A Model for the Classification of Supernovae

Brittany Spencer, C. Shane Reese
Brigham Young University

Abstract— With data collection projects such as the Dark Energy Survey underway, data from distant supernovae are becoming increasingly available. As the quantity of information increases, the ability to quickly and accurately distinguish between Type Ia and core collapse supernovae has become an essential key to understanding the nature of the evolving universe. Estimating individual supernova light curves is the first step in modern classification attempts. In this research we focus on the use of hierarchical gaussian processes to model light curves both for individual supernova and across supernova type. Properties inherent in this Bayesian non-parametric form of modeling allow curve definition at a specific point to borrow information from neighboring points and allow the data to dominate the selection of model parameters.

1 Introduction

Type Ia supernovae play a critical role in understanding the nature of the evolving universe. For example, the Dark Energy Survey intends to explore the acceleration of the universe, and thereby gain insight into the nature of dark energy. To accomplish this task, they quantify four measures of expansion, one of which is Type Ia supernovae (DES Projects, 2011). Type Ia supernova are considered “standard candles” due to the fact that all supernova of this type explode to approximately the same absolute brightness (Branch, 1992). These supernovae then function as useful galactic measuring sticks because relative distances of supernovae can be inferred from apparent brightness. However, in order to use these supernovae they must first be distinguished from core collapse supernovae. Thus, there is a crucial need for an accurate and efficient classification system.

A fundamental distinction between the classes of supernova arises in origination. Core collapse supernovae, or more broadly Type Ib, Ic and II, form when the fusion within a massive star is no longer able to repel the force of its own gravity and the star collapses on itself. This results in a violent explosion (Janka, et al., 2007). Alternatively, Type Ia, or thermonuclear supernovae, form from comparatively smaller white dwarf stars in which fusion has ceased. Candidate stars accrete mass from surrounding objects and if the star’s mass reaches a certain threshold, the temperature and pressure become such that the star explodes (Branch, 1992). This phenomenon explains why all Type IA supernova explode to approximately the same absolute brightness.

Classifying supernovae involves employing one of two approaches. Spectroscopy involves the categorization of supernova based on absorption spectra. The presence of hydrogen determines if a supernova will be classified as Type I or Type II, with other elements dictating the further distinction between Type Ia, Ib and so forth. Using the alternative method of photometry, a supernova is classified according to the shape of its light curve. Light curves are composed of flux or magnitude of light received from the supernova and the time at which the light is received.

Although these methods have been successful in varying degrees, both techniques exhibit crippling flaws. Acquiring the data required to spectroscopically confirm a supernova necessitates an immense dedication of time, making the method expensive and impractical as efforts of data collection turn to large projects such as the Dark Energy Survey. Although photometry reduces data dimensionality, these alternative methods usually fall prey to equally detrimental shortcomings. Because photometric techniques
necessitate the modeling of light curves, this method often lacks efficiency, the ability to consistently classify supernovae, or purity, the ability to classify only Type Ia supernova as Type Ia (Gjergo et al., 2012). With the increased demand for a reliable and expeditious classification technique, it is essential that a photometric system of classification be developed that overcomes these shortcomings.

The first step in the development of a functional photometric classification technique is the estimation of supernova light curves. In this project, we explore the use of a hierarchical Gaussian process to model these curves. The data used in this analysis are described in Section 2. Section 3 details the modeling method. Section 4 includes the presentation of results from our supervised classification. Section 5 contains conclusions and outlines our intended future work.

2 Data

Data for this project were provided by Rahul Biswas, a researcher at Argonne National Laboratories in the High Energy Physics Division. All data were simulated using the Supernova Analysis (SNANA) software. In total, 14910 Type Ia supernova and 72358 core collapse supernova are included in the simulated dataset. A training dataset of 1500 Type Ia and 7000 core collapse supernova was randomly selected from the main dataset. Eventually the remaining supernova will be used as a test set for unsupervised classification.

