Claire Durand

ECE 3522: Stochastic Processes in Signals and Systems

Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA 19122

# Problem Statement

This assignment revisited the concepts of mean and variance in order to take a closer look at the implications of bin size and counts on histograms and their probability distribution functions (PDF’s). Varying numbers N of random numbers between 0 and 1 were to be generated. For each N, the mean, variance, and PDF estimated using the random numbers were compared to the expected mean, variance and distribution. The mean squared error was calculated in order to better compare the distributions.

# Approach and Results

A large part of the computations were performed within a for-loop that incremented the power of 10 applicable to the number of random numbers generated. The counter *i* for this for loop was used as the index for *N*, the number of random variables generated. *N* was used for the estimated mean and variance computation according to the following equations:

For each value of , and were compared to the expected mean and variance obtained using the following equations:

Figures 1 and 2 display the difference between the expected means and variances, and the estimated means and variances. Note how the error decreases as sample size increases. This is expected as the larger the sample size, the closer the estimated mean and variance should be to the expected mean and variance.

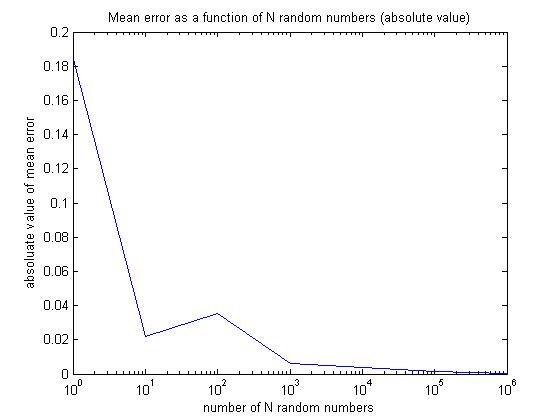


Figure - Absolute Mean Error as a function of N random numbers

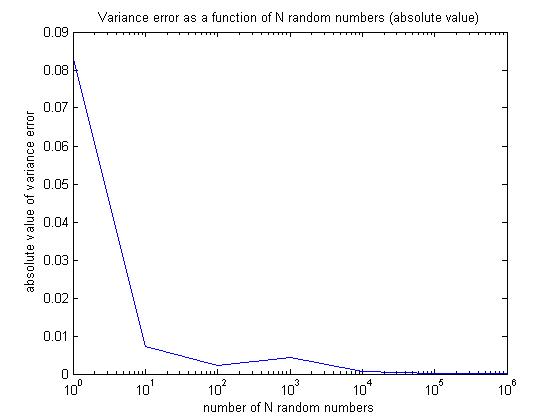


Figure - Absolute Variance Error as a function of N random numbers

In order to obtain the estimated pdf of the obtained random numbers, a normalized histogram was constructed with the *hist()* function in a fashion similar to that of Computer Assignment 4, using a bin size of 100. Knowing this bin size, and that the actual distribution should be a uniform distribution, the actual distribution should evenly divide probability between the 100 bins, creating an even probability of 0.01 for each bin. The mean squared error for the estimated pdfs versus the expected distribution is shown in Figure 3. Once again, note how the mse decreases as sample size increases.

