Title
Bayesian Nonparametric Modeling for Some Classes of Temporal Point Processes

Permalink
https://escholarship.org/uc/item/0jv9c99b

Author
Xiao, Sai

Publication Date
2015

License
CC BY-NC 4.0

Peer reviewed|Thesis/dissertation
UNIVERSITY OF CALIFORNIA
SANTA CRUZ

BAYESIAN NONPARAMETRIC MODELING FOR SOME CLASSES
OF TEMPORAL POINT PROCESSES

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

DOCTOR OF PHILOSOPHY

in

STATISTICS AND APPLIED MATHEMATICS

by

Sai Xiao

March 2015

The Dissertation of Sai Xiao
is approved:

__________________________________________
Professor Athanasios Kottas, Chair

__________________________________________
Professor Bruno Sansó

__________________________________________
Professor Abel Rodriguez

__________________________________________
Professor Juhee Lee

Tyrus Miller
Vice Provost and Dean of Graduate Studies
Table of Contents

List of Figures v
List of Tables ix
Abstract x
Dedication xiii
Acknowledgments xiv

1 Introduction 1
  1.1 Motivation and Literature Review 2
  1.2 Contribution and Organization 5

2 Background 9
  2.1 Basic Definition 9
  2.2 Homogeneous Poisson Process 10
  2.3 Inhomogeneous Poisson Process 12
  2.4 Marked Poisson Process 13
  2.5 Renewal Process 14

3 Nonparametric Bayesian Modeling for Seasonal Marked Point Processes 16
  3.1 Introduction 16
  3.2 Hurricane Data 22
  3.3 Modeling Time-Varying Intensities 27
    3.3.1 Nonparametric dynamic model for Poisson process densities 27
    3.3.2 Time series model for the total intensities 29
    3.3.3 Implementation details and posterior inference 30
    3.3.4 Analysis of dynamically evolving hurricane intensities 33
  3.4 DDP Model For Seasonal Marked Poisson Processes 37
    3.4.1 The DDP-AR model 38
    3.4.2 Standardization of hurricane damages 40
    3.4.3 Inference 43
3.4.4 Results ......................................................... 44
3.4.5 Model assessment ........................................ 54
3.5 Conclusion ...................................................... 58

4 Bayesian Nonparametric Modeling for Renewal Processes 61
4.1 Introduction ...................................................... 61
4.2 Model Formulation ........................................... 64
  4.2.1 The mixture model ........................................ 64
  4.2.2 Model properties .......................................... 68
  4.2.3 Investigating the $K$ function for the Erlang mixture model .... 71
  4.2.4 Inference .................................................. 78
4.3 Data Analysis .................................................. 79
  4.3.1 Synthetic data ............................................. 80
  4.3.2 Earthquake occurrences data ............................. 85
  4.3.3 Coal mining disaster data ............................... 88
  4.3.4 Model checking ........................................... 90
4.4 Conclusions and Further Work ............................. 92

5 Bayesian Nonparametric Modeling for Markov Renewal Processes 94
5.1 Motivation ....................................................... 94
5.2 Related Work .................................................. 97
5.3 Model Formulation ........................................... 100
  5.3.1 Prior specification ....................................... 103
  5.3.2 Simulation from prior model ............................ 105
5.4 Posterior Inference .......................................... 107
  5.4.1 Other inference ........................................... 108
5.5 Data Analysis .................................................. 109
  5.5.1 Earthquake occurrences data ............................ 109
  5.5.2 Sensitivity analysis ..................................... 117
5.6 Discussion .................................................... 119

6 Conclusions ....................................................... 124

A MCMC algorithm for the DDP-PBAR model 142
B Proof of Result 4 in Chapter 4 144
C MCMC algorithm for renewal process model 146
D MCMC algorithm for inhomogeneous renewal process model 150
### List of Figures

3.1 Left panel: The time series of annual hurricane occurrences. Right panel: Histogram (with bin width of 10 days) of hurricane occurrences over months after aggregating all hurricanes into one year. The solid and dashed lines denote the point and 95% interval estimates of the corresponding NHPP density, using the Dirichlet process mixture model discussed in Section 3.2.  

3.2 The number of hurricanes within one season aggregated by decades. In each decade, the number of hurricanes is grouped by months.  

3.3 Under the two distinct levels of data aggregation, posterior mean estimates for the hurricane intensity in 2011-2015 (left panel) and posterior densities for the probability of at least one hurricane in May for 2011-2015 (right panel).  

3.4 Boxplots of posterior samples for the average number of hurricanes in the month of September across five-year periods from 1900 to 2010.  

3.5 Posterior mean estimates (solid line) and 95% intervals (grey bands) of the hurricane intensity during 1971-2010. Points along the horizontal axis correspond to the observations.  

3.6 Boxplots of posterior samples of the mean number of hurricanes in early season (May and June) by decade (left panel), and in September by decade (right panel). In both panels, the solid dots indicate the corresponding observed numbers of hurricanes.  

3.7 Data box plots across decades for log-transformed base damages (left panel), damages normalized to current values (middle panel), and standardized damages (right panel).  

3.8 Top panel: the density of maximum wind speed conditional on June and September for all decades. Bottom panel: posterior expectation and 95% interval (red band for September; grey band for June) for the median maximum wind speed in June and September versus decade.  

3.9 Top panel: the density of logarithmic standardized damage conditional on MJJ (May-June-July) and ASO (August-September-October). Bottom panel: Posterior expectation and 95% interval (red band for ASO; grey band for MJJ) for the median standardized damage of one hurricane in MJJ and ASO.
3.10 Top panel: the density of logarithmic standardized damage in ASO given the
seven maximum wind speed categories defined in Table 3.1. Bottom panel:
Posterior expectation for the median standardized damage in ASO for the
seven maximum wind speed categories. ........................................ 51

3.11 Bivariate densities of maximum wind speed (mph) (x-axis) and logarithmic
standardized damage (y-axis) in ASO across decades. The dots correspond
to observations in ASO. .......................................................... 52

3.12 Posterior means for the conditional expectation of logarithmic standardized
damage given maximum wind speed in ASO across decades. ................. 53

3.13 Posterior Q-Q plots (mean and 95% interval) of estimated quantiles against
the theoretical uniform(0, 1) for: time (left panel), maximum wind speed
given time (middle panel), and standardized damage given time (right panel).
Results are shown for the last five decades. ................................. 56

3.14 Cross-validated posterior predictive densities in ASO of decade 2001–2010:
the left panel corresponds to maximum wind speed, and the right panel
to logarithmic standardized damage. The histograms plot the associated
observations in ASO of 2001–2010. .............................................. 57

4.1 The solid dots correspond to the CDF sample paths of G. The dashed lines
display the discretization by θ, and the thick red lines indicate the weights.
Left panel: when α = 10, few of the weights are 0. Right panel: when α = 1,
a large number of weights are 0 and only a few are significantly greater than 0. 67

4.2 K functions based on 10 samples from DP(α, G₀) where α = 200. From left
to right: G₀=Weibull(5, 0.5), G₀=Weibull(5, 1) and G₀=Weibull(5, 2). ................ 73

4.3 K functions based on 10 samples from a DP(α, G₀) where α = 200. From
left to right: G₀=Pareto(2.5, 1), G₀=Pareto(2.5, 5) and G₀=Pareto(0.01, 5). .... 74

4.4 K functions based on 10 samples from a DP(α, G₀) where G₀ = Weibull(5, 0.5).
From left to right: α = 0.2, 2, 50 respectively. ........................................ 76

4.5 K functions based on 10 samples from a DP(α, G₀) where G₀ = Weibull(5, 2).
From left to right: α = 2, 10, 50 respectively. ................................. 77

4.6 First and second simulated data sets. 4.6a and 4.6d: a snapshot of the point
process realizations for the first and second densities respectively; 4.6b and
4.6e: true data-generating function (black line), posterior mean of the inter-
arrival time density function (red line), and 95% interval bands (dash blue
lines) for the first and second densities respectively. 4.6c and 4.6f: true K
function (black line); posterior mean of the K function (red line) and the 95%
interval bands (dash blue lines) for the first and second densities respectively. 81

4.7 Third simulated data set. 4.7a: a snapshot of the simulated point process.
4.7b, 4.7c and 4.7d: true density function (black line), posterior mean of the
inter-arrival time density function (red line) and 95% interval bands (dash blue lines) for the head of the data, the tail of the data and all data respectively. 4.7e: the true K function (black line), the posterior mean of the estimated K function (red line) and the 95% interval bands (dash blue lines). 83
4.8 Fourth simulated data set. 4.7a: a snapshot of the simulated point process. 4.7b, 4.7c and 4.7d: true density function (black line), posterior mean of the inter-arrival time density function (red line) and 95% interval bands (dash blue lines) for the head of the data, the tail of the data and all data respectively. 4.7e: the true K function (black line), the posterior mean of the estimated K function (red line) and the 95% interval bands (dash blue lines). 84

4.9 4.9a: Large earthquakes in Turkey in the 20th century. 4.9b: the mean posterior predictive density of inter-arrival times (red line), 95% interval bands (blue dash lines) and the histogram of the real data. 4.9c and 4.9d: the posterior mean (red line) and 95% interval bands (blue dash lines) of hazard function (4.9c) and K function (4.9d). 87

4.10 4.10a: coal mining data from 1851 to 1875. 4.10b: the posterior mean predictive density (red line), 95% interval bands (blue dash lines) and the histogram of the real data. 4.10c: the distribution of the two largest weights and the corresponding component indices. 4.10d and 4.10e: the posterior mean (red line) and 95% interval bands (blue dash lines) of hazard function (4.10d) and K function (4.10e) given 1000 posterior samples. 89

4.11 4.11a: thinned coal mining data. 4.11b: the posterior mean predictive density (red line), 95% interval bands (blue dash lines) and the histogram of the real data. 4.11c: the distribution of the two largest weights and the corresponding component indices. 4.11d and 4.11e: the posterior mean (red line) and 95% interval bands (blue dash lines) of the hazard function (4.11d) and the K function (4.11e) given 1000 posterior samples. 91

4.12 Posterior Q-Q plots (mean and 95% interval) of estimated quantiles against the theoretical uniform(0, 1) for time in earthquake data (4.12a), coal-mining data (4.12b) and thinned coal-mining data (4.12c). 92

5.1 Simulation example. Three realizations of modulated renewal point processes from the model, with the simulated latent \( \theta(t) \), CDF \( G \) and the inter-arrival time densities over \( H(t) \) and \( t - H(t) \). 106

5.2 earthquake data. The posterior median (red dots) and 95% quantiles (grey shaded area) of \( \{ \theta(t_k) \} \), \( k = 1, \ldots, n + 1 \). 112

5.3 earthquake data. The left top panel: the inter-arrival time density from homogeneous renewal process model. The histogram is from the data. The succeeding panels from top to bottom, left to right show the inter-arrival time density at different time points with previous event occurring at the 5000th, 10000th, 15000th, 20000th, 25000th and 30000th day since the year 1900, respectively. The red solid line is the posterior median and blue dash lines are posterior 95% uncertainty band of inter-arrival time density. 113

5.4 earthquake data. The surface plot of inter-arrival time densities over \( H(t) \) and \( t - H(t) \) (in days). 114

5.5 earthquake data. Posterior estimation of \( \lambda^* \) for the earthquake data along with the observations (blacks dash lines). 115
5.6 earthquake data. Posterior Q-Q plots (mean and 95% interval) of estimated quantiles against the theoretical Uniform(0, 1) for time in earthquake data. The left one is produced from the homogeneous renewal process model and the right one is produced from the inhomogeneous renewal process model.

5.7 earthquake data. The right column is the posterior estimation of \( \{\theta(t_k)\} \) and Q-Q plots under the setting \( m = 15 \) and \( \sigma^2_{\theta} = (T/m)^2 \). The right column is the posterior estimation of \( \{\theta(t_k)\} \) and Q-Q plots under the setting \( m = 25 \) and \( \sigma^2_{\theta} = (T/m)^2 \).
List of Tables

3.1 Saffir-Simpson hurricane scale. TD: tropical depression; TS: tropical storm; HC 1 to HC 5: hurricane of category 1 to 5. The number of hurricanes corresponding to each category in the data set. ........................................ 23


4.2 Clustering patterns generated by prior models with different precision parameters. C: clustering processes. D: declustering processes. H: homogeneous Poisson processes. ................................................................. 77
Abstract

Bayesian Nonparametric Modeling for Some Classes of Temporal Point Processes

by

Sai Xiao

Model-based inferential methods for point processes have received less attention than the corresponding theory of point processes and is more scarcely developed than other areas of statistical inference.

Classical inferential methods for point processes include likelihood-based and nonparametric methods. Bayesian analysis provides simulation-based estimation of several statistics of interest for point processes. However, a challenge of Bayesian modeling, specifically for point processes, is selecting an appropriate parametric form for the intensity function. Bayesian nonparametric methods aim to avoid the narrow focus of parametric assumptions by imposing priors that can support the entire space of distributions and functions. It is naturally a more flexible and adaptable approach than those based on parametric models.

In this dissertation, we focus on developing methodology for some classes of temporal point processes modeling and inference in the context of Bayesian nonparametric methods, mainly with applications in environmental science. Firstly, we are motivated to study seasonal marked point process by an application of hurricanes occurrences. We develop nonparametric Bayesian methodology to study the dynamic evolution of a seasonal marked point process intensity under the assumption that the point process is a nonhomogeneous Poisson process. The dynamic model for time-varying intensities provides
both the intra-seasonal and inter-seasonal variability of occurrences of events. Considering
marks, we provide a full probabilistic model for the point process over the joint marks-points
space which allows for different types of inferences, including full inference for dynamically
evolving conditional mark densities given a time point, a particular time period, and even a
subset of marks. We apply this method to study the evolution of the intensity of the process
of hurricane landfall occurrences, and the respective maximum wind speed and associated
damages. We show several novel inferences which are explored for the first time in the
analysis of hurricane occurrences.

Then we look beyond Poisson processes and propose a flexible approach to mod-
eling and inference for homogeneous renewal processes. This modeling method is based
on a structured mixture of Erlang densities with common scale parameter for the renewal
process inter-arrival density. The mixture weights are defined through an underlying dis-
tribution function modeled nonparametrically with a Dirichlet process prior. This model
specification enables flexible shapes for the inter-arrival time density, including heavy tailed
and multimodal densities. Moreover, the choice of the Dirichlet process centering distribu-
tion controls clustering or declustering patterns for the point process. Finally we extend
our model to accommodate point processes with time-varying inter-arrivals, which are re-
ferred to as modulated renewal processes in the literature. We introduce time dependence
in the scale parameter of the Erlang mixture by replacing it with a latent stochastic pro-
cess. A number of synthetic data sets and real data sets are used to illustrate the modeling
approaches.

The main contribution of this thesis is to provide Bayesian nonparametric mod-
eling and inference methods for some classes of point processes, which are more flexible than existing methods. Moreover, the key complication for Bayesian inference is that the likelihood of a generic point process involves a normalizing constant which is, most of the times, analytically intractable. Discretization is very often used in existing methods to get likelihood approximations that facilitate computations, especially for models based on Gaussian process priors. Superior to these methods, our work uses the exact likelihood without approximation in all of our developed models.
To my parents, who give me endless and selfless love.
Acknowledgments

I feel indebted to many people for their help and support in the whole journey of pursuing my PhD.

My most gratitude will go to my two advisors, Athanasios Kottas and Bruno Sansó. They taught me how to think independently and deeply, encouraged and trusted me through these years. I specially thank their care and support during my maternity leave last summer. Thank Athanasios for bringing me to the world of Bayesian nonparametrics, which is like a journey to infinite possibilities. Thank Athanasios for teaching me that there is no short-cut to success and I must work and think hard to do solid research. Your rigorous attitude towards research will affect me in my future career. I truly thank Bruno for his mentoring and advising. I gain a good deal of enlightenment from his lively, clear and open mind. Thank Bruno for his pleasant and positive attitude, countless encouragement, nice and honest advices, and his efficient working style. I am very lucky to have these two great guys as my advisors. I will not be able to accomplish this thesis without their supervision.

There are many people in AMS department that I am indebted to. Thank Abel Rodriguez, who serves as my committee member for both my advancement and my thesis, for his insightful comments. Thank Juhee Lee, my another committee member, for providing useful suggestions for my thesis. Juhee’s multivariate analysis class is also one of my favorite classes. I also want to thank Raquel Prado, who offered us a wonderful time series class. AMS is a happy family with many wonderful students. Special thanks go to Maria De Yoreo, Katelyn White and Nicole Mendoza for giving me an unforgettable baby shower.

No words can express my deep gratitude to my parents, who gave me life and
support me at every turn. I feel beloved to have their support to pursue my dream, despite that going abroad to pursue a PhD means long distance separation for many years. Thank my parents-in-law who sacrificed so much to help us for a whole year. It is because of their help that I was able to go through this transition phase of my life with more confidence.

My entire process of pursuing PhD will not be so enjoyable without my husband, Jiazhong Nie, who understands me and listens to me all the time. Thank him for being sharp, reminding me and helping me to strengthen myself. Finally, thank my son, Oscar, for bringing so much joy to our lives.
Chapter 1

Introduction

A point process is a stochastic model for random events occurring in time and/or space. It includes temporal, spatial, and spatio-temporal point processes. A temporal point process is a stochastic process composed of a time series of binary events that occur in continuous time (Daley & Vere-Jones, 2003). Spatial point processes generate countable sets of events localized in space. A spatio-temporal point processes is a stochastic model for events localized in both space and time.

The theory of point processes has a long history which traces back to 1837 when the Poisson distribution firstly appeared. During these two centuries, the theory of point processes has been extensively studied. References on basic point process theory include Daley & Vere-Jones (2003), Møller & Waagepetersen (2004), Illian et al. (2008). However, model-based inferential methods for point processes received less attention and more scarcely developed than other areas of statistical inference.

This thesis focuses on modeling and inference methods for some classes of temporal
point processes. A temporal point process is different from a time series, which is a sequence of data points successively recorded. It is worth noting that the points in a point process can occur at any time in a continuum. However, the point process can be simplified to a time series of zeros and ones by discretizing the observed time window. For temporal point processes, the inference is focused on the time when the process is observed. Such models have been used in a broad range of fields such as climatology, medicine, telecommunication, environmental science, economics and epidemiology. For example, in environmental science, studies of variability in the frequencies of large hurricanes and economic damages associated with extreme weather are critically important in climate change assessment (Pielke & Pielke, 1997). In the study of cancer, the appearance times of metastases are treated as a temporal point process, and whether the intensity function involves a time trend helps researchers to better understand the kinetic mechanism for the spread of cancer (Bartoszynski et al., 1981). In criminological research, the contagion-like clustering pattern detected in the crime data can be modeled by particular point processes to improve the prediction of crimes (Mohler et al., 2011). In seismology, point process models have been used to characterize the elastic rebound behavior of earthquakes (Parsons, 2008). In telecommunication, various point process models are used to study the distribution of packet and connection inter-arrivals in network traffic data (Paxson & Floyd, 1995).

1.1 Motivation and Literature Review

Statistical inference for point processes focuses mainly on estimation for intensity functions. For a point process on the real half-line, the intensity function is defined by
conditioning on the complete history of events. Only for Poisson processes, the intensity function depends solely on the current time. The classical inferential methods for point processes include likelihood-based and nonparametric methods for spatial point processes (Diggle, 2003). Due to the development of Markov Chain Monte Carlo (MCMC) methods, Bayesian analysis provides simulation-based estimation of several statistics of interest; see Møller & Waagepetersen (2004) for the review of these methods in spatial point processes.

Bayesian analysis provides probabilistic inference for point process intensities and estimation for unknown parameters instead of merely point estimates. However, a challenge of parametric Bayesian modeling, specifically for point processes, is selecting an appropriate parametric form for the intensity functions.

Bayesian nonparametric methods aim to avoid potentially restrictive parametric assumptions. They provide priors that can support the entire space of distributions or functions, by considering prior distributions for infinite-dimensional parameters. For example, for density estimation problems, a Bayesian nonparametric model constitutes a Bayesian model on the space of all densities. To fit a finite data set, it uses a finite set of parameters to adapt to the data. From this sense, this basic idea makes the Bayesian nonparametric modeling naturally more flexible and adaptable than parametric models. The commonly used probability measures or functions over infinite dimensional space include Dirichlet process, Pitman-Yor process, Polya trees, Bernstein Polynomials and Gaussian process. For more references on Bayesian nonparametrics, see Hjort et al. (2010).

Bayesian nonparametric methods have the potential to significantly contribute to point process modeling even though the literature is relatively limited. To our knowledge,
1.2 Contribution and Organization

In this work, we focus on modeling for inhomogeneous seasonal marked Poisson processes, homogeneous renewal processes and inhomogeneous renewal processes in the context of Bayesian nonparametric methods.

The main contribution of this thesis is to provide Bayesian nonparametric modeling and inference methods for some classes of point processes, which are more flexible than existing methods. The flexibility of our models are reflected in several aspects. For seasonal marked Poisson processes, our modeling approach provides various flexible inferences. The dynamic model for time-varying intensities provides both the intra-seasonal and inter-seasonal variability of occurrences of events. Considering marks, we provide a full probabilistic model for the point process over the joint marks-points space which provides different types of inferences, including full inference for dynamically evolving conditional mark densities given a time point, a particular time period, and even a subset of marks. We applied this modeling approach in the analysis of US hurricane occurrences, our modeling approach is the first work to provide such comprehensive inferences in the existing work. For homogeneous renewal processes, the modeling approach has flexibility to enable flexible shapes of inter-arrival time density including heavy tailed and multimodal densities, which is hard to achieve in a parametric method. The model also has the capability of modeling both declustering and clustering point patterns, so it can be applicable in general circumstances. We also develop a method for inhomogeneous renewal process in order to model the time-varying inter-arrival times.

Due to the characterizations of Bayesian nonparametrics, our methods offer promise
in modeling point processes with higher flexibility and adaptability. We develop flexible mixture formulations for the intensity functions of all these point processes we consider. However, one obstacle is that the likelihood of point process involves a normalizing constant which is analytically intractable. Discretization is very often used to get likelihood approximations to facilitate the computation, especially for models based on Gaussian process priors. Superior to these methods, our work uses the exact likelihood without approximation in all of our developed models.

Here, the outline of the contents of this thesis is briefly introduced.

In Chapter 2, we introduce some basics of point processes. Then we review several point processes considered in this thesis, such as homogeneous Poisson processes, inhomogeneous Poisson processes, marked Poisson processes and renewal processes.

In Chapter 3, we are motivated to study seasonal marked point process by an application of hurricanes occurrences. Seasonal point processes refer to stochastic models for random events which are only observed in a given season. We develop nonparametric Bayesian methodology to study the dynamic evolution of a seasonal marked point process intensity. We assume the point process is a non-homogeneous Poisson process, and propose a nonparametric mixture of beta densities to model dynamically evolving temporal Poisson process intensities. Dependence structure is built through a dependent Dirichlet process prior for the seasonally-varying mixing distributions. We extend the nonparametric model to incorporate time-varying marks resulting in flexible inference for both the seasonal point process intensity and for the conditional mark distribution. The model has been applied to hurricane landfalls with reported damages along the U.S. Gulf and Atlantic coasts from
1900 to 2010. We focus on studying the evolution of the intensity of the process of hurricane landfall occurrences, and the respective maximum wind speed and associated damages.

In Chapter 4, we go beyond Poisson processes and study the Markovian dependence between points. The renewal point process is the most fundamental point process which departs from the Poisson process assumption. We propose a flexible approach to modeling and inference for homogeneous renewal processes. The model is built from a structured mixture of Erlang densities for the renewal process inter-arrival density. The Erlang mixture components have a common scale parameter, and the mixture weights are defined through an underlying distribution function modeled nonparametrically with a Dirichlet process prior. This model specification enables nonstandard shapes for the inter-arrival time density, including heavy tailed and multimodal densities. Moreover, the choice of the Dirichlet process centering distribution controls clustering or declustering patterns for the point process, which can therefore be encouraged in the prior specification. The proposed modeling approach is illustrated with several synthetic data sets, earthquake occurrences data, and coal-mining disaster data.

In Chapter 5, we further abandon the independence and identical distribution property of the inter-arrival time in homogeneous renewal process and study the inhomogeneous renewal process. We extend the model in Chapter 4 to accommodate point processes with time-varying inter-arrivals, which are referred to as modulated renewal processes in some of the literature. We introduce time dependence in the scale parameter of Erlang mixture by replacing it with a latent stochastic process. The model is illustrated with earthquake data and it shows an improvement in the goodness-of-fit to the data. We also
discuss two possible models, both of which introduce the time dependence in the underlying
distribution function which generates the mixture weights for Erlang components.

In Chapter 6, we summarize the conclusions and main contributions from Chapter
3 to 5 and discuss some possible future work.
Chapter 2

Background

There are a wide variety of classes of point processes. Each of them represents different stochastic mechanisms for the generation of point patterns. This requires the modeler or user to understand the probability theory of point processes so that an appropriate class of point processes can be chosen to describe the observed physical phenomenon or point pattern. In this chapter, a brief review of the basic theory of point processes and several common classes of temporal point processes considered in this thesis is given.

2.1 Basic Definition

A temporal point process can be defined in several ways. First of all, a realization of point processes is a series of ordered data time points localized on the time line ($\mathbb{R}^+$). Let $T_0$, $T_1$, $\ldots$ be the location of events, $T_n$ is the time of the $n$th arrival. Second, a point process can be fully specified by a counting process. Let $N(a, b]$ denote the number of events within the half-open interval $(a, b]$, $a < b$. A counting process is denoted as
\[ N = \{N(t) : t \geq 0\}, \text{ where } N(t) \text{ is defined as } N(0, t], \text{ the total number of events occurred up to and including time } t, \]

\[ N(t) = N(0, t] = \#\{i : T_i \leq t\} \text{ or } N(t) = \max\{n : T_n \leq t\} \]

The time of data points can be tracked by the counting process as below, by assuming \( T_0 = 0, \)

\[ T_n = \inf\{t : N(t) = n\}. \]

Moreover, people also use inter-arrival times or waiting times to represent point processes. The time intervals between successive events are random variables \( X_1, X_2, \ldots \) given by \( X_n = T_n - T_{n-1}. \) The arrival times of events can be tracked back by \( T_n = \sum_{i=1}^{n} X_i. \) These terms and notation will be used throughout this thesis.

### 2.2 Homogeneous Poisson Process

The Poisson process is one of the most commonly used point processes, and is always treated as the building block of more structured point processes. The homogeneous Poisson process or stationary Poisson process is the simpler version. A homogeneous Poisson process with constant intensity function \( \lambda \) can be conveniently defined by the following conditions (Daley & Vere-Jones, 2003; Grimmett & Stirzaker, 2001),

(a) The probability that two events happen at the same time is zero.

(b) The number of points in two disjoint intervals are independent random variables. If \( s < t, \) the number \( N(t) - N(s) \) in the interval \( (s, t] \) is independent of the times of events during \([0, s]).\)
(c) The number of points in each finite interval \((s, t]\) follows a Poisson distribution with mean \(\lambda(t - s)\).

