# The University of Calgary
# Department of Electrical & Computer Engineering

## <u>ENEL 529 Wireless Communications Systems</u>
## <u>Lab 1: Familiarity with MATLAB Functions Required for Lab 2</u>

Lab Day & Date: Monday, October 6, 2014 (3:00 – 4:50 pm)

Lab Report due Date: Tuesday, October 14, 2014 @ 12:00 noon.

## 1. Objective

The objective of this lab is to:

- Perform 2 exercises to become familiar with the MATLAB concepts and functions required for Lab 2.

## 2. Pre-Requisites

This document assumes that you already have a basic working knowledge of MATLAB. Otherwise, please work through the online MATLAB Tutorial.

## 3. Procedure

3.0. Log in to the account of the team leader. Create "ENEL529Lab" directory. Then create a sub-directory called "Lab1".
→ Launch MATLAB
→ Change to "Lab1" directory, where the file(s) created during this Lab will be stored. To do this,
→ click on the "Browse for Folder" icon (this is the button represented by three dots ("…") besides the folder icon on the top right corner. A "Browse for Folder" box appears
→ navigate to the "Lab1" folder.

3.1 <u>*Exercise 1 Objective:*</u> Simulation of samples of a Gaussian distributed random variable X and comparison of its theoretical vs. simulated probability density functions (pdf's).

<u>*Step 1:*</u> Generate 100 samples (i.e. total # of samples $M = 100$) of X ~ Gaussian(mean = 0, standard deviation = $\sqrt{2}$ ). <u>***Hint:***</u> 1) In MATLAB, samples of a Gaussian distributed random variable can be generated using the built-in "normrnd()" function. For example, to generate a vector called "data" of length $M$ samples of X ~ Gaussian(mean = mu, standard deviation = sigma), the MATLAB syntax is: data = normrnd(mu, sigma, 1, $M$).

<u>*Step 2:*</u> Calculate the sample mean and sample standard deviation of vector "data", denoted by $\mu_s$ and $\sigma_s$, respectively. <u>***Hint:***</u> The MATLAB built-in functions for calculating the mean and standard deviation are "mean()" and "std()", respectively,

*Step 3:* Divide the simulated samples (vector "data") into K disjoint intervals (i.e. K bins). In this exercise, assume K = 10. ***Hint:*** Use the MATLAB built-in "hist()" function. ***Note:*** The outputs of calling the "hist()" function are: i) the actual number of samples in each bin, and ii) the center of each bin. Mathematically, denote the actual number of samples in bin *j* by $m_j$, *j* = 1, 2, 3, …, K. Similarly, let $C_j$ denote the center of bin *j*, *j* = 1, 2, 3, …, K.

*Step 4:* Calculate $\Delta$, the length of each bin. Assume equal length for all the K bins so that $\Delta = C_2 - C_1 = C_3 - C_2 = \cdots = C_K - C_{K-1}$.

*Step 5:* Calculate the pdf of the simulated samples (vector "data"), referred to as simulated pdf. Let $simulated\_pdf(C_j)$ denote the value of simulated pdf at $C_j$, calculated by:

$$simulated\_pdf(C_j) = \frac{m_j}{M \text{ x } \Delta} \tag{1}$$

where *j* = 1, 2, 3, …, K.

*Step 6:* Compute the *theoretical* Gaussian pdf with parameters $\mu_s$ and $\sigma_s$ (determined in Step 2) using the formula:

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

where *j* = 1, 2, 3, …, K.

*Step 7:* Plot on the same graph the *theoretical* Gaussian pdf (i.e. result of Step 6) and the *simulated* pdf (i.e. result of Step 5). That is, superpose a plot of the *theoretical_pdf* vs. bin center on the plot of *simulated_pdf* vs. bin center. Your results should appear as shown in Fig. 1. ***Note:*** Your *simulated_pdf* result will not be identical to that shown in Fig. 1 because simulations are driven by different random seeds. **(3 points)**

