USING SURVIVAL ANALYSIS TECHNIQUES IN THE ANALYSIS
OF THE GAMMA-RAY BURST DATA

BY

BRADLEY EFRON and VAHÉ PETROSIAN

TECHNICAL REPORT NO. 167
AUGUST 1994

PREPARED UNDER THE AUSPICES
OF
PUBLIC HEALTH SERVICE GRANTS
9 R01 CA59039-18 AND 5 R01 CA55325

DIVISION OF BIOSTATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
Using Survival Analysis Techniques in the Analysis of the Gamma-Ray Burst Data

By

Bradley Efron and Vahé Petrosian

Technical Report No. 167

August 1994

Prepared Under the Auspices Of

Public Health Service Grants

9 R01 CA59039-18 and 5 R01 CA55325

Also prepared under National Science Foundation Grant DMS92-04864, and issued as Technical Report No. 459, Department of Statistics, Stanford University.

Division of Biostatistics

Stanford University

Stanford, California
USING SURVIVAL ANALYSIS TECHNIQUES IN THE ANALYSIS
OF THE GAMMA-RAY BURST DATA

Bradley Efron and Vahé Petrosian
Stanford University
Stanford, CA 94305

Abstract

Several hundred gamma-ray bursts have now been observed by BATSE, a special instrument on the Compton gamma-ray satellite. It is a matter of considerable importance to astrophysical theory whether these bursts are isotropically distributed across the sky, as they appear to be, or have some subtle clustering properties. Despite its astrophysical origins, the problem is identical to cancer-cluster questions that arise in epidemiology. Here we use some familiar survival analysis techniques, proportional hazards modelling and Mantel-Haenszel tests, to search for deviations from isotropy. The tests give mild support to a form of time-space clustering of the bursts.
1. INTRODUCTION

The observed isotropy of angular distribution of gamma-ray bursts (GRBs) has provided strong support for cosmological origin of their sources. Most popular model in this case is a neutron star-neutron star or a neutron star-black hole collision. However, such sources are expected to be isotropic and occur once every million years in a galaxy. Therefore, evidence for anisotropy or for repeated occurrence of a burst from the same location will be convincing arguments against this scenario. Several analysis of the angular distribution of GRBs have come to contradictory conclusions (Quashnock & Lamb 1993a, Nowak 1994, Narayan & Piran 1993). In our first paper on this subject (Efron & Petrosian, 1994a; EPI hereafter) we have shown that several global tests based purely on the angular position of sources show that the angular distribution of GRBs in the BATSE 1B catalog (Fishman et al, 1994) is consistent with a dozen sources being repeaters and that there is a marginal evidence for a larger scale (about 30°) clustering. However, neither of these are present in the BATSE 2B catalog indicating that this second sample is consistent with the cosmological scenario. It is not clear why the two catalogs are different, but the 2B catalog is marred by recording failures that make it difficult to assess angular accuracy. One hundred of the 326 new bursts in 2B catalog have no assessment of angular errors.

In this letter we describe tests which combine the angular position data with other characteristics of the GRBs such as time of occurrence, peak photon count rate, \( C_p \), duration etc. Quashnock & Lamb (1993b) have provided some evidence for dependence of degree of isotropy (dipole and quadpole moments) on \( C_p \) in the BATSE 1B catalog. They claim that bursts with medium intensity, \( C_p \) in the 450 to 1240 s\(^{-1}\) range, show a significant concentration toward the galactic disk. We performed the tests described in EPI on this subset of GRBs and found results similar to that for all GRBs. In §2 we perform a test on all the data that searches for correlation between \( C_p \) and various measures of angular positions. We find no evidence for any correlation.

In §3 we describe a test which searches for clustering in the time-angular position domain. As further evidence for repeaters Wang & Lingenfelter (1994) have pointed out that several sets of GRBs occurring close in time have small angular separations. Our test looks for statistical evidence for repeaters in the whole catalog. We find some marginal evidence for close sources occurring with 4 to 5 days of each other in the 1B catalog and similar evidence in the 2B catalog for repetition with time lags of up to hundreds of days.