Here, the intensity function is the mean of the number of events per unit time interval. Condition (b) is called independence property of Poisson process, by which we mean that the probability of occurrence of the next event does not depend on the history of events. Following condition (b), we can obtain that: (Ross, 1983) given that \(N(t) = n\), the \(n\) arrival times \(S_1, ..., S_n\) have the same distribution as the order statistics corresponding to \(n\) independent random variables uniformly distributed on the interval \((0, t]\), which is

\[
f(y_1, y_2, ..., y_n) = \frac{n!}{t^n}, \quad 0 < y_1 < y_2 < ... < y_n < t
\]

Condition (c) is called stationary independent increments, in the sense that the distribution of \(N(s, t]\) is stationary, only depending on the length of the interval, \((t - s)\). Also, using condition (c), it gives the probability of no events in the interval \((s, t]\),

\[
\Pr(N(s, t] = 0) = e^{-\lambda(t-s)}.
\]

Therefore, the probability of finding one event is \(\Pr(N(s, t] = 1) = 1 - e^{-\lambda(t-s)}\). This implies that the inter-arrival time between two successive events follows an exponential distribution with mean \(\lambda^{-1}\).

Based on the properties of homogeneous Poisson processes, a realization in a finite interval \((s, t]\) can be simulated in two ways. Firstly, sample the number of events \(N(s, t]\) by Poisson(\(\lambda(t - s)\)). Then, conditional on \(N(s, t]\), a set of events are independently generated by a uniform distribution on \((s, t]\). Finally, the points should be ordered to form a temporal point process. Alternatively, simulate a series of inter-arrival time interval i.i.d. from an
exponential distribution with mean $\lambda^{-1}$ and terminate the simulation when the time of the latest event exceeds the upper bound of the time window.

### 2.3 Inhomogeneous Poisson Process

A more general Poisson process, usually called non-stationary or inhomogeneous Poisson process is a Poisson process given by a time-varying intensity function $\lambda(t)$, which is defined by

$$
\lambda(t) = \lim_{h \to 0} \frac{1}{h} E[N(t, t+h)]
$$

The inhomogeneous Poisson process no longer requires stationary increments, so it allows for the possibility that events may be more likely to occur at certain times than at other times. However, the independence property still holds. Also, the number of points in any finite interval $(s, t]$ still has a Poisson distribution with mean $\Lambda(s, t) = \int_s^t \lambda(u)du$. Similarly, given the number of events in an finite interval, the events are occurred according to $\lambda(t)$ rather than being uniformly scattered. The inhomogeneous Poisson process is a more natural model for most real data. For example, the intensity function of cyclones or typhoons cannot be homogeneous because of the underlying seasonal variants. The intensity of crimes occurrences usually changes with social development and economic growth. Also, the inhomogeneous Poisson process provides a natural framework to introduce covariates into the time-varying intensity function.

By these properties of Poisson process, the likelihood of a finite realization of a Poisson process can be obtained. Let $t_1, t_2, \ldots, t_{N_T}$ be a realization of a Poisson process of intensity $\lambda(t)$ on $(0, T]$. Then we discretize $(0, T]$ into intervals of width $h$ such that in each
small intervals there can only be one event at most. So, for \( k = 1, 2, \ldots, T/h, N(I_k) = 0 \) or 1, where \( I_k = ((k-1)h, kh] \) and \( I(t_i) \) refers to the \( I_n \) containing \( t_i \). Thus, for each \( N(I_k) \), \( N(I_k) \sim \text{Poisson}(\Lambda(I_k)) \), where \( \Lambda(I_k) = \int_{I_k} \lambda(u)du \). The likelihood of Poisson process can be derived by

\[
L(t_1, t_2, \ldots, t_{NT} \mid \lambda) = \prod_{k=1}^{T/h} \text{Poisson}(N(I_k) \mid \Lambda(I_k))
\]

\[
= \prod_{k=1}^{T/h} \frac{\Lambda(I_k)^{N(I_k)}}{N(I_k)!} \exp(-\Lambda(I_k))
\]

Because \( N(I_k) \) is 1 only if \( I_k \) contain an observation and equals to 0 otherwise, we can simplify the above equation as

\[
\left\{ \prod_{k=1}^{NT} \Lambda(I(t_k)) \right\} \exp \left( -\sum_{k=1}^{T/h} \Lambda(I_k) \right)
\]

Letting the width of intervals \( h \to 0 \), the likelihood becomes

\[
L(t_1, t_2, \ldots, t_{NT} \mid \lambda) \propto \left\{ \prod_{k=1}^{NT} \lambda(t_k) \right\} \exp \left( -\int_{0}^{T} \lambda(u)du \right)
\]

(2.1)

### 2.4 Marked Poisson Process

One important extension of Poisson processes is given by marked Poisson processes. A mark corresponds to additional numerical information associated with a point. For example, there are temporal point processes such as the incidences of purchase of on-line shopping, occurrences of earthquake, arrivals of calls in the queueing system, failures of machines in the computer system and occurrences of crimes. The corresponding marks can be the total amount of merchandise purchased, magnitude of earthquake, time length of calls, duration of failure and types of crimes, respectively. The marks can have various data
types, integers, real numbers and they can even be point processes themselves. Let $X$ be a Poisson process on $T \subseteq R^d$ and the mark space be denoted as $Y$, then a marked Poisson process is expressed as $\{(\xi, y_\xi) : \xi \in X, y_\xi \in Y\}$.

The random variable $\xi$ is a Poisson($\phi$) in point space $X$ with intensity function $\phi$. The definition of MPP requires $\phi$ to be locally integrable, which means it can be integrable on any compact subset of its domain $X$. The marks are required to be mutually independent, conditional on $\xi$ (Møller & Waagepetersen, 2004). The marking theorem states that a MPP is a inhomogeneous Poisson process with intensity function given by $\varphi(t, y) = \lambda(t) f(y \mid t)$, where $\lambda(t)$ is the marginal temporal intensity function, and the conditional mark density $f(y \mid t)$ depends only on the current time point $t$.

\subsection*{2.5 Renewal Process}

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

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

\[ M(t) = F(t) + \int_0^t M(t-x)dF(x). \]

Consider a temporal point pattern \(\{0 = t_0 < t_1 < t_2 < \cdots < t_n < T\}\) observed in \((0, T]\). The likelihood can be expressed as

\[
\Pr(T_1 = t_1, T_2 = t_2, \ldots, T_n = t_n, T_{n+1} > T) = \Pr(X_1 = t_1, X_2 = t_2 - t_1, \ldots, X_n = t_n - t_{n-1}, X_{n+1} > T - t_n)
\]

\[
= \left\{ 1 - \int_{t_n}^T f(u - t_n)du \right\} \prod_{i=1}^{n} f(t_i - t_{i-1}), \tag{2.2}
\]

where \(f(\cdot)\) is the inter-arrival density function. The last term in Equation (2.2) corresponds to the probability of no arrival between \((t_n, T]\). It is also the normalizing constant of likelihood.

Moreover, for non-Poisson process, the \(K\) function can be used to determine the clustering properties of a temporal point process. The \(K\) function is initially introduced by Ripley (1977) for spatial point processes. In a temporal setting we have that \(K(t) = \lambda^{-1}E[N(t)]\), where \(\lambda\) is the number of events per unit time interval. The \(K\) function of a homogeneous Poisson process is \(t\). \(K(t) > t\), corresponds to clustering processes; while \(K(t) < t\) indicates that the point processes has a declustering pattern. For renewal process, \(\lambda = 1/E(X)\), where \(X\) is the inter-arrival time for consecutive events. The renewal function is defined as \(M(t) = E[N(t)]\), so that the \(K\) function of a renewal process is \(M(t)E(X)\). More results of renewal processes will be discussed in section 4.2.2.
Chapter 3

Nonparametric Bayesian Modeling

for Seasonal Marked Point Processes

3.1 Introduction

There are many examples of phenomena that occur every year at random times, but are limited to a specific season. Two examples of natural events with strong scientific and economic relevance are: the Atlantic hurricanes and the Pacific typhoons formed by tropical cyclones that occur between May and November; and the spawning of coho salmon that takes place from November to January. There are some situations where the observational window is limited to a given season, such as wildlife abundance in regions that are not accessible in the winter. In addition, there exist applications where interest lies in
studying a physical process during a particular season. One example is the study of extreme precipitation during the dry season in tropical environments. This can be important to guarantee water supplies and also to prevent unexpected disasters. On a different note, studying incidence of online purchase of products during the Christmas season is indispensable for retailers in order to optimize stocking, advertising, logistics, staffing, and website maintenance and support. In all these examples it is important to understand the underlying mechanism of the seasonal point process. To this end, we need a flexible statistical model that can describe the changes of the process intensity during the season. The model also has to capture the evolution of the intensities from one year to the next, borrowing strength from the whole dataset to improve the estimation in a given season. Moreover, the model should be extensible to allow for inference on possible marks associated with the occurrence of the events.

In this Chapter, we focus on the study of landfalling hurricanes recorded along the U.S. Gulf and Atlantic coasts between 1900 and 2010, and their associated maximum wind speed and damages. Hurricanes are typical seasonal extreme climate events. In light of potential societal and economic impact of climate change, the obvious question regarding hurricanes is whether there is an intensification of hurricane frequency and an increasing trend of hurricane wind speed and associated damage. A substantial part of the literature on the variability of hurricane occurrences is based on annual counts of events. For example, Elsner et al. (2004) and Robbins et al. (2011) use change point detection methods to find significant increases in storm frequencies around 1960 and 1995. Limiting the analysis to the number of hurricanes per year precludes the description of occurrence variability within
each year. Thus, it is not possible to estimate trends in hurricane occurrence during a particular period within the hurricane season, say, a given month. An alternative approach is considered in Parisi & Lund (2000) where the process of hurricane occurrences is modeled with a continuous time-varying intensity function within one year. However, in this case, the inter-annual variability is not accounted for. An approach that models intra-annual, as well as inter-annual variability is presented in Solow (1989). The model is applied to a US hurricane data set (different from the one considered here) that consists of monthly counts along the mid-Atlantic coast of the US in 1942-83. The basic assumption is that the data correspond to a Poisson process with a non-stationary intensity function. This is decomposed into a secular and a seasonal component, estimated from annual and monthly counts, respectively. The analysis indicates no trend during 1950s and a decreasing trend in 1970s for the secular component, and a stationary seasonal cycle over time.

The focus on hurricane occurrence is of great importance in a climatological context. However, the frequency of hurricanes provides only a partial measure of the threat that these phenomena represent. When exploring the association of hurricane strength with global warming, Emanuel (2005) calls for research on hurricane potential destructiveness. The disastrous impact to coastal areas draws the attention of the public, and government officials and policy makers need reliable inferences on hurricanes’ potential damage for long-term action on economic development and population growth (Pielke & Pielke, 1997). For instance, in about ten years from Hurricane Fay in 2002 to Hurricane Irene in 2011, hurricane landfalls have caused around $235 billion damages in 2013 values, and in 2005 Hurricane Katrina alone caused more than $80 billion in damage. The devastation raises
public concern about societal vulnerability to extreme climate (Katz, 2010).

The statistical literature includes some work on exploring possible trends in land-falling hurricanes total damages. Katz (2002) uses a compound Poisson process as a stochastic model for total damage. The model consists of two separate components: one for annual hurricane frequency, and a second one for individual hurricane damage. The resulting analysis suggests no upward trend for hurricane damages recorded between 1925-1995, after normalization due to societal changes. Damages are modeled using a log-normal distribution and occurrences are assumed to follow a homogeneous Poisson process, without any time-varying dynamics. Moreover, the literature includes approaches that study the effect of climate and physical factors on hurricane activity (Elsner & Jagger, 2013). Katz (2002) describes the association between hurricane damages and El Niño. Jagger & Elsner (2006) apply extreme value theory to hurricanes with extreme wind speeds. They assume a homogeneous Poisson process for the occurrences of hurricanes with wind speeds above a threshold, and a generalized Pareto distribution for maximum wind speeds. They find that the quantiles of the distribution of extreme wind speeds vary according to climate factors that affect specific regions differently. Yet another association of hurricane activity with climatic indexes is found in Jagger et al. (2011), where hurricane damages are related to the number of sunspots, as well as to the North Atlantic Oscillation and the Southern Oscillation indexes. Chavas et al. (2012) model the damage index exceedance over a certain threshold using the generalized Pareto distribution with several physical covariates, such as maximum wind speed and continental slope. Murnane & Elsner (2012) use quantile regression to study the relationship between maximum wind speed and normalized economic
losses. Essentially all the papers discussed above focus on estimating trends in hurricane damage and/or its relationship with climate factors. When the point process of hurricane occurrences is modeled, this is done under the simplistic setting of a homogeneous Poisson process.

A fundamental question that remains unanswered by the previously described work is whether the trend of hurricane damage over time is due to the increasing/decreasing frequency or to more/less destructive power of individual hurricanes. These are challenging questions, as natural variability is large and we observe only a handful of hurricanes per season. These issues motivate the presentation of a new statistical method for the analysis of the hurricane data.

In this Chapter, we propose a flexible joint model for inference on hurricane frequency, maximum wind speed and hurricane damage. Our initial assumption is that the point process of hurricane landfalls follows a non-homogeneous Poisson process. As such, the process is characterized by non-constant intensity functions indexed by the hurricane season. Notice that we refer to “intensity” using the point process terminology, and not the climate terminology, where it refers to maximum wind speed. We decompose the intensity functions into normalizing constants, which model annual hurricane frequencies, and density functions, which model normalized intensities within a season. We use a time series model for the normalizing constants. We then take advantage of the flexibility of Bayesian nonparametric methods to model the sequence of non-homogeneous density functions. The proposed approach allows for detailed inferences on both the intra-seasonal variations of hurricane occurrences, and the inter-seasonal changes of hurricane frequencies. The latter
can be considered on time frames shorter than the whole season, for example, monthly. To our knowledge, this is the first statistical analysis of hurricane behavior that takes such a comprehensive approach. Moreover, to study hurricane damage, we treat maximum wind speed and hurricane damage as marks associated with each hurricane occurrence. We extend the method described above to make inference about marks associated with the time of occurrence of the point process events. As a result, we obtain a full probabilistic description of the dynamics of the process intensities and the distribution of the marks. The application is focused on the hurricane data, but the methodology is suitable in general for time-varying seasonal marked Poisson processes.

This chapter is organized as follows. Section 3.2 describes the hurricane data and previous work relevant to this application. We perform an initial analysis of the data ignoring the year of hurricane occurrence, and using a mixture of Beta densities to model the hurricane intensity. This analysis serves to motivate the methodological development, as it clearly suggests that a simple parametric model would not capture the complex shape of the intensity function of occurrences during the hurricane season. Section 3.3 develops the methodology to incorporate dynamic evolution in the analysis, using dependent Dirichlet process mixture models. We explore the problem of data aggregation, and study different aggregation strategies. In Section 3.4, we present the extension of the model to time-varying marks and apply it to maximum wind speed and hurricane damage. Our results indicate that at the peak of the season, there is an increase in the number of hurricane occurrences, a decrease in the median maximum wind speed, and a slight decreasing trend in standardized damage associated with a particular hurricane. Section 3.5 concludes with
a general discussion.

3.2 Hurricane Data

We consider data for 239 hurricane landfalls with reported damages along the U.S. Gulf and Atlantic coasts from 1900 to 2010. The data are available from the ICAT Damage Estimator website (http://www.icatdamageestimator.com). ICAT provides property insurance to businesses and home owners for hurricane and earthquake damage in the United States. The ICAT data are consistent with the landfall summary data of the National Hurricane Center’s North Atlantic hurricane database (HURDAT). The scope of the data is restricted to landfalling hurricanes, as we emphasize the analysis of a marked point process where damage is a mark of key interest. Hurricanes are usually defined as tropical cyclones with maximum wind speed of at least 74 miles per hour (mph). With some abuse of terminology, we use “hurricanes” throughout this work to refer to all the storms in the ICAT dataset. This includes 4 tropical depressions, 63 tropical storms, 54 hurricanes of category 1, 42 hurricanes of category 2, 59 hurricanes of category 3, 14 hurricanes of category 4, and 3 hurricanes of category 5. The classification follows the Saffir-Simpson hurricane scale in Table 3.1. The dataset includes information on the landing date, base damage, normalized damage to current value, category, maximum wind speed and affected states. A detailed description of the data can be found in Pielke et al. (2008) and the ICAT website. In particular, as discussed in Pielke et al. (2008), there is an undercount of damaging storms prior to 1940. This is an important issue that needs to be considered when quantifying possible trends in the number of hurricane occurrences.
<table>
<thead>
<tr>
<th>Category</th>
<th>TD</th>
<th>TS</th>
<th>HC 1</th>
<th>HC 2</th>
<th>HC 3</th>
<th>HC 4</th>
<th>HC 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum wind speed (mph)</td>
<td>&lt; 39</td>
<td>39-73</td>
<td>74-95</td>
<td>96-110</td>
<td>111-130</td>
<td>131-155</td>
<td>&gt; 155</td>
</tr>
<tr>
<td>Counts</td>
<td>4</td>
<td>63</td>
<td>54</td>
<td>42</td>
<td>59</td>
<td>14</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 3.1: Saffir-Simpson hurricane scale. TD: tropical depression; TS: tropical storm; HC 1 to HC 5: hurricane of category 1 to 5. The number of hurricanes corresponding to each category in the data set.

In this application, we consider maximum wind speed and economic damage as marks. Maximum wind speed is defined as the maximum sustained (over one minute) surface wind speed to occur along the U.S. coast. Economic damage is reported as base damage, which is the direct total loss associated with the hurricane’s impact in the year when the hurricane occurred. In order to make all storm damages comparable, a standardization method is used to estimate the damages to a baseline year by extending the normalization method from Pielke et al. (2008); see Section 3.4.2 for details.

The time series of annual hurricane counts is shown in Figure 3.1. Evidently, hurricane occurrence depicts strong inter-annual variability. Moreover, there are indications of discontinuities, which have been thoroughly considered in the literature. In fact, significant shifts during the middle of 1940s, 1960s and in 1995 have been identified in Elsner et al. (2004) and Robbins et al. (2011). The changes in the underlying data collection methods, leading to change points in 1935 and 1960, have been explained in Landsea et al. (1999) and Robbins et al. (2011). To explore the variability within the hurricane season, Figure 3.1 plots also a histogram of hurricane occurrences ignoring the years of the events. The histogram reveals strong intra-seasonal variability, with the peak of the season around September, and
a concentration of hurricanes around June during the early part of the season. Figure 3.2 provides further insight on the variability of hurricane occurrence within the season, where we have now applied aggregation by decades. The distribution of hurricane occurrences within one season varies from decade to decade, and the inter-decadal change of hurricane occurrences varies from month to month. This indicates that the hurricane point process intensity during a given season varies over the decades. Here, we assume that such process corresponds to a non-homogeneous Poisson process (NHPP).

![Figure 3.1](image)

Figure 3.1: Left panel: The time series of annual hurricane occurrences. Right panel: Histogram (with bin width of 10 days) of hurricane occurrences over months after aggregating all hurricanes into one year. The solid and dashed lines denote the point and 95% interval estimates of the corresponding NHPP density, using the Dirichlet process mixture model discussed in Section 3.2.

There is a large body of literature on nonparametric methods to model temporal (or spatial) NHPP intensities and to tackle the analytically intractable NHPP likelihood. Some are based on the log-Gaussian Cox process model (Møller et al., 1998; Brix & Diggle,
Figure 3.2: The number of hurricanes within one season aggregated by decades. In each decade, the number of hurricanes is grouped by months.

2001; Liang et al., 2009), while others use a Gaussian Cox process model (Adams et al., 2009). An approach based on modeling the intensity function using kernel mixtures of weighted gamma process priors is developed in Wolpert & Ickstadt (1998) and Ishwaran & James (2004). The method presented in this work uses nonparametric mixtures to model a density that, up to a scaling factor, defines the NHPP intensity. The approach was originally developed in Kottas (2006) and Kottas & Sansó (2007), with different applications considered by Ihler & Smyth (2007), Ji et al. (2009), Taddy (2010), Kottas et al. (2012a), and Kottas et al. (2012b).

Let $\lambda(t)$ be the NHPP time-varying intensity, with $t$ in a bounded time window $(0, T)$. Inference proceeds by factoring the intensity function as $\lambda(t) = \gamma f(t)$, where $\gamma = \int_0^T \lambda(t) dt$ is the total intensity over $(0, T)$; note that $\gamma < \infty$ based on the local integrability of the NHPP intensity function. Hence, the likelihood function induced by the NHPP
assumption, using the observed point pattern \( \{t_1, \ldots, t_n\} \), is given by
\[
p(\{t_i\}_{i=1}^n | \gamma, f(\cdot)) \propto \exp(-\gamma)\gamma^n \prod_{i=1}^n f(t_i),
\]
indicating that \( f(t) \) and \( \gamma \) can be modeled independently. To develop inference for \( \lambda(t) \), we start by rescaling all the observations to the unit interval, thus setting \( T = 1 \). A convenient choice of distribution that will result in a conjugate prior for \( \gamma \) is the gamma distribution. Alternatively, we can use the reference prior \( p(\gamma) \propto \gamma^{-1}1_{\{\gamma > 0\}} \) (Kottas, 2006). We model \( f(t) \) using the density estimator given by the Dirichlet process (DP) mixture model (Ferguson, 1973; Antoniak, 1974). To complete the model we need to specify a mixing kernel. The kernel of choice in this case is a Beta density, which has the advantages of providing flexible shapes and, being compatible with the compact support of the intensity, avoids edge effect problems. Using the DP stick-breaking representation (Sethuraman, 1994), the model can be formulated in the following terms

\[
t_i | G, \tau \sim f(t_i | G, \tau) = \int_0^1 \text{Beta}(t_i | \mu \tau, (1 - \mu)\tau) dG(\mu), \quad G(\mu) = \sum_{j=1}^{\infty} w_j \delta_{\mu_j}(\mu)
\]

\[
z_j \overset{iid}{\sim} \text{Beta}(1, \alpha); \quad w_1 = z_1, \quad w_j = z_j \prod_{r=1}^{j-1} (1 - z_r), \quad j \geq 2; \quad \mu_j \overset{iid}{\sim} G_0 \quad (3.1)
\]

where \( G_0 \) is the DP centering distribution and \( \alpha \) is the DP precision parameter. In our case, a convenient choice for \( G_0 \) is given by the uniform distribution noting that the Beta mixture kernel is parameterized such that \( \mu \in (0, 1) \) is the mean, and \( \tau > 0 \) is a scale parameter.

We apply this model to the hurricane data ignoring the year index. As shown in Figure 3.1, the estimated density is multi-modal, non-symmetric, and has a non-standard right tail. From this analysis it is clear that a proper description of the hurricane data that assumes an underlying Poisson process requires a non-homogeneous intensity. Although the initial DP mixture model of Beta densities is flexible enough to capture non-standard shapes of intensities within a season, it is not capable of describing the evolution of intensities across
seasons. To address this problem we propose in the next section a dynamic extension of the Beta DP mixture model.

3.3 Modeling Time-Varying Intensities

We seek to model a collection of intensities evolving over years, \( \{ \lambda_k(t) : k \in K \} \), where \( K = \{ 1, 2, \ldots \} \) denotes the discrete-time index set and \( \lambda_k(t) \) is the intensity in the \( k \)th season. The model presented in the previous section uses a DP prior to mix over the mean of a Beta kernel. A temporal extension of such model will have those priors depend on \( k \). To describe the correlation between successive years, the model needs to impose dependence between the priors. As an extension of the DP prior, MacEachern (1999, 2000) proposed to model dependency across several random probability measures. The extension is based on the dependent Dirichlet process (DDP), which provides a natural way to model data varying smoothly across temporal periods or spatial regions. The construction of the DDP is based on the DP stick-breaking definition, where the weights and/or atoms are replaced with appropriate stochastic processes on \( K \). Here, we utilize the “single-p” DDP prior model, where the weights are constant over \( K \), while the atoms are realizations of a stochastic process on \( K \).

3.3.1 Nonparametric dynamic model for Poisson process densities

Denote by \( t_{i,k} \), for \( i = 1, \ldots, n_k \) and \( k = 1, \ldots, K \), the time of the \( i \)th event (hurricane landing date) in the \( k \)th season, where \( K \) is the observed number of seasons and \( n_k \) is the observation count in the \( k \)th season. Recall that \( t_{i,k} \) has been converted to the unit
interval. Following the modeling approach discussed in Section 3.2, the collection of NHPP intensities can be represented by \( \{ \lambda_k(t) = \gamma_k f_k(t) : k \in \mathcal{K} \} \). To introduce dependence on \( \mathcal{K} \), we assume a parametric time series model for \( \{ \gamma_k : k \in \mathcal{K} \} \) and a DDP mixture model for \( \{ f_k(t) : k \in \mathcal{K} \} \). The former is described in Section 3.3.2. The latter is defined as follows:

\[
f_k(t) \equiv f(t \mid G_k, \tau) = \int_0^1 \text{Beta}(t \mid \mu \tau, (1 - \mu) \tau) dG_k(\mu), \quad G_k(\mu) = \sum_{j=1}^{\infty} w_j \delta_{\mu_{j,k}}(\mu)
\]

where the weights \( \{w_j\} \), defined as in (3.1), are the same across seasons. Thus, the model assumes that observations \( t_{i,k} \) in the \( k^{th} \) season arise from a mixture of Beta distributions with component-specific means \( \mu_{j,k} \) and variances \( \mu_{j,k}(1 - \mu_{j,k}) / (\tau + 1) \). The distribution for the mean of the Beta mixture kernel is allowed to evolve over \( \mathcal{K} \), whereas \( \tau \) is common to all \( G_k \).

To impose dependence between the collection of random mixing distributions \( G_k \), we replace \( G_0 \) in (3.1) with a stochastic process for the atoms \( \{ \mu_{j,k} : k \in \mathcal{K} \} \). We thus need a discrete-time process with marginal distributions supported on (0, 1), an appealing choice for which is the positive correlated autoregressive process with Beta marginals (PBAR) developed by McKenzie (1985). For the atom \( \mu_{j,k} \), this is defined through latent random variables as follows:

\[
\mu_{j,k} = v_{j,k} u_{j,k} \mu_{j,k-1} + (1 - v_{j,k}), \tag{3.2}
\]

where \( \{v_{j,k} : k \in \mathcal{K}\} \) and \( \{u_{j,k} : k \in \mathcal{K}\} \) are mutually independent sequences of i.i.d. Beta random variables, specifically, \( v_{j,k} \overset{i.i.d.}{\sim} \text{Beta}(b, a - \rho) \) and \( u_{j,k} \overset{i.i.d.}{\sim} \text{Beta}(\rho, a - \rho) \), with \( a > 0, b > 0 \) and \( 0 < \rho < a \). Using properties for products of independent Beta random variables, it can be shown that (3.2) defines a stationary process \( \{ \mu_{j,k} : k \in \mathcal{K} \} \) with Beta\((a, b)\) marginals. Moreover, the autocorrelation function of the PBAR process is
given by \( \{ \rho ba^{-1}(a + b - \rho)^{-1} \}^m, \ m = 0, 1, \ldots \), and thus \( \rho \) controls the correlation structure of the process.