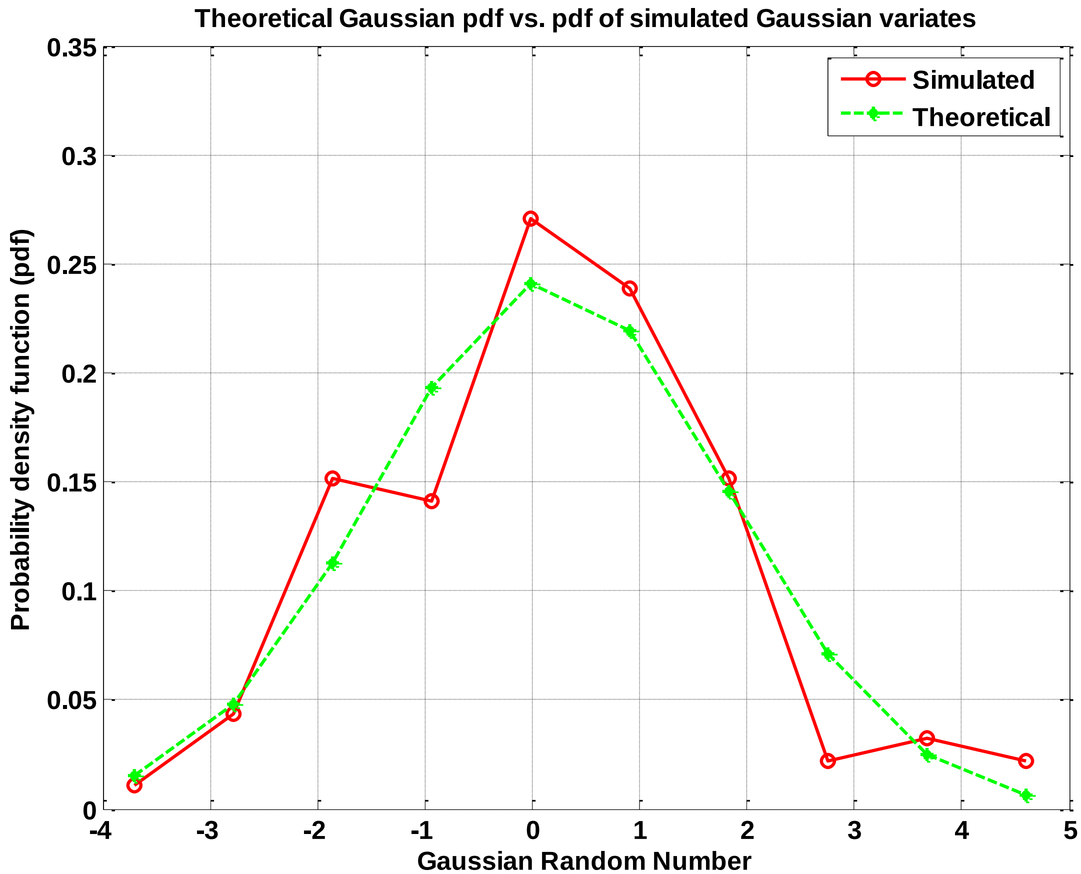


Fig. 1. Theoretical Gaussian pdf vs. pdf of simulated Gaussian random variates (100 simulated samples)

Based on visual observation of the curves in Fig. 1, the fit of the simulated pdf to the theoretical Gaussian pdf is not that good. One way to improve the fit is to increase the sample size.

*Step 8:* Repeat Step 1 to Step 7 but now change the total number of samples in Step 1 to 10,000. Your result should appear as shown in Fig. 2, which clearly shows a better fit at 10,000 samples than at 100. **(2 points)**

