Nonparametric Bayesian modeling and estimation for renewal processes

Sai Xiao, Athanasios Kottas, Bruno Sansó and Hyotae Kim *

Abstract

We propose a flexible approach to modeling for renewal processes. The model is built from a structured mixture of Erlang densities for the renewal process inter-arrival density. The Erlang mixture components have a common scale parameter, and the mixture weights are defined through an underlying distribution function modeled nonparametrically with a Dirichlet process prior. This model specification enables non-standard shapes for the inter-arrival time density, including heavy tailed and multimodal densities. Moreover, the choice of the Dirichlet process centering distribution controls clustering or declustering patterns for the point process, which can therefore be encouraged in the prior specification. Using the analytically available Laplace transforms of the relevant functions, we study the renewal function and the directly related K function, which can be used to infer about clustering or declustering patterns. From a computational point of view, the model structure is attractive as it enables efficient posterior simulation while properly accounting for the likelihood normalizing constant implied by the renewal process. A hierarchical extension of the model allows for the quantification of the impact of different levels of a factor. The modeling approach is illustrated with several synthetic data sets, earthquake occurrences data, and coal-mining disaster data.

Keywords: Dirichlet process; Erlang mixtures; Inter-arrival distribution; Markov chain Monte Carlo; Renewal function.

*S. Xiao is Data Scientist, Pinterest, San Francisco, CA. A. Kottas and B. Sansó are Professors, and Hyotae Kim is Ph.D. student, Department of Statistics, University of California, Santa Cruz, CA.

1 Introduction

Poisson processes are the most frequently used stochastic models for temporal or spatial point patterns. Their theoretical properties are fully developed, and it is possible to perform likelihood-based or Bayesian model fitting for Poisson processes with general intensity functions. In particular, the literature includes several Bayesian nonparametric modeling methods and applications; see, e.g., Møller et al. (1998), Wolpert and Ickstadt (1998), Ishwaran and James (2004), Kottas and Sansó (2007), Adams et al. (2009), Taddy (2010), Taddy and Kottas (2012), Kang et al. (2014), Xiao et al. (2015), and Rodriguez et al. (2017).

To achieve distributional flexibility beyond the Poisson assumption, as well as to model various clustering behaviors, it is desirable to consider non-Poisson point processes. Flexible inference for such processes is hampered by the difficulty of handling likelihood functions with normalizing constants that depend on the parameters of interest in ways that can make computations prohibitively expensive. One way to extend the temporal Poisson process is to consider renewal processes. Renewal processes are counting processes where the time intervals between successive arrivals are independent and identically distributed (i.i.d.) according to the inter-arrival distribution, a continuous distribution on the positive real line, $\mathbb{R}^+$, with finite expectation. The inter-arrival distribution characterizes the stochastic properties of renewal processes, including the K function which describes clustering or declustering patterns relative to the homogeneous Poisson process. More details on relevant definitions for renewal processes are given in Section 2.1. Renewal processes find a number of applications, such as modeling of earthquakes occurrences, software reliability, hardware maintenance, and queuing systems, among others.

A Weibull inter-arrival distribution is commonly used for data that exhibit clustering patterns. For example, for analysis of earthquake data, Alvarez (2005) and Epifani et al. (2014) used a Weibull distribution to model the time between different types of events (corresponding to different earthquake magnitude categories). Zhao and Nagaraja (2011) applied renewal theory to longitudinal studies of lupus and its periodical flare-ups, using exponential, gamma and Weibull inter-arrival distributions. More flexible approaches to model the renewal distribution have been considered in the literature on queueing systems, where, however, the normalizing

constant of the likelihood is ignored for the sake of tractability. The problem is thus simplified to that of density estimation of inter-arrival times. In this context, Wiper et al. (2001) developed a mixture of gamma distributions for general density estimation. Along similar lines, Ausín et al. (2007) and Thümmler et al. (2006) use mixtures of Erlang distributions. The Erlang distribution belongs to the gamma distribution family and is central to the methods proposed in this paper.

In particular, we propose a flexible Bayesian modeling approach for the renewal process inter-arrival distribution, using mixtures of Erlang distributions with mixing on the integer shape parameters and with a common scale parameter. Hence, the mixture model requires estimation of only one parameter for the kernel densities. This is a parsimonious model feature that does not undermine its flexibility, since the specific class of Erlang mixtures can approximate any continuous distribution on $\mathbb{R}^+$ (see Section 2.2). Such Erlang mixture models have been used for queueing data analysis (Tijms, 1994), and also applied in actuarial science by Lee and Lin (2010), where the mixture weights are estimated by an EM algorithm. To our knowledge, Erlang mixtures have not been used to model inter-arrival times in the point process literature.

Here, we define the mixture weights through a distribution function that is modeled non-parametrically with a Dirichlet process prior (Ferguson, 1973; Antoniak, 1974). This provides a novel Erlang mixture formulation for Bayesian nonparametric density estimation on $\mathbb{R}^+$. It extends Bernstein polynomial priors for density estimation on the unit interval, which have been explored in the Bayes nonparametrics literature following the work by Petrone (1999a,b). A key feature of our method is that the choice of the Dirichlet process centering distribution controls clustering or declustering patterns for the point process, which can therefore be informed by the prior specification. From a computational point of view, the proposed model for the inter-arrival distribution has the advantage of enabling efficient posterior simulation, while properly accounting for the renewal process likelihood normalizing constant.

Our key motivation is to develop flexible, model-based estimation for renewal processes which can adapt to different features suggested by the data, and thus overcome the potential restriction of relying on a particular parametric form for the inter-arrival distribution. Operating under the Bayesian paradigm, full uncertainty quantification is available for estimates of renewal process functionals. Moreover, we seek a modeling framework that can be elaborated to accommodate

extensions of the basic renewal process stochastic structure. Using earthquake data from four distinct regions of the Americas as a motivating example, we demonstrate such an elaboration for estimation of multiple, hierarchically related renewal processes.

The paper is organized as follows. Section 2 begins with a brief review of basic definitions for renewal processes. We then present the mixture model for the renewal process inter-arrival density, and explore model properties, focusing on its flexibility in producing general shapes for the K function. We also discuss a Markov chain Monte Carlo (MCMC) approach to posterior simulation, model inferences, and a graphical model assessment method. In Section 3, we test the model with synthetic data sets that include both declustering and clustering patterns, and in Section 4, we illustrate the methodology with coal mining disaster and earthquake data. In the context of the earthquake data analysis, we extend the modeling approach to estimate hierarchically related renewal processes. Our simulated data results show that the model can successfully capture different clustering patterns for the point process. Model assessment results also support applicability of the model on the real data sets. Finally, Section 5 concludes with a summary and discussion of possible extensions of the methodology.

2 Methodology

2.1 Background on renewal processes

A renewal process $\{N(t) : t \geq 0\}$ is defined as $N(t) = \max\{n : T_n \leq t\}$, with $T_0 = 0$ and $T_n = X_1 + X_2 + \cdots + X_n$, for $n \geq 1$, where the random variables X_i are i.i.d. from a distribution F with support in $\mathbb{R}^+$, and such that $0 < \mathrm{E}(X_i) < \infty$. Hence, T_n is the n-th arrival time and X_n is the n-th inter-arrival time. The distribution of the renewal process at any given time t is characterized by the inter-arrival distribution F. The special case of F being the exponential distribution corresponds to a homogeneous Poisson process.

Let F_k be the distribution function of T_k. From the definition of T_k and the independence of the inter-arrival times, we have $F_{k+1}(t) = \int_0^t F_k(t-u)\mathrm{d}F(u)$, for $k \geq 1$ (where $F_1 \equiv F$), and $\Pr(N(t) = k) = F_k(t) - F_{k+1}(t)$. The expected number of occurrences of the event of interest in the time interval $[0, t]$, denoted as $M(t)$, is known as the renewal function. It is defined as $M(t) =$

$\mathrm{E}(N(t)) = \sum_{k=1}^{\infty} F_k(t)$, and it satisfies the renewal equation, $M(t) = F(t) + \int_0^t M(t-u)\mathrm{d}F(u)$. Directly related to the renewal function is the K function, $K(t) = \lambda^{-1}M(t)$, where λ is the number of events per unit time interval. As discussed in Section 2.3, the K function can be used to assess clustering or declustering patterns relative to the homogeneous Poisson process.

Consider a temporal point pattern $\{0 = t_0 < t_1 < t_2 < \cdots < t_n < T\}$, observed in the time window $[0, T]$, and assumed to be a realization from a renewal process with inter-arrival distribution F and inter-arrival density function f. The upper bound, T, of the observation window is assumed fixed. The renewal process likelihood is given by

$$\begin{aligned} & \Pr(T_1 = t_1, T_2 = t_2, \ldots, T_n = t_n, T_{n+1} > T) \\ = \; & \Pr(X_1 = t_1, X_2 = t_2 - t_1, \ldots, X_n = t_n - t_{n-1}, X_{n+1} > T - t_n) \\ = \; & \left\{\prod_{i=1}^{n} f(t_i - t_{i-1})\right\}\left\{1 - \int_{t_n}^{T} f(u - t_n)\mathrm{d}u\right\}. \end{aligned} \tag{1}$$

The last term in (1) corresponds to the probability of no arrival in $[t_n, T]$, which defines the likelihood normalizing constant.

2.2 Mixture modeling for the inter-arrival density

Denote by $H(t) = t_{N(t^-)}$ the time of the last arrival prior to time point t. (By convention, $H(t) = 0$ for time points t before the first arrival.) The inter-arrival density, $f(t - H(t))$, where $t \in (H(t), \infty)$, is modeled as a mixture of Erlang densities with common scale parameter. More specifically,

$$f(t - H(t) \mid J, \theta, G) = \sum_{j=1}^{J} w_j \,\mathrm{ga}(t - H(t) \mid j, \theta^{-1}), \quad t > H(t) \tag{2}$$

where $\mathrm{ga}(\cdot \mid a, b)$ denotes the density of a gamma(a, b) distribution with mean a/b. Notice that each component of the mixture is defined by a gamma density with shape parameter given by a specified integer, that is, an Erlang density. The number of components, J, and the mixture weights, $\{w_j : j = 1, \ldots, J\}$, are both random. The latter are generated by discretizing a distribution function G: $w_1 = G(\theta)$, $w_j = G(j\theta) - G((j-1)\theta)$, for $j = 2, ..., J-1$, and $w_J = 1 - G((J-1)\theta)$. The common scale parameter, θ, for the Erlang mixture components determines

the precision of the distribution function discretization for the weights. As θ decreases, the model tends to generate more mixture components.

If we let $J \to \infty$, the support of G can be taken to be $\mathbb{R}^+$. To apply the model, we need a truncation for the number of mixture components as in (2) where J is finite, but random. Such truncation restricts the support of G to $(0, J\theta)$, resulting in a probability vector for $\{w_j : j = 1, \ldots, J\}$. We model the mixture weights nonparametrically by assigning to G a Dirichlet process (DP) prior, $G \sim \text{DP}(\alpha, G_0)$. Here, α is the DP precision parameter, and G_0 is the centering distribution defined through a Weibull distribution, Weibull(μ, ϕ), with scale and shape parameters μ and ϕ, respectively. We place hyperpriors on all DP parameters, that is, on α, μ and ϕ.

Figure 1 shows an example of the generation of weights with $\mu = 2$, $\phi = 3$, $\theta = 0.2$, $J = 20$, and two different values for the DP precision parameter, $\alpha = 10$ (left panel) and $\alpha = 1$ (right panel). In the former case, the distribution function sample path from $\text{DP}(\alpha, G_0)$ is closer to G_0, resulting in weights most of which are greater than 0; only $w_{17} = w_{20} = 0$. In the latter case, the DP sample path for G involves fewer effective atoms, and it thus produces weights most of which are nearly 0; only w_5, w_{12} and w_{20} are significantly greater than 0. Hence, the discreteness of the DP prior for G allows for a dimension reduction technique, reducing the number of Erlang mixture components to just those corresponding to the positive weights. This is an important feature of the model. Sections 2.3 and 2.4 study how different settings of α and G_0 impact the clustering/declustering patterns supported by the model.

