UNIVERSITY OF CALIFORNIA
SANTA CRUZ

# FLEXIBLE MIXTURE MODELING APPROACHES TO RENEWAL PROCESSES

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

STATISTICAL SCIENCE

by

**W. Zachary Horton**

December 2024

The Dissertation of W. Zachary Horton
is approved:

---

Professor Athanasios Kottas, Chair

---

Professor Bruno Sansó

---

Professor Paul A. Parker

---

Peter Biehl
Vice Provost and Dean of Graduate Studies

# Contents

# List of Figures

# List of Tables

## Abstract

Flexible Mixture Modeling Approaches to Renewal Processes

by

W. Zachary Horton

This dissertation develops flexible and computationally efficient Bayesian mixture modeling methods for various types of renewal processes. Renewal processes are temporal point process models whose stochastic mechanism focuses on the times between successive events, or inter-arrival times. They have been applied in a variety of fields, including system reliability, earthquake recurrence modeling, and analysis of neural spike-trains. The homogeneous renewal process assumes that the inter-arrival times are independent and identically distributed, being a generalization of the homogeneous Poisson process where inter-arrival times are exponentially distributed. Various extensions of this basic model have been proposed, of which discrete marks and time-varying hazards are relevant to this work.

We first propose a Bayesian nonparametric mixture modeling framework for homogeneous renewal process densities. Selection of the mixture kernel and prior specification are guided by specific features of renewal processes. The definition of a renewal process requires finite mean for the inter-arrival time distribution. We discuss sufficient conditions to satisfy this constraint. In addition, event clustering behavior is often of interest in analyzing renewal process point patterns. Clustering behavior is assessed through the renewal function, which can be obtained from the Laplace transform of the

inter-arrival time density, hence kernels with analytical Laplace transform expressions are preferred. We present model details using the gamma density kernel, requiring only a minor restriction on prior hyperparameters to satisfy the finite mean requirement. Motivated by the application area of earthquake recurrence modeling, we also develop a model for decreasing density shapes using a uniform mixture kernel.

Markov renewal processes are a generalization of the homogeneous case where discrete state information is observed with each event. Transitions from one state to another are governed by a Markov chain, and inter-arrival times arise conditionally from transition-specific distributions. For example, earthquake recurrence characteristics may depend on whether the observed magnitudes exceed certain thresholds. Conventional estimation methods for Markov renewal models treat each transition case independently, which facilitates convenient computation but may ignore underlying structure or similarities between cases. Using as foundation the nonparametric mixture modeling framework developed for homogeneous renewal processes, we propose a novel modeling approach for Markov renewal processes where dependence between transition cases is captured through a dependent nonparametric prior. Our proposed framework contains both the homogeneous renewal process and the conventional Markov renewal process as special limiting cases, allowing the degree and nature of dependence to be studied. This method is particularly useful in earthquake recurrence models, where the additional inferences provided by our model reveal interesting patterns in how earthquake magnitudes affect recurrence times. We explore model properties through simulated data and then compare several models applied to an earthquake dataset from Southern California.

Certain extensions of the homogeneous renewal process, such as the time-varying modulated renewal process, are defined in terms of the inter-arrival hazard rate function rather than the density. In these settings, a flexible model applied directly to the hazard function can be more easily adapted to such extensions. Additionally, prior information in some applications may be more naturally expressed on the hazard scale, which may be difficult to integrate into a density-oriented model. We propose a novel basis representation model for hazard functions, using log-logistic hazard basis functions and a nonparametric prior model for the basis coefficients. The result is a flexible and computationally efficient model for renewal process hazard functions. To demonstrate its tractability as a foundation for renewal process extensions, we formulate a nonparametric model for modulated renewal processes.

*To my wife and children, the brightest lights in my life.*

## Acknowledgments

I have been fortunate to cross paths with many wonderful people who helped me on my journey at UCSC. I am extraordinarily grateful to my advisor, Professor Athanasios Kottas. He is a skillful teacher, a kind soul, and the finest mentor a student can imagine. In particular, I am grateful for his patient guidance during hard times when my confidence and motivation waned. It is an honor to have learned from him and to be numbered among his students.

In addition to Thanasis, I want to thank professors Bruno Sansó, Raquel Prado, Robert Lund, Richard Li, and Juhee Lee for the time and work spent being my teachers. Because of their high-quality instruction, I might actually make a decent statistician someday. Specifically deserving of mention is Bruno, who taught over a third of my classes and, along with Paul Parker, was willing to serve on my dissertation committee.

It was my pleasure to befriend many talented graduate students during my time at UCSC. Yunzhe Li, Peter Trubey, and Jizhou Kang were instrumental in helping me approach difficult research questions on the office whiteboard. I fondly remember studying online with Kelsey Blackstone and Heidi Rogers during the pandemic; I would not have survived the first year without their companionship and support. I also want to thank Jacob Fontana for his work with the union and for having my back when I needed it. Mykel and Betsy Brinkerhoff, who once covered our apartment in Nicholas Cage pictures while we were away, were our first and closest friends in family student housing. My appreciation also goes out to the tireless staff who support the students,

particularly Theo-Alyce Gordon, Aaron McPherson, and Heather Delagi, who made my family and I feel so welcome.

Finally, I want to thank my family. My parents and in-laws have been an ever-present source of support, especially during the pandemic, and when we had a baby, and when our house flooded, and when we had to evacuate for the wildfires, and when our house flooded again. Hours spent on the phone with parents and siblings were a welcome respite from the stress of graduate school. Most of all, I want to express my love and appreciation for my wife, Jaelynn. If my life amounts to anything in the future, it will be because of her love, support, and sacrifice.

# Chapter 1

# Introduction

Point processes are a class of stochastic models that describe the occurrence of random events over a continuum, such as space, time, or space-time. They have found application in a wide range of fields, including seismology, ecology, finance, reliability, neuroscience, and many others. Point processes theory has been studied extensively (Cox and Isham, 1980; Daley and Vere-Jones, 2003), offering principled approaches for answering questions about event frequency, clustering, and dependence, among other behaviors. This dissertation is concerned with temporal point processes, more specifically a family known as renewal processes.

Temporal point processes are often represented through a counting process $\{N(t) : t > 0\}$ that tracks the number of events prior to time $t$ and increments by one whenever an event occurs. Inference for temporal point processes typically centers around the conditional intensity function. The conditional intensity gives the instanta-

neous rate of events at time $t$, and can be expressed by,

$$\lambda(t|\mathcal{H}(t)) = \lim_{h \to 0} \frac{\mathbb{E}[N(t+h) - N(t)|\mathcal{H}(t)]}{h}\,, \tag{1.1}$$

where $\mathcal{H}(t)$ denotes the process history up to $t$. The structure of the conditional intensity determines the properties of the corresponding point process. For example, the homogeneous Poisson process assumes a constant intensity, $\lambda(t|\mathcal{H}(t)) \equiv \lambda$, leading to event independence and Poisson distributed event counts. The nonhomogeneous Poisson process (NHPP) is a well-established extension that allows the intensity to vary over time, $\lambda(t|\mathcal{H}(t)) \equiv \lambda(t)$, but otherwise preserves the Poisson count structure.

A realization of a temporal point process over the interval $t \in (0, T)$ results in a random number of events $N(T) = n$ and a sequence of event times, denoted $\{0 < t_1 < \ldots < t_n < T\}$. Given a realized point pattern, the likelihood function is constructed from the conditional intensity:

$$\exp\left(-\int_0^T \lambda(u|\mathcal{H}(u))du\right) \prod_{i=1}^{n} \lambda(t_i|\mathcal{H}(t_i))\,. \tag{1.2}$$

For many point processes, such as the Poisson process, the likelihood is most tractable when using event times as the primary input, but other models are more naturally expressed in terms of the inter-arrival times between successive events. Inter-arrival times are obtained by $x_i = t_i - t_{i-1}$, where $t_0 = 0$ by convention, and therefore $t_i = x_1 + \ldots + x_i$. In terms of informational content, the sequence of inter-arrival times $\{x_1, \ldots, x_n\}$ is equivalent to the set of event times, but certain models can be more succinctly expressed with this alternative representation of the data. Renewal processes are one such family of models where the fundamental stochastic mechanism focuses on the structure

of inter-arrival times. The simplest of these is the homogeneous renewal process, which assumes that inter-arrival times are independent and identically distributed. An exponential inter-arrival time distribution is a special case corresponding to a homogeneous Poisson process. In the following sections, we introduce key background material for renewal processes, discuss relevant extensions, and outline the objectives of the dissertation.

## 1.1 Renewal process background

A homogeneous renewal process (HRP) is characterized by a conditional intensity where the primary argument is the elapsed time since the most recent event, $\lambda(t|\mathcal{H}(t)) \equiv h(t-t_{N(t)})$. Substituting this into Equation (1.2) yields the HRP likelihood:

$$\begin{aligned}
&\exp\left(-\int_{t_n}^{T} h(u-t_n)du - \sum_{i=1}^{n}\int_{t_{i-1}}^{t_i} h(u-t_{i-1})du\right)\prod_{i=1}^{n} h(t_i-t_{i-1}) \\
&\quad = \exp\left(-\int_{0}^{T-t_n} h(u)du - \sum_{i=1}^{n}\int_{0}^{x_i} h(u)du\right)\prod_{i=1}^{n} h(x_i)\,.
\end{aligned}$$

Letting $H(x) = \int_0^x h(u)du$, $S(x) = \exp(-H(x))$, and $f(x) = h(x)S(x)$, the likelihood can be simplified:

$$S(T-t_n)\prod_{i=1}^{n} f(x_i)\,. \tag{1.3}$$

Now written in terms of inter-arrival times, the likelihood elucidates key HRP properties. The conditional intensity $h$ acts as the inter-arrival distribution hazard function, and therefore $H$, $S$, and $f$ correspond to the cumulative hazard, survival, and probability density functions; respectively. The multiplicative contribution of each inter-arrival

time $x_i$ implies that they are independent and identically distributed (i.i.d.) given the number of events $N(T) = n$, a hallmark property of HRP models. Additionally, the leading likelihood term highlights censoring information brought about by the upper time bound $T$. Further properties and theoretical results for renewal processes can be found in sources such as Cox (1962), Tijms (1994), or Daley and Vere-Jones (2003).

The term *renewal* in the process name refers to how the inter-arrival time increment $t - t_{N(t)}$ returns to zero after each event. In other words, an event occurrence triggers the underlying mechanisms or conditions to reset as if the system were just starting. This phenomenon is embodied in realizations of HRP conditional intensities, which trace the inter-arrival hazard function $h$ until an event occurs, at which point the intensity jumps back to the initial hazard level. Figure 1.1 illustrates this behavior using a Weibull inter-arrival hazard with shape parameter 0.8. As a result, renewal process event times are not independent, except in the special case of a homogeneous Poisson process where the hazard is constant. This contrasts with a nonhomogeneous Poisson process, where event times are independent but inter-arrival times are not. Both perspectives are valuable, but in different contexts. Assumed renewal behavior is more common in areas such as earthquake recurrence models, neural spike train analysis, and queuing theory, to name a few.

The renewal function is a pivotal quantity in mathematical renewal theory and can be used to study additional properties of a renewal process, namely event clustering (Daley and Vere-Jones, 2003). It is defined as $M(t) = \mathbb{E}[N(t)] = \sum_{i=1}^{\infty} F_i(t)$, where $F_i(t)$ is the distribution function of the $i$th event time. The renewal function also satisfies the

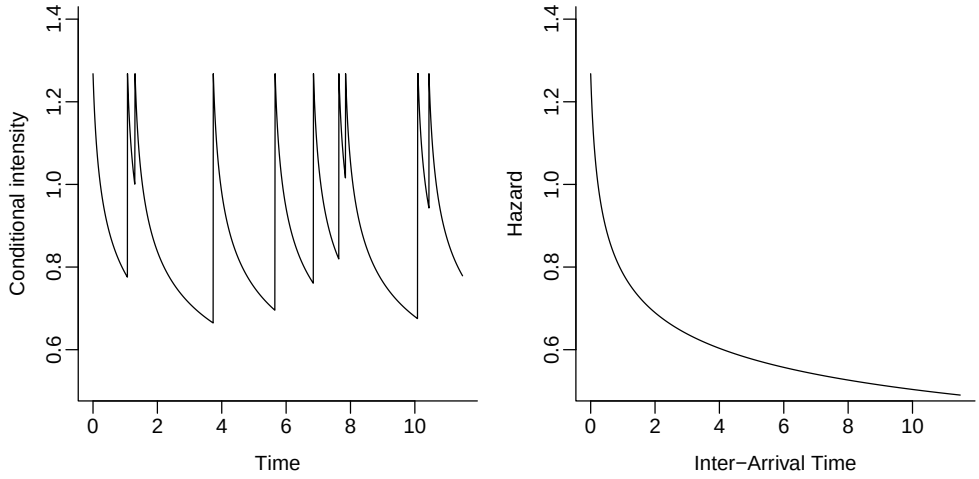


Figure 1.1: Illustration of a realized homogeneous renewal process conditional intensity function resetting after each event. The left panel shows the conditional intensity function over time, while the right panel shows the corresponding Weibull inter-arrival time hazard function.

renewal equation, given by,

$$M(t) = F(t) + \int_0^t M(t-x)dF(x)\,, \tag{1.4}$$

where $F(x) = 1-S(x)$ is the distribution function of inter-arrival times. Using Equation (1.4), the renewal function can be expressed in terms of the inter-arrival density $f$ through the Laplace domain:

$$\mathcal{L}_M(s) = \frac{\mathcal{L}_f(s)}{s(1-\mathcal{L}_f(s))}\,, \tag{1.5}$$

where $\mathcal{L}_g(s) = \int_0^\infty e^{-st}g(t)dt$ is the Laplace transform of a positive function $g$ with support on $\mathbb{R}^+$. Assuming $\mathcal{L}_f$ is available, numerical Laplace inversion routines can be applied to Equation (1.5) to recover the renewal function $M(t)$.

The renewal function carries information about event clustering behavior, often represented through the $K$-function of Ripley (1977). Letting $X$ denote the inter-arrival

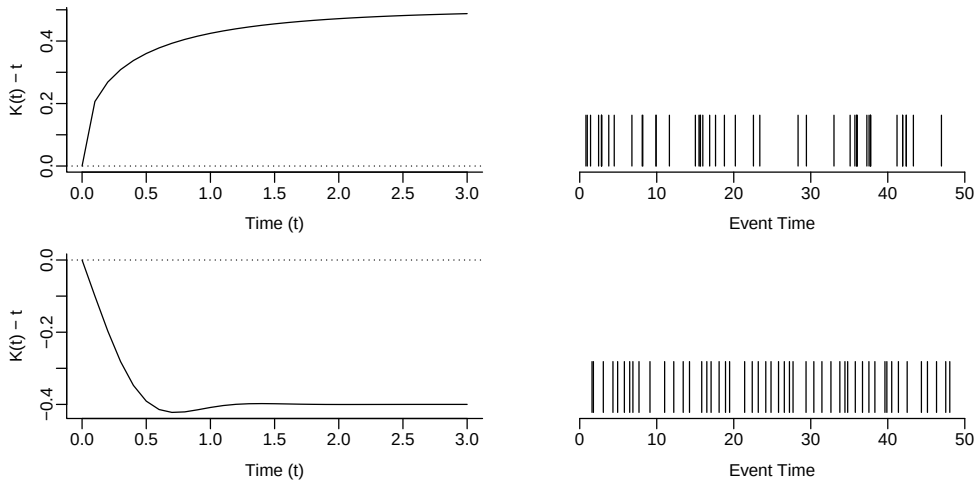


Figure 1.2: Simulated point-patterns with gamma distributed inter-arrival times and their corresponding $K$-functions. Top panels use a gamma distribution with shape 0.5, producing clustering behavior. Bottom panels use a gamma distribution with shape 5, producing declustering behavior. The line $K(t) - t = 0$ is plotted to emphasize clustering behavior relative to a homogeneous Poisson process. Both processes have inter-arrival distributions with mean of 1 and generate 50 events over $t \in (0, 50)$.

time random variable with $\mathbb{E}[X] = \int_0^\infty x f(x) dx$ denoting its expectation, the $K$-function is given by $K(t) = M(t)\mathbb{E}[X]$. The $K$-function is motivated by the elementary renewal theorem, which states that the renewal function satisfies $\lim_{t\to\infty} t^{-1} M(t) = \mathbb{E}[X]^{-1}$ provided that $\mathbb{E}[X] < \infty$. Events are said to exhibit clustering behavior when $K(t) > t$ and declustering behavior when $K(t) < t$. The special case of $K(t) = t$ corresponds to a homogeneous Poisson process, which serves as a neutral point of reference. To illustrate these behaviors, Figure 1.2 shows two simulated point-patterns with gamma distributed inter-arrivals and their corresponding $K$-functions. The quantity $K(t) - t$ is plotted instead of $K(t)$ itself to emphasize clustering relative to a homogeneous Poisson process.

It is important to note that the definition of a renewal process requires that the

inter-arrival time distribution have a finite mean, i.e., $\mathbb{E}[X] < \infty$. This condition ensures that the renewal function is well-defined and that the elementary renewal theorem holds. Although straightforward to satisfy for parametric models, care must be taken when developing nonparametric models with prior support over distributions with infinite first moment. The finite mean restriction will play a role throughout our developments in later chapters.

## 1.2 Renewal process extensions

Various extensions of the basic renewal process have been proposed. These aim to accommodate additional dependencies and more complex temporal structures, including time-varying behavior, self-exciting intensities, marked events, multivariate realizations, and many others. Two extensions relevant to this dissertation are the Markov renewal process and the modulated renewal process.

A Markov renewal process (MRP) generalizes the homogeneous case by allowing the distribution of inter-arrival times to depend on discrete state marks associated with each event. That is, realized event times $t_i$ are paired with state labels $s_i$, where $s_i \in \{1, \ldots, S\}$ can be one of $S$ states. The inter-arrival time $x_i = t_i - t_{i-1}$ can be interpreted as the time taken to *sojourn* from state $s_{i-1}$ to state $s_i$, hence $x_i$ is commonly referred to as a sojourn time in this setting. While the terminologies can be used interchangeably, we will refer to sojourn times only in the context of MRP models to be consistent with the literature, defaulting to inter-arrival times otherwise. Theoretical

properties of MRP models are deeply connected to those of semi-Markov models, for which Howard (1971) and Limnios and Oprişan (2001) are comprehensive references.

Under a Markov renewal process, the sequence of states $\{s_1, \ldots, s_n\}$ is assumed to follow a stationary Markov chain, referred to as the *embedded* Markov chain. The probability of transitioning to state $k$ conditioned on the starting state $j$ is denoted $p^{(j,k)} = \Pr(s_i = k | s_{i-1} = j)$. Given an observed state sequence, sojourn times arise from transition-specific distributions, i.e., the sojourn time $x_i$ is drawn according to a density function $f^{(j,k)}$ given that $s_{i-1} = j$ and $s_i = k$. The density $f^{(j,k)}$ is said to be the sojourn time distribution for the $j$-to-$k$ *transition case*. The MRP likelihood function for a sequence of observed times and states is expressed by,

$$\left\{ \prod_{i=1}^{n} p^{(s_{i-1},s_i)} f^{(s_{i-1},s_i)}(x_i) \right\} \left\{ \sum_{k=1}^{S} p^{(s_n,k)} S^{(s_n,k)}(x_c) \right\}, \tag{1.6}$$

where the first term accounts for both censoring due to $T$ and the unobserved state transition. Standard likelihood-based estimation for MRP models is facilitated by partitioning the data by transition cases and independently considering sojourn times within each case.

The modulated renewal process (ModRP) is a time-varying extension of the HRP, falling under the broader family of inhomogeneous renewal processes. It consists of an inter-arrival time hazard function $h$ and a modulating function $\lambda(t)$ that scales the hazard at time $t$. The most common form of the ModRP conditional intensity is multiplicative, $\lambda(t|\mathcal{H}(t)) \equiv \lambda(t)h(t - t_{N(t)})$, which shares roots with proportional hazards survival models (Cox, 1972). Other forms have been studied such as an additive

structure (Lin and Ying, 1994), but the multiplicative form is most common. The likelihood function is obtained by substituting the conditional intensity expression into Equation (1.2):

$$\exp\left(-\int_0^T \lambda(t)h(t-t_{N(t)})dt\right)\prod_{i=1}^{n}\lambda(t_i)h(t_i-t_{i-1})\,. \tag{1.7}$$

An especially desirable property of modulated renewal processes is that they contain both the HRP and the NHPP as special cases. This broad range of expressiveness makes them effective for describing complex temporal structures. While the likelihood function can be challenging to work with, success in developing a flexible and computationally efficient model makes the effort worthwhile.

For a homogeneous renewal process, the conditional intensity depends on past events only through the most recent event time. Developments in this dissertation focus on stochastic models that maintain this structure. However, frameworks that facilitate higher order dependence between events are worth mentioning. Hawkes processes (Hawkes, 1971; Hawkes and Oakes, 1974) model self-exciting behavior by incorporating contributions from all past events into the conditional intensity. Kim (2021) provides a recent review of temporal Hawkes processes and develops corresponding Bayesian nonparametric estimation methods. A different direction is to incorporate higher-order memory effects through a mixture transition density model (Le et al., 1996), such as the Bayesian nonparametric model proposed by Zheng et al. (2024).

## 1.3 Research objectives

The main objective of this dissertation is to develop stochastic modeling frameworks for renewal processes that, by leveraging Bayesian nonparametric methods, are flexible, computationally efficient, and capable of incorporating prior information in a principled manner.

We begin in Chapter 2 by developing a nonparametric mixture model for homogeneous renewal process densities using stick-breaking priors for the weights. This bears a resemblance to Dirichlet process mixture models (Antoniak, 1974), but with constraints and additional adjustments following from the renewal process structure. In particular, kernel selection is impacted both by the finite mean requirement and the desire for tractable calculation of the renewal function. After discussing several candidate mixing kernels, we develop a model using gamma densities as a general-purpose choice. In addition, motivated by applications to earthquake recurrence modeling, we propose a mixture using structured uniform distributions in order to model non-increasing density shapes. Regarding the choice of stick-breaking priors, we consider both the Dirichlet process (Ferguson, 1973; Sethuraman, 1994) and the logit stick-breaking process (Ren et al., 2011). The logit stick-breaking prior model will be a critical building block for subsequent Markov renewal process developments. We approach posterior simulation using Markov Chain Monte Carlo (MCMC), specifically Gibbs sampling. In simulation studies and an earthquake data application, we demonstrate that the proposed model can accurately estimate the inter-arrival density and outperforms the standard

parametric Weibull model.

Chapter 3 is concerned with Markov renewal processes. As will be shown, the MRP likelihood function can be rewritten in terms of sojourn times partitioned into transition cases. For a given $j$-to-$k$ transition case, estimation of the corresponding sojourn time distribution depends only on observations where the state transitions are from state $j$ to state $k$. In other words, individual observations contribute to the likelihood only through their respective transition cases, having no influence on the others The effect is that sojourn time distributions are in some sense *independent* of each other, and therefore model fitting can be viewed as performing parallel density estimation for each $j$-to-$k$ transition case. This property has historically been essential for developing efficient likelihood-based estimation methods for MRP models. However, the assumption of independent sojourn time distributions is restrictive and may not be appropriate in all settings. For many applications, it is reasonable to suppose that certain transition cases share similar sojourn time characteristics. Additionally, this structural independence assumption can be limiting when certain transition cases are rarely or never observed, leading to unreasonably uncertain inferences.

Our goal in Chapter 3 is to develop a flexible model for Markov renewal processes that facilitates borrowing of information between transition cases. We propose a modeling framework using the logit stick-breaking process mixture developed for HRP densities to describe the sojourn time distributions. To induce dependence between transition cases, we develop a hierarchical dependence structure where correlation is introduced in the mixture weights. The resulting construction supports a wide range

of dependence patterns, including the HRP and the fully independent model as special limiting cases. We discuss model assessment techniques and illustrate the proposed model through simulation studies and an earthquake data application.

Returning to the homogeneous renewal process, recall the likelihood given in Equation (1.3), which is perhaps most naturally expressed in terms of the inter-arrival time density. The majority of existing HRP models are developed from this perspective, focusing on the density function as the primary object of interest. However, there exist circumstances where directly modeling the inter-arrival hazard function is more appropriate. Prior information in certain applications, such as those adjacent to survival analysis, may be most readily available on the scale of hazards. Furthermore, one perspective may be more suitable than the other for developing individual extensions. The Markov renewal process, for example, is most readily represented from the density perspective, hence our efforts in developing a density-oriented HRP mixture model in Chapter 2. Conversely, the modulated renewal process likelihood in Equation (1.7) is formulated in terms of the HRP hazard function, suggesting that a homogeneous renewal model applied directly to inter-arrival hazards may be more easily adapted into a ModRP. As an aside, part of our motivation for utilizing Bayesian nonparametric methods is that they support modeling from either perspective, allowing prior information to be incorporated in the most appropriate manner for the circumstances.

In Chapter 4, we explore the development of flexible and computationally efficient renewal process models from the hazard perspective. Central to our treatment is the introduction of a novel basis representation model composed of log-logistic hazard

functions. Basis representation models are characterized by a linear combination of basis functions that are either entirely known or depend only on a few global parameters. This structure confines the target functional to the space spanned by the basis functions, shifting the bulk of modeling effort to the combination weights. The log-logistic basis system we propose is designed specifically for modeling hazards, and is supported by a convergence result that speaks to its flexibility. We then develop a structured Bayesian nonparametric prior for the combination weights and design a posterior sampling algorithm that capitalizes on properties of the basis model. Through simulation studies and real data examples, we examine model behavior and assess its performance on homogeneous renewal data.

As a demonstration of hazard-oriented extensions, we also formulate a modulated renewal process model, using the log-logistic basis system. We employ a basis representation model for the modulating intensity function, resulting in a flexible framework for describing time-varying renewal structures. Leveraging basis representations for both components enables the use of standard posterior sampling techniques that have otherwise been challenging to apply to these models. We explore model properties and apply it to a coal-mining disaster dataset and an earthquake recurrence dataset.

# Chapter 2

# Nonparametric Mixture Modeling for Homogeneous Renewal Process Densities

Early foundational work on renewal theory focuses on mathematical properties, particularly those related to the renewal equation and asymptotic behavior (Feller, 1941; Doob, 1948; Skellam and Shenton, 1957; Smith, 1958; Cox, 1962). Additional theoretical attention has been given to the case of Weibull distributed inter-arrival times due to its importance in reliability theory (Smith and Leadbetter, 1963; Lomnicki, 1966). Comprehensive treatments of renewal theory can be found in Tijms (1994) and Daley and Vere-Jones (2003), among others. Parametric models were the focus of early estimation and inference efforts. Popular parametric models that have been employed include the exponential distribution (Maguire et al., 1952; Jarrett, 1979), the Weibull

distribution (White, 1964; Alvarez, 2005; Epifani et al., 2014), the Pareto distribution (Singhai et al., 2009), and others (Utsu, 1984; Ogata, 1999; Zhao and Nagaraja, 2011). More recent work has focused on nonparametric estimation of the inter-arrival distribution. Pievatolo and Rotondi (2000) apply a Dirichlet process mixture with generalized gamma base distribution to earthquake change-point data. Rotondi (2010) develops an estimation method based on Pólya tree priors, also applied to earthquake recurrence. Comte and Duval (2018) present a classical nonparametric estimator based on Laguerre series expansions. Xiao et al. (2021) propose a nonparametric Erlang density mixture and apply the model to both earthquake data and industrial accident data.

In this chapter, we develop a nonparametric mixture modeling framework for the inter-arrival density of a renewal process. Inspired by Dirichlet process mixtures (Ferguson, 1973; Antoniak, 1974), our proposed framework is designed to allow practitioners to select the mixing kernel and stick-breaking prior that best suits a particular use case. A wide variety of stick-breaking priors for modeling the mixing weights can be used, but we will focus on the Dirichlet process stick-breaking construction (Sethuraman, 1994) and the logit stick-breaking process Ren et al. (2011). In selecting a mixture kernel, additional factors specific to renewal processes must be considered, namely the finite mean condition and the desire for inference on the renewal function. We will review various kernel choices, ultimately settling on the gamma kernel as a suitable general-purpose option. Our discussion is motivated in part by the study of earthquake inter-arrival times. There exist arguments in the seismology literature that underlying earthquake mechanisms should produce inter-arrival times with a non-increasing density

(Corral, 2005; Kumar et al., 2020). To that end, we will also present a structured uniform kernel that caters to applications like earthquakes where a non-increasing density shape is expected or required.

## 2.1 Mixture modeling framework

Nonparametric mixture models are powerful tools for density estimation. They possess a rich theoretical foundation and offer tremendous flexibility. Methods based on stick-breaking constructions have become especially popular for their extensibility and computational advantages (Ishwaran and James, 2001). Such *stick-breaking mixtures* are structured as countable mixtures of kernel densities, where the kernel parameters arise i.i.d. from some base distribution and the weights are generated from a prior with a stick-breaking structure. Given a random discrete mixing distribution $G = \sum_{\ell=1}^{\infty} \omega_\ell \delta_{\theta_\ell}$, a stick-breaking mixture model for a target density $f$ can be expressed as,

$$f(x|G) = \sum_{\ell=1}^{\infty} \omega_\ell f_{\mathcal{K}}(x|\theta_\ell)\,, \tag{2.1}$$

where $f_{\mathcal{K}}$ denotes a known kernel density. The component-specific kernel parameters $\boldsymbol{\theta} = \{\theta_\ell : \ell \geq 1\}$ are drawn i.i.d. from a base distribution $G_0$. The mixture weights $\boldsymbol{\omega} = \{\omega_\ell : \ell \geq 1\}$ are defined with a stick-breaking process, where independent latent variables $\boldsymbol{\nu} = \{\nu_\ell : \ell \geq 1\}$ determine the weights through $\omega_\ell = \nu_\ell \prod_{r=1}^{\ell-1}(1-\nu_r)$. The prior distribution on each $\nu_\ell$ depends on the specific stick-breaking process and is influenced by potentially several hyperparameters $\alpha$. The Dirichlet process mixture is a famous example where the mixing distribution $G$ follows a Dirichlet process (DP), resulting

in i.i.d. kernel parameters drawn from $G_0$ and weights generated from a stick-breaking procedure using beta distributed latent variables. Specifically, $\nu_\ell|\alpha \overset{i.i.d.}{\sim} \text{Beta}(1, \alpha)$ where $\alpha > 0$ is the DP precision parameter.

Despite their popularity in other areas, stick-breaking mixtures have found little use in the context of renewal processes. The Erlang mixture model proposed by Xiao et al. (2021) is a notable application of Bayesian nonparametric methods to renewal processes, but Erlang mixtures are more similar to basis expansions and differ substantially in structure from stick-breaking mixtures. In standard univariate density estimation, the choice of kernel and stick-breaking prior is often guided by prior knowledge and the specific application. Our aim is to extend this methodology to a renewal process setting, however, there are additional considerations unique to renewal processes that must be addressed when selecting a kernel and stick-breaking prior.

### 2.1.1 Kernel selection

The choice of kernel is a critical component of any mixture model. In the context of renewal processes, the kernel must satisfy two key properties: have marginal support on $\mathbb{R}^+$ and meet the finite mean condition, possibly through parameter restrictions. Additionally, the kernel should produce mixtures that are flexible, computationally tractable, and amenable to inference for the $K$-function.

The marginal support condition can be easily satisfied by selecting a kernel with support on $\mathbb{R}^+$. Bounded kernels may also be used provided that a prior is specified so that the resulting marginal density has the necessary support. For example, a uniform

kernel with support on $[0, \theta]$ can be used if $\theta$ is given a prior with support on $\mathbb{R}^+$, such as an exponential distribution. Careful treatment is required to ensure that the finite mean condition is satisfied almost surely for realizations of the mixing distribution $G$. We summarize a set of sufficient conditions in the following theorem.

**Theorem 2.1.** *Let $f$ be a density modeled by a stick-breaking mixture with kernel $f_{\mathcal{K}}$, kernel parameter(s) $\theta$ with parameter space $\Theta$, and base distribution $G_0$. Define $E_f(G) = \int_0^\infty x f(x|G)dx$ as the first moment of $f$ given mixing distribution $G$. Also let $E_{\mathcal{K}}(\theta) = \int_0^\infty x f_{\mathcal{K}}(x|\theta)dx$ be the first moment of the kernel density given $\theta$. Then, $E_f(G) < \infty$ almost surely if:*

*(1) $E_{\mathcal{K}}(\theta)$ is finite for all $\theta \in \Theta$, and,*

*(2) $\mathbb{E}[E_{\mathcal{K}}(\theta)] = \int_\Theta E_{\mathcal{K}}(\theta)dG_0(\theta) < \infty$.*

*Proof.* Begin by substituting the mixture expression $f(x|G) = \int_\Theta f_{\mathcal{K}}(x|\theta)dG(\theta)$ into the expectation $E_f(G)$:

$$E_f(G) = \int_0^\infty x \left( \int_\Theta f_{\mathcal{K}}(x|\theta)dG(\theta) \right) dx \,.$$

Assumption *(1)* requires that the kernel have finite mean for all parameter values. This allows the use of Fubini's theorem to rearrange the integrals and simplify:

$$E_f(G) = \int_\Theta \int_0^\infty x f_{\mathcal{K}}(x|\theta)dx dG(\theta) = \int_\Theta E_{\mathcal{K}}(\theta)dG(\theta) = \sum_{\ell=1}^\infty \omega_\ell E_{\mathcal{K}}(\theta_\ell) \,.$$

Now consider the expectation of each term in the sum, which can be factored as the weights and kernel parameters are independent: $\mathbb{E}[\omega_\ell E_{\mathcal{K}}(\theta_\ell)] = \mathbb{E}[\omega_\ell]\mathbb{E}[E_{\mathcal{K}}(\theta_\ell)]$.

Both terms are finite; the former by the bounded support of the weights and the latter by assumption *(2)*. Thus, the sum can be written as $\sum_{\ell=1}^{\infty} \mathbb{E}[\omega_\ell]\mathbb{E}[E_{\mathcal{K}}(\theta_\ell)] = \mathbb{E}[E_{\mathcal{K}}(\theta_\ell)] \sum_{\ell=1}^{\infty} \mathbb{E}[\omega_\ell]$. Partial sums of the weights are bounded by 1 and converge monotonically, hence,

$$\mathbb{E}[E_{\mathcal{K}}(\theta_\ell)] \sum_{\ell=1}^{\infty} \mathbb{E}[\omega_\ell] = \mathbb{E}[E_{\mathcal{K}}(\theta_\ell)]\mathbb{E}\left[\sum_{\ell=1}^{\infty} \omega_\ell\right] = \mathbb{E}[E_{\mathcal{K}}(\theta_\ell)]\,.$$

This established that $\sum_{\ell=1}^{\infty} \mathbb{E}[\omega_\ell E_{\mathcal{K}}(\theta_\ell)] < \infty$, which converges monotonically since each term is non-negative, allowing it to be rewritten as,

$$\sum_{\ell=1}^{\infty} \mathbb{E}[\omega_\ell E_{\mathcal{K}}(\theta_\ell)] = \mathbb{E}\left[\sum_{\ell=1}^{\infty} \omega_\ell E_{\mathcal{K}}(\theta_\ell)\right] = \mathbb{E}[E_f(G)]\,,$$

which implies that $\mathbb{E}[E_f(G)] < \infty$ and therefore $E_f(G) < \infty$, almost surely. □

The conditions given in Theorem 2.1 limit the choice of kernels that are viable for modeling renewal processes with nonparametric stick-breaking mixtures. Regarding condition *(1)*, kernel densities like the Weibull or gamma distributions are suitable options as they have finite mean for all parameter values. However, condition *(1)* also allows for kernels such as the Lomax or inverse-gamma distributions provided an appropriate constraint on the parameter space $\Theta$ is imposed. Condition *(2)* is more restrictive and requires that the kernel be paired with a suitable base distribution $G_0$. This ultimately devalues kernels with heavier tails like the Lomax or inverse-gamma distributions, requiring highly informative priors on the kernel parameters in addition to parameter space restrictions. The Weibull kernel is unsuitable for the same reason; although the mean is always finite, it grows extremely quickly as the shape parameter approaches 0, violating condition *(2)* for many otherwise reasonable choices of $G_0$.

Recall from Equation (1.5) that the renewal function, and the $K$-function by extension, are calculated from the Laplace transform of the inter-arrival density. Kernels with closed-form Laplace transform expressions are particularly attractive as they allow for direct calculation of the renewal function in Laplace space, requiring numerical methods only for the inverse transform. We encounter significant stability issues when numerically performing both the Laplace transform and the inverse step, discouraging use of kernels where this is mandatory, such as the Weibull, log-logistic, Lomax, and inverse-gamma distributions.

Mixture flexibility is another important consideration. The kernel mixture should support a variety of density shapes, particularly those that differ from standard parametric forms. Many distributions over $\mathbb{R}^+$ are limited to only unimodal shapes, including the log-normal and Rayleigh distributions, or are strictly decreasing, such as the exponential or half-normal distributions. In contrast, kernel densities that support both unimodal and decreasing shapes are more versatile and better suited for our purposes.

Considering these factors, the gamma distribution emerges as a strong candidate for the kernel in a nonparametric mixture model for renewal processes. We denote the gamma kernel by:

$$f_{\mathcal{K}}(x|\theta) \equiv \mathrm{Ga}(x|\kappa, \lambda) = \frac{1}{\Gamma(\kappa)\lambda^{\kappa}} x^{\kappa-1} e^{-x/\lambda}\,, \tag{2.2}$$

where $\kappa$ is the shape parameter and $\lambda$ is the scale parameter. The gamma density has a simple Laplace transform expression, $\mathcal{L}_{\mathrm{Ga}(\kappa,\lambda)}(s) = (1+\lambda s)^{-\kappa}$, making inference on the renewal function more manageable. Gamma mixtures in various forms are supported

by theoretical results affirming their flexibility (DeVore and Lorentz, 1993; Wiper et al., 2001; Lee and Lin, 2010). More importantly, the gamma kernel very easily accommodates the finite mean conditions from Theorem 2.1. The mean of the gamma distribution is $\kappa\lambda$, thus condition *(1)* is satisfied for all parameter combinations. To address condition *(2)*, we introduce relatively standard priors on the component shape parameters $\boldsymbol{\kappa} = \{\kappa_\ell : \ell \geq 1\}$ and scale parameters $\boldsymbol{\lambda} = \{\lambda_\ell : \ell \geq 1\}$. We assign $\kappa_\ell|\gamma_\kappa \overset{i.i.d.}{\sim} \text{Ga}(a_\kappa, \gamma_\kappa)$ and $\lambda_\ell|\gamma_\lambda \overset{i.i.d.}{\sim} \text{inv-Ga}(a_\lambda, \gamma_\lambda)$, where inv-Ga denotes the inverse-gamma distribution with density,

$$\text{inv-Ga}(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\beta/x}\,. \tag{2.3}$$

Condition *(2)* is satisfied by $a_\lambda > 2$, or equivalently by demanding a prior with finite second moment for the scale parameter.

Another kernel we consider is a structured uniform density, given by,

$$f_{\mathcal{K}}(x|\theta) \equiv \text{Unif}(x|0, \theta) = \frac{1}{\theta}\mathbb{I}(0 < x < \theta)\,, \tag{2.4}$$

where $\theta$ is the upper bound parameter and $\mathbb{I}(\cdot)$ denotes the indicator function that evaluates to 1 if the condition is true and 0 otherwise. The uniform kernel offers many of the same advantages as the gamma kernel. Mixtures of these structured uniform densities produce non-increasing shapes that are very flexible. Indeed, theoretical representation results indicate that all strictly non-increasing functions over $\mathbb{R}^+$ can be written as mixtures of these uniform densities (Brunner and Lo, 1989). The Laplace transform of the uniform kernel is available in closed form, $\mathcal{L}_{\text{Unif}(0,\theta)}(s) = (1 - e^{-\theta s})/(\theta s)$. In addition, the uniform kernel has an analytical expression for its survival function, which is

convenient for handling the likelihood censoring term in Equation (1.3). For the kernel parameters $\boldsymbol{\theta} = \{\theta_\ell : \ell \geq 1\}$, we assign $\theta_\ell | \eta_\theta, \gamma_\theta \overset{i.i.d.}{\sim} \text{inv-Ga}(\eta_\theta, \gamma_\theta)$, where the restriction $\eta_\theta > 1$ ensures the finite mean condition holds.

Our interest in non-increasing density shapes is motivated by the application area of earthquake recurrence modeling. Decreasing inter-arrival densities are commonly observed in earthquake datasets world-wide (Utsu, 1984; Altinok and Kolcak, 1999; Alvarez, 2005; Rotondi, 2010; Xiao et al., 2021). References in both seismology and mechanical physics have suggested that decreasing density shapes are an expected consequence of the physical processes that generate earthquakes (Corral, 2005; Kumar et al., 2020). We develop the uniform kernel model to provide a method for incorporating this domain knowledge into the mixture structure.

### 2.1.2 Stick-breaking prior selection

For our purposes in this chapter, the exact choice of stick-breaking prior is not critical. In practice, a researcher may select a stick-breaking prior based on desired inferences, computational considerations, prior knowledge, convention, or other factors. Our developments will focus on two popular stick-breaking models: the latent beta stick-breaking procedure found in Dirichlet processes and the logit stick-breaking process.

As mentioned, stick-breaking priors are structured such that the weights $\boldsymbol{\omega} = \{\omega_\ell : \ell \geq 1\}$ are generated from a sequence of latent variables $\boldsymbol{\nu} = \{\nu_\ell : \ell \geq 1\}$ with $\omega_\ell = \nu_\ell \prod_{r=1}^{\ell-1}(1 - \nu_r)$. The stick-breaking construction of the Dirichlet process, first introduced by Sethuraman (1994), assumes the latent variables are beta distributed,

$\nu_\ell|\alpha \overset{i.i.d.}{\sim} \text{Beta}(1,\alpha)$. This model has been used extensively in the literature and has many desirable properties, including multiple computational methods for posterior inference. The logit stick-breaking process, first presented by Ren et al. (2011), is an alternative construction that uses a logit transformation of the latent variables: $\log(\nu_\ell/(1-\nu_\ell))|\mu,\sigma^2 \overset{i.i.d.}{\sim} \text{N}(\mu,\sigma^2)$, where $\text{N}(\mu,\sigma^2)$ denotes the normal distribution with mean $\mu$ and variance $\sigma^2$. Computation for this model is more challenging, although recent advances in Pólya-gamma augmentation have made it more accessible (Polson et al., 2013; Rigon and Durante, 2021). Our motivation for considering this prior is in the capacity to develop certain extensions, namely the Markov renewal process model we present in Chapter 3. In the context of homogeneous renewal processes, we find both stick-breaking priors perform similarly in practice.

### 2.1.3 Hierarchical model and implementation details

To facilitate posterior inference, we adapt the standard blocked Gibbs sampler for Dirichlet process mixtures to our renewal process setting. In this section we show the hierarchical structures associated with the choices of kernel and stick-breaking prior. We limit our comments on sampling steps to high-level summaries and notable departures from the standard blocked Gibbs approach, presenting full posterior sampling details in Appendix A.

