Semiparametric Bayesian Regression with Applications in Astronomy

by

Mary Elizabeth Broadbent

Department of Statistical Science
Duke University

Date: ____________________

Approved:

Robert Wolpert, Supervisor

James O. Berger

Sayan Mukherjee

Thomas J. Loredo

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistical Science in the Graduate School of Duke University 2014
Abstract

Semiparametric Bayesian Regression with Applications in Astronomy

by

Mary Elizabeth Broadbent

Department of Statistical Science
Duke University

Date: ______________________

Approved:

_________________________
Robert Wolpert, Supervisor

_________________________
James O. Berger

_________________________
Sayan Mukherjee

_________________________
Thomas J. Loredo

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistical Science in the Graduate School of Duke University

2014
Abstract

In this thesis we describe a class of Bayesian semiparametric models, known as Lévy Adaptive Regression Kernels (LARK); a novel method for posterior computation for those models; and the applications of these models in astronomy, in particular to the analysis of the photon fluence time series of gamma-ray bursts. Gamma-ray bursts are bursts of photons which arrive in a varying number of overlapping pulses with a distinctive “fast-rise, exponential decay” shape in the time domain. LARK models allow us to do inference both on the number of pulses, but also on the parameters which describe the pulses, such as incident time, or decay rate.

In Chapter 2, we describe a novel method to aid posterior computation in infinitely-divisible models, of which LARK models are a special case, when the posterior is evaluated through Markov chain Monte Carlo. This is applied in Chapter 3, where time series representing the photon fluence in a single energy channel is analyzed using LARK methods.

Due to the effect of the discriminators on BATSE and other instruments, it is important to model the gamma-ray bursts in the incident space. Chapter 4 describes the first to model bursts in the incident photon space, instead of after they have been distorted by the discriminators; since to model photons as they enter the detector is to model both the energy and the arrival time of the incident photon, this model is also the first to jointly model the time and energy domains.
To my parents
Contents

Abstract iv
List of Tables ix
List of Figures x
Acknowledgements xiv

1 Introduction 1

1.1 Infinitely-Divisible Processes 2
1.1.1 Example: The Gamma Process as a Lévy Process 5
1.2 Lévy Adaptive Regression Kernels 6
1.3 Gamma-ray Bursts 7
1.4 Previous work in analyzing GRB light curves 9
1.5 Outline of this thesis 11

2 Parallel Thinning for Infinitely-Divisible Models 13

2.1 Introduction 13
2.2 Parallel Thinning 14
2.2.1 Chain Construction 15
2.2.2 Re-thinning the data 18
2.3 Pedagogic Examples 18
2.3.1 Simple Poisson Example 18
2.3.2 The Gamma shape parameter: Rethinning the data 21
List of Tables

2.1 Table of locations and intensities of point sources. ........................................ 29

4.1 Priors for 1-pulse Simulation ............................................................................. 74

4.2 Parameters for 1-pulse Simulation ................................................................... 77

4.3 Parameters for Two-Pulse Simulations ............................................................... 81

4.4 Parameters for Two-Pulse Simulation Experiment ............................................. 84

4.5 Summary of Experiment 1: Number of Pulses Recovered (values with probability > 0.2) ................................................................. 85

4.6 Highest posterior density intervals (95%) for the parameters of the two-pulse models, GRB 501. ................................................................. 95

B.1 Parameters for Two-Pulse Simulation Experiment ........................................... 109
List of Figures

1.1 This figure, provided by NASA, indicates the location on the sky of 2408 GRBs observed by BATSE, using an equal-area projection of Galactic coordinates. The absence of a concentration in the Galactic plane (0° latitude) indicates that the sources of GRBs are far from the Galaxy. [6]. .................................................. 9

1.2 This figure, provided by J.P. Bonnell at NASA [5], shows the heterogeneity in GRB pulse number and shape. ................................. 10

2.1 Trace plots and Posterior histograms for the distribution of $\lambda_m$ for all three chains in a pedagogical example with a single Poisson-distributed observation. .................................................. 20

2.2 Posterior traceplots and histograms for signed square root $\theta$ of the shape parameter, with (left) and without (right) parallel thinning. . . 23
   (a) Shape Parameter with Thinning .................................. 23
   (b) Shape Parameter without Thinning ................................ 23
   (c) Posterior Histogram with Thinning ............................... 23
   (d) Posterior Histogram without Thinning ........................... 23

2.3 Posterior histograms for rate parameter $\beta$, with (a) and without (b) parallel thinning. .................................................. 24
   (a) Rate parameter with Thinning .................................. 24
   (b) Rate parameter without Thinning ................................ 24

2.4 Traceplots for $\lambda_0$ and $\lambda_1$ over 20,000 iterations of the MCMC, with (a) and without (b) parallel thinning. ..................... 26
   (a) Traceplots with thinning ........................................ 26
   (b) Traceplots without thinning ...................................... 26

2.5 Posterior histogram of $\lambda_0$ and $\lambda_1$ after 20,000 iterations of Metropolis-Hastings MCMC, both with (a) and without (b) parallel thinning. . . 27
2.6 Top: Underlying model for 10 stars, convolved with the HST-ACS PSF; Bottom: Simulated photon counts for 10 stars, based on the model displayed in the top panel.

2.7 Source detection example: Model vs. Posterior Mean

2.8 The numbers of sources detected. Note that even though the correct number of sources were found quite often, the traceplot shows that mixing between the different numbers of sources is poor.

2.9 Posterior Draws and True Source Locations

2.10 Historical traceplot for the source detection problem, without parallel thinning

2.11 The numbers of sources detected. Note that the mixing is very improved relative to the previous case.

2.12 Model, vs. Posterior Mean, with Parallel Thinning

2.13 Posterior Draws and True Source Locations, with Parallel Thinning

2.14 Historical traceplot for the source detection problem, without parallel thinning

3.1 Sample light curves for three GRBs (105, 111, 218) observed by BATSE.

3.2 Plot of Norris kernels $K_N(t \mid \omega)$ of Eqn (3.2) (black) and GiG kernels $K_G(t \mid \omega)$ of Eqn (3.3) (gray), all with pulse duration of 5 seconds and peak $A = 10$ at time $T = 0$.

3.3 The four bursts discussed in this section. From top right: Triggers 3040, 2571, 1406, and 105.

3.4 Historical plot for GRB 3040

3.5 Posterior samples for the mean for GRB 3040

3.6 Posterior histogram for the number of pulses comprising GRB 3040

3.7 Posterior histogram for kernel functional form for GRB 3040

3.8 Historical plot for GRB 2571

3.9 Posterior samples for the mean for GRB 2571
3.10 Posterior histogram for the number of pulses comprising GRB 2571. 58
3.11 Posterior histogram and traceplot for kernel functional form for GRB 2571. 59
3.12 Historical plot for GRB 1406. 59
3.13 Posterior samples for the mean for GRB 1406. 60
3.14 Posterior histogram and traceplot for the number of pulses comprising GRB 1406. 60
3.15 Posterior histogram and traceplot for kernel functional form for GRB 1406. 61
3.16 Historical plot for GRB 105. 61
3.17 Posterior samples for the mean for GRB 105. 62
3.18 Posterior histogram and traceplot for the number of pulses comprising GRB 105. 62
3.19 Posterior histogram and traceplot for kernel functional form for GRB 105. 63
4.1 A comparison of the incident and observed photon spectra. 68
4.2 (a): Credible interval (95%). The solid line is the true value. (b): The 95% pointwise posterior predictive interval. 78
4.3 Log Likelihood and $\chi^2$ statistics for single-pulse simulation. 79
4.4 $J$, the number of pulses with amplitude greater than $\epsilon = 1\text{photon/cm}^2$ included. 79
4.5 Histograms for parameters when there is only one pulse. 80
4.6 Traceplot of log-likelihood. 81
4.7 Plots related to the number of pulses included in 2-pulse simulation. 81
4.8 Histograms, Pulse 1. 82
4.9 Histograms, Pulse 2. 83
4.10 (a): Credible interval for mean (95) with true mean shown in black. (b): 100 posterior draws overlaid on the data. 86
4.11 Sample draws and number of pulses. 88
4.12 Estimated pulse shapes from the method of Hakkila et al. [22], \( J \)
constrained to one .......................................................... 89
4.13 Joint Time and Energy Model Plots for Computer Experiment ........ 91
4.14 Sample draws for two-dimensional LARK model (in methods compar-
ison experiment) .......................................................... 92
4.15 Plots for GRB 501 .......................................................... 94
4.17 Temporal parameter scatterplots for the largest pulse, GRB 501 ... 96
4.18 Interesting pulse parameter correlations, GRB 501 ................. 97
4.19 Plots for GRB 537 .......................................................... 98
4.21 Plots for GRB 493 ......................................................... 100
B.1 Histograms for \( J \) for the various configurations of the parameters .. 110
B.2 Pulse volume fractions ...................................................... 111
B.3 \( V_2 = 2 \). (a): \( \Delta_T = 1 \). (b): \( \Delta_T = 2 \). (c): \( \Delta_T = 3 \). L: 95\% Credible inter-
vals with true values overlaid in black. R: Posterior 95\% predictive
intervals ................................................................. 112
B.4 \( V_2 = 3 \). (a): \( \Delta_T = 1 \). (b): \( \Delta_T = 2 \). (c): \( \Delta_T = 3 \). L: 95\% Credible inter-
vals with true values overlaid in black. R: Posterior 95\% predictive
intervals ................................................................. 113
B.5 \( V_2 = 5 \). (a): \( \Delta_T = 1 \). (b): \( \Delta_T = 2 \). (c): \( \Delta_T = 3 \). L: 95\% Credible inter-
vals with true values overlaid in black. R: Posterior 95\% predictive
intervals ................................................................. 114
B.6 \( V_2 = 10 \). (a): \( \Delta_T = 1 \). (b): \( \Delta_T = 2 \). (c): \( \Delta_T = 3 \). L: 95\%
Credible intervals with true values overlaid in black. R: Posterior 95\%
predictive intervals .......................................................... 115
B.7 \( \chi^2 \), Computer Experiment ........................................ 116
Acknowledgements

First and foremost, I would like to thank my thesis adviser, Robert Wolpert, for his guidance, instruction, and enthusiasm, all of which made this thesis possible. Tom Loredo’s patience and clarity of thought has also been invaluable, and his careful understanding of both statistics and computing taught me an incredible amount over the last three years. Jon Hakkila’s detailed knowledge of gamma-ray bursts has been helpful over the course of many teleconferences. I also greatly appreciate Sayan Mukherjee and Jim Berger for their advice and support while serving on my committee.

I owe a great deal to David McClure, Jared Murray, and particularly Andrew Cron, for their unflinchingly honest advice and mentorship. In addition, the companionship of these and many other students has made my time at Duke more fun than I would have believed. A special thanks to Nicole Dalzell, who both carefully proofread this thesis and was always a willing ear for both frustration and ideas. Without my husband, Ted Pudlik, I would have never taken an undergraduate probability class, and hence never gone to graduate school in statistics at all. His encouragement, love, and good example inspire me every day.

This work was supported by The National Aeronautics and Space Administration’s Applied Information Systems Research grant NNX09AK60G, the Statistical and Applied Mathematical Sciences Institute, and the Duke University Graduate School.
Lévy Adaptive Regression Kernels (LARK) are a Bayesian nonparametric method used for modeling non-stationary time series and spatial data. Using a Lévy process, a type of stochastic process that is also known as infinitely divisible, as a prior distribution, Bayesian inference can be done on an unknown function. LARK has been applied in various fields such as molecular spectroscopy, vulcanology, traffic pattern analysis and pollution modeling, and includes the Gaussian process as a special case [27, 8, 28, 53].

In this thesis, I use Lévy Adaptive Regression Kernels to analyze astrophysical phenomena, with a particular focus on inferring the properties of gamma-ray bursts, an astronomical phenomenon that occurs when a star, or a binary star system, becomes a black hole. The time interval when the burst emits primarily gamma-rays (generally on a timescale of minutes), called the prompt emission, is composed of an unknown number of spatio-temporal pulses with a distinct fast-rise and exponential decay shape in the time domain. I use reversible-jump MCMC to do inference on the number of distinct pulses (and hence on the dimension of the parameter space), as well as to do Bayesian model selection over different models which could describe
the phenomena.

1.1 Infinitely-Divisible Processes

The complex, trans-dimensional posterior distributions arising in the applications studied here are difficult to explore with standard MCMC algorithms. A key innovation of this work is the development of a new approach to accelerate exploration of complex posteriors. It is motivated by the well-known parallel tempering algorithm [12], but in place of tempering or annealing the posterior, we introduce the use of thinned posteriors. This approach takes advantage of a special property of the models underlying the applications studied here: infinite divisibility of the likelihood.

The concept of an infinitely-divisible random variable is introduced in many introductory probability courses. For a random variable $X$ to be infinitely divisible, for any $m \in \mathbb{N}$, there are $m$ independent and identically distributed random variables $X_1 \ldots X_m$ such that

$$X \overset{d}{=} \sum_{i=1}^{m} X_i$$

where \(\overset{d}{=}\) refers to equality in distribution. Common infinitely-divisible distributions include the gamma, normal, and Poisson distributions.

This idea can be extended to a stochastic process in one dimension, $X(t)$, by introducing the concept of stationary independent increments (SII); distributions with SII are called Lévy processes. For a one-dimensional stochastic process $X(t)$ to have the SII property, the increments $X(t_2) - X(t_1)$ for $t_1 < t_2$ must satisfy the following conditions:

1. Stationarity: The distribution of an increment cannot depend on the location of the increment, only on the total length of the increment $(t_2 - t_1)$. 

2.
2. Independence: Disjoint increments are independent: that is, for \( t_1 < t_2 < t_3 < t_4 \), \( X(t_4) - X(t_3) \perp X(t_2) - X(t_1) \), for example.

Examples of such processes include the Poisson process, the gamma process, and the Gaussian process (known also as Brownian motion). Note that the independence of the increments means that the increments must be distributed with an infinitely-divisible distribution. To see this, consider the SII process \( X(t) \) and define \( X \equiv X(t) - X(0) \) which has distribution \( F_X(\theta, t) \) where \( \theta \) is a constant parameter. For \( m \in \mathbb{N} \), set \( X_i \equiv X(it/m) - X((i-1)t/m) \).

Then
\[
X = \sum_{i=1}^{m} X_i
\]

and \( X_i \) are independent and identically-distributed (due to the two properties of the SII process), \( F_X \) must be an infinitely-divisible distribution.

The Lévy-Khinchine theorem states that for an infinitely-divisible random variable \( X \) with dimension \( d \), the natural log of the characteristic function can be written as follows:

\[
\log \mathbb{E} \exp \{ i \omega^T X \} = i \omega^T m - \frac{1}{2} \omega^T \Sigma \omega + \int_{\mathbb{R}^d} \left( e^{i \omega^T u} - 1 - i \omega^T h(u) \right) \nu(du) \quad (1.1)
\]

where \( m \in \mathbb{R}^d \), \( \Sigma \in M_{d \times d}^d \) (positive semidefinite matrices), and \( \nu(du) \) is called the Lévy measure and satisfies some regularity conditions [31]. These three terms correspond to a constant random variable with mean \( m \), a normal random variable with mean 0 and covariance matrix \( \Sigma \), and a compound Poisson random variable with the measure \( \nu(du) \). This measure is known as the Lévy measure. To see this,
recall that the characteristics for the Poisson and Normal random variables are

\[ Y \sim \text{Po}(\mu) : \quad \mathbb{E}(e^{Y \omega}) = \exp\{\mu - 1\}\mu \]
\[ Y \sim \text{No}(\mu, \Sigma) : \quad \mathbb{E}(e^{Y \omega}) = i\omega^T \mu - \frac{1}{2}\omega^T \Sigma \omega \]

\( h(u) \) is the compensator function, and is needed to be non-zero in some cases to make the integral in 1.1 integrable. Similarly, an infinitely-divisible stochastic process \( X(t) \) can be represented as the sum of a linear drift, a Brownian motion, and a compound Poisson process with the Lévy measure as its mean. In this work we consider only those processes which omit the Brownian motion and the drift. As a result of this representation, these Lévy processes have an alternative representation as a Poisson random integral. For a measure \((S, \mathcal{F}, \nu(ds))\), the Poisson random measure \( H(dx) \) is a random function from \( \mathcal{F} \) to \( \mathbb{N} \) such that for \( A \in \mathcal{F}, H(A) \sim \text{Po}(\nu(A)) \). When \( \nu(du) \) is the Lévy measure for an infinitely-divisible random variable \( X \), then

\[ H[u] = \int_{\mathbb{R}^d} uH(du) \overset{d}{=} X \quad (1.2) \]

Similarly, for \( \nu(du) \), the Lévy measure for some process \( X(t) \) with univariate domain \( S \) and range in \( \mathbb{R}^d \), we write the Poisson random integral as

\[ H[1_{\{0 < s \leq t\}}] = \int_{\mathbb{R}^d \times (0,t]} uH(du \, ds) \overset{d}{=} X(t) \quad (1.3) \]

This construction has advantages for simulation due to the representation of a Poisson random integral as the countable sum of the support of \( H \sim \text{Po}(\nu(du)ds) \), which have independent and identical distributions proportional to \( \nu(du) \).

So, in the stochastic process case, we can write

\[ H[1_{\{0 < s \leq t\}}] = \int_{\mathbb{R}^d \times (0,t]} uH(du \, ds) = \sum \{u_j \mid 0 < s_j \leq t\} \quad (1.4) \]
where \( \{s_j\} \) are distributed uniformly on \( S \) and for each \( \epsilon > 0 \), \( \{u_j | u_j > \epsilon \} \) are distributed proportionally to \( \nu(du) \) on \( (\epsilon, \infty) \). Typically in simulation of these processes, only a finite number of pairs \( (s_j, u_j) \) are drawn, creating an approximation to the true process. A concrete example of how this is done is shown in the next subsection.

1.1.1 Example: The Gamma Process as a Lévy Process

A stationary gamma process \( X(s) \sim \Gamma(\alpha(ds), \beta) \) on a domain \( S \subset \mathbb{R} \) with Lebesgue measure \( L \) is a stochastic process such that the increments have a gamma distribution

\[
X(s_2) - X(s_1) \sim \text{Ga}(\alpha(s_2 - s_1), \beta) \quad \text{for } s_2 > s_1
\]

and disjoint increments are independent. This process is discontinuous almost everywhere.

Since all increments are non-negative, it is straightforward to observe that the gamma process is monotone increasing. The Lévy-Khinchine formula described above allows us to represent the gamma process \( X(s) \) as a cumulative sum of countably many jump discontinuities, with the size of these jumps that exceed any fixed \( \epsilon > 0 \) being proportional to the Lévy intensity measure corresponding to the gamma process \( \nu \)

\[
u \sim \nu(u) = \alpha u^{-1} \exp\{-\beta u\} \mathbf{1}_{\{u > \epsilon\}}
\]

This measure \( \nu \) is not a probability density as it is not integrable. However, given some threshold \( \epsilon \), it is possible to define

\[
\nu_\epsilon(u) = \nu(u) \mathbf{1}_{\{u > \epsilon\}} = \frac{\nu(u)}{\int_\epsilon^\infty \nu(du)}
\]

Here, \( \nu_\epsilon \), the truncated Lévy measure for the gamma process, represents the distribution of discontinuities with amplitude greater than \( \epsilon \) and can be written

\[
\nu_\epsilon(u) \equiv \frac{1}{\text{E}_1(\beta \epsilon)} u^{-1} \exp\{-\beta u\} \mathbf{1}_{\{u > \epsilon\}}
\]
We can then represent the truncated Lévy process \( X(t) \) as

\[
X(t) = \sum_{0 \leq j < J} u_j 1_{\{s_j < t\}}
\]  

(1.8)

where

\[
J \sim \text{Po} \left( \int_0^\infty \nu(du) \right) \equiv \text{Po} (\alpha \text{LE}_\beta(\beta \varepsilon))
\]

(1.9)

Truncated Lévy processes can be simulated either through reversible-jump MCMC or through the Inverse Lévy Measure algorithm described in [49].