Since both J and θ impact the effective support of inter-arrival times, there is a strong interaction between these model parameters, and it is thus natural to specify a joint prior distribution for (θ, J). Specifically, we place a gamma(a_θ, b_θ) prior on θ, and given θ, assign J a discrete uniform prior distribution on $\{\lceil S_1/\theta \rceil, \ldots, \lceil S_2/\theta \rceil\}$. Here, $\lceil x \rceil$ denotes the smallest integer not less than x, and S_1 and S_2 are proxies for the lower and upper bounds, respectively, of the support for the inter-arrival times. Our default choice is to set S_1 equal to the maximum observed inter-arrival time, S, and S_2 equal to cS for fixed $c > 1$; for the data examples, we used values of c that range between 3 and 10.

We note that, for the data examples in Sections 3 and 4, model inferences are largely unaf-

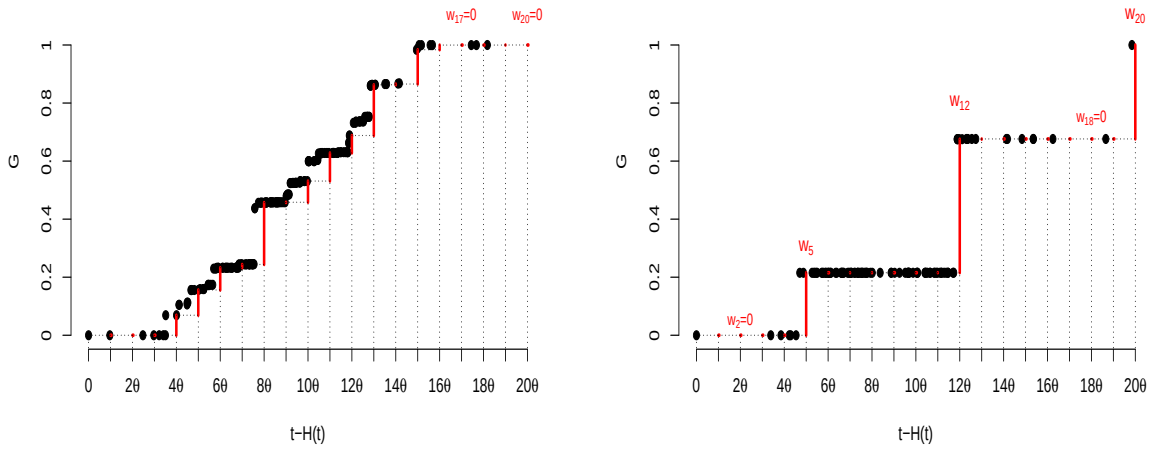


Figure 1: Mixture weights under two different specifications of the prior model: $J = 20$, $\theta = 0.2$ and $G \sim \text{DP}(\alpha, \text{Weibull}(\mu = 2, \phi = 3))$, with $\alpha = 10$ (left panel) and $\alpha = 1$ (right panel). The solid dots correspond to the distribution function sample paths for G. The dashed lines display the discretization by θ, and the thick red lines indicate the weights.

fected by fixing J to values compatible with the prior specification approach discussed above. In particular, to simplify the model fitting, J can be set equal to $\lceil S/\theta^* \rceil$, where θ^* is the prior mean (or median) for θ.

The motivation for the Erlang mixture model for the inter-arrival density is twofold: on one hand we wish to achieve flexibility, and on the other, we seek to facilitate the required computations for posterior simulation. Flexibility is guaranteed by the fact that, as $J \to \infty$ and $\theta \to 0$, the distribution function of the Erlang mixture in (2) can approximate (in a pointwise sense) any continuous distribution function on $\mathbb{R}^+$ (e.g., Lee and Lin, 2010). The DP-based model specification is particularly well suited for the inference objectives in the context of renewal processes. More specifically, as discussed in Sections 2.3 and 2.4, the model can support both clustering and declustering patterns, and this is controlled by the DP centering distribution.

At the same time, the structure of the Erlang mixture model is key for the development of an efficient MCMC posterior simulation method, which properly accounts for the likelihood

normalizing constant. Given the observed point pattern $\{0 = t_0 < t_1 < t_2 < \cdots < t_n < T\}$, the first stage of the hierarchical model for the data is written using the mixture form in (2) for the likelihood components $f(t_i - t_{i-1})$ and $f(u - t_n)$ in (1). Then, by introducing auxiliary variables $\{y_1, \cdots, y_n, y_{n+1}\}$, which given G are i.i.d. from G, we can "break" the mixture and label the component to which each observation is assigned. Of primary importance is the last configuration variable, y_{n+1}, which allows us to handle the integral in the likelihood normalizing constant in a similar fashion with the other, more standard density components of the likelihood. More details are provided in Section 2.5 and in the Appendix.

2.3 Model properties

A flexible model for renewal processes needs to be able to capture different clustering behaviors. We recall that the K function, introduced by Ripley (1977) for spatial point processes, is used to determine the clustering properties of a point process. In the context of a temporal renewal process, the K function is given by $K(t) = \lambda^{-1}M(t)$. Here, $M(t)$ is the renewal function, and λ is the number of events per unit time interval given by $\lambda = 1/\mathrm{E}(X)$, where X represents inter-arrival times. The K function can be used to assess clustering relative to the homogeneous Poisson process for which $K(t) = t$; in particular, $K(t) > t$, corresponds to clustering processes, while $K(t) < t$ indicates that the point processes has a declustering pattern. Some general properties of $M(t)$, and thus also of $K(t)$, are reported below.

Result 1. *Let $\mu_1 = E(X)$, $\mu_2 = E(X^2)$ and let c_X denote the coefficient of variation of X, where $c_X^2 = Var(X)/(E(X))^2 = (\mu_2/\mu_1^2) - 1$. Then, for sufficiently large t, $M(t) \approx t\mu_1^{-1} + 0.5(c_X^2 - 1)$.*

This result corresponds to Theorem 1.1.9 in Tijms (1994). Direct application to the K function yields $K(t) = \mu_1 M(t) \approx t + 0.5\mu_1(c_X^2 - 1)$, for large enough t. Hence, asymptotically, the sign of $c_X^2 - 1$ indicates if the point process supports declustering or clustering patterns.

Result 2. *Denote by $F_n(t)$ the distribution function of the n-th arrival time, T_n. Then, $M(t) = \sum_{n=1}^{\infty} F_n(t)$, for any $t \geq 0$. moreover, $\sum_{n=N+1}^{\infty} F_n(t) \leq \frac{F_N(t)F(t)}{1-F(t)}$, implying that,*

$$\sum_{n=1}^{N} F_n(t) \leq M(t) \leq \sum_{n=1}^{N} F_n(t) + \epsilon(t), \quad \epsilon(t) = \frac{F_N(t)F(t)}{1 - F(t)}.$$

This result is also found in Tijms (1994). Hence, provided $\epsilon(t)$ is sufficiently small, $M(t)$ can be approximated by the truncated sum $\sum_{n=1}^{N} F_n(t)$.

Result 3. *Assume that the inter-arrival density, f, has Laplace transform $L_X(s) = \int_0^\infty e^{-su} f(u)du$. Then, the Laplace transform of the renewal function, $L_M(s) = \int_0^\infty e^{-st} M(t)dt$, is given by*

$$l_M(s) = \frac{L_X(s)}{s\{1 - L_X(s)\}}. \tag{3}$$

Equation (3) is obtained using standard integration and convolution properties of the Laplace transform applied to the renewal equation, $M(t) = \int_0^t f(u)\mathrm{d}u + \int_0^t M(t-u)f(u)\mathrm{d}u$. This is a key result, as it provides a practical method to calculate the renewal function $M(t)$ using its inverse Laplace transform.

The results discussed above are applicable to general renewal functions. Next, we discuss results specific to our model. First, the mean inter-arrival time under the Erlang mixture model in (2) is given by $\mu_1 \equiv \mathrm{E}(X \mid J, \theta, G) = \theta \sum_{j=1}^{J} j w_j$. It is clear from this expression that the required finiteness for μ_1 is satisfied. Moreover,

$$c_X^2 = \frac{\{\sum_{j=1}^{J} w_j(j + j^2)\} - \{\sum_{j=1}^{J} j w_j\}^2}{\{\sum_{j=1}^{J} j w_j\}^2}. \tag{4}$$

Hence, the coefficient of variation generated from our model is fully determined by the mixture weights. Since Erlang mixtures are dense in the space of continuous distributions on $\mathbb{R}^+$ (Tijms, 1994; Lee and Lin, 2010), equation (4) can take, in principle, any positive value. This implies that the model can generate values for c_X^2 that are both below and above one, and therefore, using Result 1, our model can generate point processes with both clustering and declustering/regularity patterns, at least asymptotically.

The (approximate) form of the K function can be explored applying Result 2. For a few special cases of our model, a closed-form expression for $M(t)$ can be obtained, using the definition $M(t) = \sum_{n=1}^{\infty} F_n(t)$. In particular, for a single Erlang distribution with shape parameter equal to 2 (i.e., the special case of (2) with $w_2 = 1$) we obtain $K(t) = t + 0.5\theta\{\exp(-2t/\theta) - 1\} < t$,

for any $t > 0$, which results in declustering patterns. From numerical evalution of the approximate $K(t)$, computed using result 2, we observe that a single Erlang distribution with shape parameter greater than 2 also generates declustering processes. The opposite is observed for the mixture of an exponential distribution and an Erlang distribution with shape parameter ≥ 4, as such two-component mixtures generate clustering patterns. Section 2.4 includes results from a comprehensive empirical investigation of the K functions generated by the general form of the Erlang mixture model for the inter-arrival density, with the DP-based prior for the mixture weights. These results indicate that the model supports both clustering and declustering patterns for any $t > 0$, and not just asymptotically.

Computing the K function requires the renewal function. Using Result 3, we obtain the Laplace transform of the renewal function corresponding to the Erlang mixture for the inter-arrival density:

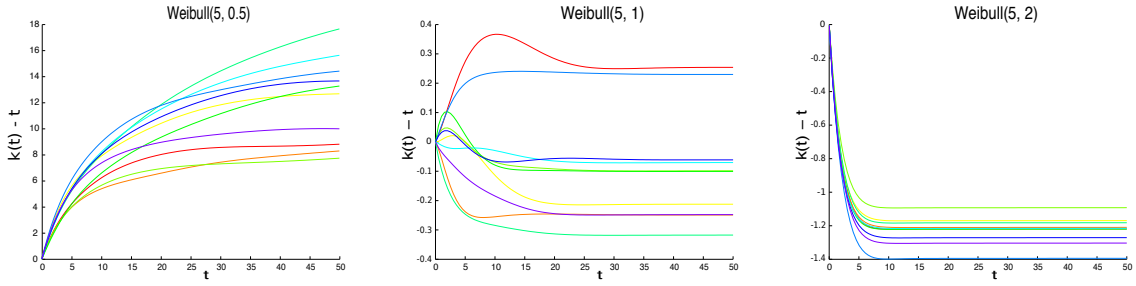
$$l_M(s) = \frac{\sum_{j=1}^{J} w_j (1+\theta s)^{-j}}{s \sum_{j=1}^{J} w_j \{1-(1+\theta s)^{-j}\}}. \tag{5}$$

This is a key result for the study of K functions arising from our model, since evaluating $M(t)$ from its definition is computationally intensive under the general version of the Erlang mixture which typically contains more than 2-3 mixture components. Instead, evaluating the renewal function of the general model using the inverse Laplace transform of Equation (5) is substantially more efficient. We use Matlab function "INVLAP" for the numerical inversion of the renewal function Laplace transform. The application of this approach is illustrated in the next section.

