The Swift satellite and redshifts of long gamma-ray bursts
(Research Note)

Z. Bagoly[1], A. Mészáros[2], L. G. Balázs[3], I. Horváth[4], S. Klose[5], S. Larsson[6], P. Mészáros[7], F. Ryde[6], and G. Tusnády[8]

[1] Laboratory for Information Technology, Eötvös University, 1117 Budapest, Pázmány P. s. 1/A, Hungary
  e-mail: zsolt.bagoly@elte.hu
[2] Astronomical Institute of the Charles University, V Holešovičkách 2, 180 00 Prague 8, Czech Republic
  e-mail: meszaros@mbox.cesnet.cz
[3] Konkoly Observatory, 1525 Budapest, POB 67, Hungary
  e-mail: balazs@konkoly.hu
[4] Department of Physics, Bolyai Military University, 1456 Budapest, POB 12, Hungary
  e-mail: horvath.istvan@zmne.hu
[5] Tautenburg Observatory, 07778 Tautenburg, Sternwarte 5, Germany
  e-mail: klose@tls-tautenburg.de
[6] Stockholm Observatory, AlbaNova, 106 91 Stockholm, Sweden
  e-mail: [stefan;felix]@stro.su.se
[7] Dept. of Astronomy & Astrophysics, Pennsylvania State University, 525 Davey Lab. University Park, PA 16802, USA
  e-mail: nnp@astro.psu.edu
[8] Rényi Institute of Mathematics, Hungarian Academy of Sciences, 1364 Budapest, POB 127, Hungary
  e-mail: tusnady@renyi.hu

ABSTRACT

Until 6 October 2005 sixteen redshifts had been measured of long gamma-ray bursts discovered by the Swift satellite. Further 45 redshifts have been measured of the long gamma-ray bursts discovered by other satellites. Here we perform five statistical tests comparing the redshift distributions of these two samples assuming as the null hypothesis an identical distribution for the two samples. Three tests (Student's *t*-test, Mann-Whitney test, Kolmogorov-Smirnov test) reject the null hypothesis at significance levels between 97.19 and 98.55%. Two different comparisons of the medians show extreme (99.78 − 99.99994)% significance levels of rejection. This means that the redshifts of the Swift sample and the redshifts of the non-Swift sample are distributed differently – in the Swift sample the redshifts are on average larger. This statistical result suggests that the long GRBs should on average be at the higher redshifts of the Swift sample.

1. Introduction

Recently, the Swift satellite (Gehrels et al. 2005) detected the gamma-ray burst GRB 050904, for which the redshift is $z = 6.29$, directly measured from the afterglow data (Haislip et al. 2005a,b; Antonelli et al. 2005; Kawai et al. 2005). The GRBs detected by the Swift satellite seem to have systematically larger average redshifts than the redshifts of GRBs detected by other satellites (Greiner 2005; Friedman & Bloom 2005). Does the redshift distribution of GRBs detected by Swift significantly differ from the other GRBs with known redshifts?

Using statistical analyses we show that this is the case. We briefly discuss the importance, meaning and the consequence of this statistical result on the redshift distribution of long GRBs. The detailed discussion of the astrophysical reasons for the obtained results is the topic of a forthcoming paper.

2. The samples

Greiner (2005) lists the observations concerning the afterglows of GRBs, and among them he also selects and lists the confirmed redshifts (see the references therein and also Friedman & Bloom 2005). For our statistical studies we used this survey and selection. There are 16 bursts with measured redshifts detected first by the Swift satellite, during the period 1 January–6 October 2005. These 16 redshifts define here the "Swift" sample, where the smallest redshift is $z = 0.26$ (GRB 050714, Prochaska et al. 2005); the largest redshift is $z = 6.29$ (GRB 050904, Kawai et al. 2005); the mean redshift is $\bar{z}_{\text{Swift}} = 2.42$; the variance is $\sigma^2_{\text{Swift}} = 2.60$; and the median is $z_{\text{Swift,med}} = 2.405$.