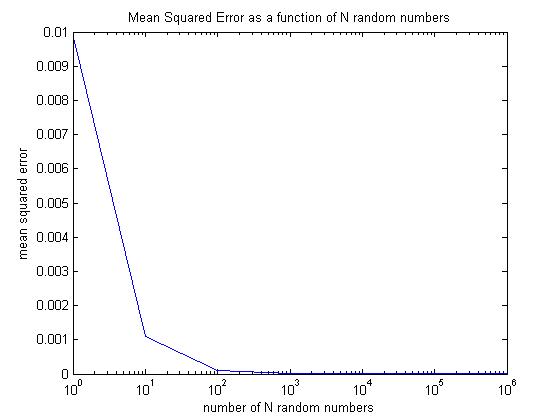


Figure - Mean Squared Error as a function of N random numbers

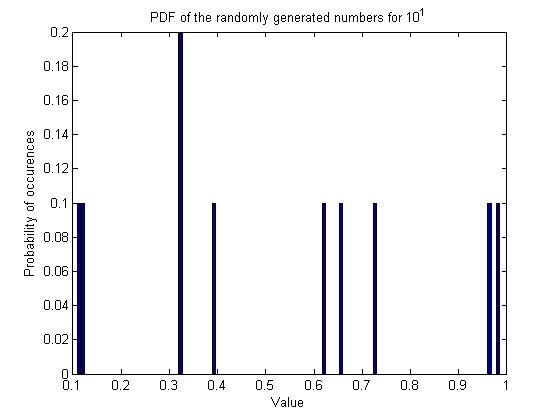


Figure - PDF of x for N = 10

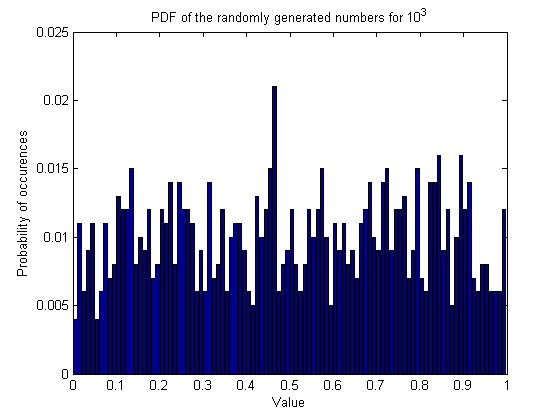


Figure - PDF of x for N = 103

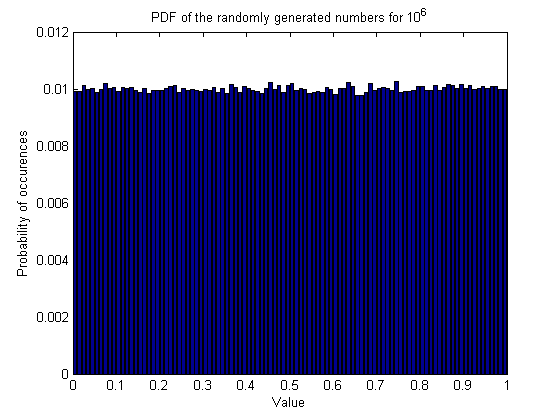


Figure - PDF of x for N = 106

Shown in Figures 4 through 6 are the PDFs of the randomly generated numbers. For N<B, there is blank space between the bars because there are not enough values in the sample space to fully populate the bins. However, once N>B, the bar graphs confirm the decrease in variance as N increases in that the counts for all of the bars approach 0.01 as N increases. This probability is to be expected since we have 100 bins, meaning that no matter their distribution, all of the numbers have to fit in anyways.

# MATLAB Code

## Computer Assignment 6 - Mean and Variance Revisited

clear; clc; clf;  
close all;

## Part 1 - Generating random numbers

% Define some parameters:  
mu = 0.5; % expected mean of the distribution  
var = 0.08333; % var = ((b-a)^2)/12 or the expected variance  
  
act\_pdf = ones(1,100) ./100; % flat distribution  
  
N = zeros(1,7);  
E\_u = zeros(1,7); % initialize mean error array  
E\_v = zeros(1,7); % initialize variance error array  
  
E\_u\_abs = zeros(1,7); % initialize abs mean error array  
E\_v\_abs = zeros(1,7); % initialize abs variance error array  
  
MSE = zeros(1,7); % initialize mse array  
  
for i = 1:7

N(i) = 10 ^ (i-1); % parametrize the number of random numbers  
  
 x = zeros(1,N(i)); % make an array for the signal  
 for n=1:N(i)  
 x(n) = rand; % Generate random numbers between 1 and N  
 end  
  
 % Estimate the expected value:  
 u = (1/N(i)) \* sum(x);  
  
 % Estimate the variance:  
 d = x - u; % difference between values and mean  
 d2 = d.^2; % square of the differences  
  
 v = (1/N(i)) \* sum(d2); % produce variance  
  
 E\_u(i) = u - mu; % obtain difference between mean and expected mean  
 E\_u\_abs(i) = abs(E\_u(i));  
 E\_v(i) = v - var; % obtain variance error  
 E\_v\_abs(i) = abs(E\_v(i));

## Estimate the PDF

% Specify number of bins (100 for range [0,1])  
 binnum = 100;  
  
 [counts, cntrs] = hist(x, binnum); % histogram of the data  
 counts\_norm = (counts)./N(i); % normalize hist using # of elements   
  
 % Plot the normalized histogram = PDF  
 figure(2+i)  
 bar(cntrs, counts\_norm);  
 xlabel('Value'); ylabel('Probability of occurences');  
 titlestr = sprintf('PDF of the randomly generated numbers for 10^%d',i-1);  
 title(titlestr);

## Compare to the actual distribution

% Find the MSE  
 diff = zeros(1,binnum); % make an array for the counts comparisons  
  
 for m=1:binnum  
 diff(m) = (act\_pdf(m)-counts\_norm(m))^2;  
 end  
  
 c = sum(diff);  
 MSE(i) = (c)/binnum;

end  
  
figure (10)  
semilogx(N,E\_u\_abs);  
xlabel('number of N random numbers'); ylabel('absoluate value of mean error');  
title('Mean error as a function of N random numbers (absolute value)');  
  
figure (11)  
semilogx(N,E\_v\_abs);  
xlabel('number of N random numbers'); ylabel('absolute value of variance error');  
title('Variance error as a function of N random numbers (absolute value)');  
  
figure(12)  
semilogx(N,MSE);  
xlabel('number of N random numbers'); ylabel('mean squared error');  
title('Mean Squared Error as a function of N random numbers');

[*Published with MATLAB® R2014a*](http://www.mathworks.com/products/matlab)

# Conclusions

As expected, as the number of samples generated increases, the mean squared error approaches zero. Additionally, as the number of generated samples increases, the probability distribution looks more like the expected distribution.