TESTING FOR REPETITION OR CLUSTERING OF GAMMA-RAY BURSTS

BY

BRADLEY EFRON and VAHÉ PETROSIAN

TECHNICAL REPORT NO. 457

AUGUST 1994

PREPARED UNDER THE AUSPICES OF

NATIONAL SCIENCE FOUNDATION GRANT DMS92–04864

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA
Testing For Repetition Or
Clustering of Gamma-Ray Bursts

By
Bradley Efron and Vahé Petrosian

Technical Report No. 457
August 1994

Prepared Under the Auspices

Of
National Science Foundation Grant DMS92-04864

Department of Statistics
Stanford University
Stanford, California
TESTING FOR REPETITION OR
CLUSTERING OF GAMMA-RAY BURSTS

By

B. Efron and V. Petrosian
Stanford University
Stanford, CA 94305

ABSTRACT

Cosmological origin of the gamma-ray bursts has been gaining popularity primarily because they manifest nearly isotropic angular distribution. Several analysis of the data using different statistical tests have claimed contradictory results on the degree and significance of deviation of the angular distribution from isotropy. In this paper we apply three tests which we believe are more powerful than those applied previously to the BATSE 1B catalog. We find that the excess of sources at small angles claimed by Quashnock & Lamb (1993a) is significant and can be interpreted as evidence about fifteen sources repeating twice. We also find some weak evidence for clustering of sources on a scale of about 30°. In the accompanying paper we discuss possible existence of possible connection between angular and temporal distributions and dependness of the angular distribution on burst intensity.
1. INTRODUCTION

The origin of gamma-ray bursts (GRBs) remains an enigma despite an impressive increase in observational quantity and quality. Figure 1 shows the sky distribution of the 260 bursts reported in the BATSE 1B Catalog (Fishman et al., 1993). The distribution looks uniform, and in fact Fishman et al. (see also Meegan et al 1991) show that it passes the usual dipole and quadrupole tests for directional uniformity. These are powerful tests for general anisotropic trends, such as a global preference for the galactic disk and have been the primary reason for the gradual gain in popularity of the cosmological origin of GRBs. However, these moments do not provide a good test of existence of local clusterings. Several recent papers have searched for clusterings using different tests.

Quashnock & Lamb (1993a), using the distribution of distances to the nearest neighbor, have shown existence of excessive number of bursts within few degrees of each other. They interpret this as evidence for burst repetition which is not expected to occur in the cosmological model. The significance of this result, however, is disputed. Nowak (1994), using some simulations, claims that the excess at short distance is not statistically significant, and Narayan & Piran (1993) show existence of comparable deviations in the distribution of distances to the farthest neighbor. These excesses are evident also in the two point correlation function but with conflicting assessment of the significances. (Narayan & Piran 1994, Hartman et al., 1993, Quashnock & Lamb 1994) Finally, Zharkova & Zharkova (1994, preprint) using a different test claim existence of a large clustering in the vicinity of Andromeda.

Figure 1. The sky distribution in galactic coordinates of the 260 gamma-ray bursts recorded in the 1993 BATSE catalogue, (Fishman et al.). The origin is the direction of the galactic center, with the equator indicating the galactic plane.
Neither one of these tests utilizes the information available in the data as fully as possible. The nearest neighbor test is limited to distribution of closest pairs. The two point correlation function averages over many pairs. For a large number of sources practical considerations dictate use of such tests, but for a small samples one can begin to examine more than the averages or minimum distances. In this paper we present results from three different tests for clustering in the BATSE 1B catalog. These test are described in sections 2-4. In section 5 w. summarize the conclusions of the three tests. There is a strong suggestion of repeated bursts, perhaps 10-15 pairs of them. Despite some intriguing hints, there is no strong evidence for other forms of clustering.

There has been two other analysis of the 1B catalog supporting the repetition and galactic origin hypothesis. These aspects of the problem are discussed in the accompanying letter. As we were preparing the manuscript for press the second BATSE catalog became available. We will begin to apply the same tests to the new and the combined samples shortly.

Section 5 summarizes the conclusions of the three tests. There is a strong suggestion of repeated bursts, perhaps 10–15 pairs of them. Despite some intriguing hints, there is no strong evidence for other forms of correlation or clustering.