There are a further 45 redshifts in Greiner's table from the period 28 February 1997–6 October 2005. In the cases of GRB 011030X and GRB 980329 these are upper limits only, and these cases were not included in the sample; in three cases (GRB 020305; GRB 991216; GRB 980326) the redshifts were only estimated, but here we considered them; for GRB 000214 the estimated redshift is between 0.37 and 0.47, and here $z = 0.42$ was used (Antonelli et al. 2000). In this way we obtain the "non-Swift" sample containing 45 redshifts. In this sample the smallest redshift is $z = 0.0085$ (GRB 980425, Galama et al. 1998); the largest redshift is $z = 4.50$ (GRB 000131, Andersen et al. 2000); the mean redshift is $\bar{z}_{\text{non-Swift}} = 1.31$; the variance is $\sigma^2_{\text{non-Swift}} = 1.11$; the median is $z_{\text{non-Swift,med}} = 1.02$.

Both samples probably only contain long GRBs (>2s), which are likely different to short and intermediate bursts (Horváth 1998; Meegan et al. 2001; Horváth 2002; Balázs et al. 2003; Hakkila et al. 2003; Borgonovo 2004). Thus, this paper deals exclusively with long GRBs.

3. The statistical tests and their results

The two samples were compared using the parametric Student's *t*-test and non-parametric rank-based statistical tests: the Mann-Whitney test, Kolmogorov-Smirnov test and two median tests. The rank-based tests have the clear advantage of being unaffected by any monotonic transformation in the z values.

The most common test for the comparison of average redshifts is Student's *t*-test. The details of this test together with the calculation of the significance level can be found in Chap. 14.2 of Press et al. (1992). Since the variances are different, one has to use the modification of the test for unequal variances. The relevant formulas are given by Eqs. (14.2.3) and (14.2.4). One obtains $t = 2.57$ for the degree-of-freedom d.o.f. = 19.75. The Student's *t*-test's significance in rejecting the null hypothesis (i.e. that the two samples have identical mean values) is 98.15%.

The Mann-Whitney test uses, instead of the values of the redshifts, their ranks. Its meaning with the relevant formulas is explained, for instance, by Lowry (1999). One considers first the two samples together, and in this common sample, containing in our case 61 objects, one sorts the redshifts into a monotonic increasing sequence. Not the values of redshifts, but only the ranks will be used in the non-Swift and in the Swift sample. This means that the null hypothesis is the assumption that in both samples the mean ranks are identical, and the difference comes from chance.

In our case the Mann-Whitney test gives a 98.55% significance level, i.e. it rejects the null hypothesis almost at the same significance level as that of the Student's *t*-test.

The Kolmogorov-Smirnov test compares the cumulative distributions of the redshifts in the two samples. The details of the test are described in Chap. 14.3 of Press et al. (1992). It gives the significance of 97.19%, which means that this test rejects the null hypothesis at this level.

Not only the mean redshifts in the two samples, but also the medians could be different. To verify this expectation we provided the following two tests.

We considered the two samples with 61 redshifts, sorted in ascending series. There are 6 non-Swift and 8 Swift GRBs with redshifts larger that $z_{\mathrm{Swift,med}} = 2.405$; there are then 39 non-Swift and 8 Swift GRBs with redshifts smaller than $z_{\mathrm{Swift,med}}$. Thus, the median of Swift sample separates the non-Swift sample in a ratio 6 : 39, instead of the expected 1 : 1.

Assuming a binomial distribution with $N = 45$ and $p = 0.5$, the expected most probable value is $Np = 22.5$ (e.g. Mészáros 1997; Balázs et al. 1998) and the expected dispersion is $\sigma^2_{\mathrm{theor.}} = Np(1-p) = 11.25$; i.e. $\sigma_{\mathrm{theor.}} = 3.4$. The difference between the most expected and the real value is $39 - 22.5 = 16.5 = 4.9\sigma_{\mathrm{theor.}}$. Because both Np and $\sigma_{\mathrm{theor.}}$ are much greater than one, a Gaussian approximation of the binomial distribution is allowed. This means that the null hypothesis (i.e. the two medians are the same) is rejected at the 4.9 sigma level, corresponding to the significance 99.99994%.