2. ANGULAR LOCATION AND INTENSITY CORRELATION

Quashnock & Lamb (1993b) suggested that the medium intensity bursts show clear deviation from anisotropy, as measured by the dipole and quadpole moments, and are clustered towards the galactic disk. In this section we investigate the relationship between intensity, \( C_p \), and angular location.

We use a regression technique called the partial likelihood model adapted from biostatistics literature (Miller 1981). In applying this model it is important to keep account
of the fact that the detection threshold is variable. Each burst is described by a pair of photon count rates \((C_{lim}, C_p)\); the threshold and peak intensity counts, respectively. BATSE actually provides as many as three \((C_{lim}, C_p)\) pairs per burst. We will consider only the bursts with known 1024 ms \((C_{lim}, C_p)\) counts. We can think of an observed burst \(x_i\) as consisting of \(C_{lim,i}, C_{p,i}\), and other information \(Z_i\), where \(Z_i\) could stand for galactic longitude, latitude, burst duration, time history, etc. We would like to use regression models to analyze the burst intensity \(C_{p,i}\) in terms of the variables \(Z_i\), but to do so we need to properly account for the data truncation \((C_p \leq C_{lim})\). Partial likelihood regression does this by means of an analysis closely related to Lynden-Bell’s method. Efron and Petrosian (1994b) explain the connection. The discussion here will be very brief. As described in Efron & Petrosian (1994b), we define the comparable set for the \(j\)th burst to be

\[
C_j = \{i : C_{lim,i} < C_{p,j} \quad \text{and} \quad C_{lim,i} \geq C_{lim,j} \}.
\] (1)

Suppose that \(z_i\) is a vector function of \(Z_i\). Let \(\beta\) be the vector of partial likelihood regression coefficients, an unknown vector of the same dimension as \(z_i\). We estimate \(\beta\) by maximizing the partial likelihood

\[
\ell(\beta) = \sum_{j=1}^{n} \ln \left( \frac{e^{\beta \cdot z_j}}{\sum_{i \in C_j} e^{\beta \cdot z_i}} \right),
\] (2)

calling the maximizer \(\hat{\beta}\). For example, if the choice \(z_i = \cos(b_i)\) gave \(\hat{\beta} > 0\), this would suggest that members of \(C_j\) having larger values of \(\cos(b_i)\) have greater probability of occurring with smaller \(C_p\); in other words the larger-intensity bursts occur nearer the galactic poles. We have performed this test on the 193 bursts in the 1B catalog and the 338 bursts in the combined catalog with known values of \(C_p\) and \(C_{lim}\) at 1024 ms. We see no such evidence. The left panels of Figure 1 show \(\ell(\beta)\) for the case where \(z_i = \cos(b_i)\). The maximizer \(\hat{\beta}\) nearly equals 0, indicating no correlation between intensity and \(\cos(\text{latitude})\). Approximate 95% and 99% confidence intervals for \(\beta\) are shown, based on Wilks’ criterion

\[
2 \cdot \ell(\beta) \geq 2 \cdot \ell(\hat{\beta}) - \chi^2_p(\alpha)
\] (3)

Here \(p\) is the dimension of \(\beta\) and \(\chi^2_p(\alpha)\) is the appropriate percentile point of a chi-squared distribution, e.g., \(\chi^2_{1,0.95} = 3.84\). Notice that the confidence intervals for the 1B catalog (Figure 1a) are quite wide, so there is the possibility that more data might show \(\beta\) to be substantially far away from zero. But as such in Figure 1b this did not happen when the 2B catalog sources were added. The right panels of Figure 1 show the partial likelihood analysis for a two dimensional vector \(z_j = (\cos(b_j), \cos(b_j)^2)\). This model would test the possibility of the brighter or the dimmer bursts being concentrated near the middle latitudes. This test also produced negative results. The value \(\beta = 0\) is well within the 84% and 50% confidence region around \(\hat{\beta}\) for the 1B and combined catalogs, respectively. Several more partial likelihood regressions were run in an attempt to correlate burst intensity with
other quantities. These included burst duration and \( \cos(\text{longitude}) \). In no case was there any evidence of a relationship. Not every possible regression relationship is expressible in the partial likelihood format, in particular not the relationships suggested by Figure 2 of Quashnock & Lamb (1993b). These are intriguing suggestions but ones that are not yet strongly supported by the data, in the 1B catalog or the combined 1B and 2B catalogs.

