an estimate of the PDF is found by using the hist function of Matlab. One hundred equal sized bins are created between 0 and 1 and the random values are sorted into these bins. This is done for 101, 103 and 106 random numbers. The MSE is calculated based on the deviation of these bin sizes from the distribution of completely uniform data. Equation 5 describes how MSE is to be calculated for the bins. Matlab code for this can be found in the appendix.



Equation 5



Figure 1

As the number of random numbers generated for our sample set of data increases, the ability to make predictions based on that data also increases. This is illustrated in figures 2, 3 and 4. Increasing from 10 to 1000, and subsequently 1000 to 1,000,000, random numbers generated for the sample data sharply decreased the MSE value calculation results. The bins also began to appear flat across the histogram, which makes sense since the data is supposed to be uniformly distributed.



Figure 2



Figure 3



Figure 4

# Matlab code

Mean & Variance

With each iteration of the following code the avg11, erravg, variance11 and errvar are plotted against the index of the iteration.

for i = 1:numRand

 avg1 = avg1 + X(i);

 avg11 = avg1 / i;

 erravg = avg11-avg;

 variance1 = variance1 + ((X(i)-avg11).^2);

 variance11 = variance1 / i;

 errvar = variance11 - variance;

end

Mean-squared Error

This code estimates the MSE of the actual bin values to their estimated value. Here, the estimated value is .01 since we are drawing uniformly from the interval [0, 1] and the bin sizes are .01.

nRandNums = 10 %# of random #s to be generated

X = rand([nRandNums 1]);

nbins = 100;

[f,x] = hist(X, nbins);

subplot(2, 1, 1)

bar(x,f/sum(f));hold on %plots histogram

mse = 0;

subplot(2, 1, 2)

for i = 1:nbins

 mse = mse+(.01-f(i)./sum(f)).^2 %calculates mse

 semilogx(i, mse/(i), 'ro');hold on %plots mse

end

# Conclusions

As we draw more and more samples from a set of data we begin to get a much clearer picture of how the data behaves. When estimating the mean and variance it took approximately 100 samples for the data to begin to converge to the expected value. It was also shown that the mean-squared error decreases dramatically with sample size. These facts can put manufacturers of expensive items that require extensive safety testing in quite the predicament. This makes being only able to afford testing a few products off the production line while requiring a very small chance of error a delicate balancing act.