ORIE 4580/5580 Simulation Modeling and Analysis

Project-Related Assignment

# Fitting Input Distributions to the Provided Data

- **Due date: Thursday November 7th, 11.59pm on Canvas website.**
- **Hand in a single assignment for your project group with the names and Net IDs of all group members. If you have not yet formed your project team then you may do this homework in pairs or on your own. Show all relevant calculations, but be brief.**
- **Include acknowledgments for any help (from others in class, online resources, etc.).**
- **Submission format: single Jupyter notebook with typeset solutions using markdown (or scanned handwritten answers, see Piazza), along with code and inline plots.**

1. Make sure that every team member has read the project description before starting this assignment. The goal of this assignment is to understand the problem, "clean" the data collected if necessary and fit input distributions.
2. Estimate the fraction of customer requests that require off-site repair. Give a confidence interval for your estimate, stating any assumptions you make.
3. Plot a histogram of the on-site repair times and fit a distribution to these data. Be sure to explain how you arrived at this distribution.
4. (a) Plot a histogram of the initial diagnosis times. Does the histogram look like any of the distributions that you studied before?
   (b) Estimate the mean and standard deviation of the initial diagnosis times for each business center separately. Do these values differ substantially?
   (c) How would you separate the initial diagnosis times into different groups by using the information that you obtained in Part b? Plot a histogram for each group.
   (d) Describe how you would model the initial diagnosis times in your simulation.
5. Estimate the fraction of calls that originate at each business center. Give a confidence interval for your estimates.

6. Fit a nonstationary Poisson process to the customer request arrival data. Assume that there are only time-of-day effects in the data. (That is, ignore day-of-week and seasonality effects.) Plot the estimated rate function. Also, give a table that shows the exact values of the rate function.

7. What distribution would you use to model the copier swap times (both at dispatch center and business centers)?

Note: There is a chance that a piece of data that you obtained from TIDInc may not be time-stationary. For example, the fraction of calls that originate at each business center may be dependent on the time of the day. Business Center 1 may be closed between 0:00-6:00, so that the fraction of calls received from Business Center 1 during this time period is zero. On the other hand, a customer location in Business Center 2 may be working a third shift.

You are not asked to check the time-stationarity of the data (of course, except for the customer request arrival data, to which you will be fitting a nonstationary Poisson process). But, in a more realistic simulation study, you would have to make these checks. At minimum, you would group the data into bins (for example, the data that were recorded between 0:00-6:00, 6:00-12:00, 12:00-18:00 and 18:00-00:00). Then, you would compute, say, the mean of the data in each bin, and check whether these means differ substantially.