SURVIVAL ANALYSIS OF THE GAMMA-RAY BURST DATA

BY

BRADLEY EFRON and VAHE PETROSIAN

TECHNICAL REPORT NO. 414

DECEMBER 1992

PREPARED UNDER THE AUSPICES

OF

NATIONAL SCIENCE FOUNDATION GRANT DMS92-04864

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA
Survival Analysis of the Gamma-Ray Burst Data

By

Bradley Efron and Vahe Petrosian

Technical Report No. 414

December 1992

Prepared Under the Auspices

Of

National Science Foundation Grant DMS92-04864

Also supported by Public Health Service Grants 5 R01 GM21215-17 and CA55325 and issued as Technical Report No. 160, Division of Biostatistics, Stanford University.

Department of Statistics
Stanford University
Stanford, California
Survival Analysis of the Gamma-Ray Burst Data

Bradley Efron and Vahe Petrosian

Abstract

Gamma-Ray bursts are short but powerful pulses of electromagnetic radiation coming from space. Specially equipped satellites have observed the arrival of several hundred bursts during the past two decades. The origin of these bursts, whether from near the earth, across the galaxy, or out of the deep cosmos, has become a major problem for astronomers. At the heart of the question lies a statistical puzzle relating to familiar issues in survival analysis: Kaplan-Meier type curves, data truncation, hazard rate analysis, Mantel-Haenszel statistics, etc. Survival analysis is used here to critically examine two different models that have been proposed for the Gamma-Ray burst data. The analysis contradicts one of these models.
1. Introduction. The gamma-ray burst data has puzzled astronomers for twenty years. Improved methods of observation have deepened the mystery rather than dispelling it. At the heart of the question lies a statistical puzzle relating to familiar issues in survival analysis. Our purpose here is to bring a full arsenal of statistical survival techniques to bear on the burst data. This includes methods familiar to biostatisticians as well as some techniques developed in the astronomy literature.

Gamma-rays are the most energetic form of electromagnetic radiation. Detectors were placed on satellites in the 1960's to monitor violations of the nuclear test-ban treaty. The detectors registered occasional powerful bursts of gamma-rays, but coming from space rather than earth. These were the enigmatic gamma-ray bursts. Hartmann (1992) provides an informative non-technical discussion of the astronomical issues.

Figure 1 shows the data from 486 bursts, a representative fraction of all the burst data collected during the past 20 years. Each observation is summarized by two numbers, \((C_{\text{lim}}, C_p)\), where \(C_p\) is the peak photon count rate (number of photons detected per second) of the burst, and \(C_{\text{lim}}\) is a limiting threshold count rate having to do with the background gamma ray activity. These count rates are a measure of the intensity of the burst and background. In what follows we shall refer to them as intensities. An event is defined as a gamma-ray burst only if \(C_p\) exceeds \(C_{\text{lim}}\). The coordinates in Figure 1 are the natural logarithms

\[
x = \ln(C_{\text{lim}}) \quad \text{and} \quad y = \ln(C_p),
\]

so the pictured data consists of 486 points \((x, y)\) satisfying \(x < y\).

The data in Figure 1 was collected on three different satellite missions. SIGNE, a French-Soviet experiment, was the earliest of the three, operating during the 1970's. SMM, for Solar Maximum Mission, collected gamma-ray data from 1980-1990. The Burst and Transient Source Experiment, BATSE, on board the Compton Gamma-Ray Observatory, has operated since April 1991. It should be noted that the values of \(C_p\) and \(C_{\text{lim}}\) for the three instruments have different relations to the true intensity of the bursts. We show the three sets of data in one Figure for convenience.

The gamma-ray bursts could arise from enormously powerful events occurring far across the cosmos, or from more moderate activity in our own galaxy, perhaps even in regions nearby the earth. The observed spectrum of radiation from some bursts and some theories favor a galactic origin, but observed distribution of the intensities and the position in the celestial sphere do not agree with a galactic origin. Therefore, astronomers are not at all certain what causes the bursts or how inherently powerful they might be. The main purpose of collecting the gamma-ray burst data is to determine, at least roughly, where the bursts are originating. The mystery is that two different aspects of the data mentioned above give two quite different answers.

Most of the ordinary “visible” matter in our galaxy is distributed in the form of a disk with thickness about \(10^3\) light years and a radius exceeding \(4 \times 10^4\) light years. The solar system is
located $2.5 \times 10^4$ light years from the center of this disk. Around the disk there exists a halo with much lower density of stars which appears to extend to or beyond $10^5$ light years.

Suppose that the bursts originate nearby, say from within a few hundred light-years of earth. This suggests the sphere model, in which sources are uniformly distributed in a sphere around us. A competitor is the disk model in which the sources are located in a much larger flat disk. Both models are illustrated by the visible stars: the bright nearby stars are roughly spherically distributed across the heavens, but the great majority of galactic stars, dim and far away, lie in the disk of the milky way. We can call this the sphere-disk model, locally spherical but globally disc-like in distribution.

![Figure 1. 486 gamma-ray bursts; $x = \ln(C_{\text{lim}})$, log of the limiting threshold intensity; $y = \ln(C_p)$, log of the peak intensity. By definition a gamma ray burst has $y > x$. The data comes from 3 satellites: SIGNE(+) 130 bursts; SMM(-), 132 bursts; and BATSE(*), 224 bursts. The units of radiation intensity depend on the equipment used, and differ for the three satellites. This is truncated data, since we cannot observe points below the truncation boundary $y = x$. The relative positions of points from the three different experiments are arbitrary because of definitismal and equipment differences. The simple calculations of the Appendix show that for sources uniformly distributed either spherically or in a disk model, the apparent burst magnitude $y = \ln(C_p)$ follows a shifted and scaled exponential distribution. The density of $y$ under these models is

$$f(y) = \theta e^{-\theta(y-y_0)} \quad \text{for} \quad y > y_0,$$

(1.2)

where $y_0$ is a lower cutoff determined by the sensitivity of the observing instrument. The scale
parameter $\theta$ depends on the model,

$$
\theta = \begin{cases} 
1.5 & \text{for the sphere model} \\
1.0 & \text{for the disk model}
\end{cases} 
$$

We can restate (1.2) in a convenient way by letting $F(y)$ be the right-sided \textit{cumulative distribution function} (cdf),

$$
F(y) = \int_y^\infty f(z)dz = e^{-\theta(y-y_0)} 
$$

and letting $H(y)$ be the \textit{cumulative hazard function}

$$
H(y) = -\ln(F(y)).
$$

Under the exponential model (1.2), $H(y)$ is linear,

$$
H(y) = \theta(y - y_0) \quad \text{for} \quad y > y_0,
$$

with $\theta$ as in (1.3). The \textit{hazard rate} for $y$ is

$$
h(y) = H'(y) = f(y)/F(y) = \theta \quad \text{for} \quad y > y_0.
$$

Estimates of the cumulative hazard rate are shown in Figure 2 for the three satellites, based on the data in Figure 1. This data is \textit{truncated} because of the definitional requirement $y > x$. Observations below the truncation boundary $y = x$ are excluded. \textit{Lynden-Bell's method}, described in Section 2, was used to construct the cumulative hazard estimates $\hat{H}(y)$. This method is closely related to the Kaplan-Meier estimate, but with some interesting differences due to observing truncated data rather than censored data.

