ENEL 529 - Lab 1

Kyle Derby MacInnis - 10053959

October 23, 2014

Abstract

The purpose of this lab is to perform two exercises in MATLAB so as to become familiar with the concepts and functions required for future laboratories.

Exercise 1 - Simulated/Theoretical Gaussian PDF

Exercise Objective

The objective of this exercise is to qualitatively compare the probability density function (pdf) of a Gaussian random variable X using a theoretical model, and a simulated model.

Procedure

1. Generate 100 samples using the built in MATLAB function:

```
MaxN = 100;                           % Number of Samples
uX = 0;                               % Guassian mean of X
stdX = sqrt(2);                       % std deviation of X

Data = normrnd(uX, stdX, 1, MaxN);    % Gaussian RND Distribution
```

2. Calculate the Sample Mean and Standard Deviation:

```
uS = mean(Data);                      % Calculate Mean of Distribution

stdS = std(Data);                     % Calculate Std Deviation of Distribution
```

3. Divide samples into K disjoint intervals:

```
K = 10;                               % # of Intervals

[m_j, C_j] = hist(Data, K);           % Split into 10 equal intervals
```

4. Calculate Δ, the length of each bin_j, assuming equal lengths:

```
delta_J = C_j(2) - C_j(1);            % Calculate Length of Intervals
```

5. Calculate the pdf of the simulated samples calculated by:

$$simulated\_pdf(C_j) = \frac{m_j}{Mx\Delta}$$

```
simulated_pdf = zeros(2, K);          % simulated pdf Container

%% Simulate pdf(C_j)
for i = 1:K
    simulated_pdf(1,i) = C_j(i);
    simulated_pdf(2,i) = m_j(i)/(MaxN*delta_J);
end
```

6. Calculate the theoretical Gaussian pdf using the sample mean and standard distribution:

$$theoretical\_pdf(C_j) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \text{EXP}\left(\frac{-(C_j - \mu_s)^2}{2\sigma_s^2}\right)$$

```
theoretical_pdf = zeros(2, K);       % theoretical pdf container

%% Theoretical pdf(C_j)
for i = 1:K
    theoretical_pdf(1,i) = C_j(i);
    theoretical_pdf(2,i) = (1/sqrt(2*pi*(stdS^2)))*exp((-(C_j(i)-uS)^2)/(2*(stdS^2)
    ));
end
```

7. Plot the two pdfs onto the same graph:

```
%% Plot Distributions (Theoretical and Simulated)
figure(1);
plot(simulated_pdf(1,:), simulated_pdf(2,:), 'r--o', theoretical_pdf(1,:),
    theoretical_pdf(2,:), 'g--+');
title('Simulated vs. Theoretical PDF for Gaussian Distribution');
xlabel('Bin Centre - C_j');
ylabel('PDF(C_j)');
```

Results

The following figures shows the initial distribution of 100 samples as explained above, as well as a distribution of 10000 samples:

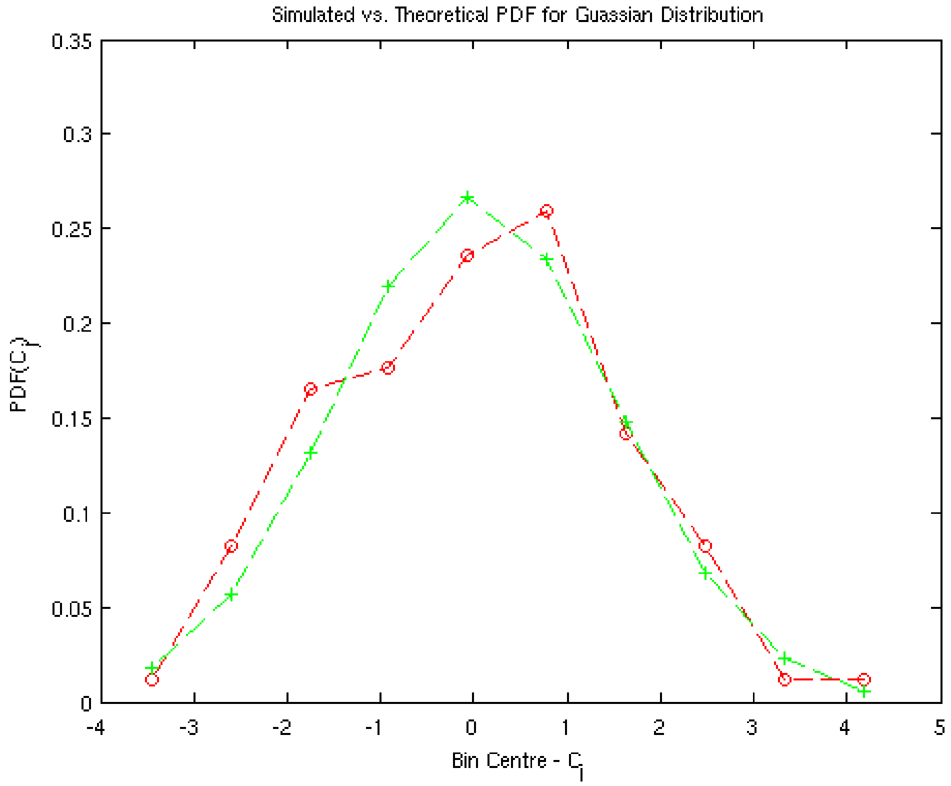

Figure 1: The Comparison between the theoretical and simulated distributions using 100 Samples.