Given an observed series of inter-arrival times $\boldsymbol{x} = \{x_1,\ldots,x_n\}$ over the observation window $(0,T)$ with censored time $x_c = T - t_n$, the likelihood of our HRP

mixture modeling framework is written as,

$$\left(\sum_{\ell=1}^{L} \omega_\ell S_{\mathcal{K}}(x_c|\theta_\ell)\right) \prod_{i=1}^{n} \sum_{\ell=1}^{L} \omega_\ell f_{\mathcal{K}}(x_i|\theta_\ell)\,, \tag{2.5}$$

where $f_{\mathcal{K}}$ is the chosen kernel density and $S_{\mathcal{K}}$ is the corresponding survival function. Notice that the mixture has been truncated to a fixed number of components $L$, and thus the set of kernel parameters $\boldsymbol{\theta} = \{\theta_\ell : \ell = 1, \ldots, L\}$ and weights $\boldsymbol{\omega} = \{\omega_\ell : \ell = 1, \ldots, L\}$ are also indexed up to $L$. To aid in posterior sampling, we introduce latent membership variables $\boldsymbol{z} = \{z_i : i = 1, \ldots, n\}$, where $z_i = \ell$ indicates that $x_i$ has been assigned to component $\ell$ with prior probability $\Pr(z_i = \ell|\boldsymbol{\omega}) = \omega_\ell$. We also introduce a latent variable $z_c$ for the censored time $x_c$ with similar prior probability $\Pr(z_c = \ell|\boldsymbol{\omega}) = \omega_\ell$. The augmented model likelihood can be expressed as,

$$S_{\mathcal{K}}(x_c|\theta_{z_c}) \prod_{i=1}^{n} f_{\mathcal{K}}(x_i|\theta_{z_i})\,. \tag{2.6}$$

Posterior sampling for the latent variables is standard, though the censored time $x_c$ requires the kernel survival function rather than the density.

Details on the stick-breaking priors have been discussed previously, but we repeat those details here for completeness. The mixture weights arise from a stick-breaking prior, $\omega_\ell = \nu_\ell \prod_{r=1}^{\ell-1}(1-\nu_r)$, where $\nu_\ell$ is generated from the chosen stick-breaking process. For the Dirichlet process, $\nu_\ell|\alpha \overset{i.i.d.}{\sim} \text{Beta}(1, \alpha)$. A gamma prior is assigned to the DP precision parameter, denoted $\alpha \sim \text{Ga}(a_\alpha, b_\alpha)$ with fixed hyperparameters $a_\alpha$ and $b_\alpha$. We use default values of $a_\alpha = 1$ and $b_\alpha = 1$, though these can be adjusted to suit particular applications. For the logit stick-breaking process, $\log(\nu_\ell/(1-\nu_\ell))|\mu, \sigma^2 \overset{i.i.d.}{\sim} \text{N}(\mu, \sigma^2)$. We assign conjugate priors $\mu \sim \text{N}(m_\mu, s_\mu^2)$ and $\sigma_\mu^2 \sim \text{inv-Ga}(a_\sigma, b_\sigma)$ with fixed hyper-

parameters. Default values of $m_\mu = 0$, $s_\mu^2 = 1$, $a_\sigma = 2.5$, and $b_\sigma = 1$ tend to perform well, but can be adjusted as needed. Fully conjugate updates for the logit stick-breaking weights are available if latent Pólya-gamma variables are introduced. We give a brief overview of this augmentation step in Chapter 3.4.1, where the logit stick-breaking process plays a more prominent role, and include implementation details for the current model in Appendix A.

Turning to the kernel parameters, the model hierarchy for the gamma kernel is structured as follows:

$$\begin{aligned} x_i|z_i, \boldsymbol{\kappa}, \boldsymbol{\lambda} &\overset{i.i.d.}{\sim} \text{Ga}(x_i|\kappa_{z_i}, \lambda_{z_i}), \quad p(x_c|z_c, \boldsymbol{\kappa}, \boldsymbol{\lambda}) = S_{\text{Ga}}(x_c|\kappa_{z_c}, \lambda_{z_c}), \\ \kappa_\ell|\gamma_\kappa &\overset{i.i.d.}{\sim} \text{Ga}(a_\kappa, \gamma_\kappa), \quad \lambda_\ell|\gamma_\lambda \overset{i.i.d.}{\sim} \text{inv-Ga}(a_\lambda, \gamma_\lambda), \\ \gamma_\kappa &\sim \text{inv-Ga}(a_\gamma^\kappa, b_\gamma^\kappa), \quad \gamma_\lambda \sim \text{Ga}(a_\gamma^\lambda, b_\gamma^\lambda), \end{aligned} \tag{2.7}$$

where $S_{\text{Ga}}$ denotes the gamma survival function. Metropolis steps are required for the shape parameters $\boldsymbol{\kappa}$, as well as the scale parameter $\lambda_\ell$ for which $z_c = \ell$. The priors on the hyperparameters are conjugate with fixed values for $a_\kappa$, $a_\lambda$, $a_\gamma^\kappa$, $b_\gamma^\kappa$, $a_\gamma^\lambda$, and $b_\gamma^\lambda$. Recall that $a_\lambda > 2$ is required to ensure the finite mean condition holds. We suggest default values of $a_\kappa = 1$, $a_\lambda = 2.5$, $a_\gamma^\kappa = 2.5$, $b_\gamma^\kappa = 1$, $a_\gamma^\lambda = 1$, and $b_\gamma^\lambda = 1$. The uniform kernel model hierarchy is similar in structure:

$$\begin{aligned} x_i|z_i, \boldsymbol{\theta} &\overset{i.i.d.}{\sim} \text{Unif}(x_i|0, \theta_{z_i}), \quad p(x_c|z_c, \boldsymbol{\theta}) = S_{\text{Unif}}(x_c|0, \theta_{z_c}), \\ \theta_\ell|\eta_\theta, \gamma_\theta &\overset{i.i.d.}{\sim} \text{inv-Ga}(\eta_\theta, \gamma_\theta), \quad \gamma_\theta \sim \text{Ga}(a_\gamma^\theta, b_\gamma^\theta), \quad (\eta_\theta - 1) \sim \text{Ga}(a_\eta^\theta, b_\eta^\theta), \end{aligned} \tag{2.8}$$

where $S_{\text{Unif}}$ denotes the uniform survival function and the prior on $\eta_\theta$ is chosen to ensure the finite mean condition holds. The updates for each $\theta_\ell$ are truncated inverse-gamma

draws, and the hyperparameters $a_\gamma^\theta$, $b_\gamma^\theta$, $a_\eta^\theta$, and $b_\eta^\theta$ are fixed. We recommend default values of $a_\gamma^\theta = 1$, $b_\gamma^\theta = 1$, $a_\eta^\theta = 2.5$, and $b_\eta^\theta = 1$.

We find that the default hyperparameter values suggested here work well in practice, with results being reasonably robust to moderate changes in these values. Indeed, the mechanisms in which hyperparameters influence results mirror those of standard Dirichlet process mixtures. For example, priors that emphasize larger values of the precision parameter $\alpha$ can produce a larger number of mixture components in the posterior. We will comment on the impact of key hyperparameter adjustments in the real data example presented in Section 2.3, particularly in the context of the uniform kernel model, but otherwise we direct the reader to the extensive literature on Dirichlet process mixtures for more detailed insight on hyperparameter sensitivity.

Selecting an appropriate number of mixture components $L$ is a common challenge in mixture modeling, with many methods available in the literature (Neal, 2000; Ishwaran and James, 2001). A simple approach is to select $L$ such that the expected value of the total weight is within some tolerance of 1, i.e., $1 - \mathbb{E}\left[\sum_{\ell=1}^{L} \omega_\ell | \alpha\right] < \epsilon$. For the Dirichlet process prior, this amounts to solving $(\alpha/(\alpha+1))^L < \epsilon$ given a desired tolerance $\epsilon$. For example, setting $\alpha = 1$ and $\epsilon = 10^{-4}$ implies a lower bound of $L > 14$. The value of $\alpha$ can be fixed, either to the prior expectation or some other informed value, or it can be marginalized over. A similar expression is available for the logit stick-breaking process using a first order approximation of the weight expectation, yielding the condition $(1 + e^{m_\mu})^{-L} < \epsilon$. In this case, a lower bound of $L > 14$ arises from $m_\mu = 0$ and $\epsilon = 10^{-4}$. For results that follow later in this chapter, we select

$L = 100$ as a conservative choice for both kernel structures and stick-breaking priors. Our reasoning relates to the different characteristics of the gamma and uniform kernels, which we discuss in more detail in Section 2.4.

### 2.1.4 Model validation

Homogeneous renewal process modeling shares many similarities with standard density estimation, and thus many of the same model checking techniques can be applied. Additional means to check the point process assumptions are available through the time-rescaling theorem (Daley and Vere-Jones, 2003). Consider a general point process with conditional intensity function $\lambda(t|\mathcal{H}(t))$ and let the cumulative conditional intensity be given by $\Lambda(t|\mathcal{H}(t)) = \int_0^t \lambda(u|\mathcal{H}(u))du$. The time-rescaling theorem states that the transformed point pattern $\{t_1^*, \ldots, t_n^*\}$, where $t_i^* = \Lambda(t_i|\mathcal{H}(t_i))$, is a realization of a unit rate Poisson process. Moreover, the random variables $u_i^* = 1 - \exp(-(t_i^* - t_{i-1}^*))$ are i.i.d. uniform on the interval $(0, 1)$. In the context of HRP modeling, the cumulative conditional intensity simplifies such that $u_i^* = F(x_i)$ where $F$ is the estimated cumulative density function (CDF) of the inter-arrival time distribution.

A visual goodness-of-fit test can be constructed by plotting the empirical quantiles of the estimated $u_i^*$ against the theoretical quantiles of the uniform distribution, which we refer to as a *time-rescaling QQ plot*. Posterior samples of the inter-arrival density can be used to calculate posterior samples of the $u_i^*$, ultimately resulting in uncertainty intervals on the QQ plot. Significant deviations from the diagonal line indicate model misspecification. That misspecification occur either due to an inappropriate

distributional assumption for the inter-arrival times, or a failure of the homogeneous renewal process assumption. Given the flexibility of the nonparametric mixture model, time-rescaling QQ plots applied to our proposed methods are more useful in diagnosing the point process assumptions compared to standard parametric HRP models.

In addition to the visual check, we also implement a numerical goodness-of-fit metric to enable more direct model comparison. We refer to this is the *ECDF Error* and define it as, $EE^{(b)} = \int_0^\infty |F^{(b)}(x) - \hat{F}(x)|dx$, where $F^{(b)}$ is the estimated inter-arrival CDF for the $b$th posterior sample and $\hat{F}$ is the empirical cumulative density function (ECDF) of the observed inter-arrival times. Smaller values of $EE^{(b)}$ indicate better model fit, and the posterior distribution of $EE^{(b)}$ can be used to compare models with uncertainty levels accounted for.

### 2.1.5 Hyperparameter effects on clustering

Here we explore the effects of hyperparameter choices on clustering behavior in the prior model. The impact of stick-breaking hyperparameters is driven mainly by the prior expected number of components, so we focus the discussion on the Dirichlet process stick-breaking prior due to the corresponding interpretation of the precision parameter $\alpha$. We generate $K$-function realizations for a given mixture kernel density by selecting values for $\alpha$ and the hyperparameters associated with $G_0$, and then simulating values for the weights and kernel parameters. Finally, the $K$-function is calculated from the resulting inter-arrival density. For the gamma kernel mixture, only the hyperparameters associated with the shape parameters $\kappa_\ell$ have an impact on clustering behavior, where

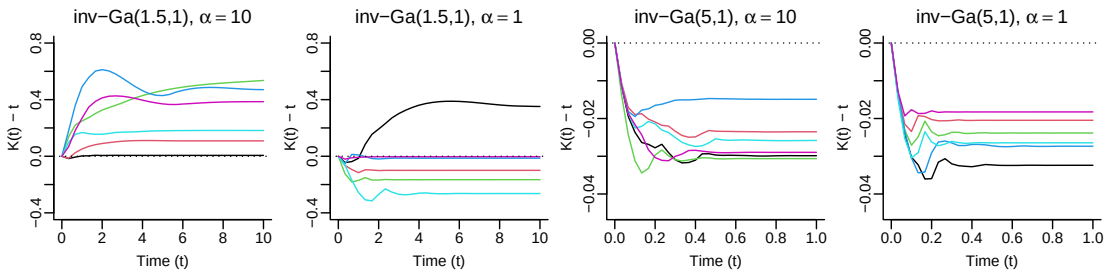


Figure 2.1: Realizations of the $K$-function from the prior mixture model with uniform kernel and Dirichlet process prior for the mixing distribution under different choices for the hyperparameters $\alpha$ and $\eta_\theta$. Each panel shows six realizations under hyperparameter values indicated in the panel title.

hyperparameters that favor larger values of $\kappa_\ell$ tend to produce declustering behavior and vice versa. In this setting, neither $\alpha$ nor the hyperparameters associated with the scale parameters $\lambda_\ell$ have a meaningful effect on whether clustering or declustering is produced.

The uniform kernel mixture exhibits more notable interactions, particularly between the shape parameter $\eta_\theta$ and $\alpha$. Figure 2.1 shows six realizations for each of four hyperparameter combinations. The shape parameter $\eta_\theta$ has a similar effect as the hyperparameters associated with $\kappa_\ell$ in the gamma kernel model. Small values of $\eta_\theta$ lead to more pronounced clustering behavior, while larger values lead to declustering. The impact of $\alpha$ appears to be associated the value of $\eta_\theta$, with smaller values biasing sampled $K$-functions towards declustering. This is perhaps to be expected; uniformly distributed inter-arrival times always exhibit declustering behavior. Small values of $\alpha$ tend to produce mixtures with fewer components, giving rise to the observed declustering behavior in the uniform kernel mixture. On the other hand, larger values of $\alpha$ yield more mixture components, allowing the inverse-gamma $G_0$ to drive clustering behavior,

namely through $\eta_\theta$. In short, when $\eta_\theta$ is small, the DP precision parameter $\alpha$ will determine whether clustering or declustering is observed, whereas when $\eta_\theta$ is large, declustering is observed regardless of the value of $\alpha$.

## 2.2 Synthetic data examples

In this section we use simulated datasets to illustrate the ability of our proposed nonparametric modeling framework to capture a wide range of renewal process behaviors. We begin by considering results for the gamma mixture model with the Dirichlet process stick-breaking prior. For this, we generate HRP point patterns from four different generative distributions, each with unique density shapes and $K$-function behaviors. The first is a Weibull distribution with shape parameter 1.5 and unit scale, denoted $\text{Wei}(1.5, 1)$. This case is characterized by a density with unimodal shape and declustering behavior in the $K$-function. Our second scenario is a Lomax distribution with shape parameter 5 and unit scale, denoted $\text{Lom}(5, 1)$, where the Lomax distribution has density $(\alpha/\lambda)(1 + x/\lambda)^{-(\alpha+1)}$ for shape $\alpha$ and scale $\lambda$. The Lomax distribution has polynomial tails and produces clustering behavior. The third is a standard half-normal distribution with density $\sqrt{2/\pi}\exp(-x^2/2)$, which exhibits declustering along with a strictly decreasing density shape. Finally, the fourth generative scenario is an evenly weighted mixture of a $\text{Ga}(2, 1)$ and a $\text{Ga}(10, 0.5)$, which possesses a multimodal shape for both the density shape and the $K$-function. For each scenario, we generate inter-arrival times using a value of $T$ such that the expected number of events is roughly

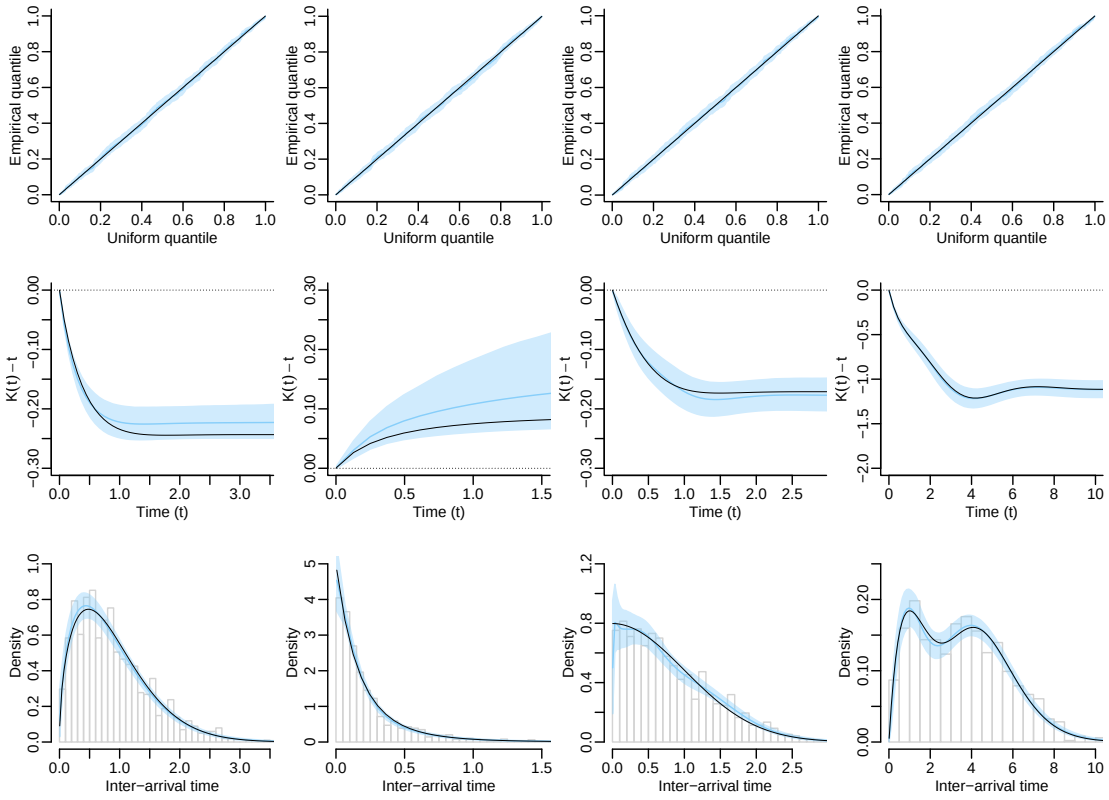


Figure 2.2: Simulated data examples fit using the gamma kernel HRP mixture model with the Dirichlet process stick-breaking prior. Generative distributions are, from left to right, Weibull, Lomax, half-normal, and an evenly weighted mixture of two gamma distributions. The top row shows time-rescaling QQ plots, the middle row shows $K$-function estimates, and the bottom row shows inter-arrival density estimates.

1,000, then fit the model using the default hyperparameter values suggested previously. The sampling algorithm is run for 5,000 iterations following a burnin period of 15,000. Figure 2.2 shows posterior estimates and 95% uncertainty intervals for the inter-arrival density, $K$-function, and the time-rescaling QQ plot for each scenario. Not only is the model able to capture the true inter-arrival density shapes, but these results highlight that the model is flexible in capturing a variety of $K$-function behaviors, including both clustering and declustering.

The uniform mixture is restricted to modeling non-increasing shapes. For this reason, and for the sake of comparison, we reuse the data generated from the Lomax and half-normal scenarios above. Notably, these distributions represent two extremes in terms of behavior, with the Lomax having polynomial tails and a tendency towards clustering, and the half-normal having exponential tails and presenting declustering behavior. For each scenario, we fit the uniform kernel mixture model with the Dirichlet process stick-breaking prior using default hyperparameter values and the same sampling configuration. Resulting inferences are displayed in Figure 2.3. Figure 2.3 shows the results of this analysis. Similar to the gamma kernel, the uniform kernel mixture is able to recover the true inter-arrival density shapes, and is able to produce both clustering and declustering behavior in the $K$-function estimates.

Though omitted here for brevity, we also repeat the simulation study using the logit stick-breaking prior for both the gamma and uniform kernel models. Results are remarkably similar to those obtained here, confirming that performance is consistent across the two stick-breaking priors. Also omitted are inferences for both kernel models using smaller datasets. We target datasets with roughly 1,000 events to be consistent with the real data example in the next section, as well as to ensure that the sampled point patterns are representative of their generative distributions. Applied to smaller datasets, the model performs comparably to standard Dirichlet process mixtures, capturing general function shapes but with increased uncertainty and some amount of overfitting.

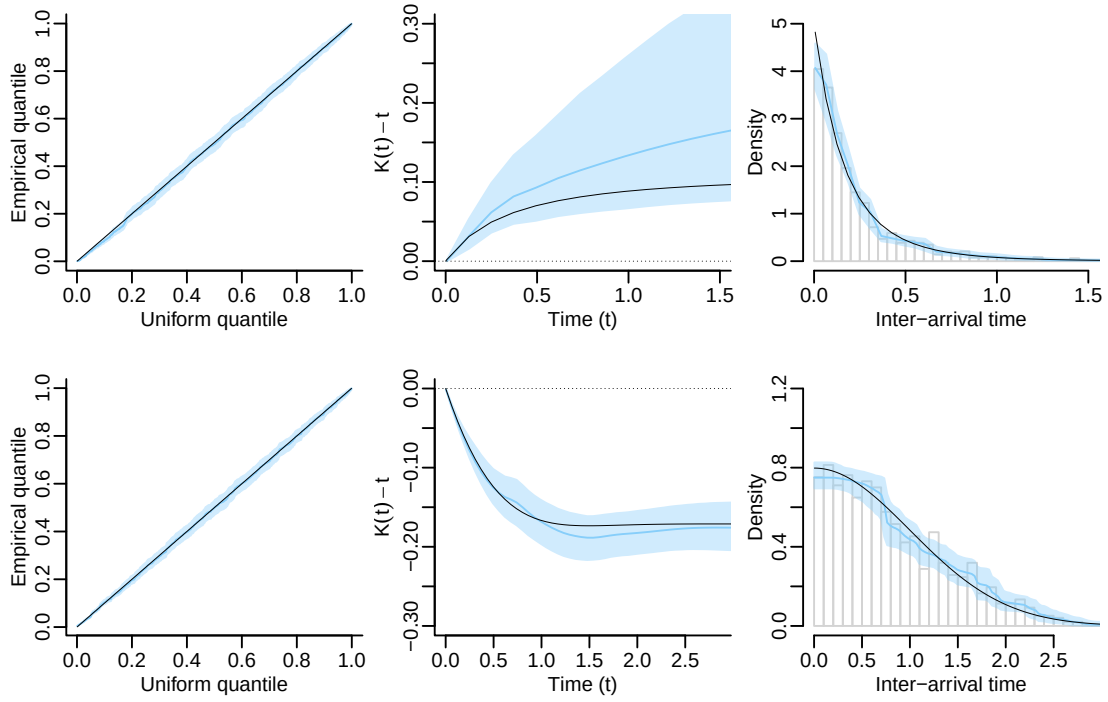


Figure 2.3: Simulated data examples fit using the uniform kernel HRP mixture model with the Dirichlet process stick-breaking prior. Generative distributions are Lomax on the top row and half-normal on the bottom. Panels from left to right show time-rescaling QQ plots, $K$-function estimates, and inter-arrival density estimates.

## 2.3 Southern California earthquake data application

Here we apply the proposed nonparametric mixture modeling framework to a dataset of Southern California earthquakes. The Southern California Earthquake Data Center (SCEDC, 2013) provides a catalog of earthquake events in Southern California and surrounding regions. Details on the network history, measurement methods, and current data collection practices can be found in Hutton et al. (2010). Our analysis focuses on larger earthquakes, having magnitude greater than 4.0 and occurring between the years 1981 and 2022. The starting year is chosen after a major upgrade in both

the instrumentation and data collection practices of the Southern California Seismic Network. In total, our dataset contains 1,248 observations with inter-arrival times ranging from 0.86 seconds to 232 days, and with 25th, 50th, and 75th percentiles of 21.6 minutes, 22.8 hours, and 14.6 days; respectively.

These observed inter-arrival times are heavily skewed with both a long tail and a pronounced concentration of values near zero. This polarized structure can be a challenge for both parametric and nonparametric models to handle, including our proposed mixture models. To adjust, we employ more informative prior distributions to account for this behavior, namely by selecting a few non-default hyperparameter values, $a_\lambda = 20$, $a_\kappa = 5$, and $a_\gamma^\theta = b_\gamma^\theta = 5$. These help to ensure that the quality of fit in the tails remains high despite the data bulk being much closer to zero. We consider fitting all four combinations of kernel and stick-breaking prior, using the same sampling configuration as in the simulated data examples. We also include results from a parametric Weibull model which is a common choice for earthquake inter-arrival modeling (Utsu, 1984; Ogata, 1999; Alvarez, 2005).

On a 2022 Apple MacBook Pro with an M2 chip and 8 gigabytes of RAM, the sampling process for both kernel models takes approximately 12.6 seconds per 1,000 posterior iterations. We use the same machine for all computations in latter chapters as well. Samples for each parameter usually converge within 5,000 iterations, though we use 15,000 burnin iterations as a more conservative choice. The effective sample size for the inter-arrival density evaluated at various points is roughly between 100 and 120 per 1,000 posterior samples, which suggests either longer sampling or thinning by a factor

Table 2.1: Posterior mean ECDF Errors for each model along with 95% credible intervals. Gamma and uniform kernel models are shown for both the Dirichlet process (DP) and logit stick-breaking processes. The parametric Weibull model is also included for comparison.

| Kernel Model | Stick-Breaking | Mean ECDF Error | 95% Interval |
|---|---|---|---|
| Weibull | N/A | 0.456 | (0.279, 0.788) |
| Gamma | DP | 0.135 | (0.072, 0.220) |
| | Logit | 0.159 | (0.091, 0.234) |
| Uniform | DP | 0.014 | (0.003, 0.048) |
| | Logit | 0.011 | (0.004, 0.032) |

of 10 may be necessary.

Figure 2.4 shows posterior estimates and 95% uncertainty intervals for the inter-arrival density, $K$-function, and the time-rescaling QQ plot. Note that 88 observations have inter-arrival times larger than 50 days, but are not shown in the data histogram in order to highlight behavior near the bulk of the data. We also show a table of ECDF Error posterior distributions for each model in Table 2.1.

The QQ plots highlight the ability of each model to capture the polarized behavior of the inter-arrival times. The Weibull model underestimates both the early spike and the long tail, also producing a $K$-function estimate that is much higher than the mixture models. In addition, the $K$-function calculation for the Weibull model is computationally intensive and rather unstable, often resulting in inaccurate credible intervals. The ECDF Error values also speak to the poor performance of the Weibull model compared to the nonparametric mixtures. The gamma kernel mixtures are able to capture the tail behavior well, but still struggle somewhat with the concentration near zero. Results from the uniform kernel mixtures are the most promising, being able to adequately capture both the tail behavior and the spike near zero. In some sense, this

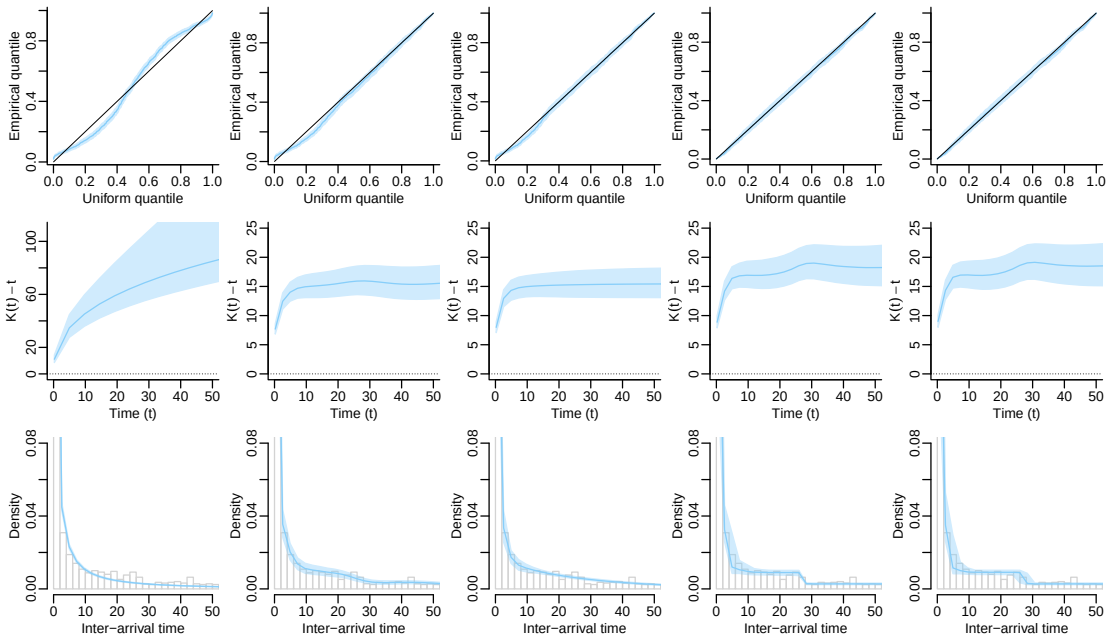


Figure 2.4: Posterior estimates and 95% uncertainty intervals for the Southern California earthquake data analysis. Each column of panels corresponds to a different combination of kernel and stick-breaking prior. From left to right, the columns show the parametric Weibull model, the gamma kernel with Dirichlet process stick-breaking, gamma kernel with logit stick-breaking, uniform kernel with Dirichlet process stick-breaking, and uniform kernel with logit stick-breaking. Time-rescaling QQ plots are shown in the top row, $K$-function estimates in the middle row, and inter-arrival density estimates in the bottom row. Note the different $K$-function scale for the Weibull model.

is not surprising given the localized structure of the uniform kernel. Finally, consistent with our findings with simulated data, it appears that the choice of stick-breaking prior in this setting has minimal impact on the quality of model fit, especially compared to the choice of kernel.

## 2.4 Discussion

Our objective in this chapter was to develop a nonparametric mixture modeling framework for homogeneous renewal processes. The methodology itself is quite similar

to standard nonparametric density estimation methods based on Dirichlet process mixture models. However, unique challenges arise in the context of renewal processes that require careful consideration, namely the finite mean condition and the desire for inference on the $K$-function. This led us to consider the gamma kernel as a general choice for the inter-arrival density kernel, and the uniform kernel as a more specialized option for strictly decreasing densities. Through simulation study and a real data application, we have shown that the proposed framework is successful in capturing a wide range of renewal process behaviors, being an improvement over standard parametric models. In addition, the earthquake data characteristics highlight the situational advantages of the uniform kernel model when a decreasing density assumption is appropriate. The choice of stick-breaking prior was comparatively less impactful than the choice of kernel. However, as will be seen in Chapter 3, the logit stick-breaking process facilitates development of certain HRP model extensions, hence our inclusion of it here.

A lingering item from this chapter is the choice of $L$. In Section 2.1.3 we discussed a simple strategy for selecting $L$ based on the expected total weight, differing slightly between stick-breaking priors. However, the uniform mixture and gamma mixture are of a very different nature, making a common approach for selecting $L$ less straightforward. The gamma kernel structure more closely resembles location-scale mixtures where mixture components are strategically placed to capture various features of the data. In contrast, the uniform kernel structure plays a role more similar to basis expansions, and we should expect that a larger number of components will be needed there. The above strategy for selecting $L$ ignores these differences, focusing only on the

prior model for the weights. Instead, we opt for the simple, yet conservative choice of $L = 100$. We find that this value is more than sufficient in practice for both kernel choices. More consequentially, this strategy carries over better to our developments in Chapter 3 where complex strategies for selecting $L$ are even more difficult to formulate. With that in mind, designing a more informed approach for selecting $L$ that accounts for different mixture behaviors is an interesting area for future research. Perhaps marginalized methods similar to Pólya urn schemes could be adapted to this setting, although extensibility remains a concern.

We conclude this chapter with comments on discretized observations. Homogeneous renewal processes are defined on continuous time, but in practice observations are often recorded on a discrete scale. For very fine discretizations of time, the use of continuous inter-arrival time densities may still be appropriate, such as with the Southern California earthquake data which is measured in seconds. However, in settings where the discretization is quite coarse, it is possible that multiple events could be observed in the same time bin. This case is challenging to handle because it violates key assumptions of the underlying point process, even for discrete-time analogs of renewal theory. One possible approach, inspired by Ogata (1999), is to consider a latent continuous-time process which treats observations as essentially rounded versions of the true, hidden event times. Such a model would require a complex latent structure in order to preserve homogeneous renewal process properties, but could be a useful direction for future research.

# Chapter 3

# Modeling Dependent Sojourn Time Distributions in Markov Renewal Processes

Markov renewal processes (MRPs) are an extension of the homogeneous case where inter-arrival time behavior can be influenced by discrete state variables observed with each event. Specifically, the state variables are assumed to follow a stationary Markov chain, referred to as the *embedded* Markov chain, and inter-arrival times arise from one of several conditional distributions, specified by the observed transition from one state to another. Times between successive earthquakes, for instance, may vary depending on whether the magnitude of the previous event was above or below a certain threshold. Markov renewal processes are used in a variety of applications, including medical data (Chou et al., 2017), acoustics (Stowell and Plumbley, 2013), finance

(Swishchuk and Islam, 2011), and, of course, earthquakes (Alvarez, 2005).

The aim of this chapter is to develop a nonparametric modeling framework for Markov renewal processes. To achieve distributional flexibility, we adapt the stick-breaking mixture model presented in Chapter 2 to the MRP setting. Furthermore, a property common to many Markov renewal models is the conditional independence of transition-specific inter-arrival time distributions, which aids in developing computational methods but can be a restrictive assumption in practice. We address this limitation by introducing an approach to modeling dependence between these transition cases, enabling the sharing of information common to multiple distributions.

## 3.1 Background and motivation

Theoretical foundations for Markov renewal processes stem from semi-Markov theory, where Lévy (1954) and Smith (1955) are notable early references. See Howard (1971) and more recently Limnios and Oprişan (2001) for comprehensive treatments of semi-Markov processes. Markov renewal theory is more distinctly developed by Pyke (1961a), Pyke (1961b), and Pyke and Schaufele (1964), where key limiting theorems are proven and details connecting semi-Markov processes to the formal renewal setting are clarified. Çinlar (1975) provides an overview of major results and applications at the time, including the popular implementation of exponential and Weibull sojourn time distributions in the context of reliability analysis. Subsequent efforts focus on asymptotic properties and adjustments for censoring (Lagakos et al., 1978; Dinse and

Larson, 1986; Greenwood and Wefelmeyer, 1996; Ouhbi and Limnios, 1999, 2003). More recently, some attention has been given to Bayesian nonparametric approaches. Warr and Woodfield (2020) directly model the sojourn time distributions using a Dirichlet process in order to efficiently and flexibly estimate first-passage time distributions. A nonparametric beta-Stacy prior model has been studied in connection with reinforced semi-Markov processes by Bulla and Muliere (2007) and Arfè et al. (2021).

A realized point pattern from a Markov renewal process over the interval $t \in (0, T)$ is composed of a sequence of event times, $\{0 < t_1 < \ldots < t_n < T\}$, and a corresponding sequence of state variables, $\{s_1, \ldots, s_n\}$ where the state space $s_i \in \mathcal{S}$ is labeled $\mathcal{S} = \{1, \ldots, S\}$ for a finite number of states $S$. The inter-arrival times $x_i = t_i - t_{i-1}$ can be interpreted as the time spend transitioning from state $s_{i-1}$ to state $s_i$. For this reason, inter-arrival times in MRP settings are commonly referred to as *sojourn times*, which we adopt for the remainder of this chapter. The state variables are assumed to follow a stationary Markov chain, where the probability of transitioning from state $j$ to state $k$ is denoted $p^{(j,k)} = \Pr(s_i = k | s_{i-1} = j)$. A sojourn time $x_i$ is conditionally drawn from a distribution with density $f^{(j,k)}$ given the observed transition from $s_{i-1} = j$ to $s_i = k$. We refer to $f^{(j,k)}$ as the sojourn time distribution for the $j$-to-$k$ transition case. The joint likelihood of the observed states and sojourn times is then given by,

$$\left\{ \prod_{i=1}^{n} p^{(s_{i-1}, s_i)} f^{(s_{i-1}, s_i)}(x_i) \right\} \left\{ \sum_{k=1}^{S} p^{(s_n, k)} S^{(s_n, k)}(x_c) \right\}, \tag{3.1}$$

where $S^{(j,k)}$ is the survival function for the sojourn time distribution $f^{(j,k)}$ and $x_c = T - t_n$ is the observed censoring time. The sum in the second term accounts for the fact

that the state of the censored observation is unknown.

Inference for Markov renewal process point patterns is often performed using likelihood-based methods. A key property of the MRP likelihood is an implied partitioning of sojourn times based on the observed transition cases. To see this more clearly, let $\boldsymbol{x}^{(j,k)} = \{x_i : s_{i-1} = j, s_i = k\}$ denote the set of sojourn times belonging to the $j$-to-$k$ transition case, and let $n^{(j,k)} = |\boldsymbol{x}^{(j,k)}|$ denote the number of corresponding observations. Introducing this notation into the likelihood function, we obtain,

$$\left\{ \prod_{j=1}^{S} \prod_{k=1}^{S} (p^{(j,k)})^{n^{(j,k)}} \left( \prod_{x_i \in \boldsymbol{x}^{(j,k)}} f^{(j,k)}(x_i) \right) \right\} \left\{ \sum_{k=1}^{S} p^{(s_n,k)} S^{(s_n,k)}(x_c) \right\} . \tag{3.2}$$

This factorized representation of the MRP likelihood highlights that the sojourn time distributions $f^{(j,k)}$ are *independent* in the sense that each density is informed only by the corresponding sojourn times in $\boldsymbol{x}^{(j,k)}$. In other words, likelihood-based estimation for MRP models essentially reduces to modeling each $j$-to-$k$ transition case separately. Granted, the censoring term contributes a small amount of crossover between some cases, but the extent of information sharing is minimal; transition case independence is the dominant structure in the likelihood.

This partitioning structure is central to the development of currently available MRP estimation methods. Alvarez (2005) presents a maximum likelihood procedure for an MRP model with Weibull sojourn time distributions. Epifani et al. (2014) approach the same model from a Bayesian perspective and design a Gibbs sampler for posterior inference that similarly exploits the independence between sojourn time distributions. Warr and Woodfield (2020) likewise leverage the conditional independence structure

in developing their nonparametric first-passage time model. Although it is helpful in designing computational strategies, separately estimating each sojourn time distribution is not always appropriate. For many applications, it is reasonable to assume that certain transition cases have similar distributional properties or some other form of dependence. This could include being agnostic to the from-state $j$ or the to-state $k$, having shared properties for self-transitions (all $j$-to-$j$ cases), or even maintaining that all sojourn time distributions are equal, as in the case of a homogeneous renewal process. Additionally, consider a setting where certain transition cases contain very few observations. Results from estimations methods that treat each case independently will be unreliable for these sparse cases. From the perspectives of both application relevance and inferential quality, there is much to be gained from allowing sojourn time distributions to borrow information from each other when supported by the data.

A motivating application for our efforts in this chapter is earthquake recurrence modeling. Many of the earliest applications and extensions of MRP models were in the context of studying seismic activity, where the magnitudes are categorized into a few bins, e.g. low, medium, and high (Patwardhan et al., 1980; Cluff et al., 1980; Altinok and Kolcak, 1999; Rotondi, 2021). Parametric Weibull models are common in this setting (Alvarez, 2005; Epifani et al., 2014), though other distributions and extensions have been explored as well (Garavaglia and Pavani, 2011; Masala, 2012; Votsi et al., 2012). The idea that certain transition cases may share distributional properties is lightly explored by Alvarez (2005), who proposes several constraints on the Weibull parameters and performs post-hoc tests to determine if the constraints are supported by the data.

When binned by magnitude, earthquake transition cases with the *high* designation are often sparse. The Turkey earthquake dataset analyzed in both Altinok and Kolcak (1999) and Alvarez (2005) is a prime example: three out of the nine transition cases contain less than four observations, and estimates for these cases are so poor that they are omitted from the analysis altogether. The Southern California dataset analyzed in Chapter 2.3 also exhibits this issue, which we explore later in this chapter. For this application, and possibly many others, it is clear that assuming independence between sojourn time distribution can undermine the inferential goals of an MRP model,

We refer to the independence of transition-specific sojourn time distributions as the *independent sojourns* property, denoting Markov renewal process models with this property as IS-MRP models. Our goal in this chapter is to develop a flexible modeling framework for Markov renewal processes that supports dependence between sojourn time distributions, i.e. a *dependent sojourns* Markov renewal process model, or DS-MRP. Rather than rely on post-hoc testing, a DS-MRP model allows for organic discovery of dependence structure. Our approach builds on the stick-breaking mixture model developed in Chapter 2 by developing a hierarchical prior structure that enables sharing of information through the mixture components.

## 3.2 Approach to modeling dependence

The mixture modeling framework developed in Chapter 2.1.3 is essential for introducing dependence between sojourn time distributions. Specifically, we make use of

a common-atoms structure to model the collection of sojourn time distributions $\{f^{(j,k)} : j,k \in \mathcal{S}\}$, as in,

$$f^{(j,k)}(x) = \sum_{\ell=1}^{L} \omega_\ell^{(j,k)} f_{\mathcal{K}}(x|\theta_\ell)\,, \tag{3.3}$$

where $\boldsymbol{\omega}^{(j,k)} = \{\omega_\ell^{(j,k)} : \ell = 1,\ldots,L\}$ denotes the set of mixture weights for the $j$-to-$k$ transition case, $f_{\mathcal{K}}(x|\theta)$ is a known kernel density function with parameter(s) $\theta$, and $\boldsymbol{\theta} = \{\theta_\ell : \ell = 1,\ldots,L\}$ denotes the set of mixture kernel parameters, which are common across all transition cases. The common-atoms structure is key to inducing dependence between sojourn distributions, as similar mixture weights $\boldsymbol{\omega}^{(j,k)}$ will lead to similar densities $f^{(j,k)}$.