2.4 Investigating the K function for the Erlang mixture model

In this section, we numerically evaluate the K function generated from our model by using the Laplace transform method. Given J, θ, G_0 and α, a realization for the K function can be simulated from the prior model by: generating a realization for the mixture weights $\{w_1, \cdots, w_J\}$ from the DP(α, G_0) prior; evaluating the Laplace transform of the renewal function $M(t)$ using Equation (5); computing $M(t)$ using the inverse Laplace transform; and computing the K function as $K(t) = \mu_1 M(t)$, where $\mu_1 = \theta \sum_{j=1}^{J} j w_j$.

We used this simulation method to investigate how the DP precision parameter α, and the



(a) (b) (c)

Figure 2: Realizations for the function $K(t) - t$ based on 10 samples from the DP(α, G_0) prior for the mixture weights, with a Weibull centering distribution and $\alpha = 200$. From left to right: $G_0 =$ Weibull$(5, 0.5)$, $G_0 =$ Weibull$(5, 1)$ and $G_0 =$ Weibull$(5, 2)$.

parameters of the Weibull DP centering distribution G_0 influence the K function, and thus the clustering patterns generated by the model. We also explored the effect of selecting different distributions for G_0. The results from some of the scenarios we have considered are discussed next.

The shape parameter of the Weibull DP centering distribution. We fix the values of $\alpha = 200$, $\theta = 2$ and $\mu = 5$ and vary the shape parameter, ϕ, of the Weibull distribution defining G_0. The DP precision parameter α is fixed to a relatively large value, implying realizations for G that are close to G_0, such that we can more effectively explore the role of the DP centering distribution parameters. The effect of decreasing the value of α is studied later in the section.

Figure 2 shows ten realizations for the K function under three different values for ϕ. We plot $K(t) - t$ for easier visual inspection; recall that $K(t) > t$ $(< t)$ corresponds to clustering (declustering), with $K(t) = t$ for the homogeneous Poisson process. For $\phi = 0.5$, the K function samples display clustering patterns; see Figure 2a. With $\phi = 1$, clustering and declustering patterns are generated, as well as K functions that are close to the one corresponding to the homogeneous Poisson process; see Figure 2b. when ϕ is increased to 2, declustering patterns are produced; see Figure 2c. Table 1 summarizes these empirical results.

Note that $\phi < 1$ (> 1) implies that the coefficient of variation for the Weibull distribution

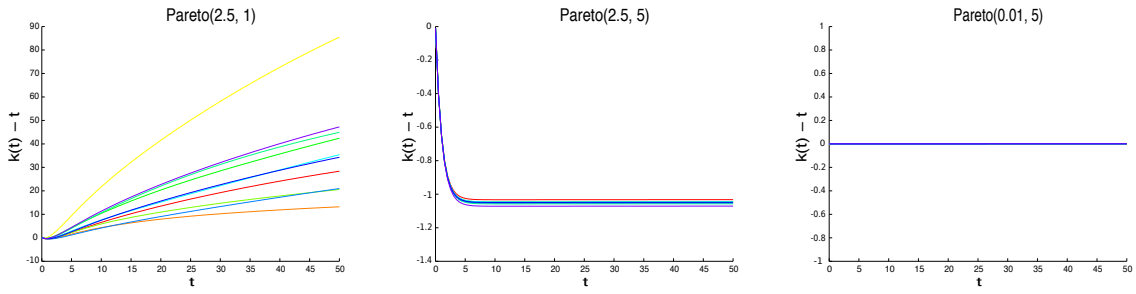
Table 1: Clustering patterns generated under different centering distributions G_0 for the DP prior. **C**: clustering; **D**: declustering; **H**: homogeneous Poisson process.

G_0	**Fixed parameters**	**Parameter(s) tested**	**C**	**D**	**H**
Weibull(μ, ϕ)	$\theta = 2, \alpha = 200, \mu = 5$	$\phi = 0.5$	✓		
		$\phi = 1$	✓	✓	✓
		$\phi = 2$		✓	
Pareto(a, b)	$\alpha = 200$	$\theta = 2,\ a = 2.5,\ b = 1$	✓		
		$\theta = 2,\ a = 2.5,\ b = 5$		✓	
		$\theta = 1,\ a = 0.01,\ b = 5$			✓

is greater than 1 (less than 1), and it also corresponds to a decreasing (increasing) hazard rate function. Our empirical investigation provides strong indication that, when the DP precision parameter takes relatively large values, the shape parameter of the Weibull centering distribution controls clustering ($\phi < 1$) or declustering ($\phi > 1$) patterns for the renewal process.

The scale parameter of the Weibull DP centering distribution. The scale parameter μ can also affect the point process pattern, albeit less evidently than the shape parameter. For instance, when $\phi > 1$ and α takes large values, then values of μ such that $\mu < \theta$ result in realizations G that concentrate mainly in $(0, \theta)$, which further results in $w_1 \to 1$. The inter-arrival density is then close to an exponential distribution with parameter $1/\theta$, corresponding to a homogeneous Poisson process. If $\mu > \theta$, then G is concentrated mostly in (θ, ∞), which produces declustering patterns.

Alternative choices for the DP centering distribution. In order to explore the influence of the tails of G_0 on the behavior of the K function, we run a set of experiments where G_0 is a Pareto distribution, Pareto(a, b), supported on $[a, \infty)$. In contrast to the exponential tails of the Weibull, the Pareto distribution has polynomial tails. The Pareto shape (tail index) parameter, b, determines the coefficient of variation, in particular, $c_X^2 = \{b(b-2)\}^{-1}$, for $b > 2$. Moreover, as b tends to infinity, the Pareto distribution concentrates all its mass at a. Again, we set $\alpha = 200$ such that DP prior realizations for G are close to the Pareto G_0 distribution,

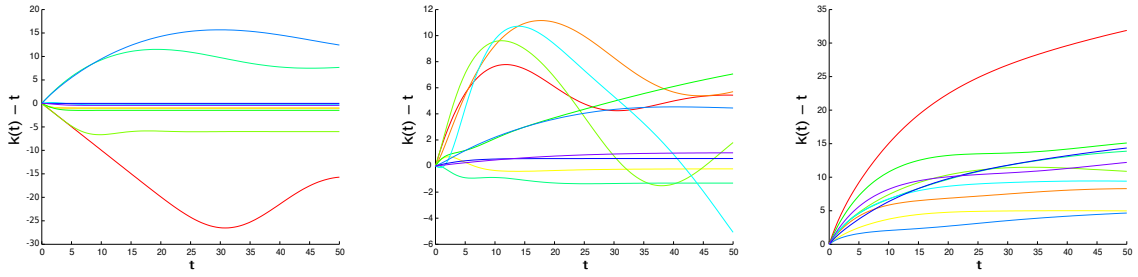


(a) (b) (c)

Figure 3: Realizations for the function $K(t) - t$ based on 10 samples from the DP(α, G_0) prior for the mixture weights, with a Pareto centering distribution and $\alpha = 200$. From left to right: $G_0 =$ Pareto$(2.5, 1)$, $G_0 =$ Pareto$(2.5, 5)$ and $G_0 =$ Pareto$(0.01, 5)$.

and $\theta = 2$. for $a = 2.5$ and $b = 1$, G_0 is a long-tailed Pareto distribution resulting in point processes with clustering patterns; see Figure 3a. For $a = 2.5$ and $b = 5$, the Pareto distribution is less dispersed and the point processes show declustering patterns, with substantially reduced variability in the K function samples; see Figure 3b. as for the Weibull case, we observe that the mixture model with the Pareto DP centering distribution can generate both clustering and declustering patterns. It is also possible to produce K functions that resemble the homogeneous Poisson process case, with very small variability. As an example, Figure 3c corresponds to $\theta = 1$, $a = 0.01$ and $b = 5$, and this is generally the case for $\theta > a$ and large b. The parameter settings and results are summarized in Table 1.

In addition to the two choices of G_0 discussed above, we considered gamma, log-normal and generalized Pareto distributions. From the empirical exploration of the K functions produced by such distributions, we conjecture that the clustering or declustering behavior of the point process is determined by the parameter that controls the tails of the DP centering distribution, which can be quantified through either values for the coefficient of variation or monotonicity of the hazard function of G_0. This observation can be used to guide the selection of G_0 for modeling purposes. Our preferred choice is the Weibull distribution, since the empirical investigation suggests a clear connection between the shape parameter and clustering/declustering patterns, and it also has an analytically available distribution function.



(a) (b) (c)

Figure 4: Realizations for the function $K(t) - t$ based on 10 samples from the DP(α, G_0) prior for the mixture weights, where $G_0 =$ Weibull$(\mu = 5, \phi = 0.5)$ and, from left to right, $\alpha = 0.2, 2, 50$.

The DP precision parameter. Here, we set $\theta = 2$ and take a Weibull$(5, 0.5)$ distribution for G_0, which is the first setting considered in the study for the role of shape parameter ϕ. In Figure 4, we plot realizations for the K function under three different values for the DP precision parameter, $\alpha = 0.2, 2, 50$. for $\alpha = 50$, we observe that the model supports only clustering patterns, which is in accordance with results discussed earlier for large α and Weibull shape parameter $\phi < 1$; compare Figure 4c and Figure 2a. The variability in the values of the K function is larger for smaller values of α, which is consistent with its role in controlling the variance of the DP prior. For small values of α, both clustering and declustering patterns become possible, although clustering is more strongly favored. Changing G_0 to a Weibull$(5, 2)$ distribution, we observe the same behavior where now declustering patterns are favored.

In the absence of prior information about clustering/declustering patterns, we recommend using a prior for α that is concentrated on small to moderate values. We also place a prior on the shape parameter of the Weibull DP centering distribution that supports values both less than and greater than 1.

2.5 The approach to posterior inference

Here, we discuss the inference methods for the inter-arrival density, as well as for other functionals of the renewal process, given the observed point pattern $\{0 = t_0 < t_1 < t_2 < \cdots < t_n < T\}$.

As mentioned in Section 2.2, key to posterior simulation is the introduction of latent variables

that identify the mixture component to which each observation is assigned. In particular, we consider variables $\{y_1, ..., y_n, y_{n+1}\}$, which given G are i.i.d. from G. The first n of these latent variables are associated with the likelihood components $f(t_i - t_{i-1})$ that comprise the product in (1). The connection is given by the following representation of the Erlang mixture model for the inter-arrival density: $\sum_{j=1}^{J} w_j \, \text{ga}(x \mid j, \theta^{-1}) = \int \left\{ \sum_{j=1}^{J} \mathbf{1}_{((j-1)\theta, j\theta]}(y) \, \text{ga}(x \mid j, \theta^{-1}) \right\} \text{d}G(y)$. Regarding the normalizing constant in (1), under the Erlang mixture model, we can write

$$1 - \int_{t_n}^{T} f(u - t_n) \text{d}u = \sum_{j=1}^{J} w_j \, C_j(t_n, \theta) = \int \left\{ \sum_{j=1}^{J} \mathbf{1}_{((j-1)\theta, j\theta]}(y) \, C_j(t_n, \theta) \right\} \text{d}G(y)$$

where $C_j(t_n, \theta) = 1 - \int_{t_n}^{T} \text{ga}(u - t_n \mid j, \theta^{-1}) \text{d}u$ is defined in terms of gamma distribution function values and depends on only one of the model parameters. Hence, the last latent variable, y_{n+1}, is used to represent the integral in the likelihood normalizing constant in a similar fashion with the more standard density components of the likelihood.

Therefore, conditional on the latent variables $\{y_1, ..., y_n, y_{n+1}\}$ and (θ, J), the first stage of the hierarchical model for the data $\{t_1, ..., t_n\}$ can be expressed as

$$\prod_{i=1}^{n} \sum_{j=1}^{J} \mathbf{1}_{((j-1)\theta, j\theta]}(y_i) \, \text{ga}(t_i - t_{i-1} \mid j, \theta^{-1}) \times \left\{ \sum_{j=1}^{J} \mathbf{1}_{((j-1)\theta, j\theta]}(y_{n+1}) \, C_j(t_n, \theta) \right\} \tag{6}$$

where $y_i \mid G \overset{i.i.d.}{\sim} G$, for $i = 1, ..., n, n+1$, and $G \mid \alpha, \mu, \phi \sim \text{DP}(\alpha, \text{Weibull}(\mu, \phi))$. The full Bayesian model is completed with priors for (θ, J) (given in Section 2.2) and for the DP parameters. in particular, we assign a gamma(a_α, b_α) prior to α, an inverse gamma, IG(a_μ, b_μ), prior to μ (with mean $b_\mu/(a_\mu - 1)$, provided $a_\mu > 1$), and a gamma(a_ϕ, b_ϕ) prior to ϕ.

