Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY
MNRAS **458,** 2024–2031 (2016)
Advance Access publication 2016 February 24
doi:10.1093/mnras/stw429

Analysis of gamma-ray burst duration distribution using mixtures of skewed distributions

M. Tarnopolski★
Astronomical Observatory, Jagiellonian University, Orla 171, PL-30-244 Kraków, Poland

Accepted 2016 February 22. Received 2016 February 22; in original form 2015 October 27

ABSTRACT

Two classes of gamma-ray bursts (GRBs) have been confidently identified thus far and are prescribed to different physical scenarios – neutron star-neutron star or neutron star-black hole mergers, and collapse of massive stars, for short and long GRBs, respectively. A third, intermediate in duration class, was suggested to be present in previous catalogues, such as Burst Alert and Transient Source Explorer (BATSE) and *Swift*, based on statistical tests regarding a mixture of two or three lognormal distributions of T_{90}. However, this might possibly not be an adequate model. This paper investigates whether the distributions of $\log T_{90}$ from BATSE, *Swift*, and *Fermi* are described better by a mixture of skewed distributions rather than standard Gaussians. Mixtures of standard normal, skew-normal, sinh-arcsinh and alpha-skew-normal distributions are fitted using a maximum likelihood method. The preferred model is chosen based on the Akaike information criterion. It is found that mixtures of two skew-normal or two sinh-arcsinh distributions are more likely to describe the observed duration distribution of *Fermi* than a mixture of three standard Gaussians, and that mixtures of two sinh-arcsinh or two skew-normal distributions are models competing with the conventional three-Gaussian in the case of BATSE and *Swift*. Based on statistical reasoning, and it is shown that other phenomenological models may describe the observed *Fermi*, BATSE, and *Swift* duration distributions at least as well as a mixture of standard normal distributions, and the existence of a third (intermediate) class of GRBs in *Fermi* data is rejected.

Key words: methods: data analysis – methods: statistical – gamma-rays: general.

1 INTRODUCTION

Mazets et al. (1981) first pointed out hints for a bimodal distribution of T_b (taken to be the time interval within which fall 80–90 per cent of the measured GRB's intensity) drawn for 143 events detected in the KONUS experiment. A bimodal structure in the distribution of durations T_{90} (time interval from 5 per cent to 95 per cent of the accumulated fluence) in Burst Alert and Transient Source Explorer (BATSE, onboard the Compton Gamma-Ray Observatory; Meegan et al. 1992) 1B data set, based on which gamma-ray bursts (GRBs) are nowadays commonly classified into short ($T_{90} < 2$ s) and long ($T_{90} > 2$ s) classes, was also found (Kouveliotou et al. 1993).While generally short GRBs are of merger origin and long ones come from collapsars, this classification is imperfect due to a large overlap in duration distributions of the two populations (Lü et al. 2010; Bromberg, Nakar & Piran 2011; Bromberg et al. 2013; Shahmoradi 2013; Shahmoradi & Nemiroff 2015; Tarnopolski 2015a,b).

Horváth (1998) discovered a third peak in the duration distribution, located between the short and long ones, in the BATSE 3B catalogue, and using multivariate clustering procedures independently the same conclusion was arrived at by Mukherjee et al. (1998). The statistical existence of the intermediate class was supported (Horváth 2002) with the use of BATSE 4B data. The evidence for a third normal component in $\log T_{90}$ was found also in *Swift* Burst Alert Telescope (BAT) data (Horváth et al. 2008, 2010; Zhang & Choi 2008; Huja, Mészáros & Řípa 2009). Other data sets, i.e. *RHESSI* (Řípa et al. 2009, 2012) and *BeppoSAX* (Horváth 2009), were both in agreement with earlier results regarding the bimodal distribution, and the detection of a third component was established on a lower, compared to BATSE and *Swift*, significance level. Hence, four different satellites provided hints about the existence of a third class of GRBs. Contrary to this, durations as observed by *INTEGRAL* have a unimodal distribution, which extends to the shortest time-scales as a power law (Savchenko, Neronov &

★E-mail: mariusz.tarnopolski@uj.edu.pl

Courvoisier 2012). Interestingly, a re-examination of the BATSE current catalogue and *Swift* data set (Zitouni et al. 2015), showed that a mixture of three-Gaussians (3-Gs) fits the *Swift* data better than a two-Gaussian (2-G), while in the BATSE case statistical tests did not support the presence of a third component.[1]

Only one data set (BATSE 3B) was truly trimodal in the sense of having three peaks (i.e. three local maxima). In the rest (i.e. BATSE 4B and current, *Swift*, *RHESSI* and *BeppoSAX*) a 3-G was found to follow the observations better than a 2-G, but those fits yielded only two peaks, so despite statistical analyses support the presence of a third normal *component*, the existence of a third physical *class* is not confirmed and may be ascribed to $\log T_{90}$ being described by a distribution different than a mixture of Gaussians, particularly a mixture of skewed distributions (Tarnopolski 2015c).

Latest numerous release is due to *Fermi* Gamma-ray Burst Monitor (GBM) observations (Gruber et al. 2014; von Kienlin et al. 2014) and consists of ∼1600 GRBs with computed durations T_{90}. Up to date, to the best of the author's knowledge, except Tarnopolski (2015c), only Horváth et al. (2012), Zhang et al. (2012), and Qin et al. (2013) conducted research on a *Fermi* subsample, consisting of 425 bursts, from the first release of the catalogue.

It was proposed (Tarnopolski 2015c), in the light of Zitouni et al. (2015), who suggested that the non-symmetry of the $\log T_{90}$ distributions is due to a non-symmetric distribution of the envelope masses of the progenitors, that a mixture of skewed distributions might be phenomenologically a better model than the commonly applied mixture of standard Gaussians. The aim of this paper is to examine whether mixtures of various skewed distributions [skew-normal (SN), sinh-arcsinh (SAS), and alpha-skew-normal (ASN)] describe the duration distribution better than a mixture of standard Gaussians. Particularly, it is verified whether two-component mixtures of skewed distributions might challenge a commonly applied 3-G model. If this is shown to be true, the existence of the intermediate class in the duration distribution will be questioned.