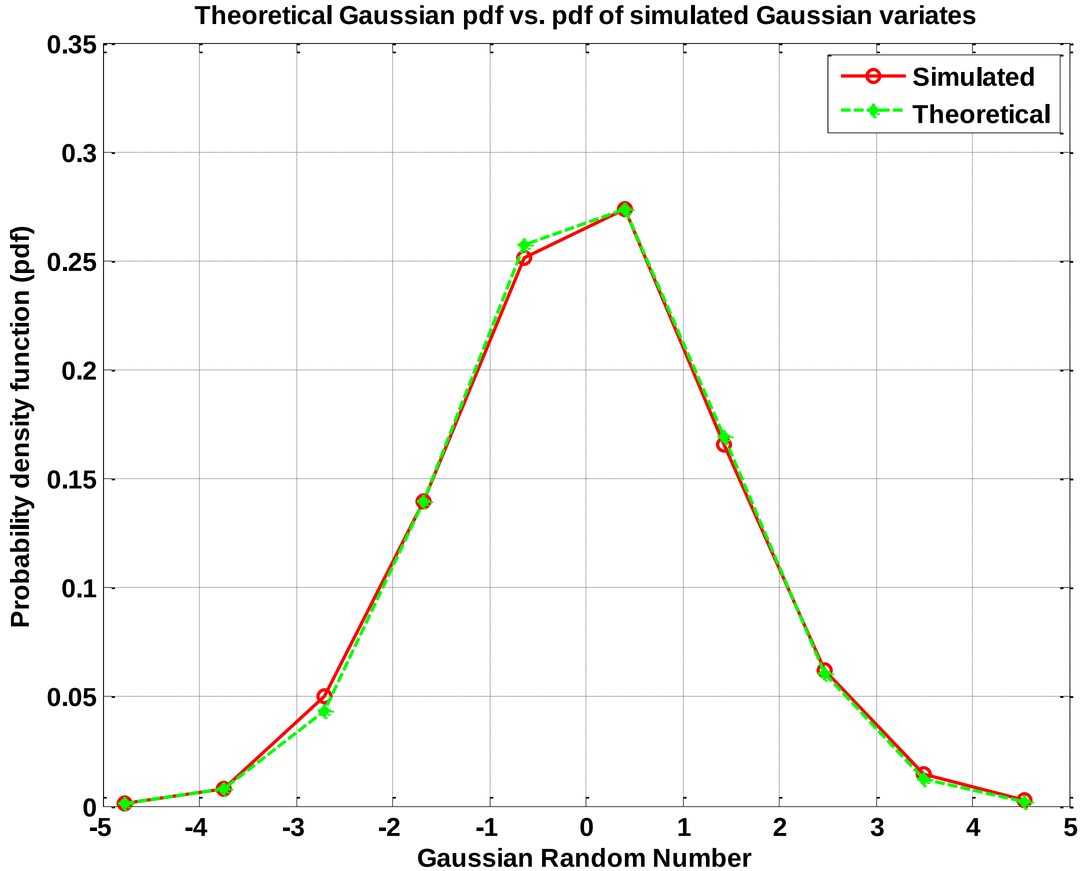


Fig. 2. Theoretical Gaussian pdf vs. pdf of simulated Gaussian random variates (10,000 simulated samples)

3.2 *Exercise 2 Objective:* Quantitative goodness-of-fit test using the Chi-square goodness-of-fit test.

Aside from visual comparison (subjective test) used in Exercise 1, it is possible to conduct a quantitative test to verify whether the pdf of the simulated samples actually follows a given theoretical pdf. One such test is the Chi-square goodness-of-fit test. The Chi-square test is a non-parametric (i.e. results are not dependent on the specific shape or parameters of the distribution) means of testing hypotheses.

The steps required for Chi-square testing of simulated Gaussian samples are outlined as follows:

*Step 1:* Generate M = 10,000 samples of X ~ Gaussian(mean = 0, standard deviation = $\sqrt{2}$). Store the samples in a vector called "data2". ***Hint:*** Use the appropriate MATLAB built-in function specified in Section 3.1.

*Step 2:* Calculate the mean and standard deviation of the simulated samples, denoted by $\mu_s$ and $\sigma_s$, respectively. ***Hint:*** Use the appropriate MATLAB built-in functions specified in Section 3.1.

*<u>Step 3:</u>* Divide the simulated samples (vector "data2") into K disjoint intervals (i.e. K bins). Assume K = 10 in this exercise. ***<u>Hint:</u>*** Use the MATLAB built-in "hist()" function. ***<u>Note:</u>*** The outputs of calling the "hist()" function are: i) the actual number of samples in each bin, and ii) the center of each bin. Mathematically, denote the actual number of samples in bin *j* by $m_j$, *j* = 1, 2, 3, …, K. Similarly, let $C_j$ denote the center of bin *j*, *j* = 1, 2, 3, …, K.