Although the DDP-PBAR prior for \( G_K = \{ G_k : k \in K \} \) is centered around a stationary process, it generates non-stationary realizations. In particular, if \( \{ \theta_k : k \in K \} \) given \( G_K \) arises from \( G_k \), then \( \mathbb{E}(\theta_k | G_k) = \sum_{j=1}^{\infty} w_j \mu_{j,k} \) and \( \text{Cov}(\theta_k, \theta_{k+1} | G_k, G_{k+1}) = \left( \sum_{j=1}^{\infty} w_j \mu_{j,k} \mu_{j,k+1} \right) - \left( \sum_{j=1}^{\infty} w_j \mu_{j,k} \right) \left( \sum_{j=1}^{\infty} w_j \mu_{j,k+1} \right) \).

The Markov chain Monte Carlo (MCMC) method for inference, discussed in Section 3.3.3 and the Appendix A, is based on a truncation approximation to the DDP prior stick-breaking representation. More specifically, \( G_k \approx \sum_{j=1}^{N} w_j \delta_{\mu_{j,k}} \), with \( w_1, \ldots, w_{N-1} \) defined as in (3.1), but \( w_N = 1 - \sum_{j=1}^{N-1} w_j \). Because the weights are constant across seasons, it is straightforward to choose the truncation level \( N \) to any level of accuracy using standard DP properties. For instance, \( \mathbb{E}\left( \sum_{j=1}^{N} w_j | \alpha \right) = 1 - \{ \alpha / (\alpha + 1) \}^N \), which can be averaged over the prior for \( \alpha \) to estimate \( \mathbb{E}\left( \sum_{j=1}^{N} w_j \right) \). Given a tolerance level for the approximation, this expression can be used to obtain the corresponding value \( N \). The truncated version of \( G_k \) is used in all ensuing expressions involving model properties and inference results.

### 3.3.2 Time series model for the total intensities

The Poisson process integrated intensities \( \{ \gamma_k \} \) can be viewed as a realization from a time series in discrete index space, with positive valued states. We adopt the state-space modeling method with exact marginal likelihood proposed by Gamerman et al. (2013). Unlike other time series models that build from a log-Gaussian distributional assumption, this approach provides a conjugate gamma prior resulting in an efficient MCMC algorithm.
for posterior simulation. The model is defined by the following evolution equation for $\gamma_k$:

$$
g_{k+1} = \frac{1}{\omega} g_k \xi_{k+1}, \quad \xi_{k+1} | \gamma_k, n_{1:k} \sim \text{Beta}(\omega a_k, (1 - \omega) a_k)
$$

where $\omega$ is a discount factor with $0 < \omega < 1$, $\xi_{k+1}$ is a random multiplicative shock, and $n_{1:k}$ denotes the information available up to time $k$.

Denote $n_0$ as the information available initially. Take the initial prior of $\gamma_0 | n_0$ as $\text{Gamma}(a_0, b_0)$. Then, the prior distribution at time $k$ is $\gamma_k | n_{1:k} \sim \text{Gamma}(a_{k|k-1}, b_{k|k-1})$, where $a_{k|k-1} = \omega a_{k-1}$ and $b_{k|k-1} = \omega b_{k-1}$. Based on the NHPP assumption, $n_k | \gamma_k \sim \text{Poisson}(\gamma_k)$, and thus the updated distribution is $\gamma_k | n_{1:k} \sim \text{Gamma}(a_k, b_k)$, where $a_k = \omega a_{k-1} + n_k$ and $b_k = \omega b_{k-1} + 1$. The smoothing updated distribution is

$$
g_{k} - \omega \gamma_{k+1} | \gamma_{k+1}, n_{1:k} \sim \text{Gamma}((1 - \omega) a_k, b_k)
$$

For MCMC posterior inference, we can obtain samples from the full conditionals of the joint vector $\gamma_1, \ldots, \gamma_K$ by first filtering the observations forward to obtain $a_k$ and $b_k$, $k = 1, \ldots, K$, and then sampling $\gamma_k$ backwards, for $k = K, \ldots, 1$, using the distribution in Equation (3.3). Using empirical Bayesian method, we estimate the discount factor $\omega$ by maximizing the joint log-likelihood function defined by the observed predictive distribution

$$
\log \prod_{k=1}^K p(n_k | n_{1:k-1}, \omega).
$$

3.3.3 Implementation details and posterior inference

Inference for the scale parameter of the Beta mixture kernel using the fully aggregated data (see Section 3.2) presented no problems and was quite robust to the choice of the gamma prior assigned to $\tau$. As discussed in more detail in Section 3.3.4, to estimate
evolving hurricane intensities using the DDP mixture model, it is necessary to apply some aggregation of the data into periods of time that comprise more than one year. In this respect, aggregating the data in decades emerges as an appropriate choice. However, the estimation of $\tau$ becomes a challenging problem, since in each decade there are still only a handful of hurricanes. In fact, a simulation analysis indicates that reliable estimation of $\tau$ requires between 50 to 100 observations per time period. This problem can be explained by the fact that $\tau$ partially controls the bandwidth of the Beta kernels, with the width of the kernels in inverse relationship with the size of $\tau$. Thus, when only a few data points are available, $\tau$ will tend to be small allowing wide kernels to use the information from most of the few available data. Such kernels can not capture the multi-modality of the seasonal hurricane intensity. We thus resort to fixing the value of $\tau$ in our analysis of the data aggregated by decade. We assume that the typical width of the Beta kernel corresponds to a month, such that $(1/12)/4$ can be used as a proxy for the corresponding standard deviation \(\{\hat{\mu}(1 - \hat{\mu})/(\tau + 1)\}^{1/2}\), yielding $\tau = 575$ when $\hat{\mu} = 0.5$. This is the value of $\tau$ used in our analysis. We note that informative priors for $\tau$ centered around this value result in similar inferences.

For the centering PBAR process of the DDP prior, we set $a = b = 1$ leading to the default choice of uniform marginal distributions for the $\mu_{j,k}$ covering the entire season between May and November. The DDP prior specification is completed with a uniform hyperprior for the PBAR correlation parameter $\rho$, and a gamma(2, 1) prior for $\alpha$. Finally, we set $N = 50$ for the truncation level in the DDP approximation; note that under the gamma(2, 1) prior for $\alpha$, $E(\sum_{j=1}^{50} w_j) \approx 0.9999578$, using the results discussed in Section
3.3.1.

We implement the DDP-PBAR model using the blocked Gibbs sampler (Ishwaran & James, 2001) with Metropolis-Hastings steps; see the Appendix A for details. Combining the posterior samples for the parameters of the DDP-PBAR model for \( \{f_k(t)\} \) and the posterior samples for the parameters of the time series model for \( \{\gamma_k\} \), a variety of inferences about hurricane intensity functionals can be obtained.

Of particular interest in our application is the average number of hurricanes within a time interval \((t_1, t_2)\) in the \(k^{th}\) season, which is given by \( \Lambda_k(t_1, t_2) = \gamma_k \int_{t_1}^{t_2} f_k(t)dt \). We can also obtain the probability of having a certain number \(x\) of hurricanes within time interval \((t_1, t_2)\) in the \(k^{th}\) season as \( \{(\Lambda_k(t_1, t_2))^x/x!\} \exp(-\Lambda_k(t_1, t_2)) \). As a consequence, the probability of having at least one hurricane within time interval \((t_1, t_2)\) in the \(k^{th}\) season is given by \( 1 - \exp(-\Lambda_k(t_1, t_2)) \). Under the DDP Beta mixture model, \( \int_{t_1}^{t_2} f_k(t)dt = \sum_{j=1}^{N} w_j \int_{t_1}^{t_2} \text{Beta}(t \mid \mu_{j,k}, (1-\mu_{j,k})\tau)dt \).

A further inferential objective is the one-step ahead prediction of the intensity function for the next season, \( \gamma_{k+1} \sum_{j=1}^{N} w_j \text{Beta}(t \mid \tilde{\mu}_{j,k+1}, (1-\tilde{\mu}_{j,k+1})\tau) \). Based on the PBAR construction in (3.2), the conditional distribution for \( \tilde{\mu}_{j,k+1} \) given \( \mu_{j,k} \) and \( v_{j,k+1} \) is a rescaled version of the \( \text{Beta}(\rho, 1-\rho) \) distribution for \( u_{j,k+1} \). Hence, for each \( j = 1, ..., N \), posterior predictive samples for the \( \tilde{\mu}_{j,k+1} \) can be readily obtained given draws for the \( \mu_{j,k} \) and \( v_{j,k+1} \); the former are imputed in the course of the MCMC, the latter can be sampled from their \( \text{Beta}(1, 1-\rho) \) distribution given the MCMC draws for \( \rho \). Therefore, combining with predictive draws for \( \gamma_{k+1} \), full inference is available for forecasting any functional of the hurricane intensity.
3.3.4 Analysis of dynamically evolving hurricane intensities

Data aggregation

The number of landfalling hurricanes with reported damages during 1900-2010 in the U.S. is 239. On average there are merely 2 or 3 hurricanes every year, with no hurricane in some years, e.g., 1922-1925 and 2009. Thus, the first practical problem we face is that of data scarcity. When modeling the data at the yearly level, the challenge is that it is difficult to analyze a process with so few realizations per year. Hence, we consider aggregating the data over periods of five and ten years, and compare the results under the two different levels of aggregation.

Figure 3.3: Under the two distinct levels of data aggregation, posterior mean estimates for the hurricane intensity in 2011-2015 (left panel) and posterior densities for the probability of at least one hurricane in May for 2011-2015 (right panel).

Aggregation over a period of time is based on the assumption that the NHPP
densities for all the years corresponding to the aggregated period are the same. For the five year aggregation we have 22 different intensities and for the decadal aggregation we have 11. Data aggregation does not effect the estimation of normalizing constants \( \{ \gamma_k \} \). In fact, we can apply the model for the \( \{ \gamma_k \} \) proposed in Section 3.3.2 to the yearly data, and then aggregate. Figure 3.3 provides results to compare the two aggregation strategies in the context of forecasting the hurricane intensity and one of its functionals in the next five years 2011-2015. Encouragingly, the results are very similar under the two levels of data aggregation.

![Boxplots of posterior samples for the average number of hurricanes in the month of September across five-year periods from 1900 to 2010.](image)

Regarding the analysis of historical data, we focus on the month of September. In fact, for the Atlantic hurricane season, August, September and October (ASO) are very important months, as 95% of Saffir-Simpson Category 3, 4, and 5 hurricane activity occurs during August to October (Landsea, 1993). In particular, September is the most frequently-
Figure 3.5: Posterior mean estimates (solid line) and 95% intervals (grey bands) of the hurricane intensity during 1971-2010. Points along the horizontal axis correspond to the occurring month. Figure 3.4 shows the estimated average number of hurricanes in September under the five year data aggregation. We observe a strong variability, in particular for the periods 1921-1925, 1966-1970 and 1991-1995. This can be attributed to the fact that during 1921-1925 there was no hurricane in September. Moreover, there was only one hurricane in September during 1966-1970, but there were 7 hurricanes in September during both 1961-1965 and 1971-1975. Finally, there was no hurricane in September during 1991-1995, but 10 hurricanes occurred in September during 1996-2000. Thus, even though the prior model is imposing some smoothness, posterior inference results are still strongly affected by the scarcity of observations, even at the level of a five year period. Our resulting inference in the
five-year aggregation level reflects the strong variability of hurricane counts in September. More specifically, the clear separation of the posterior distributions for the different periods mentioned above, gives a probabilistic assessment of significant breakpoints. These are in agreement with the change points detected in Elsner et al. (2004) and Robbins et al. (2011) for the counts over all months. However, in this work we focus on revealing possible long term trends rather than on anomaly detection. Thus, on the basis of these analyses, for the rest of this Chapter we focus on data aggregated over decades.

![Boxplots of posterior samples of the mean number of hurricanes in early season (May and June) by decade (left panel), and in September by decade (right panel). In both panels, the solid dots indicate the corresponding observed numbers of hurricanes.](image)

**Evolving hurricane intensities across decades**

Figure 3.5 presents the estimated intensity functions in the most recent four decades. The estimates fit the data very well, correctly capturing the peaks in ASO and
tails in June and November. They show some similarities between the decades, but they adapt to the characteristic of the distribution of hurricane events in each decade. An important product of our probabilistic analysis is the average number of hurricanes in a given time period, which as discussed in Section 3.3.3, requires the posterior distribution for both $\gamma_k$ and $G_k$. In Figure 3.6 we present the distributions for the mean number of hurricanes in the peak month of September and the off-season months of May and June, along with the associated observed number of hurricanes. Inference based on our model smooths the data through the decades, especially when a small number of observations are available. Overall, the distribution of the mean number of hurricanes in each decade matches the observations quite well. Both panels depict an increasing trend in the first four decades as well as during the most recent three decades. The former may be an artifact of the under-reporting during the beginning of the 20th Century. While the latter is very subtle for the off-season months, it is very strong for the month of September. In fact, the last decade depicts an average number of hurricanes in the peak of the season which is substantially higher than any other decade on record.

3.4 DDP Model For Seasonal Marked Poisson Processes

Here, we extend the DDP model, developed in the previous section, to a seasonal marked Poisson process. A marked Poisson process (MPP) refers to a Poisson process with an associated random variable or vector for each event. In our application, $\{t_{i,k} : i = 1, ..., n_k\}$ is a point pattern on $(0, T)$ and the marks can be denoted as $\{y_{i,k} : i = 1, ..., n_k\}$ on mark space $Y$. Thus, the realization from the marked point process in the $k^{th}$ decade
is \( \{(t_{i,k}, y_{i,k}) : t_{i,k} \in (0, T), y_{i,k} \in Y\} \). A MPP can be defined as a Poisson process on the joint marks-points space with intensity function \( \varphi \) on \((0, T) \times Y\). In particular, the marking theorem (Møller & Waagepetersen, 2004) states that a MPP is a NHPP with intensity function given by \( \varphi(t, y) = \lambda(t)f(y | t) \), where \( \lambda(t) \) is the marginal temporal intensity function, and the conditional mark density \( f(y | t) \) depends only on the current time point \( t \).

### 3.4.1 The DDP-AR model

We extend the methodology from Taddy & Kottas (2012) for MPPs based on joint mixture modeling on the marks-points space. This modeling approach yields flexible inference for both the marginal temporal intensity and for the conditional mark distribution. Here, it is utilized to develop a model for the collection of hurricane MPPs evolving over decades. We will refer to the full model as DDP-AR model, since in addition to the PBAR structure, it incorporates autoregressive processes to model the conditional evolution of marks over time.

The marks are given by the maximum wind speed for each hurricane and the associated economic damages. Instead of using the total dollar amount of hurricane damage, we define a standardized damage, which is calculated as a proportion of total wealth with respect to a reference region and a baseline year (see Section 3.4.2). The resulting NHPP is defined in a three dimensional space comprising time, maximum wind speed, and standardized damage. Maximum wind speed and standardized damage are transformed by taking logarithm and subtracting the global average of the log-transformed values. We denote \( y_{i,k} \) and \( z_{i,k} \) as, respectively, the transformed maximum wind speed and the transformed stan-
standardized damage of the $i^{th}$ hurricane in the $k^{th}$ decade. For the three dimensional intensity function, $\varphi_k(t,y,z)$, we use the factorization $\gamma_k f_k(t,y,z)$, where $\{\gamma_k\}$ follows the time series model presented in Section 3.3.2. Regarding the density function, we use a DDP mixture with a product of univariate kernel densities for time and marks. Thus, the dependence among time and marks is introduced by the mixing distribution. We retain the Beta kernel density for time and use Gaussian kernel densities on the log scale for the two marks, mixing on the mean of each kernel component. Hence, the DDP mixture model for $f_k(t,y,z)$ can be expressed as:

$$\int \text{Beta}(t \mid \mu_{\tau}, (1 - \mu)\tau)N(y \mid \nu, \sigma^2)N(z \mid \eta, \zeta^2) \, dG_k(\mu, \nu, \eta) \quad (3.4)$$

where $G_k(\mu, \nu, \eta) = \sum_{j=1}^{N} w_j \delta(\mu_{j,k}, \nu_{j,k}, \eta_{j,k})(\mu, \nu, \eta)$. The locations $\nu$ and $\eta$ of the normal kernels are allowed to change across decades. The scales $\sigma^2$ and $\zeta^2$ are the same across decades serving as adjusting parameters for the bandwidth of the kernels. Conditionally conjugate inverse gamma priors are assumed for $\sigma^2$ and $\zeta^2$.

Dependence across decades for maximum wind speeds and standardized damages is obtained through AR(1) processes for the respective kernel means $\{\nu_{j,k} : k \in \mathcal{K}\}$ and $\{\eta_{j,k} : k \in \mathcal{K}\}$:

$$\nu_{j,k} \mid \nu_{j,k-1} \sim N(\beta \nu_{j,k-1}, \sigma_1^2), \quad \eta_{j,k} \mid \eta_{j,k-1} \sim N(\phi \eta_{j,k-1}, \sigma_2^2)$$

with inverse gamma priors assigned to $\sigma_1^2$ and $\sigma_2^2$, and uniform priors on $(-1, 1)$ placed on $\beta$ and $\phi$. Since the DDP prior structure for $G_K = \{G_k : k \in \mathcal{K}\}$ in (3.4) extends the one for the DDP-PBAR model, we retain the result about non-stationary realizations given $G_K$, extending the argument in Section 3.3.1. When the random measures $G_k$ are integrated
out, we obtain $E(y_k) = 0$, $\text{Var}(y_k) = E(\sigma^2) + (1 - \beta^2)^{-1}E(\sigma_1^2)$ and $\text{Cov}(y_k, y_{k+1}) = \beta(1 - \beta^2)^{-1}E(\sigma_1^2)$, with analogous results for the $z_k$. These expressions can be of help for prior specification.

The MCMC method for the DDP-AR model involves an extension of the posterior simulation algorithm described in the Appendix. As the marks are associated with normal AR(1) processes and conditionally conjugate priors are used, all the parameters associated with marks have closed-form full conditionals. Finally, since the normalizing factors (required for the standardization of damages) corresponding to the period 2005–2010 are not available, the MCMC algorithm includes steps to impute the missing standardized damages for those years.

### 3.4.2 Standardization of hurricane damages

The purpose of standardizing hurricane damages is to isolate societal and spatial factors that affect the amount of damage, and are not considered in the model. There exist several methods to adjust the economic damages of past hurricanes to today’s value (Pielke et al., 2008; Schmidt et al., 2010; Collins & Lowe, 2001). Here, we define standardized damage as an extension to the method in Pielke et al. (2008).

The hurricane data set includes base damage and normalized damage. Base damage is calculated as the total landfall year dollar value of the damage caused by a hurricane. Such amount is converted to the dollar value corresponding to the latest year in the record by normalizing for inflation, wealth, and population over time. Denote inflation, wealth per

---

1 The code to implement the DDP-AR model (as well as the DDP-PBAR model) is available from the first author’s website at http://users.soe.ucsc.edu/~sxiao/research.html#software
capita, and affected county population in year t as $I_t$, $W_t$, and $P_t$, respectively. Equation (3.5) shows the normalization of the damage due to a hurricane landing in year t to values in year s:

$$\text{normalized.damage}_s = \text{base.damage}_t \times \frac{I_s}{I_t} \times \frac{W_s}{W_t} \times \frac{P_s}{P_t}.$$  (3.5)

This normalization method yields the estimated damages of all hurricanes in today’s value but in the same region, e.g., the damages caused by Katrina 2005 if it occurred under societal conditions in Louisiana affected counties in 2013.

To make hurricane damages comparable, we have to adjust for inflation, and account for the fact that much more damage will be caused if the hurricane lands in densely populated and wealthier counties than in scarcely populated and poor regions. Thus, we have to remove both a spatial and societal factor from the damage, so that the model can explore the pure association between damages and climate variability. Hence, we define standardized damage as

$$\text{standardized.damage} = \frac{\text{base.damage}_t}{I_t \cdot W_t \cdot P_t}.$$  

Such quantity can be interpreted as a base damage normalized to a reference year’s value in a reference region; in the reference year and region, the inflation factor, wealth per capita and population are all equal to 1. This method removes the difference in hurricane damages due to the landing years and locations. Neumayer & Barthel (2011) and Chavas et al. (2012) developed similar ideas normalizing damages by using $\text{base.damage}_t/\text{wealth}_t$, where wealth$_t$ is the total wealth of the affected regions. They interpret the standardized damage as a relative damage, termed actual-to-potential-loss ratio. Note that the denominator we
use, $I_t \cdot W_t \cdot P_t$, is an approximation of wealth$_t$. All inferences presented in Section 3.4.4 that involve hurricane damage refer to standardized damage. Note that, if the normalizing factors are provided, actual hurricane damages for a given affected region and year can be obtained from standardized damages. It is important to notice that the normalizing factors prior to 1925 have larger uncertainties compared to those for later periods (Pielke et al., 2008). This problem is compounded with the already mentioned issue of underreporting of hurricanes in the early part of the 20th Century. The reader should keep this in mind when interpreting the results in the following sections.

Figure 3.7: Data box plots across decades for log-transformed base damages (left panel), damages normalized to current values (middle panel), and standardized damages (right panel).

To visualize the effect of the conversion on damage values, Figure 3.7 shows three different calculations for hurricane damage and their change over decades. The base dam-
age depicts an increasing trend over decades, which disappears after normalization and standardization.

3.4.3 Inference

For a marked point process the typical inference of interest is for the distribution of the marks, conditional on time. To obtain inference about different functionals of the conditional mark distribution, we use the available posterior samples of the joint density $f_k(t, y, z)$. Specifically, conditional inference for maximum wind speed is obtained from

$$f_k(y \mid t, G_k) = \frac{f_k(y, t \mid G_k)}{f_k(t \mid G_k)} = \frac{\sum_{j=1}^N w_j \text{Beta}(t \mid \mu_{j,k}, \tau, (1 - \mu_{j,k}) \tau) N(y \mid \nu_{j,k}, \sigma^2)}{\sum_{j=1}^N w_j \text{Beta}(t \mid \mu_{j,k}, \tau, (1 - \mu_{j,k}) \tau)}$$

$$= \sum_{j=1}^N w^*_j(t) N(y \mid \nu_{j,k}, \sigma^2) \tag{3.6}$$

where $w^*_j(t) = \frac{w_j \int_{t_1}^{t_2} \text{Beta}(t \mid \mu_{j,k}, \tau, (1 - \mu_{j,k}) \tau) dt}{\sum_{j=1}^N w_j \int_{t_1}^{t_2} \text{Beta}(t \mid \mu_{j,k}, \tau, (1 - \mu_{j,k}) \tau) dt}$. Of particular importance is the distribution of maximum wind speed conditional on a specific time period, e.g., the peak season ASO or a particular month. Suppose that the time period of interest corresponds to the interval $(t_1, t_2)$. The density conditional on $(t_1, t_2)$ can be developed as

$$f_k(y_0 \mid t \in (t_1, t_2), G_k) = \lim_{\Delta y_0 \to 0} \frac{1}{\Delta y_0} \frac{\text{Pr}(y \in (y_0, y_0 + \Delta y_0), t \in (t_1, t_2) \mid G_k)}{\text{Pr}(t \in (t_1, t_2) \mid G_k)}$$

$$= \sum_{j=1}^N h^*_j N(y_0 \mid \nu_{j,k}, \sigma^2) \tag{3.7}$$

where $h^*_j = h^*_j(t_1, t_2) = \frac{w_j L^2 \int_{t_1}^{t_2} \text{Beta}(t \mid \mu_{j,k}, \tau, (1 - \mu_{j,k}) \tau) dt}{\sum_{j=1}^N w_j L^2 \int_{t_1}^{t_2} \text{Beta}(t \mid \mu_{j,k}, \tau, (1 - \mu_{j,k}) \tau) dt}$.

In Equations (3.6) and (3.7), both the weights, $w^*_{j,k}(t)$, $h^*_{j,k}$, and the mixing components $\nu_{j,k}$ change with the decade index $k$; importantly, the former are time dependent thus allowing local learning under the implied location normal mixtures. Hence, the model has the flexibility to capture general shapes for the conditional mark distribution which
are allowed to change across decades in non-standard fashion. Analogous expressions hold for the conditional distribution of standardized damage. Moreover, since Equation (3.4) provides the joint density of time, maximum wind speed, and standardized damage, we can obtain inference for a mark conditional on an interval of the other mark and an interval of time. For instance, we can explore the distribution of damage conditional on the hurricane category as defined by different intervals of maximum wind speed; see Table 3.1.

The time evolution of hurricane occurrences and the marks are controlled by autoregressive processes. One-step ahead prediction of joint time-mark distributions can be obtained by extending the method described in Section 3.3.3 with additional sampling for the \( \nu_{j,k+1} \) and \( \eta_{j,k+1} \) from the AR(1) processes that form the building blocks of the DDP prior.

### 3.4.4 Results

We applied the DDP-AR model to the full data set involving hurricane occurrences across decades and the associated maximum wind speeds and standardized damages. The hyperpriors for the time component of the DDP mixture model were similar to the ones discussed in Section 3.3.3 for the DDP-PBAR model; \( \tau \) was again fixed. For the variances of the Gaussian mixture kernels and the variances of the corresponding AR(1) processes for the DDP prior, we used \( \sigma^2 \sim IG(3, 2) \), \( \zeta^2 \sim IG(3, 10) \) and \( \sigma_1^2 \sim IG(3, 2) \), \( \sigma_2^2 \sim IG(3, 10) \). Here, the shape parameter of each inverse gamma prior is set to 3, which is the smallest (integer) value that ensures finite prior variance. The prior means were specified using the expressions for the marginal variances of maximum wind speed and standardized damage (see Section 3.4.1) with \( \beta \) and \( \phi \) replaced by their prior mean at 0. In particular, we set
\[ E(\sigma^2) = E(\sigma_1^2) = 0.5(R_y/4)^2 \quad \text{and} \quad E(\zeta^2) = E(\sigma_2^2) = 0.5(R_z/4)^2, \]

where \( R_y \) and \( R_z \) denotes the range of the \( y_{i,k} \) and \( z_{i,k} \), respectively.

The posterior distribution for the number of distinct mixing components is supported by values that range from 10 to 16. The 95\% posterior credible interval for \( \rho \) is given by \((0.73, 0.87)\), resulting in a \((0.59, 0.79)\) 95\% credible interval for the PBAR correlation. On the other hand, the 95\% posterior credible intervals for \( \beta \) and \( \phi \) are, respectively, \((-0.14, 0.79)\) and \((-0.24, 0.81)\), indicating more variability in the estimated correlation of the AR(1) centering processes for the DDP prior. Retaining the uniform priors for \( \rho \), \( \beta \) and \( \phi \), we performed a prior sensitivity analysis for the variance hyperparameters. The parameters \( \sigma^2 \) and \( \sigma_1^2 \) associated with maximum wind speed are relatively sensitive to the prior choice, while the parameters \( \zeta^2 \) and \( \sigma_2^2 \) for standardized damage are quite stable. Overall, posterior inference results are robust to moderate changes in the prior hyperparameters.