A modification of this test compares the medians of the Swift and non-Swift samples as follows. We choose 16 objects randomly from the sample of $N = 61$ events. In this subsample

the mean between the 8th and 9th objects gives the median. Let this 8th (9th) object from the subsample be in the qth (sth) position in the ordered sample of N events ($q < s$). Obviously, it can be $8 \le q \le 53$, and $9 \le s \le 54$. Then the corresponding probability for the given q and s is simply the number of the good configurations divided by the number of all configurations: $P(q,s) = \binom{q-1}{7}\binom{N-s}{7} / \binom{N}{16}$, where for non-negative integers x and $y \ge x$ we have $\binom{y}{x} = y!/(x!(y-x)!)$. For the median of the Swift sample the corresponding significance level is $\sum_{(q+s)<95} P(q,s) = 99.78\%$. (The sum must be taken for all combinations with $(q+s)/2 < 47.5$, because for these combinations the medians of the chosen subsamples are smaller than the Swift's median.) The null hypothesis is again *rejected* at this level. We verified this high significance level with 100 000 Monte-Carlo simulations, and confirmed the above result.

4. Discussion and conclusions

All five statistical tests reject the null hypothesis that the redshift distributions for the Swift and non-Swift samples are identical at a significance higher than 95%. The redshifts of the Swift sample are on average larger than that of the non-Swift sample.

Three questions emerge: I. Is this statistical result an important new result? II. What is the reason for this behavior? III. What is the impact of this result on the redshift distribution of the long GRBs?

It may seem that the result is not unexpected and is "reasonable". This point of view may follow from the fact that the Swift satellite is more sensitive, and it allows for a faster detection of fainter bursts than the other satellites. Also the number of detected GRBs during a given time interval is larger. Hence, because Swift detects more and fainter GRBs, it is not unreasonable to expect that it also detects higher redshifts on average.

Concerning this point of view two important remarks are needed.

First, it is a triviality that either an exact confirmation or an exact rejection of a theoretical expectation – done purely from the observational data by strict statistical methods – is useful. It is not clear without a detailed statistical testing, that the redshifts of the Swift sample are on average larger than that of the non-Swift sample. For example, if one takes *only* six typical redshift values (minimal, mean and maximal redshifts from both samples) and one does *not* take also the variances, then one simply cannot obtain from these six redshifts *alone* the conclusion that the two distributions have different redshifts on average. The six typical redshifts *alone* are not enough for such a claim. It is not even clear what the "most natural" parameter of the distance is: z, $1+z$, $\log(1+z)$ or even the luminosity distance? All could be "good" parameters: however they will show totally different distributions (mean, variance, skewness, etc.), and it is hard to use parametric tests only. The rank based tests are clearly preferred as they have the clear advantage of being free of such problems. The relevance of our result for the redshift distribution of long GRBs is straightforward: it excludes the possibility that the Swift sample and the non-Swift sample of GRBs originate, on average, in the same redshift range. This answers question I.

Second, from the observational point of view, it is not yet certain that fainter, long bursts are on average at larger redshifts. The redshifts of long GRBs are still poorly measured. Beyond

a few dozen of directly measured redshifts – the ones studied in this paper – only indirect estimations are known for the redshifts of long GRBs. The estimates from the BATSE data using the gamma range alone (Mészáros & Mészáros 1996; Horváth et al. 1996; Reichart & Mészáros 1997; Lamb & Reichart 2000; Schmidt 2001; Schaefer et al. 2001; Lloyd-Ronning et al. 2002; Norris 2002; Bagoly et al. 2003; Atteia 2003; Lin et al. 2004) suggest that large redshifts (up to $z = 20$, Mészáros & Mészáros 1996) for fainter GRBs may be expected. Also the analyses from the afterglow data suggest (Berger et al. 2005) that the fainter, long GRBs on average should be at higher redshifts. The very high GRBs' redshifts are discussed in detail by Lamb & Reichart (2000); the cosmological aspects of GRBs – fainter ones expected to be at higher redshifts – are studied, e.g., by Friedman & Bloom (2005). Hence, the theoretical expectation "fainter = larger redshifts on average" is reasonable indeed. But, strictly, there is no observational certainty yet, but the observational verification of the theoretical expectation has been needed. This has been provided in this article using exact statistical arguments. Thus, question II may be answered as follows: The reason for this behavior *can be* instrumental. The instruments on the Swift satellite are more sensitive than the other GRB detectors, and the fainter GRBs should indeed on average be at higher redshifts.