*<u>Step 4:</u>* Calculate the theoretical expected number of samples that fall in each bin, under the assumption that the samples follow the Gaussian distribution. To do this, proceed as follows:

i) Calculate $\Delta$, the length of each bin. Assume equal length for all the K bins so that $\Delta = C_2 - C_1 = C_3 - C_2 = \cdots = C_K - C_{K-1}$.

ii) Calculate $L_j$, the lower limit of bin *j*. Mathematically, $L_j = C_j - (\Delta/2)$.

iii) Calculate $U_j$, the upper limit of bin *j*. Mathematically, $U_j = C_j + (\Delta/2)$.

iv) Calculate $p_j$, the theoretical probability that a sample of "data2" falls in bin *j*, under the assumption that the data samples follow the Gaussian distribution with parameters $\mu_s$ and $\sigma_s$ (determined in Step 2). Let $f_X(x)$ denote the pdf of a Gaussian distributed random variable X given by:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_s^2}} \exp\left( \frac{-(x-\mu_s)^2}{2\sigma_s^2} \right) \tag{3}$$

By definition:

$$p_j = \Pr\left\{ L_j < X < U_j \right\} = \int_{L_j}^{U_j} f_X(x)dx \tag{4}$$

Substituting the expression for $f_X(x)$ (eqn. (3)) in (4) and integrating gives:

$$p_j = Q\left( \frac{L_j - \mu_s}{\sigma_s} \right) - Q\left( \frac{U_j - \mu_s}{\sigma_s} \right) \tag{5}$$

where *Q(.)* is the Q-function. ***<u>Hint:</u>*** The MATLAB built-in "qfunc(.)" function can be used to calculate *Q(.)* in (5).

v) Finally calculate $T_j$, the theoretical expected number of samples that fall in bin *j*. Mathematically, $T_j = Mp_j$, *j* = 1, 2, …, K. Recall that *M* = 10,000 in this exercise.

*<u>Step 5:</u>* Calculate the Chi-square statistic $Z_1$ using the formula [1]:

$$Z_1 = \sum_{j=1}^{K} \frac{(m_j - T_j)^2}{T_j} \tag{6}$$

*Step 6:* From Chi-square tables [1], select $Z_T$, the Chi-square threshold statistic corresponding to the value of degrees of freedom and the assumed significance level of the test. In general, the threshold statistic is selected at the specified significance level with K-*r*-1 degrees of freedom, where K is the number of bins and *r* is the number of parameters extracted from the data.

In this exercise, assume a significance level of 5% and K = 10 (from Step 3). For the Gaussian distribution, $r = 2$ (i.e., the parameters $\mu_s$ and $\sigma_s$ calculated in Step 2) and hence there are 7 (=10-2-1) degrees of freedom. At a significance level of 5% and 7 degrees of freedom, $Z_T$ = 14.07 from the Chi-square tables [1]. Softcopy of Chi-square Table is under the Lab 1 folder.

*Step 7:* Compare $Z_1$ with $Z_T$. The Chi-square test passes if $Z_1 < Z_T$ and it is concluded that the simulated data actually follows the postulated Gaussian distribution. Did the test pass? **(2 points)**

***Note:*** The test may not pass on the first run due to the fact that the samples were simulated (i.e. randomly generated). If the test does not pass, repeat Step 1 to Step 7 until the first time the test passes.

***Additional Questions and Exercises***

1.1 Prove Equation (5). **(3 points)**

Hint 1: $\int_{L_j}^{U_j} f_X(x)dx = \int_{-\infty}^{U_j} f_X(x)dx - \int_{-\infty}^{L_j} f_X(x)dx$

Hint 2: $\int_{-\infty}^{U_j} f_X(x)dx + \int_{U_j}^{\infty} f_X(x)dx = 1.0$ (the identity also holds for $L_j$)

Hint 3: Use integration by parts.

## 4. Further Readings

[1] P.M. Shankar, Introduction to Wireless Systems, Chapter 2, section 2.5.

## 5. Lab Report

Prepare a Lab Report (one per group) using ENEL529 Lab Reporting Guidelines.

**IMPORTANT NOTE:** Lab 2 involves the calculation of pdfs and verifying the goodness-of-fit test of Rayleigh distribution to the pdf of *measured* data. Both the visual and quantitative tests are required for the verification. **Hence, your understanding of the MATLAB concepts used in Lab 1 is paramount for successful completion of Lab 2.**