1.2 Lévy Adaptive Regression Kernels

By specifying a kernel function (also known as a generator) or class of kernel functions, we use the mathematical framework of Lévy processes to describe a class of random functions. The kernel function or class of functions may have parameters of their own. For an infinitely-divisible process \( X(t) \) and kernel function \( K(t \mid \theta) \) with parameters \( \theta \), we can define the LARK function

\[
F(t) = \sum \{ u_j \times K(s_j \mid \theta) 1_{\{s_j \leq t\}} \}
\]

(1.10)

where \((s_j, u_j)\) represent the countable innovations resulting from the random integral representation of the process. For tractable simulation, we use a truncation parameter \( \varepsilon \), and simulate from

\[
F_\varepsilon(t) = \sum \{ u_j K(s_j \mid \theta) 1_{\{s_j \leq t\}} 1_{\{u_j > \varepsilon\}} \}
\]

(1.11)

We use functions of this form to do Bayesian inference on unknown functions. It is possible for parameters \( \theta \) to be known or to be included in the inference. In this thesis, we focus on the applications of LARK models on the light curves (time series of photon fluence) of the prompt emissions of gamma-ray bursts, described in the next section.
1.3 Gamma-ray Bursts

Gamma-rays are highly energetic photons with energies greater than $\sim 100$ keV. Slightly less energetic are x-rays, which have energies ranging from $\sim 0.1$ keV to 100 keV. The study of astrophysical phenomena which produce photons of such energy is called high-energy astrophysics. The discussion of the astrophysics in this subsection is taken largely from Chapters 1, 2 and 11 of Melia [39], which provides both a general introduction to the field of high-energy astrophysics and gamma-ray bursts in particular.

Gamma-ray bursts (GRBs) are energetic astrophysical events characterized by a short burst of photons primarily occurring in the gamma-ray spectral regime (the “prompt emission”), followed by a long, less energetic “afterglow” that initially emits x-rays, then gradually transitions to lower energies. Afterglows have been observed sequentially in the x-ray, the ultraviolet, the visible, the infrared, and the radio regimes. The duration of the prompt emission can vary from as short as 1 ms to as long as 1000 s, while the afterglow can be on the order of months long.

Shorter GRBs also tend to be “harder,” that is, the photons typically have higher energy than the longer-duration bursts. Most gamma-ray burst scientists believe that there are two distinct classes of bursts. Long gamma-ray bursts, those with prompt emissions longer than 2 seconds, are believed to arise from supernovae of certain stars [54, 40]. As a massive star collapses into a black hole, it rotates at relativistic speeds. On the outside of the star, a disk of plasma called the accretion disk forms, and energy is expelled in a jet along the axis of this disk. The exact method by which this energy forms gamma-ray bursts is unknown. However, since the star’s core collapse and the creation of this jet takes several seconds, it is believed that this is the cause of only the long GRBs. Shorter GRBs are believed to form by the merger of two neutron stars which were orbiting each other (binary stars) into a black hole;
a torus of debris similar to the accretion disk of the longer GRBs forms, and the collapse of the binary pair into a black hole releases the gamma-rays.

Detection of such high-energy photons has historically been a challenge. The atmosphere absorbs nearly all the energy of X- and gamma-rays by a combination of the photoelectric effect, Compton scattering, and – at very high energies (> 10 MeV) – the creation of positron-electron pairs begins to absorb the majority of photons (see Landau and Lifshitz [34] for a discussion of how both energy and momentum can be conserved during this process). At the lower end of the energy spectrum, some observations can be made from ground-based instruments; while the first experimental record of cosmic radiation was observed by Hess in a manned weather balloon [25], it took a sounding rocket to identify the first cosmic sources of X-rays, as well as the cosmic background which is observed on earth as a diffuse source of X-rays [13]. But all observations of gamma-rays must be done from instruments located in space.

During the Cold War, both the USA and the Soviet Union launched satellites which looked for bursts of gamma-rays, as such bursts can indicate the occurrence of nuclear tests, which had been banned. Instead of locating gamma-rays from sources on the Earth’s surface, both the American Vela satellite and the Soviet Kosmos satellite detected gamma-ray bursts with cosmic origin [32, 38]. Only in the 1990s with the advent of higher-resolution telescopes such as BATSE and Beppo-SAX were the sources of these bursts located. The BATSE instrument, in particular, was part of the first mission to create a catalog of GRBs by scanning the entire sky at 30 keV to 1 MeV. Due to their isotropic (rotationally symmetric) distribution, it became known that gamma-ray bursts have an extragalactic source. A plot of the locations on the sky of GRBs can be seen in Figure (1.1).

The energy of a gamma ray photon is difficult to measure with precision. Instruments like BATSE work by detecting optical light produced in a crystalline detection element when a gamma ray interacts with nuclei or electrons in the crystal, losing
some or all of its energy to the production of energetic electrons in the crystal. These electrons interact with the crystal producing the optical light that is ultimately detected. The mechanics for this are similar to the mechanics through which the atmosphere can absorb energy from a photon (i.e., photoelectric effect, Compton scattering, pair creation). However, as a result of the small, finite size of the detector, a gamma ray typically only loses part of its energy in the detector. As a result, there is significant uncertainty in the energies of the gamma-rays observed by BATSE and others. We will return to this issue in much more detail in Chapter 4.

1.4 Previous work in analyzing GRB light curves

Our work focuses on analysis of the emission producing binned photon count data. An example of this data can be found in Figure 1.2. The photons are sorted by the instrument into channels based on their estimated energies; we focus on data which has four energy channels. Each channel is a time-series of photon count data, known as a light curve. In some work, these light curves are summed over the different channels to create a summed-channel light curve which includes all gamma-rays at all
energies that were detected. Analysis of the summed-channel data has advantages, in that the number of pulses is estimated coherently across energy channels. However, it also obfuscates the hard-to-soft evolution of the burst – the phenomena of higher-energy photons arriving earlier than lower-energy photons – and as a result pulses in the summed-channel space appear to have a longer duration than those in individual channels.

If there is clearly a single pulse in the summed-channel data, then one can fit any of a number of empirically-determined functional forms to the light curve. Examples of these can be found in Norris et al. [42], Lee et al. [35]. For example, the Norris kernel from [42] models the photon flux as

\[ F(t \mid t_0, \tau_1, \tau_2, A) = A \lambda \exp\{-\tau_1/(t - t_0) - (t - t_0)/\tau_2\}1_{t>t_0} \]  

where \( \lambda = \exp\{2\sqrt{\tau_1/\tau_2}\} \). Here \( \tau_0 \) represents the start time, \( A \) (for amplitude) is the maximum fluence, and time constants \( \tau_1 \) and \( \tau_2 \) govern the shape of the pulse.
By reducing the light curves to this small set of parameters, it is possible to analyze correlations and distributions of properties of the GRBs [43, 21].

Some work has also been done in analyzing the uncertainty in the number of pulses that comprise the prompt emission. Perhaps the most popular method is that described in Scargle [47], known as Bayesian blocks. After fitting a Bayesian changepoint model with horizontal segments to the data, the authors use the changepoints as indications that a pulse has begun. This gives information regarding the number of pulses, although not their shape or duration, which must be estimated separately from this analysis. Additionally this analysis of the changepoints is done to each channel separately, or to the summed-channel data (i.e., the light curves for all channels are summed together). Once the timespan of the data has been divided into pulses, properties of the pulses, such as duration and maximum fluence, can be calculated empirically [43], or functional forms can be fit to the pulses separately if there is not too much overlap in their domains [42].

We develop an approach which incorporates inference on the pulse shape parameters jointly with the number of pulses. In this thesis we also seek to incorporate shared information between the energy channels, allowing for coherent inference on the number of pulses and their shape.

1.5 Outline of this thesis

In Chapter 2, we describe a novel method to aid posterior computation in infinitely-divisible models, of which LARK models are a special case, when the posterior is evaluated through Markov chain Monte Carlo. By exploiting the infinite divisibility of these models, I create better-mixing auxiliary Markov chains, the states of which can be used as proposals for the Markov chain which represents the posterior distribution. This allows for improved posterior inference in cases where the likelihood of the model is very sharply peaked or curved; or when the dimension of the param-
eter space varies and it is difficult to move between modes in different dimensions. We demonstrate the use of this method on simulated data of various complexity, including a LARK model for the number and brightness of stars in an astronomical image.

In Chapter 3, we develop a LARK model for the time spectra of GRBs. Recall that the prompt emission of a GRB is comprised of an unknown number of spatio-temporal pulses with a distinct shape in the time domain. We use reversible-jump MCMC to do inference on the dimension of the parameter space, as well as to do Bayesian model selection over different models which could describe the phenomena. Adequate mixing of the Markov chain on this complicated parameter space is difficult to achieve without the use of aids such as those described in Chapter 2.

In Chapter 4, we analyze the spectrum of selected gamma-ray bursts in both the time and energy domains, which has not been done in previous work on GRBs. Joint inference on the time and energy dependence of the GRB pulse will allow for a more detailed understanding of the generating process of these phenomena. Additionally, the effect of the response function of the BATSE instrument, which transforms the incident spectrum, creates an inverse problem; we observe the transformed data but are interested in inference in the pre-transformation space. The methods discussed in this chapter incorporate the effect of the response function to do inference on the incident spectrum of GRBs.
2.1 Introduction

In many Bayesian inference problems, the posterior distribution is highly multimodal, and standard Markov Chain Monte Carlo (MCMC) methods often fail to move between the posterior modes. This is particularly likely in situations when there is a lot of data, and the gradient of the posterior is very high around modes. That is, it unlikely for a single Markov chain to move between two “islands” of high posterior probability.

One example of such a problem comes from astronomy. Images of the sky taken by telescopes such as the Hubble Space Telescope (HST) contain an uncertain number of stars. Modeling the number of stars and their locations in the image results in multimodality both from the label-switching problem as well as from the fact that a single “feature” on the image may be from a single star or from several stars that are not well-resolved by the telescope.

Inspired by methods in thermodynamics, one approach for addressing the problem of multimodality begins by treating the negative log-likelihood function as a Boltz-
mann energy functional at unit inverse temperature, and viewing the same model with fractional powers of the likelihood as simply having higher temperatures. In this parallel tempering approach [12], multiple MCMC chains are run in parallel at different temperatures (i.e., different fractional powers of the likelihood) ranging from zero (the prior) to unity (the posterior). Occasionally MCMC moves are proposed that swap the states of these chains. Under certain conditions parallel tempering has be shown to improve mixing times [52].

2.2 Parallel Thinning

We propose a new variation on this approach, applicable only for data following infinitely-divisible (ID) distributions. For any number $p \in [0, 1]$ it is possible to write any random variable $Y$ with an ID distribution $\mu_1(dy)$ with characteristic function (ch.f.)

$$\chi(\omega) = E \exp(i\omega Y) = \int \exp(i\omega y)\mu_1(dy)$$

as the sum $Y = \xi + \phi$ of two independent random variables $\xi, \phi$ with ch.f.s $\chi^p$ and $\chi^{1-p}$, respectively. Denote by $\mu_p(d\xi)$ and $\mu_p(d\xi \mid Y)$ the marginal and conditional (given $Y$) distributions of $\xi$. For Poisson-distributed $Y \sim \mu_1(dy) = Po(\lambda)$ with ch.f. $\chi(\omega) = \exp((e^{i\omega} - 1)\lambda)$, for example, the marginal distribution of $\xi \sim \mu_p(d\xi)$ with ch.f. $\chi^p$ is $\mu_p = Po(p\lambda)$ and the conditional distribution of $\xi$ given $Y \sim \mu_1(dy) = Po(\lambda)$ is the binomial $\mu_p(d\xi \mid Y) = Bi(Y, p)$. For Gamma-distributed $Y \sim \mu_1(dy) = Ga(\alpha, \beta)$ the conditional distribution is $\xi \mid Y \sim \mu \beta$ for Beta distributed $B \sim Be(\alpha \beta, \alpha(1-p))$ independent of $Y$.

Thinning rules are not always available in convenient closed form, but they always exist for all ID distributions. We construct Markov chains whose stationary distributions are proportional to the likelihood functions for “thinned” data-sets constructed by periodically drawing sub-samples from the conditional distribution $\mu_p(d\xi \mid Y)$. 


2.2.1 Chain Construction

Let $Y$ be a locally-compact topological semigroup. Examples of such $Y$ are the real numbers $\mathbb{R}$, the non-negative reals $\mathbb{R}_+$, the integers $\mathbb{Z}$, and the non-negative integers $\mathbb{Z}_+$, all under addition. Let $\{\mu(dy \mid \theta) : \theta \in \Theta\}$ be a parametric family of ID distributions on $Y$ for some parameter space $\Theta$ equipped with a countably-generated $\sigma$-algebra (typically $\Theta \subset \mathbb{R}^k$ for some $k$). Suppose that the distributions admit density functions $\mu(dy \mid \theta) = f(y \mid \theta)\nu(dy)$ with respect to a common reference measure $\nu(dy)$. The parameter space $\Theta$ can always be written in the form $\Theta = \mathbb{R}_+ \times \Phi = \{(\lambda, \phi)\}$ with “mass” parameter $\lambda$ that is additive under convolution, so infinite divisibility is expressed as:

$$\mu(\cdot \mid \lambda_1, \phi) \ast \mu(\cdot \mid \lambda_2, \phi) = \mu(\cdot \mid \lambda_1 + \lambda_2, \phi).$$

(2.1)

For $0 < p < 1$ denote by

$$f_p(y \mid \lambda, \phi) = f(y \mid \lambda p, \phi)$$

(2.2a)

the density of $Y$ for the “$p$-reduced” distribution with mass scaled by $p$ and denote by

$$f_p(y_1 \mid y, \theta) = \frac{f_p(y_1 \mid \theta) f_{1-p}(y - y_1 \mid \theta)}{f(y \mid \theta)}$$

(2.2b)

$$\mu_p(dy_1 \mid y, \theta) = f_p(y_1 \mid y, \theta)\nu(dy_1)$$

(2.2c)

the conditional pdf and “$p$-thinned” distribution for $Y_1 \sim \mu(dy_1 \mid \lambda p, \phi)$ given the sum $Y \equiv Y_1 + Y_2 \sim \mu(dy \mid \lambda, \phi)$ of $Y_1$ and $Y_2 \sim \mu(dy_2 \mid \lambda(1-p), \phi)$ with $Y_1 \perp Y_2$.

Fix a prior distribution $\pi(d\theta)$ on $\Theta$, a positive integer $M \in \mathbb{N} = \{1, 2, \ldots\}$, and an increasing sequence of probabilities

$$0 \leq p_1 < p_1 \cdots < p_M = 1.$$
Set \( Y_M \equiv Y \) and, for each \( m < M \), draw a “\( p_m \)-thinned” random variable\(^1\)

\[
Y_m \sim \mu_{p_m}(dy_m \mid Y, \theta). \quad (2.3)
\]

We now construct a family of coupled chains \( \{ t \mapsto \theta^t_m : t = 0, 1, \ldots \} \) on \( \Theta \) in such a way that \( \{ \theta^t_m \} \) has asymptotic distribution \( \pi(d\theta \mid Y) \propto \pi(d\theta)f(Y \mid \theta) \), the Bayesian posterior distribution for \( \theta \) given \( Y \), as \( t \to \infty \).

Let \( Q(d\theta^* \mid \theta) \) be a \( \pi \)-irreducible aperiodic recurrent Markov transition kernel on \( \Theta \) [44, Chaps. 2, 5] that (for simplicity\(^2\)) is absolutely continuous with respect to \( \pi(d\theta^*) \). Select elements \( \{ \theta^s_m : 1 \leq m \leq M \} \subset \Theta \) and construct the chains \( \{ \theta^{t+1}_m \} \) using the following two types of moves. For \( t \in \mathbb{Z}_+ = \{0, 1, \ldots \} \), find \( \{ \theta^{t+1}_m \} \) by either:

**Walk:**

1. For each \( m \in \{1, \ldots, M\} \), draw \( \theta^*_{m} \sim Q(d\theta \mid \theta^{t}_m) \);

2. For each \( m \in \{1, \ldots, M\} \), set:

\[
H^t_m \equiv \frac{f_{p_m}(Y_m \mid \theta^*_{m}) Q(d\theta^t_{m} \mid \theta^*_{m})/\pi(d\theta^t_{m})}{f_{p_m}(Y_m \mid \theta^t_{m}) Q(d\theta^*_{m} \mid \theta^t_{m})/\pi(d\theta^*_{m})} \quad (2.4)
\]

3. For each \( m \in \{1, \ldots, M\} \), independently, set \( \theta^{t+1}_{m} = \theta^*_{m} \) with probability \( \min(H^t_m, 1) \). Otherwise, set \( \theta^{t+1}_{m} = \theta^{t}_{m} \).

**Swap:**

1. Select \( 1 \leq m < M \) uniformly;

---

\(^1\) For integer-valued ID distributions \( \mu(dy \mid \theta) \) and improper prior distributions \( \pi(d\theta) \) it is typically necessary to restrict attention to \( Y \) for which the posterior \( \pi(d\theta \mid Y) \) is proper and to draw \( Y_m \) from the conditional distribution \( \mu_{p_m}(dy_m \mid Y, \theta) \), given that \( \pi(d\theta \mid Y_m) \) is proper. This does not affect Propositions 1 and 2 below.

\(^2\) More generally, it is only necessary that \( Q(d\theta^* \mid \theta)\pi(d\theta) \) and \( Q(d\theta \mid \theta^*)\pi(d\theta^*) \) be equivalent on \( \Theta \times \Theta \).
2. Set:

$$H^t \equiv \frac{f_{p_m}(Y_m \mid \theta_{m+1}^t) f_{p_{m+1}}(Y_{m+1} \mid \theta_m^t)}{f_{p_m}(Y_m \mid \theta_m^t) f_{p_{m+1}}(Y_{m+1} \mid \theta_{m+1}^t)}$$  \hspace{1cm}(2.5)$$

3. With probability \(\min(H^t, 1)\) set \(\theta_{m+1}^t = \theta_m^t, \theta_{m+1}^{t+1} = \theta_m^{t+1}\), and all other \(\theta_{m'}^{t+1} = \theta_{m'}^t\). Otherwise, set all \(\theta_{m+1}^{t+1} = \theta_{m+1}^t\).

In practice it is better to avoid numerical overflow by evaluating \(h^t \equiv \log H^t\) directly and accepting the proposal when \(E^t + h^t > 0\) for a standard exponential \(E^t \sim \text{Exp}(1)\).

**Proposition 1.** For any \(0 < p_S < 1\) the sequence \(\{\theta_m^t : t = 0, 1, \ldots\} \subset \Theta^M\) constructed by applying a “swap” move with probability \(p_S\) and a “walk” move with probability \(p_W \equiv (1 - p_S)\) at each time step \(t\), independently, is a time-homogeneous Markov chain on \(\Theta^M\) whose invariant measure has the property that

$$\theta_M^t \sim \frac{\pi(d\theta) f(Y \mid \theta)}{\int_\Theta \pi(d\theta') f(Y \mid \theta')}.$$ 

the posterior distribution for \(\theta\) given \(Y\).

**Proposition 2.** For any \(n_S \in \mathbb{N}\) the sequence \(\{\theta_m^t : t = 0, 1, \ldots\} \subset \Theta^M\) constructed by alternating one “swap” move with \(n_S\) “walk” moves is a Markov chain whose invariant measure has the property that

$$\theta_M^t \sim \frac{\pi(d\theta) f(Y \mid \theta)}{\int_\Theta \pi(d\theta') f(Y \mid \theta')}.$$ 

**Proof.** Both Propositions follow from the corresponding Metropolis-Hastings results for the “walk” moves alone [see, for example, 48, §3], once it is verified that the “swap” moves preserve the distribution of \(\theta_M^t\); this follows from the equality of the integrals of the numerator and denominator in (2.5) with respect to \(\pi(d\theta_m^t)\pi(d\theta_{m+1}^t)\).
Note that we are not taking subsamples of the data to create auxiliary chains, but are rather weakening the signal from each observation in order to promote mixing among the modes. While the former approach is feasible when the observations are conditionally independent, as in the examples in Sections 2.3.1 and 2.3.2, it will not work for correlated data, as in the Poisson time-series described in Section (2.4) and Section (3). Even when feasible, this approach is not recommended: the stationary distributions of the auxiliary chains created by taking a subset of the observations may be as “peaked” as the posterior distribution, resulting in no gain in efficiency.

