Intro to Point Patterns Project

Extreme heat scenarios such as these pose a threat to public health. For example, various studies have correlated extreme heat to increases in mortality , emergency room visits and other diseases. Given the potential negative consequences of high heat, public health officials are interest in understanding the demographics of callers who call 911 for heat-related illnesses. Ultimately, by understanding these demographics, appropriate intervention strategies can be implemented to prepare for extreme heat situations and, hopefully, reduce the occurrence of potential negative outcomes. The file `AgePP.RData` lists the ages of 1818 individuals who called 911 between the years 2006 and 2010 for a heat-related emergency. The goal of this project is to use a point process model to estimate the proportion callers of various ages.

A point pattern, point process, or Poisson process is set of N "locations" where an event occurred. In the case considered here, the "event" is a heat-related 911 call and the "location" is the age of that caller that is "located" somewhere in $0, 1, \ldots, 110$. From a statistical perspective, the number and "location" of an event are both random but the random behavior is governed by an intensity surface $\Lambda(a)$. In the point process framework, the number of events that occur N follows a Poisson distribution with mean $\sum_{a=0}^{110} \Lambda(a)$. Likewise, $\Lambda(a)$ controls the age of the caller in that the probability that a random caller is of age a is given by the normalized intensity (which is just a density function)

$$\lambda(a) = \frac{\Lambda(a)}{\sum_{a'=0}^{110} \Lambda(a')}.$$

Hence, for statistical analysis of point patterns, $\Lambda(a)$ is the primary quantity of interest.

Given observed point pattern $a_1, \ldots, a_N$ where each $a_i \in \{0, \ldots, 110\}$ with $\mathbb{P}\text{r}(a_i = a) = \Lambda(a) / \sum_{a'=0}^{110} \Lambda(a')$ and $N \sim \mathcal{P}(\sum_{a=0}^{110} \Lambda(a))$, one can show that the likelihood for $\Lambda(a)$ is given by,

$$L(\Lambda) = \exp\left\{ -\sum_{a=0}^{110} \Lambda(a) \right\} \prod_{i=1}^{N} \Lambda(a_i). \tag{1}$$

A convenient form for $\Lambda(a)$ is to let $\Lambda(a) = \delta f(a)$ where $\delta > 0$ and $\lambda(a)$ is an unknown probability density function (or, in this case, a probability mass function). That is, $\sum_{a=0}^{110} \lambda(a) = 1$.

Perform the following:

1. Given the factorization $\Lambda(a) = \delta\lambda(a)$, write out the likelihood function for δ and $\{\lambda(a) : a = 0, \ldots, 110\}$.

2. Often in point pattern analyses, the PMF $\lambda(a)$ is assumed to follow a parametric form (e.g. $\lambda(a) =$ is a Bernoulli distribution with unknown probability p). Paying particular attention to the support of $\lambda(a)$, propose and justify an appropriate parametric distribution for $\{\lambda(a) : a = 0, \ldots, 110\}$ and rewrite the likelihood in (1) given your choice.

3. Given the form of the likelihood in #1 and #2, propose a conjugate prior distribution for δ and show that it is indeed conjugate.

4. Regardless of your answer in #2, let $\lambda(a)$ be the MN$(1, p_0, \ldots, p_{110})$ where MN$(n, p_1, \ldots, p_K)$ is the multinomial distribution with parameters n and $p_1, \ldots, p_K$. Rewrite out the likelihood in (1) given this choice, propose a conjugate prior distribution for the unknown probabilities $p_0, \ldots, p_{110}$ and show that it is indeed conjugate.

5. Fit the point process model from #3 and #4 to the observed point pattern of the age of 911 callers. Plot the posterior distribution of δ and the posterior means (with 95% credible intervals) of $p_0, \ldots, p_{110}$.