Because the T_{90} distribution is detector dependent (Nakar 2007; Tarnopolski 2015a), the analysis herein is not restricted to the *Fermi* data set as it was in (Tarnopolski 2015c), but also the BATSE and *Swift* data are examined. These three data sets have been fitted to date with a mixture of standard Gaussians, but to the best of the author's knowledge no other types of distributions were applied to the observed T_{90} distributions. It may happen that due to instrument specification a three-component distribution might be a better description for some data sets, while for others a two-component one will be sufficient (Zitouni et al. 2015).

This article is organized as follows. Section 2 describes the data sets, fitting method, the properties of the distributions examined, and the method of assessing the goodness of fit. In Section 3, the results are presented. Section 4 is devoted to discussion, and Section 5 gives concluding remarks. The computer algebra system MATHEMATICA® v10.0.2 is applied throughout this paper.

[1] Adding parameters to a nested model always results in a better fit (in the sense of a lower χ^2 or a higher maximum log-likelihood, $\mathcal{L}$) due to more freedom given to the model to follow the data, i.e. due to introducing more free parameters. The important question is whether this improvement is statistically significant, and whether the model is justified.

2 DATA SETS, METHODS, AND DISTRIBUTIONS

2.1 Samples

The data sets[2] from *Fermi*,[3] BATSE,[4] and *Swift*[5] are considered herein. The BATSE current catalogue consists of 2041 GRBs, and the *Swift* data set contains 914 events. *Fermi* observed 1596 GRBs, but a data set of 1593 GRBs is used. Three durations that stand out (two shortest and one longest) were treated as outliers and excluded due to their significant separation from the remaining durations and a possibility of a strong influence on the outcome, especially on the tails of the fitted distributions (if the data are binned according to any well-established rule (see Section 2.2), the bins containing these three values are separated by empty bins from the rest of the distribution). Whereas the durations T_{90} are approximately lognormally distributed, herein their decimal logarithms, $\log T_{90}$'s, are employed; for simplicity they will be referred to as *durations* too, and whenever a phrase *normal distribution of durations* is used, it is understood in the sense of *normal distribution of logarithms of durations* ($\log T_{90}$) or, equivalently, *lognormal distribution of durations* T_{90}. This notion applies also to other distributions examined throughout this paper.

The *RHESSI* and *BeppoSAX* data sets are not examined here for the following reasons: (i) *RHESSI* has no GRB triggering and only consists of GRBs observed by other satellites; (ii) *RHESSI* is a relatively small data set (427 GRBs; Řípa et al. 2009, 2012); (iii) *BeppoSAX*, due to its relatively long (1 s) short integration time (Horváth 2009), does not contain many short GRBs.

2.2 Fitting method

Two standard fitting techniques are commonly applied: χ^2 fitting and maximum likelihood (ML) method. For the first, data need to be binned, and despite various binning rules are known (e.g. Freedman–Diaconis, Scott, Knuth etc.), they still leave place for ambiguity, as it might happen that the fit may be statistically significant on a given significance level for a number of binnings (Huja & Řípa 2009; Koen & Bere 2012; Tarnopolski 2015c). The ML method is not affected by this issue and is therefore applied herein. However, for display purposes, the binnings were chosen based on the Knuth rule.

Having a distribution with a probability density function (PDF) given by $f = f(x; \theta)$ (possibly a mixture), where $\theta = \{\theta_i\}_{i=1}^{p}$ is a set of p parameters, the log-likelihood function is defined as

$$\mathcal{L}_p(\theta) = \sum_{i=1}^{N} \ln f(x_i; \theta), \tag{1}$$

where $\{x_i\}_{i=1}^{N}$ are the data points from the sample to which a distribution is fitted. The fitting is performed by searching a set of parameters $\hat{\theta}$ for which the log-likelihood is maximized (Kendall & Stuart 1973). When nested models are considered, the maximal value of the log-likelihood function $\mathcal{L}_{\max} \equiv \mathcal{L}_p(\hat{\theta})$ increases when the number of parameters p increases.

[2] All accessed on 2015 April 29.

[3] http://heasarc.gsfc.nasa.gov/W3Browse/fermi/fermigbrst.html (Gruber et al. 2014; von Kienlin et al. 2014).

[4] http://gammaray.msfc.nasa.gov/batse/grb/catalog/current (Meegan et al. 1998; Paciesas et al. 1999).

[5] http://swift.gsfc.nasa.gov/archive/grb_table/ (Gehrels et al. 2004).

2.3 Distributions and their properties

The following distributions are considered.

A mixture of k standard normal (Gaussian) $\mathcal{N}(\mu, \sigma^2)$ distributions:

$$f_k^{(\mathcal{N})}(x) = \sum_{i=1}^{k} A_i \varphi\left(\frac{x-\mu_i}{\sigma_i}\right) = \sum_{i=1}^{k} \frac{A_i}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right), \tag{2}$$

being described by $3k-1$ free parameters: k pairs (μ_i, σ_i) and $k-1$ weights A_i, satisfying $\sum_{i=1}^{k} A_i = 1$. Skewness of each component is $\gamma_1^{(\mathcal{N})} = 0$.

A mixture of k SN distributions (O'Hagan & Leonard 1976; Azzalini 1985):

$$f_k^{(\mathrm{SN})}(x) = \sum_{i=1}^{k} 2A_i \varphi\left(\frac{x-\mu_i}{\sigma_i}\right) \Phi\left(\alpha_i \frac{x-\mu_i}{\sigma_i}\right) = \sum_{i=1}^{k} \frac{2A_i}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right) \times \frac{1}{2}\left[1 + \mathrm{erf}\left(\alpha_i \frac{x-\mu_i}{\sqrt{2}\sigma_i}\right)\right], \tag{3}$$

described by $4k-1$ parameters. Skewness of an SN distribution is

$$\gamma_1^{(\mathrm{SN})} = \frac{4-\pi}{2} \frac{\left(\zeta\sqrt{2/\pi}\right)^3}{\left(1-2\zeta^2/\pi\right)^{3/2}},$$

where $\zeta = \frac{\alpha}{\sqrt{1+\alpha^2}}$, hence the skewness $\gamma_1^{(\mathrm{SN})}$ is solely based on the shape parameter α, and is limited roughly to the interval $(-1, 1)$. The mean is given by $\mu + \sigma\zeta\sqrt{\frac{2}{\pi}}$. When $\alpha = 0$, the SN distribution is reduced to a standard Gaussian $\mathcal{N}(\mu, \sigma^2)$ due to $\Phi(0) = 1/2$.

