Statistical Inference
(EECE5612)
Spring 2022

Detectathon
(April 4-6, 2022)

Objective: The objective of this exercise is to get engaged with various forms of hypothesis testing.

Problem: A sensor system is installed at the airport entrance door to count the number of arriving customers. The airport is a busy one, open 7 days per week, 24 hours per day. Once a week, the sensor data is offloaded for analysis. When the sensor functions properly, the data reveals a Poisson process with independent, exponentially distributed arrival times $Y_n,\ n = 1, 2, \ldots$, and an average arrival rate λ_1. However, some data records reveal malfunctioning of the sensor. In those records, the usual arrival process changes into another process at an arbitrary moment in time. The new process is also a Poisson process, but with average rate $\lambda_2 > \lambda_1$. The change occurs because a manufacturing fault causes the sensor to start triggering to events unrelated to people passing through the door. If the end-of-week data analysis reveals malfunctioning of the sensor, the sensor is replaced, and a new week begins afresh.

The Matlab file `door1.mat` contains the data record for one week. The values stored in this file represent realizations $y_n,\ n = 1, 2, \ldots$, of the time (in minutes) between successive counts made by the sensor. Your tasks are the following:

- Given $\lambda_1 = 1$ arrival per minute, $\lambda_2 = 2\lambda_1$, and a time $t_0 = \alpha T$, where $T = 7 \times 24$ hours and $\alpha = 0.73$, decide which of the following is true for the data record in `door1.mat`:
  (a) The sensor functioned properly throughout the week.
  (b) The sensor malfunctioned from the start.
  (c) The sensor started malfunctioning at t_0.
- Given $\lambda_1 = 1$ arrival per minute and $\lambda_2 = 2\lambda_1$, detect the time t_0 (or the index n_0 in the sequence y_n) at which the change is most likely to have taken place.
- Determine the time of change t_0 if λ_1 and λ_2 are unknown.
- The files `door2.mat, door3.mat, ..., door10.mat` contain data records for nine other airport doors. At maintenance time, only three new sensors are available in the stockroom. At which doors should they be installed to replace the existing sensors?

Reporting: Your report should be typed, and not exceed two single-sided pages. It should be written in a professional manner. Figures and mathematical expressions should be used whenever meaningful. Figures should always have axes labeled in appropriate units (e.g. time [s]). Include any Matlab code as an appendix. Please put your name on top of the report.

Judging: Your work will be scored based on three equal categories: (1) technical correctness, (2) clarity of reporting, and (3) conciseness of the code.