2.2.2 Re-thinning the data

While not necessary to ensure convergence (i.e., for the validity of Propositions 1 and 2), empirically we find mixing, and hence convergence, to be improved by periodically replacing each thinned draw $Y_m$ with a new draw $Y_m \sim \mu_{pm}(dy_m \mid Y, \theta)$. This may be viewed as a Gibbs step on an expanded parameter space $\Theta^M \times \mathcal{Y}^{M-1}$, where the complete conditional distribution of $Y_m$ for $1 \leq m < M$ is $\mu_{pm}(dy_m \mid Y, \theta)$.

2.3 Pedagogic Examples

2.3.1 Simple Poisson Example

Let $Y \sim \text{Po}(\lambda)$ be an observation from the Poisson distribution with uncertain mean $\lambda \geq 0$. The parameter space for this example is $\Theta = \{(\lambda, \phi)\} = \mathbb{R}_+ \times \Phi$ with additive mass parameter $\lambda$ and trivial $\Phi = \{0\}$, and the distribution $\mu(dy \mid \theta)$ has density function

$$f(y \mid \theta) = \lambda^y e^{-\lambda}$$

with respect to the unit Poisson reference measure $\nu(A) = \sum_0^\infty 1_A(y)/y!$ on the nonnegative integers $\mathbb{Z}_+$. For each $0 \leq p \leq 1$ the $p$-reduced distribution is $\mu_p(dy \mid \theta) = \text{Po}(\lambda p)$ and the $p$-thinned conditional distribution is $\mu_p(dy_1 \mid y, \theta) = \text{Bi}(y, p)$.
with density

$$f_p(y_1 \mid y, \theta) = p^{y_1} (1-p)^{y-y_1} (y)_y 1_{\{0 \leq y_1 \leq y\}}$$

with respect to $\nu(dy_1)$, where $(x)_n \equiv x!/(x-n)!$. A convenient and unusual feature of this distribution is the lack of dependence of $f_p(y_1 \mid y, \theta)$ on $\theta$.

For (conjugate) gamma prior distribution $\lambda \sim \text{Ga}(\alpha, \beta)$, any $M \in \mathbb{N}$, and increasing $0 < p_1 < \cdots < p_M = 1$, set $Y_M \equiv Y$ and draw $Y_m \sim \text{Bi}(Y, p_m)$ for $1 \leq m < M$. Beginning with $\{\lambda^0_m \in \mathbb{R}_+\}$, and Markov transition kernel $Q(d\theta^* \mid \theta)$ such that $\lambda^* \sim \text{LN}(\log \lambda, 0.10^2)$, i.e. a symmetric Gaussian random walk on the log scale with median steps of size $\pm 0.10$, the Hastings log acceptance ratios are:

- **Walk:** $h^m_t = (\alpha + y_m) \log(\lambda^t_m / \lambda^*_m) + (\lambda^*_m - \lambda^*_m) (\beta + p_m)$
- **Swap:** $h^m_s = (y_{m+1} - y_m) \log(\lambda^t_m / \lambda^t_{m+1}) + (\lambda^t_{m+1} - \lambda^t_m) (p_{m+1} - p_m)$

Figures (2.1a) and (2.1b) show trace plots and histograms with $Y = 100$, $M = 3$, and $p_m = m/3$, for a single draw of $Y_1 = 25$, $Y_2 = 70$, with improper prior $\pi(d\theta) = \lambda^{-1} d\lambda$, and hence true posterior $\text{Ga}(Y = 100, 1)$. In this figure we can see that all three chains converge and that the $m = 3$ case corresponds to the true posterior (shown as a solid curve).

**Poisson Marginal Distributions**

In this section we derive the exact marginal distribution of the parameters $\lambda_m$ in the thinned chains for the Poisson example with a conjugate gamma prior distribution. Fix an observation $Y \sim \text{Po}(\lambda)$ and a number $0 < p < 1$. Consider the problem of drawing $Y_m \sim \text{Bi}(Y, p_m)$ and trying to infer $\lambda$ from the marginal model

$$Y_m \sim \text{Po}(\lambda p_m), \quad \lambda \in \mathbb{R}_+.$$  

In a Bayesian approach with conjugate prior distribution $\lambda \sim \text{Ga}(\alpha, \beta)$, the conditional (given $Y_m = y_m$) Lebesgue density function for $\lambda$ is

$$\pi_m(\lambda \mid y_m) \propto \lambda^{y_m+\alpha-1} e^{-\alpha(1-p)\lambda} 1_{\{\lambda > 0\}},$$

19
so $\lambda \mid Y_m \sim \text{Ga}(\alpha + y_m, \beta + p_m)$. The MLE is $\hat{\lambda} = y/p_m$ and the Bayes posterior mean is $\bar{\lambda} = (\alpha + y_m)/(\beta + p_m)$, with normalized conditional density given by

$$
\pi_m(\lambda \mid y_m) = \frac{(\beta + p_m)^{\alpha + y_m}}{\Gamma(\alpha + y_m)} \lambda^{\alpha + y_m - 1} e^{-(\beta + p_m)\lambda} \mathbf{1}_{\lambda > 0}.
$$

It follows that the joint pdf for $\lambda$ and $Y_m$ (for reference measure $d\lambda \nu(dy_m)$), given $Y$, is

$$
\pi_m(\lambda, y_m \mid Y) = (Y)_{y_m} p_m^{y_m} (1 - p_m)^{Y - y_m} \frac{(\beta + p_m)^{\alpha + y_m}}{\Gamma(\alpha + y_m)} \lambda^{\alpha + y_m - 1} e^{-(\beta + p_m)\lambda}
$$
for $0 < \lambda < \infty$ and $0 \leq y_m \leq Y$, and that the marginal density for $\lambda$ given $Y$ is proportional to the sum

$$
\pi_m(\lambda \mid Y) \propto \sum_{y=0}^{Y} \binom{Y}{y} p_m^y (1 - p_m)^{Y-y} \frac{(\beta + p_m)^{\alpha+y}}{\Gamma(\alpha+y)} \lambda^{\alpha+y-1} e^{-(\beta+p_m)\lambda} I_{\{\lambda > 0\}}
$$

$$
= \frac{(1 - p_m)^Y (\beta + p_m)^{\alpha} \lambda^{\alpha-1} e^{-(\beta+p_m)\lambda}}{\Gamma(\alpha)} \, _1F_1(-Y, \alpha; -\lambda(\beta + p_m)p_m \frac{1}{1 - p_m}) I_{\{\lambda > 0\}}.
$$

Here $\,_1F_1(a, b; z)$ is the confluent hypergeometric or Kummer $M$ function [1, §13], a polynomial for negative integers $a$. This is the density of the marginal stationary distribution for $\lambda_m$ in the Poisson model of Section (2.3.1) for each $1 \leq m < M$ (all independent). For $m = M$ we get the limit as $p_m \to 1$,

$$
\pi_M(\lambda \mid Y) \propto \frac{(\beta + 1)^{\alpha+Y}}{\Gamma(\alpha+Y)} \lambda^{\alpha+Y-1} e^{-(\beta+1)\lambda} I_{\{\lambda > 0\}},
$$

which is the density for the conjugate posterior distribution $\lambda \mid Y \sim \text{Ga}(\alpha + Y, \beta + 1)$.

The limit as $\beta \to 0$

$$
\pi_m(\lambda \mid Y) = \frac{(1 - p_m)^Y p_m^{\alpha} \lambda^{\alpha-1} e^{-p\lambda}}{\Gamma(\alpha)} \, _1F_1(-Y, \alpha; -\lambda p_m^2 \frac{1}{1 - p_m}) I_{\{\lambda > 0\}}
$$

is well-defined for any $Y \in \mathbb{Z}_+$, but the marginal for $\alpha = 0$ is improper since $P[Y_m = 0] > 0$.

### 2.3.2 The Gamma shape parameter: Rethinning the data

In order to construct a simple example with a multimodal posterior distribution, consider inference about a parameter vector $(\theta, \beta) \in \mathbb{R} \times \mathbb{R}_+$ upon observing gamma-distributed $Y \sim \text{Ga}(\theta^2, \beta)$ with uncertain shape parameter $\alpha = \theta^2$ and rate parameter $\beta$. The likelihood function for $(\theta, \beta)$ is

$$
f(y \mid \theta, \beta) = \frac{\beta^{\theta^2}}{\Gamma(\theta^2)} y^{\theta^2-1} \exp\{-\beta y\}.
$$
The posterior distribution of $\theta$ will be bimodal and symmetric about zero for a prior that is likewise symmetric. However, it will be difficult for a Metropolis-Hastings Markov chain to cross the deep valley between the two modes without parallel thinning. The additive mass parameter from Eqn (2.1) in Section (2.2.1) is $\alpha = \theta^2$ in this example, and the thinning distribution for $Y \sim \text{Ga}(\theta^2, \beta)$ is a scaled (by $Y$) $\text{Be}(\theta^2 p, \theta^2 (1-p))$ distribution. Note that although the thinning depends on the uncertain parameter $\theta$, this causes no trouble since the thinning may be viewed as an MCMC Gibbs step (see Section (2.2.2)).

For this numerical example, we generate one hundred independent $\text{Ga}(100, 1)$ random variables. We use independent non-conjugate prior distributions for $\theta$ and $\beta$:

$$\theta \sim \text{No}(0, \sigma = 10) \quad \beta \sim \text{Ga}(1, 1).$$

We use $M = 3$ levels of thinning with thinning parameters $p_m = 0, 0.2, \text{and } 1$ representing the prior, an intermediate chain, and the posterior respectively. A Metropolis-Hastings Markov chain Monte Carlo simulation is performed with 20,000 iterations. The thinned data are redrawn every twenty iterations as described in Section (2.2.2).

Trace plots and histograms for the parameter $\theta$ are shown in Figure (2.2). Here we see that the parallel thinning approach (left panels) reaches both modes near $-10$ and $10$ repeatedly, while a standard MCMC approach gets stuck at one of the modes (right panels). The mode reached by the unthinned chain (and hence the apparent support of the posterior distribution) depends of course on the starting location; the thinning approach visits both modes and reaches the correct posterior distribution from essentially any starting point.

Figure (2.3) shows posterior histograms for the rate parameter $\beta$ with (left panel) and without (right panel) parallel thinning, to illustrate that inference on this pa-
Figure 2.2: Posterior traceplots and histograms for signed square root $\theta$ of the shape parameter, with (left) and without (right) parallel thinning.

Parameter is unaffected by parallel thinning.

2.3.3 Extension to Mixtures of ID Distributions

Consider a random sample $Y = \{y_1, \ldots, y_n\}$ from a two-component equal mixture of Poissons with means $\lambda_0$ and $\lambda_1$. That is, for each observation $Y_i$, there is an unobserved Bernoulli variable $\zeta_i \sim \text{Bi}(1, \frac{1}{2})$ and $Y_i \mid \zeta_i \sim \text{Po}(\lambda_{\zeta_i})$.

The likelihood function of the ordered pair $(\lambda_0, \lambda_1)$ is symmetric, and so will have two distinct modes for some $\lambda_0 \neq \lambda_1$. For symmetric prior distributions this will
typically lead to a bimodal posterior distribution.

The familiar label-switching problem can of course be dealt with in several ways. In the present example a simple solution is to restrict the ordering of the parameters by using a prior distribution concentrated on the ordered set $\lambda_0 \leq \lambda_1$. Here we consider a model without such a restriction, however, to provide a transparent example of a multimodal posterior to illustrate our parallel thinning approach. Traditional Markov chain Monte Carlo methods, such as the random-walk Metropolis-Hastings algorithm, mix very slowly on multimodal posteriors. This is especially true if data are plentiful and the two modes are separated by a deep chasm on the likelihood surface. However, parallel thinning allows the chain to move easily between the two high-probability regions.

The parallel thinning approach described in Section (2.2) is not immediately applicable, because the mixture model is not infinitely divisible. This follows immediately from the observation that fractional powers of its characteristic function

$$
\chi(\omega) = E[e^{i\omega Y}] = \left[ \frac{1}{2} \exp\{\lambda_0(e^{i\omega} - 1)\} + \frac{1}{2} \exp\{\lambda_1(e^{i\omega} - 1)\} \right]
$$
are not positive definite. However, *conditioned* on the values of the mixture labels \(\{\zeta_j, j = 1 \ldots n\}\), the distribution of the \(\{Y_j\}\) are Poisson and hence ID, so a conditional thinning distribution exists for \(Y\) given the \(\zeta_j\)’s. Rethinning the data is a simple Gibbs step (see Section (2.2.1)), so it is appropriate to condition on all of the parameters when rethinning. For Poisson data, the parallel thinning can be implemented quite easily as follows.

After fixing a set of thinning parameters \(0 \leq p_1 < \cdots < p_M = 1\), thinned data sets \(\{Y_1, \ldots, Y_M\}\) are sampled with \(Y_{mj} \sim \text{Bi}(Y_j, p_m)\), and parallel Markov chains \(\{\lambda_0, \lambda_1, \ldots\}\) are constructed from the model \(Y_{mj} \sim \text{Po}(\lambda_{\zeta_j} p_m)\) with diffuse gamma priors for \(\{\lambda_i\}\) centered around 30. Simulation from the posterior is done through Metropolis-Hastings with a symmetric random walk proposal distribution, reflecting at zero.

Since full conditional distributions are available for this model it would be possible to use a Gibbs sampler, although mixing between the modes would still be slow. Because Gibbs samplers are unavailable in most practical applications with complicated mode structures, we present this simple example using a random-walk Metropolis-Hastings algorithm for the parameter vector.

Although the marginal distribution of the thinned chain, conditioned on the labels \(\zeta_j\), can be found explicitly as in Section (2.3.1), it is in general unnecessary to derive these marginal distributions (conditioned on other parameters or not) to apply this approach to practical problems, since only the unthinned chain \(\theta = \theta_M\) is of interest.

The likelihood for the \(m = M\) chain with \(p_M = 1\) is based on the entire data set, and \(\{\theta_M\}\) will have the target posterior distribution. For \(1 \leq m < M\) the \(m\)th chain is based on the reduced likelihood from \(Y_{mj} \sim \text{Po}(\lambda_{\zeta_j} p_m)\). The chains are moved in parallel for one or several steps, and occasionally a “swapping” between two adjacent chains is attempted. Additionally, at fixed intervals, each subsample \(Y_m\) is redrawn from \(\text{Bi}(Y, p_m)\). In the simple case of the two-component mixture, a single auxiliary
chain, with a corresponding thinning parameter of $p_1 = 0.5$, was used. Figures (2.4) and (2.5) show that this seems sufficient to mix between the two modes completely in a chain of moderate length.

Figure (2.4) illustrates the mixing in the chains with (left) and without (right) parallel thinning. Without parallel thinning, the chain is fixed with $\lambda_0 \approx 18$ and $\lambda_1 \approx 30$. This is but one of the two existing modes— the other of course has $\lambda_0 \approx 30$ and $\lambda_1 \approx 18$. Although the modes are equally likely a posteriori, the Metropolis-Hastings algorithm does not permit the chain to cross the valley that separates them. The inclusion of a single auxiliary chain with parallel thinning allows multiple crossings between the two modes.

Figure (2.5) show the posterior histograms for $\lambda_0$ and $\lambda_1$. The superior mixing of the parallel tempering algorithm compared with the Metropolis-Hastings algorithm alone is even more evident here: the histograms for $\lambda_0$ and $\lambda_1$ are identical, as we know from symmetry they should be.
2.4 An Example from Astronomy

Detecting astronomical objects in images is a key problem in astronomy. The data in most cases is a 2D image which represents a small segment of the sky. In this piece of the sky, there are an unknown number of stars which emit photons. Due to the resolution of the telescope, these stars can be considered point sources.

Each element of the 2D image contains the photon count for one pixel. With a perfect imaging system, for each point source, there would be a single corresponding pixel, and the number of photons in that pixel would be proportional to the apparent magnitude of the star. However, the stars are at varying distances from the telescope – in fact their locations are not necessarily known \textit{a priori} – and it is not possible to focus on them. As a result, the photons from each star are blurred into neighboring pixels.

This blurring can be modeled as a convolution of the underlying signal, and the kernel in this convolution is called the “point spread function” (PSF). For a given telescope, the PSF is well-characterized and can often be treated as known. When point sources are close together (known as the “crowded-field” regime) the convolved signals from the sources often overlap, making distinguishing them from one another
challenging.

Previous work using the principles and ideas of Bayesian statistics on this topic includes [26], [19], [11], and [9]. Previous work typically does not account for the uncertainty in the number of point sources, instead either fixing that number as if it were known, or finding point sources iteratively. We use Levy regression, first introduced in [49], in order to find all point sources simultaneously, as well as rigorously quantify our uncertainty regarding the number of sources.

As an example of the utility of parallel thinning, we use simulated data in order to compare our results to the underlying model. We use a point spread function of the Hubble Space Telescope Advanced Camera for Surveys (HST-ACS), as implemented in the software package Tiny Tim [33]. We simulate an image of ten stars with varying apparent magnitude clustered near the middle of a 30×30 grid. The sources are parameterized by a location in 2D space and an intensity. After being transformed through the point spread function, a constant background rate of 10 photons per pixel is added and the intensities in each pixel are used as Poisson means to draw the simulated data. The locations of the point sources, the convolved model, and the simulated data are shown below in Figure 2.6.

Table 2.1 shows the coordinates and intensities of the ten point sources used in Figure 2.6. Some of the smallest point sources have rates only twice the background rate. When these intensities are spread out into neighboring pixels by the point-spread function, they will be even harder to detect.

In a source detection problem, the number of stars in the field is unknown, and their number, \( K \), is a random variable in the model. Let the intensities of the \( K \) point sources be \( A_1, \ldots, A_K \), and represent their locations as \((x_{(1)}, y_{(1)}), \ldots, (x_{(K)}, y_{(K)})\). We assume that the effect of the camera is stationary across the field. This assumption fails on real charge-coupled devices, due several instrumental issues, including the non-constancy of the voltage applied across the CCD, and permanent damage done
Table 2.1: Table of locations and intensities of point sources.

<table>
<thead>
<tr>
<th>$x$</th>
<th>$y$</th>
<th>$I$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.68</td>
<td>0.75</td>
<td>1000</td>
</tr>
<tr>
<td>0.76</td>
<td>0.63</td>
<td>458</td>
</tr>
<tr>
<td>0.47</td>
<td>0.38</td>
<td>193</td>
</tr>
<tr>
<td>0.88</td>
<td>0.58</td>
<td>120.</td>
</tr>
<tr>
<td>0.28</td>
<td>0.71</td>
<td>91.6</td>
</tr>
<tr>
<td>0.93</td>
<td>0.89</td>
<td>90.9</td>
</tr>
<tr>
<td>1.38</td>
<td>1.13</td>
<td>33.6</td>
</tr>
<tr>
<td>0.88</td>
<td>1.07</td>
<td>27.7</td>
</tr>
<tr>
<td>0.53</td>
<td>0.82</td>
<td>26.2</td>
</tr>
<tr>
<td>1.02</td>
<td>0.46</td>
<td>24.8</td>
</tr>
</tbody>
</table>

to some pixels from cosmic rays. We represent the point spread function for a point source centered at $(0, 0)$ as the stationary function $P(x, y)$.