Since the hierarchical model for the data has the structure of a DP mixture model, we can utilize any of the available posterior simulation methods for DP mixtures. We opt for a marginal MCMC algorithm which builds from the Pólya urn based joint prior distribution for the y_i that arises after marginalizing G over its DP prior (e.g., Escobar and West, 1995; Neal, 2000), adding sampling from the posterior distribution of G (Gelfand and Kottas, 2002) to enable full inference for renewal process functionals. The MCMC algorithm details are provided in the Appendix.

The conditional posterior distribution for G, given (α, μ, ϕ) and $\{y_1, ..., y_n, y_{n+1}\}$, is given by a DP with updated precision parameter $\alpha^* = \alpha + n + 1$ and centering distribution $G_0^* =$

$\alpha(\alpha+n+1)^{-1}\text{Weibull}(\mu,\phi)+(\alpha+n+1)^{-1}\sum_{i=1}^{n+1}\delta_{y_i}$. Using the DP definition, this implies that the conditional posterior distribution for the vector of mixture weights $(w_1, ..., w_J)$ is a Dirichlet distribution (with parameters defined by α^* and G_0^*). Hence, given the posterior samples for (α, μ, ϕ) and $\{y_1, \cdots, y_n, y_{n+1}\}$ (obtained as described in the Appendix), we can readily obtain posterior samples for the mixture weights in (2). Those posterior samples, combined with the ones for θ and J, provide full posterior inference for the inter-arrival density over any grid of time points. Moreover, using the inverse Laplace transform of Equation (5) evaluated at the posterior samples for θ, J and $(w_1, ..., w_J)$, we obtain posterior realizations for the renewal function $M(t)$, and therefore also for the K function, $K(t) = \mu_1 M(t)$, with $\mu_1 = \theta\sum_{j=1}^{J} j w_j$.

In addition to inference for the K function, the posterior realizations for the inter-arrival density can be used to estimate the hazard rate function:

$$\lambda(t \mid H(t)) = \frac{f(t - H(t))}{1 - F(t - H(t))}, \quad t > H(t) \tag{7}$$

where F is the inter-arrival distribution function. This also defines the point process conditional intensity function. Its value can be interpreted as the probability that the next event will occur in an infinitesimal interval after time point t, conditional on the history of the process up to time t. For a renewal process, such history involves only $H(t)$, the time of the last event prior to t. Moreover, for a homogeneous renewal process, the conditional intensity function depends on t and $H(t)$ only through the time difference $t - H(t)$. We can therefore visualize the conditional intensity function in an one-dimensional plot by using $t - H(t)$ as the argument for the horizontal axis. We illustrate inference for the hazard rate (conditional intensity) function with the real data examples in Section 4.

It is noteworthy that the approach to posterior inference does not require truncation for the DP countable representation. This is due to the use of a marginal MCMC algorithm and the fact that the only functional of random distribution G needed for our inferences is the discretized version of the corresponding random distribution function that defines the mixture weights.

2.6 Model checking

The modeling approach is based on the renewal process assumption for the observed point pattern $\{0 = t_0 < t_1 < t_2 < \cdots < t_n < T\}$. A means to check this assumption, as well as

the model for the inter-arrival distribution, is given by the Time-Rescaling theorem (e.g., Daley and Vere-Jones, 2003). Assume that the point pattern is a realization from a point process with conditional cumulative intensity function $\Lambda^*(t)$, defined through the integral of the conditional intensity function over $(0, t)$ (the dependence on the history of the process up to time t is suppressed from the notation). Then, based on the Time-Rescaling theorem, the transformed point pattern $\{\Lambda^*(t_i) : i = 1, ..., n\}$ is a realization from a unit rate homogeneous Poisson process, and thus the random variables $U_i = 1 - \exp\{-(\Lambda^*(t_i) - \Lambda^*(t_{i-1}))\}$, where $\Lambda^*(0) \equiv 0$, are independently and uniformly distributed on $(0, 1)$. Therefore, graphical model assessment can be based on quantile-quantile (Q-Q) plots to evaluate agreement of the estimated U_i with the uniform distribution on $(0, 1)$.

In the context of a renewal process with conditional intensity function given by (7), we have $\Lambda^*(t_i) - \Lambda^*(t_{i-1}) = \int_{t_{i-1}}^{t_i} \lambda(s \mid t_{i-1})\mathrm{d}s = -\log\{1 - F(t_i - t_{i-1})\}$, and thus $U_i = F(t_i - t_{i-1}) = \int_{t_{i-1}}^{t_i} f(s - t_{i-1})\mathrm{d}s$. Hence, the posterior realizations for the inter-arrival density can be used to obtain posterior samples for each U_i, which result in posterior point and interval bands for the Q-Q graphs. This model assessment technique is applied to the real data examples of Section 4.

Time-rescaling diagnostics involve checking of the fit provided by the renewal process assumption for the stochastic point process mechanism, as well as of the particular model for the inter-arrival distribution. In cases of discrepancy from the uniform distribution, it is not straightforward to distinguish which of the two assumptions contributes more to lack of model fit. The theoretical support for Erlang mixtures, as well as the model flexibility arising from the nonparametric prior for the mixture weights are practically useful in this regard. A flexible model, such as the Erlang mixture, that supports general shapes for the inter-arrival distribution is more suitable than standard parametric models with respect to focusing the goodness-of-fit evaluation on the renewal process assumption.

3 Synthetic Data Examples

In this section we evaluate our model by fitting it to synthetic data corresponding to both declustering and clustering patterns (Sections 3.1 and 3.2), including comparison with a parametric Weibull model for the renewal process inter-arrival distribution (Section 3.3).

3.1 Declustering examples

We first consider two synthetic data examples that correspond to declustering patterns. The inter-arrival distribution for the first case is Weibull$(2, 1.5)$, whereas for the second it is given by a mixture of inverse Gaussian distributions, 0.42 InvGaussian$(0.5, 1) + 0.58$ InvGaussian$(2, 6.5)$. The Weibull distribution has shape parameter equal to 1.5, corresponding to a coefficient of variation that is less than 1 and a monotonically increasing hazard function, and resulting in a renewal process with a declustering pattern. A sequence of 269 data points are generated within the time window $(0, 500)$. A snap shot of the data points is shown in Figure 5a. The second inter-arrival density is motivated by the fact that inverse Gaussian distributions are a popular choice for declustering renewal process models. We generate a sequence of 366 data points within the time window $(0, 500)$, a snap shot of which is shown in Figure 5d.

In Figure 5, we compare the true $K(t) - t$ function as well as the true density function of inter-arrival times with the estimates from our model. For both functions and both data examples, the posterior mean estimate captures the shape of the corresponding function, and the posterior interval bands contain the true function. For both data examples, our proposed method accurately estimates the declustering property of the renewal processes, as demonstrated by the inference results for the corresponding K functions.

We note here that the evaluation of the K function can be a complicated task, even in full knowledge of the parametric model for the inter-arrival times. For the Weibull inter-arrival distribution, we first calculate the renewal function $M(t)$, using an approximation method from Smith and Leadbetter (1963), and then obtain the K function from $K(t) = \mu_1 M(t)$, where μ_1 is the mean of the Weibull distribution. For the mixture of inverse Gaussian distributions, we use the Laplace transform of an inverse Gaussian, which is available in closed form. The inverse Laplace transform is then obtained numerically.

3.2 Clustering examples

Here, we consider two synthetic data sets from renewal processes that exhibit clustering. The respective inter-arrival distributions are given by a Pareto$(1, 1.98)$ and a Weibull$(2.3, 0.6)$ distribution. The Pareto distribution has finite mean as its shape parameter is greater than 1. Both

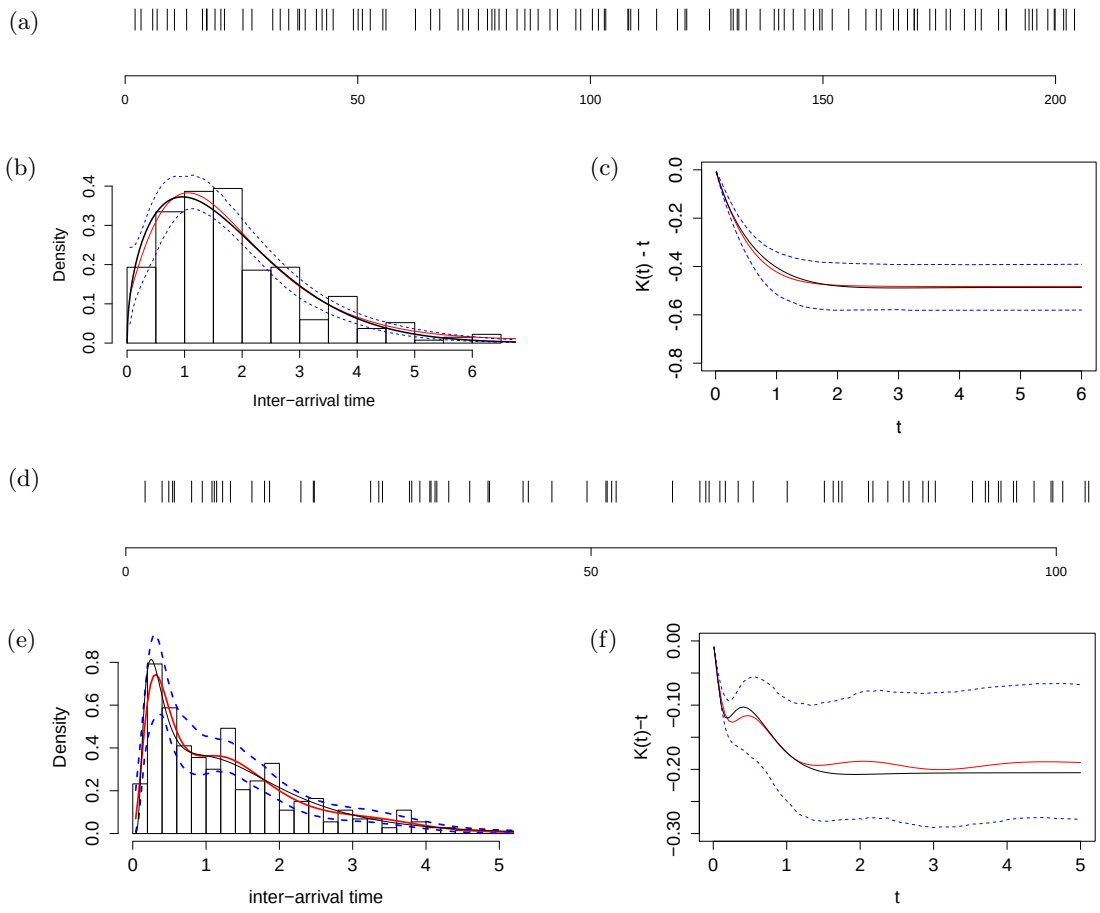


Figure 5: Declustering synthetic data examples. 5a and 5d: A snapshot of the renewal process realizations based on the Weibull and two-component inverse Gaussian mixture inter-arrival densities, respectively; 5b and 5e: Histogram of simulated inter-arrival times, true inter-arrival density (black line), posterior mean estimate (red line), and 95% interval bands (dashed blue lines) for the first and second example, respectively. 5c and 5f: True $K(t) - t$ function (black line), posterior mean estimate (red line), and 95% interval bands (dashed blue lines) for the first and second example, respectively.

