CS177, Homework 7
Due Date: Wednesday, May 28th

Please hand in a hardcopy of your homework at the beginning of class. Also upload your MATLAB functions to the appropriate EEE folder

Reading

- Olofsson Sections 2.6-2.7 on the exponential and Gaussian (normal) pdfs
- Olofsson Section 3.12 and your notes on the Poisson process
- Olofsson Section 5.4 on simulation methods

Problem 1: Exponentially Distributed Arrivals

We have a switch on a network where packets arrive in an unpredictable manner. Assume that the time T between packet arrivals is an exponential random variable with parameter λ.

1. If $\lambda = 100$ packets per microsecond (10^{-6} s), what is the probability that the time-interval T between packets will be less than 2 microseconds?
2. If $\lambda = 100$ packets per microsecond, what is the probability that the time-interval between packets will be greater than 1 millisecond (10^{-3} s)?
3. Let p be the probability that the gap in time between any pair of packets is greater than 10 microseconds. What value of λ do we need so that $p = 0.01$?

Problem 2: Disk Failures

On the class website you will find a link to a data file called `diskfail.dat`. This contains real data on the times of failures (measured in days) of disks attached to an IBM RS/6000[1] computer system called Seaborg which was operated by Lawrence Berkeley National Labs from 2001-2007. We would like to judge if this event data is well modeled by a Poisson process.

[1] A modified RS/6000 named Deep Blue was the first computer to beat world champion Garry Kasparov at chess

1. Load in the data using `eventtimes = load('diskfail.dat')` and compute the inter-arrival times (time intervals between successive events). Use the `hist` function to estimate the empirical distribution. Also compute the empirical mean inter-arrival time. Plot both the empirical distribution and an exponential distribution with parameter λ on the same plot. Choose the value for the parameter λ so that the exponential distribution has the same expectation (mean) as the empirical data. How similar is the empirical data to the exponential model of inter-arrival times?

   Hints: Calling `[n,x] = hist(data,nbins)` will give you the empirical counts `n` for bins centered at locations given by the vector `x`. To convert `n` into an estimate of probability, you will need to divide by the total number of data points before plotting it. You will need to experiment with the number of bins in order to get a nice estimate of the distribution. Plot the exponential distribution evaluated at the same set of locations x.

2. Compute the empirical probability mass function for the number of events inside an interval of w = 10 days. You can compute this by looping over different offsets and counting the number of events that fall within 10 days of that offset. Compute the mean number of counts and plot a histogram of the distribution of counts. If the events follow a Poisson process, what distribution would you expect these counts to have? Based on your estimate of the parameter λ in the previous problem and the window size w, what parameter would you expect this distribution of counts to have? What would you expect the mean number of counts per window to be? Comment on how your empirical mean compares to the expected value based on λ. How does the empirical mean change if you make w = 20 days.

Problem 3: Simulating Exponential Variables

Implement a function in MATLAB called `sim_exp.m` that simulates observations of an exponential random variable. Your function should take as input parameters n and $lambda$. The output of your function should be a vector of size $nx1$ containing n samples drawn from an exponential distribution with parameter $lambda$. Please comment your code liberally.

You can use the built in function `rand(n,1)` in order to generate n uniformly distributed random numbers. Section 5.4 in the text will explain a general method for transforming these into exponentially distributed random numbers.

1. Set lambda=5 and generate a histogram of n=1000 data points using 20 bins in the histogram. Use $hist$ to compute the histogram and plot the histogram overlayed with the true exponential pdf with parameter lambda. As in the previous problem, you will need to normalize the bin counts. Briefly comment on the results.

2. Compute the empirical mean and empirical variance of the 1000 data points you generated. Briefly comment on how these empirical numbers compare to the theoretical mean and variance of an exponential rv with parameter lambda.