For inference, we focus on the densities of maximum wind speed and logarithmic standardized damage conditional on events occurring in the early season and the peak season. Figure 3.8 shows the comparison between the maximum wind speed densities conditional on June and September in each decade. We observe that maximum wind speeds in September are higher than in June, for all decades. In the 1960s the density has a very long left-hand tail, even showing evidence of two modes. Noteworthy in the last four decades is the increasing accumulation of density on lower values of maximum wind speed. The fact that maximum wind speeds in September are decreasing is confirmed by the plot in the lower panel of Figure 3.8, where both point and interval estimates support a decreasing trend for the median maximum wind speed in September. In particular, after peaking at
more than 110 mph in the 1920s, the posterior point estimate has settled at around 85 mph in the last decade.

Figure 3.9 (top left panel) shows the density of standardized damages (on the log scale) conditional on the early season and the peak season. The densities of standardized damages in MJJ (May-June-July) are quite similar throughout all decades, while the densities in ASO show a moderate decreasing trend across decades. Figure 3.9 (bottom left panel) plots point and interval estimates for the median standardized damage in the original scale. From 1900 to 1940, the estimated median standardized damage of one hurricane in ASO is around twice as large as that in MJJ. However, from 1941 to 2010, the median standardized damage in ASO depicts significant variability, with some indication of a slight decreasing trend across decades. These results are similar to the ones reported in Katz (2002) and Pielke et al. (2008), based on essentially the same data set, albeit under different damage normalization methods. In particular, Katz (2002) normalizes the damage during 1925–1995 to 1995 values and uses a log-normal distribution to fit the damage of individual storms, finding only weak evidence of a trend in the median of log-transformed damage. Likewise, in Pielke et al. (2008) hurricane damage is normalized to 2005 values. In this case, the conclusion is that there is no long-term increasing trend in hurricane damage during the 20th century, once societal factors are removed. We also note here that Neu- mayer & Barthel (2011) detected a significant negative trend in hurricane damage. Their results are based on the same damage standardization method with the one we use, but for a different data set comprising hurricane damages from 1980–2009 in the US and Canada.
Figure 3.8: Top panel: the density of maximum wind speed conditional on June and September for all decades. Bottom panel: posterior expectation and 95% interval (red band for September; grey band for June) for the median maximum wind speed in June and September versus decade.
Figure 3.9: Top panel: the density of logarithmic standardized damage conditional on MJJ (May-June-July) and ASO (August-September-October). Bottom panel: Posterior expectation and 95% interval (red band for ASO; grey band for MJJ) for the median standardized damage of one hurricane in MJJ and ASO.
Figure 3.10 focus on the analysis of damage, conditional on the seven different types of hurricanes that occurred during ASO. The top panel reports the densities for logarithmic standardized damage conditional on the different hurricane categories. The bottom panel reports the posterior expectations for the corresponding median standardized damage. Overall, we observe that the higher the category the larger the standardized damages tend to be. Standardized damages were very similar for the hurricanes recorded in ASO of decade 1971–80, which is reflected in both types of inference shown in Figure 3.10. Standardized damages for TDs and TSs have indistinguishable distributions. Likewise, at the opposite end of the scale, damages due to HC4 and HC5 hurricanes are very similar. This is also due to the data sparseness of TDs and HC5 hurricanes (only 4 TDs and 3 HC5 hurricanes).

Figure 3.11 presents the bivariate densities of maximum wind speed and logarithmic standardized damage given the ASO period, for each decade. The last panel corresponds to the forecast density for 2011–2020. We note that only a handful of observations correspond to ASO in each particular decade. Thus, the results in Figure 3.11 are possible owing to our model’s ability to borrow strength from all the available data. Noteworthy are the positive association between maximum wind speed and damage after the third decade, and the changes in the density shapes across the decades, especially 1961–1970 and 1991-2000. We also note the decrease in maximum wind speeds, starting in 1961–1970. Overall, from 1961, both the maximum wind speed and standardized damage have a general decreasing trend. This is a reflection of the fact that fewer hurricanes with extremely high maximum wind speed have occurred in recent decades. Figure 3.12 shows the expected conditional
densities of logarithmic standardized damages given maximum wind speed in ASO. The 
data concentrate on the range where the maximum wind speed is between 50mph and 
150mph, while outside of this range the posterior expectation reverts to the mean. Regard-
ing previous related work, Murnane & Elsner (2012) modeled the relationship between wind 
speed and normalized economic loss as exponential through quantile regression methods, 
using all hurricanes in the 20th century. Our methodology allows for a more comprehensive 
investigation of the relationship between hurricane damage and maximum wind speed, in 
particular, it enables study of its dynamic evolution across decades, without the need to 
rely on specific parametric regression forms.
Figure 3.10: Top panel: the density of logarithmic standardized damage in ASO given the seven maximum wind speed categories defined in Table 3.1. Bottom panel: Posterior expectation for the median standardized damage in ASO for the seven maximum wind speed categories.
Figure 3.11: Bivariate densities of maximum wind speed (mph) ($x$-axis) and logarithmic standardized damage ($y$-axis) in ASO across decades. The dots correspond to observations in ASO.
Figure 3.12: Posterior means for the conditional expectation of logarithmic standardized damage given maximum wind speed in ASO across decades.
3.4.5 Model assessment

The modeling approach is based on the assumption of a NHPP over the joint marks-points space. To check the NHPP assumption, we use the Time-Rescaling theorem (Daley & Vere-Jones, 2003), according to which, in each decade, the cumulative intensities between successive (ordered) observations, \( \gamma_k \int_{t_{i-1,k}}^{t_{i,k}} f_k(t) dt \), are independent exponential random variables with mean one. Thus, \( 1 - \exp \left( -\gamma_k \int_{t_{i-1,k}}^{t_{i,k}} f_k(t) dt \right) \) are independent uniform(0,1) random variables. Likewise, the Poisson process assumption for the marks implies that the sets of random variables defined by the c.d.f. values of the conditional mark distributions, \( \{ F_k(y_{i,k} \mid t_{i,k}) \} \) and \( \{ F_k(z_{i,k} \mid t_{i,k}) \} \), are independent uniform(0,1) random variables. Hence, the NHPP assumption over both time and marks can be checked by using the MCMC output to obtain posterior samples for each of the three sets of random variables above, in each decade. Figure 4.12 shows the Q-Q plots of estimated quantiles for time, maximum wind speed, and standardized damage versus the theoretical uniform distribution, for the last five decades. All Q-Q plots are all close to the 45° line, especially in consideration of the limited sample sizes in each decade.

As discussed earlier, Figures 3.5 and 3.6 provide visual goodness-of-fit evidence for the model on hurricane occurrences, by comparing different types of model-based inferences to the corresponding observations. Similar evidence is provided in Figure 3.11 for the maximum wind speed and log-damage relationship. We also explored other functionals of the model obtaining similar results. In addition, we performed posterior predictive checks to study the model’s ability to predict the marks in the 11th decade, based on the data of the previous 10 decades. In particular, we implemented the model using only the 204 hurricanes...
from 1900–2000, and obtained the posterior predictive density of maximum wind speed and logarithmic standardized damage in ASO of the 11th decade (2001–2010). Figure 3.14 shows the posterior predictive densities superimposed on the histograms of corresponding observations in ASO of 2001–2010. The histogram in the left panel corresponds to 28 hurricanes, whereas the one in the right panel corresponds to only 16 hurricanes, since the damages of the other 12 hurricanes are missing. We notice that the predictions are fairly compatible with the cross-validation data.
Figure 3.13: Posterior Q-Q plots (mean and 95% interval) of estimated quantiles against the theoretical uniform(0, 1) for: time (left panel), maximum wind speed given time (middle panel), and standardized damage given time (right panel). Results are shown for the last five decades.
Figure 3.14: Cross-validated posterior predictive densities in ASO of decade 2001–2010: the left panel corresponds to maximum wind speed, and the right panel to logarithmic standardized damage. The histograms plot the associated observations in ASO of 2001–2010.
3.5 Conclusion

We have developed a Bayesian nonparametric modeling method for seasonal marked point processes and applied it to the analysis of hurricane landfalls with reported damages along the U.S. Gulf and Atlantic coasts from 1900 to 2010. Our basic assumption is that hurricane occurrences follow a non-homogeneous Poisson process, with the focus on flexible modeling for dynamically evolving Poisson process intensities. The proposed DDP-PBAR model builds from a DDP mixture prior for the normalized intensity functions based on a PBAR process for the time-varying atoms, and a parametric time-varying model for the total intensities. Inference for different Poisson process functionals can be obtained by MCMC posterior simulation. To incorporate time-varying marks into the inferential framework for our motivating application, we have extended the DDP-PBAR mixture model by adding DDP-AR components for maximum wind speed and economic damages associated with each hurricane occurrence.

In the analysis of the hurricane data, we have used aggregation to study the dynamic evolution of hurricane intensity over decades. The model uncovers different shapes across decades which however share common features with respect to the off-season in May and June and the peak month of September. The results indicate an increase in the number of landfalling hurricanes and a decrease in the median maximum wind speed at the peak of the season across decades. In the off season, both the number of hurricanes and the maximum wind speed show little variation across decades. To study economic loss as a mark, we have introduced standardized damage to adjust hurricane damages such that they are comparable both in time and space. We found a slight decreasing trend in standardized
damage of hurricanes in the peak season, which is also present conditional on the distinct hurricane categories.

With respect to the scientific context of the motivating application, our work provides a general framework to tackle different practically relevant problems. The key distinguishing feature of our approach relative to existing work involves the scope of the stochastic modeling framework under which the various inferences are obtained. As discussed in the Introduction, current work is limited to either estimating trends in hurricane occurrences at the annual level or estimating the hurricane intensity based on the fully aggregated data, thus ignoring dynamics across years. Moreover, when incorporating information on marks, existing approaches oversimplify the underlying point process structure by imposing homogeneity for the hurricane intensity. These assumptions are suspect as demonstrated with the exploratory data analysis of Section 3.2. The proposed Bayesian nonparametric methodology enables flexible estimation of dynamically evolving, time-varying hurricane intensities within each season, and therefore has the capacity to capture trends during particular periods within the hurricane season. The full inferential power of the modeling framework is realized with the extension to incorporate marks, which are included as random variables in the joint model rather than as fixed covariates as in some of the previous work. From a practical point of view, the key feature of the model for the point process over the joint marks-points space is its ability to provide different types of general conditional inference, including full inference for dynamically evolving conditional mark densities given a time point, a particular time period, and even a subset of marks.

In summary, the focus of this work has been in developing a model that can
quantify probabilistically the inter-seasonal and intra-seasonal variability of occurrence of a random process and its marks, jointly, and without restrictive parametric assumptions. The model is particularly well suited for the description of irregular long term trends, which may be present in the observations or in subsets of the records. To enhance the forecasting ability of the model, future work will consider extensions to incorporate external covariates (such as pre-season climate factors) in a similar fashion to Katz (2002), Jagger et al. (2011), and Elsner & Jagger (2013), albeit under the more general statistical modeling framework developed here.
Chapter 4

Bayesian Nonparametric Modeling for Renewal Processes

4.1 Introduction

Poisson processes are the most frequently used stochastic models for temporal or spatial point patterns. Their theoretical properties are fully known and understood, and, most importantly from an inferential point of view, they are very tractable. Indeed, it is possible to perform likelihood based or Bayesian model fitting for Poisson processes with general intensity functions. The literature includes several nonparametric modeling methods; e.g., Møller et al. (1998), Wolpert & Ickstadt (1998), Brix & Diggle (2001), Ishwaran & James (2004), Kottas (2006), Kottas & Sansó (2007), Adams et al. (2009) and Taddy & Kottas (2012). However, to achieve distributional flexibility beyond the Poisson assumption, as well as to model various clustering behaviors, it is desirable to
consider non-Poisson point processes. Flexible inference for such processes is hampered by
the difficulty of handling likelihood functions with normalizing constants that depend on
the parameters of interest in ways that can make computations prohibitively expensive.

One way to extend the temporal Poisson process on the real line to more general models
is to consider renewal processes. These are counting processes where the time intervals
between successive arrivals are identically and independently distributed according to an
arbitrary continuous distribution on the positive real line. Unlike Poisson processes, which
are memoryless, renewal processes possess a Markovian structure. Renewal processes find
a number of applications, such as modeling of earthquakes occurrences, software reliability,
hardware maintenance, and queuing systems, among others.

The objective of this work is to develop a model for renewal process which balances
both flexibility and computability. First, the model should enable flexible shapes of inter-
arrival time densities and have the capability of capturing both clustering and declustering
point patterns. Second, an efficient posterior simulation algorithm should be provided while
properly handling the normalizing constant in the likelihood. Among the existing related
work, some of them fit parametric models to renewal processes by specifying a probability
distribution for the inter-arrival times. A Weibull distribution is commonly used for data
that exhibit clustering patterns. Alvarez (2005), for example, used a Weibull distribution
for the time between events in the analysis of earthquake data. Zhao & Nagaraja (2011)
applied renewal theory to the longitudinal studies of lupus and its periodical flare-ups, using
exponential, gamma and Weibull distributions for the inter-arrival density function. These
parametric models can be only applicable in some circumstances and lack of capability to
capture nonstandard shapes of inter-arrival times. There are some more flexible approaches
to model the renewal distribution are considered in the literature on queueing systems.
However, the normalizing constant of the likelihood, corresponding to the censored waiting
time of the \( n + 1 \)st event is always ignored for the sake of tractability. The problem is
thus simplified to that of density estimation of inter-arrival times. In this context Wiper
et al. (2001) developed a mixture of gamma distributions for general density estimation.
This model is interesting, as a gamma distribution has a coefficient of variation that can
take any values. As will be seen in the next section, this is an important feature, as the
coefficient of variation controls the clustering properties of the process. Along similar lines,
Ausín et al. (2007) and Thümmler et al. (2006) use mixtures of Erlang distributions. Erlang
distributions belong to the gamma family and are central to the methods proposed in this
work.

Here, we propose to use a mixture of Erlang distributions with common scale
parameter to model the inter-arrival time. This is a parsimonious feature of the model that
does not undermine its flexibility. From a computational point of view, it has the advantage
of enabling efficient posterior simulation. The mixture of Erlang distributions with common
scale has been proposed in Tijms (1994) for queuing data analysis and applied in actuarial
science by Lee & Lin (2010), where it is used as a model for density estimation. To our
knowledge, it has never been used to model inter-arrival times in the point process literature.
In Lee & Lin (2010), the mixture weights are estimated by EM algorithm. In this work,
the mixture weights are defined through a probability distribution function that is modeled
nonparametrically with a Dirichlet process prior. Besides the flexibility and computability,
this model has two important advantages. First, for those parametric distributions whose Laplace transforms do not have closed forms, our model can be used to obtain their Laplace transforms, thus renewal function and $K$ function, empirically (see section 4.3.1). Second, in our model, the choice of the Dirichlet process centering distribution and precision parameter control clustering or declustering patterns for the point processes, which can be informed by the prior specification (see section 4.2.3).

This chapter is organized as follows. Section 4.2 introduces the model formulation, and explores the properties of the model, focusing on its flexibility in producing different $K$ functions. A purposely designed Monte Carlo approach to explore the posterior distribution of the model parameters, as well as posterior inferences drawn from the model are also introduced. In Section 4.3, we fit the proposed model to several data sets. These include two simulations with declustering patterns, two simulations with clustering patterns, and earthquake and coal mining disaster data. Our results show that the model can successfully capture different clustering patterns for the point processes. The model checking experiments also confirm applicability of the model assumption on the real-world data sets.

4.2 Model Formulation

4.2.1 The mixture model

We define $H(t)$ as the last observation up to time $t$. The density of the inter-arrival time, $f(t - H(t))$, where $t \in (H(t), \infty)$, is modeled as a mixture of Erlang distributions with
common scale parameter. This is given as

$$f(t - H(t) \mid \theta, G) = \sum_{j=1}^{J} w_j \Gamma(t - H(t) \mid j, 1/\theta), \quad t > H(t). \tag{4.1}$$

Notice that, for each component of the mixture we have a gamma density where the shape parameter $j$ is a specified integer. This corresponds to the so-called Erlang distribution. The weights, $\{w_j, j = 1, \ldots, J\}$, are nonnegative. The number of components $J$ and the weights are both random. The latter are generated by discretizing a distribution function $G$ supported on $\mathbb{R}^+$. In particular, $w_j = G(j\theta) - G((j-1)\theta)$ for $j = 1, \ldots, J-1$, and the last weight is $w_J = 1 - \sum_{j=1}^{J-1} w_j$. Thus, for a point pattern $t_0, \ldots, t_n$, we have that

$$f(t_k - t_{k-1} \mid \theta, G) = \sum_{j=1}^{J} \left[ G(j\theta) - G((j-1)\theta) \right] \frac{1}{(j-1)!\theta^j} (t_k - t_{k-1})^{j-1} e^{-\frac{(t_k-t_{k-1})}{\theta}}, \quad k = 1, \ldots, n.$$ 

Notice that $\theta$ is a scale parameter common to all mixture components. $\theta$ also determines the precision of the CDF discretization for the weights. As $\theta$ decreases, the model tends to have more Erlang mixture components.

Lee & Lin (2010) has shown that this class is dense in the space of positive continuous distributions. Assuming that $G$ is a distribution function with support on $[0, \infty)$, when $J$ is $\infty$, this class of model converges, pointwise, to $G$ as $\theta$ tends to 0. Note that in this limiting case, $J\theta \to \infty$ because the support of the distribution is on the positive real line. In this model, the specification of $G$ is given through a Dirichlet process (DP) prior (Ferguson, 1973) denoted as $G \sim \text{DP}(\alpha, G_0)$, where $\alpha$ is the precision parameter, and $G_0$ is the DP centering distribution defined on $\mathbb{R}^+$. Following the convention in the DP literature, we assume that $\alpha$ is unknown and assign to it a gamma distribution as prior. $G_0$ is assigned a Weibull distribution, denoted as Weibull($\mu, \phi$), with scale and shape parameters $\phi$ and $\mu$. 

65
respectively. A DP prior implies that \( G \) is a countable mixture of point masses, which arise i.i.d from the centering distribution \( G_0 \).

Figure 4.1 shows an example of the generation of weights with \( \mu = 2, \phi = 3, \theta = 0.2, J = 20 \). In Figure 4.1a, \( \alpha = 10 \), the CDF sample path from \( \text{DP}(\alpha, G_0) \) is closer to the CDF of \( G_0 \), the Weibull distribution, so most of the weights are greater than 0 and only \( w_{17} \) and \( w_{20} \) are equal to 0. In Figure 4.1b, \( \alpha = 1 \), the CDF sample path from the DP is similar to that of a step function. Only \( w_5, w_{12} \) and \( w_{20} \) are significantly greater than 0, and most of the weights are nearly 0. Some weights, such as \( w_2 \) and \( w_{18} \), are exactly 0. Thus, in our proposed model, not all \( J \) weights will be significantly greater than 0, especially when the number of distinct point masses, which is controlled by \( \alpha \), is small. This implies that the DP prior for \( G \) works as a data-driven dimension reduction technique, reducing the number of Erlang mixture components to just those corresponding to the positive weights. This is an important feature of our model. Regarding the role of \( G_0 \), we will see in Section 4.2.2 how different settings of \( \alpha \) and \( G_0 \) impact the clustering/declustering property of the point process.

In our approach, \( J \) is not fixed, as it is treated as a random variable. Thus, our proposed Bayesian approach requires the specification of a prior distribution for \( J \). As \( J \theta \) determines the support of the inter-arrival time, there is strong correlation between \( J \) and \( \theta \). Thus, it is natural to specify a prior distribution for \( J \) and \( \theta \) jointly. Our prior is \( \pi(\theta, J) \propto \text{Gamma}(\theta; a_\theta, b_\theta) \text{Unif}(J; \left\{ \left\lceil \frac{S_1}{\theta} \right\rceil, \ldots, \left\lceil \frac{S_2}{\theta} \right\rceil \right) \) \), where \( S_1 \) and \( S_2 \) are specified, respectively, as the lower and upper bounds for the support of the inter-arrival times. Our default choice is to set \( S_1 \) as the maximum observed inter-arrival time, \( S \). The value of
tool is provided by the calculations of the moments of the inter-arrival times. The burden of fitting the model. For the prior specification for the parameters $\mu$ and $\theta$, a very large number of components in the mixture, with the consequent high computational burden of fitting the model. For the prior specification for the parameters $\mu$ and $\theta$, a useful tool is provided by the calculations of the moments of the inter-arrival times.

$$E(X) = E(E(X|G)) = E(\sum_{j=1}^{J} w_j \theta) = E(\sum_{j=1}^{J} \int 1_{(j-1)\theta,j\theta}(y)Y_dG(y)) = E(\theta \cdot E(\left\lceil \frac{Y}{\theta} \right\rceil|G))$$

$$\in [E(E(Y|G)), E(E(Y|G) + \theta)] = [E(Y|G_0), E(Y|G_0) + E(\theta)]$$

$$\in \left[ \mu \Gamma(1 + \frac{1}{\phi}), \mu \Gamma(1 + \frac{1}{\phi}) + \frac{b_y}{a_\theta} \right], \quad (4.2)$$

We can let $E(\mu \Gamma(1 + \frac{1}{\phi})) = \frac{1}{2} E(X)$ and $E(\theta) = \frac{1}{2} E(X)$ as a guideline.
The motivation for the use of a mixture of Erlang distributions is twofold: on one hand we wish to achieve flexibility and on the other we need to facilitate computations. This model can enable flexible shapes of inter-arrival time. Additionally, it can produce different clustering behaviors of point processes, which is controlled by DP prior’s specification. In section 4.2.2, we provide detailed discuss on the properties of this model. Computability is achieved thanks to the introduction of latent variables that allows for the inclusion of the normalizing constant in the likelihood, without incurring in prohibitive computational cost. Computations for the estimation of the model parameters are performed with a Markov chain Monte Carlo method. In the same spirit of Kalli et al. (2011), we introduce auxiliary variables \( \{y_1, \cdots, y_{n+1}\} \), that are i.i.d. realizations from \( G \). These serve as configuration variables that “break” the mixture and label the component assigned to each observation. Indeed, given that \( w_j = \int 1_{(j-1)\theta,j\theta]}(y_k) dG(y_k) = G(j\theta) - G((j-1)\theta) \), the distribution of inter-arrival time \( t_k - t_{k-1} \), conditional on the auxiliary variable \( y_k \), is \( \sum_{j=1}^{J} \text{Gamma}(t_k - t_{k-1}; j, 1/\theta)1_{(j-1)\theta,j\theta]}(y_k) \). In this fashion, each observed inter-arrival time is assigned to only one Erlang component. By using the auxiliary variables, the posterior sampling can then be efficiently performed. More details are provided in section 4.2.4.

### 4.2.2 Model properties

A flexible model for renewal processes needs to be able to capture different clustering behaviors. In this section, we use \( K \) function to study the property of this model. As is discussed in section 2.5, for renewal process, \( K(t) = M(t)E(X) \). \( E(X) \) is the expected inter-arrival time. For the renewal function \( M(t) \), some of its general properties are reported below.
**Result 1.** Let $\mu_1 = E(X)$, $\mu_2 = E(X^2)$ and let $c_X$ denote the coefficient of variation of $X$, defined as $c_X^2 = \frac{\text{Var}(X)}{(E(X))^2} = \frac{\mu_2}{\mu_1^2} - 1$. Then, for large enough $t$, $M(t) \approx \frac{t}{\mu_1} + \frac{1}{2}(c_X^2 - 1)$.

The former result corresponds to Theorem 1.1.9 in Tijms (1994). Direct application to the $K$ function yields the approximation $K(t) = \mu_1 M(t) \approx t + \frac{\mu_1^2}{2}(c_X^2 - 1)$, for large enough $t$. We conclude that, asymptotically, the sign of $c_X^2 - 1$ indicates if the point process has declustering or clustering patterns. When $c_X^2 = 1$, the point process is a homogeneous Poisson process.

**Result 2.** $M(t) = \sum_{n=1}^{\infty} F_n(t)$ for any $t \geq 0$, where $F_n(t)$ is the CDF of the arrival time of the $n$-th event, $T_n$. Moreover, $\sum_{n=N+1}^{\infty} F_n(t) \leq \frac{F_N(t)F(t)}{1-F(t)}$, implying that,

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

This result is found in Tijms (1994). We note that, as long as $\epsilon(t)$ is small enough, $M(t)$ can be approximated by the truncated sum $\sum_{n=1}^{N} F_n(t)$, that we denote as $M'(t)$.

**Result 3.** Consider the renewal equation, $M(t) = F(t) + \int_{0}^{t} M(t-x)dF(x)$, and assume that $f(x)$ has Laplace transform $L_X(s)$. Then, taking the Laplace transform of both sides of the renewal equation yields

$$L_M(s) = \frac{L_X(s)}{s(1 - L_X(s))}.$$  \hspace{1cm} (4.3)

This result is important as it provides a practical method to calculate the renewal function $M(t)$ using the inverse Laplace transform of $L_M(s)$. The above results are applicable to general renewal functions. Next, we discuss several results specific to our model.
Result 4. The K function derived from our model is well defined, i.e., $K(t) < \infty$.

The proof can be found in Appendix B.

Result 5. The inter-arrival time density defined by the model proposed in Equation (4.1) produces a coefficient of variation given by

$$c_X^2 = \frac{\sum_{j=1}^{J} w_j(j^2)}{(\sum_{j=1}^{J} w_j)^2}. \quad (4.4)$$

From Equation (4.4) we have that the coefficient of variation generated from our model is fully determined by the mixture weights. Lee & Lin (2010) and Tijms (1994) show that this class of Erlang mixtures is dense in the space of positive continuous distributions. So, theoretically, Equation (4.4) can take any positive value. That implies that the model can generate coefficients of variation that are below, equal to and above one. It means that, at least asymptotically, our model can generate point processes with both clustering patterns and declustering/regularity patterns. To explore the properties of our model for finite $t$, we consider some special cases that are summarized in the following results.

Result 6. The model proposed in Equation (4.1) contains families of renewal processes with declustering as well as clustering properties.

The analytic form of the K function can be explored applying Result 2. For some special cases of our model, a simple expression for $M(t)$ can be obtained, using the definition $M(t) = \sum_{n=1}^{\infty} F_n(t)$. As an example, the Erlang distribution with shape parameter equal to 2 has $M(t) = \frac{t^2}{2\theta} + \frac{1}{4} e^{-\frac{2t}{\theta}} - \frac{1}{4}$. As $\mu_1 = 2\theta$, we have that $K(t) = t + \frac{\theta}{2}(e^{-\frac{2t}{\theta}} - 1)$, which is less than $t$ for any $t > 0$. So an Erlang distribution with shape 2 always generates declustering point processes. From numerical evalution of the formula in Result 2, we
observe that, a single Erlang distributions with shape parameter \( k \) \((k > 2)\) always generate declustering processes. The opposite is true for the mixture of an exponential distribution and an Erlang distribution with shape \( k \) \((k \geq 4)\), as this generates a clustering process. From these examples, we conclude that our model has the capability of generating point processes with declustering patterns, as well as processes with clustering patterns for \( t > 0 \), not just asymptotically.