Conditioning on $K$, the likelihood of a count in a given pixel $(x_i, y_j)$, $Y_{(i,j)}$ is distributed Poisson with mean

$$
\lambda_{(i,j)} = \sum_{k=1}^{K} A_k P(x_i - x_{(k)}, y_i - y_{(k)})
$$

We model $K, A_1 \ldots A_K$, and $(x_{(1)}, y_{(1)}) \ldots (x_{(K)}, y_{(K)})$ as the point masses in a Gamma random field. The Gamma process is discontinuous almost everywhere, and given some $\epsilon > 0$, the number of jumps in the process on a compact set exceeding magnitude $\epsilon$ has a Poisson distribution. For a stationary Gamma process with parameters $\alpha$ and $\beta$, this Poisson mean can be written

$$
K \sim \text{Po}(\alpha T E_1(\beta \epsilon))
$$

where $E_1$ is the exponential integral function [1], and $T$ is the area (in this two-dimensional case) of the compact set.

The jumps are uniformly distributed on the set, and the amplitudes of the jumps, $A_1 \ldots A_K$, are distributed

$$
f(a) = \frac{1}{E_1(\beta \epsilon)} a^{-1} \exp\{-\beta a\} 1_{\{a>\epsilon\}}
$$
This distribution for the amplitudes of the point sources is similar to the Schecter function [36] used for the distributions of luminosities in the SimClust algorithm in Longair [36]. Together, these distributions represent our priors for the number, location, and luminosity of the point sources in the image.

Inference on these parameters is done by Markov chain Monte Carlo. For a given number of point sources, $K$, we use the Metropolis-Hastings algorithm to find the posterior distributions of their locations and amplitudes. However, since the dimension of the parameter space is unknown, it is necessary to use reversible jump techniques to explore uncertainty about $K$ [18]. Reversible jump is an algorithm for exploring parameter spaces of varying dimension through transdimensional proposal distributions. Details of our implementation can be found in Appendix 1.

For the sake of comparison, we attempt to draw from the posterior distribution twice. In the first attempt, we do not use any parallel thinning, while in the second, we use Parallel Thinning to enhance mixing. From this we can see the impact on inference from the addition of the Parallel Thinning.

2.4.1 Results: Without Thinning

The parameters for the Gamma distribution in this simulation are $\alpha = 0.211$ and $\beta = 0.016$. This results in a prior mean for the number of pulses at approximately 3.6. The expected value of the brightest source \textit{a priori} is 50. The MCMC was run for 100,000 iterations, and plots below summarize the results.

It is difficult to summarize the posterior distribution of, for example, a single point source, since the number of point sources included in the model varies as the chain proceeds. One straightforward way to see how well the algorithm performed is to compare the posterior mean to the true model. This is shown in Figure 2.7. Clearly the posterior mean is very close to the true model, even before parallel thinning is used.
**Figure 2.6:** Top: Underlying model for 10 stars, convolved with the HST-ACS PSF; Bottom: Simulated photon counts for 10 stars, based on the model displayed in the top panel.
Figure 2.7: Source detection example: Model vs. Posterior Mean

Figure 2.8: The numbers of sources detected. Note that even though the correct number of sources were found quite often, the traceplot shows that mixing between the different numbers of sources is poor.

Additionally, it is of interest whether or not the correct number of sources are found. Figure 2.8 includes both a traceplot and a histogram for the number of point sources. The number of sources that are found is close to the true number in this simulation.

We can also plot the posterior locations of the point sources at different stages in the Markov chain. At every 10\textit{th} iteration (simply thinning to reduce autocorrelation), we plot the values of \((x_{(1)}, y_{(1)}) \ldots (x_{(K)}, y_{(K)})\) at that iteration on a grid representing the image. We also plot the true locations of the sources, for the sake
of comparison. This can be seen in Figure 2.9. We can see in this plots that there is posterior coverage of all of the true source locations. However, the uncertainties about these locations seem to be quite low, and it seems possible that the MCMC is not mixing well. This means that the samples may be highly autocorrelated, which would result in overconfident inference.

It is difficult to evaluate convergence of a MCMC chain when the dimension of the parameter space varies over the course of the chain. To attempt to visualize some of the information contained in the posterior distribution, we use a “historical traceplot” of the locations of the detected point sources. The plot is called historical is the fact that the points are colored according to their iteration number. Earlier iterations’ point sources are colored red, and eventually they pass through the rainbow so the final iterations are colored violet. If mixing is good, the red and purple points should be mixed, and the net color should be roughly brown. However, if there is a trend in the MCMC, then the colors will be well-separated. As can be seen below in Figure 2.10, the colors indicate that the MCMC is not mixing as well as might be desired.
2.4.2 Results: Using Parallel Thinning

We use the same parameters as in Section 2.4.2 to perform a run of the MCMC algorithm with parallel thinning. We use four auxiliary chains, at levels \( p = 0.8, 0.9, 0.95, \) and 0.97. These were chosen in order to give acceptance for “swap” moves of rates between 15 and 25%. Plots summarizing the posterior distribution follow. As shown in Figure 2.11, the correct number of point sources was correctly identified, and there is posterior mass around each of the true point sources. That the modal pos-
The posterior value of the unthinned chain was the true value was due to poor mixing of the Metropolis-Hasting algorithm which happened to be “stuck” in the correct value.

Figure 2.13 compares the model to the locations of the posterior draws. Here we clearly see that the smaller sources tend to be missed, while the larger ones are found with very high probability. The algorithm continues to easily recover the posterior mean, as shown in Figure (2.12). In addition, when comparing this figure to Figure
2.9, we can tell that the uncertainties in the locations of the sources are more clearly delineated. Finally, the historical traceplot, in Figure (2.14) is improved over the un-thinned case (Figure (2.10)).

2.5 Discussion

Parallel thinning improves mixing between the modes of a posterior distribution when the likelihood of the data is conditionally, at least, infinitely divisible. It improves mixing significantly in both simulated and scientific data. The auxiliary chains created can be run in parallel, reducing the computational load of the algorithm.
3

Modeling Gamma-Ray Burst Time Series

3.1 Introduction and Motivation

Gamma Ray Bursts (GRBs) are flashes of highly energetic gamma rays thought to be associated with supernova events, the collapse of massive rotating stars billions of light-years from Earth to form neutron stars or black holes. They were first detected in 1967 by the Vela satellites intended to detect atmospheric and space-based nuclear tests [32]. The BATSE (Burst and Transient Source Experiment) instrument [6] on NASA’s Compton Gamma-Ray Observatory continually recorded observations of GRBs, pulsars, and other transient phenomena throughout its operation from April 1991 to June 2000. The 2704 GRBs detected in that experiment have been cataloged and made available to researchers [15]; see Figure (3.1) for three examples. Each burst, typically lasting from a few seconds to a few minutes, appears to comprise a modest number (perhaps 1–20) of distinct “pulses” thought to be associated with distinct shock wave events arising from the stellar collapse. In this chapter we consider the question of how many pulses there might be, and what shape they might take.
3.2 Description of Data

In this work we consider data produced by the BATSE discriminators, one of two measurement methods on the BATSE instrument. The data from these discriminators consist of integer counts $N_i$ of photons detected within a specified energy range during a specified time interval $(t_{i-1}, t_i]$, typically 16 ms or 64 ms long. Due to bandwidth and storage restrictions on BATSE, the higher resolution time series is only available for a short interval when a burst is detected. A rolling 2s window of high-resolution (0.064s interval) data is maintained by the instrument, but unless a GRB is detected, only low-resolution data is sent to Earth. The time at which a burst is detected by the instrument is called the trigger time, and is indicated as $t = 0$ in all datasets. Typically about two minutes of high-resolution data is available per GRB.

There are four energy ranges, called “channels”, included in this type of data. Channel 1 includes photons with energies from 20 to 50 keV; Channel 2, from 50 to 100 keV; Channel 3, from 100 to 300 keV; and Channel 4, which includes all photons with energies recorded above 300 keV.

For each fixed energy band, we model the count $N_i$, the $i$th time interval for a
given energy range as independent Poisson random variables with uncertain means 
\( \lambda_i = E N_i \), and model the means as integrals

\[
\lambda_i := \int_{t_{i-1}}^{t_i} f(s) \, ds
\]

of an uncertain intensity function \( f(s) \) of time. This function is given in LARK form as

\[
f(t) = B + \sum_{j=1}^{J} A_j K(t \mid \omega_j)
\]

for background rate \( B \geq 0 \) and number \( J \in \mathbb{N} \) of pulses of magnitudes \( \{A_j\} \) and shapes determined by some family of kernel functions \( K(t \mid \omega) \); with pulse-specific parameter vectors \( \{\omega_j\} \) governing the pulses’ shapes and locations. The constant background rate \( B \), of no scientific interest, is fit empirically and then treated as known.

### 3.3 Model Selection

Most GRB plots appear asymmetric, with Fast Rise, Exponential Decay or “FRED” shapes. A popular choice for modeling the fast rise and exponential decay of the pulses that constitute a gamma ray burst is the Norris kernel

\[
K_N(t \mid \omega) \propto \exp\left\{\frac{-\tau_1}{(t-t_0)} - \frac{(t-t_0)}{\tau_2}\right\} \mathbf{1}_{\{t>t_0\}}.
\]  

(3.2)

Its introduction was based not on theoretical grounds, but rather on empirical evidence and flexibility in fitting a wide variety of FRED shapes \[42\]. Parameter vector \( \omega = (T, \tau_1, \tau_2) \) includes the peak time \( T = t_0 + \sqrt{\tau_1 \tau_2} \) at which \( K_N(t \mid \omega) \) attains its maximum (offset by \( \delta = \sqrt{\tau_1 \tau_2} \) from the onset time \( t_0 := T - \delta \)), and parameters \( \tau_1 > 0, \tau_2 > 0 \) controlling pulse rise and decay times. The kernel \( K_N(t \mid \omega) \) is normalized to have a maximum height of unity, so the \( \{A_j\} \) in Eqn (3.1) represent the peak pulse amplitudes above background.
Extending the set of functions available to fit pulses in a gamma ray burst will allow for more flexible modeling, possibly achieving adequate fit with fewer pulses. Here we introduce a five-dimensional extension of the four-dimensional Norris family of kernels, proportional to the probability density function of the generalized inverse Gaussian (GiG) probability distribution [17, 29, §9.3] that is commonly used to model hitting times of Brownian motion and other diffusions [4]. We parameterize this kernel family in the form

\[
K_G(t | \omega) \propto (t - t_0)^{p-1} \exp\{-(\tau_1/(t - t_0) + (t - t_0)/\tau_2)\}1_{t>t_0}, \quad (3.3)
\]

again normalized to have a maximum amplitude of unity. The argument vector \(\omega = (T, p, \tau_1, \tau_2)\) now includes a new shape parameter \(p \in \mathbb{R}\), along with time parameters \(\tau_1 > 0, \tau_2 > 0\) and peak time \(T \in \mathbb{R}\), which together determine the onset time \(t_0 = T - \delta(\tau)\) now offset from the peak \(T\) by

\[
\delta(\tau) := (\tau_2/2)\left[(p - 1) + \sqrt{(p-1)^2 + 4\tau_1/\tau_2}\right]. \quad (3.4)
\]

This reduces to the Norris kernel (3.2) for parameter value \(p = 1\), but for other values of \(p \in \mathbb{R}\) (not necessarily positive) it is new.

Further elaborations are possible: the pulse described by taking the GiG kernel
to a power $c$

$$K_D(t \mid \omega) \propto ((t - t_0)^{p-1} \exp\{-\tau_1/(t - t_0) + (t - t_0)/\tau_2\})^c 1_{\{t > t_0\}}, \quad (3.5)$$

for some $c \in \mathbb{R}$ also has the fast rise and exponential decay we desire, while being even more flexible than the GiG kernel. In particular, this kernel allows for more peaked behavior at the maximum flux of a pulse. We call this the the power GiG kernel (pGiG).

We will attempt to distinguish which of these models is most suitable for the gamma-ray burst data using Bayesian model selection procedures.

Since we are evaluating nonparametric models in which the number of pulses within a gamma-ray burst is unknown \textit{a priori}, it is not feasible to use a test such as the one in [30] in which the test statistic depends on the dimension of the model. This is particularly evident when one realizes that the distribution of model sizes (i.e., number of pulses) may vary based on the model’s kernel.

We are able to obtain estimates of Bayes factors for competing models from Markov Chain Monte Carlo by doing reversible-jump among the models [23]. The key challenge in implementing the reversible-jump is to create a reversible map between the different model types. Implementation of this involves a reversible-jump “switch” step which switches the model between Norris, GiG, and pGiG kernels. The reversible-jump “switch” step is a Metropolis-Hastings step with a proposal distribution from one model space to another. The details of the calculation of the Hastings ratio for this step are below.

Recall that the models are nested, with common parameters $T$, $A$, $\tau_1$ and $\tau_2$: pGiG is the “largest” with free parameters $c$ and $p$. The GiG kernel has $c = 1$, and the Norris has $c = 1$ and $p = 1$. When implementing the pGiG model, we used normal priors for $c$ and $p$. We need to add the point masses at 1 as described above. With this adaptation, the posterior probability that $p = 1$ is the posterior probability of
the Norris kernel model, and $P(c = 1) - P(p = 1)$ is the posterior probability of the GiG kernel model.

Consider the proposed switch move from $(M, \theta) \rightarrow (M^*, \theta^*)$ where $M$ is the model indicator (0 for Norris, 1 for GiG, and 2 for pGiG). The data is called $Y$. Then the Hastings ratio for this move is:

$$H = \frac{p(Y \mid \theta^*, M^*)}{p(Y \mid \theta, M)} \times \frac{p(\theta^* \mid M^*)}{p(\theta \mid M)} \times \frac{p(M^*)}{p(M)} \times \frac{q(M^*, \theta^*) \rightarrow (M, \theta)}{q(M, \theta) \rightarrow (M^*, \theta^*)}$$

Note that the likelihood of the data given $\theta$ is independent of $M$, due to the nested nature of the models, and so $p(Y \mid \theta, M) = p(Y \mid \theta)$. This results in a Hastings ratio of

$$H = \frac{p(Y \mid \theta^*, M^*)}{p(Y \mid \theta)} \times \frac{p(\theta^* \mid M^*)}{p(\theta \mid M)} \times \frac{p(M^*)}{p(M)} \times \frac{q(M^*, \theta^*) \rightarrow (M, \theta)}{q(M, \theta) \rightarrow (M^*, \theta^*)}$$

There are several methods of defining a transition kernel $q$ between two models. The simplest way is to map a GiG kernel to the Norris kernel with the same shared shape parameters, $\tau_1$ and $\tau_2$, with $p$ fixed at 1; and likewise to map a pGiG to a GiG with the same $p$, $\tau_1$, and $\tau_2$, but with a $c$ fixed to 1.

The transition back from the smaller models to the larger ones must include a random component, for the sake of ergodicity. One straightforward approach is to draw the new $p$’s from the prior distribution of $p$. Since the priors for the parameters are independent, this has the advantage of canceling out the second fraction in the above equation. Using this definition for $q$, the new Hastings ratio is

$$H = \frac{p(Y \mid \theta^*)}{p(Y \mid \theta)} \times \frac{p(\theta^* \mid M^*)}{p(\theta \mid M)} \times \frac{p(M^*)}{p(M)}$$

As we have noted above, there are many possible transition functions between models. The method described above is simple, but may not result in the highest acceptance rate, since fixing all the $p$’s in a GiG model typically has very large
consequences for the fit of the proposed Norris model. However, since it is necessary that the transition be reversible and available in closed form (for computational reasons) we find this transition kernel to be satisfactory.

3.4 The Likelihood Function

Because the number $J$ of pulses is uncertain, the space $\Theta$ of possible values of the parameter $\theta$ in this trans-dimensional model takes the form of a disjoint union:

$$\theta = ((A_1, \omega_1), \ldots, (A_J, \omega_J)) \in \Theta := \bigcup_{j=0}^{\infty} (\mathbb{R}_+ \times \Omega)^J$$

of $J$-fold Cartesian powers of the product of $\mathbb{R}_+$ (for the amplitude $A_j$ of the $j$th pulse) and the space

$$\Omega = \{ \omega = (T, p, c, \tau_1, \tau_2) \} = \mathbb{R}^2 \times \mathbb{R}_+^2$$

of possible values of the peak time $T$, the powers $p$ and $c$, and the time constants $\tau_1, \tau_2$ for the $j$th pulse (see Eqn (3.3)). The likelihood at $\theta \in \Theta$ for this Cox model upon observing photon counts $N_i$ for interval $(t_{i-1}, t_i]$, $1 \leq i \leq I$, is the product

$$\mathcal{L}(\theta) = \left\{ \prod_{i \leq I} \lambda_i^{N_i} \right\} \exp \left( - \sum_{i \leq I} \lambda_i \right)$$

of independent Poisson likelihoods with $\theta$-dependent means

$$\lambda_i = \int_{(t_{i-1}, t_i]} f(s \mid \theta) \, ds, \quad f(t \mid \theta) = B + \sum_{j=1}^{J} A_j K(t \mid \omega_j).$$

(3.6)

where $K$ is the most general kernel model. To complete the model specification, we must specify a prior distribution on $\Theta$, to induce a prior distribution on $f(t)$ and hence $\{\lambda_i\}$. 

43
3.5 The Prior Distribution

Following Wolpert et al. [51] we model $J$ as the number of jumps larger than some fixed threshold $\epsilon > 0$ for a stationary Gamma process $Z(t) \sim \mathcal{GP}(\alpha \, dt, \beta)$ on a time interval $[t_{\text{min}}, t_{\text{max}}]$ of finite length $L = (t_{\text{max}} - t_{\text{min}})$, and model the peak times and amplitudes $\{(T_j, A_j)\}_{1 \leq j \leq J}$ as the times and magnitudes of those jumps. Wolpert et al. show that $Z(t) \sim \mathcal{GP}(\alpha \, dt, \beta)$ can be represented as the stochastic integral

$$Z(t) = \int \int_{\mathbb{R}_+ \times (0,t]} u \mathcal{N}_0(du \, ds) \tag{3.7}$$

of $g(u, s) = u$ with respect to a Poisson random field $\mathcal{N}_0 \sim \text{Po}(\nu_0(du) \, ds)$ on $\mathbb{R}_+ \times \mathbb{R}$ with intensity measure $\nu_0(du) \, ds = \alpha u^{-1} \exp\{-\beta u\} \mathbf{1}_{u > 0} \, du \, ds$. More generally, they show that any probability density function $\pi_\tau(p, \tau_1, \tau_2) = \pi_\tau(\tau)$ on $\mathbb{R} \times \mathbb{R}^2_+$ induces a Lévy measure

$$\nu(du \, d\omega) = \alpha u^{-1} \exp\{-\beta u\} \mathbf{1}_{u > 0} \pi_\tau(\tau) \, du \, d\omega$$

and a Poisson random measure $\mathcal{N}(du \, d\omega) \sim \text{Po}(\nu(du \, d\omega))$ on $\mathbb{R}_+ \times \Omega$ for which

$$Z(t) = \int \int_{\mathbb{R}_+ \times \Omega} u \mathbf{1}_{\{0 < T \leq t\}} \mathcal{N}(du \, d\omega)$$

is the Gamma process $\mathcal{GP}(\alpha \, dt, \beta)$ from (3.7). The random element $f(t \mid \theta)$ from (3.6) can now be written in the form

$$f(t) = B + \sum_{j=1}^{J} A_j \, K_G(t \mid \omega_j)$$

$$= B + \int u \mathbf{1}_{\{\epsilon < u, T \in [t_{\text{min}}, t_{\text{max}}]\}} \, K_G(t \mid \omega) \, \mathcal{N}(du \, d\omega)$$