distributions have decreasing hazard rates, resulting in renewal processes with clustering. For the Pareto-based example, 556 data points were generated within the time window $(0, 500)$, with a snapshot shown in Figure 6a. For the Weibull-based example, the simulated point pattern comprised 664 data points within the time window $(0, 2000)$, with a snapshot plotted in Figure 7a. We select these two distributions to test our model's ability to handle data arising from

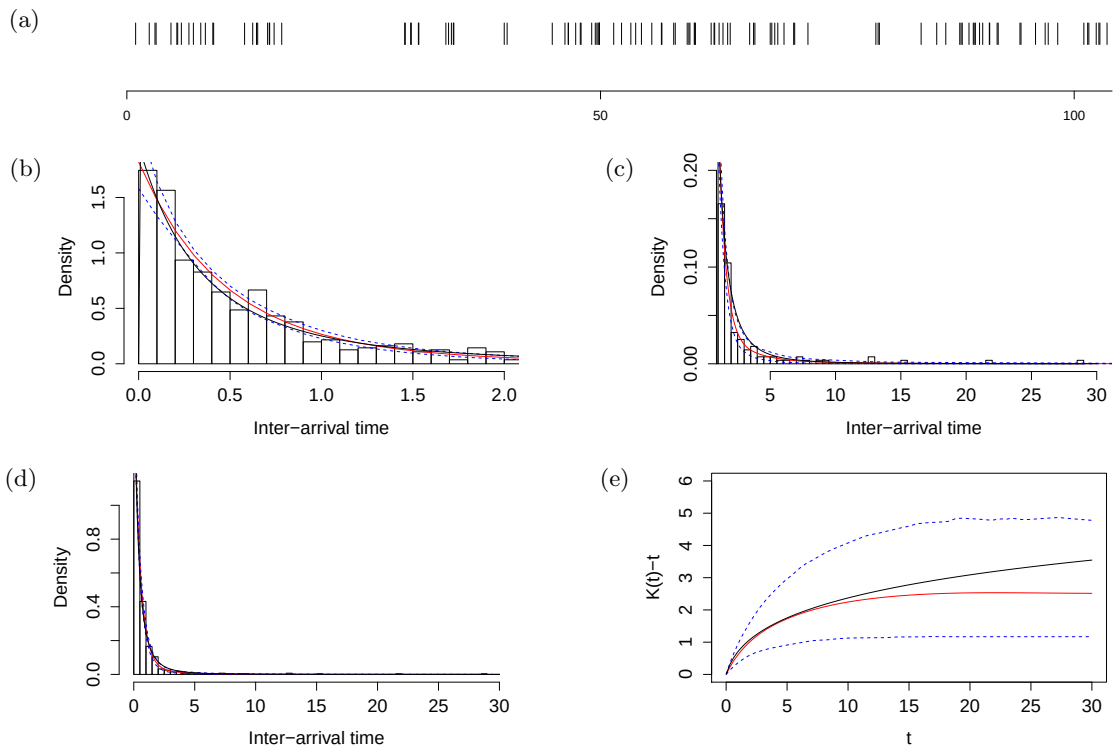


Figure 6: Clustering synthetic data example with Pareto inter-arrival density. 6a: A snapshot of the renewal process realization. 6b, 6c, and 6d: histogram of simulated inter-arrival times, true inter-arrival density function (black line), posterior mean estimate (red line), and 95% interval bands (dashed blue lines) for the head of the data, the tail of the data, and all data, respectively. 6e: True $K(t) - t$ function (black line), posterior mean estimate (red line), and 95% interval bands (dashed blue lines).

inter-arrival densities with both polynomial and exponential tail behaviors.

Figures 6 and 7 show posterior mean and interval estimates for the inter-arrival densities and $K(t) - t$ functions. To compute the true K function for the Pareto-based renewal process, we used the Laplace transform of the Pareto distribution reported in Nadarajah and Kotz (2006) and evaluated the Laplace transform of the renewal function using Equation (3). In order to more effectively visualize results for the inter-arrival density, we plot estimates for the head and tail of the data, as well as for the density over the full range of the data. As expected based on the more extreme tails of the true inter-arrival densities, the point estimates for the densities and the corresponding K functions are less accurate than the declustering examples

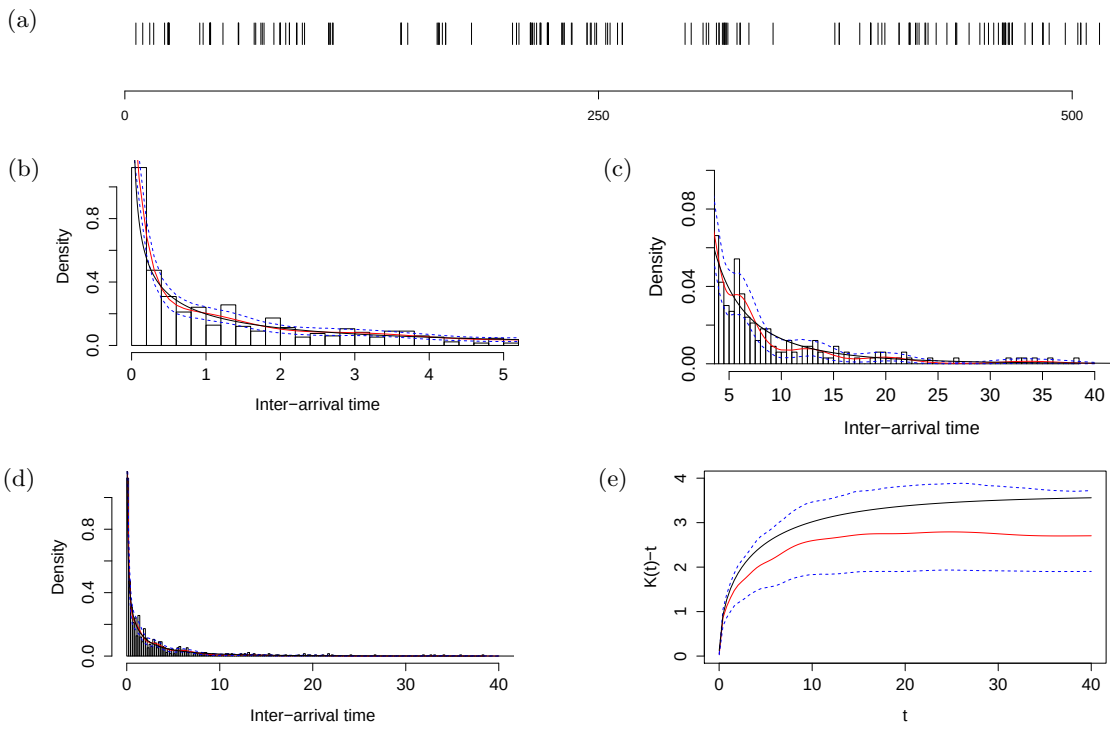


Figure 7: Clustering synthetic data example with Weibull inter-arrival density. 6a: A snapshot of the renewal process realization. 6b, 6c and 6d: histogram of simulated inter-arrival times, true inter-arrival density (black line), posterior mean estimate (red line), and 95% interval bands (dashed blue lines) for the head of the data, the tail of the data, and all data, respectively. 6e: True $K(t) - t$ function (black line), posterior mean estimate (red line), and 95% interval bands (dashed blue lines).

of the previous section. However, the posterior interval bands contain the true inter-arrival density and K function for both the Pareto-based and Weibull-based example, thus uncovering the underlying renewal process clustering structure.

3.3 Comparison with parametric Weibull model

Here, we report results from a small simulation study to compare the Erlang mixture and Weibull models for the inter-arrival distribution; as discussed in the Introduction, the latter is a commonly used parametric model for renewal processes. The posterior distribution for the two parameters of the Weibull model was sampled using a Metropolis-Hastings MCMC algorithm.

Table 2: Simulation study for comparison between the Erlang mixture and Weibull models. Posterior mean and 95% interval estimates for the TVD between the true inter-arrival distribution function and the model estimates. Results are given for the four inter-arrival distributions considered in Sections 3.1 and 3.2 – Weibull$(2, 1.5)$ (DCWeibull), 0.42 InvGaussian$(0.5, 1)$ + 0.58 InvGaussian$(2, 6.5)$ (MixtInvG), Pareto$(1, 1.98)$ (Pareto), Weibull$(2.3, 0.6)$ (CWeibull) – and for two data sets of different size.

	Erlang Mixture	Parametric Weibull
true inter-arrival distribution	True-Posterior TVD	True-Posterior TVD
DCWeibull ($T = 500$, $n = 269$)	0.036 $(0.010, 0.076)$	0.029 $(0.004, 0.069)$
mixtInvG ($T = 500$, $n = 366$)	0.034 $(0.015, 0.065)$	0.043 $(0.032, 0.070)$
pareto ($T = 500$, $n = 556$)	0.033 $(0.012, 0.060)$	0.046 $(0.032, 0.068)$
CWeibull ($T = 2000$, $n = 664$)	0.040 $(0.021, 0.065)$	0.029 $(0.006, 0.057)$
DCWeibull ($T = 1500$, $n = 856$)	0.023 $(0.010, 0.043)$	0.021 $(0.005, 0.042)$
mixtInvG ($T = 1500$, $n = 1115$)	0.034 $(0.013, 0.059)$	0.037 $(0.032, 0.046)$
pareto ($T = 1500$, $n = 1490$)	0.025 $(0.014, 0.039)$	0.050 $(0.039, 0.062)$
CWeibull ($T = 6000$, $n = 1707$)	0.027 $(0.018, 0.043)$	0.011 $(0.002, 0.027)$

We work with synthetic data from the four inter-arrival distributions considered in Sections 3.1 and 3.2. For each of the four cases, we provide results for the same data set used in Section 3.1 or 3.2, as well as for a data set with larger sample size (given in Table 2) resulting from expanding the observation time window from $(0, T)$ to $(0, 3T)$. We use the total variation distance (TVD), $\sup_x |F_{\text{true}}(x) - F_b^*(x)|$, to assess estimation performance. Here, F_{true} denotes the true inter-arrival distribution function, and F_b^*, for $b = 1, ..., B$, the posterior realizations under each model, where B is the posterior sample size. Therefore, we obtain posterior distributions for the TVD, which are summarized in Table 2 through posterior means and 95% interval estimates. We use conservative values for the effective support of each true inter-arrival distribution, such that the supremum in the TVD definition can be accurately approximated.

The results in Table 2 are compatible with results from contrasting Bayesian nonparametric and parametric models in other settings. In particular, for the cases where the data are simulated from a Weibull inter-arrival distribution, the Weibull model performs better. However, when the data generating mechanism is different from the parametric model, the Erlang mixture model

shows superior performance. Note that in the Pareto case (corresponding to polynomial tails for the inter-arrival distribution), the Weibull model does not fare better with the larger sample size. In all cases, the Erlang mixture model performance improves with increasing sample size. These results suggest that the proposed inference method for renewal processes is more robust than methods based on standard parametric models.

4 Data Applications

Here, we apply the methodology to two real data examples: the commonly studied coal mining disasters data, and times of earthquake occurrences from four different regions of the Americas. In fact, in Section 4.2, we extend the Erlang mixture model to estimate hierarchically related renewal processes corresponding to the four regions from which earthquake occurrences have been recorded.

4.1 Coal-mining disasters data

The "coal-mining disasters" data set from Jarrett (1979) is commonly used in the literature of point processes, with most approaches treating the data as a realization of an inhomogeneous Poisson process. The point pattern records the times (in days) of 191 explosions of fire-damp or coal-dust in mines leading to accidents, involving 10 or more men killed, over the time period from 15 March 1851 to 22 March 1962. To illustrate our methodology, we consider two versions of the coal-mining disasters data.