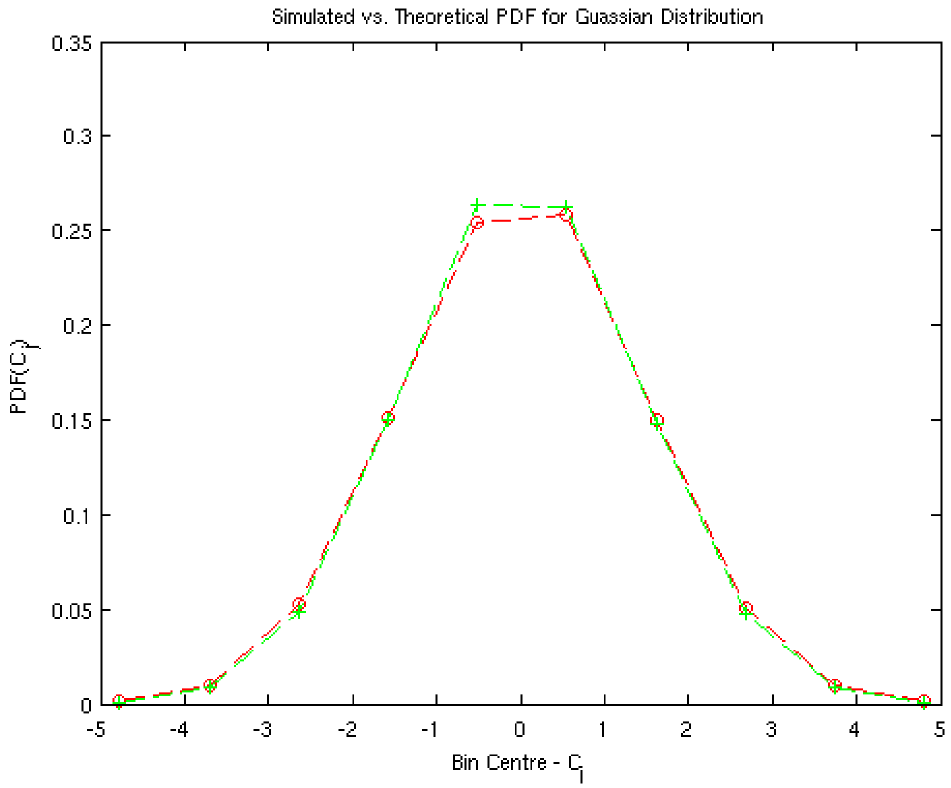

Figure 2: The Comparison between the theoretical and simulated distributions using 10000 Samples.

Exercise 2 - Chi-Square Analysis

Exercise Objective

The objective of this exercise is to perform a quantitative analysis using the Chi-Squared goodness-of-fit test to compare the simulated model with the theoretical model.

Procedure

1. Generate a Gaussian random distribution of 100 samples. Please see Exercise 1 - Step 1 for details.
2. Calculate Sample mean and Standard deviation. Please see Exercise 1 - Step 2 for details.
3. Divide sample distribution into K equal intervals. Please see Exercise 1 - Step 3 for details.
4. Calculate the theoretical number of samples that fall into each interval.
   - Calculate Δ. See Exercise 1 - Step 4 for details.
   - Calculate L_j, the Lower limit of the interval:

     $$L_j = C_j - \frac{\Delta}{2}$$

   - Calculate U_j, the Upper limit of the interval:

     $$U_j = C_j + \frac{\Delta}{2}$$

   - Calculate p_j, the probability that a sample falls into a particulate interval, assuming Gaussian distribution with pdf, $f_X(x)$:

     $$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_s^2}}\text{EXP}\left(\frac{-(x-\mu_s)^2}{2\sigma_s^2}\right)$$

     where

     $$p_j = Q\left(\frac{L_j - \mu_s}{\sigma_s}\right) - Q\left(\frac{U_j - \mu_s}{\sigma_s}\right)$$