2. The Quashnock-Lamb Test For Local Clustering.

In this section we analyze the data in Figure 1 using a modification of a test statistic suggested by Quashnock and Lamb (1993a). The modification is designed to give increased statistical power for detecting repeated bursts.

The data in Figure 1 can be represented as \( x = (x_1, x_2, \cdots, x_{260}) \), where each \( x_i \) is a bivariate point representing position on the celestial sphere. Convenient coordinates are

\[
x_i = (\ell_i, b_i), \quad i = 1, 2, \cdots, n \quad (n = 260),
\]

with \( \ell_i \) galactic longitude and \( b_i \) galactic latitude. Each burst \( x_i \) has a nearest-neighbor burst in terms of angular distance on the sphere, and we let \( A_i \) represent the angle in degrees between \( x_i \) and its nearest neighbor.

Because the BATSE sky coverage is not uniform the raw angles \( A_i \) are not true measures. They over estimate the angular distance in the regions of sky with relatively sparse coverage. To account for this we replace the \( A_i \) with the modified distance

\[
B_i = \sqrt{r(x_i)} A_i ,
\]

(2.1)

where \( r(x) \) is then BATSE sky coverage function (Figure 4 of Fishman et al. (1993) normalized part that \( \sum_{i=1}^{n} r(x_i) = n \).

The left panel of Figure 2 compares the cumulative distribution function ("cdf") of the 260 \( B_i \) values with the cdf we would expect under the null hypothesis

\[
H_0 : \text{the gamma-ray sources are non-repeating and isotropic} .
\]

(2.2)
Figure 2. Left panel: The jagged curve is the cdf of the 260 modified nearest-neighbor distances \( B_i \); (2.1); smooth dotted curve is the null hypothesis cdf assuming a uniform distribution of burst sources. Right panel: The deviation between the jagged and dotted curves is compared with 200 Monte Carlo simulations; deviation measured by Wilcoxon statistic (2.4); the actual deviation exceeded all but 5 of the 200 Monte Carlos, giving one-sided \( p \)-value \( 5.5/200 = .0274 \).

The null hypothesis cdf was computed by Monte Carlo. Simulated data sets \( x^* = (x^*_1, x^*_2, \ldots, x^*_260) \) were obtained by independently drawing the points \( x^*_i \) according to the BATSE sky coverage density \( r(x) \). This represents what BATSE would have seen if the burst sources were isotropic and non-repeating. Two hundred such data sets were generated, \( x^{*(1)}, x^{*(2)}, \ldots, x^{*(200)} \), giving 200 corresponding \( A_i, B_i \) and cdf's: \( \text{cdf}^{*(1)}, \text{cdf}^{*(2)}, \ldots, \text{cdf}^{*(200)} \). The smooth dotted curve in Figure 2 is the median curve of these 200, having value

\[
\overline{\text{cdf}}(B) = \text{median \{cdf}^{*(j)}(B), \ j = 1, 2, \ldots, 200\} \tag{2.3}
\]

at distance \( B \).

We see that the actual cdf lies suspiciously above the null hypothesis cdf, pointed out by Quashnock Lamb (1993), the greatest discrepancy occurring at a distance near 5°. The median error angle for a BATSE observation is 5.7°, so this discrepancy is consistent with the occurrence of repeated bursts. The existence of this discrepancy is not disputed. The controversy arises in the determination of the significance of the deviation observed in Figure 2 between the actual cdf and the Monte Carlo \( \overline{\text{cdf}} \)? Quashnock & Lamb (1993a) and Nowak (1994) assess significance in terms of the Kolmogorov-Smirnov test. However the Kolmogorov-Smirnov statistic is not very powerful for detecting this kind of shift toward zero observed in the nearest-neighbor distances \( B_i \). Here we will
use a Wilcoxon statistic in place of the Kolmogorov-Smirnov statistic. This gives greater statistical power for detecting shifts toward zero.