For each supernova the red-shift, denoted $z$, is given. The redshift is estimated from a simulated host-galaxy. Also included are the observation times and light flux. Consistent with common practice, the observation times are adjusted in relation to the peak flux and with regard to redshift and thereafter regarded as phase. Light flux is also adjusted for redshift. The filter, $g$, $r$, $i$, or $z$, with which each observation was obtained is also included. In this project only $i$ filters were used; we intend to incorporate the remaining filters in subsequent work.

Figure 1 contains a sample of supernova light curves from the data. Estimating models for these light curves included the challenges of deciphering signal from noise and responding to the large variation in light flux. Thus, estimating these light curves necessitated an inherently flexible yet simultaneously smooth model. We found the solution in hierarchical Gaussian processes.

3 Methods

In this project, we model supernova light curves using a functional data analysis approach. Section 3.1 contains a brief introduction to smoothing splines and their relation to gaussian processes. A general overview of Gaussian Processes is presented in Section 3.2. Section 3.3 details the structure of the hierarchical process used in this project. Section 3.4 includes a brief overview of computation mechanisms.

3.1 Smoothing Splines

Smoothing Splines are a relatively common method for modeling nonlinear effects. Specifically, a smoothing spline $f(x)$ is the
unique natural cubic spline using all distinct \(X_i\) as design points and the simultaneous minimizer of the penalized sum of squares below

\[
PSS = \sum_{i=1}^{k} (Y_i - f(X_i))^2 + \phi \int (f''(x))^2 \tag{3.1}
\]

where \(\phi\) is the smoothing parameter and the integral term on the right is the roughness penalty. Inherent in the definition of a smoothing spline is the tradeoff between smoothness and data interpolation. This tradeoff is governed by the smoothing parameter \(\phi\). When \(\phi\) is large, priority is placed on smoothness; when \(\phi\) is small, curve fit receives priority (Eubank, 1999). Approaching \(\phi\) is placed on smoothness; when \(\phi\) is small, curve fit receives priority (Eubank, 1999). Approaching nonlinear regression in this way allows for a continuous smooth solution, that is also the minimizer over all twice-differentiable functions on the interval (Green and Silverman, 1994).

Using the piecewise polynomial basis for the \(X_i\) given by de Boor (2001) and further explained by Eubank (1999), equation (1) can be expressed as

\[
PSS = (y - f)^T (y - f) + 6 \phi f^T QR^{-1} Q f \tag{3.2}
\]

where \(R\) is the symmetric tridiagonal \((n-2) \times (n-2)\) matrix defined with first and last row as

\[
(h_1, h_2, 0, \ldots, 0) \quad \text{and} \quad (0, \ldots, 0, h_{n-2}, h_{n-1})
\]

and common \(i\)th row

\[
(0, \ldots, 0, h_{i-1}, 2(h_{i-1} + h_i), h_i, 0, \ldots, 0)
\]

and \(Q^T\) is the tridiagonal \(n \times (n-2)\) with common \(i\)th row

\[
(0, \ldots, 0, h_{i-1}^{-1}, h_{i-1}^{-1} - h_i^{-1}, h_i^{-1}, 0, \ldots, 0)
\]

where \(h_i = x_{i+1} - x_i\). We label \(K = 6QR^{-1}Q^T\), which implies \(f^T K f = \int (f''(x))^2\). Using the basis described above, a closed form estimator of \(f\) is available. Specifically, taking the derivative of (2) we have

\[
\frac{\partial}{\partial f} PSS = 2(y - f) + 2\phi K f.
\]

Setting this to zero and solving for \(f\) returns

\[
f = (I + \phi K)^{-1} y, \tag{3.3}
\]
a closed form estimator for the smoothing spline.

In Bayesian framework, smoothing spline derivation requires the additional assumption of normally distributed \(Y_i\)’s and with variance \(\sigma^2_e\). The unknown vector \(f\) is given a prior distribution proportional to the partially improper gaussian process

\[
\exp \left\{ -\frac{\phi}{2\sigma^2_e} f^T K f \right\}. \tag{3.4}
\]

Conditional upon the previously stated assumptions, the posterior distribution of \(f\) is a gaussian process with mean function \((I + \phi K)^{-1} y\) and covariance matrix \(\sigma^2_e (I + \phi K)^{-1}\). Berry, Carroll, and Ruppert (2002) also point out that although \(K\) depends on the knot locations, since \(f^T K f = \int (f''(x))^2\), this distribution is independent of knot locations.