$$= B + \int u \, K_G(t \mid \omega) \, \mathcal{N}_\epsilon(du \, d\omega)$$

for $\mathcal{N}_\epsilon(du \, d\omega) = \mathbf{1}_{\{\epsilon < u, T \in [t_{\text{min}}, t_{\text{max}}]\}} \mathcal{N}(du \, d\omega) \sim \text{Po}(\nu_\epsilon(du \, d\omega))$, with finite truncated Lévy measure $\nu_\epsilon(du \, d\omega) = \mathbf{1}_{\{\epsilon < u, T \in [t_{\text{min}}, t_{\text{max}}]\}} \nu(du \, d\omega)$. This leads to a Poisson prior
distribution for the number \( J = \mathcal{N}((\epsilon, \infty) \times [t_{\text{min}}, t_{\text{max}}]) \) of pulses with amplitude above \( \epsilon \), with mean \( \mathbb{E}[J] = \mu_J := \alpha \text{LE}_1(\beta \epsilon) \) where \( \text{LE}_1(z) \equiv \int_{z}^{\infty} u^{-1} e^{-u} \, du \) is the exponential integral function \([1, p. 228]\). Conditionally on \( J \), the pulse peak locations \( \{T_j\} \) are iid uniformly distributed on the interval \([t_{\text{min}}, t_{\text{max}}]\), and the amplitudes \( \{A_j\} \) are iid with Lebesgue density

\[
\pi_A(A) := A^{-1} \exp\{-\beta A\} 1_{\{A > \epsilon\}} / \text{LE}_1(\beta \epsilon),
\]

the gamma Lévy measure \( \nu_0(du) \) normalized to be a probability measure on the interval \([\epsilon, \infty)\). Because the distribution of the approximating process

\[
Z^\epsilon(t) = \int \int 1_{\{u > \epsilon, 0 < T \leq t\}} u \mathcal{N}(du \, d\theta) = \sum_{j \leq J} A_j 1_{\{T_j \leq t\}}
\]

converges to that of \( Z(t) \sim \mathcal{GP}(\alpha \, dt, \beta) \) as \( \epsilon \to 0 \), the results about \( f(t \mid \theta) \) are insensitive to the choice of the threshold \( \epsilon \), so long as it is well below the smallest pulses regarded (by astronomers) as significant. Pulses of negligible magnitude are treated as background noise. The value we selected for \( \epsilon \) represents a peak with a maximum amplitude of 250 photons per second, far below the baseline rate of thousands of counts per second. The Gamma Process shape parameter \( \alpha \) and rate parameter \( \beta \) were selected on the basis of elicited estimates of the heights of the largest and second largest pulses in the gamma ray burst, using the inverse Lévy measure algorithm for drawing a gamma process of Wolpert and Ickstadt \([50]\), who show that the ordered amplitudes \( A_{[1]} > A_{[2]} > A_{[3]} > \cdots \) are distributed jointly as

\[
A_{[n]} = \text{LE}_1^{-1}(\tau_n / \alpha \lambda) / \beta
\]

for the event times \( \{\tau_n\} \) of a standard unit-rate Poisson process. It follows that the median of \( A_{[1]} \) and conditional median of \( A_{[2]} \) given \( A_{[1]} \) satisfy

\[
\frac{1}{2} = P[A_{[1]} > m_1], \quad \alpha \text{LE}_1(\beta m_1) = \log 2
\]

\[
\frac{1}{2} = P[A_{[2]} > m_2 \mid A_{[1]}], \quad \alpha \text{LE}_1(\beta m_2) = \log 2 + \alpha \text{LE}_1(\beta A_{[1]})
\]
From these one can infer elicited values of $\alpha$ from elicited values of the ratio $R = A_{[1]} / A_{[2]}$ at the medians, and then $\beta$ from elicited values of $A_{[1]}$ (or of the total number of photons associated with the largest pulse or of the total number of photons in the entire GRB). Our elicited choices of $R = 2$ and $L = 30\, s$, with a largest expected pulse of 12,000 photons per second, lead to $\alpha = 0.0742$ and $\beta = 6.67 \times 10^{-5}$.

For the $\tau_1$ and $\tau_2$ parameters, we used information elicited from our collaborators. Discussion with astronomers indicated two priorities in the distributions for these parameters.

1. For a given pulse, $\tau_1$ and $\tau_2$ were positively correlated.
2. The marginal distributions for $\tau_1$ and $\tau_2$ are heavy-tailed.

This suggested a hierarchical prior for these parameters. Within a GRB, there is a parameter

$$\lambda \sim \text{Ga}(a, b)$$

and

$$\tau_{i,j} \sim \text{Exp}(\lambda) \quad \text{for } i = 1, 2, \ j = 1, \ldots, J$$

Then, the distributions of the $\tau_1$ and $\tau_2$ parameters can be found as follows:

$$P(\tau_i > t) = \int_0^\infty \exp\{-\lambda t\} \lambda^{a-1} b^a \exp\{-b \lambda\} \frac{1}{\Gamma(a)} d\lambda$$

$$= \frac{b^a}{\Gamma(a)} \int_0^\infty \lambda^{a-1} \exp\{-\lambda (b + t)\} d\lambda$$

$$= \frac{b^a}{\Gamma(a)} (b + t)^{-a} \Gamma(a)$$

$$= (1 + t/b)^{-a},$$

a generalized Pareto distribution with shape parameter $a$ and scale parameter $b$. We took $a = 5$ and $b = 1$ to reflect the views of our domain scientist collaborators.
A normal prior distribution was used for the shape parameter $p$, centered at the Norris kernel value $p = 1$, with standard deviation $\xi_r = 10$.

In summary, the prior specification on $\Theta = \cup_{J<\infty}(\mathbb{R}_+ \times \Omega)^J$ is $J \sim \text{Po}(\mu_J)$ with mean $\mu_J = \alpha \text{LE}_1(\beta \epsilon)$ and, conditional on $J$, each of the $J$ components $A_j$ and $\omega_j = (T, p, \tau_1, \tau_2)$ is drawn independently from distributions $A_j \sim \pi_A(A_j) dA_j$ and $\{\omega_j\} \sim \pi_\omega(\omega) d\omega$ with density functions

$$\pi_A(A) = A^{-1} \exp\{-\beta A\} 1_{\{A>\epsilon\}}/E_1(\beta \epsilon)$$

(3.8)

for the amplitude $A_j$ and $\pi_\omega(\omega)$ determined by

$$T \sim \text{Un}([t_{\text{min}}, t_{\text{max}}])$$

(3.9a)

$$p \sim \text{No}(1, \xi_r^2)$$

(3.9b)

$$\tau_1, \tau_2 \overset{iid}{\sim} \text{Ex}(\lambda_r)$$

(3.9c)

$$\lambda_r \sim \text{Ga}(a, b)$$

(3.9d)

for $\omega = (T, p, \tau_1, \tau_2)$. Thus the prior distribution has density function with respect to the sum of $(5J)$-dimensional Lebesgue measure on each disjoint component of $\Theta = \cup_{J}(\mathbb{R} \times \Omega)^J$ given by

$$\Pi(\theta) = \frac{(\mu_J)^J}{J!} e^{-\mu_J} \prod_{j=1}^J \left[\pi_A(A_j) \pi_\omega(\omega_j)\right] = \frac{e^{-\mu_J}}{J!} \prod_{j=1}^J \nu_\epsilon(u_j, \omega_j),$$

(3.10)

where $\nu_\epsilon(u, \omega) = 1_{\{u>\epsilon, T \in [t_{\text{min}}, t_{\text{max}}]\}} \nu(u, \omega)$ is the density of the finite Lévy measure $\nu_\epsilon(du, d\omega)$.

### 3.6 Posterior Computation

Draws from the approximate posterior distribution can be found using reversible jump Markov chain Monte Carlo [18]. Recall that reversible jump is an algorithm in which at least three types of Metropolis-Hastings steps are performed: birth (in
which the addition of a new pulse is proposed), death (or its removal), walk (in which a change is proposed in the shape or location of an existing pulse), merge (combining two pulses into a single composite), and split (of a single pulse into two).

Proposals are accepted or rejected according to the Metropolis-Hastings algorithm \[24, 48\]. As in many complex problems, the posterior distribution is highly multimodal, and the reversible jump MCMC algorithm sketched above mixes very slowly among the modes \[14\]. The birth, death, and walk steps are those commonly used in reversible jump MCMC and are discussed in Appendix 1. In this section we describe our approach to the split and merge steps. We first fix five nonnegative numbers \( p_B, p_D, p_W, p_M, \) and \( p_S \) summing to one for the probabilities with which we propose each of the five move types (birth, death, walk, merge).

**Merge:**

The merge and split steps, taken with probabilities \( p_M \) and \( p_S \), are somewhat specific to this problem. A merge step replaces two pulses \( \omega' = (A', T', p', \tau_1', \tau_2') \) with \( \omega'' = (A'', T'', p'', \tau_1'', \tau_2'') \) with a single pulse \( \omega^* = (A^*, T^*, p^*, \tau_1^*, \tau_2^*) \), with the same total height and with weighted averages for the other parameters:

\[
A^* = A' + A'' \tag{3.12a}
\]
\[
T^* = \frac{A'}{A^*}T' + \frac{A''}{A^*}T'' \log \tau_1^* = \frac{A'}{A^*} \log \tau_1' + \frac{A''}{A^*} \log \tau_1'' \tag{3.12b}
\]
\[
p^* = \frac{A'}{A^*}p' + \frac{A''}{A^*}p'' \log \tau_2^* = \frac{A'}{A^*} \log \tau_2' + \frac{A''}{A^*} \log \tau_2'' \tag{3.12c}
\]

For possible merge steps we select a pair of pulses uniformly from the \( \binom{J}{2} \) possible pairs and replace both of them with a single merged pulse constructed above, placed in one of the \( J - 1 \) possible remaining order locations among the parameter vector
(as in the birth step), leading to transition probability

\[ q_{M,J \to J-1} = p_M \frac{2}{J(J-1)} \times \frac{1}{J-1} \]

It may initially seem more efficient to propose merging only adjacent pairs of pulses, rather than all \( \binom{J}{2} \) possible pairs. However, to achieve reversibility of the Markov chain, we would then have to constrain the splitting to produce only adjacent pulses. This would lead to a Hastings ratio that depends on the entire list of pulses instead of just the pair whose merger is proposed.

**Split:**

Note that the rule for merging two pulses is completely deterministic (once the pair and the location for the merged pulse is chosen). In splitting, however, we are moving from a single pulse (determined by five parameters) to two (each determined by five parameters, for a total of ten).

A split step replaces the five-dimensional pulse \( \omega^* = (A^*, T^*, p^*, \tau_1^*, \tau_2^*) \) with a ten-dimensional pair of pulses \( \omega' \) and \( \omega'' \) (see (3.11)). In order to do this while maintaining reversibility of our Metropolis-Hastings algorithm, we introduce five contrasts

\[ u = \frac{A'}{A' + A''} \sim \text{Be}(\zeta, \zeta) \quad (3.13a) \]
\[ \delta_0 = p' - p'' \sim \text{No}(0, \xi_0^2) \quad (3.13b) \]
\[ \delta_1 = \log \tau_1' - \log \tau_1'' \sim \text{No}(0, \xi_1^2) \quad (3.13c) \]
\[ \delta_2 = \log \tau_2' - \log \tau_2'' \sim \text{No}(0, \xi_2^2) \quad (3.13d) \]
\[ \delta_3 = T' - T'' \sim \text{No}(0, \xi_3^2) \quad (3.13e) \]

Construct \( \omega' \) and \( \omega'' \) from \( \omega^* \) by drawing \( u \) and \( \{\delta_k\} \) independently from the indicated
distributions and setting:

\[ A' = uA^* \quad A'' = (1-u)A^* \]  \hspace{1cm} (3.14a)

\[ p' = p^* + (1-u)\delta_0 \quad p'' = p^* - u\delta_0 \]  \hspace{1cm} (3.14b)

\[ \log \tau_1' = \log \tau_1^* + (1-u)\delta_1 \quad \log \tau_1'' = \log \tau_1^* - u\delta_1 \]  \hspace{1cm} (3.14c)

\[ \log \tau_2' = \log \tau_2^* + (1-u)\delta_2 \quad \log \tau_2'' = \log \tau_2^* - u\delta_2 \]  \hspace{1cm} (3.14d)

\[ T' = T^* + (1-u)\delta_3 \quad T'' = T^* - u\delta_3 \]  \hspace{1cm} (3.14e)

The merge step of (3.12) exactly reverses the split step of (3.14), and the split step of (3.14) can reach any possible pulse pair \((A', \omega'), (A'', \omega'')\) whose merger is \((A^*, \omega^*)\), as required for reversibility.

The Jacobian of the transformation (3.14) from \((A^*, \omega^*), (u, \delta)\) to \((A', \omega'), (A'', \omega'')\) is:

\[ \mathcal{J} = \frac{\partial(A', \omega', A'', \omega'')}{\partial(A^*, \omega^*, u, \delta)} = \begin{pmatrix}
 u & 0 & 0 & 0 & 0 & A^* & 0 & 0 & 0 & 0 \\
 0 & 1 & 0 & 0 & 0 & -\delta_0 & 1 - u & 0 & 0 & 0 \\
 0 & 0 & \tau_1'/\tau_1^* & 0 & 0 & -\delta_1 \tau_1' & 0 & (1-u)\tau_1' & 0 & 0 \\
 0 & 0 & 0 & \tau_2'/\tau_2^* & 0 & -\delta_2 \tau_2' & 0 & 0 & (1-u)\tau_2' & 0 \\
 0 & 0 & 0 & 0 & 1 & -\delta_3 & 0 & 0 & 0 & 1 - u \\
 1 - u & 0 & 0 & 0 & 0 & -A^* & 0 & 0 & 0 & 0 \\
 0 & 1 & 0 & 0 & 0 & -\delta_0 & -u & 0 & 0 & 0 \\
 0 & 0 & \tau_1''/\tau_1^* & 0 & 0 & -\delta_1 \tau_1'' & 0 & -u\tau_1'' & 0 & 0 \\
 0 & 0 & 0 & \tau_2''/\tau_2^* & 0 & -\delta_2 \tau_2'' & 0 & 0 & -u\tau_2'' & 0 \\
 0 & 0 & 0 & 0 & 0 & 1 & -\delta_3 & 0 & 0 & -u
\end{pmatrix} \]

with determinant \(|\mathcal{J}| = -A^*\tau_1'\tau_2'/\tau_1^*\tau_2^*\tau_1''\tau_2''/\tau_1^*\tau_2^*\tau_1''\tau_2''\), and so the transition probability density for particular split move, in which one of the \(J\) pulses \(\omega^*\) is picked uniformly and then split into two \((\omega', \omega'')\) which are then placed in specific order locations chosen uniformly from the \(_{J+1}^2\) possibilities, can be written:

50
\[ q_{S,J \rightarrow J+1} = p_S \frac{1}{J} \times q(u \mid \zeta) \times \prod_{k=0}^{3} q(\delta_k \mid \xi_k) \times \frac{\tau_1^* \tau_2^*}{A^* \tau_1^* \tau_2^* \tau_1'' \tau_2''} \times \frac{2}{J(J+1)} \]

where (in a minor abuse of notation) “\(q(\cdot \mid \cdot)\)” denotes each contrast’s marginal density function from (3.13).

The hyperparameters used were \(\zeta = 2\) and \((\xi_0, \xi_1, \xi_2, \xi_3) = (0.1, 1.0, 1.0, 0.5)\). These were chosen to tune acceptance rates, although the merge and split moves are more difficult to accept than the other types under any circumstance. In practice, a pulse was more likely to be “split” through mixing with more-thinned chains through parallel thinning than through accepting a reversible jump split move.

**Hastings Ratios:**

From these descriptions of the transition densities and from the prior density function given in (3.10) we can construct the Hastings Ratio, as in Section (2.2.1), for any of the five step types taking a current configuration \(\theta = \theta^t\) to a proposed one \(\theta^*\):

\[ H^t(\theta^* \mid \theta) = \frac{\mathcal{L}(\theta^*)}{\mathcal{L}(\theta)} \times \frac{\Pi(\theta^*) \cdot q_{\theta^* \rightarrow \theta}}{\Pi(\theta) \cdot q_{\theta \rightarrow \theta^*}} \]

For most of the moves there is considerable cancellation, leading to the following expressions for \(H^t\), beginning with \(\theta^t \in (\mathbb{R} \times \Omega)^J\):

- **birth:** \(H^t = \frac{\mathcal{L}(\theta^*)}{\mathcal{L}(\theta^t)} \times \frac{p_D \nu_e(A^*, \omega^*)}{p_B (J+1) b(A^*, \omega^*)} \)
- **death:** \(H^t = \frac{\mathcal{L}(\theta^*)}{\mathcal{L}(\theta^t)} \times \frac{p_B (J) b(A^*, \omega^*)}{p_D \nu_e(A^*, \omega^*)} \)
- **walk:** \(H^t = \frac{\mathcal{L}(\theta^*)}{\mathcal{L}(\theta^t)} \times \frac{\tau_1^* \tau_2^* \nu_e(A^*, \omega^*)}{\tau_1 \tau_2 \nu_e(A, \omega)} \)
- **merge:** \(H^t = \frac{\mathcal{L}(\theta^*)}{\mathcal{L}(\theta^t)} \times \frac{p_M A^* \tau_1^* \tau_2^* \tau_1'' \tau_2'' \nu_e(A^*, A') \nu_e(A^*, \omega^*)}{p_M A^* \tau_1^* \tau_2^* \tau_1'' \tau_2'' \nu_e(A', \omega')} \)
- **split:** \(H^t = \frac{\mathcal{L}(\theta^*)}{\mathcal{L}(\theta^t)} \times \frac{p_M A^* \tau_1^* \tau_2^* \tau_1'' \tau_2'' \nu_e(A', \omega') \nu_e(A^*, \omega^*)}{p_S (J+1) \tau_1^* \tau_2^* q(u \mid \zeta) \prod_{k=0}^{3} q(\delta_k \mid \xi_k) \nu_e(A^*, \omega^*)} \)
3.7 Analysis of Selected GRBs

To demonstrate the flexibility of the LARK model, we analyze several GRBs, each with a distinct shape described below. In all cases, the data in the first energy channel (from 20 to 50 kEV) was used. Of the four energy channels measured, the first channel is the one which is most likely to show evidence of a GRB. Higher energy channels often contain only the background spectrum. Plots of the data used in this analysis can be seen in Figure (3.3).

GRBs can be identified by their trigger number, or by the date on which they were observed [45]. We use the trigger numbers assigned by the BATSE instrument.

The GRB with trigger number 3040 (Figure (3.3)(a)) is the most straightforward GRB included in this analysis. It consists of a single large feature which
begins around the trigger time with the expected “fast rise” and has an exponential-
appearing decay. The entire burst lasts around 5 seconds, a typical length.

Trigger 2571 (Figure (3.3)(b)) appears as two well-separated features. These fea-
tures are on roughly the same scale, and each feature is believed to be a single pulse. However, due to the sharpness of the pulse shape, a Norris kernel may not suitably fit these pulses.

Trigger 1406 (Figure (3.3)(c)) appears as a single broad feature, and is appears from inspection to consist of a single pulse. However, the simpler models, such as the Norris kernel, may not fit such a feature with a single pulse. Model selection for Trigger 1406 is particularly important, as it demonstrates the importance of flexible kernel models.

Trigger 105 (Figure (3.3)(d)) is a complicated and bright burst, with several over-
lapping features. A single feature in the data may often be expressed either in the form of one large pulse, or as the sum of several smaller pulses. The bright-
ness of this burst, with a background photon count rate of approximately 36,000 photons/second and a peak flux of 12,000 photons/second, results in a likelihood function that is particularly peaked and makes mixing between these configurations particularly difficult.

We run separate analyses of these four GRBs. Each MCMC had 200,000 itera-
tions, with a burn-in period of 50,000 iterations. There were five auxiliary chains, with thinning parameters \( p = 0.15, 0.3, 0.45, 0.6, \) and 0.85.

3.7.1 **GRB 3040**

In a limited sense, we can assess the convergence of the Markov chain by looking at the implied estimates of the posterior mean for the time series as a function of MCMC iteration. We would expect that as the chain converges, the estimated posterior means would tend to fall into a plausible region of the parameter space.
across the time points. Evidence to the contrary may suggest a need for a longer run. For each of 100 equally-spaced iterations of the MCMC, after burnin, we create an implied mean which is then plotted over the data. Early iterations are colored in red, with later iterations in blue shades. We can see that all of the iterations are in a plausible region of the parameter space. This does not suggest a need for further MCMC iterations.