Recall from Chapter 2.1.2 that the logit stick-breaking process of Ren et al. (2011) assumes the weights $\boldsymbol{\omega}^{(j,k)}$ are given by $\omega_\ell^{(j,k)} = \nu_\ell^{(j,k)} \prod_{m=1}^{\ell-1}(1-\nu_m^{(j,k)})$, where the latent $\nu_\ell^{(j,k)}$ variables are i.i.d. following a logit-normal distribution, or in other words, the transformed variables $\psi_\ell^{(j,k)} = \log(\nu_\ell^{(j,k)}/(1-\nu_\ell^{(j,k)}))$ are normally distributed. Our approach to modeling dependence between sojourn time distributions is to assign a multivariate normal prior with length $S^2$ and carefully designed covariance to the component-specific latent variables $\boldsymbol{\psi}_\ell = \{\psi_\ell^{(j,k)} : j,k \in \mathcal{S}\}$. However, we avoid directly handling the covariance matrix by constructing a latent hierarchical structure which, when the latent variables are marginalized out, leads to a covariance matrix with favorable properties. This preserves convenient conjugate updates and allows for efficient posterior computation while supporting a wide range of dependence structures. The

hierarchical structure is defined as follows:

$$\begin{gathered}\psi_\ell^{(j,k)}|\alpha_\ell^{(j)},\beta_\ell^{(k)},\sigma_\psi^2 \sim \mathrm{N}(\alpha_\ell^{(j)}+\beta_\ell^{(k)},\sigma_\psi^2)\,,\\ \alpha_\ell^{(j)}|\mu_\ell,\sigma_\alpha^2 \sim \mathrm{N}(\mu_\ell/2,\sigma_\alpha^2)\,,\quad \beta_\ell^{(k)}|\mu_\ell,\sigma_\beta^2 \sim \mathrm{N}(\mu_\ell/2,\sigma_\beta^2)\,,\quad \mu_\ell|\sigma_\mu^2 \sim \mathrm{N}(m_\mu,\sigma_\mu^2)\,.\end{gathered} \tag{3.4}$$

Consider marginalizing this model over $\mu_\ell$, $\alpha_\ell^{(j)}$, and $\beta_\ell^{(k)}$. The marginal expectation of $\psi_\ell^{(j,k)}$ is $m_\mu$ for all $j$ and $k$, and the expectation of $\nu_\ell^{(j,k)}$ can be approximated as $(1+\exp(-m_\mu))^{-1}$ using a first-order Taylor approximation. An interesting note is that the analogous expectation under a Dirichlet process is $(1+\alpha)^{-1}$, suggesting a loose similarity between $\exp(-m_\mu)$ and the DP precision parameter $\alpha$. The marginal covariance between different transition cases is structured such that,

$$\begin{aligned}\mathrm{Cov}(\psi_\ell^{(j,k)},\psi_\ell^{(j,k)}|\sigma_\psi^2,\sigma_\alpha^2,\sigma_\beta^2,\sigma_\mu^2) &= \sigma_\psi^2+\sigma_\alpha^2+\sigma_\beta^2+\sigma_\mu^2\,,\\ \mathrm{Cov}(\psi_\ell^{(j,k)},\psi_\ell^{(j,n)}|\sigma_\alpha^2,\sigma_\mu^2) &= \sigma_\alpha^2+\sigma_\mu^2 \text{ for } k\neq n\,,\\ \mathrm{Cov}(\psi_\ell^{(j,k)},\psi_\ell^{(m,k)}|\sigma_\beta^2,\sigma_\mu^2) &= \sigma_\beta^2+\sigma_\mu^2 \text{ for } j\neq m\,,\\ \mathrm{Cov}(\psi_\ell^{(j,k)},\psi_\ell^{(m,n)}|\sigma_\mu^2) &= \sigma_\mu^2 \text{ for } j\neq m \text{ and } k\neq n\,.\end{aligned} \tag{3.5}$$

To better highlight the induced marginal dependence structure, below we show the decomposed covariance matrix for the case where $S=2$:

$$\mathrm{Cov}\left(\begin{bmatrix}\psi_\ell^{(1,1)}\\ \psi_\ell^{(1,2)}\\ \psi_\ell^{(2,1)}\\ \psi_\ell^{(2,2)}\end{bmatrix}\right) = \begin{bmatrix}\sigma_\psi^2 & 0 & 0 & 0\\ 0 & \sigma_\psi^2 & 0 & 0\\ 0 & 0 & \sigma_\psi^2 & 0\\ 0 & 0 & 0 & \sigma_\psi^2\end{bmatrix} + \begin{bmatrix}\sigma_\alpha^2 & \sigma_\alpha^2 & 0 & 0\\ \sigma_\alpha^2 & \sigma_\alpha^2 & 0 & 0\\ 0 & 0 & \sigma_\alpha^2 & \sigma_\alpha^2\\ 0 & 0 & \sigma_\alpha^2 & \sigma_\alpha^2\end{bmatrix} + \begin{bmatrix}\sigma_\beta^2 & 0 & \sigma_\beta^2 & 0\\ 0 & \sigma_\beta^2 & 0 & \sigma_\beta^2\\ \sigma_\beta^2 & 0 & \sigma_\beta^2 & 0\\ 0 & \sigma_\beta^2 & 0 & \sigma_\beta^2\end{bmatrix} + \begin{bmatrix}\sigma_\mu^2 & \sigma_\mu^2 & \sigma_\mu^2 & \sigma_\mu^2\\ \sigma_\mu^2 & \sigma_\mu^2 & \sigma_\mu^2 & \sigma_\mu^2\\ \sigma_\mu^2 & \sigma_\mu^2 & \sigma_\mu^2 & \sigma_\mu^2\\ \sigma_\mu^2 & \sigma_\mu^2 & \sigma_\mu^2 & \sigma_\mu^2\end{bmatrix}.$$

Note the implied conditioning on the variance parameters $\sigma_\psi^2$, $\sigma_\alpha^2$, $\sigma_\beta^2$, and $\sigma_\mu^2$, which are omitted for brevity. If we ensure that the order of terms in $\boldsymbol{\psi}_\ell$ is similar, iterating over each to-state $k$ before moving to the next from-state $j$, then the marginal covariance

matrix for any number of states $S$ given the variance parameters can be written as, $\sigma^2_\psi(\boldsymbol{I} \otimes \boldsymbol{I}) + \sigma^2_\alpha(\boldsymbol{J} \otimes \boldsymbol{I}) + \sigma^2_\beta(\boldsymbol{I} \otimes \boldsymbol{J}) + \sigma^2_\mu(\boldsymbol{J} \otimes \boldsymbol{J})$, where $\boldsymbol{I}$ denotes the $S \times S$ identity matrix, $\boldsymbol{J}$ denotes the $S \times S$ matrix of ones, and $\otimes$ denotes the Kronecker product. This highlights that the marginal covariance is a linear combination of four different dependence structures. The diagonal structure $\sigma^2_\psi(\boldsymbol{I} \otimes \boldsymbol{I})$ represents independent sojourn time distributions, meaning that if $\sigma^2_\psi$ is very large compared to the other variances, then the logit stick-breaking elements will be independent across transition cases. However, it should be noted that an IS-MRP model does not arise from this covariance decomposition as a limiting case; independent weights do not imply independent densities due to the common-atoms kernel structure. The structure $\sigma^2_\alpha(\boldsymbol{J} \otimes \boldsymbol{I})$ represents dependence where the elements are shared within common from-states $j$. Similarly, the structure $\sigma^2_\beta(\boldsymbol{I} \otimes \boldsymbol{J})$ represents dependence where logit stick-breaking elements are equal within common to-states $k$. Finally, the full matrix structure $\sigma^2_\mu(\boldsymbol{J} \otimes \boldsymbol{J})$ represents total dependence where elements are equal across all transition cases, which corresponds to a homogeneous renewal process. The relative magnitudes of the four marginal variances determine the estimated dependence structure of the logit stick-breaking elements.

We note that this covariance decomposition does not form a basis for all covariance matrices, but it does cover a wide range of dependence structures. Although one could opt for a more general covariance matrix, we find that the proposed structure is sufficient for most applications and has a large advantage in terms of computational ease. In addition, this more structured approach offers cleaner interpretation and better complements common inference objectives when studying MRP point patterns.

## 3.3 Hierarchical model formulation

Having developed the foundational dependence structure, our next step is to select the kernel model and present the model hierarchy in full. The choice of kernel density $f_{\mathcal{K}}$ is crucial to the flexibility and properties of the model. Our discussion from Chapter 2.1.1 is directly applicable here, suggesting the gamma and uniform kernels as strong candidates. Recall the previous discussion on earthquake inter-arrival times and the mechanistic evidence of decreasing density shapes (Corral, 2005; Kumar et al., 2020). Given our motivation for modeling earthquake point patterns, we focus our developments on the uniform kernel. As a review, the uniform kernel density for component $\ell$ is given by $\text{Unif}(x|0, \theta_\ell)$. The prior structure is unchanged from the HRP model:

$$\theta_\ell|\eta_\theta, \gamma_\theta \overset{i.i.d.}{\sim} \text{inv-Ga}(\eta_\theta, \gamma_\theta)\,, \quad \gamma_\theta \sim \text{Ga}(a^\theta_\gamma, b^\theta_\gamma)\,, \quad (\eta_\theta - 1) \sim \text{Ga}(a^\theta_\eta, b^\theta_\eta)\,, \tag{3.6}$$

where $a^\theta_\gamma$, $b^\theta_\gamma$, $a^\theta_\eta$, and $b^\theta_\eta$ are fixed hyperparameters.

To facilitate posterior computation, we introduce latent variables $\boldsymbol{z} = \{z_i : i = 1, \ldots, n\}$ such that $\Pr(z_i = \ell|s_i, s_{i-1}, \boldsymbol{\omega}) = \omega_\ell^{(s_{i-1}, s_i)}$ where $\boldsymbol{\omega} = \{\boldsymbol{\omega}^{(j,k)} : j, k \in \mathcal{S}\}$ is the full set of mixture weights. An additional latent variable $z_c$ is introduced for the censored sojourn time $x_c = T - t_n$ with $\Pr(z_c = \ell|s_n, s_c, \boldsymbol{\omega}) = \omega_\ell^{(s_n, s_c)}$. We also treat the unobserved state of the censored observation as a latent variable, denoted $s_c$, following the same Markov chain structure, $\Pr(s_c = k|s_n, \boldsymbol{p}) = p^{(s_n,k)}$ where $\boldsymbol{p} = \{p^{(j,k)} : j, k \in \mathcal{S}\}$ is the set of transition probabilities. Given these latent variables and the mixture model

structure, the augmented data likelihood is given by,

$$\left\{\prod_{j=1}^{S}\prod_{k=1}^{S}(p^{(j,k)})^{n^{(j,k)}}\left(\prod_{x_i\in\boldsymbol{x}^{(j,k)}}f_{\mathcal{K}}(x_i|\theta_{z_i})\right)\right\}\left\{p^{(s_n,s_c)}S_{\mathcal{K}}(x_c|\theta_{z_c})\right\}, \tag{3.7}$$

where $S_{\mathcal{K}}$ is the survival function for the (uniform) kernel density $f_{\mathcal{K}}$. Let $\boldsymbol{p}^{(j)} = \{p^{(j,k)} : k = 1,\ldots,S\}$ denote the set of transition probabilities starting from state $j$. For each $j \in \mathcal{S}$, we assign a Dirichlet prior of length $S$ to $\boldsymbol{p}^{(j)}$, denoted $\boldsymbol{p}^{(j)}|\alpha_p \overset{i.i.d.}{\sim} \text{Dir}_S(\alpha_p,\ldots,\alpha_p)$ where $\alpha_p$ is a fixed hyperparameter.

The mixture weights $\omega_\ell^{(j,k)}$ arise from the logit stick-breaking prior structure described in the previous section, which we repeat here for completeness:

$$\begin{gathered}\omega_\ell^{(j,k)} = \frac{\exp(\psi_\ell^{(j,k)})}{1+\exp(\psi_\ell^{(j,k)})}\prod_{r=1}^{\ell-1}\frac{1}{1+\exp(\psi_r^{(j,k)})}, \\ \psi_\ell^{(j,k)}|\alpha_\ell^{(j)},\beta_\ell^{(k)},\sigma_\psi^2 \sim \text{N}(\alpha_\ell^{(j)}+\beta_\ell^{(k)},\sigma_\psi^2), \\ \alpha_\ell^{(j)}|\mu_\ell,\sigma_\alpha^2 \sim \text{N}(\mu_\ell/2,\sigma_\alpha^2), \quad \beta_\ell^{(k)}|\mu_\ell,\sigma_\beta^2 \sim \text{N}(\mu_\ell/2,\sigma_\beta^2), \quad \mu_\ell|\sigma_\mu^2 \sim \text{N}(m_\mu,\sigma_\mu^2),\end{gathered} \tag{3.8}$$

where $m_\mu$ is a fixed hyperparameter. It is important to note that, because the final weight $\omega_L^{(j,k)}$ is determined by the sum-to-one constraint of the weights, the only $L-1$ stick-breaking elements $\psi_\ell^{(j,k)}$ are required for each transition case, hence the prior weight structure is defined for $\ell = 1,\ldots,L-1$. We assign inverse-gamma priors to the variance parameters: $\sigma_\psi^2 \sim \text{inv-Ga}(a_\sigma^\psi, b_\sigma^\psi)$, $\sigma_\alpha^2 \sim \text{inv-Ga}(a_\sigma^\alpha, b_\sigma^\alpha)$, $\sigma_\beta^2 \sim \text{inv-Ga}(a_\sigma^\beta, b_\sigma^\beta)$, and $\sigma_\mu^2 \sim \text{inv-Ga}(a_\sigma^\mu, b_\sigma^\mu)$, where $a_\sigma^\psi$, $a_\sigma^\alpha$, $a_\sigma^\beta$, $a_\sigma^\mu$, $b_\sigma^\psi$, $b_\sigma^\alpha$, $b_\sigma^\beta$, and $b_\sigma^\mu$ are fixed hyperparameters.

We conclude this section with a few notes on the fixed hyperparameter values. Recall that $a_\eta > 1$ is required for the uniform mixture model to satisfy the finite mean sufficient conditions. Considering that the mixture kernel plays a similar role in both HRP and MRP settings, we suggest the same default hyperparameter values as

discussed in Chapter 2.1.1. Regarding the variance priors, our sensitivity experiments suggest that results are quite robust to even very informative values. We suggest using an inv-Ga$(2.5, 1)$ for all marginal variance priors as a default. The impact of $m_\mu$ is similarly inconsequential, so we suggest $m_\mu = 0$ as a default. The value of $\alpha_p$ controls the influence of the Dirichlet prior on the transition probabilities. In settings where some transition cases contain very few observations, standard choices such as $\alpha_p = 1$ can have meaningful effects on the posterior. We find that a lower default value of $\alpha_p = 0.1$ allows the data in each transition case to be the strongest influence on the posterior, even when the number of observations is small. Our comments in Chapters 2.1.3 and 2.4 on selecting $L$ are applicable here as well, however, the presence of a common-atoms structure complicates the decision as more components may be needed if the sojourn time distributions are more diverse. For results that follow later in this chapter, we select $L = 200$ as a conservative choice, which we find is more than sufficient in practice.

## 3.4 Posterior inference

Here we outline notable details of the posterior inference scheme used to sample from our proposed model and discuss additional inferences of interest. Emphasis is given to the novel aspects of the model, and we refer the reader to Appendix B for a full description of the sampling algorithm.

### 3.4.1 Posterior simulation for model parameters

The modeling structures we have examined support conditionally conjugate updates for many parameters, and straightforward sampling techniques such as Metropolis-Hastings steps are available for several others. However, a level of care is required in handling the latent variables $z_c$ and $s_c$.

The update protocol for the kernel parameters follows the same structure as in Chapter 2.1.3. Updates for latent variables $\boldsymbol{z}$ and $z_c$ are also similar, though the expressions involving the mixture weights are now indexed by the corresponding $j$-to-$k$ transition case. Straightforward conjugate updates are possible for the transition probabilities $\boldsymbol{p}^{(j)}$ given the transition counts $n^{(j,k)}$, adding one to the transition case that includes $s_n$-to-$s_c$. The missing state $s_c$ is sampled from a discrete distribution based on $z_c$, the mixture weights, and the transition probabilities $\boldsymbol{p}^{(s_n)}$. We draw the marginal variances $\sigma^2_\psi$, $\sigma^2_\alpha$, $\sigma^2_\beta$, and $\sigma^2_\mu$, as well as the hierarchical variables $\alpha^{(j)}_\ell$, $\beta^{(k)}_\ell$, and $\mu_\ell$, using standard conjugate relationships.

The weights $\boldsymbol{\omega}^{(j,k)}$ are updated deterministically using the logit stick-breaking elements $\boldsymbol{\psi}^{(j,k)}$. To sample from the conditional distribution of $\psi^{(j,k)}_\ell$, we rely on the Pólya-Gamma data augmentation technique of Polson et al. (2013). This method makes use of the following property:

$$\frac{(e^\psi)^a}{(1+e^\psi)^b} = 2^{-b} e^{(a-b/2)\psi} \int_0^\infty e^{-\xi\psi^2/2} \mathrm{PG}(\xi|b,0) d\xi \,, \tag{3.9}$$

where $\mathrm{PG}(x|b,0)$ denotes the Pólya-Gamma density with shape parameter $b$ and tilt parameter 0. This property is leveraged by introducing a Pólya-Gamma random vari-

able $\xi_\ell^{(j,k)}$ for each logit stick-breaking element $\psi_\ell^{(j,k)}$. The result is a joint conditional distribution that is sampled using a Gibbs step, where the latent variable $\xi_\ell^{(j,k)}$ is drawn from a $\text{PG}(n_\ell + \ldots + n_L, \psi_\ell^{(j,k)})$ distribution, where $n_\ell = |\{i : z_i = \ell\}|$, and the logit stick-breaking element $\psi_\ell^{(j,k)}$ is drawn from a $\text{N}(n_\ell + \ldots + n_L, \xi_\ell^{(j,k)})$ distribution. Sampling from the Pólya-Gamma distribution is not trivial, but very efficient rejection algorithms are available (Polson et al., 2013; Windle et al., 2014). Implementation details connecting this to the other hierarchical weight parameters are provided in Appendix B.

### 3.4.2 Model validation and additional inferences

Here we consider additional model inferences beyond estimation of model parameters and sojourn distribution densities, focusing on model validation and quantities specific to earthquake recurrence analysis. For each of these, we provide a brief description of the computational details and discuss relevant interpretations.

#### Case-specific predictive coverage

This inference attempts to quantify how well each sojourn distribution aligns with the corresponding observations. Notably, it ignores the embedded Markov chain structure and focuses on each sojourn distribution in isolation. Readers are free to consider a variety of well explored density estimation assessment techniques, but we opt for the predictive coverage approach employed by Epifani et al. (2014). For each $j$-to-$k$ transition case, a single value is drawn from the modeled sojourn distribution $f^{(j,k)}$. This is repeated for each MCMC iteration and the final collections of draws are

used to construct posterior predictive intervals for a given coverage level, from which the observed coverage rate is calculated as the proportion of observed sojourn times falling within the corresponding predictive interval. An observed coverage rate being close to the specified coverage level indicates that the model is well-calibrated for that transition case.

In addition to predictive coverage, other factors such as predictive interval width or asymmetry can be examined to paint a more complete picture of predictive quality. As a compliment to the coverage rate, we also compute the *interval scores* introduced by Gneiting and Raftery (2007). The interval score is a proper scoring rule designed to favor predictive intervals that are both narrow and adequately calibrated. For an observation $x$ and $(1-\alpha)\times 100\%$ predictive interval $(l, u)$, the interval score is given by,

$$(u-l) + \frac{2}{\alpha}(l-x)\mathbb{I}(x<l) + \frac{2}{\alpha}(x-u)\mathbb{I}(x>u)\,, \tag{3.10}$$

where $\mathbb{I}(\cdot)$ is the indicator function. The interval score is calculated for each observation according to its transition-specific predictive interval, and then the average interval scores across all cases are reported and compared between models, with lower values being preferred. We acknowledge that this scoring rule is not typically used in the context of point processes. However, given the absence of a standard metric in this setting, we find that it conveys useful information not captured by the coverage rate alone.

**Predictive state recurrence assessments**

Consider observing Markov renewal data and let $\boldsymbol{t}^{(j)} = \{t_i : s_i = j\}$ denote the subset of event times corresponding to state $j$ with $n^{(j)} = |\boldsymbol{t}^{(j)}|$. We define the *state recurrence times* as the times between consecutive observations of state $j$, denoted $r_{i'}^{(j)} = t_{i'}^{(j)} - t_{i'-1}^{(j)}$ for $i' = 1, \ldots, n^{(j)}$. It has been shown that the state recurrence time sequence forms a homogeneous renewal process (Pyke, 1961a; Limnios and Oprişan, 2001). This property suggests a framework for assessing MRP model performance that incorporates both the Markov chain and renewal process structures. We note that state recurrence time distributions are equivalent to the concept of first-passage time distributions found in the semi-Markov process literature. Exact expressions for the Laplace transforms of these distributions are available (see Warr and Woodfield, 2020 for a recent review), but the computational burden of inverting several Laplace transform terms for each MCMC iteration is prohibitively high. Instead, we opt for a posterior predictive approach.

We begin by generating posterior predictive samples. For each MCMC iteration, we use the current parameter values to simulate a Markov renewal point pattern within the interval $(0, T)$, which involves iteratively generating states and conditionally drawing sojourn times from the corresponding distributions. From this we construct the corresponding sets of state recurrence times $\boldsymbol{r}^{(j)}$ for each state $j$. Then the empirical CDF of the simulated state recurrence times is evaluated at a dense grid of points and stored. After completing this process for each MCMC iteration, we have a collection of

empirical CDFs for each state and can report posterior predictive means and intervals for each state recurrence time distribution. In addition, we make use of the *ECDF Error* metric given in Chapter 2.1.4 to compare the empirical CDFs of the observed and simulated state recurrence times. A small ECDF Error indicates that the model predictions align well with the observed data for that state recurrence time distribution.

**Note on the $K$-function and time-rescaling QQ plots**

Absent from our inference discussion are comments on the $K$-function and time-rescaling QQ plots. Recall that the $K$-function is used to assess clustering behavior in point patterns and is often of interest in HRP modeling applications. However, there is no analogous quantity in Markov renewal theory. The Markov renewal function, unlike the homogeneous renewal function, does not summarize the entire point process, but is instead a measure of state recurrence properties (Pyke, 1961a). $K$-function estimates for the individual sojourn time distributions can be estimated, but this is an incomplete picture of overall behavior and may not accurately reflect clustering properties.

Time-rescaling QQ plots are a useful tool for assessing the goodness-of-fit of a point process model. The challenge in applying them to an MRP model lies in accounting for the Markov chain structure of the states in the conditional intensity function (Daley and Vere-Jones, 2003). Obtaining the appropriate conditional intensity function expression is a non-trivial theoretical task and is beyond the scope of this work.

**Earthquake recurrence features**

In addition to model assessment, we consider inferences specific to earthquake recurrence modeling. We briefly discuss two measures related to conditional probabilities of state transitions: the *cross state-probability* and the *transition risk profile*. The cross state-probability (CSP) is the probability of, starting at some time $t_0$ and some state $j$, transitioning to state $k$ within some time interval $\Delta t$:

$$\mathrm{CSP}^{(j,k)}_{t_0|\Delta t} = \frac{p^{(j,k)}(S^{(j,k)}(t_0) - S^{(j,k)}(t_0 + \Delta t))}{\sum_{k'} p^{(j,k')} S^{(j,k')}(t_0)} \,. \tag{3.11}$$

These are often used to assess the predictive characteristics of specific earthquake events. For example, Epifani et al. (2014) manually compare the CSP values for several earthquakes and visually check for non-uniform patterns as the time $t_0$ is moved forward. These manual comparisons are difficult to scale and appear to offer little in terms of interesting application insights, instead being more of a crude model assessment tool. However, they do inspire a similar inference quantity that we find to be more useful.

We define the transition risk profile (TRP) as the probability of transitioning to a state $k$ given that an event occurs immediately, the current state is $j$, and the time since the last event is $t$:

$$\mathrm{TRP}^{(j,k)}(t) = \Pr(s_{i+1} = k | s_i = j, x_i = t) = \frac{p^{(j,k)} f^{(j,k)}(t)}{\sum_{k'} p^{(j,k')} f^{(j,k')}(t)} \,. \tag{3.12}$$

This quantity is similar to the CSP, but is conditioned on an event happening immediately and allows for a more direct examination of the relative risk of different event types and how that risk changes over time.

## 3.5 Synthetic data examples

In this section we demonstrate the utility of the proposed DS-MRP mixture model using simulated point patterns and compare the performance against that of alternatives. The primary comparison we study is against an IS-MRP version of the model, which independently applies the nonparametric mixture model from Chapter 2 to each transition case. We also include a parametric Weibull model (which also meets the IS-MRP definition) as a benchmark for comparison.

We consider two simulation scenarios, one special case and one more general. The special case, referred to as Scenario 1, is that of a homogeneous renewal process along with an independent Markov chain of three states with even transition probabilities of $1/3$. In terms of an MRP, this is the simplest setting where all sojourn time distributions are identical. We generate a point pattern using a common $\text{Ga}(1/3, 3)$ sojourn time distribution and a time bound $T = 500$, yielding approximately 500 events. The other simulation scenario, Scenario 2, has varying sojourn time distributions between transition cases. Specifically, we construct the $j$-to-$k$ sojourn time distribution using an evenly weighted mixture of a Lomax with shape 2 and scale 1, and a $\text{Ga}(k - 2/3, 4/k)$ distribution, also with even transition probabilities. The value of $T = 1000$ is chosen to produce approximately 500 events in this scenario as well.

Table 3.1: Posterior mean estimates and 95% uncertainty intervals for the marginal variance parameters for the proposed model, fit to a point pattern generated from Scenario 1. The relative magnitude of these values reflect the prevalence of different dependence structures between sojourn time distributions.

| | $\sigma^2_\psi$ | $\sigma^2_\alpha$ | $\sigma^2_\beta$ | $\sigma^2_\mu$ |
|---|---|---|---|---|
| Posterior mean | 0.30 | 0.34 | 0.35 | 5.88 |
| 95% interval | (0.14, 0.62) | (0.13, 0.69) | (0.13, 0.65) | (3.16, 11.54) |

**Scenario 1 - Common sojourn time distributions**

This scenario is designed to assess the ability of the proposed model to recover the common sojourn time distribution structure, and to highlight the impact that information sharing across cases has on performance. We fit the model default hyperparameter values discussed previously. The MCMC algorithm is run for 5,000 iterations following a burnin period of 15,000. In Table 3.1 we show the posterior mean and 95% credible interval for the marginal variance parameters. The estimate for $\sigma^2_\mu$ is an order of magnitude larger than the other variances, indicating strong evidence for a common sojourn time distribution structure in the data.

For the sake of brevity, we move the density estimate plots for our proposed DS-MRP mixture model, as well as the corresponding IS-MRP mixture and parametric Weibull model results, to Appendix C. As a brief summary, the DS-MRP mixture model provides a good fit to the true sojourn time distribution, with the IS-MRP mixture model also performing well, but with more variability in the uncertainty intervals. The IS-MRP Weibull model, as expected, struggles slightly to capture the true distribution shape. As a numerical comparison, we report the predictive coverage rates and average interval scores for each model in Table 3.2. These results demonstrate both the need

Table 3.2: Predictive coverage rates and average interval scores for simulated sojourn times under Scenario 1 for the proposed DS-MRP mixture model with uniform kernel, the corresponding IS-MRP mixture model, and the parametric IS-MRP Weibull model. Value pairs are separated by a backslash with the coverage rate as the first value and the average interval score as the second.

| Transition case | DS-MRP mixture | IS-MRP mixture | Weibull IS-MRP |
|---|---|---|---|
| 1-to-1 | 0.932 \ 4.36 | 0.890 \ 3.19 | 0.945 \ 5.82 |
| 1-to-2 | 0.967 \ 7.03 | 0.900 \ 6.74 | 0.933 \ 7.07 |
| 1-to-3 | 0.904 \ 12.0 | 0.863 \ 12.1 | 0.932 \ 12.9 |
| 2-to-1 | 0.918 \ 5.93 | 0.885 \ 5.77 | 0.934 \ 8.71 |
| 2-to-2 | 0.964 \ 7.29 | 0.873 \ 7.48 | 0.964 \ 7.19 |
| 2-to-3 | 0.946 \ 8.60 | 0.875 \ 7.90 | 0.946 \ 8.81 |
| 3-to-1 | 0.901 \ 7.44 | 0.873 \ 7.37 | 0.930 \ 9.61 |
| 3-to-2 | 0.931 \ 10.6 | 0.897 \ 11.1 | 0.948 \ 13.0 |
| 3-to-3 | 0.908 \ 5.87 | 0.882 \ 5.42 | 0.947 \ 8.87 |

for multifaceted coverage evaluation and the performance advantages of the proposed model. In terms of coverage alone, the IS-MRP Weibull model rates are slightly closer to the desired level than the DS-MRP mixture model, while the IS-MRP mixture model is substantially lower. However, the average interval scores tell a more detailed story. In nearly all cases, the IS-MRP mixture model produces the lowest average interval scores and the Weibull model produces the highest. This suggests that the Weibull model coverage rates are created by overly wide intervals, whereas the IS-MRP mixture model has comparatively narrow intervals that are not well-calibrated. Sitting between these extremes is the DS-MRP mixture model, having decently calibrated intervals that capture the data without being overly wide.

Turning attention to predictive state recurrence checks, Table 3.3 shows the posterior means and 95% credible intervals for ECDF Error values. The comparison between the IS-MRP mixture and the parametric Weibull model shows the value added by having a flexible prior model for the sojourn time densities. The improvement of

Table 3.3: Predictive state recurrence ECDF errors for sojourn times simulated under Scenario 1. Posterior mean estimates and 95% uncertainty bounds are given for the DS-MRP mixture model, the corresponding IS-MRP mixture, and the parametric Weibull IS-MRP model.

| MRP Model | Initial State | Mean ECDF Error | 95% Interval |
|---|---|---|---|
| DS-MRP Mixture | 1 | 0.018 | (0.004, 0.070) |
| | 2 | 0.026 | (0.005, 0.091) |
| | 3 | 0.018 | (0.004, 0.067) |
| IS-MRP Mixture | 1 | 0.019 | (0.004, 0.062) |
| | 2 | 0.051 | (0.005, 0.090) |
| | 3 | 0.017 | (0.003, 0.060) |
| Weibull IS-MRP | 1 | 0.051 | (0.009, 0.185) |
| | 2 | 0.071 | (0.013, 0.247) |
| | 3 | 0.048 | (0.008, 0.167) |

the DS-MRP mixture model over the IS-MRP mixture model is less pronounced, being evident in only one of the three states. This is due in part to the common sojourn time distribution structure in Scenario 1, which results in identical recurrence time characteristics for each state.

**Scenario 2 - Varying sojourn time distributions**

This scenario employs a more complex dependence structure between sojourn time distributions. There is an element of commonality between all cases due to the Lomax mixture component, but the varying gamma components also introduce diversity between cases with different to-states $k$. We fit the same model as in Scenario 1 to a point pattern generated from Scenario 2, with the same MCMC settings. Posterior mean estimates and 95% credible intervals for the marginal variance parameters are shown in Table 3.4. These results match what we would expect for this scenario. The parameters $\sigma_\mu^2$ and $\sigma_\beta^2$ are larger than the others, suggesting the presence of both a fully

Table 3.4: Posterior mean estimates and 95% uncertainty intervals for the marginal variance parameters for the proposed model, fit to a point pattern generated from Scenario 2. The relative magnitude of these values reflect the prevalence of different dependence structures between sojourn time distributions.

| | $\sigma^2_\psi$ | $\sigma^2_\alpha$ | $\sigma^2_\beta$ | $\sigma^2_\mu$ |
|---|---|---|---|---|
| Posterior mean | 0.37 | 0.31 | 1.14 | 2.45 |
| 95% interval | (0.14, 0.89) | (0.12, 0.65) | (0.49, 2.21) | (0.93, 4.66) |

Table 3.5: Predictive state recurrence ECDF errors for sojourn times simulated under Scenario 2. Posterior mean estimates and 95% uncertainty bounds are given for the DS-MRP mixture model, the corresponding IS-MRP mixture, and the parametric Weibull IS-MRP model.

| MRP Model | Initial State | Mean ECDF Error | 95% Interval |
|---|---|---|---|
| DS-MRP Mixture | 1 | 0.037 | (0.009, 0.136) |
| | 2 | 0.027 | (0.006, 0.088) |
| | 3 | 0.033 | (0.007, 0.111) |
| IS-MRP Mixture | 1 | 0.045 | (0.009, 0.168) |
| | 2 | 0.036 | (0.006, 0.109) |
| | 3 | 0.043 | (0.008, 0.149) |
| Weibull IS-MRP | 1 | 0.052 | (0.010, 0.201) |
| | 2 | 0.047 | (0.011, 0.142) |
| | 3 | 0.052 | (0.010, 0.176) |

shared structure and a common to-state structure. The relative estimate magnitudes are not as distinct though, but this improves if more data are included.

Comparing the density estimates for the three models in Scenario 2 (provided in Appendix C), we see similar conclusions as those in Scenario 1. In particular, the failure of the Weibull model to capture the true density shape is more clearly displayed. The predictive coverage results mirror those from Scenario 1, with the DS-MRP mixture providing the best balance between coverage and interval width. A table of coverage and interval scores for this scenario is provided in Appendix C. Finally, we report the predictive state recurrence ECDF errors in Table 3.5. With this metric, DS-MRP mixture has a more pronounced advantage over the IS-MRP mixture model.

To summarize, in both simulation scenarios, the proposed DS-MRP mixture model produces favorable results compared to the IS-MRP mixture model and the parametric Weibull IS-MRP model. The density estimates are more accurate, the trade-off between predictive coverage and interval width is more reasonable, and the predictive state recurrence ECDF errors are smaller. Together, these results demonstrate the performance advantages that arise from incorporating dependence between sojourn time distributions, especially in settings with limited amounts of data. As a final check to validate these results, we conduct a simulation study using multiple datasets.

**Simulation study with multiple datasets**

Our goal for this simulation study is not to present a rigorous analysis, but rather to dispel concerns that the previous findings could be attributed to idiosyncrasies of a single dataset, as can happen when working with limited amounts of data. Though admittedly crude, our approach is to generate 100 point patterns from both Scenario 1 and Scenario 2, as well as a third scenario with the same structure as Scenario 2, but with uneven transition probabilities of $\boldsymbol{p}^{(j)} = (0.6, 0.3, 0.1)$, common for all starting states $j$. We compute the estimated state recurrence ECDF errors for each point pattern, averaged across all initial states, and display boxplots of the estimated results in Figure 3.1. There is a clear trend in these results that is consistent with the previous findings, all in favor of the proposed DS-MRP mixture model. Of course, this analysis is limited, ignoring several other factors such as the uncertainty estimates for the ECDF errors, as well as the other model evaluation metrics. Yet, even this basic assessment provides

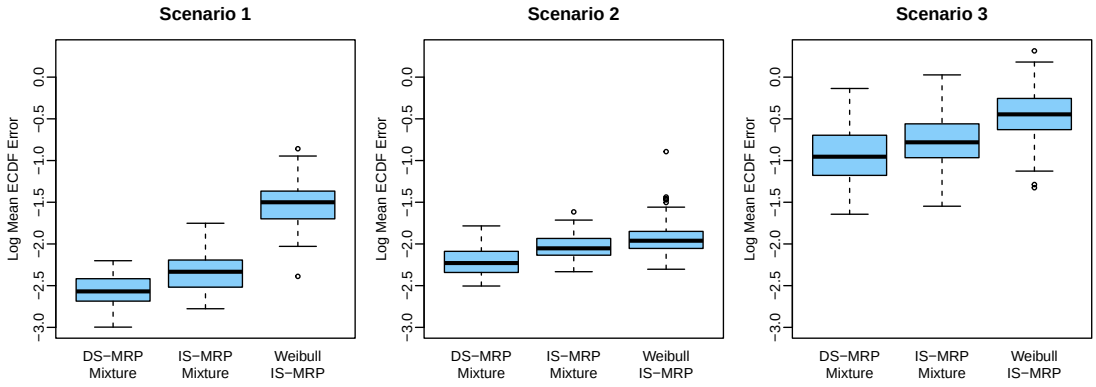


Figure 3.1: Boxplots of the estimated predictive state recurrence ECDF errors for the DS-MRP mixture model, the IS-MRP mixture model, and the parametric Weibull IS-MRP model under three different simulation scenarios. Each boxplot is based on 100 point patterns generated under the given scenario and reports the mean ECDF error estimates for each model fit.

some reassurance that the previous conclusions are sound.

## 3.6 Southern California earthquake data application

In this section we apply our proposed methodology to the Southern California earthquake dataset (SCEDC, 2013) using the proposed DS-MRP mixture model. Similar results for the IS-MRP mixture model and the parametric Weibull IS-MRP model are included in Appendix C. Recall that the dataset consists of 1,248 earthquakes, each with a magnitude of at least 4.0 and a maximum observed magnitude of 7.3. We categorize the magnitudes into three groups: *low* indicates a magnitude between 4.0 and 4.49, *medium* indicates a magnitude between 4.5 and 4.99, and *high* indicates a magnitude larger than 5.0. These cutoffs are roughly consistent with similar MRP earthquake studies (Epifani et al., 2014; Garavaglia and Pavani, 2011). From here, the data can be partitioned into nine transition cases. The largest group is the low-to-low transition case, which contains

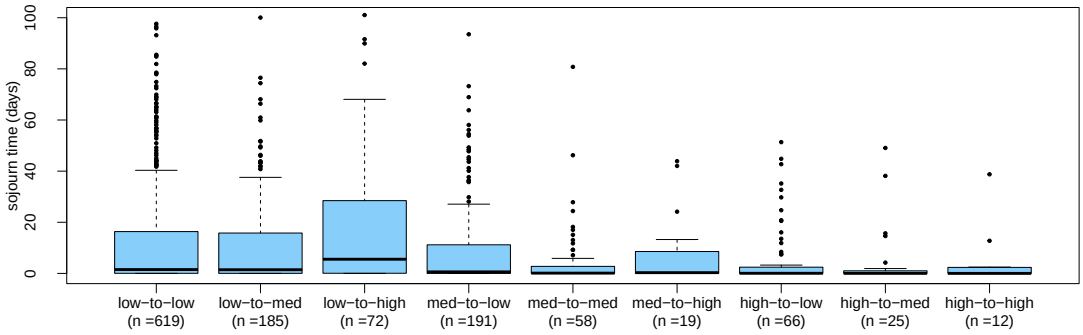


Figure 3.2: Boxplots of observed sojourn times for each transition case in the Southern California earthquake dataset. Sample sizes are included below transition case labels. Not shown are 18 observations with sojourn times greater than 100 days, the largest of which is 232 days. The low-to-low case contains 9 of these outliers, and all but 2 are in cases transitioning from the low state.

619 observations, and the smallest group is the high-to-high transition case, containing just 12 observations. Boxplots of observed sojourn times for each transition case are shown in Figure 3.2.

As an aside, which methodology is used to measure magnitude depends on a number of factors and has changed over time as technology has improved. Because magnitudes are reported using different methods, events in the full SCEDC dataset are not directly comparable across large gaps in time. Our data begins in 1981, which coincides with both a significant upgrade in monitoring technology and the establishment of a more standardized protocol for measuring earthquake magnitudes (Hutton et al., 2010). In addition, using magnitude labels rather than numerical values may further alleviate potential bias caused by mixing different measurement methods.

We fit the DS-MRP mixture model to the earthquake data using the same hyperparameters and MCMC configuration as in the simulation study. Computation time is comparable to the data analysis in Chapter 2.3, taking roughly 15 seconds per

Table 3.6: Posterior mean estimates and 95% uncertainty intervals for the marginal variance parameters for the proposed model, fit to the Southern California earthquake dataset. The relative magnitude of these values reflect the prevalence of different dependence structures between sojourn time distributions.

| | $\sigma^2_\psi$ | $\sigma^2_\alpha$ | $\sigma^2_\beta$ | $\sigma^2_\mu$ |
|---|---|---|---|---|
| Posterior Mean | 0.19 | 0.32 | 0.22 | 8.71 |
| 95% Interval | (0.09, 0.35) | (0.13, 0.58) | (0.10, 0.42) | (5.49, 12.99) |

1,000 iterations. Sampling efficiency for the sojourn densities is also similar. The estimated sojourn time densities are shown in Figure 3.3. We note that the data histograms are consistent with the decreasing density assumption, but as a sanity check, we also implemented the model with the gamma kernel, which also produces decreasing density estimates. Table 3.6 shows the posterior mean estimates and 95% uncertainty intervals for the covariance structure parameters. The large posterior mean estimate for $\sigma^2_\mu$ suggests a common sojourn time distribution structure across all transition cases, similar to the structure of Scenario 1 from Section 3.5.

To assess the quality of model fit, we turn to the posterior predictive checks described previously. Graphical and tabular results for the predictive state recurrence ECDF errors are provided in Appendix C. Figure 3.4 shows 95% posterior predictive coverage diagrams for the observed sojourn times. Given the variety of sample sizes and the heavy concentration of observations near zero, the predictive coverage rates are encouraging. Coverage rates for the small data cases are not as close to 95% as the larger cases, but they are still within a reasonable range.

In combination, these posterior predictive checks suggest that the DS-MRP mixture model is a good fit for the earthquake data. In Appendix C, we provide figures

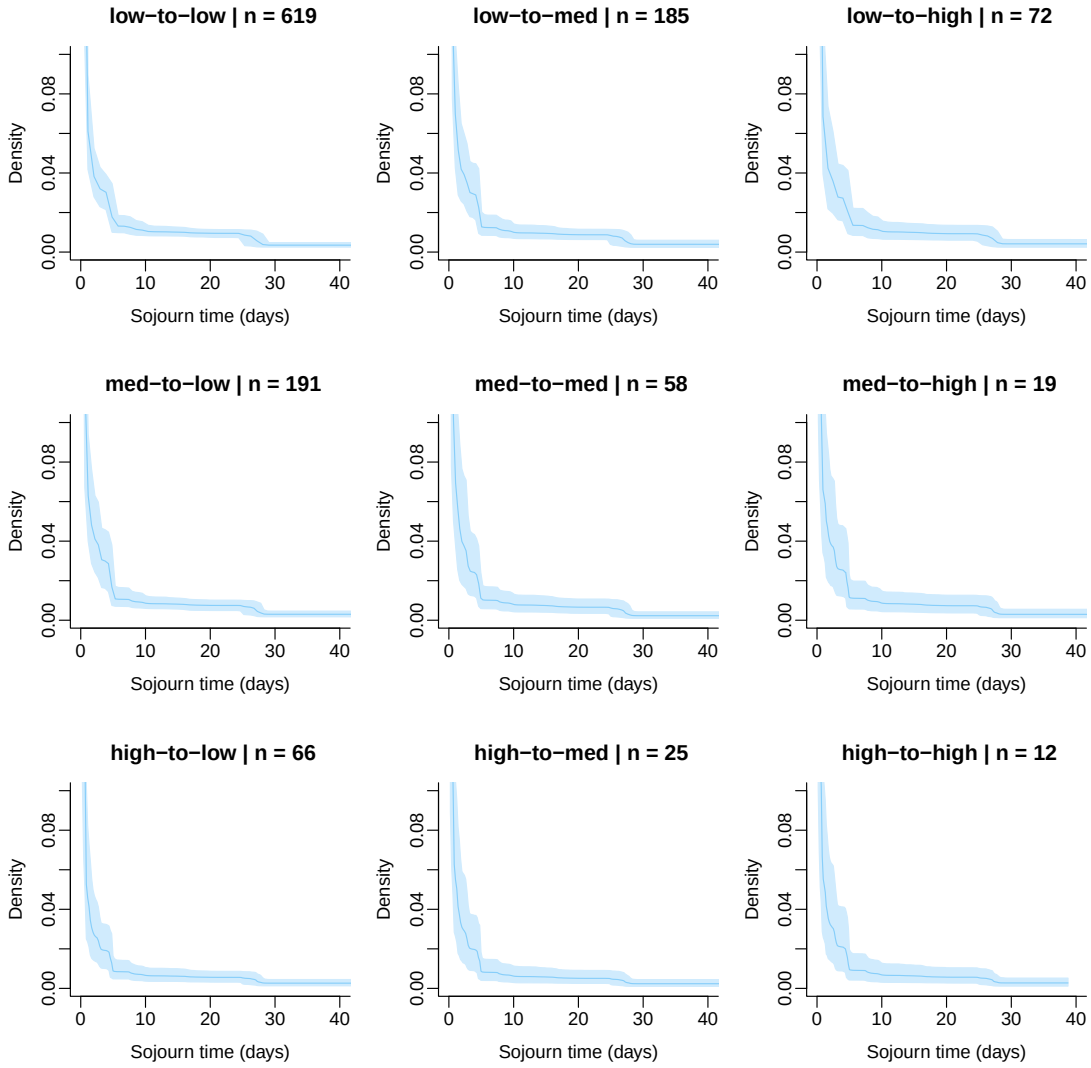


Figure 3.3: Posterior estimates and 95% uncertainty intervals for the sojourn time densities of the Southern California earthquake dataset, estimated using the DS-MRP mixture model. Mean estimates are shown as solid lines with 95% credible intervals as shaded regions.

and extended discussion comparing this model to the IS-MRP mixture model and the parametric Weibull IS-MRP model. The main takeaway is that the proposed model outperforms the other two in all measures we have considered. The Weibull model performs particularly poorly, likely due to the high concentration of observed values

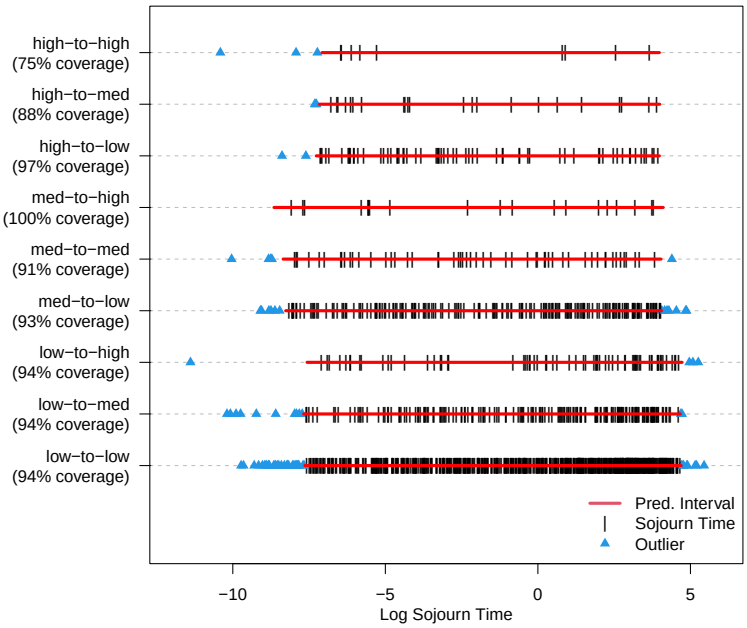


Figure 3.4: Posterior predictive coverage diagram for the Southern California earthquake dataset fit using the DS-MRP mixture model. Red lines are used to indicate the 95% predictive intervals for each transition case. Data are partitioned by transition case and are shown as blue triangles if excluded from the 95% predictive interval and as black ticks if included. Numerical coverage levels are shown beneath the transition case labels.

near zero, producing excessively large predictive coverage intervals and highly variable predictive state recurrence ECDFs. On the other hand, the IS-MRP mixture model is prone to overfitting transition cases where data is sparse, yielding low predictive coverage rates and predictive ECDF errors that are not as low as the DS-MRP model. In summary, the flexibility of the mixture model and the ability to leverage dependence across transition cases significantly improves the quality of model fit.

