ECEN321 Notes on Lab 2 Probability and Probability Densities

This document is to provide you with extra guidance and explanation for Lab 2, so read it together with the lab-material for Lab 2.

For Labs with marking required, you need to submit an individual report, explaining what you have done and achieved in the Lab. These include, explaining your working steps, show your Matlab results, and your Matlab codes. Remember to plot your graph(s) properly by providing a title, labelling the axis, and showing legends (i.e. what is what on your plot). Also remember to provide commentary for your Matlab coding, an example of which is provided below for your reference:

```
% This is to show you how to provide commentary for your Matlab coding.
% You will type % and followed by the comments on the right.
% An example is shown here for you:

popMean=2.5;     % this is the population mean
stdDev=4;        % this is the population standard deviation

```

Also remember to check out the marking rubric and breakdown of marks for Lab reports. The reports need not be formal reports; see marking rubric for what need to be included, and a sample report format used in ECEN220 Signals and Systems.

Part 1 Normal Distribution

For this part, we start by assuming that we have and we know the population mean = 2.5 and variance = 16. With these population parameters, we want to generate N = 1000 sample points, and use this as our sample. From these sample points, we can calculate the sample statistics, i.e. sample mean, variance and standard deviation.

We want to create a histogram of these 1000 sample points. Hint: use bar( ). Read the Lab Sheets for information on how to generate the histogram.

We also want to generate the theoretical pdf, using the equation given in the Lab Sheets, for both the population and sample, so that we can compare if they are similar to each other. Plot the histogram, population pdf and sample pdf all onto the same graph, as suggested in the Lab Sheets.

Then we want to plot the cdf of the population and sample. There are 2 ways to do this. Firstly, you may follow the instructions in the Lab Sheets. Secondly, you may use this equation for cdf

$$\phi\left(\frac{x-\mu}{\sigma}\right)=\frac{1}{2}\left[1+\operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

Where

$$\operatorname{erf}(x)=\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,dt$$

In Matlab, you can use the command erf( ). Look at the resulting plots, and comment on them.

Part 2 Student t Distribution

For this part we still use the same population mean and variance from Part 1.

Here, we generate a matrix of 5 by N of random variables where N =10,000. Then for each column of the matrix (i.e. 5 sample points), we generate a t-value by using the formula explained in the Lab Sheets (Read the sentences to write the equation!). Once this step is completed, we will have N numbers of t-value.

Then we generate a histogram of these 10,000 t-value, and compare it with the theoretical (empirical) pdf. (How to plot histogram? Well, you have already learnt it from Part 1) Again, there are 2 ways to generate the theoretical pdf. Firstly, you may use the equation provided in the Lab Sheets, or secondly you may ask Matlab to do it for you by using the built-in function tpdf( ). Use help tpdf( ) if you need it.

Next, we want to plot the cdf of the t-values. We use the 10,000 t-values already generated. To plot cdf using the data, you need to sort the data; see sort( ). After sorting, you need to find, or create the rank of the data; this is to say that we want to create a linear spacing of 10,000 between 0 and 1 (inclusive), so that later you can plot the sorted values vs rank. Compare this with the theoretical cdf generated by Matlab, by using tcdf( ).

Part 3 Probability and the Birthday Paradox

The purpose of this part 3 is to show you how to use simulation to investigate a problem related to probability/statistics.

There are three ways of conducting any investigations 1) Analytical investigation where mathematical derivations are used to arrive at the theoretical result, 2) Simulation where algorithm using Matlab is written to simulate the proposed scenario so that we can arrive at the simulated result, 3) Conducting actual sampling and survey where we will arrive at the actual result.

If conducted properly, all these results should agree very closely with one another.

Read and follow the lab-material where you will be doing the parts of Analytical and Simulation. If you have time, you may want to take one step further and conduct your own actual sampling and survey, eg. by using the birthday data from our 2 classes.