The Wilcoxon statistic is defined as

\[ W = \sqrt{12} \cdot n \left[ .5 - \frac{1}{n} \sum_{i=1}^{n} \text{cdf}(B_i) \right]. \quad (2.4) \]

Large values of \( W \) correspond to the distances \( B_i \) being generally too small. Since the quantity \( \text{cdf}(B_i) \) will have nearly a uniform density on \((0, 1)\) under the null hypothesis, \( W \) will have nearly a standard normal distribution \( N(0, 1) \). Instead of comparing \( W \) for the observed GRBs to a \( N(0, 1) \) distribution, we can get an exact test by comparing \( W \) to the corresponding simulated values \( W^{*1}, W^{*2}, \ldots, W^{*200} \) obtained by letting each \( x^{*(j)} \) in turn play the role of the original data \( x \). The right panel of Figure 2 shows that the actual value of \( W \) (indicated by the dashed line) exceeded all but 5 of the 200 \( W^{*j} \) values, giving a significance level or \( p \)-value \( .027 = 5.5/201 \) against the null hypothesis \( H_0 \).

This provides convincing, if not overwhelming, evidence against \( H_0 \). A still smaller \( p \)-value is obtained by considering only the 202 bursts with observed angular error < 9°. However this result can be affected by anisotropies in the BATSE distribution of angular observation errors.

3. The All-Angles Test Statistic.

A shortcoming of the above test is its use of only the nearest-neighbor distances. In this section we discuss tests that use all of pairwise distances. These test statistics turn out to be closely related to the two-point correlation function.

Let \( a_{ij} \) be the angle in degrees between bursts \( i \) and \( j \), and define the modified angular distance as in (2.1),

\[ b_{ij} = \sqrt{r(x_i)} a_{ij}, \quad (3.1) \]

where \( r(x) \) is the BATSE coverage function. We order the 33411 = \( \binom{259}{2} \) values of \( b_{ij} \) for \( i < j \) from smallest to largest, and denote this 33411 vector by \( b = (b(1), b(2), \ldots, b(33411)) \). For the data in Figure 1, \( b \) has smallest distance \( b(1) = 0° \), between the 121st and 150th bursts in the BATSE catalog, \( b(2) = 0.56° \), between the 210th and 236th bursts, etc.

We can construct tests of the null hypothesis \( H_0 \), (2.2), based on the components of \( b \). To do so, 400 independent simulated data sets \( x^{*(j)} = (x_1^{*(j)}, x_2^{*(j)}, \ldots, x_{260}^{*(j)}) \) were drawn according to the BATSE coverage function, as in Section 2. Each \( x^{*(j)} \) gave an ordered distance vector \( b^{*(j)} = (b^{*(j)}(1), b^{*(j)}(2), \ldots, b^{*(j)}(33411)) \). The attained \( p \)-value, or significance level, based on the \( k \)th component of \( b \) is defined to be

\[ p_{val}(k) = \frac{\# \{ j : b^{*(j)}(k) < b(k) \} + .5}{401}, \quad (3.2) \]
essentially the proportion of the simulated \( b^{(j)}(k) \) values less than the actual value \( b(k) \). For example \( b(5) = 0.77^\circ \) was smaller than all but 4 of the \( b^{(j)}(5) \) values, giving \( pval(5) = 4.5/401 = .011 \). Under the null hypothesis \( H_0 \) we expect a \( pval(k) = 0.5 \).

Figure 3 graphs \( pval(k) \) versus \( \bar{b}(k) \) is the median simulated \( k \)th distance for 100 representative values of \( k \):

\[
\bar{b}(k) = \text{median} \{ b^{(j)}(k), \ j = 1, 2, \cdots, 400 \}.
\]

\[
k = 1, 2, \cdots, 10, 12, 14, \cdots, 30, 34, 38, \cdots, 70, 78, 86, \cdots, 150, 166, 182, \cdots, 310, 342, \cdots,
\]
\[
630, 694, \cdots, 1270, 1398, \cdots, 2550, 2806, \cdots, 5110, 5622, \cdots, 10230.
\]

This range of \( k \), \( \bar{b}(k) \) run from \( 0.5^\circ \) to \( 67^\circ \).

The results shown in Figure 3 agree with the nearest neighbor test, but with stronger significance levels. For small values of \( k \), \( pval(k) \) often attains the smallest possible value \( 0.5/401 = .001 \). The significance suddenly evaporates at \( \bar{b}(k) = 7^\circ \). For large values of \( k \), those having \( \bar{b}(k) > 7^\circ \), the observed angles \( b(k) \) are not significantly smaller than those obtained under the null hypothesis.

