Marc Jurchak

ECE 3522: Stochastic Processes in Signals and Systems

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

# Problem Statement

In this assignment we see how many trials it takes before a dataset converges to its true average and variance. In the first problem we generate N random numbers between 1 and 0, creating a uniform distribution. We note that the average, or expected value of the distribution should be .5, and since the variance for a uniform distribution is shown below:

We expect the variance to be 1/12 or .083333. We generate N random variables and plot their average, variance and error of these measurements from numbers 1 to N. In the second problem we plot the pdf of this distribution and plot the error between this pdf and the expected value of the pdf.

# Approach and Results

For the first problem we begin by generating N = 1e6 random numbers between 0 and 1. Next we calculate the average value of these numbers from 1 to N using the following formula:

Next we plot the variance using the following formula:

We also plot the error between our calculated values and the expected values. We use the following formulae for error in average and error in variance:

We plot the average, variance, and error between these and their expected values in figures 1 and 2.

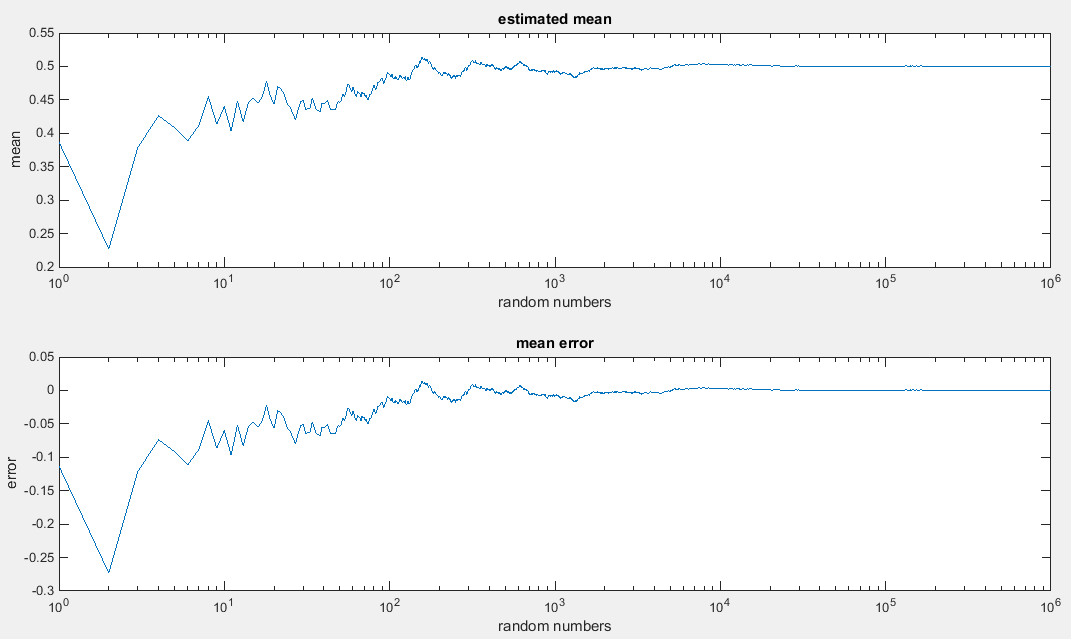


Figure : estimated mean and error

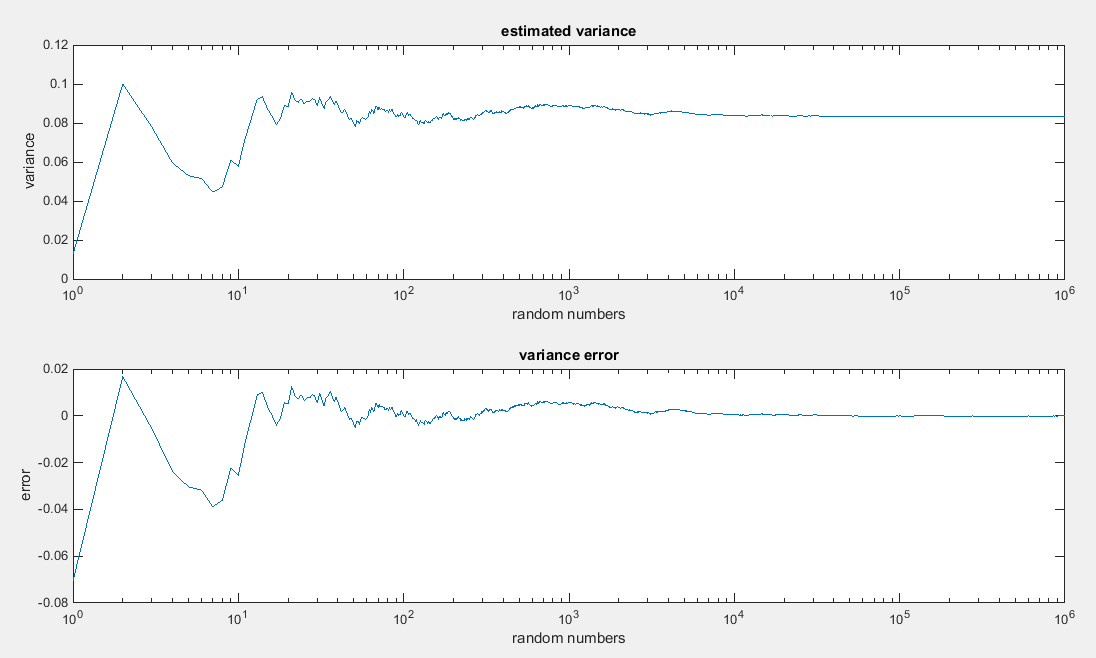


Figure : estimated variance and error

Next we construct a pdf from the data by making a histogram with bin size .01. After normalizing, we expect the probability of each bin to be .01. We use the following formula to plot the mean square error (MSE).

We plot the pdf and error of N = 10, 1000, 1e5. These plots are shown in figures 3-5.

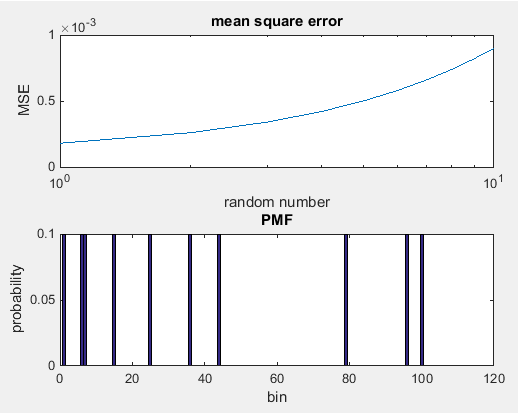


Figure : histogram and MSE when N = 10

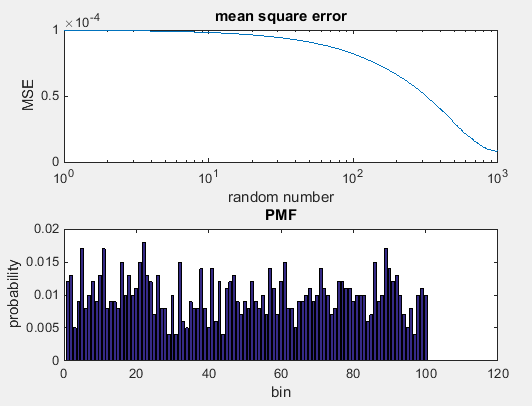


Figure : histogram and MSE when N = 1000

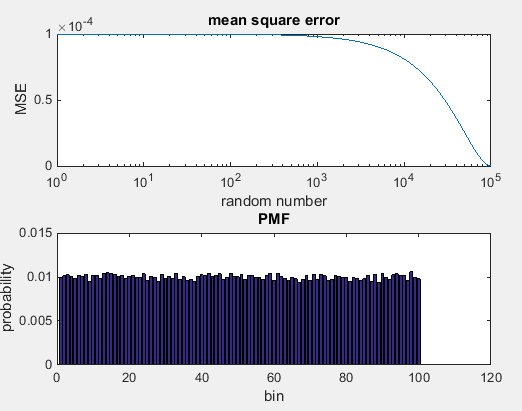


Figure :histogram and MSE when N = 1e5

# MATLAB Code

**Part 1**

clear;

clc;