**Result 7.** The inter-arrival time density in our model is \( f(x) = \sum_{j=1}^{J} w_j x^{j-1} e^{-x/\theta^j(j-1)!} \) and its Laplace transform is \( L_X(s) = \sum_{j=1}^{J} w_j \frac{1}{\theta^j(s+1/j)!} \). Using Equation (4.3), the Laplace transform of \( M(t) \) is

\[
L_M(s) = \frac{s \sum_{j=1}^{J} w_j \frac{1}{\theta^j(s+1/j)!}}{s \sum_{j=1}^{J} w_j (1 - \frac{1}{\theta^j(s+1/j)!})}.
\] (4.5)

This is an important practical result, as for general cases, which usually contain more than three mixture components, calculating \( M(t) \) from its definition is computationally intensive. Instead, evaluating the renewal function of the general model using the inverse Laplace transform of Equation (4.5) is much more computationally efficient. The application of this approach will be illustrated in the next section.

### 4.2.3 Investigating the \( K \) function for the Erlang mixture model

In this section, we numerically evaluate the \( K \) function generated from our model by using the Laplace transform method. Given \( \theta, G_0 \) and \( \alpha \), the \( K \) function can be simulated from the prior model as below. In each round of sampling,

- A realization of \( \{w_1, w_2, \ldots, w_J\} \) is generated from the \( DP(\alpha, G_0) \) and \( \theta \).
• Evaluate $M(t)$ by inverse Laplace transform, numerically.

• Evaluate the $K$ function as $K(t) = \mu_1 M(t)$, where $\mu_1 = \sum_{j=1}^{J} w_j j \theta$.

We use this simulation method to investigate how $G_0$, $\alpha$ and $\theta$ influence the $K$ function. Recall that, for $G_0$ being a Weibull distribution, $\mu$ is the scale parameter and $\phi$ is the shape parameter. We conduct a set of simulation experiments to gauge the effect of $\mu$ and $\phi$ on the clustering patterns of the simulated point processes.

**The shape parameter of the Weibull distribution** We fix the values of $\alpha$, $\theta$ and $\mu$ and vary $\phi$. The results are summarized in Table (4.1). $\alpha$ is fixed to have the, relatively large, value 200, so that $G \approx G_0$. Figure 4.2 shows ten simulated $K$ functions from three prior models. With $\phi = 0.5$, the $K$ functions always display clustering patterns, see Figure 4.2a. With $\phi = 1$, all kinds of point processes including the homogeneous Poisson processes can be generated, see Figure 4.2b. When $\phi$ is increased to 2, all simulated $K$ functions display declustering patterns, see Figure 4.2c. Table 4.1 (1) summarizes these empirical results.

It is known that $\phi < 1$ indicates that the coefficient of variation of the Weibull distribution is greater than 1, and this corresponds to a decreasing hazard rate over time. For $\phi > 1$ the coefficient of variation is less than 1, and an increasing hazard rate over time is obtained. There is not enough evidence to state that $\phi = 1$ is a critical value for the behavior of the $K$ function. However, empirically we observe that, when $\phi < 1$, there are strong indications that the point process has a clustering behavior. Intuitively, this is due to the fact that when $\phi$ is less than 1, the Weibull distribution is over-dispersed with a long
tail. The $DP(\alpha, G_0)$ can generate an even more over-dispersed distributions. The model will likely include mixture components whose coefficient of variation is larger than 1. In summary, we can conclude that, when $G_0$ is taken as a Weibull distribution, it is possible to generate clustering, declustering and homogeneous Poisson point processes, by changing the shape parameter.

![Graphs](image)

Figure 4.2: $K$ functions based on 10 samples from $DP(\alpha, G_0)$ where $\alpha = 200$. From left to right: $G_0$=Weibull(5, 0.5), $G_0$=Weibull(5, 1) and $G_0$=Weibull(5, 2).

**The scale parameter of the Weibull distribution** The scale parameter $\mu$ is also associated with the point process pattern. When $\phi > 1$ and $\alpha$ is equal to a large value, then values of $\mu$ such that $\mu < \theta$ make $G$ concentrate mainly in $(0, \theta)$, implying that $w_1 \to 1$. The inter-arrival density is then merely an exponential distribution with parameter $1/\theta$, corresponding to a homogeneous Poisson process. If $\mu > \theta$ then $G$ is mostly in $(\theta, \infty)$, which produces declustering processes.

**Pareto distribution as $G_0$** In all the numerical calculations considered so far we have assumed that $G_0$ is Weibull. In order to explore the influence of the tails of $G_0$ on the
behavior of the $K$ function, we run a set of experiments where $G_0$ is a Pareto distribution. A Pareto distribution, denoted as Pareto($a,b$), is supported in $[a, \infty)$. As opposed to a Weibull that has exponential tails, a Pareto distribution has polynomial tails. The shape parameter (tail index) $b$ determines the coefficient of variation. Moreover, as $b$ goes to infinity, the Pareto distribution concentrates all its mass at $a$. We set $a = 2.5$ and $\theta = 2$.

When $b = 1$, $G_0$ produces a long-tail Pareto distribution and thus produces point processes with clustering patterns, see Figure 4.3a. When $b = 5$, the Pareto distribution is more concentrated and the point processes show declustering patterns, see Figure 4.3b. For $a = 0.01, \theta = 1$ and $b = 5$, all simulated point processes are homogeneous Poisson processes, see Figure 4.3c. This is generally the case for $\theta > a$ and large $b$. The parameter settings and results are summarized in Table 4.1 (2). Again, as for the Weibull case, we observe that Pareto distributions can generate all three types of point processes.

In addition to the two choices of $G_0$ summarized here, we considered Gamma, log-normal and generalized Pareto distributions. From the empirical exploration of the
properties of the models produced by such distributions, we establish the conjecture that
the clustering or declustering behavior of the point process is determined by the parameter
that controls the coefficient of variation or the monotonicity of the hazard function of \( G_0 \).
For modeling purposes, it is important to consider \( G_0 \) in a family with a wide range of
coefficients of variation. Our default choice in this work is the Weibull distribution, as it
satisfies such desiderata and has a close form CDF.

\[
\begin{array}{|c|c|c|c|c|}
\hline
G_0 & \text{Fixed parameters} & \text{Parameter to be tested} & C & D \\hline
(1) Weibull (\( \mu, \phi \)) & \theta = 2, \alpha = 200, \mu = 5 & \phi = 0.5 & \checkmark & \checkmark \\hline
 & & \phi = 1 & \checkmark & \checkmark \checkmark \\hline
 & & \phi = 2 & \checkmark & \checkmark \\hline
(2) Pareto (\( a, b \)) & \alpha = 200, \theta = 2 & a = 2.5, b = 1 & \checkmark & \checkmark \\hline
 & & a = 2.5, b = 5 & \checkmark & \checkmark \\hline
 & & a = 0.01, b = 5 & \checkmark & \checkmark \\hline
\end{array}
\]

Table 4.1: Clustering patterns generated by prior models with different centering distri-
processes.

**The value of \( \alpha \).** In addition to studying the role of \( G_0 \), the centering distributions of
the DP, we investigate the influence of \( \alpha \), the precision parameter. In the following set of
experiments, \( G_0 \) is fixed to be Weibull(5,0.5) and \( \theta = 2 \). We considered \( \alpha = 0.2, 2, 50 \). From
Figure 4.4a we observe that, when \( \alpha = 0.2 \), the \( K \) function has much larger uncertainty, as
the realizations of the DP have large variations. When \( \alpha = 2 \), the realizations of the DP have
smaller variations than for smaller \( \alpha \), and so does the \( K \) function. For \( \alpha = 50 \), \( G \) is closer
to \( G_0 \), thus all resulting point processes have clustering patterns, which is the behavior of a
renewal process with interval arrival time distribution equal to a Weibull(5,0.5). Fixing \( G_0 \)
to be Weibull(5, 2) and \( \theta = 2 \) we mostly observe the declustering patterns that correspond to the chosen \( G_0 \). Only a small \( \alpha \) provides enough variability in the realization of \( G \) to produce constant \( K \) functions. This is illustrated in Figure 4.5.

The above experiments settings and results are summarized in Table 4.2. We conclude that smaller values of \( \alpha \) allow for more flexibility in the clustering behaviors of the generated point processes. Our preferred modeling strategy, in the absence of prior information about the clustering patterns, is to choose a prior distribution for \( \alpha \) that is concentrated on small values.

![Figure 4.4: K functions based on 10 samples from a DP(\( \alpha, G_0 \)) where \( G_0 = \text{Weibull}(5, 0.5) \). From left to right: \( \alpha = 0.2, 2, 50 \) respectively.](image-url)
Figure 4.5: $K$ functions based on 10 samples from a $DP(\alpha, G_0)$ where $G_0 = \text{Weibull}(5, 2)$.

From left to right: $\alpha = 2, 10, 50$ respectively.

<table>
<thead>
<tr>
<th>$G_0$</th>
<th>Fixed parameters</th>
<th>Parameter to be tested</th>
<th>C</th>
<th>D</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3) Weibull ($\mu, \phi$)</td>
<td>$\theta = 2, \mu = 5, \phi = 0.5$</td>
<td>$\alpha = 0.2$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\alpha = 2$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\alpha = 50$</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4) Weibull ($\mu, \phi$)</td>
<td>$\theta = 2, \mu = 5, \phi = 2$</td>
<td>$\alpha = 2$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\alpha = 10$</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\alpha = 50$</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.2: Clustering patterns generated by prior models with different precision parameters. **C**: clustering processes. **D**: declustering processes. **H**: homogeneous Poisson processes.
4.2.4 Inference

To perform inference on the parameters of the proposed model we consider the hierarchical representation

\[
 f(t_k - t_{k-1} | \theta, G) = \sum_{j=1}^{J} w_j \text{Gamma}(t_k - t_{k-1} | j, 1/\theta)
\]

\[
 \theta, J \sim \text{Gamma}(\theta; a_\theta, b_\theta) \text{Unif} \left( J; \left\{ \left\lfloor \frac{S_1}{\theta} \right\rfloor, \ldots, \left\lfloor \frac{S_2}{\theta} \right\rfloor \right\} \right)
\]

\[
 w_j = \int 1_{((j-1)\theta,j\theta]}(y_k)dG(y_k), \quad y_k | G \sim G
\]

\[
 G \sim \text{DP}(\alpha, G_0), \quad G_0 = \text{Weibull}(\mu, \phi)
\]

\[
 \alpha \sim \text{Gamma}(a_\alpha, b_\alpha), \quad \mu \sim \text{IG}(a_\mu, b_\mu), \quad \phi \sim \text{Gamma}(a_\phi, b_\phi).
\]

Under our model, the likelihood function of renewal process is expressed as

\[
 l(t_1, t_2, \ldots, t_n | \theta, J, w_1, w_2, \ldots, w_J) = \prod_{k=1}^{n} f(t_k - t_{k-1}) \left\{ 1 - \int_{t_n}^{T} f(u - t_n)du \right\}
\]

\[
 = \left\{ \sum_{j=1}^{J} w_j \text{Gamma}(t_1; j, 1/\theta) \right\} \left\{ \prod_{k=2}^{n} \sum_{j=1}^{J} w_j \text{Gamma}(t_{k-1}; j, 1/\theta) \right\}
\]

\[
 \times \left\{ \sum_{j=1}^{J} w_j \left(1 - \int_{t_n}^{T} \text{Gamma}(u - t_n; j, 1/\theta)du \right) \right\}. \quad (4.6)
\]

Conditional on the auxiliary variables, the joint model can be written as

\[
 l(t_1, \ldots, t_n | y_1, \ldots, y_n, y_{n+1}, \theta) = \left\{ \sum_{j=1}^{J} \text{Gamma}(t_1; j, 1/\theta) 1_{((j-1)\theta,j\theta]}(y_1) \right\}
\]

\[
 \times \left\{ \prod_{k=2}^{n} \sum_{j=1}^{J} \text{Gamma}(t_{k-1}; j, 1/\theta) 1_{((j-1)\theta,j\theta]}(y_k) \right\}
\]

\[
 \times \left\{ \sum_{j=1}^{J} \left(1 - \int_{t_n}^{T} \text{Gamma}(u - t_n; j, 1/\theta)du \right) 1_{((j-1)\theta,j\theta]}(y_{n+1}) \right\}.
\]

In this model, the normalizing constant is analytically tractable. As is shown in the last line of likelihood, the normalizing constant has an elegant expression, which is
a mixture of complementary cumulative distribution function of Erlang distribution with common scale parameter. After introducing the auxiliary variables, in each of the three components of the above expression, the sums reduce to only one term, strongly reducing the complexity of likelihood evaluations. The normalizing constant in the last term can be handled as same as the other components in the likelihood.

A Markov chain Monte Carlo (MCMC) method to explore the posterior distribution of of $\theta, J$ and $G$ is described in Appendix C. Once we have obtained draws from such distribution, it is of interest to make inference for various functionals of those quantities. In particular we are interested in inferring the shape of the $K$ function, to learn about the clustering properties of the process. We evaluate Equation (4.5) at each sample $G_b, J_b$ and $\theta_b$, $b = 1, \ldots, B$, to obtain posterior samples of the renewal function, $M_b(t \mid G_b, J_b, \theta_b, \text{data})$. From this, we have $K_b(t \mid G_b, J_b, \theta_b, \text{data}) = M_b(t \mid G_b, J_b, \theta_b, \text{data})E_b(X \mid G_b, J_b, \theta_b, \text{data})$, where $E_b(X \mid G_b, J_b, \theta_b, \text{data}) = \sum_{j=1}^{J_b}[G_b(j\theta_b) - G_b((j - 1)\theta_b)]j\theta_b$. In addition, the posterior predictive density at a given inter-arrival time point $x_0 > 0$, is calculated as $f(x_0 \mid G_b, J_b, \theta_b) = \sum_{j=1}^{J_b}[G_b(j\theta_b) - G_b((j - 1)\theta_b)]\text{Gamma}(x_0 \mid j, \theta_b)$.

### 4.3 Data Analysis

In this section we evaluate our model by fitting it to two sets of synthetic data, corresponding, respectively, to a declustering and a clustering pattern. We then consider two datasets, one for earthquake occurrences in North Anatolia, Turkey, and the second one for coal-mining disaster data.
4.3.1 Synthetic data

The first set of synthetic data has declustering patterns. We consider two different examples, corresponding to the two inter-arrival time densities with parametric forms, \( f_1(t-H(t)) = \text{Weibull}(1.5, 2) \) and \( f_2(t-H(t)) = 0.42 \text{InvGaussian}(0.5, 1) + 0.58 \text{InvGaussian}(2, 6.5) \). The Weibull distribution for the first density has shape parameter \( \phi = 1.5 \), producing a coefficient of variation that is less than 1, and a monotonically increasing hazard function, generating a point processes with a declustering pattern. A sequence of 269 data points are generated within the time window \((0, 500)\). A snapshot of the data points is shown in Figure 4.6a.

Our second density function is a mixture of inverse Gaussian distributions. This is motivated by the fact that inverse Gaussian distribution is a popular choice for declustering point process models. We generate a sequence of 366 data points within the time window \((0, 500)\) from the mixture model. A snapshot of the simulated data is shown in Figure 4.6d.

We use the MCMC method described in the Appendix C to estimate all the parameters in the proposed mixture model. We obtain 1000 posterior samples by storing every 100 iterations, following a burn-in of 2000 iterations. The simulated chains for all parameters converge within the burn-in and achieve 20-30% acceptance rate in Metropolis-Hasting sampling steps. For the first data set, we use the priors \( \theta \sim \text{Gamma}(2, 2) \), \( \mu \sim \text{IG}(3, 2) \) and \( \phi \sim \text{Gamma}(1, 0.5) \). The 95% posterior probability interval for \( \phi \) is \((0.4825, 7.1141)\) and its posterior median of \( \phi \) is 1.920, which corresponds to a declustering process. For the second data set, we give priors \( \theta \sim \text{Gamma}(1, 1) \), \( \mu \sim \text{IG}(3, 1.5) \) and \( \phi \sim \text{Gamma}(1, 0.5) \). The 95% probability interval for \( \phi \) is \((0.6553, 1.9459)\) and the posterior median is 1.1735, which also
Figure 4.6: First and second simulated data sets. 4.6a and 4.6d: a snapshot of the point process realizations for the first and second densities respectively; 4.6b and 4.6e: true data-generating function (black line), posterior mean of the inter-arrival time density function (red line), and 95% interval bands (dash blue lines) for the first and second densities respectively. 4.6c and 4.6f: true $K$ function (black line); posterior mean of the $K$ function (red line) and the 95% interval bands (dash blue lines) for the first and second densities respectively.
implies a declustering process.

We compare the true $K$ function as well as the true density of inter-arrival times with the estimates from our model. As is shown in Figures 4.6b and 4.6e for the first and second synthetic data sets respectively, the posterior predictive density captures the sampled data nicely and the posterior interval bands obviously contain the true data-generating density. The results for the $K$ functions are also positive, for all times, for both sets of simulations. In these examples our proposed method captures the declustering properties of the point process very accurately. It is important to notice that the evaluation of the true $K$ function can be a complicated task, even in the presence of full knowledge of the parametric model for the inter-arrival times. We first calculate the renewal function $M(t)$ of a Weibull distribution by using an approximation method from Smith & Leadbetter (1963) and obtain the $K$ function by $K(t) = \mu_1 M(t)$ where $\mu_1$ is the mean of the Weibull distribution. For the mixture of inverse Gaussian distributions, we use the Laplace transform of an inverse Gaussian, which is available in closed form. The inverse transform is then obtained using numerical evaluations.

The second set of synthetic data was chosen to have a clustering pattern. These simulations were generated from two parametric distributions as follows: $f_3(t - H(t)) = \text{Pareto}(1.98, 1)$, $f_4(t - H(t)) = \text{Weibull}(0.6, 2.3)$. The hazard rate of the Pareto distribution is monotonically decreasing, so the corresponding point process has a clustering property. A sequence of 556 data points were generated within the time window (0, 500). A snapshot of the data points is shown in Figure 4.7a. The fourth density function is an exponential long-tail Weibull distribution, which is commonly used as a inter-arrival time density function.
for reliability data. A sequence of 664 data points were generated within the time window $(0, 2000)$. A snapshot of data points is shown in Figure 4.8a. We choose these two density functions due to their wide use for clustering data modeling, as well as to test our model’s ability to handle data corresponding to different tail behaviors.

Figure 4.7: Third simulated data set. 4.7a: a snapshot of the simulated point process. 4.7b, 4.7c and 4.7d: true density function (black line), posterior mean of the inter-arrival time density function (red line) and 95% interval bands (dash blue lines) for the head of the data, the tail of the data and all data respectively. 4.7e: the true $K$ function (black line), the posterior mean of the estimated $K$ function (red line) and the 95% interval bands (dash blue lines).

Samples of the joint posterior distribution were obtained in a similar fashion to
Figure 4.8: Fourth simulated data set. 4.7a: a snapshot of the simulated point process. 4.7b, 4.7c and 4.7d: true density function (black line), posterior mean of the inter-arrival time density function (red line) and 95% interval bands (dash blue lines) for the head of the data, the tail of the data and all data respectively. 4.7e: the true K function (black line), the posterior mean of the estimated K function (red line) and the 95% interval bands (dash blue lines).

the samples for the first and second simulations. We used the prior Gamma(0.25, 0.5) for $\phi$ for these two synthetic data sets. For the 3rd synthetic data set, the posterior median of $\phi$ is 0.4049 and the 95% posterior probability interval is (0.1589, 0.8278). For the 4th synthetic data set, the posterior median of $\phi$ is 0.4998 and the 95% probability interval is (0.2514, 0.8891). In both cases, $\phi$ is significantly less than 1, strongly indicating the
clustering property of the synthetic point processes. Figure 4.7d and 4.8d show the true data generating density, density of samples and posterior predictive density of inter-arrival time. In order to better visualize the data fitting, we show the model fit for the head of data and the tail of the data, separately, in Figure 4.7b, 4.7c, 4.8b and 4.8c. The posterior mean of density function fits the data, both in the head and the tail quite accurately. The 95% interval bands contain the true data-generating function fairly well. Similar results are obtained for the $K$ function. As in the previous case, calculating the true $K$ function for the parametric models presented some difficulties. For the Pareto distribution, we use the Laplace transform reported in Nadarajah & Kotz (2006) and evaluated the renewal function numerically using Equation (4.3). As the shape parameter of the Pareto in this example is 1.98 ($> 1$) the mean of the distribution $\mu_1$ is finite and known.

4.3.2 Earthquake occurrences data

In seismology, earthquake occurrences were usually assumed to be a Poisson process (Cornell, 1968; Caputo, 1974). In order to model clustering foreshock-mainshock-aftershock sequences, other stochastic models such as compound Poisson process, renewal process and Markov process are used (Anagnos & Kiremidjian, 1988). Recent works use a time-dependent renewal process to accommodate the elastic behavior of earthquake reoccurrences (Parsons, 2008).

We consider a data set consisting of 76 observed large earthquakes whose magnitudes are at least 5 in Ritter scale, in North Anatolia, Turkey in the 20th Century. A graph of the occurrences is presented in Figure 4.9a. The data are recorded by the Kandilli observatory, Turkey (http://www.koeri.boun.edu.tr) and also reported in Alvarez (2005).
For this specific application, Alvarez (2005) uses a semi-Markov renewal process, which assumes the sequence of events to be a Markov chain and the distribution depends on the severity of last event. The earthquakes in Turkey are classified into three types according to the magnitude and the inter-arrival times are assumed to be Weibull distribution. Yilmaz & Çelik (2008) examine 11 distributions including the double exponential, Weibull, Log-normal, Gamma, and Pareto to model the inter-arrival times and conclude that Weibull distribution is the most appropriate distribution.

Here we assume the earthquake occurrences in Turkey follow a renewal process. We base our results on 1000 samples after thinning the resulting chain of samples by taking every 50th iteration. We used a Gamma(2, 2) prior for $\phi$, and obtained a posterior median for $\phi$ of 0.2929 with 95% interval bands of (0.1080, 1.24). This results strongly indicate that the earthquake occurrence has a clustering pattern. The parameter $\alpha$ influences the number of mixture components, as well as the value of $\theta$. Here we use the prior $\alpha \sim \text{Gamma}(2, 4)$, obtaining a posterior that is centered around 0.6. We assume a priori that $(\theta, J) \sim \text{Gamma}(1, 0.01) \text{Unif}(\{4000/\theta, \ldots, 12000/\theta\})$. The posterior distribution for $\theta$ is centered round 300, and the posterior of $J$ has a median of 35. The posterior predictive density, posterior intensity function and posterior $K$ function are shown in Figure 4.9. The posterior predictive density closely agrees with the histogram of the data.

Once the inter-arrival time density is estimated, it is easy to obtain the corresponding hazard function as

$$
\lambda(t|H(t)) = \frac{f(t - H(t))}{1 - F(t - H(t))}, \quad t \in [H(t), \infty).
$$

This can be interpreted as the probability that the next event will occur between time
\((t, \delta t)\), for a small \(\delta\), under the condition that the event has not occurred between \(H(t)\) and \(t\). Notice that this is a type of conditional intensity for the point process. It is a function of both \(t\) and \(H(t)\), but it depends only on the time difference between the current time \(t\) and the time of last event \(H(t)\). We visualize this conditional intensity function as a one-dimensional plot by using time interval \(t - H(t)\) as the x axis.

In Figure 4.9c, the hazard function shows a decreasing trend within \((0, 1000)\), indicating a clustering pattern of earthquakes. A slightly increasing trend in the hazard function within \((2000, 4000)\) is due to the extreme values on the tail of the raw data. The posterior mean and 95\% interval bands of the \(K\) function are all positive, which reinforces the inference that earthquake occurrence in Turkey has a clustering property.

Figure 4.9: 4.9a: Large earthquakes in Turkey in the 20th century. 4.9b: the mean posterior predictive density of inter-arrival times (red line), 95\% interval bands (blue dash lines) and the histogram of the real data. 4.9c and 4.9d: the posterior mean (red line) and 95\% interval bands (blue dash lines) of hazard function (4.9c) and \(K\) function (4.9d).
4.3.3 Coal mining disaster data

The coal-mining disaster data set from Jarrett (1979) is commonly used in the literature of point processes. It records 191 explosions involving 10 or more men killed. Most papers in the literature treat the data as a realization of an inhomogeneous Poisson process because the intensity function varies over time. Here, as our model makes an assumption of time-homogeneity, we will use the first 81 disasters from 15 March 1851 to 31 Dec 1875, which basically satisfy time homogeneity, see Figure 4.10a. In this case the priors for the model parameters are: \( \phi \sim \text{Gamma}(4, 2) \), \((\theta, J) \sim \text{Gamma}(1, 0.02)\text{Unif}\{800/\theta, \ldots , 8000/\theta\}) \) and \( \mu \sim \text{IG}(6, 800), \alpha \sim \text{Gamma}(2, 4) \). The posterior inference is based on 5000 MCMC iterations after 5000 burn-in iterations. The posterior \( \phi \) has a median of 1.5203 and 95% credible interval of (0.4926, 3.9219). The posterior \( \theta \) has a median of 95.3948, \( J \) has a median of 45 and \( \mu \) is centered around 110.

In Figure 4.10b, the posterior predictive density overlaps with the coal-mine data very well and the 95% interval bands are pretty narrow. In Figure 4.10c we show the posterior distribution of the mixture model weights. The largest weight, \( w^{(1)} \), has a posterior mean of 0.86. The posterior mean of the sum of two largest weights \( w^{(1)} \) and \( w^{(2)} \) is 0.967. This suggests the existence of two main mixture components. The index of the largest component is the Erlang distribution with shape 1 with a probability of 98%. Thus, the estimated inter-arrival time density is very close to an exponential distribution, which is in agreement with inter-arrival density of the homogeneous Poisson process. The posterior mean of the hazard function in Figure 4.10d is also constant over time and the posterior mean \( K \) function nearly equals \( t \). Based on these inferences, we conclude that there is
strong evidence that the 81 disasters from 15 March 1851 to 31 Dec 1875 correspond to a homogenous Poisson process.

Figure 4.10: 4.10a: coal mining data from 1851 to 1875. 4.10b: the posterior mean predictive density (red line), 95% interval bands (blue dash lines) and the histogram of the real data. 4.10c: the distribution of the two largest weights and the corresponding component indices. 4.10d and 4.10e: the posterior mean (red line) and 95% interval bands (blue dash lines) of hazard function (4.10d) and $K$ function (4.10e) given 1000 posterior samples.

To check the ability of our model to capture a declustering pattern in real-world declustering data, we retain every alternate event from the 81 coal-mining disasters, see Figure 4.11a. This data set is obtained from thinning a Poisson process, and thus, it
should have a declustering property. The prior distributions in this case are given as: \( \phi \sim \text{Gamma}(4, 2) \), \((\theta, J) \sim \text{Gamma}(1, 0.1)\text{Unif}(\{1100/\theta, \ldots, 3300/\theta\})\) and \( \mu \sim \text{IG}(3, 200) \), \( \alpha \sim \text{Gamma}(2, 1) \). The posterior inference is based on 5000 MCMC iterations after 5000 burn-in iterations. The posterior of \( \phi \) has a median of 0.9789 and 95% credible interval of (0.5168, 1.8196). The posterior of \( \theta \) has a median of 27.4779, \( J \) has a median of 96 and \( \mu \) is centered around 206.5777.