We conclude our analysis by examining the estimated transition risk profiles. Recall that these aid in understanding how the risk of transitioning into a given state varies over time following an event. In the context of earthquake recurrence, these

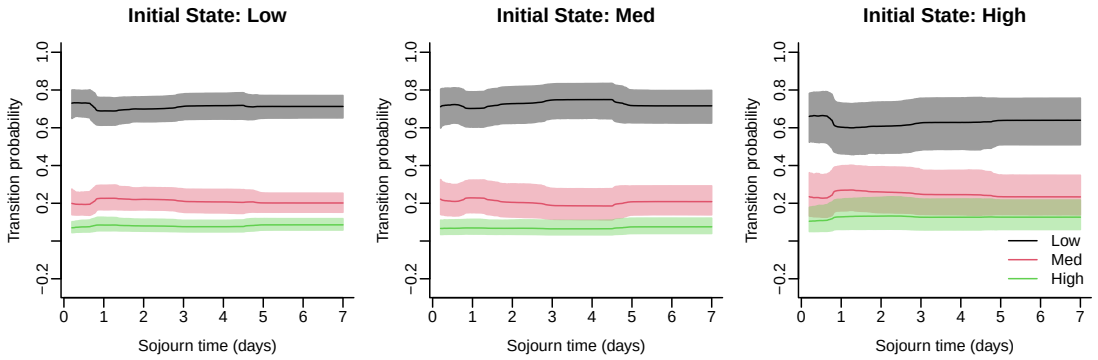


Figure 3.5: Posterior mean estimates and 95% uncertainty intervals for the transition risk profiles for the Southern California earthquake dataset. Initial states displayed are low, medium, and high, from left to right, respectively.

could be used, for example, to study whether large aftershocks are more or less likely immediately following a large earthquake. In Figure 3.5, we show the posterior mean estimates and 95% uncertainty intervals for the transition risk profiles of each initial state. Perhaps a little underwhelming at first glance, the relatively constant transition risks across time are expected in the presence of a dominant $\sigma_\mu^2$ parameter. The implication for the Southern California earthquake data is that the relative risk of a large magnitude earthquake occurring does not change significantly following an event.

## 3.7 Concluding remarks

Our goal in this chapter was to introduce a nonparametric modeling framework for Markov renewal processes that facilitates borrowing of information across transition cases. To achieve this, we combined the mixture modeling framework from Chapter 2 with a dependent prior structure, enabling organic discovery of dependence relationships between sojourn time distributions. Through simulation study and a real data applica-

tion, The proposed model demonstrated improved performance over alternatives. We focused our efforts on the uniform mixture model, consistent with the application area of earthquake recurrence modeling. However, posterior simulation for the MRP kernel parameters is very similar to that of the corresponding HRP parameters. Details for the gamma mixture model follow similarly.

We conclude by discussing a possible improvement to the state recurrence checks described in Section 3.4.2. For computational reasons, we opted for a predictive approach rather than directly computing the state recurrence densities. The ECDF Errors and plots are useful, but they are limited by the increased uncertainty inherent in predictive checks. If the computational burden can be overcome, the state recurrence densities could be used to construct time-rescaling QQ plots, which would be able to test both the density shape and the point process structure of the state recurrence times.

# Chapter 4

# The Log-logistic Basis Model for Renewal Process Hazards

The majority of methods developed for homogeneous renewal processes focus on the inter-arrival time density. This density-oriented perspective provides a tractable likelihood and facilitates many well-known estimation techniques, which was critical to our developments in Chapters 2 and 3. At the same time, an HRP can equivalently be expressed through the inter-arrival hazard function. The hazard function plays a key role in defining the HRP conditional intensity and is frequently of interest in applications adjacent to survival analysis, including earthquake recurrence modeling (Ogata, 1999; Rotondi, 2021). In these cases, prior information is often more naturally expressed on the hazard scale, which may be difficult to incorporate into a density-oriented model, particularly for nonparametric methods (such as in Xiao et al., 2021).

The fundamental objective of this chapter is to develop a flexible and com-

putationally efficient model for the inter-arrival hazard of a renewal process. We will introduce a novel basis representation for hazard functions composed of log-logistic hazards, and develop a structured Bayesian nonparametric prior for the basis coefficients. Prior information can be incorporated through the basis coefficients, and posterior simulation for the resulting hierarchical model can be implemented using latent variable augmentation and standard MCMC methods. Our model also provides a more tractable foundation for developing certain renewal process extensions that are focused on the hazard scale. As a key example, we develop a nonparametric prior probability model for modulated renewal processes, a time-varying extension of the homogeneous case.

## 4.1 Methodology for HRP hazards

### 4.1.1 Motivation and background

Although unused in the renewal process literature, Bayesian nonparametric methods have been applied to hazard estimation in the broader context of survival analysis. Neutral to the right processes, put forth by Doksum (1974), are a notable class of priors for cumulative hazards, which include the beta process (Hjort, 1990) and the gamma process (Ferguson and Phadia, 1979) as special cases. Direct hazard estimation became available with the development of corresponding kernel mixtures, most significantly the extended gamma process mixture (Dykstra and Laud, 1981; Lo and Weng, 1989; Ishwaran and James, 2004). Multivariate extensions with dependence of these methods have also been developed (Lijoi and Nipoti, 2014; Camerlenghi

et al., 2021). Posterior sampling methods for these models largely rely on variations of Blackwell-MacQueen Pólya urn schemes that marginalize over the random measures. These facilitate standard MCMC algorithms but at the cost of limiting hazard function inference to point estimates. Attempts to obtain posterior hazard credible intervals have been made, though they face challenges. An approximation based on random measure trajectory simulation has been explored, but this method only applies to certain model structures and is challenging to implement (Nieto-Barajas and Walker, 2004; James, 2005; Nieto-Barajas, 2014). A moment-based approximation technique is developed in Arbel et al. (2016) using Jacobi polynomials and importance sampling, yet closed-form expressions are available only for an exponential base measure and may suffer from numerical instability for higher-order moments.

In pursuit of a more tractable and extensible hazard modeling framework, we propose an alternative approach rooted in basis representation models. Such models represent a target function using a weighted combination of basis functions, denoted by,

$$\sum_{\ell=1}^{L} \omega_\ell h_{\mathcal{B}}(x|b_\ell, \theta)\,, \tag{4.1}$$

where $L$ is the number of basis functions, $\omega_\ell$ is the combination weight for component $\ell$, and $h_{\mathcal{B}}(x|b_\ell, \theta)$ denotes the basis function evaluated at $x$, with common parameters represented by $\theta$ and basis-specific fixed values (such as knot locations) given by $b_\ell$. Other parameters that are common to all basis functions are represented by $\theta$. This confines the target functional to the space spanned by the basis functions, often inducing smoothness as a result (e.g., smoothing splines, de Boor, 1978), and shifts the bulk of

modeling effort to the combination weights.

The Bernstein polynomial model is a well-known basis representation model that has been used for both density and NHPP intensity estimation (Petrone, 1999a,b; Zhao and Kottas, 2021). It utilizes Bernstein polynomials as basis functions, which are essentially beta densities with a specific sequence of integer-valued shape parameters. As the number of basis functions $L$ approaches infinity, the Bernstein polynomial model converges uniformly to any continuous target function over a bounded interval. The Erlang mixture model is another powerful basis system using Erlang densities, i.e., gamma densities with integer shape parameters, as basis functions. The Erlang mixture has one free parameter in the basis definition, the global scale parameter $\theta$. The Erlang density mixture converges point-wise to any continuous target density over the positive real line as $\theta \to 0$ and $L \to \infty$ (Lee and Lin, 2010). Erlang mixtures have been used for both renewal process densities (Xiao et al., 2021) and NHPP intensities (Kim and Kottas, 2022).

The Bernstein polynomial and Erlang mixture models have been used mainly for density estimation, and they suggest three key properties that we seek in a hazard basis system. First, the basis functions must be hazard functions themselves, ensuring that a weighted combination will also be a valid hazard function. Second, the basis function expressions should be available in closed form to facilitate posterior sampling. Finally, the basis system must have some reassurance of flexibility, ideally with a convergence result similar to either the Bernstein polynomial or Erlang mixture models.

### 4.1.2 Log-logistic hazard basis system

Having considered many candidates, we now propose a basis system composed of log-logistic hazard functions. This system satisfies all three desired properties, including a point-wise convergence result that we will present shortly. Let $h_{\ell\ell}(x|\alpha,\beta) = (\beta/x)\{1+(x/\alpha)^{-\beta}\}^{-1}$ denote the log-logistic hazard function. We define the log-logistic basis function as,

$$h_{\mathcal{B}}(x|b_\ell,\theta) = \frac{\theta}{b_\ell} h_{\ell\ell}(x|b_\ell, b_\ell/\theta) = \frac{1}{x}\{1+(x/b_\ell)^{-b_\ell/\theta}\}^{-1}\,, \tag{4.2}$$

where $\theta$ is a global dispersion parameter and $b_\ell$ is a known knot location specific to each basis function with $b_1 < b_2 < \cdots < b_L$. Figure 4.1 shows log-logistic basis functions evaluated over the interval $(0,1]$ with an evenly spaced grid of knots and a variety of $\theta$ values.

This basis draws inspiration from Erlang mixtures, though the role of $\theta$ is notably different. For Erlang mixtures, the parameter $\theta$ controls both the location and scale of the basis functions and, in conjunction with $L$, the effective range that can be modeled (see Xiao et al. 2021 for more discussion). In contrast, the parameter $\theta$ in our basis only affects the dispersion of each function, the general location being tied to the knots $b_\ell$ and the effective range being controlled by the largest knot $R = b_L$. In addition, the value of $\theta$ influences whether a given basis function is unimodal or strictly decreasing in shape, which occurs whenever $b_\ell \leq \theta$. Inappropriately extreme values of $\theta$ are detrimental to model flexibility. The left-most panels in Figure 4.1 show that overly large $\theta$ leads to restrictive shapes over the effective range, while the right-most panels

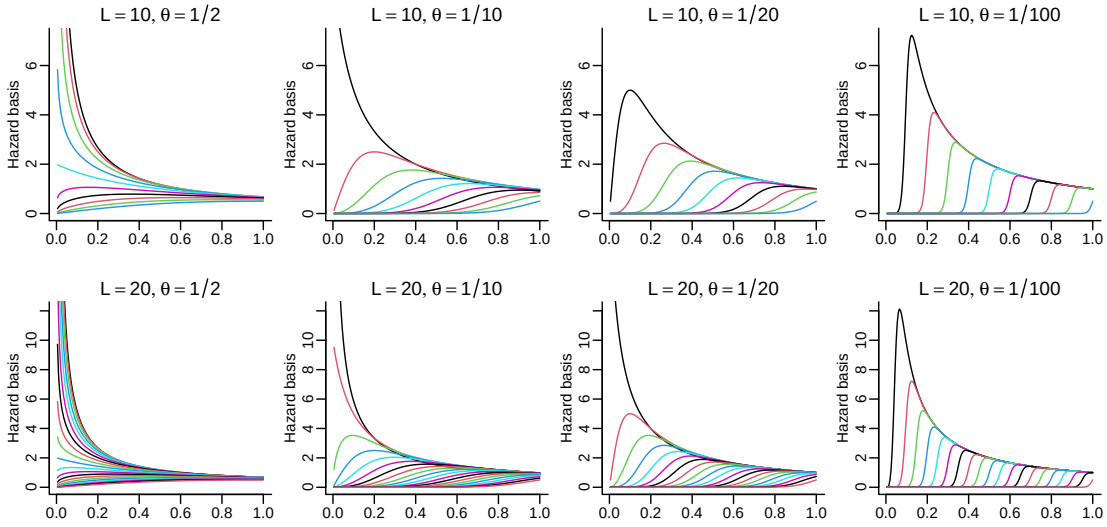


Figure 4.1: Log-logistic basis functions with an evenly spaced grid of knot locations $b_\ell$ over the interval $(0, 1]$. Top row shows evaluations with $L = 10$ basis functions and bottom row shows evaluations with $L = 20$. Left-to-right shows the parameter $\theta$ taking values of $1/2$, $1/10$, $1/20$, and $1/100$.

highlight that small $\theta$ yields functions whose contributions to the broad mixture are too granular. The Erlang mixture structure suggests that having one strictly decreasing basis function is ideal, which the log-logistic basis system achieves whenever $b_1 \leq \theta < b_2$. For simplicity and computational manageability, we fix $\theta = b_1$, which for an evenly spaced grid of $L$ knots over $(0, R]$ is $b_1 = R/L$. Estimating $\theta$ confined to the interval $[b_1, b_2)$ is possible, but consider that smaller values of $\theta$, i.e. closer to $b_1$, offer more local flexibility, and that the empirical effects of varying $\theta$ in such a small range tend to be negligible anyway.

Another key similarity between the log-logistic hazard basis and the Erlang mixture is the presence of a supporting convergence result. These results elevate their respective models and add credibility to their use in place of Bayesian nonparametric

kernel mixtures that have stronger theoretical support but may suffer from implementation challenges. The log-logistic basis convergence result is given in Theorem 4.1.

**Theorem 4.1.** *Consider a continuous, differentiable function $g(x)$ defined on $x \in \mathbb{R}^+$ such that $\lim_{x\to 0^+} g(x)$ is finite. Let $\phi(x) = \frac{d}{dx}(xg(x))$ be called the weight function and let $h_{\mathcal{B}}(x|b,\theta)$ be the log-logistic basis function as defined above. Choose any finite $R > 0$ and let $\{b_\ell = \ell R/L : \ell = 1,\ldots,L\}$ be a corresponding set of evenly spaced knot points. Then, for $x \in (0,R)$, the following point-wise convergence result holds:*

$$g(x) = \lim_{\theta\to 0^+} \lim_{L\to\infty} \sum_{\ell=1}^{L} \frac{R}{L}\phi(b_\ell)h_{\mathcal{B}}(x|b_\ell,\theta)\,.$$

Similar to the Erlang mixture and Bernstein polynomial models, Theorem 4.1 confirms that the log-logistic hazard basis is sufficiently flexible to model a wide range of functions. Indeed, the scope of convergence is all continuous and differentiable functions $g$ on $\mathbb{R}^+$, not just hazard functions. An important takeaway from this result is the unique mapping of functions $g$ to weight functions $\phi$, indicating from a practical standpoint that a prior model on the weight function induces a prior model on $g$. In Figure 4.2 we show examples of weight functions $\phi$ corresponding to Weibull hazards with different shape parameters. We note that the hazard function $h$ shown in the figure is more structured than the general $g$ from Theorem 4.1.

In order to leverage Theorem 4.1 for hazard estimation, restrictions on $\phi$ must be formulated to reduce the space of possible functions $g$ to valid hazards. A particularly useful sufficient condition is to constrain $\phi(x) \geq 0$ for all $x$, or to equivalently require $xg(x)$ be non-decreasing. This condition is not strictly necessary – there exist

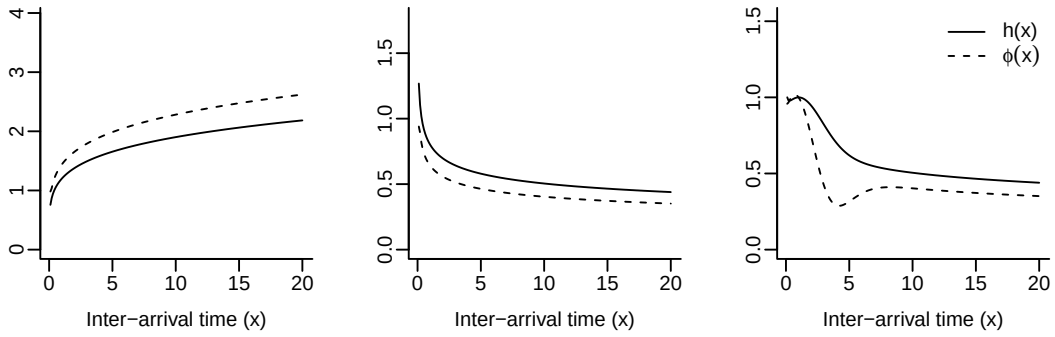


Figure 4.2: Example weight functions $\phi$ and their corresponding hazard functions $h$. Distributions shown from left to right are Wei$(1.2, 1)$, Wei$(0.8, 1)$, and an evenly weighted mixture of the two.

valid hazard functions such that $\phi(x) < 0$ for some $x$ – but it is both easy to accommodate in model development and advantageous for renewal processes. Regarding the latter, recall that renewal process inter-arrival distributions must have a finite mean. If a hazard function $h$ produces asymptotically non-increasing behavior in $xh(x)$, then $h(x)$ must decay at a rate at least as fast as $1/x$, corresponding to a survival function $S(x)$ that decays at least as slow as $1/x$, which therefore must have an infinite first moment. By restricting the weight function $\phi$ to be non-negative we remove many heavy-tailed distributions from the modeling space, though not all. Admittedly, the constraint also removes hazards where $xh(x)$ decreases temporarily but otherwise increases asymptotically. In testing a variety of shapes, we find that this mainly affects hazards with a sharp decline in the tail or with complex, multi-modal behavior, both of which are rare in practice. In fact, we found no examples of common hazard functions with finite mean that violate the constraint, including all parameter combinations of the Weibull, log-normal, gamma, and Lomax distribution families. Considering the immense modeling benefits gained, the constraint on $\phi$ arguably justifies the relatively

minor and infrequent trade-offs.

The proof of Theorem 4.1 offers insight into both the exact form of the basis functions and the role of the knot locations. It follows from two key supporting lemmas, the first of which, Lemma 4.1, establishes an integral representation for the function $g$.

**Lemma 4.1.** *Consider functions $g(x)$ and $\phi(x)$ as in Theorem 4.1. Let $\mathbb{I}(\cdot)$ denote the indicator function which equals 1 if the argument is true and 0 otherwise. Then, for $R \leq \infty$, $g$ can be represented by the following integral:*

$$g(x) = \int_0^R \frac{1}{x}\mathbb{I}(u < x)\phi(u)du\,.$$

*Proof.* Let $\Phi(x) = xg(x)$ be an antiderivative of $\phi$. The restrictions on $g$ imply that $\Phi$ is differentiable and that $\lim_{x\to 0^+} \Phi(x) = 0$. Then,

$$g(x) = \frac{1}{x}\Phi(x) = \frac{1}{x}\int_0^x \phi(u)du = \int_0^R \frac{1}{x}\mathbb{I}(u < x)\phi(u)du\,,$$

where $x < R \leq \infty$ denotes a possibly infinite upper bound. □

The integral representation in this lemma is a reasonable starting point for constructing a basis model, but the corresponding basis functions $x^{-1}\mathbb{I}(u < x)$ are discontinuous at $u = x$. For a finite weighted combination of basis functions, the discontinuity is not ideal. Lemma 4.2 addresses this issue by introducing the log-logistic basis function.

**Lemma 4.2.** *Let $h_{\mathcal{B}}(x|b_\ell, \theta)$ be the log-logistic basis function as defined in Theorem 4.1. Then, for any $b_\ell > 0$ and $b_\ell \neq x$,*

$$\lim_{\theta\to 0^+} h_{\mathcal{B}}(x|b_\ell, \theta) = \frac{1}{x}\mathbb{I}(b_\ell < x)\,.$$

*Proof.* Consider first the case that $x < b_\ell$ and thus $b_\ell/x > 1$. The limit of $(b_\ell/x)^{b_\ell/\theta}$ as $\theta$ approaches 0 grows without bound and therefore $h_{\mathcal{B}}(x|b_\ell, \theta)$ approaches 0. Now suppose that $x > b_\ell$ such that $b_\ell/x < 1$. Then the term $(b_\ell/x)^{b_\ell/\theta}$ approaches 0 as $\theta$ approaches 0 and thus $h_{\mathcal{B}}(x|b_\ell, \theta)$ approaches $1/x$. □

To summarize, the log-logistic basis functions approach the discontinuous basis functions from Lemma 4.1 as the dispersion parameter $\theta$ approaches zero, hinting at their potential as a replacement in a finite mixture model. We note that the case of $x = b$ is excluded from the lemma due to the limit approaching neither zero nor $1/x$, however this discrepancy is a removable discontinuity and does not affect the proof of Theorem 4.1. A rigorous treatment of this issue might involve formally defining a hole in the basis function at $x = b$, but it is sufficient for our purposes to note its immaterial impact and proceed with the limit as if it holds for all $x$. The proof of Theorem 4.1 is completed by combining the results from Lemmas 4.1 and 4.2.

*Proof of Theorem 4.1.* Plugging the result from Lemma 4.2 into Lemma 4.1 yields,

$$g(x) = \int_0^R \phi(u) \lim_{\theta \to 0^+} h_{\mathcal{B}}(x|u, \theta) du \,.$$

Notice that the basis function $h_{\mathcal{B}}(x|u, \theta)$ is bounded above by $1/x$ for all $u$ and $\theta$. For fixed $x$, or with respect to $u$, this bound is constant, thus for *finite* $R$ the dominated convergence theorem applies and the limit can be moved outside the integral:

$$g(x) = \lim_{\theta \to 0^+} \int_0^R \phi(u) h_{\mathcal{B}}(x|u, \theta) du \,.$$

We note that $1/x$ is unbounded as $x$ decreases, indicating that convergence is pointwise and not uniform with respect to $x$. Now observe that $\phi(u)$ and $h_{\mathcal{B}}(x|u,\theta)$ are both continuous and bounded over $u \in (0, R)$, and therefore their product is continuous and bounded over the same interval. Hence, the integrand above is Riemann integrable over $(0, R)$. Using the set of evenly spaced knot points $\{b_\ell\}$, we write the integral as the limit of a right Riemann sum:

$$g(x) = \lim_{\theta \to 0^+} \lim_{L \to \infty} \sum_{\ell=1}^{L} \frac{R}{L} \phi(b_\ell) h_{\mathcal{B}}(x|b_\ell, \theta)\,,$$

which completes the proof of Theorem 4.1. □

The proof of Theorem 4.1 helps to validate the log-logistic basis system as a suitable choice for hazard estimation, both providing a foundation for model development and satisfying an expectation of flexibility. It also illustrates the roles played by the log-logistic basis functions, the dispersion parameter $\theta$, and the knot locations $b_\ell$. On the basis of our previous discussion, going forward we fix $\theta = R/L$. Regarding the knot locations, we default to evenly spaced knots $b_\ell = \ell R/L$ for some finite $R > 0$. This choice is motivated by the Riemann sum conditions required to prove Theorem 4.1 and is simple to implement and interpret. In addition, evenly spaced knots are a standard choice for basis models like the Erlang mixture model and B-splines. However, any knot location scheme that satisfies the partitioning requirements of the Riemann sum can be used should the application demand it.

### 4.1.3 Model formulation

Our proposed approach to renewal process hazard estimation is to build a basis representation model, as in Equation (4.1), using the log-logistic basis functions defined in Equation (4.2). We proceed with $L$ basis functions placed at evenly spaced knot locations $\{b_\ell = \ell R/L \ : \ell = 1, \ldots, L\}$ over the interval $(0, R]$. For selecting $R$, we suggest a value close to $R = \max\{x_i\}$ since hazard estimation beyond the data range tends to be unreliable without additional structure. We default to the 99.5% quantile of the data range. As discussed previously, we fix the dispersion parameter $\theta = R/L$ and restrict the weight function $\phi$ to be non-negative.

The model likelihood is given by,

$$\exp\left(-\sum_{\ell=1}^{L}\omega_\ell H_{\mathcal{B}}(T-t_n|b_\ell,\theta)-\sum_{i=1}^{n}\sum_{\ell=1}^{L}\omega_\ell H_{\mathcal{B}}(x_i|b_\ell,\theta)\right)\prod_{i=1}^{n}\sum_{\ell=1}^{L}\omega_\ell h_{\mathcal{B}}(x_i|b_\ell,\theta)\,,$$

where $H_{\mathcal{B}}(x|b_\ell,\theta) = (\theta/b_\ell)\log\{1+(x/b_\ell)^{(b_\ell/\theta)}\}$ is the cumulative basis function and the combination weights are given by $\omega_\ell$. For notational brevity, we define $H^*_\ell = H_{\mathcal{B}}(T-t_n|b_\ell,\theta)+\sum_{i=1}^{n}H_{\mathcal{B}}(x_i|b_\ell,\theta)$ such that the likelihood can be simplified to,

$$\exp\left(-\sum_{\ell=1}^{L}\omega_\ell H^*_\ell\right)\prod_{i=1}^{n}\sum_{\ell=1}^{L}\omega_\ell h_{\mathcal{B}}(x_i|b_\ell,\theta)\,. \tag{4.3}$$

Note that the censoring information term from the general renewal process likelihood expression is incorporated into each $H^*_\ell$. Using aggregate data values to handle censoring is quite convenient and is another advantage of basis models for renewal process hazards over kernel mixtures, which often resort to more involved computational inference techniques.

With both the basis functions and likelihood defined, we focus our efforts on developing a suitable model for the combination weights $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_L)$. Theorem 4.1 endorses treating the weights as realizations of some function $\phi$ at the knot locations, i.e., $\omega_\ell = \phi(b_\ell)$. We propose modeling $\phi$ with the following log-Gaussian process prior:

$$\begin{aligned} \log \phi(x) | \beta_0, \beta_1, \sigma^2, \tau &\sim \mathcal{GP}\left(\mu(x), \kappa(x, x')\right) \\ \mu(x) = \beta_0 + \beta_1 \log(x); \quad \kappa(x, x') &= \frac{\sigma^2 \tau}{2} \exp\left(\frac{-|x - x'|}{\tau}\right), \end{aligned} \tag{4.4}$$

where $\mathcal{GP}$ denotes a Gaussian process, $\mu$ is the mean function, $\kappa$ is the covariance kernel, and $\beta_0$, $\beta_1$, $\sigma^2$, and $\tau$ are hyperparameters. The log-Gaussian process is both flexible and non-negative, making it a natural choice for $\phi$. In terms of the vector of mixture weights $\boldsymbol{\omega}$, this induces a multivariate log-normal prior:

$$\log(\boldsymbol{\omega}) | \beta_0, \beta_1, \sigma^2, \tau \sim \mathrm{N}_L(B\boldsymbol{\beta}, \Sigma), \tag{4.5}$$

where $\mathrm{N}_L$ denotes a multivariate normal distribution in $L$ dimensions, $\boldsymbol{\beta} = (\beta_0, \beta_1)$ is the mean parameter vector, the $L \times 2$ matrix $B$ contains a column of ones and a column of $\log(b_\ell)$ values, and $\Sigma$ is the $L \times L$ covariance matrix formed by plugging in pairs of knot locations into the covariance kernel $\kappa$. The mean function we use is motivated by the Weibull hazard, given by $\alpha\lambda x^{\alpha-1}$, which has a log-logistic basis weight function of $\log \phi(x) = \log(\alpha^2\lambda) + (\alpha - 1)\log(x)$. In other words, the nonparametric prior model is centered on Weibull distributed inter-arrival times. The Weibull connection also implies the constraint $\beta_1 > -1$, which we adopt given the role that $\beta_1$ plays in governing tail behavior.

To get a better sense of the model structure, we generate a set of weights from

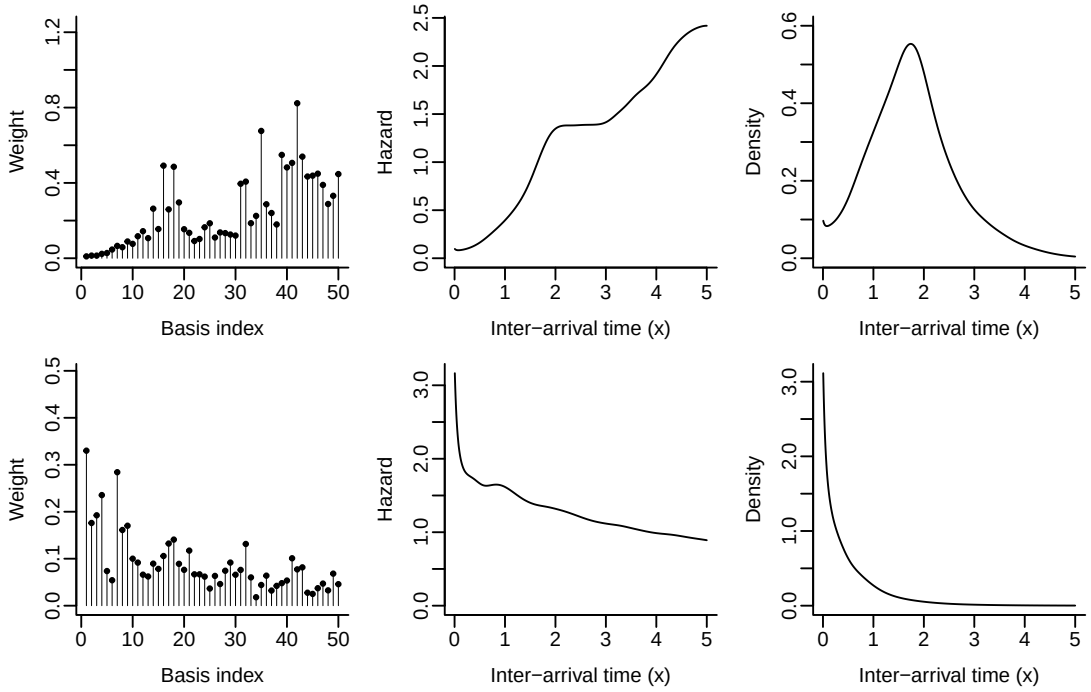


Figure 4.3: Demonstration of the log-logistic hazard basis model prior: generated weight values and corresponding hazard and density functions. Hyperparameters used for the top row are $L = 50$, $R = 5$, $\beta_0 = -2.3$, $\beta_1 = 1$, $\sigma^2 = 4$, and $\tau = 0.1$. Hyperparameters used for the bottom row are the same except $\beta_1 = -0.5$.

the prior using $L = 50$, $R = 5$, $\beta_0 = -2.3$, $\beta_1 = 1$, $\sigma^2 = 4$, and $\tau = 0.1$. Figure 4.3 shows the resulting weight values and corresponding hazard and density functions in the top row, and another simulation with $\beta_1 = -0.5$ in the bottom row. This highlights the role that $\beta_1$ plays in setting the slope of the weights and therefore the tail behavior of the resulting hazard and density. Similar plots can be generated for the other parameters, showing their effects on the weight functions. In summary, $\sigma^2$ controls the individual weight variability, $\tau$ governs the degree of similarity between neighboring weights, and $\beta_0$ influences the scale of the weights and the scale of the hazard.

The covariance $\kappa$ is the Ornstein-Uhlenbeck (OU) covariance kernel with a

marginal variance parameter $\sigma^2$ and length scale $\tau$. At first glance, a covariance kernel that produces differentiable sample paths may seem more appropriate for $\phi$ given the implied properties of $g$ from Theorem 4.1. However, our initial efforts reveal that such kernels lead to sampling convergence issues. This is in part due to the log-logistic basis functions being correlated in the sense that, for any two components $\ell$ and $\ell'$, the basis function evaluations at $x > \max(b_\ell, b_{\ell'})$ are effectively identical. Put differently, our proposed basis functions have non-local effects to the right of their knot locations, implying some degree of negative conditional correlation. The OU process is a continuous-time extension of a first-order autoregressive process and supports negative conditional covariance in the induced off-diagonal elements of $\Sigma^{-1}$. Compared to the positive values generated by radial kernels, negative conditional covariance is better aligned with the basis function behavior and matches empirical observations from posterior sampling.

For hyperparameter priors, we use conditionally conjugate normal distributions for $\beta_0$ and $\beta_1$ with means $\log(R/L)$ and 0, respectively, and a large shared variance $s_\beta = 1000^2$. Centering $\beta_0$ on $\log(R/L)$ mirrors the scaling of $\phi$ by $R/L$ in Theorem 4.1. Also note the distribution for $\beta_1$ is truncated below such that $\beta_1 > -1$. We assign a conditionally conjugate inverse gamma prior to $\sigma^2$, with shape $a_\sigma = 2.5$ to ensure a finite prior variance, and a scale parameter of $b_\sigma = 1$. Finally, we use an exponential distribution as a prior model for $\tau$ centered on $R/L$, which is the distance between adjacent knot locations. Performing sensitivity analysis on these hyperparameters highlights that inferences are fairly robust to changes in prior parameter values, which we discuss further in Section 4.1.5.

We conclude this section by considering the finite mean requirement for renewal processes. A well known result in survival analysis is that a positively valued random variable $X$ with survival function $S(x)$ has expectation $\mathbb{E}[X] = \int_0^\infty S(x)dx$. If $S(x)$ decays at a rate faster than $1/x$, or rather $S(x)$ is asymptotically bounded above by $a/x^c$ for some $a > 0$ and $c > 1$, then $\mathbb{E}[X]$ is finite. In the context of the log-logistic basis model, this implies that the finite mean restriction is satisfied as long as the total weight $\Omega = \sum_{\ell=1}^{L} \omega_\ell > 1$. Explicitly enforcing this constraint in the prior probability model is challenging. However, several structural choices in developing this model (such as strictly positive weights and the prior constraint on $\beta_1$) help alleviate the issue without direct intervention. Indeed, using a wide range of hyperparameter values, prior simulations suggest that the probability of $\Omega \leq 1$ is negligible. In practice, we find that posterior samples of $\Omega$ are essentially always larger than 1, thus simply monitoring this quantity is sufficient to ensure the finite mean requirement is satisfied.

### 4.1.4 Posterior simulation

Here, we outline a Gibbs sampling algorithm for the log-logistic basis representation model, given $N(T) = n$ observed events in the time interval $(0, T)$ and realized inter-arrival times $\boldsymbol{x} = \{x_i : i = 1, \ldots, n\}$. Considering that updates for the Gaussian process hyperparameters are standard, the majority of exposition will focus on the combination weights $\boldsymbol{\omega}$ which require more specialized treatment. In short, update steps for $\boldsymbol{\beta}$ and $\sigma^2$ are conjugate and are sampled directly, while $\tau$ is updated using a random walk Metropolis step. Complete details on the full conditional distributions are

provided in Appendix D.

For more concise and general notation, we define the mean log-weight vector $\boldsymbol{\mu} = B\boldsymbol{\beta}$ and the unscaled covariance matrix $V$ such that $\Sigma = \sigma^2 V$. The notation $\boldsymbol{\mu}(\boldsymbol{\beta})$, $V(\tau)$, and $\Sigma(\sigma^2, \tau)$ could to used to explicitly indicate particular hyperparameter dependencies, but we forgo this to reduce clutter. To facilitate updates for $\boldsymbol{\omega}$, we introduce latent membership variables $\boldsymbol{z} = \{z_i : i = 1, \ldots, n\}$, where $z_i = \ell$ indicates that event $i$ is associated with basis function $\ell$, and $\Pr(z_i = \ell | \boldsymbol{\omega}) = \omega_\ell / \Omega$, where $\Omega = \sum_{r=1}^{L} \omega_r$. The augmented model likelihood can be expressed as,

$$\exp\left(-\sum_{\ell=1}^{L} \omega_\ell H_\ell^*\right) \prod_{i=1}^{n} \Omega h_{\mathcal{B}}(x_i | b_{z_i}, \theta)\,. \tag{4.6}$$

To draw from the joint full conditional of $\boldsymbol{\omega}$ and $\boldsymbol{z}$, we employ a Gibbs step, first sampling $\boldsymbol{z}$ given $\boldsymbol{\omega}$ from a categorical distribution, then sampling $\boldsymbol{\omega}$ given $\boldsymbol{z}$. The conditional distribution of $\boldsymbol{\omega}$ given $\boldsymbol{z}$ is given by,

$$\begin{aligned} p(\boldsymbol{\omega} | \boldsymbol{z}, \boldsymbol{\beta}, \sigma^2, \tau, \boldsymbol{x}) &\propto p(\boldsymbol{\omega} | \boldsymbol{\beta}, \sigma^2, \tau) \times \left\{ \prod_{i=1}^{n} \left( \frac{\omega_{z_i}}{\Omega} \right)^{n_\ell} \right\} \times \Omega^n \exp\left(-\sum_{\ell=1}^{L} \omega_\ell H_\ell^*\right) \\ &\propto p(\boldsymbol{\omega} | \boldsymbol{\beta}, \sigma^2, \tau) \times \prod_{\ell=1}^{L} \mathrm{Ga}(\omega_\ell | n_\ell + 1, (H_\ell^*)^{-1})\,, \end{aligned} \tag{4.7}$$

where $n_\ell = |\{i : z_i = \ell\}|$ is the number of events associated with basis function $\ell$, $\mathrm{Ga}(x | \kappa, \lambda)$ denotes a gamma density evaluated at $x$ with mean given by $\kappa\lambda$. Also recall the induced prior on $\boldsymbol{\omega}$, $p(\boldsymbol{\omega} | \boldsymbol{\beta}, \sigma^2, \tau)$, is multivariate log-normal with log-scale mean vector $\boldsymbol{\mu}$ and covariance matrix $\Sigma$.

The distribution in (4.7) cannot be directly sampled, but the form lends itself to slice sampling. We introduce a set of auxiliary variables $\boldsymbol{u} = \{u_\ell : \ell = 1, \ldots, L\}$ such

that the joint conditional distribution is given by,

$$p(\boldsymbol{\omega}, \boldsymbol{u}|\boldsymbol{\beta}, \sigma^2, \tau, \boldsymbol{x}) \propto p(\boldsymbol{\omega}|\boldsymbol{\beta}, \sigma^2, \tau) \times \prod_{\ell=1}^{L} \mathbb{I}(u_\ell < \text{Ga}(\omega_\ell|n_\ell + 1, (H_\ell^*)^{-1})) \,. \tag{4.8}$$

We employ a Gibbs step to sample from this joint distribution. Given $\omega_\ell$, the slice sampling variable $u_\ell$ is sampled uniformly between 0 and $\text{Ga}(\omega_\ell|n_\ell + 1, (H_\ell^*)^{-1})$. To draw the weights $\boldsymbol{\omega}$ given $\boldsymbol{u}$, we invert the indicator function inequalities such that they take the form $\mathbb{I}(C_\ell < \omega_\ell < D_\ell)$. Graphically, $C_\ell$ and $D_\ell$ represent the lower and upper locations where the gamma density is equal to $u_\ell$, and they are computed using the Lambert-W function, which we elaborate on this further in Appendix D.

Given $\boldsymbol{u}$ and the corresponding boundary vectors $\boldsymbol{C} = \{C_\ell : \ell = 1, \ldots, L\}$ and $\boldsymbol{D} = \{D_\ell : \ell = 1, \ldots, L\}$, the full conditional distribution of $\boldsymbol{\omega}$ is a multivariate log-normal distribution where each element $\omega_\ell$ is separately truncated to be between $C_\ell$ and $D_\ell$. Sampling such a distribution has historically been a challenge, but recent advances have provided multiple efficient simulation methods. We make use of the harmonic-HMC algorithm of Pakman and Paninski (2014) to draw samples from our truncated log-normal distribution on the log-scale, i.e., drawing from a multivariate normal distribution of length $L$ with mean $\boldsymbol{\mu}$ and covariance $\Sigma$, denoted $N_L(\boldsymbol{\mu}, \Sigma)$, where each element is truncated to lie between $\log(C_\ell)$ and $\log(D_\ell)$. A code implementation is available in the `hdtg` package for `R` (Zhang et al., 2022; R Core Team, 2024). This algorithm is very efficient; we find that sampling proceeds comfortably for $L$ as large as 500, which is more than sufficient in practice.

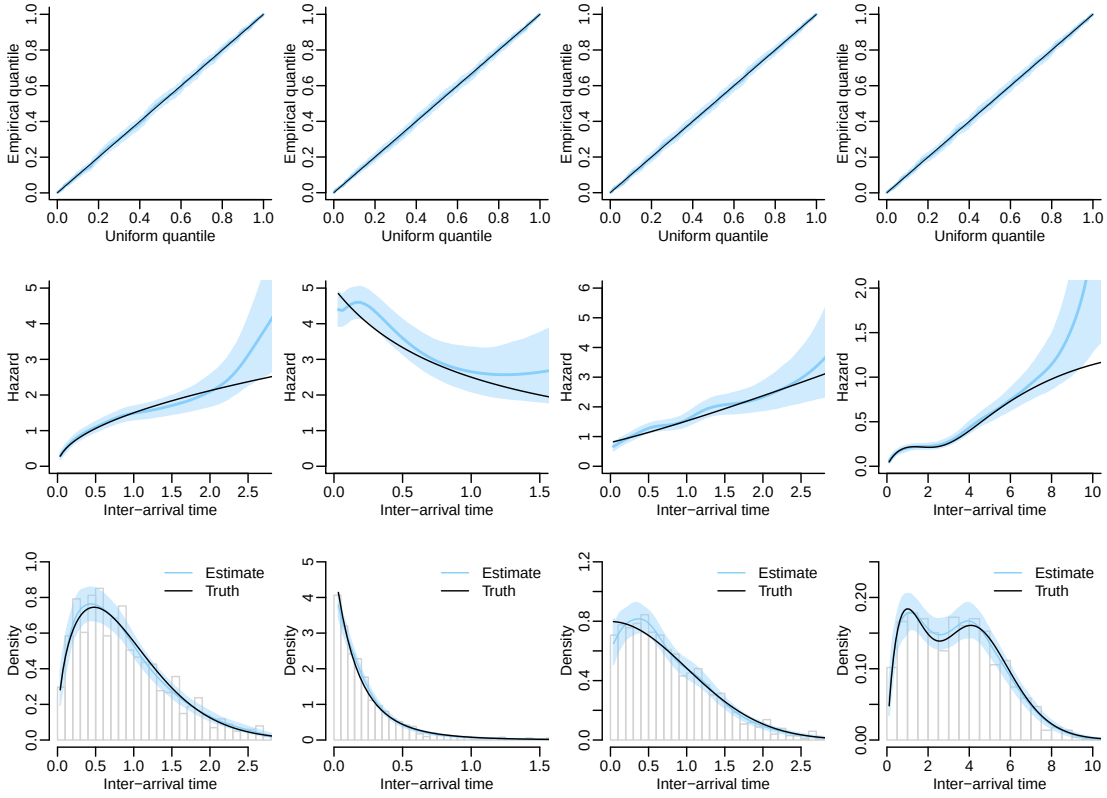


Figure 4.4: Simulated data examples fit using the log-logistic basis model with $L = 50$ components. Generative distribution are, from left to right, Weibull, Lomax, half-normal, and an evenly weighted mixture of two gamma distributions. The top row shows time-rescaling QQ plots, the middle row shows the estimated hazard functions, and the bottom row shows the estimated inter-arrival density functions.

