Homework 3

STAT 517 - Winter 2023

1. **A simplified version of the time-rescaling theorem:**. Let $t_1, t_2, \ldots, t_n$ be the occurrence times of a non-homogeneous Poisson process on $[0, T]$ with finite and absolutely continuous cumulative intensity function $\Lambda(t) < \infty$ for every $t \in [0, T]$. Define $t_i^* = F(t_i)$, where $F(t) = \Lambda(t)/\Lambda(T)$. Show that the times $t_1^*, t_2^*, \ldots, t_n^*$ correspond to the occurrence times of a homogenous Poisson process with intensity $\mu = \Lambda(T)$

2. Consider the dataset `japanearthquakes.txt` containing the year in which large earthquakes occurred in the South Kanto area of Japan.

   (a) Fit a non-homogenous Poisson process to this data that allows you to investigate the possibility of periodicity in this dataset. Is there evidence of periodicity? If so, what is the period?

   (b) Use the results you derived in Q1 above to assess the goodness of fit of this model in this dataset.

3. The Hawkes process model has been proposed as a tool for understanding the spread of information in social media. For example, in the case of Twitter data, we could use hashtags to track the spread of a particular conversation topic. Using the branching structure construction of the Hawkes process, the first tweet mentioning a given hashtag acts as a point from the *inmigrant* process, retweets/responses to these act as *first-generation offspring*, retweets/responses to these act in turn as *second-generation offspring*, and so on. Note that this setup is a bit different from the "traditional" one for Hawkes processes as we are assuming that the branching structure for the process is known.

   The data in the file `twitter.R` consists of two objects.

   - `points` is a list, with each entry in the list corresponding to a vector containing the times associated with the events in the immigrant process (first element) and each subsequent generation of offspring.
   - `parents` is another list, with the same shape as `points`, with each entry listing the parents of each point from the previous generation. (Note that the immigrants are all listed as having parent 0.)

(a) Assuming that this data is generated form a Hawkes process with an exponential excitation function, find the maximum likelihood estimator of the parameters. Make sure that you make use of the known branching structure to estimate the parameters.

(b) Is there evidence of this process being stationary? The best way to answer this question is to provide a confidence interval for the appropriate model parameter.

(c) Compute now the maximum likelihood estimators of the parameters ignoring the information contained in the branching structure. Compare these estimators to those you obtained in part (a).

(d) Based on the results above, do you think that a standard Hawkes process is a good model for this data?