---

**Figure 3.** The attained \( p \)-value \( pval(k) \), (3.2), plotted versus median \( k \)th distance for the 400 simulations, (3.3); for the 100 \( k \) values in (3.4). The dashed line indicates the 5\% significance level. The \( p \)-values are quite small for small angular separations, but suddenly become non-significant for \( \bar{b}(k) > 7^\circ \).
Figure 3 reports the p-values for 100 different tests of $H_0$. We can combine these to obtain a single overall significance level. For $K$ a subset of the values $k$ in (3.3), define

$$W(K) = \sum_{k \in K} \log(p_{\text{real}(k)}).$$

(3.5)

By comparing $W(K)$ with the corresponding values $W^{*(1)}(K), W^{*(2)}(K), \ldots, W^{*(400)}(K)$ obtained by letting each $x^{*(j)}$ play the role of the original data $x$, we obtain an overall significance level for $W(K)$.

Table 1 shows the resulting significance levels for various choices of $K$. If we are willing to confine attention to the smaller values of $k$, say the first 50, then we see strong evidence against $H_0$. It seems, therefore, very likely that there is an excess of small angles in the BATSE data, compared to what $H_0$ would predict. This is similar to saying that the observed deviation from zero of the two-point correlation function (Naragan and Pirar, (1994)) of small angles of separation is statistically significant.

As mentioned above, because the angles under question are less than or of the order of uncertainties in position measurement of BATSE, these excess of pairs at small angels can be interpreted as evidence for certain number of CRBs to be repeaters. We now estimate how many "extra" pairs with small angles are there in the BATSE data? Under the null hypothesis $H_0$, (2.3), the expected number of pairs $(x_i, x_j), i < j$, having angle $\alpha_{ij}$ less than $\theta$, say $n^{*}_{\theta}$, approximately equals

<table>
<thead>
<tr>
<th>$K$</th>
<th>angle</th>
<th>sig. level</th>
<th>$K$</th>
<th>angle</th>
<th>sig. level</th>
</tr>
</thead>
<tbody>
<tr>
<td>first 10</td>
<td>1.9°</td>
<td>.001</td>
<td>first 60</td>
<td>15.7°</td>
<td>.011</td>
</tr>
<tr>
<td>first 20</td>
<td>3.4°</td>
<td>.004</td>
<td>first 70</td>
<td>22.4°</td>
<td>.026</td>
</tr>
<tr>
<td>first 30</td>
<td>5.2°</td>
<td>.004</td>
<td>first 80</td>
<td>31.9°</td>
<td>.041</td>
</tr>
<tr>
<td>first 40</td>
<td>7.6°</td>
<td>.004</td>
<td>first 90</td>
<td>45.9°</td>
<td>.049</td>
</tr>
<tr>
<td>first 50</td>
<td>11.0°</td>
<td>.006</td>
<td>all 100</td>
<td>66.8°</td>
<td>.056</td>
</tr>
</tbody>
</table>

Table 1. Significance level for testing $H_0$ based on the combined statistic $W(K)$, (3.5) "angle" is $\tilde{b}(k)$ for the maximum $k$ in $K$.

$$n^{*}_{\theta} = \frac{1}{2} \sum_{i=1}^{n} (n-1) r(x_i) \frac{1 - \cos \theta}{2} = \frac{n \cdot (n-1)}{2} \frac{1 - \cos \theta}{2}.$$ 

(3.6)

Here we have used the facts that $\sum_{i=1}^{n} r(x_i) = n$ and that $(1 - \cos \theta)/2$ is the area of a spherical cap of radius $\theta$.

The left panel of Figure 4 graphs the difference between $n_{\theta}$, the actual number of distinct pairs having $\alpha_{ij} < \theta$, and $n^{*}_{\theta}$. The maximum $n_{\theta} - n^{*}_{\theta} = 25.9$ occurs at $\theta = 5^\circ$. The right panel graphs a similar estimate based on the nearest neighbor statistic: 260 times the difference between the jagged and smooth curves in the left panel of Figure 2. The results are similar, the maximum
difference of 29.5 occurring at $\theta = 4.75^\circ$. (Notice that some pairs can be counted twice in the nearest neighbor calculation, if each member is the other's nearest neighbor.)