The estimates $\hat{H}(y)$ loosely follow the linear form $\hat{\theta}(y - \hat{y}_0)$ except in the sparse upper regions of the observational range. The slopes $\hat{\theta}$ decrease from $\hat{\theta} = 1.38$ for SIGNE to $\hat{\theta} = 1.14$ for SMM to $\hat{\theta} = .838$ for BATSE. The simplest interpretation of this is by the sphere-disk model: the older, less sensitive instrument SIGNE was detecting gamma ray bursts from the nearby sphere, $\theta \simeq 1.5$; the newer, more sensitive instruments were detecting more distant bursts from the galactic disk, $\theta \simeq 1.0$.

Unfortunately there is a fatal objection to this scenario. If BATSE is detecting bursts mainly from the galactic disk, then the angular directions of the burst should look like the disk of the milky way, occupying a narrow band across the sky. Instead, the angular distribution appears to be uniform in the sky, just what we would expect from the sphere model. Meegen et al. (1992) show that the angular directions pass the usual tests for directional uniformity. We are left in the uncomfortable position of the angular data verifying the sphere model while the apparent magnitudes $y$ support the sphere-disk model.
Figure 2. Cumulative hazard estimates $\hat{H}(y)$ for SIGNE (dashed line), SMM(line), and BATSE (dots); based on the data in Figure 1, using Lynden-Bell's method, as explained in Section 2. The slope $\theta = 1.38$ is near the sphere-model value 1.5 for SIGNE, decreasing to $\hat{\theta} = .84$ for BATSE, near the disk-model value 1.0.

This contradiction has led to two other alternative suggestions as to the origin of the bursts. In one model the bursts originate from the halo but primarily from distances larger than the distance of the solar system from the center of the galaxy so that they appear isotropic. In this case the decrease of slope $\hat{\theta}$ mentioned above most probably reflects a decrease in the density of bursts with distance from the center. In a second model it is assumed that the bursts originate in distant galaxies in the universe which naturally explains their isotropic distribution. In this case we expect the slope $\theta$ to be 1.5 for nearby sources and to begin to decrease due to cosmological effects such as expansion, curvature of space, etc.

Does the data in Figure 1 really support the sphere-disk model? Is there compelling evidence for the existence of slope 1.5 required in the cosmological model? Section 4 presents a hazard rate analysis of the data that casts doubt on this conclusion. This analysis is based on the method of partial logistic regression, originally developed for censored data in Efron (1988). The average hazard calculation of Section 6 shows that $\theta$ is definitely less than the disk value 1.0 for the faintest bursts.
Section 5 uses Mantel-Haenszel, or log-rank tests to combine the data from the three instruments. This is a speculative tactic, but one that lets us compare the hazard rates over a very wide range of apparent magnitudes \( y \). The conclusions above, of a hazard rate declining linearly from about \( \theta = 1.5 \) to below \( \theta = 1.0 \), are particularly clear in the combined analysis.

2. Lynden-Bell's Method for Truncated Data. The cumulative hazard plots for \( y = \ln(C_p) \) in Figure 2 are based on the method of Lynden-Bell (1971). This is a clever nonparametric technique for estimating the hazard rate, density, or cdf of a random variable observed subject to data truncation. It is closely related to the Kaplan-Meier method for censored data as discussed below. Woodroofe (1985) gives a careful mathematical treatment of the ideas involved.

Lynden-Bell's method assumes that some mechanism is producing random points \((x, y)\) with \( x \) independent of \( y \). We are interested in estimating the distribution of \( y \). However we only get to observe those \((x, y)\) pairs where \( y \) lies above a truncation boundary \( t(x) \),

\[ y \geq t(x). \quad (2.1) \]

In Section 1, \( x = \ln(C_{\text{lim}}), y = \ln(C_p), \) and \( t(x) = x \). We label the observed pairs \((x_i, y_i)\) for \( i = 1, 2, \ldots, N \), so \( N = 224 \) for BATSE in Figure 1. For convenience we assume that there are no ties among the \( y \) values.

Lynden-Bell's method depends on the notion of comparable points. For each value of \( j = 1, 2, \ldots, n \), the number \( N_j \) of comparable points is defined as

\[ N_j = \#\{(x_i, y_i) : y_i \geq y_j \text{ and } t(x_i) \leq y_j\}. \quad (2.2) \]

(Note that Lynden-Bell's definition of comparable points \( C^- = N_j - 1 \) does not include the point \((x_i, y_i)\).) Figure 3 shows a simplified example with \( N = 8 \) points. For \( j = 2 \) we see that \( N_2 = 4 \). The picture has \( t(x) \) increasing in \( x \), as will be assumed throughout the paper, but this is not a necessity of the theory. It is sufficient that each \( x_i \) determines a lower truncation point \( t_i \), below which \( y_i \) cannot be observed. Definitions like (2.2) remain valid, replacing \( t(x_i) \) with \( t_i \).

The hazard rate \( h(y) \) can be estimated by a discrete distribution putting mass \( \hat{h}_j \) on \( y = y_j \), where

\[ \hat{h}_j = \ln(1 - \frac{1}{N_j}) \approx \frac{1}{N_j - \frac{1}{2}}. \quad (2.3) \]

The cumulative hazard rate and right-sided cdf estimates are

\[ \hat{H}(y_j) = \hat{H}_j = \sum_{i \leq j} \hat{h}_i \quad \text{and} \quad \hat{F}(y_j) = \hat{F}_j = e^{-\hat{H}_j}. \quad (2.4) \]

See remarks C and D of Efron (1977). Figure 2 is based on definition (2.4). The fitted straight lines in Figure 2 are least absolute deviation fits to the points \((y_j, \hat{H}_j)\).
Figure 3. A simplified example with $N = 8$ points $(x_i, y_i)$. The number of points comparable to $(x_2, y_2)$ is $N_2 = 4$, this being the number of points in the comparability rectangle, including $(x_2, y_2)$ itself.