Figure (3.5) shows a few posterior draws for GRB 3040. The dot heights each represent a photon bin count divided by the bin length (in seconds), so the $y$-axis units are photons per second. This is for visualization purposes only, as inference is done directly on the bin counts. This is because the widths of the bins vary (earlier bins are wider, as the instrument does not store high-resolution data for time intervals more than two seconds before a trigger) making visualizing the bin counts directly difficult. The colored lines in each sub-figure are the individual pulses in a particular RJ-MCMC draw, and the thick black line is the sum of all these pulses. Here we can see that a two-pulse solution fits very well, which can be confirmed by examining the posterior histogram for the number of pulses. This is shown in Figure (3.6)
Figure 3.5: Posterior samples for the mean for GRB 3040.

Figure 3.6: Posterior histogram for the number of pulses comprising GRB 3040.
This figure shows that \textit{a posteriori}, a solution with two pulses has about probability 0.85, while most of the remaining posterior mass is on a 3-pulse solution.

Interestingly, the posterior distribution for the \textit{model} – that is, which functional form is used for the kernel in the LARK algorithm – shows a preference for the most flexible kernel (pGiG). In Figure (3.7), we see that the largest model is favored roughly seven to one over the others. One might think this model could yield a one-pulse solution, since it has a more flexible form. However, there is also significant mass in the other two models, perhaps suggesting that model selection does not affect the number of pulses needed to represent this burst.

\subsection*{3.7.2 GRB 2571}

We again utilize implied estimates of posterior mean to examine convergence of the Markov chain. From Figure (3.8), it seems the posterior means of the time series have converged. The sample draws in Figure (3.9) indicate that the feature between 0 and 5 seconds is often represented by several smaller pulses instead of one large one, such as in iteration 3000 and 10500. However, in some iterations, such as iteration 7500 and 12000, this feature is represented by a single pulse. This reflects our posterior uncertainty about the nature of this feature. One feature that is visible in all sample draws represented here is the long, soft pulse that is colored in red in all panels. This pulse does not have the FRED shape that is expected of GRB pulses, but such a
pulse appears in several analyses we have done of different GRBs. This may be a reflection of changing background rates over the duration of the pulse, or it may be an indication that the kernels do not allow for flexible enough shapes to accurately represent the gamma-ray bursts.

The posterior mode for the number of pulses (Figure (3.10)) is four, although there is much more uncertainty in this GRB than in the simpler GRB 3040 (Figure (3.6)). The plausible range for the number of pulses ranges from three until seven.

Figure (3.11) shows the model choice posterior histogram and traceplot. It seems that mixing is good, but there is not much preference for the more complicated models. It seems that using more Norris kernels is preferable to a smaller solution that uses the more flexible GiG or pGiG kernels; since the Norris kernel captures roughly 65% of the posterior mass.

3.7.3 GRB 1406

Figure (3.12) and Figure (3.13) show example fits for GRB 1406. They indicate that the MCMC chain has converged to a very definite solution. As is strongly indicated in Figure (3.14), there is only one pulse in GRB 1406. The mixing of this parameter can be seen in the right panel of Figure (3.14).
Figure 3.9: Posterior samples for the mean for GRB 2571.

Figure 3.10: Posterior histogram for the number of pulses comprising GRB 2571.
The model selection part of the analysis, summarized in Figure (3.15), indicates that there is not a strong preference between the models. Given that the prior probabilities for the three model classes were equal, and Figure (3.15) shows their posterior probabilities ranging from 25% to 40%, it seems that the data gives no information that would separate these three models from one another.

3.7.4 GRB 105

This burst has the highest baseline rate of those examined here, and it also has a very high signal-to-noise ratio. This leads to challenges in mixing as the likelihood distribution is very peaked. Additionally, the complicated structure leads to many
**Figure 3.13:** Posterior samples for the mean for GRB 1406.

**Figure 3.14:** Posterior histogram and traceplot for the number of pulses comprising GRB 1406.
possible modes which must be mixed between.

It is in complicated GRBs such as 105 that posterior summary becomes a particular challenge. Since this GRB has the most complicated pattern of overlapping features, it is unsurprising that the number of pulses chosen in varies between six and ten. Interesting, there is a strong preference in the data for the simpler kernels, Norris and GiG. As shown in Figure (3.19), there is almost no posterior probability of the most complicated model, pGiG.
Figure 3.17: Posterior samples for the mean for GRB 105.

Figure 3.18: Posterior histogram and traceplot for the number of pulses comprising GRB 105.
3.8 Discussion

It seems that model selection among the three pulse shapes (kernels) was not a driving factor in any of the GRBs. No model was strongly preferred even in a single GRB, and that there was no pattern across the small number of bursts analyzed here. This is good for future work, since none of the models are based in theory; in effect, all of the models are known to be misspecified. However, increasing the complexity of the model (and the dimension of the computational problem), does not seem to improve fit or understanding.

One potential avenue to further explore in model selection is to examine a wider class of models. The Norris kernel is not the only option, and many functions have been used that do not take the form of the Norris or its expansions. The difficulty with this approach is in formulating intelligent and useful transition functions between parameterizations when the models are no longer nested; even in the straightforward discussed here, there is often difficulty accepting model-switching moves in the Metropolis-Hastings algorithm. Additionally, it will become increasingly difficult to interpret any of the pulse shape parameters when the models under comparison become increasingly different.

Another approach, which sidesteps the need to pick “the best” kernel – if one such
exists – or average over a large class of kernels which have very different parameter interpretations, is to introduce overdispersion to allow some more flexibility for model misspecification, instead of using a more complicated model. For example, instead of a Poisson likelihood, an overdispersed negative binomial model can be used for the photon counts. This will have the added flexibility from the additional variance-related parameter while maintaining the same kernel regression structure for the mean.

As can be seen in Section 3.7, there is a lot of variety in the shapes and number of pulses in a gamma-ray burst. Some, such as GRB 1406, are unambiguously comprised of a single pulse; others, such as GRB 105 and 2571, have a complicated structure which has in previous work been difficult to deconvolve. Additionally, several GRBs have a long, low pulse which does not fit the expected shape of the pulses of a GRB. Whether or not this pulse represents a physical reality or is perhaps a model specification issue is something to be determined in future work, which will be in the time \times energy domain.

The LARK model is able to flexibly fit all these GRBs and estimate the number of pulses comprising the burst in a fully Bayesian way. Along with the brief model comparison done here, which indicates that model choice does not drive inference, we are better prepared to move onto the more complicated domain spaces of the multi-channel GRB data.
Gamma-ray bursts are comprised of overlapping pulses of photons that have both an energetic and a temporal spectrum. These spectra are not independent, since higher-energy photons are often believed to arrive earlier than lower-energy photons. We use LARK models to present the first joint model for the temporo-energetic spectrum which incorporates both uncertainty in the number of pulses and in the interaction between the time and energy spectra. In addition, this flexible semiparametric Bayesian framework allows us to model GRB light curves (as the time spectra are called) in the space before they are transformed by the BATSE instrument’s detectors [6]; that is, we model the incident spectra instead of the observed spectra. By deconvolving the pulses in the pre-transformed space, the resulting inference on the pulses’ parameters are more informative of the generative process of the GRB.

4.1 Introduction to the Inverse Problem

The energies of the photons of a gamma-ray burst are recorded after they have been transformed by the telescope that used to observe them. There are five things that
can occur to a photon after it enters the detector. These were first described in Berger and Seltzer [7] and are described well in Appendix A of Loredo and Epstein [37]. The photon can deposit all of its energy into the detector, resulting in an accurate measurement of photon energy. A process called pair production can occur, in which the photon interacts with another particle and emits an electron and a positron; one or both of these created particles can escape the detector, resulting in a lower recorded energy. K-shell fluorescence, in which energy of the incident photon is used to displace an inner photon of the response medium, and Compton scattering, in which the photon transfers some of its energy to another particle, also result in altered recordings of the incident photon’s energy.

In summary, although some photons are recorded by the detector at their true incident energy, many are recorded indirectly at lower energy. There is a systematic bias in the telescope which marks photons as having a lower energy than they actually have. Since the pulses are believed to arrive earlier at higher energies, this systematic bias smears the pulse shapes in the data we observe.

Previous analyses of GRBs have focused on modeling either the time domain or the energy domain. There has not been much progress in the past modeling them together, due to the complicated relationship between the two. This relationship is exacerbated by the transformation performed by the telescope’s detectors.

The relationship between the time and energy domains, called the “response function” of the telescope, is instrument- and event-specific and is typically well-understood due to tests done on the telescope before it is sent into space.

The data we have is binned photon counts in a variety of energy bands of the observed photons.

We can model the incident gamma-ray photon intensity as a function of both time and energy. Since the response function is available, we can then transform this incident flux using the response function. This gives the observed flux as a function
of both energy and time. We can then sum over observed energies to get the model’s estimate of the observed photon count time series in each energy channel.

4.2 Method

The data are binned photon counts $Y_{ik}$, for $i = 1, 2, 3, 4$, corresponding to the four energy channels used by the discriminator, and $k = 1 \ldots N$, for the $N$ time intervals. It is believed that the data have a Poisson distribution, since the discriminators have a very small amount of downtime and the photons arrive in a Poisson process [47].

The underlying prior model for the amplitudes and incident time can be represented as the jumps exceeding some threshold of a fully-skewed $\alpha$-stable process with parameters $\alpha = 3/2, \beta = 1, \eta = 0.107$. This results in a distribution over the number of non-trivial pulses in a time interval as something that can be inferred from the data.

The number of pulses with an amplitude larger than some threshold $\epsilon$, $J_\epsilon$, are distributed Poisson a priori. We will abbreviate $J_\epsilon$ as $J$, since $\epsilon$ is typically chosen before any analysis is done; it is set such that any pulse of amplitude $\epsilon$ is regarded by astronomers as negligible in size. Because each pulse represents a single component of the GRB, there is a separate set of parameters for each pulse describing the shape, magnitude, onset time, and energy spectrum of the pulse. However, the subscripts indicating pulses have been omitted for this section.

Each pulse has an onset time, $T$, when the 100keV photons arrive at the detector (arrival time is a function of energy). We model the onset of a burst at a given energy as an instantaneous jump. Since a pulse does not appear simultaneously at all energies, a linear lag (in log energy) parameter $\gamma$ is introduced to shift the pulse onset at higher energies. When translated through the discriminators, this gives rise to a pulse which has the fast-(but not instant-) rise characteristic of a gamma-ray burst. This can be seen in Figure (4.1), which shows the incident spectra arriving
instantaneously at various energies spaced approximately on a log scale; due to the hard-to-soft evolution of the simulated burst in the figure as well as the effect of the discriminator, the observed photon counts have a smooth rise. At a given energy, as time increases past the onset of the burst, the flux falls exponentially in time with rate $\lambda$. The energy spectrum is modeled independently for each pulse.

4.2.1 Lag

The onset time of a pulse varies as a function of energy. The descriptor “hard to soft” when applied to a gamma-ray burst indicates that higher-energy photons arrive earlier than lower-energy photons in that burst. This “lag,” which can be positive or negative, must be accounted for in a model which models time and energy domains jointly. Additionally, in our model, which at a given energy level includes an instantaneous turn-on of the GRB, the lag parameter is wholly responsible for creating the smooth rise of the GRB light curve; this effect is achieved as photons arriving at different times at different energies are binned together by the discriminator.
Lee et al. [35] performed a survey of the BATSE GRBs in an effort to infer these lag parameters for a large number of GRBs. In the majority of the bursts, higher energy photons arrive at the telescope before the lower energy photons; however in about one-third of the bursts, the pulse arrived earlier in the lower energy channels than in the higher.

Oftentimes there is little spectral information available about the onset time of the GRB, due to the weakness of the signal [2]. However, for the first pulse of GRB130427A, one of the ten most energetic GRBs ever recorded, and the most energetic observed on the Fermi instrument to date, there is ample spectral data available in the time surrounding the onset of the burst [46]. Preece et al. [46] model the time lag between two distinct observed energies $E_1$ and $E_2$ as the difference of two power laws: $\tau \propto E_1^\alpha - E_2^\alpha$, where $\alpha$ is the power law index of the energy spectrum during the rise time of the pulse.

In general, past investigations into this phenomenon have focused on the time spectrum resolved into several energy bands in the observed space; that is, after the photons’ energy has been transformed by the response function of the detector [35, 46]. This means that the results found in earlier work, while suggestive, are not directly informative about the time lag in the incoming space.

To account for the above, as well as for simplicity of evaluation, we model the lag of the onset time of a pulse (as compared to its onset at 100 keV) as a linear function of the natural logarithm of the energy. This allows for behavior similar to that of the power law distribution observed in Preece et al. [46].

Thus, we can write the time component of the flux as follows:

$$\Psi(t \mid E, T, \gamma, \lambda) = \begin{cases} \exp\{-\lambda(t - T + \gamma \ln(E/100))\} & \text{if } t \geq T - \gamma \ln(E/100) \\ 0 & \text{if } t < T - \gamma \ln(E/100) \end{cases}$$

(4.1)
In addition to the time component, the energy spectrum of a pulse also varies. A common functional form for the energy spectrum of a gamma-ray burst is the Band model \cite{3}. The key feature of this model is that the photon rate is modeled by a steep power law (rate $\alpha$) at energies lower than some cutoff ($E_c = (\alpha - \beta)E_0$), and at a less steep power law (rate $\beta$) at higher energies. That is, the energy spectrum is

$$\Phi_B(E \mid \alpha, \beta, E_0) \propto \begin{cases} \frac{E_c}{100}^{\alpha-\beta} \exp\{\beta - \alpha\} \left(\frac{E}{100}\right)^\beta & \text{if } E > E_c, \\ \frac{E}{100}^\alpha \exp\{-E/E_0\} & \text{if } E \leq E_c. \end{cases}$$

(4.2)

for specified rates $\alpha < \beta < 0$.

Another common but simpler variation on this model for the energy spectra of GRBs is the Comptonized model \cite{16}. Instead of two power law decays with an exponential cutoff, there is a single power law index $\alpha$. We call this energy spectrum $\Phi$:

$$\Phi(E \mid \alpha, E_c) = \left(\frac{E}{100}\right)^\alpha \exp\{-E/E_c\}$$

(4.3)

Goldstein et al. \cite{16}, a review of the energy spectra of GRBs using the BATSE discriminator instrument, does model selection for a variety of energy spectra models, including the Band and the Comptonized models discussed here. In over 40% of the GRBs in the BATSE catalog – a plurality – the Comptonized model was selected as the best-fitting model for the energy spectrum alone.

Since our current work involves the simultaneous analysis of both the time and energy spectra, we use the simpler Comptonized model for the energy spectrum, aware of the caveat mentioned in Goldstein et al. \cite{16} that the more complicated Band model may become necessary to fit brighter bursts.
4.2.2 Volume

For more straightforward posterior computation, we parameterize the size of the pulse in terms of photon count fluence; that is, the total number of photons in a pulse, per cm$^2$ of the detector. We call this parameter $V$ as it represents the volume of a pulse in the time and energy spaces.

We model pulses in both time and energy jointly, so the product of (4.3) and (4.1) must be integrated with respect to $E$. We must choose a lower limit for the energy integration, since $\Phi$ is not integrable from 0 to $\infty$ when $\alpha < -1$. This choice is not particularly important, since photons that arrive at such low energies (below 10 keV, say), are likely to go undetected in the instrument; this fact is already represented in the response function. We use 1 keV as the threshold here. For a given pulse, we calculate the normalizing constant:

$$K(\alpha, E_c, \lambda) = \int_1^\infty \int_{T-\gamma \log\{E/100\}}^\infty \left( \frac{E}{100} \right)^\alpha \exp\{-E/E_c\}$$

$$\times \exp\{-\lambda t + \lambda T - \lambda \gamma \log\{E/100\}\} dtdE$$

$$= \int_1^\infty \left( \frac{E}{100} \right)^\alpha \exp\{-E/E_c - \lambda \gamma \log\{E/100\} + \lambda T\}$$

$$\times \int_{T-\gamma \log\{E/100\}}^\infty \exp\{-\lambda t\} dtdE$$

$$= \int_1^\infty \left( \frac{E}{100} \right)^{\alpha - \gamma \lambda} \exp\{\lambda T - E/E_c\} \left( \frac{1}{\lambda} \right) \exp\{-\lambda(T - \gamma \log\{E/100\})\}$$

$$= \int_1^\infty \left( \frac{E}{100} \right)^\alpha$$

$$\left( \frac{1}{\lambda} \right) \exp\{-E/E_c\} dE$$

$$= \left( \frac{1}{\lambda} \right) E_c^{\alpha + 1} \left( \frac{1}{100} \right)^\alpha \Gamma(\alpha + 1, E_c^{-1})$$
where
\[
\Gamma(a, x) = \int_x^\infty y^{a-1} \exp\{-y\} dy
\]
is the upper incomplete Gamma function [1, §6.5.3].

We can then write the total flux from set of pulses described by \(\theta_1 \ldots \theta_J\) as a function of both time and energy:

\[
F(t, E \mid \{\theta_j\}) = B + \sum_{j=1}^{J} V_j \Psi(t \mid E, \theta_j) \times \Phi(E \mid \theta_j)/K(\theta_j) \quad (4.4)
\]
where \(B\) is some fixed baseline.

So, each pulse is described by either seven parameters, \(\theta = (T, V, \lambda, \gamma, \alpha, \beta, E_0)\) for the Band model; or six with \(\theta = (T, V, \lambda, \gamma, \alpha, E_c)\) for the Comptonized model.

A photon that enters the detector may not be recorded at its true energy, or even at all, due to one of a number of possible reactions with other photons or the detecting material [37]. For example, Compton scattering in the detector may result in the photon being recorded at a lower energy than that of the incident photon. In general, the error in energy estimation is biased; the energies estimated by the detector are systematically lower than the true energies. The “response function” of the detector is known due to experiments performed on the ground before the telescope was launched. There is a different response function for each energy channel and each GRB, since the functions also vary as a function of the incident angle of the photons (which is measured separately). We can then transform this incident flux using the response function for energy channel \(i\), called \(R_i(E)\). This gives the observed flux as a function of both energy and time. We can then integrate over the time bin and energy channel range to obtain the model estimate of the mean value for the observed photon counts.
Then

\[ Y_{ik} \sim \text{Po}(\mu_{ik}) \]

where

\[ \mu_{ik} = \int_{t_k}^{t_{k+1}} \int_{0}^{\infty} F(t, E)R_i(E) \, dE \, dt \]

4.2.3 Priors

Priors for all parameters can be seen in Table (4.1). Some of these will be discussed below, while others where chosen simply to have convenient form and to provide support for all plausible values.

A priori, the location of the pulse in energy is taken to be uniform across the domain of the response function. The same modeling assumption is made for the location of the pulse in time, even though we know that a pulse is more likely to have occurred near \( t = 0 \), the trigger time, than significantly earlier.