Figure 4. Estimates of the excess number of small angle pairs in the BATSE data. Left Panel: The differences between the actual and expected number of burst pairs within $\theta^\circ$ as a function of $\theta$; maximum difference of 25.9 occurs at $\theta = 5^\circ$. Right Panel: Excess number of pairs estimated from the Quashnock-Lamb statistic, left panel of Figure 2; maximum 29.5 occurs at $\theta = 4.75^\circ$.

The nearest neighbor results could be interpreted as suggesting 15 isolated repeating sources, each of which repeated twice. Alternately, there might be a single source that repeated some 15 times. In the latter case though, we would expect $n_\theta - n^0_\theta$, (3.6), to be larger than 25.9, perhaps on the order of $\left(\frac{15}{2}\right) = 105$. The next section discusses a test statistic that directly checks for the existence of multiple-repeat clusters.

4. Local Density Test Statistics.

Neither the nearest neighbor nor the all-angles test statistics are particularly powerful for detecting multiply repeating sources. The two point correlation function is a powerful test for determination of the scale of clusterings when there exist a large number of clusters. We, on the other hand, are interested in detection of small number of clusters; e.g. a few sources repeating many times. We now describe a statistic designed for this purpose.

Let $A_{ik}$ denote the angular distance to the $k$th nearest neighbor of burst $i$,

$$A_{ik} = k^{th} \text{ smallest angle between } x_i \text{ and } \{x_j : j \neq i\},$$
and define

\[ B_{ik} = \sqrt{r(x_i)} A_{ik} \tag{4.1} \]

to be the modified \( k \)th nearest neighbor distance (so \( B_i \) in (2.1) equals \( B_{i1} \)). We will use test statistics based on the minimum value

\[ m(k) = \min_i \{ B_{ik} \}, \tag{4.2} \]

calling these local density statistics for reasons discussed below.

Suppose that a single source repeated a large number of times, say \( k = 20 \). Intuitively we would expect \( m(20) \) to be noticeably too small, and for the minimum to occur at a burst \( i \) belonging to the cluster of repetitions.

Exactly the same procedure as in Section 3 was used to evaluate the significance of the observed values \( m(k) \) for the BATSE data. Now the role of \( b(k) \) is played by the local density statistic \( m(k) \), with the \( p \)-value for \( m(k) \) being

\[ pval(k) = \frac{\# \{ j : m^{*(j)}(k) < m(k) \}}{401} + .5, \tag{4.3} \]

as in (3.2). Here we take \( k = 1, 2, \cdots, 100 \), instead of (3.3).

Figure 5 shows \( pval(k) \) for \( k = 1, 2, \cdots, 100 \), plotted versus the median \( \bar{m}(k) \) from 400 null hypothesis simulations,

\[ \bar{m}(k) = \text{median}\{ m^{*(j)}(k), \ j = 1, 2, \cdots, 400 \}. \tag{4.4} \]

The results are rather surprising. The \( p \)-values are well above .05 for the first 30 values of \( k \), except for \( k = 1, 2, 8, 9, 10 \). However for large values of \( k \), those having \( \bar{m}(k) > 40 \), the \( p \)-values are near .05.
Figure 5. The attained p-values $pval(k)$, (4.3), for the local density statistics $m(k)$, (4.2); $k = 1, 2, \cdots, 100$; plotted versus angular distance $\bar{m}(k)$, the median of 400 simulations $m^*(k)$; hashmarks indicate $\bar{m}(k)$ for $k = 10, 20, \cdots, 100$.

Table 2 is the equivalent of Table 1, now showing significance levels for $W(K)$ statistics (3.5) based on the local density $p$-values (4.3). Only the smallest set, $K = \{1, 2, \cdots, 10\}$, shows strong significance against $H_0$. (Notice that the angles involved are much larger than those in Table 1. Here $\bar{m}(2) = 2.2^\circ$ is already larger than the value $\bar{b}(10) = 1.9^\circ$ in Table 1.)