Definitions (2.3), (2.4) are essentially the same as those used in the Kaplan-Meier estimate. See Section 1.3 of Kalbfleisch and Prentice (1987) for example. In the Kaplan-Meier setting $N_j$ is the number of points at risk just before the $j$th event. The Lynden-Bell estimator uses the same idea, working from the smallest value of $y$ upwards. The only difference is that points with $y_i > y_j$ but $t(x_i) > y_j$ are not at risk in the truncated case, because these points can not be observed to “fail” during an infinitesimal interval including $y_j$. Notice that the $N_j$ do not necessarily decrease with $j$ in the truncated case.

Petrosian (1992) first suggested Lynden-Bell's estimator as a method of dealing with truncation in the gamma-ray burst data. Previous workers worked instead with the ratio $C_p/C_{\lim}$. The Lynden-Bell approach is more flexible and powerful, and leads to the other survival analysis techniques used in this paper. However it does involve a significant assumption: that $x$ and $y$, in this case $ln(C_{\lim})$ and $ln(C_p)$, are independently distributed variates. Section 3 shows how this assumption can be tested.

3. Independence Tests For Truncated Data. The methods used in this paper assume that $x = ln(C_{\lim})$ is independent of $y = ln(C_p)$. This is a reasonable assumption physically, but one we would like to test on the observed data. Efron and Petrosian (1992), following work by
Bhattacharya, Chernoff, and Yang (1983), proposed a class of permutation tests for independence applicable to the situation where the \((x, y)\) pairs are observed subject to truncation. We will apply one such test to the three data sets of Figure 1.

Returning to definition (2.2), let \(R_j\) be the rank of \(x_j\) among the \(x\)-values of the \(N_j\) comparable points. This can be expressed as

\[
R_j = \#\{(x_i, y_i): y_i \geq y_j \text{ and } x_i \leq x_j\}. \tag{3.1}
\]

In the example of Figure 3, \(R_2 = 2\) since \(x_2\) has the second smallest \(x\)-value among the 4 comparable points. Under \(H_0\), the hypothesis of independence, \(R_j\) has probability \(1/N_j\) of equalling 1, 2, \(\cdots\), \(N\). To state this more precisely, let \(x\) be the vector of ordered \(x\)-values, and likewise let \(y\) be the ordered vector of the \(y\)-values. Then under \(H_0\) the ranks \(R_1, R_2, \cdots, R_n\) are conditionally independent given \(x\) and \(y\), and

\[
\text{Prob}_{H_0}\{R_j = r|x, y\} = 1/N_j \quad \text{for } r = 1, 2, \cdots, N_j. \tag{3.2}
\]

Distribution (3.2) has mean and variance

\[
E_j = (N_j + 1)/2 \quad \text{and} \quad V_j = (N_j^2 - 1)/12. \tag{3.3}
\]

We will base our tests of independence on the statistic

\[
t = \sum_{j=1}^{n}(R_j - E_j)/\sqrt{\sum_{j=1}^{n}V_j}. \tag{3.4}
\]

Under \(H_0\), \(t\) has mean 0 and variance 1, and will be nearly standard normal \((N(0,1))\), for the kind of sample sizes we are dealing with here. If there is no data truncation, \(t\) is a version of Kendall’s tau statistic, a point discussed further below. Other truncated-data test statistics for independence are discussed in Efron and Petrosian (1992), but (3.4) has the advantage of staying the same when the roles of the \(x\) and \(y\) axes are interchanged in the definitions.

<table>
<thead>
<tr>
<th></th>
<th>SIGNE</th>
<th>SMM</th>
<th>BATSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(t) (3.4):</td>
<td>2.07*</td>
<td>-0.06</td>
<td>0.60</td>
</tr>
<tr>
<td>(\hat{\delta}) (3.5):</td>
<td>.561</td>
<td>.498</td>
<td>.512</td>
</tr>
<tr>
<td>(\hat{\rho}) (3.6):</td>
<td>-1.89</td>
<td>.005</td>
<td>-.036</td>
</tr>
</tbody>
</table>

Table 1. The statistic \(t\), (3.4), applied to the three data sets of Figure 1. \(H_0\) is accepted for SMM and BATSE but rejected for SIGNE at a one-sided .02 level. Dissonance statistic \(\hat{\delta}\), (3.5), and correlation estimate \(\hat{\rho}\), (3.6), suggest that the correlation between \(x\) and \(y\) is quite small, even for SIGNE. BATSE calculation includes a correction for ties in the \(x\)-values.

The top line of Table 1 shows the statistic, \(t\), (3.4), calculated for the three data sets of Figure 1. SMM and BATSE are insignificant on a \(N(0,1)\) scale, so we can accept the hypothesis of independence. However the SIGNE data rejects \(H_0\) at a one-sided .02 level. Note: The \(x\)-values
for BATSE have a large proportion of ties, so that the possible values of \( r \) in (3.2) are the tied ranks rather than \( 1, 2, \cdots, N_j \). This reduces the variance \( V_j \) but doesn't change the mean \( E_j \) in (3.3). For the BATSE data, the values of \( V_j \) used in (3.4) were the actual permutation variances of the tied ranks, rather than \((N_j^2 - 1)/12\).

We have evidence of non-zero correlation between \( x \) and \( y \) in the SIGNE data, but how large is the correlation? Line 2 of Table 1 gives an estimate of correlation, the \textit{dissonance statistic}

\[
\hat{\delta} = \frac{\sum_{j=1}^{n}(R_j - 1)}{\sum_{j=1}^{n}(N_j - 1)}. \tag{3.5}
\]

Two pairs \((x_j, y_j)\) and \((x_i, y_i)\) are dissonant if \( y_i > y_j \) and \( x_i < x_j \). The two pairs are comparable if (assuming for notational covariance that \( y_i > y_j \)) we have \( t(x_i) < y_j \), or equivalently \( x_i < t^{-1}(y_j) \). Looking at definitions (2.2) and (3.1), we see that \( \hat{\delta} \) is the proportion of comparable pairs that are dissonant. Kendall's rank correlation coefficient equals \( 1 - \hat{\delta} \) in the situation where there is no data truncation. See Section 2 of Hajek (1969).