The first involves the subset of the data corresponding to the time period from 15 March 1851 to 31 December 1875, which involves 81 events. Graphical exploration of this data subset indicates the presence of a regular pattern compatible with the assumption of time-homogeneity (see Figure 8a), which is implicit in the class of renewal processes we study. Here, the priors for the model parameters are: $\phi \sim \text{gamma}(4,\, 2)$, $(\theta, J) \sim \text{gamma}(1, 0.02)\text{Unif}(\{800/\theta, \ldots, 8000/\theta\})$, $\mu \sim \text{IG}(6, 800)$, and $\alpha \sim \text{gamma}(2, 4)$. The Erlang mixture model estimates capture the decreasing inter-arrival density suggested by the data (Figure 8b). in Figure 8c, we plot the posterior distribution for the largest mixture weight, $w^{(1)}$, and for the sum of the two largest mixture

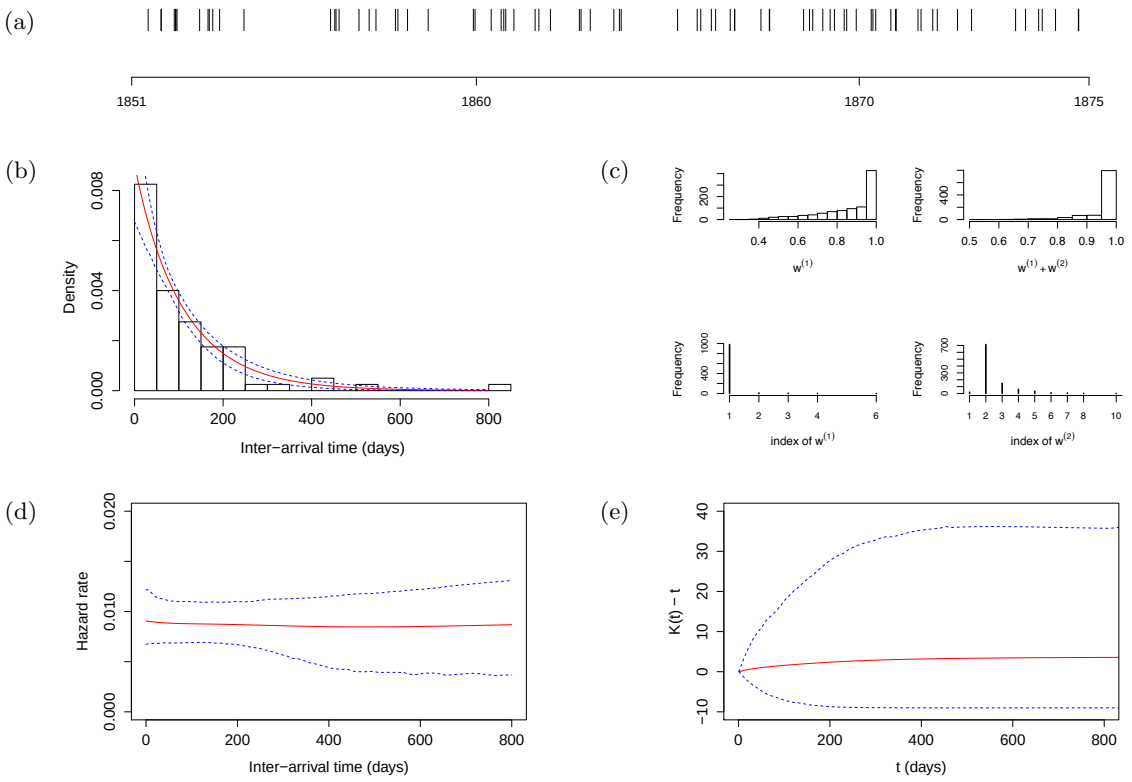


Figure 8: 8a: Subset of the coal-mining disasters data from 1851 to 1875. 8b: Posterior mean estimate (red line) and 95% interval bands (dashed blue lines) for the inter-arrival density, overlaid on the histogram of observed inter-arrival times. 8c: Posterior distribution of the two largest weights and the corresponding Erlang mixture component indexes. 8d and 8e: Posterior mean estimate (red line) and 95% interval bands (dashed blue lines) for the hazard function (8d) and the $K(t) - t$ function (8e).

weights, $w^{(1)} + w^{(2)}$. The respective posterior means are 0.86 and 0.967. Hence, given the data, the Erlang mixture for the inter-arrival density is mostly concentrated on two components. With posterior probability 0.98, the index of the largest component corresponds to the first Erlang distribution. Hence, the model favors an exponential inter-arrival distribution, which corresponds to a homogeneous Poisson process. The posterior mean estimate of the hazard rate (conditional intensity) function (Figure 8d) is essentially constant over time, and the posterior mean estimate for the K function nearly equals t, both of which provide further evidence that the particular subset of the coal-mining disasters data corresponds to a homogeneous Poisson process.

For a further illustration, we reduce the size of the point pattern by retaining every al-

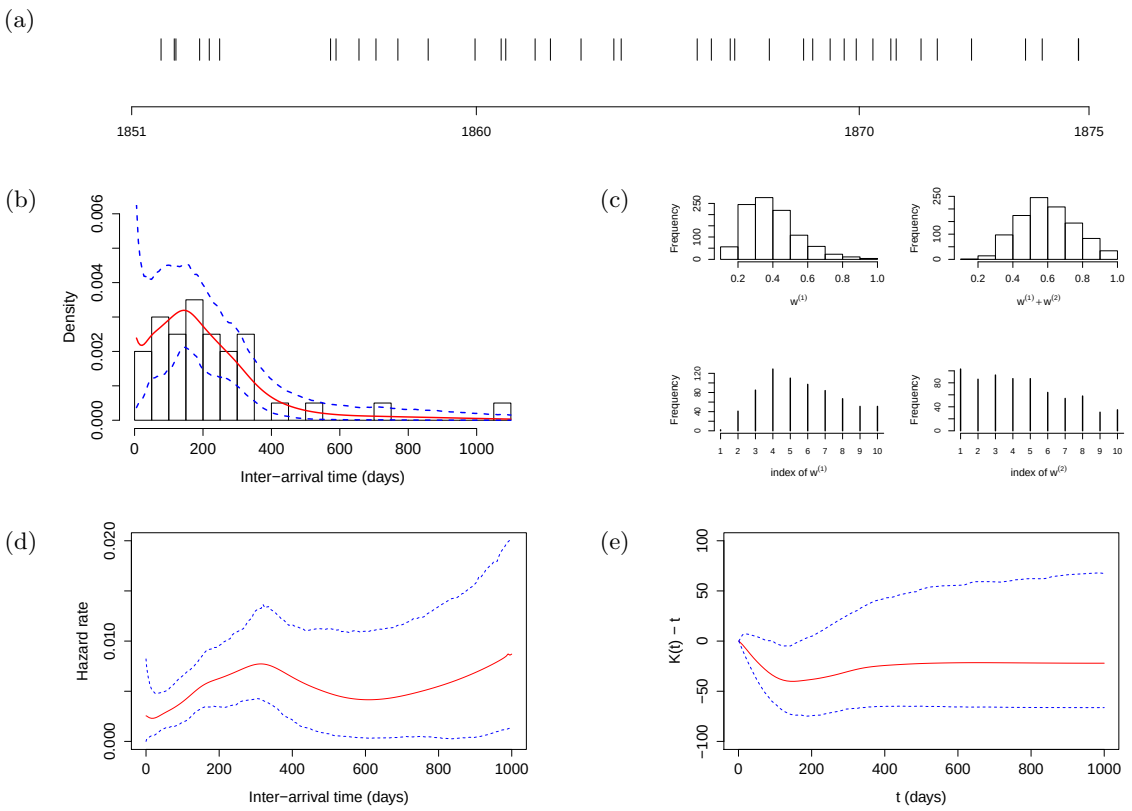


Figure 9: 9a: Thinned version of the subset of the coal-mining disasters from 1851 to 1875. 9b: Posterior mean estimate (red line) and 95% interval bands (dashed blue lines) for the inter-arrival density, overlaid on the histogram of observed inter-arrival times. 9c: Posterior distribution of the two largest weights and the corresponding Erlang mixture component indexes. 9d and 9e: Posterior mean estimate (red line) and 95% interval bands (dashed blue lines) for the hazard function (9d) and the $K(t) - t$ function (9e).

ternate event from the 81 coal-mining disasters that occurred between 1851 and 1875; see Figure 9a. this data set is obtained from thinning the previously analyzed point pattern for which the Poisson process assumption appears plausible. Therefore, the thinned point pattern should have a declustering property, which is challenging to detect given the relatively small number of observed events. The prior distributions in this case are: $\phi \sim$ gamma(4, 2), $(\theta, J) \sim$ gamma$(1, 0.1)$Unif$(\{1100/\theta, \ldots, 3300/\theta\})$, $\mu \sim$ IG$(3, 200)$, and $\alpha \sim$ gamma$(2, 1)$. Here, the Erlang mixture model estimates a unimodal inter-arrival density, with wider posterior uncertainty bands, as expected given the smaller sample size (Figure 9b). We plot again the posterior

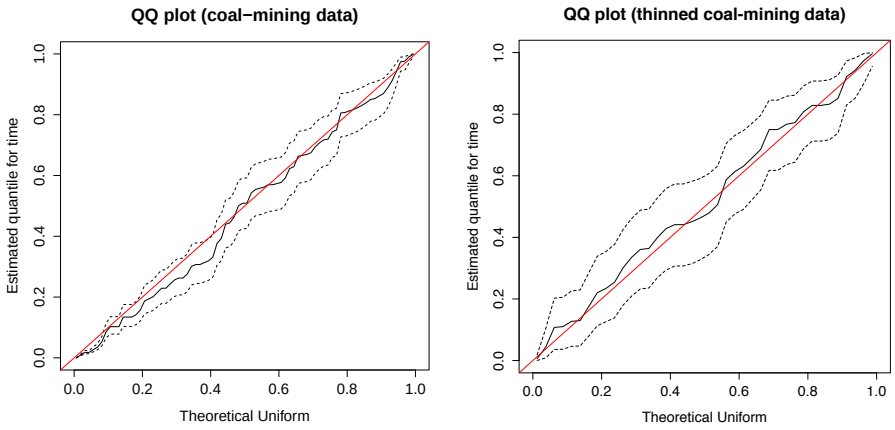


Figure 10: Coal-mining disasters data. Posterior mean and 95% interval bands for the Time-Rescaling model checking Q-Q plots for the subset of the data from 1851 to 1875 (left panel), and for the thinned version of this data subset (right panel).

distribution for the two largest mixture weights; see Figure 9c. In this case, the posterior mean for $w^{(1)} + w^{(2)}$ is 0.62, and the indexes for the two largest mixture weights are spread over several Erlang components. The estimates for the conditional intensity function reveal a time varying, non-monotonic shape (Figure 9d). Finally, Figure 9e shows the estimates for the $K(t) - t$ function. Although the posterior mean estimate supports declustering, the same conclusion can not be drawn from the entire posterior distribution as the 95% interval bands for the $K(t) - t$ function include the line at 0 over almost the entire interval of observed inter-arrival times.

Finally, Figure 10 shows results from the model checking method discussed in Section 2.6. The plots offer graphical evidence that the renewal process model with the Erlang mixture inter-arrival distribution provides a reasonable fit for the coal-mining disasters data subsets considered in this section.

4.2 A hierarchical model for earthquake occurrences data

Some of the early work in seismology on modeling earthquake occurrences was based on Poisson process assumptions (e.g., Cornell, 1968; Caputo, 1974). In order to model clustering foreshock-mainshock-aftershock sequences, other stochastic models, such as compound Poisson processes, Hawkes processes, renewal processes and Markov processes, have been explored (e.g., Anagnos

and Kiremidjian, 1988; Ogata, 1988; Zhuang et al., 2002; Parsons, 2008).

To motivate a hierarchical extension of the Erlang mixture model that will allow us to study the effect of the different levels of a factor on the distribution of inter-arrival times, we consider the times of earthquake occurrences in four distinct regions of the Americas – North America, Central America, Caribbean, South America – recorded from January 1, 1900 to present. The data is taken from the "Significant Earthquake Database" of the NOAA, National Centers for Environmental information (NGDC/WDS, 2019). Occurrences are recorded for destructive earthquakes that meet at least one of the following criteria: moderate damage (approximately $1 million or more), 10 or more deaths, magnitude 7.5 or greater, Modified Mercalli Intensity X or greater, or the earthquake generated a tsunami. The data set includes 357 earthquakes for South America, 303 for North America, 126 for Central America, and 55 for the Caribbean.