The fitted posterior predictive density is graphed with the empirical histogram in Figure 4.11b. The first two largest components weigh together about 61.97% on average. From the histograms of the indexes of \( w^{(1)} \) and \( w^{(2)} \), the most dominating component is not an exponential distribution. Instead, the weights are spread over several Erlang components. The posterior mean and 95% interval bands of the hazard function, see Figure 4.11d, continuously increase up to 350 and slightly decrease again. This reveals that the hazard function for the thinned Poisson process is time-varying and inhomogeneous. Also, the posterior mean of the \( K \) function is constantly below \( t \) as expected, see Figure 4.11e. However, the posterior 95% upper interval is above \( t \), which indicates that the thinned Poisson process is a weakly declustering point process.

### 4.3.4 Model checking

The modeling approach is based on the assumption of a renewal process. According to the Time-Rescaling theorem for point processes (Daley & Vere-Jones, 2003), the set of cumulative intensities between successive (ordered) observations, \( \{\Lambda(t_{k-1}, t_k), k = 1, \ldots, n\} \), are independent exponential random variables with mean one, where \( \Lambda(t_{k-1}, t_k) = -\ln(1 - \int_{t_{k-1}}^{t_k} f(u - t_{k-1}|G, \theta, J)du) \). Thus, \( \{1 - \exp(-\Lambda(t_{k-1}, t_k))\), \( k = 1, \ldots, n\} \) are independent
Figure 4.11: 4.11a: thinned coal mining data. 4.11b: the posterior mean predictive density (red line), 95% interval bands (blue dash lines) and the histogram of the real data. 4.11c: the distribution of the two largest weights and the corresponding component indices. 4.11d and 4.11e: the posterior mean (red line) and 95% interval bands (blue dash lines) of the hazard function (4.11d) and the $K$ function (4.11e) given 1000 posterior samples.

uniform random variables on $(0, 1)$. Hence, the renewal process assumption can be checked by using the MCMC output to obtain posterior samples for the set of random variables above. Figure 4.12 shows the Q-Q plots of estimated quantiles for time versus the theoretical uniform distribution for earthquake, coal-mining data and thinned coal-mining data.
Figure 4.12: Posterior Q-Q plots (mean and 95% interval) of estimated quantiles against the theoretical uniform(0, 1) for time in earthquake data (4.12a), coal-mining data (4.12b) and thinned coal-mining data (4.12c).

4.4 Conclusions and Further Work

In this Chapter, we propose a Bayesian nonparametric approach to modeling and inference for renewal processes. The inter-arrival time density of a renewal process is modeled as a mixture of Erlang distributions. The Erlang mixture components share a common scale parameter, and have shape parameters that correspond to fixed integers. The mixture weights are defined through a flexible distribution function modeled with a DP prior. A relevant feature of this model is that the DP prior for the CDF can generate flexible weights for the mixture components, allowing for the inter-arrival time density to have nonstandard shapes. In addition, both clustering and declustering patterns of the renewal processes can be captured by the model. These properties have carefully been studied through a number of simulated, as well as two real data examples. As is stated in the introduction, we con-
sider renewal point processes as a way to relax the Poisson process assumption, in order to handle more realistically data which may show clustering pattern (e.g. earthquakes) or declustering behavior (e.g. neural spike data). However, the renewal process assumes that all inter-arrival times follow the same distribution. Such time-homogeneity property might apply to, for example, the earthquake data, in the long term. However, in practice, it is more reasonable to believe that the rate of earthquakes occurrences vary over time. In applications of stochastic point process in reliability analysis, the state after the repair “as good as before”, implied by ordinary renewal processes, is not realistic. Thus, as future work, we plan to generalize or modulate the renewal process order to better accommodate the time-varying properties of real data.

In a generalized/modulated renewal process, the intensity function depends on two arguments: time \( t \) and the last observation up to time \( t \), denoted as \( H(t) \). The intensity function can be written as \( \lambda(t, t - H(t)) \). This approach has been considered in several papers in the neural science literature, e.g., Kass & Ventura (2001), Barbieri et al. (2001), Cunningham et al. (2008) and Koyama & Kass (2008), as well as in the literature on systems reliability, e.g., Brown & Proschan (1983) and Lawless & Thiagarajah (1996). In the present work, we have proposed a model that assumes that the process intensity is \( \lambda(t - H(t)) \), in correspondence to a homogeneous renewal process. In next chapter, we will discuss several possible extensions of the Erlang mixture model by introducing time-varying scale parameter or weights. This will make the inter-arrival time density depend on the clock time and allow for different shapes at different time points.
Chapter 5

Bayesian Nonparametric Modeling

for Markov Renewal Processes

5.1 Motivation

A renewal point process is one class of non-Poisson processes, and is commonly used to model the elastic rebound behavior of earthquakes and reliability behavior in the repairable system, etc. It assumes that the inter-arrival times are independent and identically distributed. The time homogeneity property of the distribution of successive intervals might apply to the earthquake data in a long term. However, in practice, it is more reasonable to believe that the distributions of the intervals between earthquakes occurrences vary over time. In general, for the applications of stochastic point process in reliability analysis, “as good as before after the repair”, implied by ordinary renewal processes, is not realistic. In this chapter, we extends the work in last chapter to tackle inhomogeneous
renewal processes, whose interval arrival density functions vary over time.

For a general point process, the intensity function is obtained by conditioning on the history of all past events. In Daley & Vere-Jones (2003), the conditional intensity is defined as,

\[ \lambda(t \mid H(t)) = \lim_{\Delta t \to 0} \frac{\Pr(N(t + \Delta t) - N(t) = 1 \mid H(t))}{\Delta t} . \]

where \( H(t) \) represents all historic events/spikes up to time \( t \) (including \( t \)), i.e., \( H(t) = (t_1, t_2, \ldots, t_{N(t)}) \) and \( N(t) \) is the number of events/spikes up to time \( t \) (including \( t \)). In this work, we only consider the first order conditional intensity function, so, in what follows, \( H(t) = t_{N(t)} \).

For general renewal processes, inter-arrival times are non-stationary, a property that can be obtained by assuming that the process has a non-homogeneous Markov structure. In fact, in such case, the inter-arrival time density depends on both the current time \( t \) and the last event \( H(t) \). Given the last observed event \( t_i \), let \( T_i+1 \) be the random variable representing the next event since \( t_i \). Its conditional density is

\[ f_{T_i+1}(t \mid t_i) \quad \text{or} \quad f_{i+1}(t \mid t_i), \quad \text{for } t > t_i. \]

Thus, for \( t \in \mathcal{R}^+ \), we can define a family of densities \( \{f_{i+1}(t \mid t_i), i = 0, 1, 2, \ldots\} \) to represent the time-varying inter-arrival distributions.

The conditional intensity function and inter-arrival time density have an one-to-
one relationship. \( \lambda(t \mid H(t)) \) can be derived from \( f_{i+1}(t \mid T_i = t_i) \) using its definition,

\[
\lambda(t \mid H(t)) = \lim_{\Delta t \to 0} \frac{\Pr(N(t + \Delta t) - N(t) = 1 \mid H(t))}{\Delta t} = \lim_{\Delta t \to 0} \frac{\Pr(N(t + \Delta t) - N(t) = 1 \mid T_1 = t_1, T_2 = t_2, \ldots, T_i = t_i, \text{no spike in } (t_i, t])}{\Delta t}.
\]

Then plug Equation (5.2) into (5.1), we have

\[
\lambda(t \mid H(t)) = \frac{1 - \int_{H(t)}^{t} f_{N(t) + 1}(u \mid H(t))du}{1 - \int_{H(t)}^{t} f_{N(t) + 1}(u \mid H(t))du} = \lambda(t \mid H(t)) \exp \left\{ -\int_{H(t)}^{t} \lambda(u \mid H(t))du \right\}, \quad t > H(t). \tag{5.3}
\]

Additionally, the inter-arrival time density \( f_{N(t) + 1}(t \mid H(t)) \) can be derived from \( \lambda(t \mid H(t)) \) by doing the integral of Equation (5.1) for both sides,

\[
\int_{H(t)}^{t} \lambda(u \mid H(t))du = -\ln \left( 1 - \int_{H(t)}^{t} f_{N(t) + 1}(u \mid H(t))du \right)
\]

\[
\Leftrightarrow 1 - \int_{H(t)}^{t} f_{N(t) + 1}(u \mid H(t))du = \exp \left\{ -\int_{H(t)}^{t} \lambda(u \mid H(t))du \right\} \tag{5.2}
\]

Then plug Equation (5.2) into (5.1), we have

\[
f_{N(t) + 1}(t \mid H(t)) = \lambda(t \mid H(t)) \cdot \left( 1 - \int_{H(t)}^{t} f_{N(t) + 1}(u \mid H(t))du \right)
\]

\[
= \lambda(t \mid H(t)) \exp \left\{ -\int_{H(t)}^{t} \lambda(u \mid H(t))du \right\}, \quad t > H(t). \tag{5.3}
\]

Due to the one-to-one relationship (5.1) and (5.3), the likelihood can be expressed by using either conditional intensity function or inter-arrival time densities. Say \((t_1, t_2, \ldots, t_n)\) are \(n\) observations in the time window \((0, T]\. In the point process literature, the likelihood is commonly defined in terms of conditional function as follows,

\[
p(t_1, t_2, \ldots, t_n) = \prod_{i=1}^{n} \lambda(t_i \mid t_{i-1}) \exp \left\{ -\int_{0}^{T} \lambda(u \mid H(u))du \right\} \tag{5.4}
\]
By plugging the equations (5.3) and (5.2) into (5.4), the likelihood can be easily expressed also in terms of inter-arrival time densities.

\[
p(t_1, t_2, \ldots, t_n) = \prod_{i=1}^{n} f_{T_i}(t_i | t_{i-1}) \exp \left\{ - \int_{t_i}^{T} \lambda(u | H(u)) du \right\}
\]

\[
= \prod_{i=1}^{n} f_{T_i}(t_i | t_{i-1}) \left( 1 - \int_{t_i}^{T} f_{T_{n+1}}(u | t_n) du \right)
\]

(5.5)

Alternatively, we can derive the likelihood directly by

\[
p(t_1, t_2, \ldots, t_n) = p(T_1 = t_1, T_2 = t_2, \ldots, T_n = t_n \text{ and } T_{n+1} > T)
\]

\[
= \int_{T}^{\infty} p(T_1 = t_1, T_2 = t_2, \ldots, T_n = t_n, T_{n+1} = u) du
\]

\[
= \prod_{i=1}^{n} f_{T_i}(t_i | t_{i-1}) \int_{T}^{\infty} f_{T_{n+1}}(u | t_n) du
\]

\[
= \prod_{i=1}^{n} f_{T_i}(t_i | t_{i-1}) \left( 1 - \int_{t_i}^{T} f_{T_{n+1}}(u | t_n) du \right)
\]

(5.6)

To sum up, the conditional intensity function \( \lambda(t | H(t)) \) and a set of \( \{ f_{i+1}(t | t_i) \} \) have a one-to-one correspondence. The two likelihood forms (5.4) and (5.5) are equivalent.

Specifying a conditional intensity function fully specifies the inter-arrival time density, and vice versa. Thus, if the likelihood in terms of conditional intensity function is complicated, we can switch to an expression in terms of inter-arrival time (and vice versa).

### 5.2 Related Work

There are a lot of related works to generalize or modulate homogeneous renewal processes to accommodate real data behavior. In the field of reliability analysis, a probabilistic model is commonly used to estimate the expected failure times by considering different states following a repair. The state “as good as new” or “perfect repair” is usu-
ally implied by an ordinary renewal process due to its homogeneity and Markov property and the state “as bad as old” or “minimal repair” is implied by an inhomogeneous Poisson process, due to its memoryless property (Yanez et al., 2002; Lindqvist, 2010). To tackle realistic phenomena such as states “better than old but worse than new”, “better than new” and “worse than old”, a probabilistic model for a more general non-Poisson process is needed. Another stream of related works is point process models applied to neural spike train analysis. It is generally believed that the inhomogeneous (time-varying) Poisson process is sufficient to model the multi-trial pooled data, but not applicable for the within-trial neural spike data (Kass et al., 2005). A Poisson process has the property of inhomogeneity, but also inherits the property of independent increments. However, for single-trial spike data, some neurons exhibit inability to fire another spike in a relative refractory period and some neurons display bursting or clustering behavior. These history dependent behaviors suggest a non-Poisson process model.

We classify these works into two categories. One category of works models the conditional intensity function directly and uses the likelihood form (5.4). Lawless & Thiragarajah (1996) proposed to model the log of conditional intensity function as a linear model by incorporating several functions of current time $t$ and history time points $H(t)$ as explanatory variables. Kass & Ventura (2001) proposed a general model, named inhomogeneous Markov interval (IMI) model. The main idea is to model the conditional intensity of point process conditional on the clock time and its lapse time since the previous spike time, $\lambda(t|H(t)) = \lambda(t, t - H(t))$. They approximate the continuous time as a binary time series and fit the binary sequence by using a cubic spline-based regression model incorpo-
rating functions of $t$ and $(t - H(t))$ as explanatory variables. Rao & Teh (2011) proposed a multiplicative IMI model, $\lambda(t, t - H(t)) = \lambda(t) h(t - H(t))$. It starts with a Gamma-interval renewal process whose hazard function is $h(t - H(t))$ and modulate it with an inhomogeneous Poisson process $\lambda(t)$. They adopted similar technique in Adams et al. (2009) to make the likelihood tractable via uniformization. The restriction is that uniformization requires the hazard function to be bounded, which means the model is only applicable when the renewal interval distribution is Gamma distribution with shape parameter greater than 1. So the generalization of this model to handle clustering point process would be a problem.

The other category of works models the inter-arrival time density directly and uses the likelihood form (5.5). Yanez et al. (2002) proposed a so-called generalized renewal process by introducing a repair effectiveness parameter to stretch or shrink the inter-arrival time to accommodate states such as “worse than old” and “better than new” and etc. Barbieri et al. (2001) developed a framework of modeling conditional intensity function $\lambda(t | H(t))$. They firstly give $\lambda(t)$ a parametric form and model the inter-arrival time density by letting rescaled inter-spike arrival time $\int_{H(t)}^{t} \lambda(u)du$ follow a Gamma or Inverse-Gaussian interval process. Then the conditional intensity function is derived from the one-to-one relationship (5.1). Following this idea, Cunningham et al. (2008) extended the inhomogeneous Gamma-interval process by giving a flexible Gaussian process model to allow different possible functions for $\lambda(t)$.

Again, the inference of the continuous model is complicated. In most of those related works, they either approximate the continuous time by discretization and deal with a discrete time model, or evaluate the normalizing constant in the likelihood of the con-
continuous time point process approximately. In Kass & Ventura (2001), a continuous time is approximated by discretization and then converted as a binary time series. The estimation of all parameters uses maximum likelihood estimation method based on the approximated likelihood. In Barbieri et al. (2001), as long as the rescaled inter-spike arrival time is not exponentially distributed, there is no closed form of the point process likelihood from this model and the inference is performed without considering the normalizing constant in the likelihood. Similarly, Cunningham et al. (2008) simplifies the normalizing constant terms as intervals of an inhomogeneous Poisson process to preserve tractability and also discretize the continuous intensity function for computational purpose. Overall, only in Rao & Teh (2011) the exact likelihood is used, other works always approximate the likelihood function.

In our work, we model a time-varying density function depending on $t - H(t)$ and $H(t)$. We then show that this allows the normalizing constant to be tractable in the next section. Also, our model is extended from the homogeneous renewal process model, so it maintains the flexibility to model inter-arrival time densities of various shapes and the capability of handling both clustering and declustering point patterns.

5.3 Model Formulation

For homogeneous renewal processes, we use mixture of Erlang with common scale parameter to model the inter-arrival time density. Here, we extend it to model inhomogeneous inter-arrival time densities by letting the scale parameter depend on the previous time point. Given the last event time $t_{k-1}$, the density function of the next event (the $k^{th}$ event) is a mixture of Erlang distribution with the common scale parameter depending
on $t_{k-1}$, denoted by $\theta(t_{k-1})$. The number of mixture components is also time dependent, denoted by $J_{k-1}$. The mixing weights are generated by discretizing the CDF $G$ by $\theta(t_{k-1})$, so the weights also depend on $t_{k-1}$ and we denote them as $w_j(t_{k-1})$. The time dependent inter-arrival time density is expressed as,

$$f_{T_k}(t \mid t_{k-1}) = \sum_{j=1}^{J_{k-1}} w_j(t_{k-1}) \Gamma(t - t_{k-1} \mid j, 1/\theta(t_{k-1})), \quad t > t_{k-1}$$

(5.7)

where $w_j(t_{k-1}) = G(j\theta_{t_{k-1}}) - G((j - 1)\theta(t_{k-1}))$ and $w_{J_{k-1}} = 1 - \sum_{j=1}^{J_{k-1}} w_j$. So, given the previous event time point, the probability of the time of the next event corresponds to a mixture of Erlang, with a different scale parameter and a different number of mixture components.

To simplify the model, we unify the number of mixture components $\{J_0, J_1, \ldots, J_n\}$ to a single $J$. Then, we need to guarantee that $J$ is big enough for all inter-arrival time densities. We give a unified prior for $J$, which is $\text{Unif}\left\{\frac{S_1}{\rho^2}, \ldots, \frac{\rho S_1}{\rho^2}\right\}$, where $\rho > 1$, $S_1$ is assigned to be the observed maximum time interval, and $\theta^*$ is assigned to be approximately two standard deviations below the mean of the stochastic process $\theta(t)$. As is discussed in section 4.2, $\rho$ can be tuned to guarantee that the weights sum up to nearly 1, by checking the prior simulation and posterior inference.

We use a Gaussian process (GP) to model the logarithm of $\theta(t)$. The mean of GP is $\mu$ with some constant variance. The GP serves as a latent process for the logarithmic scale parameters of the Erlang distribution. However, for point process applications, there is one realization of the point processes, and there is even only one observation $t_k$ for each inter-arrival time density $f_k(t \mid t_{k-1})$. So, the estimation of the GP parameters is challenging. To solve this issue, we approximate the Gaussian process by using a convolution
of a white noise latent process $x(t)$ and a smoothing kernel $k(t)$. The kernel function is chosen as $k(\gamma) \propto \exp\{-\frac{1}{2\sigma^2}||\gamma||^2\}$. By convolving a continuous white noise process $x(t)$, the correlation function of the resulting Gaussian process is the exponential correlation $\exp\{-\frac{1}{2\sigma^2}||\gamma||^2\}$. Here, we restrict $x(t)$ to a few time points $v_1, \ldots, v_m$ over the time window $(0,T)$. Define $x = \{x_1, \ldots, x_m\}$ where $x_i = x(v_i), i = 1, \ldots, m$ and let $x \sim N(0,\sigma^2 I_m)$. The approximated Gaussian process for $\log \theta_i$ is

$$
\log (\theta(t)) = \mu + \sum_{i=1}^{m} x_i k(t - v_i)
$$

In this case, only $m$ random variables $x(v_1), \ldots, x(v_m)$ control the whole process $\theta(t)$ even though $\theta(t)$ is continuous. Here, $m$ will be fixed to be a much smaller value compared with the number of observations. This dimension reduction idea is critically important for this model. It will facilitate the posterior estimation of parameters and make the computation more efficient.

The biggest distinction between model (5.7) and the model in the last Chapter is the replacement of the single scale parameter $\theta$ with a stochastic process $\theta(t)$. Also, the scale parameter determines the segmentation of $G$ when generating mixture weights for the Erlang components. So when the scale parameter is small, there tend to be more Erlang mixture components, and vice versa.

One nice feature of this model is that it includes inhomogeneous Poisson processes and homogeneous renewal processes as its special cases. When $\theta(t)$ is very smooth and is close to the mean of the GP at any time point, this model reduces to homogeneous renewal process model. Moreover, if every observation corresponds to the first component of mixture of Erlang, that means the inter-arrival time follows an exponential distribution.
with time-varying rate parameter. According to the memoryless property of exponential
distribution, the independence between successive events breaks. In such case, this model
actually generates inhomogeneous Poisson processes.

5.3.1 Prior specification

To complete this model, we need to assign hyper priors for all model parameters. In
the DP prior, as discussed in section 4.2.3, we still use Weibull distribution as the centering
distribution with shape parameter $\phi$ and scale parameter $\lambda$. For the precision parameter
$\alpha$, a gamma distribution $\text{Gamma}(a_{\alpha}, b_{\alpha})$, which results in a gamma full conditional. Both
$\phi$ and $\lambda$ are positive random variables, so $\phi$ is given a gamma distribution with shape
parameter $a_{\phi}$ and rate parameter $b_{\phi}$ and $\lambda$ is given a inverse-gamma distribution with
shape parameter $a_{\lambda}$ and scale parameter $b_{\lambda}$. For the parameters in the process convolution,
we give a dispersed normal distribution for the mean parameter $\mu$, $\mu \sim \text{Norm}(0, \sigma^2_{\mu})$. The
scale parameter $\sigma^2_{\mu}$ is given a conjugate inverse gamma distribution. Overall, the whole
hierarchical model is

$$f(t_k | t_{k-1}) = \sum_{j=1}^{J} w_j(t_{k-1}) \text{Gamma}(t_k - t_{k-1} | j, 1/\theta(t_{k-1})) \quad i = 1, \ldots, n$$

$$w_j(t_{k-1}) = G(j\theta(t_{k-1})) - G((j-1)\theta(t_{k-1}))$$

$$G \sim DP(\alpha, G_0), \quad G_0 = \text{Weibull}(\phi, \lambda)$$

$$\log(\theta(t_k)) = \mu + \sum_{i=1}^{m} x_i k(t_k - v_i) \quad (5.8)$$

$$x_i \sim \text{Norm}(0, \sigma^2_x), \quad i = 1, \ldots, m.$$  

$$\phi \sim \text{Gamma}(a_\phi, b_\phi), \quad \lambda \sim \text{IG}(a_\lambda, b_\lambda), \quad \alpha \sim \text{Gamma}(a_\alpha, b_\alpha).$$

$$\mu \sim \text{Norm}(0, \sigma^2_\mu), \quad \sigma^2_x \sim \text{IG}(a_x, b_x), \quad J \sim \text{Unif}\{S_1, \ldots, \rho S_1\}$$

The range parameter $\sigma^2_\phi$ and number of knots $m$ are fixed. The positions of knots $\{v_i\}$ are equally spaced over $(0, T]$. For prior specification of other hyper parameters, we either elicit values from expert information or empirically from moments of inter-arrival times. By assuming that the scale parameters are invariant over time, so that we could reduce our model to the homogeneous renewal process. Then the priors for $\mu$ and $\lambda$ can be specified through the expectation of inter-arrival times in Equation (4.2) from data as is discussed in section 4.2. The prior of $\phi$ can be specified from the expert suggestion. For clustering data set, $\phi$ should have a prior with mean centering less than 1, while for declustering data set, $\phi$ can have a prior with mean centering with a value greater than 1. Without any prior information, we suggest to assign a prior with mean around 1 and with a large variance. Because $\alpha$ plays the role of controlling the number of distinct mixture component in DP prior, the expected prior number of components in Antoniak (1974) can be used as a guide to specify $\alpha$’s prior. Intuitively, clustering data set is more likely to require a
multimodal DP so that $\alpha$ is suggested to be assigned with a prior of small mean. The prior specification for $\sigma^2_x$ is difficult. One conservative way is to assume that all observations are equal to their corresponding scale parameters, so that $\sigma^2_x$’s prior can be specified roughly from the variance of logarithmic inter-arrival times. All the hyper parameters and the fixed values of $\sigma^2_\theta$ and number of knots $m$ should be given a careful sensitivity analysis over a range of proper values and check whether the posterior inferences are stable.

5.3.2 Simulation from prior model

In this part, we explore the potential flexibility of this model by simulating a modulated renewal process from the prior model. Given the last event, the distribution of the next event time point is determined. To draw a point process containing $n$ samples, firstly, we need to set $\mu$ and $\sigma^2_x$. Assign $m$ knots, $v_1, \ldots, v_m$ equally spaced from $(0, T]$. Get $n$ points realization of $G$ from the DP prior $DP(\alpha, G_0)$, denoted as $\{y_1, \ldots, y_n\}$. Here, we assume $t_0 = 0$. For $k = 1$, get $\theta(t_{k-1})$ from $e^{\mu + \sum_{i=1}^{m} x_i k (t_{k-1} - v_i)}$. Sample $t_k - t_{k-1}$ from $\text{Gamma} \left( \lceil \frac{y_k}{\theta(t_{k-1})} \rceil, \theta(t_{k-1}) \right)$ and obtain the $k$th time point $t_k$. Given the $k$th time point, perform the last step recursively to get $n$ points until $(n + 1)$th point exceeds the bound $T$.