Under \( H_0 \), \( \hat{\delta} \) is an almost unbiased estimate of \( \delta = .5 \), there being equal probability of dissonance or consonance. If \( x \) and \( y \) are correlated then \( \hat{\delta} \) estimates \( \delta > .5 \) of the correlation is negative and \( \delta < .5 \) of the correlation is positive. In the case where \((x, y)\) is bivariate normal, and there is no data truncation, the Pearson correlation coefficient \( \rho \) is given by

\[
\rho = \sin[\pi(\frac{1}{2} - \delta)]. \tag{3.6}
\]

The estimate for SIGNE is \( \hat{\rho} = \sin[\pi(\frac{1}{2} - \hat{\delta})] = -0.189 \), indicating only a small negative correlation, even though it is significantly non-zero.

If correlation is a real concern, we could base our analyses on the transformed data

\[(X, y) = (x - cy, y) \tag{3.7}\]

for some choice of the constant \( c \), rather than on the original \((x, y)\) pairs. The choice \( c = -.03 \) makes the \( t \) statistic (3.4) approximately 0 for the SIGNE data. The SIGNE cumulative hazard estimate in Figure 2 was recalculated after this change. The resulting estimate \( \hat{F}(y) \) was almost indistinguishable from the dashed curve in Figure 2. Correction (3.7) was not used for any other analysis in this paper.

4. Hazard Rate Modelling. Astronomers are interested in certain hazard rate models for the distribution of \( y = \ln(C_p) \). In the galactic origin (or sphere-disk) model the hazard rate should have an upper asymptote of 1.5 for large values of \( y \), a lower asymptote of 1.0 for small values of \( y \), with a smooth transition between 1.5 and 1.0 at some intermediate interval on the \( y \) scale. This section tests various hazard rate models for the data in Figure 1, using the method of partial logistic regression developed in Efron (1988).
Table 2 shows the first 10 and last 2 bins of a discretization of the BATSE data into 101 bins of width .05. The \( k \)th bin, \( y \in (l_{k}, u_{k}) \), contains \( s_{k} \) of the BATSE \( y \)-values, for example \( s_{10} = 6 \), so \( \sum_{k} s_{k} = 224 \), the total number of observed BATSE bursts. The column labelled \( n \) is the number of comparable points for the \( k \)th bin. This is an averaged version of definition (2.2), averaging over the bin width,
\[
n_{k} = \#\{(x_{i}, y_{i}) : y_{i} \geq l_{k} \text{ and } t(x_{i}) \leq l_{k}\} + \frac{1}{2}\#\{(x_{i}, y_{i}) : l_{k} \leq t(x_{i}) < u_{k}\}
\]
(4.1)
For the gamma-ray burst data, \( t(x) = x \).

The discrete hazard rate \( \pi_{k} \) for the \( k \)th bin is the probability that \( y \) falls into that bin given that \( y \) exceeds the bin’s lower limit,
\[
\pi_{k} = \text{Prob}\{y \in (l_{k}, u_{k}) | y > l_{k}\}. \tag{4.2}
\]
this is related to the continuous hazard rate (1.7) by
\[
e^{-\int_{l_{k}}^{u_{k}} h(y)dy} = 1 - \pi_{k}, \tag{4.3}
\]
or equivalently
\[
h_{k} = -[\ln(1 - \pi_{k})]/w_{k}, \tag{4.4}
\]

<table>
<thead>
<tr>
<th>bin</th>
<th>lo</th>
<th>up</th>
<th>s</th>
<th>n</th>
<th>( \hat{\pi} )</th>
<th>( \hat{h} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.45</td>
<td>5.50</td>
<td>1</td>
<td>17.0</td>
<td>0.059</td>
<td>1.21</td>
</tr>
<tr>
<td>2</td>
<td>5.50</td>
<td>5.55</td>
<td>2</td>
<td>31.0</td>
<td>0.065</td>
<td>1.33</td>
</tr>
<tr>
<td>3</td>
<td>5.55</td>
<td>5.60</td>
<td>3</td>
<td>96.0</td>
<td>0.031</td>
<td>0.63</td>
</tr>
<tr>
<td>4</td>
<td>5.60</td>
<td>5.65</td>
<td>5</td>
<td>160.0</td>
<td>0.031</td>
<td>0.63</td>
</tr>
<tr>
<td>5</td>
<td>5.65</td>
<td>5.70</td>
<td>7</td>
<td>174.0</td>
<td>0.040</td>
<td>0.82</td>
</tr>
<tr>
<td>6</td>
<td>5.70</td>
<td>5.75</td>
<td>7</td>
<td>188.5</td>
<td>0.037</td>
<td>0.76</td>
</tr>
<tr>
<td>7</td>
<td>5.75</td>
<td>5.80</td>
<td>3</td>
<td>184.5</td>
<td>0.016</td>
<td>0.33</td>
</tr>
<tr>
<td>8</td>
<td>5.80</td>
<td>5.85</td>
<td>3</td>
<td>187.5</td>
<td>0.016</td>
<td>0.32</td>
</tr>
<tr>
<td>9</td>
<td>5.85</td>
<td>5.90</td>
<td>4</td>
<td>191.0</td>
<td>0.021</td>
<td>0.42</td>
</tr>
<tr>
<td>10</td>
<td>5.90</td>
<td>5.95</td>
<td>6</td>
<td>188.5</td>
<td>0.332</td>
<td>0.65</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td>2</td>
<td>2.0</td>
<td>0.000</td>
<td>0.00</td>
</tr>
<tr>
<td>101</td>
<td>10.45</td>
<td>10.50</td>
<td>2</td>
<td>2.0</td>
<td>1.300</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Table 2. the first 10 and last two bins of the BATSE data, discretized into bins of width .05; \( s_{5} = 7 \) of the BATSE \( y \)-values lie in bin 5, \( y \in (5.65, 5.70) \); \( n_{5} = 174.0 \) comparable points for bin 5, definition (4.1); discrete hazard rate estimate \( \hat{\pi}_{5} = s_{5}/n_{5} = .020 \) corresponds to continuous hazard rate \( \hat{h}_{5} = -[\ln(1 - \pi_{5})]/\cdot 5 = 0.82 \).

where \( w_{k} = u_{k} - l_{k} \) and \( h_{k} = \int_{l_{k}}^{u_{k}} h(y)dy/w_{k} \). Table 2 shows the nearly unbiased estimates \( \hat{\pi}_{k} = s_{k}/n_{k} \) for the \( \pi_{k} \), and the corresponding \( h_{k} \) estimates
\[
\hat{h}_{k} = -[\ln(1 - \pi_{k})]/.05 = -20 \cdot \ln(1 - s_{k}/n_{k}). \tag{4.5}
\]
In a partial logistic regression the discrete hazards $\pi_k$ are modelled as smooth functions on the $y$-scale. For example we might use a cubic model

$$\logit(\pi_k) = \log\left(\frac{\pi_k}{1 - \pi_k}\right) = \beta_0 + \beta_1 Y_k + \beta_2 Y_k^2 + \beta_3 Y_k^3,$$  \hfill (4.6)

where $Y_k = (\text{lo}_k + \text{up}_k)/2$, the midpoint of the $k$th bin. The model is fit to the data $(\text{lo}_k, n_k)$ using a standard logistic regression program, assuming that the $s_k$ are independent binomial observations, with $n_k$ draws and success probability $\pi_k$,

$$s_k \sim \text{bi}(n_k, \pi_k) \quad \text{independently for} \quad k = 1, 2, \ldots, K = 101.$$  \hfill (4.7)

Efron (1988) discusses why assumption (4.7) is applicable, the reasoning being similar to that for the proportional hazards model.