<table>
<thead>
<tr>
<th>$K$</th>
<th>angle</th>
<th>sig. level</th>
<th>$K$</th>
<th>angle</th>
<th>sig. level</th>
</tr>
</thead>
<tbody>
<tr>
<td>first 10</td>
<td>13.4°</td>
<td>.019</td>
<td>first 60</td>
<td>49.2°</td>
<td>.089</td>
</tr>
<tr>
<td>first 20</td>
<td>23.5°</td>
<td>.066</td>
<td>first 70</td>
<td>54.1°</td>
<td>.091</td>
</tr>
<tr>
<td>first 30</td>
<td>31.3°</td>
<td>.128</td>
<td>first 80</td>
<td>58.9°</td>
<td>.099</td>
</tr>
<tr>
<td>first 40</td>
<td>37.9°</td>
<td>.099</td>
<td>first 90</td>
<td>63.3°</td>
<td>.084</td>
</tr>
<tr>
<td>first 50</td>
<td>43.6°</td>
<td>.089</td>
<td>all 100</td>
<td>67.9°</td>
<td>.076</td>
</tr>
</tbody>
</table>

Table 2. Same as Table 1, except using the local density statistics $m(k)$; "angle" is $\bar{m}(k)$ for the maximum $k$ in $K$.

The equivalent of Figure 5 for the 202 GRBs, with observed angular error $< 9^\circ$ looks the same, except that the p-values are much smaller for angular distances in the range of $40^\circ$ to $60^\circ$. This is an intriguing result, suggesting an extended source of gamma-ray bursts, but again it is suspect because of its dependence on the BATSE observational error statistics.
Figure 6 indicates the bursts $x(k)$ giving the minima $m(k)$ in (4.2), for $k = 1, 2, \cdots, 100$. Most occur in the region having longitude between $60^\circ$ and $90^\circ$, and latitude between $0^\circ$ and $-30^\circ$: a tendency that was even stronger for the 202 now accurately known positions. Zharkov and Zharkov’s (1994) statistics also highlight this region, which they note is near Andromeda’s location $(115^\circ, -20^\circ)$!

![Figure 6. Location of the bursts $x(k)$ achieving the minimum distance $m(k)$ in (4.2); stars indicate $x(k)$ for $1 \leq k \leq 10$, most of these being near $(90^\circ, -30^\circ)$; +’s indicate $x(k)$ for $11 \leq k \leq 100$.](image)

We can interpret the bursts $x(k)$ in Figure 6 as points of highest estimated burst density. To see this, let $n_\theta(x_i)$ be the number of points $x_j, j \neq i$, lying within angular distance $\theta$ of $x_i$, and let $n_\theta^*(x_i) = (n - 1) \cdot r(x_i)[1 - \cos(\theta)]/2$ be the approximate expected number of such points under the null hypothesis $H_0$, (2.3). An obvious estimate of local burst density near $x_i$ is

$$\hat{d}_\theta(x_i) = \frac{n_\theta(x_i)}{n_\theta^*(x_i)} = \frac{2}{(n - 1) \cdot r(x_i)[1 - \cos(\theta)]} \cdot \frac{n_\theta(x_i)}{4(n - 1) r(x_i) \theta^2}.$$ 

At $\theta = A_{ik}$ we have $n_\theta(x_i) = k$, giving $k$th nearest-neighbor density estimate

$$\hat{d}_\theta(x_i) = \frac{4}{(n - 1) \cdot r(x_i) A_{ik}^2} = \frac{4k}{(n - 1)} B_{ik}^2,$$

(4.1). Therefore $m(k)$ as defined in (4.2) is inversely proportional to the greatest $k$th nearest-neighbor density estimate, among all 260 bursts. (Similarly $B_i$ in (2.1) is inversely proportional to $\hat{d}_\theta(x_i)$ for $\theta$ equal to the nearest neighbor distance $A_i$.)

11
Quashnock and Lamb (1993y) also consider as a separate group the midrange bursts, those having burst intensities near the middle of the BATSE observations. An analysis similar to that of Figure 5 was run for the 85 bursts having peak 1024 brightness between 450 and 1400. Now pval(k) was below .05 only for large separations, \( \bar{m}(k) > 50^\circ \), the smallest p-value being .0075 at 67°. The W test for all of the k values gave an overall p-value of only .062. The cluster in Figure 6 was also quite evident for this data set.