The volumes of the individual pulses larger than \( \epsilon \) are modeled by the largest jumps of a fully-skewed \( \alpha \)-stable process. The distribution of the pulses, which proportional to is the \( \epsilon \)-truncated Lévy distribution associated with the fully-skewed \( \alpha \)-stable distribution with \( \alpha = 3/2 \),

\[ f(v) = 3\frac{\eta}{\sqrt{8\pi}} V^{-5/2} 1_{\{V > \epsilon\}} \]  

We set \( \alpha = 3/2 \), to reflect the observation that photon fluence decays as a power law with index \(-5/2\) [41]. Using an \( \epsilon \) of 1 photon/cm\(^2\), we set \( \eta \) so that the expected number of pulses larger than \( \epsilon \) is three, which was given as a plausible prior by our astronomer collaborators. This results in a value of \( \eta = 0.1074 \).
Table 4.1: Priors for 1-pulse Simulation

<table>
<thead>
<tr>
<th>Interpretation</th>
<th>Parameter</th>
<th>Prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start time (s)</td>
<td>$T$</td>
<td>$\text{Unif}(-12, 25)$</td>
</tr>
<tr>
<td>Pulse volume (photons/cm$^2$)</td>
<td>$V$</td>
<td>$f(v) = 3^{0.1074} \frac{V^{-5/2}}{\sqrt{8\pi}} \mathbb{1}_{V&gt;\varepsilon}$</td>
</tr>
<tr>
<td>Linear time lag (s)</td>
<td>$\gamma$</td>
<td>$\text{No}(0, \sigma = 1.2)$</td>
</tr>
<tr>
<td>Time decay scale (s$^{-1}$)</td>
<td>$\lambda$</td>
<td>$\frac{1}{\text{Ga}(\alpha = 1, \beta = 1)}$</td>
</tr>
<tr>
<td>Power law index</td>
<td>$\alpha$</td>
<td>$-\text{Exp}(\text{scale} = 1.5)$</td>
</tr>
<tr>
<td>Cutoff energy (keV)</td>
<td>$E_c$</td>
<td>$\text{No}(300, \sigma = 100)$</td>
</tr>
</tbody>
</table>

4.3 Details of Implementation

4.3.1 RJ-MCMC

The model is fit using Markov chain Monte Carlo (MCMC). The dimension of the parameter space is unknown because we do not know the number of pulses comprising the gamma-ray burst. For this reason, reversible-jump MCMC is used \cite{18}. In addition to the typical Metropolis-Hastings step, reversible-jump involves several types of transdimensional step. Some of these steps add new pulses to the model (by “birthing” an entirely new pulse, or by “splitting” an existing pulse into two), and some decrease the number of pulses in the model, either by the “death” of a pulse or by “merging” two pulses into one. At each iteration of the Markov chain, one type of step is taken. This results in posterior samples of varying dimension, from which the posterior distribution for the number of pulses, $J$, can be determined.

We use the traditional birth, death and walk steps described in Appendix A. In addition we also use “merge” and “split” steps, which will merge two pulses or split a single pulse into two; the algorithm for these steps is described in Chapter 3. As it is hard for the Markov chain to explore both cases with the traditional birth, death and walk steps, the addition of split and merge steps improves mixing in cases where a feature in a GRB may be resolved as either one or two pulses. In this case the iterations are divided with probabilities $p_B, p_D, p_W, p_S$, and $p_M$ which all sum to one.
4.3.2 Parallel Thinning

Transdimensional Markov chains typically have lower acceptance rates than those with fixed dimension. In fixed dimension Metropolis-Hastings steps it is always possible to achieve a desired acceptance rate by proposing small enough steps, but that is not possible for jumps across dimensions. Additionally, the posterior distribution for the pulse structure in a GRB may be highly multimodal. In order to efficiently mix between modes, a new tempering algorithm known as parallel thinning is used [10]. This algorithm creates auxiliary chains using only a fraction of the available data. This results in faster mixing in these auxiliary chains, and it is possible to construct a Metropolis Hastings step to “swap” two adjacent chains while maintaining the correct limiting distribution. Details of the algorithm can be seen in Chapter 2.

4.3.3 Adaptive Metropolis-Hastings

We use a modification of the walk step of the Metropolis-Hastings algorithm which results in more effective mixing for this model. Each walk step affects only a single pulse, by innovating on every parameter of the pulse. Instead of having the standard deviations of the innovation proposals constant throughout the MCMC, we allow the standard deviations to vary between pulses as a function of the pulse volume at the current iteration. Since smaller pulses are less likely to affect the overall Poisson mean function, larger steps should be taken for smaller values of $V$. For a single Metropolis-Hastings step for a single pulse with parameters $(T, V, \gamma, \lambda, \alpha, E_c)$, we perform the transition as follows.

Suppose we are doing inference on parameters $X = (T, \gamma, \lambda, \alpha, E_c)$ and $V$. $V$ is a scalar and $X$ is a real-valued vector of dimension $p = 5$. We need to define a transition kernel $q$
\[ q \left( \begin{pmatrix} X \\ V \end{pmatrix} \to \begin{pmatrix} X^* \\ V^* \end{pmatrix} \right) \] (4.6)

which is reversible.

Consider the transition kernel
\[ \begin{pmatrix} X^* \\ V^* \end{pmatrix} \sim \mathcal{N}_0 \left( \begin{pmatrix} X \\ V \end{pmatrix}, \begin{pmatrix} \sigma_X V^{-1} I_p & 0 \\ 0 & \sigma_V V^{-1} \end{pmatrix} \right) \] (4.7)

for some \( \sigma_X, \sigma_V > 0 \), and where \( I_p \) denotes the identity matrix in \( \mathbb{R}^p \). Clearly this is reversible, since the normal distribution has full support.

The ratio
\[ \frac{q \left( \begin{pmatrix} X^* \\ V^* \end{pmatrix} \to \begin{pmatrix} X \\ V \end{pmatrix} \right)}{q \left( \begin{pmatrix} X \\ V \end{pmatrix} \to \begin{pmatrix} X^* \\ V^* \end{pmatrix} \right)} \] (4.8)

is needed in the Metropolis-Hastings algorithm and can be simplified:

\[
\begin{align*}
q \left( \begin{pmatrix} X^* \\ V^* \end{pmatrix} \to \begin{pmatrix} X \\ V \end{pmatrix} \right) & = \frac{V^*}{\sigma_X 2\pi}^{p/2} \exp \left\{ -\frac{\sum_{i=1}^{p} (X_i^* - X_i)^2 V^*}{2\sigma_X^2} \right\} \frac{\sqrt{V^*}}{\sigma_V \sqrt{2\pi}} \exp \left\{ -\frac{(V^* - V)^2 V^*}{2\sigma_V^2} \right\} \\
& = \left( \frac{V^*}{V} \right)^{(p+1)/2} \exp \left\{ -\frac{\sum_{i=1}^{p} (X_i^* - X_i)^2 V}{2\sigma_X^2} \right\} \frac{\sqrt{V}}{\sigma_V \sqrt{2\pi}} \exp \left\{ -\frac{(V^* - V)^2 V}{2\sigma_V^2} \right\}
\end{align*}
\]

Thus the transition described above is reversible and can be used in a Metropolis-Hastings scheme without concern for the ergodicity of the chain.

4.3.4 Fitting a Single Pulse

In this section we generate and analyze data simulated from our model that appears similar to a GRB. The data is drawn on a time interval of −12 to 25 seconds, with 0s
Table 4.2: Parameters for 1-pulse Simulation

<table>
<thead>
<tr>
<th>Interpretation</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start time (s)</td>
<td>( T )</td>
<td>5</td>
</tr>
<tr>
<td>Pulse volume (photons/cm(^2))</td>
<td>( V )</td>
<td>10</td>
</tr>
<tr>
<td>Linear time lag (s)</td>
<td>( \gamma )</td>
<td>1</td>
</tr>
<tr>
<td>Time decay (s(^{-1}))</td>
<td>( \lambda )</td>
<td>0.5</td>
</tr>
<tr>
<td>Power law index</td>
<td>( \alpha )</td>
<td>-0.94</td>
</tr>
<tr>
<td>Cutoff energy (keV)</td>
<td>( E_c )</td>
<td>150</td>
</tr>
</tbody>
</table>

representing the trigger time. As in many of the BATSE GRBs, the bin lengths have duration 1.024s at times before \(-2s\), and 0.064 after \(-2s\). A single pulse with the parameters described in Table (4.2) is transformed by the discriminators for GRB 1406. The data are then drawn from Poisson distributions with the means described by this transformed pulse, along with a fixed background rate. This background rate is the empirical rate from GRB 1406, and as in all our LARK scenarios, it is treated as known throughout the analysis.

No parallel thinning was used on this simulated dataset, as there is no indication that the model was multimodal in the energy spectrum parameters. Additionally, no overdispersion was allowed in these simulations, as the data are known to be Poisson.

The MCMC was 100,000 iterations long. Only each 5th iteration was kept, in an attempt to decrease the autocorrelation between the draws. The true value can be seen as the solid line in Figure (4.2)(a), which shows the posterior credible interval for the simulation. Figure (4.2)(b) shows the 95% posterior predictive interval, which indicates that the form of the mean function is well-recovered. Figure (4.3) shows the log-likelihood and \(\chi^2\) statistics as the MCMC progresses. Although this chain, as in many reversible-jump implementations, shows signs of “stickiness” (high autocorrelation between nearby iterations), we see that there is no long-term trend in these traceplots.

Figure (4.4) shows the posterior summary for \( J \), the number of pulses included
Figure 4.2: (a): Credible interval (95%). The solid line is the true value. (b): The 95% pointwise posterior predictive interval.

in the simulation. The first panel, (a), shows the traceplot for $J$. We see that the mixing between dimensions is quite good, as the chain explores dimensions as large as three regularly. Panel (b) shows the posterior histogram, which reassures us that the true number of pulses is recovered. The third panel indicates that even when
4.3.5 Simulating Two Pulses

In this section we simulate data in which one of the pulses does not have the characteristic FRED shape; that is, one of the pulses has a slower rise than decay. Recall that the model includes an instantaneous turn-on in the incident time spectra; however a slow rise in the binned-energy data can be obtained through a large linear lag parameter. This emphasizes the need for two-dimensional forward fitting which includes the influence of the discriminator in understanding the shape of GRB pulses. In addition, we add the wrinkle that there are two pulses, which are visually distinct but overlapping. The MCMC was 100,000 iterations long. The acceptance rate for the walk steps (which was about 40% of the steps) was 15%; the acceptance rates for
Figure 4.5: Histograms for parameters when there is only one pulse

the transdimensional steps were between 3 and 4%. These steps, although they have a low acceptance rate, are critical both in determining the correct number of pulses and in providing intermediate steps for the MCMC chain to move efficiently through the state space. Additionally three auxiliary chains were used in parallel thinning.

Posterior summary is difficult when the dimension of the posterior has uncertainty associated with it. Figure (4.7) shows the posterior summary for $J$, the number of
pulses. This is encouraging. Similarly Figure (4.6) shows the log-likelihood traceplot, which does indicate some small drift but remains in the range of about 10 units. The true number of pulses is well-recovered, as shown by the posterior summary for
the dimension in Figure (4.7). We know that under the true model, the largest pulse contains 77% of the photons, which is well-within the range shown in Figure (4.7)(c).

The pulse parameters are acceptably recovered, when we condition on the true number of pulses. Posterior histograms, with the true values indicated by vertical lines, can be seen in Figure (4.8) and Figure (4.9).

Figure (4.10) shows the credible interval for the mean flux, overlaid on the data.
The second subplot shows 100 draws overlaid on the data. The early draws are red, and the later draws are green; all the draws are partially transparent. The last panel shows a single draw of the MCMC along with $\chi^2$ statistics for each channel.
Table 4.4: Parameters for Two-Pulse Simulation Experiment

<table>
<thead>
<tr>
<th>Interpretation</th>
<th>Parameter</th>
<th>Pulse 1</th>
<th>Pulse 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start time (s)</td>
<td>$T$</td>
<td>5</td>
<td>Varies</td>
</tr>
<tr>
<td>Pulse volume (photons/cm²)</td>
<td>$V$</td>
<td>10</td>
<td>Varies</td>
</tr>
<tr>
<td>Linear time lag (s)</td>
<td>$\gamma$</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Time decay (s⁻¹)</td>
<td>$\lambda$</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Power law index</td>
<td>$\alpha$</td>
<td>$-1$</td>
<td>$-1$</td>
</tr>
<tr>
<td>Cutoff energy (keV)</td>
<td>$E_c$</td>
<td>350</td>
<td>350</td>
</tr>
</tbody>
</table>

4.3.6 Computer Experiment 1

In order to test the capabilities of the LARK model on simulated data, we perform a more structured experiment. We draw data from a model which uses two pulses with the same spectral model, but different volumes and turn-on times. The first pulse always has the same parameters, while the $V$ and $T$ parameters for the second pulse vary. The fixed values can be found in Table (4.4). The key goal here is to determine how close two pulses can be before the LARK model merges them into a single pulse.

The $E_c$ value is higher for these pulses than for those in earlier sections. The number of photons with energy greater than $E_c$ decreases exponentially with energy. It is desirable for inference that all channels have a high signal-to-noise ratio; a pulse with a low $E_c$ will peter out before any photons from the signal are sorted into Channel 4. Thus a higher value of $E_c$ causes more signal in Channel 4, the highest-energy band of the discriminator data. This is not an unusual value for this parameter ([16]), which can be as much as 1000 keV.

We test three values of $\Delta T = T_2 - T_2$: 1, 2, and 3 seconds; and vary $V_2$ from 2, 3, 5, 10 photons/cm². This comes to a total of twelve runs.

For each run, we begin the MCMC with one pulse starting at the true value, and a second pulse of volume 1 photon per cm² at time $T = 0$s. This allows us to avoid a long burn-in period. We then proceed for 100,000 iterations. Summary and
Table 4.5: Summary of Experiment 1: Number of Pulses Recovered (values with probability > 0.2)

<table>
<thead>
<tr>
<th>$V_2\Delta_T$</th>
<th>1s</th>
<th>2s</th>
<th>3s</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 photon/cm²</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3 photon/cm²</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5 photon/cm²</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>10 photon/cm²</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

discussion of each of these runs can be found in Appendix 2.

We are able to get a strong indication of the limits of the LARK model for detecting the correct number of pulses when the pulses are either small or overlapping. These results can be seen in Table (4.5). Here we see that the $V_2 = 2$ photons/cm² case does not result in recovery of the true number of pulses regardless of the separation between the pulses. However, as the second pulse becomes larger and/or the time separation of the pulses increases, we are better able to recover the true number of pulses.
Figure 4.10: (a): Credible interval for mean (95) with true mean shown in black. (b): 100 posterior draws overlaid on the data.
4.4 Comparison of One- and Two-dimensional Methods on Simulated Data

To compare the results of Experiment 1 to the single-channel fits, we fit the one-dimensional model from Chapter 3 to the simulated data for $V_2 = 5$ photon/cm$^2$ and $\Delta T = 2s$. We performed identical analyses on both the individual channels and the summed channels. We then compare these results to the results of the algorithm described in Hakkila et al. [22]

4.4.1 One-Dimensional LARK

The MCMC was run for 20,000 iterations per channel. A value of $\epsilon = 100$ photons/s was used, and the model was fixed to use only the Norris kernel, without any model selection or overdispersion.

In Figure (4.11), the first column shows sample draws from the MCMC chain. We see that recovery of the photon flux seems reasonable in all four energy channel light curves, as well as in the summed-channel light curve.

The second column indicates the posterior distribution for the number of pulses. The four channels are not consistent in the number of pulses recovered, and in particular we only recover the two large pulses in Channel 2 and the summed-channel analyses. In the other channels, as we can see in the third column of Figure (4.11), the largest pulse (by count fluence) contains almost all of the photons. The incoherence of these results emphasizes the need for joint analysis of the time and energy domains.

4.4.2 Comparative model

Dr. Hakkila used the algorithm from Hakkila et al. [22] to fit the individual channels using the Norris kernels [20]. Parameter estimates were consistent with those found by the one-dimensional LARK. The estimated mean functions for the four channels
Figure 4.11: Sample draws and number of pulses
as well as the summed-channel light curve can be seen in Figure (4.12). Chi-squared values were satisfactory to the point that this method would not detect that there were actually two pulses that comprise the mean function.

4.4.3 Two-Dimensional LARK

We compare the results of the single-channel methods to results to a longer version of the $V_2 = 5, \Delta_T = 2$ case of Experiment 1. On the same data, we evaluate the two-dimensional model for 100,000 iterations, discarding the first 25,000 iterations as burnin. Figure 4.13 shows traceplots and posterior summaries for various quantities. Subplots (a) and (b) demonstrate that the true number of pulses is recovered handily, and (d) indicates that $\chi^2$ values for the various posterior draws are well within what is expected for this data. Subplot (e) shows the posterior scatterplot for the largest and second largest pulses, with the true value indicated by a +; we see that the total volume is correctly estimated even though there is often some tradeoff between the two largest pulses.

That the correct number of pulses was recovered despite the one-dimensional
methods fitting reasonably well with a single pulse is indicative of the need for joint energetic and temporal fitting of the gamma-ray bursts. Additionally, methods such as the one-dimensional LARK often return incoherent numbers of pulses which make it difficult to perform analyses of the hard-to-soft evolution of the burst.
Figure 4.13: Joint Time and Energy Model Plots for Computer Experiment
4.5 Example: GRBs

We now turn our attention to using the model to analyze actual gamma-ray bursts. It is not expected that all bursts will be well-described by the instant turn-on model, but a sufficient number can be found that the necessity of modeling in the incident photon space becomes clear. For an initial analysis, we focus on gamma-ray bursts with relatively short (< 20s) durations and relatively good signal-to-noise ratios in at least three of the energy channels. For these reasons we choose the GRBs with trigger numbers 493, 501, and 537. These bursts all appear visually as a single pulse, although some (493) are known to not fit well under the Norris kernel models [20]. For each burst, a MCMC with 100,000 iterations (the first 25,000 being discarded as burnin) was performed.

4.5.1 Overdispersion

The model described in Section (4.2) is necessarily an approximation to the true process generating the data. To compensate for model misspecification, an overdispersion parameter, \( r \), is introduced. This parameter does not depend on the time interval or the energy channel. To allow for overdispersion of the data, instead of

\[ \text{Figure 4.14: Sample draws for two-dimensional LARK model (in methods comparison experiment)} \]
modeling $Y_{ij}$ as Poisson, we can model it as negative binomial:

$$Y_{ij} \sim \text{NB} \left( n = r \mu_i, p = \frac{r}{1 + r} \right).$$

Then $\mathbb{E}(Y_{ij}) = \mu_{ij}$, as before, but instead of constraining the variance of $Y_{ij}$ to be $\mu_{ij}$, as in the Poisson case,

$$\text{Var}(Y_{ij}) = \mu_{ij} \frac{1 + r}{r}$$

Thus, when $r$ is small, the variance of $Y_i$ is inflated beyond that of a Poisson model. This parameterization is not affected by merging two adjacent bins (and thus to any re-binning, by induction), because

$$\text{NB}(n_1, p) + \text{NB}(n_2, p) = \text{NB}(n_1 + n_2, p).$$

When applicable (i.e., when the data is modeled as overdispersed instead of Poisson), the ratio $p = \frac{r}{1 + r}$, varies between 0 and 1, and is given a uniform prior on that range.

### 4.5.2 GRB 501

Figures for burst 501 can be seen in Figure (4.15). This burst is fit with between three and four pulses, with a preference for three. In Figure (4.15)(b), we see that almost all of the photon fluence is contained in the largest two pulses. The overdispersion is less than in GRB 493, indicating that the Poisson model with instant turn-on (which we know to be an approximation) fits reasonably well. Figure (4.16) shows the distinct two-(large)-pulse fits obtained by this model very clearly. Figure (4.16)(a) shows the credible interval for the mean, which we see goes through the middle of the time series. The median is shown on this plot as a solid line. Figure (4.16)(b), the posterior predictive intervals, which include the effect of overdispersion from the negative binomial model, show that the coverage of this model is fairly good. The likelihood is very sticky, due to the difficulty of the transdimensional MCMC, and the effective sample size based on the $\chi^2$ statistic was only 16.0.
We can then do inference on some of the parameters of the pulses for this burst. Table (4.6) gives the 95% highest posterior density intervals (HPDs) for the pulse parameters, conditioned on the model finding three pulses. This conditioning is done to help with the label-switching problem. Here they are ordered in decreasing volume.

We see some interesting results here; for example, the sign of the lag parameter, $\gamma$, is not consistent between the pulses (clearly positive for the larger two, and negative for the smallest pulse). Additionally, information about the spectral parameters $E_c$ and $\alpha$ are recovered only in the largest pulse; the HPDs for the other two are similar to the highest-prior-density intervals.

Pairwise scatterplots of some of the parameters of the largest pulse can be seen in Figure (4.17) and Figure (4.18). Due to the small effective sample size, these scatterplots show some “bunching” indicative of poor mixing. Figure (4.17)(b) shows a strong relationship between the volume of the pulse and its decay time; while scale $1/\lambda$ and $\gamma$ are uncorrelated in Figure (4.17)(c).