### 4.1.5 Synthetic data examples

In this section we explore properties and performance of the log-logistic basis model applied to simulated data. Our simulations are designed to illustrate the flexibility of the model in capturing a variety of hazard function shapes. We sample HRP point patterns using the same generative procedure as the simulations in Chapter 2.2, which we repeat here for clarity: a Weibull with shape parameter 1.5 and unit scale, a Lomax with shape parameter 5 and unit scale, a standard half-normal, and an evenly weighted

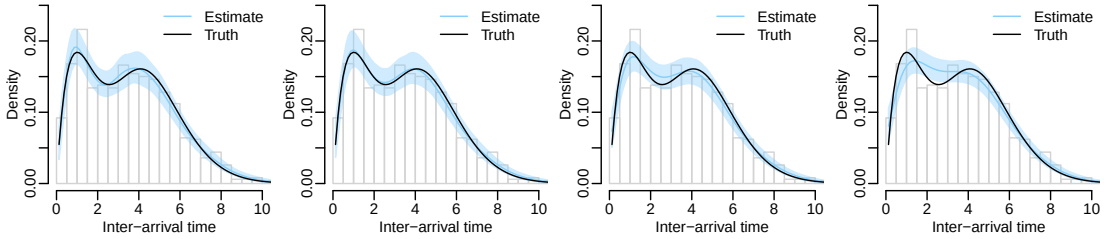


Figure 4.5: Sensitivity analysis for the log-logistic basis model using the gamma mixture generative distribution. Each panel shows density estimates for the model fit with $a_\sigma$ set to 2.5, 10, 50, and 200, left to right respectively, and default values for the other hyperparameters.

mixture of a $\text{Ga}(2, 1)$ and a $Ga(10, 0.5)$. Recall the Lomax distribution has density $(\alpha/\lambda)(1 + x/\lambda)^{-(\alpha+1)}$ where $\alpha$ is the shape parameter and $\lambda$ is the scale. The standard half-normal distribution has density $\sqrt{2/\pi}\exp(-x^2/2)$. For each scenario, we generate inter-arrival times using a value of $T$ such that the expected number of events is roughly 1,000. We fit the model using $L = 50$ basis functions and default hyperparameter values. The MCMC algorithm is run for 5,000 iterations following a burnin period of 15,000. Figure 4.4 shows posterior estimates for the hazard function, the implied inter-arrival time density, and time-rescaling QQ plots for each scenario. The overall conclusion from Figure 4.4 is that the log-logistic basis model is able to effectively capture a variety of hazard function shapes. Posterior intervals for the hazard and density functions capture the generative functions well, and the QQ plots indicate a good fit for all scenarios.

We note that these results are relatively robust to changes in the hyperparameter values, particularly those related to $\beta_0$ and $\beta_1$. To demonstrate the impact of more informative priors on $\sigma^2$, we repeat the simulations using the gamma mixture generative distribution and set $a_\sigma$ to increasingly large values, starting at the default 2.5 and

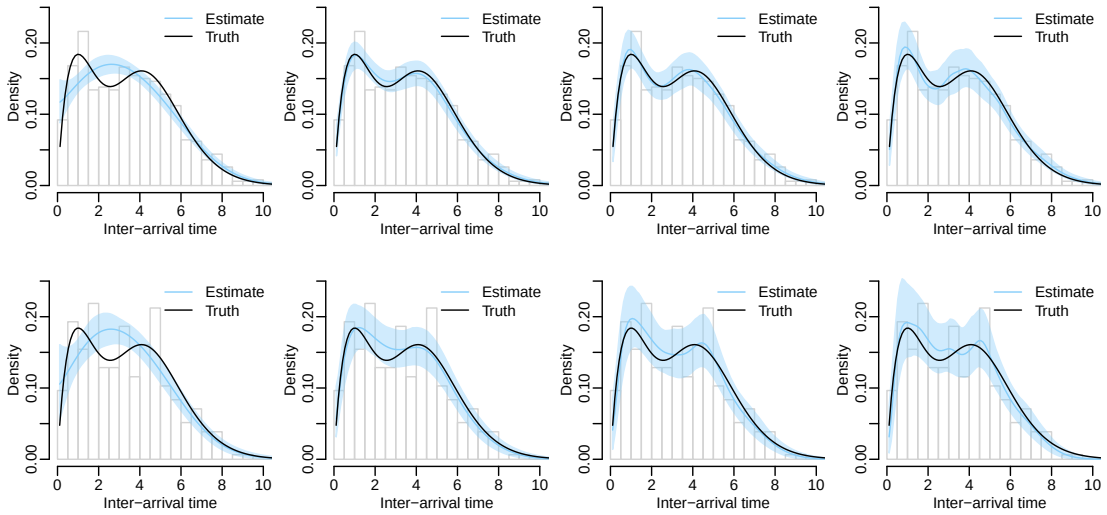


Figure 4.6: Sensitivity analysis for the log-logistic basis model using the gamma mixture generative distribution. Each column shows density estimates for the model fit with $L$ set to 5, 20, 50, and 100, left to right respectively, and default values for the other hyperparameters. The top row shows results for a dataset with approximately 2,000 events, while the bottom row shows results for a dataset with around 200 events.

going up to 200. Figure 4.5 shows the results of this sensitivity analysis, focusing on estimates of the inter-arrival density, which show more clearly where discrepancies are most pronounced. Although the effect is gradual, priors that emphasize large values of $\sigma^2$ lead to more variable weight functions and consequently fewer significant features in the density estimates. In settings where overfitting is a concern, this behavior can be leveraged to produce more conservative estimates.

By far the most impactful hyperparameter to select is the number of basis functions $L$. Larger $L$ allows for more flexible hazard estimates, but also increases the computational burden. We study the impact of $L$ on the model fit by repeating the gamma mixture simulation with $L$ set to 5, 20, 50, and 100. Figure 4.6 displays estimated inter-arrival densities for each value of $L$ in the top row. Additionally, in the

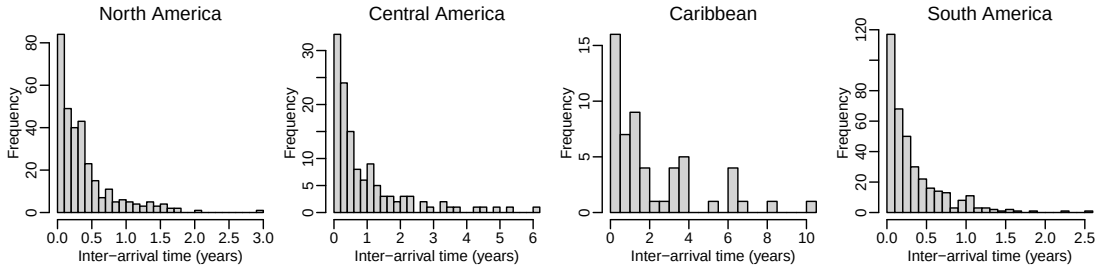


Figure 4.7: Histograms of inter-arrival times for the NOAA Significant Earthquake Database data for four regions: North America, Central America, the Caribbean, and South America.

bottom row we show the same simulation study but where $T$ has been adjusted such that the expected number of events is near 300. Unsurprisingly, when $L$ is too small, the model is unable to capture the complexity of the generative distribution. However, the quality of fit rapidly improves as $L$ increases.

### 4.1.6 Earthquake data application

We now consider applying the log-logistic hazard model to a real earthquake dataset. These data are taken from the *Significant Earthquake Database* published and maintained by NOAA National Centers for Environmental Information (NGDC/WDS, 2021). Earthquakes recorded in this database are rather destructive, satisfying at least one of the following criteria: $1 million in damages, over 10 deaths, magnitude of at least 7.5, Modified Mercalli Intensity X or greater, or the earthquake generated a tsunami. We consider earthquake events occurring in four regions: North America, Central America, the Caribbean, and South America. The data contain events from Jan. 1, 1900 to Dec. 31 2019 and include 313 earthquakes for North America, 126 for Central America, 55 for the Caribbean, and 367 for South America. Figure 4.7 shows histograms of the inter-

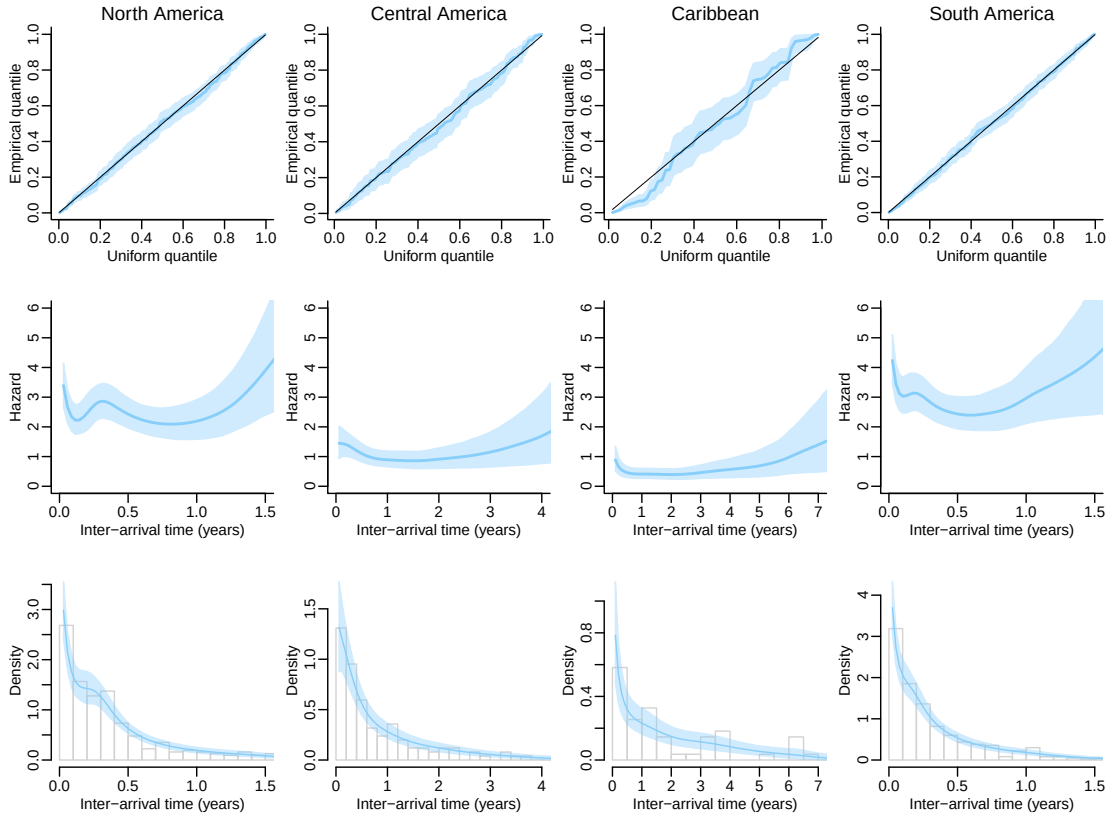


Figure 4.8: Posterior inference for the hazards (middle), density functions (bottom) and time-rescaling QQ plots (top) for the NOAA Significant Earthquake Database data.

arrival times for each region. The longest inter-arrival time in this dataset is 3,719 days in the Caribbean region, occurring between an aftershock of the San Fermín earthquake in Puerto Rico on Nov. 12, 1918 and the Cumaná earthquake off the coast of Venezuela on Jan. 17, 1929.

We fit the model separately to each region, using $L = 50$ basis functions and default hyperparameters mentioned previously. Using the same machine as in previous chapters, the sampling algorithm for the North America region takes approximately 0.5 seconds per 1,000 posterior iterations. Samples for each parameter usually converge

within 5,000 iterations, though we use 15,000 burnin iterations as a more conservative choice. The effective sample size for the inter-arrival density evaluated at various points is roughly between 100 and 120 per 1,000 posterior samples, thus we thin samples by a factor of 10. Figure 4.8 shows the posterior mean of the hazard and density functions for each region, along with time-rescaling QQ plots and 95% posterior uncertainty intervals. The density estimates and QQ plots suggest that the model effectively captures the data distribution in each region, with perhaps the exception of the Caribbean region, where the data are unusually concentrated close to zero. The hazard function magnitudes also clearly show differences in overall rates of large earthquakes between regions, with South and North America being the most active. Also, recall from our discussion in Chapter 2.1.1 that there are mechanistic reasons to expect earthquake inter-arrival densities to be decreasing. Despite no model structure to enforce this behavior, the model produces decreasing densities for these data.

## 4.2 Bayesian nonparametric inference for modulated renewal processes

Many extensions have been proposed to incorporate time-varying behavior into the homogeneous renewal process. A modulated renewal process (ModRP) is formed by scaling the inter-arrival hazard function by a modulating intensity $\lambda(t)$, yielding the ModRP conditional intensity expression:

$$\lambda(t|\mathcal{H}(t)) \equiv \lambda(t)h(t - t_{N(t)}), \tag{4.9}$$

where $\lambda$ and $h$ are required both to be non-negative and locally integrable. The resulting inter-arrival time distribution is no longer stationary, being expressed by a conditional inter-arrival time density:

$$f(x|t_{prev}) = \lambda(t_{prev} + x)h(x) \exp\left(-\int_0^x \lambda(t_{prev} + u)h(u)du\right), \tag{4.10}$$

where $t_{prev}$ is the time of the most recent event.

Modulated renewal processes have been used in several application areas, including earthquake modeling (Lin and Fine, 2009), neural spike train analysis (Kass and Ventura, 2001; Liu and Lengyel, 2023), medical event data (Dabrowska and Ho, 2006), and system reliability (Lawless and Thiagarajah, 1996; Kobbacy et al., 1998). They were first introduced by Cox (1972), stemming from similar work on proportional hazard models. Subsequent efforts focus on asymptotic properties and improving on the partial likelihood estimation methods used originally (Berman, 1981; Oakes and Cui, 1994; Lawless and Thiagarajah, 1996). An additive form of the conditional intensity has also been studied (Lin and Ying, 1994), though the multiplicative form is more common. Lin and Fine (2009) present an estimation method based on martingale equations that is applicable to both the additive and multiplicative forms. Rao and Teh (2011) propose a Bayesian semiparametric method where the modulating function is modeled by a transformed Gaussian process and the hazard component follows a gamma distribution. The associated posterior sampling algorithm is cleverly designed around uniformization, the thinning approach used to simulate point process data (Daley and Vere-Jones, 2003), and leverages the Gaussian process structure.

In this chapter, we take an alternative approach to ModRP modeling. We develop a fully nonparametric model for the ModRP conditional intensity, by combining the log-logistic basis system introduced in the previous section for the hazard, with a structured basis mixture for the modulating intensity. We aim to provide a more flexible hazard model than the gamma assumption used by Rao and Teh (2011) while retaining similar flexibility in the modulating intensity. In addition, by directly modeling both components using basis mixtures, a posterior sampling algorithm is more straightforward to develop and can be implemented using standard MCMC methods.

### 4.2.1 ModRP properties

We follow the nomenclature of Rao and Teh (2011) in describing the components of a modulated renewal process: the function $h$ is referred to as the hazard function, and $\lambda(t)$ is called the modulating intensity or modulating function. When $h(x) = 1$, the model reduces to an NHPP with intensity $\lambda(t)$, whereas an HRP with inter-arrival hazard $h(x)$ emerges when $\lambda(t) = 1$, hence the terminology. However, there is an interplay between $h$ and $\lambda$ that obfuscates their roles outside these special cases. In this section we explore that interplay and the way it manifests in simulated datasets. This will build intuition for the model structure and provide a foundation for the development of our nonparametric ModRP model.

To begin, we simulate data with constant hazard and modulating intensity functions. Figure 4.9 shows the generative functions under three different scenarios as well as simulated event times for each. The density being shown is the exponential

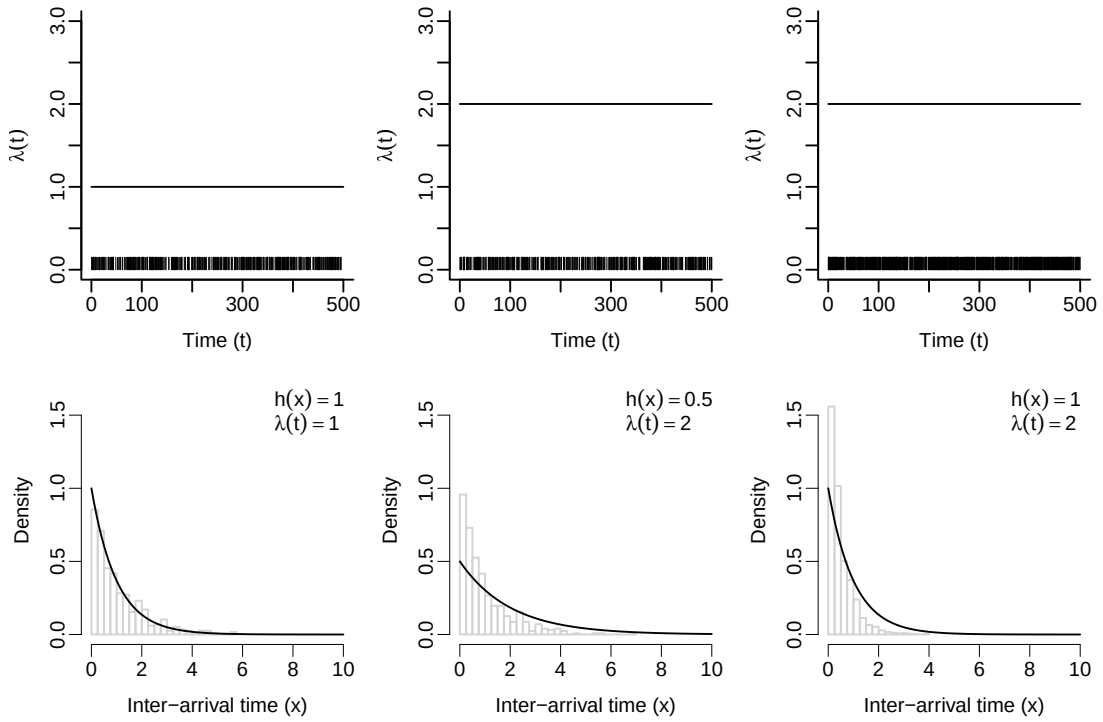


Figure 4.9: Simulated data from a modulated renewal process with constant hazard and modulating intensity functions and upper time bound $T = 500$. The top panels show the modulating intensities with small ticks showing the observed event times. The bottom panels show the exponential density corresponding to the constant hazard function with a histogram of the observed inter-arrival times. The interplay between the hazard and modulating intensity is evident from the inconsistency between the observed inter-arrival times and the exponential density.

density associated with the constant hazard function. The left-most column shows the special case of a point pattern from a homogeneous Poisson process with unit rate, and the implied density from $h$ expectedly matches the observed inter-arrival times. However, $h$ does not match the observations for the middle and right-most columns. The middle column has the same unit conditional intensity as the left-most column and therefore produces a similar point pattern, yet has different values for the constituent functions. This highlights a fundamental consequence of the multiplicative ModRP structure in Equation (4.9): the scales of $h$ and $\lambda$ are not identifiable. We emphasize

Table 4.1: Simulation results for ModRP data with Wei$(2, 1)$ hazard function, Beta$(2, 2)$ modulating density scaled to $(0, T)$, and various values of $\Lambda$ and $T$. The number of data points $n$ and the 90th percentile of the inter-arrival times are shown.

| $T$ | $\Lambda$ | $n$ observations | 90th percentile |
|---|---|---|---|
| 50 | 50 | 59 | 1.41 |
| 50 | 100 | 75 | 1.13 |
| 50 | 200 | 109 | 0.80 |
| 200 | 50 | 115 | 2.76 |
| 200 | 100 | 152 | 2.25 |
| 200 | 200 | 222 | 1.59 |

that this is not a modeling limitation, but a fundamental overspecification of the point process.

Rao and Teh (2011) take a heuristic approach to addressing this issue by fixing the mean of the gamma distribution associated with the hazard function $h$, but this is overly restrictive for our purposes and does not directly target the source of the problem. To elaborate on the scale ambiguity, consider factoring the modulating intensity as $\lambda(t) = \Lambda f_\lambda(t)$, where $\Lambda = \int_0^T \lambda(t)dt$ is the total modulating intensity over the time interval $(0, T)$, and $f_\lambda(t)$ is a density over same interval. Going forward, we refer to $f_\lambda$ as the modulating density. The value of $\Lambda$ is not identifiable with respect to the scale of $h$ and must be fixed in some way. Theoretically, any positive value of $\Lambda$ may be chosen, but there is a corresponding impact on the effective support of the inter-arrival times. To better understand this, we simulate data from a variety of ModRP scenarios with a Wei$(2, 1)$ hazard function $h(x) = 2x$, a Beta$(2, 2)$ modulating density scaled to $(0, T)$, and several values of $\Lambda$ and $T$. Simulated point patterns are drawn using the uniformization algorithm described in Rao and Teh (2011). We examine both the number of data points and the effective support of the inter-arrival times as measured

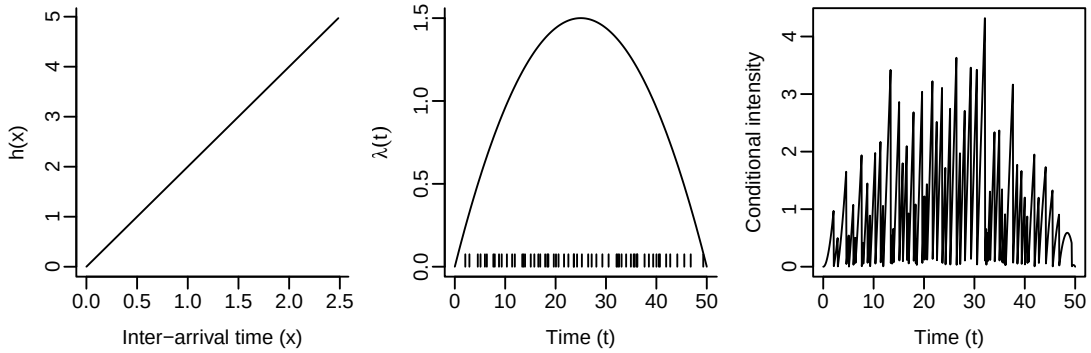


Figure 4.10: Simulated data from a modulated renewal process with Wei$(2, 1)$ hazard function, Beta$(2, 2)$ modulating density scaled to the interval $(0, T)$, total intensity $\Lambda = 50$, and upper time bound $T = 50$. Shown are the hazard function, modulating density, and realized conditional intensity.

by the 90th percentile. The results are shown in Table 4.1, and an example simulation for $\Lambda = 50$ and $T = 50$ is shown in Figure 4.10. Note that, because the inter-arrival times are non-stationary, we plot the realized conditional intensity as a more meaningful data visualization.

Both $T$ and $\Lambda$ have positive effects on the sample size $n$ as expected, but the relationship is no longer linear as it would be for a Poisson process. The effective support is interesting to study in preparation for extending our log-logistic basis model to the ModRP case. Recall from the previous section that the value of $R$, the upper bound of inter-arrival times for our log-logistic basis model, is chosen close to the largest observed inter-arrival time. In contrast to the HRP case, our exploration of ModRP properties shows that the effective support of simulated inter-arrival times is not a direct result from the hazard function $h$, making the choice of $R$ less clear. As a baseline, the 90th percentile of the Wei$(2, 1)$ distribution is approximately 1.52. From Table 4.1, we see that the simulated 90th percentiles are closest to this value when $\Lambda = T$. Indeed, we

find empirically that, under $\Lambda = T$, the effective support of the simulated inter-arrival times roughly aligns with that implied by the hazard $h$ for effectively any choice of $f_\lambda$ and all but the heaviest-tailed choices for $h$. This suggests that, given the choice of $\Lambda = T$, selecting $R$ based on the upper percentiles of the observed inter-arrival times is likely to be effective for the ModRP case. These insights are crucial for choosing appropriate configuration parameters for our nonparametric ModRP model.

### 4.2.2 Model formulation

Our approach to modeling a modulated renewal process is to assign basis mixture models to the constituent functions. For the hazard function $h$, we utilize the log-logistic basis system introduced previously:

$$h(x) = \sum_{\ell=1}^{L} \omega_\ell^h h_{\mathcal{B}}(x|b_\ell, \theta)\,, \tag{4.11}$$

where $L$ is the fixed number of basis functions, $\theta = R/L$ is the global dispersion parameter, and $b_\ell = \ell R/L$ form an evenly spaced grid of knots. The prior model for the weights $\boldsymbol{\omega}^h = \{\omega_\ell^h : \ell = 1, \ldots, L\}$ is identical to the original HRP model, using a log-Gaussian process to describe the weight function with hyperparameters $\beta_0$, $\beta_1$, $\sigma^2$, and $\tau$ (see Section 4.1.3).

Turning to the modulating intensity, we address the scale identifiability issue by factoring $\lambda(t) = \Lambda f_\lambda(t)$ and fixing the total intensity $\Lambda$. Given our discussion on the effective support, we set $\Lambda = T$, and select $R$ based on an upper percentile of the observed inter-arrival times. To model the modulating density $f_\lambda$, we use the Bernstein

polynomial basis system scaled to the interval $(0, T)$:

$$f_\lambda(t) = \sum_{m=1}^{M} \omega_m^f f_{\mathcal{B}}(t|m, M)\,, \tag{4.12}$$

where $M$ is the fixed number of basis functions, and $f_{\mathcal{B}}(t|m, M)$ denotes the Bernstein polynomial basis function evaluated at $t \in (0, T)$. Bernstein polynomial basis functions are defined in terms of beta densities, which we rescale to support the interval $(0, T)$:

$$f_{\mathcal{B}}(t|m, M) = \frac{1}{T}\text{Beta}\left(\frac{t}{T}\middle|m, M - m + 1\right)\,, \tag{4.13}$$

where $\text{Beta}(t|a, b)$ denotes the beta density evaluated at $t$ with shape parameters $a$ and $b$. The weights $\boldsymbol{\omega}^f = \{\omega_1^f, \ldots, \omega_M^f\}$ are given by increments of a random distribution function $G$ with support on $(0, T)$, which is assigned a Dirichlet process prior:

$$\omega_m^f = G(mT/M) - G((m-1)T/M); \quad G|\alpha \sim \mathcal{DP}(\alpha, G_0)\,, \tag{4.14}$$

where $\mathcal{DP}(\alpha, G_0)$ denotes the Dirichlet process with precision parameter $\alpha$ and base distribution $G_0$. By selecting a uniform base distribution for $G_0$ such that $G_0(t) = t/T$ for $t \in (0, T)$, we obtain an implied (symmetric) Dirichlet prior for the vector of weights:

$$\boldsymbol{\omega}^f|\alpha \sim \text{Dir}_M(\alpha/M, \ldots, \alpha/M)\,, \tag{4.15}$$

where $\text{Dir}_M$ denotes an $M$-dimensional Dirichlet distribution. We assign a $\text{Ga}(a_\alpha, b_\alpha)$ prior to $\alpha$ with moderate values for hyperparameters $a_\alpha$ and $b_\alpha$, generally near 1, though our results tend not to be sensitive to this choice.

It should be noted that this model structure can only be used to model a modulating intensity over the bounded interval $(0, T)$—it does not extend to the entire

positive real line. This is a limitation both of the beta densities and the choice to factor $\lambda(t)$ into $\Lambda$ and $f_\lambda(t)$. Restricting model support to the observation window $(0, T)$ is a common practice in NHPP modeling, which is arguably also meaningful in the current context given the similarities between an NHPP intensity and a ModRP modulating intensity.

With priors assigned to both components, the likelihood function for our proposed modulated renewal process model is given by,

$$\begin{aligned}
&\exp\left(-\int_0^T \lambda(t)h(t-t_{N(t)})dt\right) \times \prod_{i=1}^{n} \lambda(t_i)h(x_i) \\
&= \exp\left(-\sum_{m=1}^{M}\sum_{\ell=1}^{L} \omega_m^f \omega_\ell^h \Lambda \int_0^T f_{\mathcal{B}}(t|m,M) h_{\mathcal{B}}(t-\mathcal{H}(t)|b_\ell,\theta)dt\right) \\
&\quad \times \prod_{i=1}^{n}\left\{\sum_{m=1}^{M}\sum_{\ell=1}^{L} \omega_m^f \omega_\ell^h \Lambda f_{\mathcal{B}}(t_i|m,M) h_{\mathcal{B}}(x_i|b_\ell,\theta)\right\},
\end{aligned}$$

where $t_i$ denotes the time of event $i$, $x_i = t_i - t_{i-1}$ is the corresponding inter-arrival time, and $N(T) = n$ is the total number of observed events. This expression can be simplified by recognizing that the basis functions are parameter free. Let $h^*_{m,\ell}(t,x) = \Lambda f_{\mathcal{B}}(t|m,M)h_{\mathcal{B}}(x|b_\ell,\theta)$ be shorthand for the product of the basis evaluations for given $m$ and $\ell$. Additionally, we define partial cumulative basis terms $H^*_{m,l,i} = \int_0^{x_i} h^*_{m,\ell}(x + t_{i-1}, x)dx$, and denote the corresponding aggregate terms as $H^*_{m,l} = \sum_{i=1}^{n} H^*_{m,l,i} + \int_0^{T-t_n} h^*_{m,\ell}(x+t_n, x)dx$. The likelihood function can then be written as,

$$\exp\left(-\sum_{m=1}^{M}\sum_{\ell=1}^{L} \omega_m^f \omega_\ell^h H^*_{m,\ell}\right) \times \prod_{i=1}^{n}\left\{\sum_{m=1}^{M}\sum_{\ell=1}^{L} \omega_m^f \omega_\ell^h h^*_{m,\ell}(t_i,x_i)\right\}. \tag{4.16}$$

An important feature of this likelihood is that the $H^*_{m,\ell}$ and $h^*_{m,\ell}(t_i,x_i)$ terms are fixed quantities given the data, meaning that the values can be precomputed and reused for

each iteration of MCMC sampling. This is a huge computational advantage and a key motivation for our use of basis representation models for both functions. The integral involved in each $H^*_{m,l,i}$ cannot be solved analytically, but is easily computed numerically given that the integrand is bounded and smooth. Gaussian quadrature is efficient and quite accurate, though a variety of other methods could be used as well.

### 4.2.3 Posterior simulation

Gibbs sampling can be used effectively to sample from the posterior distribution of the proposed ModRP model. Similar to the HRP log-logistic basis model, we introduce latent component membership variables $z_i = (z_i^f, z_i^h)$ for each event time $t_i$ such that,

$$\Pr(z_i = (m,\ell)|\boldsymbol{\omega}^f, \boldsymbol{\omega}^h) = \frac{\omega_m^f \omega_\ell^h}{\sum_{m=1}^M \sum_{\ell=1}^L \omega_m^f \omega_\ell^h} = \omega_m^f \omega_\ell^h (\Omega^h)^{-1}\,, \tag{4.17}$$

where $\Omega^h = \sum_{\ell=1}^L \omega_\ell^h$ denotes the total hazard weight. The augmented data likelihood is then given by,

$$\exp\left(-\sum_{m=1}^M \sum_{\ell=1}^L \omega_m^f \omega_\ell^h H^*_{m,\ell}\right) \times \prod_{i=1}^n \Omega^h h^*_{z_i}(t_i, x_i)\,, \tag{4.18}$$

where $h^*_{z_i}$ is shorthand for $h^*_{z_i^f, z_i^h}$. Using bivariate membership variables in this way yields a conditional structure that essentially separates the sampling for the hazard parameters from the modulating intensity parameters. Indeed, the hazard weights $\boldsymbol{\omega}^h$ are still drawn from a multivariate truncated log-normal distribution using slice sampling uniform variables, having only slightly modified parameter expressions compared to the HRP model in order to account for the modulating intensity weights. Likewise,

the modulating intensity weights $\boldsymbol{\omega}^f$ are updated using slice sampling variables and a truncated Dirichlet distribution. The hyperparameters for the hazard weights are conditionally independent of the modulating intensity weights and vice versa, enabling the use of sampling protocols developed for both the HRP log-logistic model and the Bernstein polynomial density model. Derivations for the full conditional distributions and corresponding sampling algorithms for all parameters are provided in Appendix E.

### 4.2.4 Approach to model checking

A commonly used method of model checking for temporal point processes is to lean on the time-rescaling theorem (Daley and Vere-Jones, 2003). Applying this to a ModRP setting is similar to the approach discussed for the HRP model in Chapter 2.1.4, but we include a brief review here for completeness and to highlight a few key differences.

The theorem is concerned with rescaling event times by the cumulative conditional intensity. Letting $\Lambda(t|\mathcal{H}(t)) = \int_0^t \lambda(s|\mathcal{H}(s))ds$ denote the cumulative conditional intensity, the rescaled event times are given by $t_i^* = \Lambda(t_i|\mathcal{H}(t_i))$. Unlike the HRP case, the integral for a ModRP conditional intensity is not generally expressed succinctly in terms of the cumulative hazard or cumulative modulating intensity and must be computed numerically. Assuming the observed point pattern arises from a process with the given conditional intensity, then the rescaled times form a Poisson process with unit rate and the transformed inter-arrival times are exponentially distributed. It is common to transform the rescaled inter-arrival times by $u_i^* = 1 - \exp\{-(t_i^* - t_{i-1}^*)\}$, which are uni-

formly distributed under the assumed process, and then to visually compare quantiles of these transformed times to the corresponding uniform quantiles in a QQ plot.

Time-rescaling QQ plots simultaneously check both the quality of the statistical model fit and whether the stochastic model assumptions implied by the conditional intensity are met. In the case of our ModRP model, the flexibility of the basis models for both functions implies that the QQ plots are more useful for checking whether the modulated renewal process structure is appropriate. Similar visualizations could be constructed for flexible HRP and NHPP models as a way of checking whether these special cases offer equivalent performance.

### 4.2.5 Synthetic data examples

Here we illustrate the performance of our proposed nonparametric ModRP model using synthetic data based on known hazard and modulating intensity functions. We consider the following five sets of generative functions:

1. $\lambda(t) = 1$ and $h(x) = 1$, corresponding to a homogeneous Poisson process.

2. $\lambda(t) = 1$ and $h(x) = 2x$, corresponding to a homogeneous renewal process with $\text{Wei}(2, 1)$ distributed inter-arrival times.

3. $\lambda(t) = (0.33)\sin(t/100) + 1$ and $h(x) = 1$, corresponding to a nonhomogeneous Poisson process with a sinusoidal intensity.

4. $\lambda(t) = (0.33)\sin(t/100) + 1$ and $h(x) = 2x$, combining the previous two scenarios into a non-trivial modulated renewal process.

5. $\lambda(t) = (t/500)(1 + \sin(t/100)) + 0.25$ and the hazard $h$ corresponding to an even mixture of $\text{Ga}(2, 1/4)$ and $\text{Ga}(4, 1/2)$ densities.

For each scenario, we sample event times with time bound $T = 1000$, yielding approximately 1,000 events for each simulation. We then fit our model using $M = 30$ modulating density basis functions, $L = 30$ hazard basis functions, and hyperparameter values of $a_\alpha = 1$, $b_\alpha = 1$, $a_\sigma = 2.5$, and $b_\sigma = 1$, as well as the default values suggested previously for the log-Gaussian process hyperparameters. The MCMC algorithm is run for 5,000 iterations following a burnin period of 15,000.

Following from the previous discussion on identifiability, these generative functions produce equivalent processes when proportionally rescaled. Practically, this means that recovering the generative functions is not possible without additional scale information, namely the value of $\Lambda$. For example, the scales given in scenario 5 produce a total intensity of $\Lambda \approx 1406$. We generate data from this scenario and fit our model using both the default choice of $\Lambda = T$ and the informed choice of $\Lambda = 1406$. Because visuals play a key role in diagnosing the impact of $\Lambda$, we start with scenario 5 as it is the most complex in terms of function shape. In Figure 4.11 we display time-rescaling QQ plots and compare the estimated hazard and modulating intensity functions for both choices. Not surprisingly, using the informed $\Lambda$ recovers the generative functions well, while the default choice misses the scale but otherwise captures the shapes. The QQ plots in Figure 4.11 are virtually indistinguishable, confirming again that $\Lambda$ is not identifiable and indicating that the choice of $\Lambda$ has no discernible impact on the quality of model fit.

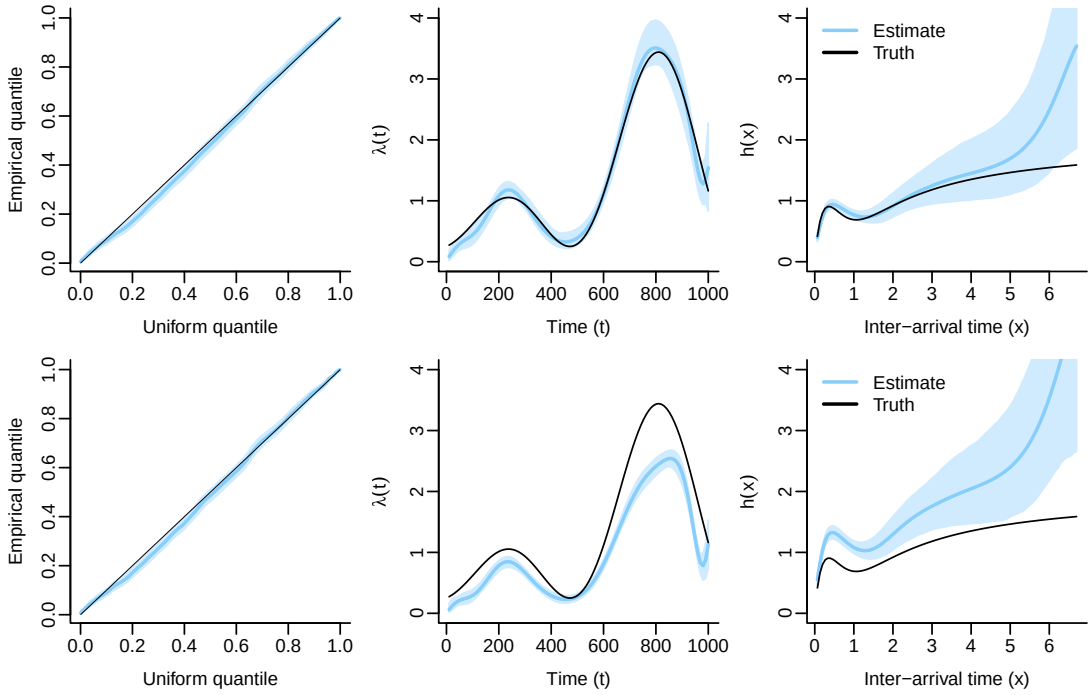


Figure 4.11: Results of fitting the nonparametric ModRP model to synthetic data generated from scenario 5. The top row shows model results using the informed choice of $\Lambda = 1406$, while the bottom row shows results using the default choice of $\Lambda = T$. Middle and right-most panels show the estimated modulating intensity and hazard functions, respectively, and the left-most panels show time-rescaling QQ plots.

We also show posterior inference for conditional inter-arrival time densities in Figure 4.12 for several values of previous times $t_{prev}$. These are interesting on their own as they provide insight on the inter-arrival time dynamics, but they also corroborate the overall invariance of the model fit to the choice of $\Lambda$. The conditional density expression in Equation (4.10) involves multiplying the hazard and modulating intensity, which is itself unaffected by rescaling between the two functions. In other words, these quantities are identifiable and should not depend on the choice of $\Lambda$. This is confirmed by the very similar estimated densities between the top row (informed $\Lambda$) and bottom row (default $\Lambda$) in Figure 4.12. In practice, minor differences appear in areas where the modulating

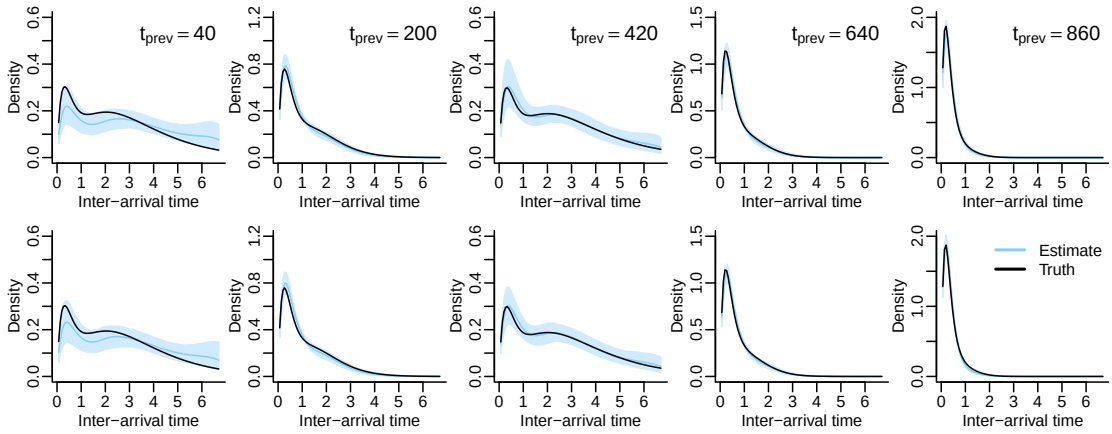


Figure 4.12: Posterior inference for conditional inter-arrival time densities using the nonparametric ModRP model fit to synthetic data generated from scenario 5. The top row shows model results using the informed choice of $\Lambda = 1406$, while the bottom row shows results using the default choice of $\Lambda = T$. The values of $t_{prev}$ are shown at the top of each column.

intensity is near extreme values, particularly on the lower side. This is due to the fact that the priors on the basis coefficients are not scaled with the choice of $\Lambda$, and so have slightly different levels of influence in these regions. Given the results in Figures 4.11 and 4.12, we conclude that the choice of $\Lambda$ is essentially arbitrary for the purposes of model fitting and that a default choice of $\Lambda = T$ is generally sufficient. However, for the sake of visual consistency in comparing our model estimates to the generative functions, we use the *correct* values of $\Lambda$ for the remaining scenarios.