Hence, we conjecture that the difference in the average redshifts originates – either partly or fully – from instrumental selection effects. For a more concrete answer a detailed study of the instrumental properties of the considered satellites is needed, which is planned to be provided in a forthcoming paper. It should quantify the importance of the instrumental biases.

For the impact on the redshift distributions, from the results discussed above three possibilities can be envisaged for the GRB distribution:

A. The long subgroup of GRBs is unique (no further subgrouping is needed), and its redshift distribution is represented by the non-Swift sample; the redshifts of the Swift sample are biased. We believe that this possibility is strongly disfavored, because the distribution of GRBs before the Swift era were strongly biased (Mészáros et al. 2004), and in the non-Swift sample only the brightest GRBs should be present. Hence, it is unlikely that the non-Swift sample represents the true intrinsic redshift distribution of the unique long GRBs (Mészáros et al. 2004). Berger et al. (2005) also reject this possibility.

B. The long subgroup of GRBs is unique, and its redshift distribution is represented by the Swift sample; the redshifts of the non-Swift sample are biased. It was shown (Jakobsson et al. 2006) that the Swift redshift data may represent the real intrinsic redshift distribution of long GRBs, but one cannot exclude the possibility that the exact GRB redshift distribution is neither properly represented by the Swift sample nor by the non-Swift sample, e.g. the redshift distributions of the BATSE sample (Mészáros & Mészáros 1996; Schaefer et al. 2001; Lloyd-Ronning et al. 2002; Bagoly et al. 2003) may be quite different to that of the Swift sample. Lin et al. (2004) even claim that the majority of GRB in the BATSE sample are at $z > 10$.

There are 16 Swift GRBs with measured z and optical transient (OT) from the 32 Swift GRBs with OT observation (50%), while for the non-Swift sample there are 38 GRBs with z and OT from the total of 65 GRBs with OT (58%, Greiner 2005). Seven non-Swift events have measured z without an OT. The remarkably higher OT detection rate for the Swift bursts ($\simeq 43\%$ vs. the non-Swift $\simeq 29\%$) indicates that the lower redshift measurement success ratio could be an indication of an observational bias from the cosmic reddening: i.e. for the significantly higher Swift redshifts a spectral line may drop out of the *V* and *R* bands, and are detectable only in the *J*, *K* bands or at even longer wavelengths. This clearly reduces the success of the redshift determination. The OT search efficiency is clearly <100% due to observational constraints, so the $\simeq 43\%$ rate of the OT detection in the Swift era may suggest that the majority of the GRBs are within the optically transparent cosmological region.

C. The long subgroup of GRBs is not necessarily unique, i.e. the different redshifts in our samples may originate from different types of sources with different redshift distributions: e.g. the non-uniqueness of the long subgroup is also an alternative (Borgonovo 2004). Also Berger et al. (2005) consider, as a possibility, that the Swift GRBs with higher redshifts represent a luminous subgroup of the long class alone. On the other hand, recent statistical analyses show that three subgroups of GRBs are enough to explain the statistical properties of the BATSE sample (Horváth et al. 2006). Hence the long subclass should not be further divided. This means that the observed differences might be accounted for by the different observational strategies of the different experiments.

However, the different energy sensitivity of the different spacecrafts cannot satisfactorily explain the differences in the redshifts. The Swift and non-Swift trigger energy ranges overlap: in the non-Swift samples there are 27 "softer" and 6 "harder" GRBs ("softer" and "harder" mean that the observed trigger energy is below or above the average Swift trigger energy), while 12 GRBs have the trigger energy within the Swift band. Only $\simeq 13\%$ of GRBs were triggered above the Swift energy band, and the median z of this small subsample is higher ($z_{med} = 1.64$) than the non-Swift median $z_{non-Swift,med} = 1.02$. The higher trigger energy may select GRBs of higher redshifts, however, this effect could be caused by the small sample size.

