

## Homework 03 - Partial solutions

Instructor: Abel Rodriguez

TA: Aparna Venkat

*Note: This is only a high-level sketch of the solutions. If you catch any typos or have questions, post them in the discussion board.*

1. Since  $\Lambda$  is absolutely continuous on a bounded interval  $[0, T]$ , it is also differentiable. Call  $\Lambda'(t) = \lambda(t)$ . I will denote the likelihood of  $\{t_i^*\}_{i=1}^n$  as  $P^*$  and the likelihood of  $\{t_i\}_{i=1}^n$  using  $P$ . We have,

$$P(t_1, \dots, t_n) = e^{-\Lambda(T)} \prod_{i=1}^n \lambda(t_i).$$

By performing a multivariate random variable transformation transformation,

$$\begin{aligned} P^*(t_1^*, \dots, t_n^*) &= P(F^{-1}(t_1), \dots, F^{-1}(t_n)) |\mathcal{J}| \\ &= e^{-\Lambda(T)} \prod_{i=1}^n \lambda(F^{-1}(t_i)) |\mathcal{J}| \end{aligned}$$

where  $\mathcal{J}$  is the Jacobian of this transformation. It is easy to see that we only need to care about the diagonal elements of the Jacobian. Thus,

$$\begin{aligned} P^*(t_1^*, \dots, t_n^*) &= e^{-\Lambda(T)} \prod_{i=1}^n \lambda(F^{-1}(t_i)) |(F^{-1})'(t_i)| \\ &= e^{-\Lambda(T)} \prod_{i=1}^n \lambda(F^{-1}(t_i)) \frac{\Lambda(T)}{\lambda(F^{-1}(t_i))} \\ &= e^{-\Lambda(T)} (\Lambda(T))^n, \end{aligned}$$

which shows the desired result that the transformed times  $t_i^*$  are a homogeneous Poisson process on the interval  $[0, 1]$ .

2. Pick any periodic function that is a valid intensity i.e., non-negative. For example,

$$\lambda(t) = \exp\{A \cos(\omega t + \phi) + B\}.$$

Then write the likelihood,

$$\begin{aligned} \mathcal{L} &= e^{-(\Lambda(T) - \Lambda(T_0))} \prod_{i=1}^n \lambda(t_i), \\ \Lambda(x) &= \int_0^x \lambda(t) dt. \end{aligned}$$

We need to be careful with  $\Lambda(T_0)$  here as the observations are not on  $(0, T]$  but on  $(T_0, T]$ . Now, we just maximize it numerically. Note that we need to calculate  $\Lambda(T) - \Lambda(T_0)$  as well. For this particular model, this needs to be done numerically as well. We can estimate the function  $\hat{\lambda}$ .

For goodness-of-fit, we perform the transformation as in Q1,  $t^* = \hat{\Lambda}(t)/(\hat{\Lambda}(T) - \hat{\Lambda}(T_0))$ . Then, we test if  $\{t_i^*\}_{i=1}^n$  come from a standard uniform distribution using Kolmogorov-Smirnov test and/or Q-Q plots. See Lab 3 for examples of these.

3. As notation, let  $X_{j,k}^{(i)}$  denote the arrival time of sample  $k$  in generation  $j$  which spawned from immigrant  $i$ . Assuming immigrants are the 0-th generation,  $X_{0,1}^{(i)}$  is the arrival time of immigrant  $i$ . Suppose we observe for time  $[0, T]$  and count  $m$  immigrants. Then,

$$P(X_{0,1}^{(1)}, \dots, X_{0,1}^{(m)}) = e^{-\mu T} \mu^m$$

as we know immigrants come from a homogeneous Poisson point process with rate  $\mu$ .

Now, take any  $j$ -th generation,  $j > 0$ . The number of arrivals in this generation (descended from  $i$ ),  $N_{ij}$ , is given by  $\text{Poisson}(\alpha)$  and their arrival times are given by  $\text{Exponential}(\beta)$ . Suppose there are  $N_{ij} = n_{ij}$  arrivals belonging to the  $j$ -th generation descended from immigrant  $i$ . Then,

$$\begin{aligned} P(X_{j,1}^{(i)}, \dots, X_{j,n_{ij}}^{(i)}) &= P(X_{j,1}^{(i)}, \dots, X_{j,n_{ij}}^{(i)} \mid N_{ij} = n_{ij}) P(N_{ij} = n_{ij}) \\ &= P(N_{ij} = n_{ij}) \prod_{k=1}^{n_{ij}} \beta \exp \left\{ -\beta \left( X_{j,k}^{(i)} - \text{parent}(X_{j,k}^{(i)}) \right) \right\} \\ &= \frac{e^{-\alpha N_{j-1,i}} \alpha^{n_{ij}}}{n_{ij}!} \beta^{n_{ij}} \exp \left\{ -\beta \sum_{k=1}^{n_{ij}} \left( X_{j,k}^{(i)} - \text{parent}(X_{j,k}^{(i)}) \right) \right\} \end{aligned}$$

where  $\text{parent}(X_{j,k}^{(i)})$  is the parent in generation  $j-1$  that spawned  $X_{j,k}^{(i)}$ . We need to do this for all generations. Eventually the log-likelihood will look like<sup>1</sup>

$$\begin{aligned} \ell &= -\mu T + N_{\text{immigrants}} \log \mu + N_{\text{descendants}} (\log \alpha \beta) - N_{\text{arrivals}} \alpha \\ &\quad - T_{\text{spawn}} \beta, \end{aligned}$$

where  $N_{\text{immigrants}}$  is the total number of immigrants;  $N_{\text{descendants}}$  is the total number of descendants who came from the  $N_{\text{immigrants}}$  immigrants;  $N_{\text{arrivals}} = N_{\text{immigrants}} + N_{\text{descendants}}$  is the total number of arrivals, and;  $T_{\text{spawn}}$  is the time it took for each child to spawn from its parent (this is summed over all children).

This gives us the MLE,

$$\begin{aligned} \hat{\mu} &= \frac{N_{\text{immigrants}}}{T} \\ \hat{\alpha} &= \frac{N_{\text{descendants}}}{N_{\text{arrivals}}} \\ \hat{\beta} &= \frac{N_{\text{descendants}}}{T_{\text{spawn}}} \end{aligned}$$

---

<sup>1</sup>Here we are assuming that arrivals in the fourth generation do not produce any children. If instead you assume that the children of the fourth generation are not observed, then the coefficient on  $\alpha$  will be  $N_{\text{arrivals}} - N_4$  where  $N_4$  is the number of arrivals in the fourth generation.

And if  $\hat{\alpha} < 1$ , then its variance is,

$$\begin{aligned}\text{Var}(\hat{\alpha}) &= \alpha(1 - \alpha) \\ \implies \widehat{\text{Var}}(\hat{\alpha}) &= \hat{\alpha}(1 - \hat{\alpha})\end{aligned}$$