Four polynomial models were fit to the data of Figure 1: the constant hazard rate model $\logit(\pi_k) = \beta_0$, the linear model

$$\logit(\pi_k) = \beta_0 + \beta_1 Y_k,$$  \hfill (4.8)

the quadratic model $\logit(\pi_k) = \beta_0 + \beta_1 Y_k + \beta_2 Y_k^2$, and the cubic model (4.6). Each model was fit to each of the three data sets, and also to the combined version of the data described in Section 5. As in a standard logistic regression, each fit gives maximum likelihood estimates for the $\beta$ parameters, estimates of their standard errors, and also a maximized log likelihood based on the binomial model (4.7).

Table 3 shows twice the log-likelihood differences for successive models, say $D = 2 \times (\hat{L}_1 - \hat{L}_0)$. Under the null hypothesis that the larger model is no improvement over the smaller model, $D$ has an approximate chi-square distribution with one degree of freedom. We can compare the differences seen in Table 2 with the appropriate chi-square percentiles $D^{(0.90)} = 2.71$, $D^{(0.95)} = 3.84$, $D^{(0.99)} = 6.63$.

We see that the difference between the linear model (4.8) and the model of constant hazard rate is quite significant, except for SMM. The estimated linear coefficients $\hat{\beta}_1$ were positive in all four cases, indicating a greater hazard rate at the high end of the $y$ scale. Going on to quadratic or cubic models gave no further significant improvements. The linear model leads to a Gompertz distribution for $y$, see equation (5.8) of Efron (1988).

The left panel of Figure 4 graphs the fitted linear model $\logit(\pi) = \hat{\beta}_0 + \hat{\beta}_1 Y$ for the combined data. The dashed lines are the discrete hazard rates corresponding to the sphere and disk models (1.2), (1.3),

$$\pi_{\text{sphere}} = .0723 \quad \text{and} \quad \pi_{\text{disk}} = .0488,$$  \hfill (4.9)

11
Table 3. Likelihood analysis based on partial logistic regressions: tabulated quantity is $2 \cdot (\hat{\ell}_1 - \hat{\ell}_0)$, where $\hat{\ell}_1$ and $\hat{\ell}_0$ are the maximized binomial log likelihoods (4.7), larger and smaller models respectively; “constant” is model $\text{logit}(\pi_k) = \beta_0$, “linear” is $\text{logit}(\pi_k) = \beta_0 + \beta_1 Y_k$ etc., as in (4.6). Linear model is a strongly significant improvement over model with constant hazard rate, except for SMM data; no other significant differences. Significance of two-level model is discussed below. Method of combining the three data sets is explained in Section 5. All of these results based on bin width .05, as in Table 2. Using smaller bin widths gave similar results.

obtained from $1 - \pi = \exp\{-.05\theta\}$, (4.3). The fitted hazard rate $\hat{\pi}$ is below $\pi_{\text{disk}}$ at the low end of the $y$ scale, increasing above $\pi_{\text{sphere}}$ at the high end. The calculations of Section 6 strongly support a hazard rate less than the disk value 1.0 for small values of $y$.

Figure 4's left panel also shows the best-fitting two-level model for the combined data. As suggested by the sphere-disk model, the two-level model assigns two values to the discrete hazard rate (4.2),

Figure 4. Left panel: fitted discrete hazards $\pi_k$ for the combined data; linear model (4.8), light solid line; two-level model (4.1) with $\hat{k} = 84$, heavy solid line; dashed lines indicate values of the discrete hazard corresponding to the sphere and disk models. Right panel: maximized log-likelihood for the two-level model as a function of the break point; lower dashed line is $3.84/2$ below maximum, the .05 significance level. The likelihood function is strongly multimodal. The horizontal axis in both panels refers to the adjusted burst magnitude $\bar{y}$ in (5.3).
logit(π_k) = β_0 \quad \text{for} \quad k \leq k_0 \quad (4.10)

and

logit(π_k) = β'_0 \quad \text{for} \quad k \geq k_0 + 9,

with linear interpolation on the logit scale between k_0 and k_0 + 9.

The values of β_0, β'_0 and the break-point \( \hat{k}_0 \) used in Figure 4 was estimated by maximum likelihood. Each choice of k_0 gives a log-likelihood \( \hat{L}(k_0) \) obtained by maximizing the binomial likelihood (4.7) over the choice of \( β_0 \) and \( β_1 \) in (4.10). The right panel of Figure 4 graphs \( \hat{L}(k_0) \) as a function of the break-point \( Y_{k_0} \). The maximizing value \( \hat{k}_0 \) was the one used in the left panel. The choice of interpolation constant 9 in (4.10), which wasn’t crucial to the analysis, was big enough to smooth out local irregularities in the likelihood function \( \hat{L}(k_0) \).

The likelihood function \( \hat{L}(k_0) \) is distinctly non-monotonic. The secondary peak at \( Y_{k_0} = 3.4 \) almost reaches the .05 significance level \( \hat{L}(k_0) = 3.84/2 \). The same analysis applied to the separate data sets gave similarly multi-peaked likelihoods for SMM and BATSE. SIGNE showed a single strong maximum at \( Y = 5.15 \).

The two-level model is not contained in the polynomial models. This makes it difficult to interpret the maximized log-likelihoods seen in Table 3. For the combined data, the linear model and the two-level model achieve almost the same maximum likelihood. How can we choose between them?