Thus, in our opinion, both the B and C cases are possible, and a further detailed study is needed. This we will provide in a separate paper. In answer to question III, however, our statistical arguments relying on the observational data alone suggest that the average redshifts of the long GRBs can be very large ($z \simeq (2.4-2.6)$ or even larger).

References

Antonelli, L. A., Piro, L., Vietri, M., et al. 2000, ApJ, 545, L39
Antonelli, L. A., Grazian, A., D'Avanzo, P., et al. 2005, GCN Circ., 3924
Atteia, J.-L. 2003, A&A, 407, L1
Andersen, M. I., Hjorth, J., Pedersen, H., et al. 2000, A&A, 364, L54
Bagoly, Z., Csabai, I., Mészáros, A., et al. 2003, A&A, 398, 919
Balázs, L. G., Mészáros, A., & Horváth, I. 1998, A&A, 339, 1
Balázs, L. G., Bagoly, Z., Horváth, I., Mészáros, A., & Mészáros, P. 2003, A&A, 401, 129
Berger, E., et al. 2005 [arXiv:astro-ph/0505107]
Borgonovo, L. 2004, A&A, 418, 487
Friedman, A. S., & Bloom, J. S. 2005, ApJ, 627, 1
Galama, T. J., Vreeswijk, P. M., Pian, E., et al. 1998, IAU Circ., 6895, 1
Gehrels, N., Chincarini, G., Giommi, P., et al. 2004, ApJ, 611, 1005
Greiner, J. 2005, http://www.mpe.mpg.de/~jcg/grbgen.html
Haislip, J., Reichart, D., Cypriano, E., et al. 2005a, GCN Circ., 3914
Haislip, J., Nysewander, M., Reichart, D., et al. 2005b, GCN Circ., 3919
Hakkila, J., Giblin, T. W., Roiger, R. J., et al. 2003, ApJ, 582, 320

Horváth, I. 1998, ApJ, 508, 757
Horváth, I. 2002, A&A, 392, 791
Horváth, I., Mészáros, P., & Mészáros, A. 1996, ApJ, 470, 56
Horváth, I., Balázs, L. G., Bagoly, Z., Ryde, F., & Mészáros, A. 2006, A&A, 447, 23
Jakobsson, P., Levan, A., Fynbo, J. P. U., et al. 2006, A&A, 447, 897
Kawai, N., Yamada, T., Kosugi, G., et al. 2005, GCN Circ., 3937
Lamb, D. Q., & Reichart, D. E. 2000, ApJ, 536, 1
Lin, J. R., Zhang, S. N., & Li, T. P. 2004, ApJ, 605, 819
Lloyd-Ronning, N. M., Fryer, C. L., & Ramirez-Ruiz, E. 2002, ApJ, 574, 554
Lowry, R. 1999, `http://faculty.vassar.edu/lowry/ch11a.html`
Meegan, C. A., et al. 2001, Current BATSE Gamma-Ray Burst Catalog, `http://gammaray.msfc.nasa.gov/batse/grb/catalog`

Mészáros, A. 1997, A&A, 328, 1
Mészáros, A., & Mészáros, P. 1996, ApJ, 466, 29
Mészáros, A., Bagoly, Z., Balázs, L. G., Horváth, I., & Klose, S. 2004, in Proc. of the Third Roma Symposium Gamma-Ray Bursts in the Afterglow Era, ed. M. Feroci, F. Frontera, N. Masetti, & L. Piro (ASP Publ. Co.), 118
Norris, J. P. 2002, ApJ, 579, 386
Press, W. H., et al. 1992, Numerical Recipes, Second edition (Cambridge: Cambridge Univ. Press)
Prochaska, J. X., Chen, H.-W., Bloom, J. S., et al. 2005, GCN Circ., 3679
Reichart, D. E., & Mészáros, P. 1997, ApJ, 483, 597
Schaefer, B. E., Deng, M., & Band, D. L. 2001, ApJ, 563, L123
Schmidt, M. 2001, ApJ, 552, 36