5. Time-Space Clustering. The bursts in the BATSE catalog are ordered according to their time of observation \( t_1 < t_2 < t_3 < \cdots < t_{260} \). This gives each burst 3 coordinates,

\[
x_i = (t_i, u_i, v_i) \quad i = 1, 2, \cdots, 260,
\]

\((u, v)\) representing galactic longitude and latitude as before. A test for clustering in the \((t, u, v)\) coordinates is developed here, based on a survival analysis technique called the Mantel-Haenszel test. See Miller (Section 2.2, 1981) and also Efron and Petrosian (1994).

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 observational standard error in degrees, defined as the square root of the BATSE catalog entry Stat. Loc. Error plus 4°. (4° is the assessment of systematic error.) The \( \sigma_i \) range from 4.01° to 26.50°, with median 5.66°. We also define \( r_{ik} = \sigma_i^2 + \sigma_k^2 \), a measure of variance for \( a_{ik} \), the angular separation between \( x_i \) and \( x_k \).

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 \( \{x_1, x_2, \cdots, x_{j-1}\} \) that occurred before time \( t_j \). We will use the measure of closeness

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

If \( c \) exceeds 320 days, the total time span of the BATSE catalog, 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 to 320 days.

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

\[
E_j^o = \sum_{k=j}^{260} r(x_k) s_{jk} / \sum_{k=j}^{260} r(x_k) \quad \text{and} \quad V_j^o = \sum_{k=j}^{260} r(x_k) (s_{jk} - E_j^o)^2 / \sum_{k=j}^{260} r(x_k),
\]

where \( r(\cdot) \) is the BATSE coverage function. The Mantel-Haenszel statistic is defined to be

\[
M(c) = \sum_{j=2}^{260} (S_j - E_j^o) / (\sum_{j=2}^{260} V_j^o)^{1/2}.
\]
Under the null hypothesis $H_0$ that past events have no influence on the future, $M(c)$ is approximately standard normal,

$$H_0 : M(c) \sim N(0,1).$$

(5.5)

<table>
<thead>
<tr>
<th>$c$</th>
<th>$M(c)$</th>
<th>(pval)</th>
<th>$c$</th>
<th>$M(c)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.13</td>
<td></td>
<td>10</td>
<td>-0.11</td>
</tr>
<tr>
<td>3</td>
<td>1.18</td>
<td></td>
<td>20</td>
<td>-0.34</td>
</tr>
<tr>
<td>4</td>
<td>2.59</td>
<td>.005</td>
<td>30</td>
<td>-0.22</td>
</tr>
<tr>
<td>5</td>
<td>2.08</td>
<td>.019</td>
<td>100</td>
<td>-0.65</td>
</tr>
<tr>
<td>6</td>
<td>1.54</td>
<td></td>
<td>320</td>
<td>-1.27</td>
</tr>
<tr>
<td>7</td>
<td>1.11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Values of the Mantel-Haenszel statistic (5.4) for different choices of the time lag $c$. $M(c)$ is significantly positive only for $c = 4$ or 5 days.

Table 3 shows $M(c)$ for several choices of $c$. The only choices that make $c$ significantly positive are $c = 4$ or 5 days. The significance at $c = 4$ is quite strong, $p$-value .005 according to (5.5), though this value does not take into account the fact that we maximized $M(c)$ over a range of possible choices of $c$. Taken at face value, Table 3 suggests a time-space clustering, on a 4 or 5 day time scale.

6. Angular Location and Burst Intensity. Quashnock and Lamb (1993x) suggest that the median-intensity bursts are clustered toward the galactic center. Here we will investigate the relationship between intensity and angular location using a regression technique called the partial likelihood model, also adapted from the biostatistics literature. See Chapter 6 of Miller (1981).

Each burst records two photon counts $C_{lim}$ and $C_p$, the threshold and peak intensity counts respectively. BATSE actually provides as many as three $(C_{lim}, C_p)$ pairs per burst, but here we will consider only the 1024ms counts. Of the 260 bursts, there are $n = 193$ having 1024ms $(C_{lim}, C_p)$ counts, and these are the only bursts considered in this section. For convenient notation we define $(s_i, t_i)$ to be $(C_{lim}, C_p)$ for the $i$th burst, $i = 1, 2, \ldots, 193$. We always have

$$s_i < t_i$$

(6.1)

since $C_{lim} < C_p$ by the definition of what constitutes an observed gamma-ray burst.