Here, the point patterns are $\{0 < t_{1,r} < t_{2,r} < \cdots < t_{n_r,r} < T\}$, where $r = 1, ..., 4$ indicates the region. We extend the Erlang mixture model structure such that the inter-arrival density for region r is given by (2) with region-specific scale parameters θ_r and weights generating distributions G_r, which are assigned hierarchical priors. In particular, we assume

$$\theta_r \mid \gamma \overset{i.i.d.}{\sim} \text{gamma}(a_\theta, \gamma) \quad \text{and} \quad G_r \mid \alpha_r, \mu, \phi \overset{ind.}{\sim} \text{DP}(\alpha_r, G_0 = \text{Weibull}(\mu, \phi)), \quad r = 1, ..., 4,$$

where $\alpha_r \mid \beta \overset{i.i.d.}{\sim} \text{gamma}(a_\alpha, \beta)$, for $r = 1, ..., 4$, with random β and γ. Assuming a common DP centering distribution, G_0, and different precision parameters, α_r, leverages the role of the DP prior parameters in the context of our modeling approach. This model formulation retains the flexible Erlang mixture inter-arrival density for each region, while allowing for borrowing of strength across regions through the hierarchical priors for hyperparameters θ_r and α_r. The model can include also region-specific number of mixture components, although a common J strikes a good balance between model parsimony and flexibility.

The hierarchical model is implemented with a relatively straightforward extension of the posterior simulation method presented in Section 2.5 and in the Appendix. Regarding the specific priors for the earthquake data analysis, we set $a_\alpha = 2$, $a_\theta = 10$, and place gamma$(5, 28)$ and gamma$(7, 6)$ priors on β and γ, respectively. Moreover, we use a gamma$(0.08, 0.1)$ prior for ϕ, and an IG$(2, 500)$ prior for μ. Finally, we set $J = 200$, using the prior specification approach discussed in Section 2.2.

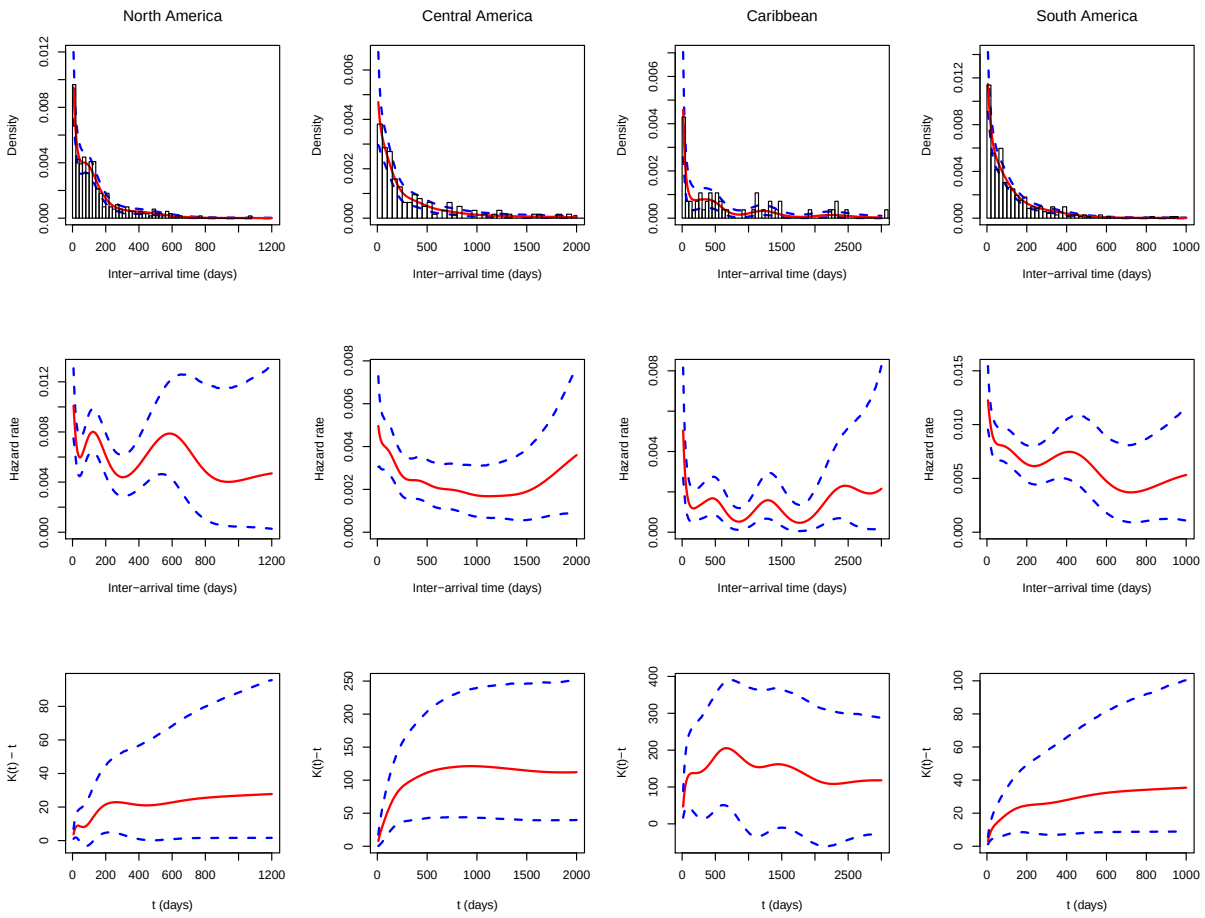


Figure 11: Earthquakes data. Posterior mean estimate (red line) and 95% interval bands (dashed blue lines) for the inter-arrival density (top row), the hazard function (middle row), and the $K(t) - t$ function (bottom row). The top row panels include the observed inter-arrival times. The columns correspond to the four different regions of the Americas.

Figure 11 plots estimates for renewal process functionals for each of the four regions. The model captures well the distinct and non-standard inter-arrival density shapes suggested by the data, and it also estimates non-linear hazard rate functions across the four regions. Model inferences support clustering patterns essentially for all four regions, the only possible exception being the Caribbean for which the posterior mean estimate for the $K(t) - t$ function takes decisively positive values, but the 95% interval bands include the line at 0 for inter-arrival time values greater than about 1000 days. Central America stands out in terms of marginal posterior

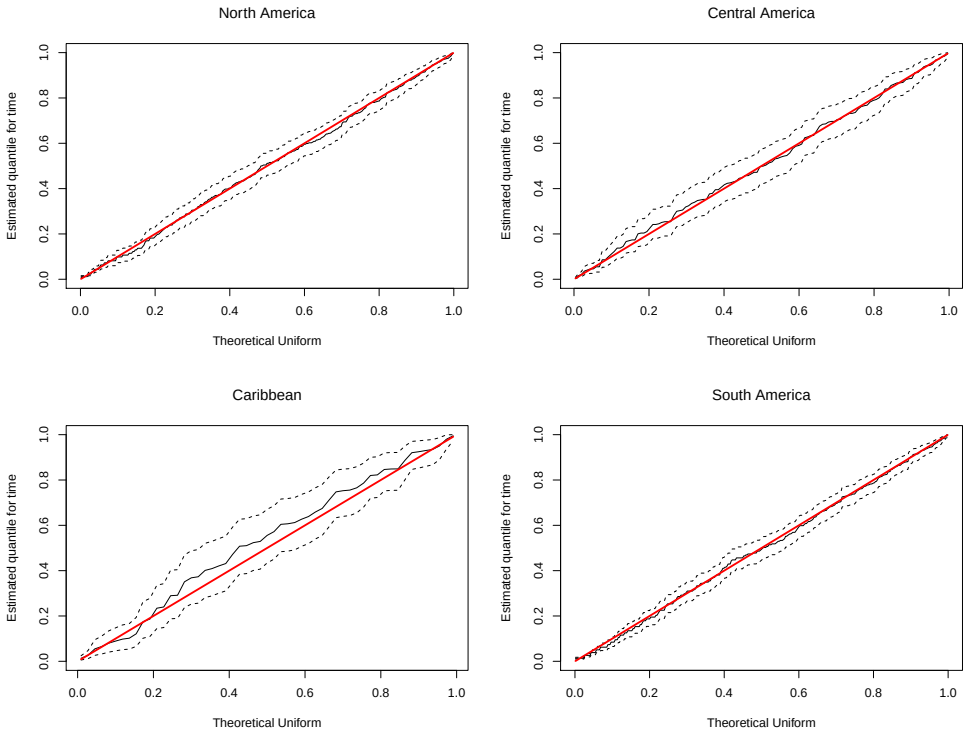


Figure 12: Earthquakes data. Posterior mean and 95% interval bands for the Time-Rescaling model checking Q-Q plots for the data from each of the four regions of the Americas.

densities for the θ_r and α_r parameters (not shown), as they are supported by larger values compared to the other three regions. This is compatible with the role of these model parameters noting that the data for Central America suggest a heavier tailed inter-arrival distribution than North and South America, and at the same time, a more standard inter-arrival density than the multimodal density for the Caribbean. Finally, the Time-Rescaling model checking method provides graphical support for good model fit (see Figure 12).

5 Discussion

We have developed a Bayesian nonparametric approach to modeling and inference for renewal processes. The inter-arrival time density of the renewal process is modeled as a mixture of Erlang distributions with common scale parameter. The mixture weights are defined through

increments of a distribution function which is modeled with a Dirichlet process prior. The nonparametric prior for the mixture weights allows non-standard shapes for the inter-arrival density, as well as for other key functionals of the renewal process, such as the K function and the conditional intensity function. In particular, both clustering and declustering patterns of the renewal processes can be captured by the model. The inferential flexibility of the modeling approach has been studied through synthetic data sets, as well as two real data examples.

Modeling with renewal point processes provides a means to relax the Poisson process assumption, in particular, in studying temporal point patterns with clustering or declustering properties. However, the standard form of renewal processes considered in this work is based on the assumption that all inter-arrival times follow the same distribution. Such time homogeneity property might apply to, for instance, earthquake data in the long term. However, in practice, it is more reasonable to assume that the rate of earthquake occurrences varies over time. In applications of point processes in reliability analysis, the "as good as before, after the repair" property, implied by homogeneous renewal processes, may be suspect. A class of stochastic models that has been explored for analysis of repairable systems is the trend-renewal process (Lindqvist et al., 2003; Lindqvist, 2006; Xu et al., 2017). Trend-renewal processes include both inhomogeneous Poisson processes and homogeneous renewal processes as special cases.

The scope of renewal process models can also be extended by allowing the conditional intensity function in (7) to depend on both time t and $H(t)$, the time of the last event prior to t. In a generalized/modulated renewal process (Cox, 1972), the intensity function is typically written as a function of t and $t - H(t)$. Such point process models have been explored in the neuronal science literature (e.g., Kass and Ventura, 2001; Barbieri et al., 2001; Koyama and Kass, 2008), as well as in the literature on systems reliability (e.g., Brown and Proschan, 1983; Lawless and Thiagarajah, 1996), although inference is typically based on potentially restrictive parametric forms for the inter-arrival density and the corresponding conditional intensity function, $\lambda(t, t - H(t))$.

A possible extension of our modeling approach for modulated renewal processes involves introduction of time-varying weights in the Erlang mixture model for the inter-arrival density. This can be accomplished with a time-dependent nonparametric prior for the random distribu-

tion function used to define the mixture weights. Such an extension will allow the inter-arrival density to depend on the clock time t (in addition to $t - H(t)$), retaining the model's ability to capture general shapes for the renewal function and the conditional intensity function that are not restricted by a particular parametric form. The challenge involves developing model formulations for the mixture weights that achieve a good balance between inferential flexibility and computational feasibility, especially for point patterns of small to moderate size.