### 3.2 Gaussian Processes

Gaussian processes are a special case of stochastic processes. A stochastic process implies the existence of a function that governs the placement of the data (Hida & Itôsuda, 1993). The process is Gaussian if any finite realization of random variables also follows a Gaussian distribution.

We define \(h(t)\) as a Gaussian process on \(t_1, \ldots, t_n\). Then

\[
h(t) = \left( \begin{array}{c} h(t_1) \\ \vdots \\ h(t_n) \end{array} \right) \sim \mathcal{N}_n \left( \left( \begin{array}{c} \mu(t_1) \\ \vdots \\ \mu(t_n) \end{array} \right), \Sigma \right)
\]

where \(\mu\) is the mean function and \(\Sigma\) is the covariance function. A thorough overview of Gaussian Processes is given by Barber (2012); only a brief introduction, borrowing partly from his work, is presented here.

The covariance function relates points to surrounding ones. For example if \(t\) and \(t'\) are close we expect the output at \(y\) and \(y'\) to be similar. A possible covariance function is \(\Sigma_{ij} = \exp\left\{-\phi ||t_i - t_j||^2\right\}\), where \(||t_i - t_j||\) denotes the Euclidean distance between locations \(t_i\) and \(t_j\) and \(\phi\) is a smoothing parameter. On the other hand, if many draws are taken from the Gaussian Process and averaged at each design point, these averages tend to \(\mu\).

As an example, consider a model for the the 6 randomly selected points in Figure 2.
Gaussian Processes define highly flexible smooth curves that are close approximations to the data.

Six realizations from a Gaussian Process are given in various colors. We note several key features of this process. First, draws from a Gaussian process result in smooth curves over the space. Despite the smoothness of these curves, they remain incredibly flexible due to the lack of monotonic restriction. In addition, these draws also closely approximate the data. This property is evident in Figure 3 where 10000 realizations are given. The realizations from the process are given in gray with the mean curve in blue. In locations with a relatively high density of data, observed variability across curves is greatly reduced.

Using Gaussian Processes to model supernova light curves has attractive features. We can expect these curves to fit well, be flexible enough to account for differences across supernova and simultaneously remain smooth. In addition, we can think of individual light curves for a class of supernova as draws from a Gaussian process with the same mean curve but conditional on differing data. We can also model mean curves with a Gaussian process.

3.3 Hierarchical Gaussian Processes

The Gaussian process employed in this project is considered hierarchical because the mean functions, smoothing parameters, and variance measures are not fixed but assigned priors and estimated from the data. Appealing to the underlying hierarchical structure of the data also implies the added benefit of borrowing strength across curves.

To ease computation while incorporating the underlying hierarchical structure, the flux-phase space was re-expressed at 25 unique design points. For each of the $k$ supernova, class parameters of interest include

- Supernova specific light curves, $	heta_{iT} = (\theta_{iT}(z_1), \ldots, \theta_{iT}(z_{25}))$
- Supernova specific variances, $\sigma_{iT}^2$
- Mean light curve, $\mu_T = (\mu_T(z_1), \ldots, \mu_T(z_{25}))$
- Design point specific variances, $\tau_{jT}^2$
- Smoothing parameter, $\phi_T$

where the subscript $T$ corresponds to either Type Ia or core collapse, $i = 1, \ldots, k$, and $j = 1, \ldots, 25$.