Figure 5.1 shows three realizations of point processes simulated from our model conditional on a realization of $G$. We set $\mu = 2, \sigma^2_x = 1, T = 1000, m = 15, \alpha = 5, \phi = 2, \lambda = 20$ and $\sigma^2_\theta = \left( \frac{2T}{m} \right)^2$. The $\{\theta_t\}$ over $(0, T]$ and the given realization of $G$ are shown. The surface plot shows that the inter-arrival time densities have different shapes over time. The density functions can have unimodal, decreasing and event multimodal shapes.
Figure 5.1: Simulation example. Three realizations of modulated renewal point processes from the model, with the simulated latent $\theta(t)$, CDF $G$ and the inter-arrival time densities over $H(t)$ and $t - H(t)$. 
5.4 Posterior Inference

The hierarchical model is expressed in the form of (5.8). To facilitate the posterior inference, as the same manner as stated in the last Chapter, \( \{y_k\} \) is introduced as auxiliary variables and are i.i.d realizations from CDF \( G \). The mixing weights can be expressed as,

\[ w_j = \int 1_{(j-1)\theta(t_{k-1}), j\theta(t_{k-1})]}(y_k) dG(y_k), \quad y_k \sim G \]

After incorporating the latent variables, the likelihood can be written as

\[
p(t_1, \ldots, t_n \mid y_1, \ldots, y_n, y_{n+1}, \theta(t_0), \theta(t_1), \ldots, \theta(t_n)) \\
\propto \left\{ \sum_{j=1}^{J} \Gamma(t_1; j, \theta(t_0)) 1_{(j-1)\theta(t_0), j\theta(t_0)}(y_1) \right\} \\
\times \left\{ \prod_{k=2}^{n} \sum_{j=1}^{J} \Gamma(t_k - t_{k-1}; j, \theta(t_{k-1})) 1_{(j-1)\theta(t_{k-1}), j\theta(t_{k-1})}(y_k) \right\} \\
\times \left\{ \sum_{j=1}^{J} \left( 1 - \int_{t_n}^{T} \Gamma(u - t_n; j, \theta(t_n)) du \right) 1_{(j-1)\theta(t_{n}), j\theta(t_{n})}(y_{n+1}) \right\} \\
\propto \text{Gamma}(t_1; \left[ \frac{y_1}{\theta(t_0)} \right], \theta(t_0)) \left\{ \prod_{k=2}^{n} \text{Gamma}(t_k - t_{k-1}; \left[ \frac{y_k}{\theta(t_{k-1})} \right], \theta(t_{k-1})) \right\} \\
\times \left( 1 - \int_{t_n}^{T} \Gamma(u - t_n; \left[ \frac{y_{n+1}}{\theta(t_n)} \right], \theta(t_n)) du \right),
\]

where \( \theta = \{\theta(t_0), \theta(t_1), \ldots, \theta(t_n)\} \) can be calculated from the regression equation by

\[
\log(\theta(t_{k-1})) = \mu + K_{k}x, \quad \text{where} \quad K_{k} = \{K_{k1}, \ldots, K_{km}\}, \quad K_{kj} = \exp\{-\frac{1}{2\sigma^2}||v_j - t_{k-1}||^2\}\]

and \( x = \{x_1, \ldots, x_m\} \). Therefore, the data likelihood function is actually conditional on \( x = \{x_1, \ldots, x_m\} \). Thus, after combining with the priors, the joint posterior distribution is
finally expressed as

\[ p(y_1, \ldots, y_n, y_{n+1}, x_1, \ldots, x_m, J, \phi, \lambda, \alpha, \mu, \sigma_y^2 \mid t_1, \ldots, t_n) \propto \]

\[ \Gamma(t_1; \left\lceil \frac{y_1}{e^{\mu + K_1 x}} \right\rceil, e^{\mu + K_1 x}) \left\{ \prod_{k=2}^{n} \Gamma(t_k - t_{k-1}; \left\lceil \frac{y_k}{e^{\mu + K_k x}} \right\rceil, e^{\mu + K_k x}) \right\} \cdot \]

\[ \left( 1 - \int_{t_n}^{T} \Gamma(u - t_n; \left\lceil \frac{y_{n+1}}{e^{\mu + K_{n+1} x}} \right\rceil, e^{\mu + K_{n+1} x}) du \right) \cdot \text{Unif}(J \mid \{ \frac{S_1}{\theta}, \ldots, \frac{p S_1}{\theta^*} \}) \cdot \gamma \cdot \text{Gamma}(\phi \mid a_\phi, b_\phi) \cdot \text{IG}(\lambda \mid a_\lambda, b_\lambda) \cdot \gamma \cdot \text{Norm}(\mu \mid 0, \sigma_y^2) \cdot \text{IG}(\sigma_y^2 \mid a_x, b_x) \]

The DP prior for \( y_1, \ldots, y_{n+1} \) can be developed by the Polya urn characterization of the DP or stick-breaking construction by using truncated version of \( G \). In this work, we use the latter. Let \( G_N \) be a finite truncation approximation to the DP mixture \( G \), where \( N \) is the truncation level. Then \( y_k \sim i.i.d. \) \( G_N, k = 1, \ldots, n+1 \). Using stick-breaking construction of DP prior, \( G_N = \sum_{l=1}^{N} \gamma_l \). The atoms \( \gamma_l, l = 1, \ldots, N \) are i.i.d. as \( G_0 \). The weights arise from stick-breaking mechanism.

We adopt MCMC algorithm to simulate posterior samples iteratively from their full conditional. The parameters involved in the DP construction are updated by using block Gibbs sampler. The MCMC algorithm is detailed in Appendix D.

### 5.4.1 Other inference

For modulated renewal processes, one of the inference of interest is to show that the inter-arrival time density is inhomogeneous over time. In other words, it is useful to shown that \( f(t \mid H(t)) \) might have different distributions given that the \( H(t) \) is different. To obtain the \( f(t \mid H(t)) \), we use the posterior samples of \( x \) and \( \mu \) to get the corresponding \( \theta(H(t)) \).

\[ \theta(H(t)) = \exp(\mu + K'x), \text{ where } K' = (K'_1, \ldots, K'_m) \text{ and } K'_j = \exp\{-\frac{1}{2\sigma^2}||v_j - H(t)||^2\}. \]
Then from the posterior samples of mixture weights and atoms \( \{ p_l, \gamma_l, l = 1, \ldots, N \} \), we can obtain \( G \). The \( f(t \mid H(t)) \) can be derived by

\[
f(t \mid H(t)) = \sum_{l=1}^{N} p_l \text{Gamma}\left(t - H(t) \mid \frac{\gamma_l}{\theta(H(t))}, \theta(H(t))\right).
\]

The intensity function is always a typical inference of interest for point process. In this work, we model a conditional intensity \( \lambda(t \mid H(t)) \), which is a function of two variables \( H(t) \) and \( t \). Note that from this model, what we obtain is a conditional intensity function \( \lambda(t \mid H(t)) \), unfortunately the marginal intensity \( \lambda(t) \) cannot be derived analytically. One way of presenting the conditional intensity function is to show \( \lambda^*(t) \), which is defined piece-wise as follows,

\[
\lambda^*(t) = \begin{cases} 
\frac{f_1(t)}{1 - \int_0^t f_1(u)du} & 0 < t \leq t_1 \\
\frac{f_i(t|t_{i-1})}{1 - \int_{t_{i-1}}^{t} f_i(u|t_{i-1})du} & t_{i-1} < t \leq t_i, \quad i = 2, 3, \ldots, n. \\
\frac{f_{n+1}(t|t_n)}{1 - \int_{t_n}^{T} f_{n+1}(u|t_n)du} & t_n < t \leq T, \quad n = N(T). 
\end{cases}
\]

(5.10)

We can thus obtain a one dimensional plot by fixing \( H(t) \) based on the data \( (t_1, \ldots, t_n) \) over \( (0, T] \).

Moreover, the goodness of fit of the model to the data is evaluated by using the method discussed in section 4.3.4.

### 5.5 Data Analysis

#### 5.5.1 Earthquake occurrences data

We run our model for the Turkey Earthquake data which is used in Chapter 4. We give a quite dispersed prior for \( \mu, \mu \sim \text{Norm}(0, 10^2) \). \( \phi \) is assign with Gamma(1, 1).
Let the prior mean of $\lambda$ be around the mean of observed inter-arrival times, so $\lambda \sim \text{InvGamma}(3, E(t - H(t)))$. $\sigma_x^2$ is given InvGamma(3, 20). In the process convolution, the number of knots $m$ is fixed to be 6 and the range parameter is $(2T/m)^2$. Under this prior specification, the simulated sum of weights $\sum_{j=1}^{J} w(t_k)$ from the prior model has a 95% uncertainty interval of $(1, 1)$ for all $k = 1, \ldots, n + 1$. The MCMC algorithm discussed in Appendix D is used to fit the model. After 10000 burning-ins and thinning from 50000 iterations of samples, 1000 posterior samples are finally obtained. The posterior inference is quite robust over a proper range of the priors of $\mu, \phi, \lambda$ and $\sigma_x^2$. The sensitivity analysis is discussed in section 5.5.2. From the posterior samples, the sum of weights $\sum_{j=1}^{J} w(t_k)$ for $k = 1, \ldots, n + 1$ are all close to 1. The worst case is, when $k = 65$, $\sum_{j=1}^{J} w(t_{65})$ has a 95% uncertainty interval of $(0.9271, 1)$. Most of $\sum_{j=1}^{J} w(t_k)$ have 95% uncertainty intervals of $(0.99, 1)$. Figure 5.2 shows the posterior estimated scale parameters $\{\theta_k\}, k = 1, \ldots, n + 1$. The x-axis value of dots corresponds to the time of each observed earthquake. We observe that scale parameters are large during 1900-1920, and then decreases significantly and reaches the lowest values around 1940s, and is leveled up after around the year 1970. Compared with the point process of earthquake data, our inference is in agreement with the fact that there are less earthquakes recorded during 1900-1920, and the earthquakes were most clustered from 1940s to 1960s but more scattered after the year 1970. Thus, it shows that a large-valued scale parameter corresponds to a relatively large inter-arrival, while a small-valued scale parameter corresponds to a relative small inter-arrival.

This fact is also reflected in our inference of conditional inter-arrival time densities. In Figure 5.3, we show a few inter-arrival time densities $f(t \mid H(t))$ when $H(t)$ is 5000th,
10000th, 15000th, 20000th, 25000th and 30000th day respectively. The mean and dispersion of the densities are changing over time although they all have decreasing shapes. When $H(t) = 25000$, the distribution is more right skewed, which means the it is very likely that the next event will occur immediately; while when $H(t) = 5000$, the distribution is quite dispersed. The inter-arrival time density yield by the homogeneous renewal process is also shown for comparison. We can observe that the inter-arrival time density in the homogeneous renewal process model is an density reflecting the average distribution. To view the family of densities more completely, Figure 5.4 shows the surface plot of inter-arrival time densities over $H(t)$ and $t - H(t)$. As is shown that the densities are decreasing at any time, however, they have different means and dispersion levels.

Figure 5.5 shows the posterior estimation of $\lambda^*$ from Equation (5.10) from 1935 to 1945. All over this time window, the intensity function is always decreasing immediately after an earthquake, that means the earthquake occurrences have a clustering point pattern. However, the conditional intensity function decreases much faster right after the previous earthquake from 1943 to 1945 than from 1935 to 1943. This distinguishes from homogeneous renewal processes, where the conditional intensity function given any previous time is its hazard function shown in Figure 4.9c. The shape of the posterior estimated conditional intensity function is in good agreement with the observed data, especially from 1935 to 1940 when the earthquakes were quite absent and from 1943 to 1945 when the earthquakes were occurred frequently.

To assess the goodness-of-fit of the model to the earthquake data, we make use of the time rescaling theorem again as is discussed in section 4.3.4 and plot the Q-Q plot of
posterior estimated quantiles of the rate-rescaled probability densities versus the theoretical uniform distribution, see Figure 5.6. The left one is the Q-Q plot by the homogeneous renewal process model and the right one is the Q-Q plot by the inhomogeneous Markov renewal process model. In the plots, the quantiles up to 10th percentile, which corresponds small inter-arrival times, is better estimated by the inhomogeneous renewal process model. This means the latent time-varying process \( \{\theta(t_k)\} \) helps the model to capture those clustering earthquakes over the year 1939 to 1949. However, the uncertainty band is wider due to the variation over \( \{\theta(t_k)\} \).

Figure 5.2: earthquake data. The posterior median (red dots) and 95% quantiles (grey shaded area) of \( \{\theta(t_k)\}, k = 1, \ldots, n + 1 \).
Figure 5.3: earthquake data. The left top panel: the inter-arrival time density from homogeneous renewal process model. The histogram is from the data. The succeeding panels from top to bottom, left to right show the inter-arrival time density at different time points with previous event occurring at the 5000th, 10000th, 15000th, 20000th, 25000th and 30000th day since the year 1900, respectively. The red solid line is the posterior median and blue dash lines are posterior 95% uncertainty band of inter-arrival time density.
Figure 5.4: earthquake data. The surface plot of inter-arrival time densities over $H(t)$ and $t - H(t)$ (in days).
Figure 5.5: earthquake data. Posterior estimation of $\lambda^*$ for the earthquake data along with the observations (black dash lines).
Figure 5.6: earthquake data. Posterior Q-Q plots (mean and 95% interval) of estimated quantiles against the theoretical Uniform(0, 1) for time in earthquake data. The left one is produced from the homogeneous renewal process model and the right one is produced from the inhomogeneous renewal process model.
5.5.2 Sensitivity analysis

In this section, we perform sensitivity analysis for the selected values for the number of knots and the range parameter in the process convolution for the latent process \{\theta(t_k)\}, and assess the prior sensitivity for the parameter \(\sigma_x^2\). For other parameters \(\mu, \lambda, \phi\) which are also in the homogeneous renewal process model, their posterior estimation results are quite robust to appropriate changes in their hyper parameters.

Firstly, we consider two alternative sets of fixed values for the number of knots \(m\) and the range parameter \(\sigma_\theta^2\) in the kernel. In the section 5.5.1, we choose \(m = 6\), here we try to set \(m\) to be 15 and 25, but decrease the range parameter. In the first case, \(m = 15\) and \(\sigma_\theta^2 = (T/m)^2\). In the second case, \(m = 25\) and \(\sigma_\theta^2 = (T/m)^2\). From Figure 5.7, we find that with more number of knots in process convolution, the uncertainty of the inferences is also increased. With 15 number of knots and smaller range parameter \((T/m)^2\), the posterior estimated \{\theta(t_k)\} has more variation over time compared with Figure 5.2, especially the bump around 1980s. From the goodness-of-fit assessment, the 50-100th percentile of the Q-Q plot, which corresponds to comparatively large inter-arrival times, is improved. While, the model has worse agreement in the 0-50th percentile, which means the inter-arrivals between clustering earthquakes are not well fitted. In the second case, with 25 number of knots, the process \{\theta(t_k)\} is very wiggly due to overfitting the data and the model over-estimates the intermediate inter-arrivals from the Q-Q plot. Overall, for this data set, giving 25 knots is not suggested considering the size of data.

Then, we consider three different priors for \(\sigma_x^2\), InvGamma(3, 4), InvGamma(3, 20) and InvGamma(3, 1000). We found that \(\sigma_x^2\) is very sensitive to the prior specification and
is hard to learn from the data. The posterior density of $\sigma_x^2$ is almost as same as the prior distribution in all three cases. In the first case, the prior mean is even smaller than the variance of the logarithmic inter-arrival times, so it makes $\sigma_x^2$ too small and the stochastic process $\{\theta(t_k)\}$ is too smoothing to capture its variation over time. From the goodness-of-fit assessment, the model cannot capture the clustering events. However, the posterior inferences show little difference under the two prior with large mean values. From this result, giving $\sigma_x^2$ a prior of large mean is suggested.

Figure 5.7: earthquake data. The right column is the posterior estimation of $\{\theta(t_k)\}$ and Q-Q plots under the setting $m = 15$ and $\sigma_\theta^2 = (T/m)^2$. The right column is the posterior estimation of $\{\theta(t_k)\}$ and Q-Q plots under the setting $m = 25$ and $\sigma_\theta^2 = (T/m)^2$. 

118
5.6 Discussion

We have discussed how an homogeneous renewal process model is extended to an inhomogeneous renewal process model. The main idea is to replace the scale parameter in the mixture of Erlang by a underlying stochastic process \( \theta(t) \). We use a process convolution of a white noise process and a smooth kernel to construct the latent process \( \theta(t) \). Finally, the probability of the next event is a mixture of Erlang distribution with common scale parameter depending on the previous event time. This can be interpreted that the probability distribution of the next event is determined immediately after the previous event occurs.

Besides this model, there are also a few possible ways of modeling inhomogeneous renewal process via introducing the temporal dependence in the distribution function \( G \) which generates the mixture weights. In the homogeneous renewal process model, the weight in mixture of Erlang \( w_j \) is generated by \( G(j\theta) - G((j - 1)\theta) \). Here, we consider to develop continuously dependent measures \( \{ G_t \} \). Given the previous event \( H(t) \), let weights be generated from \( G_{H(t)} \). So, the inter-arrival time densities can be expressed as

\[
f(t \mid H(t)) = \sum_{j=1}^{J} w_j(H(t)) \text{Gamma}(t - H(t) \mid j, 1/\theta) \quad t > H(t) \quad (5.11)
\]

where the weights depend on the time of last event. \( w_j(H(t)) = G_{H(t)}(j\theta) - G_{H(t)}((j - 1)\theta) \) and \( \sum_{j=1}^{J} w_j(H(t)) \to 1 \). We propose two models, both of which adopt the same idea as above but construct the collection of random measures \( \{ G_t \} \) differently.

Model using DDP

In the homogenous renewal process model, we place a DP prior on distribution function \( G \). Following the idea of generalizing DP to the DDP (MacEachern, 1999), we start
with the stick-breaking construction of DP and apply a “single-p” construction to obtain a collection of random measures \( \{G_t\} \) with time dependent atoms,

\[
G_t = \sum_{l=1}^{\infty} p_l \delta_{g_l(t)}, \quad t \in \mathbb{R}^+
\]

where \( \sum_{l=1}^{\infty} p_l = 1 \) and the atoms \( \{g_l(t)\} \) are dependent on \( t \). Borrowing the idea of ANOVA-DDP with continuous covariates (De Iorio et al., 2004), we propose to introduce the dependence across \( \{G_t\} \) by assigning a logistic function for the atoms \( \{g_l(t)\} \) and treating \( t \) as a covariate. Specifically, the construction of the dependence across the atoms is in the form of

\[
g_l(t) = 2 \beta_{l0} \cdot \frac{\exp(\beta_{l1} t)}{1 + \exp(\beta_{l1} t)} \tag{5.12}
\]

where \( \beta_l = (\beta_{l0}, \beta_{l1}) \) follows the distribution \( G \) with a Dirichlet process prior \( DP(\alpha, G_0) \). \( G_0 \) is a bivariate distribution. \( G_0 = G_{0,0} \times G_{0,1} \). \( \beta_0 \sim G_{0,0} \) and \( \beta_1 \sim G_{0,1} \) are mutually independent and independent with time \( t \). Marginally, for any \( t \), the random measure \( G_t \) follows a Dirichlet process. Here, \( \beta_{l0} \in (0, +\infty) \) and \( \beta_{l1} \in (-\infty, +\infty) \). \( \beta_{l1} \) can be treated as the slope of \( t \). We placed a sigmoid function as the link function for the linear equation \( \beta_{l1} t \) because \( g_l(t) \) should be positive on real line. \( \beta_{l0} \) plays the role of representing the magnitude of \( g_l(t) \).

\( G_{0,1} \) is a given by \( N(\mu_0, \sigma_0^2) \). When \( \mu_0 \approx 0 \) and \( \sigma_0^2 \to 0 \), \( \beta_{l1} \approx 0 \) for all \( l = 1, \ldots \). In this case, the Equation (5.12) is simplified as \( g_l(t) = \beta_{l0} \), so \( G_t = G_{0,0} \). Therefore, the model reduces to the homogeneous renewal process model as its special case. The choice of \( G_{0,0} \) includes Weibull, Pareto and Generalized Pareto distribution, as is discussed in Chapter 4.
Model with GWP

Geometric Weight Prior (GWP) is a random probability measure analogous to DP, but it is simpler. As we know, the weights in DP prior in the stick-breaking construction is defined as

\[ p_l = v_l \prod_{j=1}^{l-1} (1 - v_j), \quad v_l \sim \text{Beta}(1, \alpha) \]  

(5.13)

The geometric weights replace random \( \{v_l\} \) within weights in Equation (5.13) with their expectation \( \omega = E(v_i) = (1 + \alpha)^{-1} \) and place a prior on \( \omega \). In different words, the infinite random weights in DP are removed and instead there is only one random variable \( \omega \in (0, 1) \) in GWP. Hence, \( G \) is a random measure with the geometric weight prior GWP, which is denoted by \( G \sim \text{GWP}(\omega, G_0) \). A realization of \( G \) from GWP(\( \omega, G_0 \)) is

\[ G = \sum_{l=1}^{\infty} p_l \delta_{Z_l} \quad \text{where} \quad p_l = \omega (1 - \omega)^{l-1} \]

with \( Z_l \sim G_0 \) and \( \omega \in (0, 1) \). It retains the same property as DP prior. The expectation of \( G \) is its centering distribution, i.e. \( E(G) = G_0 \). Because \( \omega = (1 + \alpha)^{-1} \), in the same manner as \( \alpha \) works in DP, as \( \omega \) decreases, the discrete distribution \( P \) is closer to its centering distribution \( P_0 \). Different from Dirichlet process, its construction determines that the weights are ordered such that \( w_i > w_{i+1} \). GWP also has full support as DP prior (Ongaro & Cattaneo, 2004).

Due to these nice properties of GWP and its simpler construction for the weights, we propose to replace DP prior for the distribution function \( G \) with GWP and introduce
temporal dependence through $\omega$. The model is then specified as

$$ f(t_k \mid t_{k-1}) = \sum_{j=1}^{J_k} w_j(t_{k-1}) \text{Gamma}(t_k - t_{k-1} \mid j, 1/\theta) \quad i = 1, \ldots, n $$

$$ w_j = \int 1_{[j/\theta, j/\theta]}(y_k) dG_{t_{k-1}}(y_k), \quad y_k \sim G_{t_{k-1}}, $$

$$ G_t \sim GWP(\omega(t), G_0), \quad G_0 = \text{Weibull}(\phi, \lambda) $$

Then the temporal dependence in \{G_t\} is introduced by imposing a latent stochastic process \{\omega(t)\}. This process can be constructed again, by a process convolution of a latent process $x(t)$ and a smoothing kernel $k(t)$. The $\omega(t)$ is of the form,

$$ \omega(t) = \frac{\sum_{i=1}^{m} k(t - v_i) x_i}{\sum_{i=1}^{m} k(t - v_i)} $$

where $x_i = x(v_i)$, and $v_1, \ldots, v_m$ are $m$ equally spaced time points over $(0, T)$. For each $x_i$, it follows Beta($ab, (1 - a)b$) with mean $a$ and variance $\frac{a(1-a)}{b+1}$. The kernel function is $k(\gamma) \propto \exp\{-\frac{1}{2\sigma^2}\|\gamma\|^2\}$. We normalize the $\omega(t)$ to guarantee that $\omega(t)$ is between $(0, 1)$. The mean of the process \{\omega(t)\} is $a$ but it is a non stationary process because its variance varies over time.

Overall, the model in section 5.3 introduces the time dependence in the scale parameter of mixture of Erlang, while these two models introduce the time dependence in the distribution function $G$. Marginally, for any $t$, all these models can reduce to homogeneous renewal process model and hold its flexibility property. The difference lies in the time dependence structure, which enables the changes of densities over time.

To summarize this work, our main idea considers the inter-arrival time density in the form of $f(t \mid H(t)) = f(t - H(t), H(t))$, which makes the normalizing constant in the likelihood analytically tractable. Then, according to the one-to-one relationship
between inter-arrival time density and conditional intensity function, \( f(t - H(t), H(t)) \) can be transformed to \( \lambda(t - H(t), t) \) without losing any information of the point process. This differs from the ideas utilized in neuroscience, Kass & Ventura (2001) and Rao & Teh (2011). They simplify the intensity function by the multiplicative form \( \lambda_1(t)\lambda_2(t - H(t)) \). \( \lambda_1(t) \) is interpreted as the firing rate and \( \lambda_2 \) indicates the recovery to fire after generating an action. Under the assumption of the multiplicative form, the intensity function can reduce to some special cases such as in homogeneous Poisson process with marginal intensity function \( \lambda_1(t) \) and homogeneous renewal process with hazard function \( \lambda_2(t - H(t)) \). Apparently, this multiplicative form of conditional intensity is a restrictive form rather than a general one, and it might not be suitable for other applications. Our model is more general without any assumption of the form of conditional intensity function, however, the price we pay is the marginal intensity function is not tractable from our models.
Chapter 6

Conclusions

In this thesis, we focus on some classes of temporal point processes ranging from Poisson processes to non-Poisson processes, including seasonal marked inhomogeneous Poisson processes, homogeneous renewal processes and inhomogeneous renewal processes. The main contribution of our work is that we provide flexible models for these point processes.

Originally motivated by the analysis of US hurricane occurrences, we start our research with seasonal point processes. The main feature of our model is that we focus on dynamic evolving intensities within a season and across seasons simultaneously. In the DDP-PBAR model, the intra-seasonal variability is modeled by the inhomogeneous Poisson process model by using DP mixture model with Beta distribution as the kernel, and the inter-seasonal variability is captured by dependent Dirichlet process prior for the seasonally-varying mixing distributions. Such a model provides various flexible inferences such as long-term trend of intensity functions in a certain time period across decades. Then, we extend our model by incorporating associated marks. We extended the DDP-
PBAR mixture model by adding DDP-AR components for marks to obtain a model for inhomogeneous Poisson processes over the joint marks-points space, which leads to different types of general conditional inferences. Overall, this is a comprehensive model for seasonal marked point processes from both theoretical and practical points of view.

One opportunity for further enhancements for this model is its forecasting ability. In the hurricane application, the predictive power of this model is somewhat limited especially when only a handful of observations are available in certain months. We have explored the predictive distributions of many functionals of the process and found that most of the times those distributions revert to the mean. To improve the predictive power, the future work will consider extensions to incorporate external covariates in a similar fashion to Katz (2002), Jagger et al. (2011), and Elsner & Jagger (2013). The covariates may fall into two categories. One type of covariates that change across seasons but stay constant within season, such as annual environmental index, can be introduced to the time series model for the total intensities. The other type of covariates that change within the season, such as the daily temperature, can be introduced to the density function by incorporating a regression model in the mixture weights in the same spirit of Rodríguez & Dunson (2011), Dunson & Park (2008) and Reich & Fuentes (2007).

From Chapter 4, we switch our focus to non-Poisson processes. Firstly, we work with homogeneous renewal processes and our objective is to provide a flexible model for its inter-arrival density. The model is built with a mixture of Erlang distributions with common scale parameter and a DP prior is imposed for the distribution function to generate the mixture weights. The resulting model has the capability of generating flexible shapes
of inter-arrival times and can enable both clustering and declustering point patterns by given different choices of centering distribution and precision parameter in the DP prior. Besides these nice properties, another feature of this model is that the model structure enables efficient posterior simulation while properly accounting for the normalizing constant of the renewal process in the likelihood. In Chapter 5, we further refine this model for inhomogeneous renewal processes by replacing the single scale parameter with a latent stochastic process of scale parameters in mixture of Erlang distributions, so that the inter-arrival time density also depends on the clock time.

For the inhomogeneous renewal process, our flexible mixture formulation can generate flexible shapes for the inter-arrival time density at any time point, but the changes of shapes of inter-arrival time densities over time might be relatively limited. The latent distribution function $G$ is invariant over time and only the time-varying scale parameters are responsible for the inhomogeneity among inter-arrival time densities. When the scale parameter increases, the mean and variance of each Erlang component increases, while the number of mixture component decreases. So it is unrealistic for this model to allow any types of changes of shapes over time and it may only allow very smooth changes. In the discussion section of Chapter 5, we propose two alternative models both of which introduce the time dependence to the latent distribution function $G$, which is possible to enable more flexible changes of shapes over time. Future work include more careful study on these two models.

Moreover, for the inhomogeneous renewal process, we build models for inter-arrival time density directly and obtain conditional intensity function by using their one-to-one
relationship. Thus, the resulting inferences concentrate more on inter-arrival time density.

For conditional intensity function, the only inference we can obtain is a piece-wise intensity function conditional on an observed point process, as is shown in Equation (5.10). To obtain various inferences such as the marginal intensity function $\lambda(t)$ or a two-dimensional function $\lambda(t, t - H(t))$, developing models for conditional intensity function directly in the spirit of Kass & Ventura (2001) will be considered as future work.
Bibliography


Kuo, L. & Yang, T. Y. (1996). Bayesian computation for nonhomogeneous poisson


Schmidt, S., Kemfert, C. & Hoeppe, P. (2010). The impact of socio-economics and
climate change on tropical cyclone losses in the usa. Regional Environmental Change 10, 13–26.