We can think of an observed burst $x_i$ as consisting of $s_i, t_i$, and other information $Z_i$, where $Z_i$ includes galactic longitude, latitude, burst duration, time history, etc. We would like to use regression models to analyze the burst intensity $t_i$ in terms of the explanatory variables $Z_i$, but to do so we need to properly account for the data truncation (6.1). Partial likelihood regression does this by means of an analysis closely related to Lynden-Bell's method. Efron and Petrosian (1994) explain the connection. The discussion here will be very brief.
As with Lynden-Bell’s method, we define the comparable set for the $j$th burst to be

$$C_j = \{i : s_i < t_j \text{ and } t_i \geq t_j\}. \quad (6.2)$$

If $t$ is thought of as time, then $C_j$ indexes those bursts $(s_i, t_i)$ that might possibly have been observed at $t_j$ (since their truncation limit was less than $t_i$ and they did not occur “before” $t_i$).

Suppose that $z_i$ is a vector function of $Z_i$. In the first two cases below, $z_i = \cos(v_i)$ where $v_i$ is galactic latitude, and $z_i = (\cos(v_i), \cos(v_i)\cos(v_i^2))$. 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

$$\ell(\beta) = \sum_{j=1}^{n} \ln \left( \frac{e^{\beta^t z_i}}{\sum_{i \in C_j} e^{\beta^t z_i}} \right), \quad (6.3)$$

calling the maximizer $\hat{\beta}$. For example if the choice $z_i = \cos(v_i)$ gave $\hat{\beta} > 0$, this would suggest that members of $C_j$ having bigger values of $\cos(v_i)$ have greater probability of occurring “earlier” as we move up along the intensity scale; in other words that larger-intensity bursts occur nearer the galactic poles.

This is not what actually happened. The left panel of Figure 7 shows $\ell(\beta)$ for the case where $z_i = \cos(v_i)$. The maximizer $\hat{\beta}$ nearly equals 0, indicating no relationship between intensity and $\cos(\text{latitude})$. Approximate 95% and 99% confidence intervals for $\beta$ are shown, based on Wilks’ criterion

$$2 \cdot \ell(\beta) \geq 2 \cdot \ell(\hat{\beta}) - \chi^2_p(\alpha) \quad (6.4)$$

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 are quite wide, so there is still the possibility that more data might show $\beta$ to be substantially far away from zero.

The right panel of Figure 7 shows the partial likelihood analysis for $z_j = (\cos(v_j), \cos(v_j)\cos(v_j^2)$). This model would allow the brighter or the dimmer, bursts to be concentrated near the middle latitudes, but this did not happen. The value $\beta = 0$ is well within the 84% confidence region around $\hat{\beta}$.

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 and Lamb (1993x). These are intriguing suggestions but ones that are not yet strongly supported by the data, at least not that in the 1993 catalog.
Figure 7. Partial likelihood analysis of the relationship between galactic latitude and burst intensity; left panel: $z_j = \cos(v_j), v_j$ = latitude; right panel $z_j = (\cos(v_j), \cos(v_j)^2)$. 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 intensity measurements.

7. Conclusions. Here are some conclusions from these analyses:

- There is strong evidence for a small number of repeat bursts, perhaps 15 pairs out of the 260 BATSE bursts.
- There is no evidence of multiple bursts where the multiplicity is more than 2 or 3.
- There is moderately strong evidence for a time-space clustering, with the time window being on the order of 4 days.
- There is borderline evidence for a diffuse cluster of bursts emanating from the region indicated in Figure 6.
- There is no evidence of relationships between burst intensity and angular location.

Temporary INSERT HERE:

5. CONCLUSION
Using the nearest neighbor angular distribution and a new test statistics we show that the Quashnock & Lamb claim of existence of excess number of nearby sources is significant.

This can be interpreted as a strong evidence for certain number of bursts, perhaps about 15 pairs out of the 260 bursts in BATSE 1B catalog, to be repeaters.

There is no evidence of multiple bursts repetition where the multiplicative is more than 3.

There is a marginal evidence for existence of a large scale clustering in the region indicated in Figure 6.

In the accompanying paper we discuss the temporal - angular correlation and intensity angular position relation of same bursts. In subsequent papers we intend to apply the same methods to the larger catalog which has became available recently.

REFERENCES


Hartman et al. (1993)