We fit our model to data generated from scenarios 1 through 4 using the same configuration parameters as for scenario 5. The results for the estimated hazards and modulating intensities are shown in Figure 4.13. Note that we forgo the time-rescaling QQ plots for these scenarios to save space and focus on the functional estimates. The estimated QQ plots for all scenarios are very close to the unit line, as expected given the flexibility of our model and the known ModRP structure of the data. Overall,

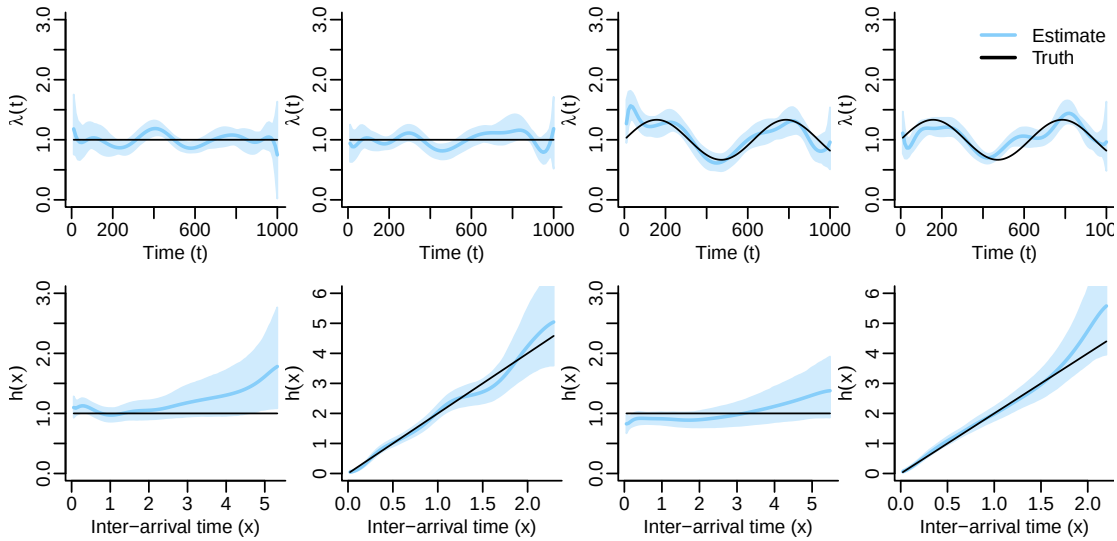


Figure 4.13: Results of fitting the nonparametric ModRP model to synthetic data generated from scenarios 1 through 4, shown in columns left to right. The top row shows the estimated hazard functions, and the bottom row shows the estimated modulating intensity functions.

the model successfully recovers the generative functions for all scenarios, including the special cases, though the modulating intensity functions exhibit some overfitting for the constant intensity scenarios. This is in part due to the large number of basis functions used to estimate a simple constant function. We omit them here for brevity, but fitting the model with a much smaller value of $M$ ($\sim 10$) yields smoother modulating intensity estimates, but also reduces the resolution of the sinusoidal intensity in scenarios 3 through 5. Similar observations follow from reducing $L$ in estimating the hazard, though it seems that the hazard estimates are less sensitive to large $L$. It should be noted that when $M$ is too large relative to the number of events $n$, the truncated Dirichlet sampling steps can produce invalid truncation bounds. Sampling from a truncated Dirichlet is a challenging problem in general, and while a more sophisticated approach could be used, it is easier and more effective to simply reduce $M$ in practice.

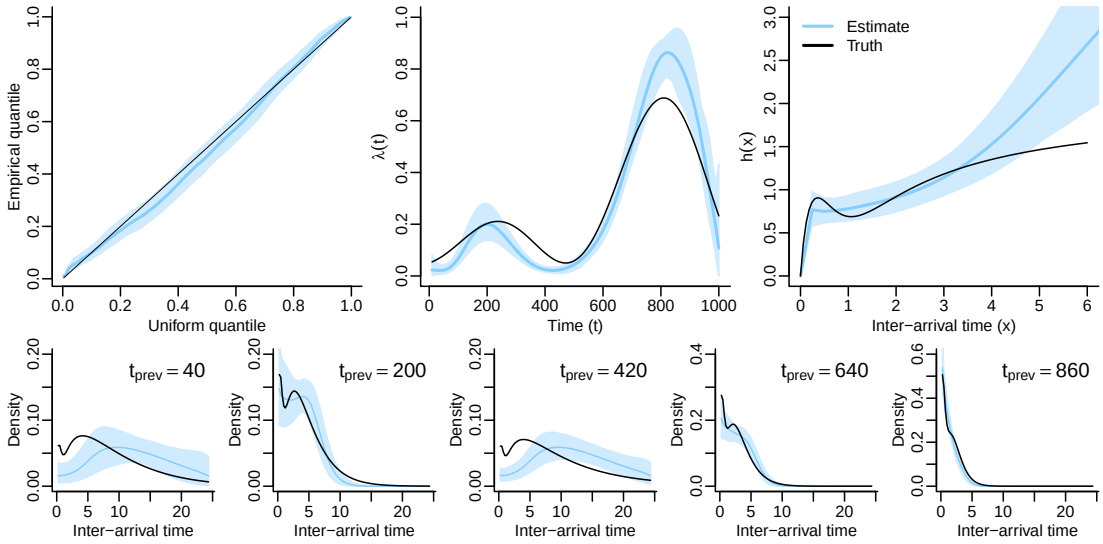


Figure 4.14: Results of fitting the nonparametric ModRP model to synthetic data generated from scenario 5 with a reduced modulating intensity producing a smaller number of events. The top row from left to right shows the time-rescaling QQ plot, the estimated hazard function, and the estimated modulating intensity. The bottom row shows the estimated conditional inter-arrival time densities for several values of $t_{prev}$.

Recall that these generated datasets consist of approximately 1,000 events each. To explore model performance in the presence of fewer events, we generate data from scenario 5 where the true modulating intensity has been scaled down by a factor of 5, producing around 300 events. In Figure 4.14, we show the estimated hazard and modulating intensity function, as well as a time-rescaling QQ plot and several conditional inter-arrival time densities. In terms of recovering the truth, the model does not perform as well as in the previous scenarios with more data. The modulating intensity shape is captured well enough—the deviation near $t = 500$ has more to do with idiosyncrasies in the generated data than the model fit—but the multimodal shape of the hazard is missed. Yet, the model fit is still reasonable, recovering most of the underlying structure and the overall pattern of the various functions.

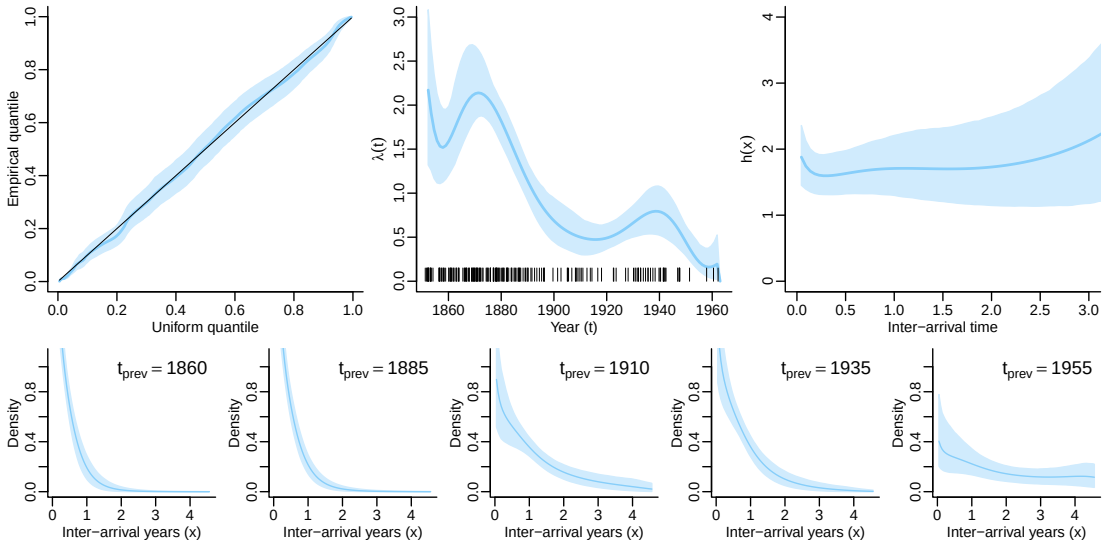


Figure 4.15: Results of fitting the nonparametric ModRP model to the coal-mining disasters dataset. The top row from left to right shows the time-rescaling QQ plot, the estimated modulating intensity, and the estimated hazard function. The bottom row shows the estimated conditional inter-arrival time densities for several values of $t_{prev}$.

### 4.2.6 Real data examples

#### Coal-mining disasters

The coal-mining disasters dataset from Jarrett (1979) is a classic target for point process modeling that consists of 191 events recorded between 1851 and 1962. The point pattern is known to exhibit time-varying intensity behavior, corresponding to improvements in safety regulation and technology over the years. Our goal with this analysis is to understand if the NHPP structure is sufficient to capture the underlying dynamics or if a more complex ModRP model is needed. We note that this dataset is frequently used as an illustration for NHPP models. For instance, both Taddy and Kottas (2012) and Kim and Kottas (2022) develop nonparametric mixture models for NHPP intensities and apply them to the coal-mining disasters data.

We fit our model using $M = 20$, $L = 20$, and the same hyperparameter values and MCMC configuration as in the synthetic data examples. Figure 4.15 shows the estimated hazard and modulating intensity functions, as well as a time-rescaling QQ plot and several conditional inter-arrival time densities. Looking at our results, in particular the flatness of the estimated hazard function, it seems that the NHPP assumption is reasonable. Moreover, our estimated modulating intensity is very similar in terms of shape to the intensity functions in Taddy and Kottas (2012) and Kim and Kottas (2022). Of course, the scale of the modulating intensity does not match, but this is to be expected given the identifiability issues discussed earlier. Rao and Teh (2011) also analyze this dataset using their proposed ModRP model, yet the modulating intensity function estimated there does not mirror the other NHPP results as well as our method does.

**NOAA significant earthquakes**

Recall the NOAA significant earthquake dataset presented in Section 4.1.6, which consists of over 800 destructive earthquakes between 1900 and 2019 split between four regions of the American continents. With our proposed methodology, we aim to understand the underlying structure of these events and how they differ across regions. We fit the model to each region separately using the same configuration parameters as for the coal-mining dataset. Computation time and sampling efficiency are effectively the same as for the HRP analysis in Section 4.1.6. A major driver for computational savings here comes from the dimensionality of the covariance matrix $\Sigma$, which depends

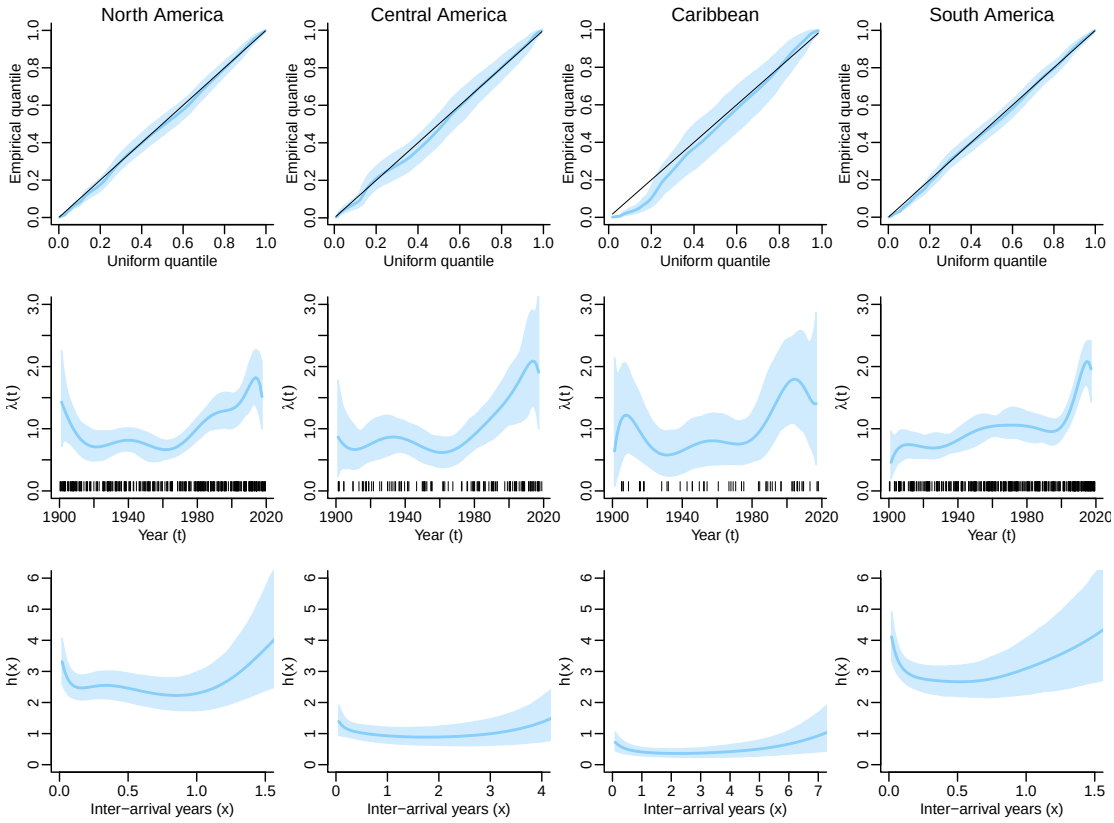


Figure 4.16: Results of fitting the nonparametric ModRP model to the NOAA significant earthquake dataset. The specific regions are indicated at the top of each column. The time-rescaling QQ plot, estimated modulating intensity, and estimated hazard function are shown in the top, middle, and bottom rows, respectively.

on the number of basis functions, not the number of observations, and therefore scales well with larger datasets. About 20% of the compute time is spent precomputing the $H^*$ terms.

Figure 4.16 shows the estimated hazard and modulating intensity functions, as well as a time-rescaling QQ plot for each region. The inter-arrival hazard shapes (and their relative scales) are broadly consistent with those from the HRP data analysis in Section 4.1.6. Interestingly, the modulating intensity estimates seems to suggest a

rise in earthquake activity starting around 1980. This is consistent with advances in detection technology implemented in California around that time (Hutton et al., 2010), and we suppose that similar upgrades could have been made in other regions as well, thus potentially explaining the increase in detected events. Despite this rise in the intensity, the time-rescaling QQ plots are roughly similar to the HRP model results, suggesting that the HRP assumption may be at least tolerable in practice.

## 4.3 Concluding remarks

Our goal in this chapter was to develop a flexible and computationally efficient model for the inter-arrival hazard function of a renewal process. Inspired by basis representation models such as Erlang mixtures and Bernstein polynomial models, we proposed the log-logistic basis system and developed a corresponding nonparametric prior probability model for the hazard function. To illustrate the utility of a hazard-oriented renewal process model, we developed a nonparametric model for modulated renewal processes by combining the log-logistic basis model with a nonparametric intensity model. In both simulated and real data examples, our proposed methodologies performed well, recovering true generative functions and providing meaningful insights into the underlying dynamics of the real data.

The log-logistic hazard model has the capacity to be applied in many other contexts. An obvious extension is to survival analysis, where hazard functions are often of interest. Covariates could be incorporated through a structure similar to

the Cox proportional hazards model, which shares roots with modulated renewal processes. Indeed, the modulated renewal process originally developed by Cox (1972) is formulated such that the inter-arrival hazard is modulated by a regression-like function $\exp\{\beta_1 x_1(t) + \ldots + \beta_p x_p(t)\}$ where $x_1(t), \ldots, x_p(t)$ are time-varying covariates. A similar structure could be used to incorporate covariates into the ModRP model through the modulating intensity: $\lambda(t) = \lambda_0(t) \exp\{\beta_1 x_1(t) + \ldots + \beta_p x_p(t)\}$, where $\lambda_0(t)$ is a baseline intensity.

A notable drawback of the proposed HRP model is the difficulty of obtaining inference for the $K$-function. The implied density function that arises from representing the hazard as a log-logistic basis mixture does not have a closed-form expression for the Laplace transform, thus requiring numerical approximation routines to perform both the Laplace transform and the inverse transform. Not only is this a significant computational burden, overflows and precision issues frequently arise and lead to inaccurate results. Future work focusing on developing more efficient and reliable methods for obtaining the $K$-function when the Laplace transform is not analytically tractable would be a valuable contribution to the field.

The ModRP model we develop performs well in practice, but there are several areas for improvement. We mentioned previously the issue of truncated Dirichlet sampling. The Gibbs sampling approach is tractable, but struggles with instability when $M$ is large. A promising alternative is to use multivariate slice sampling, requiring more effort to determine the appropriate latent variable structure. Another area for improvement would be making $L$ and $M$ random variables, but this seems quite dif-

ficult due to computational limitations. A major source of computational savings for our current model comes from the ability to precompute and reuse basis evaluations. Allowing $L$ and $M$ to vary would necessitate recomputing these evaluations at each iteration, bumping up compute time by many orders of magnitude. Rather than pursue this direction, perhaps a more practical approach would be to explore regularizing the basis coefficients through penalties or shrinkage priors.

# Chapter 5

# Conclusion

The main objective of this dissertation is to develop stochastic modeling frameworks for various types of renewal processes. Our developments aim to provide flexible and computationally efficient methods that are capable of incorporating prior information in a principled manner.

In Chapter 2, we have developed a nonparametric mixture modeling framework for homogeneous renewal process densities. This framework is based on Dirichlet process mixtures, but with specific constraints on the mixture kernel and prior specification to ensure that the renewal process finite mean condition is met. We arrive at the gamma density kernel as a suitable general-purpose choice that requires only a minor restriction on the prior hyperparameters. In addition, motivated by the application area of earthquake recurrence modeling, we present details for a structured uniform mixture kernel. This kernel mixture produces flexible density estimates within a decreasing shape constraint. We consider two stick-breaking prior models for the weights, one

following the Dirichlet process and the other following logit stick-breaking, ultimately finding that results are similar in a homogeneous renewal process context.

Chapter 3 is concerned with Markov renewal processes, a marked generalization of the homogeneous renewal process where discrete state information is observed with each event. These stochastic models are particularly relevant in earthquake recurrence modeling, where several key models in the literature are presented in this context. The MRP likelihood function is structured such that observed sojourn times are partitioned into transition cases, each with its own sojourn time distribution. Independent treatment of each transition case is common in the literature, but can struggle in settings with limited amounts of data and ignores possible dependencies that may exist between cases. In combination with the homogeneous renewal process model to represent sojourn time densities, we propose a dependent nonparametric prior for the mixture components that facilitates borrowing of strength between transition cases. Compared with the traditional independent estimation approach, the resulting model has better posterior predictive performance and can provide additional insights into the underlying dependence structure.

In Chapter 4, we have considered modeling renewal processes from the perspective of the inter-arrival hazard function. We introduced a novel log-logistic hazard basis system for directly modeling hazard functions. The proposed basis system is supported by a convergence result that confirms its flexibility, and suggests a structured nonparametric prior for the basis coefficients. Performance of the proposed hazard model is similar to that of flexible density-based models, but the hazard-based approach offers

two key advantages: prior information about the inter-arrival hazard can be more naturally incorporated, and the model can be more easily adapted to time-varying renewal process extensions. As a key example, we have also developed a nonparametric model for modulating renewal processes. This model combines the hazard basis system with a Bernstein polynomial representation for the modulating intensity function. A key advantage of this model is in computation: having fixed basis representations for both functions facilitates a more standard MCMC sampling approach that is computationally efficient.

Throughout the development of these models, we have observed a few themes that highlight some of the more unexpected outcomes of the dissertation. Perhaps most prominent is the importance of flexible methods. It was surprising to find that nonparametric mixture models have received limited attention in the renewal process literature, despite their popularity in other areas. This applies also to the earthquake applications, where parametric models are still somewhat common. In both simulation and real data applications, we have seen that the nonparametric models can provide more accurate and informative results than their parametric counterparts. However, the preference for more structured models is understandable, as the theoretical details for larger models can be difficult to manage, and the computational burden can be substantial, especially for extensions. In this regard, accessibility to computationally efficient methods is one of our more meaningful contributions.

The other surprising outcome worth noting is the effectiveness of the hazard-oriented approach. Although we had hoped to develop a flexible basis model for hazard

functions, discovering the convergence result that supports the log-logistic basis system was unexpected, as was the emergence of a potential prior model from the proof. Furthermore, despite the model being placed on the hazard scale, we found that the implied estimate of the density function was competitive with other density-based models. The approach to posterior simulation was also more straightforward than expected, a boon that we attribute to the fixed basis functions and recent advances in truncated multivariate normal sampling. Between the computational tractability, the theoretical foundation, and our empirical results, we believe that the log-logistic hazard basis system is a promising development that will prove useful in a variety of applications, even beyond renewal processes.

# Appendix A

# Posterior Computation for the HRP Kernel Mixture Model

Here we present details for posterior computation for the HRP kernel mixture model presented in Chapter 2. Our approach is based on Gibbs sampling, following a somewhat standard blocked Gibbs algorithm. Many of the full conditional distributions arise from conjugate prior relationships, or otherwise produce expressions corresponding to well-known distributions. For parameters with full conditional expressions that are not straightforward to sample, we use random walk Metropolis updates with a log-normal proposal distribution. We implement the adaptive batching scheme of Roberts and Rosenthal (2009) to tune the proposal variance, targeting an acceptance rate between 35% and 50%.

## A.1 Full conditionals for gamma kernel parameters

**Sampling $\gamma_\kappa$**

The full conditional for $\gamma_\kappa$ is given by,

$$p(\gamma_\kappa|\boldsymbol{\kappa}) \propto p(\gamma_\kappa)\prod_{\ell=1}^{L} p(\kappa_\ell|\gamma_\kappa) \propto \gamma_\kappa^{-a_\gamma^\kappa-1}\exp\left(-\frac{b_\gamma^\kappa}{\gamma_\kappa}\right)\prod_{\ell=1}^{L}\gamma_\kappa^{-a_\kappa}\exp\left(\frac{-\kappa_\ell}{\gamma_\kappa}\right),$$

which is proportional to an inv-Ga$(a^*, b^*)$ distribution with shape $a^* = a_\gamma^\kappa + La_\kappa$ and rate $b^* = b_\gamma^\kappa + \sum_{\ell=1}^{L}\kappa_\ell$. Sampling from this distribution is straightforward.

**Sampling $\gamma_\lambda$**

The full conditional for $\gamma_\lambda$ is given by,

$$p(\gamma_\lambda|\boldsymbol{\lambda}) \propto p(\gamma_\lambda)\prod_{\ell=1}^{L} p(\lambda_\ell|\gamma_\lambda) \propto \gamma_\lambda^{a_\gamma^\lambda-1}\exp\left(\frac{-\gamma_\lambda}{b_\gamma^\lambda}\right)\prod_{\ell=1}^{L}\gamma_\lambda^{a_\lambda}\exp\left(-\frac{\gamma_\lambda}{\lambda_\ell}\right),$$

which is proportional to a Ga$(k^*, l^*)$ distribution with shape $k^* = a_\gamma^\lambda + La_\lambda$ and scale $l^* = (\frac{1}{b_\gamma^\lambda} + \sum_{\ell=1}^{L}\frac{1}{\lambda_\ell})^{-1}$. Sampling from this distribution is straightforward.

**Sampling $\kappa_\ell$**

The full conditional for $\kappa_\ell$ is given by,

$$\begin{aligned} p(\kappa_\ell|\boldsymbol{z}, z_c, \boldsymbol{\lambda}, \gamma_\kappa, \boldsymbol{x}, x_c) &\propto p(\kappa_\ell|\gamma_\kappa)p(x_c|\kappa_\ell)^{\mathbb{I}(z_c=\ell)}\prod_{i\in\boldsymbol{z}_\ell} p(x_i|\kappa_\ell) \\ &\propto \text{Ga}(\kappa_\ell|a_\kappa, \gamma_\kappa)(1 - F_{\text{Ga}}(x_c|\kappa_\ell, \lambda_\ell))^{\mathbb{I}(z_c=\ell)}\prod_{i\in\boldsymbol{z}_\ell}\text{Ga}(x_i|\kappa_\ell, \lambda_\ell), \end{aligned}$$

where $F_{\text{Ga}}$ is the CDF of the Gamma distribution, $\mathbb{I}(\cdot)$ is the indicator function, and $\boldsymbol{z}_\ell = \{i : z_i = \ell\}$ is the set of indices corresponding with observations assigned to

component $\ell$. This is not a conjugate update, so we use a random walk Metropolis step to sample from this conditional distribution.

**Sampling $\lambda_\ell$**

The full conditional for $\lambda_\ell$ is given by,

$$\begin{aligned} p(\lambda_\ell|\boldsymbol{z}, z_c, \boldsymbol{\kappa}, \gamma_\lambda, \boldsymbol{x}, x_c) &\propto p(\lambda_\ell|\gamma_\lambda) p(x_c|\lambda_\ell)^{\mathbb{I}(z_c=\ell)} \prod_{i \in \boldsymbol{z}_\ell} p(x_i|\lambda_\ell) \\ &\propto \lambda_\ell^{-a_\lambda - 1} \exp\left(-\frac{\gamma_\lambda}{\lambda_\ell}\right) (1 - F_{\text{Ga}}(x_c|\kappa_\ell, \lambda_\ell))^{\mathbb{I}(z_c=\ell)} \prod_{i \in \boldsymbol{z}_\ell} \lambda_\ell^{\kappa_\ell - 1} \exp\left(-\frac{x_i}{\lambda_\ell}\right), \end{aligned}$$

where $n_\ell = |\boldsymbol{z}_\ell|$ is the number of observations assigned to component $\ell$. If $z_c \neq \ell$, then this expression is proportional to an inv-Ga$(a^*, b^*)$ distribution with shape $a^* = a_\lambda + n_\ell \kappa_\ell$ and rate $b^* = \gamma_\lambda + \sum_{i \in \boldsymbol{z}_\ell} x_i$, which is straightforward to sample from. When $z_c = \ell$, the term involving $x_c$ is present and the expression is no longer conjugate, so we use a random walk Metropolis step for this case.

## A.2 Full conditionals for uniform kernel parameters

**Sampling $\eta_\theta$**

The full conditional for $\eta_\theta$ is given by,

$$p(\eta_\theta|\boldsymbol{\theta}, \gamma_\theta) \propto p(\eta_\theta) \prod_{\ell=1}^{L} p(\theta_\ell|\eta_\theta, \gamma_\theta) \propto \text{Ga}(\eta_\theta - 1|a_\eta^\theta, b_\eta^\theta) \prod_{\ell=1}^{L} \text{inv-Ga}(\theta_\ell|\eta_\theta, \gamma_\theta)\,.$$

This is not an easily sampled density, so we use a random walk Metropolis step.

**Sampling $\gamma_\theta$**

The full conditional for $\gamma_\theta$ is given by,

$$p(\gamma_\theta|\boldsymbol{\theta}, \eta_\theta) \propto p(\gamma_\theta) \prod_{\ell=1}^{L} p(\theta_\ell|\eta_\theta, \gamma_\theta) \propto \gamma_\theta^{a_\gamma^\theta - 1} \exp\left(-\frac{\gamma_\theta}{b_\gamma^\theta}\right) \prod_{\ell=1}^{L} \gamma_\theta^{\eta_\theta} \exp\left(-\frac{\gamma_\theta}{\theta_\ell}\right),$$

which is proportional to a $\text{Ga}(k^*, l^*)$ distribution with shape $k^* = a_\gamma^\theta + \eta_\theta L$ and scale $l^* = \left(\frac{1}{b_\gamma^\theta} + \sum_{\ell=1}^{L} \frac{1}{\theta_\ell}\right)^{-1}$. Sampling from this distribution is straightforward.

**Sampling $\theta_\ell$**

The full conditional for $\theta_\ell$ is given by,

$$\begin{aligned} p(\theta_\ell|\eta_\theta, \gamma_\theta, \boldsymbol{z}, z_c, \boldsymbol{x}, x_c) &\propto p(\theta_\ell|\eta_\theta, \gamma_\theta) p(x_c|\theta_\ell)^{\mathbb{I}(z_c=\ell)} \prod_{i \in \boldsymbol{z}_\ell} p(x_i|\theta_\ell) \\ &\propto \theta_\ell^{-n_\ell - \eta_\theta - 1} \exp\left(-\frac{\gamma_\theta}{\theta_\ell}\right) \mathbb{I}\left(\max_{i \in \boldsymbol{z}_\ell} x_i < \theta_\ell\right) \left((1 - \frac{x_c}{\theta_\ell})\mathbb{I}(x_c < \theta_\ell)\right)^{\mathbb{I}(z_c=\ell)}. \end{aligned}$$

For the case $z_c \neq \ell$, this is proportional to an inv-$\text{Ga}(a^*, b^*)$ distribution with shape $a^* = n_\ell + \eta_\theta$ and rate $b^* = \gamma_\theta$, truncated with a lower bound of $\max_{i \in \boldsymbol{z}_\ell} x_i$. This can be sampled directly using the inverse CDF method, which is made more stable by working on the log-scale. When, $z_c = \ell$, the term involving $x_c$ is present and the expression is no longer conjugate. We can avoid a random walk Metropolis step by introducing an auxiliary uniform variable $u$ such that joint conditional distribution is given by,

$$p(\theta_\ell, u|-) \propto \text{inv-Ga}(\theta_\ell|n_\ell + \eta_\theta, \gamma_\theta) \mathbb{I}\left(\max_{i \in \boldsymbol{z}_\ell} x_i < \theta_\ell\right) \mathbb{I}(x_c < \theta_\ell) \mathbb{I}\left(u < 1 - \frac{x_c}{\theta_\ell}\right),$$

where the original full conditional is recovered by marginalizing over $u$. We sample from this expression using a Gibbs step, drawing $u$ from a $\text{Unif}(0, 1 - x_c/\theta_\ell)$ distribution and

then sampling $\theta_\ell$ from an inv-Ga$(n_\ell + \eta_\theta, \gamma_\theta)$ distribution truncated above which ever is larger between $\max_{i \in \boldsymbol{z}_\ell} x_i$ and $x_c/(1-u)$.

## A.3 Full conditionals for weight parameters

**Sampling $z_i$ and $z_c$**

The full conditional for $z_i$ is given by $\Pr(z_i = \ell | \boldsymbol{\theta}, \boldsymbol{\omega}, \boldsymbol{x}) \propto \omega_\ell f_{\mathcal{K}}(x_i|\theta_\ell)$, where $f_{\mathcal{K}}$ is the selected kernel density and $\theta_\ell$ are the corresponding kernel parameters. The full conditional for $z_c$ is similarly given by $\Pr(z_c = \ell | \boldsymbol{\theta}, \boldsymbol{\omega}, x_c) \propto \omega_\ell S_{\mathcal{K}}(x_c|\theta_\ell)$, where $S_{\mathcal{K}}$ is the selected kernel survival function. These are discretely valued with $\ell \in \{1, \ldots, L\}$, so we sample from these directly by calculating the expressions for each component and normalizing.

**Sampling $\boldsymbol{\omega}$ under the Dirichlet process prior**

Under the constructive definition of the Dirichlet process, the weights $\boldsymbol{\omega}$ arise from stick-breaking with beta latent variables (Sethuraman, 1994). When truncated to a mixture of $L < \infty$ components, the prior on the weights is a generalized Dirichlet distribution with probability density denoted by,

$$\text{GDir}_L(\boldsymbol{\omega}|\boldsymbol{a}, \boldsymbol{b}) = \left( \prod_{\ell=1}^{L-1} \frac{\Gamma(a_\ell + b_\ell)}{\Gamma(b_\ell)\Gamma(a_\ell)} \right) \omega_L^{b_{L-1}-1} \left( \prod_{\ell=1}^{L-1} \omega_\ell^{a_\ell - 1} \left( \sum_{r=\ell}^{L} \omega_r \right)^{b_{\ell-1} - (a_\ell + b_\ell)} \right),$$

where $\boldsymbol{a}$ and $\boldsymbol{b}$ are $(L-1)$-dimensional vectors of positive shape parameters. When truncated, the Dirichlet process stick-breaking prior is equivalent to a generalized Dirichlet distributed with $a_\ell = 1$ and $b_\ell = \alpha$ for $\ell \in \{1, \ldots, L-1\}$. The full conditional for $\boldsymbol{\omega}$ is

then given by,

$$p(\boldsymbol{\omega}|\alpha, \boldsymbol{z}, z_c) \propto p(\boldsymbol{\omega}|\alpha)p(\boldsymbol{z}|\boldsymbol{\omega}) \propto \mathrm{GDir}_L(\boldsymbol{\omega}|\mathbf{1}, \boldsymbol{\alpha}) \prod_{\ell=1}^{L} \left( \omega_\ell^{\mathbb{I}(z_c=\ell)} \prod_{i=1}^{n} \omega_\ell^{\mathbb{I}(z_i=\ell)} \right) ,$$

where $\mathbf{1} = \{1, \ldots, 1\}$ and $\boldsymbol{\alpha} = \{\alpha, \ldots, \alpha\}$ are $(L-1)$-dimensional vectors. This is proportional to a $\mathrm{GDir}(\boldsymbol{\omega}|\boldsymbol{a}^*, \boldsymbol{b}^*)$ distribution with shape parameters $a_\ell^* = 1 + n_\ell^*$ and $b_\ell^* = \alpha + \sum_{r=\ell+1}^{L} n_r^*$, where $n_\ell^* = \mathbb{I}(z_c = \ell) + \sum_{i=1}^{n} \mathbb{I}(z_i = \ell)$ is the number of observations assigned to component $\ell$. This can be sampled directly using standard methods.

**Sampling $\alpha$**

Recall that $\alpha$ is the precision parameter for the Dirichlet process stick-breaking prior. The full conditional for $\alpha$ is a standard result given by,

$$p(\alpha|\boldsymbol{\omega}) \propto p(\alpha)p(\boldsymbol{\omega}|\alpha) \propto \alpha^{a_\alpha - 1} \exp\left(-\frac{\alpha}{b_\alpha}\right) \alpha^{L-1} \omega_L^\alpha .$$

This is proportional to a $\mathrm{Ga}(a_\alpha + L - 1, b_\alpha - \log(\omega_L))$, which is straightforward to sample from, though some care is required to prevent numerical underflow issues when taking the log of $\omega_L$.

**Sampling $\boldsymbol{\omega}$ under the logit stick-breaking prior**

Under the logit stick-breaking prior, the weights $\boldsymbol{\omega}$ arise from stick-breaking with transformed normal latent variables, given by,

$$\omega_\ell = \frac{\exp(\psi_\ell)}{1 + \exp(\psi_\ell)} \prod_{r=1}^{\ell-1} \frac{1}{1 + \exp(\psi_r)} .$$

We compute the weights deterministically from the latent variables $\boldsymbol{\psi}$ and draw the latent variables from the conditional distribution given by,

$$p(\psi_\ell|\mu,\sigma^2,\boldsymbol{z},z_c) \propto \mathrm{N}(\psi_\ell|\mu,\sigma^2)\prod_{\ell=1}^{L}\left(\omega_\ell^{\mathbb{I}(z_c=\ell)}\prod_{i=1}^{n}\omega_{z_i}^{\mathbb{I}(z_i=\ell)}\right)$$
$$\propto \exp\left(\frac{-(\psi_\ell-\mu)^2}{2\sigma^2}\right)\exp(\psi_\ell)^{n_\ell}\left(1+\exp(\psi_\ell)\right)^{-\sum_{r=\ell}^{L}n_r}\,,$$

where $n_\ell = \mathbb{I}(z_c=\ell)+\sum_{i=1}^{n}\mathbb{I}(z_i=\ell)$ is the number of observations assigned to component $\ell$. This is not a standard distribution, but as mentioned in the main text, we can avoid performing random walk Metropolis updates by introducing auxiliary Pólya-Gamma variables (Polson et al., 2013). The procedure relies on the following mixture identity:

$$\frac{(e^\psi)^a}{(1+e^\psi)^b} = 2^{-b}e^{m\psi}\int_0^\infty e^{-\xi\psi^2/2}\mathrm{PG}(\xi|b,0)d\xi\,,$$

where $m=a-b/2$ and $\mathrm{PG}(b,0)$ denotes the Pólya-Gamma distribution with shape $b$ and tilt $z=0$. Using this, the full conditional distribution of $\psi_\ell$ can be augmented by a latent Pólya-Gamma variable $\xi_\ell$:

$$p(\psi_\ell,\xi_\ell|-) \propto \mathrm{PG}\left(\xi_\ell\Big|\sum_{r=\ell}^{L}n_r,0\right)\exp\left(\frac{-(\psi_\ell-\mu)^2}{2\sigma^2}\right)\exp\left(m_\ell\psi_\ell-\frac{1}{2}\xi_\ell\psi_\ell^2\right)\,,$$

where $m_\ell = n_\ell - \sum_{r=\ell}^{L}n_r/2$. We draw from this joint density using a Gibbs step. The auxiliary Pólya-Gamma variable $\xi_\ell$ is drawn from a conjugate $\mathrm{PG}(b^*,z^*)$ distribution with shape $b^*=\sum_{r=\ell}^{L}n_r$ and tilt $z^*=\psi_\ell$. Efficient methods for sampling Pólya-Gamma variables are presented in both Polson et al. (2013) and Windle et al. (2014). Conditioned on $\xi_\ell$, the stick-breaking element $\psi_\ell$ is sampled from a conjugate $\mathrm{N}(m^*,v^*)$ distribution with $v^*=(1/\sigma^2+\xi_\ell)^{-1}$ and $m^*=v^*\left(\frac{\mu}{\sigma^2}+m_\ell\right)$.

### Sampling $\mu$ and $\sigma^2$

Recall that $\mu$ and $\sigma^2$ are the mean and variance parameters for the logit stick-breaking prior. The full conditional for $\mu$ and $\sigma^2$ is given by,

$$p(\mu, \sigma^2|\boldsymbol{\psi}) \propto p(\mu, \sigma^2)p(\boldsymbol{\psi}|\mu, \sigma^2) \propto \mathrm{N}(\mu|m_\mu, s_\mu^2)\text{inv-Ga}(\sigma^2|a_\sigma, b_\sigma) \prod_{\ell=1}^{L-1} \mathrm{N}(\psi_\ell|\mu, \sigma^2)\,.$$

This is a standard normal-inverse-gamma conjugate update, which can be sampled directly using standard methods. Our approach is to use a Gibbs step, sampling $\sigma^2$ from an inv-Ga$(a^*, b^*)$ distribution with shape parameter $a^* = a_\sigma + L/2$ and rate parameter $b^* = b_\sigma + \sum_{\ell=1}^{L-1}(\psi_\ell - \mu)^2/2$, and then sampling $\mu$ from a $N(m^*, v*)$ distribution with $v^* = \frac{1}{1/s_\mu^2 + L/\sigma^2}$ and $m^* = v^*(m_\mu/s_\mu^2 + \sum_{\ell=1}^{L-1} \psi_\ell/\sigma^2)$.

# Appendix B

# Posterior Computation for the MRP Mixture Model

Here we present details for posterior simulation for the DS-MRP mixture model described in Chapter 3. Most model parameters have conditionally conjugate full conditional distributions, which suggests Gibbs sampling as a natural choice for posterior simulation. In the case of parameters which are not conditionally conjugate, we use a random walk Metropolis update with log-normal proposal distribution and proposal variance dynamically tuned using the adaptive batching method of Roberts and Rosenthal (2009). Given that the MRP model is an extension of the uniform mixture presented in Chapter 2, we can leverage the posterior simulation methods developed there. In particular, the updates for $\boldsymbol{\theta}$, $\eta_\theta$, and $\gamma_\theta$ are identical to those given in Appendix A, so we omit them here.

## B.1 Full conditionals for data-level variables

**Sampling $z_i$ and $z_c$**

The full conditional for $z_i$ is given by $\Pr(z_i = \ell|\boldsymbol{\theta}, \boldsymbol{\omega}, \boldsymbol{x}, \boldsymbol{s}) \propto \omega_\ell^{(s_{i-1},s_i)} f_{\mathcal{K}}(x_i|\theta_\ell)$, where $f_{\mathcal{K}}$ is the (uniform) kernel density and $\theta_\ell$ are the corresponding kernel parameters. The full conditional for $z_c$ is similarly given by, $\Pr(z_c = \ell|\boldsymbol{\theta}, \boldsymbol{\omega}, \boldsymbol{x}, \boldsymbol{s}) \propto \omega_\ell^{(s_{i-1},s_i)} S_{\mathcal{K}}(x_c|\theta_\ell)$, where $S_{\mathcal{K}}$ is the corresponding kernel survival function. These are straightforward to sample directly.

**Sampling $s_c$**

The full conditional for $s_c$ is given by,

$$\Pr(s_c = k|z_c, \boldsymbol{\omega}, \boldsymbol{p}, \boldsymbol{s}) \propto \Pr(s_c = k|s_n, \boldsymbol{p}) \Pr(z_c = \ell|s_c = k, \boldsymbol{\omega}) \propto p^{(s_n,k)} \omega_{z_c}^{(s_n,k)} .$$

This is a discrete distribution with $S$ possible outcomes, which is straightforward to sample directly.

**Sampling $\boldsymbol{p}^{(j)}$**

Recall that $\boldsymbol{p}^{(j)} = \left\{p^{(j,1)},\ldots,p^{(j,S)}\right\}$. The full conditional for $\boldsymbol{p}^{(j)}$ is given by,

$$p(\boldsymbol{p}^{(j)}|\boldsymbol{s}, s_c) \propto p(\boldsymbol{p}^{(j)}) p(\boldsymbol{s}, s_c|\boldsymbol{p}^{(j)}) \propto \prod_{k=1}^{S} \left\{ (p^{(j,k)})^{\alpha_p + n^{(j,k)} - 1} \right\} ,$$

where $n^{(j,k)} = \mathbb{I}(s_n = j)\mathbb{I}(s_c = k) \sum_{i=1}^{n} \mathbb{I}(s_{i-1} = j)\mathbb{I}(s_i = k)$ denotes the number of observed transitions from state $j$ to state $k$. This is a Dirichlet distribution with parameters $\alpha_p + n^{(j,k)}$ for each $k$, which is simple to sample from.

## B.2 Full conditionals for hierarchical weight parameters

**Sampling $\sigma_\mu^2$**

The full conditional for $\sigma_\mu^2$ is given by,

$$\begin{aligned} p(\sigma_\mu^2|\boldsymbol{\mu}) &\propto p(\sigma_\mu^2)\prod_{\ell=1}^{L-1} p(\mu_\ell|\sigma_\mu^2) \\ &\propto (\sigma_\mu^2)^{-a_\sigma^\mu-1}\exp\left(-\frac{b_\sigma^\mu}{\sigma_\mu^2}\right)(\sigma_\mu^2)^{-(L-1)/2}\exp\left(-\frac{1}{2\sigma_\mu^2}\sum_{\ell=1}^{L-1}(\mu_\ell-m_0)^2\right), \end{aligned}$$

where $\boldsymbol{\mu} = \{\mu_1, \ldots, \mu_{L-1}\}$. This is proportional to an inv-Ga$(a^*, b^*)$ distribution with $a^* = a_\sigma^\mu + (L-1)/2$ and $b^* = b_\sigma^\mu + \sum_{\ell=1}^{L-1}(\mu_\ell - m_0)^2/2$. Sampling from this distribution is straightforward.