Table 4 is an argument against the two-level model. Linear logistic regressions (4.7), (4.8) were fit to portions of the discretized combined data: just those bins k with \( k_1 \leq k \leq k_2 \). The two-level model in the left panel of Figure 4 has its break-point at \( \hat{k}_0 = 84 \), the total number of bins for the combined data being \( K = 137 \). Suppose that this model is correct. Then if we choose \( k_2 \leq 84 \), the linear coefficient \( β_1 \) in (4.8) should be zero. Similarly, \( β_1 \) should be zero if we choose \( k_1 \geq \hat{k}_0 + 9 = 93 \). Instead, Table 4 shows positive values of \( \hat{β}_1 \) for most choices of \( k_1 \) and \( k_2 \). The two-level model does not seem believable, at least not for the combined data.

<table>
<thead>
<tr>
<th>( k_1 )</th>
<th>( k_2 )</th>
<th>( \hat{β}_1 ) (se)</th>
<th>( k_1 )</th>
<th>( k_2 )</th>
<th>( \hat{β}_1 ) (se)</th>
<th>( k_1 )</th>
<th>( k_2 )</th>
<th>( \hat{β}_1 ) (se)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>.07 (.39)</td>
<td>120</td>
<td>137</td>
<td>1.20 (1.78)</td>
<td>60</td>
<td>80</td>
<td>-.31 (.32)</td>
</tr>
<tr>
<td>1</td>
<td>40</td>
<td>.25 (.15)</td>
<td>100</td>
<td>137</td>
<td>.77 (.42)</td>
<td>50</td>
<td>90</td>
<td>-.01 (.11)</td>
</tr>
<tr>
<td>1</td>
<td>60</td>
<td>.19 (.07)</td>
<td>80</td>
<td>137</td>
<td>.41 (.15)</td>
<td>40</td>
<td>100</td>
<td>.24 (.08)</td>
</tr>
<tr>
<td>1</td>
<td>80</td>
<td>.13 (.04)</td>
<td>60</td>
<td>137</td>
<td>.28 (.10)</td>
<td>30</td>
<td>110</td>
<td>.15 (.06)</td>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td>.16 (.03)</td>
<td>40</td>
<td>137</td>
<td>.23 (.06)</td>
<td>20</td>
<td>120</td>
<td>.14 (.04)</td>
</tr>
<tr>
<td>1</td>
<td>120</td>
<td>.16 (.03)</td>
<td>20</td>
<td>137</td>
<td>.15 (.04)</td>
<td>10</td>
<td>130</td>
<td>.15 (.03)</td>
</tr>
</tbody>
</table>

Table 4. Linear logistic regressions (4.7), (4.8) applied just to bins k with \( k_1 \leq k \leq k_2 \); combined data, bin width .05, \( K = 137 \) bins; best-fit two-level model on the left of Figure 4 has break at \( \hat{k}_0 = 84 \). The estimated slopes \( \hat{β}_1 \) are generally positive for all choices of \( k_1 \) and \( k_2 \), disagreeing with the two-level model.
5. Combining The Three Experiments. The analysis in Section 4 referred to a combined version of the data that incorporates the results from all three satellites. A simple model was used to combine the data in Figure 1. The parameters of this model were fit by use of the Mantel-Haenszel or log-rank test, as explained below.

We can imagine the same gamma-ray burst being observed by both BATSE and SMM, giving measurements \((x, y)\) and \((\tilde{x}, \tilde{y})\) respectively. The physical model assumes that there are two constants \(A\) and \(B\) such that

\[
\tilde{x} = x - B \quad \text{and} \quad \tilde{y} = y - A.
\]  
(5.1)

The constant \(A\) allows for a difference in the two instruments' sensitivities. BATSE is more sensitive than SMM, so \(A\) is positive. The constant \(B\) allows for differences between the two instruments that affect the setting for the threshold sensitivity. Among these are the sensitivity, cosmic and background intensities, etc. \(C_{\text{lim}}\) and \(C_p\) are multiplied by \(e^{-A}\) when observed by SMM rather than BATSE, (1.1). Notice that

\[
\tilde{y} - \tilde{x} = y - x - C, \quad C = A - B.
\]  
(5.2)

If \(y - x\) is less than \(C\), then \(\tilde{y} - \tilde{x} < 0\), which means the burst would not be counted by SMM. The Mantel-Haenszel analysis described later gives the estimate \(C = 2.10\). Taken literally this means that the BATSE stars in Figure 1 lying below \(y = x + 2.10\), 81% of the BATSE data, would not have been counted as gamma-ray bursts by SMM. All of the constants \((A, B)\) for combining BATSE, SMM, and SIGNE, are given in Table 5.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C = A - B</th>
</tr>
</thead>
<tbody>
<tr>
<td>BATSE-SMM:</td>
<td>2.82</td>
<td>.723</td>
<td>2.10</td>
</tr>
<tr>
<td>SMM-SIGNE:</td>
<td>1.58</td>
<td>.252</td>
<td>1.33</td>
</tr>
<tr>
<td>BATSE-SIGNE:</td>
<td>4.40</td>
<td>.975</td>
<td>3.43</td>
</tr>
</tbody>
</table>

Table 5. Combination constants \((A, B)\), (5.1), obtained from the Mantel-Haenszel analysis. The values of \(B\) imply that 81% of the BATSE bursts would have been missed by SMM, and 72% of the SMM bursts would have been missed by SIGNE.

The combined data set is defined in terms of the constants in Table 5. Write the data in Figure 1 as \((x_{Sj}, y_{Sj})\), where \(S\) indexes the three satellites SIGNE, SMM, BATSE, and \(j\) indexes the individual bursts for each satellite. The combined data set used in Table 3 and Figure 4 is

\[
(\tilde{x}_{Sj}, \tilde{y}_{Sj}) = \begin{cases} 
(x_{Sj}, y_{Sj}) & \text{for } S = \text{SIGNE}, \; j = 1, 2, \ldots, 130 \\
(x_{Sj} - .252, y_{Sj} - 1.58) & \text{for } S = \text{SMM}, \; j = 1, 2, \ldots, 132 \\
(x_{Sj} - .975, y_{Sj} - 4.40) & \text{for } S = \text{BATSE}, \; j = 1, 2, \ldots, 223, 
\end{cases}
\]  
(5.3)
One outlying BATSE point was dropped from consideration here, the point represented by the star at (4.96, 5.47) in Figure 1. Altogether there are 485 points in the combined data set.

The physical model underlying (5.1) is very simple: that the less sensitive instrument is recording only a certain percentage of the bursts recorded by the more sensitive instrument, say \(100 \cdot p\) percent, these being the most obvious ones. "Most obvious" means those points lying furthest above the truncation boundary \(y = z\).

