ECE-597 Probability, Random Processes, and Estimation
Computer Assignment #1

Due: Friday March 20, 2015

The Matlab files on the class website approximate the **cdf** (gauss_cdf.m) and **pdf** (gauss_pdf.m)of a Gaussian density with a mean of μ and a variance of σ.

The empirical formula used to approximate the cdf is

$$F_X(x|x_1, x_2, x_3, \ldots, x_n) = \frac{\text{number of samples } x_1, x_2, x_3, \ldots, x_n \text{ no greater than } x}{n}$$

while the empirical formula used to approximate the pdf is

$$f_X(x)\Delta x = P(x < X \le x + \Delta x) \approx \frac{\text{number of data values in } (x, x + \Delta x]}{\text{total number of data values}}, \Delta x << 1$$

In addition, the true cdf and pdf are plotted on the same graph so you can compare the estimates from the true values. The inputs to both files are **Npoints** (the number of sample points to generate), **Nbins** (the number of bins, which controls the size of Δx), μ (the mean), and σ (the standard deviation.) You should study these pieces of code and try to determine how they work. You are responsible for this!

Note: *In what follows you will find that the estimated cdf's are smoother than the estimated pdf's. This is because the pdf is the derivative of the pdf, so what looks like a small error in the cdf will be amplified when the pdf is computed.*

1) Run the m-files as they are for various values of **Npoints** and **Nbins** assuming $\mu = 0$ and $\sigma = 1$. For large values of **Npoints**, approximately 100,000 and moderate values of **Nbins**, approximately 100, there should be a fairly good match between the estimated and analytical functions. You should turn at least three pairs (one pdf and one cdf) of **neatly labeled** plots for this part.

2) Run the simulations for **Npoints** = 10,000 and **Nbins** = 100 for the cases of (**m**,**sigma**) = (0,1), (5,2), and (3,3). You should be able to verify the correct ranges of values for these plots since you know that it is very unlikely that any points will lie outside three standard deviations from the mean, and the mean should be at the center of the plot.

3) Modify both sections of code to work with a uniform (on [0,1]) density (lookup the Matlab command **rand**.) You need estimates of both the pdf and cdf here for a range of the values of **Npoints** and **Nbins**. You only need two arguments for this piece of code. You should turn at least three pairs of **neatly labeled** plots for this part. You may want to change the scale on these plots, since they will probably look very bad with lots of oscillations, but the oscillations are generally not very large. They just look large when Matlab chooses the axes.

4) Starting from a unform random variable **X** generate an exponential random variable **Y**. The pdf and cdf for an exponential random variable is

$$
\begin{aligned}
f_Y(y) &= \frac{1}{\mu} e^{-y/\mu} u(y) \\
F_Y(y) &= (1 - e^{-y/\mu}) u(y)
\end{aligned}
$$

Plot the true and estimated pdfs. Your m-files functions should have three arguments, (**Npoints**, **Nbins**, μ), where **Npoints** is the number of samples to generate, **Nbins** is the number of sample bins, and μ is the parameter in the exponential density. Your routine should both generate the exponential random variables and determine both the estimated and true pdfs and plot them. You should get very good agreement between your estimate and the true values if **Npoints** is large enough. Turn in at least three different plots for **Npoints = 1e5**, **Nbins = 50** and $\mu = 0.1, 1, 10$.

5) Starting from a unform random variable **X** generate a Rayleigh random variable **Y**. The pdf and cdf for a Rayleigh random variable is

$$
\begin{aligned}
f_Y(y) &= \frac{y}{\sigma^2} e^{-y^2/2\sigma^2} u(y) \\
F_Y(y) &= (1 - e^{-y^2/2\sigma^2}) u(y)
\end{aligned}
$$

Plot the true and estimated pdfs. Your m-files functions should have three arguments, (**Npoints**, **Nbins**, σ), where **Npoints** is the number of samples to generate, **Nbins** is the number of sample bins, and σ is the parameter in the Rayleigh density. Your routine should both generate the Rayleigh random variables and determine both the estimated and true pdfs and plot them. You should get very good agreement between your estimate and the true values if **Npoints** is large enough. Turn in at least three different plots for **Npoints = 1e5**, **Nbins = 50** and $\sigma = 0.1, 1, 10$.