**Sampling $\sigma_\alpha^2$**

The full conditional for $\sigma_\alpha^2$ is given by,

$$\begin{aligned} p(\sigma_\alpha^2|\boldsymbol{\alpha},\boldsymbol{\mu}) &\propto p(\sigma_\alpha^2)\prod_{j=1}^{S}\prod_{\ell=1}^{L-1} p(\alpha_\ell^{(j)}|\mu_\ell,\sigma_\alpha^2) \\ &\propto (\sigma_\alpha^2)^{-a_\sigma^\alpha-1}\exp\left(-\frac{b_\sigma^\alpha}{\sigma_\alpha^2}\right)(\sigma_\alpha^2)^{-S(L-1)/2}\exp\left(-\frac{1}{2\sigma_\alpha^2}\sum_{j=1}^{S}\sum_{\ell=1}^{L-1}(\alpha_\ell^{(j)}-\mu_\ell/2)^2\right), \end{aligned}$$

where $\boldsymbol{\alpha} = \{\alpha_1^{(1)}, \ldots, \alpha_{L-1}^{(S)}\}$. This is proportional to an inv-Ga$(a^*, b^*)$ distribution with $a^* = a_\sigma^\alpha + S(L-1)/2$ and $b^* = b_\sigma^\alpha + \sum_{j=1}^{S}\sum_{\ell=1}^{L-1}(\alpha_\ell^{(j)} - \mu_\ell/2)^2/2$. Sampling from this distribution is straightforward.

**Sampling $\sigma^2_\beta$**

The full conditional for $\sigma^2_\beta$ is given by,

$$\begin{aligned} p(\sigma^2_\beta|\boldsymbol{\beta},\boldsymbol{\mu}) &\propto p(\sigma^2_\beta)\prod_{k=1}^{S}\prod_{\ell=1}^{L-1} p(\beta^{(k)}_\ell|\mu_\ell,\sigma^2_\beta) \\ &\propto (\sigma^2_\beta)^{-a^\beta_\sigma-1}\exp\left(-\frac{b^\beta_\sigma}{\sigma^2_\beta}\right)(\sigma^2_\beta)^{-S(L-1)/2}\exp\left(-\frac{1}{2\sigma^2_\beta}\sum_{k=1}^{S}\sum_{\ell=1}^{L-1}(\beta^{(k)}_\ell-\mu_\ell/2)^2\right), \end{aligned}$$

where $\boldsymbol{\beta}=\{\beta^{(1)}_1,\ldots,\beta^{(S)}_{L-1}\}$. This is proportional to an inv-Ga$(a^*,b^*)$ distribution with $a^*=a^\beta_\sigma+S(L-1)/2$ and $b^*=b^\beta_\sigma+\sum_{k=1}^{S}\sum_{\ell=1}^{L-1}(\beta^{(k)}_\ell-\mu_\ell/2)^2/2$. Sampling from this distribution is straightforward.

**Sampling $\sigma^2_\psi$**

The full conditional for $\sigma^2_\psi$ is given by,

$$\begin{aligned} p(\sigma^2_\psi|\boldsymbol{\psi},\boldsymbol{\alpha},\boldsymbol{\beta}) &\propto p(\sigma^2_\psi)\prod_{j=1}^{S}\prod_{k=1}^{S}\prod_{\ell=1}^{L-1} p(\psi^{(j,k)}_\ell|\alpha_\ell,\beta_\ell,\sigma^2_\psi) \\ &\propto (\sigma^2_\psi)^{-a^\psi_\sigma-1}\exp\left(-\frac{b^\psi_\sigma}{\sigma^2_\psi}\right)(\sigma^2_\psi)^{-S^2(L-1)/2} \\ &\quad\times\exp\left(-\frac{1}{2\sigma^2_\psi}\sum_{j=1}^{S}\sum_{k=1}^{S}\sum_{\ell=1}^{L-1}(\psi^{(j,k)}_\ell-\alpha^{(j)}_\ell-\beta^{(k)}_\ell)^2\right), \end{aligned}$$

where $\boldsymbol{\psi}=\{\psi^{(1,1)}_1,\ldots,\psi^{(S,S)}_{L-1}\}$. This is proportional to an inv-Ga$(a^*,b^*)$ distribution with $a^*=a^\psi_\sigma+S^2(L-1)/2$ and $b^*=b^\psi_\sigma+\sum_{j=1}^{S}\sum_{k=1}^{S}\sum_{\ell=1}^{L-1}(\psi^{(j,k)}_\ell-\alpha^{(j)}_\ell-\beta^{(k)}_\ell)^2/2$. Sampling from this distribution is straightforward.

**Sampling $\mu_\ell$**

The full conditional distribution for $\mu_\ell$ is given by,

$$p(\mu_\ell|\boldsymbol{\alpha},\boldsymbol{\beta},\sigma^2_\mu,\sigma^2_\alpha,\sigma^2_\beta) \propto \mathrm{N}(\mu_\ell|m_\mu,\sigma^2_\mu) \times \prod_{j=1}^{S} \mathrm{N}\left(\alpha_\ell^{(j)}|\mu_\ell/2,\sigma^2_\alpha\right) \times \prod_{k=1}^{S} \mathrm{N}\left(\beta_\ell^{(k)}|\mu_\ell/2,\sigma^2_\beta\right) .$$

This is a conjugate relationship, where the full conditional is a $\mathrm{N}(m^*,v^*)$ distribution with $v^* = \left(\frac{1}{\sigma^2_\mu} + \frac{S}{4\sigma^2_\alpha} + \frac{S}{4\sigma^2_\beta}\right)^{-1}$ and $m^* = v^*\left(\frac{m_\mu}{\sigma^2_\mu} + \sum_{j=1}^{S}\frac{\alpha_\ell^{(j)}}{2\sigma^2_\alpha} + \sum_{k=1}^{S}\frac{\beta_\ell^{(k)}}{2\sigma^2_\beta}\right)$. Sampling from this distribution is straightforward.

**Sampling $\alpha_\ell^{(j)}$**

The full conditional distribution for $\alpha_\ell^{(j)}$ is given by,

$$p(\alpha_\ell^{(j)}|\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\psi},\sigma^2_\alpha,\sigma^2_\psi) \propto \mathrm{N}(\alpha_\ell^{(j)}|\mu_\ell/2,\sigma^2_\alpha) \prod_{k=1}^{S} \mathrm{N}\left(\psi_\ell^{(j,k)}|\alpha_\ell^{(j)}+\beta_\ell^{(k)},\sigma^2_\psi\right) .$$

This is a conjugate relationship, where the full conditional is a $\mathrm{N}(m^*,v^*)$ distribution with $v^* = \left(\frac{1}{\sigma^2_\alpha} + \frac{S}{\sigma^2_\psi}\right)^{-1}$ and $m^* = v^*\left(\frac{\mu_\ell}{2\sigma^2_\alpha} + \sum_{k=1}^{S}\frac{\psi_\ell^{(j,k)}-\beta_\ell^{(k)}}{\sigma^2_\psi}\right)$. Sampling from this distribution is straightforward.

**Sampling $\beta_\ell^{(k)}$**

The full conditional distribution for $\beta_\ell^{(k)}$ is given by,

$$p(\beta_\ell^{(k)}|\boldsymbol{\mu},\boldsymbol{\alpha},\boldsymbol{\psi},\sigma^2_\beta,\sigma^2_\psi) \propto \mathrm{N}(\beta_\ell^{(k)}|\mu_\ell/2,\sigma^2_\beta) \prod_{j=1}^{S} \mathrm{N}\left(\psi_\ell^{(j,k)}|\alpha_\ell^{(j)}+\beta_\ell^{(k)},\sigma^2_\psi\right) .$$

This is a conjugate relationship, where the full conditional is a $\mathrm{N}(m^*,v^*)$ distribution with $v^* = \left(\frac{1}{\sigma^2_\beta} + \frac{S}{\sigma^2_\psi}\right)^{-1}$ and $m^* = v^*\left(\frac{\mu_\ell}{2\sigma^2_\beta} + \sum_{j=1}^{S}\frac{\psi_\ell^{(j,k)}-\alpha_\ell^{(j)}}{\sigma^2_\psi}\right)$. Sampling from this distribution is straightforward.

**Sampling $\psi_\ell^{(j,k)}$ and $\psi_\ell^{(k,j)}$**

The full conditional distribution for $\psi_\ell^{(j,k)}$ is given by,

$$p(\psi_\ell^{(j,k)}|\boldsymbol{z},\boldsymbol{\alpha},\boldsymbol{\beta},\sigma_\psi^2) \propto p(\psi_\ell^{(j,k)}|\alpha_\ell^{(j)},\beta_\ell^{(k)},\sigma_\psi^2)\prod_{r=\ell}^{L}\prod_{i\in\boldsymbol{z}_r^{(j,k)}} p(z_i|\boldsymbol{\psi})$$

$$\propto \mathrm{N}(\psi_\ell^{(j,k)}|\alpha_\ell^{(j)}+\beta_\ell^{(k)},\sigma_\psi^2)\prod_{r=\ell}^{L}\left(\omega_r^{(j,k)}\right)^{n_r^{(j,k)}}$$

$$\propto \exp\left(\frac{-(\psi_\ell^{(j,k)}-\alpha_\ell^{(j)}-\beta_\ell^{(k)})^2}{2\sigma_\psi^2}\right)\exp\left(\psi_\ell^{(j,k)}\right)^{n_\ell^{(j,k)}}\left\{1+\exp\left(\psi_\ell^{(j,k)}\right)\right\}^{-\sum_{r=\ell}^{L} n_r^{(j,k)}},$$

where $\boldsymbol{z}_r^{(j,k)}$ denotes the set of indices $i$ such that $z_i = r$, $s_i = k$, and $s_{i-1} = j$, with $n_r^{(j,k)} = |\boldsymbol{z}_r^{(j,k)}|$ denoting the size. This is not an easily recognizable distribution, however, as mentioned in the main text, we implement Pólya-Gamma data augmentation to avoid random walk Metropolis updates. The procedure relies on the following mixture identity:

$$\frac{(e^\psi)^a}{(1+e^\psi)^b} = 2^{-b}e^{m\psi}\int_0^\infty e^{-\xi\psi^2/2}\mathrm{PG}(\xi|b,0)d\xi\,,$$

where $m = a - b/2$ and $\mathrm{PG}(b,0)$ denotes the Pólya-Gamma distribution with shape $b$ and tilt $z = 0$. Using this, the full conditional distribution of $\psi_\ell^{(j,k)}$ can be augmented by a latent Pólya-Gamma variable $\xi_\ell^{(j,k)}$:

$$p(\psi_\ell^{(j,k)},\xi_\ell^{(j,k)}|-) \propto \mathrm{PG}\left(\xi_\ell^{(j,k)}\middle|\sum_{r=\ell}^{L} n_r^{(j,k)},0\right)\exp\left(\frac{-(\psi_\ell^{(j,k)}-\alpha_\ell^{(j)}-\beta_\ell^{(k)})^2}{2\sigma_\psi^2}\right)$$
$$\times\exp\left(m_\ell^{(j,k)}\psi_\ell^{(j,k)}-\frac{1}{2}\xi_\ell^{(j,k)}\psi_\ell^{(j,k)^2}\right),$$

where $m_\ell^{(j,k)} = n_\ell^{(j,k)} - \sum_{r=\ell}^{L} n_r^{(j,k)}/2$. We draw from this distribution using a Gibbs step, first sampling $\xi_\ell^{(j,k)}$ from its full conditional:

$$p(\xi_\ell^{(j,k)}|\psi_\ell^{(j,k)},\boldsymbol{z}) \propto \mathrm{PG}\left(\xi_\ell^{(j,k)}\middle|\sum_{r=\ell}^{L} n_r^{(j,k)},0\right)\exp\left(\frac{1}{2}\xi_\ell^{(j,k)}\psi_\ell^{(j,k)^2}\right).$$

This is a conjugate relationship, proportional to a $\text{PG}(b^*, z^*)$ distribution with shape $b^* = \sum_{r=\ell}^{L} n_r^{(j,k)}$ and tilt $z^* = \psi_\ell^{(j,k)}$. Sampling from this distribution is not trivial, but efficient rejection samplers are available (Polson et al., 2013; Windle et al., 2014). The full conditional for $\psi_\ell^{(j,k)}$ given $\xi_\ell^{(j,k)}$ is then written as,

$$\begin{aligned} p(\psi_\ell^{(j,k)}|\xi_\ell^{(j,k)}, \boldsymbol{z}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\psi^2) &\propto p\left(\psi_\ell^{(j,k)}|\xi_\ell^{(j,k)}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\psi^2\right) p\left(\xi_\ell^{(j,k)}|\psi_\ell^{(j,k)}, \boldsymbol{z}\right) \\ &\propto \exp\left(\frac{-(\psi_\ell^{(j,k)} - \alpha_\ell^{(j)} - \beta_\ell^{(k)})^2}{2\sigma_\psi^2}\right) \exp\left(m_\ell^{(j,k)}\psi_\ell^{(j,k)} - \frac{1}{2}\xi_\ell^{(j,k)}\psi_\ell^{(j,k)^2}\right). \end{aligned}$$

With some algebraic manipulation, this can be shown to be proportional to a $\text{N}(m^*, v^*)$ distribution with $v^* = \left(1/\sigma_\psi^2 + \xi_\ell^{(j,k)}\right)^{-1}$ and $m^* = v^*\left(\frac{\alpha_\ell^{(j)} + \beta_\ell^{(k)}}{\sigma_\psi^2} + m_\ell^{(j,k)}\right)$. Finally, recall that the weights $\omega_\ell^{(j,k)}$ are updated deterministically from $\boldsymbol{\psi}$ by,

$$\omega_\ell^{(j,k)} = \frac{\exp(\psi_\ell^{(j,k)})}{1 + \exp(\psi_\ell^{(j,k)})} \prod_{r=\ell+1}^{L} \frac{1}{1 + \exp(\psi_r^{(j,k)})}.$$

# Appendix C

# Additional Illustrations of the MRP Mixture Model

Here we provide additional illustrations of results obtained from the DS-MRP mixture model presented in Chapter 3. We include figures and tables for both simulated and real data examples that were excluded from the main text for brevity.

## C.1 Simulated data examples

Recall from Chapter 3.5 that we generate three-state MRP point patterns under two scenarios, selecting values of $T$ such that approximately 500 events are generated. Figure C.1 shows posterior density estimates obtained from fitting the DS-MRP mixture model to a point pattern generated under scenario 1. Figures C.2 and C.3 show corresponding results obtained from fitting the IS-MRP mixture and parametric Weibull models, respectively. As mentioned in the main text, the DS-MRP mixture

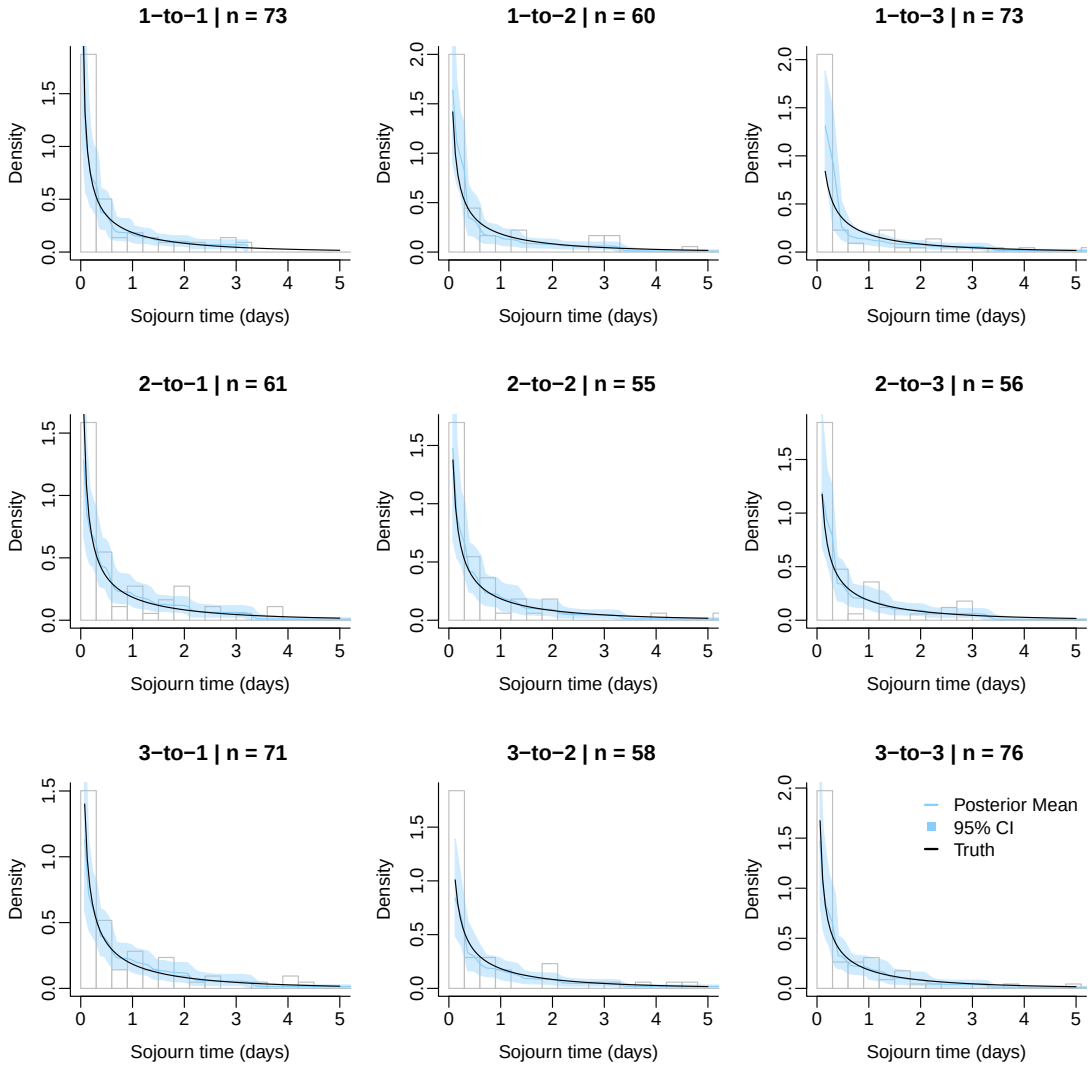


Figure C.1: Sojourn time density estimates and 95% credible intervals obtained from fitting the DS-MRP mixture model to a simulated point pattern generated under scenario 1.

provides a good fit, matching the generative sojourn time distributions. The IS-MRP mixture model also performs well, but has more variability in the uncertainty bands. The Weibull model, as expected, struggles slightly to capture the generative sojourn time density shape.

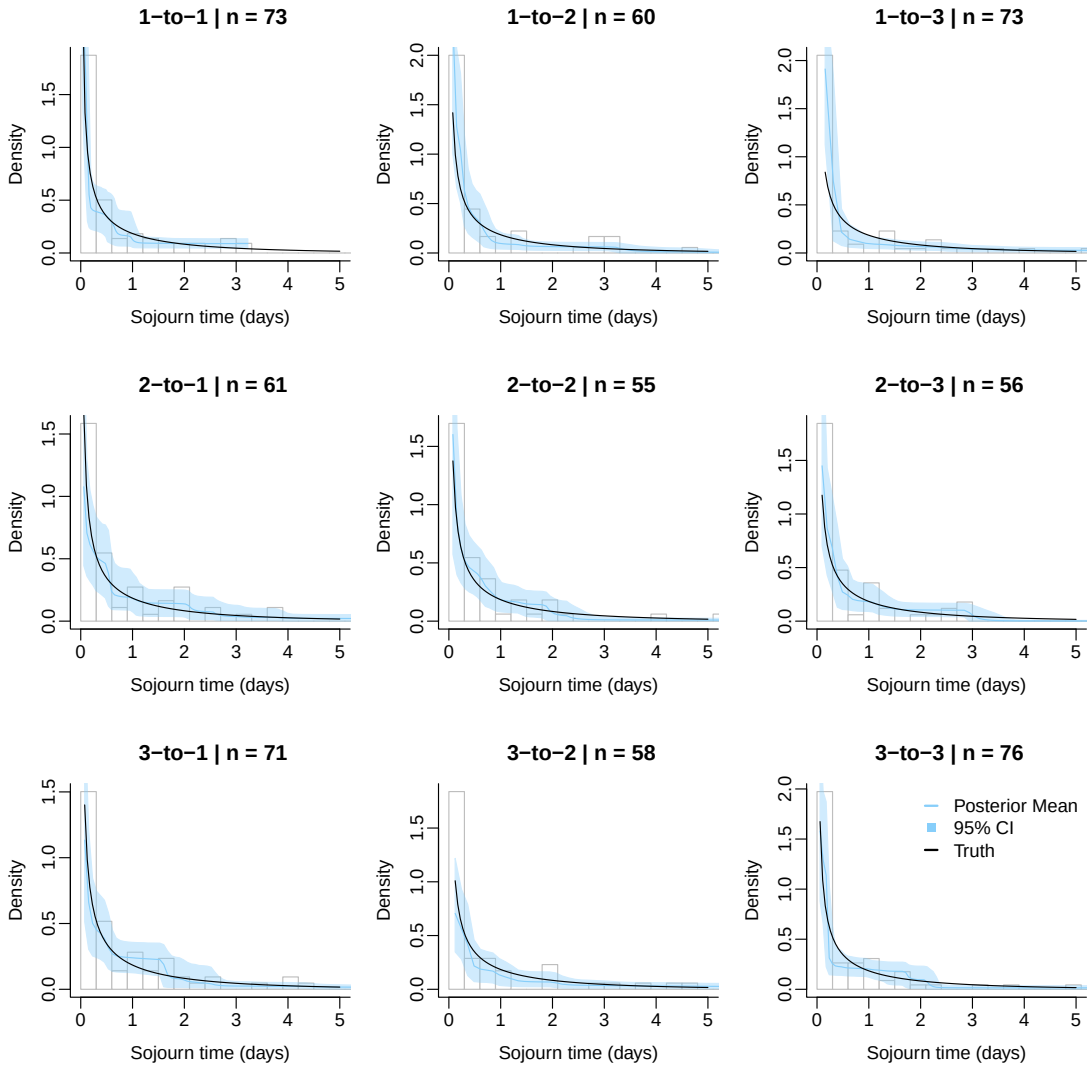


Figure C.2: Sojourn time density estimates and 95% credible intervals obtained from fitting the IS-MRP mixture model to a simulated point pattern generated under scenario 1.

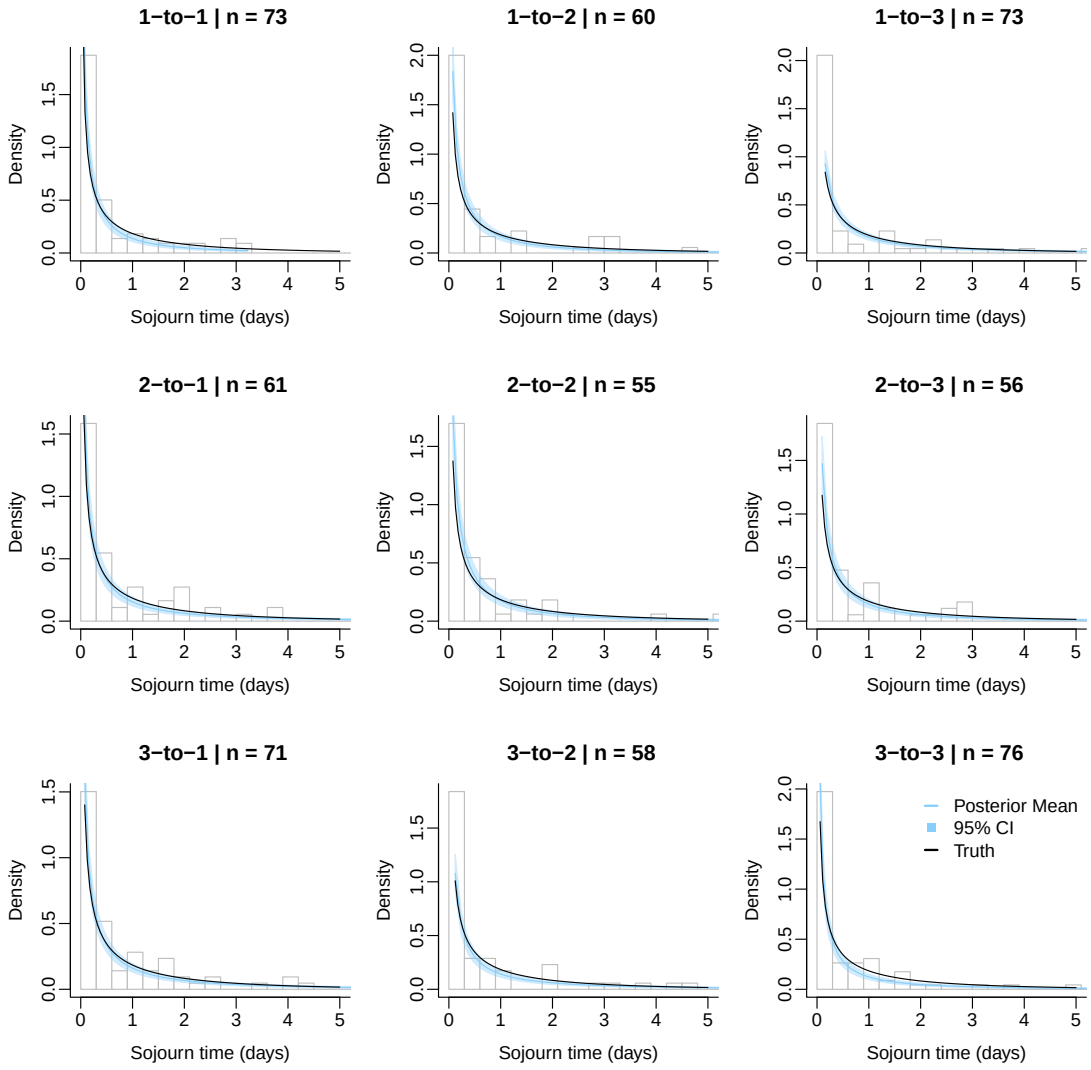


Figure C.3: Sojourn time density estimates and 95% credible intervals obtained from fitting the parametric Weibull mixture model to a simulated point pattern generated under scenario 1.

Table C.1: Predictive coverage rates and average interval scores for simulated sojourn times under Scenario 2 for the proposed DS-MRP model with uniform kernel, the corresponding IS-MRP mixture model, and the parametric Weibull model. Value pairs are separated by a backslash with the coverage rate as the first value and the average interval score as the second.

| Transition case | Uniform mixture DS-MRP | Uniform mixture IS-MRP | Parametric Weibull IS-MRP |
|---|---|---|---|
| 1-to-1 | 0.788 \ 6.16 | 0.864 \ 5.98 | 0.939 \ 9.12 |
| 1-to-2 | 0.955 \ 15.1 | 0.910 \ 15.0 | 0.940 \ 15.1 |
| 1-to-3 | 0.961 \ 10.6 | 0.941 \ 10.9 | 0.961 \ 10.6 |
| 2-to-1 | 0.828 \ 5.61 | 0.862 \ 4.45 | 0.914 \ 7.42 |
| 2-to-2 | 0.986 \ 9.40 | 0.942 \ 7.70 | 1.000 \ 9.19 |
| 2-to-3 | 0.930 \ 13.1 | 0.965 \ 13.4 | 0.965 \ 13.8 |
| 3-to-1 | 0.900 \ 8.54 | 0.883 \ 8.42 | 0.917 \ 8.91 |
| 3-to-2 | 0.974 \ 11.3 | 0.974 \ 11.4 | 0.974 \ 11.4 |
| 3-to-3 | 1.000 \ 8.39 | 0.927 \ 7.62 | 0.982 \ 7.95 |

Figures C.4, C.5, and C.6 show the posterior density estimates for the DS-MRP mixture, IS-MRP mixture, and Weibull IS-MRP models, respectively, for a point pattern generated under scenario 2. As with the previous scenario, the DS-MRP mixture model provides a good fit to the generative sojourn time distributions while the Weibull model struggles to capture the shape. We also include predictive coverage values and average interval scores for the three models in Table C.1, which were omitted from the main text.

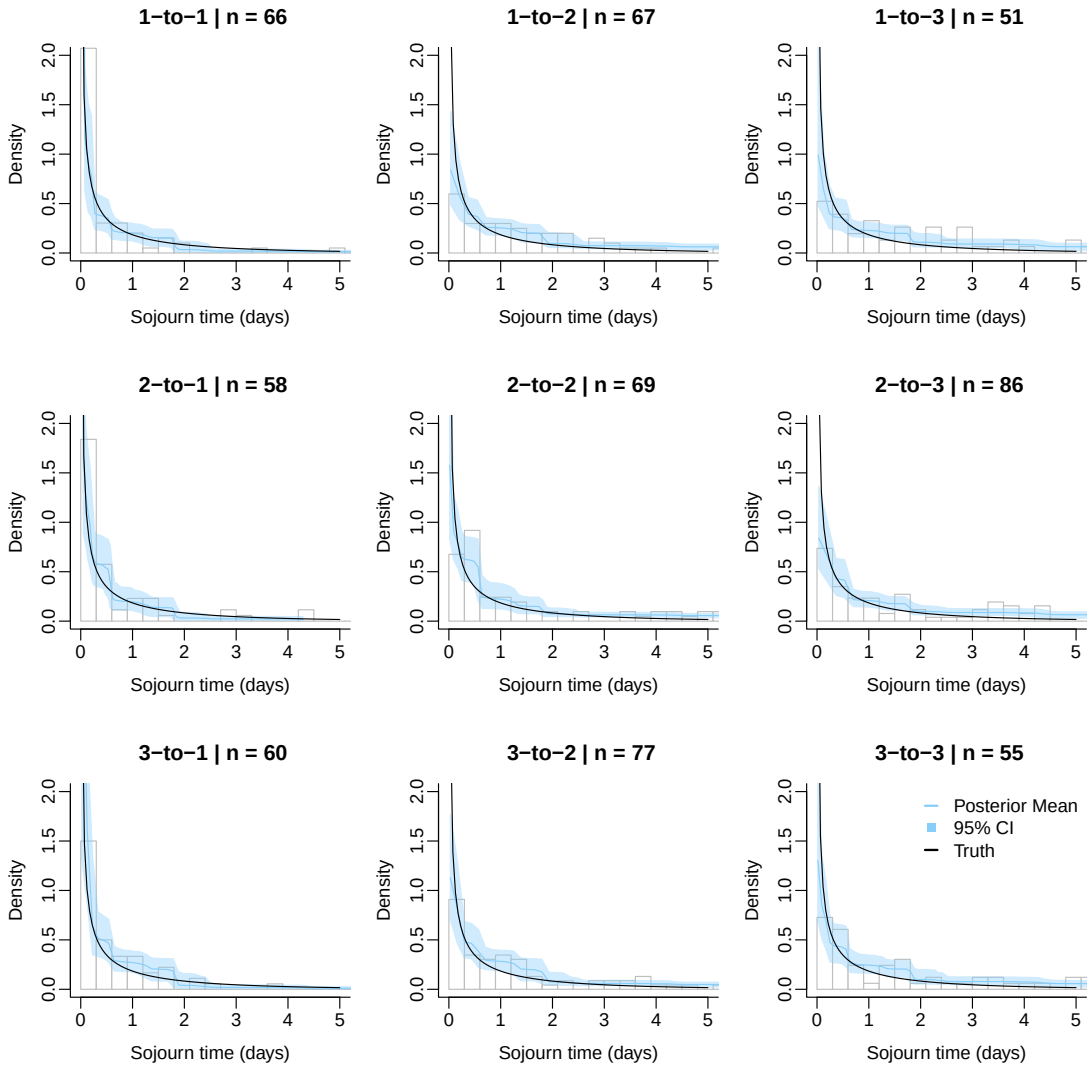


Figure C.4: Sojourn time density estimates and 95% credible intervals obtained from fitting the DS-MRP mixture model to a simulated point pattern generated under scenario 2.

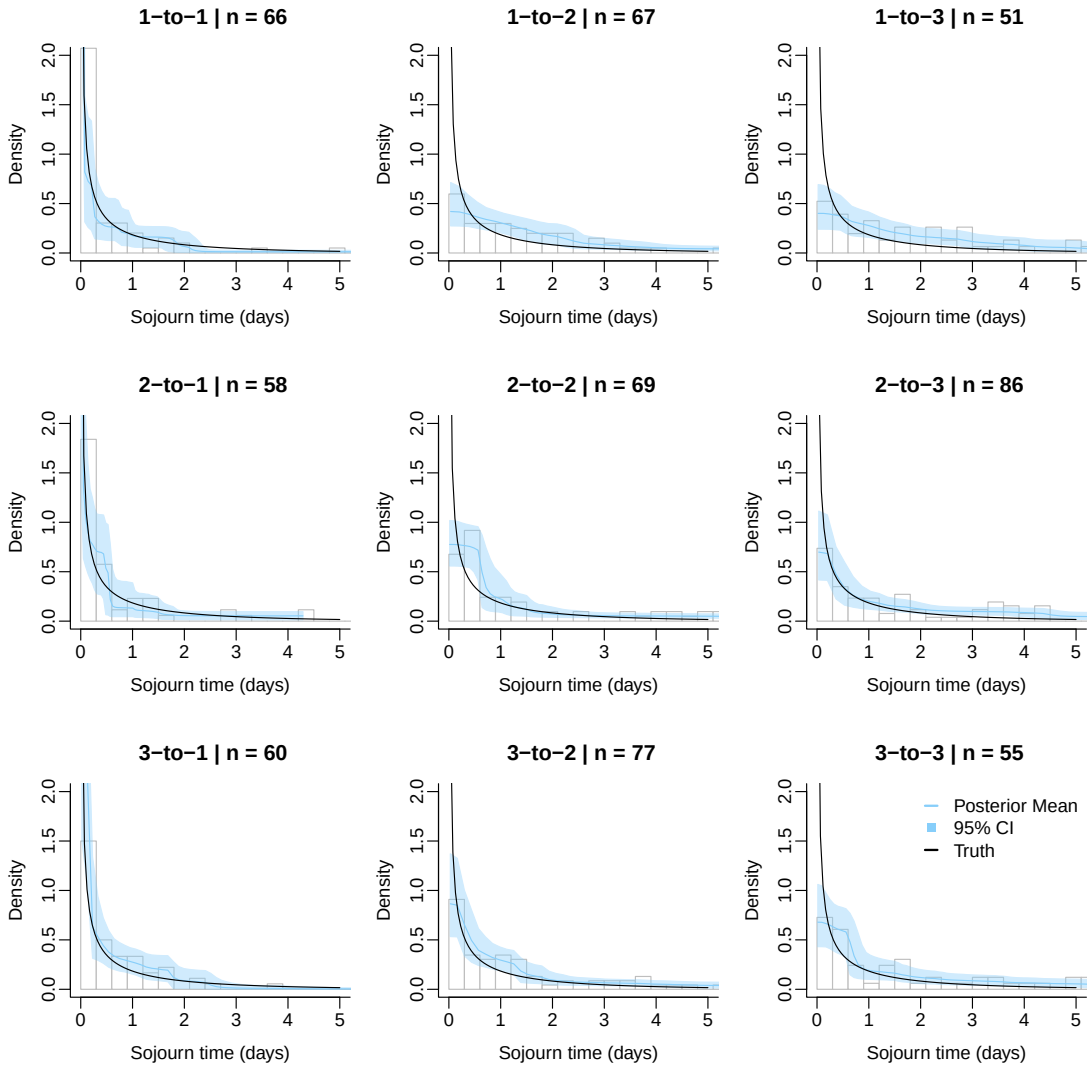


Figure C.5: Sojourn time density estimates and 95% credible intervals obtained from fitting the IS-MRP mixture model to a simulated point pattern generated under scenario 2.

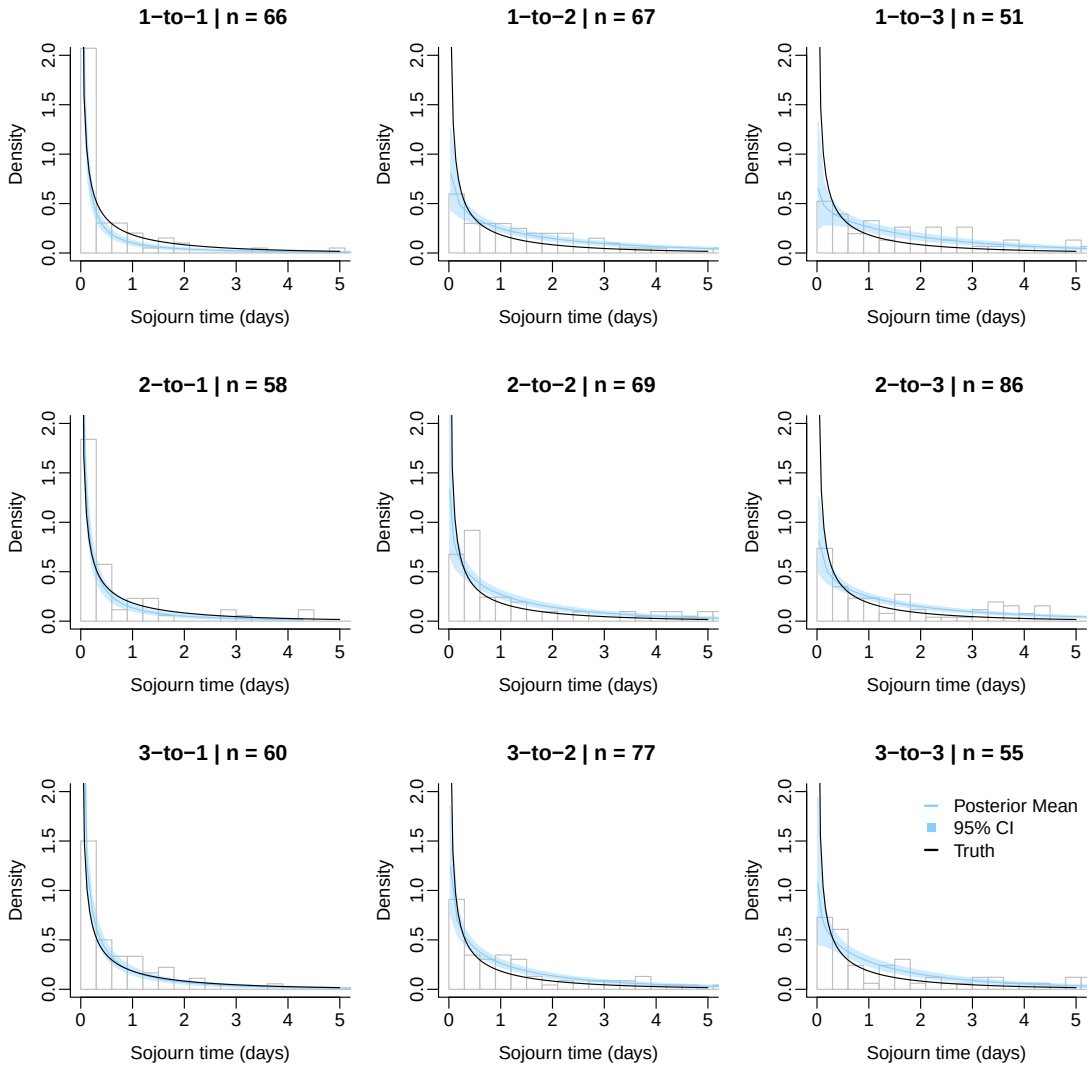


Figure C.6: Sojourn time density estimates and 95% credible intervals obtained from fitting the parametric Weibull mixture model to a simulated point pattern generated under scenario 2.

## C.2 Southern California earthquake data analysis

Chapter 3.6 presents results from fitting the DS-MRP mixture model to a dataset of Southern California earthquakes. To complete our analysis, we fit the IS-MRP mixture model and the parametric Weibull IS-MRP model to the same point pattern and compare results. We begin by showing the posterior density estimates for the two omitted models in Figures C.7 and C.8. With an uneven distribution of transition cases, the differences between models are more pronounced than in the simulated examples. As mentioned in the main text, the DS-MRP model recovers a strong fully common dependence structure, hence the general shape and uncertainty bands for that model are consistent across transition cases. In contrast, the IS-MRP mixture model fit is quite poor for the transition cases with high magnitude states, and estimates shapes are not consistent between cases.

Regarding posterior predictive performance, we compare predictive coverage diagrams for all three models in Figure C.9. Corresponding coverage percentage values and average intervals scores are provided in Table C.2. These results reinforce the conclusions drawn from the simulation studies. The Weibull model produces excessively large predictive intervals, especially for transitions with high magnitude states. The uniform mixture models perform better in terms of interval scores, but the intervals for the IS-MRP mixture are too narrow and have reduced coverage as a result. Even more so than with the simulations, the DS-MRP mixture model strikes an attractive balance between coverage and interval width for this point pattern.

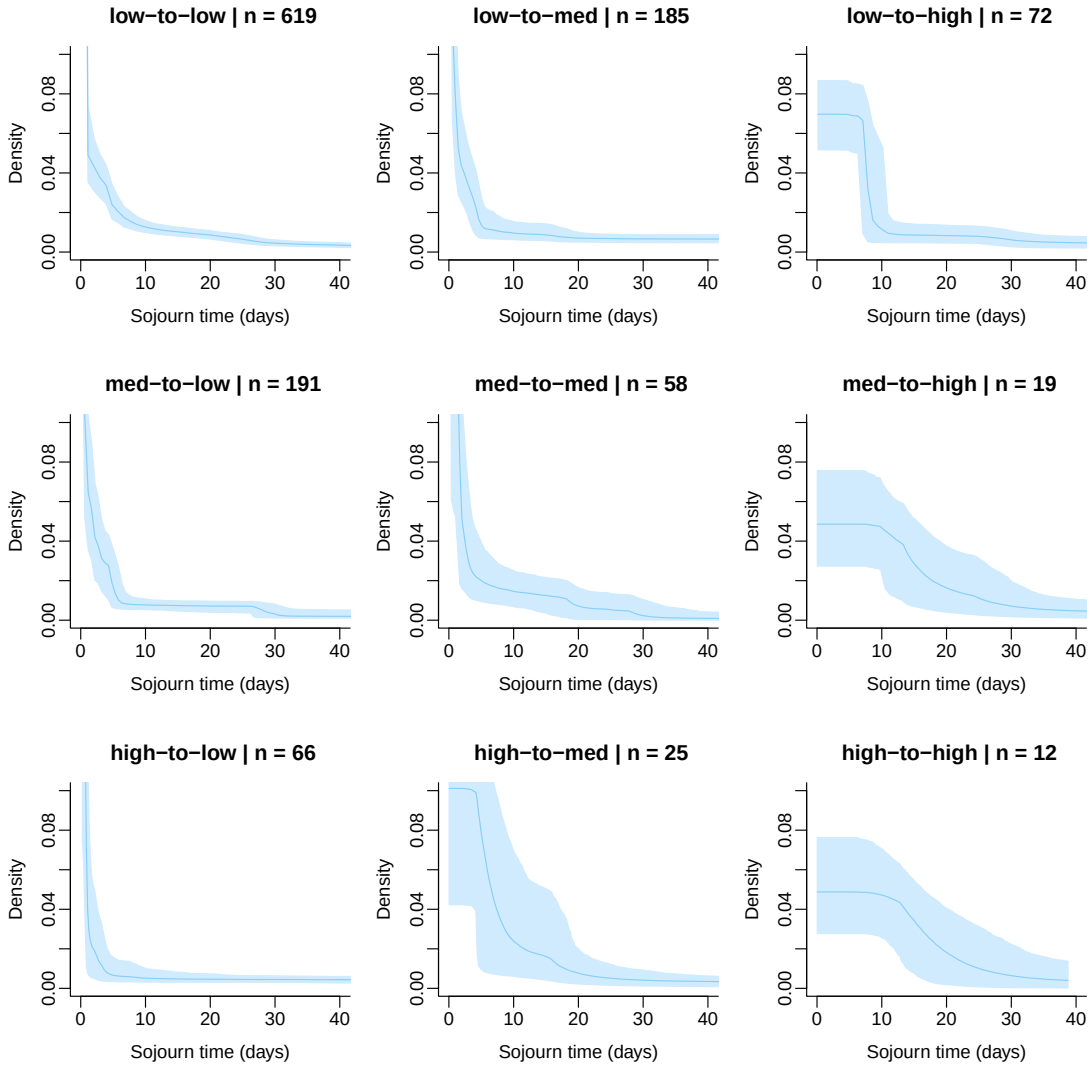


Figure C.7: Posterior estimates and 95% uncertainty intervals for the sojourn time densities of the Southern California earthquake dataset, estimated using the IS-MRP mixture model. Mean estimates are shown as solid lines with 95% credible intervals as shaded regions.