The data has the following Normal sampling density,

$$f(Y|\theta) = \frac{\exp \left\{ -\frac{1}{2} \sum_{i=1}^k (Y_i - \theta_i)^\top \Sigma_i^{-1} (Y_i - \theta_i) \right\}}{(2\pi)^{mk/2} \prod_{i=1}^k |\Sigma_i|^{1/2}}$$

thereby satisfying the condition given in Section...
3.2. The following priors are assigned

\[
\begin{align*}
\pi(\theta_{\mu}\mid \mu) & \sim N_m(\mu, [T^{-1}_T + \phi K_s]^{-1}) \\
\pi(\mu_T\mid \phi_T) & \sim N_n(0, \phi K_s^{-1}) \\
\pi(\phi_T\mid \phi_T) & \sim \text{Gamma}(a_\phi, b_\phi) \\
\pi(\sigma^2_T\mid a_\sigma, b_\sigma) & \sim \text{InvGam}(a_\sigma, b_\sigma) \\
\pi(\tau^2_T\mid \alpha_T, \beta_T) & \sim \text{InvGam}(\alpha_T, \beta_T) \\
\pi(\alpha_T\mid c_{\alpha}, d_{\alpha}) & \sim \text{InvGam}(c_{\alpha}, d_{\alpha}) \\
\pi(\beta_T\mid c_\beta, d_\beta) & \sim \text{InvGam}(c_\beta, d_\beta)
\end{align*}
\]

where \(a_\phi, b_\phi, a_\sigma, b_\sigma, c_{\alpha}, d_{\alpha}, c_\beta, \) and \(d_\beta\) are known constants and the matrix \(T\) is a diagonal matrix with the design point specific variances, \(\tau^2_T\), as elements. Complete conditional distributions for the above parameters are included in the appendix. In addition, for all variables except \(\alpha_T\) the prior distributions for the parameters are conjugate. This allows for simplified derivation and computation.

3.4 Computation

The number of parameters requiring estimation is clearly intractable. As such, a standard successive substitution MCMC algorithm was used. Due to the conjugacy of nearly all priors, most parameters could be drawn from pre-existing random number generators. For the only parameter without a conjugate prior, \(\alpha_T\), Metropolis-Hastings sampling was incorporated into the algorithm. Over the course of the project over 25000 iterations have been evaluated; MCMC chains for the results presented here reached 5000 iterations. Parameter convergence was evaluated through standard output diagnostics.

4 Results

Using the hierarchical structure outlined above, models were estimated for each supernova in the training sample as well as for the two mean curves. Figure 4 contains the models for the sample of Type Ia light curves shown previously. Models for a sample of core collapse light curves are given in Figure 5. Individual curves are in blue, mean curves, specific to type, are in red, and posterior predictive distributions are given in gray.

---

Figure 4: Type Ia Supernova Light Curves

Models for Type Ia light curves are given in blue with the posterior predictive distributions given in gray. The mean Type Ia curve is shown in red.

Figure 5: Core Collapse Supernova Light Curves

Models for core collapse light curves are given in blue with the posterior predictive distributions given in gray. The mean core collapse curve is shown in red.
As they should be, these models are smooth yet simultaneously adapt to relationships inherent in the data. Modeling in this way successfully recognizes various curve shapes, highlighting one of the process’ strong points. However, modeling differences in curve scale is an aspect that we are still investigating. Current models borrow too much strength from the mean curve which reduces the variation in individual curves. On a more positive note, posterior predictive distributions are successful in capturing a majority of the data, indicating overall the fits of the models are good. Also, on average, Type Ia curves have more pronounced peaks than core collapse curves. This result is consistent with results in previous photometric analysis. Figure 6 highlights this difference with a comparison of mean curves. The

Figure 6

Although there is variation in the mean curves, several locations indicate potential differences in mean curves that may eventually be useful for classification purposes.

Type Ia mean curve is given in black with the core collapse mean curve in red. These mean curves are given on a standardized scale, whereas individual curves are given in the original scale. Although Type Ia curves exhibit significantly more variation, the comparison of mean curves highlights several areas that may lead to eventual differentiation of class type.

These models are a work in progress. Full implementation of the hierarchical structure is still being explored. Additionally, one area that occupies our current efforts is the dilemma of aligning curve peaks before modeling light curves. This procedure would resemble current practices in other photometric classification attempts. In addition, this process would potentially introduce greater differences in mean curves and thus allow for more fluid classifications. However, our previous attempts to first align curve peaks and subsequently estimate curves introduce a discontinuity at the peak of each curve. Future efforts will further investigate this issue.