Appendix A

MCMC algorithm for the DDP-PBAR model

The DDP-PBAR model for the data \( \{ t_{i,k} \} \) can be expressed as follows:

\[
 t_{i,k} \mid G_k, \tau \sim \int \text{Beta}(\mu \tau, (1 - \mu) \tau) dG_k(\mu), \quad i = 1, \ldots, n_k; \quad k = 1, \ldots K
\]

\[
 G_k(\mu) = \sum_{j=1}^{N} w_j \delta_{\mu_{j,k}}(\mu)
\]

\[
 z_j \sim \text{Beta}(1, \alpha), \quad w_1 = z_1; \quad w_j = z_j \prod_{r=1}^{j-1} (1 - z_r), \quad j = 1, \ldots, N - 1; \quad w_N = 1 - \sum_{j=1}^{N-1} w_j
\]

\[
 \mu_{j,k} = v_{j,k} u_{j,k} \mu_{j,k-1} + (1 - v_{j,k}), \quad v_{j,k} \sim \text{Beta}(1, 1 - \rho), \quad u_{j,k} \sim \text{Beta}(\rho, 1 - \rho)
\]

We use an MCMC algorithm to draw posterior samples of \( (\{ \mu_{j,k} \}, \{ v_{j,k} \}, \{ w_j \}, \rho, \alpha) \), including blocked Gibbs sampling steps for the DDP prior parameters (Ishwaran & James, 2001). Configuration variables \( \{ L_{i,k} \} \) are introduced to indicate the mixture component to which each observation is allocated. We use \( n^* \) to denote the number of distinct values in the \( \{ L_{i,k} \} \), and \( L^* = \{ L_j^*: j = 1, \ldots, n^* \} \) for the set of distinct values.
The first step is to update the atoms \( \{\mu_{j,k}\} \), which depends on whether \( j \) corresponds to an active component or not. When \( j \notin L^* \), \( \mu_{j,1} \sim \text{Unif}(0,1) \), and for \( k = 2, \ldots, K \), \( \mu_{j,k} \) is drawn from \( p(\mu_{j,k} \mid \mu_{j,k-1}, v_{j,k}, \rho) \), which is a scaled Beta distribution arising from the PBAR process:

\[
p(\mu_{j,k} \mid \mu_{j,k-1}, v_{j,k}, \rho) = \frac{1}{v_{j,k} \mu_{j,k-1}} \text{Beta}(\frac{\mu_{j,k} + v_{j,k} - 1}{v_{j,k} \mu_{j,k-1}} \mid \rho, 1 - \rho)
\]

where \( \mu_{j,k} \in (1 - v_{j,k}, \min\{1, 1 - v_{j,k} + v_{j,k} \mu_{j,k-1}\}) \). When \( j \in L^* \), the posterior full conditional for \( \mu_{j,1} \) is proportional to \( \prod_{i=1}^{N} \text{Beta}(t_{i,1} \mid \mu_{j,1}, \tau, (1 - \mu_{j,1}) \tau)p(\mu_{j,2} \mid \mu_{j,1}, v_{j,2}, \rho)p(\mu_{j,1}) \). For \( k = 2, \ldots, K - 1 \), the full conditional for \( \mu_{j,k} \) is proportional to \( \prod_{i=1}^{N} \text{Beta}(t_{i,k} \mid \mu_{j,k}, \tau, (1 - \mu_{j,k}) \tau)p(\mu_{j,k+1} \mid \mu_{j,k}, v_{j,k+1}, \rho)p(\mu_{j,k} \mid \mu_{j,k-1}, v_{j,k}, \rho) \). Finally, the full conditional for \( \mu_{j,K} \) is proportional to \( \prod_{i=1}^{N} \text{Beta}(t_{i,K} \mid \mu_{j,K}, \tau, (1 - \mu_{j,K}) \tau)p(\mu_{j,K} \mid \mu_{j,K-1}, v_{j,K}, \rho) \). We use Metropolis-Hastings steps to update the \( \mu_{j,k} \), with the proposal distribution taken to be \( p(\mu_{j,k} \mid \mu_{j,k-1}, v_{j,k}, \rho) \).

The sampling of weights \( \{w_{j}\} \), configuration variables \( \{L_{i,k}\} \), and \( \alpha \) can be implemented using standard updates under the blocked Gibbs sampler. Updating the latent variables \( \{v_{j,k}\} \) involves only the PBAR process. The full conditionals are given by

\[
p(v_{j,k} \mid \mu_{j,k}, \mu_{j,k-1}, \rho) \propto \frac{1}{v_{j,k}} \text{Beta}(\frac{\mu_{j,k} + v_{j,k} - 1}{v_{j,k} \mu_{j,k-1}} \mid \rho, 1 - \rho)\text{Beta}(v_{j,k} \mid 1, 1 - \rho),
\]

where \( v_{j,k} \in \left(1 - \mu_{j,k}, \min\{1, \frac{1 - \mu_{j,k}}{\mu_{j,k-1}}\}\right) \), and sampling from each of them was implemented with a Metropolis-Hastings step based on \( \text{Beta}(1, 1 - \rho) \) as the proposal distribution. Finally, the PBAR correlation parameter \( \rho \) is also sampled using a Metropolis-Hastings step.
Appendix B

Proof of Result 4 in Chapter 4

Result 4: The $K$ function derived from our model is well defined, i.e., $K(t) < \infty$.

Proof. First, note that $K(t) = \mu_1 M(t)$. $M(t)$ can be evaluated via result 2 and 3. To guarantee that $K(t)$ is well defined, we have to prove that $\mu_1 = E(X)$ is always finite, even for the untruncated version of the model. When $J \to \infty$, $f(x) = \sum_{j=1}^{\infty} w_j \text{Gamma}(x|j, 1/\theta)$, where $w_j = G(j\theta) - G((j-1)\theta)$. So $\mu_1 = E(X) = E(E(X | G)) = E(\sum_{j=1}^{\infty} [G(j\theta) - G((j-1)\theta)]j\theta)$.

Define a series $\{a_n\}$, $a_n = \sum_{j=1}^{n} [G(j\theta) - G((j-1)\theta)]j\theta$. Then $\mu_1 = E(a_\infty)$. Since $\{a_n\}$ is the sequence of partial sum of an infinite series, obviously, $\{a_n\}$ is a monotonic
increasing sequence. Say a random variable $Z \sim G$. When $n \to \infty$,

\[
\sum_{j=1}^{\infty} [G(j\theta) - G((j-1)\theta)] \cdot j\theta = \sum_{j=1}^{\infty} G((j-1)\theta \leq Z \leq j\theta) \cdot j\theta \\
= \theta \cdot \sum_{j=1}^{\infty} G(j-1 \leq \frac{Z}{\theta} \leq j) \cdot j \\
= \theta \cdot \mathbb{E}(\lceil \frac{Z}{\theta} \rceil \mid G) \\
\leq \theta \cdot \mathbb{E}(\frac{Z}{\theta} + 1 \mid G) = \mathbb{E}(Z \mid G) + \theta = \mathbb{E}(G_0) + \theta.
\]

So $\{a_n\}$ is upper bounded by $\mathbb{E}(G_0) + \theta$, which is finite. According to monotone convergence theorem, $\{a_n\}$ has a finite limit, and then $a_\infty$ is finite. Thus, $\mu_1 = \mathbb{E}(a_\infty)$ is always finite and $K(t)$ is well defined.
Appendix C

MCMC algorithm for renewal process model

We provide the details for the steps of the MCMC used to simulate from the posterior distribution of the parameters of our developed model.

Their joint posterior of \((\theta, J)\) is

\[
p(\theta, J|y_1, \cdots, y_n, y_{n+1}, t_1, \cdots, t_n) \propto \text{Gamma}(\theta; a_\theta, b_\theta)\text{Unif}(J; \{\lceil \frac{S_1}{\theta} \rceil, \ldots, \lceil \frac{S_2}{\theta} \rceil\})
\]

\[
\times \left\{\sum_{i=1}^{J} \text{Gamma}(t_1; i, \theta)1_{((i-1)\theta, i\theta]}(y_1)\right\} \left\{\prod_{k=2}^{n} \sum_{i=1}^{J} \text{Gamma}(t_k - t_{k-1}; i, \theta)1_{((i-1)\theta, i\theta]}(y_k)\right\}
\]

\[
\times \left\{\sum_{i=1}^{J} \left(1 - \int_{t_n}^{T} \text{Gamma}(u - t_n; i, \theta)du\right)1_{((i-1)\theta, i\theta]}(y_{n+1})\right\}
\]

The Metropolis Hasting algorithm is used to sample \((\theta, J)\) jointly. The proposal distribution is \(Q(\theta, J)\). By change of variables, we have \(\log \theta\) and \(\logit \left(\frac{J \theta - S_1}{S_2 - S_1}\right)\) to follow Gaussian distributions but \(J\) should be converted to integer from the continuous proposed sample.
The acceptance rate is expressed as,

\[
\text{acceptance rate} = \frac{P(\theta^*, J^* | y_1, \ldots, y_n, y_{n+1}, t_1, \ldots, t_n) Q(\theta_t, J_t | \theta^*, J^*)}{P(\theta_t, J_t | y_1, \ldots, y_n, y_{n+1}, t_1, \ldots, t_n) Q(\theta^*, J^* | \theta_t, J_t)}
\]

Here we show how to get the value of full conditional \( P \) with \( \theta \) and \( J \). Through the latent variables \( y_1, \ldots, y_{n+1} \), we can get the index of mixture component assigned to \( t_1, \ldots, t_n \) and normalizing constant term by \( y \) variables.

The acceptance rate is expressed as,

\[
\text{acceptance rate} = \frac{P(\theta^*, J^* | y_1, \ldots, y_n, y_{n+1}, t_1, \ldots, t_n) Q(\theta_t, J_t | \theta^*, J^*)}{P(\theta_t, J_t | y_1, \ldots, y_n, y_{n+1}, t_1, \ldots, t_n) Q(\theta^*, J^* | \theta_t, J_t)}
\]

After incorporating the likelihood, the posterior full conditionals of latent variables are as follows. For \( y_1 \):

\[
\begin{align*}
p(y_1 | \{y_j : j \neq k, j = 1, \ldots, n, n + 1\}, \alpha, n) &= \frac{\alpha}{\alpha + n} G_0(y_1 | \mu, \phi) + \frac{1}{\alpha + n} \sum_{j=2}^{n+1} \delta_{y_j}(y_1) \\
&\propto \frac{\alpha}{\alpha + n} g_0(y_1 | \mu, \phi) + \frac{1}{\alpha + n} \sum_{j=2}^{n+1} \delta_{y_j}(y_1) \\
&= \frac{\alpha}{\alpha + n} g_0(y_1 | \mu, \phi) \sum_{i=1}^{J} \left( \frac{\alpha}{\alpha + n} \right) \delta_{y_j}(y_1) \\
&\propto \frac{\alpha}{\alpha + n} g_0(y_1 | \mu, \phi) \sum_{i=1}^{J} \Gamma(\alpha | \mu, \phi) \frac{1}{\alpha} \delta_{y_j}(y_1) \\
&= \frac{\alpha g_0(y_1 | \mu, \phi) \sum_{i=1}^{J} G_i(t_1, \theta) \delta_{y_j}(y_1)}{\alpha \sum_{i=1}^{J} G_i(t_1, \theta) \int_{\alpha}^{\mu} g_0(y_1 | \mu, \phi) dy_1} + \sum_{j=2}^{n+1} \left( \frac{\alpha}{\alpha + n} \right) \delta_{y_j}(y_1)
\end{align*}
\]
where $G_i(t_1, \theta) = \text{Gamma}(t_1; i, \theta)$.

For $y_k$, $k = 2, \cdots, n$:

$$p(y_k | \{ y_j : j \neq k, j = 1, \cdots, n, n+1 \}, \alpha, n, \theta, t_k, t_{k-1})$$

$$\propto \left[ \frac{\alpha}{\alpha + n} g_0(y_k | \mu, \phi) + \frac{1}{\alpha + n} \sum_{j=1, j \neq k}^{n+1} \delta_{y_j}(y_k) \right] \sum_{i=1}^{J} \text{Gamma}(t_k - t_{k-1}; i, \theta) \mathbf{1}_{(i-1)\theta, \theta]}(y_k)$$

$$= \frac{\alpha g_0(y_k | \mu, \phi) \sum_{i=1}^{J} G_i(t_k; t_{k-1}, \theta) \mathbf{1}_{(i-1)\theta, \theta]}(y_k) + \sum_{j=1, j \neq k}^{n+1} \left( \sum_{i=1}^{J} G_i(t_k; t_{k-1}, \theta) \mathbf{1}_{(i-1)\theta, \theta]}(y_j) \right) \delta_{y_j}(y_k)}{\alpha \sum_{i=1}^{J} G_i(t_k; t_{k-1}, \theta) \int_{(i-1)\theta}^{\theta} g_0(y_k | \mu, \phi) dy_k + \sum_{j=1, j \neq k}^{n+1} \left( \sum_{i=1}^{J} G_i(t_k; t_{k-1}, \theta) \mathbf{1}_{(i-1)\theta, \theta]}(y_j) \right) \delta_{y_j}(y_k)}$$

where $G_i(t_k; t_{k-1}, \theta) = \text{Gamma}(t_k - t_{k-1}; i, \theta)$.

For $y_{n+1}$:

$$p(y_{n+1} | \{ y_j : j = 1, \cdots, n \}, \alpha, n, \theta, t_n, T)$$

$$\propto \left[ \frac{\alpha}{\alpha + n} g_0(y_{n+1} | \mu, \phi) + \frac{1}{\alpha + n} \sum_{j=1}^{n} \delta_{y_j}(y_{n+1}) \right] \sum_{i=1}^{J} \left( 1 - \int_{T_n}^{T} \text{Gamma}(T - t_n; i, \theta) \right) \mathbf{1}_{(i-1)\theta, \theta]}(y_{n+1})$$

$$= \frac{\alpha g_0(y_{n+1} | \mu, \phi) \sum_{i=1}^{J} C_i(t_n; \theta) \mathbf{1}_{(i-1)\theta, \theta]}(y_{n+1}) + \sum_{j=1}^{n} \left( \sum_{i=1}^{J} C_i(t_n; \theta) \mathbf{1}_{(i-1)\theta, \theta]}(y_j) \right) \delta_{y_j}(y_{n+1})}{\alpha \sum_{i=1}^{J} C_i(t_n; \theta) \int_{(i-1)\theta}^{\theta} g_0(y_{n+1} | \mu, \phi) dy_{n+1} + \sum_{j=1}^{n} \left( \sum_{i=1}^{J} C_i(t_n; \theta) \mathbf{1}_{(i-1)\theta, \theta]}(y_j) \right) \delta_{y_j}(y_{n+1})}$$

where $C_i(t_n; \theta) = 1 - \int_{T_n}^{T} \text{Gamma}(T - t_n; i, \theta)$.

The centering distribution of Dirichlet process is $G_0(\mu, \phi)$, which we choose Weibull distribution, where $\text{Weibull}(\mu, \phi) = \frac{\phi}{\mu} \left( \frac{x}{\mu} \right)^{\phi-1} e^{-\left( x/\mu \right)^\phi}$ with mean $\mu \Gamma(1 + 1/\phi)$. The priors for $\mu$ and $\phi$ are $\mu \sim \text{IG}(a_\mu, b_\mu)$ and $\phi \sim \text{Gamma}(a_\phi, b_\phi)$. Both $\mu$ and $\phi$ can be sampled from their full conditionals by Metropolis Hasting.

$$\log p(\mu | y_1, \cdots, y_n, y_{n+1}, \phi) \propto - (a_\mu + n^* \phi + 1) \log \mu - \frac{b_\mu}{\mu} - \sum_{j=1}^{J} \frac{y_j^{1+ \phi}}{\mu^\phi}$$

$$\log p(\phi | y_1, \cdots, y_n, y_{n+1}, \mu) \propto (a_\phi - 1) \log \phi - b_\phi \phi + n^* \log \phi - n^* (\phi - 1) \log \mu + (\phi - 1) \sum y_j^\phi - \frac{\sum y_j^{1+ \phi}}{\mu^\phi}$$

The $n^*$ is the number of distinct latent variable $y_i$ and is also the number of non-zero weights in the model. $y^*$ indicates the distinct values in the $\{ y_1, \cdots, y_n, y_{n+1} \}$.

The precision parameter of Dirichlet process $\alpha$ can be sampled by introducing the auxiliary variable $\eta$ and use Gibbs sampler to draw from $p(\eta | \alpha, \text{data})$ and $p(\alpha | \eta, n^*, \text{data})$ as introduced by Escobar & West (1995).
All the updates of the above parameters $\theta, J, y, \alpha, \mu, \phi$ are based on posteriors marginalizing $G$. In order to get the full posterior inference for the $G$ and $w$, we need to sample $G$ from $G|\theta, J, y, \alpha, \mu, \phi$, which is a new Dirichlet process $DP(\tilde{\alpha}, \tilde{G}_0)$ where

$$\tilde{\alpha} = \alpha + n + 1 \text{ and } \tilde{G}_0(\cdot; \mu, \phi) = \frac{\alpha}{\alpha + n + 1} G_0(\cdot; \mu, \phi) + \frac{1}{\alpha + n + 1} \sum_{k=1}^{n+1} \delta_{y_k} (\cdot).$$

From the posterior distribution $G$, we can obtain a realization from Dirichlet distribution with parameter $(d_1, \cdots, d_J)$. For $i = 1, \cdots, J - 1$,

$$d_i = \tilde{\alpha} \left[ \tilde{G}_0(i\theta) - \tilde{G}_0((i-1)\theta) \right] = \alpha [G_0(i\theta; \mu, \phi) - G_0((i - 1)\theta; \mu, \phi)] + \sum_{k=1}^{n+1} 1_{y_k \in ((i-1)\theta, i\theta)}$$  \hspace{1cm} (C.1)

and $d_J = (\alpha + n + 1) - \sum_{i=1}^{J-1} d_i$. In practice, the number of distinct values in $y$ is much less than $J$, especially for clustering data. A lot of Dirichlet elements $d_i$ are nearly zero, when $i\theta$ is in the tail of $G_0$. So to make the MCMC more computationally efficient, we ignore the $d_i$ if it is too small compared to other Dirichlet elements so that the dimension of $w$ is significantly reduced to $J'$ and the sampled weights are $w^*$.  

149
Appendix D

MCMC algorithm for inhomogeneous renewal process model

A Gibbs sampler is used to simulate posterior samples from the posterior distribution of the parameters of inhomogeneous renewal process model. The hierarchical model can be found in (5.8). Here, we provide the MCMC details to complete the posterior inference in section 5.4.

The $y_k$ has mixture distribution, represented as $y_k = \sum_{l=1}^{N} p_l \delta(\gamma_l)$. For posterior inference, we introduce the configuration variables $L = (L_1, \ldots, L_{n+1})$ to present which mixture component is assigned to the mixture parameter. Each $L_k$ takes values in $\{1, \ldots, N\}$ such that $L_k = l$ iff $y_k = \gamma_l$ for $k = 1, \ldots, n+1$ and $l = 1, \ldots, N$. So, the MCMC steps include:
1. updating $\gamma_l, l = 1, \ldots, N$. The full conditional distribution is

$$p(\gamma_l \mid \theta, \phi, \lambda, L, u, \sigma^2, t_1, \ldots, t_n, T) \propto \text{Weibull}(\gamma_l \mid \phi, \lambda) \times \left\{ \sum_{j=1}^{J} 1_{((j-1)\theta_{t_0} \gamma_{L_1} | \gamma_{L_1} \Gamma(t_1 \mid j, 1/\theta_{t_0})} \right\} \times \left\{ \prod_{k=2}^{n} \sum_{j=1}^{J} 1_{((j-1)\theta_{t_{k-1}} \gamma_{L_k} | \gamma_{L_k} \Gamma(t_k - t_{k-1} \mid j, 1/\theta_{t_{k-1}})} \right\} \times \left\{ \sum_{j=1}^{J} 1_{((j-1)\theta_{t_n} \gamma_{L_{n+1}} \mid \gamma_{L_{n+1}} \Gamma(t_{n+1} \mid j, 1/\theta_{t_n})} \left( 1 - \int_{t_n}^{T} \Gamma(u - t_n \mid j, 1/\theta_{t_n}) du \right) \right\}$$

Let $n^*$ be the number of distinct values of vector $L$. The distinct values are \{\(L^*_l : l = 1, \ldots, n^*\)\}.

For each $\gamma_{L,0}$ and $\gamma_{L,1}, l = 1, \ldots, N$,

if $l \notin \{L^*_r : r = 1, \ldots, n^*\}$, $\gamma_l \sim \text{Weibull}(\phi, \lambda)$.

if $l \in \{L^*_r : r = 1, \ldots, n^*\}$,

$$p(\gamma_l \mid \theta, \phi, \lambda, L, u, \sigma^2, t_1, \ldots, t_n, T) \propto \text{Weibull}(\gamma_l \mid \phi, \lambda) \times 1(\gamma_{L_k} < \min (J\theta_{t_k})) \times \left\{ \prod_{L_1 \neq l}^{J} \sum_{j=1}^{J} 1_{((j-1)\theta_{t_0} \gamma_{L_1} \mid \gamma_{L_1} \Gamma(t_1 \mid j, 1/\theta_{t_0})} \right\} \times \left\{ \prod_{k=2}^{n} \sum_{L_k \neq l}^{J} \sum_{j=1}^{J} 1_{((j-1)\theta_{t_{k-1}} \gamma_{L_k} \mid \gamma_{L_k} \Gamma(t_k - t_{k-1} \mid j, 1/\theta_{t_{k-1}})} \right\} \times \left\{ \prod_{L_{n+1} \neq l}^{J} \sum_{j=1}^{J} 1_{((j-1)\theta_{t_n} \gamma_{L_{n+1}} \mid \gamma_{L_{n+1}} \Gamma(t_{n+1} \mid j, 1/\theta_{t_n}) \left( 1 - \int_{t_n}^{T} \Gamma(u - t_n \mid j, 1/\theta_{t_n}) du \right) \right\}$$

In the latter case, $\gamma_l$ can be sampled by Metropolis Hasting algorithm.

2. sampling the weights $p(p \mid L, t_1, \ldots, t_n, T)$ using standard updates in block Gibbs
sampler routine.

3. updating $L_k, k = 1, \ldots, n + 1$ from a discrete distribution on \{1, \ldots, N\} with probabilities $\tilde{p}_{lk} \propto p_l \cdot \text{likelihood involving } L_k$.

When $k = 1$, $\tilde{p}_{lk} \propto p_l \sum_{j=1}^J 1_{((j-1)\theta, j\theta]}(\gamma_{L_k}) \Gamma(t_1 | j, 1/\theta)$.

When $k = 2, \ldots, n$, $\tilde{p}_{lk} \propto p_l \sum_{j=1}^J 1_{((j-1)\theta, j\theta]}(\gamma_{L_k}) \Gamma(t_k - t_{k-1} | j, 1/\theta)$.

When $k = n + 1$, $\tilde{p}_{lk} \propto p_l \sum_{j=1}^J 1_{((j-1)\theta, j\theta]}(\gamma_{L_{n+1}}) \left(1 - \int_{t_n}^T \Gamma(u | j, 1/\theta) \, du\right)$.

4. sampling $\alpha$ using standard updates in block Gibbs sampler routine.

5. The updates of $\mu$ is by sampling from its full conditional as below with Metropolis Hasting algorithm.

$$p(\mu | x_1, \ldots, x_m, \sigma^2_x, \gamma, L, t_1, \ldots, t_n, T) \propto \Gamma(t_1; \left[\frac{\gamma L_1}{e^{\mu + K_1 x}}\right], e^{\mu + K_1 x}) \cdot \left\{ \prod_{k=2}^n \Gamma(t_k - t_{k-1}; \left[\frac{\gamma L_k}{e^{\mu + K_k x}}\right], e^{\mu + K_k x}) \right\} \cdot \left(1 - \int_{t_n}^T \Gamma(u | t_n; \left[\frac{\gamma_{L_{n+1}}}{e^{\mu + K_{n+1} x}}\right], e^{\mu + K_{n+1} x}) \, du\right) \cdot \text{Norm}(\mu | a_\mu, b_\mu)\}$$

6. The full conditional of $x = \{x_1, \ldots, x_m\}$ is

$$p(x_1, \ldots, x_m | \sigma^2_x, \gamma, L, t_1, \ldots, t_n, T) \propto \Gamma(t_1; \left[\frac{\gamma L_1}{e^{\mu + K_1 x}}\right], e^{\mu + K_1 x}) \cdot \left\{ \prod_{k=2}^n \Gamma(t_k - t_{k-1}; \left[\frac{\gamma L_k}{e^{\mu + K_k x}}\right], e^{\mu + K_k x}) \right\} \cdot \left(1 - \int_{t_n}^T \Gamma(u | t_n; \left[\frac{\gamma_{L_{n+1}}}{e^{\mu + K_{n+1} x}}\right], e^{\mu + K_{n+1} x}) \, du\right) \cdot \exp\{-\frac{1}{2\sigma^2_x} x' x\}$$

Use Metropolis Hasting algorithm, we propose a new vector $x^* = \{x_1^*, \ldots, x_m^*\}$ from multivariate Gaussian distribution centered at location which is $x = \{x_1, \ldots, x_m\}$ and
with covariance matrix $\sigma_p^2 I_m$. The acceptance ratio is 

$$\frac{p(x^*_1, \ldots, x^*_m | \sigma^2_p, \{\gamma_{L_k}, k = 1, \ldots, n + 1\}, t_1, \ldots, t_n, T)}{p(x_1, \ldots, x_m | \sigma^2_p, \{\gamma_{L_k}, k = 1, \ldots, n + 1\}, t_1, \ldots, t_n, T)}.$$ 

$\sigma_p^2$ can be tuned to get an appropriate acceptance rate.

7. The full conditional of $J$ depends on the posterior sample of $\{\theta_k\}$, $k = 1, \ldots, n + 1$. After sampling of $x = \{x_1, \ldots, x_m\}$, $\{\theta_k\}$ can be calculated by the equation $\log \theta_{t_k-1} = \mu + K_k x$. The full conditional of $J$ is

$$p(J | \sim) \propto \text{Unif}(S_1/\theta^\ast, \ldots, \rho S_1/\theta^\ast) \prod_{k=1}^{n+1} 1(\gamma_{L_k} < \min(J \theta_k))$$

8. The full conditional of $\sigma^2_x$ is,

$$p(\sigma^2_x | \sim) \propto \sigma_x^{-m} \exp\{-\frac{1}{2\sigma^2_x} x' x\} \cdot \text{IG}(\sigma^2_x | a_x, b_x) \sim \text{IG}(a_x + \frac{m}{2}, b_x + \frac{1}{2} x' x)$$

9. The full conditionals of $\phi$ and $\lambda$ are using Metropolis Hasting algorithm and their full conditionals are

$$p(\phi | \sim) \propto \text{Gamma}(\phi | a_\phi, b_\phi) \prod_{l=1}^{N} \text{Weibull}(\gamma_l | \phi, \lambda)$$

$$p(\lambda | \sim) \propto \text{IG}(\lambda | a_\lambda, b_\lambda) \prod_{l=1}^{N} \text{Weibull}(\gamma_l | \phi, \lambda)$$

As explained in Chapter 4, the full posterior inference of weights and CDF $G$ can be performed in the same manner described in Appendix C.