Data Analysis and Simulation Homework 2 Due: 2/24/2015

1. Select two large traces from Microsoft iotta traces: http://iotta.snia.org/traces/
   There are 36 traces available there but I would like you to select from the following set: $mds_1, proj_3, stg_0, mds_0, web_2, ts_0, usr_0, rsrch_0, wdev_0$. Pick two traces from which you can extract the following information (this is just an example): arrival time stamp and service time of the request. You may need to extrapolate the service time. For *both* inter-arrival and service times:

   - Provide the mean, variance, standard deviation, and C.V. of the inter-arrival and service processes. Plot the transient measures (i.e., across time) of these metrics.
   - Plot the pdf (or its approximation using histograms) and cdf of the inter-arrival times and the service times as indicated by the file size (assume that service times are linearly related to the file size). Use different size bins and comment on the visual/actual effect. Be creative.
   - Plot the autocorrelation function (as a function of the lag) of the inter-arrival times and service times of the trace.

2. Assume that the trace drives a single server queue with infinite buffer that operates under the FCFS service discipline.

   - Provide the following metrics: mean job delay time (service time plus wait time in the queue), system utilization, waiting queue length, and autocorrelation of the *departure* process.
   - Now drive the single server queue using *exponential* service times. Use the same mean for the service times as the mean of the trace and provide the same measures as before. Compare and comment.
   - For a range of mean inter-arrival times that will result in system utilization that ranges in [0.1...0.9], plot the mean waiting queue length and mean delay times as a function of system utilizations, and autocorrelation of the departure process. Compare the results under the original trace and the trace with `exponential` service times in last question. Comment on your findings.

   When using simulation use the next-even simulation model as described in Chapter 5.

Try to be creative, this is part of research.