Figure 1a. Partial likelihood analysis of the relationship between galactic latitude and burst intensity; left panel: \( z_j = \cos(b_j), b_j = \text{latitude} \); right panel \( z_j = (\cos(b_j), \cos(b_j)^2) \). In both cases the null hypothesis of no relationship, \( \beta = 0 \), is accepted. Likelihood-based confidence intervals and confidence regions include \( \beta = 0 \) for moderate values of the coverage probability. Based on the 193 BATSE bursts having 1024ms peak and threshold photon count rates.

3. TIME-SPACE CLUSTERING

As mentioned in §1, Wang & Lingenfelter (1994) have shown that several GRBs which occur close in time have very small angular separations. In this section we perform tests on all the data designed to reveal whether or not the position of past bursts affect the position of future ones. The bursts in the BATSE catalog are ordered according to their time of observation \( t_1 < t_2 < t_3 < \cdots < t_n \). This gives each burst 3 coordinates,

\[
x_i = (t_i, l_i, b_i) \quad i = 1, 2, \ldots, n,
\]

with \( (l, b) \) representing galactic longitude and latitude. A test for clustering in the \((t, l, b)\) with coordinates is developed here, based on a survival analysis technique called the Mantel-Haenszel test. See Miller (1981, Section 2.2) and also Efron and Petrosian (1994b).
Each burst $x_i$ has an associated error of observation, which plays a role in the Mantel-Haenszel test. Let $\sigma_i$ be $x_i$'s angular position observational standard error in degrees; $\sigma_i^2 = \sigma^2 + \eta^2$, where $\eta$ is the BATSE catalog entry Stat. Loc. Error and 4° is the assessment of the systematic error. In the 1B catalog the $\sigma_i$ range from 4.01° to 26.50°, with median 5.66°. The 2B catalog gives similar statistics for the $\sigma_i$, but with one important caveat: One hundred of the 325 new bursts in the 2B catalog suffered from recording errors. The angular errors, $\sigma_i$, are not known for these 100 bursts. We did not include them in our analysis. We also define $\tau_{ik}^2 = \sigma_i^2 + \sigma_k^2$, which is a measure of variance for $a_{ik}$, the angular separation between $x_i$ and $x_k$.

![Figure 1b](image)

Figure 1b. Same as Figure 1a but based on 339 BATSE bursts in the combined 1B and 2B catalogs having 1024 ms data.

The Mantel-Haenszel test compares the event $x_j$ that actually occurred at time $t_j$ with all the events $x_k, k \geq j$, that might have occurred. The comparison is made in terms of the closeness of $x_k$ to the set of events $\{x_1, x_2, \ldots, x_{j-1}\}$ that occurred before time $t_j$. We will use the measure of closeness

$$s_{jk} = \sum_{i:t_i-c \leq t_i \leq t_j} e^{-\frac{1}{2} \frac{\tau_{ik}^2}{\tau_{ik}^2}}$$

(5)

which roughly speaking is the number of bursts that occurred less than $c$ days before $x_j$ and lay within one standard error $\tau_{ik}$ of the position of event $x_k$. Note that in absence of errors $\tau_{ik} = 0$ and the Gaussian is replaced by a delta function, and only events occurring at exactly same angular position can be considered as repeaters. If $c$ exceeds the total time span of the observation, then $s_{jk}$ sums a Gaussian distances measure over all values.
of \( t_i < t_j \). If \( c = 5 \) days, for example, then \( s_{jk} \) includes only those events \( x_i \) occurring no more than 5 days before the \( j \)th event. We tried values of \( c \) ranging from 2 days to the total duration of the observations.