   - Calculate $T_j = Mp_j$:

```
L_j2 = zeros(1, K2);                    % Lower Bound of Intervals
U_j2 = zeros(1, K2);                    % Upper Bound of Intervals
p_j2 = zeros(1, K2);                    % Probability of Intervals
T_j2 = zeros(1, K2);                    % # of Samples in Intervals

%% Calculate Lower and Upper Bounds
for i=1:K2
    L_j2(1,i) = C_j2(1,i) - (delta_j2/2);     % Calculate Lower Bound
    U_j2(1,i) = C_j2(1,i) + (delta_j2/2);     % Calculate Upper Bound
end

%% Calculate Probability and # of Samples
for i=1:K2
    p_j2(1,i) = qfunc((L_j2(1,i)-uS2)/stdS2)-qfunc((U_j2(1,i)-uS2)/stdS2);
    T_j2(1,i) = MaxN2*p_j2(1,i);
end
```

5. Calculate the Chi-Square statistic Z_1:

   $$Z_1 = \sum_{j=1}^{K}\frac{(m_j - T_j)^2}{T_j}$$

```
Z1 = 0;                                   % Initialize (Chi-Square) Z1 to 0

%% Calculate Chi-Square Value
for i=1:K2
    Z1 = Z1 + ((m_j2(1,i) - T_j2(1,i))^2/T_j2(1,i));
end
```

6. Compare Z_1 with Z_T. If $Z_1 < Z_T$ then test passes:

```
Zt = 14.07;                               % Threshold Chi-Square Value from Tables

%% Compare Chi-Square Value to Test Distribution
if Z1 < Zt
   display(Z1);
   display(Zt);
   display('The Test Passed and the Distribution is Gaussian!');
else
   display(Z1);
   display(Zt);
    display('The Test did not pass, the distribution is not Gaussian.');
end
```

Results

Upon completion and execution of the code, the test passed verification after one run. The sample distribution is indeed, for all intents-and-purposes, Gaussian.

Additional Questions and Exercises

Prove the following equation:

$$p_j = Q\left(\frac{L_j - \mu_s}{\sigma_s}\right) - Q\left(\frac{U_j - \mu_s}{\sigma_s}\right)$$

Proof:

1. $Q(x)$ gives the probability that a value will be greater than x standard deviations above the mean and is defined as:

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} \text{EXP}\left(-\frac{u^2}{2}\right) du$$

$$Q(x) = 1 - Q(-x) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \text{EXP}\left(-\frac{u^2}{2}\right) du$$

2. The probability that a value will be greater than the lower limit L_j standard deviations is:

$$Q\left(\frac{L_j - \mu_s}{\sigma_s}\right) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{L_j - \mu_s}{\sigma_s}} \text{EXP}\left(-\frac{u^2}{2}\right) du$$

3. The probability that a value will be greater than the upper limit U_j is given by:

$$Q\left(\frac{U_j - \mu_s}{\sigma_s}\right) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{U_j - \mu_s}{\sigma_s}} \text{EXP}\left(-\frac{u^2}{2}\right) du$$

4. Since, both are cumulative from $-\infty$ then the probability difference, in respect to $-\infty$ would be:

$$Q\left(\frac{L_j - \mu_s}{\sigma_s}\right) - Q\left(\frac{U_j - \mu_s}{\sigma_s}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{U_j - \mu_s}{\sigma_s}} \text{EXP}\left(-\frac{u^2}{2}\right) du - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{L_j - \mu_s}{\sigma_s}} \text{EXP}\left(-\frac{u^2}{2}\right) du$$

5. Simplifying the equation looks like:

$$Q\left(\frac{L_j-\mu_s}{\sigma_s}\right)-Q\left(\frac{U_j-\mu_s}{\sigma_s}\right)=\frac{1}{\sqrt{2\pi}}\int_{\frac{L_j-\mu_s}{\sigma_s}}^{\frac{U_j-\mu_s}{\sigma_s}}\text{EXP}\left(-\frac{u^2}{2}\right)du$$

6. Since the integral is the cumulative distibution function (cdf) from the lower limit to the upper limit, the result is the probability that a value would be found in that range, hence:

$$p_j(x)=P[L_j\leq x\leq U_j]=\frac{1}{\sqrt{2\pi}}\int_{\frac{L_j-\mu_s}{\sigma_s}}^{\frac{U_j-\mu_s}{\sigma_s}}\text{EXP}\left(-\frac{u^2}{2}\right)du=Q\left(\frac{L_j-\mu_s}{\sigma_s}\right)-Q\left(\frac{U_j-\mu_s}{\sigma_s}\right)$$

Remarks on the Lab

The lab provided a nice introduction to the basic techniques necessary for the remainder of this course. It acted a great refresher to MATLAB, and showcased a good amount of basic functionality. The lab also helped to refresh on the topics of both the Chi-Square Goodness-of-fit test, and the ideas of Gaussian distributions. It was not too difficult, and most of the lab was pretty straight forward. The hardest part of the lab was the proof.