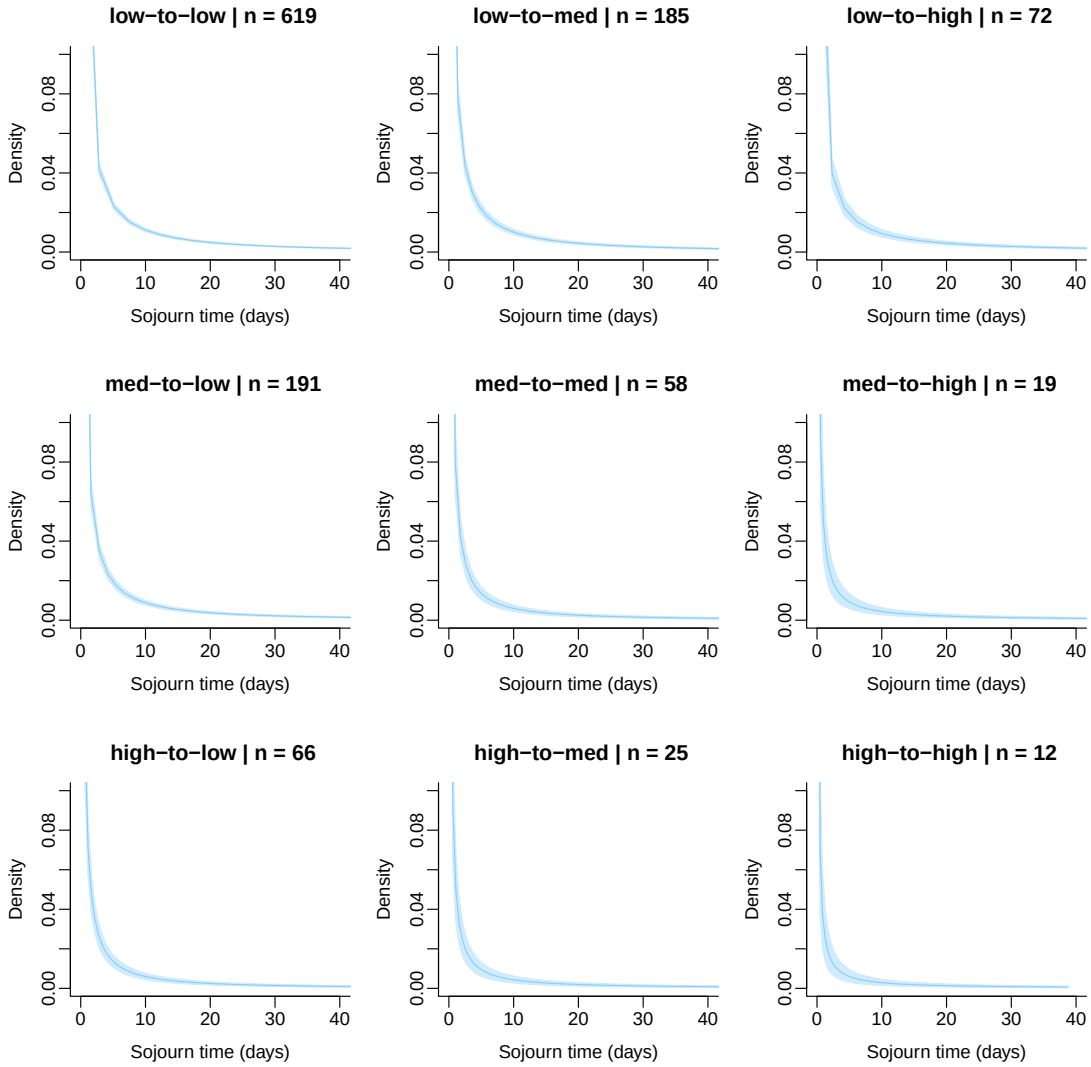


Figure C.8: Posterior estimates and 95% uncertainty intervals for the sojourn time densities of the Southern California earthquake dataset, estimated using the parametric Weibull MRP model. Mean estimates are shown as solid lines with 95% credible intervals as shaded regions.

Table C.2: Predictive coverage rates and average interval scores for the Southern California earthquake dataset, comparing the DS-MRP mixture model, the corresponding IS-MRP mixture model, and the parametric Weibull IS-MRP model. Value pairs are separated by a backslash with the coverage rate as the first value and the average interval score as the second.

| Transition case | DS-MRP mixture | IS-MRP mixture | Weibull IS-MRP |
|---|---|---|---|
| low-to-low | 0.940 \ 1.33 | 0.717 \ 1.23 | 0.997 \ 1.87 |
| low-to-med | 0.930 \ 1.10 | 0.849 \ 1.03 | 1.000 \ 2.65 |
| low-to-high | 0.944 \ 1.98 | 0.694 \ 1.81 | 1.000 \ 11.0 |
| med-to-low | 0.937 \ 1.01 | 0.832 \ 1.23 | 1.000 \ 1.70 |
| med-to-med | 0.914 \ 0.74 | 0.638 \ 0.73 | 1.000 \ 2.67 |
| med-to-high | 1.000 \ 0.57 | 0.421 \ 0.55 | 1.000 \ 400.1 |
| high-to-low | 0.970 \ 0.53 | 0.742 \ 0.50 | 1.000 \ 1.76 |
| high-to-med | 0.880 \ 0.54 | 0.280 \ 0.57 | 1.000 \ 13.0 |
| high-to-high | 0.750 \ 0.54 | 0.250 \ 0.54 | 1.000 \ 15230.4 |

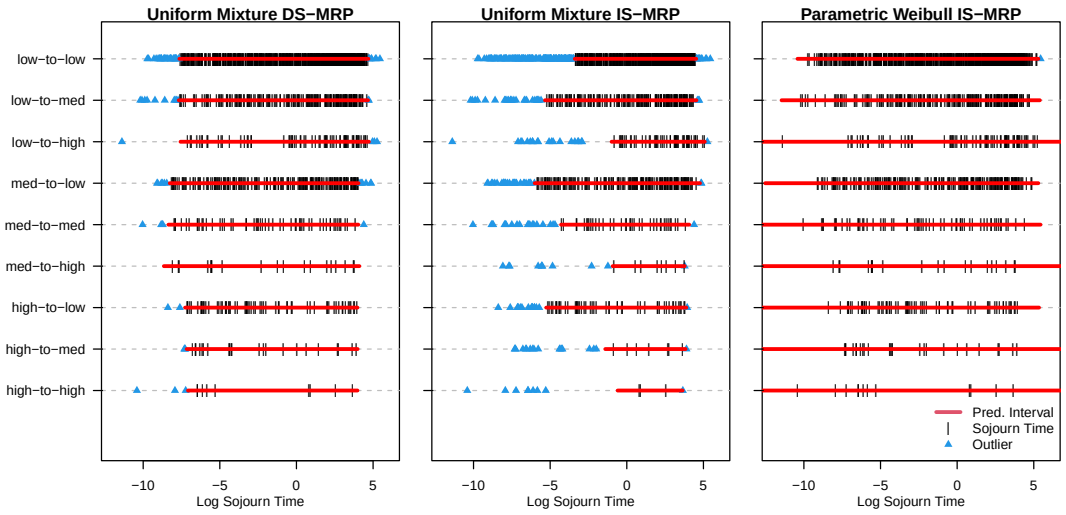


Figure C.9: Posterior predictive coverage diagrams for the Southern California earthquake dataset, fit using the DS-MRP mixture model (left), IS-MRP mixture model (middle), and parametric Weibull MRP model (right). Red lines are used to indicate the 95% predictive intervals for each transition case. Data are partitioned by transition case and are shown as blue triangles if excluded from the 95% predictive intervals and as black ticks if included.

Table C.3: Predictive state recurrence ECDF errors for the southern California earthquake dataset. Posterior mean estimates and 95% uncertainty bounds are given for the uniform mixture DS-MRP model, the corresponding IS-MRP mixture, and the parametric Weibull model.

| MRP Model | Initial State | Mean ECDF Error | 95% Interval |
|---|---|---|---|
| DS-MRP Mixture | Low | 0.10 | (0.01, 0.31) |
| | Med | 2.87 | (1.24, 5.53) |
| | High | 7.02 | (2.10, 17.64) |
| IS-MRP Mixture | Low | 0.16 | (0.04, 0.41) |
| | Med | 3.55 | (1.55, 6.70) |
| | High | 20.92 | (3.19, 22.37) |
| Weibull IS-MRP | Low | 1.68 | (0.27, 5.42) |
| | Med | 20.92 | (0.85, 74.21) |
| | High | 95.39 | (3.64, 1128.66) |

Continuing with model comparison, we examine the predictive state recurrence ECDFs for each model, shown graphically in Figure C.10. Numerical summaries of the ECDF Error posterior distributions are provided in Table C.3. In all cases, the DS-MRP mixture model outperforms both the IS-MRP mixture and parametric Weibull IS-MRP models, with the Weibull performing the worst. One potential reason for the poor performance of the Weibull distribution is found in the simulated predictive point patterns, which have far fewer events than the observed data (approximately 500 events versus 1,200). Indeed, around 2% of posterior predictive point patterns have no observed events with the high magnitude state, hence the very wide predictive intervals in the ECDF plot. Overall, these results indicate that the DS-MRP mixture model is not only a good fit to the Southern California earthquake data, but performs better than the other models in terms of visual results and predictive performance.

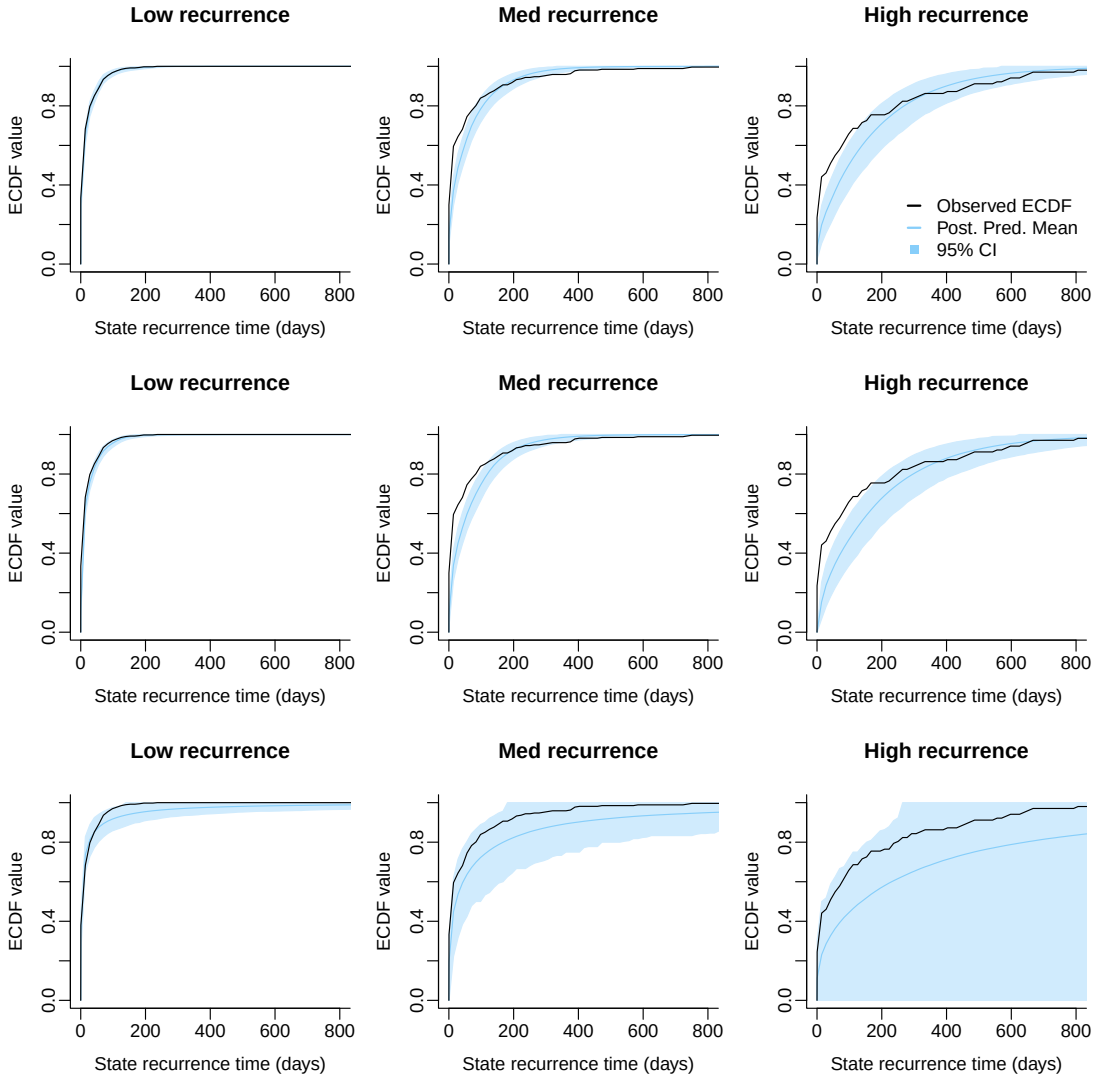


Figure C.10: Predictive state recurrence ECDFs for the Southern California earthquake dataset, fit using the DS-MRP mixture model (top), IS-MRP mixture model (middle), and parametric Weibull MRP model (bottom).

# Appendix D

# Posterior Computation for the Log-logistic Hazard Basis Model

Here we show posterior computation details for the log-logistic HRP hazard basis model presented in Chapter 4.1. Most model parameters have full conditional distributions that are straightforward to sample from. For the remaining parameters, we use random walk Metropolis updates with a log-normal proposal distribution. We implement the adaptive batching scheme of Roberts and Rosenthal (2009) to tune the proposal variance, targeting an acceptance rate between 35% and 50%. As a reminder from Chapter 4.1.4, we use intermediary variables $\boldsymbol{\mu}$, $\Sigma$, and $V$ to simplify notation.

**Sampling $z_i$**

The full conditional for $z_i$ is given by $\Pr(z_i = \ell|\boldsymbol{\omega}, \boldsymbol{x}) \propto \omega_\ell h_{\mathcal{B}}(x_i|b_{z_i}, \theta)$ where $h_{\mathcal{B}}$ is the log-logistic hazard function. This is discretely valued with $\ell \in \{1, \ldots, L\}$ and can be

sampled directly.

**Sampling $\boldsymbol{\beta}$**

The full conditional for $\boldsymbol{\beta}$ is given by,

$$p(\boldsymbol{\beta}|\boldsymbol{\omega}, \sigma^2, \tau) \propto p(\boldsymbol{\beta})p(\boldsymbol{\omega}|\boldsymbol{\beta}, \sigma^2, \tau) \propto \mathrm{N}_2(\boldsymbol{\beta}|\boldsymbol{m}_\beta, s_\beta^2 \boldsymbol{I})\mathrm{N}_L(\log(\boldsymbol{\omega})|B\boldsymbol{\beta}, \Sigma)\,,$$

where $B$ is the $L \times 2$ matrix containing a column of ones and a column of $\log(b_\ell)$, $\boldsymbol{m}_\beta = (R/L, 0)$ is the prior mean, $s_\beta^2$ is the prior variance, and $\boldsymbol{I}$ is the identity matrix. This is a standard multivariate normal conjugate update. The full conditional for $\boldsymbol{\beta}$ is a $N_2(\boldsymbol{m}^*, S^*)$ distribution with covariance $S^* = \left[(1/s_\beta^2)\boldsymbol{I} + B^T\Sigma^{-1}B\right]^{-1}$ and mean $\boldsymbol{m}^* = S^*\left[(1/s_\beta^2)\boldsymbol{m}_\beta + B^T\Sigma^{-1}\log(\boldsymbol{\omega})\right]$.

**Sampling $\sigma^2$**

The full conditional for $\sigma^2$ is given by,

$$\begin{aligned} p(\sigma^2|\boldsymbol{\omega}, \boldsymbol{\beta}, \tau) &\propto p(\sigma^2)p(\boldsymbol{\omega}|\boldsymbol{\beta}, \sigma^2, \tau) \\ &\propto (\sigma^2)^{-a_\sigma - 1}\exp\left(-\frac{b_\sigma}{\sigma^2}\right)(\sigma^2)^{-L/2}\exp\left(-\frac{1}{2\sigma^2}(\log(\boldsymbol{\omega}) - \boldsymbol{\mu})^T V^{-1}(\log(\boldsymbol{\omega}) - \boldsymbol{\mu})\right)\,. \end{aligned}$$

This is proportional to an inv-Ga$(a^*, b^*)$ distribution with $a^* = a_\sigma + L/2$ and $b^* = b_\sigma + \frac{1}{2}(\log(\boldsymbol{\omega}) - \boldsymbol{\mu})^T V^{-1}(\log(\boldsymbol{\omega}) - \boldsymbol{\mu})$. Sampling from this distribution is straightforward.

**Sampling $\tau$**

The full conditional for $\tau$ is given by,

$$p(\tau|\boldsymbol{\omega}, \boldsymbol{\beta}, \sigma^2) \propto p(\tau)p(\boldsymbol{\omega}|\boldsymbol{\beta}, \sigma^2, \tau) \propto \mathrm{Exp}(\tau|R/L)\mathrm{N}_L(\log(\boldsymbol{\omega})|\boldsymbol{\mu}, \Sigma)\,,$$

where $\text{Exp}(\tau|\theta)$ denotes the exponential distribution with mean $\theta$. This is not an easily sampled density, so we use a random walk Metropolis step with log-normal proposal distribution and an adaptive proposal variance, following the batching scheme of Roberts and Rosenthal (2009).

**Sampling $\boldsymbol{\omega}$**

Following our presentation in Chapter 4.1.4, we make use of auxiliary variables $\boldsymbol{u} = \{u_1, \ldots, u_L\}$ to sample the full conditional for $\boldsymbol{\omega}$. The joint full conditional is given by,

$$p(\boldsymbol{\omega}, \boldsymbol{u}, |\boldsymbol{z}, \boldsymbol{\beta}, \sigma^2, \tau) \propto p(\boldsymbol{\omega}|\boldsymbol{\beta}, \sigma^2, \tau) \prod_{i=1}^{n} \mathbb{I}(u_\ell < \text{Ga}(\omega_\ell | n_\ell + 1, (H_\ell^*)^{-1}))\,,$$

which we sample from using a Gibbs step. Immediate from this expression is the uniform full conditional for $\boldsymbol{u}$, which can be sample directly. To draw the weights $\boldsymbol{\omega}$ given $\boldsymbol{u}$, we invert the indicator function inequalities such that they take the form $\mathbb{I}(C_\ell < \omega_\ell < D_\ell)$. The values of $C_\ell$ and $D_\ell$ correspond to the locations where the above gamma density is equal to $u_\ell$, which can be calculated using the Lambert-W function.

The Lambert-W function $W_k(y) = x$ is a complex-valued function that gives solutions to the equation $y = xe^x$. Although not available in closed form, values of $W_k(y)$ can be computed using widely available numerical techniques. The real-valued solutions lie along the so-called principal branch with $k = 0$, and another branch with $k = -1$. For our purposes, the solution for $x$ for $y = \text{Ga}(x|\kappa, \lambda)$ is given by,

$$x = -\lambda(\kappa - 1) W_k \left( \frac{-1}{\kappa - 1} \left(\lambda \Gamma(\kappa) y\right)^{\frac{1}{\kappa - 1}} \right) .$$

For $\kappa > 1$, the lower bound $C_\ell$ is given by the primary branch $W_0$ and the upper bound

$D_\ell$ is given by the $W_{-1}$ branch. When $0 < \kappa < 1$, the density is strictly decreasing, thus the lower bound $c_\ell = 0$ and the only solution for $D_\ell$ is given by the primary branch $W_0$. If $\kappa = 1$, the density is that of an exponential distribution, which has closed expressions for the bounds $C_\ell = 0$ and $D_\ell = -\lambda \log(y\lambda)$.

With values determined for $C_\ell$ and $D_\ell$, the full conditional for $\omega_\ell$ is a multivariate log-normal distribution with mean parameter $\mu$ and covariance $\Sigma$, truncated such that each $\omega_\ell$ lies in the interval $(C_\ell, D_\ell)$. As discussed in Chapter 4.1.4, we sample from this distribution using the harmonic-HMC algorithm of Pakman and Paninski (2014).

# Appendix E

# Posterior Computation for the Nonparametric Modulated Renewal Process Model

Here we show posterior computation details the nonparametric modulated renewal process model presented in Chapter 4.2. As an application of the log-logistic hazard basis model, the sampling details for the hazard function parameters are similar. Indeed, the full conditionals for $\boldsymbol{\beta}$, $\sigma^2$, and $\tau$ given $\boldsymbol{\omega}^h$ are identical to those given in Appendix D. The full conditional for $\boldsymbol{\omega}^h$ is slightly different, given by,

$$p(\boldsymbol{\omega}^h|\boldsymbol{\beta},\sigma^2,\tau,\boldsymbol{z})\prod_{\ell=1}^{L}\text{Ga}\left(\omega_\ell^h\middle|n_\ell^h+1,\left\{\sum_{m=1}^{M}\omega_m^f H_{m,\ell}^*\right\}^{-1}\right).$$

This expression yields different values for the bounds $C_\ell$ and $D_\ell$, but otherwise the sampling procedure remains the same. What remains to be shown in this appendix are sampling details for the modulating function parameters $\boldsymbol{\omega}^f$ and $\alpha$, as well as the

bivariate membership variables $\boldsymbol{z}$.

**Sampling $z_i$**

The full conditional for $z_i$ is given by $\Pr(z_i = (m, \ell)|\boldsymbol{\omega}^f, \boldsymbol{\omega}^h, \boldsymbol{x}) \propto \omega^f_m \omega^h_\ell h^*_{m,\ell}(t_i, x_i)$. This is discretely valued and can be sampled directly.

**Sampling $\alpha$**

The full conditional for $\alpha$ is given by,

$$p(\alpha|\boldsymbol{\omega}^f) \propto p(\alpha)p(\boldsymbol{\omega}^f|\alpha) \propto \text{Ga}(\alpha|a_\alpha, b_\alpha)\text{Dir}_L(\boldsymbol{\omega}^f|\alpha/M, \ldots, \alpha/M)\,.$$

This is not an easily sampled density, so we use a random walk Metropolis step with log-normal proposal distribution and an adaptive proposal variance, following the batching scheme of Roberts and Rosenthal (2009).

**Sampling $\boldsymbol{\omega}^f$**

The full conditional for $\boldsymbol{\omega}^f$ is given by,

$$\begin{aligned} p(\boldsymbol{\omega}^f|\boldsymbol{\omega}^h, \alpha, \boldsymbol{z}, \boldsymbol{x}) &\propto p(\boldsymbol{z}|\boldsymbol{\omega}^f, \boldsymbol{\omega}^h)p(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{\omega}^f, \boldsymbol{\omega}^h)p(\boldsymbol{\omega}^f|\alpha) \\ &\propto \text{Dir}_M\left(\boldsymbol{\omega}^f|n^f_1 + \alpha/M, \ldots, n^f_M + \alpha/M\right) \prod_{m=1}^{M} \exp\left(-\omega^f_m \sum_{\ell=1}^{L} \omega^h_\ell H^*_{m,\ell}\right)\,. \end{aligned}$$

Similar to the hazard basis weights, we introduce uniform variables $\boldsymbol{u}^f = \{u^f_1, \ldots, u^f_M\}$ to sample from this full conditional, using a Gibbs step to draw the weights $\boldsymbol{\omega}^f$ given $\boldsymbol{u}^f$ and vice versa. The full conditional for $u^f_m$ is a uniform distribution over the interval from 0 to $\exp\left(-\omega^f_m \sum_{\ell=1}^{L} \omega^h_\ell H^*_{m,\ell}\right)$. The full conditional for $\boldsymbol{\omega}^f$ given $\boldsymbol{u}^f$ is

a $\text{Dir}_M\left(n_1^f + \alpha/M + u_1^f, \ldots, n_M^f + \alpha/M + u_M^f\right)$ distribution, truncated such that each $\omega_m^f$ is bounded above by $\frac{-\log(u_m^f)}{\sum_{\ell=1}^{L} \omega_\ell^h H_{m,\ell}^*}$.

Sampling from a truncated Dirichlet distribution is challenging. We find that updating each element using a Gibbs step is reasonably tractable. Consider a general set of weights $\{\omega_1, \ldots, \omega_M\}$ arising from a truncated Dirichlet distribution with parameters $\{\alpha_1, \ldots, \alpha_M\}$, lower bounds $\{C_1, \ldots, C_M\}$, and upper bounds $\{D_1, \ldots, D_M\}$. Such a distribution has non-empty support if and only if $0 \leq C_m < D_m \leq 1$ for all $m$, and $\sum_{m=1}^{M} C_m < 1 < \sum_{m=1}^{M} D_m$. To derive the full conditional of a single element, we must account for the sum constraint. As such, we explicitly specify that $\omega_M = 1 - \sum_{m=1}^{M-1} \omega_m$. Without loss of generality, consider the distribution of $\omega_1$ given $\omega_2, \ldots, \omega_{M-1}$. The sum constraint implies that $\omega_1 < 1 - \sum_{m=2}^{M-1} \omega_m$. Notice that $1 - \sum_{m=2}^{M-1} \omega_m = \omega_1 + \omega_M = \omega_1^*$ is a fixed quantity, representing the total available weight for $\omega_1$ given the other weights. Indeed, for an unconstrained Dirichlet distribution, the conditional distribution is given by,

$$\frac{\omega_1}{\omega_1^*} \big| \omega_2, \ldots, \omega_{M-1} \sim \text{Beta}(\alpha_1, \alpha_M)\,.$$

Next we consider the impact of the truncation bounds, of which two are applicable for $\omega_1$, namely $C_1 < \omega_1 < D_1$ and $C_M < \omega_M < D_M$. Combining both sets of constraints, the boundaries applied to the above conditional distribution are expressed as,

$$\max\left(\frac{C_1}{\omega_1^*}, 1 - \frac{D_M}{\omega_1^*}\right) < \frac{\omega_1}{\omega_1^*} < \min\left(\frac{D_1}{\omega_1^*}, 1 - \frac{C_M}{\omega_1^*}\right)\,.$$

Putting this back in the context of our Gibbs step for the intensity weights, each weight $\omega_m^f$ can be drawn conditionally through the following procedure:

1. Calculate the (current) available weight $\omega_m^* = \omega_m + \omega_M$.

2. Find the conditional lower bound $C^* = \max\left\{0, 1 + \frac{\log(u_M^f)}{\omega_m^* \sum_{\ell=1}^{L} \omega_\ell^h H_{m,\ell}^*}\right\}$

3. Find the conditional upper bound $D^* = \min\left\{1, \frac{-\log(u_m^f)}{\omega_m^* \sum_{\ell=1}^{L} \omega_\ell^h H_{m,\ell}^*}\right\}$

4. Draw a value $y_m$ from a Beta$(n_m^f + \alpha/M, n_M^f + \alpha/M)$ distribution truncated between $C^*$ and $D^*$.

5. Save $\omega_m^f = y_m \omega_m^*$ and update $\omega_M^f = (1 - y_m)\omega_m^*$.

# References

Altinok, Y. and Kolcak, D. (1999). "An application of the semi-Markov model for earthquake occurrences in North Anatolia, Turkey." *Journal of the Balkan Geophysical Society*, 2:90–99.

Alvarez, E. E. (2005). "Estimation in stationary Markov renewal processes, with application to earthquake forecasting in Turkey." *Methodology and Computing in Applied Probability*, 7:119–130.

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

Arbel, J., Lijoi, A., and Nipoti, B. (2016). "Full Bayesian inference with hazard mixture models." *Computational Statistics and Data Analysis*, 93:359–372.

Arfè, A., Peluso, S., and Muliere, P. (2021). "The semi-Markov beta-Stacy process: a Bayesian non-parametric prior for semi-Markov processes." *Statistical Inference for Stochastic Processes*, 24:1–15.

Berman, M. (1981). "Inhomogeneous and modulated gamma processes." *Biometrika*, 68:143–152.

Brunner, L. J. and Lo, A. Y. (1989). "Bayes methods for a symmetric unimodal density and its mode." *The Annals of Statistics*, 17:1550–1566.

Bulla, P. and Muliere, P. (2007). "Bayesian nonparametric estimation for reinforced Markov renewal processes." *Statistical Inference for Stochastic Processes*, 10:283–303.

Camerlenghi, F., Lijoi, A., and Prünster, I. (2021). "Sirvova; analysis via hierarchically dependent mixture hazards." *Annals of Statistics*, 49:863–884.

Çinlar, E. (1975). "Markov renewal theory: a survey." *Management Science*, 21:727–752.

Chou, M., Parlar, M., and Zhou, Y. (2017). "Optimal timing to initiate medical treatment for a disease evolving as a semi-Markov process." *Journal of Optimization Theory and Applications*, 175:194–217.

Cluff, L. S., Patwardhan, A. S., and Coppersmith, K. J. (1980). "Estimating the probability of occurrences of surface faulting earthquakes on the Wasatch fault zone, Utah." *Bulletin of the Seismological Society of America*, 70:1463–1478.

Comte, F. and Duval, C. (2018). "Statistical inference for renewal processes." *Scandinavian Journal of Statistics*, 45:164–193.

Corral, A. (2005). "Time-decreasing hazard and increasing time until the next earthquake." *Physical Review E*, 71:017101.

Cox, D. (1972). "The statistical analysis of dependencies in point processes." In Lewis, P. (ed.), *Stochastic point processes.* (pp. 55–66). New York: Wiley.

Cox, D. and Isham, V. (1980). *Point Processes.* London: Chapman and Hall.

Cox, D. R. (1962). *Renewal Theory.* London: Methuen.

Dabrowska, D. and Ho, W. (2006). "Estimation in a semiparametric modulated renewal process." *Statistica Sinica*, 16:93–119.

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

de Boor, C. (1978). *A Practical Guide to Splines.* New York: Springer.

DeVore, R. and Lorentz, G. (1993). *Constructive Approximation.* Grundlehren der mathematischen Wissenschaften. Berlin: Springer.

Dinse, G. E. and Larson, M. G. (1986). "A note on semi-Markov models for partially censored Data." *Biometrika*, 73:379–386.

Doksum, K. (1974). "Tailfree and neutral random probabilities and their posterior distributions." *The Annals of Probability*, 2:183–201.

Doob, J. L. (1948). "Renewal theory from the point of view of the theory of probability." *Transactions of the American Mathematical Society*, 63:422–438.

Dykstra, R. L. and Laud, P. (1981). "A Bayesian nonparametric approach to reliability." *Annals of Statistics*, 9:356–367.

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.

Feller, W. (1941). "On the integral equation of renewal theory." *The Annals of Mathematical Statistics*, 12:243–267.

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

Ferguson, T. S. and Phadia, E. G. (1979). "Bayesian nonparametric estimation based on censored data." *The Annals of Statistics*, 7:163–186.

Garavaglia, E. and Pavani, R. (2011). "About earthquake forecasting by Markov renewal processes." *Methodology and Computing in Applied Probability*, 13:155–169.

Gneiting, T. and Raftery, A. E. (2007). "Strictly proper scoring rules, prediction, and estimation." *Journal of the American Statistical Association*, 102:359–378.

Greenwood, P. E. and Wefelmeyer, W. (1996). "Empirical estimators for semi-Markov processes." *Mathematical Methods in Statistics*, 5:299–315.

Hawkes, A. G. (1971). "Spectra of some self exciting and mutually exciting point processes." *Biometrika*, 58:83–90.

Hawkes, A. G. and Oakes, D. (1974). "A cluster process representation of a self-exciting process." *Journal of Applied Probability*, 11:493–503.

Hjort, N. L. (1990). "Nonparametric Bayes estimators based on beta processes in models for life history data." *The Annals of Statistics*, 18:1259–1294.

Howard, R. (1971). *Dynamic Probabilistic Systems, Volume II: Semi-Markov and Decision Processes*. New York: Wiley.

Hutton, K., Woessner, J., and Hauksson, E. (2010). "Earthquake Monitoring in Southern California for Seventy-Seven Years (1932–2008)." *Bulletin of the Seismological Society of America*, 100:423—-446.

Ishwaran, H. and James, L. F. (2001). "Gibbs sampling methods for stick-breaking priors." *Journal of the American Statistical Association*, 96:161–173.

Ishwaran, H. and James, L. F. (2004). "Computational methods for multiplicative intensity models using weighted gamma processes." *Journal of the American Statistical Association*, 99:175–190.

James, L. F. (2005). "Bayesian Poisson process partition calculus with applications Bayesian Lévy moving averages." *The Annals of Statistics*, 33:1771–1799.

Jarrett, R. G. (1979). "A note on the intervals between coal-mining disasters." *Biometrika*, 66:191–193.

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

Kim, H. (2021). *Flexible Bayesian Modeling and Inference Methods for Hawkes Pro-*

*cesses*. Doctoral dissertation, University of California, Santa Cruz, CA. Retrieved from https://escholarship.org/uc/item/7b05w3hf.

Kim, H. and Kottas, A. (2022). "Erlang mixture modeling for Poisson process intensities." *Statistics and Computing*, 32:1–15.

Kobbacy, K., Fawzi, B., Percy, D., and Ascher, H. (1998). "A full history proportional hazards model for preventative maintenance scheduling." *Quality and Reliability Engineering International*, 13:187–198.

Kumar, P., Korkolis, E., Benzi, R., Denisov, D., Niemeijer, A., Schall, P., Toschi, F., and Trampert, J. (2020). "On interevent time distributions of avalanche dynamics." *Scientific Reports*, 10:1–11.

Lagakos, S. W., Sommer, C. J., and Zelen, M. (1978). "Semi-Markov models for partially censored data." *Biometrika*, 65:311–317.

Lawless, J. and Thiagarajah, K. (1996). "A point-process model incorporating renewals and time trends, with application to repairable systems." *Technometrics*, 38:131–138.

Le, M., Martin, R., and Raftery, A. (1996). "Modeling flat stretches, bursts, and outliers in time series using mixture transition distribution models." *Journal of the American Statistical Association*, 91:1504–1515.

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.

Lévy, P. (1954). "Processus semi-Markoviens." In Gerretsen, J. C. H. and de Groot, J. (eds.), *Proceedings of the International Congress of Mathematicians: 1954 Volume III.* (pp. 416–426). Amsterdam: North-Holland Publishing Company.

Lijoi, A. and Nipoti, B. (2014). "A class of hazard rate mixtures for combining survival data from different experiments." *Journal of the American Statistical Association*, 109:802–814.

Limnios, N. and Oprişan, G. (2001). *Semi-Markov Processes and Reliability.* Boston: Birkhäuser.

Lin, D. and Ying, Z. (1994). "Semiparametric analysis of the additive risk model." *Biometrika*, 81:61–71.

Lin, F. and Fine, J. (2009). "Pseudomartingale estimating equations for modulated renewal process models." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71:3–23.

Liu, D. and Lengyel, M. (2023). "Bayesian nonparametric (non-)renewal processes for analyzing neural spike train variability." In Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., and Levine, S. (eds.), *Advances in Neural Information Processing Systems*, volume 36. Curran Associates, Inc.

Lo, A. Y. and Weng, C.-S. (1989). "On a class of Bayesian nonparametric estimates: II. Hazard rate estimates." *Annals of the Institute of Statistical Mathematics*, 41:227–245.

Lomnicki, Z. A. (1966). "A note on the Weibull renewal process." *Biometrika*, 53:375–381.

Maguire, B. A., Pearson, E. S., and Wynn, A. H. A. (1952). "The time intervals between industrial accidents." *Biometrika*, 39:168–180.

Masala, G. (2012). "Earthquakes occurrences estimation through a parametric semi-Markov approach." *Journal of Applied Statistics*, 39:81–96.

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

NGDC/WDS (2021). "National Geophysical Data Center/World Data Service: NCEI/WDS Global Significant Earthquake Database." *NOAA National Centers for Environmental Information.*

Nieto-Barajas, L. E. (2014). "Bayesian semiparametric of short– and long–term hazard ratios with covariates." *Computational Statistics & Data Analysis*, 71:477–490.

Nieto-Barajas, L. E. and Walker, S. G. (2004). "Bayesian nonparametric survival analysis via Lévy driven Markov processes." *Statistica Sinica*, 14:1127–1146.

Oakes, D. and Cui, L. (1994). "On semiparametric inference for modulated renewal processes." *Biometrika*, 81:83–90.

Ogata, Y. (1999). "Estimating the hazard of rupture using uncertain occurrence times of paleoearthquakes." *Journal of Geophysical Research: Solid Earth*, 104:17995–18014.

Ouhbi, B. and Limnios, N. (1999). "Nonparametric estimation for semi-Markov processes based on its hazard rate functions." *Statistical Inference for Stochastic Processes*, 2:151–173.

Ouhbi, B. and Limnios, N. (2003). "Nonparametric reliability estimation of semi-Markov processes." *Journal of Statistical Planning and Inference*, 109:155–165.

Pakman, A. and Paninski, L. (2014). "Exact Hamiltonian Monte Carlo for truncated multivariate Gaussians." *Journal of Computational and Graphical Statistics*, 23:518–542.

Patwardhan, A. S., Kulkarni, R. B., and Tocher, D. (1980). "A semi-Markov model for characterizing recurrence of great earthquakes." *Bulletin of the Seismological Society of America*, 70:323–347.

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

Petrone, S. (1999b). "Random Bernstein polynomials." *Scandanaivan Journal of Statistics*, 26:373–393.

Pievatolo, A. and Rotondi, R. (2000). "Analysing the interevent time distribution to identify seismicity phases: a Bayesian nonparametric approach to the multiple-changepoint problem." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 49:543–562.

Polson, N. G., Scott, J. G., and Windle, J. (2013). "Bayesian inference for logistic

models using Pólya–gamma latent variables." *Journal of the American Statistical Association*, 108:1339–1349.

Pyke, R. (1961a). "Markov renewal processes: definitions and preliminary properties." *The Annals of Mathematical Statistics*, 32:1231–1242.

Pyke, R. (1961b). "Markov renewal processes with finitely many states." *The Annals of Mathematical Statistics*, 32:1243–1259.

Pyke, R. and Schaufele, R. (1964). "Limit theorems for Markov renewal processes." *The Annals of Mathematical Statistics*, 35:1746–1764.

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rao, V. and Teh, Y. W. (2011). "Gaussian process modulated renewal process." In Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., and Weinberger, K. (eds.), *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc.

Ren, L., Du, L., Carin, L., and Dunson, D. (2011). "Logistic stick-breaking process." *Journal of Machine Learning Research*, 12:203—-239.

Rigon, T. and Durante, D. (2021). "Tractable Bayesian density regression via logit stick-breaking priors." *Journal of Statistical Planning and Inference*, 211:131–142.

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

Roberts, G. O. and Rosenthal, J. S. (2009). "Examples of adaptive MCMC." *Journal of Computational and Graphical Statistics*, 18:349–367.

Rotondi, R. (2010). "Bayesian nonparametric inference for earthquake recurrence time distributions in different tectonic regimes." *Journal of Geophysical Research: Solid Earth*, 115:B01302.

Rotondi, R. (2021). "Bayesian Analysis of Seismic Events." In Balakrishnan, N., Colton, T., Everitt, B., Piegorsch, W., Ruggeri, F., and Teugels, J. (eds.), *Wiley StatsRef: Statistics Reference Online*. Hoboken, NJ: Wiley.

SCEDC (2013). "Southern California Earthquake Data Center." *Caltech*, Dataset.

Sethuraman, J. (1994). "A constructive definition of Dirichlet priors." *Statistica Sinica*, 4:639–650.

Singhai, R., Joshi, S. D., and Bhatt, R. K. P. (2009). "Offered-load model for Pareto inter-arrival network traffic." In *2009 IEEE 34th Conference on Local Computer Networks*. (pp. 364–367). Zurich, Switzerland.

Skellam, J. G. and Shenton, L. R. (1957). "Distributions associated with random walk and recurrent events." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 19:64–118.

Smith, W. L. (1955). "Regenerative stochastic processes." *Proceedings of the Royal Society of London: Series A (Mathematical and Physical Sciences)*, 232:6–31.

Smith, W. L. (1958). "Renewal theory and its ramifications." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 20:243–284.

Smith, W. L. and Leadbetter, M. R. (1963). "On the renewal function for the Weibull distribution." *Technometrics*, 5:393–396.

Stowell, D. and Plumbley, M. D. (2013). "Segregating event streams and noise with a Markov renewal process model." *Journal of Machine Learning Research*, 14:2213–2238.

Swishchuk, A. and Islam, M. (2011). "The geometric Markov renewal processes with application to finance." *Stochastic Analysis and Applications*, 29:684–705.

Taddy, M. and Kottas, A. (2012). "Mixture modeling for marked Poisson processes." *Bayesian Analysis*, 7:335–362.

Tijms, H. (1994). *Stochastic Models: An Algorithmic Approach*. Chichester: Wiley.

Utsu, T. (1984). "Estimation of parameters for recurrence models of earthquakes." *Bulletin of the Earthquake Research Institute, University of Tokyo*, 59:53–66.

Votsi, I., Limnios, N., Tsaklidis, G., and Papadimitriou, E. (2012). "Estimation of the expected number of earthquake occurrences based on semi-Markov models." *Methodology and Computing in Applied Probability*, 14:685–703.

Warr, R. L. and Woodfield, T. B. (2020). "Bayesian nonparametric estimation of first passage distributions in semi-Markov processes." *Applied Stochastic Models in Business and Industry*, 36:237–250.

White, J. (1964). "Weibull renewal analysis." *Society of Automotive Engineers Technical Paper*, 640624.

Windle, J., Polson, N. G., and Scott, J. G. (2014). "Sampling Pólya-Gamma random variates: alternate and approximate techniques." *arXiv*, eprint arXiv:1405.0506.

Wiper, M., Insua, D. R., and Ruggeri, F. (2001). "Mixtures of gamma distributions with applications." *Journal of Computational and Graphical Statistics*, 10:440–454.

Xiao, S., Kottas, A., Sansó, B., and Kim, H. (2021). "Nonparametric Bayesian modeling and estimation for renewal processes." *Technometrics*, 63:100–115.

Zhang, Z., Chin, A., Nishimura, A., and Suchard, M. A. (2022). "hdtg: an R package for high-dimentional truncated normal simulation." *arXiv preprint arXiv:2210.01097*.

Zhao, C. and Kottas, A. (2021). "Modelling for Poisson process intensities over irregular spatial domains." *arXiv preprint arXiv:2106.04654*.

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

Zheng, X., Kottas, A., and Sansó, B. (2024). "Mixture modeling for temporal point processes with memory." *arXiv preprint arXiv:2407.03774*.