Figure 4.15: Plots for GRB 501
Figure 4.16: Plots for GRB 501. (a): Posterior predictive intervals (95%) 
(b): Credible interval for mean (95%)

Table 4.6: Highest posterior density intervals (95%) for the parameters of the two-pulse models, GRB 501.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Pulse 1</th>
<th>Pulse 2</th>
<th>Pulse 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$ (s)</td>
<td>$(-0.542, -0.344)$</td>
<td>$(-0.542, -0.344)$</td>
<td>$(-0.542, -0.344)$</td>
</tr>
<tr>
<td>$V$ (photons/s)</td>
<td>$(10.2, 19.3)$</td>
<td>$(10.2, 19.3)$</td>
<td>$(10.2, 19.3)$</td>
</tr>
<tr>
<td>$\gamma$ (s)</td>
<td>$(0.5923, 0.875)$</td>
<td>$(0.5923, 0.875)$</td>
<td>$(0.5923, 0.875)$</td>
</tr>
<tr>
<td>$\lambda$ (s$^{-1}$)</td>
<td>$(0.344, 0.511)$</td>
<td>$(0.344, 0.511)$</td>
<td>$(0.344, 0.511)$</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>$(-0.229, 0)$</td>
<td>$(-0.229, 0)$</td>
<td>$(-0.229, 0)$</td>
</tr>
<tr>
<td>$E_c$ (keV)</td>
<td>$(53.1, 71.5)$</td>
<td>$(53.1, 71.5)$</td>
<td>$(53.1, 71.5)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4.17: Temporal parameter scatterplots for the largest pulse, GRB 501

Figure (4.18) indicates the correlation of various energy-spectrum parameters. Here we see that $V$ and $\alpha$, nor are $V$ and $E_c$, are not strongly correlated; however, there is a negative correlation between $E_c$ and $\alpha$. This is what we would expect: as the cutoff energy increases, increasing the number of high-energy photons, $\alpha$ decreases in order to maintain the correct “volume” of photons in the pulse.
Figure 4.18: Interesting pulse parameter correlations, GRB 501

4.5.3 GRB 537

GRB 537 had a very high variance inflation parameter (Figure (4.19)(c)), but as a result a LARK model was able to fit the data very well with few pulses (Figure (4.19)(a)) In this case there appear to be only two, roughly equally-sized pulses. The model appears to fit well to this burst, although it does not put any weight on the single-pulse model. The sample size for this simulation was 28.1.
Figure 4.19: Plots for GRB 53798

(a): $J$ Histogram

(b): Cumulative Volume by $J$

(c): $(1 + r)/r$, Overdispersion
Figure 4.20: Plots for GRB 537. (a): Posterior predictive intervals (95%)  
(b): Credible interval for mean (95%)
4.5.4 GRB 493

Trigger 493 is fit with a large number of pulses, as seen in the posterior histogram in Figure (4.21)(a), a large number of which contribute significantly to the total photon fluence (Figure (4.21)(b)). The model fit, as measured by the overdispersion factor in Figure (4.21)(c), which indicates that the variance is inflated by between 5 and 15% relative to the Poisson model, indicates that there is some misspecification in the underlying model. Our flexible approach, which allows for some overdispersion as well as multiple pulses, allows us to get satisfactory posterior mean intervals and predictive intervals, seen in Figure (4.22). It seems from the credible interval that a large number of pulses are needed to obtain a slower rise time than is allowed by our instant turn-on time model. In future work this can be incorporated into a LARK framework, either by expanding the class of kernels used or by incorporating model choice into our Bayesian problem.

\[ (1 + r)/r, \text{ Overdispersion} \]

**Figure 4.21:** Plots for GRB 493
4.6 Discussion

In this chapter we describe and apply a novel Bayesian nonparametric regression model to the problem of joint spectro- and temporal- modeling of the prompt emissions of gamma-ray bursts. We demonstrate the efficacy of this method in recovering the spectral and temporal parameters in simulation studies, and perform a simulation study which indicates the value of the LARK methods in determining the number of pulses comprising the burst. Furthermore, we compare the performance of our two-dimensional model and a pair of single-channel methods on some simulated data. Finally, we demonstrate that an instant turn-on model is able to fit some
gamma-ray bursts when the effect of the discriminators and the spectral evolution of the pulse are taken into effect.

Future work should include a generalization of the model to allow a non-instantaneous turn-on of the burst at a given energy level. In addition, the underlying Lévy process may be changed to an $\alpha$-stable process, which will more closely mimic the known distribution of pulse volumes. Analysis of higher-resolution data will also allow more subtle energy spectrum modeling; on BATSE, the medium-energy resolution (MER) data, which is sorted into 16 energy bins as opposed to the current four, comes with additional statistical challenges.

Due to the effect of the discriminators on BATSE and other instruments, it is important to model the gamma-ray bursts in the incident space. The model described in this chapter is the first to model bursts in the incident photon space, instead of after they have been distorted by the discriminators; since to model photons as they enter the detector is to model both the energy and the arrival time of the incident photon, this model is also the first to jointly model the time and energy domains. In addition, the Bayesian nonparametric nature of this method results on doing inference on the number of bursts, along with quantification of the uncertainty in this value.
Appendix A

Reversible Jump MCMC

Reversible-jump MCMC is a technique for sampling from a parameter space of unknown dimension, which occurs in many Bayesian non-parametric statistical models. It is also useful when Bayesian model selection is being done between models of either the same or differing sizes when more analytic methods are not available.

The basic reversible-jump algorithm for a parameter space of varying dimension involves three types of steps: “birth,” “death,” and “walk.” The birth and death steps are transdimensional and change the dimension of the parameter space, or affect the interpretation of the parameters, while the walk step is a traditional Metropolis-Hastings or Gibbs step. For each iteration of the Markov chain, one type of step is taken. In the context of LARK models, where each addition to the parameter space represents a kernel used in the regression, we also introduce “split” and “merge” steps which combine and divide kernels.

In model selection, there are traditional “walk” steps, which explore the parameter space conditioned on the model, and trans-model “switch” steps which attempt to move the chain to a different model.
A.1 RJ-MCMC for Model Selection

Consider the proposed switch move from \((M, \theta) \rightarrow (M^*, \theta^*)\) where \(M\) is the model indicator (0 for Norris, 1 for GiG, and 2 for pGiG). The data is called \(Y\). Then the Hastings ratio for this move is:

\[
H = \frac{p(Y \mid \theta^*, M^*)}{p(Y \mid \theta, M)} \times \frac{p(\theta^* \mid M^*)}{p(\theta \mid M)} \times \frac{p(M^*)}{p(M)} \times \frac{q(M, \theta) \rightarrow (M, \theta)}{q(M^*, \theta^*) \rightarrow (M^*, \theta^*)}
\]

Note that the likelihood of the data given \(\theta\) is independent of \(M\), due to the nested nature of the models, and so \(p(Y \mid \theta, M) = p(Y \mid \theta)\).

This results in a Hastings ratio of

\[
H = \frac{p(Y \mid \theta^*)}{p(Y \mid \theta)} \times \frac{p(\theta^* \mid M^*)}{p(\theta \mid M)} \times \frac{p(M^*)}{p(M)} \times \frac{q(M, \theta) \rightarrow (M, \theta)}{q(M^*, \theta^*) \rightarrow (M^*, \theta^*)}
\]

A.2 RJ-MCMC for LARK models

Birth Steps

In the birth step, we propose adding a new pulse to the \(J\) existing pulses. The parameters for this new pulse are drawn from the birth distribution \(\theta^* \sim b(\cdot)\). We call the list of jumps before adding the new pulse \(\Theta = \langle \theta_1 \ldots \theta_J \rangle\) and the list after adding the proposed pulse \(\Theta^* = \langle \theta_1 \ldots \theta_J, \theta_* \rangle\). This birth distribution must have the same support as the prior, but can otherwise be of any form.

We accept the proposed move from \(\Theta\) to \(\Theta^*\) using the Metropolis-Hastings ratio:

\[
\frac{\mathcal{L}(Y \mid \Theta^*) \Pi(\Theta^*) q(\Theta^*, \Theta)}{\mathcal{L}(Y \mid \Theta) \Pi(\Theta) q(\Theta, \Theta^*)}
\]  

(A.1)

where \(\Pi(\cdot)\) is the prior for a set of pulses, \(\mathcal{L}(Y \mid \cdot)\) is the likelihood of the data given a set of pulses, and \(q(a, b)\) is the transition probability from state \(a\) to \(b\).
Note that the priors for $\Theta_*$ and $\Theta$ are functions of $J$, as well as functions of the pulse parameters themselves. The number of pulses greater than $\epsilon$ is

$$J \sim \text{Po}(\mu_\epsilon)$$  \hspace{2cm} (A.2)

where

$$\mu_\epsilon = \int_\epsilon^\infty \nu(du)$$  \hspace{2cm} (A.3)

and $\nu(du)$ is the Lévy measure for the prior process. For simplicity in notation, for a given pulse described by parameters $\theta$, we say $\theta \sim \pi(\cdot)$ a priori. Then, the prior distribution for $\Theta_*$ can be written:

$$\Pi(\Theta \mid J, \theta_1 \ldots \theta_J) = \prod_{j=1}^J \pi(\theta_j) \times \frac{\mu_\epsilon^J}{J!} \exp\{-\mu_\epsilon\}$$  \hspace{2cm} (A.4)

and we can write the ratio

$$\frac{\Pi(\Theta_* \mid J + 1, \theta_1, \ldots, \theta_J, \theta_{J+1})}{\Pi(\Theta \mid J, \theta_1 \ldots \theta_J)} = \frac{\prod_{j=1}^J \pi(\theta_j)\pi(\theta_*) \times \frac{\mu_\epsilon^{J+1}}{(J+1)!} \exp\{-\mu_\epsilon\}}{\prod_{j=1}^J \pi(\theta_j) \times \frac{\mu_\epsilon^J}{J!} \exp\{-\mu_\epsilon\}}$$  \hspace{2cm} (A.5)

$$= \frac{\pi(\theta_*) \frac{\mu_\epsilon}{J+1}}{1}$$  \hspace{2cm} (A.6)

The transition from $\Theta$ to $\Theta_*$, written $q(\Theta, \Theta_*)$, is $p_B \times b(\theta_*)/(J + 1)$. The factor of $J + 1$ represents the arbitrary choice to place $\theta_*$ at the end of the list of pulses to create $\Theta_*$. The transition $q(\Theta_*, \Theta)$ represents the probability of moving from $\Theta_*$ to $\Theta$. This is a death step, which will be described in more detail shortly. The probability of choosing a death step is $p_D$, and a pulse is chosen uniformly from $\theta_1, \ldots, \theta_J, \theta_*$ to be removed. In order to move from $\Theta_*$ to $\Theta$, $\theta_*$ must be selected to be removed from the model. The probability that $\theta_*$ is chosen is $1/(J + 1)$ and thus $q(\Theta_*, \Theta) = \frac{p_D}{J+1}$. 

105
The ratio \( q(\Theta_*, \Theta) / q(\Theta, \Theta_*) \) is then \( \frac{p_D}{p_B \times b(\theta_*)} \).

Thus the Metropolis-Hastings ratio can be written

\[
H = \frac{\mathcal{L}(Y | \Theta_*)}{\mathcal{L}(Y | \Theta)} \times \frac{\pi(\theta_*)}{\pi(\theta)} \frac{\mu_e}{J + 1} \times \frac{p_D}{p_B \times b(\theta_*)}
\]

As in the traditional Metropolis-Hastings procedure, the proposed move is accepted with probability \( \max(1, H) \).

**Death Steps**

When performing a death step, as alluded to above, we remove a pulse \( \theta_k \) chosen uniformly at random from the list of pulses \( \Theta \) to create \( \Theta_* \). Here again \( J \) is the number of pulses in \( \Theta \).

The ratio of priors \( \Pi(\Theta_*) / \Pi(\Theta) \) is found in the same way as in the birth step to yield:

\[
\frac{\Pi(\Theta_*)}{\Pi(\Theta)} = \frac{\prod_{1 \leq j \leq J, j \neq k} \pi(\theta_j) \times \frac{\mu_e^{J-1}}{(J-1)!}}{\prod_{1 \leq j \leq J} \pi(\theta_j) \times \frac{\mu_e^J}{J!}} = \frac{J}{\pi(\theta_k) \mu_e} \tag{A.7}
\]

The transition kernel for the death move \( q(\Theta_*, \Theta) \) is \( p_D / J \). The transition kernel for the reverse move, the birth of \( \theta_k \), is found in the same way as described in the previous section, with the caveat that \( \Theta \) has \( J - 1 \) pulses instead of \( J \):

\[
q(\Theta, \Theta_*) = \frac{p_B b(\theta_k)}{J}
\]

So the Metropolis-Hastings ratio for a death step is

\[
H = \frac{\mathcal{L}(Y | \Theta_*)}{\mathcal{L}(Y | \Theta)} \times \frac{J}{\pi(\theta_k) \mu_e} \times \frac{p_D}{p_B b(\theta_k)} \tag{A.9}
\]

and the move is accepted with probability \( \max(1, H) \).
Walk Steps

Walk steps are performed at each iteration with probability $p_W$. This step is no different from a standard Metropolis-Hastings update. It is not a problem from a theoretical perspective to attempt to innovate on more than one pulse at once, but since the other steps described also affect only a single pulse, we use the same approach for the walk steps. Typically, a single pulse, $\theta_k$, is chosen uniformly at random from the pulses in $\Theta$, and its parameters are changed with some innovation kernel $\theta_s \sim p(\cdot \mid \theta_k)$. A common choice for $p$ is a Normal distribution centered around $\theta_k$. In this case, the Metropolis-Hastings ratio is very simple:

$$
H = \frac{\mathcal{L}(Y \mid \Theta_s)}{\mathcal{L}(Y \mid \Theta)} \times \frac{\pi(\theta_s)}{\pi(\theta_k)} \times \frac{p(\theta_k \mid \theta_s)}{p(\theta_s \mid \theta_k)} \tag{A.10}
$$

A.2.1 Gamma Process used in Chapters 2 and 3

In Chapter 2, we use reversible-jump for LARK to estimate the number of stars visible in an astronomical image. In Chapter 3, RJ-MCMC is used to analyze gamma-ray burst single-channel data. In both cases, the number and shape of objects is modeled as a gamma process. For the gamma process, with rate parameter $\alpha(dt)$ on a compact set of length $L$ and rate parameter $\beta$, as described in Section (1.1.1), the number of pulses (or stars), $J$, is modeled as follows

$$
J \sim \text{Po}(\alpha LE_1(\beta \epsilon)) \tag{A.11}
$$

and the amplitude parameters associated with the jumps of the process have distribution

$$
f(u) = \frac{1}{E_1/\beta \epsilon} u^{-1} \exp\{-u\} 1_{\{u > \epsilon\}}
$$
A.2.2 \textit{\alpha\text{-stable Process used in Chapter 4}}

In Chapter 4, an alpha-stable process is used to model the incident photon spectrum of gamma-ray bursts. The truncated process is modeled using RJ-MCMC. For a fully-skewed alpha-stable process with parameters \((\alpha, 1, \gamma(dt), 0)\) on a domain of length \(L\), the number of pulses larger than \(\epsilon\) is distributed

\[J_{\epsilon} \sim \text{Po}(2\gamma L c e^{-\alpha})\]

where

\[c = \frac{1}{\pi} \Gamma(\alpha) \sin \left(\frac{\pi \alpha}{2}\right)\]

and the volumes of these pulses have independent Pareto distributions

\[f(v) = \frac{\alpha \epsilon^\alpha}{v^{\alpha+1}} 1_{\{v > \epsilon\}}\]
Appendix B

Simulation Experiment

Table B.1: Parameters for Two-Pulse Simulation Experiment

<table>
<thead>
<tr>
<th>Interpretation</th>
<th>Parameter</th>
<th>Pulse 1</th>
<th>Pulse 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start time (s)</td>
<td>$T$</td>
<td>5</td>
<td>Varies</td>
</tr>
<tr>
<td>Pulse volume (photons/cm$^2$)</td>
<td>$V$</td>
<td>10</td>
<td>Varies</td>
</tr>
<tr>
<td>Linear time lag</td>
<td>$\gamma$</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Time decay ($s^{-1}$)</td>
<td>$\lambda$</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Power law index</td>
<td>$\alpha$</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>Cutoff energy (keV)</td>
<td>$E_c$</td>
<td>350</td>
<td>350</td>
</tr>
</tbody>
</table>

As a test of the two-dimensional inverse problem model described in Chapter 4, we perform a computer experiment, where we systematically vary the parameters of two simulated pulses and observe how well LARK recovers their properties.

The two pulses used in the simulation have the same time and energy spectra, and vary only in amplitude and the time which separates their onset. These parameters can be seen in Table (B.1). We test three values of $\Delta T = T_2 - T_1$: 1, 2, and 3 seconds; and vary $V_2$ from 2, 3, 5, 10 photons/cm$^2$.

In Figure (B.1) we see that some of the runs succeeded in recovering the two pulses separately, instead of merging them into one. None of the smallest pulses,
\[ V_2 \mid \Delta_T = T_2 - T_1 \]

2 photon/cm²

3

5

10

Figure B.1: Histograms for \( J \) for the various configurations of the parameters

where \( V_2 = 2 \) photons/cm² resulted in a successful recovery of the true number of pulses. Additionally, in the first column of Figure (B.1) we see that if the second pulse is significantly smaller than the first pulse, than the smaller pulse is is not well-recovered. When \( \Delta_T = 1 \)s, only when the two pulses are equally sized \((V_2 = 10 \text{ photons/cm}^2)\) are the correct number of pulses recovered. In many of the simulations, however, the true number of pulses was identified.

Figure (B.2) shows the cumulative volume of the pulses in each of the twelve simulations. Here we see that typically, more volume is assigned to the largest pulse than to the smaller pulses.

In the next four figures (Figure (B.3) – Figure (B.6)), we see both the credible intervals and the posterior predictive intervals (both 95%). Coverage is good, indi-
\[ V_2 | \Delta T = T_2 - T_1 \]

2 photon/cm²

\[ V_3 \]

\[ V_5 \]

\[ V_{10} \]

\textbf{Figure B.2:} Pulse volume fractions.

dicating that model fit is satisfactory. Figure (B.7) gives \( \chi^2 \) statistic values for various iterations, also indicates satisfactory model fit. The red line at 1720 on this plot indicates how many degrees of freedom the true model has. Clearly all of the iterations indicate that the fits are adequate using \( \chi^2 \) test, which is a standard approach in astronomical lightcurve fitting.
Figure B.3: $V_2 = 2$. (a): $\Delta_T = 1$. (b): $\Delta_T = 2$. (c): $\Delta_T = 3$. L: 95% Credible intervals with true values overlaid in black. R: Posterior 95% predictive intervals.
Figure B.4: $V_2 = 3$. (a): $\Delta_T = 1$. (b): $\Delta_T = 2$. (c): $\Delta_T = 3$. L: 95% Credible intervals with true values overlaid in black. R: Posterior 95% predictive intervals.
Figure B.5: $V_2 = 5$. (a): $\Delta_T = 1$. (b): $\Delta_T = 2$. (c): $\Delta_T = 3$. L: 95% Credible intervals with true values overlaid in black. R: Posterior 95% predictive intervals.
Figure B.6: $V_2 = 10$. (a): $\Delta_T = 1$. (b): $\Delta_T = 2$. (c): $\Delta_T = 3$. L: 95% Credible intervals with true values overlaid in black. R: Posterior 95% predictive intervals.
Figure B.7: $\chi^2$, Computer Experiment
Bibliography


Biography

Mary Elizabeth Broadbent was born on February 20, 1990 in Louisville, KY. She received her B.A. in Chemistry and Mathematics *cum laude* from Amherst College in 2010. In 2012, she earned a M.S. in Statistical Science *en route* to her Ph.D. in Statistical Science, which she obtained under the supervision of Robert L. Wolpert in 2014.