Appendix: MCMC posterior simulation

Here, we provide the details of the MCMC algorithm discussed in Section 2.5. Again, key to posterior simulation is introduction of the latent variables $\{y_1, ..., y_n, y_{n+1}\}$, under which the augmented likelihood for the data, $\{0 = t_0 < t_1 < \cdots < t_n < T\}$, assumes the form in (6).

After marginalizing G over its DP prior, the resulting joint prior for the latent variables, given α, μ and ϕ, implies the following prior full conditional for each y_i, $i = 1, ..., n+1$,

$$p(y_i \mid \{y_k : k \neq i\}, \alpha, \mu, \phi) = \frac{\alpha}{\alpha + n} \text{Weibull}(y_i \mid \mu, \phi) + \frac{1}{\alpha + n} \sum_{k \neq i} \delta_{y_k}(y_i).$$

The posterior full conditional for each y_i can be obtained by combining its prior full conditional above with the relevant term from (6). More specifically, the posterior full conditional for y_i, $i = 1, ..., n$, is a mixed distribution: y_i is set equal to y_k, where $k \neq i$, with probability

$$\frac{\text{ga}(x_i \mid \lceil y_k \theta^{-1} \rceil, \theta^{-1})}{\alpha \sum_{j=1}^{J} q_{0j} \, \text{ga}(x_i \mid j, \theta^{-1}) + \sum_{k \neq i} \text{ga}(x_i \mid \lceil y_k \theta^{-1} \rceil, \theta^{-1})} \, ,$$

where $x_i = t_i - t_{i-1}$, and $q_{0j} \equiv q_{0j}(\theta, \mu, \phi) = G_0(j\theta) - G_0((j-1)\theta)$ (available in closed form under the Weibull choice for G_0); and, with the remaining probability, y_i is drawn from distribution proportional to $\sum_{j=1}^{J} \{q_{0j} \, \text{ga}(x_i \mid j, \theta^{-1})\} \text{TWeib}_j(y_i \mid \mu, \phi, \theta)$, where $\text{TWeib}_j(y_i \mid \mu, \phi, \theta)$ denotes the Weibull(μ, ϕ) distribution truncated such that $(j-1)\theta < y_i \leq j\theta$. The posterior full conditional for y_{n+1} has similar structure with $\text{ga}(x_i \mid \lceil y_k \theta^{-1} \rceil, \theta^{-1})$ replaced by $C_{\lceil y_k \theta^{-1} \rceil}(t_n, \theta)$, and $\text{ga}(x_i \mid j, \theta^{-1})$ replaced by $C_j(t_n, \theta)$.

The posterior full conditional for (θ, J) involves conditioning on all the latent variables and it can therefore be expressed as

$$p(\theta, J \mid ..., \text{data}) \propto p(\theta, J) \, c_{I_{n+1}}(t_n, \theta) \prod_{i=1}^{n} \text{ga}(x_i \mid I_i, \theta^{-1}) \, \mathbf{1}(\max\{I_1, ..., I_{n+1}\} \leq J) \, ,$$

where $p(\theta, J)$ is the prior for (θ, J), given in Section 2.2, and $I_i = \lceil y_i \theta^{-1} \rceil$, for $i = 1, ..., n+1$. Based on the DP prior structure, the posterior full conditional for (μ, ϕ) is given by

$$p(\mu, \phi \mid ..., \text{data}) \propto \text{IG}(\mu \mid a_\mu, b_\mu)\, \text{gamma}(\phi \mid a_\phi, b_\phi) \prod_{\ell=1}^{n^*} \text{Weibull}(y_\ell^* \mid \mu, \phi) \ ,$$

where n^* is the number of distinct elements of $\{y_1, ..., y_n, y_{n+1}\}$, and y_ℓ^*, $\ell = 1, ..., n^*$, are the distinct values. We use Metropolis-Hastings steps to update μ, ϕ, and (θ, J). finally, α is sampled with the data augmentation method developed in Escobar and West (1995).

Acknowledgments

The authors wish to thank the Editor, an Associate editor, and two referees for several useful comments. This research was supported in part by the National Science Foundation under awards SES 1024484 and DMS 1513076.

References

Adams, R. P., Murray, I., and MacKay, D. J. (2009), "Tractable Nonparametric Bayesian Inference in Poisson Processes with Gaussian Process Intensities," in *Proceedings of the 26th International Conference on Machine Learning*, Montreal, Canada.

Alvarez, E. (2005), "Estimation in Stationary Markov Renewal Processes, with Application to Earthquake Forecasting in Turkey," *Methodology and Computing in Applied Probability*, 7, 119–130.

Anagnos, T. and Kiremidjian, A. (1988), "A review of earthquake occurrence models for seismic hazard analysis," *Probab. Eng. Mech.*, 3, 3–11.

Antoniak, C. (1974), "Mixtures of Drichlet processes with applications to Bayesian nonparametric problems," *Annals of Statistics*, 2, 1152–1174.

Ausín, M., Lillo, R. E., and Wiper, M. P. (2007), "Bayesian control of the number of servers in a GI/M/c queueing system," *Journal of Statistical Planning and Inference*, 137, 3043–3057.

Barbieri, R., Quirk, M. C., Frank, L. M., Wilson, M. A., and Brown, E. (2001), "Construction and analysis of non-Poisson stimulus-response models of neural spike train activity," *Journal of Neuroscience Methods*, 105, 25–37.

Brown, M. and Proschan, F. (1983), "Imperfect Repair," *Journal of Applied Probability*, 20, 851–859.

Caputo, M. (1974), "Analysis of seismic risk, engineering seismology and earthquake engineering," *NATO Advanced Study Institutes Series, Series E: Applied Sciences*, 3, 55–86.

Cornell, C. (1968), "Engineering seismic risk analysis," *Bull. Seism. Soc. Am.*, 58, 1583–1606.

Cox, D. R. (1972), "The statistical analysis of dependencies in point processes," in *Stochastic Point Processes; Statistical Analysis, Theory and Applications*, ed. Lewis, P. A. W., Wiley, pp. 55–66.

Daley, D. J. and Vere-Jones, D. (2003), *An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods*, New York: Springer, 2nd ed.

Epifani, I., Ladelli, L., and Pievatolo, A. (2014), "Bayesian estimation for a parametric Markov renewal model applied to seismic data," *Electronic Journal of Statistics*, 8, 2264–2295.

Escobar, M. D. and West, M. (1995), "Bayesian density estimation and inference using mixtures," *Journal of the American Statistical Association*, 90, 577–588.

Ferguson, T. S. (1973), "A Bayesian analysis of some nonparametric problems," *Annals of Statistics*, 1, 209–230.

Gelfand, A. E. and Kottas, A. (2002), "A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models," *Journal of Computational and Graphical Statistics*, 11, 289–305.

Ishwaran, H. and James, L. F. (2004), "Computational Methods for Multiplicative Intensity Models Using Weighted Gamma Processes: Proportional Hazards, Marked Point Processes, and Panel Count Data," *Journal of the American Statistical Association*, 99, 175–190.

Jarrett, R. G. (1979), "A Note on the Intervals Between Coal-Mining Disasters," *Biometrika*, 66, 191–193.

Kang, J., Nichols, T. E., Wager, T. D., and Johnson, T. D. (2014), "A Bayesian hierarchical spatial point process model for multi-type neuroimaging meta-analysis," *The Annals of Applied Statistics*, 8, 1800–1824.

Kass, R. and Ventura, V. (2001), "A spike train probability model," *Neural Computation*, 13, 1713–1720.

Kottas, A. and Sansó, B. (2007), "Bayesian mixture modeling for spatial Poisson process intensities, with applications to extreme value analysis," *Journal of Statistical Planning and Inference*, 137, 3151–3163.

Koyama, S. and Kass, R. (2008), "Spike train probability models for stimulus-driven leaky integrate-and-fire neurons," *Neural Computation*, 20, 1776–1795.

Lawless, J. F. and Thiagarajah, K. (1996), "A Point-Process Model Incorporating Renewals and Time Trends, with Application to Repairable Systems," *Technometrics*, 38, 131–138.

Lee, S. C. K. and Lin, X. S. (2010), "Modeling and evaluating insurance losses via mixtures of Erlang distributions," *North American Actuarial Journal*, 14, 107–130.

Lindqvist, B. H. (2006), "On the statistical modeling and analysis of repairable systems," *Statistical Science*, 21, 532–551.

Lindqvist, B. H., Elvebakk, G., and Heggland, K. (2003), "The trend-renewal process for statistical analysis of repairable systems," *Technometrics*, 45, 31–44.

Møller, J., Syversveen, A. R., and Waagepetersen, R. P. (1998), "Log Gaussian Cox processes," *Scandinavian Journal of Statistics*, 25, 451–482.

Nadarajah, S. and Kotz, S. (2006), "On the Laplace Transform of the Pareto Distribution," *Queueing Syst. Theory Appl.*, 54, 243–244.

Neal, R. (2000), "Markov chain sampling methods for Dirichlet process mixture models," *Journal of Computational and Graphical Statistics*, 9, 249–265.

NGDC/WDS (2019), "National Geophysical Data Center / World Data Service: NCEI/WDS Global Significant Earthquake Database," NOAA National Centers for Environmental Information, doi: 10.7289/V5TD9V7K (Accessed: 2019-03-01).

Ogata, Y. (1988), "Statistical models for earthquake occurrences and residual analysis for point processes," *Journal of the American Statistical Association*, 83, 9–27.

Parsons, T. (2008), "Earthquake recurrence on the south Hayward fault is most consistent with a time dependent, renewal process," *Geophysical Research Letters*, 35.

Petrone, S. (1999a), "Bayesian density estimation using Bernstein polynomials," *Canadian Journal of Statistics*, 27, 105–126.

— (1999b), "Random Bernstein polynomials," *Scandinavian Journal of Statistics*, 26, 373–393.

Ripley, B. (1977), "Modelling spatial patterns," *Journal of the Royal Statistical Society B*, 39, 172–212.

Rodriguez, A., Wang, Z., and Kottas, A. (2017), "Assessing systematic risk in the S&P500 index between 2000 and 2011: A Bayesian nonparametric approach," *The Annals of Applied Statistics*, 11, 527–552.

Smith, W. L. and Leadbetter, M. R. (1963), "On the Renewal Function for the Weibull Distribution," *Technometrics*, 5, pp. 393–396.

Taddy, M. (2010), "Autoregressive Mixture Models for Dynamic Spatial Poisson Processes: Application to Tracking the Intensity of Violent Crime," *Journal of the American Statistical Association*, 105, 1403–1417.

Taddy, M. A. and Kottas, A. (2012), "Mixture Modeling for Marked Poisson Processes," *Bayesian Analysis*, 7, 335–362.

Thümmler, A., Buchholz, P., and Telek, M. (2006), "A Novel Approach for Phase-Type Fitting with the EM Algorithm," *IEEE Transactions on Dependable and Secure Computing*, 3, 245–258.

Tijms, H. C. (1994), *Stochastic models: an algorithmic approach*, John Wiley & Sons Chichester.

Wiper, M., Insua, D. R., and Ruggeri, F. (2001), "Mixtures of Gamma Distributions with Applications," *Journal of Computational and Graphical Statistics*, 10, pp. 440–454.

Wolpert, R. L. and Ickstadt, K. (1998), "Poisson/Gamma random field models for spatial statistics," *Biometrika*, 85.

Xiao, S., Kottas, A., and Sansó, B. (2015), "Modeling for seasonal marked point processes: An analysis of evolving hurricane occurrences," *The Annals of Applied Statistics*, 9, 353–382.

Xu, Z., Hong, Y., Meeker, W. Q., Osborn, B. E., and Illouz, K. (2017), "A multi-level trend-renewal process for modeling systems with recurrence data," *Technometrics*, 59, 225–236.

Zhao, Y. and Nagaraja, H. N. (2011), "Fisher information in window censored renewal process data and its applications," *Annals of the Institute of Statistical Mathematics*, 63, 791–825.

Zhuang, J., Ogata, Y., and Vere-Jones, D. (2002), "Stochastic declustering of space-time earthquake occurrences," *Journal of the American Statistical Association*, 97, 369–380.