Let \((x_j, y_j)\) represent the burst measurements for the more sensitive instruments and \((X_k, Y_k)\) the measurements for the less sensitive instrument. The constants \(A\) and \(B\) were obtained as follows. (i) Choose a trial value of \(C\); (ii) define

\[
J = \{ j : y_j - x_j > C \};
\]  

(5.4)

(iii) compute \(A = \min_J \{ y_j - C \} - \min \{ Y \}, \) (iv) set \((\bar{x}_j, \bar{y}_j) = (x_j - C - A, y_j - A)\); (v) run a two-sample Mantel-Haenszel (log-rank) test comparing the set of numbers \(\{ \bar{y}_j, y \in J \}\) with the set \(\{ Y \}\); (vi) vary the trial value of \(C\) until the Mantel-Haenszel statistic equals 0.

Miller (Section 2.2, 1981), gives a nice treatment of the Mantel-Haenszel statistic, including its application to survival analysis. In terms of his notation, the numbers at risk before the \(k\)th event, \(n_{k1}\) and \(n_{k2}\), are determined in the gamma-ray context by definitions like (2.3) that take truncation into account.

Having Chosen \(C\), the model suggests that the less sensitive instrument is recording the most obvious \(100p\) percent of the better instrument's bursts, where, in terms of (5.4),

\[
p = \#J/\#\{(x_j, y_j)\}.
\]  

(5.5)

In the SMM-BATSE case, \#J = 43 and \#\{(x_j, y_j)\} = 223, so \(p = .19\). This would be the end of the comparison story, except that we also allow the two instruments to differ by a translation constant \(A\) on both the \(x\) and \(y\) scales. Step (iii) estimates \(A\) by equalizing the left end-points of the two data sets \(\{ Y \}\) and \(\{ \bar{y}_j, j \in J \}\). This works well in our context because the exponential shape of the densities allows their lower endpoint to be accurately located.

Figure 5 shows the Mantel-Haenszel test statistics as a function of \(C\). It is clear that the estimate \(C = 1.33\) for SMM versus SIGNE and \(C = 2.10\) for BATSE versus SMM do not enjoy a great deal of statistical accuracy. Nevertheless they produce quite reasonable-looking results.

The cumulative hazard plots for the three satellites appear very different in Figure 2. Figure 6 shows the same three curves as in Figure 2, except that they have been translated in accordance with model (5.3). Now the hazard plots for the three satellites look quite compatible.
Figure 5. Mantel-Haenszel test statistics for SMM versus SIGNE (left panel) and BATSE versus SMM (right panel). The estimated zero-values $C = 1.33$ and $C = 2.10$ are subject to considerable statistical error.

The horizontal axis in Figure 6 is $\tilde{y}_{S_j}$, the *adjusted burst magnitude* defined in (5.3). The *adjusted hazard* $\tilde{H}_{S_j}$ is plotted vertically, where $\tilde{H}_{S_j}$ is defined in terms of the cumulative hazards $\hat{H}_{S_j}$ for the three satellites, shown in Figure 2,

$$
\tilde{H}_{S_j} = \begin{cases} 
\hat{H}_{S_j} & \text{for } S = BATSE \\
\hat{H}_{S_j} + 1.75 & \text{for } S = SM\text{M} \\
\hat{H}_{S_j} + 3.24 & \text{for } S = SIGNE .
\end{cases}
$$

(Actually the $\hat{H}$ estimate for BATSE used in (5.6) differs from that in Figure 2 because of the deletion of the one outlying point in (5.3).)

The adjusted magnitude scale $\tilde{y}$ in (5.3) shifts the BATSE magnitudes 2.82 units left of those for SMM. The first SMM adjusted magnitude, say $\tilde{y}_{SMM,1}$, occurs at a point where the BATSE cumulative hazard $\hat{H}$ equals 1.75. The definition in (5.6) makes $\tilde{H}_{SMM}$ equal $\tilde{H}_{BATSE}$ at this point. In fact $\tilde{y}_{SMM,1}$ occurs at the 81st percentile of the BATSE magnitudes. This means that definition (5.6) is equivalent to taking the SMM distribution to be the upper 19% of the BATSE distribution, as our model suggested. A similar argument applies to the upward shift of 3.24 applied to $\tilde{H}_{SIGNE}$ in (5.6).
Figure 6. Cumulative hazards when the three data sets are made comparable; same curves as in Figure 2, but translated to fit each other according to (5.3). Here, unlike Figure 2, the three curves look quite compatible. Horizontal axis is adjusted burst magnitude $\bar{y}_{sj}$, (5.3); vertical axis is adjusted hazard $\bar{H}_{sj}$, (5.6).

6. Cumulative Event Rates. Figure 4 suggests that the hazard rate for $y = \ln(C_p)$ is less than the disk-model value 1.0 at the low end of the $y$ scale, and greater than the sphere-model value 1.5 at the high end of the $y$ scale. Is this really true? Here we will use another method, cruder than the partial logistic regressions but also less dependent on parametric assumptions, to answer this question. We will look at the data separately for the three satellites in order to avoid the uncertainties associated with combining the data.

Let $(s_k, n_k)$ be the discretized data defined at the beginning of Section 4, as shown for BATSE in Table 2. The $k$th cumulative event rate from below is defined to be

$$P_k = \sum_{j \leq k} s_k / \sum_{j \leq k} n_k.$$  (6.1)

The binominal model (4.7) says that $P_k$ is an unbiased estimate of the average hazard rate

$$\pi_k = \sum_{j \leq k} n_j \pi_j / \sum_{j \leq k} n_j,$$  (6.2)

with standard error

$$\sigma_k = [\sum_{j \leq k} n_j \pi_j (1 - \pi_j) / (\sum_{j \leq k} n_j)^2]^{1/2}.$$  (6.3)
Because of Jensen’s inequality, $\sigma_k$ is less than the standard error $\sigma_k^{\uparrow}$ applying to the case where for $j \leq k$ all the $\pi_j$ equal $\pi_k^{\uparrow}$,

$$\sigma_k^{\uparrow} = [\pi_k^{\uparrow}(1 - \pi_k^{\uparrow})/\sum_{j \leq k} n_j]^{1/2} \quad (6.4)$$

We can also define cumulative event rates from above, say

$$P_k = \sum_{j \geq k} s_k / \sum_{j \geq k} n_k, \quad (6.5)$$
making the obvious changes in definitions (6.2)–(6.4).