A mixture of k SAS distributions (Jones & Pewsey 2009):

$$f_k^{(\mathrm{SAS})}(x) = \sum_{i=1}^{k} \frac{A_i}{\sigma_i}\left[1 + \left(\frac{x-\mu_i}{\sigma_i}\right)^2\right]^{-\frac{1}{2}} \times \beta_i \cos h\left[\beta_i \sin \mathrm{h}^{-1}\left(\frac{x-\mu_i}{\sigma_i}\right) - \delta_i\right] \times \exp\left[-\frac{1}{2}\sin \mathrm{h}\left[\beta_i \sin \mathrm{h}^{-1}\left(\frac{x-\mu_i}{\sigma_i}\right) - \delta_i\right]^2\right], \tag{4}$$

being described by $5k-1$ parameters. It turns out that skewness of the SAS distribution increases with increasing δ, positive skewness corresponding to $\delta > 0$. Tailweight decreases with increasing β, $\beta < 1$ yielding heavier tails than the normal distribution, and $\beta > 1$ yielding lighter tails. With $\delta = 0$ and $\beta = 1$, the SAS distribution reduces to a standard Gaussian, $\mathcal{N}(\mu, \sigma^2)$. Skewness of a SAS distribution is

$$\gamma_1^{(\mathrm{SAS})} = \frac{1}{4}\left[\sin \mathrm{h}\left(\frac{3\delta}{\beta}\right) P_{3/\beta} - 3\sin \mathrm{h}\left(\frac{\delta}{\beta}\right) P_{1/\beta}\right],$$

where

$$P_q = \frac{e^{1/4}}{\sqrt{8\pi}}\left[K_{(q+1)/2}(1/4) + K_{(q-1)/2}(1/4)\right].$$

Here, K is the modified Bessel function of the second kind. The mean is given by $\mu + \sigma \sinh(\delta/\beta) P_{1/\beta}$.

A mixture of k ASN distributions (Elal-Olivero 2010):

$$f_k^{(\mathrm{ASN})}(x) = \sum_{i=1}^{k} A_i \frac{\left(1-\alpha_i\frac{x-\mu_i}{\sigma_i}\right)^2 + 1}{2+\alpha_i^2} \varphi\left(\frac{x-\mu_i}{\sigma_i}\right) = \sum_{i=1}^{k} A_i \frac{\left(1-\alpha_i\frac{x-\mu_i}{\sigma_i}\right)^2 + 1}{2+\alpha_i^2} \frac{1}{\sqrt{2\pi}\sigma_i} \times \exp\left(-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right), \tag{5}$$

described by $4k-1$ parameters. Skewness of an ASN distribution is

$$\gamma_1^{(\mathrm{ASN})} = \frac{12\alpha^5 + 8\alpha^3}{(3\alpha^4 + 4\alpha^2 + 4)^{3/2}},$$

and is limited roughly to the interval $(-0.811, 0.811)$. The mean is given by $\mu - \frac{2\alpha\sigma}{2+\alpha^2}$. For $\alpha \in (-1.34, 1.34)$ the distribution is unimodal, and bimodal otherwise.

2.4 Assessing the likelihood of the fits

If one has two fits such that $\mathcal{L}_{p_2,\mathrm{max}} > \mathcal{L}_{p_1,\mathrm{max}}$, then twice their difference, $2\Delta\mathcal{L}_{\mathrm{max}} = 2(\mathcal{L}_{p_2,\mathrm{max}} - \mathcal{L}_{p_1,\mathrm{max}})$, is distributed like $\chi^2(\Delta p)$, where $\Delta p = p_2 - p_1 > 0$ is the difference in the number of parameters (Kendall & Stuart 1973; Horváth 2002). If a p-value associated with the value of $\chi^2(\Delta p)$ does not exceed the significance level α, one of the fits (with higher $\mathcal{L}_{\mathrm{max}}$) is statistically better than the other. For instance, for a 2-G and a 3-G, $\Delta p = 3$, and despite that, according to Footnote 1, $\mathcal{L}_{\mathrm{max,\,3-G}} > \mathcal{L}_{\mathrm{max,\,2-G}}$ holds always, twice their difference provides a decisive p-value.

It is crucial to note that it follows from the above description that this method is not suitable for situations when the model with the higher $\mathcal{L}_{\mathrm{max}}$ has fewer parameters, i.e. $\Delta\mathcal{L}_{\mathrm{max}} > 0$ and $\Delta p < 0$. Moreover, while all of the skewed distributions considered herein contain the standard Gaussian as their special case, what makes them nested models, but e.g. the SAS and SN distributions are not nested, hence no direct comparison can be performed for them with this approach.

For nested as well as non-nested models, the Akaike information criterion (AIC; Akaike 1974; Burnham & Anderson 2004; Biesiada 2007; Liddle 2007) may be applied. The AIC is defined as

$$\mathrm{AIC} = 2p - 2\mathcal{L}_{\mathrm{max}}. \tag{6}$$

A preferred model is the one that minimizes AIC. The formulation of AIC penalizes the use of an excessive number of parameters, hence discourages overfitting. It prefers models with fewer parameters, as long as the others do not provide a substantially better fit. The expression for AIC consists of two competing terms: the first measuring the model complexity (number of free parameters) and the second measuring the goodness of fit (or more precisely, the lack of thereof). Among candidate models with AIC_i, let $\mathrm{AIC}_{\mathrm{min}}$ denote the smallest. Then,

$$Pr_i = \exp\left(-\frac{\Delta_i}{2}\right), \tag{7}$$

where $\Delta_i = \mathrm{AIC}_i - \mathrm{AIC}_{\mathrm{min}}$, can be interpreted as the relative (compared to $\mathrm{AIC}_{\mathrm{min}}$) probability that the i-th model minimizes the AIC.