The actual score observed at \( t_j \) is \( S_j = s_{jj} \). Suppose that the events \((x_1, \ldots, x_{j-1})\) have no influence on which event \( x_k, k \geq j \), actually occurs at time \( t_j \). Then the expected value and variance of \( S_j \) are

\[
E_j^o = \sum_{k=j}^{n} r(x_k)s_{jk}/\sum_{k=j}^{n} r(x_k) \quad \text{and} \quad V_j^o = \sum_{k=j}^{n} r(x_k)(s_{jk} - E_j^o)^2/\sum_{k=j}^{n} r(x_k),
\]

where \( r(x) \) is the BATSE coverage function and \( n \) is the total number of bursts (see EPI). The Mantel-Haenszel statistic is defined to be

\[
M(c) = \frac{\sum_{j=2}^{n}(S_j - E_j^o)/(\sum_{j=2}^{n} V_j^o)^{1/2}}. \tag{7}
\]

Under the null hypothesis \( H_0 \) that past events have no influence on the future, \( M(c) \) is expected to have nearly a Gaussian distribution with zero mean and dispersion unity;

\[
H_0 : M(c) \sim N(0,1) \tag{8}
\]

absolute values of \( M(c) \) exceeding 2 would usually indicate significant evidence against \( H_0 \) though some caution is needed because many values of \( c \) are being looked at.

Figure 2 shows \( M(c) \) as a function of lag time \( c \). The solid line is for the 1B catalog with total duration of 320 days. For \( c = 4 \) or 5 days \( M(c) > 2 \). The significance at \( c = 4 \) is quite strong, \((p\text{-value .005 according to (8))}, \) though this value does not take into account the fact that we examined \( M(c) \) over a range of possible choices of \( c \). Taken at face value, this would suggest a time-space clustering, on a 4 or 5 day time scale. The dashed line shows the \( M(c) \) values for the combined catalog. The position errors \( \sigma_i \) are not known for a fair fraction of this new bursts. We have eliminated these from our analysis. The effects of this truncation is this worrisome but unknown. Nevertheless repeating the above test for the combined catalog we find that the significant deviation from \( H_0 \) is still present at 4 to 5 days of delay. Furthermore, there seems to be mildly significant evidence against \( H_0 \) at long delays \( c \approx 300 \) days.

4. CONCLUSION

We have searched the BATSE 1B data for evidence of either anistropy on repetition which could negate the cosmological origin of the GRBs. In the first paper, EPI, we found some convincing evidence for a small number of GRBs having repeated twice and a marginal evidence for a larger scale clustering in the 1B catalog but no such evidence in the 2B or the combined catalogs. These test were carried out without consideration of the other characteristics of GRBs. In this paper we have carried out two additional test.
We first search for a relation between the intensity (peak photon count rate, \( C_p \)) and degree of isotropy of the angular distributions of GRBs. We find no evidence for such relationships. However, our results does not directly contradict Queshnock & Lamb's (1993b) claims that mid-intensity bursts are concentrated toward the galactic disk.

Figure 2. Variation of \( M(c) \) from (7) with \( c \) for the BATSE 1B catalog (solid) and combined 1B and 2B Catalogs (dashed). Only the bursts with known positional uncertainties; \( n = 265 \) for 1B and \( n = 485 \) for the combined catalogs, respectively.

Next we test for presence of repeaters in the joint temporal and angular data. Evidence for this based on short angular separation for few bursts occurring close in time has been presented by Wang & Lingenfelter (1994). We show that there is some marginal evidence for clustering of close sources within 4 to 5 days of each other in the 1B catalog. This evidence is present in the combined 1B and 2B catalogs which in addition shows statistically evidence for repetition at longer time delays. However the absence of angular errors for 100 of the 2B bursts make this last result problematical.

REFERENCES

Efron, B. & Petrosian, V. 1994b,