[mean\_est, mean\_error, var\_est, var\_error] = uniform\_dist\_mean\_var(1e6);

figure(1)

subplot(2,1,1)

semilogx(mean\_est)

title('estimated mean')

ylabel('mean')

xlabel('random numbers')

subplot(2,1,2)

semilogx(var\_est)

title('mean error')

ylabel('error')

xlabel('random numbers')

figure(2)

subplot(2,1,1)

semilogx(mean\_error)

title('estimated variance')

ylabel('variance')

xlabel('random numbers')

subplot(2,1,2)

semilogx(var\_error)

title('variance error')

ylabel('error')

xlabel('random numbers')

**uniform\_dist\_mean\_var function**

function [estimated\_mean, estimated\_variance, error\_mean, error\_variance] = uniform\_dist\_mean\_var(N)

numbers = zeros(1,N);

for i = 1:N

numbers(i) = rand;

end

estimated\_variance = zeros(1,N);

total\_v = 0;

total\_var = zeros(1,N);

total\_var = numbers - .5;

estimated\_mean = zeros(1,N);

total\_mean = 0;

for i = 1:N

total\_mean = sum(numbers(1:i));

estimated\_mean(i) = (1/i) \* (total\_mean);

total\_mean = 0;

total\_v = sum(total\_var(1:i).^2);

estimated\_variance(i) = (1/i) \* total\_v;

total\_v = 0;

end

error\_mean = zeros(1,N);

error\_variance = zeros(1,N);

for i = 1:N

error\_mean(i) = estimated\_mean(i) - .5;

error\_variance(i) = estimated\_variance(i) - (1/12);

end

end

**Problem 2**

clear;

clc;

clf;

MSE = uniform\_pdf\_mse(10);

PMF = uniform\_pdf(10);

figure(1)

subplot(2,1,1)

semilogx(MSE)

title('mean square error')

ylabel('MSE')

xlabel('random number')

subplot(2,1,2)

bar(PMF)

title('PMF')

ylabel('probability')

xlabel('bin')

MSE = uniform\_pdf\_mse(1000);

PMF = uniform\_pdf(1000);

figure(2)

subplot(2,1,1)

semilogx(MSE)

title('mean square error')

ylabel('MSE')

xlabel('random number')

subplot(2,1,2)

bar(PMF)

title('PMF')

ylabel('probability')

xlabel('bin')

MSE = uniform\_pdf\_mse(100000);

PMF = uniform\_pdf(100000);

figure(3)

subplot(2,1,1)

semilogx(MSE)

title('mean square error')

ylabel('MSE')

xlabel('random number')

subplot(2,1,2)

bar(PMF)

title('PMF')

ylabel('probability')

xlabel('bin')

**uniform\_pdf function**

function [PMF] = uniform\_pdf(N)

numbers = zeros(1,N);

for i = 1:N

numbers(i) = rand;

end

B = 100;

histogram = histcounts(numbers,B);

PMF = histogram ./ length(numbers);

end

**uniform\_pdf\_mse function**

function [MSE] = uniform\_pdf\_mse(N)

numbers = zeros(1,N);

for i = 1:N

numbers(i) = rand;

end

B = 100;

MSE = zeros(1,N);

for j = 1:N

histogram = histcounts(numbers(1:j),B);

PMF = histogram ./ length(numbers);

ideal\_PMF = zeros(1,100);

ideal\_PMF(1:100) = .01;

difference = PMF - .01;

MSE(j) = (1/B)\*sum(difference.^2);

end

end

# Conclusions

One very important fact that we found from this experiment is that it takes many trials to find general information about the data. In any of the plots above, it took about 1e4 data points for the plots to “settle” to their expected values. This means that in real world tests, assuming they are fair, it takes many trials to show a faithful representation of probability. Another important fact is the early data can be very deceptive. For example in figure 1 it took 100 trials before the average even approached .5. Another thing we learned is that we never reach the “true” expected value, even after a million trials. Looking at any of the error plots shows that we can approach zero, but never truly get to it. Figure 3 shows a very important facet of histograms: With few trials, many of the bins are zero, which makes the error increase drastically. We see when N = 1000 and 1e5 that the error degrades to zero as the bins become closer to a straight line.