What is essential in assessing the goodness of a fit in the AIC method is the difference, $\Delta_i = \mathrm{AIC}_i - \mathrm{AIC}_{\mathrm{min}}$, not the absolute

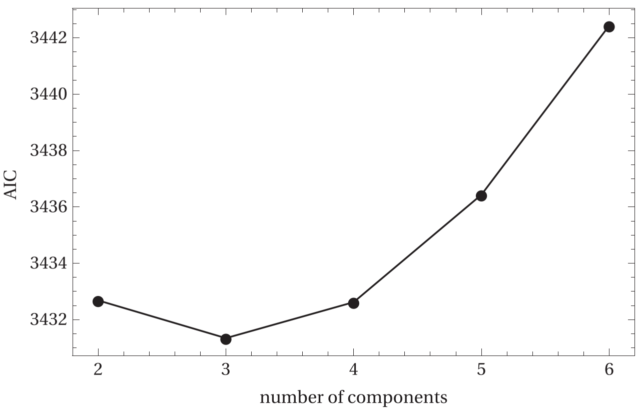

Figure 1. AIC versus number of components in a mixture of standard normal distributions. The minimal value corresponds to a 3-G.

value[6] of an AIC_i. If $\Delta_i < 2$, then there is substantial support for the i-th model (or the evidence against it is worth only a bare mention), and the proposition that it is a proper description is highly probable. If $2 < \Delta_i < 4$, then there is strong support for the i-th model. When $4 < \Delta_i < 7$, there is considerably less support, and models with $\Delta_i > 10$ have essentially no support (Burnham & Anderson 2004; Biesiada 2007). It is important to note that when two models with similar $\mathcal{L}_{\max}$ are considered, the Δ_i depends solely on the number of parameters due to the $2p$ term in equation (6). Hence, when $\Delta_i/2\Delta p < 1$, the relative improvement is due to actual improvement of the fit, not to increasing the number of parameters only.

Finally, AIC tries to select a model that most adequately describes reality (in the form of the data under examination). This means that in fact the model being a real description of the data is never considered.

3 RESULTS

3.1 Finding the number of components – standard Gaussian case

First, a mixture of standard Gaussians given by equation (2) is fitted using the ML method, i.e. maximizing $\mathcal{L}$ given by equation (1). The mixtures range from $k = 2$ to $k = 6$ components. The AIC is calculated by means of equation (6). The preferred model is the one with the lowest AIC, and it follows from Fig. 1 that among the Gaussian models examined, a mixture of three components is the most plausible to describe the observed distribution of *Fermi* $\log T_{90}$. The same conclusion is drawn for the BATSE and *Swift* data sets. Hence, as it is expected that the other PDFs [SN, SAS, and ASN given by equation (3)–(5)] will be more flexible in fitting the data, the forthcoming analysis is restricted to two or three components for distributions being a mixture of unimodal PDFs (SN and SAS), and to one, two, or three components for ASN, as its one bimodal component may turn out to follow the data well enough.

[6] The AIC value contains scaling constants coming from the log-likelihood $\mathcal{L}$, and so Δ_i are free of such constants (Burnham & Anderson 2004). One might consider $\Delta_i = \mathrm{AIC}_i - \mathrm{AIC}_{\min}$ a rescaling transformation that forces the best model to have $\Delta_{\min} := 0$.

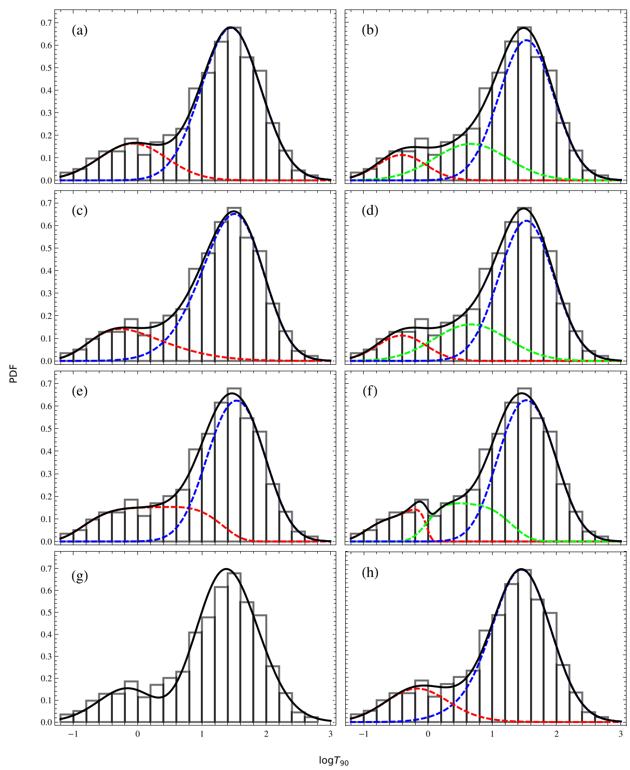

Figure 2. Distributions fitted to $\log T_{90}$ data gathered by *Fermi*. Colour dashed curves are the components of the (black solid) mixture distribution. The panels show a mixture of (a) two standard Gaussians, (b) three standard Gaussians, (c) 2-SN, (d) 3-SN, (e) 2-SAS, (f) 3-SAS, (g) 1-ASN, and (h) 2-ASN distributions.

3.2 Fitting the distributions

3.2.1 Fermi

The following distributions are examined: a 2-G and 3-G, a two- and three-SN (2-SN and 3-SN), a two- and three-SAS (2-SAS and 3-SAS), a one- and two-ASN (1-ASN and 2-ASN). The results in graphical form are displayed in Fig. 2, whereas the fitted parameters are gathered in Table 1, which contains also the values of $\mathcal{L}_{\max}$, AIC and relative probability, given by equation (1), (6), and (7), respectively. For completeness, a mixture of three ASN distributions was also fitted to the data, and turned out to be the worst among the fits obtained, with AIC = 3496.548 (i.e. higher by about 40 than the highest AIC, corresponding to a 1-ASN, from Table 1). To visualize the relative goodness-of-fits, the values of AIC and the relative probabilities are shown in Fig. 3. The minimal AIC is obtained by a 2-SN distribution. There is also a 37.7 per cent probability that a 2-SAS distribution describes the data. Both distributions consist of two components and are bimodal. The third lowest AIC was attained by a 3-G distribution with a probability of being correct equal to 21 per cent (corresponding to $\Delta_{3-\mathrm{G}} = 3.119$, which is a somewhat weaker support than the 2-SAS has relative to a 2-SN). While the 2-G exhibits a significant 10.8 per cent probability of being the correct distribution, it is only the fourth among the eight tested, with considerably less support (i.e. $\Delta_{2-\mathrm{G}} = 4.459$). The remaining four (2-ASN, 3-SAS, 3-SN, and 1-ASN) have only a few per cent of chance for describing the duration distribution, therefore are unlikely to be a proper model.