Figure 7 shows the cumulative event rates $P_k$ for SIGNE, SMM, and BATSE, from below on the left and from above on the right. The horizontal axis is the proportion of the events involved, say

$$\text{prop}_k = \sum_{j \leq k} s_j / N \quad \text{or} \quad \text{prop}_k = \sum_{j \geq k} s_j / N, \quad (6.6)$$

for (6.1) or (6.5) respectively. Here $N$ is the total number of events, $N = 130$ for SIGNE, $N = 132$ for SMM, $N = 224$ for BATSE. The dotted curves in Figure 7 show $P_k \pm 1.645\sigma_k^{\uparrow}$, these being approximate 90% confidence intervals for $\pi_k^{\uparrow}$. The estimated standard error $\hat{\sigma}_k^{\uparrow}$ is based on the conservative formula (6.4),

$$\hat{\sigma}_k^{\uparrow} = [P_k(1 - P_k)/\sum_{i \leq k} n_i]^{1/2}. \quad (6.7)$$

The BATSE cumulative event rates from below are particularly interesting. The confidence intervals for $\pi_k^{\uparrow}$ lie significantly below the disk-model value $\pi_{\text{disk}} = .0488$, (4.9). This rules out the possibility that the hazard rate $h(y)$ is approaching 1.0 as a lower asymptote as $y$ gets small. The implication of the linear model in Figure 4, that the hazard rate smoothly declines below the disk-model value 1.0 for dim events (small values of $y$) seems vindicated by this less parametric analysis.

At the high end of the $y$-scale, $h(y)$ never significantly exceeds the sphere-model value 1.5. The confidence intervals for $\pi_k^{\uparrow}$ associated with the cumulative event rates from above never lie entirely above $\pi_{\text{sphere}} = .0723$.

It is not necessary to discretize the data in order to calculate the cumulative event rates. For any point $y$ on the $y$-scale we can define a continuous analogue of (6.1),

$$P_y = N_y / T_y, \quad (6.8)$$

where $N_y = \# \{y_i < y\}$, the number of events occurring before $y$, and $T_y$ is the total time on test before $y$. 

18
Figure 7. Cumulative event rates for SIGNE (top), SMM (middle), BATSE (bottom); from below (6.1), on the left; from above (6.5), on the right; horizontal axis is proportion of events involved, (6.6). Dotted curves are approximate 90% confidence intervals $P_k \pm 1.645\sigma_k$, (6.7). The BATSE cumulative event rates from below show that $\pi_k$ is less than $\pi_{\text{isk}} = .0488$ (4.9) for small values of $y$. There is no significant evidence that $\pi_k$ exceeds $\pi_{\text{sphere}} = .0723$ at the upper end of the $y$ scale.
\[ T_y = \sum_{i: t(x_i) < y} \left[ \min(y_i, y) - t(x_i) \right]. \]  \hspace{1cm} (6.9)

As the width \( w \) of the discretizing bins goes to zero, \( P_k/w \) defined by (6.1) approaches \( P_y \), for \( k \) chosen so that bin \( k \) includes \( y \). Statistic (6.8) is often used in the study of exponential distributions and Poisson processes. See for example Equation 15, Chapter 18 of Johnson and Kotz (1975). The partial logistic regressions of Section 4 can also be carried out continuously, but the calculations are more complicated than (6.9), as discussed in Section 5 of Efron (1988).

The binomial assumption (4.7) has been taken literally in deriving (6.2)–(6.4), ignoring the fact that the \( n_k \) themselves are random variables in our context. This is easy to justify from a likelihood or Bayesian point of view, but not by standard frequentist calculations. Remark A of Efron (1988) suggests that \( \hat{\sigma}_k^2 \) will tend to overestimate the actual standard error of \( \pi_k \).

Here is a summary of our conclusions about \( h(y) \), the hazard rate of the apparent burst magnitude \( y = \ln(C_y) \):

1. \( h(y) \) is not constant as a function of \( y \);
2. the preferred model for the hazard rate is the linear logistic (4.7);
3. the sphere-disk two-level model (4.10) is in some important ways contradicted by the data;
4. at the low end of the \( y \)-scale, the hazard rate is less than the disk-model value 1.0, as much as 40% less;
5. the data supports, or at least does not contradict, the sphere-model value 1.5 for the hazard at the upper end of the \( y \) scale.

References


Appendix

The Sphere and Disk Models. The exponential distributions (1.2), (1.3) are derived from simple homogeneity arguments. Suppose that a gamma ray burst occurs at distance $R$ from the earth, with true intensity $L$. The apparent intensity observed here is determined by the inverse square law

$$C_p = L/R^2. \quad (A.1)$$

Constants relating to geometrical considerations or instrument sensitivity have been absorbed into the definition of $L$ in (A.1).

Let $r_0$ indicate a distance so large that even the largest possible burst intensity, say $L_{\text{max}}$, gives an apparent intensity $C_p = L_{\text{max}}/r_0^2$ below the detection limit of our instrument. Obviously we need only consider events with $R < r_0$. Given $R < r_0$, suppose that the distance $R$ to a gamma ray burst has cdf following the power law

$$\text{Prob}\{R < r | R < r_0\} = \left(\frac{r}{r_0}\right)^\gamma. \quad (A.2)$$

The infinite sphere model has $\gamma = 3$ while the infinite disk model with thickness $\ll r$ has $\gamma = 2$. This says nothing more than that a sphere's volume is proportional to its radius cubed, and a disk's volume is proportional to its radius squared.

Define $M = \ln(L)$ and $Q = -2\ln(R)$. Then (A.1) gives

$$y = \ln(C_p) = M + Q. \quad (A.3)$$

If $R$ is distributed according to the power law (A.2), then it is easy to calculate that $Q$ has a shifted exponential density

$$f^Q(q) = \theta e^{-\theta(q-q_0)} \quad \text{for} \quad q > q_0, \quad (A.4)$$

where

$$\theta = \frac{\gamma}{2} \quad \text{and} \quad q_0 = -2\ln(r_0). \quad (A.5)$$

The values of $\theta$ agree with these in (1.3).

Finally, we assume that the true intensity $L$ is independent of $R$, or equivalently that $M$ and $Q$ are independent. Then $y = M + Q$ is the sum of a shifted and scaled exponential random variable $Q$ and an independent random variable $M$. Moreover we only get to observe values of $y$ greater than some minimum value $y_0$. By the definition of $r_0$, the maximum possible signal $M_1 = \ln(L_1)$ will be undetectable at distance $r_0$. Equivalently,

$$y_0 > M_1 + q_0. \quad (A.6)$$

Taken together, (A.3), (A.6) and the independence of $M$ and $Q$ imply that the conditional density of $y$ given $y > y_0$ is of the exponential form (1.2).