5 Conclusions and Future Work

Hierarchical Gaussian process formulation allows for flexibility in curve definition coupled with a simultaneous recognition of similarity. Models for supernova light curves obtained from this process are smooth and relatively close approximations of data. In addition, several regions in the flux-phase space yield potential differences in mean light curves that may lead to eventual supernova class differentiation.

Finally, these models have only been classified in a supervised state. Future effort will focus on developing a means of classification for the models fit through the hierarchical Gaussian process outlined above. Specifically, we plan to use a Dirichlet process for classification, much like Dahl did in his 2006 classification of genes. We are hopeful that this unsupervised classification will yield an efficient and accurate means of classifying supernovae.
A Appendix

Complete Conditional Distributions

\[ \theta | \mu, \Sigma, T_i, Y_i, \phi \sim \]
\[ N_m \left( \mathbf{I}^{-1} \mathbf{1} + \phi \mathbf{K}_i \right)^{-1} \left[ \Sigma_i^{-1} Y_i + \mathbf{I}^{-1} \mu + \phi \mathbf{K}_i \right], \]
\[ \left[ \Sigma_i^{-1} + \mathbf{I}^{-1} + \phi \mathbf{K}_i \right]^{-1} \]
\[
\text{for } i = 1, \ldots, k
\]

\[ \mu | \phi, \theta_k, \cdots, \theta_k, T_k, \phi \sim \]
\[ N_m \left[ k T_k^{-1} + (k + 1) \phi \mathbf{K}_k \right]^{-1} \left[ k T_k^{-1} + (k + 1) \phi \mathbf{K}_k \right] \theta, \]
\[ \left[ k T_k^{-1} + (k + 1) \phi \mathbf{K}_k \right]^{-1} \]

where
\[ \theta = \left( \frac{1}{k} \sum_{i=1}^{k} \theta_i(z_1), \cdots, \frac{1}{k} \sum_{i=1}^{k} \theta_i(z_m) \right) \]

\[ \phi | \theta_1, \cdots, \theta_k \sim \]
\[ G \left( \alpha_\phi + \frac{m(1+k)}{2}, \right) \]
\[ \left[ \frac{1}{\beta_\phi} + \frac{1}{2} \mu^T \mathbf{K}_i \mu + \frac{1}{2} \sum_{i=1}^{k} (\theta_i - \mu) \right]^{-1} \]

\[ \sigma_i^2 | \theta_1, \cdots, \theta_k \sim \]
\[ IG \left( \alpha_\sigma + \frac{m}{2}, \left( \frac{1}{\beta_\sigma} + \frac{1}{2} n_i (Y_i - \theta_i)'(Y_i - \theta_i) \right) \right]^{-1} \]
\[ \text{for } i = 1, \ldots, k \]

\[ \tau_j^2 | \theta_1, \cdots, \theta_k, \mu \sim \]
\[ IG \left( \alpha_\tau + \frac{k}{2}, \left( \frac{1}{\beta_\tau} + \frac{1}{2} \sum_{i=1}^{k} (\theta_i - \mu) \right) \right]^{-1} \]
\[ \text{for } j = 1, \ldots, m \]

\[ [\alpha_\tau \tau_1 \cdots \tau_m, \beta_\tau] \sim \]
\[ \frac{1}{\Gamma(\alpha_\tau)^{\frac{m}{2}}} \frac{1}{\beta_\tau^{\frac{m}{2}}} \left( \prod_{j=1}^{m} (\tau_j^2)^{-(\alpha_\tau+1)} \right) \exp \left\{ -\frac{1}{\alpha_\tau} \right\} \]

\[ [\beta_\tau \tau_1 \cdots \tau_m, \alpha_\tau] \sim IG \left( c_\beta + \alpha_\tau m, \left( \frac{1}{d_\beta} + \sum_{j=1}^{m} (\tau_j^2)^{-1} \right)^{-1} \right) \]

References