Table 1. Parameters of the fits to the *Fermi* data. Label corresponds to labels from Fig. 2. The smallest AIC is marked in bold, and p is the number of parameters in a model.

Label	Dist.	i	μ_i	σ_i	α_i	δ_i	β_i	A_i	$\mathcal{L}_{\max}$	AIC	ΔAIC	Pr	p
(a)	2-G	1	− 0.073	0.525	–	–	–	0.215	− 1711.342	3432.683	4.459	0.108	5
		2	1.451	0.463	–	–	–	0.785					
(b)	3-G	1	− 0.409	0.379	–	–	–	0.107	− 1707.672	3431.343	3.119	0.210	8
		2	0.668	0.570	–	–	–	0.231					
		3	1.530	0.426	–	–	–	0.662					
(c)	2-SN	1	− 0.735	0.954	2.819	–	–	0.208	− 1707.112	**3428.224**	0	1	7
		2	1.865	0.664	− 1.507	–	–	0.792					
(d)	3-SN	1	− 0.415	0.379	0.019	–	–	0.107	− 1707.672	3437.343	9.119	0.010	11
		2	0.726	0.573	− 0.127	–	–	0.231					
		3	1.515	0.426	0.044	–	–	0.662					
(e)	2-SAS	1	1.537	0.468	–	− 0.014	1.068	0.685	− 1706.089	3430.177	1.953	0.377	9
		2	2.158	6.146	–	− 2.367	7.756	0.315					
(f)	3-SAS	1	0.434	1.063	–	0.370	2.111	0.214	− 1704.248	3436.497	8.273	0.016	14
		2	0.473	0.402	–	− 4.161	2.680	0.111					
		3	1.529	0.468	–	0.020	1.087	0.675					
(g)	1-ASN	1	0.744	0.590	− 1.712	–	–	1	− 1725.038	3456.077	27.853	$<10^{-6}$	3
(h)	2-ASN	1	0.087	0.499	0.535	–	–	0.186	− 1710.427	3434.853	6.629	0.036	7
		2	1.150	0.483	− 0.667	–	–	0.814					

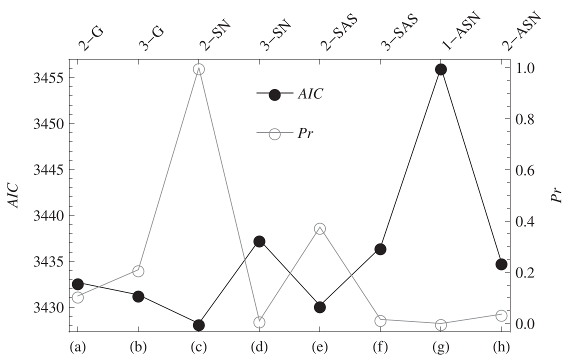

Figure 3. AIC and relative probability (Pr) for the *Fermi* models.

3.2.2 BATSE and Swift

The results are slightly different for the BATSE and *Swift* data sets, and are displayed in graphical form in Figs 4 and 5. Here, instead of fitting a 1-ASN and a 2-ASN, a 2-ASN and a 3-ASN distributions are fitted, because the 1-ASN yielded an AIC so large that a comparison with other models would be uninsightful.[7] For both samples, the minimal AIC is obtained for a mixture of three standard Gaussians, hence the results of all the previous analyses are confirmed (Horváth 2002; Horváth et al. 2008; Zhang & Choi 2008; Horváth 2009; Huja & Řípa 2009; Huja et al. 2009; Zitouni et al. 2015). However, for the second best models (2-SAS and 2-SN for BATSE and *Swift*, respectively), the ΔAIC is $\approx$1, corresponding to a relative probability of 57.9 per cent and 63.2 per cent for BATSE and *Swift*, respectively (see Tables 2 and 3). This is a substantial

[7] For BATSE, $\mathrm{AIC}_{1-\mathrm{ASN}} = 5029.805$, being higher by almost 100 than the highest AIC, corresponding to a 3-ASN, and for *Swift* $\mathrm{AIC}_{1-\mathrm{ASN}} = 2029.240$, being by about 4 bigger than the highest AIC (also corresponding to a 3-ASN), and by almost 35 higher than the lowest AIC (attained for a 3-G); compare with Tables 2 and 3.

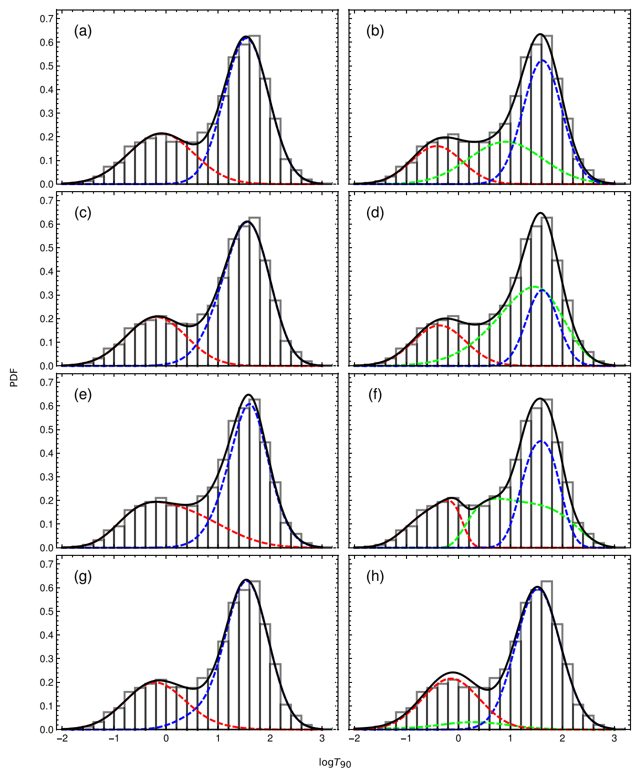

Figure 4. Distributions fitted to $\log T_{90}$ data from the BATSE current catalogue. Colour dashed curves are the components of the (black solid) mixture distribution. The panels show a mixture of (a) two standard Gaussians, (b) three standard Gaussians, (c) 2-SN, (d) 3-SN, (e) 2-SAS, (f) 3-SAS, (g) 2-ASN, and (h) 3-ASN distributions.

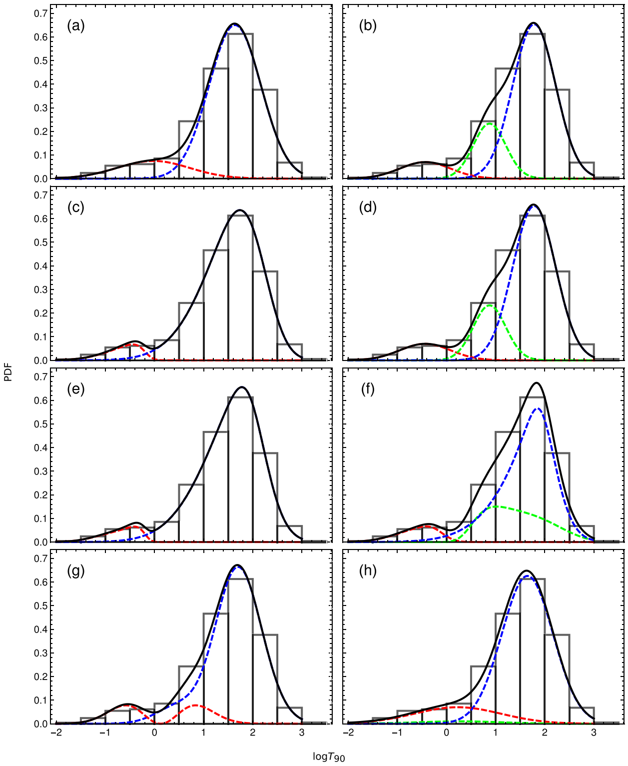

Figure 5. Distributions fitted to $\log T_{90}$ data observed by *Swift*. Colour dashed curves are the components of the (black solid) mixture distribution. The panels show a mixture of (a) two standard Gaussians, (b) three standard Gaussians, (c) 2-SN, (d) 3-SN, (e) 2-SAS, (f) 3-SAS, (g) 2-ASN, and (h) 3-ASN distributions.

support for these two-component models (Burnham & Anderson 2004; Biesiada 2007), hence they cannot be ruled out (see also Figs 6 and 7). The next lowest, i.e. third and fourth, AIC for the BATSE data correspond to a 2-ASN and a 2-G, while the *Swift* data set is well described by a 2-SAS or 2-ASN distribution. The rest of the models examined have a relative probability of being a better description of the data than a 3-G distribution less than 10 per cent. The 3-ASN has a negligible relative probability for both data sets.

4 DISCUSSION

Since (Horváth 1998), fitting a mixture of standard (i.e. non-skewed) Gaussians to the duration distribution of GRBs is a common practice. Nearly all of the catalogues examined showed that a 3-G fit is statistically more significant than a 2-G. This has been the basis of justifying the possibility of a third, intermediate in duration, class of GRBs, but might be ascribed simply to a higher flexibility of the fitted PDF due to a noticeably higher number of parameters. In many works, a model consisting of 3-Gs was called a trimodal, what is incorrect, as a trimodal distribution is characterized by three modes, hence three peaks recognized through local maxima (Schilling, Watkins & Watkins 2002). This was the case only in the BATSE 3B data set (Horváth 1998), where 797 GRBs were examined. However, in BATSE current catalogue ($\sim$2000 GRBs), no such structure was detected (Horváth 2002; Zitouni et al. 2015) – it appears that the peak related to an intermediate class was smeared out when more data was gathered. Other catalogues, e.g. *Swift*, also exhibit a bimodal distribution, although apparently skewed. The presumed intermediate class was proposed to be linked to X-ray flares, or are related to long GRBs through some physically meaningful parameters or set of parameters (Veres et al. 2010). Recently it was suggested (Zitouni et al. 2015) that the duration distribution

Table 2. Parameters of the fits to the BATSE data. Label corresponds to labels from Fig. 4. The smallest AIC is marked in bold, and p is the number of parameters in a model.

Label	Dist.	i	μ_i	σ_i	α_i	δ_i	β_i	A_i	$\mathcal{L}_{\max}$	AIC	ΔAIC	Pr	p
(a)	2-G	1	−0.095	0.627	–	–	–	0.336	−2448.329	4906.659	3.844	0.146	5
		2	1.544	0.429	–	–	–	0.664					
(b)	3-G	1	−0.420	0.487	–	–	–	0.196	−2443.407	**4902.815**	0	1	8
		2	0.907	0.705	–	–	–	0.316					
		3	1.615	0.372	–	–	–	0.488					
(c)	2-SN	1	−0.193	0.578	0.001	–	–	0.300	−2446.991	4907.981	5.166	0.076	7
		2	1.889	0.609	−1.351	–	–	0.700					
(d)	3-SN	1	−0.372	0.505	0.019	–	–	0.217	−2443.016	4908.033	5.218	0.074	11
		2	1.575	0.307	0.152	–	–	0.539					
		3	1.972	0.982	−2.219	–	–	0.244					
(e)	2-SAS	1	−0.231	1.003	–	0.343	1.237	0.395	−2442.953	4903.906	1.091	0.579	9
		2	1.600	0.354	–	−0.058	0.872	0.605					
(f)	3-SAS	1	−0.120	0.575	–	−0.734	1.430	0.208	−2441.530	4911.060	8.245	0.016	14
		2	−1.192	2.802	–	3.365	4.416	0.409					
		3	1.592	0.414	–	−0.036	1.223	0.383					
(g)	2-ASN	1	0.116	0.596	0.577	–	–	0.295	−2445.935	4905.869	3.054	0.217	7
		2	1.199	0.457	−0.857	–	–	0.705					
(h)	3-ASN	1	−0.414	0.771	−1.156	–	–	0.059	−2457.621	4937.243	34.428	$<10^{-7}$	11
		2	1.701	0.434	0.403	–	–	0.646					
		3	−0.162	0.548	−0.031	–	–	0.295					

Table 3. Parameters of the fits to the *Swift* data. Label corresponds to labels from Fig. 5. The smallest AIC is marked in bold, and p is the number of parameters in a model.

Label	Dist.	i	μ_i	σ_i	α_i	δ_i	β_i	A_i	$\mathcal{L}_{\max}$	AIC	ΔAIC	Pr	p
(a)	2-G	1	−0.026	0.740	–	–	–	0.139	−999.848	2009.695	14.315	0.001	5
		2	1.638	0.528	–	–	–	0.861					
(b)	3-G	1	−0.435	0.519	–	–	–	0.091	−989.654	**1995.308**	0	1	8
		2	0.875	0.332	–	–	–	0.194					
		3	1.785	0.437	–	–	–	0.715					
(c)	2-SN	1	−0.199	0.622	−4.514	–	–	0.059	−991.112	1996.348	1.040	0.632	7
		2	2.208	0.915	−2.327	–	–	0.941					
(d)	3-SN	1	−0.424	0.519	−0.026	–	–	0.091	−989.654	2001.308	6.000	0.050	11
		2	0.890	0.332	−0.054	–	–	0.194					
		3	1.776	0.437	0.026	–	–	0.715					
(e)	2-SAS	1	−0.271	0.435	–	−1.044	1.364	0.057	−989.692	1997.385	2.077	0.354	9
		2	1.790	0.539	–	−0.311	0.942	0.943					
(f)	3-SAS	1	−0.397	0.435	–	−0.386	1.072	0.068	−988.293	2004.586	9.278	0.010	14
		2	0.808	1.085	–	0.801	1.687	0.250					
		3	1.861	0.395	–	−0.334	0.823	0.682					
(g)	2-ASN	1	0.126	0.503	3.035×10^6	–	–	0.134	−994.295	2002.590	7.282	0.262	7
		2	1.244	0.535	−1.028	–	–	0.866					
(h)	3-ASN	1	−0.583	0.957	−1.091	–	–	0.024	−1001.719	2025.438	30.130	$<10^{-6}$	11
		2	1.516	0.523	−0.252	–	–	0.821					
		3	−0.017	0.887	−0.277	–	–	0.155					

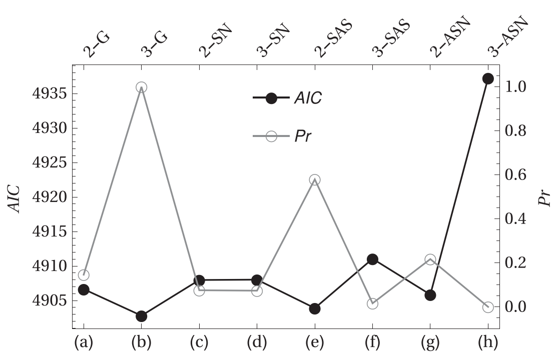

Figure 6. AIC and relative probability (Pr) for the BATSE models.

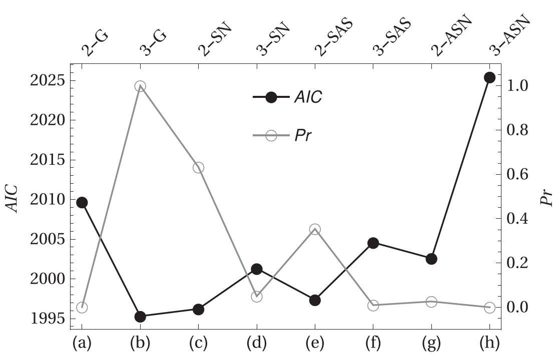

Figure 7. AIC and relative probability (Pr) for the *Swift* models.

corresponding to the collapsar scenario (associated to long GRBs) might not be necessary symmetric, its reason being a non-symmetric distribution of envelope masses of the progenitors. Therefore, mixtures of skewed distributions were tested herein, and it was found that a 2-SN (having the minimal AIC) and 2-SAS distributions are the best candidates to describe the observed $\log T_{90}$ distribution in the *Fermi* sample. These two models yield $\Delta_{2-\mathrm{SAS}} < 2$, which implies a substantial support for the 2-SAS model compared to a 2-SN model Burnham & Anderson (2004), corresponding to a probability of 37.7 per cent. Nevertheless, both of these two most plausible models are a mixture of only two skewed components. The model with the third smallest AIC is a 3-G with $\Delta_{3-\mathrm{G}} = 3.119$, which gives strong support for the 3-G model, although somewhat weaker than the preferred 2-SN and 2-SAS. The corresponding likelihood of the 3-G model is 21 per cent. The model with the fourth smallest AIC is a 2-G, with $\Delta_{2\text{-G}} = 4.459$, which means considerably less support, corresponding to a likelihood of 10.8 per cent. Other models yielded probabilities not higher than 3.6 per cent, hence are unlikely to describe the data well.

In the case of BATSE and *Swift*, the results are slightly different. The best model for describing their duration distribution is indeed a 3-G, however a strong support (ΔAIC ≈ 1) for a 2-SAS and a 2-SN distributions indicates that models with two skewed components cannot be ruled out, although despite being of complexity comparable to the 3-G distribution (i.e. having one parameter more and one less than the 3-G), they do not introduce a third component that might be thought to come from a third class. Hence, these two-component models are of simpler interpretation, especially when the possibility that the distribution of envelope masses is non-symmetric is considered. Moreover, for the BATSE data set, a 2-ASN and a 2-G are models with the third and fourth lowest AIC, with a relative probability of 21.7 per cent and 14.6 per cent, respectively. For *Swift*, a 2-SAS has a favourable ΔAIC ≈ 2, while a 2-ASN yielded a relative probability of 26.2 per cent, both being a considerable support. In all cases, the distributions fitted are bimodal, hence the existence of a third, intermediate in duration, GRB class is unlikely to be present in these catalogues, as well as in the *Fermi* sample.

It is important to note that in *Fermi* the sensitivity at very soft and very hard GRBs was higher than in BATSE (Meegan et al. 2009). Soft GRBs are intermediate in duration, and hard GRBs have short durations. Hence, an increase in intermediate GRBs relative to long ones might be expected as a consequence of improving instruments, yet the third class remains elusive (e.g. Tarnopolski 2015c). *Swift* is more sensitive in soft bands than BATSE was, hence its data set has a low fraction of short GRBs. Therefore, the group populations inferred from *Fermi* observations are reasonable considering the characteristics of the instruments.

5 CONCLUSIONS

Mixtures of various statistical distributions were fitted to the observed GRB durations of *Fermi*, BATSE, and *Swift*. It was found, based on the AIC, that for *Fermi* the most probable among the tested models is a two-component skew-normal distribution (2-SN). The second most plausible, with a relative probability of 37.7 per cent, is a two-component sinh-arcsinh distribution (2-SAS). A 3-G has a relative probability of 21 per cent of being correct. It is concluded that an elusive intermediate GRB class is unlikely to be present in the *Fermi* duration distribution, which is better described by a two-component mixture of skewed rather than symmetric distributions, hence the third class appears to be a statistical effect, and not a physical phenomenon.

For BATSE and *Swift*, a 3-G was found to describe the distributions best, however due to the small ΔAIC the preference of a 3-G over a 2-SAS and a 2-SN, respectively, is not strong enough to rule out the latter models. Also, a considerable support is shown by a 2-ASN and a 2-G in the case of BATSE, and a 2-SAS and a 2-ASN in the case of *Swift*. This corroborates the possibility of a non-existence of a third, intermediate GRB class, and gives evidence that the commonly applied mixture of standard normal distributions may not be a proper model, as some skewed distributions describe the data at least as well (BATSE and *Swift*), or considerably better (*Fermi*).

ACKNOWLEDGEMENTS

The author acknowledges support in a form of a special scholarship of Marian Smoluchowski Scientific Consortium Matter-Energy-Future from KNOW funding, grant number KNOW/48/SS/PC/2015, and wishes to thank the anonymous reviewer for useful comments that lead to significant improvement of the paper.

REFERENCES

Akaike H., 1974, IEEE Trans. Autom. Control, 19, 716
Azzalini A., 1985, Scand. J. Stat., 12, 171
Biesiada M., 2007, J. Cosmol. Astropart. Phys., 2, 3
Bromberg O., Nakar E., Piran T., 2011, ApJ, 739, L55
Bromberg O., Nakar E., Piran T., Sari R., 2013, ApJ, 764, 179
Burnham K. P., Anderson D. R., 2004, Sociol. Methods Res., 33, 261
Elal-Olivero D., 2010, Proyecciones J. Math., 29, 224
Gehrels N. et al., 2004, ApJ, 611, 1005
Gruber D. et al., 2014, ApJS, 211, 12
Horváth I., 1998, ApJ, 508, 757
Horváth I., 2002, A&A, 392, 791
Horváth I., 2009, Ap&SS, 323, 83
Horváth I., Balázs L. G., Bagoly Z., Veres P., 2008, A&A, 489, L1
Horváth I., Bagoly Z., Balázs L. G., de Ugarte Postigo A., Veres P., Mészáros A., 2010, ApJ, 713, 552
Horváth I., Balázs L. G., Hakkila J., Bagoly Z., Preece R. D., 2012, in Beloborodov A. et al., eds, Proc. Sci., Classification of Fermi and Swift GRBs. Munich, p. 046 (Available at http://pos.sissa.it/cgi-bin/reader/conf.cgi?confid=152)
Huja D., Řípa J., 2009, Balt. Astron., 18, 311
Huja D., Mészáros A., Řípa J., 2009, A&A, 504, 67
Jones M. C., Pewsey A., 2009, Biometrika, 96, 761
Kendall M., Stuart A., 1973, The Advanced Theory of Statistics. Griffin, London
Koen C., Bere A., 2012, MNRAS, 420, 405
Kouveliotou C., Meegan C. A., Fishman G. J., Bhat N. P., Briggs M. S., Koshut T. M., Paciesas W. S., Pendleton G. N., 1993, ApJ, 413, L101
Liddle A. R., 2007, MNRAS, 377, L74
Lü H.-J., Liang E.-W., Zhang B.-B., Zhang B., 2010, ApJ, 725, 1965
Mazets E. P. et al., 1981, Ap&SS, 80, 3
Meegan C. A., Fishman G. J., Wilson R. B., Horack J. M., Brock M. N., Paciesas W. S., Pendleton G. N., Kouveliotou C., 1992, Nature, 355, 143
Meegan C. A. et al., 1998, in Meegan C. A., Preece R. D., Koshut T. M., eds, AIP Conf. Proc. Vol. 428, The 4B BATSE Gamma-Ray Burst Catalog. Am. Inst. Phys., New York, p. 3
Meegan C. A. et al., 2009, ApJ, 702, 791
Mukherjee S., Feigelson E. D., Babu G. J., Murtagh F., Fraley C., Raftery A., 1998, ApJ, 508, 314
Nakar E., 2007, Phys. Rep., 442, 166
O'Hagan A., Leonard T., 1976, Biometrika, 63, 201
Paciesas W. S. et al., 1999, ApJS, 122, 465
Qin Y. et al., 2013, ApJ, 763, 15
Řípa J., Mészáros A., Wigger C., Huja D., Hudec R., Hajdas W., 2009, A&A, 498, 399
Řípa J., Mészáros A., Veres P., Park I. H., 2012, ApJ, 756, 44
Savchenko V., Neronov A., Courvoisier T. J.-L., 2012, A&A, 541, A122
Schilling M. F., Watkins A. E., Watkins W., 2002, Am. Stat., 56, 223
Shahmoradi A., 2013, ApJ, 766, 111
Shahmoradi A., Nemiroff R. J., 2015, MNRAS, 451, 126
Tarnopolski M., 2015a, Ap&SS, 359, 20
Tarnopolski M., 2015b, MNRAS, 454, 1132
Tarnopolski M., 2015c, A&A, 581, A29
Veres P., Bagoly Z., Horváth I., Mészáros A., Balázs L. G., 2010, ApJ, 725, 1955
von Kienlin A. et al., 2014, ApJS, 211, 13
Zhang Z.-B., Choi C.-S., 2008, A&A, 484, 293
Zhang F.-W., Shao L., Yan J.-Z., Wei D.-M., 2012, ApJ, 750, 88
Zitouni H., Guessoum N., Azzam W. J., Mochkovitch R., 2015, Ap&SS, 357, 7